

Universidad del Bío-Bío

Profesor Guía: Sr. Arnaldo Jélvez C.

Facultad de Ingeniería

Departamento de Ingeniería Industrial



**“APLICACIÓN DE UN MODELO PREDICTIVO DE FUGA DE
CLIENTES UTILIZANDO DATA MINING EN VTR
GLOBALCOM S.A. ZONA SUR”**

Trabajo de titulación presentado en conformidad a los requisitos para
obtener el título de Ingeniero Civil Industrial

CONCEPCIÓN, DICIEMBRE 2011

MAURICIO E. MORENO ECHEVERRÍA

VÍCTOR F. J. OVALLE RETAMAL

DEDICATORIAS

De: Mauricio E. Moreno Echeverría

Dedico esta investigación a mi familia, en especial a mi Madre por su amor y gran esfuerzo que ha puesto en brindarme una buena educación desde pequeño, logrando además inculcar principios y valores que me han ayudado de apoco a lograr el cumplimiento de metas y desafíos.

A mi hermana Carolina, por su apoyo incondicional en todos estos años de vida. A mis sobrinos Alonso, Cristóbal y Daniela por brindarme tantas alegrías. A mi cuñado Luis por aconsejarme tantas veces en temas de la universidad, y a Pilar, por su amor constante e incondicional durante todos estos años importantes.

*"Así que no temas, porque yo estoy contigo; no te angusties, porque yo soy tu Dios. Te fortaleceré y te ayudare; te sostendré con mi diestra victoriosa".
(Isaías 41-10).*

Mauro.

De: Víctor F. J. Ovalle Retamal

Quiero dedicar este trabajo con todo mi corazón a mi papá y mamá, por todo el amor, esfuerzo, preocupación, valores, enseñanzas de vida y educación entregadas desde niño hasta lo que soy hoy en día, ya que sin la ayuda y confianza depositada hacia mí, jamás hubiese conseguido alcanzar este logro.

A mis hermanas Johana, Paola y sus familias por su apoyo, a mis abuelos Juan y Rosa, y a mi tío Juan Miguel por aguantarme dos años en su hogar y además por la preocupación entregada. A mi tía María Cristina, por siempre confiar en mí y ser un apoyo para con mi madre a pesar de la distancia.

A mis padrinos Alfonso & Irma de San Javier, por siempre estar al tanto de mí y por estar al lado de mis padres en buenos y malos momentos.

A la familia Bravo Muñoz, quienes me recibieron el año 2009 en su hogar por hacerme sentir como uno más de ellos, por la confianza entregada, preocupación y cariño muchas gracias.

Víctor Ovalle Retamal.

AGRADECIMIENTOS

Primero que todo, queremos agradecer a Dios por darnos las capacidades y herramientas que nos permitieron llegar al final de esta aventura.

A nuestras familias por su apoyo irrestricto, amigos, compañeros y personas que fuimos conociendo en el camino y que aportaron tanto a nuestra formación académica, como humana.

A nuestro profesor guía Sr. Arnaldo Jélvez Caamaño por su apoyo constante en la realización de la investigación. Por su comprensión en los momentos de dificultad, cooperación y sobretodo, por su calidez humana.

A Ricardo Tapia y Francisco Cantero, por darnos la oportunidad de desarrollar nuestro trabajo, a Rodrigo Velásquez, Nicolás Von Caprivi, Daniel Jorquera, y en general a todo el personal VTR Globalcom S.A. por el apoyo entregado a lo largo de estos meses, y por la comprensión que tuvieron para con nosotros.

Por último, agradecer al profesor Sr. Nelson Caro (Q.E.P.D.) por la buena formación que nos entregó en los cursos de Gestión, por inculcar siempre que la base para desarrollarse de buena forma en cualquier ámbito de la vida, es la de mantener buenas relaciones humanas.

RESUMEN

Esta investigación tiene como objetivo aplicar un modelo de predicción de fuga de clientes en la compañía VTR Globalcom S.A. Zona Sur. Específicamente se han elegido las plazas de Concepción y Temuco para la aplicación del modelo.

La investigación se basó en las metodologías y herramientas de Data Mining para la determinación del modelo de predicción de fuga de clientes. Para ello se utilizó la base de datos de la compañía, en donde se pudo extraer información demográfica de clientes fugados / actuales, y de características de servicios contratados por estos últimos. El periodo de tiempo elegido para el desarrollo de la investigación se comprende desde Enero de 2009 hasta Junio de 2011.

Posteriormente, en conjunto con analistas de la propia compañía, se eligieron las variables que a priori podrían ser más influyentes para la generación de los modelos. Luego mediante el proceso KDD (Knowledge Discovery in Databases) se logró limpiar y transformar las variables que luego fueron incluidas en las dos técnicas específicas utilizadas de Data Mining.

La primera técnica utilizada fue Análisis Cluster (o Análisis de Conglomerados), técnica estadística multivariante que permitió generar perfiles de clientes fugados, donde se establecen características tales como: edad, gse, antigüedad de servicios, niveles de deuda, y otras dieciséis variables relevantes para el estudio.

La segunda técnica utilizada fue Regresión Logística Multivariante, técnica que permitió estudiar la relación entre la variable dependiente y las variables independientes que se estimaron importantes para el estudio. Aquí se realizaron análisis univariantes, bivariantes y evaluación de posibles interacciones o modificaciones de tipo efecto y/o confusión. Luego de la construcción de los modelos para ambas plazas, se realizó la validación con meses de prueba (Abril 2011 – Junio 2011), en donde se pudo contrastar el modelo con respecto a los meses que se

utilizaron para la construcción de estos mismos (Enero 2009 – Marzo 2011). En general ambos modelos tienen un nivel de acierto global superior a un 70%.

Posterior a la generación de ambos modelos, se estableció un plan de acción en base a Marketing Relacional, debido a las debilidades presentes con respecto a la generación de estrategias en post de retención y captación de clientes. La estrategia en sí, propone capacitar al personal de ventas y atención al cliente para las plazas de Concepción y Temuco, logrando generar una herramienta de apoyo que permita fidelizar al corto plazo de mejor manera a los clientes de ambas plazas.

La estructura de esta investigación, se divide en los siguientes capítulos:

En el capítulo 1 se detalla lo que respecta a la presentación del tema. Información sobre origen del tema y justificación, objetivos, alcances de estudio y metodología propuesta.

En el capítulo 2 se presentan los antecedentes de la compañía. Se detalla información sobre organigrama, participación en el mercado, desconexión de servicios, entre otros datos.

En el capítulo 3 se muestra el marco teórico utilizado en la investigación. Se detalla información sobre el proceso KDD (Knowledge Discovery in Databases) y Data Mining.

La información sobre las técnicas de Data Mining específicas utilizadas en la investigación se encuentran en el capítulo 4, aquí se detalla sobre Análisis Cluster y sobre Regresión Logística Multivariante.

La aplicación de las técnicas de Data Mining, junto a la construcción de modelos, y posterior validación con los meses de prueba, se detalla en el capítulo 5.

Por último, en el capítulo 6 se detalla el plan de acción elaborado para la compañía en base a los resultados de las técnicas utilizadas.

ÍNDICE DE CONTENIDOS

INTRODUCCIÓN	2
CAPITULO 1: PRESENTACIÓN DEL TEMA.....	4
1.1 Origen del tema	4
1.2 Justificación del tema	4
1.3 Objetivos del estudio.....	5
1.4 Alcances o Ámbito de Estudio	6
1.5 Metodología.....	7
CAPITULO 2: ANTECEDENTES DE LA EMPRESA	9
2.1 Antecedentes de la empresa.....	9
• Sueño VTR.....	10
• Las seis líneas de acción estratégicas “6 GOs”	10
• Organigrama.....	11
2.2 Participación en el mercado.....	12
2.3 Desconexión de servicios	15
2.4 Reclamos más frecuentes	18
2.5 Motivos más frecuentes de desconexión	20
2.6 Servicios actuales con mayor demanda.....	22
2.7 Grupo socioeconómico de clientes actuales.....	23
CAPÍTULO 3: MARCO TEÓRICO	25
3.1 ¿Qué es Churn?	25
3.2 KDD y sus etapas.....	27
Fase 1: Fase de integración y recopilación de datos	29
Fase 2: Selección, limpieza y transformación.....	29
Fase 3: Fase de Minería de datos	29
Fase 4: Fase de Evaluación e interpretación	30
Fase 5: Fase de Difusión y uso.....	30

3.3	Data Mining	31
3.3.1.	Relación con otras disciplinas	32
3.3.2.	Tarea de Data Mining.....	34
3.3.4.	Técnicas de Data Mining.....	38
I.	Técnicas algebraicas y estadísticas	38
II.	Técnicas Bayesianas	39
III.	Técnicas basadas en conteos de frecuencia y tablas de contingencias	40
IV.	Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas ...	40
V.	Técnicas relacionales. Declarativas y estructurales.....	40
VI.	Técnicas basadas en redes neuronales artificiales.....	41
VII.	Técnicas basadas en núcleo y máquinas de soporte vectorial	42
VIII.	Técnicas estocásticas y difusas.....	42
IX.	Técnicas basadas en casos, en densidad o distancia	43
 CAPÍTULO 4: TÉCNICAS DE DATA MINING UTILIZADAS EN EL PROYECTO		44
4.1	¿Qué es el Análisis Cluster?.....	44
4.2	Aplicaciones del Análisis Cluster	45
4.3	Proceso del Análisis Cluster.....	48
	Primer Paso: Objetivos del Análisis Cluster	48
	Segundo Paso: Diseño de investigación mediante análisis cluster	51
I.	Detección de datos atípicos.....	51
II.	Medidas de similitud	52
III.	Tipificación de los datos	56
A.	Estandarización por variables.....	56
B.	Estandarización por la observación.....	57
	Tercer Paso: Supuestos del Análisis Cluster.....	57
I.	Representatividad de la muestra	57
II.	Impacto de la multicolinealidad.....	58
	Cuarto Paso: Obtención de conglomerados y valoración del ajuste conjunto	58
I.	Métodos Jerárquicos.....	60
A.	Vínculo único vecino más próximo (Single linkage o Nearest neighbor)	60
B.	Vínculo completo (Complete linkage o furthest Neighbour).....	61
C.	Vínculo medio (Average Linkage).....	61
D.	Método del centroide	61
E.	Vinculación Intra-Grupos	62
F.	Método de Ward	62
G.	Algoritmo de Howard-Harris	63
II.	Métodos no Jerárquicos	64
A.	Algoritmo de las H-medias.....	65
B.	Algoritmo de las K-medias.....	65
	Quinto Paso: Interpretación de los conglomerados	67
	Sexto Paso: Validación y Perfil de los Grupos	68

•	Validez interna	68
•	Validez externa	69
4.4	Regresión Logística Binaria.....	69
4.4.1	El modelo de Regresión Logística.....	72
4.4.2	Estimación de los parámetros.....	74
4.4.3	Bondad de Ajuste: Contraste de hipótesis.....	76
I.	Desviación (Deviance)	76
II.	Prueba Chi-Cuadrado	77
III.	Prueba de Hosmer-Lemeshow.....	78
4.4.4	Análisis del modelo	78
 CAPÍTULO 5: APLICACIÓN DE TECNICAS DE DATA MINING		80
5.1	Datos	80
5.2	Análisis Cluster	81
5.2.1	Introducción.....	81
5.2.2	Selección de datos y variables a utilizar.....	82
5.2.3	Transformación de las variables	84
5.2.4	Elección del número de conglomerados para cada plaza en estudio.....	90
•	Plaza Concepción.....	92
•	Plaza Temuco.....	96
•	Resumen de resultados	99
5.2.5	Aplicación del Método No Jerárquico (K-Medias).....	100
•	Plaza Concepción.....	101
•	Plaza Temuco.....	105
•	Análisis de resultados.....	109
5.3	Regresión Logística Multivariante	111
5.3.1	Introducción.....	111
5.3.2	Selección de Datos y variables a utilizar	112
5.3.3	Transformación de las variables	114
5.3.4	Análisis univariado de las variables.....	115
•	Plaza Concepción.....	116
•	Plaza Temuco.....	119
•	Resultados y Conclusiones.....	122
5.3.5	Exploración de las relaciones bivariantes.....	124
•	Plaza Concepción.....	127
•	Plaza Temuco.....	131
•	Resultados y Conclusiones.....	135
5.3.6	Evaluación de interacciones o modificaciones de efecto y/o confusión.....	137
•	Plaza Concepción.....	139
•	Plaza Temuco.....	143
•	Análisis de resultados.....	146

•	Interacción	146
•	Confusión.....	146
•	Conclusión general	147
5.3.7	Construcción del modelo de Regresión Logística Multivariante	148
•	Plaza Concepción.....	149
•	Plaza Temuco.....	151
•	Curva COR.....	153
5.3.8	Validación de modelos.....	153
•	Plaza Concepción.....	154
•	Plaza Temuco.....	155
 CAPITULO 6: MARKETING RELACIONAL APLICADO A UN PLAN DE ACCIÓN		157
6.1	Introducción.....	157
6.2	Evolución de Marketing	159
6.2.1	Enfoque transaccional	162
6.2.2	Del Marketing Transaccional al Marketing Relacional	163
6.3	Marketing Relacional.....	165
6.3.1	Definición.....	165
6.3.2	Características e importancia del Marketing Relacional.....	167
6.3.3	Modelo del Marketing Relacional	168
I.	Gestión de la base de clientes	170
II.	Gestión de la lealtad	171
6.3.4	Premisas en el desarrollo del Marketing Relacional.....	172
6.3.5	La estrategia del Marketing Relacional	174
I.	Estrategia basada en el valor del cliente	178
II.	Conocimiento de la base de clientes.....	178
A.	Calidad de la base de datos.....	179
B.	La segmentación por comportamiento de compra.....	179
6.3.6	Lealtad.....	181
I.	Personalidad Pública de la marca	183
II.	Satisfacción del Cliente.....	184
III.	Costos de cambios	187
IV.	Voz del cliente.....	190
6.3.7	Plan de Marketing Relacional Eficaz.....	191
6.3.8	Como obtener clientes reales y rentables	195
I.	Plan de comunicación en el cultivo de la relación.....	196
II.	Programa de recuperación e clientes antiguos.....	199
6.3.9	Indicadores clave del éxito dentro de un programa de Marketing Relacional.....	202
6.3.10	Plan de Acción.....	204
I.	Definición.....	204
II.	Elementos del Plan de Acción	205
A.	Objetivos	205

B.	Metas	206
C.	Indicadores	207
D.	Actividades	207
E.	Responsable	207
6.4	Plan de acción VTR Globalcom S.A. Zona Sur	208
I.	Justificación	208
II.	Diseño del Plan de capacitación Personal de Ventas y Atención al Cliente VTR Globalcom S.A. Zona Sur	209
6.5	Conclusiones	215
	CONCLUSIONES GENERALES	217
	BIBLIOGRAFÍA	220
	REFERENCIAS BIBLIOGRÁFICAS EN LÍNEA	221
	ANEXOS	224

ÍNDICE DE FIGURAS, GRÁFICOS Y TABLAS

ÍNDICE DE FIGURAS	Página
Figura N° 2.1: Organigrama Nacional y Gerencia Zona Sur	11
Figura N° 3.1: Fases del proceso de Descubrimiento de Conocimiento, KDD	28
Figura N° 3.2: Disciplinas que contribuyen a la minería de datos	32
Figura N° 3.3: Ejemplo de regresión lineal	39
Figura N° 3.4: Red neuronal para el problema de jugar un cierto deporte	42
Figura N° 4.1: Diagrama de decisión de los pasos 1-3	49
Figura N° 4.2: Pasos 4-6 del diagrama de decisión del análisis Cluster	59
Figura N° 4.3: Representación Gráfica del modelo lineal	70
Figura N° 4.4: Representación gráfica de la función logística	72
Figura N° 5.1: Etapas de Análisis Cluster	81
Figura N° 5.2: Diagrama de bases de datos a utilizar en Análisis Cluster	82
Figura N° 5.3: Herramienta Análisis Cluster de K-Medias	100
Figura N° 5.4: Etapas de Regresión Logística Multivariante	111
Figura N° 5.5: Diagrama sobre datos a utilizar en RLM	112
Figura N° 5.6: Interacción y Confusión sobre la variable dependiente y la variable independiente PERMA_MESES	137
Figura N° 6.1: Factores de la creciente importancia del Marketing Relacional	160
Figura N° 6.2: Modelo de la Estrategia de Marketing Relacional	169
Figura N° 6.3: Alternativas Estratégicas	177
Figura N° 6.4: Dimensiones del cliente	180

Figura N° 6.5: Pilares para aumentar la lealtad de sus clientes	182
Figura N° 6.6: Áreas básicas de Creación de Valor	191
Figura N° 6.7: Plan eficaz de Marketing Relacional	194
Figura N° 6.8: Círculo de la Riqueza en Marketing	195
Figura N° 6.9: Preguntas que se deben realizar al diseñar un plan de acción	204
Figura N° 6.10: Elementos de un plan de acción	205
ÍNDICE DE GRÁFICOS	Página
Gráfico N° 2.1: Mercado Nacional Servicio de TV Cable (2008-2010)	12
Gráfico N° 2.2: Mercado Nacional Servicio de Telefonía (2008-2010)	13
Gráfico N° 2.3: Mercado Nacional Servicio de Internet (2008-2010)	14
Gráfico N° 2.4: Desconexiones por servicios – Plaza Concepción	15
Gráfico N° 2.5: Desconexiones por servicios – Plaza Temuco	16
Gráfico N° 2.6: Desconexiones totales por localidad – Plaza Concepción	16
Gráfico N° 2.7: Desconexiones totales por localidad – Plaza Temuco	17
Gráfico N° 2.8: Reclamos comerciales más frecuentes – Plaza Concepción	18
Gráfico N° 2.9: Reclamos comerciales más frecuentes – Plaza Temuco	19
Gráfico N° 2.10: Reclamos técnicos más frecuentes – Plaza Concepción	19
Gráfico N° 2.11: Reclamos técnicos más frecuentes – Plaza Temuco	20
Gráfico N° 2.12: Motivos de desconexión más frecuentes – Plaza Concepción	21
Gráfico N° 2.13: Motivos de desconexión más frecuentes – Plaza Temuco	21
Gráfico N° 2.14: Servicios actuales con mayor demanda – Plaza Concepción	22

Gráfico N° 2.15: Servicios actuales con mayor demanda – Plaza Temuco	22
Gráfico N° 2.16: GSE para clientes actuales – Plaza Concepción	23
Gráfico N° 2.17: GSE para clientes actuales – Plaza Temuco	24
Gráfico N° 5.1: Método Vinculación fuera de grupos – Plaza Concepción	93
Gráfico N° 5.2: Método Vecino más lejano – Plaza Concepción	94
Gráfico N° 5.3: Método Agrupación de centroides – Plaza Concepción	95
Gráfico N° 5.4: Método Vinculación fuera de grupos – Plaza Temuco	96
Gráfico N° 5.5: Método Vecino más lejano – Plaza Temuco	97
Gráfico N° 5.6: Método Agrupación de centroides – Plaza Temuco	98
Gráfico N° 5.7: Porcentaje de casos en cada conglomerado – Plaza Concepción	101
Gráfico N° 5.8: Porcentaje de casos en cada conglomerado – Plaza Temuco	105
Gráfico N° 5.9: Distribución por GSE – Estado Activo y Fuga – Plaza Concepción	116
Gráfico N° 5.10: Distribución por Nivel de deuda – Estado Activo y Fuga – Plaza Concepción	116
Gráfico N° 5.11: Distribución por Permanencia – Estado Activo y Fuga – Plaza Concepción	116
Gráfico N° 5.12: Distribución por Edad – Estado Activo y Fuga – Plaza Concepción	117
Gráfico N° 5.13: Distribución por Sexo – Estado Activo y Fuga – Plaza Concepción	117
Gráfico N° 5.14: Distribución por Promociones – Estado Activo y Fuga – Plaza Concepción	117
Gráfico N° 5.15: Distribución por Reclamos asociados a servicios – Estado Activo y Fuga – Plaza Concepción	118
Gráfico N° 5.16: Distribución por Canal de pago utilizado – Estado Activo y Fuga – Plaza Concepción	118
Gráfico N° 5.17: Distribución por Canal de entrada de servicios – Estado Activo y Fuga – Plaza Concepción	118

Gráfico N° 5.18: Distribución por GSE – Estado Activo y Fuga – Plaza Temuco	119
Gráfico N° 5.19: Distribución por Nivel de deuda – Estado Activo y Fuga – Plaza Temuco	119
Gráfico N° 5.20: Distribución por Permanencia – Estado Activo y Fuga – Plaza Temuco	119
Gráfico N° 5.21: Distribución por Edad – Estado Activo y Fuga – Plaza Temuco	120
Gráfico N° 5.22: Distribución por Sexo – Estado Activo y Fuga – Plaza Temuco	120
Gráfico N° 5.23: Distribución por Promociones – Estado Activo y Fuga – Plaza Temuco	120
Gráfico N° 5.24: Distribución de clientes por Reclamos asociados a servicios – Estado Activo y Fuga – Plaza Temuco	121
Gráfico N° 5.25: Distribución por Canal de pago utilizado – Estado Activo y Fuga – Plaza Temuco	121
Gráfico N° 5.26: Distribución por Canal de entrada – Estado Activo y Fuga – Plaza Temuco	121
ÍNDICE DE TABLAS	Página
Tabla N° 5.1: Variables utilizadas en la investigación de Análisis Cluster	83
Tabla N° 5.2: Clasificación de variables dicotómicas y policótomicas en Análisis Cluster	84
Tabla N° 5.3: Variables y cambios de atributos utilizados en Análisis Cluster	85
Tabla N° 5.4: Diferencias Básicas de cada uno de los métodos de conglomeración	89
Tabla N° 5.5: Inconvenientes de cada uno de los métodos de conglomeración	90
Tabla N° 5.6: Método Vinculación fuera de grupos – Plaza Concepción	92
Tabla N° 5.7: Método Vecino más Lejano – Plaza Concepción	93
Tabla N° 5.8: Método Agrupación de centroides – Plaza Concepción	94
Tabla N° 5.9: Método Vinculación fuera de grupos – Plaza Temuco	96
Tabla N° 5.10: Método más alejado – Plaza Temuco	97

Tabla N° 5.11: Método Agrupación de centroides – Plaza Temuco	98
Tabla N° 5.12 : Centros iniciales del conglomerado cuatro – Plaza Concepción	102
Tabla N° 5.13 : Centros finales del conglomerado cuatro – Plaza Concepción	103
Tabla N° 5.14: Historial de iteraciones para cambios en los centros del conglomerado cuatro – Plaza Concepción	104
Tabla N° 5.15: Distancia entre el centro del conglomerado cuatro y los restantes formados – Plaza Concepción	104
Tabla N° 5.16: Centros iniciales del conglomerado dos – Plaza Temuco	106
Tabla N° 5.17: Centros finales del conglomerado dos – Plaza Temuco	107
Tabla N° 5.18: Historial de iteraciones para cambios en los centros del conglomerado dos – Plaza Temuco	108
Tabla N° 5.19: Distancia entre el centro del conglomerado dos y los restantes formados – Plaza Temuco	108
Tabla N° 5.20: Variables utilizadas en RLM	113
Tabla N° 5.21: Variables y cambios de atributos utilizados en RLM	114
Tabla N° 5.22: Prueba Chi-Cuadrado para la variable dependiente y variable GSE – Plaza Concepción	127
Tabla N° 5.23: Regresión Logística Simple para la variable dependiente y variable GSE – Plaza Concepción	127
Tabla N° 5.24: Prueba Chi-Cuadrado para la variable dependiente y variable DEUDA – Plaza Concepción	127
Tabla N° 5.25: Regresión Logística Simple para la variable dependiente y variable DEUDA – Plaza Concepción	128
Tabla N° 5.26: Prueba Chi-Cuadrado para la variable dependiente y variable PERMA_MESES – Plaza Concepción	128
Tabla N° 5.27: Regresión Logística Simple para la variable dependiente y variable PERMA_MESES – Plaza Concepción	128
Tabla N° 5.28: Prueba Chi-Cuadrado para la variable dependiente y variable EDAD_C – Plaza Concepción	129
Tabla N° 5.29: Regresión Logística Simple para la variable dependiente y variable EDAD_C – Plaza Concepción	129

Tabla N° 5.30: Prueba Chi-Cuadrado para la variable dependiente y variable RECLAM – Plaza Concepción	129
Tabla N° 5.31: Regresión Logística Simple para la variable dependiente y variable RECLAM – Plaza Concepción	130
Tabla N° 5.32: Prueba Chi-Cuadrado para la variable dependiente y variable CPAGO – Plaza Concepción	130
Tabla N° 5.33: Regresión Logística Simple para la variable dependiente y variable CPAGO – Plaza Concepción	130
Tabla N° 5.34: Prueba Chi-Cuadrado para la variable dependiente y variable C_ENTRADA – Plaza Concepción	131
Tabla N° 5.35: Regresión Logística Simple para la variable dependiente y variable C_ENTRADA – Plaza Concepción	131
Tabla N° 5.36: Prueba Chi-Cuadrado para la variable dependiente y variable GSE – Plaza Temuco	131
Tabla N° 5.37: Regresión Logística Simple para la variable dependiente y variable GSE – Plaza Temuco	132
Tabla N° 5.38: Prueba Chi-Cuadrado para la variable dependiente y variable DEUDA – Plaza Temuco	132
Tabla N° 5.39: Regresión Logística Simple para la variable dependiente y variable DEUDA – Plaza Temuco	132
Tabla N° 5.40: Prueba Chi-Cuadrado para la variable dependiente y variable PERMA_MESES – Plaza Temuco	133
Tabla N° 5.41: Regresión Logística Simple para la variable dependiente y variable PERMA_MESES – Plaza Temuco	133
Tabla N° 5.42: Prueba Chi-Cuadrado para la variable dependiente y variable EDAD_C – Plaza Temuco	133
Tabla N° 5.43: Regresión Logística Simple para la variable dependiente y variable EDAD_C – Plaza Temuco	134
Tabla N° 5.44: Prueba Chi-Cuadrado para la variable dependiente y variable RECLAM – Plaza Temuco	134
Tabla N° 5.45: Regresión Logística Simple para la variable dependiente y variable RECLAM – Plaza Temuco	134
Tabla N° 5.46: Prueba Chi-Cuadrado para la variable dependiente y variable CPAGO – Plaza Temuco	135
Tabla N° 5.47: Regresión Logística Simple para la variable dependiente y variable CPAGO – Plaza Temuco	135

Tabla N° 5.48: Regresión Logística Multivariante – Variables PERMA_MESES, GSE y multiplicativa – Plaza Concepción	139
Tabla N° 5.49: Regresión Logística Multivariante - Variables PERMA_MESES y GSE – Plaza Concepción	139
Tabla N° 5.50: Regresión Logística Multivariante – Variables PERMA_MESES, DEUDA y multiplicativa – Plaza Concepción	139
Tabla N° 5.51: Regresión Logística Multivariante – Variables PERMA_MESES y DEUDA – Plaza Concepción	140
Tabla N° 5.52: Regresión Logística Multivariante – Variables PERMA_MESES, EDAD_C y multiplicativa – Plaza Concepción	140
Tabla N° 5.53: Regresión Logística Multivariante – Variables PERMA_MESES y EDAD_C – Plaza Concepción	140
Tabla N° 5.54: Regresión Logística Multivariante – Variables PERMA_MESES, RECLAM y multiplicativa – Plaza Concepción	141
Tabla N° 5.55: Regresión Logística Multivariante – Variables PERMA_MESES y RECLAM – Plaza Concepción	141
Tabla N° 5.56: Regresión Logística Multivariante – Variables PERMA_MESES, CPAGO y multiplicativa – Plaza Concepción	141
Tabla N° 5.57: Regresión Logística Multivariante – Variables PERMA_MESES y CPAGO – Plaza Concepción	142
Tabla N° 5.58: Regresión Logística Multivariante – Variables PERMA_MESES, C_ENTRADA y multiplicativa – Plaza Concepción	142
Tabla N° 5.59: Regresión Logística Multivariante – Variables PERMA_MESES y C_ENTRADA – Plaza Concepción	142
Tabla N° 5.60: Regresión Logística Multivariante – Variables PERMA_MESES, GSE y multiplicativa – Plaza Temuco	143
Tabla N° 5.61: Regresión Logística Multivariante – Variables PERMA_MESES, GSE y multiplicativa – Plaza Temuco	143
Tabla N° 5.62: Regresión Logística Multivariante – Variables PERMA_MESES, DEUDA y multiplicativa – Plaza Temuco	143
Tabla N° 5.63: Regresión Logística Multivariante - Variables PERMA_MESES y DEUDA – Plaza Temuco	144
Tabla N° 5.64: Regresión Logística Multivariante - Variables PERMA_MESES, EDAD_C y multiplicativa – Plaza Temuco	144
Tabla N° 5.65: Regresión Logística Multivariante - Variables PERMA_MESES y EDAD_C – Plaza Temuco	144

Tabla N° 5.66: Regresión Logística Multivariante - Variables PERMA_MESES, RECLAM y multiplicativa – Plaza Temuco	145
Tabla N° 5.67: Regresión Logística Multivariante - Variables PERMA_MESES y RECLAM – Plaza Temuco	145
Tabla N° 5.68: Regresión Logística Multivariante – Variables PERMA_MESES, CPAGO y multiplicativa – Plaza Temuco	145
Tabla N° 5.69: Regresión Logística Multivariante – Variables PERMA_MESES y CPAGO – Plaza Temuco	146
Tabla N° 5.70: Resultado del análisis de Confusión – Plaza Concepción	146
Tabla N° 5.71: Resultado del análisis de Confusión – Plaza Temuco	147
Tabla N° 5.72: Modelo de RLM para plaza de Concepción	149
Tabla N° 5.73: Resumen de sensibilidad, especificidad y porcentaje global de acierto – Plaza Concepción	150
Tabla N° 5.74: Modelo de RLM para plaza de Temuco	151
Tabla N° 5.75: Resumen de sensibilidad, especificidad y porcentaje global de acierto – Plaza Temuco	152
Tabla N° 5.76: Validación del modelo de RLM con tres niveles de corte – Plaza Concepción	154
Tabla N° 5.77: Validación del modelo de RLM con tres niveles de corte – Plaza Temuco	155
Tabla N° 6.1: Marketing Relacional v/s Marketing Transaccional	164
Tabla N° 6.2: Criterios de evaluación según los clientes	184
Tabla N° 6.3: Indicadores fundamentales en las etapas de la relación con el cliente	203
Tabla N° 6.4: Plan de Acción VTR Globalcom S.A. Zona Sur (Plaza Concepción & Temuco)	211
Tabla N° 6.5: Contenidos programas de Capacitación	212
Tabla N° 6.6: Cronograma de Actividades	213
Tabla N° 6.7: Costo Capacitación	214

ÍNDICE DE ANEXOS	Página
ANEXO A: Centros iniciales y finales de los conglomerados uno, dos y tres – Plaza Concepción	225
ANEXO B: Cambios en los centros de los conglomerados uno, dos y tres – Plaza Concepción	225
ANEXO C: Matriz de distancias euclídeas entre los conglomerados uno, dos y tres – Plaza Concepción	226
ANEXO D: Centros iniciales y finales de los conglomerados uno y tres – Plaza Temuco	226
ANEXO E: Cambios en los centros de los conglomerados uno y tres – Plaza Temuco	227
ANEXO F: Matriz de distancias euclídeas entre los conglomerados uno y tres – Plaza Temuco	227
ANEXO G: Curva COR – Modelo de RLM – Plaza Concepción	227
ANEXO H: Niveles de corte, sensibilidad, especificidad y 1 - especificidad – Modelo de RLM – Plaza Concepción	228
ANEXO I: Curva COR – Modelo de RLM – Plaza Temuco	228
ANEXO J: Niveles de corte, sensibilidad, especificidad y 1 - especificidad – Modelo de RLM – Plaza Temuco	229

INTRODUCCIÓN

En la actualidad existe una fuerte competencia en el negocio de telecomunicaciones debido al cambio abrupto de tecnologías, ofertas y demandas establecidas por el mercado. En nuestro país son múltiples las empresas que ofrecen servicios y productos en esta área. Internet, Telefonía y TV Cable son los principales servicios oferentes donde cada una de ellos cuenta con características únicas como precio, calidad en el servicio, nivel de cobertura, entre otras.

Para cualquier empresa es fundamental la cartera de clientes ya que es la principal fuente de activos que permite generar utilidades. En una primera instancia las empresas necesitan definir esta cartera con apoyo de publicidad y promociones, y en una segunda instancia se genera la necesidad de retener por el mayor tiempo posible a los clientes.

Con ayuda de estrategias de marketing, inversiones en publicidad y con generaciones de ofertas, es posible en el corto plazo captar nuevos clientes, pero el poder retenerlos requiere esfuerzos adicionales.

La retención busca identificar a los potenciales clientes con mayor tendencia de fuga debido a múltiples factores, generadas muchas veces por mejores ofertas y productos ofrecidos por la competencia, deficiencia en los servicios y productos entregados, o simplemente por factores económicos.

Las fugas pueden ser clasificadas en dos tipos: fugas voluntarias, en donde el cliente se desafilia por iniciativa propia, sin intervención de la empresa y fugas involuntarias, cuando la empresa es responsable del término de los acuerdos generados entre ambas partes, frecuente cuando existe mal uso de los servicios ofrecidos o cuando se generan deudas impagas en el tiempo.

De los antecedentes expuestos anteriormente, la presente investigación se enfocará en poder aplicar métodos de predicción de fugas de clientes para una institución del sector telecomunicaciones considerando fugas de tipo voluntarias e

involuntarias. La base para el desarrollo de la investigación es Data Mining (Minería de Datos) que es la extracción no trivial de información que se encuentra implícita en los datos.

Por último, se utilizarán dos técnicas específicas para el desarrollo del trabajo, las cuales son Análisis Cluster y Regresión Logística Multivariante. La información sobre las técnicas empleadas, su aplicación y todo el detalle necesario, se verá en extenso en los capítulos propios de la investigación.

CAPITULO 1: PRESENTACIÓN DEL TEMA

1.1 Origen del tema

El tema a desarrollar “Aplicación de un modelo predictivo de fuga de clientes utilizando Data Mining en VTR Globalcom S.A. Zona Sur” surge de reuniones técnicas realizadas con el académico del Departamento de Ingeniería Industrial y profesor patrocinante, Sr. Arnaldo Jélvez. Él propone la idea de aplicar herramientas de Data Mining en alguna empresa de servicio de la región. Resultado de esto, y luego de establecer conversaciones con la empresa VTR, el tema es propuesto formalmente al Gerente Zonal VTR Zona Sur, Sr. Ricardo Tapia, y posteriormente al Sr. Francisco Cantero, Sub-Gerente de Marketing y Clientes VTR Zona Sur.

1.2 Justificación del tema

Conocer, diferenciar, caracterizar y predecir la conducta de los clientes es de suma complejidad e importancia para las empresas, dado que los beneficios marginales que estos últimos van aportando con el paso del tiempo, crecen de forma significativa, por ello se tiene claridad que es de suma importancia retener a un cliente, que por el contrario captar uno nuevo.

En la actualidad, grandes bases de datos están presentes en las empresas con una variada gama de información de sus clientes que aumenta con el tiempo y muchas veces es desaprovechada, ya sea por factores de costos, limitantes de tiempo, capacitación de personal u otros factores. De lo anterior, es necesario comprender y entender él o los factores que permitan predecir la conducta de los clientes de manera tal de saber qué vender, cuándo y cómo vender.

Para ello, uno de los caminos recomendables por la literatura es utilizar esas grandes bases de datos presentes en las compañías a través de la herramienta de Data Mining (Minería de datos), que puede extraer información con el objetivo de generar conocimiento, y que en particular para este trabajo, ayudar a predecir dado una serie de factores las tendencias de fuga de clientes. Data Mining o también conocida como Minería de datos, prepara, sondea, explora y recopila aquella

información útil para el desarrollo de la empresa contribuyendo a la toma de decisiones tácticas y estratégicas logrando con ello fidelizar al cliente por el mayor tiempo posible.

1.3 Objetivos del estudio

- **Objetivo General**

- ✓ Proponer un modelo predictivo de fuga de clientes dentro de la empresa VTR Globalcom S.A. Zona Sur, específicamente en las plazas de Concepción y Temuco utilizando técnicas de Data Mining.

- **Objetivos Específicos**

- ✓ Estudiar el comportamiento de clientes de la ciudad de Concepción y Temuco respecto a pago, consumo de servicios, n° de reclamos, entre otras características.
- ✓ Identificar las variables que definen el problema, su importancia y relación para el modelo predictivo.
- ✓ Determinar factores que permitan clasificar a los nuevos y antiguos clientes como potencialmente fugables por medio de conglomerados, y además, generar un modelo predictor que estime la probabilidad de fuga como función de otros factores.
- ✓ Diseñar estrategias y/o planes de acción que ayuden a fidelizar de mejor forma a los clientes, logrando reducir la tasa de fuga en las plazas ya antes mencionadas.

1.4 Alcances o Ámbito de Estudio

El presente estudio recopila información proveniente de la base de datos de información de clientes de la empresa VTR Globalcom S.A., correspondiente a las plazas de Concepción y Temuco, con el propósito de encontrar un modelo predictivo que estime la probabilidad de fuga de clientes.

Con respecto a la base de datos existente en VTR Globalcom S.A., se tomará para el estudio un período de tiempo comprendido entre Enero 2009 y Junio 2011, contando con información de productos contratados por clientes, datos demográficos, sistemas de pago, n° reclamos, tipos de reclamos, entre otros factores.

Debido a la complejidad de la base de datos existente desde el punto de vista del análisis, magnitud y heterogeneidad del tipo de cliente, se recolectó información de software especializado para Data Mining para instrucción y capacitación de los alumnos memoristas como son Excel 2010, Access 2010, SPSS v19.

Con respecto a la aplicación del modelo, éste se realizará a través de dos etapas. En la primera, se utilizará Análisis Cluster que permitirá clasificar en distintos conglomerados a clientes fugados, logrando tener patrones de fugas que permitan diseñar estrategias y/o planes de acción para clientes actuales, lo que permitirá evitar fugas futuras de estos últimos. En una segunda etapa, se verá la relación que existe entre la variable dicotómica dependiente (Activo / Fuga) y las múltiples variables independientes existentes en las bases de datos, permitiendo encontrar cuales son las que más influyen en fuga de clientes para las plazas de Concepción y Temuco.

Luego de la aplicación de los modelos, se plantearán estrategias y/o planes de acción que permitan mejorar el servicio post-venta, logrando aumentar el grado de satisfacción del cliente y por otra parte, aumentar la tasa de retención para la compañía.

Dentro de la investigación será común asociar fugas de servicios con fugas de clientes, dado que la fuga de él o de los servicios contratados está directamente relacionada con la fuga del cliente.

Fuga Servicios → Fugas Clientes

Por último, es importante señalar que se establece el compromiso de reserva de información para fines solo académicos, autorizando la publicación de esta investigación en algún congreso o simposio dentro o fuera del país.

1.5 Metodología

La investigación a desarrollar en la empresa VTR Globalcom S.A. comprenderá una serie de etapas claramente definidas. En una primera etapa será necesario acceder y conocer la base de datos con respecto a información de clientes y servicios contratados de las plazas de Concepción y Temuco, con el objetivo de dimensionar la cantidad de clientes y servicios contratados que se posee desde los primeros años en que comenzó a operar la compañía con una base de datos.

Luego del conocimiento de la base de datos, se deberá analizar la información detalladamente para luego filtrar y determinar él o los factores más significativos a estudiar (con ayuda de expertos de la misma empresa), los cuales serán incluidos en los modelos predictivos.

En una tercera etapa se aplicara la técnica de Análisis Cluster. Esta técnica permitirá clasificar en grupos y subgrupos a los clientes dada ciertas características en común, permitiendo categorizar el riesgo de fuga.

Posteriormente se dará paso a la aplicación del modelo de Regresión Logística Multivariante, logrando determinar qué variables independientes son las que influyen realmente en la fuga, logrando determinar además con qué porcentaje de precisión actúa el modelo.

Para finalizar, se diseñaran estrategias y/o planes de acción de marketing, que permitan evitar la fuga de clientes y que de paso, logren fidelizar a estos últimos por el mayor tiempo posible.

CAPITULO 2: ANTECEDENTES DE LA EMPRESA

2.1 Antecedentes de la empresa

VTR es la empresa líder en la industria de las telecomunicaciones de Chile, con una participación del mercado disgregado en 48% TV Cable, 18% Telefonía y un 38% en servicios de Internet, cuyo 80% pertenece a Liberty Global y un 20% a CorpGroup (Grupo Saieh, 20%).

Actualmente ofrece servicios integrados de comunicaciones y entretenimiento, con más de 2.5 millones de hogares cubiertos por su red HFC (Híbrida de fibra óptica y coaxial), que le permite ser líder en televisión digital y banda ancha de altas velocidades. Es también la segunda compañía de telefonía fija residencial, y dueña del Carrier de larga distancia 111 y de la empresa bazuca.com

La compañía está presente en más de 1 millón de hogares chilenos, con presencia entre Arica y Coyhaique. Es gestora de la primera cadena de noticias 24 horas en Chile, en sociedad con CNN, promotora de una red de canales locales y temáticos a nivel nacional e impulsora del canal ViveDeportes. El ideal de la compañía siempre ha sido satisfacer los intereses y preferencias de los chilenos, apoyando al mismo tiempo la identidad local y el desarrollo deportivo del país.

VTR ocupa el puesto número 6, en el ranking de las mejores empresas para trabajar en Chile (Great Place to Work) con cerca de 4.000 colaboradores y una de las 18 empresas más responsables socialmente (Ranking RSE Pro-Humana).

- **Sueño VTR**

- ✓ La clásica Misión y Visión de las compañías fue reemplazada por el Sueño VTR. “En VTR existimos para crear experiencias únicas de entretenimiento y comunicación para nuestros clientes y que así todos disfrutemos de lo que nos gusta de la vida.”

- **Las seis líneas de acción estratégicas “6 GOs”**

La estrategia se lleva a la acción a través de seis líneas transversales a toda la compañía.

- ✓ **GO FLAT:** Desarrollar y extender los planes telefónicos con tarifas planas que contribuyan a la simplicidad, claridad y transparencia en la relación con los clientes.
- ✓ **GO DIGITAL:** Avanzar en el acceso de los clientes actuales y potenciales a los servicios de televisión digital, aumentando progresivamente la penetración de esta tecnología.
- ✓ **GO WIRELESS:** Avanzar en el desarrollo de su presencia en el mercado de los servicios inalámbricos.
- ✓ **GO BROADBAND:** Consolidar el liderazgo de la compañía tanto en la oferta de velocidad como en la experiencia de uso del servicio Internet Banda Ancha.
- ✓ **DO GOOD:** Incorporar criterios sustentabilidad y responsabilidad social a todas las actividades de la compañía.
- ✓ **GET IT GREAT:** Lograr los máximos estándares de excelencia en el servicio a clientes mediante planes de inversión y mejora en los ámbitos de procesos, tecnología, infraestructura y capital humano.

- Organigrama

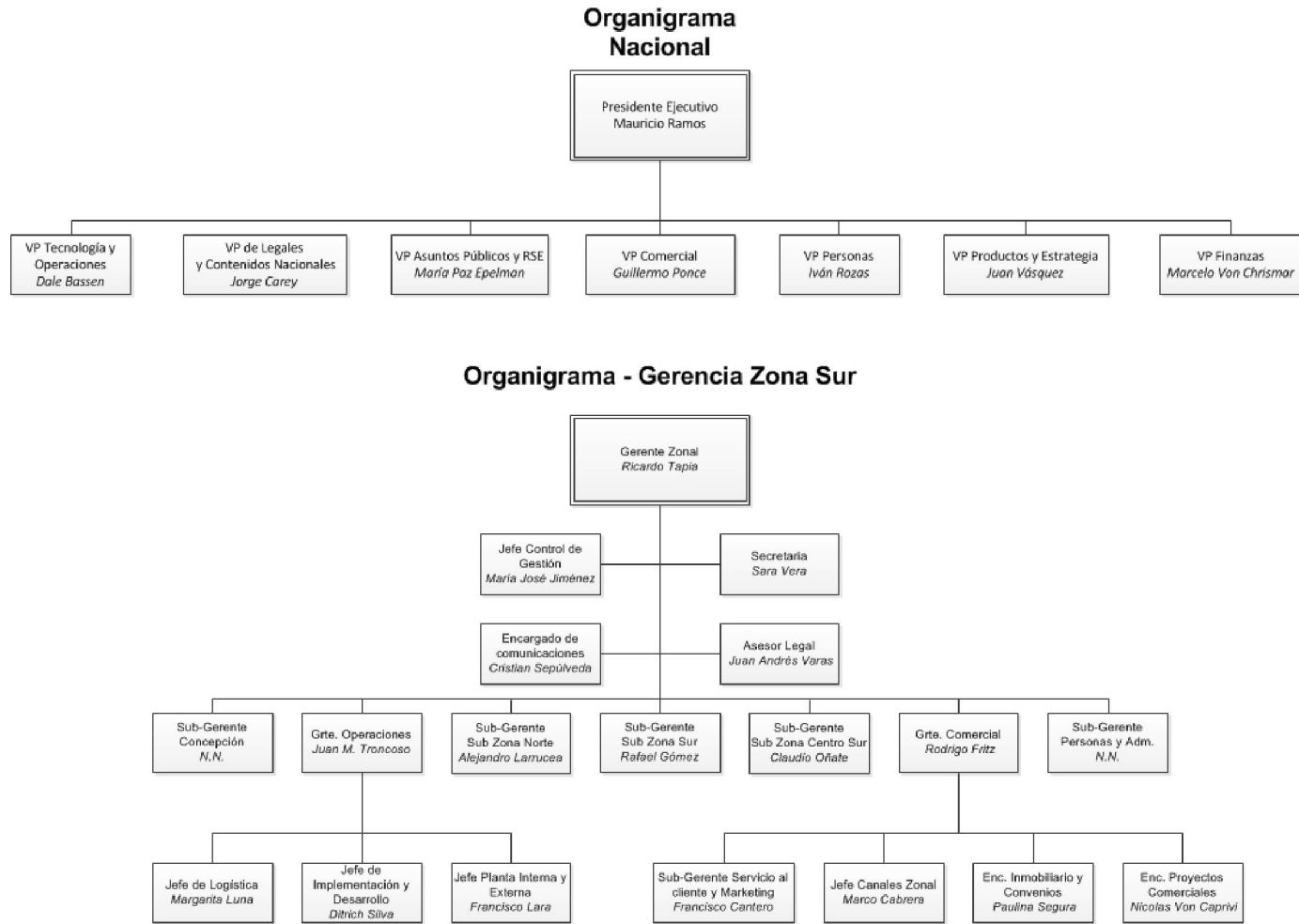


Figura N° 2.1: Organigrama Nacional y Gerencia Zona Sur

2.2 Participación en el mercado

A continuación se puede apreciar el nivel de participación que tiene la empresa VTR Globalcom S.A. en servicios de TV Cable, Internet y Telefonía entre el año 2008 y 2010. De igual modo, se puede ver el nivel de participación por parte de la competencia en los mismos servicios señalados anteriormente.

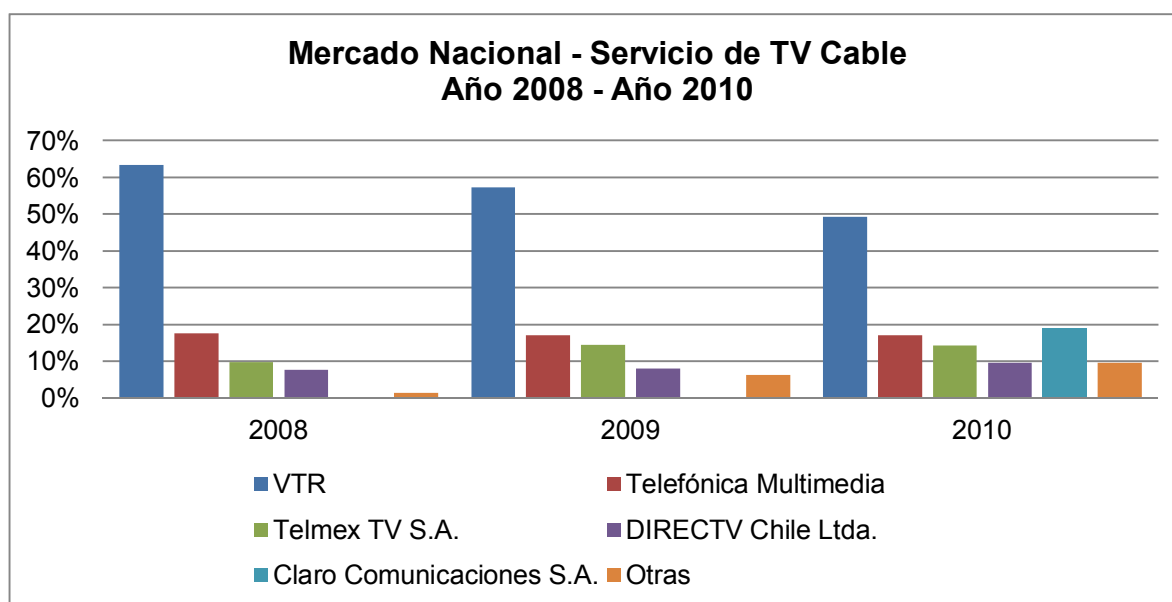


Gráfico N° 2.1: Mercado Nacional Servicio de TV Cable (2008-2010)

Fuente: SUBTEL

Del Gráfico N° 2.1 se observa que VTR desde el año 2008 al año 2010 ha sido el principal proveedor de servicios de TV Cable en nuestro país. En el año 2008 la participación en este mercado fue cercana a un 63%, lo que al paso de los años ha ido en una baja considerable llegando en el año 2010 a un 48%. Se observa que la competencia más cercana (Telefónica y Claro) alcanzan una participación de un 17% aproximadamente de este mercado, porcentaje que ha ido en aumento con el paso de los años.

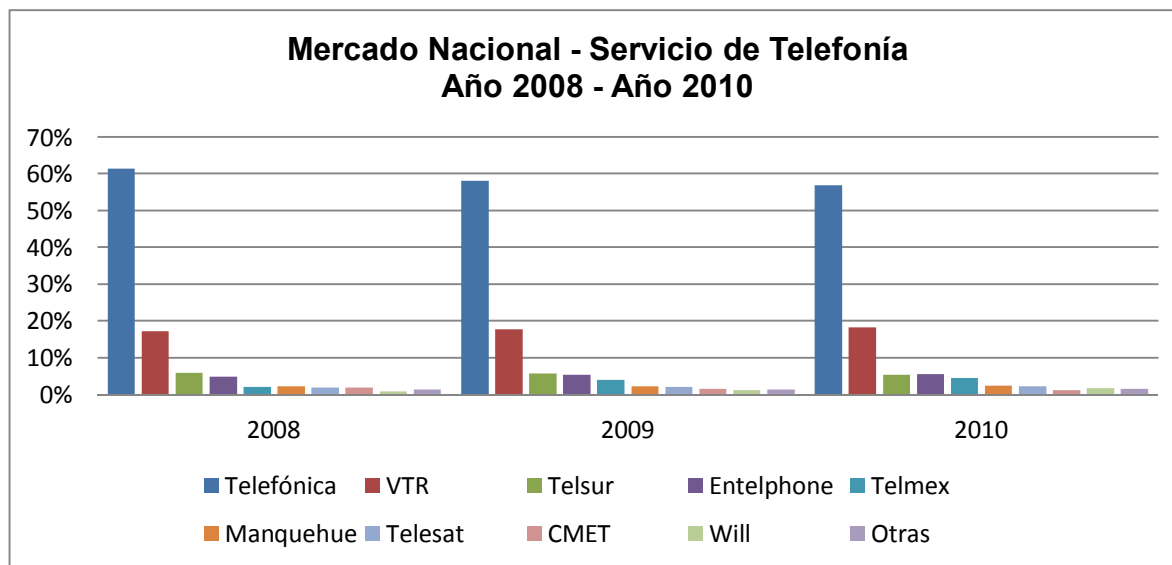


Gráfico N° 2.2: Mercado Nacional Servicio de Telefonía (2008-2010)

Fuente: SUBTEL

Del Gráfico N° 2.2 se observa que VTR tiene aproximadamente un 18% de participación en este mercado, siendo este porcentaje constante entre los años 2008 y 2010. En este mercado, la empresa Telefónica es líder absoluta en esta área, llegando a un nivel de participación de un 61% en el año 2008, y teniendo con el paso de los años una leve baja, llegando a tener un 57% de participación en el año 2010.

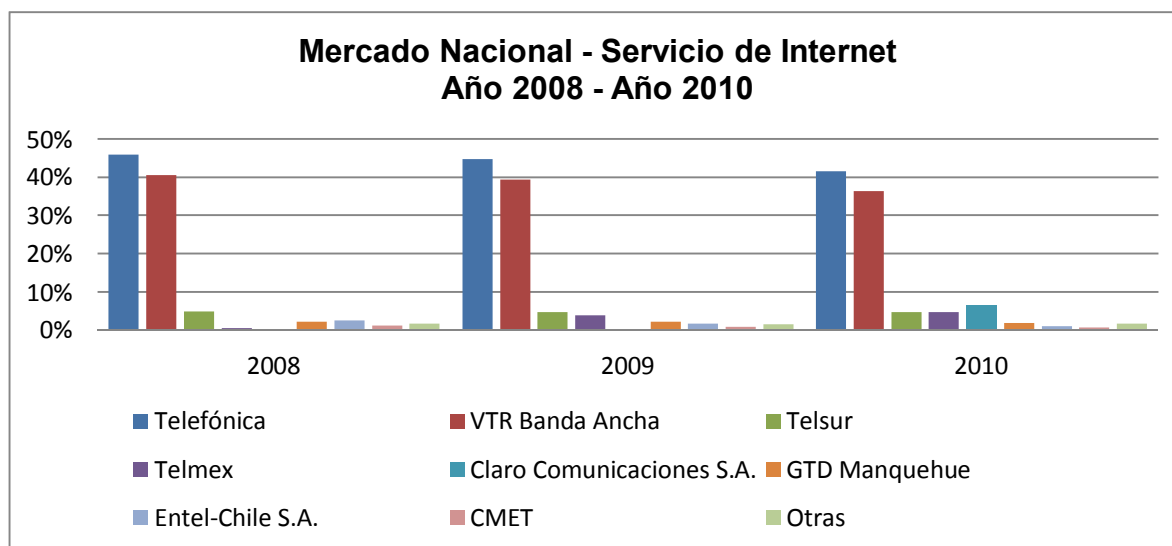


Gráfico N° 2.3: Mercado Nacional Servicio de Internet (2008-2010)

Fuente: SUBTEL

Del Gráfico N° 2.3 se detalla que VTR tiene una participación en este mercado de un 41% en el año 2008, sufriendo una baja en su participación en años posteriores llegando a un 36% en el año 2010. De igual modo que sucede en el mercado de Telefonía, la empresa Telefónica es líder en el área de Internet llegando a tener una participación de un 46% en el año 2008 y que al paso de los años ha ido en baja, llegando en el año 2010 a un 42%.

2.3 Desconexión de servicios

En los Gráficos N° 2.4 y N° 2.5 que a continuación se presentan, se puede observar el porcentaje de desconexiones de servicios de TV Cable, Internet y Telefonía para las plazas de Concepción y Temuco. La información presente en ambos gráficos corresponde a datos recopilados entre el año 2008 y el año 2010.

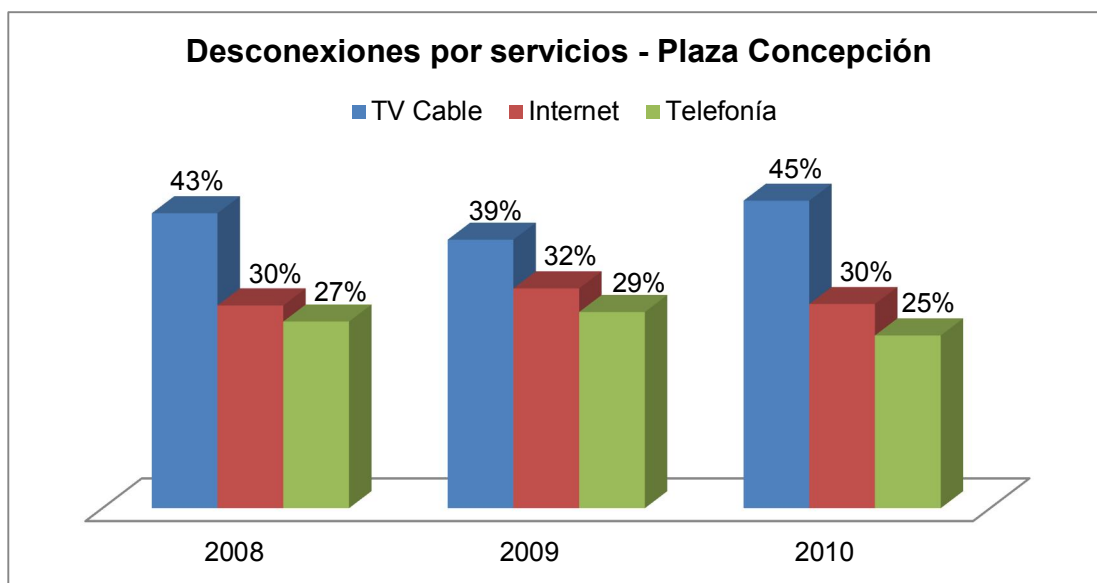


Gráfico N° 2.4: Desconexiones por servicios – Plaza Concepción

Fuente: VTR Globalcom S.A.

En el Gráfico N° 2.4 se observa que para los años 2008, 2009 y 2010, el servicio que sufre el mayor porcentaje de desconexiones para la plaza de Concepción es TV Cable, luego le sigue el servicio de Internet y por último, el servicio de Telefonía.

En el Gráfico N° 2.5 correspondiente al porcentaje de desconexiones de servicios para la plaza de Temuco, se observa el mismo comportamiento.

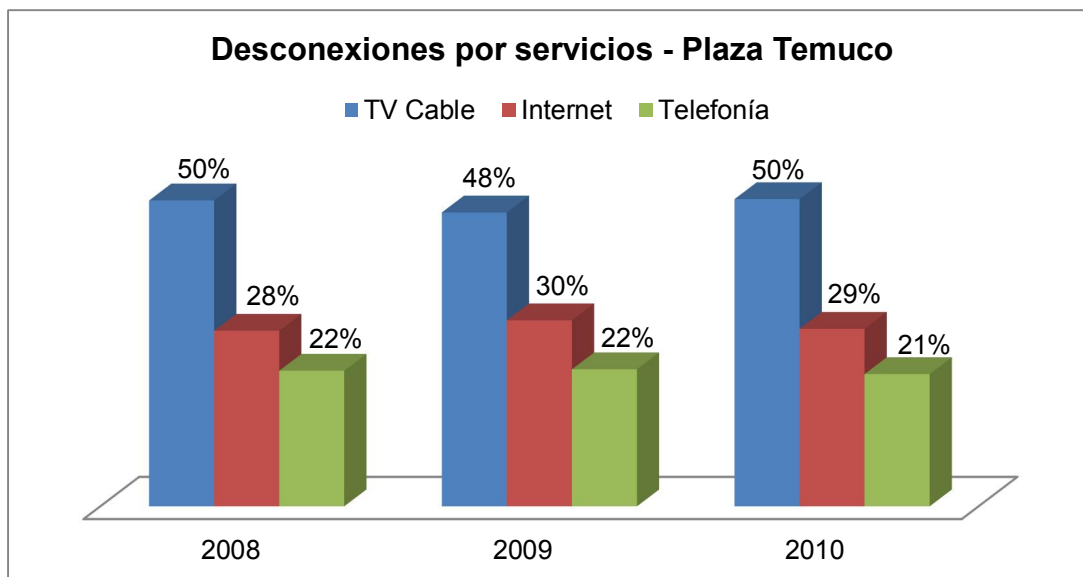


Gráfico N° 2.5: Desconexiones por servicios – Plaza Temuco

Fuente: VTR Globalcom S.A.

Otro análisis que se detalla a continuación es sobre el porcentaje de desconexiones totales (TV Cable + Internet + Telefonía) por localidad que se han generado entre los años 2008 y 2010.

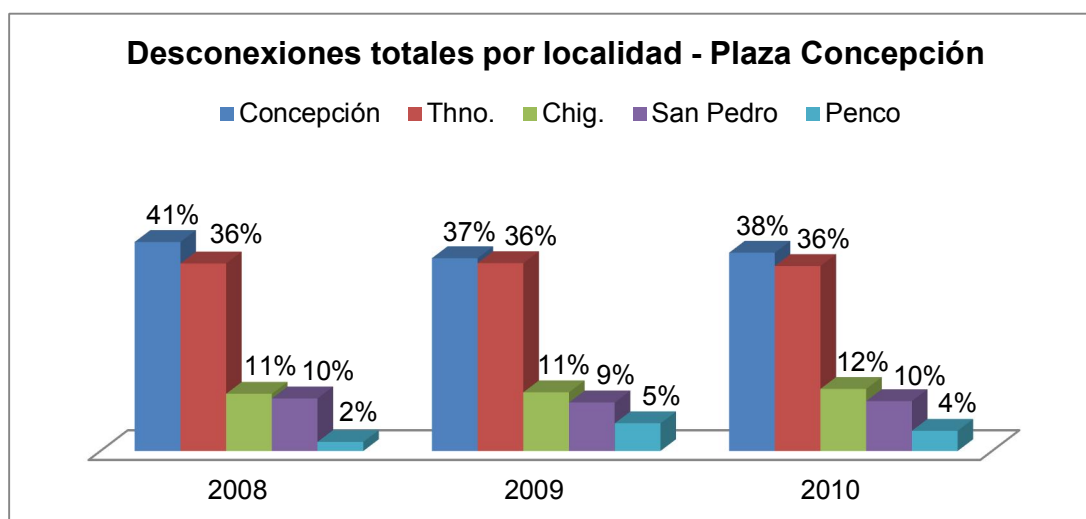


Gráfico N° 2.6: Desconexiones totales por localidad – Plaza Concepción

Fuente: VTR Globalcom S.A.

Del Gráfico N° 2.6 se detalla que para la plaza de Concepción, compuesta por las localidades de Concepción, Talcahuano, Chiguayante, Penco y San Pedro, el mayor porcentaje de desconexiones se produce en la localidad de Concepción, seguida muy de cerca por Talcahuano. De forma contraria, la localidad con el menor porcentaje de desconexiones es Penco.

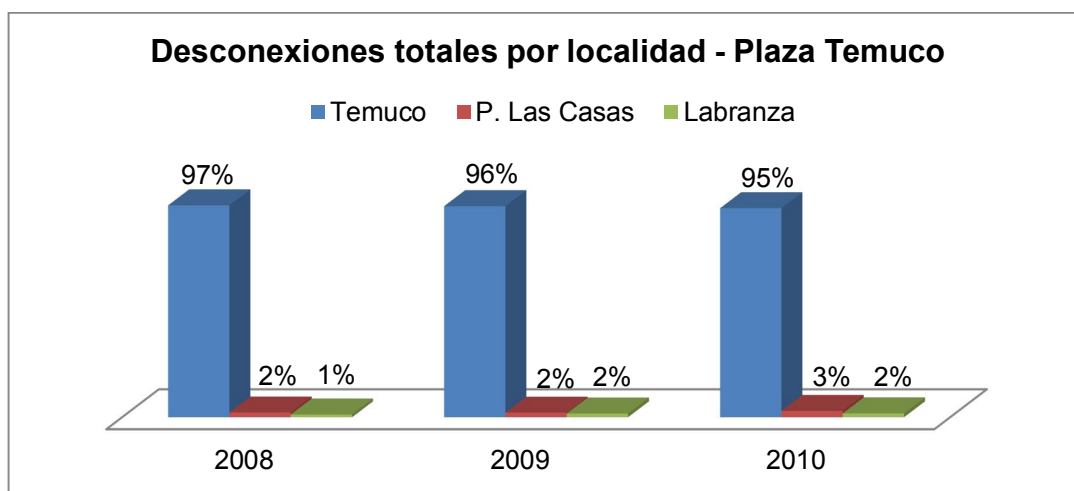


Gráfico N° 2.7: Desconexiones totales por localidad – Plaza Temuco

Fuente: VTR Globalcom S.A.

Del Gráfico N° 2.7, se observa que para la plaza de Temuco, compuesta por las localidades de Temuco, Labranza y Padre Las Casas, el mayor porcentaje de desconexiones se produce en la localidad de Temuco. De forma totalmente opuesta, el menor porcentaje de desconexiones se producen en las localidades de Padre Las Casas y Labranza. Esto último sucede por la pequeña magnitud de servicios contratados en estas dos últimas localidades, en comparativa con la localidad de Temuco.

¿Por qué se generan las desconexiones de servicios?. Básicamente sucede por la alta competencia que se genera en este mercado y por la rápida tendencia al cambio de preferencias y/o gustos por parte de los clientes. Según Cognus¹ el

1. Cognus, empresa consultora fundada en el año 2007, Chile.

abandono por parte de los clientes en el sector telecomunicaciones es de un 30%, considerado el más alto de todos los tipos de mercados.

Por ello es necesario estudiar el comportamiento de los servicios desconectados, para lograr posteriormente generar estrategias y/o planes de acción que permitan disminuir el número de desconexiones que actualmente sufre la compañía.

2.4 Reclamos más frecuentes

Los reclamos por no conformidad de servicios generados por los clientes pueden ser de dos tipos. El primero de ellos de carácter comercial, y el segundo de carácter técnico. Los Gráficos N° 2.8, N° 2.9, N° 2.10 y N° 2.11 que se detallan a continuación, contienen el desglose de los seis reclamos comerciales y técnicos más frecuentes para ambas plazas, cabe señalar que para la elaboración de los gráficos se utilizó información comprendida entre el año 2008 y el 2010.

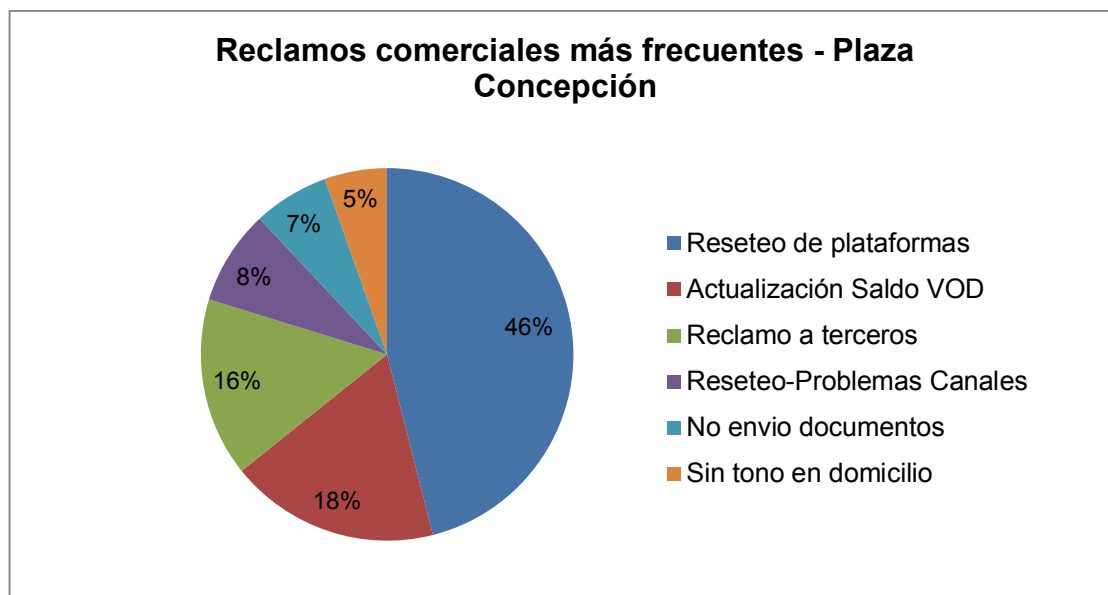


Gráfico N° 2.8: Reclamos comerciales más frecuentes – Plaza Concepción

Fuente: VTR Globalcom S.A.

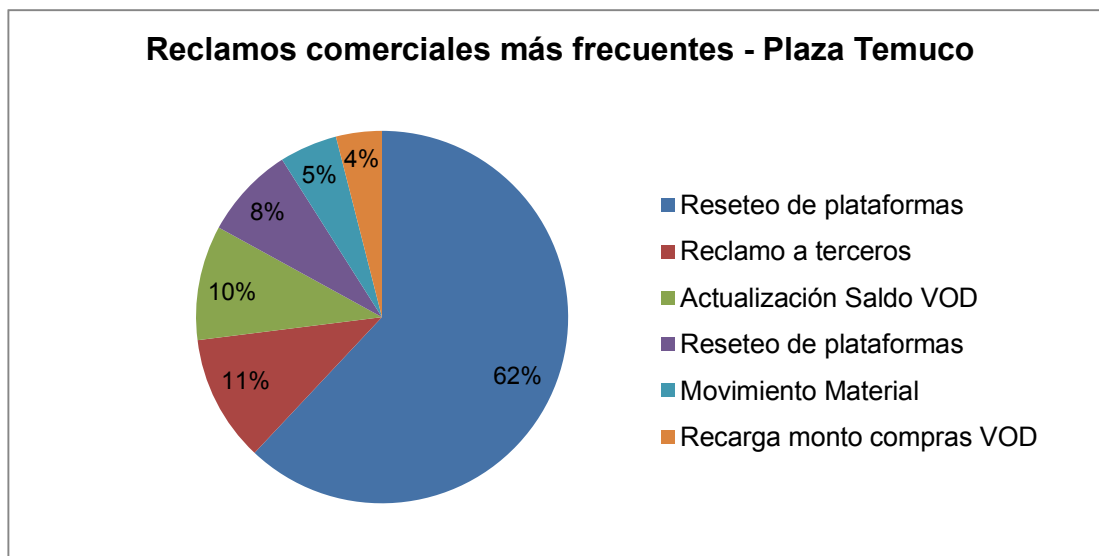


Gráfico N° 2.9: Reclamos comerciales más frecuentes – Plaza Temuco

Fuente: VTR Globalcom S.A.

De los Gráficos N° 2.8 y N° 2.9, se observa que el reclamo comercial más frecuente para ambas plazas es el reseteo de plataformas con un 46% para Concepción y un 62% para Temuco.

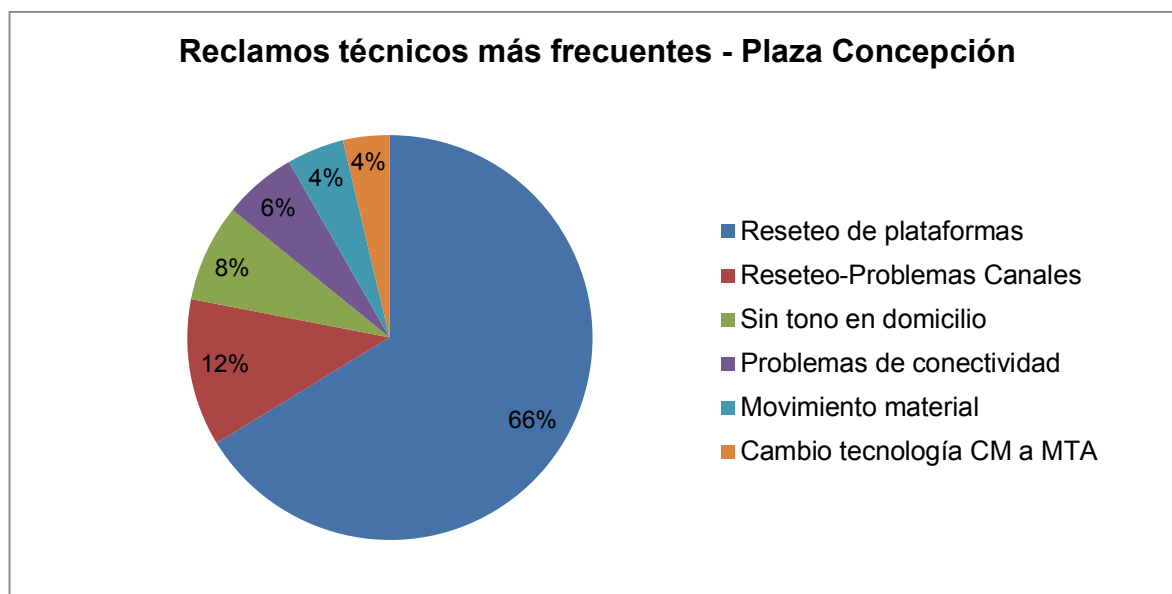


Gráfico N° 2.10: Reclamos técnicos más frecuentes – Plaza Concepción

Fuente: VTR Globalcom S.A.

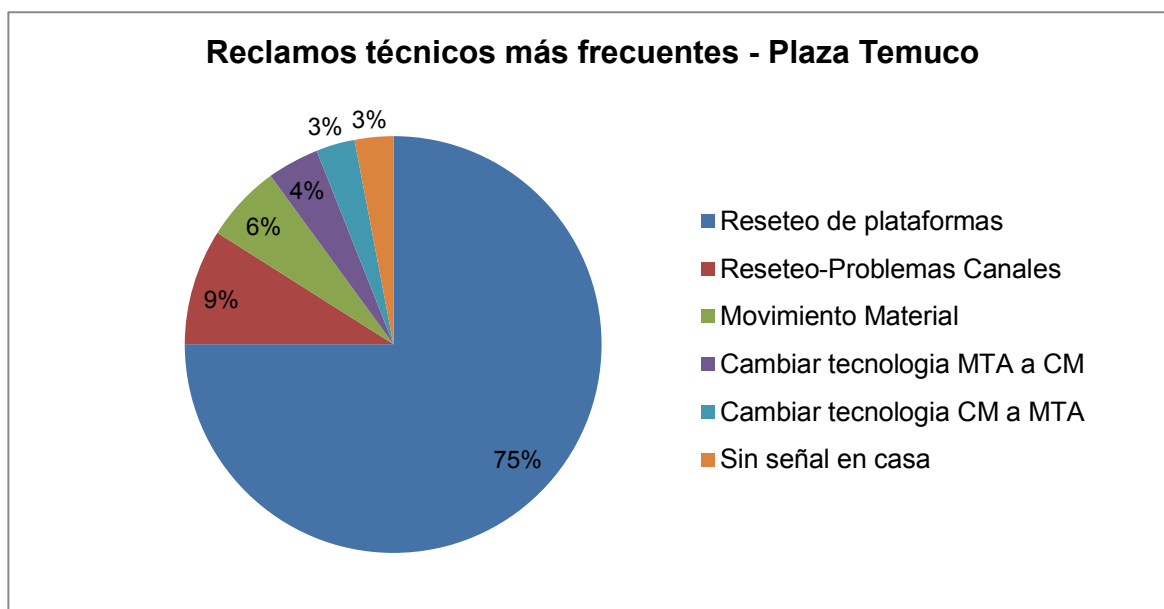


Gráfico N° 2.11: Reclamos técnicos más frecuentes – Plaza Temuco

Fuente: VTR Globalcom S.A.

De los Gráficos N° 2.10 y N° 2.11, se detalla que el reclamo técnico más frecuente para ambas plazas es el reseteo de plataformas con un 66% para Concepción y un 75% para Temuco. En general este reclamo es el que se genera con mayor frecuencia tanto en comerciales y técnicos.

2.5 Motivos más frecuentes de desconexión

A continuación, en los Gráficos N° 2.12 y N° 2.13 se pueden observar los siete motivos más frecuentes de desconexión de servicios para las plazas de Concepción y Temuco. La información empleada para la elaboración de estos gráficos está comprendida entre el año 2008 y 2010, excluyendo el periodo desde Marzo hasta Junio del año 2010 para la plaza de Concepción, debido que en este intervalo de tiempo el motivo con mayor frecuencia de desconexión de servicios fue producto del terremoto ocurrido el 27F.

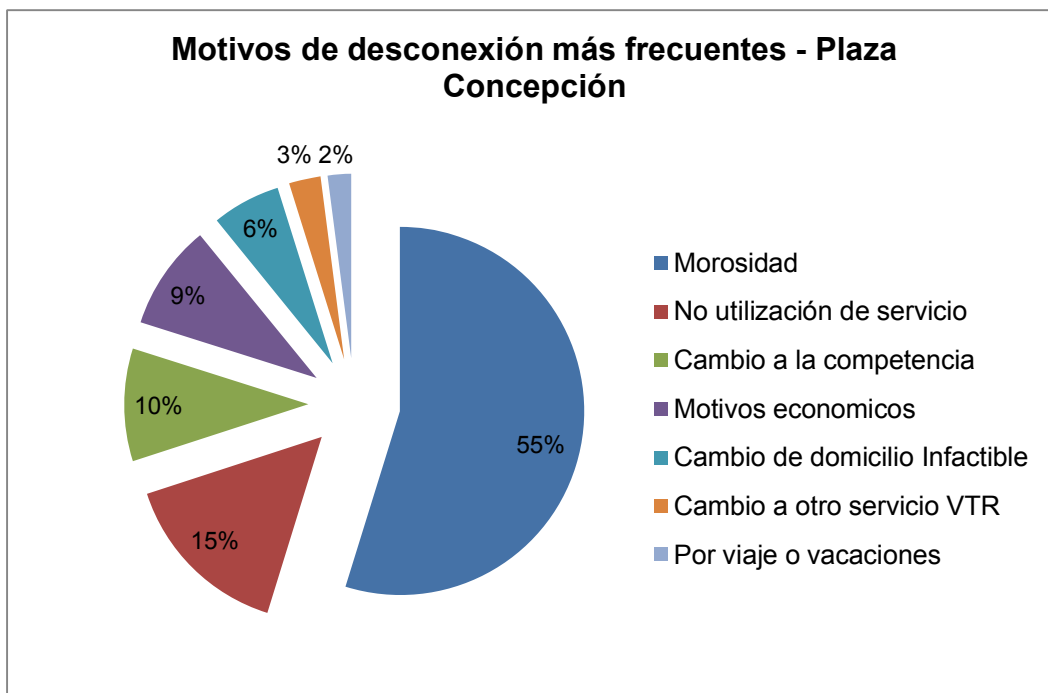


Gráfico N° 2.12: Motivos de desconexión más frecuentes - Plaza Concepción

Fuente: VTR Globalcom S.A.

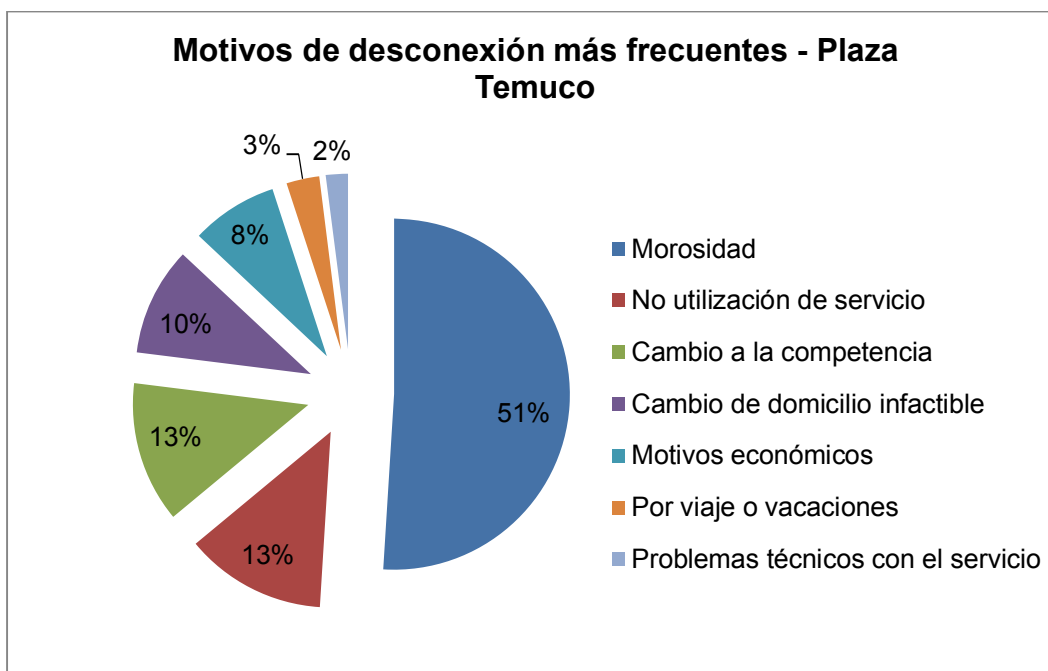


Gráfico N° 2.13: Motivos de desconexión más frecuentes – Plaza Temuco

Fuente: VTR Globalcom S.A.

De los Gráficos N° 2.12 y N° 2.13 se observa que el motivo más frecuente de desconexión de servicios para ambas plazas es Morosidad (el no pago de servicios por parte de los clientes), llegando a un 55% para la plaza de Concepción y un 51% para la plaza de Temuco. Luego le sigue No utilización de servicio con un 15% para la plaza de Concepción y un 13% para la plaza de Temuco. Por último, el motivo de desconexión menos frecuente para ambas plazas es Por viajes o vacaciones, llegando a un 2% para la plaza de Concepción y un 3% para la plaza de Temuco.

2.6 Servicios actuales con mayor demanda

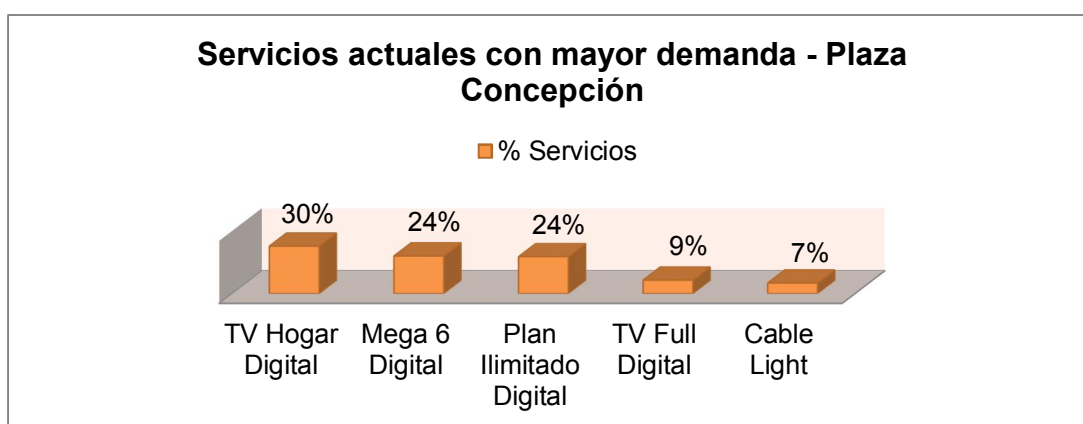


Gráfico N° 2.14: Servicios actuales con mayor demanda – Plaza Concepción

Fuente: VTR Globalcom S.A.

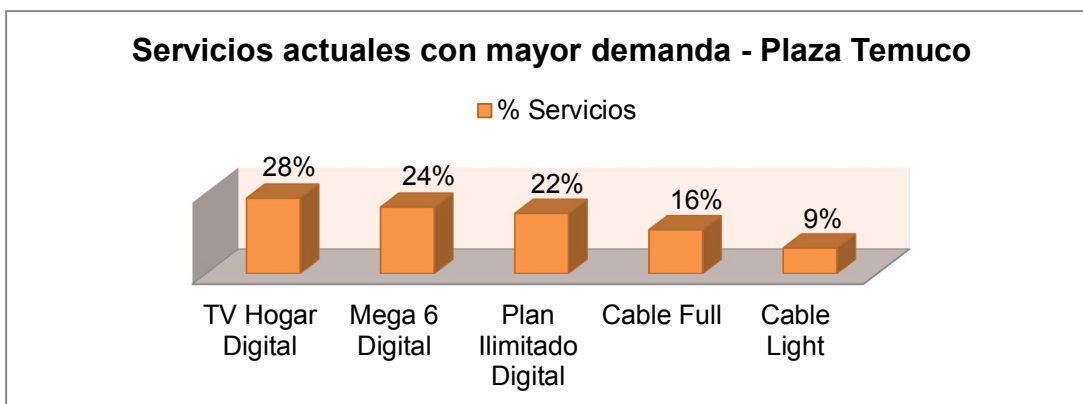


Gráfico N° 2.15: Servicios actuales con mayor demanda – Plaza Temuco

Fuente: VTR Globalcom S.A.

Los servicios mayormente demandados para ambas plazas como se puede observar en los Gráficos N° 2.14 y N° 2.15 son los servicios de TV Hogar Digital con un 30% para la plaza de Concepción y un 28% para la plaza de Temuco. Por otra parte, el servicio con menor demanda presente en los gráficos es Cable Light, con un 7% para la plaza de Concepción y un 9% para la plaza de Temuco. Se aprecia además que los tipos de servicios demandados en general son los mismos para ambas plazas.

2.7 Grupo socioeconómico de clientes actuales

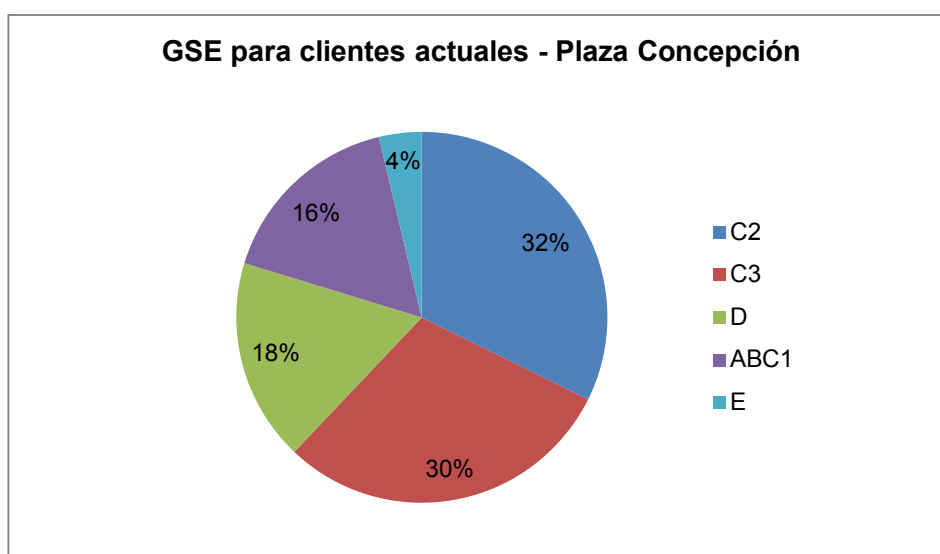


Gráfico N° 2.16: GSE para clientes actuales - Plaza Concepción

Fuente: VTR Globalcom S.A.

Del Gráfico N° 2.16 se puede observar que el 62% de la cartera de clientes para la plaza de Concepción la compone clientes con GSE C2 y C3, un 18% D, 16% ABC1 y 4% E. Esta información puede ser alta relevancia a la hora de dirigir un plan de acción para evitar fugas posteriores de clientes en la compañía.

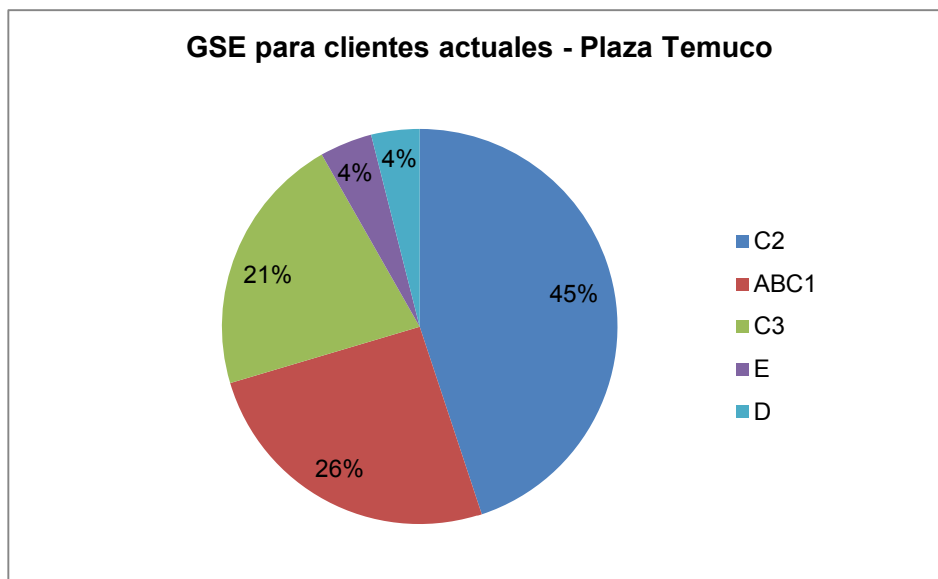


Gráfico N° 2.17: GSE para clientes actuales - Plaza Temuco

Fuente: VTR Globalcom S.A.

Del Gráfico N° 2.17 se observa que el 71% de la cartera de clientes para la plaza de Temuco la compone clientes con GSE C2 y ABC1, 21% C3 y 4% D y E. En esta plaza el grupo socioeconómico de la principal parte de la cartera de clientes es distinta a la plaza de Concepción, ya que incorpora al segmento con un mayor nivel adquisitivo, lo que implicaría crear estrategias focalizadas para cada una de las plazas.

CAPÍTULO 3: MARCO TEÓRICO

3.1 ¿Qué es Churn?

Hoy en día la retención de clientes para el negocio de las telecomunicaciones es vital, y es así que en los últimos años, el costo de adquisición de un nuevo cliente ha ido aumentando cada vez más debido a agresivas campañas de marketing realizadas por pequeñas y medianas empresas que crecen agresivamente, y por el aumento de costos asociados a la captación de nuevos clientes que incurren las grandes empresas. Según Matisson (1997) (En: Molina, 2009: Aplicación de técnicas de minería de datos para predicción del Churn de clientes en una empresa de Telecomunicaciones, Pág. 14), antes el foco de todas las empresas, o en su mayoría, era la “calidad del servicio”; en cambio hoy la gestión del Churn (fuga) se ha convertido en uno de los temas más críticos del negocio.

Customer Churn (fuga de clientes) se refiere a la proporción de clientes que se van o dejan al proveedor de un servicio durante un periodo de tiempo determinado. Y es así que el abandono de los clientes se clasifica de dos maneras según Molina (2009).

- **Voluntario:** cuando el cliente es quien toma la decisión de cambiar de proveedor de un servicio.
- **Involuntario:** cuando la empresa decide terminar el contrato por factores relativos al cliente entre las cuales pueden ser por incumplimiento de pago, incumplimiento de contrato, mal uso del producto o servicio, entre otros.

La fuga de clientes o servicios (Churn) es un término que se restringía solo al entorno de marketing, sin embargo hoy involucra a todos los sectores dentro de una compañía no importando el perfil o el tamaño y es por ello que se encuentra presente en las estrategias de las compañías. Es así como en el sector de las Telecomunicaciones, uno de los más afectados en esta materia, las tasas de Churn superan en algunos casos el 30%, la Banca y seguros oscila entre el 7% y el 18%, por último Retail, Distribución o Turismo y Ocio fluctúa entre el 5% y el 15%².

Para obtener una descripción de la situación de abandono de los clientes existen los Mapas de Abandono que entregan un estado de situación útil para un diagnóstico en profundidad. Esto hace referencia a la segmentación que es un paso crítico en la predicción del Churn para identificar los perfiles de los clientes propensos a la fuga y a quienes se debe retener, condicionado al valor del cliente, de tal manera de saber cuánto pierde la compañía si el cliente se va, y también saber cuánto se puede llegar a invertir para retener o evitar la fuga. Las estrategias de prevención del Churn nunca deben perder de vista ese factor, como elemento crítico, ya que depende de este si vale la pena que el cliente deba ser o no retenido.

La segmentación y caracterización de clientes fugados permiten establecer el roadmap (camino a seguir) del Churn, y sistemas de alertas para saber los momentos de si un cliente deje la compañía. Mediante el uso de técnicas analíticas o técnicas de Data Mining (minería de datos), se puede obtener un modelo que permita saber cuándo un cliente manifieste tendencia a la fuga. Información tal como datos demográficos, comportamiento de consumo, minutos de tráficos cursados, facturación en un periodo, o tasas de reclamos son factores para la construcción de un modelo de Data Mining (minería de datos).

Es necesario en cualquier estrategia de prevención del Churn, saber el por qué se marchan, cuales son los indicadores que desencadena el proceso de fuga. Solo conociendo porque se van los clientes será posible evitar su marcha actuando exactamente como el cliente demanda, con planes de retención adecuados.

Muchas compañías se equivocan al centrar su atención en el factor precio como medida para evitar el Churn, ya que éste no es el detonante para la marcha de clientes en la mayoría de los casos. Es hora de que las empresas comprendan que es en el nivel de servicio al cliente, y no en los precios, donde generalmente se esconde la explicación de la fuga. Un estudio realizado en el “Top 10” de las empresas europeas de asistencia en carretera revela que variaciones de hasta 40% en las tarifas de unas y otras no influyen en su cuota de mercado; las hay más caras con más clientes y más asequibles con menos participación de mercado G. Giuliani

(2004). (En: Molina, 2009: Aplicación de técnicas de minería de datos para predicción del Churn de clientes en una empresa de Telecomunicaciones, Pág. 15).

3.2 KDD y sus etapas

KDD (Knowledge Discovery in Databases) es un término muy utilizado y relacionado con la minería de datos, que consta de una serie de fases iterativas (secuenciales) que corresponden al proceso global cuyo objetivo es la extracción o descubrimiento de conocimiento en base a datos, el cuál la Minería de Datos es uno de los pasos claves Hernández (2004).

Fayyad (1996. En: Hernández, 2004, Pág. 13) define KKD, *“como el Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en esta última instancia, comprensibles a partir de los datos”*.

Hernández (2004), señala que dentro de las propiedades propiamente tal del conocimiento extraído se puede mencionar:

- **Válido:** hace referencia a que los patrones deben seguir siendo precisos para datos nuevos (con un cierto grado de certidumbre), y no sólo para aquellos que han sido usados en su obtención.
- **Novedoso:** que aporte algo desconocido tanto para el sistema y preferiblemente para el usuario.
- **Potencialmente útil:** la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.
- **Comprensible:** la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. De hecho una información incomprensible no proporciona conocimiento (al menos desde el punto de vista de la utilidad).

La secuencia de fases para la extracción de conocimiento se realiza en forma iterativa e interactiva. Iterativa porque la salida de alguna de las fases puede hacer volver a pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad. Interactiva porque el usuario, o más bien el experto en el dominio del problema, debe ayudar en la preparación de los datos y validación del conocimiento extraído Hernández (2004). Hernández (2004) explica cinco fases en el proceso KDD que quedan graficadas en la Figura N° 3.1 y son:

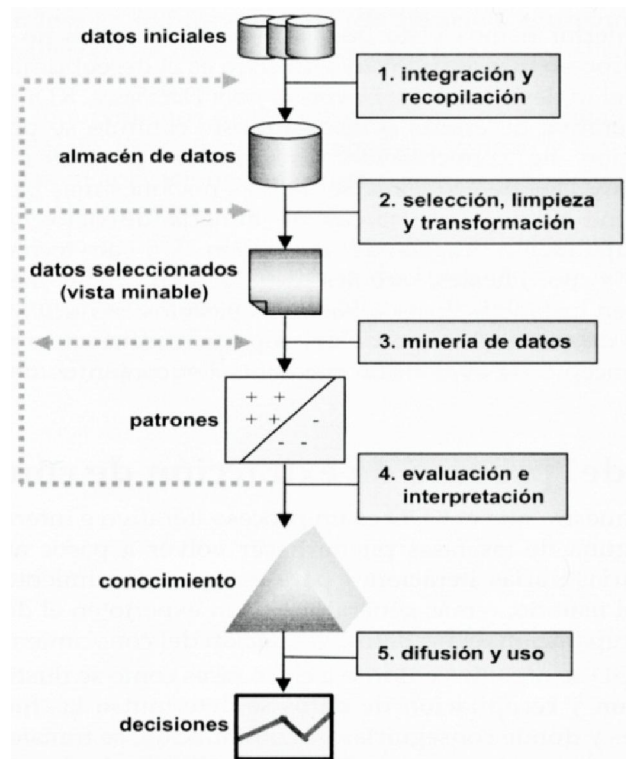


Figura N° 3.1: Fases del proceso de Descubrimiento de Conocimiento, KDD

Fuente: Hernández, 2007, Pág. 20

Fase 1: Fase de integración y recopilación de datos

Durante esta etapa, se determinan las fuentes de información que pueden ser útiles y donde conseguirlas, las cuales pueden residir en diferentes bases de datos, ya sea internas como externas. Es acá también, donde se analizan los datos, a fin de transformarlos a un formato común, frecuentemente mediante un almacén de datos que consiga unificar de manera operativa toda la información recogida, detentando y resolviendo las inconsistencias, facilitando la navegación y visualización de estos, para discernir qué aspectos resultan interesantes ser estudiados, con el propósito de alcanzar los objetivos definidos.

Fase 2: Selección, limpieza y transformación

A lo largo de esta etapa, se reúne los datos objetivos a los cuales se les aplicará la Minería de Datos. La cuidadosa selección de los atributos relevantes es crucial para la tarea de minería de datos, ya que se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Luego en la limpieza, se decide sobre aquellos datos que no se ajustan al comportamiento general del resto (Outliers), eliminar los registros incoherentes, duplicados e incoherentes y también, decidir el tratamiento sobre los datos faltantes o perdidos por último en la transformación, se construye nuevos atributos a partir de los ya existentes para mejorar su poder predictivo.

Fase 3: Fase de Minería de datos

Es la fase más característica de la secuencia KDD, donde su objetivo es descubrir situaciones anómalas y/o interesantes, tendencias, patrones y secuencias en los datos. El Input de la minería de datos, son los datos pre-procesados en las dos fases anteriores del procesos KDD y a partir de ellos, construir un modelo que permita producir nuevo conocimiento que sea útil para el usuario. Su modelo es una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos o para explicar situaciones

pasadas. Para ello, es necesario tomar una serie de decisiones antes de empezar el proceso:

- Determinar qué tipo de tarea de minería es el más apropiado.
- Elegir el tipo de modelo.
- Elegir el algoritmo de minería que resuelve la tarea y obtenga el tipo de modelo que estamos buscando.

Fase 4: Fase de Evaluación e interpretación

En esta etapa, se evalúan los patrones descubiertos en la fase anterior para el apoyo de toma de decisiones donde la calidad de estos debe ser medido. Para ello, se utilizan técnicas de evaluación que permiten determinar la precisión de los modelos (Validación simple, Validación cruzada, Validación cruzada con n-pliegues y bootstrapping). Por otro lado, hay que evaluar el contexto donde el modelo será utilizado, ya que no siempre la precisión es una buena medida, dado que podría no reflejar el contexto real de la aplicación. Ante esto, se hace necesario contrastar el conocimiento obtenido ante los expertos de la empresa, y si es necesario se vuelve a las fases anteriores para una nueva iteración.

Fase 5: Fase de Difusión y uso

Validado el modelo, se hace uso del nuevo conocimiento y se hace partícipe de él a todos los posibles usuarios, de manera tal que el nuevo conocimiento extraído se haga parte del Know-How de la organización.

3.3 Data Mining

Hoy en día la demanda de información estratégica no sólo atañe a las grandes corporaciones, sino también a las pequeñas, las que desean obtener ventaja competitiva o bien posicionarse en el mercado.

Para responder a esta demanda de conocimiento, surge el concepto de Data Mining del cuál la literatura entrega definiciones del área que las engloban, en mayor o mayor medida lo que es, como trabaja y en que consiste el Data Mining.

Fayyad (1996 En: Hernández, 2005, Pág. 6) lo define como *“el paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre-procesados”*.

Witten & Frank (2000, En: Hernández, 2004, Pág. 5), lo define como *“es el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos”*.

Según Molina (2009, Pág. 23) *“se refiere al proceso de extraer conocimiento de bases de datos. Su objetivo es descubrir situaciones anómalas y/o interesantes, tendencias, patrones y secuencias en los datos. Su input son los datos pre-procesados en las fases anteriores de la metodología, el objetivo es construir un modelo a partir de ellos, el cual pueda producir nuevo conocimiento que sea útil para el usuario”*.

Dentro de Data Mining (Minería de Datos) hay tipos de tareas con sus propios requisitos, las que pueden ser resueltas por un algoritmo de Data Mining específico. Las distintas tareas pueden ser descriptivas (agrupamiento o clustering, reglas de asociación secuenciales y las correlaciones) o predictivas (clasificación y la regresión).

3.3.1. Relación con otras disciplinas

El desarrollo que ha alcanzado Data Mining lo catapulta a un campo multidisciplinar de desarrollo paralelo o como prolongación de otras tecnologías. Es así, que los avances alcanzados se producen en las siguientes áreas relacionadas.

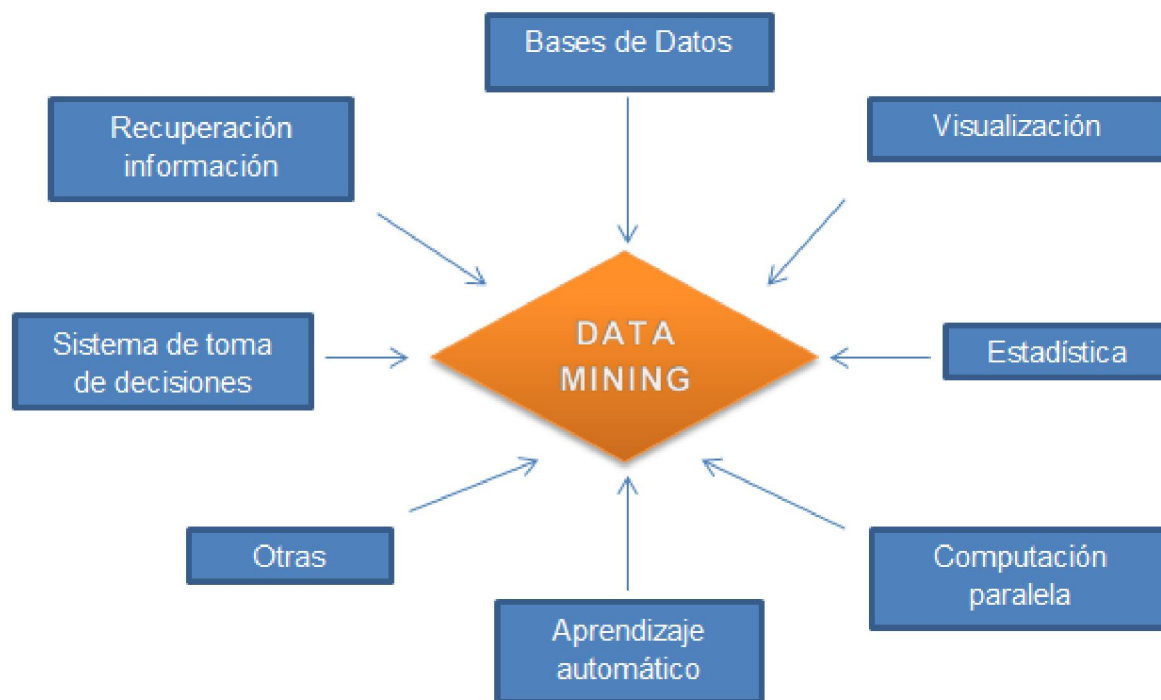


Figura N° 3.2: Disciplinas que contribuyen a la minería de datos

Fuente: Hernández, 2007, Pág. 14

- **Las bases de datos:** conceptos como los almacenes de datos y el procesamiento analítico en línea (OLAP) tienen una gran relación con Data Mining aunque en esta último caso, no se trata de obtener informes avanzados a base de agregar los datos de cierta manera compleja pero predefinida, sino de extraer conocimiento novedoso y comprensible.
- **La recuperación de información:** consiste en obtener información desde datos textuales, por lo que su desarrollo histórico se ha basado en el uso efectivo de bibliotecas (recientemente digitales) y en la búsqueda por

Internet. Una tarea típica es encontrar documentos a partir de palabras claves lo cual puede verse como un proceso de clasificación de los documentos en función de estas palabras clave. Para ello se usan medidas de similitud entre los documentos y la consulta.

- **Las estadística:** ha proporcionado muchos de los conceptos, algoritmos y técnicas que se utilizan en Data Mining, como por ejemplo, la media, la varianza y las distribuciones, el análisis univariante y multivariante, la regresión lineal y no lineal, la teoría del muestreo, la validación cruzada, la modelización paramétrica y no paramétrica, las técnicas bayesianas, etc.
- **El aprendizaje automático:** es el área de la inteligencia artificial que se ocupa de desarrollar algoritmos (y programas) capaces de aprender y constituye junto con la estadística, el corazón del análisis inteligente de los datos. Los principios seguidos en el aprendizaje automático y en la minería de datos son los mismos: la máquina aprende un modelo a partir de ejemplos y lo usa para resolver el problema.
- **Los sistemas para la toma de decisión:** son herramientas y sistemas informatizados que asisten a los directivos en la resolución de problemas y en la toma de decisiones. El objetivo es proporcionar la información necesaria para realizar decisiones efectivas en el ámbito empresarial o en tareas de diagnóstico (por ejemplo en medicina).
- **La visualización de datos:** el uso de técnicas de visualización permite al usuario descubrir, intuir o entender patrones que serían más difíciles de “ver” a partir de descripciones matemáticas o textuales de los resultados. Existen técnicas de visualización, como por ejemplo, las gráficas (diagrama de barra, gráficas de dispersión, histogramas, etc.), las icónicas (basadas en figuras, colores, etc.), las basadas en pixeles (cada dato se representa como un único pixel), las jerárquicas (dividiendo el área de representación en regiones dependiendo de los datos) y muchas otras.

- **La computación paralela y distribuida:** actualmente, muchos sistemas de bases de datos comerciales incluyen tecnologías de procesamiento paralelo, distribuido o de computación en grid. En estos sistemas el costo computacional de las tareas más complejas de minería de datos se reparte entre diferentes procesadores o computadores. Su éxito se debe en parte a la explosión de los almacenes de datos (su adaptación distribuida) y de la minería de datos, en los que las prestaciones de los algoritmos de consulta son críticas. Una de las principales ventajas del procesamiento paralelo es precisamente la escalabilidad de los algoritmos, lo que lo hace idóneo para estas aplicaciones.
- **Otras disciplinas:** dependiendo del tipo de datos a ser minados o del tipo de aplicación, la minería de datos usa también técnicas de otras disciplinas como el lenguaje natural, el análisis de imágenes, el procesamiento de señales, los gráficos por computadora, etc.

3.3.2. Tarea de Data Mining

Hernández (2004), señala que dentro de Data Mining, se distinguen diversos tipos de tareas, cada una de las cuales puede considerarse como un tipo de problema a ser resuelto por un algoritmo de Data Mining. Esto significa que cada tarea tiene sus propios requisitos, y que el tipo de información obtenida con una tarea puede diferir mucho de la obtenida con otra.

Las distintas tareas pueden ser predictivas o descriptivas, tal como se muestra a continuación:

- **Clasificación:** tarea predictiva, quizás la más utilizada donde en ella, cada registro pertenece a una clase determinada. El resto de los atributos se utiliza para predecir la clase. El objetivo es maximizar la función de precisión de la clasificación de los nuevos registros. La clasificación se ocupa de resultados discretos: si o no. Es útil en problemas como clasificación de los candidatos a

crédito; detectar reclamos fraudulentos por los seguros; identificación de procedimientos médicos; etc.

Ejemplo: consideremos un oftalmólogo que desea disponer de un sistema que le sirva para determinar la conveniencia o no de recomendar la cirugía ocular a sus pacientes. Para ello dispone de una base de datos de sus antiguos pacientes clasificados en operados satisfactoriamente o no en función del tipo de problema que padecían (miopía y su grado, o astigmatismo) y de su edad. El modelo encontrado se utiliza para clasificar nuevos pacientes, es decir, para decidir si es conveniente operarlos o no.

- **Regresión:** consiste en aprender una función que asigna a cada registro un valor real. Esta es la principal diferencia respecto a la clasificación; el valor a predecir es numérico. El objetivo de este caso es minimizar el error entre el valor predicho y el real. Puede servir para estimar el número de niños en una familia, estimar el ingreso total, etc.

Ejemplo: un empresario quiere conocer cuál es el costo de un nuevo contrato basándose en los datos correspondientes a contratos anteriores. Para ello usa una fórmula de regresión lineal, ajustando con los datos pasados la función lineal y usándola para predecir el costo en el futuro.

- **Agrupamiento:** es la tarea descriptiva por excelencia y consiste en obtener grupos “naturales” a partir de los datos. Se habla de grupos y no de clases, porque a diferencia de la clasificación, en lugar de analizar datos etiquetados con una clase, los analiza para generar esta etiqueta. Los datos son analizados basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizando la similitud entre los distintos grupos. Es decir, se forman grupos tales que los objetos de un mismo grupo son muy similares entre sí y, al mismo tiempo, son muy diferentes a los objetos de otro grupo. Al agrupamiento se le suele llamar segmentación, ya que parte o segmenta los datos en grupos que pueden ser o no disjuntos.

Ejemplo: Una librería que ofrece sus servicios a través de Internet usa el agrupamiento para identificar grupos de clientes en base a sus preferencias de compras que le permita dar un servicio más personalizado. Así, cada vez que un cliente se interesa por un libro, el sistema identifica a que grupo pertenece y le recomienda otros libros comprados por clientes de su mismo grupo.

- **Correlaciones:** tarea descriptiva que se usa para examinar el grado de similitud de los valores de dos variables numéricas. Una fórmula estándar para medir la correlación lineal es el coeficiente r , el cual es un valor real comprendido entre -1 y 1. Si r es las variables están perfectamente correlacionadas, caso contrario (r igual a -1) la correlación es negativa. Mientras si es 0 no hay correlación. Esto quiere decir que cuando r es positivo, las variables tienen un comportamiento similar (ambas crecen o decrecen al mismo tiempo) y cuando r es negativo si una variable crece la otra decrece. El análisis de correlaciones, sobre todo las negativas, puede ser muy útil para establecer reglas de ítems correlacionados, como se muestra en el siguiente ejemplo.

Ejemplo: Un inspector de incendios que desea obtener información útil para la prevención de incendios probablemente esté interesado en conocer correlaciones negativas entre el empleo de distintos grosores de protección del material eléctrico y la frecuencia de ocurrencia de incendios.

- **Reglas de asociación:** Tarea descriptiva muy similar a las correlaciones que tiene como objetivo identificar relaciones no explícitas entre atributos categóricos. Pueden ser de muchas formas, aunque la formulación más común es del estilo “si el atributo X toma el valor d entonces el atributo Y toma el valor “ b ”. Las reglas de asociación no implican una relación causa-efecto, es decir, puede no existir una causa para que los datos estén asociados. Este tipo de tarea se utiliza frecuentemente en el análisis de la cesta de la compra, para identificar productos que son frecuentemente

comprados juntos, información está que puede usarse para ajustar los inventarios, para la organización física del almacén o en campañas publicitarias. Las reglas se evalúan usando dos parámetros (precisión y soporte).

Ejemplo: Una compañía de asistencia sanitaria desea analizar las peticiones de servicios médicos solicitados por sus asegurados. Cada petición contiene información sobre las pruebas médicas que fueron realizadas al paciente durante una visita. Toda esta información se almacena en una base de datos en la que cada petición es un registro cuyos atributos expresan si se realiza o no cada una de las posibles pruebas médicas que pueden ser realizadas a un paciente. Mediante reglas de asociación, un sistema encontraría aquellas pruebas médicas que frecuentemente se analizan juntas por ejemplo, que un 70 por ciento de las veces que se pide un análisis de orina también se solicita uno de sangre, y esto ocurre en dos de cada diez pacientes. La precisión de esta regla es del 70 por ciento y el soporte del 20 por ciento.

- **Asociación Secuencial:** se usa para determinar patrones secuenciales en los datos. Estos patrones se basan en secuencias temporales de acciones y difieren de las reglas de asociación en que las relaciones entre los datos se basan en el tiempo.

Ejemplo: Una tienda de venta de electrodomésticos y equipos de audio analiza las ventas que ha efectuado usando análisis secuencial y descubre que el 30 por ciento de los clientes que compraron un televisor hace seis meses compraron un DVD en los siguientes dos meses.

3.3.4. Técnicas de Data Mining

Todo problema requiere de métodos, técnicas o algoritmos para sean resueltas y Data Mining no es exenta a ello. Resulta interesante conocer que, además que una tarea puede tener muchos métodos diferentes para ser resuelta y teniendo el mismo métodos (o al menos el mismo tipo de técnicas) puede resolver un gran abanico de tareas.

A continuación, se verán en modo de reseña, los tipos de técnicas existentes para llevar a cabo las tareas de Data Mining vistas en el punto anterior según lo que señala Hernández (2004).

I. Técnicas algebraicas y estadísticas

Basadas en expresar modelos y patrones mediante fórmulas algebraicas, funciones lineales, funciones no lineales, distribuciones o valores estadísticos tales como medias, varianzas, correlaciones, etc. Frecuentemente, estas técnicas, cuando obtienen un patrón, lo hacen a partir de un modelo ya predeterminado, del cual se estiman unos coeficientes o parámetros, de ahí el nombre de técnicas paramétricas. Algunos de los algoritmos más conocidos son la regresión lineal, la regresión logarítmica y la regresión logística. Además, los discriminantes lineales y no lineales basados en funciones predefinidas es decir, discriminantes paramétricos, entran dentro de esta categoría.

De manera general, la fórmula para una regresión lineal se presenta como $y = c_0 + c_1x_1 + \dots + c_nx_n$, donde x_i son los atributos predictores e y la salida (Variable dependiente). Si los atributos son modificados en la función de regresión por alguna otra función (cuadrados, inversa, logaritmos, combinaciones de variables...), es decir $y = c_0 + f_1(x_1) + \dots + f_n(x_n)$, la regresión se dice no lineal. Se pueden incorporar variables locales o transformaciones en las variables predictores y en la salida, permitiendo flexibilizar este tipo de técnicas. El abanico de técnicas se dispara aún más cuando consideramos técnicas no paramétricas.

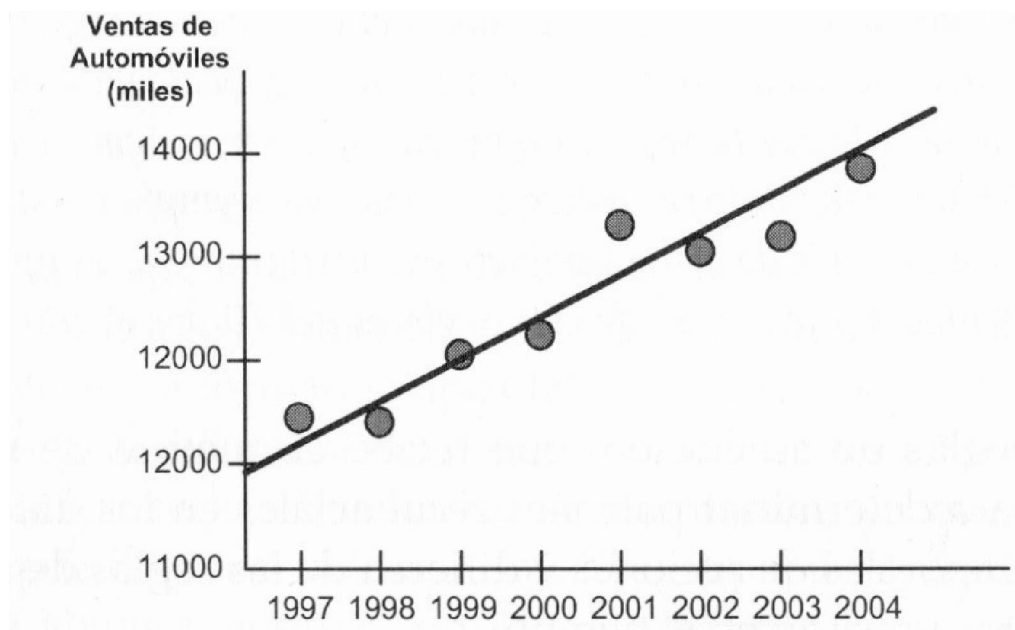


Figura N° 3.3: Ejemplo de regresión lineal

Fuente: Hernández, 2007, Pág. 28

II. Técnicas Bayesianas

Se basan en estimar la probabilidad de pertenencia (a una clase o grupo), mediante la estimación de las probabilidades condicionales inversas o a priori, utilizando para ello el Teorema de Bayes. Algunos algoritmos muy populares son el clasificador bayesiano naïve, los métodos basados en máxima verosimilitud y el algoritmo EM. Las redes bayesianas generalizan las topologías de las interacciones probabilísticas entra variables y permiten representar gráficamente dichas interacciones.

La regla de Bayes establece que si tenemos una hipótesis H sustentada para una evidencia E , entonces:

$$p(H | E) = \frac{p(E | H) \cdot p(H)}{p(E)} \quad (1)$$

1. Es la regla básica para realizar inferencia permitiendo actualizar la creencia que se tiene de un suceso o conjunto de sucesos a la luz de nuevos datos u observaciones.

III. Técnicas basadas en conteos de frecuencia y tablas de contingencias

Estas técnicas se basan en contar la frecuencia en la que dos o más sucesos se presentan conjuntamente. Cuando el conjunto de sucesos posibles es muy grande, existen algoritmos que van comenzando por pares de sucesos e incrementando los conjuntos sólo en aquellos casos en que las frecuencias conjuntas superen un cierto umbral. Su implementación se divide normalmente en dos fases: la extracción de los conjuntos de ítems que cumplan con la cobertura requerida desde los datos, y la generación de las reglas a partir de estos conjuntos. Ejemplos de estos algoritmos son el algoritmo "A priori" y similares. El funcionamiento de este algoritmo se basa en la búsqueda de los conjuntos de ítems con determinada cobertura. Para ello, en primer lugar se construyen simplemente los conjuntos formados por solo un ítem que superan la cobertura mínima. Este conjunto de conjuntos se utiliza para construir el conjunto de conjuntos de dos ítems, y así sucesivamente hasta que se llegue a un tamaño en el cual no existan conjuntos de ítems con la cobertura requerida.

IV. Técnicas basadas en árboles de decisión y sistemas de aprendizaje de reglas

Estos corresponden a condiciones organizadas en forma jerárquica, a modo de árbol. Son muy útiles para encontrar estructuras en espacios de alta dimensionalidad y en problemas que mezclen datos categóricos y numéricos. Esta técnica que, además de su presentación en forma de reglas, se basan en dos tipos de algoritmos:

- i. Divide y vencerás, como el ID3/C4.5 o el CART
- ii. Separa y vencerás, como el CN2.

V. Técnicas relacionales. Declarativas y estructurales

La característica principal de este conjunto de técnicas es que representan los modelos mediante lenguajes declarativos como los lenguajes lógicos, funcionales

o lógico-funcionales. Las técnicas de ILP (programación lógica inductiva) son las más representativas y las que han dado nombre a un conjunto de técnicas denominadas minería de datos relacional.

VI. Técnicas basadas en redes neuronales artificiales

Se trata de técnicas que aprenden un modelo mediante el entrenamiento de los pesos que conectan un conjunto de nodos o neuronas. La topología de la red y los pesos de las conexiones determinan el patrón aprendido. Existen innumerables variantes de organización: perceptrón simple, redes multicapa, redes de base radial, redes de Kohonen, etc., con no menos algoritmos diferentes para cada organización; el más conocido es el de retro propagación (backpropagation).

Las redes neuronales pueden utilizarse para el reconocimiento de patrones, la comprensión de información y la reducción de la dimensionalidad, el agrupamiento, la clasificación, la visualización, etc.

Una red neuronal puede verse como un grafo dirigido con muchos nodos (elementos del proceso) y arcos entre ellos (sus interconexiones). Cada uno de estos elementos funciona independientemente de los demás, usando datos locales (la entrada y la salida del nodo) para dirigir su procesamiento. En la Figura N° 3.4 se muestra la organización de una red neuronal.

Una red neuronal consta de una capa de entrada, en la que cada nodo corresponde a una variable independiente a examinar, unos nodos internos organizados en una o varias capas ocultas y una capa de salida con los nodos de salida (los posibles valores de las variables objetivo). Cada nodo de la capa de entrada está conectado a cada nodo de la capa oculta. Los nodos de la capa oculta pueden estar conectados a nodos de otra capa oculta o a los nodos de la capa de salida. Cada arco está etiquetado por un peso de conexión (w) y en cada nodo hay una función de activación que indica el efecto de ese nodo sobre

los valores de los atributos de una instancia en los nodos de entrada y los nodos de salida determinan la predicción para dicha instancia.

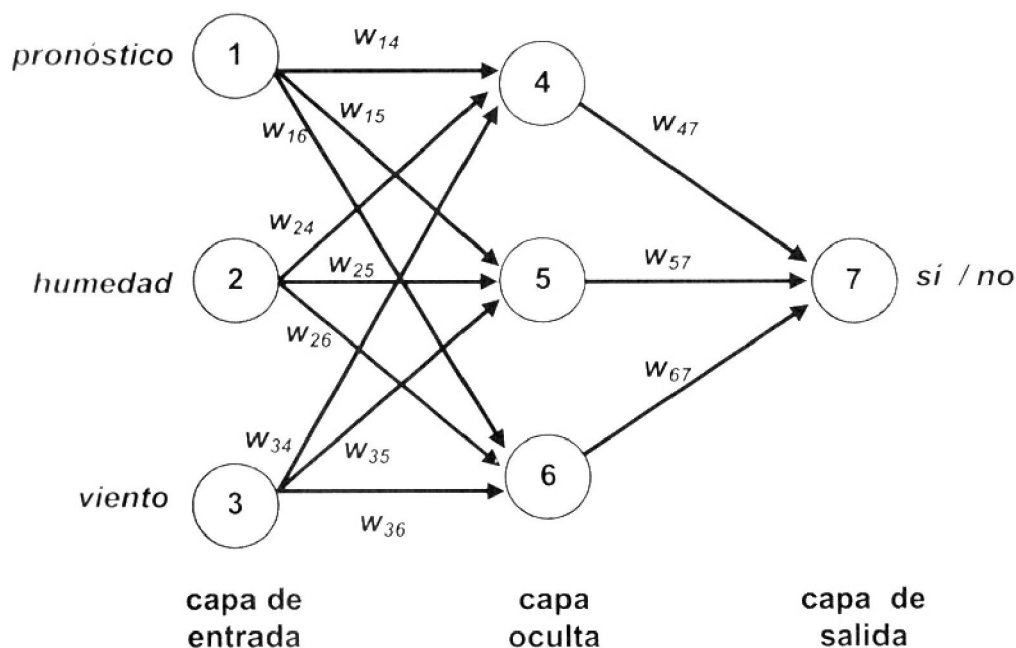


Figura N° 3.4: Red neuronal para el problema de jugar un cierto deporte

Fuente: Hernández, 2007, Pág. 32

VII. Técnicas basadas en núcleo y máquinas de soporte vectorial

Se trata de técnicas que intentan maximizar el margen entre los grupos o las clases formadas. Para ello se basan en unas transformaciones que pueden aumentar la dimensionalidad. Estas transformaciones se llaman núcleo (kernels). Existen muchísimas variantes, dependiendo del núcleo utilizado y de la manera de trabajar con el margen.

VIII. Técnicas estocásticas y difusas

Bajo él, se incluyen la mayoría de las técnicas que, junto a las redes neuronales, forman lo que se denomina computación flexible (Soft Computing). Son técnicas en las que o bien los componentes aleatorios son fundamentales,

como el simulated annealing, los métodos evolutivos y genéricos, o bien al utilizar funciones de pertenencia difusas (Fuzzy).

IX. Técnicas basadas en casos, en densidad o distancia

Son métodos que se basan en distancias al resto de elementos, ya sea directamente, como los vecinos más próximos (los casos más similares), de una manera más sofisticada, mediante la estimación de funciones de densidad. Además de los vecinos más próximos, algunos algoritmos muy conocidos son los jerárquicos, como Two-Step o COBWEB, y los no jerárquicos, como K medias.

En el aprendizaje basado en instancias, cada nueva instancia se compara con las existentes usando una métrica de distancia, y la instancia más próxima se usa para asignar su clase a la instancia nueva. La variante más sencilla de este método de clasificación es conocido como “el vecino más próximo” (Knearest-neighbors), usa los k vecinos más próximos, en cuyo caso la clase mayoritaria de estos k vecinos se asigna a la nueva instancia.

CAPÍTULO 4: TÉCNICAS DE DATA MINING UTILIZADAS EN EL PROYECTO

4.1 ¿Qué es el Análisis Cluster?

Según Hair, Anderson, Tatham & Black (2007), Análisis Cluster³, es la denominación de un grupo de técnicas multivariantes cuyo principal propósito es agrupar objetos basándose en las características que poseen. El Análisis Cluster clasificar objetos (es decir, encuestados, productos u otras entidades) de tal forma que cada objeto es muy parecido a los que hay en el conglomerado con respecto a algún criterio de selección predeterminado. Los conglomerados de objetos resultantes deberán mostrar un alto grado de homogeneidad interna (dentro del conglomerado) y un alto grado de heterogeneidad externa (entre conglomerados). Por tanto, si la clasificación es acertada, los objetos dentro de los conglomerados estarán próximos cuando se representen gráficamente, y los diferentes grupos estarán muy alejados.

Según Pérez (2008), el Análisis Cluster es un método estadístico multivariante de clasificación automática de datos. A partir de una tabla de casos-variables, trata de situar los casos (individuos) en grupos homogéneos, conglomerados o cluster, no conocidos de antemano pero sugeridos por la propia esencia de los datos, de manera que individuos que puedan ser considerados similares sean asignados a un mismo cluster, mientras que individuos diferentes (Disimilares) se localicen en cluster distintos. La diferencia esencial con el análisis discriminante estriba en que en este último es necesario especificar previamente los grupos por un camino objetivo, ajeno a la medida de las variables en los casos de la muestra, por lo tanto el Análisis Cluster define grupos tan distintos como sea posible en función de los propios datos.

El Análisis Cluster por lo tanto, es un método que permite descubrir asociaciones y estructuras en los datos que no son evidentes a priori pero que pueden ser útiles una vez que sean encontrados. Los resultados de un Análisis

Cluster pueden contribuir a la definición formal de un esquema de clasificación tal como una taxonomía para un conjunto de objetos, a sugerir modelos

3. También denominado de grupos, de conglomerados o tipologías. La denominación cluster está bastante extendida y se asocia con técnicas que nada tienen que ver con otra acepción que la expresión "Análisis de grupos".

estadísticos para describir poblaciones, a asignar nuevos individuos a las clases para diagnóstico e identificación, etc.

Las principales notas características del Análisis Cluster son Luque (2000):

- No hay distinción entre variables dependientes e independientes.
- Se persigue establecer grupos homogéneos internamente y heterogéneos entre ellos.
- Se pueden agrupar casos o individuos pero también variables o características, a diferencia del análisis factorial, que se centra en variables.
- Se trata de técnicas descriptivas, no de técnicas explicativas.
- Implícitamente se admite que en la población o conjunto de elementos a agrupar, y para el conjunto de características o variables que se dispone, existe la posibilidad de clasificar.

4.2 Aplicaciones del Análisis Cluster

Las aplicaciones de este análisis abarcan todas las ramas del conocimiento como por ejemplo, Biología estableciendo clasificaciones de seres vivos, ya sea plantas o animales e identificar especies y familias; en Psicología para identificar tipologías de personalidad, de comportamientos paranoicos o esquizofrénicos; en Sociología, para definir grupos sociales con arreglo a unas características determinadas; en Medicina, para establecer tipos de síntomas o de trastornos; en Arqueología, para describir diferentes objetos lícitos o herramientas de diversa índole, y así en otras ramas diferentes Luque (2000).

El Análisis Cluster se ha aplicado, en que investigaciones y advertencias y recomendaciones las cuales se detallarán a continuación:

En marketing por ejemplo:

- Estudiar tipologías de comportamientos de compra y de consumidores.
- Obtención de segmentos en un mercado determinado.
- Estudios de imagen y de necesidades con respecto a un producto.
- Lograr clasificaciones de productos, marcas o empresas y de esta forma entender mejor sus parecidos y sus diferencias identificando grupos estratégicos de determinada orientación estratégica de empresas o clases de productos competitivos.
- Identificación de grupos de ciudades y, dentro de cada uno de ellos, de la ciudad más representativa que pueda servir como laboratorio de pruebas para acciones de marketing como probar un diseño nuevo, una campaña de publicidad o una actividad promocional.

Con respecto a las investigaciones, a modo de ejemplo se pueden mencionar las siguientes:

- Para establecer grupos estratégicos de empresas del sector de la distribución comercial y además analizar su evolución.
- Para identificar los valores relevantes para una región concreta, entre los que destacan: familia, honradez, amistad o paz; y otros que no se consideran tan importantes (política, sentimiento, vivir al día, religión).
- Para identificar segmentos de consumidores con denominaciones de origen de Navarra y de Aragón.
- La satisfacción en el sector servicios es un tema complejo ya que se ha utilizado el Análisis Cluster para identificar grupos de usuarios de un determinado servicio publicado con diferente nivel de satisfacción.

- Para la caracterización del electorado potencial de un partido.

Finalmente resulta necesario mencionar algunas advertencias y recomendaciones al momento de aplicar Análisis Cluster.

1. Los datos deben estar en escalas comparables y, si es necesario, estandarizarlos.
2. La determinación del número de grupos debe inspirarse en el conocimiento teórico y práctico existente sobre el fenómeno en estudio. Las consideraciones de tipo operativo también sirven de orientación: así un número excesivamente reducido de grupos aportaría una información pobre, no generaría un incremento del conocimiento sobre la cuestión en estudio, mientras que si se trata de un elemento demasiado grande de grupos complica la interpretación y conduce a la confusión.
3. La distancia elegida puede constituir, por si misma, un elemento determinante en el proceso de agrupación.
4. Otro tanto ocurre con el orden de los casos en los procedimientos no jerárquicos. Es decir, el orden condiciona la composición de los grupos; si existe sospecha de la presencia de este efecto habría que proceder cambiando el orden.
5. Los procedimientos jerárquicos han perdido parte de su popularidad a favor de los no jerárquicos como consecuencia del desarrollo de programas informáticos. La clave está por lo tanto, en la selección de los puntos de origen; por lo demás les afectan en menor medida los valores raros, el tipo de distancia utilizada o la inclusión de variables relevantes.
6. El tamaño relativo de los grupos tiene su relevancia. No es apropiada una solución con grupos formados por uno o muy pocos elementos, en tanto que otros contienen muchos elementos.

4.3 Proceso del Análisis Cluster

El Análisis Cluster constituye una fase dentro de un proceso de investigación en el que ya se habrían fijado los objetivos, las hipótesis o los procedimientos de datos que se considerarán pertinentes. Incluso para la obtención de la información ya se deben contemplar los requerimientos de los análisis de datos que se piensan adecuados según nuestros objetivos de investigación. Para ello debemos destacar entonces los siguientes pasos que llevan a alcanzar dichos objetivos, según Hair, Anderson, Tatham & Black (2007).

Primer Paso: Objetivos del Análisis Cluster

El objetivo fundamental del Análisis Cluster es la obtención de un conjunto de objetos en dos o más grupos basándose en su similitud para un conjunto de características especificadas. Al formar grupos homogéneos, se alcanzan los siguientes objetivos graficados en la Figura N° 41.

1. **Descripción de una taxonomía:** el uso más tradicional del análisis cluster ha sido para propósitos exploratorios y la formulación de una taxonomía, es decir, de una clasificación de objetos realizada empíricamente.
2. **Simplificación de los datos:** en el curso de la obtención de una taxonomía, el análisis cluster también obtiene una perspectiva simplificada de las observaciones. Con una estructura definida, las observaciones pueden agruparse para análisis ulteriores.

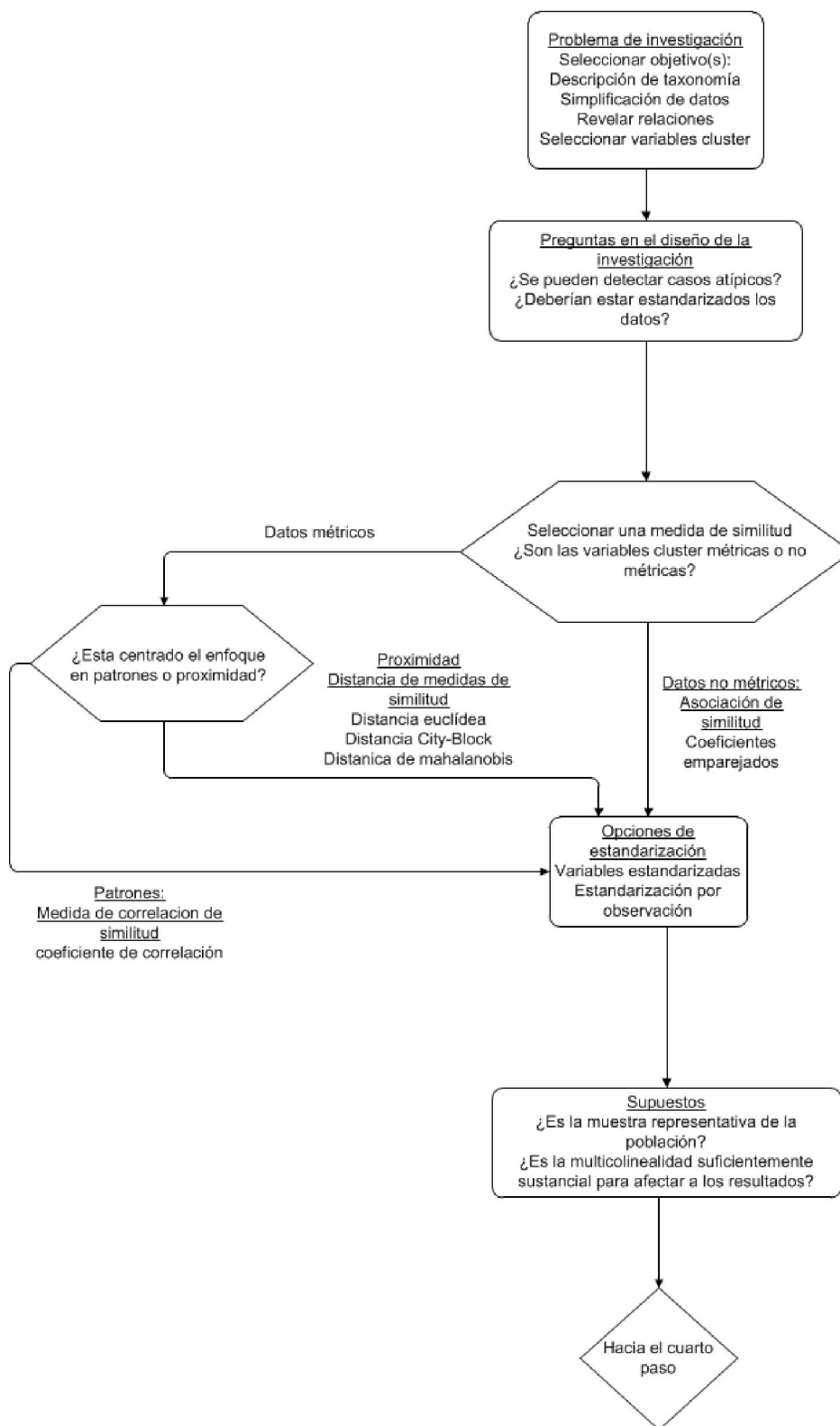


Figura N° 4.1: Diagrama de decisión de los pasos 1-3

Fuente: Hair, Anderson, Tatham & Black, 2007, Pág. 499

3. Identificación de relación: con los conglomerados definidos y la estructura subyacente de los datos representada en dichos conglomerados, se tiene un medio para revelar relaciones entre las observaciones que quizá no fuese posible con las observaciones individuales. Mientras se utilizan análisis tales como el discriminante para identificar relaciones empíricamente, o los grupos están sujetos a métodos cualitativos, la estructura simplificada del análisis cluster muchas veces representa relaciones o similitudes y diferencias no reveladas previamente.

En cualquier aplicación, los objetivos del análisis cluster no pueden separar de la selección de variables utilizadas para caracterizar los objetos a agrupar. Tanto si el objetivo es exploratorio como confirmatorio, se ha restringido efectivamente los resultados posibles por las variables elegidas para el uso. Los conglomerados derivados reflejan la estructura inherente de los datos solo como definida por las variables.

La selección de las variables a incluir en el valor teórico del análisis cluster, debe hacerse con relación a consideraciones teóricas, conceptuales y prácticas. Cualquier aplicación del análisis cluster debe descansar en cierta lógica en función de la cual se seleccionan las variables. Tanto si dicha lógica se basa en una teoría explícita, investigación pasada o suposición, el investigador debe darse cuenta de la importancia de incluir sólo aquellas variables que:

- i. Caracterizan los objetos que se están agrupando.
- ii. Se refieren específicamente los objetivos del análisis cluster.

Según Hair, Anderson, Tatham & Black, (2007), *“la técnica del análisis cluster no tiene un medio para diferenciar las variables relevantes de las irrelevantes. Solo obtiene los grupos de objetos más consistentes, aunque diferenciados, para todas las variables. La inclusión de una variable irrelevante aumenta la posibilidad de que se creen atípicos sobre esta variable, que pueden tener un efecto importante sobre los resultados. Por tanto, uno nunca debería incluir variables indiscriminadamente*

sino en su lugar, elegir las variables utilizando el objetivos de investigación como criterio de selección”.

Segundo Paso: Diseño de investigación mediante análisis cluster

Con los objetivos definidos y las variables seleccionadas, se debe tratar tres incógnitas antes de empezar el proceso de partición:

- i. ¿Pueden detectarse los atípicos y, si es posible, deberían ser destruidos?
- ii. ¿Cómo deberían medirse la similitud de los objetos?
- iii. ¿Deberían estandarizarse los datos?

La importancia de estos supuestos y las decisiones hechas en los últimos pasos se hacen aparentes cuando nos damos cuenta de que aunque el análisis cluster está buscando una estructura de los datos, debe imponerse en realidad una estructura a partir de una metodología seleccionada. El análisis cluster no puede evaluar todas las posibles particiones porque, incluso para un problema relativamente pequeño de partición de veinte cinco objetos en cinco conglomerados no solapados, existen $2,4 \cdot 10^{15}$ particiones posibles.

I. Detección de datos atípicos

En su búsqueda de una estructura, el análisis cluster es muy sensible a la inclusión de variables irrelevantes. Pero también, es sensible a los datos atípicos (datos que son muy diferentes al resto). Los datos atípicos pueden representar tanto:

- i. Observaciones verdaderamente aberrantes que no son representativas de la poblacional en general.
- ii. Una muestra reducida del grupo de la población que provoca una mala representación del grupo o los grupos de la muestra.

En ambos casos, los datos atípicos distorsionan la verdadera estructura y hacen que los conglomerados deducidos no sean representativos de la verdadera estructura de la población, por esta razón, siempre es necesaria una representación preliminar de estos por medio de un diagrama de perfil. Sin embargo, ante la presencia de una gran base de datos, la lectura de este diagrama se dificulta, es por eso que se realiza un estudio de perfil único que los distingue del resto de las observaciones.

Cualquiera que sea el medio utilizado, las observaciones identificadas como atípicas pueden evaluarse a efectos de su representatividad respecto de la población y eliminarlos del análisis si se consideran no representativos.

II. Medidas de similitud

La similitud entre objetos es una medida de correspondencia, o parecido, entre objetos que van a ser agrupados.

La similitud entre objetos puede medirse de varias formas, pero tres métodos dominan las aplicaciones del análisis cluster: medidas de correlación, medidas de distancia y medidas de asociación. Cada uno de los métodos representa una perspectiva particular de similitud, dependiendo tanto de sus objetivos como del tipo de datos. Tanto las medidas de distancia como la correlación exigen datos métricos, mientras que las medidas de asociación son para datos no métricos.

A. Medidas de correlación

La medida de similitud entre objetos que probablemente nos viene a la mente en primer lugar es el coeficiente de correlación entre un par de objetos medido sobre varias variables. En efecto, en lugar de hacer la correlación entre dos conjuntos de variables, invertimos la matriz de las "X variables" de los objetos de tal forma que las columnas representan los objetos y las filas representan las variables. Por tanto, el coeficiente de correlación entre las dos

columnas de números es la correlación (o similitud) entre los perfiles de los objetos.

B. Medidas de distancia

Representan la proximidad de las observaciones respecto a las otras para las variables del valor teórico del análisis cluster, son las medidas de similitud más utilizadas. Las medias de distancia son en realidad medidas de diferencia, donde los valores elevados indican una menor similitud. La distancia se convierte en medida de similitud utilizando una relación inversa.

Miden la distancia entre dos objetos de forma que, cuanto mayor (Rep. Menor) sea su valor, más (Rep. Menos) diferentes son los objetos y menor (Rep. Mayor) la probabilidad de que los métodos de clasificación los pongan en el mismo grupo. Ejemplo de esta medida la distancia euclidiana estándar, también llamada distancia métrica.

- **Distancia Euclidiana:** es la distancia geométrica en un espacio de unas dimensiones determinadas, adecuada especialmente para ejes ortogonales. Con dos dimensiones es la hipotenusa del triángulo rectángulo, mientras que para i dimensiones la distancia entre dos elementos X e Y es la raíz cuadrada de la suma de las diferencias al cuadrado para cada dimensión:

$$d(X,Y) = \sqrt{\sum_i (X_i - Y_i)^2} \text{ y } d(X,Y) = \sum_i (X_i - Y_i) \quad (2)$$

Además, se utiliza la distancia Euclidiana al cuadrado que es igual pero sin sacar la raíz cuadrada; o la distancia Euclidiana media que se obtiene dividiendo por el número de sumando o variables sobre las que se calculan las diferencias.

2. Ecuación para calcular la distancia geométrica.

- **Distancia de Minkowsky (City-Block):** esta se expresa como:

$$d(X, Y) = \left[\sum_i (X_i - Y_i)^n \right]^{1/n} \quad (3)$$

Para n mayor o igual a 1

Para $n=2$, es la distancia Euclidiana.

Para $n=1$, se trata de una distancia denominada de *city-block* o de *Manhattan* y consiste en sumar las diferencias absolutas entre las variables; en muchas ocasiones proporciona resultados similares a la distancia Euclidiana. Su expresión es:

$$d(X, Y) = \sum_i |X_i - Y_i| \quad (4)$$

- **Distancia de Chebychev:** Su expresión es la siguiente

$$d(X, Y) = \max |X_i - Y_i| \quad (5)$$

- **Distancia de Mahalanobis:** esta distancia tiene una métrica distinta a la de la Euclidiana. Es recomendable para situaciones en las que se produce multicolinealidad. Coincide con la Euclidea para variables estandarizadas cuando las correlaciones son nulas. Se expresión es la siguiente:

$$d(X, Y) = (X_i - Y_i) W^{-1} (X_i - Y_i) \quad (6)$$

W : Matriz de covarianzas

-
3. Representa la distancia a recorrer entre dos puntos con un trazado urbanístico perpendicular.
 4. Valor absoluto de la distancia de Minkowsky.
 5. Es el máximo valor absoluto de la distancia Minkowsky.
 6. Considera sólo la distancia de una observación desde los valores medios de las var. Indep.

C. Medidas de asociación

Tienen un carácter cualitativo y se obtienen a partir de la existencia de coincidencias, de acuerdos o desacuerdos. La media del parecido entre objetos a través de sus características cualitativas también se suele denominar medidas de similitud que, al contrario de la distancia, a mayor similitud mayor parecido. La media de la similitud debe satisfacer las siguientes condiciones:

- La similitud entre A y B es igual que la distancia entre B y A : $S_{AB} = S_{BA}$.
- La similitud de A consigo mismo o de B consigo mismo es igual y mayor que la similitud existente entre A y B : $S_{AA} = S_{BB} > S_{AB}$ para $A \neq B$.
- Al igual que en la distancia, a veces se añade una tercera condición conocida como la desigualdad del triángulo: si S_{AB} y S_{BC} son grandes, entonces S_{AC} también lo es.

Las medidas de similitud se obtienen bien interrogando directamente a personas sobre su percepción, bien midiendo las características o la posesión de atributos. En este sentido, un caso particular es la utilización de variables binarias (con valores 0, no posesión de un atributo y 1, posesión de un atributo) a partir de las cuales se construyen índices sabiendo que las posibilidades que se pueden presentar como:

p = Número de coincidencias positivas (1,1)

n = Número de coincidencias negativas (0,0)

d = Número de diferencias (1,0)

e = Número de diferencias (0,1)

$$t = p + n + d + e \quad (7)$$

La combinación de estas situaciones da lugar a una amplia gama de Índices, según se prime las coincidencias positivas y/o negativas o las diferencias de uno u otro tipo. Algunos de estos índices son: el índice de Sokal y Michener, Roger y Tanimoto, Sokal y Sneath.

III. Tipificación de los datos

Con la medida de similitud seleccionada, luego se debe tratar sólo con una pregunta más: ¿Deberían tipificarse los datos antes de calcular las similitudes? Al contestar esta, se debe considerar varios asuntos. En primer lugar, la mayoría de las distancias medidas son bastantes sensibles a las diferentes escalas o magnitudes de las variables.

A. Estandarización por variables

La manera más común de estandarización es la conversión de cada variable a unas puntuaciones estándar (también conocidas como puntuaciones Z) restando la media y dividiendo por la desviación típica de cada variable. Esta es una opción que tienen todos los programas informáticos y muchas veces incluso está incluida en el procedimiento de análisis cluster. Es una forma general de una función de distancia normalizada que utiliza una medida de distancia Euclídeana susceptible de transformación normalizada de los datos originales. Este proceso convierte cada puntuación de los datos originales en un valor estandarizado con una media de 0 y desviación estándar de 1. Esta transformación, a cambio, elimina el sesgo introducido por las diferencias en las mediciones de varios atributos o variables utilizadas en el análisis.

7. Ecuación general de las medidas de asociación.

B. Estandarización por la observación

Contando con una base de encuestados, se puede aplicar análisis cluster y obtener conglomerados, pero una posibilidad muy distinta es que se obtengan conglomerados de gente que dice que sus respuestas son importantes, o de gente que dice que algo tenía poca importancia y quizá algunos conglomerados intermedios. Lo que estamos viendo son los efectos de tipo respuesta en los conglomerados. Los efectos de tipo de respuesta son las pautas sistemáticas de respuesta a un conjunto de cuestiones, tales como los que siempre dicen sí (respuesta favorable para todas las cuestiones) o los que siempre dicen no (respuesta desfavorable para todas las cuestiones).

Tercer Paso: Supuestos del Análisis Cluster

El análisis cluster, no es una técnica de inferencia estadística en la que se analizan parámetros de una muestra en la medida en que puedan ser representativos de una población. Por el contrario, es una metodología objetiva de cuantificación de las características estructurales de un conjunto de observaciones. Como tal, tiene fuertes propiedades matemáticas pero no fundamentos estadísticos. Las exigencias de normalidad, linealidad y homocedasticidad que son tan importantes en otras técnicas realmente tienen poco peso en este tipo de análisis. Sin embargo, los esfuerzos deben centrarse en dos asuntos críticos que son: la representatividad de la muestra y la multicolinealidad.

I. Representatividad de la muestra

Rara vez se tiene un censo con el fin de utilizarlo en el análisis cluster. Normalmente se obtiene una muestra y los conglomerados se derivan en la esperanza de que representen la estructura de la población. Es así que se debe confiar en que la muestra obtenida es verdaderamente representativa de la población. Como ya se ha mencionado, los datos atípicos pueden ser en realidad producto de una muestra escasa de grupos divergentes que, cuando se descartan, introducen sesgos en la estimación de la estructura. Por tanto, el análisis cluster solo

es tan bueno como la representatividad de la muestra. Por consiguiente, todos los esfuerzos deberían dirigirse a asegurar que la muestra es representativa y que los resultados son generalizables para la población a estudiar.

II. Impacto de la multicolinealidad

La multicolinealidad actúa como proceso de ponderación no aparente, sin embargo afecta el análisis. Por esta razón, se anima a examinar las variables utilizadas en el análisis cluster en busca de una multicolinealidad substancial y, si se encuentra, o bien reducir las variables al mismo número en cada conjunto o bien utilizar una de las distancias medidas, como la distancia de Mahalanobis⁴, que compensa esta correlación.

Cuarto Paso: Obtención de conglomerados y valoración del ajuste conjunto

Con las variables seleccionadas y la matriz de similitud calculada, comienza el proceso de partición señalado en la Figura N° 4.2. Es acá donde se debe seleccionar en primer lugar el algoritmo de aglomeración utilizado en la formación de conglomerados y a continuación tomar la decisión del número de conglomerados que se van a formar. Ambas decisiones tienen implicaciones substanciales no solo sobre los resultados que se obtendrán, sino también sobre la interpretación que se puede derivar de los resultados.

Lo primero que hay que contestar en esta fase es, ¿Qué procedimiento debería utilizarse para colocar objetos similares en grupos o conglomerados? Esto es, ¿Que algoritmo de obtención de conglomerados o qué conjunto de reglas es más apropiado? El criterio esencial de todos los algoritmos es, sin embargo, que intenta maximizar las diferencias entre los conglomerados relativos a la variación dentro de los conglomerados.

4. Considera sólo la distancia de una observación desde los valores medios de las variables independientes.

Los algoritmos de obtención de los conglomerados se clasifican en dos grandes grupos que son: Jerárquicos y no Jerárquicos.

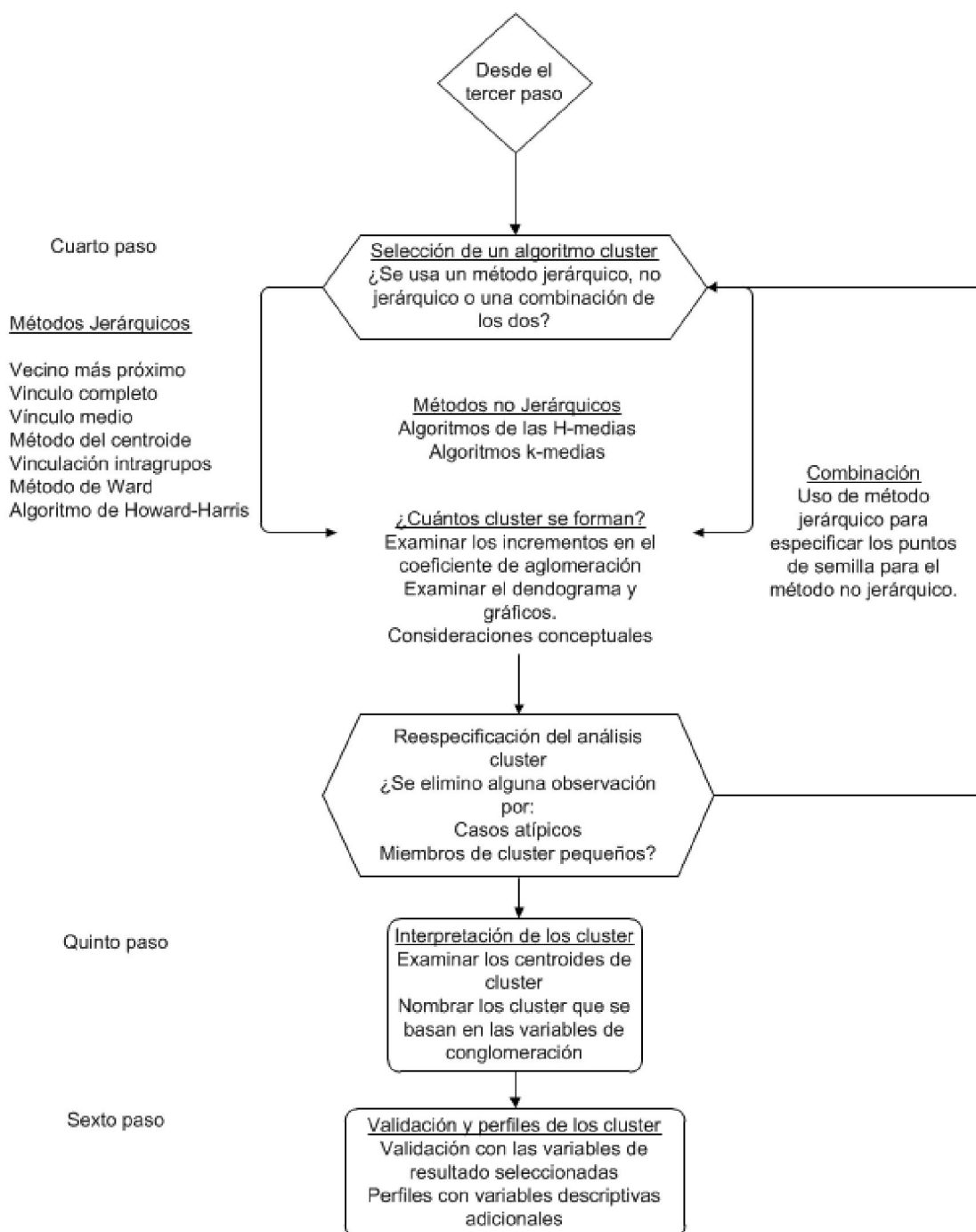


Figura N° 4.2: Pasos 4-6 del diagrama de decisión del análisis Cluster

Fuente: Hair, Anderson, Tatham & Black, 2007, Pág. 509

I. Métodos Jerárquicos

Suponen desarrollar una jerarquía, la formación de grupos constituye un proceso secuencias que se representa gráficamente. Dada una población, se trata de establecer una jerarquía de partes, delimitando un número de subconjuntos de forma que entre ellos no tengan elementos comunes (Disimilares) y que cada subconjunto este incluido en otro (Jerarquía).

Dentro de los procedimientos jerárquicos la literatura señala los siguientes:

- **Ascendentes (Joining)**, comienza con n cluster de un objeto cada uno. En cada paso del algoritmo se recalculan las distancia entre los grupos existentes y se unen los 2 grupos más similares o menos disimilares. El algoritmo acaba con 1 cluster conteniendo todos los elementos. Ejemplo de este método son: Método de Ward, Distancia mínima (Single Linkage) y Distancia máxima (Complete Linkage).
- **Descendentes (Divisive)**, que consisten justamente en lo contrario, partir del conjunto poblacional e ir dividiendo en subconjuntos hasta llegar al elemento. Es decir, comienzan con 1 cluster que engloba a todos los elementos haciendo que en cada paso del algoritmo se divide el grupo más heterogéneo. El algoritmo acaba con n cluster de un elemento cada uno.

Ahora bien, a continuación se detallará acerca de los procedimientos más utilizados que son:

A. Vínculo único vecino más próximo (Single linkage o Nearest neighbor)

Donde los grupos se unen considerando la menor de las distancias existentes entre los miembros más cercanos de distintos grupos. Con ello permite crear grupos más homogéneos pero a su vez, permite cadenas de alineamientos entre sujetos muy lejanos. A continuación se identifica la distancia mínima siguiente y habrá un tercer individuo que se incorpore al

grupo o bien se formará un nuevo grupo con dos nuevos componentes; así se continúa hasta que todos los objetos están comprendidos en un solo grupo. Los grupos deben estar nítidamente definidos para que proporcione buenos resultados.

B. Vínculo completo (Complete linkage o furthest Neighbour)

Es similar al anterior salvo que ahora como distancia entre dos grupos se toman la mayor existente (vecino más lejano) entre dos objetos cualquiera, cada uno de un grupo diferente.

C. Vínculo medio (Average Linkage)

En este caso no solamente intervienen dos individuos o elementos, sino que para calcular la distancia entre grupos se recurre a la media pero este adopta muchas variantes. Así, puede ser la distancia media entre pares de objetos de los dos grupos sin ponderar (Unweighted pair-group average), o ponderando por el tamaño de los grupos o números de elementos de cada uno, sobre todo en el caso de grupos de tamaño muy diferente (weighted pair-group average). Con este procedimiento se considera más información, no solamente la derivada de las situaciones extremas de distancia mínima o máxima.

D. Método del centroide

El centroide de un grupo es el punto medio en un espacio multidimensional determinado por las dimensiones o variables que se consideran en nuestro análisis. A medida que se producen incorporaciones a un grupo y el número de elementos que lo integran varía, el centroide también se modifica. Los métodos que se basan en el centroide toman la distancia entre grupos como la distancia entre sus centroides o centros de gravedad. Precisamente al considerar un punto medio, los valores extraños o raros no influyen tanto en este método. También tiene variantes según se pondere, es decir, según se

considere el tamaño de los grupos o no. Se utiliza el ponderado cuando los grupos se estiman que son sensiblemente diferentes.

$$d^2(X, Y) = \sum_i (\bar{X}_i - \bar{Y}_i)^2 \quad (8)$$

Donde \bar{X}_i y \bar{Y}_i son las medias de las variables X_i y Y_i .

E. Vinculación Intra-Grupos

Este método de vinculación, aprovecha la información de todos los miembros de los dos conglomerados que se comparan uniéndolos previamente. La distancia entre dos conglomerados se calcula como:

$$d_{AB} = \frac{1}{C_{n_A+n_B}^2} \sum_{i,j \in A \cup B} d_{ij} \quad (9)$$

Siendo $n_A = \text{Card}(A)$, $n_B = \text{Card}(B)$ y $C_{n_A+n_B}^2$ las combinaciones sin repetición de $n_A + n_B$ elementos tomados de dos en dos.

Tanto este método como el método de Ward, que se comenta a continuación, son sensibles a posibles transformaciones monótonas de los datos.

F. Método de Ward

Forma parte de los denominados métodos de la varianzas porque utiliza un análisis de la varianza para evaluar las distancias entre grupos. En este caso se intenta minimizar la suma de los cuadrados de los grupos que se pueden formar en cada paso, los grupos se van formando de manera que se produzca el menor aumento en las sumas de los cuadrados. Es un procedimiento que

8. Fórmula que se ocupa para calcular la distancia entre puntos medios de los conglomerados.

9. Es la distancia promedio existente entre todos los miembros del conglomerado unidos a ambos.

tiende a producir grupos pequeños equilibrados en cuanto al número de elementos que los integran.

La idea central es ir agrupando de forma jerárquica elementos de modo que se minimice una determinada función objetivo, con ello perseguirá la minimización de la variación intra-grupal de la estructura formada. Este método tiende a generar conglomerados demasiados pequeños y demasiados equilibrados en tamaño, además de que es un método menos sensible a outliers.

El método, en si busca minimizar $\sum_r SSW_r$

$$SSW_r = \sum_{m=1}^{n_r} \sum_{j=1}^p (X_{rjm} - \bar{X}_{rj})^2 \quad (10)$$

Donde X_{rjm} denota el valor de la variable X_j en el m-ésimo elemento el grupo r .

En cada paso del algoritmo une los grupos r y s que minimizan:

$$SSW_t - SSW_r - SSW_s = \frac{n_r n_s}{n_r + n_s} d_{rs}^2 \quad (11)$$

Con $t = r \cup s$ y d_{rs}^2 la distancia entre los centroides de r y s

G. Algoritmo de Howard-Harris

A diferencia de los anteriores este es un procedimiento de tipo descendente ya que va formando grupos a partir de otros y de forma secuencial utilizando el criterio de minimizar la varianza intra-grupos en cada subdivisión. Su forma de operar es la siguiente:

10. Corresponde a las sumas de cuadrados intra-grupo para cada grupo r .
 11. Formula general del método de Ward.

- i. Selecciona la variable que tienen mayor varianza, distinguiendo dos grupos según estén por encima o por debajo de la medias.
- ii. Comprueba que cada elemento forma parte del grupo que le es más afín. Para ello calcula la distancia Euclideana del elemento a los valores medios del grupo para asignarlo al más próximo, aunque no sea al que provisionalmente se asignó, en tal caso se trata de una reasignación. Así se consolidan los dos primeros grupos.
- iii. Calculada la suma de cuadrados para cada grupo o la suma de las diferencias entre los valores de las variables y los valores medios del grupo, se selecciona el que tenga dicha suma mayor y se elige la variable con mayor varianza dividiendo, como antes se comentó, en dos grupos según los elementos tengan valores superiores o inferiores a la media de tal variable.
- iv. Ahora con tres grupos se procede igual que en el punto ii para comprobar y, si es necesario, reasignar hasta que no se produzca reasignaciones o se llegue a un número de iteraciones establecido.
- v. Para obtener un grupo más se procede de igual forma.
- vi. El proceso finaliza cuando:
 - Se consigue un número de grupos determinado.
 - El tamaño de los grupos no llega a un mínimo establecido.
 - No se consigue una reducción significativa de la suma de cuadrados, en suma, de la variabilidad de los grupos.

II. Métodos no Jerárquicos

Son aquellos que asignan los casos a grupos diferenciados que el propio análisis configura, sin que unos dependan de otros. Es así, que resulta intuitivo

suponer que una clasificación correcta debe ser aquella en que la dispersión dentro de cada grupo formado sea el menor posible (*Criterio de Varianza*) y lleva a seleccionar una configuración cuando la suma de las varianzas dentro de cada grupo (Varianza Residual) sea mínima. La literatura ha propuesto diversos algoritmos de clasificación no jerárquica basados en minimizar progresivamente esta varianza, que difieren en la elección de los cluster provisionales que necesita el arranque del proceso y en el método de asignación de individuos a los grupos., los cuales se mencionarán a continuación, Pérez (2005).

A. Algoritmo de las H-medias

Parte de una primera configuración arbitraria de grupos con su correspondiente media, eligiendo un primer individuo de arranque de cada grupo asignando posteriormente cada caso al grupo cuya media es más cercana. Una vez que todos los casos han sido ubicados, calcula de nuevo las medias o centroides y la toma en lugar de los primeros individuos como una mejor aproximación de los mismos, repitiendo el proceso mientras la varianza residual vaya disminuyendo. La partición de arranque define el número de cluster que, lógicamente puede disminuir si ningún caso es asignado a alguno de ellos.

B. Algoritmo de las K-medias

Es el más importante, y parte al igual que el algoritmo de las H-medias, de unas medias arbitrarias y mediante pruebas sucesivas, contrasta el efecto que sobre la varianza residual tiene la asignación de cada uno de los casos a cada uno de los grupos. El valor mínimo de varianza determina una configuración de nuevos grupos con sus respectivas medias. Se asignan otra vez todos los casos a estos nuevos centroides en un proceso que se repite hasta que ninguna transferencia puede ya disminuir la varianza residual, o se alcance otro criterio de parada (un número limitado de pasos de iteración) o simplemente, que la diferencia obtenida entre los centroides de dos pasos

consecutivos sea menor que un valor prefijado. El procedimiento configura los grupos maximizando, a su vez, la distancia entre sus centros de gravedad y como la varianza total es fija, minimizar la residual hace máxima la factorial o inter-grupos. Puesto que minimizar la varianza residual es equivalente a conseguir que sea mínima la suma de distancias al cuadrado desde los casos a la media del Cluster al que van a ser asignados, es esta distancia Euclideana al cuadrado la utilizada por el método.

Este tipo de método es conveniente utilizarlo cuando los datos a clasificar son muchos y/o para refinar una clasificación obtenida utilizando un método jerárquico.

Ahora bien, este tipo de agrupación no jerárquica, presenta tres ventajas importantes:

- i. El procedimiento exige que en un principio, se infiera el número de agrupamientos que van a existir.
- ii. Selección de los conglomerados iniciales de los agrupamientos influye mucho sobre el procedimiento. Además si se deja que el software elija los conglomerados a su elección, a menudo depende del orden en que se leen los datos, de modo que por ejemplo, dos investigadores podrían realizar un análisis por agrupación sobre el mismo conjunto de datos y producir agrupamientos completamente diferentes.
- iii. Con frecuencia, el procedimiento no es factible desde el punto de vista del cálculo porque hay precisamente demasiadas elecciones posibles, no solo para el número de agrupamientos, sino también para las ubicaciones de estos últimos.

Quinto Paso: Interpretación de los conglomerados

En primer lugar hay que conocer algo tan obvio como el número y composición de los mismos. Si se ha seguido un procedimiento no jerárquico esto es inmediato, puesto que incluso se define el número de grupos a priori, eso si bien lleva a una solución, no se puede comprobar si es la mejor entre las posibles.

Ahora bien, en caso de haber optado por un procedimiento jerárquico, el número de los conglomerados no es algo tan evidente. Normalmente, la representación gráfica de cómo se forman los grupos pone en relación la distancia de unión entre dos elementos con la distancia mayor existente; por tanto el número de grupos depende de la distancia a la que se haga el corte para analizar. Si el corte se hace a distancias pequeñas el número de grupos será mayor que si se toman distancias grandes, ya que todos los elementos estarán comprendidos en pocos grupos.

Concretando el número de conglomerados y su composición, la interpretación de cada uno de ellos se efectúa considerando las características de los elementos que lo componen y analizando si poseen o representan determinadas características en mayor medida que otras, recurriendo a las estadísticas descriptivas por grupo de las variables de partida. El centroide de un conglomerado es un buen referente para la descripción, sin olvidar que en realidad los grupos muestran tendencias dentro de la población que se estudia y no es fácil que sean agrupaciones incuestionables o puras. Por otro lado, recurrir a variables diferentes de las utilizadas para el análisis por ejemplo, relacionadas con características socioeconómicas o psicográficas de los integrantes del conglomerado, ayuda a la interpretación y a extraer conclusiones. Todo lo anterior permite calificar o poner nombres a los grupos.

La interpretación en los no jerárquicos se enriquece recurriendo a un análisis de la varianza para examinar las diferencias entre los grupos. Si el análisis ha cumplido con su objetivo de conseguir una buena clasificación de elementos, la variabilidad dentro de un conglomerado será pequeña y la variabilidad entre

conglomerados será grande. La comparación de las diferencias al cuadrado entre conglomerados con las de dentro del conglomerado nos permite obtener un ratio F y un nivel de significación. Las variables con un F grande y el nivel de significación pequeño difieren entre los distintos conglomerados. Sin embargo, debe quedar claro que aquí el test F debe utilizarse con precaución y, desde luego, con un sentido descriptivo, puesto que se han calculado precisamente para maximizar la diferencia entre ellos y no a partir de ninguna hipótesis a priori. En suma, no se debe utilizar como prueba de igualdad de la medias de los conglomerados.

Algunas ideas que pueden ser útiles en la interpretación de los resultados son las siguientes:

- Realizar ANOVA y MANOVAS para ver qué grupos son significativamente distintos y en qué variables lo son.
- Realizar un Análisis factorías o de Componentes Principales para representar, gráficamente los grupos obtenidos y observar las diferencias existentes entre ellos.
- Calcular perfiles medios por grupos y compararlos.

Sexto Paso: Validación y Perfil de los Grupos

Una vez obtenidos los grupos e interpretado los resultados conviene, siempre que sea posible, proceder a la validación de los mismo con el fin de averiguar, por un lado, hasta qué punto los resultados obtenidos son extrapolables a la población de la que vienen los objetos seleccionados y, por el otro lado, por qué han aparecido dichos grupos. Esta validación se puede realizar de forma externa o interna.

- **Validez interna**

Se puede establecer utilizando procedimientos de validación cruzada. Para ello se dividen los datos en dos grupos y se aplica el algoritmo de clasificación

a cada grupo comparando los resultados obtenidos en cada grupo. Por ejemplo, si el método utilizado es el de las k-medias se asignaría cada objeto de uno de los grupos al cluster más cercano obtenido al clasificar los datos el otro grupo y se mediría el grado de acuerdo entre las clasificaciones obtenidas utilizando los dos métodos.

- **Validez externa**

Se puede realizar comparando los resultados obtenidos con un criterio externo (por ejemplo, clasificaciones obtenidas por evaluadores independientes o analizando en los grupos obtenidos, el comportamiento de variables no utilizadas en el proceso de clasificación) o realizando un Análisis Cluster con una muestra diferente de la realizada.

4.4 Regresión Logística Binaria

El modelo de regresión logística es un procedimiento por medio del cual se intenta analizar las relaciones de asociación entre una variable dependiente dicotómica (binaria o dummy) Y y una o varias variables independientes (regresores o predictores) X_n , cuantitativas o categóricas, todo ello a fin de lograr los siguientes objetivos:

- Determinar la existencia o ausencia de relación entre una o más variables independientes y la variable dependiente.
- Medir la magnitud de dicha relación y estimar o predecir la probabilidad de que se produzca (o no) el suceso definido por la variable dependiente en función de los valores que adaptan las variables independientes.

Por tanto, la regresión logística binaria es uno de los modelos lineales generalizados más frecuentes, ya que modeliza una probabilidad (probabilidad de compra de un producto entre varios, probabilidad de fraude, probabilidad de morosidad, etc.) en este caso la variable de respuesta tiene dos (o más)

posibilidades, cada una con su respectiva probabilidad, siendo la suma de probabilidades igual a uno, Hernández (2004).

El análisis de regresión logística es una herramienta muy flexible en cuanto a la naturaleza de las variables explicativas, pues estas pueden ser de escala y categóricas.

Si se supone que se tiene la variable de estudio codificada de la siguiente manera:

$$Y = \begin{cases} 0, & \text{no ocurre el suceso} \\ 1, & \text{si ocurre el suceso} \end{cases}$$

Además, si se considera que solo se tiene una variable explicativa X, en estas condiciones se podría considerar un modelo de regresión lineal con el propósito de ver que dificultades puede surgir:

$$y_i = p_i = a + b \cdot x_i + u_i \quad (12)$$

Si se estima este modelo y se representa gráficamente la recta de regresión queda como en la figura

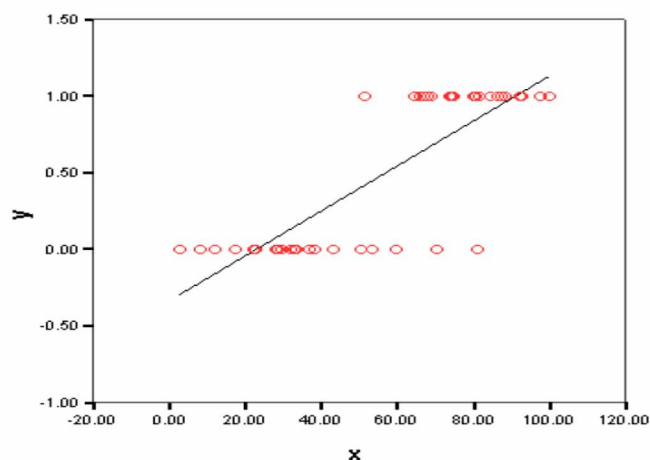


Figura N° 4.3: Representación Gráfica del modelo lineal
Fuente: Luque, 2000, Pág. 34

12. Fórmula de regresión lineal básica.

Se puede observar que la línea de regresión no está acotada en el intervalo $[0,1]$ y, por lo tanto, ya no va a representar una probabilidad.

Además, consideraciones de índole matemática lleva a la conclusión de que los residuos no van a ser homocedásticos y, por tanto, la técnica de estimación por mínimos cuadrados dejará de ser un método óptimo de estimación.

Una forma que se tiene de garantizar que los valores pronosticados estén en el intervalo $[0,1]$ es considerar la siguiente transformación:

$$p(a/x) = F(x \cdot b) \quad (13)$$

Donde F es una función de distribución:

$$F : R \rightarrow R$$

Una función de distribución es una función real de variable real, de forma que verifica que:

- Está acotada en el intervalo $[0,1]$

$$0 \leq F(x) \leq 1 \forall x$$

- Es monótona no decreciente

$$x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$$

Y además, está definida en todo R , tomando los siguientes valores:

$$F(-\infty) = 0$$

$$F(+\infty) = 1$$

13. Fórmula que considera la transformación para poner a prueba una función acotada $[0,1]$.

En general, la gráfica de una función de distribución es como la figura.

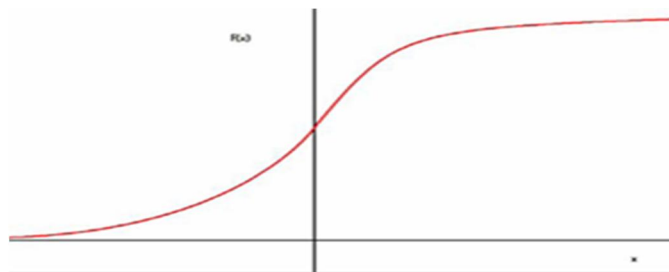


Figura N° 4.4: Representación gráfica de la función logística

Fuente: Luque, 2000, Pág. 34

Si se utiliza la función de distribución logística, el análisis se denomina regresión Logística, y si se utiliza la función de distribución normal se denomina regresión Probit.

4.4.1 El modelo de Regresión Logística

El modelo de regresión logística parte de la hipótesis de que los datos siguen el siguiente modelo:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k + u = x \cdot b + u \quad (14)$$

Con el fin de simplificar la notación, se define Z como:

$$Z = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k \quad (15)$$

Por lo tanto, el modelo se puede representar como:

$$\ln\left(\frac{p}{1-p}\right) = z + u \quad (16)$$

-
- 14. Hipótesis que pone a prueba el origen de la fórmula de regresión logística.
 - 15. Transformación de una ecuación a un valor Z .
 - 16. Expresión logarítmica de la ecuación.

Donde p es la probabilidad de que ocurra el suceso de estudio. Operando algebraicamente sobre el modelo, se tiene:

$$\begin{aligned} \ln\left(\frac{p}{1-p}\right) &= z \\ \frac{p}{1-p} &= e^z \\ p &= (1-p) \bullet e^z \\ p &= e^z - p \bullet e^z \\ p + p \bullet e^z &= e^z \\ p(1+e^z) &= e^z \\ p &= \frac{e^z}{1+e^z} \end{aligned} \quad (17)$$

Como la función de distribución logística es:

$$F(x) = \frac{e^x}{1+e^x} \quad (18)$$

Por tanto, se puede reescribir el modelo de forma mucho más compacta:

$$p = \frac{e^z}{1+e^z} = F(z) = F(x \bullet b) \quad (19)$$

De donde se deduce que el modelo de regresión logística es, en principio, un modelo de regresión no lineal, pero es lineal en escala logarítmica atendiendo a su definición original:

$$\begin{aligned} \ln\left(\frac{p}{1-p}\right) &= z \\ \ln(p) - \ln(1-p) &= z \\ \ln(p) - \ln(1-p) &= b_0 + b_1 \bullet x_1 + b_2 \bullet x_2 + \dots + b_k \bullet x_k \end{aligned} \quad (20)$$

-
- 17. Operación algebraica para conseguir el modelo de regresión logística binaria.
 - 18. Función de distribución logística binaria.
 - 19. Forma compacta de la función logística binaria.
 - 20. Comprobación de que el modelo de regresión logística binaria, proviene de un modelo de regresión lineal.

Es decir, la diferencia de la probabilidad de que ocurra un suceso respecto de que no ocurra es lineal pero en escala logarítmica. Por tanto, el significado de los coeficientes, aunque guardando una cierta relación con el modelo de regresión lineal es algo más complejo de interpretar.

Las dos formas más importantes de expresar el modelo de regresión logística:

$$\ln(p) - \ln(1 - p) = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k \quad (21)$$

$$\frac{p}{1 - p} = e^{b_0} \cdot e^{b_1 \cdot x_1} \cdot e^{b_2 \cdot x_2} \dots e^{b_k \cdot x_k} \quad (22)$$

4.4.2 Estimación de los parámetros

Sea una muestra de n elementos, donde se ha observado la variable respuesta Y (que solo puede tomar dos valores: cero y uno) y la variable x .

La función de probabilidad de una observación cualquiera es:

$$\begin{aligned} P(Y = 1 / x) &= p \\ P(Y = 0 / x) &= 1 - p \end{aligned} \quad (23)$$

Por lo tanto:

$$P(Y / x) = p^y \cdot (1 - p)^{1-y} \quad (24)$$

Entonces la función de probabilidades de la muestra es:

$$P(y_1, y_2, \dots, y_n) = \prod_i p_i^{y_i} \cdot (1 - p)^{1-y_i} \quad (25)$$

-
21. Una de las forma de representar el modelo de regresión logística binaria (Logit).
 22. Forma compactada de representar el modelo ya antes mencionado. (Cociente de probabilidades).
 23. Formulación de probabilidad elemental.
 24. Transformación algebraica para estimar parámetros.
 25. Función de la probabilidad de la muestra llamada (Likelihood).

Esta expresión recibe el nombre de verosimilitud de la muestra (Likelihood). Tomando logaritmos, se tiene lo siguiente:

$$\log P(Y) = \sum_i^n y_i \cdot \log\left(\frac{p_i}{1-p_i}\right) + \sum_i^n \log(1-p_i) \quad (26)$$

Expresando p en función de los parámetros que se desea estimar:

$$L(B) = \sum_i^n y_i \cdot x_i \cdot b - \sum \log(1 + e^{x_i \cdot b}) \quad (27)$$

Resulta evidente que aunque se derive y se establezca la condición de máximo, no se puede despejar los coeficientes b .

La solución es:

$$B_a = B_0 + \left(-\frac{\partial^2 L(B)}{\partial B \cdot \partial B'} \right)^{-1} \cdot \left(\frac{\partial L(B)}{\partial B} \right) \quad (28)$$

Esta solución establece como encontrar una solución B_a , la ordenada en el origen, y $\beta_1 = (\beta_1, \dots, \beta_k)$. A veces, se utilizan también como parámetros $\exp(\beta_0)$ y $\exp(\beta_j)$, que se denominan los Odds ratios o Ratios de Probabilidades. Estos valores indican cuanto se modifican las probabilidades por unidad de cambio en las variables x .

$$O_i = \frac{p_i}{1-p_i} = \exp(\beta_0) \cdot \prod_{j=1}^k \exp(\beta_j)^{x_j} \quad (29)$$

Si se supone considerar dos elementos que tienen valores iguales en todas las variables menos en una. Sean $(x_{i1}, \dots, x_{ih}, \dots, x_{ik})$ los valores de las variables para el

26. Fórmula de Likelihood añadida con logaritmos.
 27. Fórmula de Likelihood expresadas en parámetros "P".
 28. Solución no derivable ante la presencia de parámetros "P".
 29. Fórmulas de los Odds Ratio o llamados también ratios de probabilidades.

primer elemento y $(x_{j_1}, \dots, x_{j_h}, \dots, x_{j_k})$ para el segundo, y todas las variables son las mismas en ambos elementos menos en la variable h donde $x_{ih} = x_{jh} + 1$. Entonces, el Odds ratio para estas dos observaciones es:

$$\frac{O_i}{O_j} = e^{\beta_h} \quad (30)$$

“e” indica cuanto se modifica el ratio de probabilidades cuando la variable x_j aumenta en una unidad. Si se considera $p_i = 0,5$ en el modelo logit, entonces.

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 \cdot x_{1i} + \dots + \beta_k \cdot x_{ki} = 0 \quad (31)$$

Es decir:

$$x_{1i} = -\frac{\beta_0}{\beta_1} - \sum_{j=2}^k \frac{\beta_j \cdot x_{ji}}{\beta_1} \quad (32)$$

4.4.3 Bondad de Ajuste: Contraste de hipótesis

Este tipo de medidas de bondad del ajuste se basa en contrastar la hipótesis nula H_0 de que el modelo seleccionado ajusta bien los datos por medio de un estadístico con una distribución conocida, Johnson (1998).

I. Desvianza (Deviance)

El estadístico desvianza (D) se define como una función del logaritmo neperiano del cociente de la función de verosimilitud del modelo seleccionado y la del modelo saturado. Un modelo saturado es aquel que contiene tantos

30. Odds Ratio, considerando dos elementos que tienen en valores iguales menos en una.

31. Formula del modelo logit, considerando una probabilidad $p_i = 0,5$.

32. x_{1i} representa el valor de x_1 que hace igualmente probable que un elemento cuyas restante variables son x_{2i}, \dots, x_{ki} , pertenezca a la primera o la segunda población.

parámetros como datos y que perfectamente los valores observados. La desviación tiene la siguiente expresión:

$$D = -2 \ln \left[\frac{\text{Verosimilitud del modelo seleccionado}}{\text{Verosimilitud del modelo saturado}} \right]$$

La cantidad entre corchetes se denomina razón de verosimilitud y el propio estadístico D es también llamado test o prueba de razón de verosimilitud. Este estadístico se distribuye aproximadamente como una X^2 con $N - p$ grados de libertad, donde N es el número de observaciones y p el número de parámetros contenidos en el modelo.

II. Prueba Chi-Cuadrado

Esta prueba se basa en la obtención de un estadístico X^2 que mide el nivel de discordancia que puede existir al comparar, para cada uno de los diferentes patrones de predictores existentes, el número de respuestas (afirmativas) observadas con la probabilidad estimada por el modelo. Por patrón de predictores se entiende cada una de las diferentes combinaciones de valores que pueden adoptar las variables independientes incluidas en el modelo. El estadístico X^2 , cuando el número de patrones predictores $M < N$, es:

$$X^2 = \sum_{i=1}^M \frac{m_i (y_i - \hat{p}_i)^2}{\hat{p}_i (1 - \hat{p}_i)} \quad (33)$$

Donde m_i es el número de casos incluidos en cada patrón de predictores, y_i la opción de la variable respuesta y \hat{p}_i la probabilidad estimada por el modelo para el patrón de covariables i . Para grandes muestras el estadístico se distribuye, obviamente, como una Chi-Cuadrado con $M - p$ grados de libertad.

33. Fórmula general de la prueba Chi-Cuadrado.

En presencia de variables continuas, el número de patrones de predictores es muy notable que sea igual al número de observaciones muestrales $M \approx N$. En estos casos la prueba X^2 tomaría la expresión:

$$X^2 = \sum_{i=1}^N \frac{(y_i - \hat{p}_i)^2}{\hat{p}_i(1 - \hat{p}_i)} \quad (34)$$

Puesto que n_i sería igual a 1. Hosmer y Lemeshow (1989) advierten de la obtención de valores p incorrectos cuando $M \approx N$; sin embargo, estos autores sugieren que, en los casos en que el modelo ajustado es el correcto, se puede utilizar la prueba X^2 con $N - p$ grados de libertad con unos resultados razonables.

III. Prueba de Hosmer-Lemeshow

Este estadístico se obtiene calculando el estadístico Chi-Cuadrado de Pearson, construyendo una tabla de contingencia a la que aplica un contraste tipo X^2 .

4.4.4 Análisis del modelo

Los residuos del modelo (residuos de Pearson), se definen como:

$$e_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}} \quad (35)$$

Y, si el modelo es correcto, serán variables de media cero y varianza unidad que pueden servir para hacer el análisis de dicho modelo. El estadístico $X_0^2 = \sum i^{e_i^2}$ permite realizar un contraste global de la bondad del ajuste. Se distribuye

34. Fórmula para la Prueba Chi-Cuadrado de variables continuas.

35. Fórmula general de los residuos de Pearson para análisis del modelo.

asintóticamente como una χ^2 con $(n - k - 1)$ grados de libertad, donde $k + 1$ es el número de parámetros en el modelo.

En lugar de los residuos de Pearson:

$$d_i = -2 \left(y_i \log \hat{p}_i + (1 - y_i) \log (1 - \hat{p}_i) \right) \quad (36)$$

36. Formula general de los residuos de Pearson para análisis del modelo.

CAPÍTULO 5: APLICACIÓN DE TECNICAS DE DATA MINING

5.1 Datos

Para el desarrollo de la investigación se han utilizado principalmente dos bases de datos. La primera de ellas cuenta con información de características propias de él o de los servicios contratados por el cliente, y la segunda con información demográfica de estos últimos. También se ha obtenido otras informaciones provenientes de otras bases de datos, las cuales se han conseguido con procesos de cruce de información con herramientas ofimáticas como Microsoft Access 2010 y Microsoft Excel 2010.

La información recopilada para el trabajo está comprendida entre Enero de 2009 hasta Junio de 2011, contando con un universo de 100.000 datos de información en conjunto para las plazas de Concepción y Temuco.

Cabe señalar que por motivo del terremoto ocurrido en Febrero del año 2010 en nuestro país, el periodo con información en las bases de datos comprendido entre Marzo a Junio del mismo año quedara fuera del estudio para la plaza de Concepción, ya que en estos meses la información existente en las bases está de alguna manera distorsionada, y no refleja la normalidad de los periodos como ocurre en los meses restantes que se han tomado como base para la investigación.

5.2 Análisis Cluster

5.2.1 Introducción

El Análisis Cluster o también llamado Análisis de Conglomerados, es una técnica estadística multivariante, cuyo objetivo es dividir un conjunto de datos observados en distintas categorías, permitiendo que los perfiles de los objetos pertenecientes a un mismo grupo sean lo más similares entre sí (cohesión interna del grupo) y los perfiles de los objetos de clusters diferentes sean distintos (aislamiento externo del grupo).

Análisis Cluster es un método de interdependencia de datos. A diferencia de métodos dependientes como estudios de regresión, regresión logística o análisis de correlación, que buscan estudiar cómo se relaciona una variable de tipo dependiente (y) con otras de carácter independientes (x_1, x_2, \dots, x_n), este método de forma sencilla logra agrupar datos con características similares.

Para la utilización de esta técnica, es importante definir las etapas necesarias para la aplicación.

- 1.- Selección de la muestra ó población de datos
- 2.- Selección y transformación de variables a utilizar
- 3.- Selección de concepto de distancia o similitud y medición de las mismas
- 4.- Selección y aplicación del criterio de agrupación
- 5.- Determinación de la estructura correcta

Figura N° 5.1: Etapas de Análisis Cluster

Fuente: Mahía, 2010, Pág. 13

5.2.2 Selección de datos y variables a utilizar

La selección de datos fue realizado en conjunto con analistas de la empresa, logrando observar y estudiar el amplio universo de información que se puede extraer desde las bases de datos. Con la observación de la información, se dio paso al análisis de esta misma, logrando comprender a grandes rasgos las características propias de las bases de datos y de cómo poder orientar el trabajo en base a la técnica de Análisis Cluster.

Es de importancia señalar que la selección de datos, se basó en características de tipo demográficas que han tenido los clientes fugados de la compañía, y de paso, las características que han tenido los servicios dados de baja.

Posterior a la selección de datos, se utilizó una fase existente en el proceso KDD (específicamente, fase de limpieza de datos) la cual permitió eliminar datos atípicos (edades no coherentes, grupos socioeconómicos sin información, deudas negativas, entre otras) logrando contar con una base de datos limpia y sin incoherencias. Luego de este proceso, se pudo establecer el universo de datos con que se trabajó en esta técnica, contabilizando 86.000.



Figura N° 5.2: Diagrama de bases de datos a utilizar en Análisis Cluster

Por otro parte, la selección de las variables es de suma importancia para el correcto desarrollo del trabajo, ya que si en esta etapa no se logra definir cuáles son las más influyentes para la investigación, es probable que todo el esfuerzo carezca de sentido o en un peor caso de una base estadística que sostenga la técnica empleada.

Por ello, en busca de lograr definir estas variables se pidió la opinión de expertos en esta área. Gracias a la ayuda de analistas de VTR (Rodrigo Velázquez y Nicolás Von Caprivi) se logró definir veinte variables que según ellos en una primera instancia son relevantes para esta técnica.

A continuación se detallan las variables a utilizar y una breve descripción de las mismas.

Tabla N° 5.1: Variables utilizadas en la investigación de Análisis Cluster

VARIABLE	DESCRIPCIÓN
EDAD	Edad del cliente
SEXO	Sexo del cliente
GSE	Grupo Socioeconómico del cliente
RECO	Reconexión del servicio
MOTDX	Motivos de la desconexión del servicio
DEUDA	Niveles de deuda asociadas al servicio
NUM_RECLA	Números de reclamos asociados al servicio
TIPO_RECLA	Tipos de reclamos asociados al servicio
ANTIGUE_SER	Antigüedad del servicio
PROMO_RETEN	Promociones de retención asociados al servicio
CAT_CLIENTE	Categoría asociada al tipo de cliente
TIPO_MULT	Tipos de multiservicios contratados
CANAL_ENTRADA	Canal entrada del servicio
CANTF001	Cantidad de servicios de Telefonía Contratados
CANT2WAY	Cantidad de servicios de Internet Contratados
CANTC001	Cantidad de servicios de TV Cable Contratados
PILETA	Cambios de servicios contratados luego de una desconexión
GTAR_PRO	Empaquetamiento de servicios contratados
GTAR_VALOR	Valoración del empaquetamiento de servicios contratados
CANAL_PAGO	Medios utilizados para el pago de servicios

5.2.3 Transformación de las variables

El término variable fue definido por D’Ary, Jacobs y Razavieh en 1982 como las características o atributos que permiten diferentes valores. Ejemplo de aquello es la estatura, temperatura, clima, etc. Existen diversas formas de clasificación de variables, pero su clasificación fue de acuerdo con el sujeto de estudio en cuestión y el uso de las mismas.

Para el desarrollo de Análisis Cluster las variables fueron clasificadas como categóricas. Estas últimas clasifican a los sujetos distribuyéndolos en grupos, de acuerdo a algún atributo previamente establecido, por ejemplo, grupo socioeconómicos, categoría del cliente, deudas, etc.

A su vez, estas variables se pueden subdividir en dos tipos: variables dicotómicas que poseen dos categorías, por ejemplo MASCULINO-FEMENINO (SEXO), y variables policotómicas que establecen tres o más categorías, como por ejemplo ABC1, C2, C3, D, E (GSE).

La utilización de este tipo de variables obedece exclusivamente a facilitar el manejo de información dentro de la investigación. A continuación se establece un cuadro resumen con respecto a las variables específicas utilizadas.

Tabla N° 5.2: Clasificación de variables dicotómicas y policotómicas en Análisis Cluster

TIPO DE VARIABLES	
DICOTÓMICAS	• SEXO
	• CAT_CLIENTE
POLICOTÓMICAS	• EDAD
	• GSE
	• RECO
	• MOTDX
	• DEUDA
	• NUM_RECLA
	• TIPO_RECLA
	• ANTIGUE_SER
	• PROMO_RETEN
	• TIPO_MULTI
	• CANAL_ENTRADA
	• CANTC001
• CANT2WAY	
• CANTF001	
• PILETA	
• GTAR_PRO	
• GTAR_VALOR	
• CANAL_PAGO	

Luego de establecer las variables que se utilizaron en Análisis Cluster, es necesario realizar un proceso de transformación de variables para manipular de mejor forma los datos, ya que esto logró categorizar los datos existentes y de paso, facilitó la interpretación de resultados.

Tabla N° 5.3: Variables y cambios de atributos utilizados en Análisis Cluster

VARIABLE	ATRIBUTO
EDAD	18 años <= 27 años -----> 1
	28 años <= 37 años -----> 2
	38 años <= 47 años -----> 3
	48 años <= 57 años -----> 4
	58 años <= 67 años -----> 5
	68 años o más -----> 6
SEXO	Femenino -----> 1
	Masculino -----> 2
GSE	ABC1 -----> 1
	C2 -----> 2
	C3 -----> 3
	D -----> 4
	E -----> 5
RECO	Si Reconexión -----> 1
	No Reconexión -----> 2
MOTDX	Morosidad -----> 1
	Cambio a la competencia -----> 2
	Económico -----> 3
	No utilización de servicios -----> 4
	Traslado de servicios -----> 5
	Fin de promociones -----> 6
	Otros motivos -----> 7

DEUDA	\$0 <= 29.999 -----> 1
	\$30.000 <= 59.999 -----> 2
	\$60.000 <= 89.999 -----> 3
	\$90.000 o más -----> 4
NUM_RECLA	Sin reclamos -----> 0
	1 reclamo -----> 1
	2 reclamos -----> 2
	3 o más -----> 3
TIPO_RECLA	Sin registro -----> 0
	Técnicos -----> 1
	Comerciales -----> 2
	Técnicos & Comerciales -----> 3
ANTIGUE_SER	1 mes <= 12 meses -----> 1
	13 meses <= 24 meses -----> 2
	25 meses <= 48 meses -----> 3
	49 meses <= 73 meses -----> 4
	74 meses o más -----> 5
PROMO_RETEN	2 meses Dcto. Económico -----> 1
	6 meses Dcto. Económico -----> 2
	12 meses Dcto. Económico -----> 3
	6 meses Prod. Gratis -----> 4
	12 meses Prod. Gratis -----> 5
	Sin promoción -----> 6
CAT_CLIENTE	Residencial -----> 1
	Convenios -----> 2

TIPO_MULT	CableFónico -----> 1
	CableInternet-----> 2
	FonoInternet -----> 3
	Sólo Cable -----> 4
	Sólo Fono -----> 5
	Sólo Internet -----> 6
	Triple Play -----> 7
CANAL_ENTRADA	Ceac In -----> 1
	Ceac Out -----> 2
	Dealer Pto. -----> 3
	Dealer Terreno -----> 4
	Sucursal -----> 5
	Terreno -----> 6
	Otros -----> 7
CANTC001	Sin registro -----> 0
	1 servicio -----> 1
	2 servicios o más -----> 2
CANT2WAY	Sin registro -----> 0
	1 servicio -----> 1
	2 servicios o más -----> 2
CANTF001	Sin registro -----> 0
	1 servicio -----> 1
	2 servicios o más -----> 2
PILETA	3P a Cable-Fono -----> 1
	3P a Cable-Internet -----> 2
	3P a Fono-Internet -----> 3
	CableFónico a Cable -----> 4
	CableFónico a Fono -----> 5
	CableInternet a Cable -----> 6
	CableInternet a Internet -----> 7
	FonoInternet a Fono -----> 8
	FonoInternet a Internet -----> 9
	Fin de pileta -----> 10

GTAR_PRO	Cable Full -----> 1
	Cable Hogar -----> 2
	Internet 2 MB -----> 3
	PT350 -----> 4
	PTI -----> 5
	Internet 1 MB -----> 6
	110 Min. Controlado -----> 7
	Internet 4 MB -----> 8
	Otros -----> 9
GTAR_VALOR	Alto -----> 1
	Medio -----> 2
	Bajo -----> 3
CANAL_PAGO	Agencias -----> 1
	Otros Canales -----> 2
	Canales Interno -----> 3
	Agencias - Otros Canales -----> 4
	Agencias - Canales Interno -----> 5
	Canales Interno - Otros Canales -----> 6
	Agencias - Canales Interno - Otros Canales -----> 7

Si bien la mayor parte de las variables no requiere una descripción detallada, hay algunas que se han definido que podrían causar preguntas sobre su significado.

De forma breve se explicarán cuatro variables las cuales se detallan a continuación.

- **PROMO_RETEN**, son promociones ofrecidas a clientes cuando estos últimos dan de baja de forma voluntaria los servicios. Son focalizados y pretenden como la descripción lo dice RETENER al cliente de una potencial fuga.
- **CANAL_ENTRADA**, corresponde al canal que se utilizó para contratar el servicio dado de baja. Aquí podemos encontrar diversos tipos, desde la contratación vía telefónica, sucursales o vía internet. Los canales más utilizados son los que se presentan en la Tabla N° 5.3, en donde se categorizan los siete canales de entrada más utilizados.

- **PILETA**, corresponde a los cambios ocurridos en el paquete de productos contratados post desconexión de un servicio. Por ejemplo, si se contrata un TRIPLE PACK (TV Cable, Telefonía e Internet), y luego se desconecta Internet, habrá un cambio de PILETA de 3P a Cable-Fono. De igual modo ocurre con las otras combinatorias de servicios que pueden existir.
- **CANAL_PAGO**, son los medios utilizados por los clientes para efectuar el pago de la deuda de servicios contratados. Comúnmente se clasifican en canales internos y canales externos. Los canales internos son propios de la compañía, y los externos están clasificados en agencias (Servipag y Sencillito) y otros (Tiendas de Retail, Bancos, Etc.).

Para el desarrollo de esta técnica de Data Mining, existen principalmente dos métodos de conglomeración que comúnmente se utilizan. A continuación se describe un cuadro resumen sobre las diferencias básicas y sus principales inconvenientes que posee cada una.

Tabla N° 5.4: Diferencias Básicas de cada uno de los métodos de conglomeración

Fuente: Ayuga, 2010, Pág. 18

Métodos Jerárquicos	Métodos No Jerárquicos
Comienza con las observaciones y no precisa determinar a priori el número de cluster.	Comienza con una partición inicial de conglomerados. A priori se determina el número y composición de los conglomerados
La asignación de objetos es definitiva.	El procedimiento es iterativo y permite la reasignación de objetos.
Operar con una matriz de similitudes.	Operan con la matriz de datos originales.

Tabla N° 5.5: Inconvenientes de cada uno de los métodos de conglomeración

Fuente: Ayuga, 2010, Pág. 19

Métodos Jerárquicos	Métodos No Jerárquicos
Si la estructura de la población es desconocida resulta difícil escoger el algoritmo.	Dificultad en conocer a priori el número real de los conglomerados existentes en la muestra
Es difícil operar e interpretar los gráficos con más de 200 datos.	Formar todas las particiones posibles para escoger la óptima es muy complejo
Mayor cantidad de atípicos en este punto.	Complejidad en los análisis.
Una mala partición no puede modificarse.	Una mala decisión inicial sobre el n° de composición de los grupos ocasiona una errónea clasificación.

5.2.4 Elección del número de conglomerados para cada plaza en estudio

Para el correcto cálculo del número de conglomerados que cada plaza necesita, fue necesario primero establecer el método de medida de distancia y de conglomeración.

Como la investigación se realizó en base a variables categóricas establecidas anteriormente, el método de medida de distancia elegido fue Chi-Cuadrado según la recomendación de los autores⁵ para este tipo de variables.

5. Rosana Ferrero. Master en Estadística Aplicada. Universidad de Granada, España.

Los métodos de conglomeración que fueron utilizados son los siguientes:

- a) **Vinculación fuera de grupos**
- b) **Vecino más lejano**
- c) **Agrupación de centroides**

Es de importancia señalar que existen múltiples métodos de conglomeración, pero puntualmente se eligieron estos tres que según los autores⁵ proporcionan buenos resultados al utilizar el método de medida de distancia mencionado anteriormente. Las formas y combinatorias en esta etapa pueden ser múltiples, pero aplicarlas todas carece de sentido, ya que en gral. los métodos llegan a resultados muy cercanos entre sí.

A continuación para cada método de conglomeración empleado se detalla una tabla resultante del análisis del software estadístico SPSS v19, en la cual aparecen las etapas de conglomeración (que siempre son igual $n-1$, donde n corresponde al número de variables definidas), y luego un historial de los coeficientes de conglomeración que miden la distancia entre las uniones de los cluster en cada etapa. Este valor inicialmente siempre es pequeño, aumentando de forma gradual hasta llegar a $n-1$ etapas. Por último dentro de la tabla, se encuentra las diferencias de coeficientes de conglomeración, la cual nos permite ver el salto más significativo desde una etapa a otra (comúnmente se le llama a esta técnica “método del codazo”).

El método del codazo (o “*The bigger Jump*” en inglés) es una técnica algebraica que permite determinar el número de cluster recomendado para el método empleado de forma fácil y rápida. Siguiere que se debe observar la mayor diferencia de coeficientes dentro de la tabla, luego posicionarse en la etapa anterior a la cual le corresponda esa diferencia de coeficientes y observar el número de la etapa. Este valor es el número de cluster (conglomerados) recomendados por el método.

5. Rosana Ferrero, Master en Estadística Aplicada, Universidad de Granada, España.

Por último, la información sobre los coeficientes de conglomeración se presenta de forma gráfica, en donde se puede ver con mayor claridad el salto que se origina en las distancias de coeficientes, lo que permite establecer con mayor claridad esta técnica utilizada.

- **Plaza Concepción**

- a) **Vinculación fuera de grupos**

Tabla N° 5.6: Método Vinculación fuera de grupos – Plaza Concepción

Etapa	Coeficientes	Dif. Coeficientes
1	37,49	
2	58,02	20,53
3	62,51	4,49
4	81,82	19,31
5	106,95	25,14
6	122,86	15,91
7	129,65	6,78
8	142,52	12,88
9	149,16	6,64
10	157,41	8,25
11	157,92	0,51
12	165,22	7,30
13	171,21	5,99
14	179,84	8,64
15	199,02	19,17
16	222,08	23,06
17	234,63	12,55
18	247,92	13,30
19	271,23	23,31

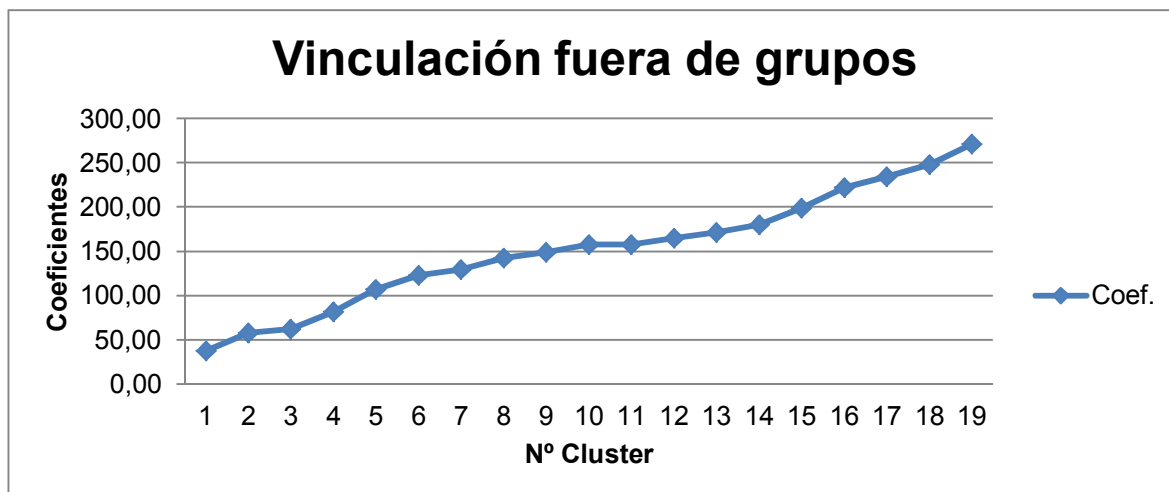


Gráfico N° 5.1: Método Vinculación fuera de grupos – Plaza Concepción

b) Vecino más lejano

Tabla N° 5.7: Método Vecino más Lejano – Plaza Concepción

Etapa	Coeficientes	Dif. Coeficientes
1	37,49	
2	61,37	23,88
3	62,51	1,14
4	89,75	27,24
5	121,48	31,72
6	137,97	16,49
7	145,95	7,97
8	156,78	10,83
9	161,61	4,83
10	170,21	8,60
11	172,65	2,44
12	186,62	13,97
13	207,17	20,55
14	213,74	6,56
15	243,67	29,94
16	274,51	30,84
17	293,56	19,05
18	295,86	2,29
19	318,77	22,91

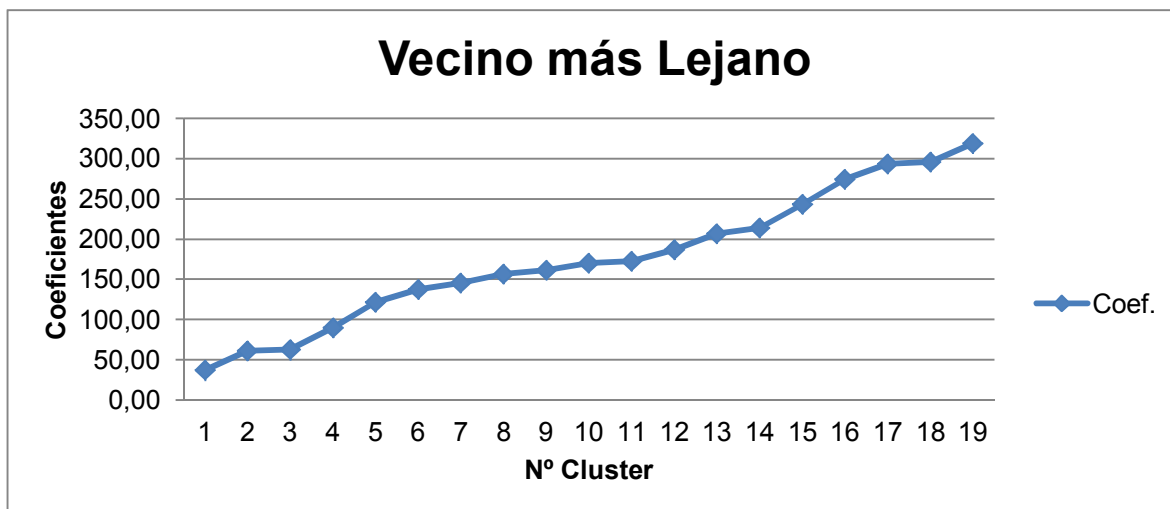


Gráfico N° 5.2: Método Vecino más lejano – Plaza Concepción

c) Agrupación de centroides

Tabla N° 5.8: Método Agrupación de centroides – Plaza Concepción

Etapa	Coeficientes	Dif. Coeficientes
1	37,49	
2	48,65	11,16
3	62,51	13,87
4	67,76	5,25
5	88,45	20,70
6	89,43	0,97
7	90,36	0,94
8	97,05	6,69
9	98,76	1,71
10	102,85	4,09
11	106,56	3,71
12	107,94	1,38
13	108,09	0,16
14	113,06	4,97
15	129,50	16,44
16	149,14	19,64
17	153,25	4,11
18	154,11	0,86
19	171,62	17,51

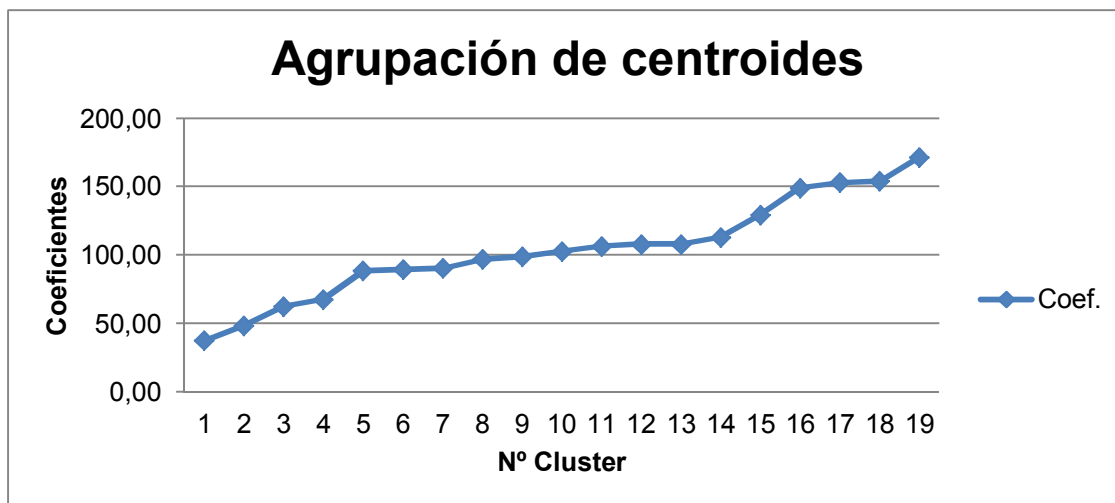


Gráfico Nº 5.3: Método Agrupación de centroides – Plaza Concepción

- Plaza Temuco

a) Vinculación fuera de grupos

Tabla N° 5.9: Método Vinculación fuera de grupos – Plaza Temuco

Etapa	Coefficientes	Dif. Coeficientes
1	12,12	
2	20,74	8,61
3	31,04	10,31
4	55,21	24,16
5	68,26	13,05
6	73,87	5,62
7	75,58	1,71
8	80,74	5,16
9	84,49	3,75
10	92,77	8,28
11	95,45	2,68
12	97,92	2,47
13	98,89	0,97
14	103,86	4,97
15	108,90	5,03
16	124,03	15,13
17	138,59	14,56
18	140,93	2,33
19	152,28	11,35

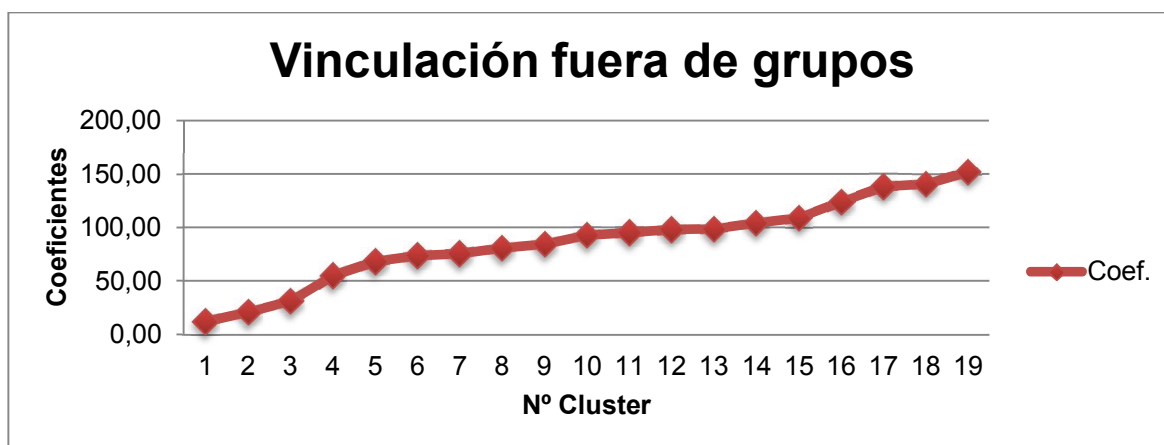


Gráfico N° 5.4: Método Vinculación fuera de grupos – Plaza Temuco

b) Vecino más alejado

Tabla N° 5.10: Método más alejado – Plaza Temuco

Etapa	Coefficientes	Dif. Distancias
1	12,12	
2	26,70	14,57
3	45,95	19,26
4	62,21	16,26
5	85,71	23,50
6	86,35	0,64
7	90,97	4,63
8	95,45	4,48
9	98,91	3,46
10	99,18	0,27
11	102,54	3,36
12	108,22	5,68
13	118,24	10,02
14	124,65	6,41
15	138,88	14,23
16	159,29	20,41
17	171,08	11,78
18	180,64	9,57
19	200,85	20,21

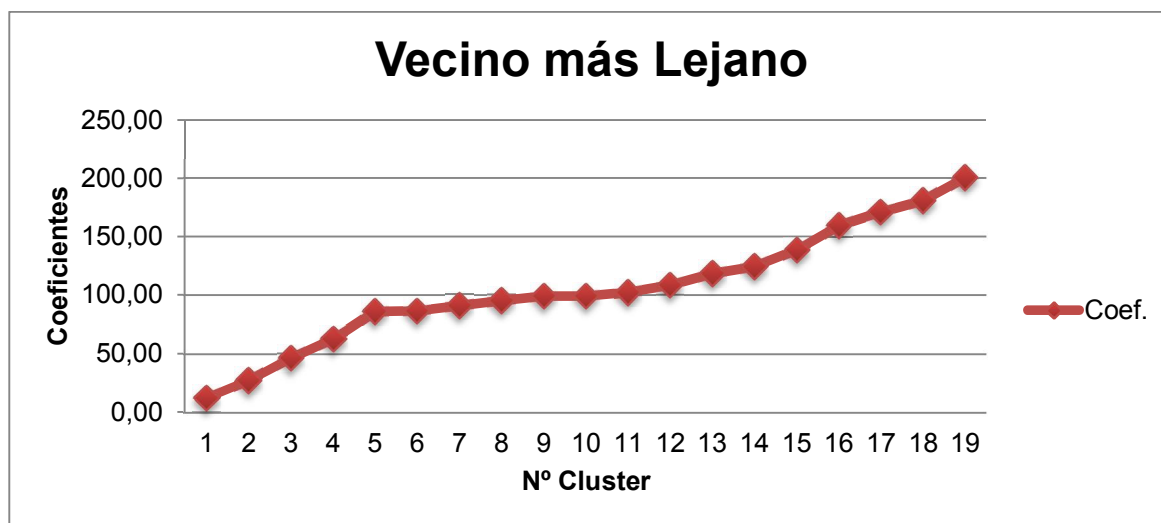


Gráfico N° 5.5: Método Vecino más lejano – Plaza Temuco

c) Agrupación de centroides

Tabla N° 5.11: Método Agrupación de centroides – Plaza Temuco

Eta	Coeficientes	Dif. Distancias
1	12,12	
2	27,71	15,58
3	38,65	10,94
4	52,84	14,19
5	53,08	0,25
6	54,88	1,80
7	56,34	1,46
8	60,04	3,70
9	60,74	0,69
10	62,03	1,30
11	62,46	0,43
12	63,18	0,71
13	64,12	0,95
14	65,21	1,08
15	67,74	2,54
16	78,36	10,62
17	89,20	10,84
18	91,23	2,03
19	100,65	9,42

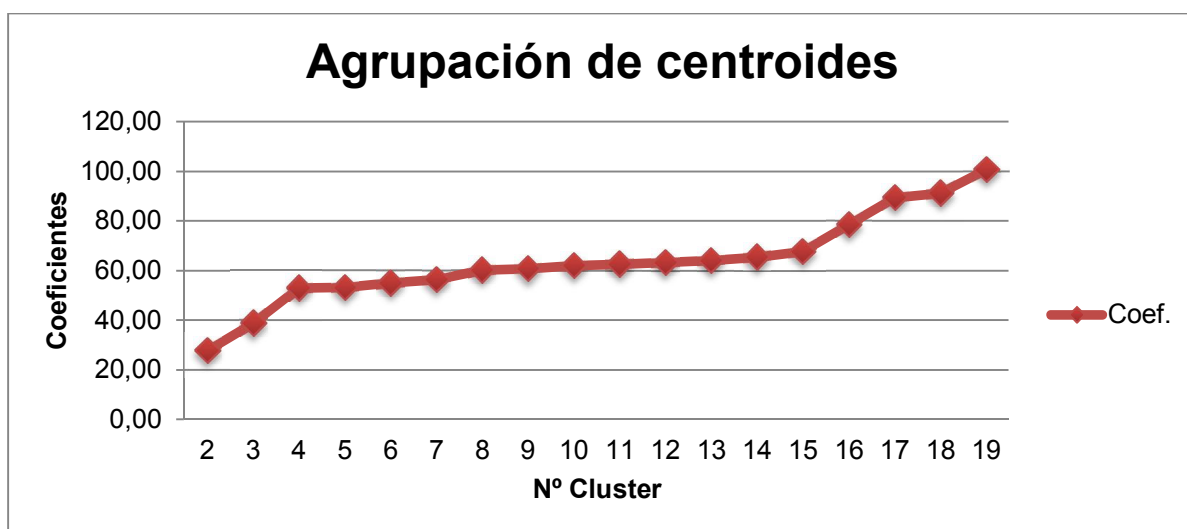


Gráfico N° 5.6: Método Agrupación de centroides – Plaza Temuco

- **Resumen de resultados**

Para la plaza de Concepción, las tres técnicas empleadas indicaron que el número de conglomerados a utilizar para esta fue de cuatro. Por otra parte, para la plaza de Temuco, son dos técnicas empleadas que indicaron que el número de conglomerados a utilizar fue de tres, y la otra restante que fue de cuatro. Por simple mayoría (dos sobre tres técnicas) el número de conglomerados a utilizar para esta plaza fue de tres.

Si bien ningún método empleado es mejor que otro, es una buena base el poder comparar las distintas técnicas permitiendo establecer un rango de soluciones al problema.

Conociendo el número de conglomerados recomendados a utilizar, se aplicó el método K-Medias (Método No Jerárquico) para interpretar los resultados del Análisis Cluster para cada plaza de estudio.

5.2.5 Aplicación del Método No Jerárquico (K-Medias)

El Análisis Cluster de K-Medias es una herramienta específica que permite asignar n objetos a un número fijo de k conglomerados ($k > 1$) cuyas características no se conocen con certeza, pero que se basa en un conjunto de variables específicas. Es de utilidad cuando se trabaja con un gran conjunto de datos (miles).

El cálculo del número fijo de k conglomerados para cada plaza, fue visto y estudiado anteriormente, estableciéndose dichos valores.

El primer paso del Análisis Cluster de K-Medias es construir centros de conglomerados iniciales, posterior a esto se inicia la asignación de casos basándose en la distancia de los centros, y luego, actualiza las posiciones de los centros de los conglomerados basándose en los valores medios de los casos. Todo este trabajo se realiza por medio de sucesivas iteraciones, hasta llegar al resultado óptimo.

Se forma simple se observa cómo funciona la técnica de K-Medias. Los múltiples puntos que representan datos son re-ordenados con respecto a las distancias de los centros de conglomerados que se originan luego de las sucesivas iteraciones, permitiendo al final conseguir conglomerados finales distintos entre sí.

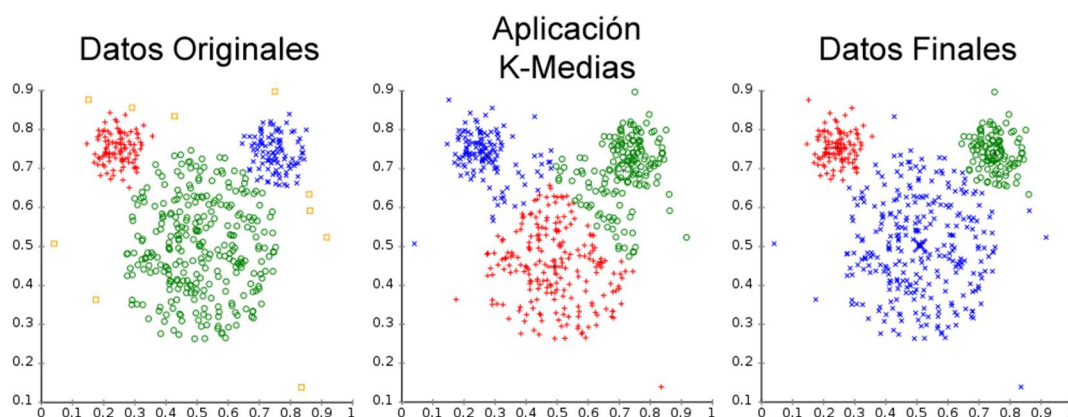


Figura N° 5.3: Herramienta Análisis Cluster de K-Medias

La aplicación de la técnica de Análisis Cluster de K-Medias fue realizada en el software estadístico SPSS v19.

Por último, es de importancia señalar que para cada plaza se detalla a continuación el cluster más representativo, dejando los resultados de los restantes en los **ANEXOS A, B, C, D, E y F** de la investigación.

- **Plaza Concepción**

El total de desconexiones estudiadas fue de 63.850 para la plaza de Concepción. Del resultado de la aplicación de Análisis Cluster de K-Medias se establece la siguiente conformación de conglomerados.

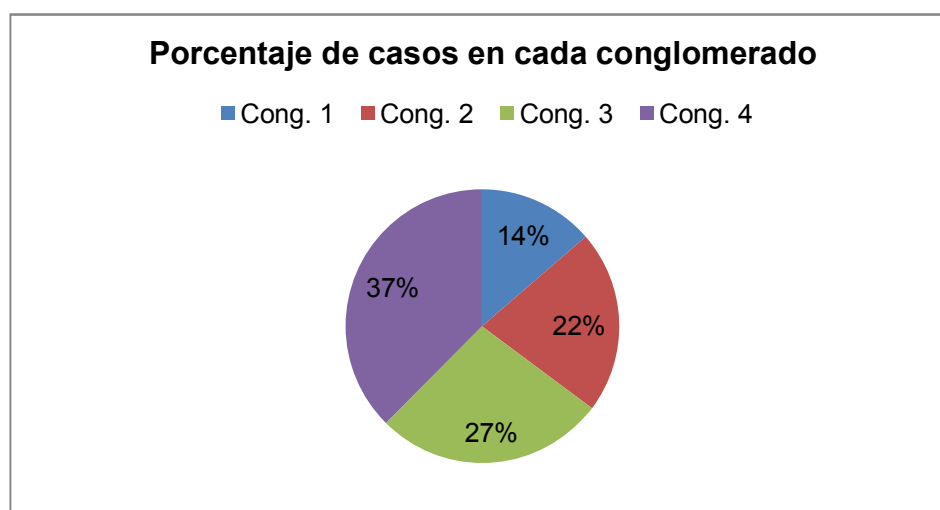


Gráfico N° 5.7: Porcentaje de casos en cada conglomerado – Plaza Concepción

La conformación de los centros iniciales del conglomerado más representativo mediante la herramienta estadística se detalla en la Tabla N° 5.12, en esta primera instancia hay una ordenación inicial de datos que corresponde justo antes de iniciar la primera iteración.

Los valores que aparecen tanto en la Tabla N° 5.12 y N° 5.13, corresponden a los valores de las medias de cada variable.

Tabla N° 5.12: Centros iniciales del conglomerado cuatro – Plaza Concepción

Centros iniciales del conglomerado	
	Conglomerado
	4
MOTDX	7
DEUDA	1
EDAD	1
SEXO	1
GSE	2
NUM_RECLA	0
TIPO_RECLA	0
ANTIGUE_SER	1
CANAL_ENTRADA	1
CANAL_PAGO	1
CANTC001	1
CANTF001	0
CANT2WAY	0
PILETA	10
GTAR_PRO	1
GTAT_VALOR	1
PROMO_RETEN	6
CAT_CLIENTE	1
TIPO_MULT	4
RECO	2

Posteriormente, se inició el proceso de iteración logrando de forma sucesiva cambiar los centros iniciales del conglomerado cuatro. El número máximo de iteraciones utilizado es de diez, ya que luego de este valor generalmente no se logran cambios significativos en los centros.

Luego de las iteraciones se logró la conformación de los centros finales del conglomerado cuatro. De esta tabla, se pueden apreciar las características de composición que reúne el cluster.

Tabla N° 5.13: Centros finales del conglomerado cuatro – Plaza Concepción

Centros finales del conglomerado	
	Conglomerado
	4
MOTDX	3
DEUDA	2
EDAD	3
SEXO	2
GSE	3
NUM_RECLA	0
TIPO_RECLA	0
ANTIGUE_SER	2
CANAL_ENTRADA	5
CANAL_PAGO	4
CANTC001	1
CANTF001	0
CANT2WAY	0
PILETA	10
GTAR_PRO	2
GTAT_VALOR	2
PROMO_RETEN	6
CAT_CLIENTE	1
TIPO_MULT	4
RECO	2

En la Tabla N° 5.14 se observan los cambios en los centros del conglomerado cuatro a medida que se generan las iteraciones sucesivas por medio del software estadístico utilizado. Se puede apreciar que en las primeras iteraciones los cambios fueron significativos, pero a medida que aumentan las iteraciones los cambios van siendo de menor significancia.

Tabla N° 5.14: Historial de iteraciones para cambios en los centros del conglomerado cuatro – Plaza Concepción

Historial de iteraciones	
Iteración	4
1	6,677
2	,857
3	,412
4	,334
5	,088
6	,030
7	,014
8	,008
9	,016
10	,020

En la Tabla N° 5.15 se observa la distancia euclídea entre el centro del conglomerado cuatro y los restantes formados. Se establece que la mayor distancia se genera con el conglomerado uno, y que la menor ocurre con el conglomerado dos.

Tabla N° 5.15: Distancia entre el centro del conglomerado cuatro y los restantes formados – Plaza Concepción

Distancia euclídea	
Conglomerado	4
1	9,562
2	4,802
3	7,122
4	0

- **Plaza Temuco**

El total de desconexiones estudiadas fue de 22.383 para la plaza de Temuco. Del resultado de la aplicación de Análisis Cluster de K-Medias se establece la siguiente conformación de conglomerados.

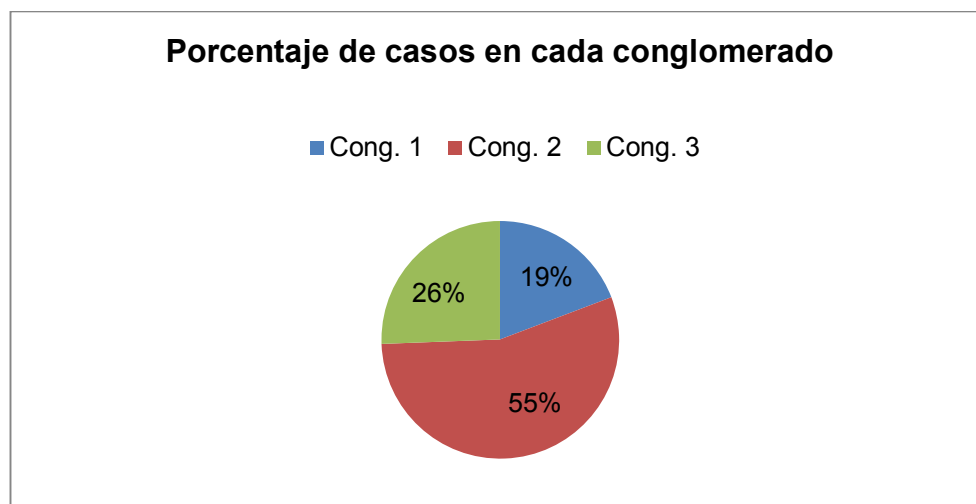


Gráfico N° 5.8: Porcentaje de casos en cada conglomerado – Plaza Temuco

De igual modo como sucede en la plaza de Concepción, la primera etapa fue la conformación de los centros iniciales del conglomerado más representativo mediante la herramienta estadística. Dicha información se detalla en la Tabla N° 5.16.

Tabla N° 5.16: Centros iniciales del conglomerado dos – Plaza Temuco

Centros iniciales del conglomerado	
	Conglomerado
	2
MOTDX	7
DEUDA	1
EDAD	2
SEXO	1
GSE	2
NUM_RECLA	3
TIPO_RECLA	2
ANTIGUE_SER	3
CANAL_ENTRADA	1
CANAL_PAGO	7
CANTC001	1
CANTF001	0
CANT2WAY	0
PILETA	10
GTAR_PRO	1
GTAT_VALOR	1
PROMO_RETEN	6
CAT_CLIENTE	1
TIPO_MULTA	4
RECO	2

Posteriormente, se inició el proceso de iteración logrando de forma sucesiva cambiar los centros de este conglomerado más representativo en estudio.

Luego de las iteraciones, se logró la conformación de los centros finales del conglomerado dos, información que se detalla en la Tabla N° 5.17. De igual modo como se observó en la plaza de Concepción, se pueden apreciar las características de composición que reúne este cluster.

Tabla N° 5.17: Centros finales del conglomerado dos – Plaza Temuco

Centros finales del conglomerado	
	Conglomerado
	2
MOTDX	2
DEUDA	2
EDAD	3
SEXO	2
GSE	2
NUM_RECLA	1
TIPO_RECLA	1
ANTIGUE_SER	2
CANAL_ENTRADA	5
CANAL_PAGO	4
CANTC001	0
CANTF001	1
CANT2WAY	1
PILETA	9
GTAR_PRO	3
GTAT_VALOR	2
PROMO_RETEN	6
CAT_CLIENTE	1
TIPO_MULT	3
RECO	2

En la Tabla N° 5.18 se observan los cambios en los centros finales del conglomerado a medida que se generan las iteraciones sucesivas por medio del software estadístico. Se puede apreciar que desde la octava iteración ya no existe ningún cambio en los centros de conglomerados para el cluster formado.

Tabla N° 5.18: Historial de iteraciones para cambios en los centros del conglomerado dos – Plaza Temuco

Historial de iteraciones	
Iteración	2
1	6,603
2	1,443
3	,502
4	,396
5	,075
6	,011
7	,002
8	,000

En la Tabla N° 5.19 se observa la distancia euclídea entre los centros del conglomerado dos y los restantes formados. Se establece que la mayor distancia se genera con el conglomerado uno, y que la menor ocurre con el conglomerado tres.

Tabla N° 5.19: Distancia entre el centro del conglomerado dos y los restantes formados – Plaza Temuco

Distancia euclídea	
Conglomerado	2
1	7,735
2	0
3	7,543

- **Análisis de resultados**
- **Plaza Concepción**

El conglomerado más representativo para la plaza de Concepción es el número cuatro, con una representación de un **37%** del total de desconexiones analizadas.

➤ **Cluster Cuatro**

El tipo de cliente que se presenta en este cluster es de tipo residencial, la edad del mismo fluctúa entre los 38 y 47 años de edad, con sexo de tipo masculino y grupo socioeconómico C3. El motivo de la desconexión se debe a problemas de carácter económico, con deudas que fluctúan entre \$30.000 y \$59.999. Este conglomerado no presenta ningún tipo de reclamo acumulado hasta la fecha de desconexión.

La antigüedad de él o los servicios al momento de la desconexión fluctúa entre los 13 y los 24 meses de contratación del mismo. El canal de entrada utilizado para la contratación del servicio es por medio de sucursales propias de la empresa, y los canales de pago utilizados en este conglomerado son medios externos como agencias (Servipag y Sencilito) y otros canales, como pago por medio de bancos o utilizando convenios con empresas del retail.

Solo se presenta el servicio contratado de TV Cable, y no de servicios de Telefonía e Internet, de esto último, luego de la desconexión se pone fin a la piqueta de productos. El empaquetamiento de producto en este conglomerado es Cable Hogar, que tiene un valor de categorización de nivel medio para la empresa. Este conglomerado no presenta promociones de retención de ningún tipo al momento de la desconexión, y luego de la desconexión no existe intención por parte del cliente en volver a reconectar el servicio dado de baja.

- **Plaza Temuco**

El conglomerado más representativo para la plaza de Temuco es el número dos, con una representación de un **55%** del total de desconexiones analizadas.

- **Cluster Dos**

El tipo de cliente que se presenta en este cluster es de tipo residencial, la edad del mismo fluctúa entre los 38 y 47 años de edad, con sexo de tipo masculino y grupo socioeconómico C2. El motivo de la desconexión se debe a cambios a la competencia, con deudas que fluctúan entre \$30.000 y \$59.999. Este conglomerado presenta un reclamo acumulado hasta la fecha de desconexión, y su carácter es de tipo técnico.

La antigüedad de él o los servicios al momento de la desconexión fluctúa entre los 13 y los 24 meses de contratación del mismo. El canal de entrada utilizado para la contratación del servicio es por medio de sucursales propias de la empresa, y los canales de pago utilizados en este conglomerado son medios externos como agencias (Servipag y Sencilito) y otros canales, como pago por medio de bancos o utilizando convenios con empresas del retail.

Los clientes poseen un Tipo Multiservicio de Internet y Telefonía, y la desconexión del servicio para este cluster es Telefonía, pasando luego de la desconexión a un paquete de un servicio (Internet). El empaquetamiento de producto en este conglomerado es Internet 2MB, que tiene un valor de categorización de nivel medio para la empresa. Este conglomerado no presenta promociones de retención de ningún tipo al momento de la desconexión, y luego de la desconexión no existe intención por parte del cliente en volver a reconectar el servicio dado de baja.

5.3 Regresión Logística Multivariante

5.3.1 Introducción

La regresión logística multivariante se define por la existencia de una variable dependiente (y) con dos estados y múltiples variables de tipo independiente (x_1, x_2, \dots, x_n) que pueden ser de carácter categóricas o cuantitativas. El principal objetivo de esta técnica de Data Mining es obtener una estimación no sesgada o ajustada de la relación entre la variable dependiente (o resultado) y múltiples variables independientes que se estime importantes para el estudio.

Para el trabajo desarrollado, la variable dependiente dicotómica se define como sigue. Su primer estado es de tipo ACTIVO, y el segundo estado es FUGA. Se define ACTIVO como el estado de no desconexión de servicios de TV Cable, Internet y Telefonía, y como FUGA a la desconexión propiamente tal.

$$y = \begin{cases} 0, & \text{servicio con estado ACTIVO} \\ 1, & \text{servicio con estado FUGA} \end{cases}$$

Los pasos que se han utilizado para la elaboración de esta parte del capítulo, son definidos en la sgte. figura. Cabe señalar que los detalles y características de esta técnica de Data Mining fueron detallados en capítulos anteriores.

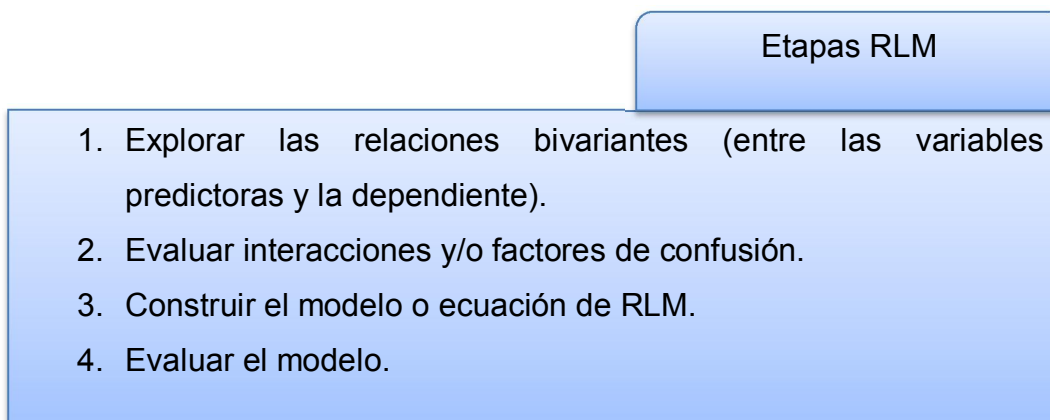


Figura Nº 5.4: Etapas de Regresión Logística Multivariante

Fuente: Aguayo, 2010, Pág. 1

5.3.2 Selección de Datos y variables a utilizar

La selección de datos necesarios para el desarrollo de RLM corresponde a información sobre características demográficas de clientes fugados y actuales que ha tenido la compañía, y también, de características de servicios fugados y actuales asociados a estos mismos clientes. El periodo de tiempo que se ha elegido para la recolección de datos, corresponde a Enero de 2009 a Junio de 2011.

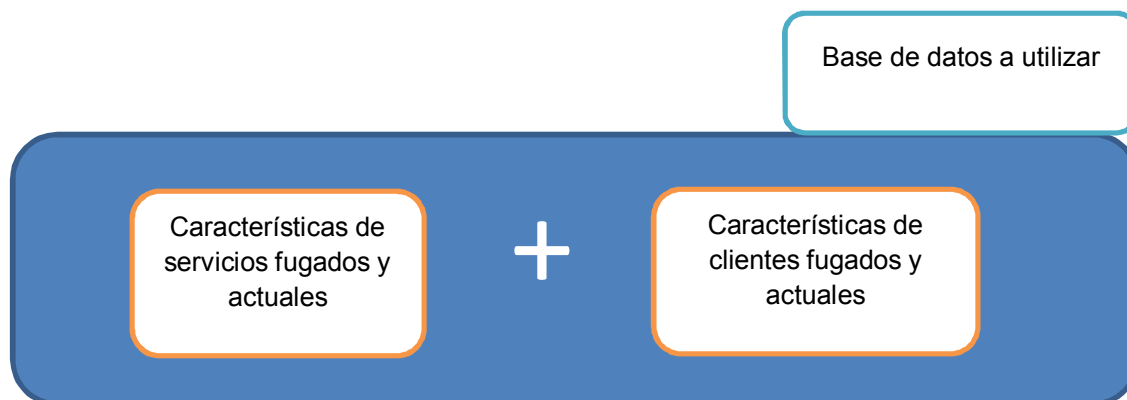


Figura N° 5.5: Diagrama sobre datos a utilizar en RLM

Posterior a la selección de datos, se utilizó una fase existente en el proceso KDD llamada limpieza de datos. Esta fase permitió eliminar datos atípicos (edades no coherentes, grupos socioeconómicos sin información, deudas negativas, entre otras), logrando al final de la limpieza, una base de datos sin incoherencias que podrían de algún modo distorsionar el análisis que se quiere lograr. Luego de este proceso, se pudo establecer el universo de datos con que se trabajó en esta técnica, contabilizando 35.000.

Luego de la aplicación de la fase perteneciente a KDD, es necesario definir qué variables a priori podrían ser influyentes para la construcción del modelo. Es por ello, que con ayuda de analistas de VTR Globalcom S.A. fueron elegidas nueve variables independientes que en una primera instancia podrían ser significativas para el estudio.

A continuación se detallan las variables independientes utilizadas en RLM y una breve descripción de las mismas.

Tabla N° 5.20: Variables utilizadas en RLM

VARIABLE	DESCRIPCIÓN
GSE	Grupo socioeconómico del cliente
DEUDA	Valor de la deuda actual
PERMA_MESES	Permanencia en meses de serv. contratados
EDAD_C	Edad del cliente
SEXO	Sexo del cliente
PROMO_ASOC	Promociones asociadas a serv. contratados
RECLAM	Reclamos asociados a serv. contratados
CPAGO	Canales de pago utilizados
C_ENTRADA	Canal de entrada a la compañía

No se ampliara mayormente en la descripción y detalles específicos de algunas variables dado que la gran mayoría no lo requiere. Pero existen algunas que podrían causar preguntas o dudas sobre su significado, y que a continuación se detallan.

- **DEUDA**, corresponde al valor actualizado de la deuda para los servicios activos y el último registro de los servicios fugados disponibles en la base de datos. La actualización de los valores se hizo con fecha de junio de 2011.
- **PERMA_MESES**, tiempo de permanencia en meses de él o los servicios contratados por el cliente.
- **PROMO_ASOC**, existencia o no de promociones ofrecidas al cliente para prolongar su tiempo en la compañía. Pudiendo ser descuentos económicos en servicios o gratuidad de estos últimos.
- **RECLAM**, existencia o no de reclamos asociados a los servicios contratados por el cliente.

- **CPAGO**, canales de pago utilizados por el cliente para la cancelación de servicios contratados. Puede ser de tipo interno o externo.
- **C_ENTRADA**, canal de entrada de él o los servicios contratados. Se optó por colapsar las categorías policotómicas a dicotómicas logrando dos tipos de clasificación. La primera de ellas con nombre “Push” (Dealer Terreno + Terreno) y la segunda “Pull” (Dealer Pto. + Sucursal).

5.3.3 Transformación de las variables

De igual forma como se hizo con Análisis Cluster, las variables para el desarrollo de RLM se clasificaron en categóricas. La utilización de este tipo de clasificación obedece exclusivamente a facilitar el manejo de información dentro de la investigación. A continuación se establece la forma de categorización empleado en esta sección del capítulo.

Es de importancia señalar que muchas de las variables utilizadas en RLM originalmente son de tipo policotómicas, las cuales fueron empleadas en Análisis Cluster. Para el caso de RLM es recomendable colapsar o agrupar categorías para transformarlas en dicotómicas, ya que de esta forma el trabajo resulta más simple y con mejor pronóstico en el modelo.

Tabla N° 5.21: Variables y cambios de atributos utilizados en RLM

VARIABLE	ATRIBUTO
GSE	ABC1 + C2 -----> 0
	C3 + D + E -----> 1
DEUDA	\$0 - \$14.999 -----> 0
	\$15.000 o más -----> 1
PERMA_MESES	0 a 25 meses -----> 0
	26 meses o más -----> 1

EDAD_C	Mayor a 40 años -----> 0
	Menor o igual a 40 años -----> 1
SEXO	Masculino -----> 0
	Femenino -----> 1
PROMO_ASOC	Existencia de promoción -----> 0
	No Existencia de promoción -----> 1
RECLAM	Existencia de reclamos -----> 0
	No Existencia de reclamos -----> 1
CPAGO	C. Externo -----> 0
	C. Interno -----> 1
C_ENTRADA	Canales Push -----> 0
	Canales Pull -----> 1

5.3.4 Análisis univariado de las variables

Posterior a la transformación de las variables, fue necesario realizar un análisis univariado de cada una de ellas para poder cuantificar la importancia o no de incluirlas posteriormente en el modelo.

Para ello fue necesario realizar una comparación de distribución de cada una de las variables independientes en función de la variable dependiente ESTAD_SERV (ACTIVO / FUGA).

Se establece como importante cualquiera que presente una diferencia absoluta igual o mayor a 5% en la comparación de ambas distribuciones. Si hay una diferencia en valor absoluto menor a 5%, se establece que esa variable no es significativa para ser incluida en el modelo.

• **Plaza Concepción**

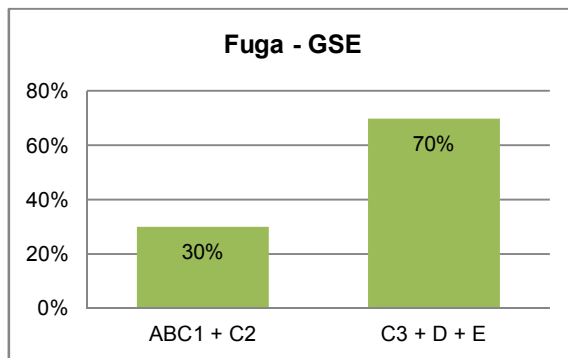
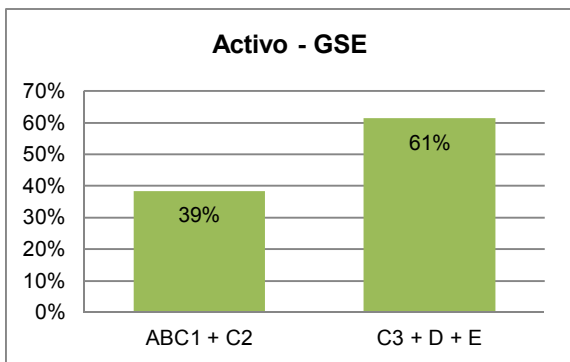


Gráfico N° 5.9: Distribución por GSE – Estado Activo y Fuga – Plaza Concepción

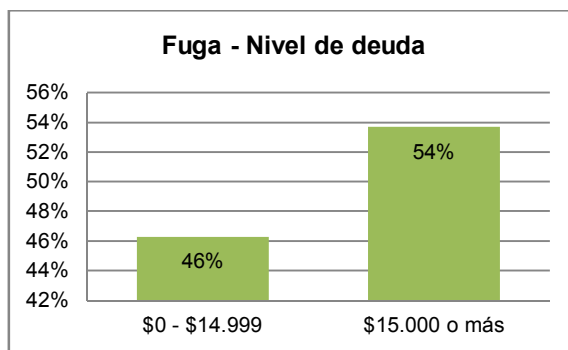
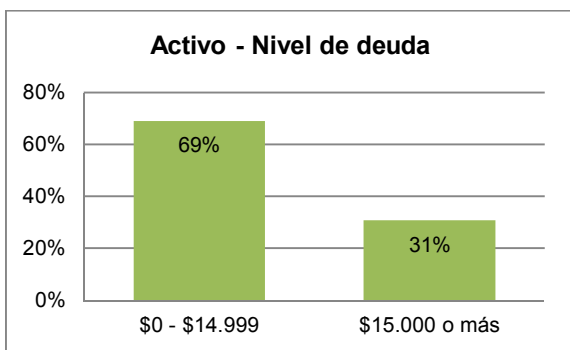


Gráfico N° 5.10: Distribución por Nivel de deuda – Estado Activo y Fuga – Plaza Concepción

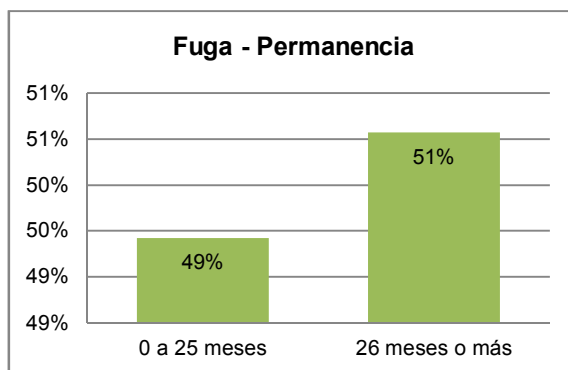
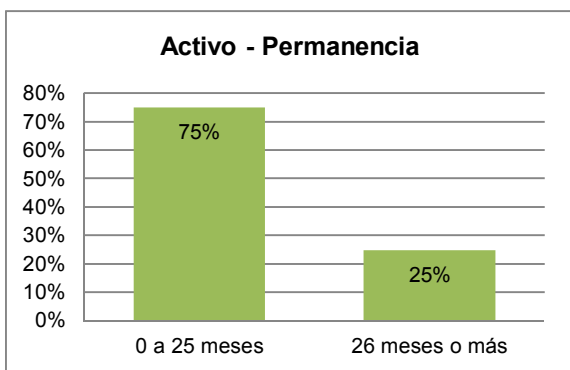


Gráfico N° 5.11: Distribución por Permanencia – Estado Activo y Fuga – Plaza Concepción

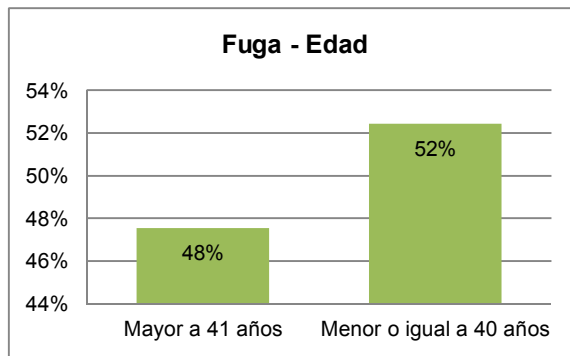
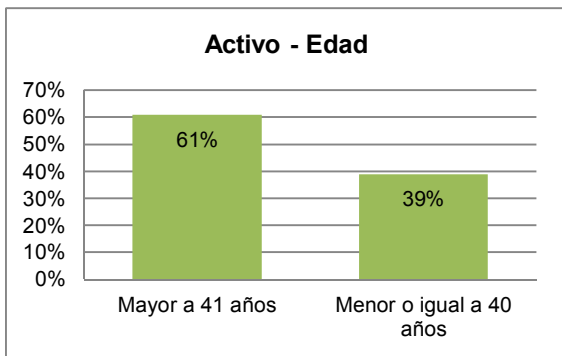


Gráfico N° 5.12: Distribución por Edad – Estado Activo y Fuga – Plaza Concepción

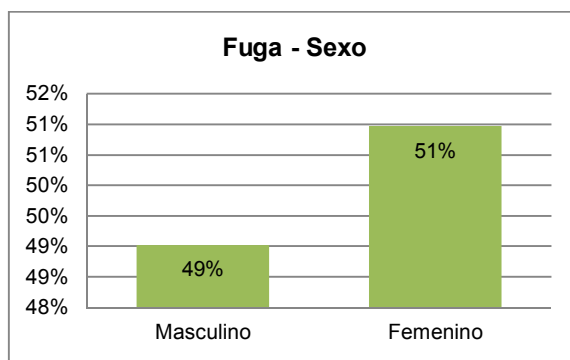
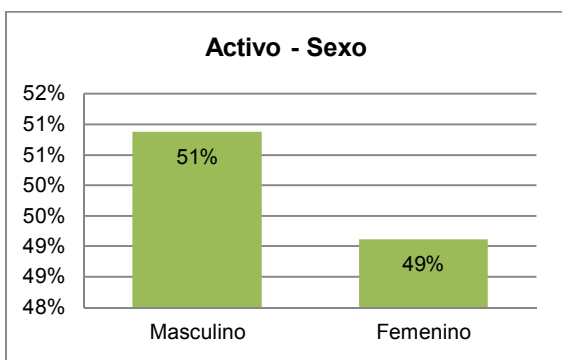


Gráfico N° 5.13: Distribución por Sexo – Estado Activo y Fuga – Plaza Concepción

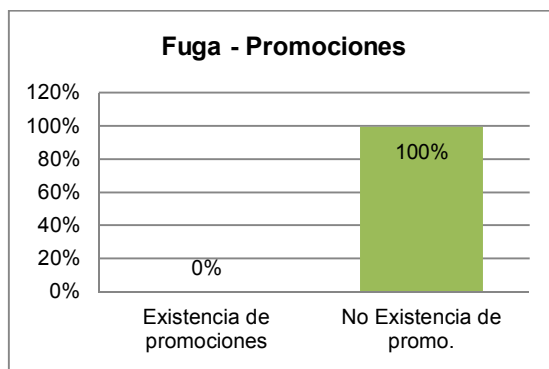


Gráfico N° 5.14: Distribución por Promociones – Estado Activo y Fuga – Plaza Concepción

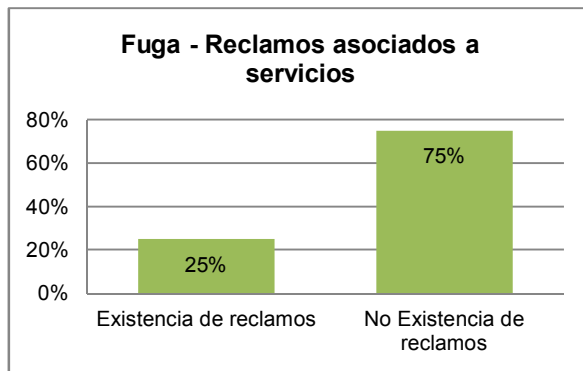


Gráfico N° 5.15: Distribución por Reclamos asociados a servicios – Estado Activo y Fuga – Plaza Concepción

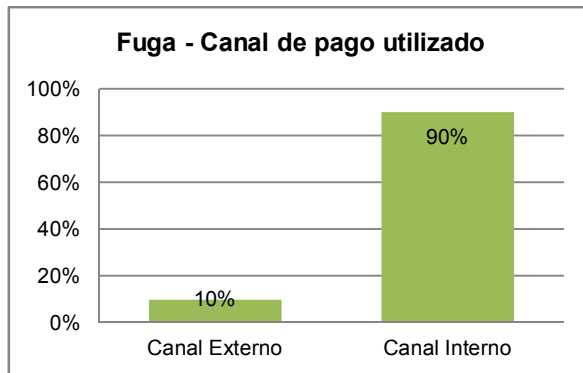
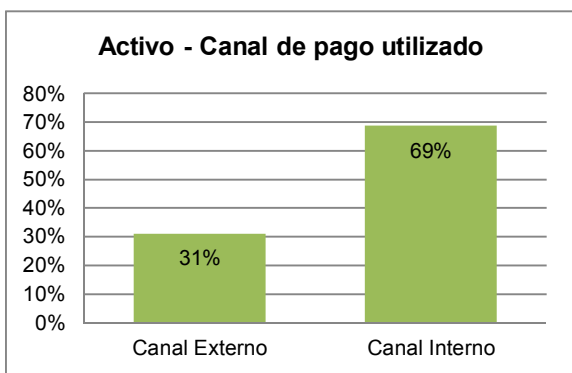


Gráfico N° 5.16: Distribución por Canal de pago utilizado – Estado Activo y Fuga – Plaza Concepción

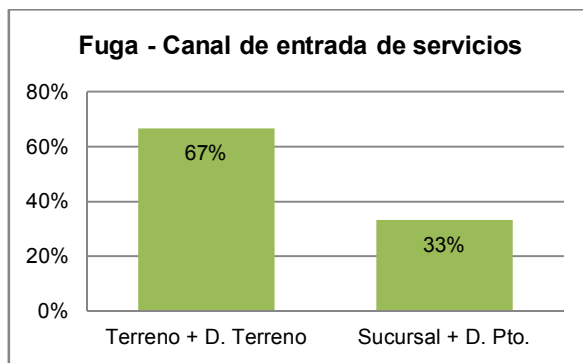
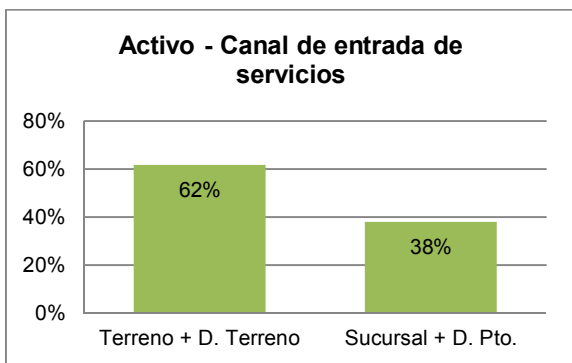


Gráfico N° 5.17: Distribución por Canal de entrada de servicios – Estado Activo y Fuga – Plaza Concepción

• **Plaza Temuco**

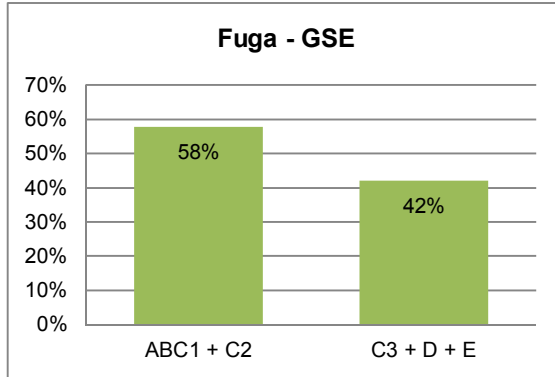
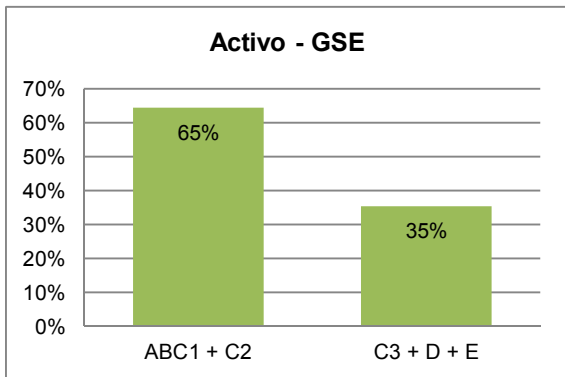


Gráfico N° 5.18: Distribución por GSE – Estado Activo y Fuga – Plaza Temuco

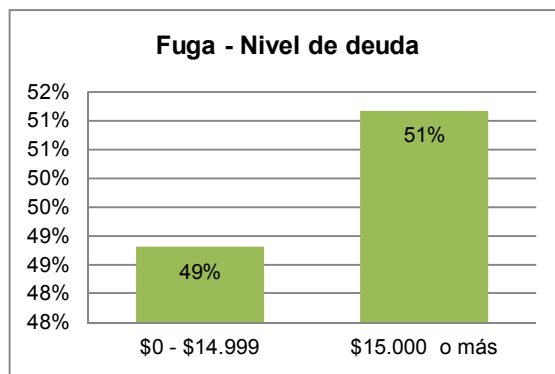
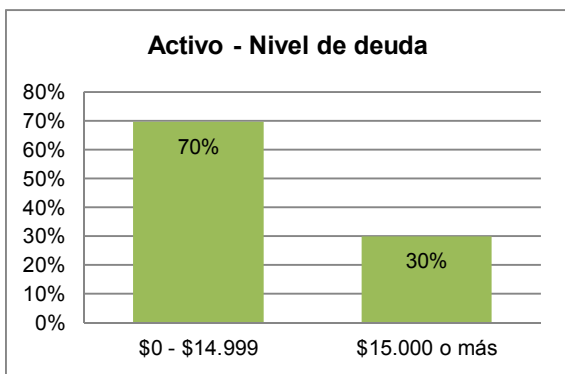


Gráfico N° 5.19: Distribución por Nivel de deuda – Estado Activo y Fuga – Plaza Temuco

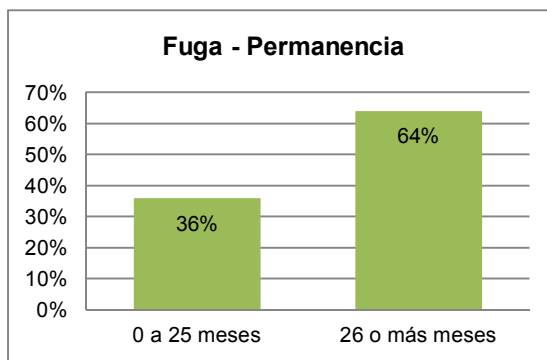
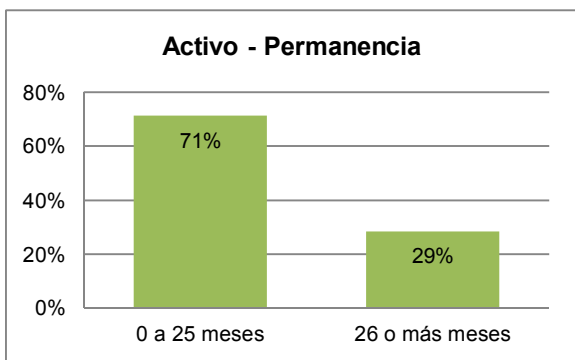


Gráfico N° 5.20: Distribución por Permanencia – Estado Activo y Fuga – Plaza Temuco

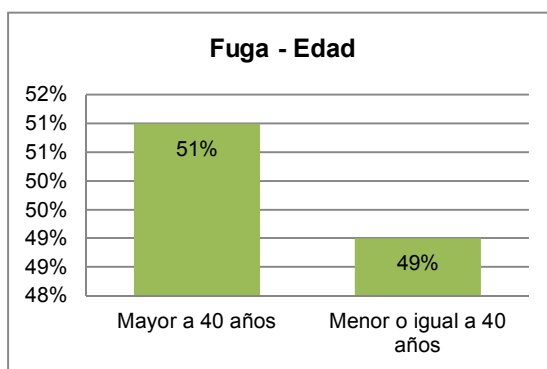
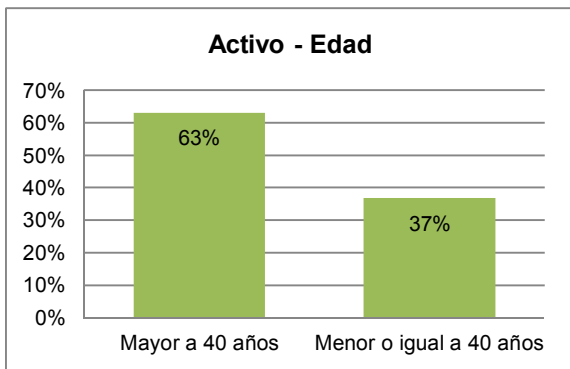


Gráfico N° 5.21: Distribución por Edad – Estado Activo y Fuga – Plaza Temuco

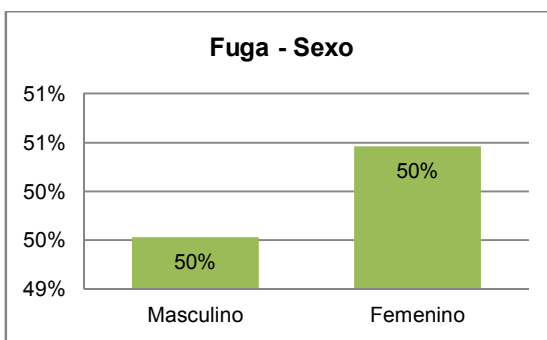
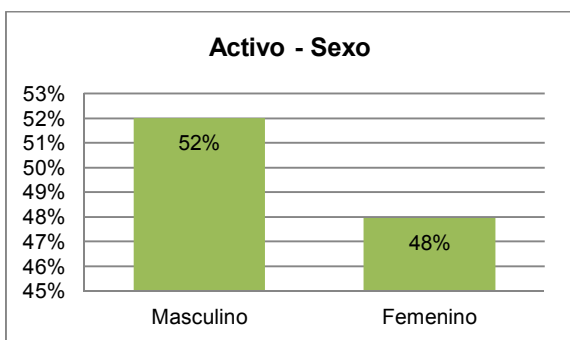


Gráfico N° 5.22: Distribución por Sexo – Estado Activo y Fuga – Plaza Temuco

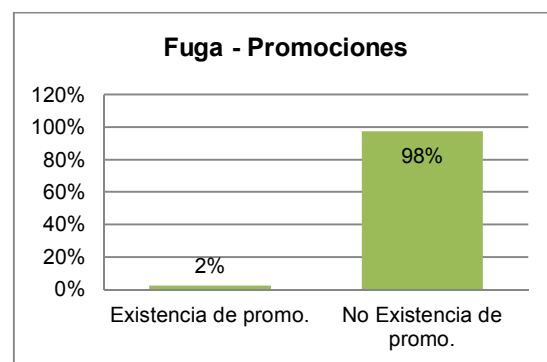


Gráfico N° 5.23: Distribución por Promociones – Estado Activo y Fuga – Plaza Temuco

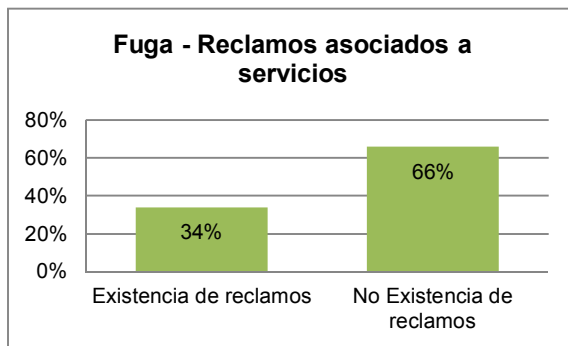
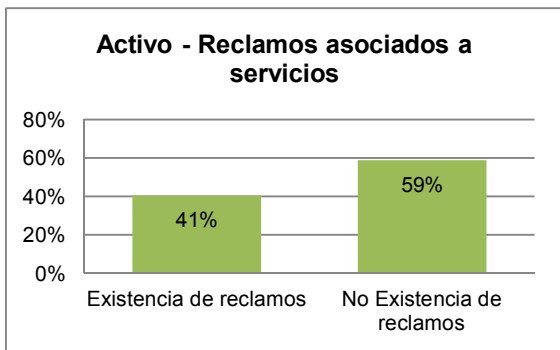


Gráfico N° 5.24: Distribución de clientes por Reclamos asociados a servicios – Estado Activo y Fuga – Plaza Temuco

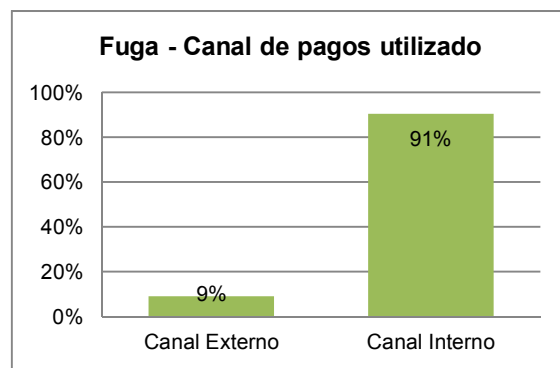
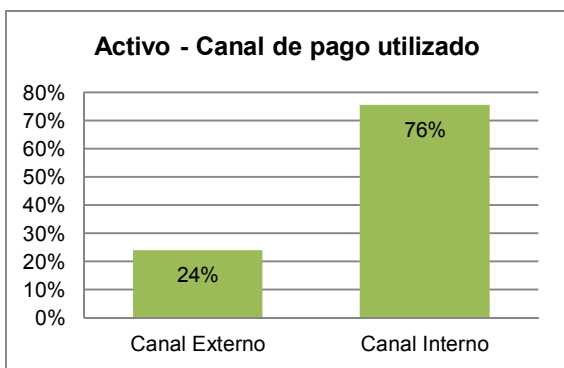


Gráfico N° 5.25: Distribución por Canal de pago utilizado – Estado Activo y Fuga – Plaza Temuco

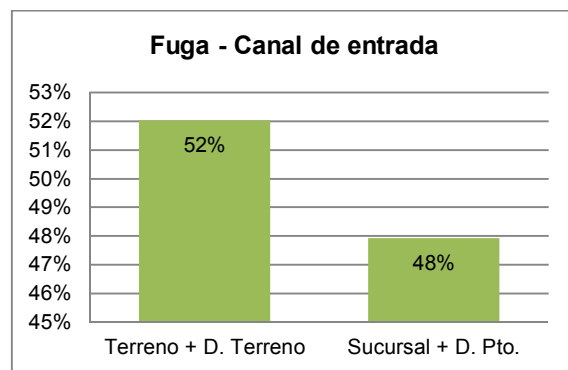
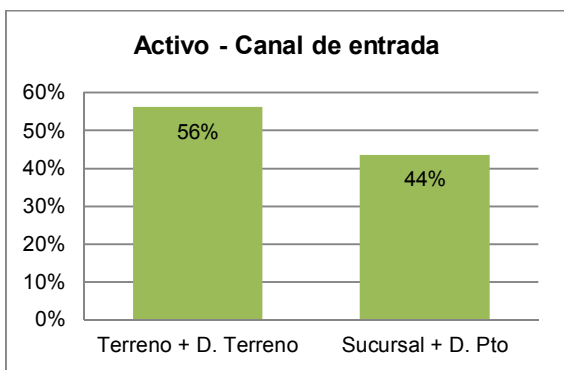


Gráfico N° 5.26: Distribución por Canal de entrada – Estado Activo y Fuga – Plaza Temuco

- **Resultados y Conclusiones**

- **Plaza Concepción**

- La diferencia de distribución de la variable GSE de los estados ACTIVO y FUGA es de un 9%.
- La diferencia de distribución de la variable DEUDA de los estados ACTIVO y FUGA es de un 23%.
- La diferencia de distribución de la variable PERMA_MESES de los estados ACTIVO y FUGA es de un 26%.
- La diferencia de distribución de la variable EDAD_C de los estados ACTIVO y FUGA es de un 13%.
- La diferencia de distribución de la variable SEXO de los estados ACTIVO y FUGA es de un 2%.
- La diferencia de distribución de la variable PROMO_ASOC de los estados ACTIVO y FUGA es de un 1%.
- La diferencia de distribución de la variable RECLAM de los estados ACTIVO y FUGA es de un 8%.
- La diferencia de distribución de la variable CPAGO de los estados ACTIVO y FUGA es de un 21%.
- La diferencia de distribución de la variable C_ENTRADA de los estados ACTIVO y FUGA es de un 5%.

Del análisis anterior, se resume lo siguiente.

- Las variables **GSE, DEUDA, PERMA_MESES, EDAD_C, RECLAM, CPAGO y C_ENTRADA** son variables relevantes de incluir en el modelo, ya que la comparación de distribuciones para cada variable es igual o superior en valor absoluto a 5%.

- Las variables **SEXO** y **PROMO_ASOC** por el contrario, no son relevantes de incluir en la generación del modelo, dado que la comparación de distribuciones para cada variable es menor en valor absoluto a 5%.

Plaza Temuco

- La diferencia de distribución de la variable GSE de los estados ACTIVO y FUGA es de un 7%.
- La diferencia de distribución de la variable DEUDA de los estados ACTIVO y FUGA es de un 21%.
- La diferencia de distribución de la variable PERMA_MESES de los estados ACTIVO y FUGA es de un 35%.
- La diferencia de distribución de la variable EDAD_C de los estados ACTIVO y FUGA es de un 12%.
- La diferencia de distribución de la variable SEXO de los estados ACTIVO y FUGA es de un 2%.
- La diferencia de distribución de la variable PROMO_ASOC de los estados ACTIVO y FUGA es de un 0%.
- La diferencia de distribución de la variable RECLAM de los estados ACTIVO y FUGA es de un 7%.
- La diferencia de distribución de la variable CPAGO de los estados ACTIVO y FUGA es de un 15%.
- La diferencia de distribución de la variable C_ENTRADA de los estados ACTIVO y FUGA es de un 4%.

Del análisis anterior, se resume lo siguiente.

- Las variables **GSE, DEUDA, PERMA_MESES, EDAD_C, RECLAM** y **CPAGO** son variables relevantes de incluir en el modelo, ya que la comparación de distribuciones para cada variable es igual o superior en valor absoluto a 5%.

- Las variables **SEXO**, **PROMO_ASOC** y **C_ENTRADA** por el contrario, no son relevantes de incluir en la generación del modelo, dado que la comparación de distribuciones para cada variable es menor en valor absoluto a 5%.

5.3.5 Exploración de las relaciones bivariantes

Antes de construir el modelo fue necesario explorar la asociación entre la variable dependiente (y) y las múltiples variables independientes (x_1, x_2, \dots, x_n) que son candidatas a ser incorporadas.

$$X_i \rightarrow Y$$

Para ello, en una primera etapa se analizó mediante tablas de contingencia las variables independientes (resultantes del análisis univariado) con la variable dependiente mediante una prueba Chi-Cuadrado.

Posteriormente, se realizó un análisis de regresión logística simple permitiendo ver la existencia de relación entre la variable dependiente y la independiente por medio del estadístico de Wald.

Prueba de independencia Chi-Cuadrado

La prueba de independencia Chi-Cuadrado permitió determinar si existe o no relación entre dos variables categóricas. Esta prueba indica si existe o no una relación directa entre las variables, pero no puede indicar el grado o nivel de relación.

H₀: Las variables son independientes.

H₁: Las variables están relacionadas entre sí.

Cada prueba realizada cuenta con un valor establecido de alfa (α), cuyo valor hace referencia al nivel de confianza que se desea que tenga la prueba Chi-

Cuadrado. Para el desarrollo de la investigación el nivel de confianza utilizado es de 95%, lo cual establece que el valor de α sea de 5%.

La prueba Chi-Cuadrado parte de la hipótesis que las variables son independientes, es decir, que no existe relación alguna entre ellas y que por tanto, ninguna de ellas ejerce influencia sobre la otra. Su objetivo es comprobar la hipótesis mediante el nivel de significación, por lo que si el valor de la significación es mayor o igual $\alpha = 5\%$, la hipótesis será aceptada, pero si es menor la hipótesis será rechazada.

Regresión logística simple

Este análisis permitió ver si la variable independiente influye de forma significativa en la probabilidad de ocurrencia del estudio.

H0: La variable independiente no influye significativamente sobre la variable dependiente.

H1: La variable independiente influye significativamente sobre la variable dependiente.

Las hipótesis presentadas anteriormente están planteadas conceptualmente y se necesita plantear la hipótesis de forma operativa. Esto significa que se deben encontrar parámetros estadísticos que pueden ser evaluados para que, según valores obtenidos, se pueda rechazar o no la hipótesis nula.

Un modelo de regresión logística univariado es validado si b_1 es significativamente distinto de cero. b_1 es el coeficiente de regresión logística muestral y es un estimador de B_1 , que es el coeficiente de regresión logística poblacional.

Cuando b_1 es significativamente distinto de cero, indica que es muy poco probable que B_1 sea cero.

Las hipótesis alternativas planteadas son las siguientes.

$$H_0 \quad B_1 = 0$$

$$H_1 \quad B_1 \neq 0$$

Estadístico de Wald

Por medio de este estadístico se logró contrastar de forma operativa las hipótesis anteriores. Se distribuye según una Chi-Cuadrado con un grado de libertad si la variable independiente es de tipo numérico, y si es cualitativa los grados de libertad será igual al número de categorías menos uno.

Si el valor experimental de Wald, bajo los supuestos de la hipótesis nula, es menor que el nivel de significancia fijado en el contraste, se rechaza la hipótesis nula y se concluye que la variable independiente influye en la probabilidad de las características de la variable dependiente.

Los resultados se presentan en tablas generadas en SPSS, en donde para cada variable se detalla el valor de su coeficiente, error estándar, parámetro de Wald, grados de libertad, valor de significancia (p) y el valor del parámetro ODDS RATIO. El intervalo de confianza utilizado es de 95%.

- **B:** Parámetro b estimado de la ecuación.
- **S.E.:** Error estándar de las estimaciones.
- **Wald:** Prueba de significación estadística que testea la hipótesis nula (b iguales a cero).
- **gl:** Grados de libertad de cada variable.
- **Sig.:** Nivel de significación del Wald (probabilidad de error al descartar la hipótesis nula).
- **Exp(B):** Muestra cuanto aumenta (o disminuye) el "ODDS RATIO". El valor 1 indica que la variable no influye.

- **Plaza Concepción**
- **Variable GSE**

Tabla N° 5.22: Prueba Chi-Cuadrado para la variable dependiente y variable GSE – Plaza Concepción

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	174,952	1	,000

Tabla N° 5.23: Regresión Logística Simple para la variable dependiente y variable GSE – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
GSE	-,373	,028	174,245	1	,000	,688	,651	,728
Constante	-,196	,016	143,058	1	,000	,822		

- **Variable DEUDA**

Tabla N° 5.24: Prueba Chi-Cuadrado para la variable dependiente y variable DEUDA – Plaza Concepción

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	1231,392	1	,000

Tabla N° 5.25: Regresión Logística Simple para la variable dependiente y variable DEUDA – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
DEUDA	,957	,028	1205,380	1	,000	2,605	2,468	2,750
Constante	-,725	,018	1600,289	1	,000	,484		

- **Variable PERMA_MESES**

Tabla N° 5.26: Prueba Chi-Cuadrado para la variable dependiente y variable PERMA_MESES – Plaza Concepción

Pruebas de chi-cuadrado			
	Valor	Gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	1623,872	1	,000

Tabla N° 5.27: Regresión Logística Simple para la variable dependiente y variable PERMA_MESES – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,126	,028	1572,461	1	,000	2,831	2,818	3,261
Constante	-,742	,018	1796,305	1	,000	,476		

- **Variable EDAD_C**

Tabla N° 5.28: Prueba Chi-Cuadrado para la variable dependiente y variable EDAD_C – Plaza Concepción

Pruebas de chi-cuadrado			
	Valor	Gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	416,083	1	,000

Tabla N° 5.29: Regresión Logística Simple para la variable dependiente y variable EDAD_C – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
EDAD_C	,547	,027	413,276	1	,000	1,728	1,639	1,821
Constante	-,573	,018	973,961	1	,000	,564		

- **Variable RECLAM**

Tabla N° 5.30: Prueba Chi-Cuadrado para la variable dependiente y variable RECLAM – Plaza Concepción

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	189,983	1	,000

Tabla N° 5.31: Regresión Logística Simple para la variable dependiente y variable RECLAM – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
RECLAM	-,408	,030	188,916	1	,000	,665	,627	,705
Constante	-,205	,016	169,340	1	,000	,815		

- **Variable CPAGO**

Tabla N° 5.32: Prueba Chi-Cuadrado para la variable dependiente y variable CPAGO – Plaza Concepción

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	1494,120	1	,000

Tabla N° 5.33: Regresión Logística Simple para la variable dependiente y variable CPAGO – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
CPAGO	-1,426	,039	1351,669	1	,000	,240	,223	,259
Constante	-,053	,015	12,800	1	,000	,948		

- **Variable C_ENTRADA**

Tabla N° 5.34: Prueba Chi-Cuadrado para la variable dependiente y variable C_ENTRADA – Plaza Concepción

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	56,351	1	,000

Tabla N° 5.35: Regresión Logística Simple para la variable dependiente y variable C_ENTRADA – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
C_ENTRADA	,209	,028	56,283	1	,000	1,232	1,167	1,302
Constante	-,458	,022	417,316	1	,000	,633		

- **Plaza Temuco**
- **Variable GSE**

Tabla N° 5.36: Prueba Chi-Cuadrado para la variable dependiente y variable GSE – Plaza Temuco

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	23,306	1	,000

Tabla N° 5.37: Regresión Logística Simple para la variable dependiente y variable GSE – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
GSE	-,281	,058	23,256	1	,000	,755	,674	,846
Constante	,614	,046	179,251	1	,000	1,848		

- **Variable DEUDA**

Tabla N° 5.38: Prueba Chi-Cuadrado para la variable dependiente y variable DEUDA – Plaza Temuco

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	226,106	1	,000

Tabla N° 5.39: Regresión Logística Simple para la variable dependiente y variable DEUDA – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
DEUDA	,886	,060	221,368	1	,000	2,425	2,158	2,725
Constante	,084	,036	5,357	1	,021	1,088		

- **Variable PERMA_MESES**

Tabla N° 5.40: Prueba Chi-Cuadrado para la variable dependiente y variable PERMA_MESES – Plaza Temuco

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	631,860	1	,000

Tabla N° 5.41: Regresión Logística Simple para la variable dependiente y variable PERMA_MESES – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,493	,061	598,165	1	,000	4,149	3,948	5,015
Constante	-,244	,039	38,516	1	,000	,784		

- **Variable EDAD_C**

Tabla N° 5.42: Prueba Chi-Cuadrado para la variable dependiente y variable EDAD_C – Plaza Temuco

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	74,987	1	,000

Tabla N° 5.43: Regresión Logística Simple para la variable dependiente y variable EDAD_C – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
EDAD_C	,498	,058	74,516	1	,000	1,645	1,469	1,842
Constante	,229	,037	37,993	1	,000	1,257		

- **Variable RECLAM**

Tabla N° 5.44: Prueba Chi-Cuadrado para la variable dependiente y variable RECLAM – Plaza Temuco

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	25,473	1	,000

Tabla N° 5.45: Regresión Logística Simple para la variable dependiente y variable RECLAM – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
RECLAM	-,294	,058	25,413	1	,000	,745	,665	,836
Constante	,552	,036	236,170	1	,000	1,736		

- **Variable CPAGO**

Tabla N° 5.46: Prueba Chi-Cuadrado para la variable dependiente y variable CPAGO – Plaza Temuco

Pruebas de chi-cuadrado			
	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	212,994	1	,000

Tabla N° 5.47: Regresión Logística Simple para la variable dependiente y variable CPAGO – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
CPAGO	-1,122	,079	200,049	1	,000	0,325	,279	,380
Constante	,620	,031	391,536	1	,000	1,859		

- **Resultados y Conclusiones**

Plaza Concepción

De las pruebas de independencia Chi-Cuadrado entre la variable dependiente y las variables independientes se observa que en todas el valor de significación es menor a 0,05 para las variables **GSE, DEUDA, PERMA_MESES, EDAD_C, RECLAM, CPAGO y C_ENTRADA** lo que permite concluir que la hipótesis nula (Las variables son independientes) se rechaza, y se acepta la hipótesis alternativa (Las variables están relacionadas entre sí).

Del análisis de regresión logística simple se observa que en todas las pruebas el valor de significación es menor a 0,05 para las variables **GSE, DEUDA, PERMA_MESES, EDAD_C, RECLAM, CPAGO y C_ENTRADA** con lo que se

concluye que la hipótesis nula (La variable independiente no influye significativamente sobre la variable dependiente) se rechaza, y se acepta la hipótesis alternativa (La variable independiente influye significativamente sobre la variable dependiente).

Plaza Temuco

De las pruebas de independencia Chi-Cuadrado entre la variable dependiente y las variables independientes se observa que en todas el valor de significación es menor a 0,05 para las variables **GSE, DEUDA, PERMA_MESES, EDAD_C, RECLAM y CPAGO** lo que permite concluir que la hipótesis nula (Las variables son independientes) se rechaza, y se acepta la hipótesis alternativa (Las variables están relacionadas entre sí).

Del análisis de regresión logística simple se observa que en todas las pruebas el valor de significación es menor a 0,05 para las variables **GSE, DEUDA, PERMA_MESES, EDAD_C, RECLAM y CPAGO** con lo que se concluye que la hipótesis nula (La variable independiente no influye significativamente sobre la variable dependiente) se rechaza, y se acepta la hipótesis alternativa (La variable independiente influye significativamente sobre la variable dependiente).

Por último se observa que ambas pruebas, tanto Chi-Cuadrado y Regresión Logística Simple, concluyen lo mismo con respecto a las variables significantes para ser incluidas en el modelo.

5.3.6 Evaluación de interacciones o modificaciones de efecto y/o confusión

Para la evaluación de posibles interacciones o modificaciones de tipo efecto y/o confusión se eligió como base de estudio la variable independiente PERMA_MESES y la variable dependiente ESTAD_SERV (ACTIVO / FUGA), para luego explorar el papel de las restantes variables independientes sobre las anteriores.

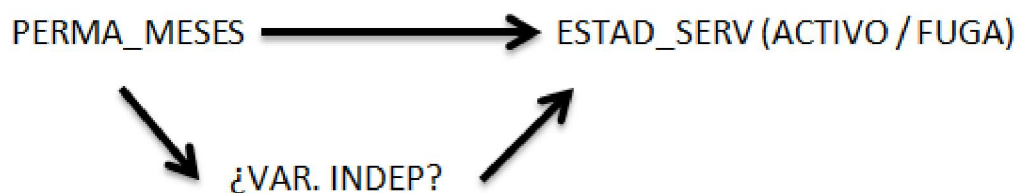


Figura N° 5.6: Interacción y Confusión sobre la variable dependiente y la variable independiente PERMA_MESES

Dentro de las posibles modificaciones que se pueden generar, están las que se detallan a continuación.

- **Interacción**, ocurre cuando un término multiplicativo compuesto por dos variables independientes interactúan en el efecto sobre la v. dependiente. Si la significancia de esta variable es $<0,05$ se considera estadísticamente significativa y se incluye en la ecuación del modelo.
- **Confusión**, se genera cuando el valor OR que evalúa la fuerza de asociación entre la v. independiente y la v. dependiente cambia de forma importante cuando se introduce en la ecuación de RLM la tercera variable. Si el cambio porcentual es igual o mayor a 10%, se establece que esa variable introducida debe incluirse en la ecuación del modelo. El valor OR en las tablas se señala con nombre $Exp(B)$.

La información que se detalla, proviene del análisis de regresión logística simple realizado en la exploración de las relaciones bivariantes. Se muestra los valores Chi-Cuadrado, Significancia y OR para la variable independiente PERMA_MESES. Los valores OR serán utilizados en el análisis de confusión.

PLAZA CONCEPCIÓN

PERMA_MESES → ACTIVO / FUGA

X^2 : 1623,872, significancia: 0,000

OR: 2,831 (IC95% 2,818 – 3,261)

PLAZA TEMUCO

PERMA_MESES → ACTIVO / FUGA

X^2 : 631,86, significancia: 0,000

OR: 4,149 (IC95% 3,948 – 5,015)

A continuación se detallan los modelos en donde se incluye una tercera variable, lo que permitirá evaluar interacciones o modificaciones de tipo efecto y/o confusión.

El análisis de resultados y conclusiones se detalla posteriormente.

- Plaza Concepción
- PERMA_MESES - GSE → ACTIVO / FUGA

Tabla N° 5.48: Regresión Logística Multivariante – Variables PERMA_MESES, GSE y multiplicativa – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,156	,036	1041,213	1	,000	3,179	2,963	3,410
GSE	-,482	,039	154,566	1	,000	,618	,572	,666
GSE * PERMA_MESES	,025	,060	3,054	1	,065	1,026	,912	1,154
Constante	-,592	,021	803,091	1	,000	,553		

Tabla N° 5.49: Regresión Logística Multivariante – Variables PERMA_MESES y GSE – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,165	,029	1641,556	1	,000	3,207	3,032	3,393
GSE	-,471	,030	253,671	1	,000	,624	,589	,661
Constante	-,595	,020	922,342	1	,000	,552		

- PERMA_MESES - DEUDA → ACTIVO / FUGA

Tabla N° 5.50: Regresión Logística Multivariante – Variables PERMA_MESES, DEUDA y multiplicativa – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,397	,039	1304,242	1	,000	4,045	3,750	4,364
DEUDA	1,200	,036	1083,788	1	,000	3,320	3,091	3,566
DEUDA * PERMA_MESES	-,622	,059	2,542	1	,080	,537	,478	,602
Constante	-1,259	,025	2511,628	1	,000	,284		

Tabla N° 5.51: Regresión Logística Multivariante – Variables PERMA_MESES y DEUDA – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,133	,029	1502,951	1	,000	3,105	2,932	3,288
DEUDA	,965	,029	1136,018	1	,000	2,624	2,481	2,776
Constante	-1,149	,022	2671,487	1	,000	,317		

• **PERMA_MESES - EDAD_C → ACTIVO / FUGA**

Tabla N° 5.52: Regresión Logística Multivariante – Variables PERMA_MESES, EDAD_C y multiplicativa – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,151	,038	902,217	1	,000	3,160	2,932	3,407
EDAD_C	,633	,035	318,886	1	,000	1,883	1,757	2,019
EDAD_C * PERMA_MESES	,142	,059	2,932	1	,067	1,153	1,026	1,295
Constante	-1,059	,026	1687,369	1	,000	,347		

Tabla N° 5.53: Regresión Logística Multivariante – Variables PERMA_MESES y EDAD_C – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,210	,029	1719,782	1	,000	3,355	3,168	3,552
EDAD_C	,684	,028	580,748	1	,000	1,982	1,875	2,095
Constante	-1,086	,023	2170,314	1	,000	,338		

• **PERMA_MESES - RECLAM → ACTIVO / FUGA**

Tabla N° 5.54: Regresión Logística Multivariante – Variables PERMA_MESES, RECLAM y multiplicativa – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,326	,034	1482,657	1	,000	3,767	3,521	4,030
RECLAM	-,179	,039	21,399	1	,000	,836	,775	,902
PERMA_MESES * RECLAM	-,643	,062	105,973	1	,000	,526	,465	,594
Constante	-,689	,021	1104,406	1	,000	,502		

Tabla N° 5.55: Regresión Logística Multivariante – Variables PERMA_MESES y RECLAM – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,135	,029	1577,639	1	,000	3,110	2,940	3,289
RECLAM	-,432	,031	196,398	1	,000	,649	,611	,690
Constante	-,619	,019	1020,610	1	,000	,538		

• **PERMA_MESES - CPAGO → ACTIVO / FUGA**

Tabla N° 5.56: Regresión Logística Multivariante – Variables PERMA_MESES, CPAGO y multiplicativa – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,233	,033	1437,135	1	,000	3,432	3,220	3,658
CPAGO	-1,264	,051	602,289	1	,000	,283	,256	,313
CPAGO * PERMA_MESES	-,461	,080	2,996	1	,071	,631	,539	,738
Constante	-,502	,019	676,536	1	,000	,605		

Tabla N° 5.57: Regresión Logística Multivariante – Variables PERMA_MESES y CPAGO – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,158	,030	1530,931	1	,000	3,183	3,004	3,373
CPAGO	-1,464	,040	1336,023	1	,000	,231	,214	,250
Constante	-,476	,019	647,726	1	,000	,622		

• **PERMA_MESES - C_ENTRADA → ACTIVO / FUGA**

Tabla N° 5.58: Regresión Logística Multivariante – Variables PERMA_MESES, C_ENTRADA y multiplicativa – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,018	,047	461,878	1	,000	2,768	2,523	3,038
C_ENTRADA	,146	,037	15,964	1	,000	1,158	1,077	1,244
C_ENTRADA * PERMA_MESES	,170	,059	3,080	1	,064	1,185	1,055	1,331
Constante	-,836	,029	803,175	1	,000	,434		

Tabla N° 5.59: Regresión Logística Multivariante – Variables PERMA_MESES y C_ENTRADA – Plaza Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,127	,028	1569,590	1	,000	3,186	2,919	3,263
C_ENTRADA	,211	,029	53,559	1	,000	1,235	1,167	1,307
Constante	-,878	,026	1163,061	1	,000	,416		

- Plaza Temuco
- PERMA_MESES - GSE → ACTIVO / FUGA

Tabla N° 5.60: Regresión Logística Multivariante – Variables PERMA_MESES, GSE y multiplicativa – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,768	,110	257,325	1	,000	5,859	4,721	7,271
GSE	-,408	,079	26,485	1	,000	,665	,569	,777
GSE * PERMA_MESES	-,299	,134	2,086	1	,096	,742	,570	,965
Constante	-,022	,058	,143	1	,705	,978		

Tabla N° 5.61: Regresión Logística Multivariante – Variables PERMA_MESES, GSE y multiplicativa – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,569	,063	629,507	1	,000	4,803	4,249	5,429
GSE	-,514	,063	65,997	1	,000	,598	,528	,677
Constante	,035	,052	,459	1	,498	1,036		

- PERMA_MESES - DEUDA → ACTIVO / FUGA

Tabla N° 5.62: Regresión Logística Multivariante – Variables PERMA_MESES, DEUDA y multiplicativa – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,793	,081	494,442	1	,000	6,010	5,131	7,039
DEUDA	1,229	,083	219,669	1	,000	3,418	2,905	4,021
DEUDA * PERMA_MESES	-,682	,128	3,230	1	,061	,506	,394	,649
Constante	-,768	,055	197,010	1	,000	,464		

Tabla N° 5.63: Regresión Logística Multivariante - Variables PERMA_MESES y DEUDA – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,534	,063	596,627	1	,000	4,637	4,100	5,244
DEUDA	,952	,064	223,364	1	,000	2,592	2,288	2,937
Constante	-,648	,049	175,596	1	,000	,523		

• **PERMA_MESES - EDAD_C → ACTIVO / FUGA**

Tabla N° 5.64: Regresión Logística Multivariante - Variables PERMA_MESES, EDAD_C y multiplicativa – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,346	,078	294,152	1	,000	3,843	3,295	4,482
EDAD_C	,425	,079	28,881	1	,000	1,530	1,310	1,787
EDAD_C * PERMA_MESES	,529	,131	2,591	1	,081	1,697	1,313	2,192
Constante	-,442	,054	66,110	1	,000	,643		

Tabla N° 5.65: Regresión Logística Multivariante - Variables PERMA_MESES y EDAD_C – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,546	,062	617,952	1	,000	4,693	4,154	5,301
EDAD_C	,626	,062	101,597	1	,000	1,870	1,656	2,113
Constante	-,537	,050	117,342	1	,000	,584		

• **PERMA_MESES - RECLAM → ACTIVO / FUGA**

Tabla N° 5.66: Regresión Logística Multivariante - Variables PERMA_MESES, RECLAM y multiplicativa – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,795	,083	469,479	1	,000	6,019	5,117	7,080
RECLAM	-,200	,084	5,656	1	,017	,819	,694	,965
PERMA_MESES * RECLAM	-,605	,127	22,828	1	,000	,546	,426	,700
Constante	-,179	,048	13,944	1	,000	,836		

Tabla N° 5.67: Regresión Logística Multivariante - Variables PERMA_MESES y RECLAM – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,545	,062	619,275	1	,000	4,690	4,153	5,297
RECLAM	-,470	,063	55,354	1	,000	,625	,552	,707
Constante	-,093	,044	4,394	1	,036	,912		

• **PERMA_MESES - CPAGO → ACTIVO / FUGA**

Tabla N° 5.68: Regresión Logística Multivariante – Variables PERMA_MESES, CPAGO y multiplicativa – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,534	,069	498,308	1	,000	4,638	4,054	5,307
CPAGO	-1,355	,132	105,894	1	,000	,258	,199	,334
CPAGO * PERMA_MESES	,155	,175	2,488	1	,075	1,168	,829	1,647
Constante	-,062	,042	2,145	1	,143	,940		

Tabla N° 5.69: Regresión Logística Multivariante – Variables PERMA_MESES y CPAGO – Plaza Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
PERMA_MESES	1,559	,063	607,654	1	,000	4,753	4,199	5,380
CPAGO	-1,268	,086	217,469	1	,000	,281	,238	,333
Constante	-,071	,041	3,012	1	,083	,931		

- **Análisis de resultados**
- **Interacción**

Luego de analizar los modelos generados (donde se incluye una variable multiplicativa) se observa que solo hay interacción cuando se introduce la variable PERMA_MESES * RECLAM dado que es el único caso que la variable multiplicativa tiene significancia estadística <0,05. Por ello, para la generación del modelo tanto para la plaza de Concepción y Temuco se debe incluir esta variable multiplicativa.

- **Confusión**

El análisis de confusión se realiza luego de sacar del modelo la variable multiplicativa. Luego se observa que en todos los modelos generados hay presencia de confusión dado que existe un cambio porcentual absoluto igual o superior al 10% del OR de la variable independiente PERMA_MESES, con respecto al valor de la misma variable resultante del análisis de regresión logística simple.

Tabla N° 5.70: Resultado del análisis de Confusión – Plaza Concepción

Variable indep. base	Variable Introducida	OR	OR regresión log. simple	Cambio %
PERMA_MESES	GSE	3,207	2,831	13%
PERMA_MESES	DEUDA	3,105	2,831	10%
PERMA_MESES	EDAD_C	3,355	2,831	19%
PERMA_MESES	RECLAM	3,11	2,831	10%
PERMA_MESES	CPAGO	3,183	2,831	12%
PERMA_MESES	C_ENTRADA	3,186	2,831	13%

Tabla N° 5.71: Resultado del análisis de Confusión – Plaza Temuco

Variable indep. base	Variable Introducida	OR	OR regresión log. simple	Cambio %
PERMA_MESES	GSE	4,803	4,149	16%
PERMA_MESES	DEUDA	4,637	4,149	12%
PERMA_MESES	EDAD_C	4,693	4,149	13%
PERMA_MESES	RECLAM	4,69	4,149	13%
PERMA_MESES	CPAGO	4,753	4,149	15%

- **Conclusión general**

Del análisis de interacción y confusión se establece que las variables **GSE, DEUDA, EDAD_C, PERMA_MESES, CPAGO, RECLAM, C_ENTRADA** y **PERMA_MESES * RECLAM** deben ser incluidas en el modelo para la plaza de Concepción.

Del mismo análisis se establece que las variables **GSE, DEUDA, EDAD_C, PERMA_MESES, RECLAM, CPAGO** y **PERMA_MESES * RECLAM** deben ser incluidas en el modelo para la plaza de Temuco.

5.3.7 Construcción del modelo de Regresión Logística Multivariante

Luego de las conclusiones que se han generado tanto en los análisis univariante, bivariante y evaluación de posibles interacciones o modificaciones de tipo efecto y/o confusión, se está en condiciones de generar los modelos de regresión logística multivariante para ambas plazas.

Los resultados del modelo generado se presentan en una tabla resumen donde se detalla el valor de coeficiente, error estándar, parámetro de Wald, grados de libertad, valor de significancia (p), el valor del parámetro ODDS RATIO y su respectivo intervalo de confianza (95%) para cada covariable.

Para este trabajo si el valor de probabilidad está comprendido entre 0 y 0,5 el servicio se clasifica como $y = 0$ (Activo), mientras que si la probabilidad calculada está comprendida entre 0,5 y 1 el servicio se clasifica como $y = 1$ (Fuga).

Posterior a la generación de modelos para ambas plazas, es necesario conocer los niveles de sensibilidad y especificidad que generan ambos, y también el porcentaje global de acierto.

- La sensibilidad es la probabilidad de clasificar correctamente a un servicio fugado.

$$\% \text{ Sensibilidad} = \text{Fugas Positivas} / \text{Total Servicios Fugados}$$

- La especificidad es la probabilidad de clasificar correctamente a un servicio activo.

$$\% \text{ Especificidad} = \text{Activos Negativos} / \text{Total Servicios Activos}$$

- **Plaza Concepción**

Las variables independientes incluidas en el modelo de RLM son las siguientes:

- **GSE:** Grupo socioeconómico del cliente.
- **DEUDA:** Nivel de deuda existente para cada servicio.
- **PERMA_MESES:** Antigüedad del servicio contratado.
- **EDAD_C:** Edad del cliente.
- **RECLAM:** Existencia de reclamo en el servicio contratado.
- **CPAGO:** Canal de pago utilizado para pago de servicios.
- **C_ENTRADA:** Canal de entrada para la contratación de él o los servicios por parte del cliente.
- **PERMA_MESES * RECLAM:** Variable multiplicativa compuesta por la interacción entre ambas variables independientes.

Tabla N° 5.72: Modelo de RLM para plaza de Concepción

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
GSE	-,274	,032	73,082	1	,000	,761	,714	,810
DEUDA	,823	,030	740,196	1	,000	2,277	2,146	2,416
PERMA_MESES	1,467	,038	1509,015	1	,000	4,338	4,028	4,671
EDAD_C	,721	,030	561,015	1	,000	2,056	1,937	2,183
RECLAM	-,210	,041	25,822	1	,000	,811	,748	,879
CPAGO	-1,390	,042	1117,249	1	,000	,249	,230	,270
C_ENTRADA	,185	,031	34,863	1	,000	1,203	1,131	1,279
PERMA_MESES * RECLAM	-,643	,067	91,437	1	,000	,526	,461	,600
Constante	-1,164	,038	923,550	1	,000	,312		

Se puede observar que todas las covariables incluidas en el modelo mantienen la significación estadística $<0,05$ en el contraste de hipótesis que las relaciona con la variable dependiente ESTAD_SERV.

Con respecto a los coeficientes de significación:

- $Exp(\beta) > 1$ para las variables **DEUDA**, **PERMA_MESES**, **EDAD_C** y **C_ENTRADA**, lo que implica que estas variables son factores que incrementan la probabilidad de ocurrencia de fuga de servicios.
- $Exp(\beta) < 1$ para las variables **GSE**, **RECLAM**, **CPAGO** y **PERMA_MESES * RECLAM**, lo que implica que estas variables son factores que disminuyen la probabilidad de ocurrencia de fuga de servicios.

En general los errores estándar (E.T. en la tabla) para cada covariable son bajos y además se observa que los intervalos de confianza son acotados permitiendo que la estimación de los estimadores sea menos imprecisa.

Tabla N° 5.73: Resumen de sensibilidad, especificidad y porcentaje global de acierto – Plaza Concepción

Observado		Pronosticado		Porcentaje correcto
		ESTAD_SERV		
ESTAD_SERV		0	1	
	0	10840	2637	80,4
	1	4044	5708	58,5
Porcentaje global				71,2

Se puede apreciar que el modelo clasifica de forma correcta 5708 casos (de los 9752 totales) de fugas ($y=1$), por lo que su **sensibilidad** es de **58,5%**, por otra parte, clasifica de forma correcta 10840 casos (de los 13477 totales) activos ($y=0$), por lo que su **especificidad** es de **80,4%**, porcentaje más alto que el de sensibilidad.

El modelo de forma global clasifica de forma correcta al **71,2%** de los casos analizados.

- **Plaza Temuco**

Las variables independientes incluidas en el modelo de RLM son las siguientes:

- **GSE:** Grupo socioeconómico del cliente.
- **DEUDA:** Nivel de deuda existente para cada servicio.
- **PERMA_MESES:** Antigüedad del servicio contratado.
- **EDAD_C:** Edad del cliente.
- **RECLAM:** Existencia de reclamo en el servicio contratado.
- **CPAGO:** Canal de pago utilizado para pago de servicios.
- **PERMA_MESES * RECLAM:** Variable multiplicativa compuesta por la interacción entre ambas variables independientes.

Tabla N° 5.74: Modelo de RLM para plaza de Temuco

Variables en la ecuación								
	B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
							Inferior	Superior
GSE	-,368	,068	29,659	1	,000	,692	,606	,790
DEUDA	,836	,066	158,012	1	,000	2,306	2,025	2,627
PERMA_MESES	2,060	,090	520,439	1	,000	7,844	6,572	9,363
EDAD_C	,623	,066	89,898	1	,000	1,865	1,640	2,122
RECLAM	-,217	,089	5,959	1	,015	,805	,676	,958
CPAGO	-1,176	,091	166,566	1	,000	,308	,258	,369
PERMA_MESES * RECLAM	-,715	,134	28,394	1	,000	,489	,376	,637
Constante	-,460	,076	37,132	1	,000	,631		

Al igual que para la plaza de Concepción, se observa que todas las covariables incluidas en el modelo mantienen la significación estadística <0,05 en el contraste de hipótesis que las relaciona con la variable dependiente ESTAD_SERV.

Con respecto a los coeficientes de significación:

- $Exp(\beta) > 1$ para las variables **DEUDA**, **PERMA_MESES** y **EDAD_C**, lo que implica que estas variables son factores que incrementan la probabilidad de ocurrencia de fuga de servicios.
- $Exp(\beta) < 1$ para las variables **GSE**, **RECLAM**, **CPAGO** y **PERMA_MESES** * **RECLAM**, lo que implica que estas variables son factores que disminuyen la probabilidad de ocurrencia de fuga de servicios.

En general los errores estándar (E.T. en la tabla) para cada covariable son bajos y además se observa que los intervalos de confianza son acotados permitiendo que la estimación de los estimadores sea menos imprecisa.

Tabla N° 5.75: Resumen de sensibilidad, especificidad y porcentaje global de acierto – Plaza Temuco

Observado		Pronosticado		Porcentaje correcto
		ESTAD_SERV 0	1	
ESTAD_SERV	0	1142	922	55,3
	1	551	2661	82,8
Porcentaje global				72,1

Se puede apreciar que el modelo clasifica de forma correcta 2661 casos (de los 3212 totales) de fugas ($y=1$), por lo que su **sensibilidad** es de **82,8%**, por otra parte, clasifica de forma correcta 1142 casos (de los 2064 totales) activos ($y=0$), por lo que su especificidad es de **55,3%**, porcentaje más bajo que el de sensibilidad.

Finalmente el modelo de forma global clasifica de forma correcta al **72,1%** de los casos analizados.

- **Curva COR**

Clásicamente, la exactitud de una prueba diagnóstica se ha evaluado en función de dos características: la sensibilidad y la especificidad⁶. Sin embargo, éstas varían en función del criterio elegido como punto de corte, para el caso de este trabajo, entre los servicios activos y los fugados. Una forma más global de conocer la calidad de la prueba en el espectro completo de puntos de corte es mediante el uso de curvas ROC (receiver operating characteristics) que constituyen una herramienta fundamental y unificadora en el proceso de evaluación y uso de las pruebas diagnósticas.

La curva ROC proporciona una visión general del espectro completo de posibilidades de sensibilidad y especificidad, ya que de forma gráfica se puede observar todos los pares de sensibilidad / especificidad resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados. En el eje y de coordenadas se sitúa la sensibilidad o fracción de verdaderos positivos, y en el eje x se sitúa la fracción de falsos positivos o 1-especificidad.

Los resultados en detalle de la curva COR, se incluyen en el **ANEXO G, H, I y J**.

5.3.8 Validación de modelos

Posterior a la construcción de los modelos, es importante ver el nivel de validación que tienen estos últimos. La validación es ver si “trabaja” de igual forma con otros datos distintos de aquellos con los que se ha generado.

Para evaluar esto último, se realizó una toma de muestra aleatoria simple de 50 casos de servicios fugados para los meses de Abril, Mayo y Junio de 2011 para ambas plazas.

Es de importancia señalar que se han utilizado tres niveles de corte para evaluar el nivel de acierto entregado por los modelos.

6. Jenicek M, Cleroux R. 1987.

- **Plaza Concepción**

Parámetros del modelo:

$$\beta_0 = -1,164, \beta_1 = -0,274, \beta_2 = 0,823, \beta_3 = 1,467, \beta_4 = 0,721, \\ \beta_5 = -0,210, \beta_6 = -1,390, \beta_7 = 0,185, \beta_8 = -0,643$$

$$P = \frac{1}{1 + e^{-z-c}} \quad (37)$$

Siendo:

$$z_c = -1,164 - 0,274 * GSE + 0,823 * DEUDA + 1,467 * PERMA_MESES + \\ 0,721 * EDAD_C - 0,210 * RECLAM - 1,390 * CPAGO + 0,185 * C_ENTRADA - \\ 0,643 * PERMA_MESES * RECLAM$$

Tabla Nº 5.76: Validación del modelo de RLM con tres niveles de corte – Plaza Concepción

	Abril	Mayo	Junio
Fugas (M. A. S)	50	50	50
Nº Aciertos (N.C. 0,3)	19	12	26
Nº Aciertos (N.C. 0,4)	13	12	22
Nº Aciertos (N.C. 0,5)	12	12	21
Nº Aciertos promedio	15	12	23

De la Tabla Nº 5.76 se observa que el modelo tiene un mayor número de aciertos en el mes de Junio de 2011, muy superior con respecto a los otros meses. De forma promedio, tiene un nivel de acierto de 23 casos analizados para Junio, 12 para Mayo y 15 para Abril. Además, a medida que el nivel de corte disminuye, se puede observar que el nivel de acierto aumenta.

37. Ecuación de regresión logística multivariante para predicción de fuga de clientes para plaza de Concepción

- **Plaza Temuco**

Parámetros del modelo:

$$\beta_0 = -0,460, \beta_1 = -0,368, \beta_2 = 0,823, \beta_3 = 2,060, \beta_4 = 0,623, \\ \beta_5 = -0,217, \beta_6 = -1,176, \beta_7 = -0,716$$

$$P = \frac{1}{1 + e^{-z-t}} \quad (38)$$

Siendo:

$$z_t = -0,460 - 0,368 * GSE + 0,823 * DEUDA + 2,060 * PERMA_MESES + \\ 0,623 * EDAD_C - 0,217 * RECLAM - 1,176 * CPAGO - 0,716 * \\ PERMA_MESES * RECLAM$$

Tabla N° 5.77: Validación del modelo de RLM con tres niveles de corte – Plaza Temuco

	Abril	Mayo	Junio
Fugas (M. A. S)	50	50	50
Nº Aciertos (N.C. 0,3)	28	26	28
Nº Aciertos (N.C. 0,4)	22	21	21
Nº Aciertos (N.C. 0,5)	17	14	18
Nº Aciertos promedio	22	20	22

De la Tabla N° 5.77 se observa que el modelo tiene un mejor número de aciertos que el modelo generado para la plaza de Concepción. Esto concuerda con los porcentajes de sensibilidad y especificidad obtenidos en la construcción del mismo. Se puede ver que el número de aciertos promedio es muy similar para todos los meses, existiendo una leve mejoría en el número de aciertos promedio para los meses de Abril y Junio. Por último, como sucede de igual forma con la plaza de

38. Ecuación de regresión logística multivariante para predicción de fuga de clientes para plaza de Temuco

Concepción, a medida que disminuye el nivel de corte, se observa que hay un mayor número de aciertos.

CAPITULO 6: MARKETING RELACIONAL APLICADO A UN PLAN DE ACCIÓN

6.1 Introducción

Día a día las empresas buscan atraer nuevos clientes basados en modernas estrategias de captación, ofreciendo para ello, una gama de productos que cada cierto tiempo van renovando. La creación de necesidades va aumentando y es aquí donde las empresas de todo tipo colocan todos sus esfuerzos en hacer la vida más fácil de sus clientes. La publicidad que forma parte de la estrategia de marketing, se utiliza para crear imagen de marca como es el *marketing directo* para hablarle al consumidor más de cerca y el marketing relacional, que busca la fidelización permitiendo con ello:

- **Una mayor competitividad y eficacia:** de la empresa gracias a los vínculos de unión que mantiene con sus clientes, que son un muro para la competencia, y a la flexibilidad y rapidez para adaptarse a las preferencias de cada cliente.
- **Segmentar a los clientes:** con información real y significancia, lo que posibilita una revisión de la misión inicial de la empresa. Esta segmentación permite un crecimiento y globalización más efectivo a través de la búsqueda de clientes posibles que compartan los rasgos característicos de los mejores clientes.

Los argumentos en que se sustentan y justifican los enfoques de fidelización resaltan los beneficios para la empresa que los proponen. Se habla de retener a los clientes, de mejorar la productividad de las actuaciones de marketing y de rentabilidad.

El marketing relacional enfocado al “cliente individual”, hace más eficaz la respuesta de la empresa al mercado fragmentado, y permite la constante reorientación de la segmentación, es decir, de la estrategia para afrontar el crecimiento y la globalización del mercado, poniendo énfasis en la atención y cultivo de la relación entre el cliente y la empresa. El marketing debe trabajar conjuntamente con los clientes para establecer asociaciones que vayan más allá de los intercambios tradicionales para convertirse en una parte integral de la

empresa. El cliente es la base de la vida de cualquier tipo de empresa, y esta para ser competitiva y crecer con rentabilidad, debe adaptarse a sus necesidades hasta el punto de conseguir integrarlo dentro de su misma organización, desde el desarrollo de nuevos productos hasta la definición operativa de los procesos de producción, o de entrega de los servicios.

Dada esta nueva postura de relación con el cliente, se ha llegado a través de distintas líneas de desarrollo conceptual y de práctica empresarial que según Alet (1996) son:

- **Comunicación:** Marketing directo que ha visto incrementar su importancia ante la saturación y descontrol de los medios masivos, y crecimiento de las necesidades diversificadas de clientes.
- **Distribución:** Asociación entre los poseedores de canales y empresas que comparten la base de datos de clientes.
- **Marketing y Servicios:** Consta de la importancia de la satisfacción del cliente en relación a la frecuencia y calidad de los contactos con él.
- **Marketing de productos y servicios industriales o de alta tecnología:** es la creación y mantenimiento de la relación con cada uno de los miembros de la organización de los clientes, para adaptarse a sus deseos y necesidades cambiantes en un contexto competitivo.
- **Calidad:** es fundamental en la satisfacción del cliente y tienen un papel preponderante en el desarrollo de productos y servicios dentro del funcionamiento de la empresa.
- **Producción flexible, JIT:** Los sistemas productivos se orientan a conseguir la máxima flexibilidad, en lugar de la obtención de economías de escala. Las eficiencias se analizan dentro del proceso integral desde el pedido a la entrega al cliente final, no simplemente en el mismo proceso de la fábrica.

- **Estrategia de la empresa:** La estrategia de la empresa consiste en el avance progresivo de los acuerdos y las relaciones entre las empresas, desde alianzas estratégicas hasta la red de empresas. Se desarrollará la estrategia de la empresa en torno al cliente tratando de hacer los procesos de marketing más disciplinados, encarados a objetivos comunes de toda la empresa, y donde el marketing tiene un papel clave como motor de la orientación al cliente.

La creación de valor para el cliente parte necesariamente del reconocimiento de la importancia de las relaciones, las cuales afectan tanto al contenido como al resultado de las transacciones. Se desarrollará la estrategia de la empresa en torno al cliente, tratando de hacer los procesos de marketing más disciplinados, encarados a objetivos comunes de toda la empresa, y donde el marketing tiene un papel clave como motor de la orientación al cliente.

6.2 Evolución de Marketing

La gestión del marketing en la última década, se ha enfrentado a cambios drásticos en el entorno de la empresa que han puesto un reto importante en su desarrollo, llevando a la aparición de factores nuevos, y a cambios en la importancia de los existentes. Especialmente se ha generado una complejidad que dificulta en gran medida las posibilidades de gestión de las variables comerciales de una forma conocida, para poder alcanzar una ventaja competitiva sostenible en mercados en rápida transformación.

En la Figura N°6.1, se sintetiza los factores más significativos del contexto actual y futuro del marketing relacional.

I. Evolución del consumidor

Su comportamiento de compra es cada vez más difuso y difícil de seguir. Se constata un cambio de los roles de compra familiares (mayor importancia de la decisión conjunta hombre-mujer, compra mayor de los niños por falta de tiempo de los padres, mayor compra directa de los hombres por la incorporación de las mujeres al mercado laboral), son utilizados distintos sistemas de compra más

adaptados al estilo de vida, y se modifica el peso de las variables de decisión de compra apareciendo básicamente el tiempo, que provoca la exigencia creciente de más y mejores servicios.

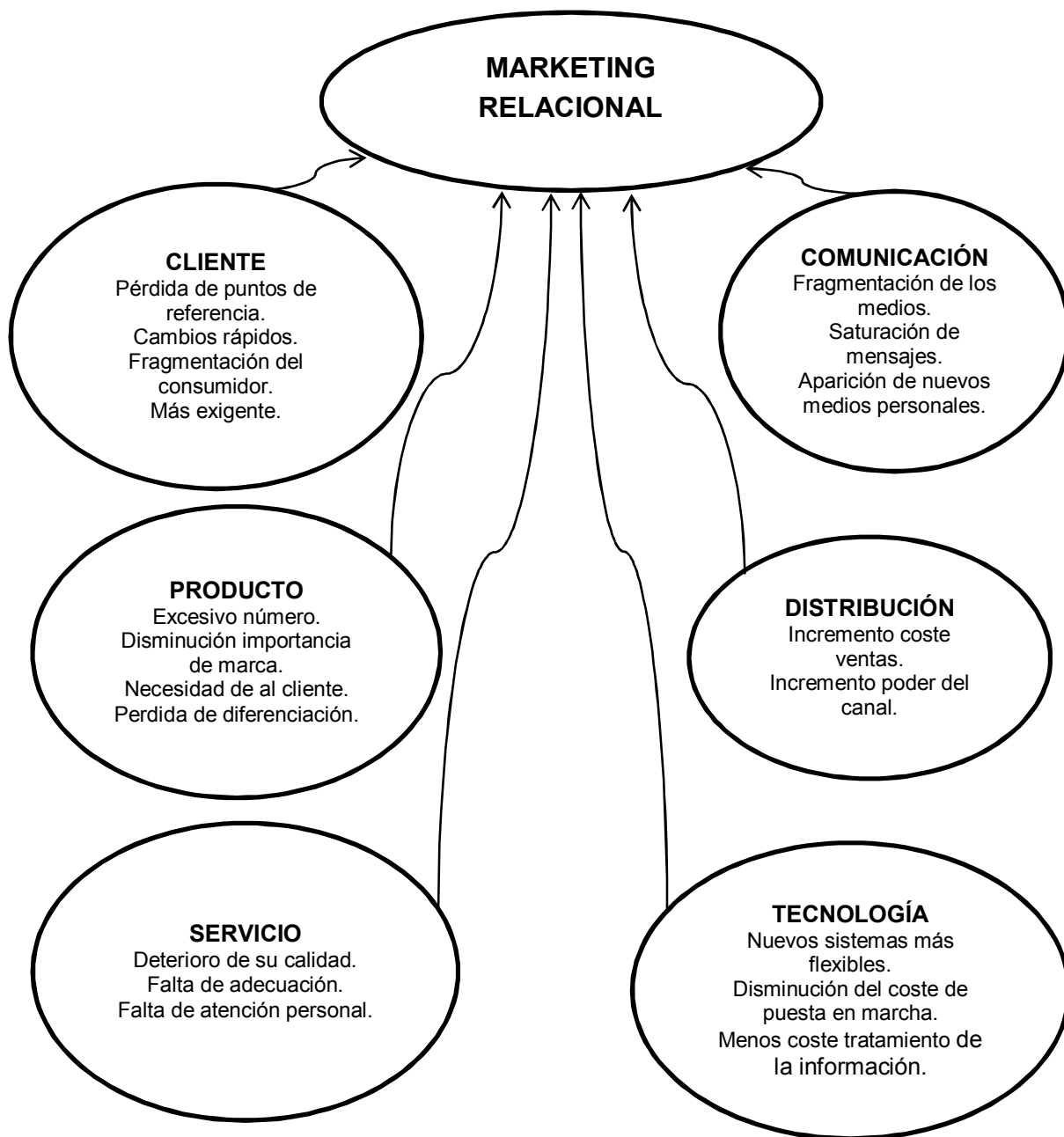


Figura N° 6.1: Factores de la creciente importancia del Marketing Relacional

Fuente: Alet, 1996, Pág. 21

II. Evolución del poder de la marca

Una proporción sustancial de consumidores están perdiendo su vinculación con la marca porque los productos tienden a ser cada vez más similares en calidad y, por lo tanto, pueden ser comprados sobre la base del precio. El factor principal de la falta de diferencia está en que el desarrollo de nuevos productos se ha quedado en simples extensiones de productos.

La disponibilidad de ofertas especiales es tan persuasiva que el cliente ha sido educado para aprovecharlas y romper las barreras de la marca. Además, cuanto más familiar es con una categoría de productos, más probable es que aumente su tendencia a cambiar entre marcas, por la sencilla razón de que ya no le aporta ningún diferencial de garantías importantes.

III. Evolución de los medios

Con el aumento del número de emisoras de televisión y radio, periódicos y revistas, se amplía enormemente la elección de los consumidores. En consecuencia, se pierde la seguridad de la audiencia de cada medio, por constantes traspasos entre distintas opciones. Queda ya lejos, por ejemplo, que prácticamente la mitad del país estuviese viendo un mismo programa.

IV. Avances tecnológicos

La creación de una base de datos, capaz de recoger todas las transacciones con los clientes, resulta ya accesible para muchas empresas. Ha evolucionado el marketing, haciendo posible la recopilación de cualquier información y su rápido tratamiento. El ordenador aporta los datos relevantes para tener éxito, tanto para seleccionar los clientes, como para valorar lo que podemos invertir en cada uno de ellos.

V. Evolución de los sistemas de ventas y de la distribución

Frente a los costos decrecientes de los avances tecnológicos, la gestión de ventas es cada vez más difícil y costosa. El coste de la fuerza de ventas ha

crecido por encima de la inflación en la última década y su eficiencia se ha visto reducida ante la dificultad en alza de llegar a los decisores.

Por otro lado está la complejidad creciente de poder realizar la venta, por la sofisticación de los productos que requieren una mayor formación y atención por parte de los vendedores, y por el incremento del número de productos que deben gestionar ante la diversificación que ha sido utilizada como válvula de escape (sería mejor hablar de trampa) del estancamiento de las líneas de productos existentes.

6.2.1 Enfoque transaccional

En la literatura encontramos diversos autores aportando algún tipo de definición en lo que respecta a marketing. Por ejemplo Philip Kotler (2001), define marketing como “un proceso social a través del cual individuos y grupos obtienen lo que necesitan y lo que desean mediante la creación, oferta y libre intercambio de productos y servicios valioso con otros”.

Ahora bien, en cuanto a una definición gerencial, el marketing a menudo se ha definido como “El arte de vender productos”. Sin embargo, muchos se sorprenden al enterarse de que la parte más importante del marketing no es vender, vender es solo la punta del iceberg del marketing. Peter Druck, un importante teórico en administración, lo expresa de la siguiente manera:

“Suponemos que siempre habrá cierta necesidad de vender, pero el objetivo del marketing es volver superflua la actividad de vender. El propósito del marketing es conocer y entender al cliente tan bien que el producto o servicio se ajuste perfectamente a él y se venda solo. En teoría, el resultado del marketing debe ser un cliente que está listo para comprar. Lo único que se necesita entonces, es poner a su disposición el producto o servicio”⁷.

Pues bien, Lambin (1995) define Marketing, como el proceso de planificar y ejecutar la concepción o diseño del producto, el precio, la información y la

7. Machan, Dyan (1998, En: Soap? cars? what's the difference?, 1998, Vol. 162 N° 5)

distribución de ideas y servicios para generar transacciones que satisfagan tanto los objetivos de las personas como los de las organizaciones.

Dentro de este enfoque transaccional, frecuentemente las empresas parecían actuar según un paradigma cuyas principales características son las siguientes:

- Ni el proveedor ni el cliente tienen memoria. Cada vez que se quiere vender algo, el proceso parte de cero.
- El cliente es anónimo.
- Cada acto de venta tiene que ser rentable por sí mismo.
- Si se pierde a algún cliente, en el mercado se encontrarían abundantes clientes potenciales a los que será fácil captar.
- Es más fácil y barato captar un cliente nuevo que esforzarse por retener a los clientes que ya lo son.
- Fundamentalmente, la empresa habla y el cliente es el que escucha. Se envían mensajes masivos y poco diferenciados.

6.2.2 Del Marketing Transaccional al Marketing Relacional

La relación con el cliente, que se ve menoscaba en el enfoque transaccional, es la piedra angular sobre la que se construye la estrategia de marketing relacional, orientada a crear, mantener y explotar dicha relación. En la Tabla N° 6.1 se realiza una comparación en la evolución de un punto de vista Transaccional al uno Relacional que deja un mayor número de dividendos a las empresas.

Tabla Nº 6.1: Marketing Relacional v/s Marketing Transaccional

Fuente: Alet, 1996, Pág. 39

CRITERIOS	MARKETIN TRANSACCIONAL	MARKETING RELACIONAL
MARKETING-MIX	Tradicional 4Ps	4Ps+Servicio con punto de vista cliente
ENFOQUE	Mercado genérico	Base de clientes
OBJETIVO	Venta puntual	Venta continuada
FACTORES CLAVES	Economías de escala Participación de mercado Resultados por producto	Economías de mercados y de alcance, JIT. Lealtad del cliente Resultados por cliente
CALIDAD	Técnica / Interna	Percibida por el cliente considerando proceso y relación
COMPORTAMIENTO DE COMPRA <ul style="list-style-type: none"> Sensibilidad precio Costos de cambio 	Muy alta Bajos	Baja (diferencia de la relación) Altos (vínculos establecidos)
PRODUCTO/SERVICIO	Diversificación / Extensión de líneas de producto.	Servicios adicionales al cliente Ventas complementarias al cliente
ORGANIZACIÓN MARKETING <ul style="list-style-type: none"> Figura fundamental Papel departamento marketing Función de marketing 	Product Manager Reducido, poca interacción Marketing	Costumer manager y Trade manager Substantial, importancia estratégica Toda la empresa
COMUNICACIÓN	PUBLICIDAD GENERAL	MARKETING DIRECTO
MERCADOS	---Productos de Consumo----- -----Productos de consumo duradero----- -----Productos Industriales----- -----Servicios-----	

6.3 Marketing Relacional

6.3.1 Definición

La literatura proporciona una variada gama de definiciones acerca del marketing de relaciones, comenzando por la definición clásica que establece Philip Kotler.

“Marketing es un proceso social a través del cual individuos y grupos obtienen lo que necesitan y lo que desean mediante la creación, oferta y libre intercambio de productos y servicios valiosos con otros” Kotler (2001).

Ahora bien, entre las definiciones generales que se han propuesto sobre el marketing de relaciones se pueden señalar las siguientes:

“El marketing de relaciones se refiere a todas las actividades de marketing dirigidas al establecimiento, desarrollo y mantenimiento de relaciones de intercambio fructífera con los proveedores, con organizaciones laterales, con los compradores y con los interiores de la propia organización” Hunt y Morgan (1995, En: “The Structure of commitment in Exchange”, Journal of Marketing, Vol. 59, Enero, P.78).

Para Etzel y Walker (2000, En: Stanton W, 2000, Pág. 678): *“El marketing de relaciones es una emergente estructura disciplinar para crear, desarrollar y sostener intercambios de valor entre las partes implicadas, de manera que las relaciones de intercambio evolucionarían para conseguir continuos y estables vinculaciones en la oferta de la cadena”*.

“Marketing relacional es el proceso social y directivo de establecer y cultivar relaciones con los clientes, creando vínculos con beneficios para cada una de las partes incluyendo a vendedores, prescriptores y cada uno de los interlocutores fundamentales para el mantenimiento y explotación de la relación” Alet (1996).

Pues bien, el marketing relacional está basado en varias áreas claves que llevan a alcanzar una relación a largo plazo entre los clientes y la empresa:

- **Identificar y cualificar a los clientes actuales y potenciales**, y actualizar continuamente la base de datos para almacenar información relevante que permita un aprendizaje sobre las necesidades de los clientes actuales y potenciales.
- **Adaptar los programas de marketing**, los productos y servicios para que se adecuen a esas necesidades individuales específicas.
- **Integrar el plan de comunicaciones** dirigidas al consumidor individual para establecer un dialogo efectivo.
- **Controlar y gestionar la relación con cada cliente**, a lo largo de su vida, mejorando su valor para la empresa.

Es debido a lo anterior que Kotler (1996, En: Alet, 1996, Pág. 35) agrega lo siguiente:

....."La transacción es parte de un concepto más amplio, conocido como marketing de relaciones. Los buenos ejecutivos de marketing tratan de construir la confianza del consumidor a largo plazo, a través de mantener buenas relaciones con los clientes, distribuidores, comerciantes y suministradores, por medio de la promesa y del cumplimiento del suministro de alta calidad y de ofrecimiento de un buen servicios y precios razonables a la otra parte a lo largo del tiempo....El marketing de relaciones disminuye los costos y el tiempo de las transacciones y, en algunos casos, permite pasar de la transacción negociada a la simple rutina.

.....el resultado final del marketing de relaciones es alcanzar lo que se denomina como redes de marketing. La red de marketing la constituyen la compañía y las empresas con las cuales se han establecido relaciones sólidas y de confianza mutua en los negocios. De forma progresiva, el marketing está pasando de maximizar los beneficios de cada transacción individual a maximizar el beneficio de las relaciones con otras partes. El principio en que se basa es: construye buenas relaciones y a ello le seguirán transacciones rentables"

6.3.2 Características e importancia del Marketing Relacional

Con respecto a las características podemos citar las siguientes:

- **Interactividad:** el cliente toma cuando quiere la iniciativa del contacto, como receptor y como emisor de comunicaciones.
- **La direccionalidad de las acciones y su correspondiente personalización:** las empresas pueden dirigir mensajes distintos a cada cliente, adecuados precisamente a las circunstancias de ese cliente.
- **La memoria:** el registro en memoria de la identidad, datos, característica, preferencias y detalles de las interacciones anteriormente mantenidas con cada cliente.
- **La receptividad:** las empresas deben hablar menos y escuchar más y así permitir que sea el cliente quien decida si quiere o no mantener una relación, quien define el modo de comunicación y si quiere mantenerla o terminarla.
- **Orientado a sus clientes:** la empresa debe centrarse más en el consumidor, sus necesidades y los procesos que siguen para satisfacerlo. Poner más énfasis en la participación del cliente que en la participación del mercado.
- **B2B y B2C:** el marketing relacional no se aplica solamente a las situaciones business to consumer, sino también de manera más importante a las relaciones business to business y tanto más en la comercialización de servicios que en la productos tangibles.

Ahora bien, su importancia radica en los siguientes puntos que se detallará a continuación:

- **Híper competencia:** el actual contexto comercial llega a constatar que frecuentemente es hasta 5 veces más caro identificar y captar un cliente

nuevo que mantener satisfecho y fiel a un cliente que ya lo es, es decir, la importancia radica en mantener a los clientes.

- **Mercado maduro:** el número de clientes potenciales es limitado, y en algunos casos puede estar incluso disminuyendo.
- **Rentabilidad:** se puede calcular el valor de un cliente satisfecho, teniendo en cuenta su rendimiento previsto anual, multiplicado por el número de años en que se espera que siga siendo cliente. Una base de clientes satisfecho tiene un valor medible.
- **Expectativas:** frecuentemente el valor y rentabilidad de un cliente satisfecho aumenta con el transcurso del tiempo.

6.3.3 Modelo del Marketing Relacional

Este modelo permite el desarrollo de un esquema manejable que defina y oriente los distintos campos de actuación, de manera de poder establecer, mantener y desarrollar las relaciones con los clientes. En el modelo es posible distinguir dos grandes áreas como la señala la Figura N° 6.2.

- Gestión de la base de los clientes.
- Gestión de la lealtad.

Ambas áreas se encuentran apoyadas por herramientas de bases de datos y orientadas por el criterio del valor de la vida del cliente. La base de datos es una herramienta que permite recoger, tratar de forma sistemática y explotar toda la información sobre los clientes de la empresa.

Lo que la empresa intenta es fundar y mantener relaciones con su base de clientes a través de los factores generadores de lealtad, realizando adecuaciones de las diferentes ofertas a las etapas de la relación y la fase de vida propia de cada cliente.

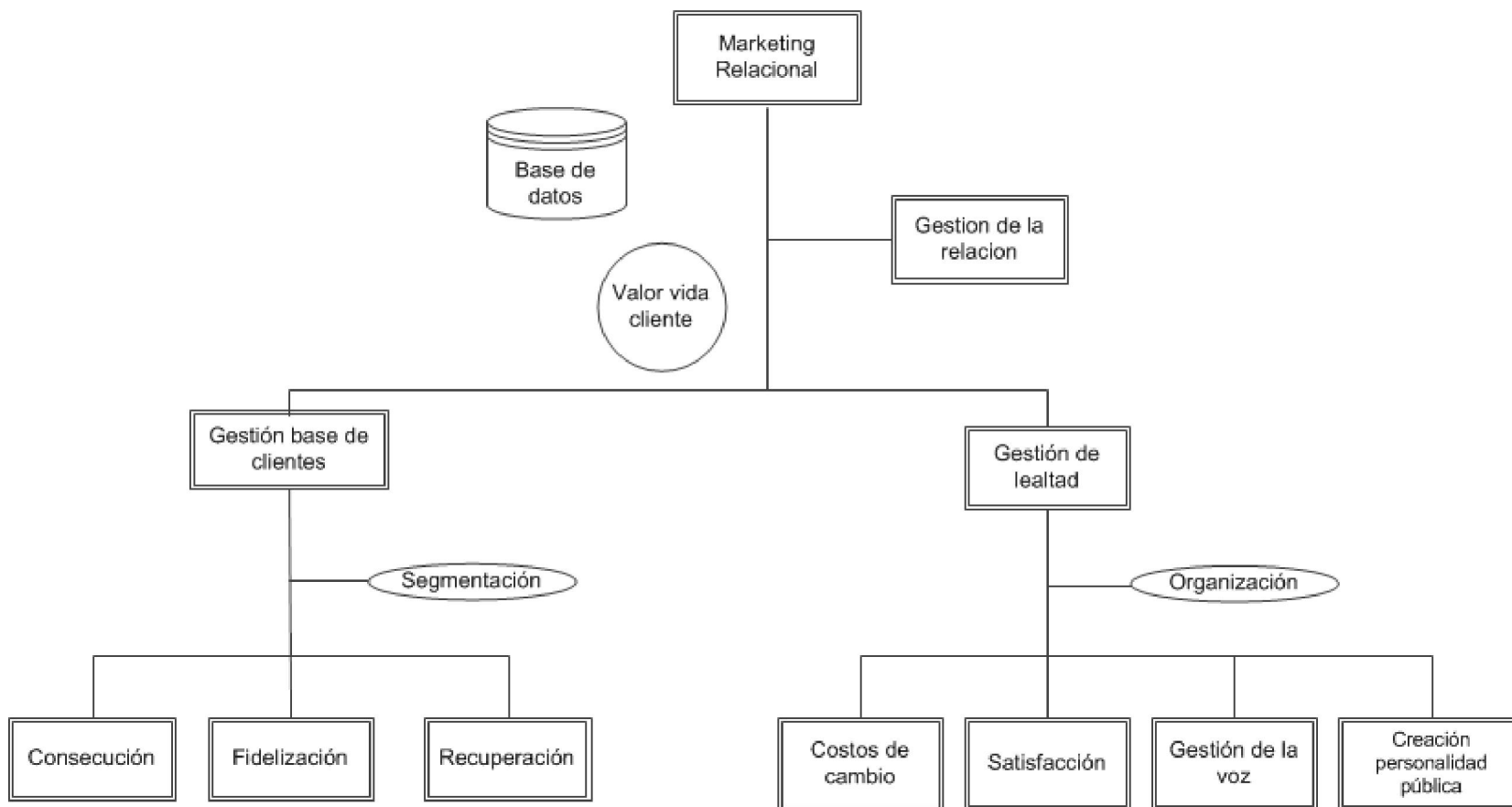


Figura N° 6.2: Modelo de la Estrategia de Marketing Relacional

Fuente: Alet, 1996, Pág. 37

I. Gestión de la base de clientes

Maneja todos los clientes (potenciales, actuales y antiguos) de la empresa, de tal forma que define su mercado y su historia. La gestión se adecua a las diferentes etapas y características de la relación con el cliente, distinguiéndose básicamente tres fases de actuación.

- a) Consecución de clientes.
- b) Fidelización de clientes y explotación de la relación.
- c) Recuperación de clientes perdidos.

Por un lado, es fundamental la consideración de cuáles son los distintos clientes que la empresa debe atender a lo largo de todo el proceso de creación de valor para el cliente. Por otro lado, el éxito depende de la aceptación y el apoyo de cada uno de los socios en cada fase del proceso, desde los trabajadores, pasando por los proveedores, a prescriptores o incluso accionistas.

Otros de los aspectos que deben ser considerados, los cuales permiten mejorar la gestión de la relación con el cliente, son los siguientes:

- Segmentos o grupos de clientes con características comunes que permiten establecer un plan de marketing adecuados a estos.
- Ciclo de vida o etapas de la vida del cliente. Permiten ver y reaccionar a las ofertas de la empresa, por sobre las características de sus respectivos segmentos.
- Fases de la venta: recoge la etapa de una relación concreta entre la empresa y el cliente.

II. Gestión de la lealtad

La lealtad de valorar todos los instrumentos al alcance de la empresa para cuidar la base de clientes, de tal forma que se optimice tanto la duración del cliente en la empresa como su rentabilidad temporal.

La gestión de la lealtad se descompone en:

- a) **Creación de la personalidad pública de la marca:** comunicando las diferencias de valor que aporta respecto a los competidores, y cuidando todos los elementos de imagen que atraen al cliente.
- b) **Incremento de la satisfacción del cliente:** mediante la gestión de las variables clave de la generación de expectativas y resultados, que aporten valor al cliente.
- c) **Aumento de los costos de cambio relativos:** la empresa debe tratar de obtener unos costos de cambio intravendedor considerablemente más bajos que los costos intervendedores.
- d) **Gestión de la voz:** que permita abrir una vía de comunicación fácil, cómoda y eficaz del cliente hacia la empresa.

Estas dos áreas se apoyan en la herramienta de la base de datos, guiadas por el criterio fundamental del valor de vida del cliente.

Antes de entrar en la gestión de la base de clientes, es imprescindible tratar en profundidad la base de datos como herramienta base o soporte para recoger y explotar toda la información sobre los clientes de la empresa.

La empresa trata de fundar y mantener relaciones con su base de clientes a través de los factores de la lealtad, adecuando las distintas ofertas a las etapas de la relación y la fase de vida propia de cada cliente.

6.3.4 Premisas en el desarrollo del Marketing Relacional

a) **La cooperación y la colaboración:** hablar de cooperación implica un compromiso de la empresa con el cliente y el único modo para cumplir es tratarlo individualmente y convertirlo en un aparte integral de este. La colaboración sobreviene cuando la empresa escucha y el cliente habla, cuando la empresa invita a cada cliente a participar activamente en la elaboración concreta del producto o servicio, cuando la empresa reconoce al cliente personalmente y le transmite cuál es su importancia.

b) **El valor Vitalicio del cliente y la participación por cliente:**

- **El valor vitalicio del cliente (VVC):** el objetivo final que cada empresa busca es maximizar el valor vitalicio del cliente (VVC), que está en función de sus compras futuras; es decir la estimación del margen de contribución de cada cliente durante el periodo de tiempo que se mantiene como tal. El VVC, depende de su vida media y de su contribución anual media. La clave en el VVC es el tratar a los clientes como el activo principal de la empresa.

Ahora bien, existen tres formas básicas para aumentar el valor vitalicio de los clientes (VVC):

a. **Menor costo de captación por cliente** (tanto si son nuevos clientes o si antes ya lo hayan sido).

b. **Buscar incrementar las ventas por cliente:**

- i. Por incremento de la vida media del cliente.
- ii. Por ventas cruzadas.
- iii. Por ventas complementarias.
- iv. Por incremento en la frecuencia de compra.
- v. Disminuir los costos por cliente (a lo largo de su vida como cliente).

- **La participación de cada cliente:** la información individual es fundamental para poder conocer las peculiaridades de cada cliente. Las estrategias diseñadas en base a un sistema de información que poseen las empresas son básicas a la hora de identificar y cualificar a los clientes actuales y potenciales: su comportamiento histórico de compra en relación con la empresa y los competidores, sus características personales (demográficas, estilos de vida) y para identificar el tipo de relación que establecen con la empresa y la respuesta a las acciones de esta.

- c) **La percepción de utilidad del cliente:** su percepción es la única forma de llegar a conocerlo. Además la empresa debe percibir que la transmisión de información particular le aportará beneficios financieros y especialmente afectivos y emocionales.

- d) **Dialogar con cada cliente:** las relaciones entre las personas están basadas en la comunicación y el dialogo, de un modo personalizado y fluido en ambas direcciones, posibilitada por la tecnología digital y multimedia. Sin la aparición de los nuevos medios de comunicación no se podría hablar del Marketing Relacional como la principal tendencia de los próximos años.

- e) **Gestionar la relación de forma personalizada:** habitualmente los clientes pasan distintas etapas en su relación con la empresa y son:
 - **Cliente posible:** cumplimiento de características iniciales que reflejan la eventualidad de convertirse en clientes.
 - **Cliente potencial:** cliente posible en un plazo corto de tiempo, al cumplir características o interés que lo hacen prever así.
 - **Comprador:** ha realizado una primera adquisición a la empresa.
 - **Cliente:** la repetición de su compra demuestra una consolidación en la relación entre empresa y comprador.
 - **Socio o cliente asociado:** el cliente que apoya a la empresa y se siente implicado con ella. Normalmente actuará como prescriptor y recomendado.

Cada una de las fases en relación con el cliente implica un mayor o menor grado de conocimiento de sus particularidades y una adaptación diferenciada de la oferta comercial de la empresa. Esta adaptación viene determinada por el conocimiento que se tenga del cliente, pero también por la importancia que tenga para la empresa. Basándose en el principio de Pareto, este afirma que el 80% de los ingresos de la empresa son generados por el 20% de los clientes, y es por tanto la misión de las empresas el de conocer a este 20% y tratarlos de forma privilegiada y diferenciada.

6.3.5 La estrategia del Marketing Relacional

La estrategia consiste en determinar un sistema de objetivos, políticas y planes de acción, claramente especificados, que configuran una orientación, como la movilización de recursos para aprovechar las oportunidades identificadas y disminuir los riesgos futuros. Por ello la estrategia nos permite concentrarnos en dos áreas básicas esenciales:

- a) La decisión de inversión en producto/mercados que define el alcance de la estrategia en términos de productos y mercados, y la asignación de las inversiones en ellos.
- b) El desarrollo de una ventaja sostenible para competir en dichos mercados que determina la forma de competir en función de las capacidades primordiales, los objetivos marcados y los recursos asignados.

La obtención de una ventaja competitiva señala Michael Porter "...surge del valor que la empresa es capaz de crear para sus compradores que excede el coste de crearlos. El valor es lo que los compradores están dispuestos a pagar, y el valor superior se obtiene por la oferta de precios inferiores que los competidores para beneficios equivalentes, o de proveer beneficios únicos que compensan más que los precios superiores".

La ventaja competitiva de acuerdo con Porter, ha sido clasificada hasta el momento en tres tipos.

- a) **Menor costo:** esta estrategia va asociada con estrategias tendientes a alcanzar un volumen elevado por obtención de economías de escala, avances en la curva de experiencia y economías de alcance.
- b) **Estrategia diferencia:** intenta adoptar un mayor número de empresas aunque muchas de ellas caen en una posición poco clara y, por lo tanto, débil.
- c) **Ocupación de un nicho protegido:** son los que hace referencia P. Kotler y son:
 - i. De suficiente tamaño y capacidad de compra para ser rentable.
 - ii. Con capacidad de compra
 - iii. De poco interés para sus competidores.
 - iv. La empresa tiene que tener las capacidades y recursos para servir al nicho efectivamente.
 - v. La empresa debe constituir un fondo de comercio de clientes suficiente para defenderse contra cualquier competidor importante.

La estrategia de Marketing Relacional posee fundamentos distintos y un enfoque diferencial respecto a la estrategia de nicho, permitiendo alcanzar la ventaja competitiva sostenible: vínculos de unión y rapidez de reacción. Estos vínculos reflejan la lealtad del cliente con la empresa, apoyados en una mayor satisfacción así como también en costos de cambios adicionales para el cliente, logrando además protegerse de la entrada de nuevos competidores.

Al tener un mayor conocimiento del cliente, la empresa puede reaccionar mejor y evitar que un nicho desaparezca. Por medio de una comunicación bidireccional es posible recopilar información de los clientes y detectar de forma rápida cualquier cambio que se pueda estar produciendo, tanto en sus necesidades o en algunas insatisfacciones y a través de la comunicación continuada será posible ir adecuando de forma continua los productos y expectativas.

Las estrategias de Marketing Relacional apuntan a alcanzar niveles mayores de intimidad con el cliente de modo de entregar un producto o servicio adaptado a sus necesidades y que estos a su vez sean entregados sin inconvenientes y a precios competitivos. Es así, que al aplicar una estrategia relacional debe hacerse en el contexto del tipo de relación con el o los clientes:

- 1) **Frecuencia potencial de las transacciones:** la cual viene dada tanto por la duración de los beneficios y lo complejo del proceso de decisión como del sistema de entrega del producto o servicios.
- 2) **Duración e intensidad del proceso de entrega del producto o servicio:** pone énfasis en que la entrega de valor puede ser tanto o más importante que la entrega misma del producto o servicio, por lo que se hace necesario un enfoque personalizado con el cual se transmita confianza.

A continuación se señalarán los tipos de estrategias en el campo del Marketing Relacional

Citamos a Ted Levitt (1995, En: Alet, 1996, Pág. 41) que nos dice lo siguiente:

“Una estrategia que no haga referencia explícita a los clientes y su entorno competitivo dejará, con toda seguridad, de generar y mantener un nivel satisfactorio de clientes y de conciencia competitiva de su empresa, en especial en los importantes rincones en los que se lleva a cabo el trabajo real.”

La estrategia de marketing relacional trata de gestionar la base de clientes de tal forma de que se puedan identificar oportunidades de negocio y de explotárlas de forma eficiente, extrayendo su máximo potencial. En la estrategia de marketing relacional, los clientes son vistos como el auténtico recurso escaso que se tienen que optimizar.

La empresa puede elegir entre distintas alternativas de estrategia distinguiendo entre la que centran su atención en la participación de mercado frente a las que atienden a la base de clientes. La atención de la participación de mercado está vinculada con la consecución de nuevos clientes, tanto como aumento del mercado, como por el aumento de la misma participación.

Tradicionalmente la estrategia de la empresa se orienta hacia un planteamiento ofensivo para alcanzar nuevos mercados, o el aumento de su participación dentro de ellos. El resultado es que las empresas dedican mucho esfuerzo a la adquisición de nuevos clientes que a la retención de estos. En el caso de la estrategia defensiva, se orienta a la reducción de la pérdida de clientes y de cambio de marcas. El objetivo de la estrategia defensiva es la maximización de la lealtad o retención de clientes y se compone según muestra la Figura N° 6.3.

- Aumentar la satisfacción del cliente.
- Construir costos de cambio.
- Favorecer la voz del cliente.
- Conseguir una personalidad pública más fuerte.

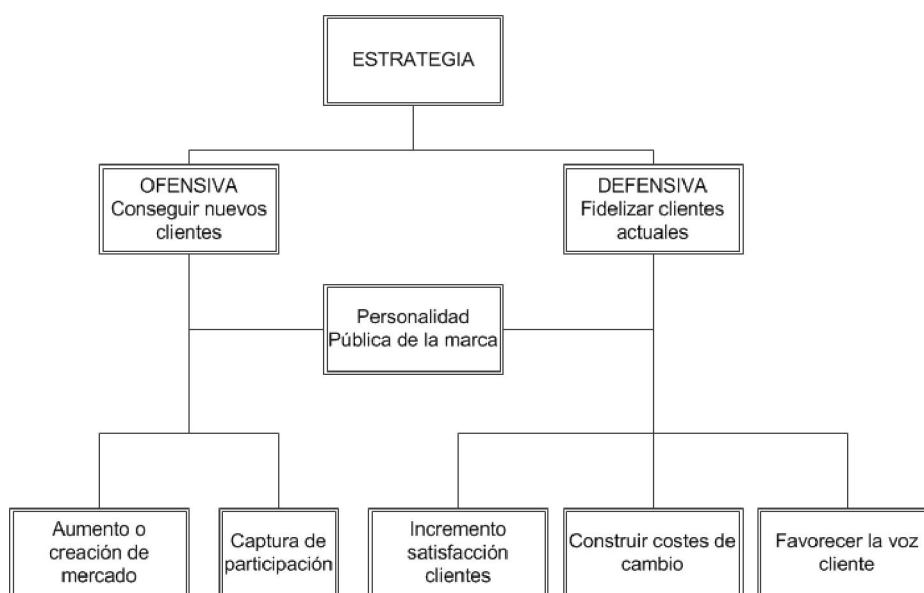


Figura N° 6.3: Alternativas Estratégicas

Fuente Alet, 1996, Pág. 50

I. Estrategia basada en el valor del cliente

La creación de valor para el cliente es uno de los principios que debería seguir toda la empresa para la obtención de rentabilidad y estabilidad, es decir, su propio valor. El beneficio del empresario es más que la suma de los beneficios de cada uno de sus clientes.

Por lo tanto, el enfoque central será el cliente para así comprender la magnitud de las relaciones económicas de la empresa, dentro del marco del marketing relacional, donde se conocen los clientes potenciales, actuales y antiguos.

La estrategia de marketing relacional busca la obtención de la rentabilidad actual y futura de la empresa a través de la satisfacción superior de sus clientes, construyendo relaciones duraderas con cada uno de ellos, las que permitan generar un valor añadido específico para cada uno de ellos e individualmente. Este valor surge de una combinación de producto, comunicación y servicio personalizado, al considerar, de hecho, a cada cliente como un segmento de mercado unipersonal, de tal manera que:

“La segmentación unipersonal es un concepto estratégico, y no únicamente de marketing, pues afecta a aspectos de la organización y, sobre todo, supone un cambio en la filosofía que rige la relación de la empresa con el mercado. La segmentación unipersonal permite la creación de una ventaja competitiva duradera frente a los competidores tradicionales”.

II. Conocimiento de la base de clientes

Realizar una estrategia de marketing relacional pasa necesariamente por el desarrollo de una base de datos de clientes, que nos aportará información fundamental para conocer datos de cada uno de ellos.

El salto cualitativo de los datos relacionados entre sí, permite sacar conclusiones relevantes y una información que se convierte en un conocimiento

especialmente importante sobre los clientes. El propósito de la base de datos es en esencia, hacer posible establecer una relación personal entre la empresa y cada cliente. De su análisis extraerá la empresa conclusiones fundamentales sobre el diseño de productos y procesos y podrá adaptarse a las características más significativas de los grupos identificados como diferentes y con potencial de negocio. Sin duda alguna, la base de datos aporta la gran diferencia entre el marketing transaccional y el relacional.

Un sistema integrado de marketing apoyado en una base de datos es una inversión estratégica que debe ser apoyada y enriquecida continuamente. El conocimiento que se puede obtener de los clientes a través de la explotación de la base de datos es tal que puede aportar una ventaja competitiva sostenible. Además, la base de datos es el fundamento para la planificación, ejecución y medición de las acciones de marketing. Sobre ella se realiza la segmentación, la localización de nuevos clientes, el diseño de nuevos productos y la mínima previsión de ventas.

A. Calidad de la base de datos

La calidad de la información de la base de datos debe ser cuidada en forma especial debido a la gran importancia que tiene en los resultados obtenidos. La calidad se asegura a lo largo de cada una de las fases o procesos siguientes: introducción, normalización, de duplicación y actualización.

La falta de calidad trae importantes trastornos a la gestión de la base de clientes y la comunicación con cada cliente. Entre los más habituales están: Duplicación, eliminación de clientes, Falta de datos, errores en los datos.

B. La segmentación por comportamiento de compra

La gran ventaja de trabajar con una base de datos es la posibilidad de segmentar a los clientes de una manera óptima. Con la información contenida en la base de datos se produce un salto importante al pasar de apoyarse en

predictores del comportamiento (deseo, intención, imagen) al comportamiento de compra en sí mismo. Se valora la capacidad de compra futura en función de cuánto y cuándo nos ha comprado, dónde, como, y por qué. Lo anterior se resume en la Figura N° 6.4.

TIPO DE CLIENTE

- Estilo de vida.
- Etapa de la vida.
- Lugar de residencia.

COMPORTEAMIENTO DE COMPRA

- Presencia.
- Frecuencia de compra.
- Valor de las compras.
- Tipos de producto.



- Capacidad de crédito
- Acontecimientos.
- Grado satisfacción.
- Reactividad

- Lealtad a la empresa.
- Sistema de compra.
- Canal de compra.

Figura N° 6.4: Dimensiones del cliente

Fuente: Alet, 1996 Pág. 81

6.3.6 Lealtad

Alet (1996), señala que *“la lealtad es una medida de la vinculación del cliente a la marca o empresa y refleja la probabilidad de que un cliente cambie a otra marca, en especial cuando se modifica alguna característica o funcionamiento o precio”*.

¿De qué manera se puede afectar la lealtad?; a través del comportamiento de compra del cliente, es decir de la sensibilidad a la marca. La sensibilidad a la marca según Alet (1996), *“es una variable psicológica que indica el papel que juega la marca en el proceso que precede a la compra”*.

Las variables que afectan a la decisión de compra están relacionada, por el contrario, con la situación específica del producto/mercado y el conocimiento del cliente en dicho contexto de decisión. Así, las variables importantes son las siguientes:

- a) **Criterios de elección de los productos:** en la medida que el cliente exige niveles más altos de calidad más sensible es a la marca.
- b) **Creencias de existencia de diferencias entre las distintas marcas:** si no hay diferencias no tiene sentido el apego a una o varias marcas.
- c) **Competencia percibida del mismo comprado en la elección:** el desconocimiento por parte del cliente elimina la posibilidad de identificación y referencia de la marca.
- d) **Grado de implicación del consumidor:** varía en la medida que el cliente se muestre interesado cada vez más en los productos o servicios.

¿Cómo una empresa puede obtener una mayor lealtad por parte de sus clientes? La respuesta a la pregunta, se encuentra en la Figura N° 6.5 y radica en conseguir cuatro objetivos que son: Personalidad pública de la marca, satisfacción del cliente, costos de cambio y voz.

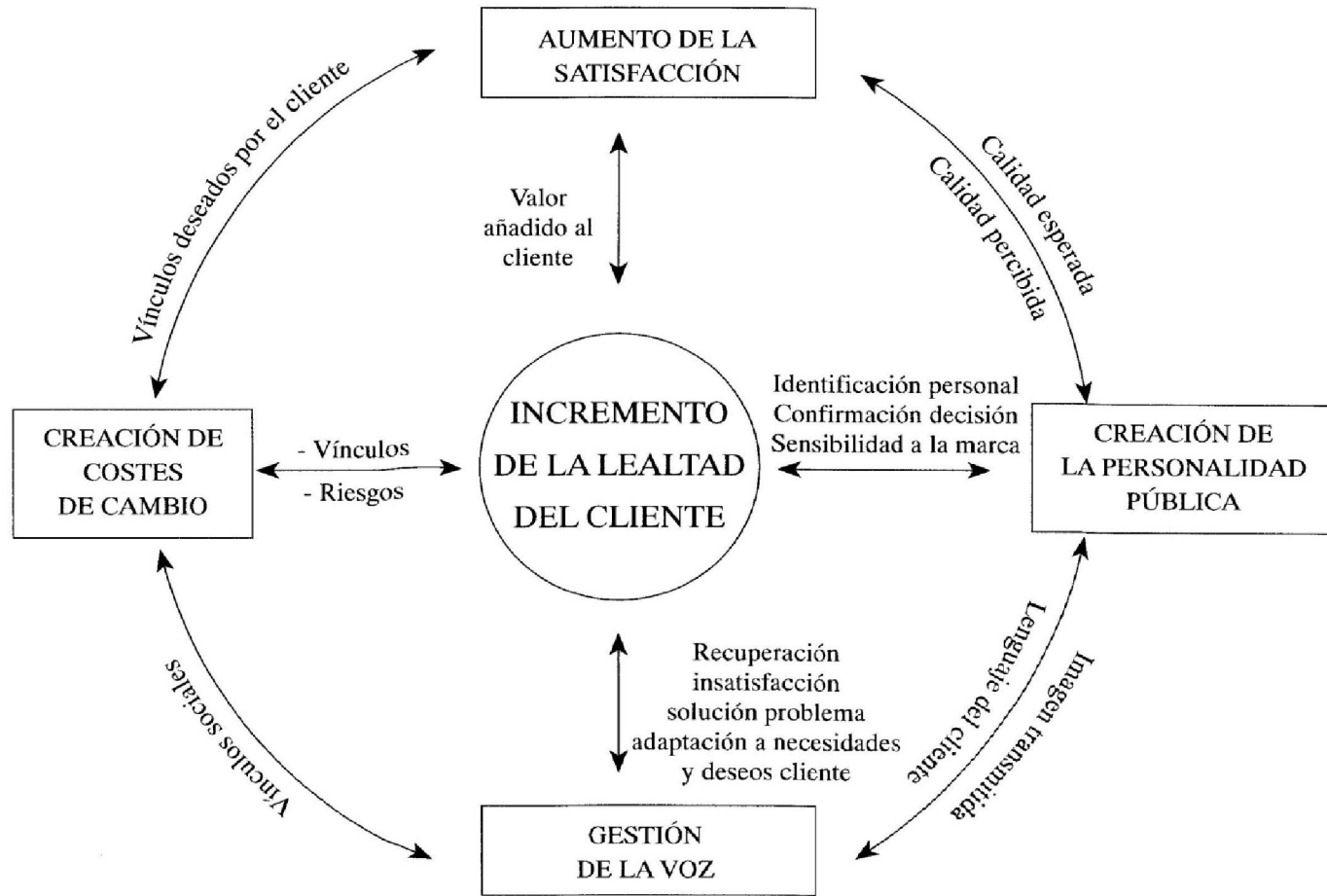


Figura N° 6.5: Pilares para aumentar la lealtad de sus clientes

Fuente: Alet, 1996, Pág. 87

I. Personalidad Pública de la marca

Según Alet (1996), “la personalidad pública plantea la creación del conjunto de representaciones deseadas que una persona o grupo de personas ha de asociar a una marca, u organización”. Es decir, la personalidad pública de la marca, recoge los atributos deseados de la imagen de la marca, forma coherente, apropiada y atractiva, para triunfar en el mercado en la consecución y lealtad de los clientes”.

A medida que el consumidor se ve saturado de mensaje, este comienza a filtrarlos, influyendo netamente sus percepciones a la hora de decidir cuáles son los mensajes realmente recibidos. Para Alet (1996), la creación de la personalidad pública aporta sensibilidad a la marca ya que:

- i. **Dota de credibilidad** en la comunicación de las diferencias de la marca.
- ii. **Aporta confianza** al cliente en situaciones de riesgo.
- iii. **Traslada la personalidad privada** del consumidor a la personalidad pública de la marca como proyección personal adecuada.
- iv. **Despierta el interés** del cliente por los atributos de la marca y los beneficios que puede obtener.
- v. **Aporta una información útil** y coherente que puede proporcionar al cliente competencia suficiente para decidir en base a la marca que recoge los atributos básicos de sus deseos.

En resumen, la personalidad pública de la marca, crea una patente de atributos principales que no puede ser tomada por los competidores y a su vez le sirve como barrera de entrada protectora de la lealtad de sus clientes.

II. Satisfacción del Cliente

Si las empresas no dimensionan la real importancia a la satisfacción de sus clientes, jamás podrán alcanzar clientes leales.

Alet (1996), señala que *“la satisfacción del cliente es el resultado de la valorización que realiza sobre la calidad percibida en el servicio o producto entregado por la empresa”*. Además Kotler (2001) señala que la satisfacción del cliente *“es el grado en que el desempeño percibido de un producto concuerda con las expectativas del comprador”*.

Un factor importante a considerar es lo difícil que resulta medir la calidad del servicio en las empresas. En la Tabla N° 6.2 Alet (1996) indica cuales son:

Tabla N° 6.2: Criterios de evaluación según los clientes

Fuente: Alet, 1996, Pág. 98

Dimensión	Definición
Fiabilidad	- Capacidad de realizar el servicio prometido correctamente.
Sensibilidad	- Deseo de ayudar a los clientes y de proveer un servicio rápido
Tangibles	- Apariencia de las instalaciones físicas, equipo, personal y materiales de comunicación.
Competencia	- Posesión de las capacidades requeridas y el conocimiento para realizar el servicio.
Cortesía	- Educación, respeto, consideración y amabilidad del personal de contacto.
Credibilidad	- De confianza, honestidad credibilidad del proveedor de servicio.
Seguridad	- Sin riesgo, peligro o duda.
Acceso	- Facilidad de contacto y de aproximación.
Comunicación	- Mantener a los clientes informados en un lenguaje que puedan entender, y escucharlos.
Comprensión del cliente	- Hacer el esfuerzo de conocer a los clientes y sus necesidades

Junto al servicio al cliente, la calidad y el marketing juegan un papel importante en la entrega de valor. Es así que existe una estrecha relación entre estos tres factores de manera que el marketing relacional logre su gran objetivo que es “Obtener clientes leales”.

La entrega de un servicio superior, puede llegar a significar una ventaja competitiva con respecto al resto de las empresas. La calidad también permite diferenciarse del resto, ya que es el medio que le permite mantener por largos períodos de tiempo su posesión. Según Gutiérrez (1997), *“la calidad la define el cliente, es el juicio que éste tiene sobre un producto o servicio y resulta por lo general en la aprobación o rechazo del producto.”* Alet (1996) es aún más preciso y señala que calidad del servicio es *“la percepción que tienen un cliente acerca de la correspondencia entre el desempeño y las expectativas, relacionados con el conjunto de elementos secundarios, cuantitativos y cualitativos, de un producto o servicio principal”*.

Dado lo anterior, resulta conveniente investigar acerca de las expectativas que un cliente puede tener y como deben ser analizadas. El primer escalón radica “en el nivel deseado” es decir el que quiere recibir y el segundo escalón es “el nivel adecuado” que es una valoración del nivel percibido.

Realizar una buena gestión dada las expectativas del cliente resulta en ventas que según Alet (1996), son:

- 1) **Controlar sus actividades** de comunicación de marketing y así asegurarse que están vendiendo lo correcto y no están dando promesas que no podrán ser cumplidas.
- 2) **Elegir a sus intermediarios** como distribuidores, agentes los que son controlados estrechamente de modo de asegurar que crean las expectativas correctas.

- 3) Controlar el rendimiento** versus las expectativas, es mejor prometer menos que no estar seguro de poder cumplir y así ir entregando más de lo prometido.

Para que una empresa sea capaz de obtener un cliente leal, es necesario que esta maneje los atributos valorados en la calidad del servicio y su enfoque dirigido a lo que realmente es valioso para el cliente. Resulta necesario por tanto, estar en permanente contacto con una relación profunda, conocer en detalle las necesidades actuales y potenciales de los clientes, de modo de adecuarse de manera eficiente a sus preferencias y valores. Los clientes que están altamente satisfechos son menos sensibles a los precios, siguen siendo de la empresa por más tiempo y hablan a favor de esta y sus productos.

Dentro de un análisis en el que se busca entregar un paquete de valor al cliente surge el concepto de producto total. Para los clientes, el producto es un conjunto complejo de satisfacciones de valor. Alet (1996), define producto total en base a un desglose en cuatro niveles que son:

- i. **Producto genérico:** es la cosa rudimentaria sustantiva sin la cual no existe la posibilidad de entrar al juego de la participación del mercado.
- ii. **Producto esperado:** representa las expectativas mínimas del cliente. Incorpora componentes de servicio, recambios, accesorios, instalación o postventa.
- iii. **Producto aumentado:** es la oferta que aporta beneficios inesperados o diferenciales, que distinguen a la empresa de la competencia.
- iv. **Producto potencial:** se refiere a todo lo que todavía queda por hacer, lo que es posible efectuar con potencial factible de atraer y mantener clientes.

Es así que el concepto de producto total, permite reconciliar la perspectiva del producto que tiene la empresa con la perspectiva del cliente acerca de la entrega de valor que recibe.

III. Costos de cambios

Estos sirven de soporte a la lealtad, debido a que hacen más costosa la salida de un cliente o el cambio a otra empresa de la competencia. Si el cambio resulta muy costoso para el cliente queda en evidencia que la reducción de la base de clientes será menor, por lo que la empresa debe tenerlos en consideración de modo que sus clientes permanezcan por más tiempo con ella.

Existen múltiples tipos de cambios, donde están los propios del cliente y los vinculados al tipo de producto/servicio.

A. Costos de cambio personales del cliente

- **Costos emocionales:** están dados por la vinculación del cliente con la empresa o por la tradición de la empresa a lo largo del tiempo.
- **Hábitos del cliente:** aparece un costo de cambio cuando se presenta un producto alternativo que implica una modificación en el comportamiento del individuo y en el de la empresa.
- **Esfuerzos de aprendizaje necesarios:** son los costos personales que implican los estudios de nuevos productos, en términos de horas y energía de los clientes.
- **Ventajas económicas asociadas a la lealtad:** este es el más evidente del Marketing Relacional, ya que establece la lealtad a través de un condicionamiento de las ventajas económicas que el cliente obtiene por mantener una relación a lo largo del tiempo con la empresa.
- **Riesgos sociales y psicológicos:** en cualquier situación de cambio existen temores y sensaciones de riesgo por miedo a perder algo o que algo no salga como se esperaba.

- **Cambios del papel del usuario:** corresponder a la percepción propia del cliente, en la valoración del cambio que significaría si es que cambia de empresa.
- **Contratos establecidos con empresas de productos o servicios complementarios o relacionados:** el cliente puede tener un compromiso indirecto con una empresa, lo que le implicaría altos costos de cambio.

B. Costos de cambio asociados al producto

- **Costos de búsqueda:** en la identificación y cualificación de los proveedores alternativos.
- **Costos de rediseño o reformulación:** para poder incorporar el producto/servicio en el proceso de producción o consumo.
- **Costos de formación o de aprendizaje nuevo:** incluye los costos de formación propios, además de los tiempos muertos o perdidos y los costos de mala calidad en el proceso de aprendizaje.
- **Costos de cambios de papeles de algunos miembros de la organización:** pueden surgir resistencia al cambio lo que significaría costos adicionales de reestructuración.
- **Riesgo de fracaso:** si el cambio no funciona bien puede implicar pérdidas en el mercado.
- **Inversiones en equipo relacionado:** el nuevo producto puede no requerir de las maquinas o herramientas que se utilizaban antes, lo que significaría la adquisición de nuevos equipos.
- **Costos de marcha atrás en el recambio:** si el cambio no funciona adecuadamente, se han de reconocer los costos implícitos en la

vuelta al producto y sistemas anteriores, que aumentaron probablemente con el paso del tiempo.

- **Costos contractuales:** muchas empresas mantienen relaciones con los clientes vinculados a contratos, lo que significa dar cumplimiento a tales, de lo contrario deberá hacerse cargo de lo que significa romper un acuerdo contractual.

Alet (1996), señala que los costos de cambio tienen una importancia fundamental en la lealtad por cuanto determinan la actitud inicial ante la probabilidad de búsqueda o valoración de alternativas.

¿Cuáles son los tipos de vínculos para crear lazos con los clientes?, existen tres tipos de vínculos que la empresa puede establecer y son: financieros, sociales y estructurales.

- i. **Vínculos Financieros:** determinan el primer nivel de los vínculos, los basados en ventajas económicas referidas a descuentos especiales en precio, o pro premio por volumen de compra acumulado. Por lo general contemplan un plan de retención de clientes, donde se hace un reconocimiento especial del cliente con programas de compra o uso frecuente.
- ii. **Vínculos Sociales:** Estos vínculos surgen de una relación cuidada, personalizada, con detalles afectivos de reconocimiento honesto y sincero. Más que nada lo que la empresa debe hacer es fortalecer la relación que mantiene con sus clientes de manera que se mantenga enterados de cuáles son sus necesidades y deseos individuales y así personalizar sus productos y servicios.
- iii. **Vínculos estructurales:** son el último nivel, el que garantiza el mayor nivel de compromiso y lealtad, donde la empresa y el cliente adaptan sus sistemas o su organización para desarrollar la relación.

IV. Voz del cliente

Con los cambios que se han producido en las sociedades, hoy en día las empresas se enfrentan a consumidores más educados, con menos miedo de decir lo que piensa y de quejarse si el producto o servicios que está percibiendo no es lo que realmente esperaba.

En general las quejas son vistas negativamente porque conforman esfuerzos negativos, reportan castigos psicológicos a los miembros implicados de la organización y muestran, en definitivas, los puntos débiles a los errores cometidos en la gestión por la empresa.

Ante una situación de insatisfacción existen distintos tipos de respuesta. La primera es la voz, que es una respuesta directa a la empresa; luego privado, que es un boca en boca negativo; y la tercera son respuestas hacia otras entidades como asociaciones de consumidores.

Además, se detectan cuatro tipos de clientes según Alet (1996) que son:

- **Pasivos:** que se quejan por debajo de la media y por lo general son jóvenes.
- **Voceadores:** que son los de mayor porcentaje, los más interesantes para la empresa, ya que se quejan directamente con esta y en menor cantidad con privados y terceras personas.
- **Irritados:** sus quejan son mayoritariamente privadas, aunque de igual manera se quejan directamente con la empresa.
- **Activistas:** que se quejan por sobre la media en las tres dimensiones.

En definitiva, escuchar a los clientes permite detectar si la empresa presenta algún problema en su funcionamiento, y así conocer cuáles son las variables de satisfacción del cliente más importantes para él, como también la

obtención de ideas para mejorar el funcionamiento y tener una valoración mayor de los empleados y distintos componentes del producto/servicios.

6.3.7 Plan de Marketing Relacional Eficaz

A través de un plan de marketing Relacional, la empresa puede obtener la creación de valor dentro de un marco de crecimiento con rentabilidad. Básicamente, lo que plantea este tipo de plan es el establecimiento de relaciones continuada a lo largo del tiempo, de manera de establecer y ampliar las relaciones con los clientes.

En la Figura N° 6.6, se presentan dos fases de acción que se concretan en las dos áreas de marketing básicas en crear valor para la empresa:

- Front-End: conversión de no clientes en clientes al menor costo.
- Back-End: mantenimiento y valorización de los clientes actuales.

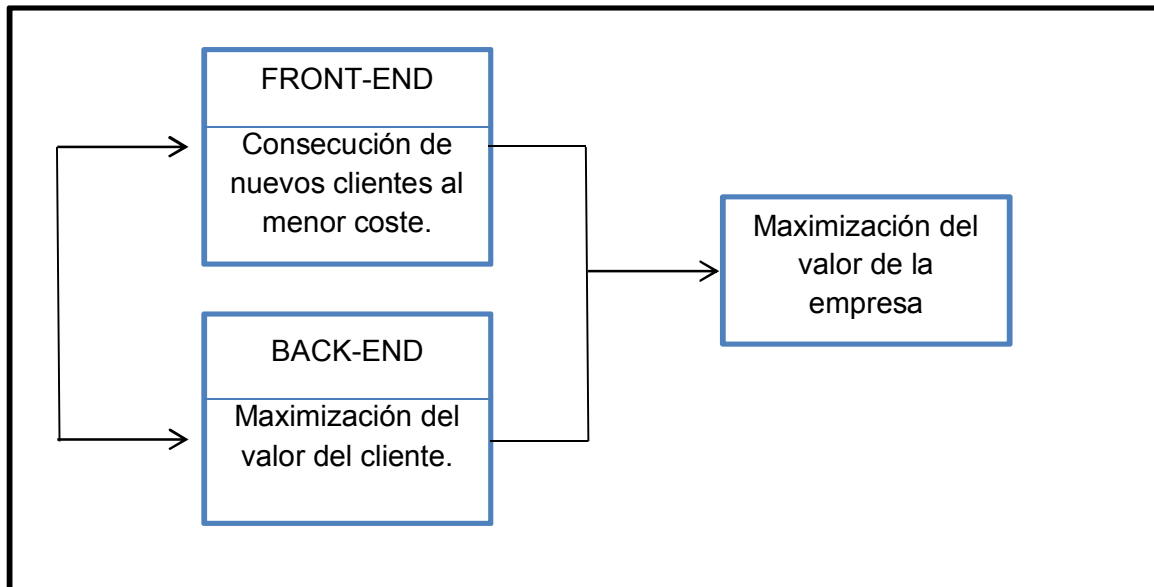


Figura N° 6.6: Áreas básicas de Creación de Valor

Fuente: Alet, 1996, Pág. 137

Es así, que en el cuidado y explotación de estas dos áreas y en sus misma interrelación donde podemos conseguir aumentar el valor del cliente y, consiguientemente de la empresa.

El valor de la empresa se aumenta por tres grandes bloques de acción:

1. Consecución de clientes al menor costo:

- a) Desarrollo de las fuentes más eficientes de obtención de nuevos clientes.
- b) Recuperación de clientes antiguos deseables.

2. Aumento de las ventas por cliente:

a) Aumento de la vida media del cliente en la empresa, a partir de:

- Mayor satisfacción del cliente.
- Creación de costos de cambio o barreras de salida.
- Recuperación de clientes insatisfechos.

b) Realización de ventas cruzadas:

- Desarrollo de servicios de valor añadido a media del cliente.

c) Obtención de ventas complementarias

d) Aumento de la frecuencia de compra.

3. Disminución de los costos por cliente:

a) Disminución de costos comerciales

- Segmentación adecuada
- Adecuación de la oferta
- Determinación de la frecuencia óptima de envíos.
- Integración de los medios utilizados.
- Implicación del cliente en la realización de los productos y servicios.

b) Aumento de la conversión de demanda en ventas efectivas

- Mejora de procesos de generación y gestión de leads.
- Servicio rápido y satisfactorio.
- Reducción de la tasa de pérdida en ofertas de continuidad.
- Disminución de devoluciones.

c) Gestión eficiente:

- En creación y producción de las campañas comerciales., en Fulfillment⁸.
- Con estructura simple, coordinada.
- Integración de los programas comerciales de cada medio, canal, y tarea comercial, a partir de una base de datos común.

Todo este conjunto de campos de acción se concentran en un Plan de Marketing dirigido a la consecución de los objetivos fijados dentro de la estrategia general de la empresa, que queda reflejado en la Figura N° 6.7. Este plan, recoge todos los detalles del Plan de Marketing que afectan drásticamente a los resultados de la empresa y que, bien realizados, llevan a un incremento sustancial del valor de la empresa.

Con respecto al Plan de Marketing, el dilema se traslada a invertir más por un nuevo cliente, o aportar un mayor valor a éste y dependerá de la situación competitiva y los objetivos económicos y financieros fijados por la empresa.

Alet (1996) señala que: en el marco de una visión dinámica, la empresa debe concentrarse en aumentar el valor del cliente para entrar en el “Círculo de Confianza”, que enfoca las funciones en las dos fases de acción de Marketing: conseguir nuevos clientes y aumento del valor del cliente. Lo anterior se señala en la Figura N° 6.8.

8. Etapa entre un cliente y un proveedor que se da cuando este último debe cumplir en tiempo y forma con todo lo prometido.

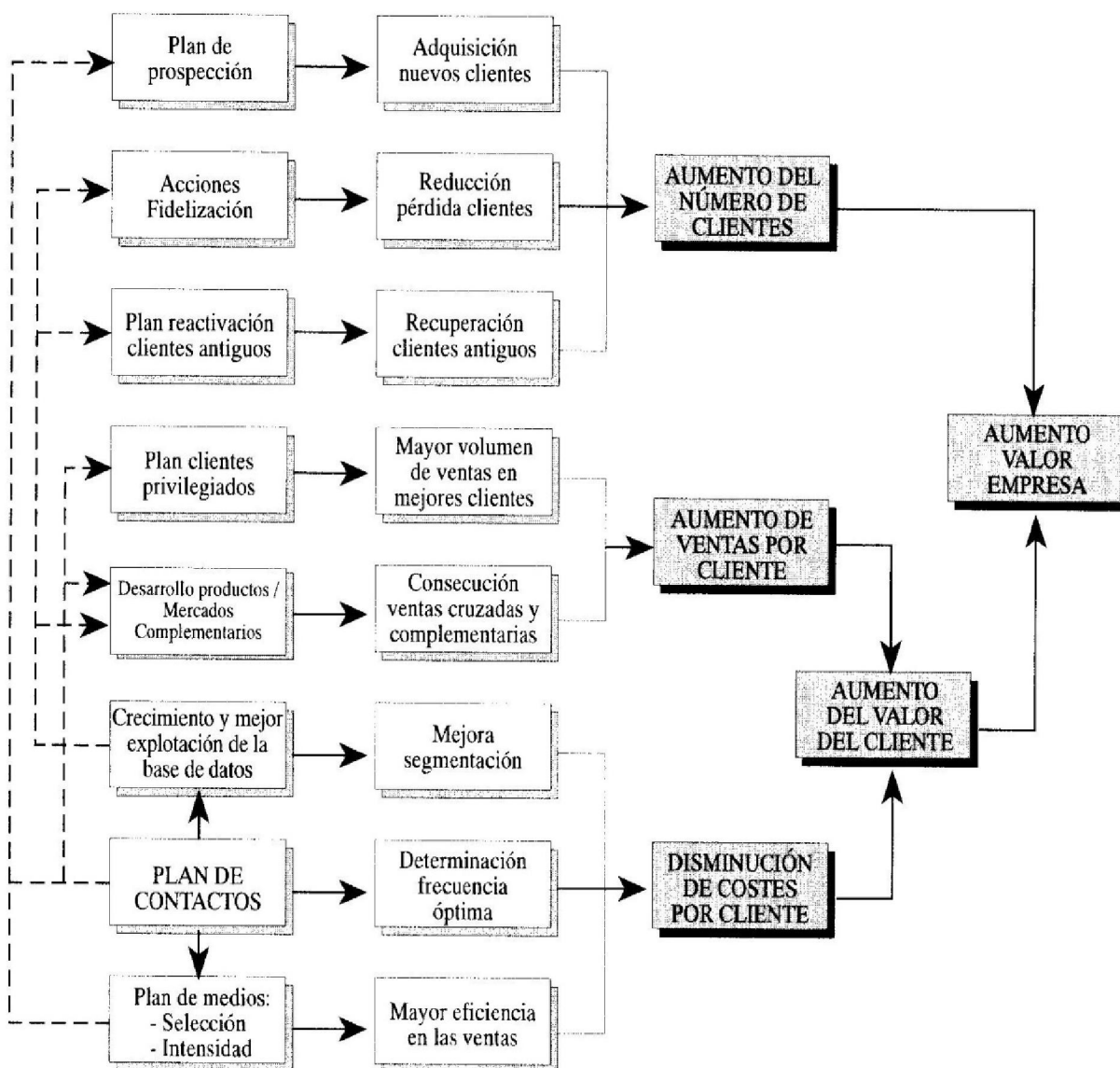


Figura N° 6.7: Plan eficaz de Marketing Relacional

Fuente: Alet, 1996, Pág. 139

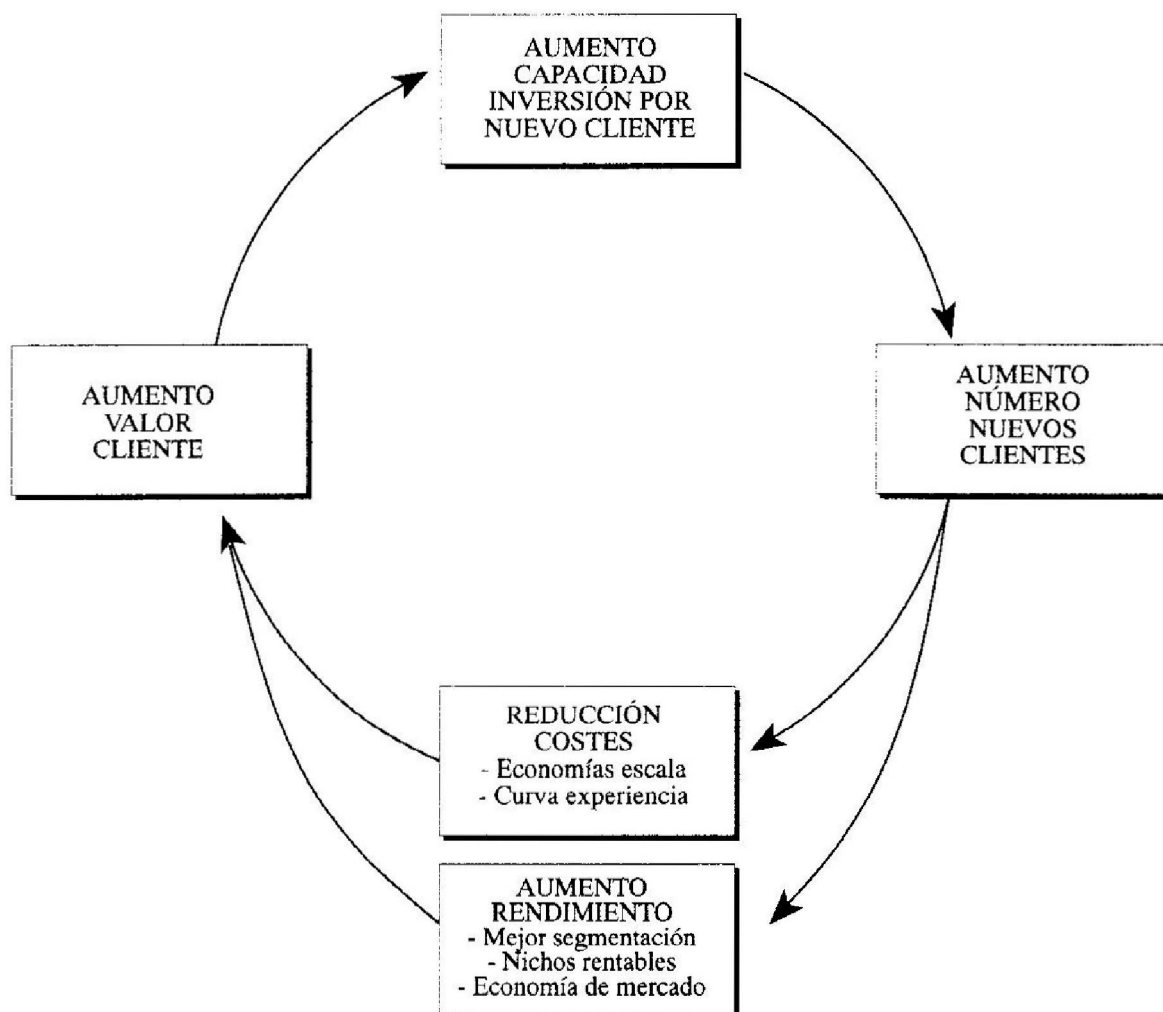


Figura N° 6.8: Círculo de la Riqueza en Marketing

Fuente: Alet, 1996, Pág. 140

6.3.8 Como obtener clientes reales y rentables

Una vez se tiene un cliente, entra en acción el marketing *Back-end*, orientado a maximizar el valor del cliente, con programas de continuidad, con el tratamiento de los clientes como el activo más importante. Los clientes, como activo, tienen que ser evaluados por los beneficios que aportan a lo largo de su vida económica para la empresa.

El objetivo básico es el aumento del valor de los clientes, a través de distintas políticas Alet (1996) que son:

- Retención: mantenimiento de los clientes activos en la empresa.
- Mayor volumen promedio de compra por pedido.
- Mayor frecuencia de compra.
- Mayor potencial de consumo por creación de nuevas líneas de productos, consiguiendo ventas cruzadas.

Por tanto, el cultivo de la relación va dirigido a la fidelización del cliente y el aumento de su valor para la empresa, a través de una serie de acciones para mantener una relación continuada que permita conocerse mejor mutuamente, realizar nuevas ventas y mantener el compromiso que se ha establecido.

Ahora bien, podemos señalar según Alet (1996) que para conseguir el objetivo final que es obtener clientes leales y rentables, se realiza a través de un plan de comunicación de la relación y programas para recuperar clientes perdidos.

I. Plan de comunicación en el cultivo de la relación

Existe un gran abanico de acciones de comunicación posibles para fidelizar y aumentar el valor de los clientes.es conveniente realizar.

A continuación se indicará diversos tipos de acciones que sirven de muestra de referencia de las acciones posibles dentro de este campo.

A. Creación de un Club

El club es una de las formas más completas de realizar el Marketing Back-End, por cuanto, dentro de él, se establecer una serie de vínculos entre la empresa y el cliente y un dialogo de forma natural y casi permanente.

Implica además, un compromiso importante de futuro de la empresa en la creación de una orientación en torno al cliente, con un reconocimiento de los

múltiples contactos, respuestas y compras que acontecen a lo largo de la relación con éste.

El fin que con lleva crear un club, es el establecimiento de vínculos profundos entre la empresa y el cliente, para obtener una relación de lealtad. De la relación establecida, se obtienen ventajas para la empresa de ventas cruzadas y complementarias, mayor conocimiento de los clientes y una mayor permanencia de estos en la empresa. Por ende, el club se convierte en un escudo protector ante las acciones de la competencia.

B. Material de acompañamiento

Es conveniente acompañar el producto vendido o pedido de una información que reafirme que el cliente ha tomado la decisión correcta y que ha sido una buena elección. Es importante dar testimonios de clientes satisfechos, garantías de la calidad de los componentes o del proceso de fabricación.

C. Agradecimiento

El agradecimiento por sí mismo es algo poco usual en las comunicaciones de la empresa. Se aprovecha normalmente para combinarlo con oferta de nuevos productos y servicios. Sin embargo, el decir simplemente gracias de forma natural y sincera, aporta un elemento importante de vinculación suficientemente importante que merece desvincularlo de nuevas ofertas comerciales.

D. Venta Cruzada

En la venta cruzada, se trata de ofrecer productos de un tipo distinto a los que el cliente posee hasta el momento. Por ejemplo, un seguro contra daños a un vehículo.

E. Ventas complementarias

Es igualmente interesante ofrecer productos complementarios al producto entregado, como en el caso de las actualizaciones de Anti Virus.

En el caso de venta por catálogo e usual presentar una oferta especial con el envío del artículo, dirigida exclusivamente al comprador del artículo, dando valor a la especificidad de la oferta y su carácter personal y valioso.

F. Ofertas especiales

Resulta rentable colocar ofertas reservadas a clientes especiales, como la persona a la que nos dirigimos, con preciso especiales o artículos con cantidad limitada, utilizando la exclusividad o la idea de oportunidad. Cuando más específica y afinada sea la oferta especial para los clientes, mayor será la probabilidad de éxito en la obtención de ventas adicionales.

En ciertos casos la oferta especial es una ventaja que no aporta ingresos económicos a la empresa, pero ofrece un claro beneficio al cliente para que continúe consumiendo los productos o servicios de la empresa.

G. Información sobre los productos y servicios de la empresa

Resulta útil presentar información fundamental sobre la empresa, sus productos, sus servicios, su gente, en especial para los clientes que nos compran por primera vez y regularmente. Se trata de una información que acerque la empresa al cliente y que le haga participe de su vida empresarial.

H. Mantener el contacto

El plan de contacto de la empresa debe reconocer el ciclo de compra del cliente sus necesidades y los productos consiguientes a ofrecer, para diseñar la frecuencia, medio y contenido idóneos para una comunicación efectiva.

Cuando el cliente sepa más de la empresa, habrá una alta probabilidad que siga comprando y se mantenga fiel a esta.

I. Servicio de Post-Venta

La empresa tiene que continuar la acción de venta facilitando el uso del producto o servicio por parte del cliente. En el caso de productos complejos es primordial el apoyo de la red de ventas, servicio técnico, o número de teléfonos donde realizar consultas. Después de la venta, la empresa continúa dando soporte al usuario para una correcta y fácil utilización de sus productos.

II. Programa de recuperación e clientes antiguos

Existen razones evidentes para considerar que un antiguo cliente puede ser más rentable que uno nuevo. Las diferencias más destacables según Alet (1996), son entre otras:

- i. Son más accesibles ya abiertos a responder a los esfuerzos de la empresa.
- ii. Mantienen un sentimiento de lealtad o afinidad a la empresa como antiguo amigo.
- iii. Después de su insatisfacción con la empresa, pueden haber tenido experiencias de un nivel similar en empresas competidoras, por lo que pueden aceptar nuevamente su oferta.
- iv. Si se sabe la razón de la partida, tras haber resuelto el problema originario la empresa puede comunicarle la novedad y lo que significa para el antiguo cliente y su problema particular.
- v. Se dispone de una información que puede ayudar a desarrollar un programa específico para obtener una oferta y servicios relevantes para el antiguo cliente, lo cual es imposible o muy caro en el caso de un presunto cliente.

A. Análisis de las razones de la pérdida de clientes

La pérdida de un cliente es el comportamiento de no lealtad que queremos evitar. El análisis de la pérdida de clientes es una forma de conocer los factores que provocan la no lealtad y su importancia relativa

Según Desouza (1996: Citado en Alet (1996), Pág. 172) existen ocho razones de pérdida de un cliente con distintas matizaciones:

1. **Precio:** el cliente cambia a un competidor con un precio menor. Este precio puede ser analizado tanto en términos absolutos como en términos relativos (Calidad/Precio).
2. **Producto:**
 - a) El cliente se va a un competidor que ofrece un nuevo producto superior.
 - b) Uno de los productos entregados no ha satisfecho las expectativas ante un funcionamiento irregular, por lo que el cliente cambia de marca.
 - c) Falta de actualización y servicios post-venta.
 - d) El cliente se mueve en un nivel de decisión de alternativas de elección, entre las que está el producto considerado.
3. **Servicio:** el cliente deja la empresa por un servicio pobre o defectuoso. Pueden distinguirse varios tipos:
 - a) Errores en el funcionamiento con entrega incorrecta.
 - b) Mala atención personal por parte de un representante de la empresa.
 - c) Errores en la asistencia técnica.
 - d) Problemas en la gestión de capacidad.
4. **Mercado:** el cliente se pierde por salirse del mercado donde concurre el producto:

- a) Traslado geográfico
 - b) Cierre de la empresa o fallecimiento del cliente.
 - c) Cambio de actividad del cliente empresa.
 - d) Cambios en el estilo de vida, etapa de la vida, problemas de salud, etc.
5. **Tecnología:** el cliente pasa a utilizar un producto basado en una tecnología o sistemas distintos. Sería una situación donde se va a productos sustitutos o alternativos.
6. **Organización:** los clientes son perdidos por razones políticas más allá de aspectos técnicos o de servicio:
- a) Incorporación del cliente a un grupo competidor.
 - b) Nuevas políticas restrictivas de compra.
 - c) Absorción de partes del negocio por el propio cliente, a través de una organización interna.
7. Vínculos personales del cliente con el nuevo proveedor
8. Pérdida de la parte poseída del cliente
- a) Ante el proveedor principal
 - b) Frente a otro proveedor

B. Plan de recuperación de clientes

La recuperación debe realizarse con un planteamiento previsional, tratando de detectar la fuga probable por un comportamiento previo de queja, su grado de satisfacción mostrado en las encuestas de satisfacción sistemáticas en los momentos adecuados de la relación, su no respuesta en un plazo de tiempo a nuestras comunicaciones, o la misma antigüedad del bien o servicio adquirido.

El plan de recuperación contemplaría los siguientes pasos:

- 1) Reconocimiento de la no lealtad del cliente al haber adquirido otra marca.
- 2) Comprensión de las razones de su no lealtad.
- 3) Reacción tan pronto como sea posible sobre la solución del déficit o problemas que llevaron al cliente a dejar la marca.
- 4) Comunicaciones periódicas indicando los avances en las áreas consideradas importantes para el cliente,
- 5) Comunicaciones con el antiguo cliente equivalentes a la de prospecto cualificado, cuando entra en la fase de probable nueva compra, dentro de sus ciclo de compra habitual o la indicada por el mismo.

6.3.9 Indicadores clave del éxito dentro de un programa de Marketing Relacional

Para dirigir adecuadamente un programa de marketing Relacional y conseguir obtener el deseado beneficio debemos cuidar de los factores claves de éxito comercial, que dependerán tanto del negocio en que participe la empresa, como de la estrategia que se haya decidido seguir.

El análisis de las distintas etapas de la relación y su trasposición a los indicadores claves se muestran en la Tabla N° 6.3.

Tabla N° 6.3: Indicadores fundamentales en las etapas de la relación con el cliente

Fuente: Alet, 1996, Pág. 175

ETAPA	INDICADOR	CÁLCULO
Identificación Presuntos	<ul style="list-style-type: none"> - % de respuestas - Respuestas - Costo por respuesta 	<ul style="list-style-type: none"> - Identificados/Audiencia - Inversión/Respuesta
Cualificación	<ul style="list-style-type: none"> - Numero de presuntos con potencial inmediato, corto, medio y largo plazo. 	<ul style="list-style-type: none"> - Clasificación potencial en el tiempo
Ventas compradores	<ul style="list-style-type: none"> - % de conversión - Valor de compra - Tipos de productos - Costo por pedido 	<ul style="list-style-type: none"> - Ventas/Cualificados - VPM - Número líneas de producto - Inversión/Pedido
Relación clientes	<ul style="list-style-type: none"> - Compradores con varias compras - Mas anteriores venta - Aportación de amigos 	<ul style="list-style-type: none"> - % Compradores repetidores/Activos - Pedidos y ventas/cliente - Referencias/cliente
Asociación socios	<ul style="list-style-type: none"> - Cadena de la amistad aporta nuevos clientes - Participa en el desarrollo y mejora de nuestros productos y servicios 	<ul style="list-style-type: none"> - Nuevos compradores a través de clientes. - Ideas obtenidas de clientes
Recuperación antiguos clientes	<ul style="list-style-type: none"> - Clientes recuperados por causa. - Clientes calmados 	<ul style="list-style-type: none"> - Número y % recuperados por causa. - % de calmados/Insatisfechos iniciales

6.3.10 Plan de Acción

I. Definición

El plan de acción es una herramienta que facilita llevar a cabo los fines planteados por una organización, priorizando las iniciativas más importantes para cumplir con objetivos y metas trazados, es decir describe el medio por el cuál se lograrán los objetivos.

Es así, que el Plan de Acción permite organizar y orientar estratégicamente acciones, capital humano, procesos, instrumentos y recursos disponibles hacia el logro de objetivos y metas así como definir indicadores que facilitan el seguimiento y evaluación de las acciones y sirven de guía para la toma oportuna de decisiones.



Figura N° 6.9: Preguntas que se deben realizar al diseñar un plan de acción

Fuente: Elaboración Propia

Un plan debe contener la siguiente información:

- Definir los objetivos específicos.
- Establecer las metas necesarias para alcanzar cada uno de los objetivos específicos.
- Establecer indicadores que permitan medir el logro de la meta.
- Determinar las actividades que deben desarrollar.
- Especificar el despacho judicial responsable de llevar a cabo las actividades.
- Si se necesita coordinar la realización de alguna actividad, se debe indicar con cual despacho, institución u organización.

II. Elementos del Plan de Acción

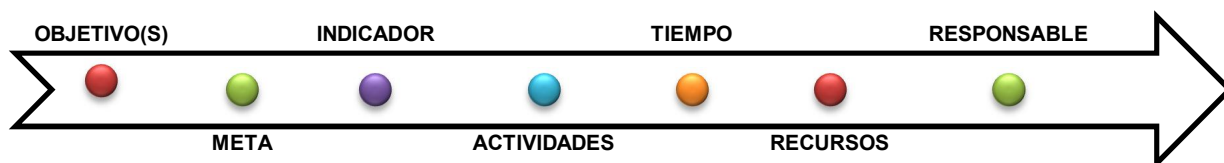


Figura N° 6.10: Elementos de un plan de acción

Fuente: Elaboración Propia

A. Objetivos

Un objetivo consiste en un deseo de lo que se quiere lograr (atender a una problemática o necesidad). Este propósito debe ser expresado en forma clara y concisa. Todo objetivo por tanto, debe responder la pregunta ¿Para qué?

Dentro de sus características se encuentran:

- **Precisos:** de tal forma que se pueda realizar una buena planificación y adecuada evaluación de los objetivos.

- **Adecuados en el tiempo:** que se puedan cumplir en un período razonable de trabajo.
- **Flexibles:** que sean posibles de modificar cuando se presenten situaciones inesperadas.
- **Motivadores:** que sean para las personas un reto posible de alcanzar. Que los involucrados se sientan identificados con el desarrollo de actividades.
- **Participativos:** los mejores resultados se logran cuando los responsables del cumplimiento de los objetivos pueden participar en su establecimiento.
- **Factibles:** deben ser reales, prácticos y posibles de lograr.
- **Convenientes:** que sus logros apoyen los propósitos y misiones básicas de la organización.
- **Obligatorios:** una vez establecidos debe existir una obligación razonable, para lograr su cumplimiento.

B. Metas

Reflejan lo que se va a lograr en un tiempo determinado, respondiendo a cuanto se desea alcanzar. La principal característica de una meta es que debe explicar claramente, ¿Qué se quiere?, ¿Cuánto se quiere? Y ¿Para cuándo se quiere?

La fijación de metas hace posible la medición de los resultados y la evaluación del grado de cumplimiento y eficiencia logrados por el responsable. En otras palabras, la meta es la cuantificación del objetivo específico ubicado en el tiempo y lugar.

Las metas definen la cantidad de lo que se desea alcanzar a través de los objetivos específicos, pueden especificarse en términos absolutos o relativos.

Las metas deben ser:

- Realistas: que se pueden alcanzar con los recursos disponibles.
- Precisas: debidamente explícitas.
- Periódicas: definir la fecha en la que se desean alcanzar.
- Medibles: señalar unidades de medida.
- Coherentes: vinculadas con los objetivos.

C. Indicadores

Son parámetros de medida por medio de los cuales se determinan el logro de la meta y por consiguiente el cumplimiento de los objetivos específicos.

Los indicadores tienen que ser medibles en cantidad y tiempo. El conjunto de indicadores permite analizar, evaluar y justificar las actividades y los resultados planteados.

D. Actividades

Las actividades son todas aquellas tareas o eventos destinados al cumplimiento de las metas previstas. Señalan los pasos lógicos o el camino que se debe seguir para contribuir al logro de las metas.

Las actividades deben responder a las preguntas ¿Que se debe hacer? Y ¿Cómo se debe hacer?

E. Responsable

Se debe señalar quien concretamente es responsable de realizar la actividad. Los responsables han de tener la capacidad de realizar la actividad planteada.

6.4 Plan de acción VTR Globalcom S.A. Zona Sur

I. Justificación

La decisión de realizar un Plan de Acción, radica en la creación de un plan de capacitación a personal de ventas y atención al cliente de las Plazas de Concepción y Temuco, de manera que estos adquieran los conocimientos necesarios en Marketing Relacional. Lo anterior se justifica debido al desconocimiento que cuenta este personal, evidenciado en la escasez de estrategias de retención, aumento de la satisfacción del cliente y la poca lealtad que tienen estos últimos, esto reflejado en los resultados del estudio de Análisis Cluster realizado en esta investigación, donde se observa que para ambas plazas el mayor nº de desconexiones de servicios ocurre cuando estos tienen una antigüedad que fluctúa promedio entre los 13 y 24 meses.

Realizada la capacitación, el personal de Ventas y Atención al cliente seleccionado, será capaz de comprender, entender y mejorar las estrategias de retención de clientes antiguos y captación, también aumentará la participación de estos en importantes reuniones realizadas en las Oficinas de Marco Polo, entregando sus impresiones, acotaciones y nuevas ideas en post de aumentar la cartera de clientes, logrando que las relaciones con estos últimos, sea muy difícil de romper por parte de los competidores.

El recinto de capacitación, será las oficinas de Marco Polo 8880, comuna de Hualpén, donde cuenta con una sala de capacitación propia, equipada con equipo digital, audiovisual, proyector, pizarra y computadores necesarios para cada uno de los participantes, además de una cafetería donde realizar los intermedios (Coffe Break).

Con respecto a los expositores, estos serán ejecutivos con un alto grado de capacitación y experiencia en las oficinas de Marco Polo y Santiago pertenecientes a diversos departamentos, de manera de generar un conocimiento transversal del Marketing Relacional en diversas líneas de la compañía.

II. Diseño del Plan de capacitación Personal de Ventas y Atención al Cliente VTR Globalcom S.A. Zona Sur

1. Objetivo

Conocer y comprender conceptos actualizados de Marketing Relacional y la importancia que este tiene dentro de las compañías.

2. Beneficios del Plan de acción

- El personal capacitado será capaz de comprender las herramientas que posee el Marketing relacional y como estos ayudan a alcanzar los objetivos y metas de la empresa.
- Elevará el nivel de conocimiento y toma de decisiones.
- Logrará un alto nivel de efectividad en el personal capacitado, al momento de diseñar e implementar estrategias establecidas por parte de la empresa en retención y captación de nuevos clientes.

3. Cobertura

Capacitar a las secciones que se encuentran directamente relacionados con venta y atención al cliente de las Plazas de Concepción y Temuco.

4. Procedimiento de implementación

- Aprobado el plan de capacitación, se procederá a la selección de Personal de Venta de las plazas Concepción y Temuco como así también de Atención al cliente, por parte del Gerente Zonal y Gerente Comercial Zona Sur
- Se estructurará un grupo conformado por un máximo de 30 personas entre las Plazas de Concepción y Temuco, con el propósito de que los recursos técnicos sean aprovechados de manera óptima.

- Se establecerán horarios fijos, en los cuales se llevará a cabo la capacitación de manera de realizar la rotación necesaria y así evitar escasez de personal.
- Iniciada la capacitación, cada módulo contará con un responsable que será quien evaluará a los participantes, aplicando un sistema de evaluación que considere adecuada. Esta comprenderá aspectos conversados en la capacitación, adaptación y participación de los participantes.
- Durante todo el proceso se llevara un proceso tanto teórico como práctico, en los cuales cada responsable de los módulos determinará el módulo de evaluación y actividades de reforzamiento.
- La duración de proceso de capacitación, comprenderá una duración de veinte horas desglosadas en tres semanas. Posterior a la capacitación se realizará un seguimiento a cada plaza.

5. Contenido de los módulos a desarrollar en la capacitación

La modalidad de la capacitación comprenderá un parte teórica y una parte práctica de un 70% y 30% respectivamente. La parte teórica comprenderá el desarrollo de cuatro módulos, en los cuales los responsables de cada uno de ellos será personal con un alto grado de capacitación perteneciente a la Oficinas de Marco Polo 8880, Hualpén y Casa Central, Santiago.

El detalle del Plan de Acción, queda plasmado en la Tabla N° 6.4, donde se detalla los objetivos a cumplir, las actividades, metas, indicadores, los recursos involucrados, período de ejecución y responsables.

Además, en la Tabla N° 6.5 se detalla el contenido de los módulos, en la Tabla N° 6.6 el cronograma de actividades y en la Tabla N° 6.7 los costos involucrados en la capacitación.

Tabla N° 6.4: Plan de Acción VTR Globalcom S.A. Zona Sur (Plaza Concepción & Temuco)

Fuente: Elaboración Propia

OBJETIVO	ACTIVIDAD	METAS	INDICADORES	RECURSOS			PERIODO DE EJECUCIÓN	RESPONSABLE
				HUMANOS	TÉCNICOS	FINANCIEROS		
Capacitar al personal de ventas y atención al cliente Plaza Concepción y Temuco	Diseño del proyecto y selección de instructores que impartirán los módulos	Lograr que a lo menos cuatro expositores de la compañía, desarrollen la capacitación.	Aumentar en un 30% el número de expositores de la Plaza de Concepción.	<ul style="list-style-type: none"> - Gerente comercial - Sub-gerente Servicio al Cliente y Marketing - Jefe Control Gestión. - Enc. Proyectos Comerciales - Personal de Ventas y Atención al Cliente Plaza Concepción y Temuco 	Sala de capacitación, Marco Polo 8880, Hualpén. Equipo de proyección digital. Computadores de escritorio. Material de estudio.	Insumos para alimentación. Impresión de material de estudio. Dinero para traslado desde la Plaza de Temuco.	Tres semanas	Gerente Zona Sur & Gerente Comercial
	Diseño de los contenidos a tratar en los módulos	Generar cuatro módulos de estudio, donde el tema central será marketing Relacional e incluir aplicaciones y experiencias en compañías de diversos rubros.	Mejorar en un 40% el conocimiento en relación con Marketing Relacional.					
	Creación de actividades complementarias y recreativas durante la capacitación	Identificar a lo menos ocho personas capacitadas que sobresalgan del resto de sus compañeros.	Incrementar el número de participantes en las reuniones estratégicas o de análisis en un 40%.					

Tabla N° 6.5: Contenidos programas de Capacitación

Fuente: Elaboración Propia

MÓDULO	ASPECTOS	RESPONSABLE	DURACIÓN
Módulo I Introducción al Marketing Relacional	<ul style="list-style-type: none"> • Introducción • Objetivos del Marketing Relacional. • ¿Por qué ha crecido el Marketing Relacional? 	Gerente Comercial	2 Horas
Módulo II Marketing Relacional y Programas de Fidelización	<ul style="list-style-type: none"> • ¿Qué buscamos por medio de la Fidelización? • Programas de captura, Fidelización, Rentabilización y recuperación de clientes. • Ventajas y Desventajas de la Fidelización. • Objetivos y requisitos para realizar un programa de fidelización. 	Jefe Control de Gestión	6 horas
Módulo III Gestión de Relaciones con Clientes	<ul style="list-style-type: none"> • Modelos de gestión de clientes. • Estrategias de gestión de clientes. • ¿Qué es el valor de vida de un cliente? • ¿Cómo obtener clientes leales y rentables? • Conceptos fundamentales del CRM. • CRM como estrategia y soporte de negocio. 	Sub-Gerente Servicio al Cliente y Marketing	8 horas
Módulo IV Plan de Marketing Relacional & Inteligencia de Clientes	<ul style="list-style-type: none"> • ¿Qué es un Plan estratégico? • Las razones actuales de sus usos. • Cuadro de mando relacional. • Técnicas fundamentales de análisis y segmentación de clientes. • Cómo establecer la efectividad de una acción de Marketing Relacional. 	Encargado de Nuevos Proyectos	4 horas
TOTAL DE HORAS PROGRAMA DE CAPACITACIÓN			20 horas

Tabla N° 6.6: Cronograma de Actividades

Fuente: Elaboración Propia

MODULO	OBJETIVO	RESPONSABLE	DURACIÓN HORAS	Semana 1			Semana 2			Semana 3		
				D1	D2	D3	D4	D5	D6	D7	D8	D9
Introducción al Marketing Relacional	<ul style="list-style-type: none"> Comprender la dimensión y objetivos que trabaja el Marketing Relacional. 	Gerente Comercial	2									
Marketing Relacional y Programas de Fidelización	<ul style="list-style-type: none"> Comprender los fundamentos del Marketing Relacional y su importancia dentro de las empresas. 	Jefe Control de Gestión	6									
Gestión de Relaciones con Clientes	<ul style="list-style-type: none"> Entregar herramientas para la implementación de relaciones con clientes. Comprender el CRM desde la perspectiva del capital relacional que representa una marca. 	Sub-Gerente Servicio al Cliente y marketing	8									
Plan de Marketing Relacional & Inteligencia de Clientes	<ul style="list-style-type: none"> Entregar herramientas que permitan el diseño de planes de Marketing de Relaciones velando por la orientación a resultados. Abordar el control de gestión de clientes desde una perspectiva estratégica de las marcas basado en cuadros de mando. 	Encargado de Nuevos Proyectos	4									

Tabla N° 6.7: Costo Capacitación

Fuente: Elaboración Propia

SECCIÓN	DESCRIPCIÓN	VALOR UNITARIO	CANTIDAD	TOTAL
Abarrotes	Café-Gold Premium 170G	\$ 2.355	5	\$ 11.775
	Té-Lipton 100 Un.	\$ 2.490	3	\$ 7.470
	Azúcar 1.5 Kg	\$ 1.000	4	\$ 4.000
	Endulzante 270 CC	\$ 2.450	3	\$ 7.350
Panadería	Pan-Ideal 650g	\$ 1.529	20	\$ 30.580
	Queso-Colún Gauda 500 G	\$ 3.190	8	\$ 25.520
	Jamón-Arztía Pechuga de pavo 250 G	\$ 1.197	6	\$ 7.182
	Paté-Bonanza cerdo 100G	\$ 599	4	\$ 2.396
Confitería	Nick	\$ 239	15	\$ 3.585
	Líder-Coco	\$ 240	15	\$ 3.600
	Frac-Vainilla	\$ 299	15	\$ 4.485
	McKay-Tritón	\$ 340	15	\$ 5.100
	Costa-Chocochips	\$ 349	15	\$ 5.235
Alfajores	NutraBien-Brownie 4 Un.	\$ 1.440	15	\$ 21.600
	Lagos del Sur-6 Un.	\$ 990	15	\$ 14.850
Snack	Lay's 450 G.	\$ 1.820	10	\$ 18.200
Bebidas	Pack: Pepsi-Bilz-Pap 3Litros c/u	\$ 3.690	12	\$ 44.280
	Agua Mineral Cachantún 6 Un. 500cc	\$ 1.930	6	\$ 11.580
Artículos de Oficina	Cuaderno Torre 100 Hojas	\$ 890	30	\$ 26.700
	Papel Fotocopia Carta 1 Resma	\$ 2.790	30	\$ 83.700
	Marcador Artel (20 unidades)	\$ 711	20	\$ 14.220
	Tijera 1 Un.	\$ 460	6	\$ 2.760
	Carpeta con elástico	\$ 790	30	\$ 23.700
	Bolígrafo (Negro-Rojo-Azul)	\$ 690	40	\$ 27.600
	Papel Krafft-Pliego	\$ 120	30	\$ 3.600
	Marcador de Pizarra 4 colores	\$ 2.290	6	\$ 13.740
Expositores	Bono Incentivo	\$ 100.000	7	\$ 700.000
Lugar	Marco Polo N° 8880, Hualpén	No aplica	No Aplica	No Aplica
Personal de Atención	2 Hombres, 2 Mujeres	\$ 100.000	4	\$ 400.000
Bono Locomoción	Expositores Santiago	\$ 50.000	2	\$ 100.000
	Plaza Concepción y Temuco	\$ 50.000	30	\$ 1.500.000
TOTAL				\$ 3.124.808

6.5 Conclusiones

A lo largo de este capítulo, se desarrolló un plan de acción para la Zona Sur de VTR Globalcom S.A., involucrando personal de ventas y atención al cliente de las Plazas de Concepción y Temuco, como también las localidades de cada una de las plazas ya mencionadas.

La idea de su creación nace producto de resultados obtenidos del Análisis Cluster realizado anteriormente en la investigación, en donde se establece que por parte de los clientes hay una baja permanencia en la compañía, lo que sumado además a las pocas promociones de retención generadas en los meses que se han tomado como base para el desarrollo del trabajo, han evidenciado que existe un desconocimiento general por parte del equipo de VTR en lo que respecta a Marketing Relacional, puesto que los planes de retención en ningún momento han tenido la finalidad de aumentar la satisfacción del cliente, y de paso su lealtad con la compañía.

La creación y posterior puesta en marcha del plan, será comandada por ejecutivos de alto grado en la Zona Sur, como es el gerente Zona Sur y Gerente Comercial. Su diseño, ajusta un mixture ideal entre la parte teórica y práctica, evidenciada en el desarrollo de los ítems y actividades comprendidas de cada uno de los módulos.

Consideramos que el personal seleccionado terminada la capacitación, será capaz de comprender, analizar, diseñar & inferir acciones relacionadas con Marketing Relacional dentro de la empresa, con el propósito de elevar el grado de satisfacción de los clientes, aumentando la lealtad de estos durante el tiempo.

Además de lo anterior, el personal capacitado será capaz de integrar un departamento o área especializada en diseñar estrategias para retención de clientes activos y captación, como así también elevar el nivel de satisfacción de estos con respecto a los productos que poseen en sus hogares, por medio de

ofertas y promociones concretas, imposibles de igualar por parte de competencia, manteniendo un contacto periódico con sus clientes.

CONCLUSIONES GENERALES

La utilización de ambas técnicas (que permitieron generar ambos modelos), obedece a poder entregar a la compañía dos análisis distintos con respecto a la utilización de las mismas bases de datos.

La técnica de Análisis Cluster permite generar perfiles de clientes potencialmente fugables para la compañía, estableciéndose números de conglomerados y porcentajes de representación de estos últimos para las plazas de Concepción y Temuco. Con estos perfiles, en donde se incluyen veinte variables en cada uno de ellos, se puede observar las características demográficas de clientes, y de servicios que se han fugado. Con esta valiosa información, la compañía podrá tener distintos perfiles de conglomerados en donde a cada uno de ellos se le podrá asignar un nivel de riesgo, lo que ayudará a tomar decisiones de marketing y/o planes de acción para disminuir la tasa de fuga de clientes a lo largo del tiempo.

Por su parte la técnica de regresión logística multivariante permite identificar los factores de riesgo que influyen en la fuga de servicios para cada plaza. El modelo generado para la plaza de Concepción incluye ocho variables (**GSE, DEUDA, EDAD_C, PERMA_MESES, CPAGO, RECLAM, C_ENTRADA y PERMA_MESES * RECLAM**), de las cuales las variables **DEUDA, PERMA_MESES, EDAD_C y C_ENTRADA** al tener un coeficiente de regresión mayor a uno ($Exp(\beta) > 1$) incrementan la probabilidad de ocurrencia de fuga de servicios. Por otra parte, el modelo generado para la plaza de Temuco incluye siete variables (**GSE, DEUDA, EDAD_C, PERMA_MESES, CPAGO, RECLAM y PERMA_MESES * RECLAM**), de las cuales las variables **DEUDA, PERMA_MESES y EDAD_C** al tener un coeficiente de regresión mayor a uno ($Exp(\beta) > 1$) incrementa la probabilidad de ocurrencia de fuga de servicios.

El nivel de sensibilidad del modelo generado para la plaza de Concepción es de 58,5%, inferior al nivel de especificidad que es de 80,4%. El modelo de forma global clasifica de forma correcta al 71,2% de los casos analizados.

Por otra parte, el nivel de sensibilidad del modelo generado para la plaza de Temuco es de 82,8%, muy superior al nivel de especificidad que es de 55,3%. El modelo de forma global clasifica de forma correcta al 72,1% de los casos analizados.

Finalmente, de Regresión Logística Multivariante, es posible mediante la formulación matemática establecer en la actualidad la probabilidad de fuga de ciertos clientes, como por ejemplo, los que cumplan con los perfiles o patrones de fuga realizado en Análisis Cluster, lo que permitiría ser aún más exacto en la aplicación de esta técnica, y por ende darle un mayor sentido a todo el fruto de esta investigación.

Si bien ambos modelos clasifican de forma correcta sobre un 70% de los casos analizados, es recomendable obtener y/o recabar información de otros factores de riesgo, que por múltiples motivos no han sido considerados en este estudio de regresión logística multivariante. Factores como nivel de utilización de servicios, nº de desconexiones y/o interrupciones de servicios, etc., podrían brindar valiosa información para la generación de futuros modelos.

Con respecto al Plan de Acción propuesto a VTR, podemos decir que su origen es en base a la escasez de estrategias de retención de clientes, aumento de satisfacción y lealtad, evidenciado en los resultados del análisis cluster realizado en esta investigación, donde el tiempo de antigüedad promedio fluctúa entre los 13 y 24 meses.

La base de este plan de acción, es realizar una capacitación al personal de ventas y atención al cliente de las plazas de Concepción y Temuco, donde los contenidos y actividades complementarias estarán bajo supervisión del Gerente Zonal y Gerente Comercial. Con respecto a la capacitación en términos generales, este se realizará en Concepción, con relatores de la compañía tanto de Concepción, Temuco y Santiago capacitados en Marketing Relacional.

Una vez realizada la capacitación, los relatores podrán identificar los mejores valores que participaron en esta, y así estos podrán integrar un nuevo

equipo o departamento de retención de clientes, que tendrá la misión de diseñar estrategias para elevar la satisfacción de estos y aumentar el nivel de fidelidad hacia la compañía. Además de lo anterior, se pretende aumentar el nivel de participación de personal de Ventas y Atención al cliente en reuniones de estrategia comercial en la compañía, ya que además de tener los conocimientos presentados en la capacitación, estos permiten entregar una visión del cliente más cercana que permita alcanzar los objetivos zonales y corporativos.

BIBLIOGRAFÍA

GUJARATI D. 2003. Econometría. México, McGraw. 972p.

GUTIÉRREZ H. 2001. Calidad Total y Productividad. México. McGraw-Hill. 403p

HAIR J, ANDERSON R, TATHAM R, BLACK W. 2007. Análisis Multivariante. España, Pearson Education. 799p.

HERNÁNDEZ ORALLO, JOSÉ. 2004. Introducción a la minería de datos. España, Pearson. 680p.

JENICEK M, CLEROUX R. 1987. Epidemiología: Principios, Técnicas y Aplicaciones. 1ª Edición. España, Salvat. 393p.

JOHNSON D. 2000. Métodos Multivariados aplicados al análisis de datos. México. International Thomson Editores. 566p.

KOTLER, P. 2001. Dirección de Marketing. La edición del milenio. México, Pearson Education. 787p.

LAMBIN, J. 1995. Mercadeo: Ventas-Dirección y administración. Madrid, McGraw. 610p.

MACHAN, DYAN. 1998. Soap? cars? what's the difference?, Forbes Vol. 162 N° 5, p14.

MOLINA S. T. 2009. Aplicación de técnicas de minería de datos para predicción del Churn de clientes en una empresa de Telecomunicaciones. Memoria de magister en Tecnologías de Información y Gestión. Santiago, Pontificia Universidad Católica de Chile. Facultad de Ingeniería Industrial.

PÉREZ C. 2008. Técnicas estadísticas con SPSS 12. Madrid. Pearson Prentice Hall. 802p.

STANTON W., 2000. Fundamentos de Marketing. México, Mc Graw-Hill. 102p

REFERENCIAS BIBLIOGRÁFICAS EN LÍNEA

AGUAYO, MARIANO & LORA, ESTRELLA. Cómo hacer una regresión Logística Binaria “paso a paso” (I). [En línea] <http://www.fabis.org/html/archivos/docuweb/Regres_log_1r.pdf> [Consulta: 07 de Julio 2011]

AGUAYO, MARIANO & LORA, ESTRELLA. Cómo hacer una regresión logística binaria “paso a paso” (II): Análisis Multivariante. [En línea] <http://www.fabis.org/html/archivos/docuweb/regresion_logistica_2r.pdf> [Consulta: 07 de Julio 2011]

AYUGA, ESPERANZA. UNIVERSIDAD POLITECNICA DE MADRID. Análisis de Conglomerados. [En línea] <http://ocw.upm.es/estadistica-e-investigacion-operativa/matematicas-y-estadistica-aplicada/contenidos/ocw/anal_multivar/mat_clase/anal_mult_2.pdf> [Consulta: 05 de Septiembre 2011]

BLIGOO. Innovación bajo Incertidumbre. [En línea] <<http://eqkafati.bligoo.com/content/view/211213/Tasa-de-fuga-o-churn-y-como-prevenir-el-abandono-de-nuestros-clientes.html>> [Consulta: 02 de Julio 2011]

BURGUEÑO, M.J, GARCÍA-BASTOS, J.L., GONZÁLEZ-BUITRAGO, J.M. Las curvas ROC en la investigación de las pruebas diagnósticas. [En línea] <http://epidemiologiaanimalhumana.com/yahoo_site_admin/assets/docs/M10417071.277232805.pdf> [Consulta: 20 de Septiembre 2011]

CENTRO DE ESTADÍSTICA Y SOFTWARE MATEMÁTICO. Análisis de Conglomerados. [En línea] <<http://www.cesma.usb.ve/~lbravo/co6111/clase4.pdf>> [Consulta: 20 de Julio 2011]

COGNUS. Fuga de Clientes. [En línea] <<http://www.cognus.cl/content/view/313490/fuga-de-clientes.html>> [Consulta: 20 de Octubre 2011]

DATUM INTERNACIONAL. Cluster Analysis. [En línea]
<http://www.datum.com.pe/datum_english/cluster.pdf> [Consulta: 18 Junio 2011]

ENCICLOPEDIA VIRTUAL. Diccionario de economía y Finanzas. [En línea]
<<http://www.eumed.net/cursecon/dic/logist.htm>> [Consulta: 10 de Noviembre 2011]

FERRERO, ROSANA. Análisis Cluster. [En línea]
<<http://es.scribd.com/doc/44649023/SPSS-Act2-ej3>> [Consulta: 08 de Agosto 2011]

FLORES, LUZ. Análisis estadístico de los factores de riesgo que influyen en la enfermedad angina de pecho. [En línea]
<http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/basic/flores_ml/t_completo.pdf>
[Consulta: 09 de Septiembre 2011]

MAHÍA, RAMÓN. UDI de Econometría e Informática - Dpto. Economía Aplicada. Universidad Autónoma De Madrid. Análisis Cluster. [En línea]
<http://www.uam.es/personal_pdi/economicas/rmc/documentos/cluster.PDF>
[Consulta: 05 de Septiembre 2011]

MINISTERIO DE LA PROTECCIÓN SOCIAL. Que es un Plan de Acción. [En Línea] <http://www.minproteccionsocial.gov.co/documentos_y_publicaciones/plan_de_accion.pdf> [Consulta: 17 de Agosto 2011]

PERÉZ, ANTONIO. Un modelo para sistemas complejos basado en técnicas de elementos finitos bidimensionales. [En línea]
<<http://descargas.cervantesvirtual.com/servlet/SirveObras/jlv/35716118989159721976613/023141.pdf?incr=1>> [Consulta: 05 de Septiembre 2011]

SALCEDO. C. 2004. Estimación de la ocurrencia de incidencias en declaraciones de pólizas de importación. [En Línea]
<http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/basic/salcedo_pc/enPDF/cap2.pdf> [Consulta: 10 de Agosto 2011]

SCRIBD. Plan de Acción [En línea] <<http://es.scribd.com/doc/6490905/Plan-de-Accion>> [Consulta: 20 de Agosto 2011]

UNIVERSIDAD DE GRANADA. Departamento de Estadística. [En línea]. <http://www.ugr.es/~bioestad/_private/cpfund7.pdf> [Consulta: 15 de Junio 2011]

UNIVERSIDAD CARLOS III DE MADRID. Departamento de Estadística. [En Línea] <<http://halweb.uc3m.es/esp/personal/personas/imolina/midocencia/tecnicasinvestigacion0910/slidestema4medidasproximidad.pdf>> [Consulta: 12 de Agosto 2011]

UNIVERSIDAD DE GRANADA. Introducción al Análisis Cluster, Consideraciones Generales <<http://www.ugr.es/~gallardo/pdf/cluster-1.pdf>> [Consulta: 15 de Agosto 2011]

UNIVERSIDAD DE SALAMANCA. Departamento de Estadística. [En línea]. <<http://biplot.usal.es/alumnos/ciencias/2estadistica/multivar/cluster.pdf>> [Consulta: 13 de Junio 2011].

UNIVERSITEIT UTRECHT. Spss Cluster Analysis. [En línea] <<http://www.cs.uu.nl/docs/vakken/wo2/SPSS/spss8.pdf>> [Consulta: 05 de Agosto 2011]

ANEXOS

ANEXO A: Centros iniciales y finales de los conglomerados uno, dos y tres – Plaza Concepción

	Centros iniciales de conglomerados			Centros finales de conglomerados		
	1	2	3	1	2	3
MOTDX	7	1	1	3	2	3
DEUDA	1	2	3	2	2	2
EDAD	4	6	1	3	3	3
SEXO	1	1	2	1	1	2
GSE	5	3	3	3	3	3
NUM_RECLA	0	1	3	1	1	1
TIPO_RECLA	0	2	2	1	1	1
ANTIGUE_SER	1	5	3	2	2	2
CANAL_ENTRADA	6	7	1	5	5	5
CANAL_PAGO	1	7	7	4	4	4
CANTC001	1	0	1	1	0	1
CANTF001	1	1	1	1	0	1
CANT2WAY	0	0	1	1	1	1
PILETA	4	10	1	2	10	2
GTAR_PRO	8	9	3	8	6	3
GTAT_VALOR	3	3	3	3	2	2
PROMO_RETEN	6	6	6	6	6	6
CAT_CLIENTE	1	1	2	1	1	1
TIPO_MULTI	1	5	7	7	6	7
RECO	2	2	2	2	2	2

ANEXO B: Cambios en los centros de los conglomerados uno, dos y tres - Plaza Concepción

Historial de iteraciones			
Iteración	Cambio en los centros de los conglomerados		
	1	2	3
1	6,482	6,295	6,395
2	1,410	1,205	,673
3	1,439	,557	,621
4	,906	,259	,131
5	,357	,106	,069
6	,297	,104	,116
7	,214	,100	,072
8	,265	,173	,018
9	,224	,155	,016
10	,034	,051	,006

ANEXO C: Matriz de distancias euclídeas entre los conglomerados uno, dos y tres - Plaza Concepción

Distancias entre los centros de los conglomerados finales			
Conglomerado	1	2	3
1	0	8,215	5,441
2	8,215	0	8,322
3	5,441	8,322	0

ANEXO D: Centros iniciales y finales de los conglomerados uno y tres – Plaza Temuco

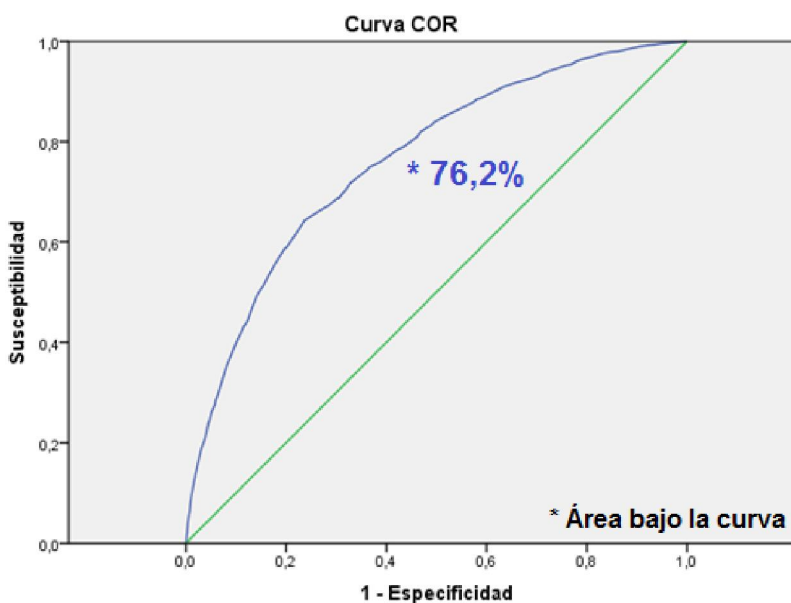
	Centros iniciales de los conglomerados		Centros finales de los conglomerados	
	1	3	1	3
MOTDX	3	1	3	3
DEUDA	1	4	2	2
EDAD	6	2	3	3
SEXO	1	2	2	1
GSE	1	3	2	2
NUM_RECLA	3	0	1	1
TIPO_RECLA	2	0	1	1
ANTIGUE_SER	3	1	2	2
CANAL_ENTRADA	7	7	5	5
CANAL_PAGO	7	1	4	4
CANTC001	1	1	0	1
CANTF001	1	1	1	1
CANT2WAY	0	1	0	1
PILETA	4	3	10	2
GTAR_PRO	9	1	9	3
GTAT_VALOR	3	1	2	2
PROMO_RETEN	6	6	6	6
CAT_CLIENTE	1	1	1	1
TIPO_MULT	1	7	5	7
RECO	2	1	2	2

ANEXO E: Cambios en los centros de los conglomerados uno y tres - Plaza Temuco

Historial de iteraciones		
Iteración	Cambio en los centros de los conglomerados	
	1	3
1	6,497	5,345
2	1,239	1,863
3	,358	1,268
4	,415	1,122
5	,056	,214
6	,004	,025
7	,001	,003
8	,000	,000

ANEXO F: Matriz de distancias euclídeas entre los conglomerados uno y tres - Plaza Temuco

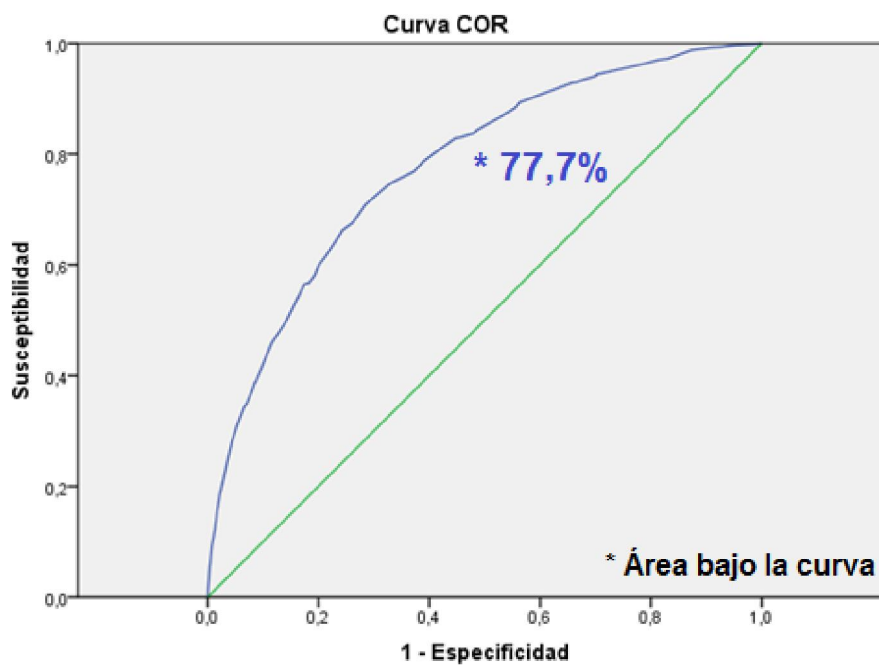
Distancias entre los centros de los conglomerados finales		
Conglomerado	1	3
1	0	6,696
3	6,696	0



ANEXO G: Curva COR – Modelo de RLM – Plaza Concepción

ANEXO H: Niveles de corte, sensibilidad, especificidad y 1 - especificidad – Modelo de RLM – Plaza Concepción

Nivel de corte	Sensibilidad	Especificidad	1 - Especificidad
,000	1,000	,000	1,000
,102	,990	,091	,909
,201	,943	,269	,731
,305	,831	,512	,488
,398	,719	,670	,330
,406	,719	,671	,329
,503	,590	,803	,197
,500	,585	,804	,196
,603	,407	,896	,104
,707	,265	,948	,052
,808	,094	,989	,011
,841	,084	,991	,009
,858	,066	,992	,008
,874	,047	,995	,005
1,000	,000	1,000	,000



ANEXO I: Curva COR – Modelo de RLM – Plaza Temuco

ANEXO J: Niveles de corte, sensibilidad, especificidad y 1 - especificidad – Modelo de RLM – Plaza
Temuco

Nivel de corte	Sensibilidad	Especificidad	1 - Especificidad
,000	1,000	,000	1,000
,108	,998	,012	,988
,200	,993	,088	,912
,307	,944	,298	,702
,399	,894	,437	,563
,500	,828	,553	,447
,508	,801	,592	,408
,603	,662	,758	,242
,698	,567	,817	,183
,801	,373	,920	,080
,898	,158	,983	,017
,911	,125	,986	,014
,928	,091	,993	,007
,946	,044	,997	,003
1,000	,000	1,000	,000