



**UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERIA INDUSTRIAL**

**MODELAMIENTO PREDICTIVO PARA EL AUMENTO DE CONSUMO DE TARJETA  
DE CRÉDITO SOBRE EL ANÁLISIS DE COMPORTAMIENTO TRANSACCIONAL DE  
CLIENTES DE UNA INSTITUCIÓN FINANCIERA**

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL**

**CARLOS ELÍAS ROCO BENAVIDES**

**PROFESOR GUÍA:  
ALEJANDRA PUENTE CHANDÍA**

**MIEMBROS DE LA COMISIÓN:  
LUIS ABURTO LAFOURCADE  
CÉSAR ARAYA HERNÁNDEZ**

**SANTIAGO DE CHILE  
JULIO 2010**

RESUMEN DE LA MEMORIA  
PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL INDUSTRIAL  
POR: CARLOS ELÍAS ROCO BENAVIDES  
FECHA: 02/07/2010  
PROF. GUÍA: SRTA. ALEJANDRA PUENTE

## **MODELAMIENTO PREDICTIVO PARA EL AUMENTO DE CONSUMO DE TARJETA DE CRÉDITO SOBRE EL ANÁLISIS DE COMPORTAMIENTO TRANSACCIONAL DE CLIENTES DE UNA INSTITUCIÓN FINANCIERA**

A nivel de la industria bancaria, del total de tarjetas de crédito circulantes, un 33% de éstas presentan actividad. BancoEstado a Octubre de 2009 posee un total de 400.000 tarjetas, de las cuales sólo un 25% son utilizadas. Lo anterior motiva a tomar acciones para mejorar este indicador, ya que tarjeta de crédito constituye uno de los productos importantes dentro del mix de oferta, tanto por su modelo de negocio de alta rentabilidad como por los beneficios asociados a su mayor utilización, entre los cuales destaca la vinculación de los clientes.

El objetivo principal de este trabajo es generar modelos de propensión para incentivar el uso de tarjetas de crédito, principalmente en segmentos de bajo o nulo consumo. Además se busca caracterizar la dinámica de uso de tarjeta de crédito y los perfiles de consumo de los clientes.

La metodología utilizada para enfrentar este problema tiene como base el proceso KDD. Dentro de éste se utilizaron las variables transaccionales **RFMN** con las cuales se segmentó a los clientes en función de su consumo. Adicionalmente se generaron **cadenas de Markov** que permitieron caracterizar la dinámica de evolución transaccional de los clientes a través del tiempo. En una etapa posterior se perfiló a los clientes con respecto a los tipos de compra y se incorporó esta información en los modelos de propensión, los cuales buscan encontrar las variables que explican el aumento de consumo de un cliente. Los algoritmos testeados fueron **Regresión Logística, Perceptrón Multicapa y Árboles de decisión**.

Resultado de lo anterior se obtuvo una segmentación con un grupo de clientes inactivos y cinco *clusters* de clientes con actividad, los cuales se definieron en relación al consumo en la tarjeta de crédito. Dentro del segmento de clientes inactivos se descubrió que sólo un porcentaje cercano al 20% de los clientes volverán a ser clientes activos en el período siguiente. Así también se aprecia que de forma transversal en todos los segmentos activos, el número de rubros distintos donde una persona compra es una de las variables relevantes al momento de determinar si una persona aumentará su consumo. Por otra parte se observa que los clientes de menor consumo utilizan la tarjeta principalmente para pedir avances en varias cuotas, mientras que los de mayor consumo lo hacen para comprar y hacer avances en efectivo.

Dentro de las conclusiones de este trabajo se destaca que para el segmento de los clientes inactivos, si este estado se mantiene por más de 8 meses, solamente se deben dirigir esfuerzos de marketing a clientes que poseen deuda en BancoEstado. Para los que presentan inactividad menor a 8 meses, se debe fomentar la adquisición y uso de otros medios de pago (CuentaRUT, Cuenta Corriente, Chequera Electrónica, CuentActiva) ya que esto incrementa la probabilidad de que aumente su consumo en el periodo siguiente. Para los segmentos de clientes activos, se concluye que se deben diseñar campañas que apunten a incrementar el número de rubros donde una persona compra y en particular incentivar el uso de la tarjeta en rubros de alta frecuencia, por ejemplo supermercados y farmacias. En función de lo anterior, se plantean distintas acciones para cada segmento, entre ellas: Campañas informativas de promociones vigentes a los clientes con alta probabilidad de aumento de consumo y acciones activas como metas de consumo para clientes con baja o mediana probabilidad.

Entre los trabajos futuros se plantea utilizar la metodología en otras áreas de la empresa donde también se cuente con datos transaccionales o bien en otros medios de pago, y se plantea la posibilidad de reformular la variable dependiente de manera generar modelos de fuga que pudiesen complementar a los existentes en el área.

## Agradecimientos

Esta es quizás la sección más difícil de escribir de todo este trabajo, las palabras de agradecimiento a todos los que han estado involucrados en este proceso siempre van a ser pocas:

Me gustaría agradecer en primer lugar a mi familia, por su apoyo incondicional y por todo el esfuerzo realizado para que pudiera cumplir con mi meta de ser un profesional. En particular mi madre y padre fueron muy importantes en esta etapa que ahora termina, sufriendo en cada uno de los momentos difíciles y sonriendo cuando las cosas marcharon bien. A mis hermanos, por toda su alegría y compañía en este tiempo, a pesar de que Jessenia no está con nosotros físicamente sé que nos acompaña siempre junto a Dios. A mis familiares por su preocupación y alegría de vivir, la que de verdad da fuerzas para seguir siempre adelante.

A ti Brenda, por tu amor, compañía y apoyo incondicional. Siempre estuviste a mi lado dándome palabras de aliento y por qué no decirlo, presionándome para que las cosas salieran bien. Tú conoces mejor que cualquiera todo lo que tuvimos que pasar para llegar hasta este punto y es por eso que agradezco tanto que siempre hayas estado ahí. Muchas gracias por estar conmigo como mi compañera de universidad, pero por sobre todo como mi polola.

A mis amigos y amigas. Los aprecio y les agradezco por estar ahí cuando se les necesita y por qué no decirlo, cuando se requieren distracciones!

A mis profesores y personas que aportaron directamente en este trabajo, por toda su colaboración, por su tiempo y principalmente por la confianza depositada en mí.

Finalmente quisiera agradecer a todos aquellos que de alguna manera u otra me acompañaron en este camino.

A todos, simplemente...

Gracias.

## ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN.....	11
1.1. ANTECEDENTES GENERALES.....	11
1.2. PROBLEMA DE NEGOCIO Y JUSTIFICACIÓN.....	13
1.3. OBJETIVOS.....	15
1.3.1. OBJETIVO GENERAL.....	15
1.3.2. OBJETIVOS ESPECÍFICOS.....	15
1.4. ALCANCES.....	16
2. MARCO CONCEPTUAL.....	16
2.1. CRM.....	16
2.2. METODOLOGÍA KDD.....	17
2.3. VARIABLES TRANSACCIONALES RFM.....	19
2.4. CADENAS DE MARKOV.....	21
2.5. ANÁLISIS DE CONGLOMERADOS TIPO K MEDIAS.....	24
2.6. ÁRBOLES DE DECISIÓN.....	25
2.7. REGRESIÓN LOGÍSTICA.....	27
2.8. RED NEURONAL.....	28
2.9. INDICADORES PARA MEDIR LA CALIDAD DE LOS MODELOS.....	30
3. METODOLOGÍA.....	33
3.1. DETALLE METODOLOGÍA.....	33
4. VARIABLES DE MODELAMIENTO.....	34
4.1. DESCRIPCIÓN DE DATOS.....	34
4.1.1. ANÁLISIS UNIVARIADO.....	35
4.1.2. ANÁLISIS MULTIVARIADO.....	38
4.2. SELECCIÓN DE DATOS.....	39
4.2.1. SELECCIÓN DEL LARGO DE LOS PERÍODOS.....	40
4.2.2. DEFINICIÓN DE LAS VENTANAS DE TIEMPO.....	41
4.2.3. DEFINICIÓN DE CLIENTES INACTIVOS.....	42
4.3. TRANSFORMACIÓN.....	42
4.4. PREPROCESAMIENTO.....	43
4.4.1. TRATAMIENTO DE VALORES FALTANTES Y OUTLIERS.....	43

5. MINERÍA DE DATOS.....	46
5.1. SEGMENTACIÓN .....	46
5.1.1. SELECCIÓN DEL NÚMERO DE CLUSTERS .....	46
5.1.2. SEGMENTACIÓN DE CLIENTES POR PERÍODO. ....	48
5.2. CADENAS DE MARKOV Y ELECCIÓN DE LA SEGMENTACIÓN .....	49
5.3. COMPROBACIÓN DE COMPORTAMIENTO MARKOVIANO.....	50
5.4. CADENAS DE MARKOV: MATRIZ DE TRANSICIÓN.....	55
5.5. CADENAS DE MARKOV: ANÁLISIS DE 3 PERÍODOS. ....	56
6. MODELOS DE PROPENSIÓN PARA INCENTIVAR USO DE TARJETA .....	57
6.1. CARACTERIZACIÓN SEGMENTOS OBTENIDOS .....	58
6.1.1. ANÁLISIS DE SEGMENTOS INTERESANTES EN TÉRMINOS DE MONTOS TRANSACCIONADOS.....	59
6.1.2. PERFILES DE CONSUMO DE CADA SEGMENTO.....	60
6.2. DESCRIPCIÓN DE LOS MODELOS DE PROPENSIÓN.....	71
6.3. SELECCIÓN DE DATOS .....	71
6.4. PREPROCESAMIENTO .....	72
6.5. TRANSFORMACIÓN .....	72
6.6. DATA MINING.....	73
6.6.1. BASES TRAIN Y TEST DE MODELO DE PROPENSIÓN .....	73
6.7. RESULTADOS DE LOS MODELOS DE PROPENSIÓN .....	74
6.7.1. RESULTADOS EN SEGMENTO INACTIVOS.....	74
6.7.2. RESULTADOS EN SEGMENTO MUY BAJO CONSUMO .....	77
6.7.3. RESULTADOS EN SEGMENTO BAJO CONSUMO .....	79
6.7.4. RESULTADOS EN SEGMENTO INTERMEDIO.....	81
6.7.5. RESULTADOS EN SEGMENTO ALTO CONSUMO .....	83
6.8. DISCUSIÓN DE LOS RESULTADOS DE LOS MODELOS .....	85
6.8.1. DESEMPEÑO DE LOS MODELOS.....	85
6.8.2. COMPARACIÓN DE LAS REGLAS DE DECISIÓN DE LOS MODELOS ..	86
6.8.3. PERFILES DE CLIENTES QUE SE OBSERVAN A PARTIR DE LAS REGLAS DE DECISIÓN.....	89
7. CONCLUSIONES Y RECOMENDACIONES.....	90
7.1. CONCLUSIONES CON RESPECTO A LOS OBJETIVOS PLANTEADOS .....	90

7.2. CONCLUSIONES DE LAS TÉCNICAS UTILIZADAS.....	91
7.3. CONCLUSIONES DE LOS RESULTADOS DE LOS MODELOS.....	93
8. TRABAJOS FUTUROS.....	96
9. BIBLIOGRAFÍA.....	97
10. ANEXOS.....	98
ANEXO A: Evolución del número de clientes que poseen Tarjeta de Crédito.....	98
ANEXO B: Histogramas de variables RFM posterior al tratamiento de outliers .....	99
ANEXO C: Centros de los <i>clusters</i> obtenidos según las combinaciones de variables RFMN.....	101
ANEXO D: Errores de transición comparando distintas estimaciones. ....	104
ANEXO E: Estadísticos descriptivos de las distintas segmentaciones realizadas .....	110
ANEXO F: Detalle de tablas truncadas en sección 6.1.2 .....	118
ANEXO G: Esquemas de los modelos de Árboles de decisión para los distintos segmentos.....	126
ANEXO H: Diccionario de Variables Árboles de Decisión.....	134

## ÍNDICE DE FIGURAS

Figura 1: Participación Colocaciones Marzo 2009.....	11
Figura 2: Distribución sectorial de la Cartera de Colocaciones a marzo de 2009 .....	12
Figura 3: Gráfico Rentabilidad BancoEstado.....	12
Figura 4: Stock y Cuentas con Movimiento, Octubre 2009 .....	13
Figura 5: Penetración de Tarjeta de Crédito, Octubre 2009 .....	14
Figura 6: Evolución de la rentabilidad de un cliente de la Industria Bancaria a través del tiempo .....	17
Figura 7: Etapas del proceso KDD .....	19
Figura 8: Variables RFM.....	20
Figura 9: Ejemplo matriz de Transición .....	21
Figura 10: Ejemplo Cadena de Markov .....	22
Figura 11: Ejemplo análisis de 3 periodos.....	23
Figura 12: Matriz de Transición .....	24
Figura 13: Perceptrón.....	29
Figura 14: Red Neuronal Multicapa .....	30
Figura 15: Matriz de Confusión .....	31
Figura 16: Ejemplo Curvas ROC .....	32
Figura 17: Ejemplo curva de ganancia .....	32
Figura 18: Muestra de clientes seleccionados.....	36
Figura 19: Histograma variable <i>Recency</i> .....	37
Figura 20: Histograma variable <i>Frequency</i> .....	37
Figura 21: Histograma variable <i>Monetary Value</i> .....	38
Figura 22: Muestra de clientes seleccionados tras la aplicación de los filtros.....	40
Figura 23: Histograma Variable <i>Monetary Value</i> .....	44
Figura 24: Histograma Variable N (número de transacciones).....	45
Figura 25: Gráfico número de <i>clusters</i> versus Distancia (variable RM).....	47
Figura 26: Gráfico número de <i>clusters</i> versus Distancia (variables NM).....	47
Figura 27: Matriz de transición de un periodo .....	55
Figura 28: Matriz de transición de 2 Periodos .....	56
Figura 29: Matriz de transición de 3 Periodos .....	57
Figura 30: Caracterización de los Segmentos.....	59
Figura 31: Curvas de Ganancia de Información para el Modelo del Segmento Inactivos .....	75
Figura 32: Curva de Ganancia de Información del Modelo para el Segmento de Muy Bajo Consumo .....	78
Figura 33: Curvas de ganancia de información para el Modelo del Segmento de Bajo Consumo .....	80

Figura 34: Curvas de Ganancia de Información para el Modelo del Segmento Intermedio .....	82
Figura 35: Curvas de Ganancia de Información para el Modelo del Segmento de Alto Consumo .....	84
Figura 36: Evolución del número de clientes con Tarjeta de Crédito .....	98
Figura 37: Histograma variable Recency (excluidos outliers).....	99
Figura 38: Histograma variable F (excluidos outliers) .....	99
Figura 40: Histograma variable N (excluidos outliers) .....	100
Figura 41: Esquema del modelo de Árbol de decisión para el segmento Inactivos.....	127
Figura 42: Esquema del modelo de Árbol de decisión para el segmento Muy bajo consumo.....	130
Figura 43: Esquema modelo de Árbol de decisión para el segmento Bajo consumo...	131

#### ÍNDICE DE TABLAS

Tabla 1: Definición variables RFM y N .....	20
Tabla 2: Definición de Indicadores de comparación entre modelos .....	31
Tabla 3: Detalle de las variables transaccionales a utilizar .....	35
Tabla 4: Estadísticos descriptivos variables RFM .....	36
Tabla 5: Correlaciones entre variables .....	38
Tabla 6: Definición de los períodos .....	42
Tabla 7: Estadísticos Descriptivos variables RFM (periodo 7) .....	43
Tabla 8: Estadísticos descriptivos periodo 7(outliers tratados).....	45
Tabla 9: Número de <i>clusters</i> óptimo según variables.....	48
Tabla 10: Ejemplo de los centros de <i>clusters</i> según variables FNM .....	48
Tabla 11: Errores de las distintas segmentaciones .....	52
Tabla 12: Descripción de los segmentos con respecto a las variables RFMN .....	53
Tabla 13: Resultado de la subdivisión del segmento 4 de la segmentación FNM.....	54
Tabla 14: Segmentos finales y estadísticos descriptivos promedio de las variables FNM .....	55
Tabla 15: Estadísticos descriptivos de cada segmento.....	58
Tabla 16: Tamaño e importancia de los segmentos.....	60
Tabla 17: Agrupación General.....	61
Tabla 18: Agrupación por Tipo de compra .....	61
Tabla 19: Agrupación por Rubro de compra.....	62
Tabla 20: Agrupación General segmento Muy bajo consumo .....	63
Tabla 21: Agrupación General segmento Bajo consumo .....	63
Tabla 22: Agrupación General segmento Intermedio .....	63
Tabla 23: Agrupación General segmento Alto consumo .....	64
Tabla 24: Agrupación General segmento <i>Heavy Users</i> .....	64
Tabla 25: Agrupación por Tipo de compra segmento Muy Bajo Consumo .....	65



Tabla 26: Agrupación por Tipo de compra segmento Bajo Consumo .....	65
Tabla 27: Agrupación por Tipo de compra segmento Intermedio.....	66
Tabla 28: Agrupación por Tipo de compra segmento Alto Consumo .....	66
Tabla 29: Agrupación por Tipo de compra segmento <i>Heavy Users</i> .....	67
Tabla 30: Agrupación por Rubro de compra segmento Muy Bajo Consumo.....	68
Tabla 31: Agrupación por Rubro de compra segmento Bajo Consumo .....	68
Tabla 32: Agrupación por Rubro de compra segmento Intermedio .....	69
Tabla 33: Agrupación por Rubro de compra segmento alto consumo.....	69
Tabla 34: Agrupación por Rubro de compra segmento <i>Heavy Users</i> .....	70
Tabla 35: Comparación de los indicadores de los distintos modelos Segmento Inactivo .....	75
Tabla 36: Comparación de los indicadores de distintos modelos.....	78
Tabla 37: Comparación de indicadores de distintos modelos segmento Bajo consumo	80
Tabla 38: Comparación de los indicadores de distintos modelos Segmento Intermedio	82
Tabla 39: Comparación de los indicadores de distintos modelos Segmento Alto Consumo.....	84
Tabla 41: Centros de <i>clusters</i> según la variable Frecuency.....	101
Tabla 42: Centros de <i>clusters</i> según la variable Monetary.....	101
Tabla 43: Centros de <i>clusters</i> según la variable N.....	101
Tabla 44: Centros de <i>clusters</i> según variables RF.....	102
Tabla 45: Centros de <i>clusters</i> según variables RM.....	102
Tabla 46: Centros de <i>clusters</i> según variables FM.....	102
Tabla 47: Centros de <i>clusters</i> según variables FN.....	102
Tabla 48: Centros de <i>clusters</i> según variables RN.....	103
Tabla 49: Centros de <i>clusters</i> según variables NM.....	103
Tabla 50: Centros de <i>clusters</i> según variables RFM.....	103
Tabla 51: Centros de <i>clusters</i> según las variables RNM.....	103
Tabla 53: Centros de <i>clusters</i> según variables RFNM.....	104
Tabla 54: Errores de transición con distintas estimaciones (variable R) .....	104
Tabla 55: Errores de transición con distintas estimaciones (variable F).....	105
Tabla 56: Errores de transición con distintas estimaciones (variable M).....	105
Tabla 57: Errores de transición con distintas estimaciones (variable N) .....	105
Tabla 58: Errores de transición con distintas estimaciones (variables RF) .....	106
Tabla 59: Errores de transición con distintas estimaciones (variables RM) .....	106
Tabla 60: Errores de transición con distintas estimaciones (variables FM).....	106
Tabla 61: Errores de transición con distintas estimaciones (variables FN) .....	107
Tabla 62: Errores de transición con distintas estimaciones (variables RN).....	107
Tabla 63: Errores de transición con distintas estimaciones (variables NM) .....	107
Tabla 64: Errores de transición con distintas estimaciones (variables RFM) .....	108
Tabla 65: Errores de transición con distintas estimaciones (variables RNM).....	108
Tabla 66: Errores de transición con distintas estimaciones (variables RFN).....	108
Tabla 67: Errores de transición con distintas estimaciones (variables FNM) .....	109

Tabla 68: Errores de transición con distintas estimaciones (variables RFMN).....	109
Tabla 69: Descripción de los segmentos con respecto a las variables RFMN (variable R) .....	110
Tabla 70: Descripción de los segmentos con respecto a las variables RFMN (variable F) .....	110
Tabla 71: Descripción de los segmentos con respecto a las variables RFMN (variable M) .....	111
Tabla 72: Descripción de los segmentos con respecto a las variables RFMN (variable N) .....	112
Tabla 73: Descripción de los segmentos con respecto a las variables RFMN (variables RF) .....	112
Tabla 74: Descripción de los segmentos con respecto a las variables RFMN (variables RM) .....	113
Tabla 76: Descripción de los segmentos con respecto a las variables RFMN (variables FN) .....	114
Tabla 77: Descripción de los segmentos con respecto a las variables RFMN (variables RN).....	115
Tabla 78: Descripción de los segmentos con respecto a las variables RFMN (variables NM) .....	115
Tabla 79: Descripción de los segmentos con respecto a las variables RFMN (variables RFM) .....	116
Tabla 80: Descripción de los segmentos con respecto a las variables RFMN (variables RNM).....	116
Tabla 81: Descripción de los segmentos con respecto a las variables RFMN (variables RFN).....	117
Tabla 82: Descripción de los segmentos con respecto a las variables RFMN (variables RFNM).....	117
Tabla 83: Detalle categoría OTROS en tabla 25 .....	118
Tabla 84: Detalle categoría OTROS en tabla 26 .....	119
Tabla 85: Detalle categoría OTROS en tabla 27 .....	119
Tabla 86: Detalle categoría OTROS en tabla 28.....	120
Tabla 87: Detalle categoría OTROS en tabla 29 .....	120
Tabla 89: Detalle categoría OTROS en tabla 31 .....	122
Tabla 90: Detalle categoría OTROS en tabla 32 .....	123
Tabla 91: Detalle categoría OTROS en tabla 33 .....	124
Tabla 92: Detalle categoría OTROS en tabla 34 .....	125
Tabla 93: Diccionario variables segmento Inactivos.....	134
Tabla 94 : Diccionario variables segmento Muy bajo consumo.....	136
Tabla 95: Diccionario variables segmento Bajo consumo .....	138
Tabla 96: Diccionario variables segmento Intermedio.....	139
Tabla 97: Diccionario variables segmento Alto consumo .....	141

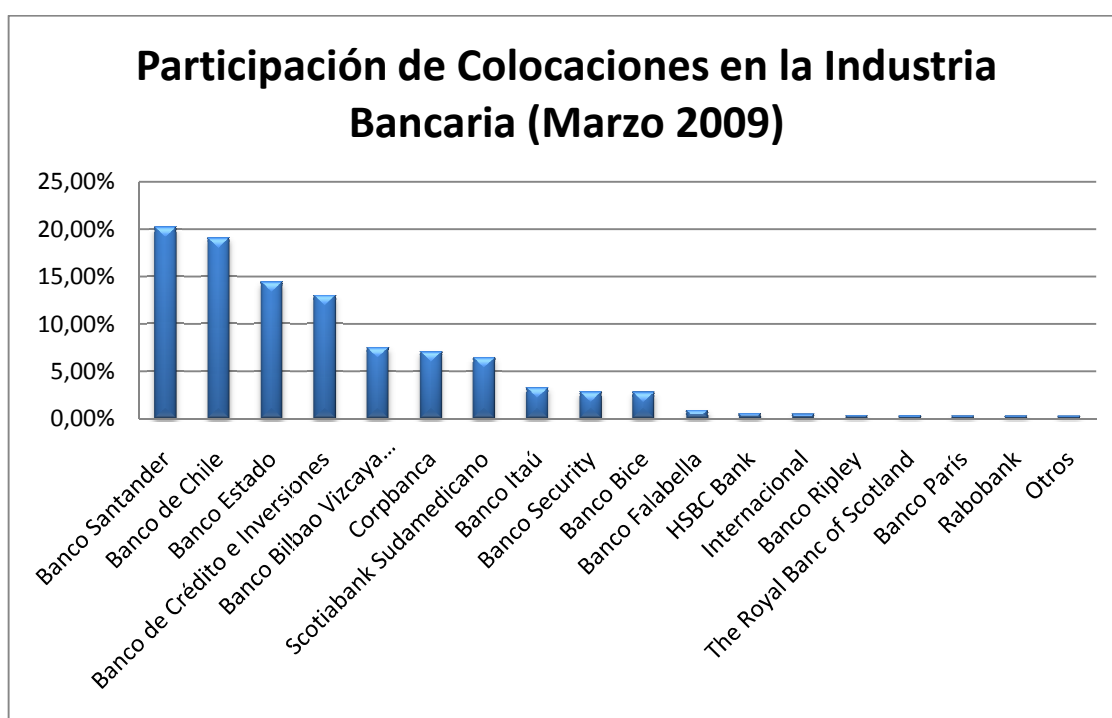
## 1. INTRODUCCIÓN

Este trabajo se desarrolla dentro del contexto de la industria bancaria en Chile, en particular en BancoEstado por lo cual se presentan antecedentes generales que permiten contextualizar y justificar el desarrollo de una metodología que aborde el problema de negocio de interés. Además se establecen los objetivos generales y específicos que se pretenden lograr, y los alcances y limitaciones de este trabajo.

### 1.1. ANTECEDENTES GENERALES

La industria bancaria chilena consta de 25 entidades, siendo 5 de ellas las que a marzo del 2009 concentran casi el 75% de las colocaciones totales:

Figura 1: Participación Colocaciones Marzo 2009

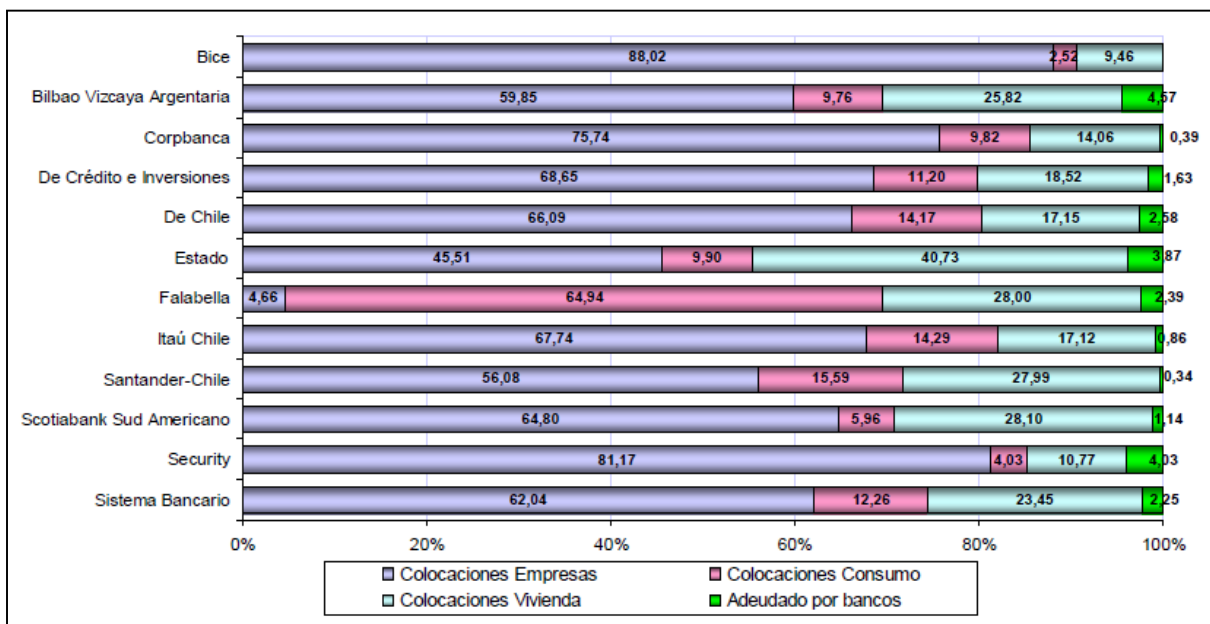


Fuente: Panorama de la Industria Bancaria, Marzo 2009, SBIF

Las colocaciones de las distintas instituciones se vieron disminuidas durante el año 2008 debido al impacto de la crisis económica a nivel mundial, sin embargo BancoEstado lidera el crecimiento de las colocaciones (3,67%) en todas las carteras llegando a una participación de mercado del 14.43%.

Las colocaciones de BancoEstado tienen un perfil distinto al resto del sistema, teniendo especial participación en vivienda, lo cual se aprecia en la siguiente figura:

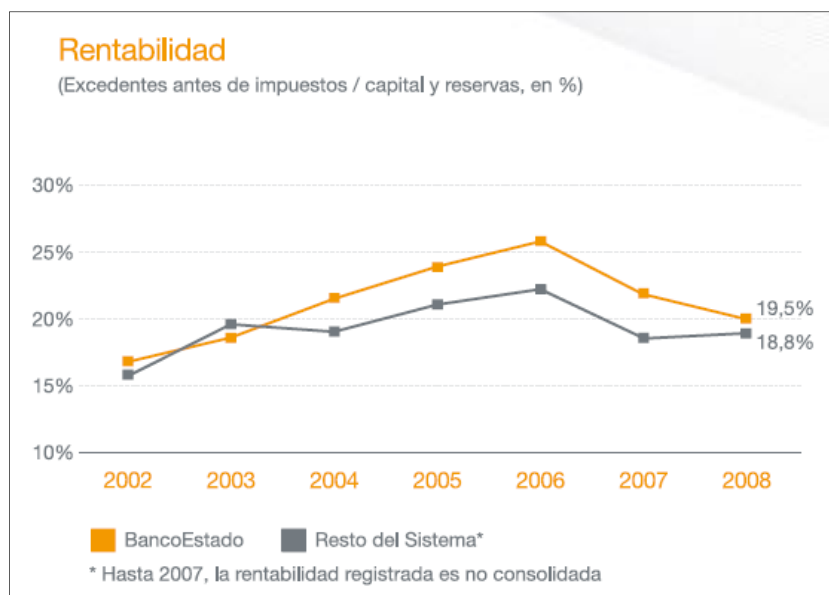
**Figura 2: Distribución sectorial de la Cartera de Colocaciones a marzo de 2009**



Fuente: Panorama de la Industria Bancaria, Marzo 2009, SBIF

Con respecto su rentabilidad el siguiente gráfico muestra que ésta ha sido superior al promedio de la industria desde el año 2004, situándolo en una sólida posición frente a sus competidores.

**Figura 3: Gráfico Rentabilidad BancoEstado**



Fuente: Memoria 2008 BancoEstado

Para los más de 8 millones de clientes que de la Subgerencia Banca Personas, BancoEstado posee 4 grandes líneas de productos:

- Ahorro
- Seguros
- Créditos
- Medios de Pago

Dentro de los medios de pago con los que cuenta se pueden nombrar:

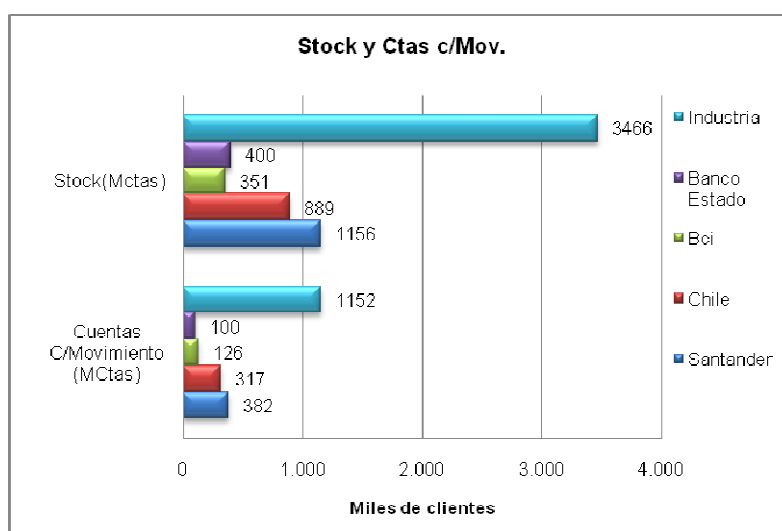
- Tarjeta de Crédito
- Cuenta Corriente
- Chequera electrónica
- Tarjeta cuota fija

El tema a desarrollar se relaciona con el producto medio de pago y en particular con Tarjeta de crédito, producto en donde BancoEstado ocupa el tercer lugar en participación de mercado(11,6%) y el primer lugar en crecimiento (49,9%) en el periodo Octubre 2008-Octubre 2009 en cuanto a stock de tarjetas de créditos (número de tarjetas circulantes).<sup>1</sup>

## 1.2. PROBLEMA DE NEGOCIO Y JUSTIFICACIÓN

BancoEstado posee un stock a octubre del año 2009 que asciende a 400 mil tarjetas de crédito, de las cuales sólo un 25% se encuentran activas. Si se compara este porcentaje con el de la industria, se observa que para la competencia relevante este valor es en promedio un 33%, en particular los bancos Santander y Chile tienen 33,1% y 35,7% respectivamente lo cual se aprecia en la siguiente ilustración:

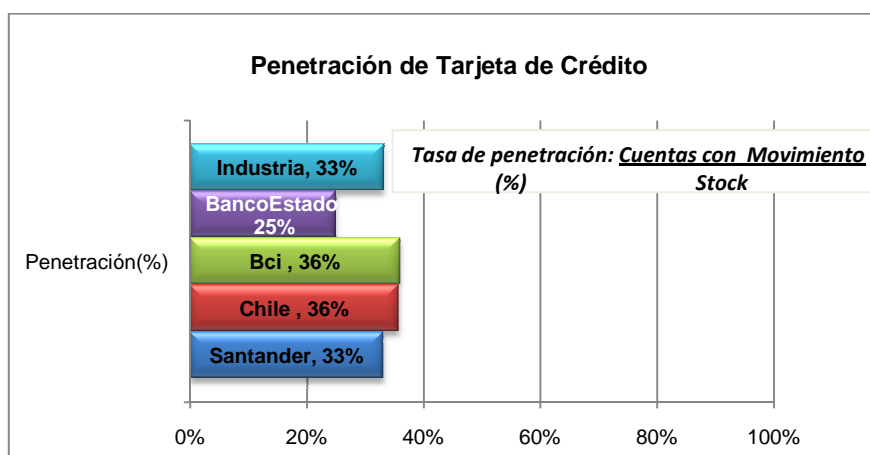
Figura 4: Stock y Cuentas con Movimiento, Octubre 2009



Fuente: INFORME DE ACTIVIDAD DE TARJETAS DE CREDITO BANCOESTADO Y SISTEMA FINANCIERO A OCTUBRE 2009

<sup>1</sup> Fuente: Superintendencia de Bancos e Instituciones Financieras(SBIF)

Figura 5: Penetración de Tarjeta de Crédito, Octubre 2009



Fuente: INFORME DE ACTIVIDAD DE TARJETAS DE CREDITO BANCOESTADO Y SISTEMA FINANCIERO A OCTUBRE 2009

El menor porcentaje de este índice de penetración de tarjeta de crédito en BancoEstado con respecto a la industria justifica y hace necesario estudiar formas de incentivar el uso de este producto entre los clientes de menor consumo y en particular en aquellos clientes inactivos (75%).

Uno de los intereses por incentivar el uso de la tarjeta de crédito entre los clientes lo constituye el modelo de negocio de este producto, que genera variados ingresos de los cuales un porcentaje es recibido por la institución:

- Cobro de mantención: Por usar la tarjeta de crédito se cobra mensualmente un cargo fijo asociado, el que debe ser cancelado siempre que se utiliza la tarjeta.
- Comisiones a los comercios asociados al pago con tarjeta de crédito: Cada compra que se realiza con este medio de pago en cualquier tipo de comercio tiene incorporada una comisión con cargo al contratante.
- Cobro por realizar un avance: Por cada avance que se realiza, se cobra un valor fijo.
- Intereses por avances: Los avances en una o varias cuotas, tienen un interés asociado al número de cuotas en que se realiza.
- Cobros por no pago de tarjeta: Cuando una persona debe pagar un avance y lo realiza posteriormente a la fecha de vencimiento, la cuota pasa a un proceso de *Revolving* en el cual se aplican intereses superiores a los iniciales y que son del orden de un crédito de consumo o superior dependiendo de las características de cada caso.

- Cobro anual por posesión de tarjeta de crédito: Anualmente se debe pagar un monto por tener tarjeta de crédito.

Como se observa el modelo de negocio de tarjeta de crédito posee numerosas formas de generar ganancias, y por esto es considerada como uno de los productos más rentables tanto de la industria bancaria como del retail.

Además de la rentabilidad propia de la tarjeta de crédito, se señala que un cliente activo genera una serie de beneficios:

- La tarjeta de crédito es un producto vinculante, ya que una vez que un cliente comienza a hacer uso de ésta, es difícil que deje de transaccionar, generando un círculo virtuoso en el cual se vincula más al cliente.
- Se crea historia de los clientes en cuanto a gustos y preferencias de compra.
- Es un producto que requiere una cantidad mínima de fuerza de ventas para ser rentable, en comparación, por ejemplo, con la venta de créditos de consumo.

En síntesis, la tarjeta de crédito es un producto de alta importancia para el banco y por lo tanto cualquier acción que incentive el uso de ésta por parte de los clientes de bajo consumo y en particular de los inactivos generará mejoras en la rentabilidad.

**El problema de negocio a resolver radica en que el 75% de los clientes que cuentan con tarjeta de crédito, no la utilizan.**

**Así es como nace el presente tema de memoria, donde se busca mediante modelos de propensión incentivar el uso de tarjeta de crédito principalmente entre los segmentos de bajo consumo y en particular lograr activar a clientes inactivos.**

### **1.3. OBJETIVOS**

#### **1.3.1. OBJETIVO GENERAL**

Diseñar modelos de propensión para incentivar el uso de tarjeta de crédito entre los clientes con menor consumo y en particular entre los clientes inactivos.

#### **1.3.2. OBJETIVOS ESPECÍFICOS**

- Caracterizar la dinámica de uso de tarjeta de crédito a través del tiempo y los perfiles de consumo de los segmentos que son definidos.
- Crear modelos de propensión para identificar a los clientes que potencialmente pueden aumentar el uso de su tarjeta de crédito.

## 1.4. ALCANCES

- La información es confidencial, por tanto no se mostrarán datos a nivel de cliente.
- El modelo de cadenas de Markov entregará probabilidades por grupos de clientes, no se obtendrán valores a nivel desagregado.
- No se contempla el diseño de campañas a realizar a cada segmento, ya que esto depende de las políticas de la institución en cuestión, sin embargo se entregarán recomendaciones generales con respecto al tipo de estrategias a utilizar en función de los resultados obtenidos.

## 2. MARCO CONCEPTUAL

En este capítulo se describe brevemente los distintos conceptos y metodologías que se utilizarán para abordar las distintas etapas de este trabajo, entre las cuales destacan: Metodología KDD, variables transaccionales RFMN, cadenas de Markov y algoritmos de predicción utilizados en *Data Mining*.

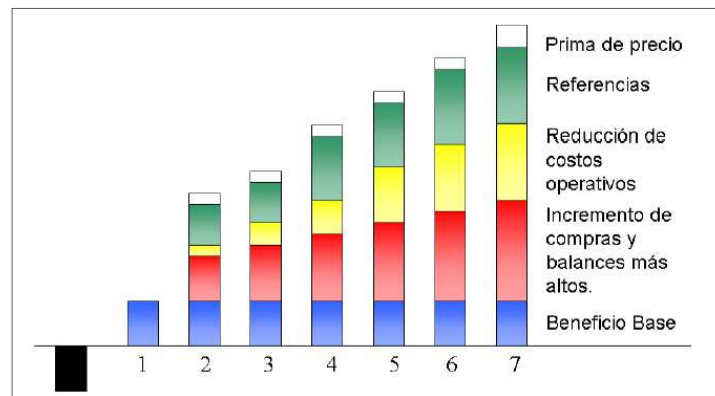
### 2.1. CRM

Hoy en día la relación cliente-empresa no se debe mirar como un proceso de simple venta de un producto o servicio, este concepto ha cambiado y se entiende que las relaciones que se puedan formar con los clientes a lo largo del tiempo tienen un alto impacto con respecto al beneficio que se puede extraer de ellas.

El concepto de CRM (*Customer Relationship Management*) consiste en estrategias enfocadas a construir relaciones duraderas y rentables con los clientes. Un conocimiento acabado de éstos basado en su comportamiento transaccional puede ayudar a construir estrategias más efectivas que permitan generar ofertas adecuadas a las necesidades de los clientes, entregadas en el momento oportuno y comunicadas por canales más efectivos; lo que se resume en la entrega de una oferta personalizada. Realizar estas acciones genera relaciones de largo plazo con los clientes, lo que se traduce en una mayor rentabilidad tal como se muestra en el siguiente gráfico [13]:



Figura 6: Evolución de la rentabilidad de un cliente de la Industria Bancaria a través del tiempo



Fuente: "Quality comes to services, Harvard Business Review"

El cliente en un principio se traduce en costos para la institución, pero a medida que transcurre el tiempo y comienza a adquirir nuevos productos y servicios su rentabilidad crece sostenidamente haciendo mucho más beneficioso mantener y fidelizar a los miembros de la cartera en vez de realizar esfuerzos en captar nuevos clientes.

CRM propone una serie de etapas que permiten crear lazos de confianza entre la empresa y los clientes, entre las cuales se encuentran:

- Levantar conocimiento acerca de los clientes.
- Entender y diferenciar a los clientes.
- Crear propuestas valiosas de acuerdo a sus necesidades.
- Mejorar la experiencia de los clientes con la empresa.

Este trabajo se enfoca en el segundo tópico: ya que se realizará una segmentación que permita diferenciar a los clientes en función de su transaccionalidad; se modelará su evolución a lo largo del tiempo; se perfilará a los mismos para observar diferencias entre grupos y finalmente se crearán modelos de propensión que permitan entender algunas de las variables que hacen que un cliente incremente su consumo con la tarjeta de crédito [9].

## 2.2. METODOLOGÍA KDD

La metodología o proceso KDD (*Knowledge Discovery in Databases*) puede ser traducido como "Descubrimiento del conocimiento en las bases de datos" y es definido como "El proceso no-trivial de identificar patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles dentro de los datos".

Este proceso se caracteriza por ser iterativo, de modo que no se trata de un sistema que analiza automáticamente los datos y entrega como resultado un conocimiento útil,

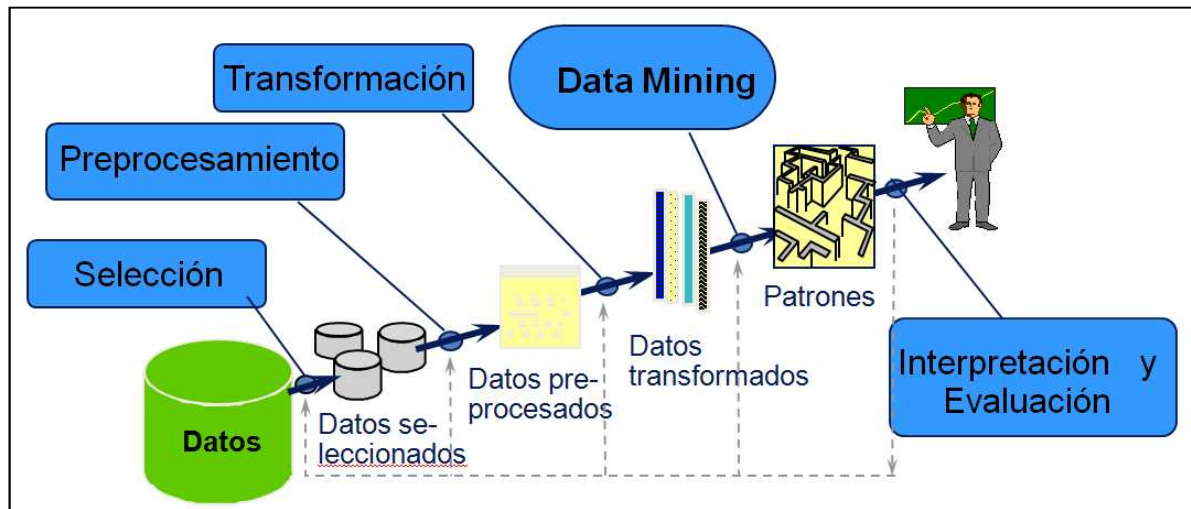
sino que requiere el desarrollo de una serie de etapas para obtener los resultados esperados.

El proceso KDD se compone principalmente de 5 etapas, las cuales son las siguientes:

1. Selección de Datos: Se asume la realización de una etapa previa donde se recopilan datos sobre el proyecto en estudio. Más específicamente, en esta etapa se seleccionan aquellos datos que son relevantes para el estudio. La selección puede abarcar tablas, atributos y datos. Con una adecuada elección de éstos últimos mejora el desempeño de los algoritmos a utilizar, requiriendo menor tiempo de procesamiento y facilita el entendimiento del conocimiento obtenido.
2. Pre-procesamiento: En esta etapa se toman los datos seleccionados y se “limpian”, eliminando datos incoherentes, registros duplicados, y aplicando tratamiento de datos fuera de rango (*outliers*). El pre-procesamiento es la fase donde generalmente se ocupa la mayor parte del tiempo para analizar y preparar los datos y por tal motivo es muy importante ya que una buena preparación de los datos es determinante en la calidad de los resultados.
3. Transformación: En esta etapa se enriquece la base de datos mediante la creación y/o modificación de las variables existentes. Por ejemplo: estandarización de los datos; creación de nuevos atributos a partir de los que ya se tiene (mediante operaciones algebraicas); re-codificación de atributos; transformación de variables categóricas a variables con valores 0 o 1, etc. En esta etapa también se hace agrupamiento de atributos, dependiendo de las características de los mismos.
4. Data Mining: Es la etapa central del proceso KDD, donde se extrae el conocimiento de las bases de datos mediante distintos modelos que permiten identificar patrones en los datos que anteriormente no eran observables.
5. Interpretación y Evaluación: En esta etapa se analiza el resultado de todo el proceso, interpretando y evaluando el resultado de los algoritmos usados. Es aquí donde se extrae para el usuario final los patrones más importantes, para esto es necesario establecer parámetros que le permitan al usuario comparar la calidad de un patrón con respecto a otro. Un ejemplo de parámetro de comparación es la precisión del modelo, comparando el nivel de error de cada uno de ellos.

Las etapas del proceso KDD pueden ser vistas gráficamente con el siguiente esquema:

Figura 7: Etapas del proceso KDD



Fuente: Apuntes curso IN643: Introducción a minería de datos.

### 2.3. VARIABLES TRANSACCIONALES RFM

La creación de estas variables necesita datos del comportamiento histórico de las transacciones de los clientes, sirven para crear segmentaciones desde la perspectiva transaccional y son la base para modelos de proyección del comportamiento de clientes.

Al contar con los datos transaccionales se puede observar cuándo fue la última vez que realizó una transacción en un período determinado y con esto se define el *Recency* (R). Conociendo las fechas en que realizaron las transacciones, se puede calcular un promedio de diferencia de días entre éstas y así se determina la variable *Frequency* (F). Si se tiene el dato del monto de cada una de las transacciones analizadas, se calcula el valor promedio de las mismas, con lo cual se determina el *Monetary Value* (M).

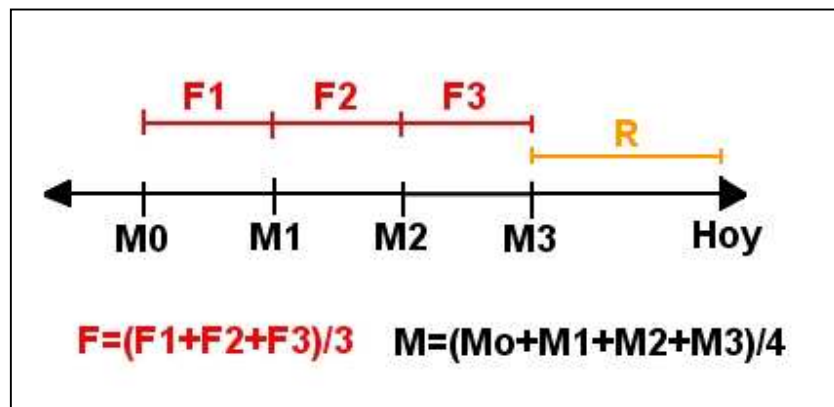
Con el objetivo de complementar la información entregada por estas 3 variables y de este modo generar segmentaciones más robustas se incorpora la variable N definida como el número de transacciones realizadas.

Tabla 1: Definición variables RFM y N

Variable	Definición
R ( <i>Recency</i> )	Tiempo transcurrido desde la última transacción hasta el momento de la evaluación (días)
F ( <i>Frequency</i> ):	Número de días promedio entre una transacción y otra <sup>2</sup> .
M ( <i>Monetary Value</i> ):	Monto promedio de las transacciones (UF)
N	Número de transacciones en el período de tiempo

Se puede observar la definición de las 3 primeras variables de manera gráfica en la siguiente figura:

Figura 8: Variables RFM



FUENTE: Memoria SEGOVIA R., C.A. 2005

Algunas de las ventajas y desventajas de utilizar estas variables son [1]:

#### Ventajas

- Las variables RFM contienen la información más relevante del comportamiento transaccional de un cliente.
- Es de fácil comprensión.
- Permiten identificar y observar la evolución transaccional de los clientes.
- Permiten focalizar las medidas de fidelización de los clientes.

<sup>2</sup> Dado que se trabajará con ventanas de tiempo sucesivas la variable F considerará el promedio días entre transacciones de un período de tiempo, pero incluirá en este promedio el tiempo transcurrido entre la última transacción del período anterior y la primera del período de análisis, de modo de no perder información.

## Desventajas

- No siempre se cuenta con la confiabilidad de los datos.
- Disponibilidad de los datos (debe haber una cantidad razonable de historia transaccional)
- Manejo de gran cantidad de datos dispersos
- La variable F definida para este trabajo como el promedio de días entre transacciones, no siempre permite identificar entre dos clientes que tengan la misma frecuencia, pero que difieren en gran medida en el número de transacciones<sup>3</sup>.

Si en otro contexto se necesitara mejorar las segmentaciones, se podría complementar el uso de estos índices con variables que puedan ser relevantes como edad, renta, sexo, etc.

[1], [11], [15]

## 2.4. CADENAS DE MARKOV

Las cadenas de Markov son un proceso estocástico, es decir, modelos de probabilidad para procesos que evolucionan en el tiempo. Estas son utilizadas para observar el comportamiento de un sistema a lo largo del tiempo, por ejemplo, seguir los distintos estados por los cuales transita una persona hasta llegar a una fase final. La gran ventaja de este modelo es que el estado al cual evolucionará el sistema, solo depende del estado actual en el cuál se encuentra y no de la historia pasada. Una cadena de Markov tiene dos elementos que la definen:

- Estados: corresponde a los distintos niveles de algún parámetro entre los cuales un sistema puede evolucionar. En este caso, los estados están representados por los distintos segmentos de RFM por los cuales puede pasar una persona [7]. Los estados normalmente son representados de manera gráfica como nodos.
- Probabilidad de transición: Corresponden a las probabilidades de pasar de un estado a otro en un período determinado. Estas probabilidades quedan representadas mediante la llamada “matriz de transición” P.

Figura 9: Ejemplo matriz de Transición

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix}$$

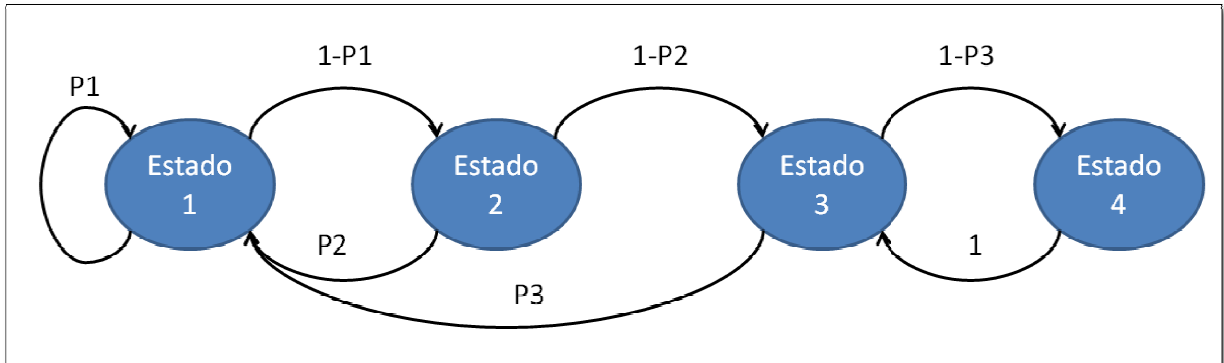
FUENTE: Elaboración propia

---

<sup>3</sup> Para resolver este problema se incluye la variable N: número de días entre transacciones.

Las cadenas de Markov gráficamente se pueden ver de la siguiente manera:

Figura 10: Ejemplo Cadena de Markov



FUENTE: Elaboración propia

Los distintos nodos representan diferentes estados por los cuales puede pasar el sistema. También se observan las probabilidades asociadas al paso de un estado a otro. Por ejemplo, la probabilidad de pasar del estado 2 al 3 es  $(1-P_2)$  en un período.

Los procesos de Markov cumplen con dos propiedades básicas:

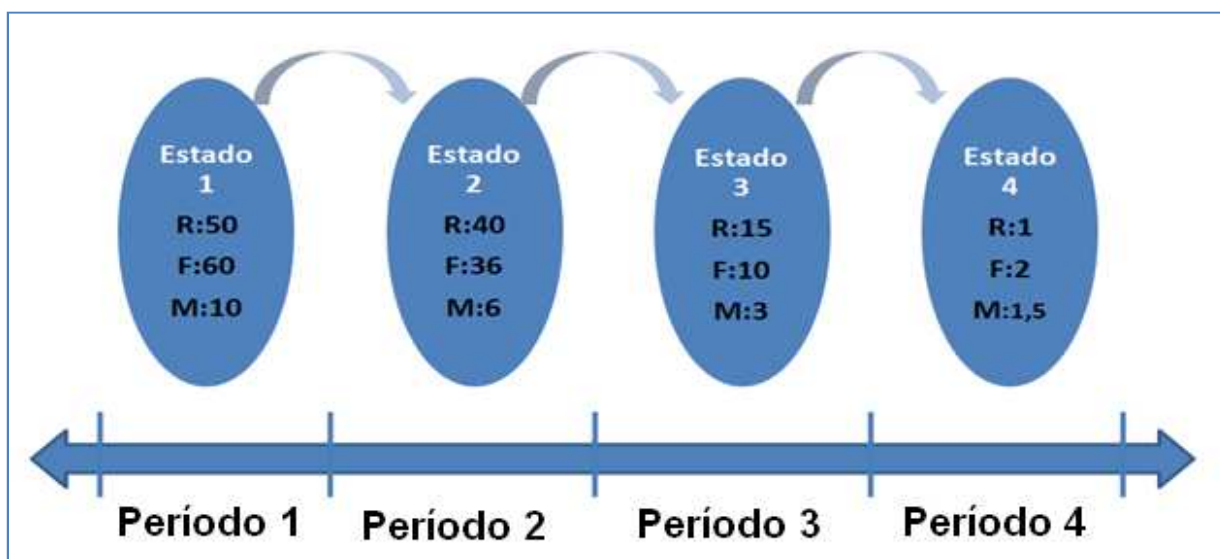
- La suma de las probabilidades de transición desde un estado fijo a los demás es igual a 1.
- Para saber cuáles son las probabilidades de transición de aquí a  $N$  períodos más, basta multiplicar  $P$  por si misma  $N$  veces.

Esta metodología se utilizará para modelar la evolución de los clientes a lo largo del tiempo a través de los distintos estados o segmentos creados en función de las variables RFMN [9].

En el presente trabajo las cadenas de Markov se utilizan de la siguiente manera: Los estados de la cadena serán los segmentos RFMN que deben ser definidos en la sección de segmentación. Las probabilidades de transición representan la probabilidad de que los clientes cambien desde un estado RFMN a otro entre dos períodos de tiempo, permitiendo observar la evolución transaccional de los clientes [7].

Para apreciar esta evolución en un período de tiempo más extenso se realizará un análisis de 3 períodos: se observa el estado inicial y final de los clientes, el cuál corresponderá a 3 períodos después del inicial. Una explicación gráfica de este análisis se muestra en la siguiente figura:

Figura 11: Ejemplo análisis de 3 periodos



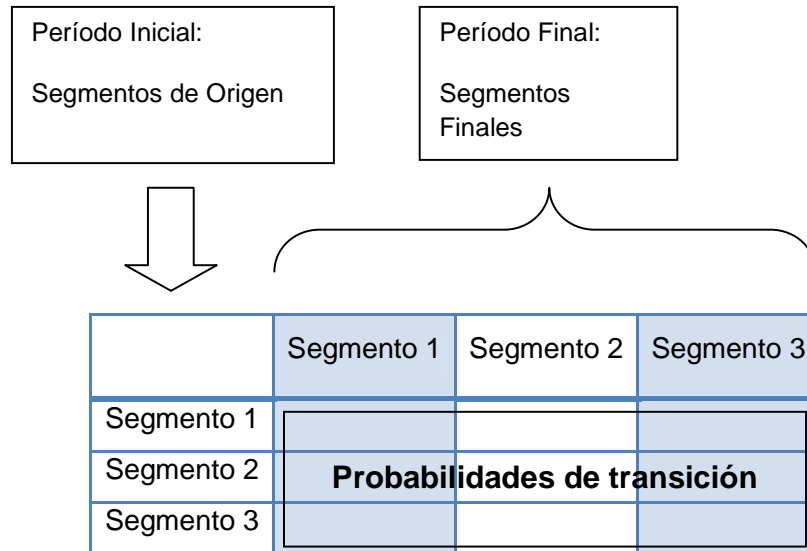
Fuente: Elaboración propia

Mediante el análisis de 3 periodos se pueden detectar “camino” que llevan a un cliente de bajo consumo a convertirse en un *Heavy User* (cliente de alto consumo). Se comienza en un estado inicial (Estado 1) y se observa la evolución de los clientes luego de 3 periodos (Estado 4) obteniéndose las trayectorias antes señaladas.

La forma intuitiva de representar una cadena de Markov es mediante un diagrama que muestra nodos y flechas, que representan estados y transiciones respectivamente<sup>4</sup>. Estos diagramas pueden ser confusos para mostrar gran cantidad de estados y transiciones, por lo cual la representación de la cadena de Markov, en este caso, se realizará mediante matrices de transición.

<sup>4</sup> Figura 10

Figura 12: Matriz de Transición



## 2.5. ANÁLISIS DE CONGLOMERADOS TIPO K MEDIAS

El análisis de conglomerados es una técnica multivariante cuya finalidad es dividir un conjunto de objetos en grupos o *clusters* de forma que los perfiles de los objetos del mismo grupo sean muy similares entre sí, y los de los objetos de *segmentos* diferentes sean distintos.

Esta técnica a diferencia de otras es menos exigente en cuanto a los supuestos que deben cumplir los datos, como de linealidad y simetría. Además no asume ninguna distribución específica para las variables y permite el uso de varios métodos de estimación de la matriz de distancias.

Existen dos tipos de análisis de conglomerados: el análisis de conglomerados jerárquico y el análisis de conglomerados de K medias. El método jerárquico permite determinar el número óptimo de conglomerados existente en los datos sin requerir un conocimiento a priori del número de *clusters* a obtener.

El método de K medias permite procesar un número ilimitado de casos, pero sólo da la posibilidad de utilizar un método de aglomeración y requiere que se proponga previamente el número de conglomerados que se desea obtener.

El proceso de asignación de casos a conglomerados según K medias se describe a continuación:

- a. Particionar el conjunto de observaciones en k grupos iniciales arbitrarios definidos por el usuario.
- b. Recorrer todas las observaciones, asignándolas al conglomerado cuyo centroide esté a menor distancia. Cada vez que se reasigna una observación a un conglomerado



distinto del que la contenía se deben recalculan los centroides del que pierde la observación y del que la recibe.

Si el conglomerado A (que consiste en  $n_A$  observaciones) pierde la observación  $x_i$  y si el conglomerado B (con  $n_B$  observaciones) recibe a  $x_i$ , los centroides respectivos  $x_A$  y  $x_B$  se modifican de la siguiente forma:

$$x'_A = \frac{1}{n_A - 1} (n_A x_A - x_i)$$

$$x'_B = \frac{1}{n_B + 1} (n_B x_B + x_i)$$

c. Repetir el paso b hasta que no haya más reasignaciones.

El método de K medias dentro de su procedimiento utiliza la distancia euclidiana para medir la distancia entre los casos, la cual se calcula de la siguiente manera:

$$d_{ii'} = \sqrt{\sum_j (X_{ij} - X_{i'j})^2}$$

Donde X se refiere a las puntuaciones obtenidas por el caso i y el caso i' (i es distinto a i') en cada una de las  $j=1, \dots, p$  variables incluidas en el análisis.

El método de K medias dentro de esta memoria se utilizará para realizar la segmentación de los clientes. Para esto se ocupará la técnica de manera exploratoria, probando con distintos valores de K (conglomerados iniciales) hasta encontrar cuál es número más apropiado de grupos a formar, en función de criterios de negocios y juicio de expertos.

## 2.6. ÁRBOLES DE DECISIÓN

Los árboles de decisión son modelos de predicción utilizados en *data mining* que generan estructuras en forma de árbol para generar reglas para la clasificación de un conjunto de datos. Se utilizan para descubrir patrones, los cuales se recogen y se organizan en modelos que se utilizarán posteriormente para hacer predicciones.

Básicamente están formados por:

- Nodos: nombres o identificadores de los atributos.
- Ramas: posibles valores del atributo asociado al nodo.
- Hojas: conjuntos ya clasificados de ejemplos y etiquetados con el nombre de una clase.

Existen distintos tipos de arboles de decisión, los más importantes son los árboles de clasificación y de regresión.

Algunas de las principales características de estos modelos son:

- Son de tipo supervisado (se utiliza el conocimiento a-priori del comportamiento de un conjunto de observaciones) y fáciles de entender e interpretar. Son modelos “transparentes”, en los cuales se puede seguir cada clasificación obtenida mediante las reglas lógicas.
- Pueden clasificar utilizando tanto datos categóricos como continuos. No existe una restricción de los tipos de datos necesarios para poder construir el árbol de decisión.
- Son fácilmente convertidos en un conjunto de reglas lógicas, útiles para implementar al momento que se necesitan reutilizar las reglas obtenidas para un determinado árbol de decisión.
- Se pueden generar las reglas lógicas obtenidas, en código SQL de manera de integrarlas en los sistemas de bases de datos de una empresa.
- Permiten identificar aquellos atributos que entregan una mayor cantidad de información al modelo.

A continuación se describen brevemente algunos tipos de árboles de decisión:

#### **C 4.5**

Es un algoritmo desarrollado por Ross Quinlan (1993). Es recursivo, y se basa en la estrategia "divide y vencerás" (aplicando el algoritmo a conjuntos de datos cada vez más pequeños). Pertenece a la familia TDIDT (*Top Down Induction Decision Trees*). Permite trabajar tanto con atributos discretos como continuos. Utiliza la ganancia proporcional (*gain ratio*), de esta manera se consigue evitar que las variables con mayor número de posibles valores salgan beneficiadas en la selección. Además el algoritmo C4.5 incorpora una poda del árbol de clasificación una vez que este ha sido introducido. La poda está basada en la aplicación de un test de hipótesis que trata de responder a la pregunta de si merece la pena expandir o no una determinada rama.

#### **CHAID (*Chi square- Automatic Interaction Detection*)**

Basado en el algoritmo “*Automatic Interaction Detection*” (AID) publicado por J.A. Hartigan en el año 1975, este algoritmo solo es capaz de tratar con variables predictoras discretas. No realiza una fase de post poda para evitar el sobreentrenamiento, sino que es en la misma fase de construcción del árbol cuando se decide parar. Los Árboles basados en la metodología CHAID, generan 2 o más ramas a partir de un nodo.

#### **CART (“*Classification and Regression tree*”)**

Este algoritmo, desarrollado por Breiman el año 1984, en base a CHAID propuesto por J.A. Hartigan el año 1975, utiliza el criterio basado en *GINI Index* (mide el grado de impureza de un nodo) para el caso de la clasificación. Cuando se pretende construir un árbol de regresión, los criterios se basan en la mínima suma de las desviaciones cuadráticas. Para realizar la post-poda se realiza una estimación del error, ya sea mediante un conjunto de datos diferentes del que se ha utilizado para construir el árbol

o bien aplicando una metodología de validación cruzada. La ramificación de los atributos (*splits*) es binaria. El método tiende a crear *splits* desbalanceados, agrupando una clase mayoritaria en un nodo, y el resto en otros nodos. Los árboles basados en la metodología CART, permiten generar únicamente dos ramas a partir de un nodo.

## 2.7. REGRESIÓN LOGÍSTICA

Regresión logística es un modelo de clasificación supervisado que necesita el entrenamiento con una base destinada para ello que contenga los ejemplos de los cuales ya se conoce la respuesta para así clasificar nuevos casos. .

El objetivo principal que busca resolver esta técnica es el de modelar cómo influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos.

A diferencia de otros métodos, como puede ser la regresión lineal, la regresión logística puede hacer predicciones sobre valores que no necesariamente son dicotómicos (0 o 1), es decir, que la salida que presenta es un número que está en el intervalo [0,1] y que puede ser interpretado como una probabilidad de ocurrencia. Es ahí donde se debe estimar si la probabilidad que entrega el modelo se interpreta como un 1 o un 0 (ocurrencia o no ocurrencia) para los casos dicotómicos.

El modelo Logit nace de un modelo de regresión modificado:

$$Y = a + b * Xi$$

Donde Xi es la variable independiente

Luego, si se desea que la variable Y pueda tomar cualquier valor, se puede construir la siguiente función:

$$\text{Ln} \frac{P}{1 - P}$$

Ahora, si se reemplaza Y por esta función, entonces queda:

$$\text{Ln} \frac{P}{1 - P} = a + b * Xi$$

Haciendo un manejo algebraico se obtiene:

$$\text{Probabilidad Ocurrencia} = \frac{1}{1 + e^{(-a-b*Xi)}}$$

Esta ecuación es precisamente la que se conoce como modelo logístico, donde el número de factores que influye puede ser más de uno. Generalizando, se obtiene que la ecuación de la Regresión Logística es:

$$P(x) = \frac{e^{a+\sum_i^n b * X_i}}{1 + e^{a+\sum_i^n b * X_i}}$$

Como se puede apreciar, la regresión logística ocupa esta ecuación internamente y va adaptando los valores de los distintos betas dependiendo de la importancia que tiene cada una de las variables en la clasificación que se desea obtener.

## 2.8. RED NEURONAL

El modelo de Red Neuronal es una modelación matemática que asimila el sistema nervioso (mediante nodos y conexiones), y su función es la de procesar información identificando patrones y observaciones de distinta clase.

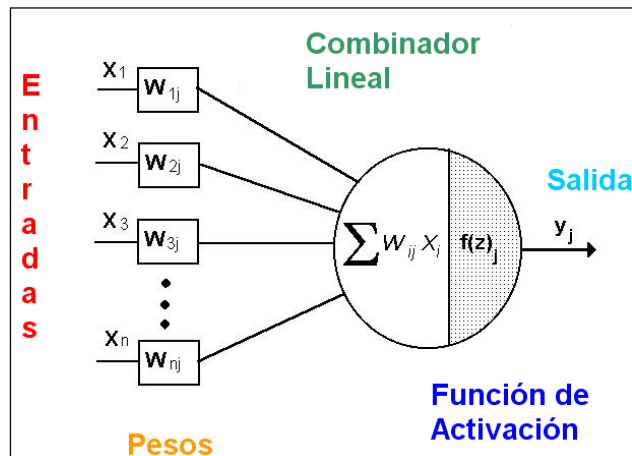
### El Perceptrón

Este modelo es un tipo de red neuronal que se caracteriza por su capacidad de “aprender” a reconocer patrones. El Perceptrón es la versión más simple de las redes neuronales, en la cual se tiene una capa de “receptores” que recibe los patrones y se encarga de clasificar una observación. Además existe una neurona de salida que clasifica las observaciones en dos tipos 0 o 1, dependiendo de los patrones encontrados. Para realizar esta clasificación es necesario utilizar una función de activación que transforme dichos patrones en salidas 0 o 1.

Este algoritmo es análogo al funcionamiento de la unidad básica del sistema nervioso, la neurona y lo componen los siguientes elementos:

- Unidades de entrada: Son las señales o variables que estimulan al Perceptrón, se representan como  $X_i$  con  $i$  en el intervalo  $[1, N]$  donde  $N$  es el número de variables que el modelo considera.
- Conexiones: Son las que transmiten la señal de estímulo, a cada una de estas conexiones se asocia un peso  $W_i$  que entrega la importancia que tiene dicha variable para la predicción.
- Conminador lineal: Toma las entradas con sus respectivos pesos y las transforma en una señal.
- Función de activación: Recibe la señal desde el conminador e inhibe o excita a la neurona para entregar una salida determinada.
- Umbral: Punto de corte donde se determina si la neurona se activa o no.

Figura 13: Perceptrón



FUENTE: WEBER, R.; L'HUILLIER, G. 2008 [12]

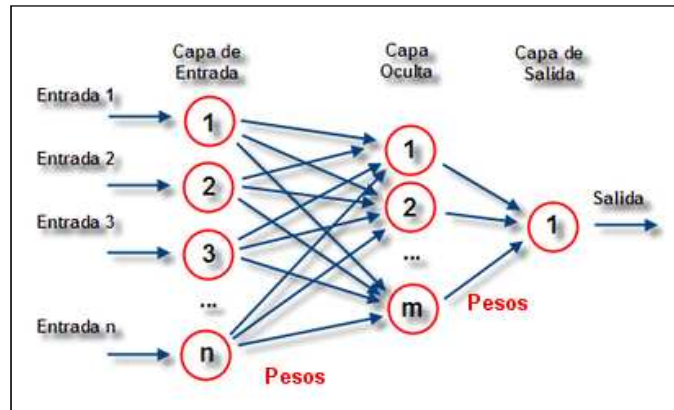
El perceptrón es un modelo de clasificación supervisado, esto quiere decir, que para “aprender” a clasificar observaciones, antes debe entrenarse con un número importante de casos, de los cuales se conoce la clasificación. De este modo el modelo, cada vez que se equivoca en una clasificación, ajusta los pesos asignados a cada variable y luego continúa con el siguiente ejemplo hasta que adquiere un cierto nivel de aprendizaje. El problema del perceptrón es que por ser un modelo comparativamente simple, tiene dificultades para clasificar observaciones que son formadas mediante funciones no lineales, debido a esto es que se crea el Perceptrón Multicapa o más conocido como Red Neuronal Multicapa, que es el modelo que se utilizará en este caso.

### Red Neuronal Multicapa

La red neuronal multicapa es un tipo de modelo de clasificación que se basa en el perceptrón, pero a diferencia de éste permite obtener buenas clasificaciones de las observaciones que provienen de problemas no lineales.

La red neuronal multicapa se diferencia del perceptrón principalmente en su estructura, ya que además de tener una capa de entrada, que recibe las señales y la capa de salida que entrega la clasificación, internamente tiene una o más “Capas Ocultas” que permiten que el modelo pueda adaptar los pesos mejorando la clasificación.

Figura 14: Red Neuronal Multicapa



FUENTE: WEBER, R.; L'HUILLIER, G. 2008 [12]

Los parámetros que se pueden ajustar en una Red Neuronal Multicapa son los siguientes:

- Cantidad de Neuronas en cada una de las capas.
- Función de activación.
- Tasa de aprendizaje: Parámetro que determina cuán rápido aprende la red neuronal en cada iteración (modificación alta o baja en los pesos). Este parámetro puede presentar dos inconvenientes: Una baja tasa de aprendizaje puede hacer que si se comenzó próximo a un mínimo local, la red converja a éste. Por otro lado, si la tasa de aprendizaje es muy alta, el modelo puede estar iterando sin encontrar un punto de convergencia.
- Momentum: Se utiliza para mejorar el aprendizaje de la red, ya que ayuda a controlar la situación en que una alta tasa de aprendizaje modifique de forma drástica los pesos.
- Épocas. Cantidad de veces que el modelo recorre la base completa de ejemplos que se le están mostrando durante el período de aprendizaje. Mayor número de épocas implica un menor el error de clasificación, por otro lado con un bajo número de ejemplos se corre el riesgo de que el modelo se “sobreajuste” a los datos (memorice los ejemplos y no los patrones).

## 2.9. INDICADORES PARA MEDIR LA CALIDAD DE LOS MODELOS

Para medir el desempeño de los modelos en estudio se deben analizar una serie de índices que permiten compararlos entre sí. Los indicadores de precisión y sensibilidad, se definen en base a la matriz de confusión, la cual entrega la información de los errores y aciertos de la clasificación y en donde cada cuadrante de la matriz tiene el siguiente significado:

- A: Predicción acertada, valor predicho negativo y valor real negativo
- B: Predicción errónea, valor predicho positivo dado que el valor real es negativo
- C: Predicción errónea, valor predicho negativo dado que el valor real es positivo
- D: Predicción acertada, valor predicho positivo y valor real positivo.

Figura 15: Matriz de Confusión

Matriz de confusión		Predicción	
		Negativo	Positivo
Real	Negativo	A	B
	Positivo	C	D

Fuente: Elaboración propia

En base a la matriz de confusión se pueden definir los siguientes indicadores:

Tabla 2: Definición de Indicadores de comparación entre modelos

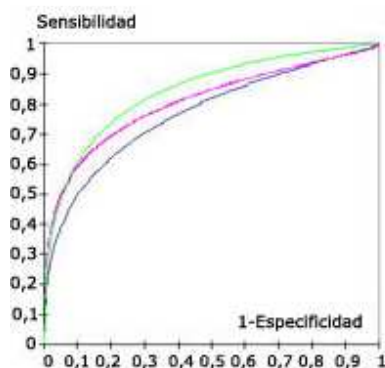
Indicador	Fórmula
Eficiencia	$\frac{(A + D)}{(A + B + C + D)}$
Precisión	$\frac{D}{(B + D)}$
Sensibilidad	$\frac{D}{(C + D)}$
Verdadero Positivo (VP)	$\frac{D}{(C + D)}$
Verdadero Negativo (VN)	$\frac{A}{(A + B)}$
Falso Negativo (FN)	$\frac{C}{(C + D)}$
Falso Positivo (FP)	$\frac{B}{(A + B)}$

### Curvas ROC

Las curvas ROC (*Receiver Operating Characteristic*) son una alternativa a la matriz de confusión, que entrega similar información de manera gráfica. En el eje x se representa

[1-especificidad (FP)] y en el eje y [sensibilidad (VP)] para distintos puntos de corte, generando de esta manera los indicadores.

Figura 16: Ejemplo Curvas ROC



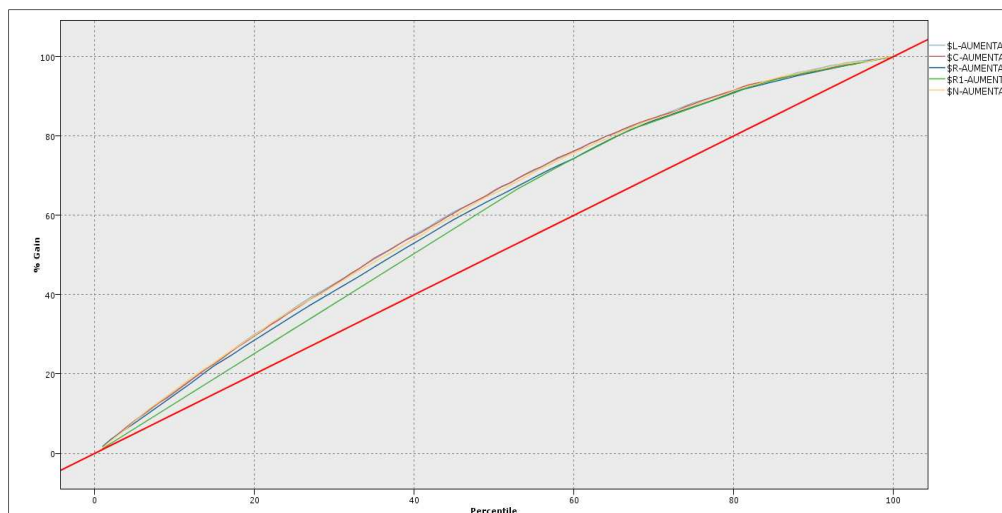
FUENTE: WEBER, R.; L'HUILLIER, G. 2008[12]

El área bajo la curva da cuenta del poder discriminador del modelo, es decir, de la capacidad de clasificar correctamente un elemento. Cuando el área bajo la curva es mayor que 0,5 significa que el uso del modelo es mejor que el azar. Mientras más cercano a 1, mejor será la evaluación del modelo.

### Curvas Gain (Curvas de ganancia acumulada)

Estas curvas dan cuenta de la reducción del aumento de la entropía que resultaría de dividir los datos según un atributo A. La ganancia de información da preferencia a árboles con profundidad mínima, ya que "ordena" los atributos de mayor a menor importancia, con lo que se consigue una reducción de la entropía con un tamaño mínimo del árbol.

Figura 17: Ejemplo curva de ganancia



FUENTE: WEBER, R.; L'HUILLIER, G. 2008[12]



### 3. METODOLOGÍA

En este punto se muestra cómo el marco conceptual, así como la metodología expuesta anteriormente serán aplicados en este trabajo.

#### 3.1. DETALLE METODOLOGÍA

A continuación se detallan las etapas del proceso KDD y la manera en que serán utilizadas en el presente trabajo:

1. Selección de datos: En primer lugar se definirá a los clientes que entrarán dentro de este análisis, para lo cual se debe precisar cuáles serán los clientes con tarjetas de crédito vigentes, definir el período de tiempo a analizar y también el mínimo de historia transaccional que se exigirá a cada cliente para asegurar una correcta definición de las variables RFMN. Luego se obtiene la información transaccional asociada a cada uno de los clientes en el período definido.
2. Pre-procesamiento de datos: Se analizará el caso de los valores perdidos y *outliers*, dándole un tratamiento a los mismos. En esta etapa también se hará un análisis de las variables seleccionadas: Análisis univariado, para describir las variables transaccionales con las que se está trabajando.
3. Transformación: En esta etapa se crearán nuevas variables a partir de las existentes y se harán las transformaciones necesarias a los atributos con que se cuenta para obtener las variables RFMN deseadas para cada cliente. También se extraerán variables que sean relevantes de modo de poder complementar las variables RFMN con éstas.
4. *Data Mining*: Se realizarán los distintos experimentos, es decir: construir distintas cadenas de Markov, variando:
  - El número de variables RFMN a ocupar probando distintas combinaciones. Se analizarán los resultados de utilizar cada variable por sí sola (R, F, M o N), en pares (RF, RM, FM, RN, etc.) o bien todas juntas (RFMN).
  - La forma de estimar las probabilidades de transición entre los estados.

En esta etapa los resultados deben ser validados mediante los errores asociados a cada modelo. Finalmente se realizará un análisis de 3 períodos con las cadenas de Markov, de modo de describir el comportamiento transaccional de los clientes a través del tiempo.

5. Análisis de resultados de la segmentación y cadenas de Markov: Se caracterizará a cada uno de los segmentos obtenidos con variables de consumo para definir perfiles que diferencien a los distintos *clusters*.
6. Nuevo proceso KDD: Una vez concluido los resultados de las cadenas de Markov, se diseñarán modelos de propensión, desarrollando un nuevo proceso KDD dentro de esta etapa:
  - Luego de caracterizar los grupos, se definirán los unos y ceros para los modelos correspondientes a cada segmento, en donde los unos serán los clientes que aumentan su consumo al período siguiente y los ceros serán los clientes que lo mantienen o disminuyen.
  - Se aplicarán los distintos modelos: Regresión Logística, Red Neuronal y Árboles de decisión para concluir cuál modelo obtiene mejores resultados en función de los distintos tipos de errores, la aplicación de criterios de negocio y juicio de expertos.
7. Análisis de resultados de los modelos de propensión: Se analizan los resultados de los modelos de propensión de cada segmento mediante matrices de confusión y curvas de ganancia, así como el análisis de las variables relevantes según criterios de negocio con lo cual se generarán recomendaciones de estrategias a seguir.

#### **4. VARIABLES DE MODELAMIENTO**

En este capítulo se presentan las variables que se utilizarán en este trabajo. Además se mostrará un análisis descriptivo de las mismas y se detallará la definición de los períodos a utilizar, la división de las bases de *train* y *test*.

##### **4.1. DESCRIPCIÓN DE DATOS**

El banco en el cual se desarrolla el presente trabajo, cuenta al julio del 2009 con más de 350 mil tarjetas de crédito vigentes. Para cada una de estas, se conserva el registro detallado de todas las transacciones asociadas a dichas tarjetas desde el año 2005, sin embargo se considerarán válidos sólo los datos registrados desde el año 2007, asegurando así un mejor nivel de calidad de éstos.

Además de los datos transaccionales de la tarjeta de crédito, se cuenta con información de tipo socio-demográfica, así como datos de deuda neta en la institución bancaria, en el sistema financiero, tenencia de productos, datos transaccionales de otros medios de pago entre otros. Estos serán utilizados en los modelos de propensión (parte final de este trabajo) que se realizarán.

A continuación se muestra una tabla con las principales variables transaccionales a utilizar:

**Tabla 3: Detalle de las variables transaccionales a utilizar**

Variable	Tipo	Descripción
Rut	Entero	Identificador del cliente
Cst_R	Entero	Días transcurridos desde la última compra
Cst_F	Real	Tiempo promedio (días) entre compras
Cst_M	Real	Monto (UF) promedio de cada transacción
Cst_R_F	Real	R/F
N_txn	Entero	Número de transacciones de cada cliente
Antigüedad	Entero	Antigüedad en meses de cada cliente
mnt_txn	Real	Monto (UF) total transacciones
Min_fecha	Fecha	Fecha primera compra
Máx_fecha	Fecha	Fecha última compra
Dif_dias	Entero	Días de diferencia entre la última compra y la primera

La variable Cst\_R\_F (R/F) mide el *Recency* de un cliente con respecto a su *Frequency*, lo que permite observar si un cliente se está atrasando con respecto a su frecuencia ( $R/F > 1$ ), se mantiene ( $R/F = 1$ ) en ella o bien se está adelantando a ella ( $R/F < 1$ ).

#### 4.1.1. ANÁLISIS UNIVARIADO

En el presente trabajo, se utilizarán las variables transaccionales RFM para segmentar a los clientes y con esto crear las cadenas de Markov, por lo que solo estas variables son relevantes y merecen un análisis.

Del total de 400 mil clientes a octubre del 2009, el análisis univariado se aplicará al grupo de clientes con más de un año de antigüedad y que haya presentado actividad entre enero del año 2007 y junio del año 2009, con lo cual se obtiene un universo de clientes de más de 180 mil.



Fuente: Elaboración propia

#### Variables RFM:

En la Tabla 3 se observan los estadísticos descriptivos de las tres variables relevantes de este trabajo, correspondientes a todo el período de estudio (01-01-2007 hasta 30-06-2009)

**Tabla 4: Estadísticos descriptivos variables RFM**

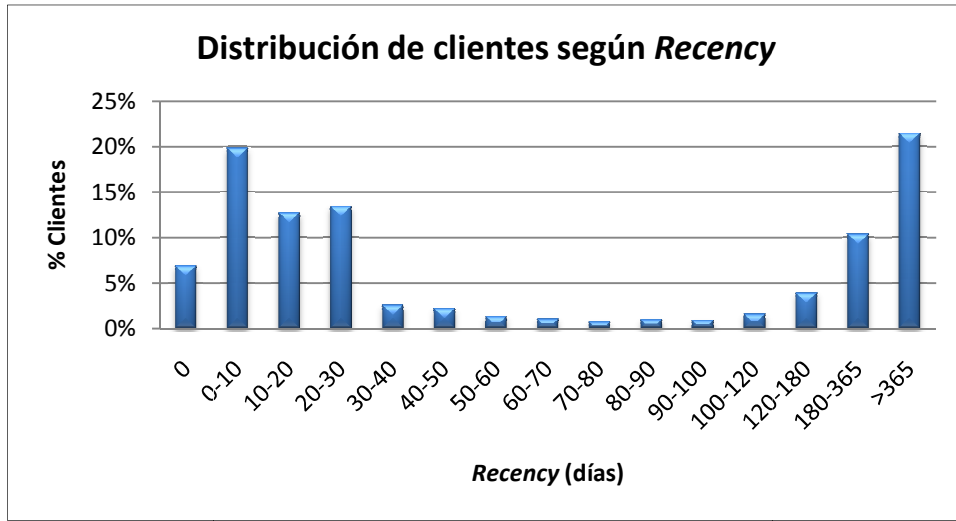
	<b>R</b>	<b>F</b>	<b>M</b>
<b>Mínimo</b>	0	0	0
<b>Máximo</b>	911	450,0	118,569
<b>Promedio</b>	177,786	19,568	2,246
<b>Moda</b>	0	0	0,150
<b>Mediana</b>	26	14,931	1,703

Se observa que el promedio de *Recency* es alto con respecto al período completo analizado, esto indicaría que en promedio los clientes no han realizado transacciones hace casi 6 meses. Este valor se debe a que en este análisis aún no se han incorporado filtros a los clientes ni se ha realizado un tratamiento de los datos ya que corresponde a un análisis exploratorio inicial. Los filtros y procesamientos se realizan en etapas posteriores de este trabajo.

A continuación se presenta los histogramas de cada una de estas variables.

Recency

Figura 19: Histograma variable *Recency*



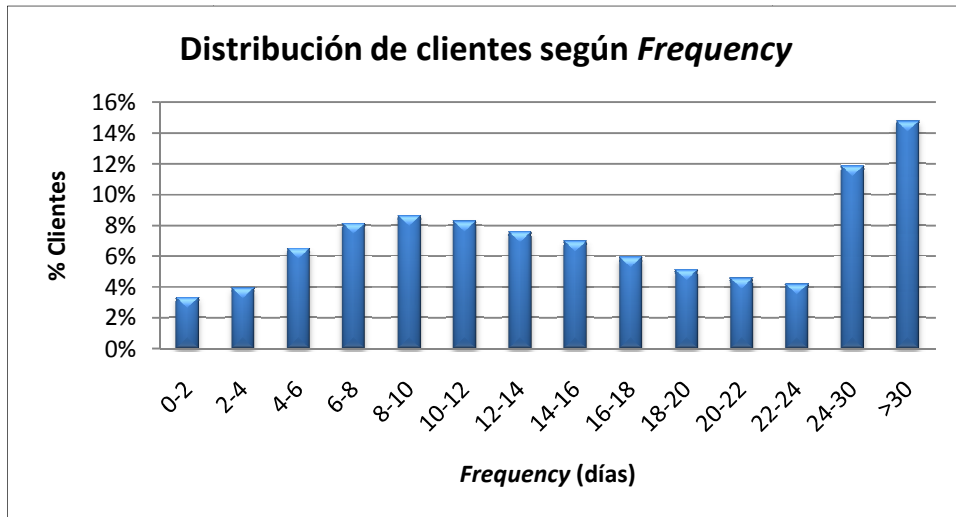
Fuente: Elaboración propia

Existe una marcada diferencia entre dos tipos de clientes: unos que tienen un alto *Recency* y otros que no presentan transacciones hace más de 100 días. El primer grupo representa cerca del 50% mientras que el segundo presenta cerca de un 40% de los clientes.

Frequency

La figura 20 muestra la distribución de los clientes según *Frequency* en donde la frecuencia promedio de las transacciones se encuentra entre 8 y 14 días, tramos que presentan a cerca del 25% del total de los clientes analizados y luego comienza a decaer lentamente hacia otros intervalos de frecuencia.

Figura 20: Histograma variable *Frequency*

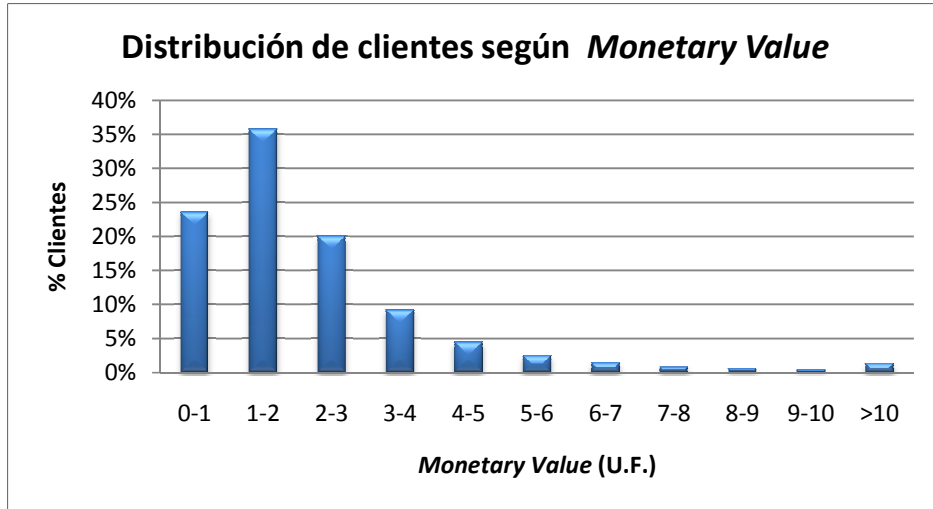


Fuente: Elaboración propia

### Monetary Value

La figura 21 muestra la distribución de los clientes según *Monetary Value*: el valor promedio de las distintas transacciones se encuentra entre 0 y 7 U.F. por transacción y mayores montos corresponden a una cantidad mínima de casos. El porcentaje acumulado entre 0 y 7 UF, contiene cerca del 96,9% de los casos.

Figura 21: Histograma variable *Monetary Value*



Fuente: Elaboración propia

### 4.1.2. ANÁLISIS MULTIVARIADO

Al igual que en el análisis anterior, se considera al mismo grupo de personas dado que también corresponde a un análisis exploratorio inicial. Las correlaciones de todas las variables se muestran a continuación en la Tabla 4. A pesar de que se utilizará solo las variables RFM, se muestra las correlaciones con otras, para poder inferir alguna relación de dependencia importante.

Tabla 5: Correlaciones entre variables

	R	F	M	cst_R_F	N_txn	Antigüedad	mnt_txn	dif_dias
R	1,00	-0,01	0,02	0,34	-0,35	0,18	-0,25	<b>-0,75</b>
F	-0,01	1,00	-0,07	-0,14	-0,23	0,03	-0,25	0,03
M	0,02	-0,07	1,00	0,01	0,01	0,03	0,35	0,01
cst_R_F	0,34	-0,14	0,01	1,00	-0,10	0,04	-0,08	-0,27
N_txn	-0,35	-0,23	0,01	-0,10	1,00	0,22	<b>0,74</b>	0,45
Antigüedad	0,18	0,03	0,03	0,04	0,22	1,00	0,18	0,52
mnt_txn	-0,25	-0,25	0,35	-0,08	<b>0,74</b>	0,18	1,00	0,35
dif_dias	<b>-0,75</b>	0,03	0,01	-0,27	0,45	0,52	0,35	1,00

A continuación se analizan las cuatro relaciones entre variables que presentan una mayor correlación:

- R con dif\_dias: La correlación es de -0,75, esto se debe a que el período analizado es constante y de 30 meses, por lo que a una mayor diferencia de días entre la primera y última transacción, menor es el *Recency* del cliente.
- N\_txn con mnt\_txn: La correlación es de 0,74. Esta correlación también es esperable, dado que es razonable que a un mayor número de transacciones, la suma total del monto de estas sea mayor.
- Antigüedad con dif\_dias: La correlación es de 0,52. Esta correlación es esperable, dado que la diferencia de días se calcula entre la primera y la última transacción, por lo que es más probable que una persona más antigua tenga una mayor diferencia de días.
- N\_txn con Antigüedad: La correlación es de 0,45. Esta correlación es alta dado que se espera que, para una ventana de tiempo fija, persona con más antigüedad haya realizado más transacciones que una nueva.

Las correlaciones entre las variables RFM son bajas<sup>5</sup>, por lo que la información transaccional de los clientes se puede resumir perfectamente en estas tres: la baja correlación asegura que la información que está capturando cada una de estas variables es independiente, esto dado que M captura el monto de las transacciones, mientras que R y F la información de la actividad de cada cliente.

## 4.2. SELECCIÓN DE DATOS

Para seleccionar los clientes que se incluyen en este análisis, se utilizan los siguientes criterios:

Período de tiempo:

- El período a analizar se estableció entre las fechas 01-01-2007 y 30-06-2009, lo cual constituye 30 meses de registros de datos. Esto dado que se tiene pocos registros de las transacciones de tarjetas de crédito hasta el año 2006.

Tipo de Transacciones:

- Se escogió 3 tipos de transacciones: Avances, Compras, Recaudación SII, ya que son voluntarias por parte de los clientes.

Clientes:

- Se busca a los clientes con una tarjeta de crédito vigente a la fecha y se obtienen sus transacciones históricas (con esa u otra tarjeta anterior que puede o no estar vigente a la fecha de este trabajo).

---

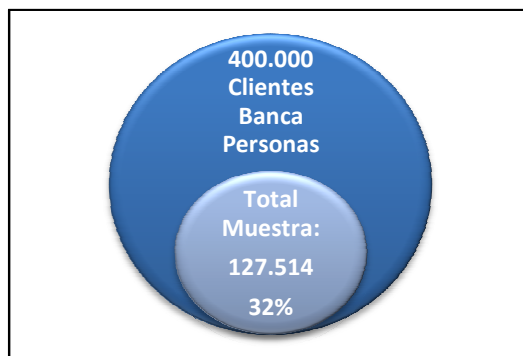
<sup>5</sup> Ver correlación entre las 3 variables en la Tabla 4

- Deben pertenecer al segmento “Personas”.
- Tener más de un año de antigüedad con tarjeta de crédito.
- Cupo mayor a cero.
- Aprobar filtro de morosidad.

Una vez aplicados los criterios anteriormente nombrados, el número de clientes que son seleccionados dentro del análisis corresponde a 127.514, considerando tanto a personas que tuvieron transacciones en el período de tiempo así como aquellas personas que no, pero que cumplen con los criterios expuestos.

De todas las variables, para hacer la segmentación se escogen cuatro de éstas: R, F, M (cst\_R, cst\_F, cst\_M) y además se agrega la variable “número de transacciones” (N).

ión de los filtros



Fuente: Elaboración propia

#### 4.2.1. SELECCIÓN DEL LARGO DE LOS PERÍODOS

Dado que se desea utilizar cadenas de Markov para observar la evolución de los clientes a lo largo del tiempo, se deben definir ventanas de tiempo para obtener transiciones entre períodos y con esto realizar dicho análisis.

Se cuenta con información transaccional de 30 meses, por lo que el largo de los períodos puede tener variadas alternativas:

- 2 Meses.
- 3 Meses.
- 4 Meses.
- 5 Meses.
- 6 Meses.
- 10 Meses.
- 15 Meses.



Para la elección del largo de los períodos, es necesario tener presente los siguientes aspectos:

- El número de transiciones entre períodos debe ser suficiente para poder generalizar comportamientos, por lo que no es posible, por ejemplo, elegir ventanas de tiempo de 10 o 15 meses, ya que solo se contaría con 1 o 2 transiciones y por lo tanto no se podría generalizar el comportamiento de dichas transiciones para todo el intervalo. En función de lo anterior, el número de transiciones óptimo es de al menos 5.
- Es deseable que los períodos de tiempo sean divisores de 12, permitiendo observar el comportamiento del período en el año anterior y comparar comportamientos estacionales. Con lo cual, sería recomendable que los períodos fuesen de largo 2, 3, 4 o 6 meses.
- Los períodos deben tener un tiempo mínimo razonable para que una persona pueda haber realizado una transacción. Por ejemplo la elección de 2 o 3 meses no sería apropiada ya que, en promedio una persona demora 3 meses en comenzar a utilizar la tarjeta de crédito una vez que la ha adquirido.

Considerando los criterios anteriores, la elección del período de tiempo a analizar en cada ventana será de 4 meses por las siguientes razones:

- Al tener períodos de 4 meses, se obtienen 7 períodos distintos y consecuentemente 6 transiciones, lo que permite generalizar comportamientos (por ejemplo obtener promedios)
- Períodos de 4 meses otorga una visión estacional del comportamiento de los clientes, ya que es posible observar el mismo período en el año anterior y con esto se puede hacer comparaciones entre los mismos períodos.
- El largo escogido constituye un tiempo razonable para haber realizado transacciones: del total de clientes, solo el 6,57% tiene una frecuencia de compra mayor a 120 días, por lo tanto la información que se pierde con este corte es menor.
- Con ventanas de tiempo de 4 meses se puede realizar 4 análisis de 3 períodos, por lo cual ya se cuenta con una percepción más general del comportamiento de los clientes.

#### **4.2.2. DEFINICIÓN DE LAS VENTANAS DE TIEMPO**

En la sección anterior se definió el largo adecuado de los períodos (4 meses), dado esto, ahora se puede definir las ventanas de tiempo a analizar. Para esto, se eliminaron los datos de los 2 primeros meses de información tanto por la antigüedad de estos como por la definición del largo de los períodos (7 períodos de 4 meses). Las ventanas de tiempo son las siguientes:

**Tabla 6: Definición de los períodos**

<b>Período</b>	<b>Inicio</b>	<b>Término</b>
<b>1</b>	01-Marzo-2007	30-Junio-2007
<b>2</b>	01-Julio-2007	31-Octubre-2007
<b>3</b>	01-Noviembre-2007	28-Febrero-2008
<b>4</b>	01-Marzo-2008	30-Junio-2008
<b>5</b>	01-Julio-2008	31-Octubre-2008
<b>6</b>	01-Noviembre-2008	28-Febrero-2009
<b>7</b>	01-Marzo-2009	30-Junio-2009

### **4.2.3. DEFINICIÓN DE CLIENTES INACTIVOS**

Dado que se definió a los clientes que entrarán en el análisis y que corresponden a un total de 127.514 personas que cumplen con los criterios de selección. Dentro de este grupo, se definen los clientes inactivos como aquello que en alguna de las ventanas de tiempo anteriormente señaladas no cuenten con ninguna transacción en los cuatro meses que se estén analizando en el momento. Dada esta definición una persona puede ser considerada como inactiva en una ventana de tiempo y en otra no. También existirán clientes que permanecerán en estado inactivo durante los 28 meses analizados y otros que nunca estarán en esta condición.

### **4.3. TRANSFORMACIÓN**

Una vez definidas las ventanas de tiempo a analizar, se toma la información de cada uno de esos meses y se realiza el cálculo de las variables R, F, M y N para cada uno de estos períodos.

Antes de comenzar a visualizar los datos con los que se trabaja, se debe señalar que la variable F es la única de las cuatro que considera la última transacción antes del período en el cual se está calculando, por lo cual el valor de esta variable podría ser mayor a 122 días, que es la ventana de tiempo con la que se está trabajando.

Debido a que la variable F puede tener valores mayores a 122 días, existe el inconveniente de que los primeros periodos tendrán a la variable F con este valor como máximo, dado que no se cuenta con información anterior, mientras que a medida que se avanza en los períodos, el valor de esta variable puede ser mayor (por ejemplo una persona del período 6 podría tener un valor de 625 días en la variable F). El problema que surge es que cuando se desee hacer una segmentación, los distintos períodos no podrían ser comparables debido a esta variable. Para resolver este inconveniente, se opta en dejar a la variable F con un valor máximo igual a 122 días independiente del período, con esto una persona que tiene un valor de F mayor a 122 simplemente la variable se truncará.

A continuación se presenta un resumen los estadísticos descriptivos de las variables transaccionales usadas, para el último período de tiempo (período 7):

Tabla 7: Estadísticos Descriptivos variables RFM (período 7)

Estadísticos Descriptivos	Mínimo	Máximo	Media	Desv. St.
R	0	121	14,5	16,6
F	0,5	122	40,9	39,3
M	0,0	152,1	2,7	4,6
N	1	235	9,1	12,7

El valor máximo de la variable F es de 122 ya que todos los valores mayores fueron truncados por los motivos antes explicados.

La creación de las variables R, F, M y N se realizó para cada uno de los 7 períodos de tiempo.

Se observa que los valores de M y N son altos en comparación a lo que se espera para un período de tiempo de 4 meses, ya que aún no se han sacado del estudio los *outliers* de estas variables. Esta labor se realizará en la etapa siguiente (preprocesamiento), en donde estos casos serán tratados.

#### 4.4. PREPROCESAMIENTO

##### 4.4.1. TRATAMIENTO DE VALORES FALTANTES Y OUTLIERS

Por la construcción de las variables R, F, M y N, las cuales provienen de los datos transaccionales de los clientes, no se tiene valores faltantes dentro de la base de datos.

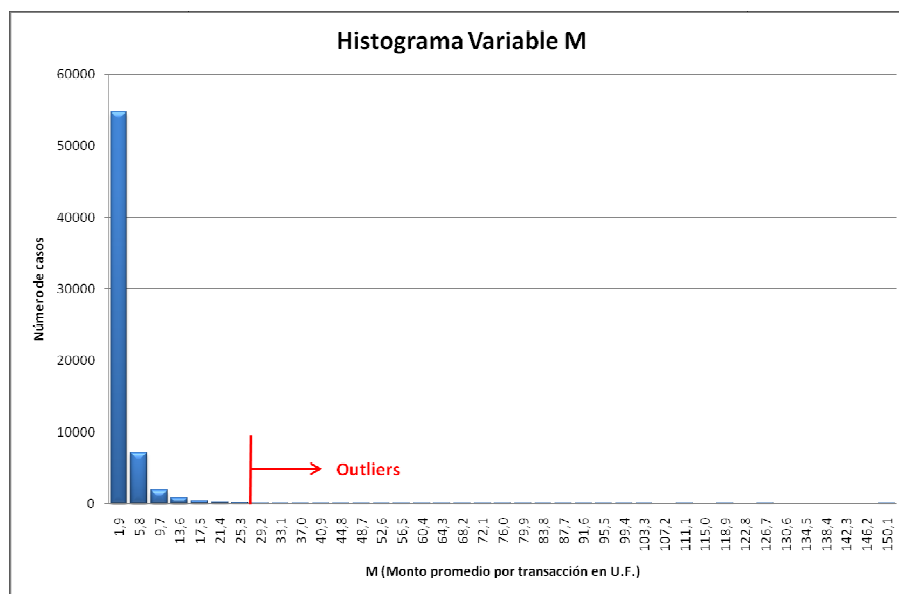
Dado que se definió que el largo de los períodos a analizar es de 4 meses, para hacer el análisis de *outliers*, se debe considerar alguno de esos períodos como referencia. Se escoge utilizar tanto para hacer este análisis como para la segmentación el último período de datos (7): 01-Marzo-2009 al 30-Junio-2009.

Luego de analizar las variables relevantes para la segmentación (R, F, M y N), se llega a la conclusión de que para las variables R y F no hay problemas de valores fuera de rango. Sin embargo es distinto para la variables M y N las cuales presentan casos con valores altos en comparación al promedio de las observaciones. Tras estos resultados, se deja fuera del análisis a aquellas observaciones cuyos valores de M o N sean mayores a la media de la variable más 5 veces la desviación estándar de la misma, eliminando del análisis a los llamados valores extremos de cada variable.

Para la variable M el punto de corte que define los *outliers* es de 25,7 U.F., con lo cual quedan fuera del análisis 438 casos, que representan al 0.67% del total de 65.857 clientes activos en el período.

A continuación se observa un gráfico con la distribución de la variable M y el punto de corte para considerar los *outliers* del período:

**Figura 23: Histograma Variable *Monetary Value***

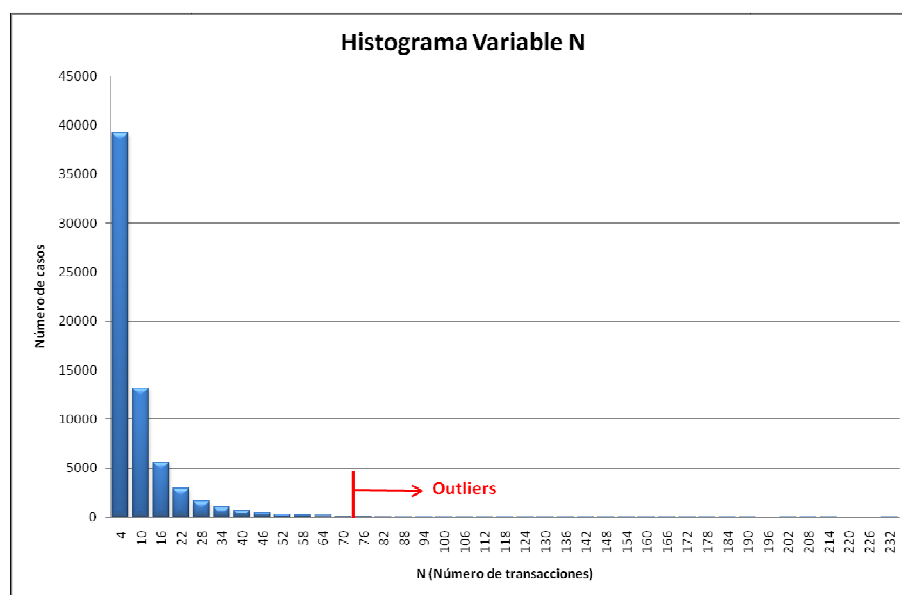


Fuente: Elaboración propia

Así mismo, para la variable N el punto de corte para considerar un caso como *outlier* corresponde a 72 transacciones en la ventana de tiempo de 4 meses. En el período 7 el número de *outliers* corresponde a 462 casos, que representan al 0.7% del total de 65.857 clientes activos en el período.

Se muestra un gráfico con la distribución de la variable N y el punto de corte para considerar los *outliers* del período:

**Figura 24: Histograma Variable N (número de transacciones)**



Fuente: Elaboración propia

Con estos dos criterios finalmente se eliminan 883 casos *outliers* (hay intersección entre los dos criterios) que representan al 1,3% del total de 65.857 clientes activos en el período. Por lo tanto para el período 7 el número total de clientes activos con que se cuenta es de de 64.974.

Este mismo tratamiento se hizo para cada uno de los 7 períodos señalados de modo que las ventanas de tiempo queden con observaciones comparables entre sí.

Dado que ya se ha realizado el tratamiento de *outliers* en los períodos, se analizan nuevamente los estadísticos descriptivos de los datos con los cuales finalmente se trabajará. Se muestran a continuación los estadísticos descriptivos de los clientes activos del período 7:

**Tabla 8: Estadísticos descriptivos periodo 7(outliers tratados)**

Estadísticos Descriptivos	Mínimo	Máximo	Media	Desv. St.
<b>R</b>	0	121	27,8	30,2
<b>F</b>	0,5	122	40,9	39,1
<b>M</b>	0,0	25,7	2,5	3,1
<b>N</b>	1	72	8,5	10

Se distingue que los estadísticos descriptivos han cambiado, en especial en la variable R (*Recency*), que anteriormente era cercana a los 14 días y tras el tratamiento de

*outliers* ahora es aproximadamente de 28 días, es decir, en promedio los clientes activos compran una vez al mes.<sup>6</sup>

## 5. MINERÍA DE DATOS

En esta sección se mostrará cómo se hace la segmentación de los clientes según las variables R, F, M y N. Una vez obtenidos estos *clusters*, estos definirán los estados de las Cadenas de Markov que modelarán el comportamiento transaccional de los clientes a lo largo del tiempo.

### 5.1. SEGMENTACIÓN

En este trabajo, para observar la evolución de los clientes a lo largo del tiempo se crearán cadenas de Markov que permitan analizar dicho comportamiento. Para esto es necesario crear *clusters* con las variables R, F, M y N, los cuales definirán los Estados de estas cadenas.

Para segmentar a los clientes se utiliza K medias en dos formas distintas, en primer lugar se hace un análisis no jerárquico en el cual, para cada una de las variables (R, F, M y N) y la combinación de ellas (RF, RM, RN, FM, FN, MN, RFM, RMN, RFN, FNM, RFNM) se busca el número óptimo de *clusters* a ocupar en cada caso. En segundo lugar, se utiliza para obtener la pertenencia de cada cliente a su grupo en base al número de segmentos determinado anteriormente.

#### 5.1.1. SELECCIÓN DEL NÚMERO DE CLUSTERS

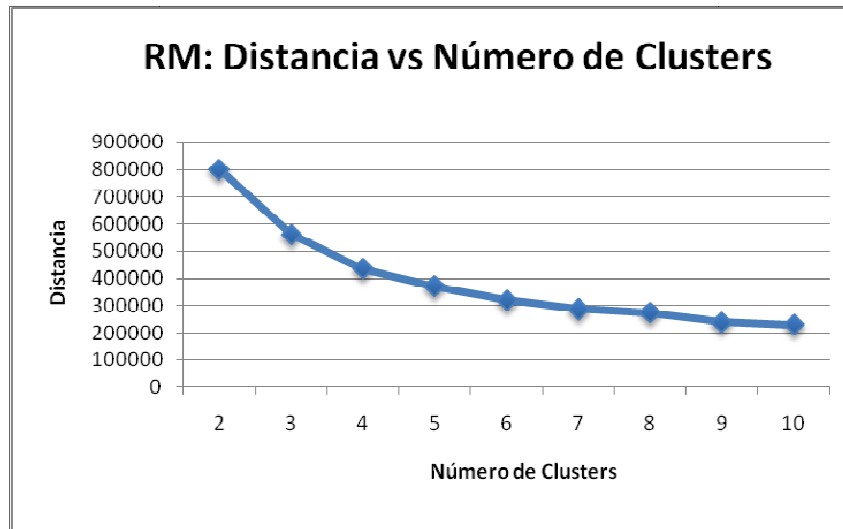
Para determinar el número de segmentos apropiado para cada una de las variables (R, F, M y N) o bien para las combinaciones de ellas (RF, RM, RN, FM, FN, MN, RFM, RMN, RFN, FNM, RFNM) se utiliza K medias de forma no jerárquica. En este proceso se toma una o varias variables y se comienza a iterar con distintos valores de K, partiendo desde 2 *clusters* hasta 10, guardando en cada iteración la distancia de los clientes a su respectivo centro, luego se suman las distancias totales para cada número de segmentos y se grafica el número de grupos versus la distancia total obtenida para tratar de identificar algún punto en la curva donde exista un marcado cambio en la trayectoria, lo cual indicaría que con ese número de *clusters* existe una disminución considerable de las distancias totales.

Para las distintas segmentaciones no es posible observar un claro quiebre en la curva, un ejemplo de esto se muestra a continuación en los gráficos de las duplas RM y NM:

---

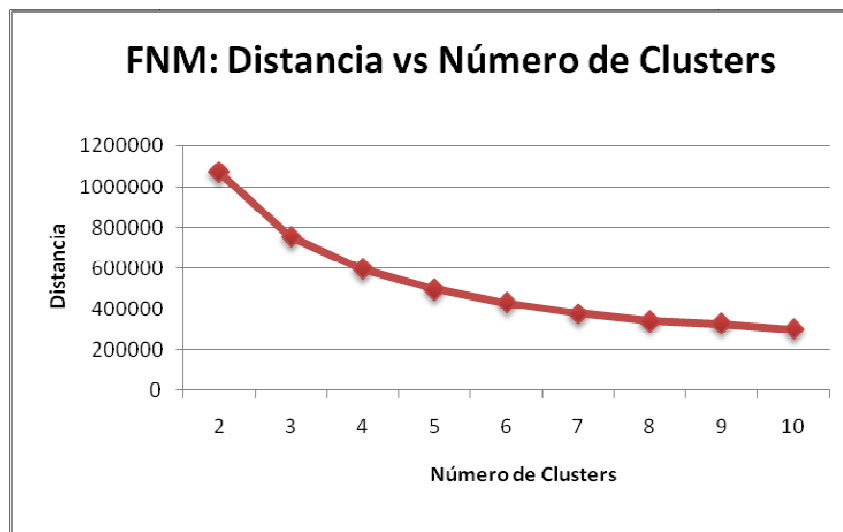
<sup>6</sup> Los histogramas de las variables M y N después del preprocesamiento, así como de R y N se encuentran en el ANEXO B

Figura 25: Gráfico número de *clusters* versus Distancia (variable RM)



Fuente: Elaboración propia

Figura 26: Gráfico número de *clusters* versus Distancia (variables NM)



Fuente: Elaboración propia

En estos dos casos, no se puede afirmar gráficamente cuál es el número óptimo de *clusters*, por lo que se define el siguiente criterio: Al igual que en el método anterior, se suma la distancia total para distinto número de *clusters* y se realiza lo siguiente: se compara el número actual de grupos versus el agregar uno adicional, cuando el aporte marginal (en reducción de distancia) de agregar un segmento es menor a un 10%, entonces se detiene el proceso.

Bajo este criterio, se obtiene que el número de segmentos óptimo por cada combinación de variables es el siguiente:

Tabla 9: Número de *clusters* óptimo según variables

Variable(s)	Nro. de <i>Clusters</i>
R	4
F	4
M	4
N	4
RF	3
RM	4
FM	4
FN	4
RN	5
NM	3
RFM	3
RNM	5
RFN	3
FNM	4
RFNM	3

En la mayoría de los casos el número óptimo de segmentos es 3 o 4, a excepción de las segmentaciones con RN y RNM donde el número óptimo de grupos es 5.

### 5.1.2. SEGMENTACIÓN DE CLIENTES POR PERÍODO.

En función del número óptimo de *clusters* obtenidos en la sección anterior para cada combinación de variables, se determina la segmentación de los clientes tomando como entrada dicha información.

Para realizar la segmentación, se tomará nuevamente como referencia el último período de tiempo (período 7) y a partir del número óptimo de segmentos se obtiene los centros de los grupos para cada combinación de variables.

Se presenta a continuación un ejemplo de los resultados de los centro de *clusters*.

Tabla 10: Ejemplo de los centros de *clusters* según variables FNM

CLUSTER	F	N	M
1	116,6	2	4,6
2	7,5	20	1,6
3	61,1	3	2,9
4	24,5	5	2,1



## 5.2. CADENAS DE MARKOV Y ELECCIÓN DE LA SEGMENTACIÓN

En la sección anterior se determinó la pertenencia de los clientes a los distintos segmentos mediante K medias, al cual se ingresó el valor de los centros de los *clusters* a ocupar en cada caso (dependiendo de la combinación de variables a utilizar).

La clasificación se realiza en cada uno de los 7 períodos y con esto es posible crear diferentes cadenas de Markov con las cuales se observa la transición de los clientes entre segmentos de un período a otro mediante la matriz de transición<sup>7</sup>.

Para evaluar las segmentaciones, se utiliza los siguientes errores:

Error Absoluto (MAD): El error absoluto analiza el porcentaje de equivocaciones del modelo con respecto al total de casos.

$$\text{Error Absoluto (MAD)} = \frac{\sum_i \sum_j \text{ABS}((N'_{ij}(t,t+\Delta t)) - (N_{ij}(t,t+\Delta t)))}{N(t)}$$

Donde:

$(N'_{ij}(t, t + \Delta t))$  : Número de clientes que el modelo dice que deberían pasar del estado i al estado j, desde período t al al (t+Δt)

$(N_{ij}(t, t + \Delta t))$  : Número de clientes que en realidad pasaron del estado i al j, desde período t al al (t+Δt)

$N(t)$  : Número total de clientes que había en el período t

$\text{ABS}(x - y)$  : Módulo de la resta entre x e y.

Error Ponderado por el número de transiciones: Este error dada su definición es igual al promedio de los errores porcentuales. Pondera los errores porcentuales por el peso que tiene el número de transiciones sobre el total. Se encarga del problema que hay al comparar experimentos con distinto número de *clusters* o estados.

Error ponderado por Transiciones (Promedio errores porcentuales) =

$$\sum_i \sum_j \frac{\text{ABS}((N'_{ij}(t,t+\Delta t)) - (N_{ij}(t,t+\Delta t)))}{(N_{ij}(t,t+\Delta t))} \times \frac{1}{\text{Nro.de Transiciones}}$$

Este error considera las equivocaciones en cada una de las casillas de la matriz de transición y no el error absoluto, ya que si se tomara este último siempre las clasificaciones con mayor número de estados tendrán un error menor.

<sup>7</sup> Explicación en el punto 2.4 de este trabajo.

Este error puede tomar valores mayores a 100%.

Error Ponderado por Distancias: Este error al igual que en el caso anterior, sirve para comparar clasificaciones con distinto número de *clusters*. Se pondera el error absoluto por la distancia entre los centros de los segmentos analizados en ese momento. La distancia utilizada es la euclidiana.

$$\text{Error ponderado por distancias} = \frac{\sum_i \sum_j \text{ABS}(N'_{ij}(t, t+\Delta t) - N_{ij}(t, t+\Delta t))}{N(t)} \times \text{Distancia}(i, j)$$

Este error tampoco tiene un tope máximo de 100%, pudiendo alcanzar valores mayores.

### 5.3. COMPROBACIÓN DE COMPORTAMIENTO MARKOVIANO

Se intenta comprobar la propiedad de pérdida de memoria de las cadenas de Markov, la cual propone que la última información es la necesaria para precedir el comportamiento del sistema en el futuro y no la información pasada.

Para demostrar este comportamiento se procede de la siguiente manera: Se calcula la matriz de transición para cada una de las distintas segmentaciones antes señaladas y se verá cuál es la mejor forma de estimar las transiciones entre distintas alternativas. Si el sistema cumple con la propiedad de pérdida de memoria, entonces las probabilidades de la última transición deberán ser las que mejor predigan las probabilidades de transición del período siguiente.

Para verificar esta propiedad se compararán los errores cometidos al proyectar la transición siguiente utilizando las probabilidades de la última transición con respecto a los errores que se obtienen de las siguientes alternativas de predicción:

- Promedio de todas las transiciones de las personas y calcular las probabilidades.
- Tendencia de las transiciones de las personas en los dos últimos periodos y calcular las probabilidades.
- Misma transición pero en el año anterior y calcular las probabilidades.

Luego de verificar esta propiedad, se buscará dentro de las segmentaciones la que mejor se ajusta al comportamiento deseado: tener un bajo error en la predicción de las probabilidades y que permita distinguir entre distintos tipos de clientes con respecto a su transaccionalidad.

En el ANEXO D se muestra los resultados de los errores de predecir la última transición (transición desde el período 6 al 7) con cada una de las alternativas de predicción anteriormente señaladas, para cada segmentación.

Estos resultados muestran que en gran parte de los casos para los 3 tipos de error la penúltima transición (del período 5 al 6), es la que mejor predice la transición deseada:

del período 6 al 7. Cabe destacar que el tipo de error MAD (error global del modelo con respecto a casos posibles) fue el menor en todas las segmentaciones.

El error ponderado por número de transiciones fue menor en 10 de las 15 segmentaciones (67%) al predecir ocupando la penúltima transición como la última información conocida.

En función de lo anterior se concluye que la última transición de la que se tiene información es la que mejor predice la siguiente, con lo que el sistema cumple con la propiedad de pérdida de memoria de las cadenas de Markov. Ahora se debe decidir cuál de las segmentaciones que se han realizado es la que más se adecua a los objetivos del trabajo.

Para elegir la segmentación que se utilizará para caracterizar a los clientes con respecto a su transaccionalidad a lo largo del tiempo, se debe cumplir con dos criterios principales:

- El error de las probabilidades de transición del sistema debe ser bajo con respecto a los cuales está siendo comparado.
- La segmentación debe permitir separar a los clientes con respecto a su nivel de transaccionalidad con tarjeta de crédito, siendo éste el criterio principal.

Dado que ya se comprobó que los menores errores se producen al ocupar la última transición para predecir, dentro de este caso, el criterio que debe primar es el segundo, ya que busca que la segmentación describa el comportamiento transaccional de los clientes y por esto, debe ser capaz de diferenciar los distintos segmentos. Este criterio, si bien puede ser menos objetivo que el primero, toma en cuenta las características del negocio en cuestión, ya que por ejemplo no serviría una segmentación que tuviese un 0,1% de error, pero que solo tuviese 2 segmentos, dado que no aporta información valiosa.

Para observar una comparación del error cometido en la predicción de las probabilidades de transición se expondrá a continuación dos tipos de error:

- Error Promedio = Promedio( $\frac{ABS(Probabilidad Real_{ij} - Probabilidad Estimada_{ij})}{Probabilidad Real_{ij}}$ )

Donde:

$ABS(x - y)$ : Módulo de la resta entre  $x$  e  $y$ .

$i, j$ : Corresponden al segmento inicial y final, respectivamente en una transición. También se asocia a un casillero dentro de la matriz de transición.

- Error Cuadrático Medio  
$$= \sqrt{\sum_{i=0}^n \sum_{j=0}^n (Probabilidad Estimada_{ij} - Probabilidad Real_{ij})^2}$$

Donde

$i, j$ : corresponden al segmento inicial y final, respectivamente en una transición. También se asocia a un casillero dentro de la matriz de transición.

Ahora se puede observar cada uno de estos errores para las distintas segmentaciones:

**Tabla 11: Errores de las distintas segmentaciones**

<b>Segmentación</b>	<b>Error Promedio</b>	<b>Error Cuadrático Medio</b>
<b>R</b>	7,66%	5,66%
<b>F</b>	10,57%	9,93%
<b>M</b>	8,20%	7,80%
<b>N</b>	21,74%	14,60%
<b>RF</b>	6,89%	3,96%
<b>RM</b>	7,54%	5,64%
<b>FM</b>	8,98%	6,24%
<b>FN</b>	12,61%	9,75%
<b>RN</b>	11,93%	10,77%
<b>NM</b>	17,47%	10,78%
<b>RFM</b>	6,90%	3,97%
<b>RNM</b>	12,03%	10,54%
<b>RFN</b>	7,03%	4,04%
<b>FNM</b>	12,79%	9,85%
<b>RFNM</b>	7,05%	4,01%
<b>Promedio</b>	<b>10,63%</b>	<b>7,84%</b>

Las segmentaciones no superan en error promedio el 22% y en error cuadrático medio el 15%.

A continuación se muestra la segmentación que a priori fue una de las que tuvo un mejor desempeño en cuanto a distinguir a los clientes en función de su transaccionabilidad.

Segmentación realizada con variables las FNM:

Tabla 12: Descripción de los segmentos con respecto a las variables RFMN

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	11.037	17,0%	0,0	121,0	42,9	33,8	89,0	122,0	116,6	9,9
2	9.880	15,2%	0,0	121,0	37,0	32,1	42,7	88,8	61,1	12,6
3	25.637	39,5%	0,0	121,0	28,3	29,3	6,0	43,8	24,5	8,4
4	18.420	28,3%	0,0	121,0	13,0	19,7	0,5	27,4	7,5	3,2

CLUSTER	M [U.F]				N [número de transacciones en el período]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,6	5,0	1,0	9,0	1,6	1,1	0,4
2	0,0	25,5	2,9	3,4	1,0	19,0	2,7	1,8	0,7
3	0,0	25,3	2,1	2,3	1,0	27,0	5,4	2,4	1,3
4	0,0	21,7	1,6	1,3	1,0	72,0	20,0	12,1	5,0

Esta segmentación permite diferenciar a los clientes con un número bajo de transacciones con respecto a los que tienen un gran número de éstas (variable N). Además, el tamaño de los segmentos es balanceado, es decir, no existen segmentos muy pequeños ni muy grandes.

El problema de esta segmentación es que el grupo 4 (de alto consumo) posee un 28% del total de clientes activos del período, cifra muy alta considerando el tipo de clientes que son, por lo cual este segmento se debe desagregar para obtener información del perfil de compra de estos clientes.

Para dividir el segmento 4 en 2 subsegmentos se utiliza nuevamente las distintas combinaciones de las variables R, F, M y N para así determinar la mejor división, con respecto al aumento de error (dado que se espera que el error suba) y la separación de los clientes en dos subsegmentos distintos: Alto consumo y *Heavy Users*

Se realiza la subdivisión de este *cluster* con las variables antes señaladas y se determina que la segmentación que mejor separa a los *Heavy Users* del resto de alto consumo es la obtenida con las variables FNM. La nueva segmentación queda de la siguiente manera<sup>8</sup>:

<sup>8</sup> Para conocer los estadísticos descriptivos de las otras segmentaciones consultar ANEXO D

Tabla 13: Resultado de la subdivisión del segmento 4 de la segmentación FNM<sup>9</sup>

CLUSTER	TAMAÑO		R				F			
			[días]				[días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	11.037	17,0%	0,0	121,0	42,9	33,8	89,0	122,0	116,6	9,9
2	9.880	15,2%	0,0	121,0	37,0	32,1	42,7	88,8	61,1	12,6
3	25.637	39,5%	0,0	121,0	28,3	29,3	6,0	43,8	24,5	8,4
4	14.885	22,9%	0,0	121,0	15,0	21,1	0,5	27,4	8,5	2,7
5	3.535	5,4%	0,0	87,0	4,6	7,2	0,6	26,3	3,3	1,3

CLUSTER	M				N				
	[U.F.]				[número de transacciones en el período]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,6	5,0	1,0	9,0	1,6	1,1	0,4
2	0,0	25,5	2,9	3,4	1,0	19,0	2,7	1,8	0,7
3	0,0	25,3	2,1	2,3	1,0	27,0	5,4	2,4	1,3
4	0,1	21,7	1,6	1,4	1,0	32,0	15,1	5,2	3,8
5	0,0	18,8	1,3	1,0	28,0	72,0	40,5	11,2	10,1

A partir del antiguo *cluster 4* se obtiene el segmento de *Heavy Users* que corresponde al 5,4% del total de clientes activos los que se caracterizan por un alto grado de transacciones, con un promedio de 10 mensuales a diferencia de las 4 realizadas por el actual segmento 4.

Con respecto al error de predicción de esta nueva segmentación, se observa que antes de dividir el antiguo segmento 4, el error promedio era de 12,79% y el error cuadrático medio de 9,85%; posterior a la división la segmentación definitiva tiene los siguientes errores: error promedio de 16,83% y un error cuadrático medio de 13,82%. Finalmente, a pesar del aumento de los errores se elige esta configuración, dado que permite separar a los clientes con respecto a su transaccionalidad con tarjeta de crédito y en específico permite identificar a los *Heavy Users* (segmento 5)<sup>10</sup>.

Una vez definida la segmentación definitiva, es posible construir la matriz de transición para los distintos segmentos, entre los períodos. Los 5 segmentos obtenidos más el segmento de clientes inactivos (los cuales no fueron considerados en la segmentación) constituirán los 6 estados de las cadenas de Markov.

<sup>9</sup> En el ANEXO E se puede consultar los estadísticos descriptivos de las otras segmentaciones realizadas

<sup>10</sup> Se tomó esta decisión en base a criterio de expertos.

#### 5.4. CADENAS DE MARKOV: MATRIZ DE TRANSICIÓN.

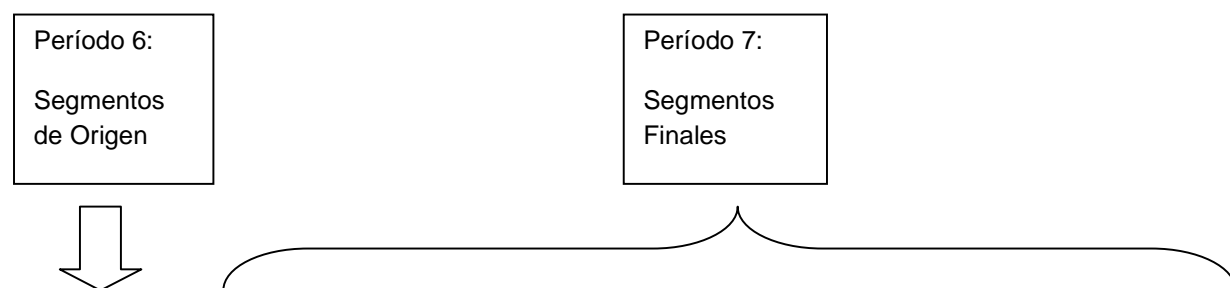
Dado que ya se ha definido la segmentación final a utilizar, a continuación se muestran los estadísticos descriptivos de 3 las variables que describen los segmentos (FNM) y el nombre que tendrá cada uno de estos:

Tabla 14: Segmentos finales y estadísticos descriptivos promedio de las variables FNM

<i>CLUSTER</i>	<b>F</b>	<b>N</b>	<b>M</b>
<b>0: Inactivo</b>	0	0	0
<b>1: Muy Bajo Consumo</b>	116,6	1,1	4,6
<b>2: Bajo Consumo</b>	61,1	1,8	2,9
<b>3: Intermedio</b>	24,5	2,4	2,1
<b>4: Alto Consumo</b>	8,5	5,2	1,6
<b>5: Heavy Users</b>	3,3	11,2	1,3

A continuación se muestra la matriz de transición P para la evolución desde el período 6 al 7, con lo cual se construye la cadena de Markov:

Figura 27: Matriz de transición de un periodo



<b>Segmentos</b>	<b>Inactivo</b>	<b>Muy Bajo Consumo</b>	<b>Bajo Consumo</b>	<b>Intermedio</b>	<b>Alto Consumo</b>	<b>Heavy Users</b>
<b>Inactivo</b>	86,21%	10,07%	2,42%	1,04%	0,21%	0,05%
<b>Muy Bajo Consumo</b>	31,95%	16,92%	20,49%	26,08%	4,25%	0,32%
<b>Bajo Consumo</b>	18,13%	12,80%	22,83%	38,53%	7,28%	0,43%
<b>Intermedio</b>	9,62%	6,03%	13,74%	53,96%	15,98%	0,67%
<b>Alto Consumo</b>	3,62%	1,62%	4,12%	32,85%	51,49%	6,30%
<b>Heavy Users</b>	0,91%	0,15%	0,51%	5,96%	39,21%	53,25%

Fuente: Elaboración Propia

Gran parte de los *Heavy Users* se mantiene dentro de los segmentos de alto consumo en el período siguiente, lo que habla del grado de vinculación que genera el la tarjeta de crédito dentro de los clientes. Así mismo, en el caso del segmento de clientes inactivos, se observa que un alto porcentaje de ellos continúa en este estado al período siguiente.

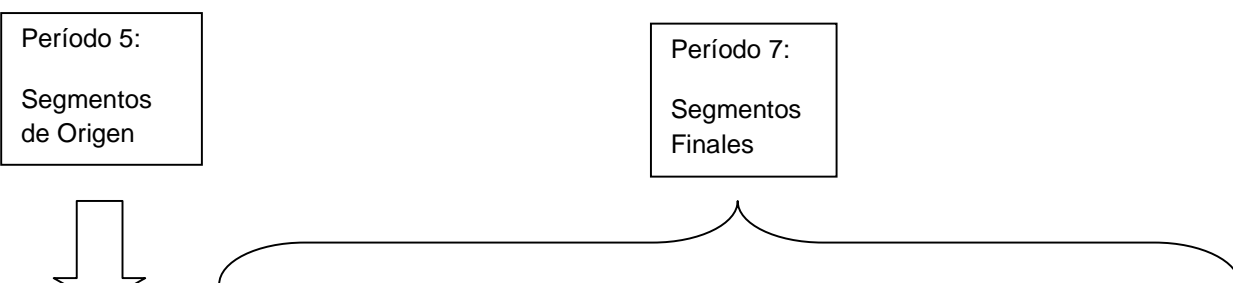
Además dentro de los segmentos de consumo intermedio se puede notar una transición en términos porcentuales que no es despreciable y por tanto a priori se estima que en dichos segmentos se podría incentivar el aumento de consumo más efectivamente.

## 5.5. CADENAS DE MARKOV: ANÁLISIS DE 3 PERÍODOS.

Lo que se busca en esta sección es encontrar rutas de los clientes desde segmentos de bajo consumo hacia estados mayores. Para esto se hace un análisis de tres períodos en el cual se observa el comportamiento de un cliente desde un estado inicial y se mira su estado final tres períodos después (en la sección anterior se observó solamente el análisis de un período)

Con los datos históricos con que se cuenta, el análisis de tres períodos más actual que se puede realizar comienza con el estado inicial en el período 4 y termina en el 7. A continuación se muestran las matrices de transición de 2 y 3 períodos que terminan en el período 7:

Figura 28: Matriz de transición de 2 Periodos



Segmentos	Inactivo	Muy Bajo Consumo	Bajo Consumo	Intermedio	Alto Consumo	Heavy Users
Inactivo	82,32%	7,76%	3,41%	4,73%	1,58%	0,20%
Muy Bajo Consumo	31,52%	21,22%	16,91%	24,13%	5,73%	0,48%
Bajo Consumo	20,39%	15,46%	19,03%	35,71%	8,73%	0,68%
Intermedio	11,64%	7,97%	13,69%	48,20%	17,33%	1,17%
Alto Consumo	5,82%	2,75%	5,46%	32,50%	46,28%	7,19%
Heavy Users	2,29%	0,79%	1,14%	9,92%	37,01%	48,84%

Fuente: Elaboración propia



Figura 29: Matriz de transición de 3 Periodos

Segmentos	Inactivo	Muy Bajo Consumo	Bajo Consumo	Intermedio	Alto Consumo	Heavy Users
Inactivo	78,70%	7,95%	4,18%	6,44%	2,35%	0,39%
Muy Bajo Consumo	31,14%	21,00%	14,80%	25,20%	7,16%	0,70%
Bajo Consumo	21,78%	16,05%	18,16%	33,99%	9,38%	0,64%
Intermedio	14,25%	9,24%	13,25%	44,12%	17,71%	1,44%
Alto Consumo	7,66%	4,01%	6,54%	32,92%	41,54%	7,32%
Heavy Users	3,32%	1,37%	1,21%	11,16%	35,23%	47,70%

Fuente: Elaboración propia

De las matrices de transición de 2 y 3 períodos se observa que la distribución de las mismas no varía mayormente con respecto a la matriz de transición de un período (modelo estacionario), por lo que es consistente señalar que los clientes de grupos de alto consumo en general continúan en estos estados y el grupo de inactivos mantiene en su mayoría esta condición ya sea de un período a otro o bien después de tres períodos. Así mismo, se observa que los segmentos de consumo moderado, tienen una movilidad que permite pensar en crear acciones que fomenten el aumento de consumo de aquellos clientes.

## 6. MODELOS DE PROPENSIÓN PARA INCENTIVAR USO DE TARJETA

En este capítulo se abordan los modelos de propensión que permiten identificar los clientes a los cuales se puede incentivar para que aumenten su consumo con tarjeta de crédito. En la sección de cadenas de Markov se realizó el ejercicio de observar la evolución transaccional que tienen los clientes a nivel agregado durante el tiempo. En esta etapa se individualizará a los clientes más propensos a aumentar su consumo y de este análisis nacerán propuestas que serán expuestas en las conclusiones de este trabajo.

## 6.1. CARACTERIZACIÓN SEGMENTOS OBTENIDOS

Luego de realizar la segmentación de los clientes, se expone las características de cada uno de los *clusters* mediante los estadísticos descriptivos para el último período:

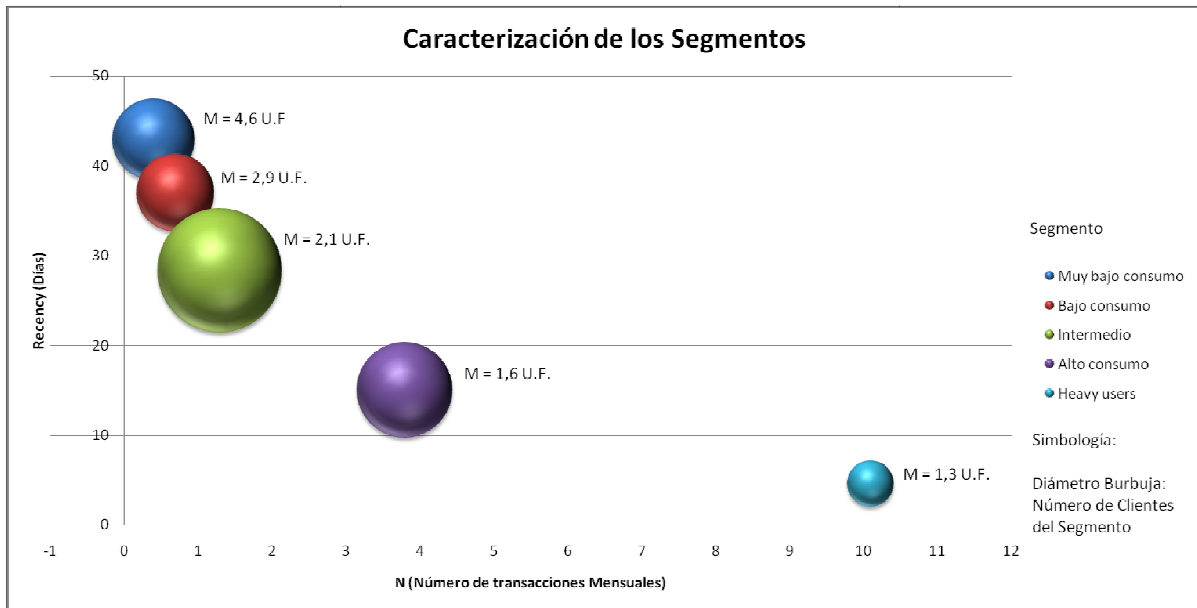
Tabla 15: Estadísticos descriptivos de cada segmento

CLUSTER	TAMAÑO		R				F			
			[días]				[días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	11.037	17,0%	0,0	121,0	<b>42,9</b>	33,8	89,0	122,0	<b>116,6</b>	9,9
2	9.880	15,2%	0,0	121,0	<b>37,0</b>	32,1	42,7	88,8	<b>61,1</b>	12,6
3	25.637	39,5%	0,0	121,0	<b>28,3</b>	29,3	6,0	43,8	<b>24,5</b>	8,4
4	14.885	22,9%	0,0	121,0	<b>15,0</b>	21,1	0,5	27,4	<b>8,5</b>	2,7
5	3.535	5,4%	0,0	87,0	<b>4,6</b>	7,2	0,6	26,3	<b>3,3</b>	1,3

CLUSTER	M				N				
	[U.F.]				[número de transacciones en el período]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	<b>4,6</b>	5,0	1,0	9,0	<b>1,6</b>	1,1	0,4
2	0,0	25,5	<b>2,9</b>	3,4	1,0	19,0	<b>2,7</b>	1,8	0,7
3	0,0	25,3	<b>2,1</b>	2,3	1,0	27,0	<b>5,4</b>	2,4	1,3
4	0,1	21,7	<b>1,6</b>	1,4	1,0	32,0	<b>15,1</b>	5,2	3,8
5	0,0	18,8	<b>1,3</b>	1,0	28,0	72,0	<b>40,5</b>	11,2	10,1

Las principales diferencias entre estos grupos se encuentran en el número de transacciones (N) así como en el monto de estas (M). Lo cual se puede observar en el siguiente gráfico, donde el tamaño de las burbujas representa el número relativo de clientes del segmento:

Figura 30: Caracterización de los Segmentos



Fuente: Elaboración propia

Se distinguen los 5 grupos identificados en la segmentación, siendo el más grande el de consumo Intermedio. El grupo *Heavy Users* se caracteriza por ser pequeño en número, pero a nivel de transaccionalidad está muy por encima del resto de los segmentos.

Se debe destacar que a medida que los clientes tienen más transacciones el monto promedio de éstas va disminuyendo. Este comportamiento es lógico, ya que una persona que compra en un rubro solo una vez al mes gasta un monto mayor que una persona que compra varias veces y que por tanto estará dividiendo su consumo en un número mayor de transacciones.

Es necesario analizar la conveniencia o no de incentivar el uso de tarjeta de crédito en términos de montos transaccionados por cliente, ya que podría ocurrir que el consumo mensual de una persona que utiliza la tarjeta una vez al mes sea mayor que el de un cliente que hace uso frecuente de la misma y por lo tanto no se justificaría incentivar el uso de la tarjeta de crédito. En la siguiente sección se efectúa el análisis del punto antes expuesto.

### 6.1.1. ANÁLISIS DE SEGMENTOS INTERESANTES EN TÉRMINOS DE MONTOS TRANSACCIONADOS

En esta sección analiza la relación existente entre el número de transacciones (N) y el monto promedio de éstas (M), ya que en la sección anterior se observó que en los segmentos con mayor número de transacciones en el período, menor es el monto promedio de las mismas.

A continuación la tabla 16 muestra el valor total que representa un cliente promedio de cada uno de los segmentos, en términos de ingresos (NxM)

Tabla 16: Tamaño e importancia de los segmentos

Segmento	N [Número de transacciones]	M [U.F.]	Valor Cliente Promedio: N x M [U.F.]	Tamaño Segmento [Número de clientes]	Tamaño Segmento [Porcentaje relativo]	Valor Segmento: N x M x Tam.Seg [U.F.]
Muy Bajo Consumo	1,6	4,6	7,36	11.037	17,00%	81.232,3
Bajo Consumo	2,7	2,9	7,83	9.880	15,20%	77.360,4
Intermedio	5,4	2,1	11,34	25.637	39,50%	290.723,6
Alto Consumo	15,1	1,6	24,16	14.885	22,90%	359.621,6
Heavy Users	40,5	1,3	52,65	3.535	5,40%	186.117,8

En la tabla anterior se observa que el valor monetario de un cliente (en U.F.) aumenta a medida que se incrementa la cantidad de transacciones por período, a pesar de que el monto promedio de éstas disminuya.

Lo anterior justifica la importancia de incentivar el uso de tarjeta de crédito a pesar del *trade off* entre el número de transacciones y el monto de éstas.

Se debe destacar que un cliente de alto consumo realiza compras por montos de más del doble que los clientes del segmento Intermedio al igual que los clientes del *cluster* de *Heavy Users* en comparación con los de Alto consumo, reafirmando nuevamente la importancia de fomentar el consumo con tarjeta de crédito en los distintos segmentos.

### 6.1.2. PERFILES DE CONSUMO DE CADA SEGMENTO

La segmentación que se ha ocupado para realizar los análisis anteriores creó en base al comportamiento transaccional de los clientes, esta sección tiene como objetivo realizar 3 agrupaciones del conjunto total de transacciones, buscando diferenciar la el tipo de consumo de cada uno de los segmentos. La información que entregan estas agrupaciones será usada en los modelos de propensión<sup>11</sup>. A continuación se exponen las 3 agrupaciones realizadas con las transacciones de los clientes:

<sup>11</sup> Ya que usar variables transaccionales sería redundante.

**Tabla 17: Agrupación General**

<b>Agrupación General</b>
AVANCES
COMPRAS
RECAUDACION SII

Esta agrupación permite identificar si un segmento tiene un perfil de compra o bien de pedir dinero por medio de avances.

La segunda agrupación desagrega las categorías de avances y compras, con lo que se forma la siguiente:

**Tabla 18: Agrupación por Tipo de compra**

<b>Tipo de uso</b>	<b>Agrupación por Tipo de compra</b>
COMPRAS	COMPRAS
	COMPRAS 3CPC
	COMPRA N CUOTAS
	COMPRA CUOTA FIJA
	COMPRAS INTERNACIONALES
	COMPRA CASINO NAC-INT
	COMPRA BENCINA
AVANCES	AVANCE EN EFECTIVO
	AVANCE EN CUOTAS
	AVANCE INTERNACIONAL
RECAUDACION SII	RECAUDACION SII

Esta agrupación abre las categorías de compras y avances y las clasifica con respecto al TIPO de transacción que realiza la persona. Esta desagregación permite diferenciar entre compras/avances en efectivo o en distintas modalidades de cuotas, distinguiendo también las compras y avances internacionales, así como transacciones algunos otros rubros (casinos y compra de bencina).

Finalmente se presenta una tercera agrupación de las transacciones, que se caracteriza por segmentarlas con respecto al RUBRO específico donde se realizan:

Tabla 19: Agrupación por Rubro de compra

Tipo de uso	Agrupación por Rubro de compra
AVANCES	AVANCES
COMPRAS	GRANDES TIENDAS
	VESTUARIO Y CALZADO
	SUPERMERCADOS
	OTROS CONSUMOS
	MUEBLES Y ARTICULOS ELECTRONICOS
	LINEAS AEREAS
	RECAUDACION <sup>12</sup>
	COMBUSTIBLES
	FARMACIAS
	RECAUDACION SECTOR PUBLICO <sup>13</sup>
	ACCESORIOS Y SERVICIOS AUTOMOTRICES
	CLINICAS Y PRESTADORES
	OTROS T&E <sup>14</sup>
	OTROS SERVICIOS <sup>15</sup>
	OTROS CONSUMO INTERNACIONAL
	RESTAURANTES
	HOTELES
	EDUCACION
	OPERADORES RENT A CAR
	CASINOS DE JUEGO DE AZAR
COMPRAS INTERNET VARIAS	

Como se señalaba anteriormente, esta agrupación depende del RUBRO de las transacciones y permite distinguir las preferencias de los clientes en las compras con tarjeta de crédito dentro de los distintos sectores del mercado.

En esta agrupación mantiene la categoría Avances, dado que no se puede asociar a alguno de los otros rubros.

<sup>12</sup> Pago servicios básicos, pago ambulancias, casas de reposo, AFP, Isapres, etc.

<sup>13</sup> Conservador de bienes raíces, notarías, poder ejecutivo, etc.

<sup>14</sup> Entretención en general (pubs, bares, piscinas, teatro, etc.)

<sup>15</sup> Escuelas de conductores, arquitectos, abogados, gasfitería, salones de belleza, etc.

Una vez realizadas estas agrupaciones, se realiza un perfilamiento de los 5 segmentos activos en función éstas para identificar diferencias entre los clientes de los distintos segmentos.

Se muestra a continuación la agrupación General de las transacciones para cada uno de los *clusters* de clientes activos<sup>16</sup>:

**Tabla 20: Agrupación General segmento Muy bajo consumo**

AGRUPACION GENERAL	Número de transacciones [#]	Monto transacciones [U.F.]	Número de transacciones [%]	Monto transacciones [%]	M (Monto / # trans.) [U.F.]
COMPRAS	7.036	39.787,10	40,1%	53,6%	5,65
AVANCES	10.491	34.355,29	59,8%	46,3%	3,27
RECAUDACION SII	20	23,87	0,1%	0,0%	1,19
<b>total</b>	<b>17.547</b>	<b>74.166</b>	<b>100,0%</b>	<b>100,0%</b>	<b>4,23</b>

**Tabla 21: Agrupación General segmento Bajo consumo**

AGRUPACION GENERAL	Número de transacciones [#]	Monto transacciones [U.F.]	Número de transacciones [%]	Monto transacciones [%]	M (Monto / # trans.) [U.F.]
AVANCES	16.898	39.088,26	63,8%	54,5%	2,31
COMPRAS	9.547	32.610,73	36,1%	45,5%	3,42
RECAUDACION SII	27	17,95	0,1%	0,0%	0,66
<b>Total</b>	<b>26.472</b>	<b>71.717</b>	<b>100,0%</b>	<b>100,0%</b>	<b>2,71</b>

**Tabla 22: Agrupación General segmento Intermedio**

AGRUPACION GENERAL	Número de transacciones [#]	Monto transacciones [U.F.]	Número de transacciones [%]	Monto transacciones [%]	M (Monto / # trans.) [U.F.]
AVANCES	98.441	163.477,75	71,8%	59,8%	1,66
COMPRAS	38.490	109.697,41	28,1%	40,1%	2,85
RECAUDACION SII	261	97,75	0,2%	0,0%	0,37
<b>Total</b>	<b>137.192</b>	<b>273.273</b>	<b>100,0%</b>	<b>100,0%</b>	<b>1,99</b>

<sup>16</sup> Las tablas están ordenadas por el Monto transacciones (%)

Tabla 23: Agrupación General segmento Alto consumo

AGRUPACION GENERAL	Número de transacciones [#]	Monto transacciones [U.F.]	Número de transacciones [%]	Monto transacciones [%]	M (Monto / # trans.) [U.F.]
AVANCES	183.875	241.505,25	81,6%	67,7%	1,31
COMPRAS	40.945	114.638,19	18,2%	32,2%	2,80
RECAUDACION SII	627	411,24	0,3%	0,1%	0,66
total	225.447	356.555	100,0%	100,0%	1,58

Tabla 24: Agrupación General segmento *Heavy Users*

AGRUPACION GENERAL	Número de transacciones [#]	Monto transacciones [U.F.]	Número de transacciones [%]	Monto transacciones [%]	M (Monto / # trans.) [U.F.]
COMPRAS	128.143	144.913,56	89,5%	77,5%	1,13
AVANCES	14.576	41.648,82	10,2%	22,3%	2,86
RECAUDACION SII	478	430,27	0,3%	0,2%	0,90
total	143.197	186.993	100,0%	100,0%	1,31

De las tablas anteriores se aprecia que en los cuatro primeros *clusters* hay un aumento en el número de avances a medida que se observan los segmentos de mayor consumo. A priori este comportamiento no era algo intuitivo, dado que se esperaba que los segmentos de menor consumo tuvieran mayor número de avances que los de mayor consumo, sin embargo, dado que en esta agrupación no se hace distinción entre los distintos tipos de avances, no se puede responder aún a este comportamiento<sup>17</sup>.

Los segmentos muestran diferencias en el monto de los avances, siendo los segmentos de menor consumo los que solicitan avances de mayores montos.

Finalmente, se observa que el segmento de *Heavy Users* tiene un perfil de consumo diferente a los demás grupos, dado que la mayoría de sus transacciones, tanto en monto como en número, son compras y no avances. Esto da cuenta de un segmento que utiliza la tarjeta de crédito de forma habitual.

A continuación se observa un perfilamiento de los segmentos con respecto al TIPO de transacciones que realizan<sup>18</sup>:

<sup>17</sup> La respuesta a este comportamiento se da en el análisis de la agrupación de transacciones por TIPO de compra

<sup>18</sup> Las tablas están ordenadas por el Monto transacciones (%)



Tabla 25: Agrupación por Tipo de compra segmento Muy Bajo Consumo

AGRUPACION POR TIPO DE COMPRA	Número de transacciones	Monto transacciones	Número de transacciones	Monto transacciones	M (Monto / # trans.)
	[#]	[U.F.]	[%]	[%]	[U.F.]
AVANCE EN CUOTAS	2.236	24.705,21	12,7%	33,3%	11,05
AVANCE EN EFECTIVO	4.760	14.837,22	27,1%	20,0%	3,12
COMPRA N CUOTAS	1.972	13.213,66	11,2%	17,8%	6,70
COMPRAS 3CPC	3.664	9.717,09	20,9%	13,1%	2,65
COMPRAS	3.292	6.105,22	18,8%	8,2%	1,85
OTROS <sup>19</sup>	1.623	5.587,86	9,2%	7,5%	3,44
<b>Total</b>	<b>17.547</b>	<b>74.166,26</b>	<b>100,0%</b>	<b>100,0%</b>	<b>4,23</b>

Tabla 26: Agrupación por Tipo de compra segmento Bajo Consumo

AGRUPACION POR TIPO DE COMPRA	Número de transacciones	Monto transacciones	Número de transacciones	Monto transacciones	M (Monto / # trans.)
	[#]	[U.F.]	[%]	[%]	[U.F.]
AVANCE EN EFECTIVO	8.135	18.925,71	30,7%	26,4%	2,33
AVANCE EN CUOTAS	1.342	13.416,69	5,1%	18,7%	10,00
COMPRAS 3CPC	5.987	12.656,06	22,6%	17,6%	2,11
COMPRA N CUOTAS	2.262	11.633,64	8,5%	16,2%	5,14
COMPRAS	6.285	8.674,93	23,7%	12,1%	1,38
OTROS <sup>20</sup>	2.461	6.409,91	9,3%	8,9%	2,60
<b>Total</b>	<b>26.472</b>	<b>71.716,94</b>	<b>100,0%</b>	<b>100,0%</b>	<b>2,71</b>

<sup>19</sup> Detalle categoría OTROS en ANEXO F

<sup>20</sup> Detalle categoría OTROS en ANEXO F

Tabla 27: Agrupación por Tipo de compra segmento Intermedio

AGRUPACION POR TIPO DE COMPRA	Número de transacciones	Monto transacciones	Número de transacciones	Monto transacciones	M (Monto / # trans.)
	[#]	[U.F.]	[%]	[%]	[U.F.]
AVANCE EN EFECTIVO	34.831	75.133,34	25,4%	27,5%	2,16
COMPRAS 3CPC	30.819	52.339,78	22,5%	19,2%	1,70
COMPRAS	47.675	51.042,75	34,8%	18,7%	1,07
COMPRA N CUOTAS	9.165	37.589,39	6,7%	13,8%	4,10
AVANCE EN CUOTAS	3.413	33.423,60	2,5%	12,2%	9,79
OTROS <sup>21</sup>	11.289	23.744,04	8,2%	8,7%	2,10
<b>Total</b>	<b>137.192</b>	<b>273.272,91</b>	<b>100,0%</b>	<b>100,0%</b>	<b>1,99</b>

Tabla 28: Agrupación por Tipo de compra segmento Alto Consumo

AGRUPACION POR TIPO DE COMPRA	Número de transacciones	Monto transacciones	Número de transacciones	Monto transacciones	M (Monto / # trans.)
	[#]	[U.F.]	[%]	[%]	[U.F.]
COMPRAS	101.570	96.750,83	45,1%	27,1%	0,95
AVANCE EN EFECTIVO	37.524	86.271,97	16,6%	24,2%	2,30
COMPRAS 3CPC	52.213	72.392,25	23,2%	20,3%	1,39
COMPRA N CUOTAS	12.335	41.295,25	5,5%	11,6%	3,35
AVANCE EN CUOTAS	2.624	25.265,37	1,2%	7,1%	9,63
OTROS <sup>22</sup>	19.181	34.579,02	8,5%	9,7%	1,80
<b>Total</b>	<b>225.447</b>	<b>356.554,68</b>	<b>100,0%</b>	<b>100,0%</b>	<b>1,58</b>

<sup>21</sup> Detalle categoría OTROS en ANEXO F

<sup>22</sup> Detalle categoría OTROS en ANEXO F

Tabla 29: Agrupación por Tipo de compra segmento *Heavy Users*

AGRUPACION POR TIPO DE COMPRA	Número de transacciones [#]	Monto transacciones [U.F.]	Número de transacciones [%]	Monto transacciones [%]	M (Monto / # trans.) [U.F.]
COMPRAS	84.005	76.133,79	58,7%	40,7%	0,91
COMPRAS 3CPC	29.832	36.653,11	20,8%	19,6%	1,23
AVANCE EN EFECTIVO	13.475	33.211,30	9,4%	17,8%	2,46
COMPRA N CUOTAS	5.094	16.405,17	3,6%	8,8%	3,22
COMPRAS INTERNACIONALES	4.046	7.792,87	2,8%	4,2%	1,93
OTROS <sup>23</sup>	6.745	16.796,41	4,7%	9,0%	2,49
<b>Total</b>	<b>143.197</b>	<b>186.992,65</b>	<b>100,0%</b>	<b>100,0%</b>	<b>1,31</b>

En las tablas anteriores se observa un comportamiento razonable: los clientes de los segmentos de menor consumo tienen un mayor número de avances y compras en varias cuotas, a diferencia de los segmentos de mayor consumo donde las compras y avances se realizan principalmente en una cuota.

Es en esta sección donde se da respuesta al comportamiento observado en el análisis anterior (agrupación general), donde se constató que el número de avances en *clusters* de mayor consumo era mayor (en porcentaje) que en los segmentos menores. Esto se justifica ya que los *clusters* de mayor consumo realizan una mayor cantidad de avances en efectivo (una cuota) mientras que en los segmentos de consumo inferior los avances se realizan en varias cuotas, lo cual no se podía distinguir en la agrupación general.

Nuevamente se aprecia que el segmento de *Heavy Users* está constituido por un grupo de personas que principalmente realizan compras, ya que si se suma los porcentajes de compra a una cuota o 3 cuotas precio contado, más del 70% del número de transacciones se atribuye a estas categorías y así también más del 60% del monto.

A continuación se muestra el perfil de compra de los segmentos con respecto a la agrupación por RUBRO de compra<sup>24</sup>:

<sup>23</sup> Detalle categoría OTROS en ANEXO F

<sup>24</sup> Las tablas están ordenadas por el Monto transacciones (%)

Tabla 30: Agrupación por Rubro de compra segmento Muy Bajo Consumo

AGRUPACION POR RUBRO DE COMPRA	Número de transacciones	Monto transacciones	Número de transacciones	Monto transacciones	M (Monto / # trans.)
	[#]	[U.F.]	[%]	[%]	[U.F.]
AVANCES	7.036	39.787,10	40,1%	53,6%	5,65
GRANDES TIENDAS	1.778	7.337,17	10,1%	9,9%	4,13
MUEBLES Y ARTICULOS ELECTRONICOS	836	5.863,68	4,8%	7,9%	7,01
LINEAS AEREAS	603	3.671,67	3,4%	5,0%	6,09
VESTUARIO Y CALZADO	1.519	3.117,45	8,7%	4,2%	2,05
OTROS CONSUMOS	989	2.507,33	5,6%	3,4%	2,54
OTROS <sup>25</sup>	4.786	11.881,86	27,3%	16,0%	2,48
Total	17.547	74.166,26	100,0%	100,0%	4,23

Tabla 31: Agrupación por Rubro de compra segmento Bajo Consumo

AGRUPACION POR RUBRO DE COMPRA	Número de transacciones	Monto transacciones	Número de transacciones	Monto transacciones	M (Monto / # trans.)
	[#]	[U.F.]	[%]	[%]	[U.F.]
AVANCES	9.547	32.610,73	36,1%	45,5%	3,42
GRANDES TIENDAS	2.698	7.948,31	10,2%	11,1%	2,95
MUEBLES Y ARTICULOS ELECTRONICOS	933	4.499,67	3,5%	6,3%	4,82
VESTUARIO Y CALZADO	2.504	4.303,83	9,5%	6,0%	1,72
SUPERMERCADOS	2.781	4.137,48	10,5%	5,8%	1,49
LINEAS AEREAS	609	3.677,10	2,3%	5,1%	6,04
OTROS <sup>26</sup>	7.400	14.539,82	28,0%	20,3%	1,96
Total	26.472	71.716,94	100,0%	100,0%	2,71

<sup>25</sup> Detalle categoría OTROS en ANEXO F

<sup>26</sup> Detalle categoría OTROS en ANEXO F

Tabla 32: Agrupación por Rubro de compra segmento Intermedio

AGRUPACION POR RUBRO DE COMPRA	Número de transacciones	Monto transacciones	Número de transacciones	Monto transacciones	M (Monto / # trans.)
	[#]	[U.F.]	[%]	[%]	[U.F.]
AVANCES	38.490	109.697,41	28,1%	40,1%	2,85
GRANDES TIENDAS	12.144	27.473,82	8,9%	10,1%	2,26
SUPERMERCADOS	17.089	20.658,86	12,5%	7,6%	1,21
VESTUARIO Y CALZADO	11.599	18.472,81	8,5%	6,8%	1,59
RECAUDACION	15.004	17.855,84	10,9%	6,5%	1,19
MUEBLES Y ARTICULOS ELECTRONICOS	3.704	13.557,21	2,7%	5,0%	3,66
OTROS <sup>27</sup>	39.162	65.556,96	28,5%	24,0%	1,67
<b>Total</b>	<b>137.192</b>	<b>273.272,91</b>	<b>100,0%</b>	<b>100,0%</b>	<b>1,99</b>

Tabla 33: Agrupación por Rubro de compra segmento alto consumo

AGRUPACION POR RUBRO DE COMPRA	Número de transacciones	Monto transacciones	Número de transacciones	Monto transacciones	M (Monto / # trans.)
	[#]	[U.F.]	[%]	[%]	[U.F.]
AVANCES	40.945	114.638,19	18,2%	32,2%	2,80
SUPERMERCADOS	39.242	40.185,95	17,4%	11,3%	1,02
GRANDES TIENDAS	19.422	34.817,40	8,6%	9,8%	1,79
RECAUDACION	29.623	32.368,41	13,1%	9,1%	1,09
VESTUARIO Y CALZADO	16.520	23.970,88	7,3%	6,7%	1,45
OTROS CONSUMOS	17.379	18.606,71	7,7%	5,2%	1,07
OTROS <sup>28</sup>	62.316	91.967,15	27,6%	25,8%	1,48
<b>Total</b>	<b>225.447</b>	<b>356.554,68</b>	<b>100,0%</b>	<b>100,0%</b>	<b>1,58</b>

<sup>27</sup> Detalle categoría OTROS en ANEXO F

<sup>28</sup> Detalle categoría OTROS en ANEXO F

Tabla 34: Agrupación por Rubro de compra segmento *Heavy Users*

AGRUPACION POR RUBRO DE COMPRA	Número de transacciones [#]	Monto transacciones [U.F.]	Número de transacciones [%]	Monto transacciones [%]	M (Monto / # trans.) [U.F.]
AVANCES	14.576	41.648,82	10,2%	22,3%	2,86
SUPERMERCADOS	35.893	32.295,83	25,1%	17,3%	0,90
RECAUDACION	18.193	19.841,51	12,7%	10,6%	1,09
GRANDES TIENDAS	9.947	15.582,34	6,9%	8,3%	1,57
COMBUSTIBLES	14.474	10.134,18	10,1%	5,4%	0,70
OTROS CONSUMOS	10.741	9.743,44	7,5%	5,2%	0,91
OTROS <sup>29</sup>	39.373	57.746,54%	27,5%	30,9%	1,47
<b>Total</b>	143.197	186.992,65	100,0%	100,0%	1,31

Antes de realizar el análisis de esta agrupación, se debe señalar que los AVANCES no se encuentran desagregados, por lo cual los valores de este ítem serán mayores a los demás en casi todos los segmentos.

Los *clusters* de menor consumo tienden a realizar transacciones de mayores montos, por ejemplo en el rubro Supermercados el monto promedio de transacción para el segmento de bajo consumo es de 1,5 U.F. mientras que en el segmento de *Heavy Users* tiene por monto promedio 0,9 U.F., lo cual era esperable ya que los segmentos de más bajo consumo normalmente en este rubro hacen compras de mayor cantidad pero más distantes en tiempo una de otra.

Si se observa con atención las tablas del grupo de *Heavy Users* versus las tablas de los segmentos de menor consumo, se aprecia que este segmento realiza gran parte de sus compras en rubros que podrían ser llamados “cotidianos”: Supermercados, Farmacias, Combustibles, etc. Es decir, las transacciones de los segmentos de alto consumo son de carácter diario o habitual. Lo mismo se observa en el rubro Recaudación, dado que éste agrupa el pago de servicios básicos y otros utilizando la tarjeta de crédito. Todo lo anterior indica que los segmentos de alto consumo utilizan la tarjeta de crédito para todo tipo de compras sin importar el rubro.

En los segmentos de menor consumo, la tarjeta de crédito es usada para la compra de bienes o artículos de mayor valor o durabilidad. Esto se puede constatar en las tablas de los segmentos de menor consumo, donde las categorías “Grandes Tiendas”, “Muebles y artículos electrónicos” y “Líneas aéreas” son los que capturan los mayores montos, efecto que no ocurre en los segmentos de mayor consumo.

Ejemplos de esta diferencia en el uso de la tarjeta de crédito se observa en los rubros Supermercados y Farmacias los cuales adquieren importancia en términos de

<sup>29</sup> Detalle categoría OTROS en ANEXO F

porcentaje de número de transacciones a medida que se analizan los segmentos de mayor consumo.

Los datos analizados no permiten describir el comportamiento de los clientes del segmento Inactivos, ya que no existen transacciones de las cuales obtener información.

Dado que ya se tiene una descripción desde distintos enfoques para los segmentos, en la siguiente sección incorporará esta información para crear los modelos de propensión.

## **6.2. DESCRIPCIÓN DE LOS MODELOS DE PROPENSIÓN**

Los modelos de propensión que se desarrollan tienen como objetivo identificar patrones en los clientes que permitan llevarlos a segmentos de mayor consumo.

A partir de la segmentación, se obtienen 6 *clusters* de clientes:

1. Segmento Inactivo
2. Segmento Muy bajo consumo
3. Segmento Bajo consumo
4. Segmento Intermedio
5. Segmento Alto consumo
6. Segmento *Heavy Users*

A partir de estos segmentos, los modelos propensión buscan encontrar los patrones de consumo de los clientes que permitan predecir si una persona de un segmento determinado en el período siguiente aumentará o no su consumo.

Como se ha podido apreciar, los segmentos tienen diferencias en cuanto a la forma en que utilizan la tarjeta de crédito tanto en los rubros en que lo hacen, así como la cantidad de veces que la ocupan. Para abordar estas diferencias es que se crean distintos modelos para cada uno de los segmentos.

Dado que se busca encontrar los patrones de aquellos clientes que aumentan su consumo (especialmente los del segmento inactivo), hacer estos modelos tiene sentido para todos los segmentos excepto para el segmento de *Heavy Users*, ya que estos clientes no tienen un segmento superior al cual puedan evolucionar.

## **6.3. SELECCIÓN DE DATOS**

Los datos utilizados son los resultantes de las secciones anteriores del presente trabajo, sin embargo existe una diferencia entre los modelos, en cuanto a los datos a utilizar:

Para el modelo del segmento de inactivos se utiliza datos transaccionales de otros medios de pago: Cuenta Corriente, CuentaRut, Cuenta Vista y otros.

En los demás segmentos, además de esta nueva fuente de datos, se ocupa las agrupaciones expuestas en el punto 6.1.2 de este trabajo para identificar perfiles de

consumo en los clientes. En estos modelos se omite el uso de variables transaccionales directamente usadas en la segmentación de los *clusters* de modo que de evitar redundancias.

A estas fuentes de datos se debe agregar todos los datos de tenencias de productos, saldos de productos, deudas en el banco, deuda SBIF, share de consumo, demográficos, pago automático de tarjeta, campañas realizadas, etc.

#### 6.4. PREPROCESAMIENTO

Dado que en las etapas anteriores ya se realizó el preprocesamiento de los datos, no es necesario buscar *outliers* o datos faltantes en las variables transaccionales adicionales que se utilizan (saldos de productos, deudas netas, etc.).

Con respecto a las variables de perfil de consumo de los clientes, cada transacción tiene asociado un código de rubro, del cual se realizan las tres agrupaciones expuestas en el punto anterior, debido a esto tampoco se observan datos faltantes u *outliers*.

#### 6.5. TRANSFORMACIÓN

En esta etapa se crea nuevas variables para ser incluidas en los modelos de propensión de cada segmento, las variables de mayor importancia dentro de las transformaciones realizadas son:

- Número de transacciones y montos en otros medios de pago, esto por cada producto y tipo de transacción. Estas variables son utilizadas principalmente en el modelo de clientes inactivos.
- Porcentaje de consumo con tarjeta de crédito en distintos rubros, para determinar el perfil de compra de cada cliente. Estas variables se utilizan en el resto de los modelos de clientes activos.
- Se crea quintiles para variables importantes con muchas categorías y mediante tablas de contingencia se analiza si estas nuevas variables tienen correlación con la variable dependiente y de este modo incluirlas.
- Se crean variables *delta* que capturan diferencias entre períodos para una misma variable, permitiendo observar la evolución de compra de los clientes.

En esta etapa se define la variable dependiente AUMENTA, que toma valores 0 o 1, donde el valor 1 señala cuando la persona evoluciona a algún segmento de mayor consumo en el período siguiente y 0 cuando una persona mantiene o baja su consumo.

Esta variable es definida de esta manera ya que una definición alternativa podría presentar los siguientes problemas:

- Definir como los unos de la variable dependiente a los clientes que desde un segmento llegan a uno de mayor consumo en específico: Esta definición tiene dos problemas principales, en primer lugar se desearía que ese segmento de



destino fuese el de *Heavy Users*, sin embargo los casos que llegan a ese grupo desde *clusters* de consumo inferior son mínimos por lo que no se tiene datos suficientes con los cuales entrenar el modelo, ya que es más probable que estos clientes incrementen su consumo lentamente en vez de evolucionar se forma explosiva. El segundo lugar la proporción de clientes inactivos que llegan a segmentos de consumo intermedios y mayores es mínima (menos del 3%) por lo que al entrenar los modelos éstos no lograrán predecir de manera correcta los clientes del segmento inactivo que pueden comenzar a utilizar su tarjeta, fallando en uno de los objetivos principales de este trabajo, que es aumentar la tasa de penetración en las tarjetas de crédito.

## 6.6. DATA MINING

En esta sección se desarrolla la parte principal de los modelos de propensión, es donde se buscan los patrones que explican el aumento de consumo de los clientes. Para obtener estos patrones, los modelos deben ser entrenados en una *base llamada train* y luego se evalúa el poder de predicción de éstos en la base de *test*.

### 6.6.1. BASES TRAIN Y TEST DE MODELO DE PROPENSIÓN

En este trabajo se han definido 7 períodos o ventanas de tiempo y por lo tanto 6 transiciones. Dado que el período 1 no cuenta con información anterior para formar la variable F, se decide eliminar este período y así también la transición desde el período 1 al 2 en la base que se ocupará de *train*. En estos modelos se busca predecir lo que ocurre en las transiciones (si un cliente aumenta o no su consumo) y por lo tanto las bases de *train* y *test* contendrán información de éstas:

Base de *Train*:

- Transición del período 3 al 4
- Transición del período 4 al 5
- Transición del período 5 al 6

Base de *Test*:

- Transición del período 6 al 7

No se considera la información de la transición del período 2 al 3 en la base de *train* dado que es una transición muy antigua con respecto a la que se quiere comprobar.

Debido a que una persona puede evolucionar de un segmento a otro en más de una oportunidad se considerará solo la transición más reciente por cada cliente.

## **6.7. RESULTADOS DE LOS MODELOS DE PROPENSIÓN**

Para entrenar cada uno de los modelos de propensión desarrollados se balanceó las bases de *train* en cuanto a la proporción de unos y ceros de la variable AUMENTA<sup>30</sup>. Luego de haber entrenado los modelos, se procede a aplicarlos en la base de *test* y se evalúa los resultados. En cada modelo se probó distintas configuraciones de variables y transformaciones hasta llegar a aquellas que tienen un sentido desde un enfoque del negocio.

Los tipos algoritmos utilizados en cada uno de los modelos son 3: regresión logística, red neuronal y árboles de decisión, en específico en estos últimos los utilizados son CRT, Chaid, C5.0<sup>31</sup>

En la siguiente sección se muestran los resultados de los modelos de los distintos segmentos.

### **6.7.1. RESULTADOS EN SEGMENTO INACTIVOS**

Este modelo en particular es el más importante de los realizados, ya que el problema de negocio principal que busca resolver este trabajo es el de activar a aquellos clientes que durante algún tiempo no han tenido transacciones.

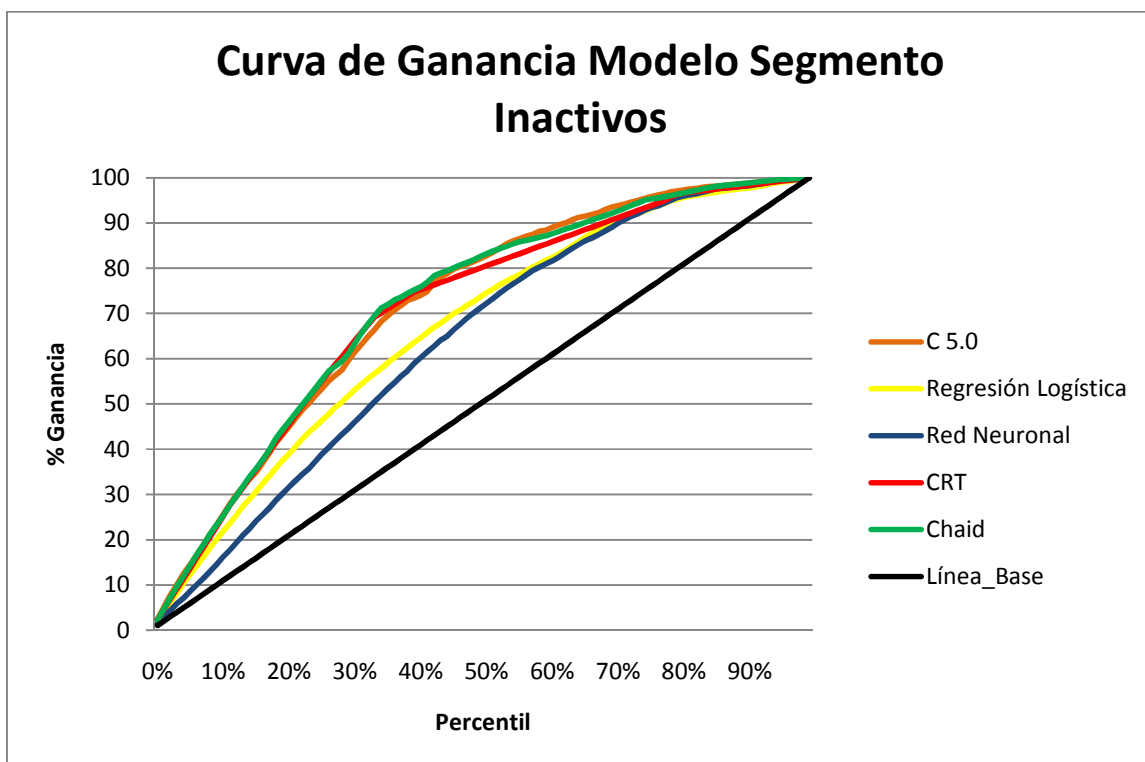
En la siguiente figura se muestra la curva de ganancia de los modelos:

---

<sup>30</sup> En base a criterio de expertos y bibliografía de minería de datos [12]

<sup>31</sup> Versión comercial para SPSS del árbol C4.5

Figura 31: Curvas de Ganancia de Información para el Modelo del Segmento Inactivos



Fuente: Elaboración propia

Se puede notar que para este segmento los árboles de decisión tienen un desempeño mejor en comparación a los otros dos modelos. A continuación se muestran los índices de los distintos de comparación:

Tabla 35: Comparación de los indicadores de los distintos modelos Segmento Inactivo

Modelo	Sensibilidad	Precisión	Eficiencia
Regresión Logística	68,5%	21,2%	60,5%
Red Neuronal	77,7%	19,0%	51,3%
Chaid	76,6%	27,5%	69,0%
C 5.0	79,9%	23,9%	62,1%
CRT	74,7%	25,8%	66,9%

El nivel de sensibilidad para la mayoría de los modelos ronda al 75%, sin embargo en cuanto a la precisión y eficiencia de los mismos, se observa que el modelo que ocupa Chaid tiene un desempeño que lo distingue por sobre los demás. También se debe destacar que el árbol de decisión que se genera tiene un sentido desde el punto de

vista del negocio con las reglas de decisión que se crean, por lo que este es el modelo elegido para este segmento<sup>32</sup>.

Al analizar el árbol Chaid resultante del modelo, las conclusiones más importantes con respecto a las variables que determinan a los clientes que aumentan su consumo son las siguientes:

- Aparece como una variable importante el cupo utilizado de la tarjeta en el período anterior a la inactividad, reflejando si una persona lleva más de un período como inactivo o no. Esta variable discrimina y hace la distinción en tres grupos, pero siempre bajo la regla que a mayor cupo utilizado antes de la inactividad mayor es la probabilidad de volver a activarse.
  - En el caso de los clientes que sí utilizaban bastante la tarjeta antes de la inactividad, el porcentaje de activación es sobre el 70% y esta cifra aumenta si el cliente no está sobre-endeudado, si ha pagado la tarjeta (el cupo utilizado disminuye de un período a otro) y si se le han hecho campañas en los últimos 2 períodos. Así también si una persona realiza giros con otros medios de pago (cuenta corriente por ejemplo) su probabilidad de activación aumenta.
  - En el segmento que el cupo utilizado del período anterior de la inactividad es intermedio, la probabilidad de activarse aumenta si la persona presenta transferencias y giros en otros medios de pago, si no está sobre-endeudada en el banco y si tiene muchos productos contratados (esto habla de la vinculación del cliente)
  - Finalmente está el grupo donde el cupo de la tarjeta en el período anterior es cero, es decir, que la persona lleva al menos dos períodos sin utilizar la tarjeta. En este grupo, la probabilidad de que un cliente se active es muy baja (35%), sin embargo hay una variable que dentro de este grupo discrimina muy bien a los que podrían activarse: el promedio de la deuda de consumo del banco en el período.
    - En este caso la variable señala que si una persona no tiene este tipo de deuda en el banco, simplemente no se activará (habla de que son personas totalmente desvinculadas del banco).
    - Cuando esta deuda sea pequeña solamente hay dos casos en que la persona puede activarse: cuando la tarjeta es cuota fija o bien, si es que es tarjeta de crédito normal pero adicionalmente debe tener depósitos y transferencias con otros medios de pagos (cuenta corriente).
    - Cuando la deuda comercial del banco es menor o igual a \$470.000 y la tarjeta de crédito es normal, se activa solamente si el cliente presenta pagos y transferencias con otros medios de pago. En el

---

<sup>32</sup> Consultar ANEXO G

caso que la tarjeta es de cuota fija, la probabilidad es alta y aumenta si es que la persona tiene alta deuda en el banco tanto en el período como en el anterior (alta vinculación del cliente).

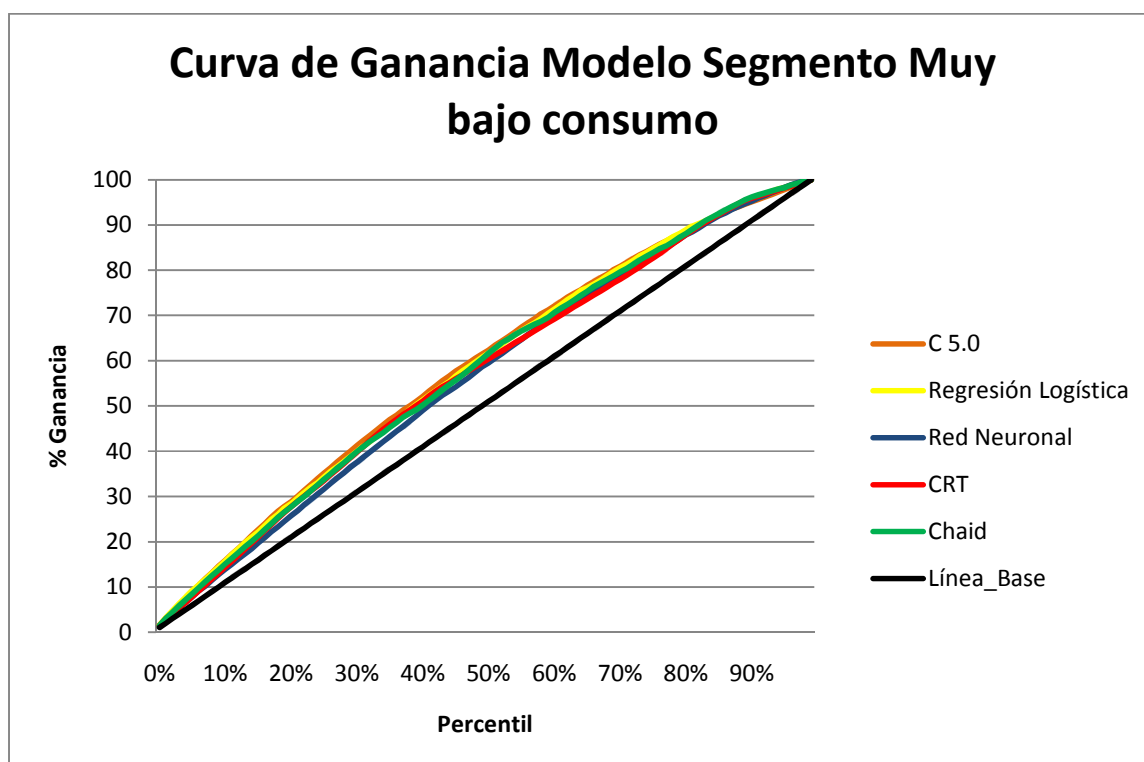
- El último grupo es el que tiene alta deuda de consumo en el banco (más de \$470.000), este grupo por su grado de endeudamiento tiene una menor probabilidad de activarse, sin embargo los que lo hacen son personas que tienen alta cantidad de pagos en otros medios, y que tienen hasta 3 años de antigüedad. En el grupo de gente más antigua la variable importante es si tienen alta cantidad de giros en otros rubros o bien contrataron un seguro de salud. En el caso de las personas que tienen pocos pagos, la única forma de activarse es que un producto de consumo haya sido contratado en el período y que tenga más de 2 medios de pago.

#### **6.7.2. RESULTADOS EN SEGMENTO MUY BAJO CONSUMO**

Este segmento al igual que el de inactivos también es importante, ya que en la sección de cadenas de Markov se pudo apreciar que casi un 50% de los clientes de este *cluster* de un período a otro mantiene o baja su consumo. Por lo mismo este es otro de los puntos importantes ya que representa a uno de los segmentos de bajo consumo que se desea incentivar con respecto al consumo con tarjeta de crédito.

En cuanto a los modelos aplicados, a continuación se puede ver el gráfico de ganancias:

Figura 32: Curva de Ganancia de Información del Modelo para el Segmento de Muy Bajo Consumo



Fuente: Elaboración propia

En la figura anterior se puede notar que los distintos modelos generan respuestas bastante parecidas y no hay una marcada supremacía de alguno de ellos.

A continuación se muestra la tabla de índices de estos modelos para elegir el definitivo:

Tabla 36: Comparación de los indicadores de distintos modelos

Modelo	Sensibilidad	Precisión	Eficiencia
Regresión Logística	56,0%	62,9%	60,6%
Red Neuronal	78,2%	56,9%	58,5%
Chaid	63,9%	61,7%	61,3%
C 5.0	76,2%	59,4%	61,2%
CRT	51,5%	63,5%	60,1%

De la tabla anterior se puede destacar que los desempeños son bastante parecidos en los distintos índices, sin embargo, se aprecia que el modelo que ocupa Chaid tiene un desempeño similar en cada uno de ellos. Nuevamente se privilegia este modelo por las reglas de decisión que toma dicho árbol ya que tienen sentido desde el punto de vista

del negocio. En base a lo anterior, se descartó por ejemplo el modelo con C5.0 ya que el árbol resultante no tiene el sentido esperado en cuanto a las decisiones que toma.<sup>33</sup>

Al analizar el árbol Chaid resultante del modelo, las conclusiones más importantes con respecto a las variables que discriminan a los clientes que aumentan su consumo son las siguientes:

- Cuando el porcentaje de avances en varias cuotas comienza a ser la parte más importante de las transacciones de una persona, lo más probable es que ese cliente no aumente su consumo al período siguiente.
- Cuando este aumento de porcentaje es moderado, aumentará el consumo de una persona solo si tiene contratado un crédito de consumo o bien es un cliente nuevo.
- Cuando no aumenta este porcentaje de avances en varias cuotas la variable más importante es el número de transacciones del período anterior: mientras más transacciones haya tenido, más probable es que aumente su consumo. Así, cuando el número de transacciones es 2 o más, entonces hay alta probabilidad de que aumente su consumo. Para el caso de los clientes que en el período anterior tienen pocas transacciones, las variables que se hacen relevantes son:
  - El número de rubros donde compró en el período: a mayor número, mayor probabilidad.
  - Numero de medios de pago: a mayor número, mayor probabilidad.
  - Número de productos del cliente: a mayor número, mayor probabilidad. (clientes más vinculados)
  - Edad: a menor edad, mayor probabilidad, así también si es cliente nuevo.
  - Rubros varios: mientras mayor sea la diversidad de rubros en los cuales compra el cliente, mayor es la probabilidad.
  - Si tiene cupo disponible en medios de pago, aumenta la probabilidad (cliente no sobre-endeudado)

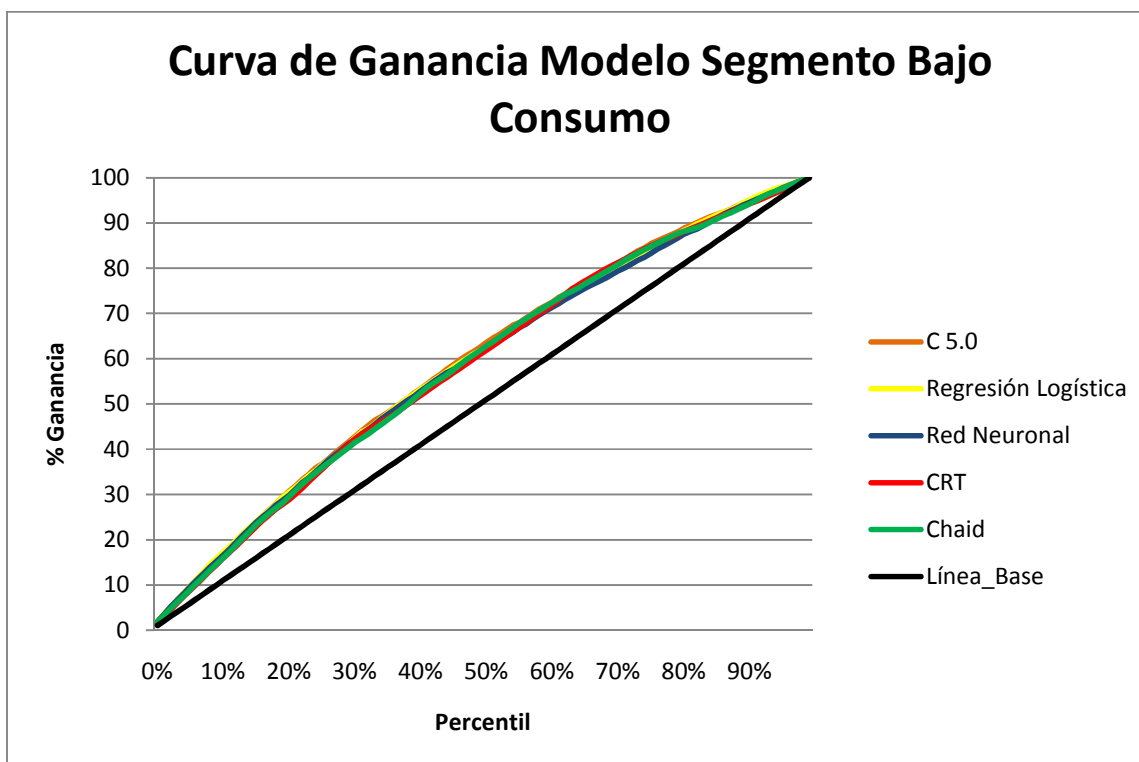
### **6.7.3. RESULTADOS EN SEGMENTO BAJO CONSUMO**

A continuación se muestra la curva de ganancia de los distintos modelos que fueron probados para el *cluster* de bajo consumo:

---

<sup>33</sup> Consultar ANEXO G

Figura 33: Curvas de ganancia de información para el Modelo del Segmento de Bajo Consumo



Fuente: Elaboración propia

Al igual que en el modelo del segmento anterior, se observa que gráficamente no existen mayores diferencias en cuanto al desempeño de los modelos. A continuación se presenta la tabla de resumen con los principales indicadores:

Tabla 37: Comparación de indicadores de distintos modelos segmento Bajo consumo

Modelo	Sensibilidad	Precisión	Eficiencia
Regresión Logística	57,7%	58,6%	61,6%
Red Neuronal	62,2%	56,5%	60,4%
Chaid	51,8%	59,1%	61,1%
C 5.0	57,1%	59,4%	62,1%
CRT	72,7%	55,4%	60,3%

En cuanto los índices de eficiencia y precisión existe poca varianza entre el desempeño de los distintos modelos. Sin embargo, se observan diferencias en el índice de sensibilidad, donde el modelo que ocupa CRT tiene un desempeño superior al resto. Este modelo fue revisado y se observa que las reglas de decisión creadas tienen



sentido desde el punto de vista del negocio, por lo tanto se escoge el modelo CRT para el segmento de bajo consumo.<sup>34</sup>

Al analizar el árbol CRT resultante del modelo, las conclusiones más importantes con respecto a las variables que discriminan a los clientes que aumentan su consumo son las siguientes:

- La principal variable que define un aumento de la probabilidad es el número de rubros donde compra la persona: a mayor número de rubros, mayor probabilidad de aumentar el consumo. Además de esta variable, se hacen relevantes las siguientes:
  - Número de transacciones en el período anterior: a mayor número, mayor probabilidad.
  - Aumento en el porcentaje de avances a varias cuotas: a mayor aumento porcentual, menor es la probabilidad.
  - Porcentaje de compras en supermercados: a mayor porcentaje, mayor probabilidad.
  - Número de productos contratados en el período (tenencia): a mayor número, mayor probabilidad.
  - Cupo utilizado en tarjeta de crédito: a menor cupo utilizado, mayor probabilidad de aumentar el consumo (clientes no sobre-endeudados)
  - Edad: a menor edad mayor es la probabilidad, así también si es cliente nuevo.

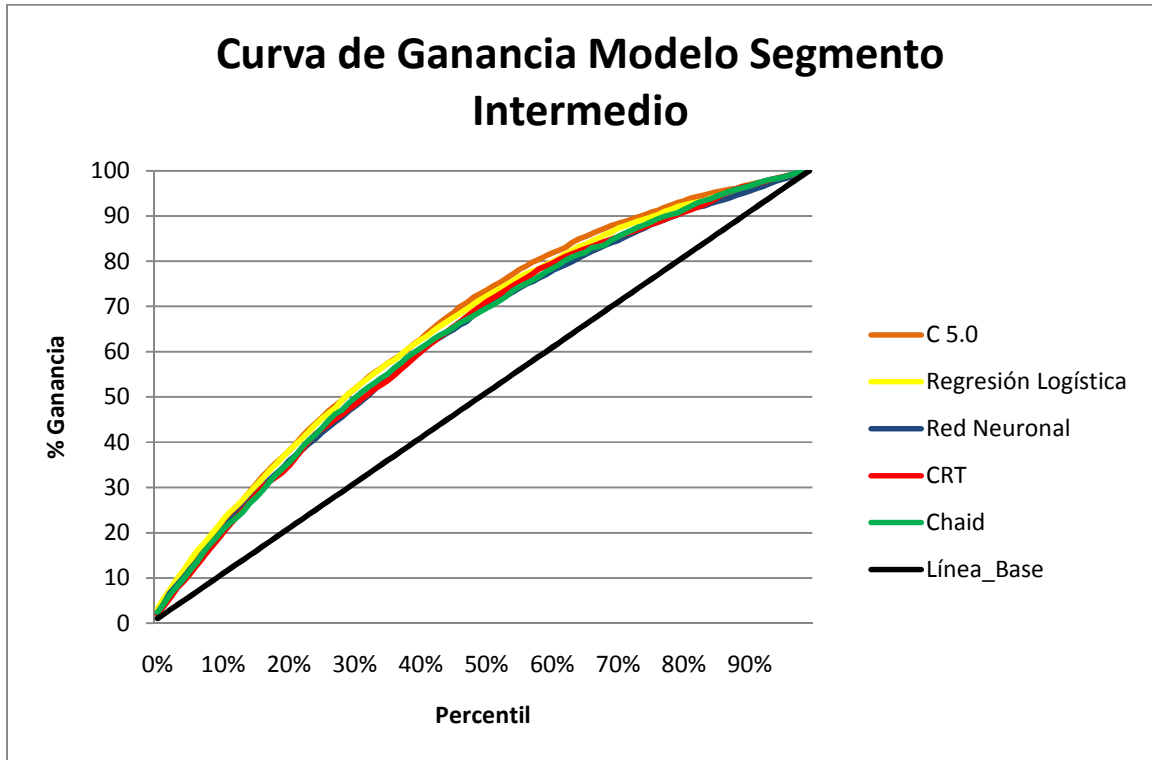
#### **6.7.4. RESULTADOS EN SEGMENTO INTERMEDIO**

Este modelo tiene un consumo promedio mayor a todos los vistos anteriormente y además es el que considera a un mayor número de clientes dentro de los modelos de segmentos activos. A continuación se muestra la curva de ganancia para cada uno de los modelos:

---

<sup>34</sup> Consultar ANEXO G

Figura 34: Curvas de Ganancia de Información para el Modelo del Segmento Intermedio



Fuente: Elaboración propia

Este modelo en comparación con los anteriores de clientes activos tiene una curva de ganancia que es mejor (mayor área bajo la curva), sin embargo los modelos no se diferencian mayormente.

Tabla 38: Comparación de los indicadores de distintos modelos Segmento Intermedio

Modelo	Sensibilidad	Precisión	Eficiencia
Regresión Logística	61,1%	25,8%	64,3%
Red Neuronal	62,9%	24,0%	60,6%
Chaid	63,3%	24,2%	60,9%
C 5.0	64,0%	25,5%	62,9%
CRT	61,0%	24,3%	61,8%

En cuanto a la sensibilidad de cada modelo se destaca el desempeño del árbol C 5.0, el cual también en los índices de precisión y eficiencia presenta buenos resultados.

Dado que el desempeño de este modelo no es claramente superior, se analiza las reglas de clasificación de los modelos y se concluye el árbol C 5.0 es el que tiene más sentido desde la perspectiva del negocio<sup>35</sup>.

Al analizar el árbol C 5.0 resultante del modelo, las conclusiones más importantes con respecto a las variables que discriminan a los clientes que aumentan su consumo son las siguientes:

- La principal variable nuevamente es el número de rubros distintos donde ha comprado la persona en el período: a mayor número, mayor probabilidad
- Cupo de la tarjeta utilizado: a menor cupo utilizado mayor es la probabilidad (clientes no sobre-endeudados)
- Cliente nuevo: si el cliente es nuevo aumenta la probabilidad.
- Contratación de PAT en el período actual y periodo anterior: si contrata un PAT (Pago Automático de Cuentas), aumenta la probabilidad
- Aumento de porcentaje de Avance en varias cuotas: A mayor porcentaje de aumento, menor la probabilidad.
- Deuda en SBIF: a menor deuda, mayor probabilidad (clientes no sobre-endeudados)

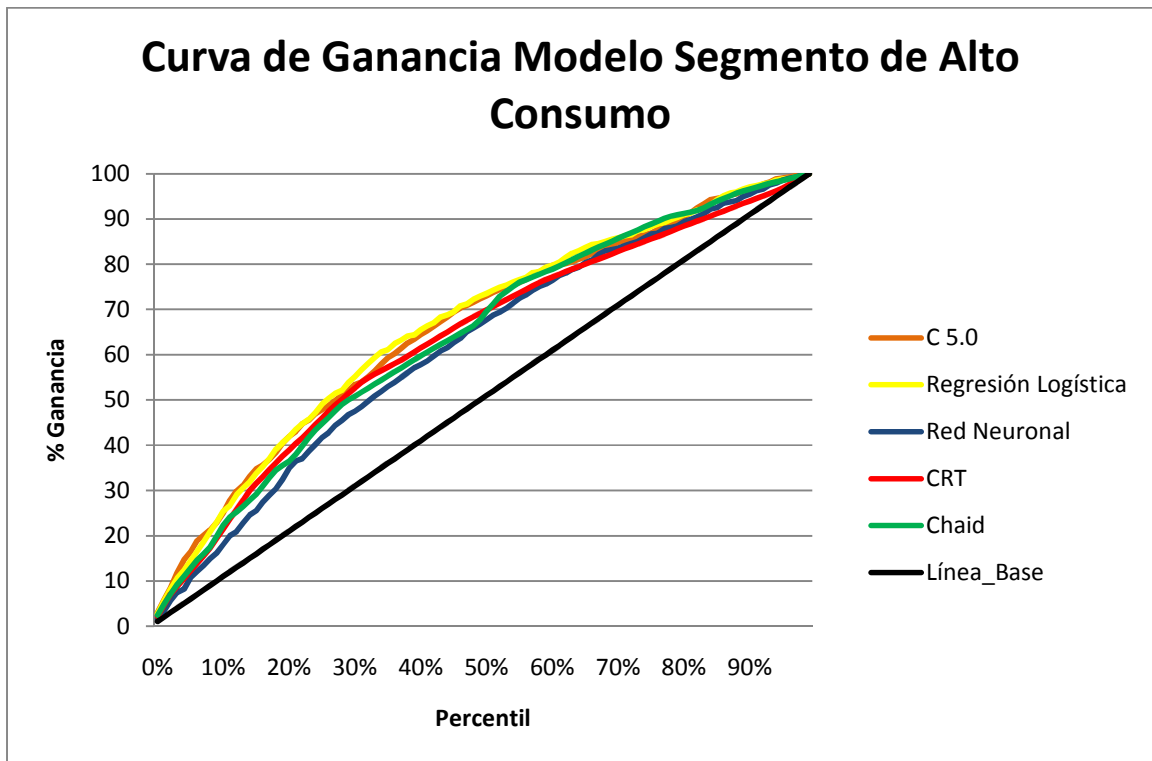
#### **6.7.5. RESULTADOS EN SEGMENTO ALTO CONSUMO**

El modelo de este segmento es distinto a los demás, puesto que se realiza con uno de los *clusters* de mayor consumo de toda la muestra, por lo mismo los clientes de este segmento solo pueden aumentar su consumo volviéndose *Heavy Users*, siendo este paso difícil dado que las características de estos últimos se diferencian considerablemente del resto de clientes. A continuación se muestra la curva de ganancia de los modelos:

---

<sup>35</sup> Consultar ANEXO G

Figura 35: Curvas de Ganancia de Información para el Modelo del Segmento de Alto Consumo



Fuente: Elaboración propia

Se puede notar que la curva de ganancia de estos modelos es similar entre ellos, por lo mismo es necesario observar los índices de cada uno para decidir cuál es el elegido:

Tabla 39: Comparación de los indicadores de distintos modelos Segmento Alto Consumo

Modelo	Sensibilidad	Precisión	Eficiencia
Regresión Logística	61,2%	10,8%	65,9%
Red Neuronal	68,4%	8,3%	50,6%
Chaid	59,9%	9,2%	60,1%
C 5.0	67,1%	9,6%	58,3%
CRT	58,6%	9,8%	63,4%

En este caso se aprecia que en sensibilidad los dos modelos que destacan son la red neuronal y C 5.0, sin embargo al ver los índices de precisión y eficiencia, es este último

modelo el que hace la diferencia. Además de lo anterior, se debe señalar que las reglas que ocupa este árbol son acorde a la lógica esperada y por esto es el modelo elegido<sup>36</sup>.

Como se señaló anteriormente este modelo presenta como característica el contar con un solo grupo al cual puede evolucionar, ya que la única opción es elevar el consumo a *Heavy Users*. El índice de precisión da cuenta de lo anterior, ya que las bases de *train* y *test* cuentan con pocos casos de transición entre estos segmentos (Alto consumo a *Heavy Users*).

Al analizar el árbol C 5.0 resultante del modelo, las conclusiones más importantes con respecto a las variables que discriminan a los clientes que aumentan su consumo son las siguientes:

- Nuevamente la variable de número de distintos rubros donde el cliente compra se hace relevante: a mayor número de rubros, mayor probabilidad.
- Si la tarjeta es de cuota fija la probabilidad de aumentar consumo es muy baja (es extraño que un cliente de alto consumo tenga este tipo de tarjeta)
- Deuda consumo en el banco: a menor deuda, mayor probabilidad (clientes no sobre-endeudado)
- Cliente nuevo: si el cliente es nuevo, la probabilidad aumenta
- Si es dueño de cuenta corriente, la probabilidad sube
- Si tiene compras en los rubros de supermercados o en restaurantes, la probabilidad sube.

## **6.8. DISCUSIÓN DE LOS RESULTADOS DE LOS MODELOS**

### **6.8.1. DESEMPEÑO DE LOS MODELOS**

Con respecto al desempeño de los modelos, se puede señalar lo siguiente:

El modelo del segmento de clientes inactivos es uno de los que presenta los mejores resultados, y por esto se puede confiar en el grado de predicción del mismo, permitiendo identificar a los clientes que podrían aumentar su nivel de consumo y de este modo no se derrocharían esfuerzos incentivando a clientes que nunca se activarán. Con respecto a lo anterior, mediante el análisis realizado con las cadenas de Markov se observa que aproximadamente un 80% de ellos no responderá a estímulos de parte del banco por lo que los esfuerzos deben ser realizados en el 20% restante que ahora se puede identificar mediante el modelo.

---

<sup>36</sup> Consultar ANEXO G

Los modelos de los segmentos de muy bajo consumo y bajo consumo no presentaron resultados destacables, ya que si bien los modelos predicen mejor que el azar, la ganancia que éstos generan es menor a la deseada. Sin embargo, las reglas de decisión que con estos modelos se generan son valiosas para identificar claves que promueven el aumento de consumo en los clientes.

Los modelos de los segmentos de consumo Intermedio y Alto son mejores que los recién señalados, pero inferiores al modelo del segmento inactivo. Estos modelos tienen una gran diferencia entre sí y dado que la cantidad de clientes que aumenta su nivel de consumo en el segmento intermedio de un período a otro es alrededor de un 16,66%, mientras que en el *cluster* de alto consumo esta cifra es del 6.3%. Por lo anterior se señala que si se tuviera que elegir entre cuál de estos 2 segmentos elegir para invertir esfuerzos, la elección debería ser hecha por el *cluster* de consumo intermedio, ya que presenta una mayor movilidad de los clientes que lo componen, lo cual se observa en la sección de las cadenas de Markov.

Entre los factores que pueden haber influido en el desempeño de los modelos se pueden nombrar:

- Tamaño de las ventanas de tiempo elegidas: dado que en el análisis de 3 períodos se observa que no hay gran variación con respecto al análisis de un período, se señala que un tamaño mayor en las ventanas de tiempo no debiese influir en los resultados, sin embargo esta decisión no se puede desestimar como una de las posibles fuentes de error.
- Segmentación escogida: La segmentación se realizó íntegramente con las variables transaccionales RFMN, quizás la incorporación de otras variables podría haber ayudado a mejorar esta segmentación y con ello los resultados podrían haber sido mejores.
- La decisión de dejar solamente la última información de una transición de un cliente como la que se usó para entrenar los modelos quizás pueda haber influido en la calidad de los mismos.

Otra definición de los ceros y unos de la variable dependiente podría haber llevado a resultados distintos, sin embargo, al igual que en los demás casos queda en el ámbito de la especulación.

## **6.8.2. COMPARACIÓN DE LAS REGLAS DE DECISIÓN DE LOS MODELOS**

Al comparar las reglas de decisión generadas en cada uno de los segmentos se aprecian las siguientes diferencias:

En el segmento de inactivos, dado que no se utilizan las variables del tipo de consumo de los clientes, las reglas de decisión no tienen esta componente y por lo mismo se distingue de los demás modelos. Las variables que mejor discriminan en este segmento, a diferencia de los otros son:

- El cupo utilizado en el período anterior, esta variable se encuentra en los modelos de otros grupos, pero nunca con respecto al período anterior. En este modelo se busca saber si una persona antes de volverse inactivo tenía cupo utilizado o no, lo que lleva a inferir si la persona lleva más de dos períodos inactiva o no.
- Las campañas realizadas en los últimos dos períodos. Estas dos variables cobran gran importancia en este modelo, a diferencia de los demás, donde éstas no se encuentran en posiciones altas del árbol. Estas variables hablan de la importancia de realizar campañas a los clientes para que vuelvan a activarse.
- Transferencias, giros y pagos con otros medios. Estas variables cobran gran importancia en este grupo, dado que señalan si una persona está desvinculada del banco o si utiliza otros medios de pago que éste le ofrece. Dichas variables no son tan importantes en los demás segmentos.

En el modelo del segmento de muy bajo consumo las reglas que lo diferencian de los otros grupos son:

- El número de transacciones en el período anterior. Esta variable es relevante para este segmento y para el de bajo consumo, permitiendo conocer el grado de consumo que tenía una persona antes de entrar en este *cluster*.
- El número de productos que posee un cliente. Esta variable solamente es ocupada en este modelo y da cuenta del grado de vinculación de los clientes, mientras más productos tiene más vinculado está y por lo mismo será más probable que aumente su consumo.

El modelo del segmento de bajo consumo presenta las siguientes variables que lo distinguen con los demás:

- Número de transacciones en el período anterior. Da cuenta del nivel de consumo anterior al estado actual de un cliente.
- Número de productos contratados en el período. Esta variable sólo es utilizada en este modelo y habla del grado de vinculación que está adquiriendo un cliente en este período.

En el modelo del segmento de consumo Intermedio las variables que lo diferencian del resto son:

- Contratación de PAT (Pago Automático de Cuentas). Esta variable es muy importante, ya que cuando una persona contrata un PAT es sinónimo de que quiere comenzar a pagar cuentas con su tarjeta de crédito, lo cual habla de un cliente que comienza a utilizar la tarjeta de forma habitual.

Finalmente se señala que el segmento de alto consumo es bastante distinto a los demás, dado que hay gran cantidad de variables que son utilizadas solo en él:

- El tipo de tarjeta (cuota fija o normal). Esta variable no cobra gran importancia en los demás segmentos.
- Deuda de consumo en el banco. Si un cliente está endeudado con el banco, menor es la probabilidad que aumente su consumo.
- Compras en rubros específicos, como restaurantes. Estas variables preguntan específicamente por consumo en algún rubro, ya que en su mayoría los clientes de alto consumo ya usan la tarjeta de forma "habitual"

Con respecto a las similitudes de los segmentos, estas variables son las que más se repiten:

- Número de rubros distintos donde compra. Esta variable es relevante en prácticamente todos los modelos, habiendo una sutil diferencia y es que el valor de esta variable debe ir en aumento hacia los segmentos de mayor consumo para ser relevante.
- El aumento del porcentaje de avances en varias cuotas en las transacciones de un cliente. Esta variable también aparece en casi todos los modelos. Al aumentar este porcentaje, menor es la probabilidad de aumentar el consumo.
- Cupo utilizado de la tarjeta. Esta variable da cuenta del grado de endeudamiento de los clientes, a mayor cupo utilizado menor es la probabilidad de aumentar el consumo.
- Edad y si es cliente nuevo. Los clientes entre 20 y 39 años tienen mayor probabilidad de aumentar su consumo, al igual que los clientes nuevos.



### **6.8.3. PERFILES DE CLIENTES QUE SE OBSERVAN A PARTIR DE LAS REGLAS DE DECISIÓN.**

Modelo del segmento inactivo.

Se observan 3 tipos de clientes distintos que incrementan su consumo:

- Clientes que normalmente utilizaban su tarjeta, pero que por algún motivo dejaron de hacerlo. Estos son los clientes a los que se debe incentivar, ya que es muy probable que salgan de este estado si se le realizan campañas, más si es que reúnen condiciones como estar al día en las cuentas y no sobre endeudados.
- Clientes que antes de la inactividad usaban la tarjeta de forma moderada. Esta persona responde de menor manera, pero si es que se trata de clientes vinculados mediante una gran cantidad de productos contratados, es probable que vuelvan a transaccionar.
- Clientes que llevan 8 meses como inactivos. Estos clientes es muy difícil que vuelvan a transaccionar a excepción de los que poseen deuda en el banco y aquellos que cuentan con otros medios de pago.

Para el modelo de muy bajo consumo los clientes que se observan son los siguientes:

- Cliente que hace avances en varias cuotas de forma reiterada. Este es un tipo de cliente que solo utiliza la tarjeta de crédito para pedir dinero prestado y por lo mismo es que es improbable que aumente su consumo.
- Cliente que hace avances en varias cuotas de forma moderada. Este cliente probablemente aumentará su consumo si es que tiene un crédito de consumo o es nuevo en el banco.
- Cliente que no utiliza la tarjeta para pedir avances en varias cuotas. Dependiendo del número de transacciones del período anterior, este cliente probablemente aumente su consumo, esta probabilidad aumenta si realiza compras en varios rubros y tiene un gran número de productos (cliente vinculado). Si este cliente no está sobre-endeudado, también es probable que aumente su consumo.

En el segmento de bajo consumo se distinguen dos tipos de clientes:

- Cliente que compra en varios rubros. Este cliente fácilmente aumentará su consumo, especialmente si en el período anterior presentó gran cantidad de transacciones.
- Cliente que utiliza la tarjeta para hacer avances en varias cuotas. Este cliente realiza pocas compras y muchos avances en varias cuotas, aumentará su consumo siempre que no se encuentre sobre-endeudado

El segmento Intermedio presenta los siguientes perfiles de clientes:

- Cliente que compra en varios rubros. Este cliente probablemente aumentará su consumo. Esta probabilidad se incrementa si es que no se encuentra sobreendeudado.
- Cliente que utiliza la tarjeta para hacer avances en varias cuotas. Este cliente realiza pocas compras y muchos avances en varias cuotas. Este cliente no es probable que aumente su consumo.
- Cliente antiguo que comienza a utilizar la tarjeta para pagar cuentas mediante PAT. Este cliente probablemente aumentará su consumo en el período siguiente y la probabilidad aumenta si no se encuentra endeudado.

Finalmente el segmento de clientes de alto consumo presenta los siguientes perfiles de clientes:

- Cliente que transacciona en muchos rubros distintos. Este tipo de cliente aumenta su consumo mientras en más rubros realice compras de forma habitual. En particular, si la persona no tiene grandes deudas, esta probabilidad aumenta.

## **7. CONCLUSIONES Y RECOMENDACIONES**

Una de las recomendaciones que se puede hacer a área donde se realizó este trabajo es la de mejorar la confiabilidad de las bases de datos con que deben realizar los estudios. Dado que se observó que para un mismo cliente puede haber discrepancia de uno de sus atributos si es que obtiene dicha información desde dos orígenes distintos. Además de lo anterior, sería aconsejable aumentar los recursos de los servidores existentes o bien reemplazarlos por máquinas más actuales, dado que el espacio disponible para la creación de tablas se puede completar rápidamente si no se realiza el mantenimiento adecuado de las bases.

Las conclusiones de este trabajo se hacen tanto a nivel general así como con respecto a los resultados obtenidos en las distintas secciones del presente informe.

### **7.1. CONCLUSIONES CON RESPECTO A LOS OBJETIVOS PLANTEADOS**

- El objetivo principal del presente trabajo se cumplió, ya que se diseñaron modelos de propensión para los distintos segmentos que mediante sus reglas de decisión permiten hacer recomendaciones para incentivar el uso de la tarjeta de

crédito, siendo en particular el modelo del segmento de clientes inactivos el que presenta un mejor desempeño en base a los indicadores utilizados.

- Se logró caracterizar la dinámica de uso de tarjeta de crédito a través del tiempo mediante el uso de las cadenas de Markov (análisis de 3 períodos).
- Se perfiló a los clientes con respecto a la utilización que le dan a la tarjeta de crédito mediante la agrupación de sus transacciones de tres formas distintas observando las diferencias de uso de la tarjeta entre los distintos segmentos.
- Se crearon cinco modelos de propensión que permiten, mediante la probabilidad que entregan como salida, identificar a los clientes a los cuales se debe incentivar en el uso de la tarjeta de crédito.

## **7.2. CONCLUSIONES DE LAS TÉCNICAS UTILIZADAS**

Con respecto a las metodologías aplicadas y etapas del modelo KDD se puede concluir lo siguiente:

- La base de datos con la que se hace un modelamiento es determinante para los resultados de éste, así se hace necesario realizar una buena elección de los datos y variables a utilizar. En una etapa posterior, pero decisiva, el preprocesamiento y transformación de los datos pueden ser concluyentes con respecto a los resultados que se obtienen del trabajo.

La segmentación utilizando las variables RFM no siempre es capaz de separar de manera correcta a los distintos grupos de clientes, ya que la variable F solo considera el promedio de días entre transacciones y en algunos casos no separa entre los clientes con alto número y bajo número de éstas. Por lo tanto es recomendable mejorar esta segmentación incorporando una nueva variable, N que corresponde al número de transacciones en el período, ya que en conjunto con las otras variables permite distinguir entre clientes frecuentes y clientes esporádicos dentro de un período de tiempo. Para aplicar esta metodología es importante verificar que la correlación entre las variables N y F sea baja (en este caso la correlación es -0.23).

- Se comprobó que el modelamiento utilizando cadenas de Markov permite observar la dinámica de comportamiento transaccional de los clientes mediante el análisis de 3 períodos, identificando las probabilidades de transición entre los distintos segmentos y proporcionando información relevante sobre los segmentos con que se trabaja. Un ejemplo de ello ocurre al observar que los clientes desde

los niveles de bajo consumo no evolucionan a segmentos mayores de forma explosiva, sino que lo hacen lentamente pasando primero por los grupos de consumo intermedio.

- Se comprobó empíricamente que la propiedad de pérdida de memoria de una cadena de Markov es aplicable a este caso, dado que la información de la última transición es la que mejor predice el comportamiento futuro de los segmentos.
- La definición del largo de los períodos a analizar es de gran importancia con respecto a la cantidad de datos que se puedan obtener y esto a su vez, se ve reflejado en los resultados obtenidos. La decisión en cada caso debe ser hecha considerando las características del problema que se desea abordar.
- La segmentación que se usa de base para la cadena de Markov definitiva, no solo debe ser escogida según el criterio de minimización de error en la predicción de las probabilidades de transición, sino también es necesario considerar la aplicación de criterios de negocio para que segmentación sea útil y considere, por ejemplo, un número razonable de segmentos de mercado.
- La elección de las variables utilizadas en los modelos es de vital importancia para el desempeño de los mismos, ya que muchas veces un modelo puede tomar variables que estén correlacionadas con la dependiente, pero que en definitiva no tienen ninguna relación con la misma si se mira desde el contexto del negocio, por esto muchas veces es preferible eliminar variables que no tengan sentido directo para que no ensucien los resultados obtenidos al no poder interpretarlas.
- Cuando los resultados de un modelo deben ser interpretados para ver si son aplicables a la realidad, los modelos de árboles de decisión y regresión logística son buenas herramientas, ya que cada uno de ellos permite ver las reglas de decisión aplicadas y con esto juzgar si éstas van acorde al contexto del negocio.

### 7.3. CONCLUSIONES DE LOS RESULTADOS DE LOS MODELOS

Con respecto a los resultados obtenidos de la aplicación de los modelos se puede concluir:

- Los clientes que en un principio están en el segmento de *Heavy Users* terminan después de tres períodos en este mismo segmento o en el de Alto consumo (47,70% y 35,23% respectivamente), por lo que se comprueba que la tarjeta de crédito es un producto vinculante: una vez que una persona la utiliza regularmente, es poco probable que deje de hacerlo.
- Las cadenas de Markov muestran que de los clientes del segmento inactivos, casi un 80% difícilmente se vuelve a activar luego de pasados 8 o 12 meses de inactividad por lo que prácticamente podrían ser considerados como clientes fugados de este producto y por lo tanto los recursos y esfuerzos deben ser orientados hacia el 20% restante.
- En el perfilamiento realizado a los segmentos de clientes activos con respecto a sus hábitos de consumo, se concluye que los segmentos de menor consumo ocupan la tarjeta de crédito principalmente para realizar avances en varias cuotas, mientras que los de mayor consumo lo realizan en una cuota. Se puede destacar que los segmentos de menor consumo ocupan la tarjeta de forma ocasional, para rubros de montos mayores: Muebles, Grandes tiendas entre otros, mientras que en los segmentos de mayor consumo la tarjeta de crédito comienza a ocuparse con mayor frecuencia, usándola para rubros de compra habitual como Supermercados, Farmacias y pago de cuentas (Recaudación).
- Con respecto a los resultados del modelo de clientes inactivos, se puede concluir que dentro de los clientes que ya llevan dos períodos inactivos, los únicos en los que vale la pena poner algún esfuerzo para activarlos son aquellos que tienen alguna deuda de consumo con BancoEstado. En el caso de los clientes sin deuda en el banco, no vale la pena invertir dinero y esfuerzo en activarlos, ya que son clientes desvinculados que difícilmente se activarán.

Existe una concordancia entre los resultados arrojados por las cadenas de Markov y el modelo de propensión del segmento de clientes inactivos, ya que ambos coinciden en cuanto a que un cliente con más de 8 meses de inactividad es muy difícil que aumente su consumo.

- Los resultados entregados por los modelos de propensión indican que los segmentos con modelos de mejor desempeño son el de inactivos y el de los

clientes de consumo intermedio y alto, siendo el primero de estos tres el que mejor grado de predicción tiene. Los modelos de los segmentos de muy bajo y bajo consumo entregaron resultados mejores que el azar, pero que no cumplen con los desempeños deseados.

- En el segmento de clientes inactivos la presencia de transacciones en otros medios de pago (CuentaRUT, Cuenta Corriente, Chequera Electrónica, CuentActiva) genera que la probabilidad de aumentar el consumo sea alta. Así también, este fenómeno se observa cuando los clientes tienen créditos vigentes y un alto número de productos contratados (mayor vinculación).
- En cuanto a las acciones a tomar con los resultados de cada modelo, se propone dos líneas de acción:
  - Para los modelos de segmentos activos, se propone que a los clientes que sean considerados con una alta probabilidad de aumento de consumo al siguiente período no se generen acciones activas, ya que esto implicaría ocupar recursos en clientes que, no importando las acciones del banco, de todas formas lo más probable es que aumenten su consumo. Para este grupo, se propone diseñar acciones en las cuales simplemente se haga recordatorios a los usuarios de los comercios en los cuales hay descuentos y promociones (para así también aumentar el número de rubros donde compran), sin incurrir en campañas más costosas. Por el contrario, a los clientes que tienen una probabilidad intermedia o baja de aumentar su consumo se propone que sean diseñadas campañas activas que incentiven el uso de la tarjeta, para que su consumo se mantenga alto y no decaiga. Dentro de estas campañas activas, se puede señalar que un club de puntos podría ser un incentivo para que las personas no bajen su consumo, ya que de mantenerlo podrían conseguir algún tipo de premio o beneficio. Así también campañas de compras en muchas cuotas precio contado pueden incentivar a estos clientes. Finalmente la oferta de otros productos puede aumentar el grado de vinculación de los clientes con lo que probablemente aumenten su consumo.
  - Para el segmento de clientes inactivos la propuesta en la forma de tratar a los clientes propensos es distinta: Dado que un cliente está en el segmento inactivo, significa que al menos lleva cuatro meses en ese estado, si alguno de ellos presenta una alta probabilidad de aumentar su consumo, no se puede correr el riesgo de que en el período siguiente continúe como inactivo, ya que se observó que un cliente con más de 8 meses sin uso de la tarjeta es casi imposible volver a activarlo, por esto es que se propone que a todos los clientes que tengan una probabilidad

media-alta de salir de este estado se les realicen campañas activas. Estos clientes con probabilidad media-alta corresponden al 20% de los clientes que puede volver a activarse, por lo que se pueden crear campañas de ofrecimiento de otros medios de pago ya que la utilización de éstos aumenta la vinculación de los clientes y se definió como una de las variables importantes para su aumento de consumo. Así también se puede ofrecer beneficios como el no pago de cargo de mantención de la tarjeta por un determinado tiempo, o la baja de intereses de la misma para que así la persona comience a utilizarla y se quizás se convierta en un comprador habitual con ella. A los clientes con baja deuda en el banco (menor a \$470.000) se les pueden hacer ofrecimientos de créditos de consumo, ya que al tener uno contratado crece la probabilidad de aumento de consumo.

- A medida que se observan los segmentos de mayor consumo, el número de rubros en que una persona hace transacciones, se hace más relevante para determinar si un cliente subirá su nivel de consumo. Dado esto es que se recomienda hacer campañas que incentiven el uso de la tarjeta de crédito en rubros de “compra habitual” tales como Supermercados, Farmacias, Combustibles, Otros consumos, etc. pues se observa que cuando una persona tiene algún porcentaje de sus compras en estos rubros, la probabilidad de aumentar su consumo es siempre alta, así también, se pudo ver que aproximadamente el 90% de las personas que contratan un Pago Automático de Tarjeta (PAT) al período siguiente suben de nivel, por lo que esta podría ser una buena herramienta para incentivar el consumo
- A medida que una persona aumenta el porcentaje de avances en varias cuotas que realiza dentro de sus transacciones se observa que la probabilidad de aumentar su consumo decae, así también se observa que cuando una persona tiene bastante cupo utilizado de la tarjeta de crédito es menos probable que aumente su consumo.
- Durante el análisis de los distintos modelos, se repite el patrón de que una persona joven o adulto joven (entre 20 y 39 años) es más propensa a subir su nivel de consumo que una persona adulta, por lo tanto, se debe tener en cuenta esta variable al momento de decidir a quién hacer algún tipo de campaña.

## 8. TRABAJOS FUTUROS

Luego de haber desarrollado este trabajo se proponen algunas aristas que pudiesen ser considerados para trabajos posteriores:

- Se podría realizar una nueva segmentación que incluya adicionalmente distintas variables demográficas, puesto que no fueron relevantes en los modelos, pero sí podrían generar segmentaciones más robustas.
- Probar si la metodología desarrollada en este trabajo puede ser utilizada en otras áreas del banco donde también haya una componente de transaccionalidad, por ejemplo en el rubro de microempresas o bien con otros medios de pago.
- Así también, se podría probar si esta metodología tiene los mismos resultados en toda la industria bancaria para este medio de pago, o bien si los resultados son un enfoque particular que se apega a la realidad de esta institución.
- Automatizar este trabajo en los sistemas del banco de modo que realice periódicamente, actualizando la pertenencia de los clientes a los *clusters*, validando o reformulando el número de segmentos a utilizar y con esto proveer una herramienta de soporte a la decisión de cuáles serán los clientes a ser considerados dentro de las campañas.
- Dado que los modelos desarrollados en este trabajo están entrenados para encontrar a los clientes que aumentan su consumo en el período siguiente, sería interesante definir la variable dependiente como 1 cuando un cliente baja su consumo y de este modo utilizar los resultados que de esto puedan surgir para mejorar los modelos de fuga existentes en el área.



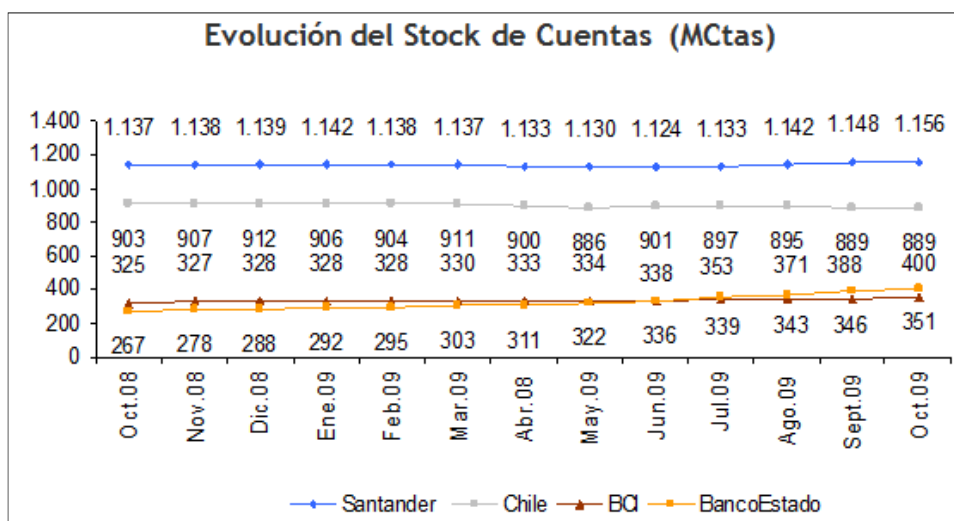
## 9. BIBLIOGRAFÍA

- [1] ABURTO, L. Resultados de Marketing Directo Utilizando Conceptos de Segmentación. RFM: Recency, Frequency, Monetary Value [en línea] <http://www.penta.cl/analytics/papers/rfm.pdf> [consulta:11 enero]
- [2] BERRY, M y LINOFF, G. 2004. Data mining techniques: for marketing, sales, and customer relationship management. 2ª edición, Wiley . 672p.
- [3] BRIONES V., D.A. 2002. Modelo de predicción de fuga de clientes para BancoEstado. Memoria Ingeniería Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.62p.
- [4] FRITIS, N; POLIT, E; RAMÍREZ, M; YOUNG, W. Modelos de Segmentación. Material de estudio, curso IN58B.
- [5] HAENLEIN, M; KAPLAN, A.; BEESER, A. 2007 .A model to determine customer lifetime value in retail banking context. European Management Journal Vol. 25, 221-234.
- [6] MARKER, J. 1998. Studying Policy Retention Rates Using Markov Chains. Casualty actuarial society v85.
- [7] PFEIFER, P.E. y CARRAWAY, R.L. 2000. Modeling Customer Relationship as Markov Chains. Journal Interactive marketing v14. no. 2:43-55.
- [8] SANZANA Z., G.A. 2006. Modelo de predicción de cierre voluntario de cuentas corrientes en una institución bancaria. Memoria Ingeniería Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.74p.
- [9] SEGOVIA R., C.A. 2005. Caracterización del proceso de fuga de clientes de un retail banking utilizando información transaccional. Memoria Ingeniería Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.83p.
- [10] TWO CROWS CORPORATION. 1999. Introduction to Data Mining and Knowledge Discovery, tercera edición, <http://www.twocrows.com>
- [11] URZÚA S., P.A. 2007. Estimación de customer Lifetime value mediante técnicas supervisadas de data mining en una empresa de retail. Memoria Ingeniería Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.111p.
- [12] WEBER, R.; L'HUILLIER, G. 2008. Apuntes de clases de cátedra y clases auxiliares curso IN643: Introducción a la Minería de Datos.
- [13] WEBER, R. ;MIRANDA J. 2008.Sistemas de Predicción de Fugas Voluntarias de Clientes. Caso de estudio Instituto Milenio. Sistemas Complejos de Ingeniería.
- [14] Clasificación: Aprendizaje Supervisado, <http://elvex.ugr.es/etexts/spanish/proyecto/cap5.pdf>

## 10. ANEXOS

### ANEXO A: Evolución del número de clientes que poseen Tarjeta de Crédito

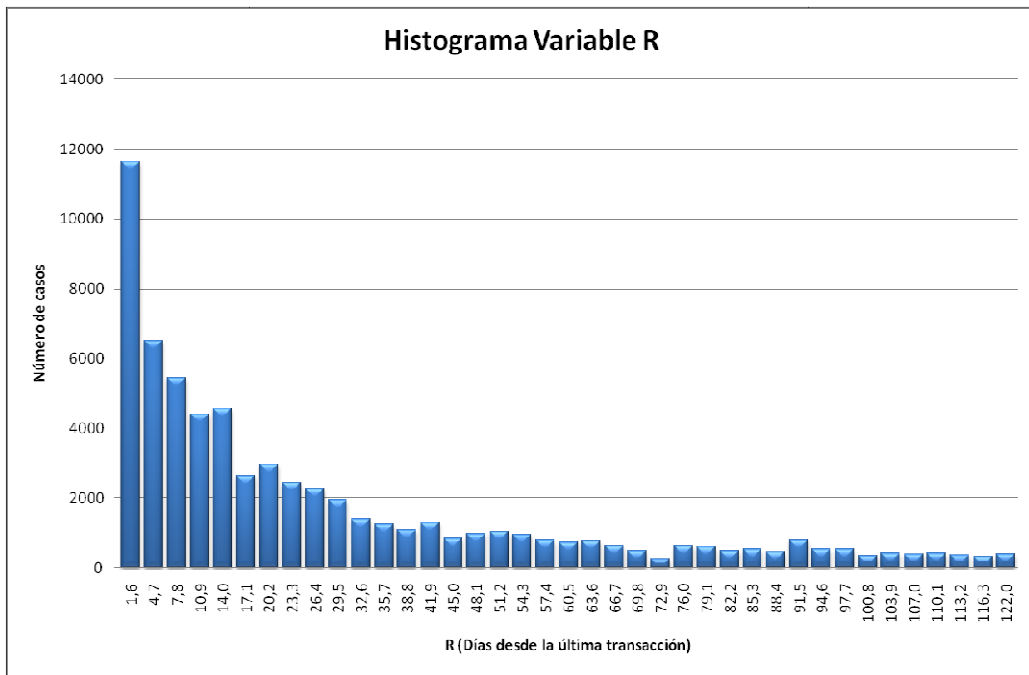
Figura 36: Evolución del número de clientes con Tarjeta de Crédito



Fuente: INFORME DE ACTIVIDAD DE TARJETAS DE CREDITO BANCOESTADO Y SISTEMA FINANCIERO A OCTUBRE 2009 DICIEMBRE 2009, BancoEstado

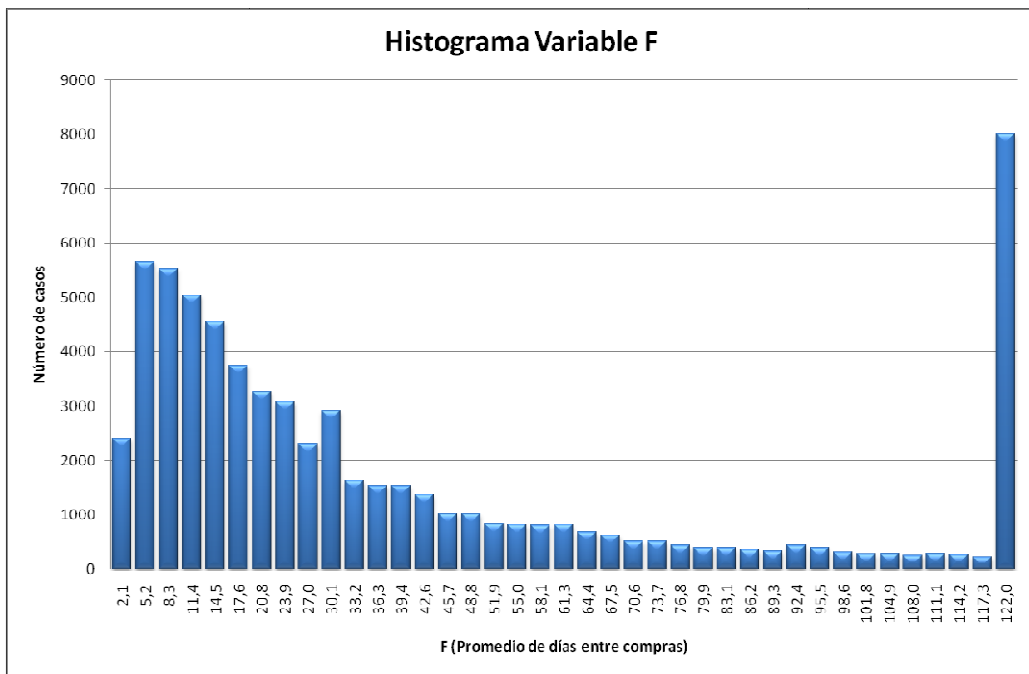
**ANEXO B: Histogramas de variables RFM posterior al tratamiento de outliers**

**Figura 37: Histograma variable Recency (excluidos outliers)**



Fuente: Elaboración propia

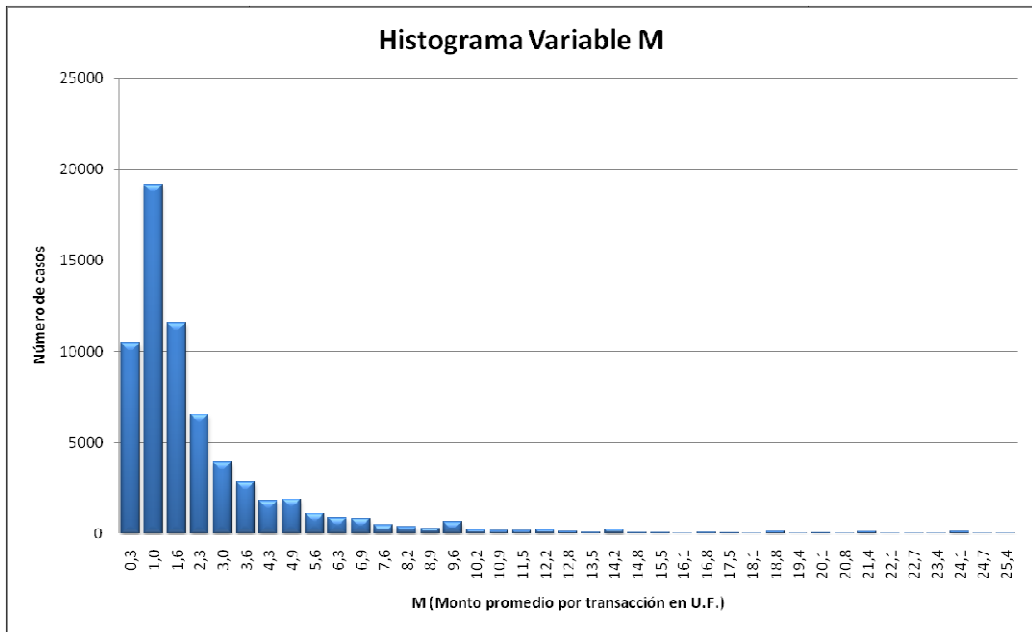
**Figura 38: Histograma variable F (excluidos outliers)**



Fuente: Elaboración propia

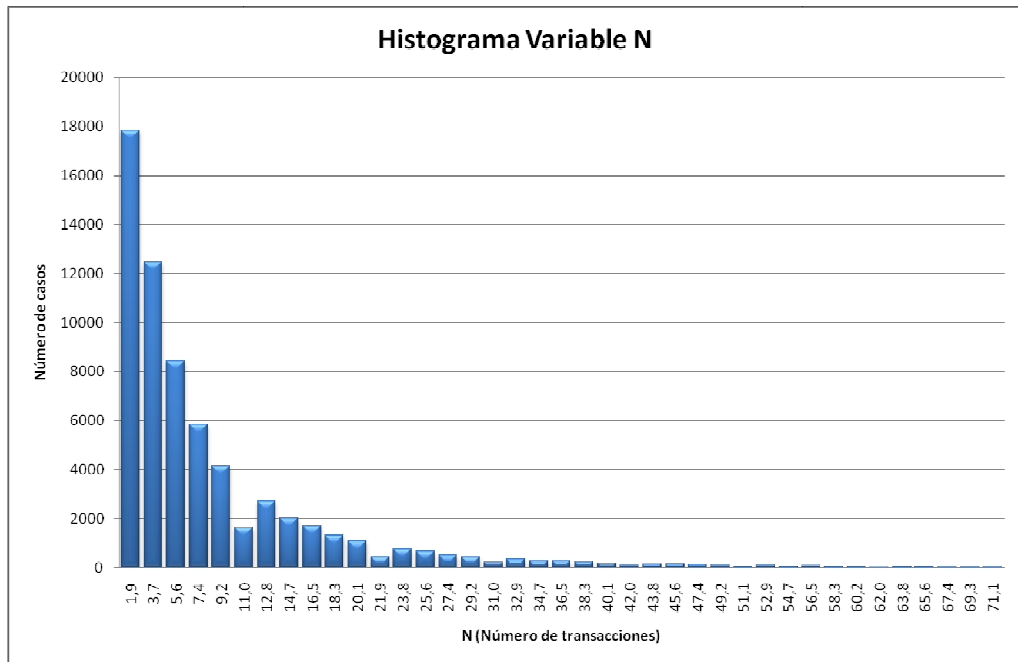
En el histograma anterior se observa una gran cantidad de casos con valor  $F=122$ , ya que se agrupó y dio este valor a todos los casos con  $F \geq 122$ .

Figura 39: Histograma variable M (excluidos outliers)



Fuente: Elaboración propia

Figura 40: Histograma variable N (excluidos outliers)



Fuente: Elaboración propia

**ANEXO C: Centros de los *clusters* obtenidos según las combinaciones de variables RFMN**

Tabla 40: Centros de *Clusters* según la variable Recency

<b>CLUSTER</b>	<b>R</b> <b>[días]</b>
<b>1</b>	27,7
<b>2</b>	98,8
<b>3</b>	6,6
<b>4</b>	59,3

Tabla 41: Centros de *clusters* según la variable Frequency

<b>CLUSTER</b>	<b>F</b> <b>[días]</b>
<b>1</b>	11,4
<b>2</b>	69,8
<b>3</b>	34,3
<b>4</b>	118,4

Tabla 42: Centros de *clusters* según la variable Monetary

<b>CLUSTER</b>	<b>M</b> <b>[U.F]</b>
<b>1</b>	19,7
<b>2</b>	4,3
<b>3</b>	1,2
<b>4</b>	10,1

Tabla 43: Centros de *clusters* según la variable N

<b>CLUSTER</b>	<b>N</b> <b>[número de transacciones en el periodo]</b>
<b>1</b>	14
<b>2</b>	53
<b>3</b>	4
<b>4</b>	30

Tabla 44: Centros de *clusters* según variables RF

<b>CLUSTER</b>	<b>R</b> <b>[días]</b>	<b>F</b> <b>[días]</b>
1	40	110,3
2	12	20,8
3	77	33,8

Tabla 45: Centros de *clusters* según variables RM

<b>CLUSTER</b>	<b>R</b> <b>[días]</b>	<b>M</b> <b>[U.F]</b>
1	28	2,601
2	99	3,897
3	7	2,019
4	59	3,095

Tabla 46: Centros de *clusters* según variables FM

<b>CLUSTER</b>	<b>F</b> <b>[días]</b>	<b>M</b> <b>[U.F]</b>
1	11,3	1,7
2	118,2	4,6
3	68,9	3,1
4	33,8	2,3

Tabla 47: Centros de *clusters* según variables FN

<b>CLUSTER</b>	<b>F</b> <b>[días]</b>	<b>N</b> <b>[número de transacciones en el periodo]</b>
1	7,5	20
2	24,4	5
3	116,6	2
4	61,0	3

Tabla 48: Centros de *clusters* según variables RN

<b>CLUSTER</b>	<b>R</b> [días]	<b>N</b> [número de transacciones en el periodo]
1	99	2
2	59	3
3	7	8
4	28	5
5	4	34

Tabla 49: Centros de *clusters* según variables NM

<b>CLUSTER</b>	<b>N</b> [número de transacciones en el periodo]	<b>M</b> [U.F.]
1	17	1,6
2	43	1,3
3	4	2,8

Tabla 50: Centros de *clusters* según variables RFM

<b>CLUSTER</b>	<b>R</b> [días]	<b>F</b> [días]	<b>M</b> [U.F.]
1	40	110,3	4,3
2	12	20,8	1,9
3	77	33,8	2,7

Tabla 51: Centros de *clusters* según las variables RNM

<b>CLUSTER</b>	<b>R</b> [días]	<b>N</b> [número de transacciones en el periodo]	<b>M</b> [U.F.]
1	96	2	3,8
2	55	4	3,0
3	7	8	2,2
4	26	5	2,6
5	4	34	1,4

Tabla 52: Centros de *clusters* según las variables RFN

<b>CLUSTER</b>	<b>R</b> [días]	<b>F</b> [días]	<b>N</b> [número de transacciones en el periodo]
1	76	33,9	3
2	12	20,7	12
3	40	110,0	2

Tabla 53: Centros de *clusters* según variables RFNM

<b>CLUSTER</b>	<b>R</b> [días]	<b>F</b> [días]	<b>N</b> [número de transacciones en el periodo]	<b>M</b> [U.F.]
1	76	33,9	3	2,7
2	12	20,7	12	1,9
3	40	110,0	2	4,3

## ANEXO D: Errores de transición comparando distintas estimaciones.

Tabla 54: Errores de transición con distintas estimaciones (variable R)

<b>Variable(s): R</b> [días]	<b>MAD</b>	<b>Error ponderado por número de transiciones</b>	<b>Error ponderado por distancia</b>
<b>Última transición</b>	<b>3,71%</b>	<b>8,51%</b>	172,63%
<b>Promedio transiciones</b>	25,63%	21,93%	618,67%
<b>Promedio 2 últ. transiciones</b>	7,09%	9,27%	<b>168,40%</b>
<b>Tendencia 2 últ. transiciones</b>	15,48%	18,65%	453,35%
<b>Transición año anterior</b>	27,67%	23,11%	623,12%



Tabla 55: Errores de transición con distintas estimaciones (variable F)

Variable(s): F [días]	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>5,15%</b>	11,86%	188,90%
Promedio transiciones	25,95%	24,86%	559,00%
Promedio 2 últimas transiciones	7,89%	<b>11,76%</b>	<b>168,62%</b>
Tendencia 2 últimas transiciones	17,84%	23,69%	446,34%
Transición año anterior	27,72%	25,14%	578,35%

Tabla 56: Errores de transición con distintas estimaciones (variable M)

Variable(s): M [U.F]	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>3,76%</b>	<b>11,10%</b>	<b>11,98%</b>
Promedio transiciones	25,99%	20,50%	20,16%
Promedio 2 últ. transiciones	7,96%	14,43%	13,64%
Tendencia 2 últ. transiciones	13,94%	17,95%	17,97%
Transición año anterior	27,89%	22,49%	21,41%

Tabla 57: Errores de transición con distintas estimaciones (variable N)

Variable(s): N [número de transacciones en el periodo]	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>5,84%</b>	22,89%	45,93%
Promedio transiciones	25,96%	23,76%	49,66%
Promedio 2 últ. transiciones	9,35%	22,27%	45,83%
Tendencia 2 últ. transiciones	14,26%	29,09%	48,52%
Transición año anterior	27,22%	<b>17,86%</b>	<b>36,42%</b>

Tabla 58: Errores de transición con distintas estimaciones (variables RF)

Variable(s): RF	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>3,66%</b>	<b>8,31%</b>	<b>165,77%</b>
Promedio transiciones	25,64%	21,83%	615,29%
Promedio 2 últ. transiciones	7,11%	9,23%	167,35%
Tendencia 2 últ. transiciones	15,41%	18,35%	442,20%
Transición año anterior	27,69%	23,17%	620,94%

Tabla 59: Errores de transición con distintas estimaciones (variables RM)

Variable(s): RM	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>4,70%</b>	<b>8,57%</b>	<b>120,62%</b>
Promedio transiciones	25,51%	22,41%	346,07%
Promedio 2 últ. transiciones	7,90%	9,06%	142,15%
Tendencia 2 últ. transiciones	14,41%	12,75%	168,83%
Transición año anterior	27,17%	22,43%	321,70%

Tabla 60: Errores de transición con distintas estimaciones (variables FM)

Variable(s): FM	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>4,46%</b>	<b>10,43%</b>	<b>160,70%</b>
Promedio transiciones	25,96%	22,59%	616,29%
Promedio 2 últimas transiciones	7,74%	10,67%	198,38%
Tendencia 2 últimas transiciones	14,56%	15,91%	285,18%
Transición año anterior	27,77%	22,15%	610,97%

Tabla 61: Errores de transición con distintas estimaciones (variables FN)

Variable(s): FN	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>6,00%</b>	13,62%	200,17%
Promedio transiciones	25,84%	22,10%	555,53%
Promedio 2 últimas transiciones	8,70%	<b>12,54%</b>	<b>190,01%</b>
Tendencia 2 últimas transiciones	17,73%	25,68%	442,56%
Transición año anterior	27,70%	24,04%	575,84%

Tabla 62: Errores de transición con distintas estimaciones (variables RN)

Variable(s): RN	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>5,91%</b>	<b>12,46%</b>	<b>171,19%</b>
Promedio transiciones	25,92%	21,14%	399,54%
Promedio 2 últimas transiciones	8,98%	13,89%	196,59%
Tendencia 2 últimas transiciones	15,22%	17,02%	222,81%
Transición año anterior	27,23%	21,61%	365,03%

Tabla 63: Errores de transición con distintas estimaciones (variables NM)

Variable(s): NM	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>5,57%</b>	17,71%	43,77%
Promedio transiciones	26,04%	22,01%	49,66%
Promedio 2 últimas transiciones	9,07%	19,68%	43,52%
Tendencia 2 últimas transiciones	13,63%	18,93%	44,94%
Transición año anterior	27,18%	<b>16,96%</b>	<b>36,53%</b>

Tabla 64: Errores de transición con distintas estimaciones (variables RFM)

Variable(s): RFM	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>3,65%</b>	<b>8,31%</b>	<b>165,82%</b>
Promedio transiciones	25,64%	21,85%	615,53%
Promedio 2 últimas transiciones	7,11%	9,24%	167,73%
Tendencia 2 últimas transiciones	15,42%	18,38%	442,94%
Transición año anterior	27,69%	23,17%	621,09%

Tabla 65: Errores de transición con distintas estimaciones (variables RNM)

Variable(s): RNM	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>5,73%</b>	<b>12,76%</b>	<b>185,86%</b>
Promedio transiciones	25,90%	21,46%	404,90%
Promedio 2 últimas transiciones	8,73%	13,91%	198,11%
Tendencia 2 últimas transiciones	15,82%	18,08%	257,16%
Transición año anterior	27,23%	21,26%	364,45%

Tabla 66: Errores de transición con distintas estimaciones (variables RFN)

Variable(s): RFN	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>3,73%</b>	<b>8,51%</b>	173,73%
Promedio transiciones	25,63%	21,93%	618,62%
Promedio 2 últimas transiciones	7,09%	9,26%	<b>168,30%</b>
Tendencia 2 últimas transiciones	15,51%	18,63%	454,86%
Transición año anterior	27,67%	23,10%	623,86%

Tabla 67: Errores de transición con distintas estimaciones (variables FNM)

Variable(s): FNM	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>6,02%</b>	13,78%	200,26%
Promedio transiciones	25,84%	22,17%	555,45%
Promedio 2 últimas transiciones	8,71%	<b>12,55%</b>	<b>189,67%</b>
Tendencia 2 últimas transiciones	17,79%	25,87%	444,14%
Transición año anterior	27,69%	24,08%	575,75%

Tabla 68: Errores de transición con distintas estimaciones (variables RFMN)

Variable(s): RFNM	MAD	Error ponderado por número de transiciones	Error ponderado por distancia
Última transición	<b>3,71%</b>	<b>8,51%</b>	172,63%
Promedio transiciones	25,63%	21,93%	618,67%
Promedio 2 últimas transiciones	7,09%	9,27%	<b>168,40%</b>
Tendencia 2 últimas transiciones	15,48%	18,65%	453,35%
Transición año anterior	27,67%	23,11%	623,12%

Las tablas anteriores dan cuenta de que la última transición es la que mejor predice la transición futura, esto se verifica ya que éstas presentan en general el menor error.

## ANEXO E: Estadísticos descriptivos de las distintas segmentaciones realizadas

### Segmentación con variable R

Tabla 69: Descripción de los segmentos con respecto a las variables RFMN (variable R)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	6.278	9,7%	80,0	121,0	98,8	11,5	0,5	122,0	61,9	43,4
2	8.745	13,5%	44,0	79,0	59,3	10,4	1,1	122,0	58,3	41,7
3	15.937	24,5%	18,0	43,0	27,7	7,5	1,5	122,0	46,2	39,0
4	34.014	52,4%	0,0	17,0	6,6	5,2	1,6	122,0	30,1	33,5

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,6	3,9	4,7	1,0	65,0	1,9	1,9	0,5
2	0,0	25,7	3,1	3,8	1,0	61,0	3,5	3,6	0,9
3	0,0	25,5	2,6	3,1	1,0	68,0	5,7	5,4	1,4
4	0,0	25,1	2,0	2,3	1,0	72,0	12,2	11,8	3,1

### Segmentación con variable F

Tabla 70: Descripción de los segmentos con respecto a las variables RFMN (variable F)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	10.272	15,8%	0,0	121,0	42,7	33,9	94,1	122,0	118,4	7,4
2	7.576	11,7%	0,0	121,0	38,2	32,5	52,2	94,0	69,8	12,0
3	16.417	25,3%	0,0	121,0	29,9	29,4	22,9	52,0	34,3	8,1
4	30.709	47,3%	0,0	121,0	19,0	25,4	0,5	22,8	11,4	5,7

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,6	5,1	1,0	9,0	1,6	1,1	0,4
2	0,0	25,5	3,0	3,5	1,0	16,0	2,5	1,7	0,6
3	0,0	25,2	2,3	2,6	1,0	33,0	4,0	2,2	1,0
4	0,0	25,3	1,7	1,6	1,0	72,0	14,6	11,5	3,7

### Segmentación con variable M

Tabla 71: Descripción de los segmentos con respecto a las variables RFMN (variable M)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	913	1,4%	0,0	121,0	52,7	37,0	2,0	122,0	96,3	36,2
2	3.081	4,7%	0,0	121,0	46,3	36,4	2,0	122,0	78,4	42,7
3	12.230	18,8%	0,0	121,0	32,6	31,8	1,0	122,0	53,3	42,2
4	48.750	75,0%	0,0	121,0	24,9	28,4	0,5	122,0	34,4	35,0

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	14,9	25,7	19,7	2,9	1,0	38,0	1,8	2,1	0,4
2	7,2	14,9	10,1	2,1	1,0	45,0	2,8	3,3	0,7
3	2,8	7,2	4,3	1,2	1,0	71,0	5,6	6,5	1,4
4	0,0	2,8	1,2	0,6	1,0	72,0	9,7	10,7	2,4

## Segmentación con variable N

**Tabla 72: Descripción de los segmentos con respecto a las variables RFMN (variable N)**

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	46.886	72,2%	0,0	121,0	34,8	32,1	1,0	122,0	53,4	39,2
2	12.810	19,7%	0,0	117,0	11,3	13,6	0,5	85,2	10,2	6,7
3	4.002	6,2%	0,0	87,0	6,1	8,7	1,2	37,0	4,6	2,4
4	1.276	2,0%	0,0	84,0	3,5	6,3	0,6	17,7	2,4	1,0

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	2,8	3,5	1,0	9,0	3,8	2,4	1,0
2	0,1	21,7	1,6	1,4	10,0	22,0	14,4	3,6	3,6
3	0,0	18,8	1,4	1,1	23,0	41,0	29,7	5,3	7,4
4	0,1	12,6	1,2	0,8	42,0	72,0	53,0	8,6	13,2

## Segmentación con variables RF

**Tabla 73: Descripción de los segmentos con respecto a las variables RFMN (variables RF)**

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	13.146	20,2%	0,0	121,0	40,2	32,0	62,4	122,0	110,3	17,2
2	9.925	15,3%	38,0	121,0	76,5	22,5	0,5	99,0	33,8	20,9
3	41.903	64,5%	0,0	49,0	12,3	11,2	1,5	73,5	20,8	15,8



CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,3	4,8	1,0	12,0	1,8	1,3	0,4
2	0,0	25,3	2,7	3,3	1,0	65,0	3,4	3,2	0,8
3	0,0	25,1	1,9	1,9	2,0	72,0	11,8	10,9	2,9

## Segmentación con variables RM

Tabla 74: Descripción de los segmentos con respecto a las variables RFMN (variables RM)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	6.285	9,7%	79,0	121,0	98,8	11,5	0,5	122,0	61,9	43,3
2	8.738	13,4%	44,0	79,0	59,3	10,3	1,1	122,0	58,3	41,8
3	15.974	24,6%	17,0	43,0	27,7	7,6	1,5	122,0	46,2	39,0
4	33.977	52,3%	0,0	17,0	6,5	5,2	1,6	122,0	30,0	33,5

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,6	3,9	4,7	1,0	65,0	1,9	1,9	0,5
2	0,0	25,7	3,1	3,8	1,0	61,0	3,5	3,6	0,9
3	0,0	25,5	2,6	3,1	1,0	68,0	5,7	5,4	1,4
4	0,0	25,1	2,0	2,3	1,0	72,0	12,2	11,8	3,1

## Segmentación con variables FM

Tabla 75: Descripción de los segmentos con respecto a las variables RFMN (variables FM)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	10.369	16,0%	0,0	121,0	42,7	33,9	93,5	122,0	118,2	7,7
2	7.710	11,9%	0,0	121,0	38,2	32,4	51,0	93,6	68,9	12,0
3	16.489	25,4%	0,0	121,0	29,7	29,2	22,3	51,4	33,8	8,0
4	30.406	46,8%	0,0	121,0	19,0	25,4	0,5	22,6	11,3	5,6

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,6	5,1	1,0	9,0	1,6	1,1	0,4
2	0,0	25,5	3,1	3,5	1,0	16,0	2,5	1,7	0,6
3	0,0	25,2	2,3	2,6	1,0	33,0	4,1	2,2	1,0
4	0,0	25,3	1,7	1,6	1,0	72,0	14,7	11,5	3,7

## Segmentación con variables FN

Tabla 76: Descripción de los segmentos con respecto a las variables RFMN (variables FN)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	11.037	17,0%	0,0	121,0	42,9	33,8	89,0	122,0	116,6	9,9
2	9.947	15,3%	0,0	121,0	36,9	32,0	42,7	88,8	61,0	12,6
3	25.694	39,5%	0,0	121,0	28,3	29,3	6,0	43,8	24,4	8,4
4	18.296	28,2%	0,0	121,0	13,0	19,6	0,5	27,4	7,5	3,2

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,6	5,0	1,0	9,0	1,6	1,1	0,4
2	0,0	25,5	2,9	3,3	1,0	19,0	2,7	1,8	0,7
3	0,0	25,3	2,1	2,3	1,0	27,0	5,4	2,4	1,3
4	0,0	21,7	1,6	1,4	1,0	72,0	20,1	12,1	5,0

## Segmentación con variables RN

**Tabla 77: Descripción de los segmentos con respecto a las variables RFMN (variables RN)**

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	6.278	9,7%	80,0	121,0	98,8	11,5	0,5	122,0	61,9	43,4
2	8.732	13,4%	44,0	79,0	59,3	10,3	1,1	122,0	58,4	41,7
3	15.683	24,1%	17,0	44,0	27,9	7,6	2,4	122,0	47,6	39,1
4	28.519	43,9%	0,0	19,0	7,1	5,2	4,7	122,0	34,6	34,1
5	5.762	8,9%	0,0	49,0	4,4	5,2	1,5	37,0	4,3	2,4

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,6	3,9	4,7	1,0	65,0	1,9	1,9	0,5
2	0,0	25,7	3,1	3,8	1,0	48,0	3,5	3,5	0,9
3	0,0	25,5	2,6	3,2	1,0	38,0	5,2	4,3	1,3
4	0,0	25,1	2,1	2,4	1,0	22,0	8,0	5,2	2,0
5	0,0	18,8	1,4	1,0	21,0	72,0	34,0	12,0	8,5

## Segmentación con variables NM

**Tabla 78: Descripción de los segmentos con respecto a las variables RFMN (variables NM)**

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	48.795	75,1%	0,0	121,0	34,0	31,8	0,5	122,0	51,9	39,2
2	13.304	20,5%	0,0	109,0	10,0	12,5	0,9	77,4	8,8	5,9
3	2.875	4,4%	0,0	87,0	4,2	6,9	0,6	27,4	3,1	1,4

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	2,8	3,5	1,0	11,0	4,1	2,7	1,0
2	0,1	20,7	1,6	1,2	11,0	30,0	17,1	5,2	4,3
3	0,0	18,8	1,3	1,0	31,0	72,0	43,2	10,7	10,8

## Segmentación con variables RFM

Tabla 79: Descripción de los segmentos con respecto a las variables RFMN (variables RFM)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	13.148	20,2%	0,0	121,0	40,2	32,0	62,4	122,0	110,3	17,2
2	9.926	15,3%	38,0	121,0	76,5	22,5	0,5	99,0	33,8	20,9
3	41.900	64,5%	0,0	49,0	12,3	11,2	1,5	73,5	20,8	15,8

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,3	4,8	1,0	12,0	1,8	1,3	0,4
2	0,0	25,3	2,7	3,3	1,0	65,0	3,4	3,2	0,8
3	0,0	25,1	1,9	1,9	2,0	72,0	11,8	10,9	2,9

## Segmentación con variables RNM

Tabla 80: Descripción de los segmentos con respecto a las variables RFMN (variables RNM)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	7.206	11,1%	76,0	121,0	96,1	12,9	0,5	122,0	61,1	43,2
2	9.097	14,0%	40,0	76,0	55,1	9,6	1,1	122,0	58,3	41,5
3	15.276	23,5%	16,0	41,0	26,0	6,7	2,5	122,0	46,1	38,7
4	27.639	42,5%	0,0	18,0	6,8	5,0	4,7	122,0	34,7	34,2
5	5.756	8,9%	0,0	49,0	4,4	5,2	1,5	37,0	4,3	2,4

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,6	3,8	4,6	1,0	65,0	2,0	2,1	0,5
2	0,0	25,7	3,0	3,7	1,0	48,0	3,6	3,6	0,9
3	0,0	25,5	2,6	3,1	1,0	33,0	5,5	4,4	1,4
4	0,0	25,1	2,2	2,5	1,0	22,0	8,1	5,2	2,0
5	0,0	18,8	1,4	1,0	21,0	72,0	34,1	12,0	8,5

### Segmentación con variables RFN

Tabla 81: Descripción de los segmentos con respecto a las variables RFMN (variables RFN)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	13.223	20,4%	0,0	121,0	40,0	32,0	62,0	122,0	110,0	17,4
2	10.054	15,5%	37,0	121,0	76,0	22,7	0,5	99,0	33,9	20,8
3	41.697	64,2%	0,0	51,0	12,2	11,1	1,5	73,0	20,7	15,6

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,3	4,8	1,0	12,0	1,8	1,3	0,4
2	0,0	25,3	2,7	3,3	1,0	65,0	3,3	3,1	0,8
3	0,0	25,1	1,9	1,9	2,0	72,0	11,8	10,9	3,0

### Segmentación con variables RFNM

Tabla 82: Descripción de los segmentos con respecto a las variables RFMN (variables RFNM)

CLUSTER	TAMAÑO		R [días]				F [días]			
			MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD
1	13.229	20,4%	0,0	121,0	40,0	32,0	62,0	122,0	110,0	17,4
2	10.052	15,5%	37,0	121,0	76,0	22,7	0,5	99,0	33,9	20,8
3	41.693	64,2%	0,0	51,0	12,2	11,1	1,5	73,0	20,7	15,6

CLUSTER	M [U.F.]				N [número de transacciones en el periodo]				
	MIN	MAX	PROM	DESV STD	MIN	MAX	PROM	DESV STD	Promedio N Mensual
1	0,0	25,7	4,3	4,8	1,0	12,0	1,8	1,3	0,4
2	0,0	25,3	2,7	3,3	1,0	65,0	3,3	3,1	0,8
3	0,0	25,1	1,9	1,9	2,0	72,0	11,8	10,9	3,0

## ANEXO F: Detalle de tablas truncadas en sección 6.1.2

Agrupación por Tipo de consumo

Tabla 83: Detalle categoría OTROS en tabla 25

AGRUPACION POR TIPO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
COMPRA CUOTA FIJA	1.093	3.802,64	6,2%	5,1%	3,48
COMPRAS INTERNACIONALES	432	1.426,45	2,5%	1,9%	3,30
AVANCE INTERNACIONAL	40	244,67	0,2%	0,3%	6,12
COMPRA BENCINA	17	62,22	0,1%	0,1%	3,66
COMPRA CASINO NAC – INT	21	28,00	0,1%	0,0%	1,33
RECAUDACION SII	20	23,87	0,1%	0,0%	1,19
Total OTROS	1.623	5.587,86	9,2%	7,5%	3,44

Tabla 84: Detalle categoría OTROS en tabla 26

AGRUPACION POR TIPO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
COMPRA CUOTA FIJA	1.826	4.727,33	6,9%	6,6%	2,59
COMPRAS INTERNACIONALES	454	1.180,02	1,7%	1,6%	2,60
AVANCE INTERNACIONAL	70	268,32	0,3%	0,4%	3,83
COMPRA BENCINA	60	166,18	0,2%	0,2%	2,77
COMPRA CASINO NAC – INT	24	50,10	0,1%	0,1%	2,09
RECAUDACION SII	27	17,95	0,1%	0,0%	0,66
<b>Total OTROS</b>	<b>2.461</b>	<b>6.409,91</b>	<b>9,3%</b>	<b>8,9%</b>	<b>2,60</b>

Tabla 85: Detalle categoría OTROS en tabla 27

AGRUPACION POR TIPO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
COMPRA CUOTA FIJA	8.113	16.237,40	5,9%	5,9%	2,00
COMPRAS INTERNACIONALES	2.278	5.601,16	1,7%	2,0%	2,46
AVANCE INTERNACIONAL	246	1.140,47	0,2%	0,4%	4,64
COMPRA BENCINA	193	386,53	0,1%	0,1%	2,00
COMPRA CASINO NAC – INT	198	280,73	0,1%	0,1%	1,42
RECAUDACION SII	261	97,75	0,2%	0,0%	0,37
<b>Total OTROS</b>	<b>11.289</b>	<b>23.744,04</b>	<b>8,2%</b>	<b>8,7%</b>	<b>2,10</b>

Tabla 86: Detalle categoría OTROS en tabla 28

AGRUPACION POR TIPO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
COMPRA CUOTA FIJA	11.566	18.855,11	5,1%	5,3%	1,63
COMPRAS INTERNACIONALES	5.362	10.621,56	2,4%	3,0%	1,98
AVANCE INTERNACIONAL	797	3.100,85	0,4%	0,9%	3,89
COMPRA BENCINA	305	922,68	0,1%	0,3%	3,03
COMPRA CASINO NAC – INT	524	667,58	0,2%	0,2%	1,27
RECAUDACION SII	627	411,24	0,3%	0,1%	0,66
<b>Total OTROS</b>	<b>19.181</b>	<b>34.579,02</b>	<b>8,5%</b>	<b>9,7%</b>	<b>1,80</b>

Tabla 87: Detalle categoría OTROS en tabla 29

AGRUPACION POR TIPO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
COMPRA CUOTA FIJA	4.423	6.576,36	3,1%	3,5%	1,49
AVANCE EN CUOTAS	538	6.334,76	0,4%	3,4%	11,77
AVANCE INTERNACIONAL	563	2.102,75	0,4%	1,1%	3,73
COMPRA CASINO NAC – INT	584	955,27	0,4%	0,5%	1,64
RECAUDACION SII	478	430,27	0,3%	0,2%	0,90
COMPRA BENCINA	159	396,99	0,1%	0,2%	2,50
<b>Total OTROS</b>	<b>6.745</b>	<b>16.796,41</b>	<b>4,7%</b>	<b>9,0%</b>	<b>2,49</b>



Agrupación por Rubro de consumo

Tabla 88: Detalle categoría OTROS en tabla 30

AGRUPACION POR RUBRO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
SUPERMERCADOS	1.255	2.316,93	7,2%	3,1%	1,85
RECAUDACION SECTOR PUBLICO	457	1.666,99	2,6%	2,2%	3,65
ACCESORIOS Y SERVICIOS AUTOMOTRICES	332	1.481,39	1,9%	2,0%	4,46
CLINICAS Y PRESTADORES	197	1.143,07	1,1%	1,5%	5,80
RECAUDACION	581	1.059,46	3,3%	1,4%	1,82
OTROS T&E	197	637,20	1,1%	0,9%	3,23
EDUCACION	63	614,06	0,4%	0,8%	9,75
FARMACIAS	505	530,33	2,9%	0,7%	1,05
OTROS CONSUMOS INTERNACIONAL	155	450,24	0,9%	0,6%	2,90
OPERADORES Y RENT A CAR	57	444,89	0,3%	0,6%	7,81
HOTELES	76	427,14	0,4%	0,6%	5,62
OTROS SERVICIOS	196	405,04	1,1%	0,5%	2,07
COMBUSTIBLES	517	382,33	2,9%	0,5%	0,74
RESTAURANTES	149	208,31	0,8%	0,3%	1,40
CASINOS Y JUEGOS DE AZAR	43	93,35	0,2%	0,1%	2,17
COMPRAS INTERNET VARIAS	6	21,13	0,0%	0,0%	3,52
<b>Total OTROS</b>	<b>4.786</b>	<b>11.881,86</b>	<b>27,3%</b>	<b>16,0%</b>	<b>2,48</b>

Tabla 89: Detalle categoría OTROS en tabla 31

AGRUPACION POR RUBRO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
OTROS CONSUMOS	1.498	2.936,94	5,7%	4,1%	1,96
RECAUDACION SECTOR PUBLICO	668	1.833,15	2,5%	2,6%	2,74
RECAUDACION	1.097	1.650,34	4,1%	2,3%	1,50
ACCESORIOS Y SERVICIOS AUTOMOTRICES	412	1.598,37	1,6%	2,2%	3,88
CLINICAS Y PRESTADORES	294	1.112,45	1,1%	1,6%	3,78
EDUCACION	99	1.019,20	0,4%	1,4%	10,29
FARMACIAS	1.037	928,11	3,9%	1,3%	0,89
COMBUSTIBLES	1.063	717,46	4,0%	1,0%	0,67
OTROS T&E	286	558,89	1,1%	0,8%	1,95
HOTELES	96	523,92	0,4%	0,7%	5,46
OTROS CONSUMOS INTERNACIONAL	167	419,39	0,6%	0,6%	2,51
OTROS SERVICIOS	244	409,33	0,9%	0,6%	1,68
RESTAURANTES	285	324,66	1,1%	0,5%	1,14
CASINOS Y JUEGOS DE AZAR	106	265,35	0,4%	0,4%	2,50
OPERADORES Y RENT A CAR	44	229,09	0,2%	0,3%	5,21
COMPRAS INTERNET VARIAS	4	13,16	0,0%	0,0%	3,29
<b>Total OTROS</b>	<b>7.400</b>	<b>14.539,82</b>	<b>28,0%</b>	<b>20,3%</b>	<b>1,96</b>

Tabla 90: Detalle categoría OTROS en tabla 32

AGRUPACION POR RUBRO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
OTROS CONSUMOS	10.422	13.413,06	7,6%	4,9%	1,29
LINEAS AEREAS	2.192	12.065,99	1,6%	4,4%	5,50
RECAUDACION SECTOR PUBLICO	2.838	6.751,98	2,1%	2,5%	2,38
ACCESORIOS Y SERVICIOS AUTOMOTRICES	1.722	5.363,32	1,3%	2,0%	3,11
FARMACIAS	6.626	5.105,79	4,8%	1,9%	0,77
COMBUSTIBLES	7.069	4.629,31	5,2%	1,7%	0,65
CLINICAS Y PRESTADORES	1.427	3.649,50	1,0%	1,3%	2,56
EDUCACION	412	3.509,93	0,3%	1,3%	8,52
OTROS T&E	1.468	2.343,21	1,1%	0,9%	1,60
OTROS CONSUMOS INTERNACIONAL	804	2.028,95	0,6%	0,7%	2,52
OTROS SERVICIOS	1.271	1.782,44	0,9%	0,7%	1,40
RESTAURANTES	1.771	1.685,61	1,3%	0,6%	0,95
HOTELES	396	1.209,86	0,3%	0,4%	3,06
OPERADORES Y RENT A CAR	160	1.123,10	0,1%	0,4%	7,02
CASINOS Y JUEGOS DE AZAR	553	793,01	0,4%	0,3%	1,43
COMPRAS INTERNET VARIAS	31	101,90	0,0%	0,0%	3,29
<b>Total OTROS</b>	<b>39.162</b>	<b>65.556,96</b>	<b>28,5%</b>	<b>24,0%</b>	<b>1,67</b>

Tabla 91: Detalle categoría OTROS en tabla 33

AGRUPACION POR RUBRO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
MUEBLES Y ARTICULOS ELECTRONICOS	5.108	13.775,56	2,3%	3,9%	2,70
LINEAS AEREAS	2.028	11.419,46	0,9%	3,2%	5,63
FARMACIAS	15.903	11.112,82	7,1%	3,1%	0,70
COMBUSTIBLES	16.672	10.854,31	7,4%	3,0%	0,65
RECAUDACION SECTOR PUBLICO	3.858	8.281,07	1,7%	2,3%	2,15
ACCESORIOS Y SERVICIOS AUTOMOTRICES	2.262	5.942,22	1,0%	1,7%	2,63
EDUCACION	623	5.468,25	0,3%	1,5%	8,78
CLINICAS Y PRESTADORES	2.343	5.291,81	1,0%	1,5%	2,26
RESTAURANTES	4.385	3.940,05	1,9%	1,1%	0,90
OTROS CONSUMOS INTERNACIONAL	1.502	3.266,98	0,7%	0,9%	2,18
OTROS T&E	2.758	3.157,78	1,2%	0,9%	1,14
OTROS SERVICIOS	2.624	3.085,31	1,2%	0,9%	1,18
HOTELES	725	2.394,96	0,3%	0,7%	3,30
OPERADORES Y RENT A CAR	238	2.133,45	0,1%	0,6%	8,96
CASINOS Y JUEGOS DE AZAR	1.233	1.772,92	0,5%	0,5%	1,44
COMPRAS INTERNET VARIAS	54	70,20	0,0%	0,0%	1,30
<b>Total OTROS</b>	<b>62.316</b>	<b>91.967,15</b>	<b>27,6%</b>	<b>25,8%</b>	<b>1,48</b>

Tabla 92: Detalle categoría OTROS en tabla 34

AGRUPACION POR RUBRO CONSUMO	Número de transacciones (#)	Monto transacciones (UF)	Número de transacciones (%)	Monto transacciones (%)	M (monto / # trans.)
VESTUARIO Y CALZADO	7.535	9.736,19	5,3%	5,2%	1,29
FARMACIAS	12.806	8.225,72	8,9%	4,4%	0,64
MUEBLES Y ARTICULOS ELECTRONICOS	2.835	6.718,59	2,0%	3,6%	2,37
RECAUDACION SECTOR PUBLICO	2.216	5.374,15	1,5%	2,9%	2,43
LINEAS AEREAS	688	3.905,14	0,5%	2,1%	5,68
EDUCACION	406	3.795,52	0,3%	2,0%	9,35
CLINICAS Y PRESTADORES	1.817	3.198,31	1,3%	1,7%	1,76
RESTAURANTES	3.601	3.147,88	2,5%	1,7%	0,87
OTROS SERVICIOS	1.854	3.067,90	1,3%	1,6%	1,65
ACCESORIOS Y SERVICIOS AUTOMOTRICES	1.154	2.894,02	0,8%	1,5%	2,51
OTROS T&E	1.770	1.915,03	1,2%	1,0%	1,08
OTROS CONSUMOS INTERNACIONAL	964	1.707,46	0,7%	0,9%	1,77
HOTELES	448	1.595,18	0,3%	0,9%	3,56
CASINOS Y JUEGOS DE AZAR	1.111	1.482,35	0,8%	0,8%	1,33
OPERADORES Y RENT A CAR	152	926,41	0,1%	0,5%	6,09
COMPRAS INTERNET VARIAS	16	56,68	0,0%	0,0%	3,54
<b>Total OTROS</b>	<b>39.373</b>	<b>57.746,54%</b>	<b>27.5%</b>	<b>30.9%</b>	<b>1.47</b>

**ANEXO G: Esquemas de los modelos de Árboles de decisión para los distintos segmentos**

Figura 41: Esquema del modelo de Árbol de decisión para el segmento Inactivos

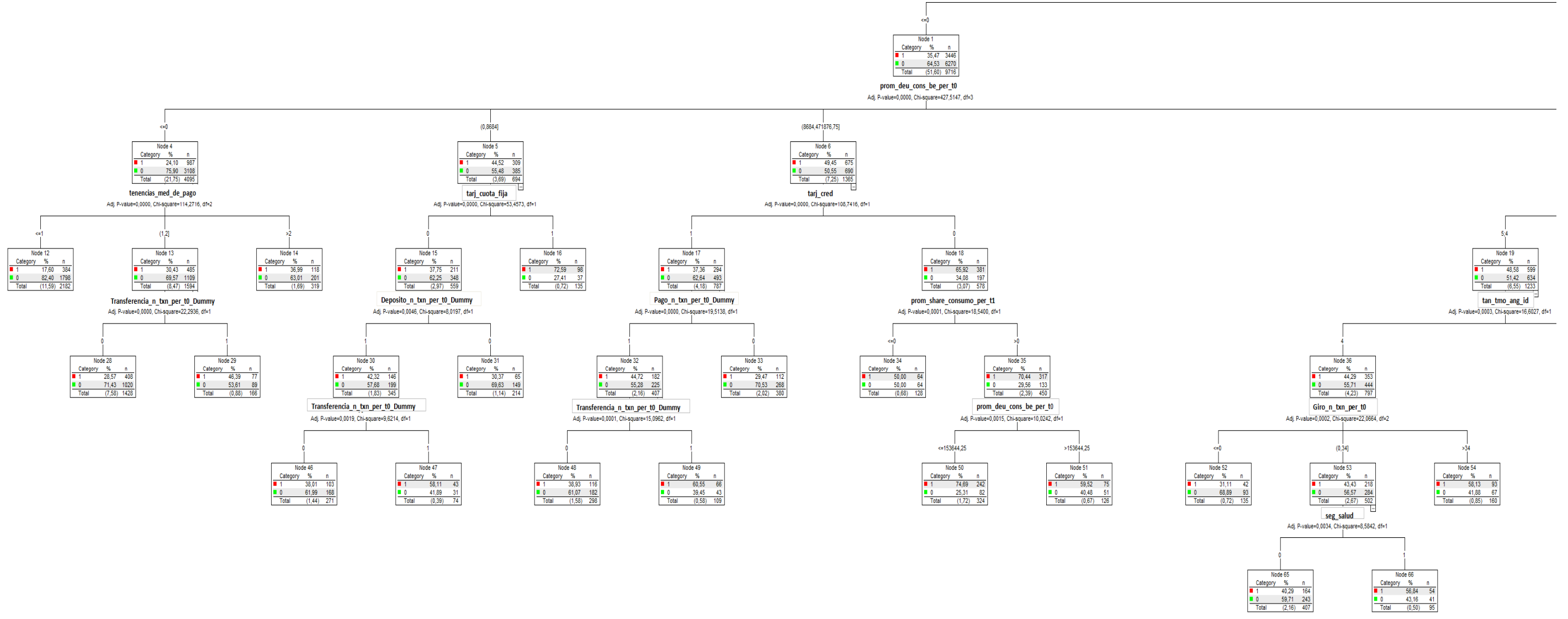


Figura 41: Esquema del modelo de Árbol de decisión para el segmento Inactivos (continuación)

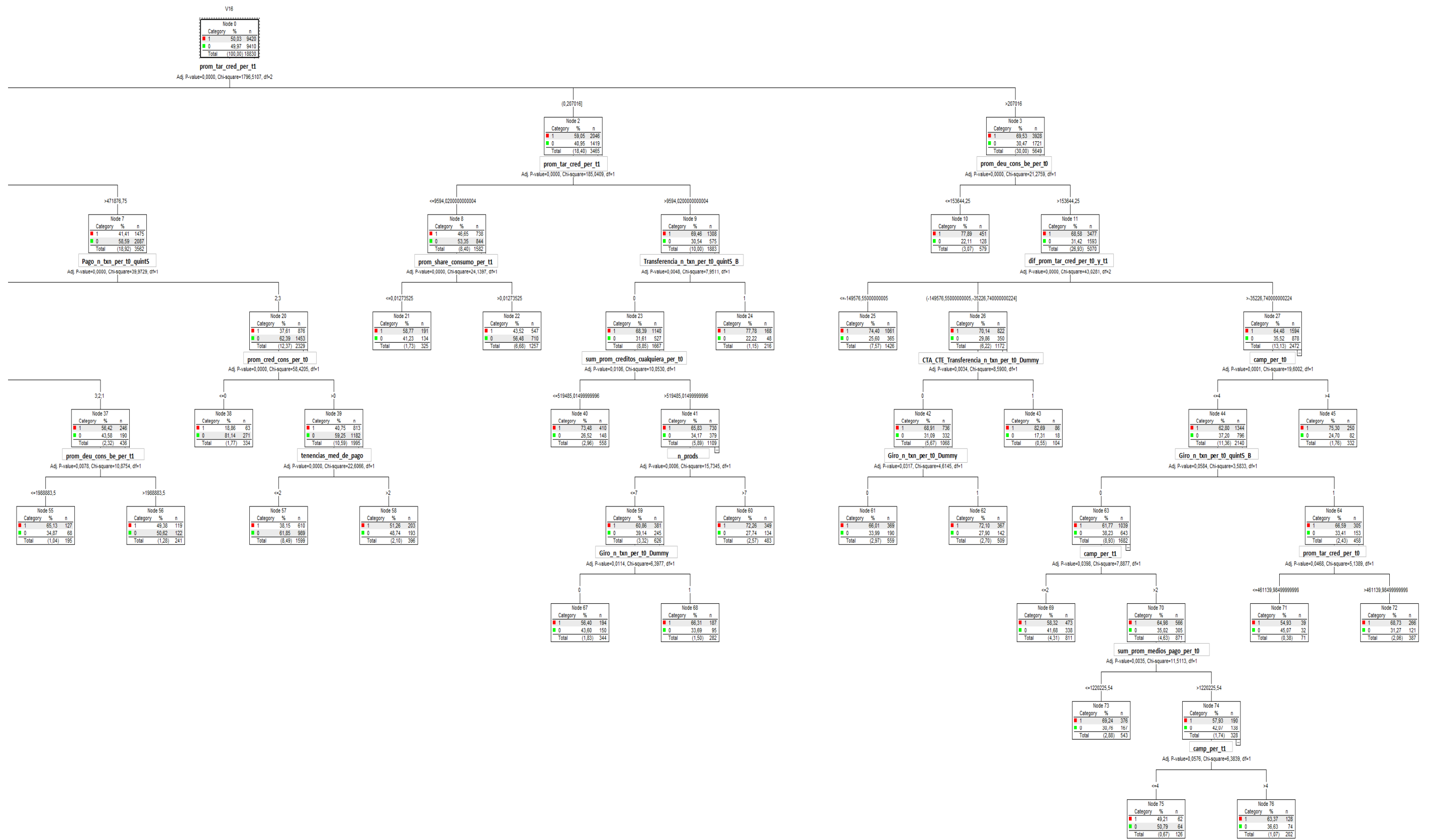




Figura 42: Esquema del modelo de Árbol de decisión para el segmento Muy bajo consumo

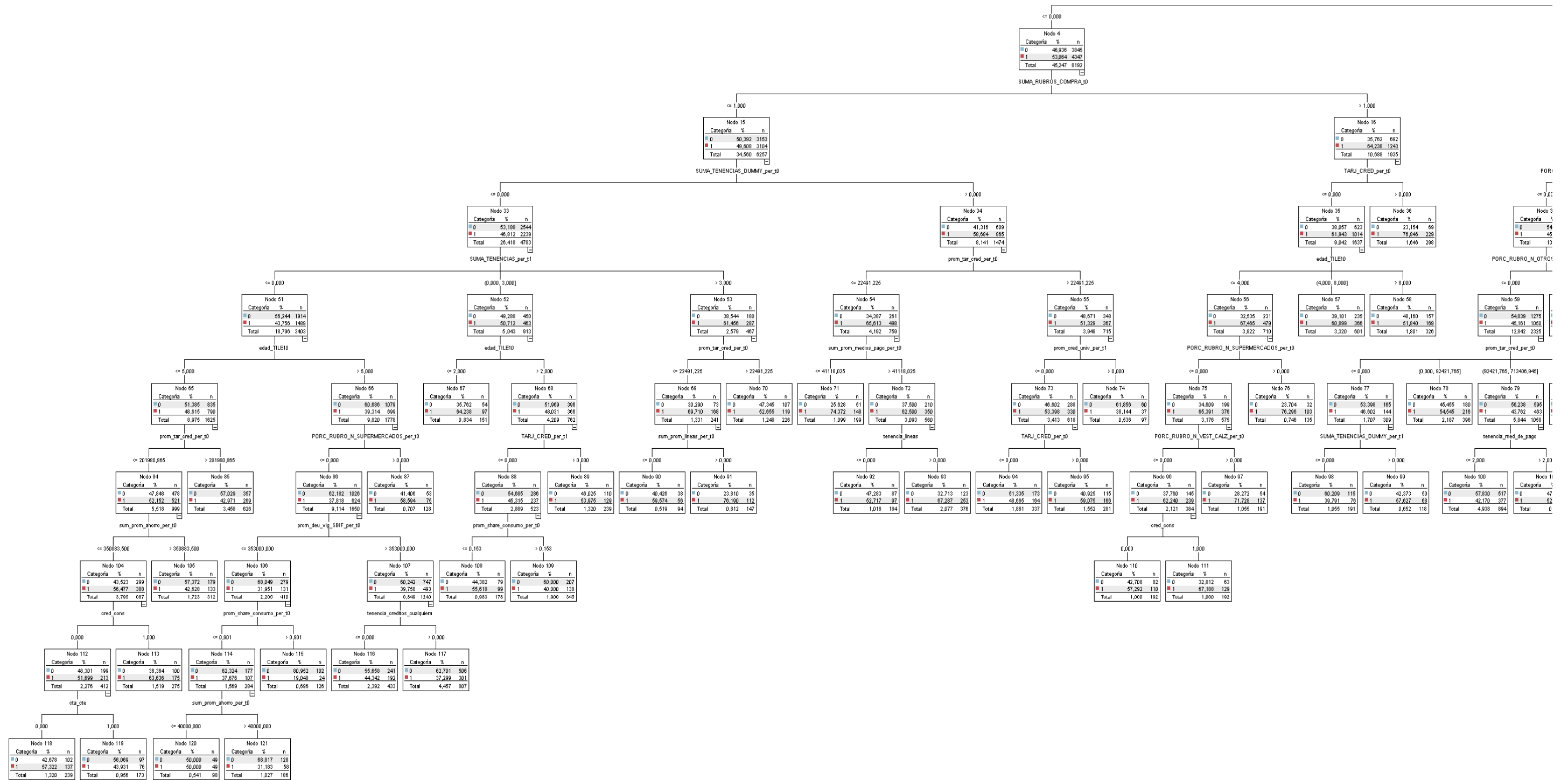


Figura 42: Esquema del modelo de Árbol de decisión para el segmento Muy bajo consumo (continuación)

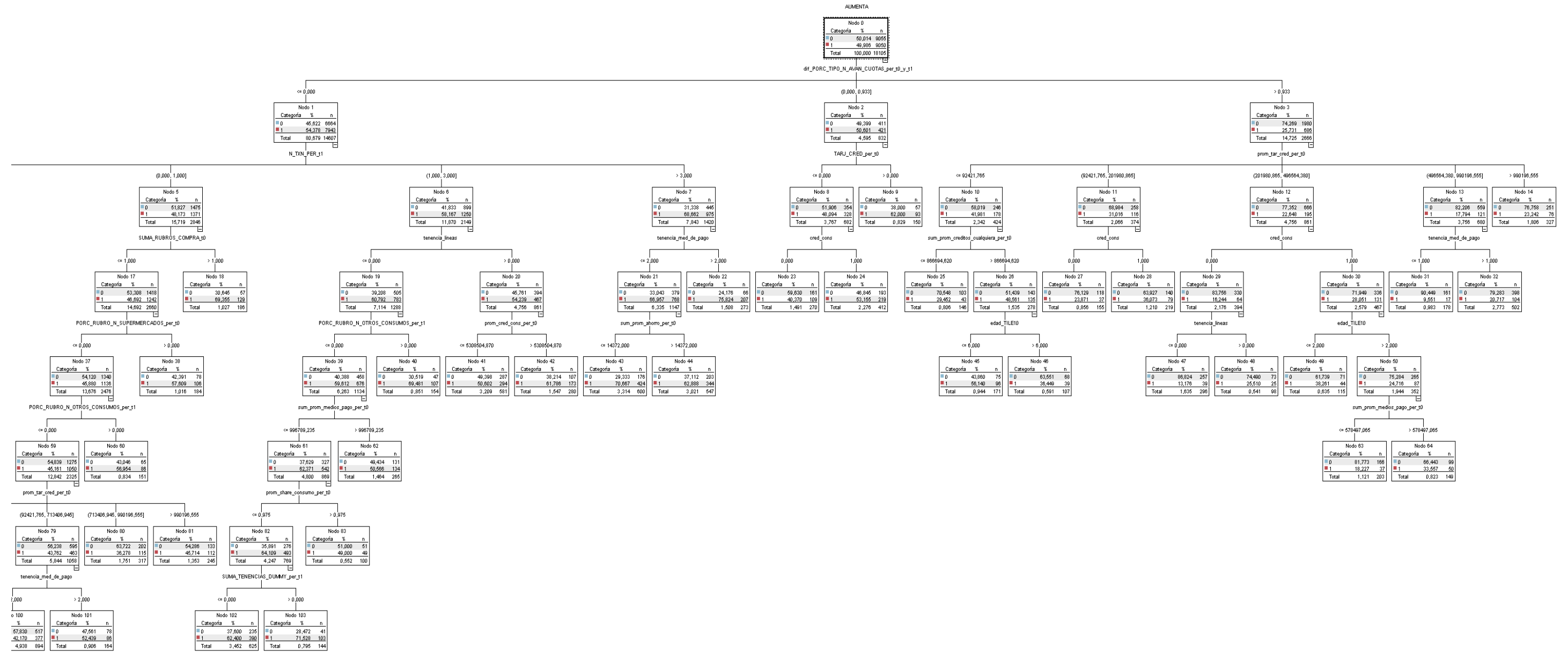


Figura 413: Esquema modelo de Árbol de decisión para el segmento Bajo consumo

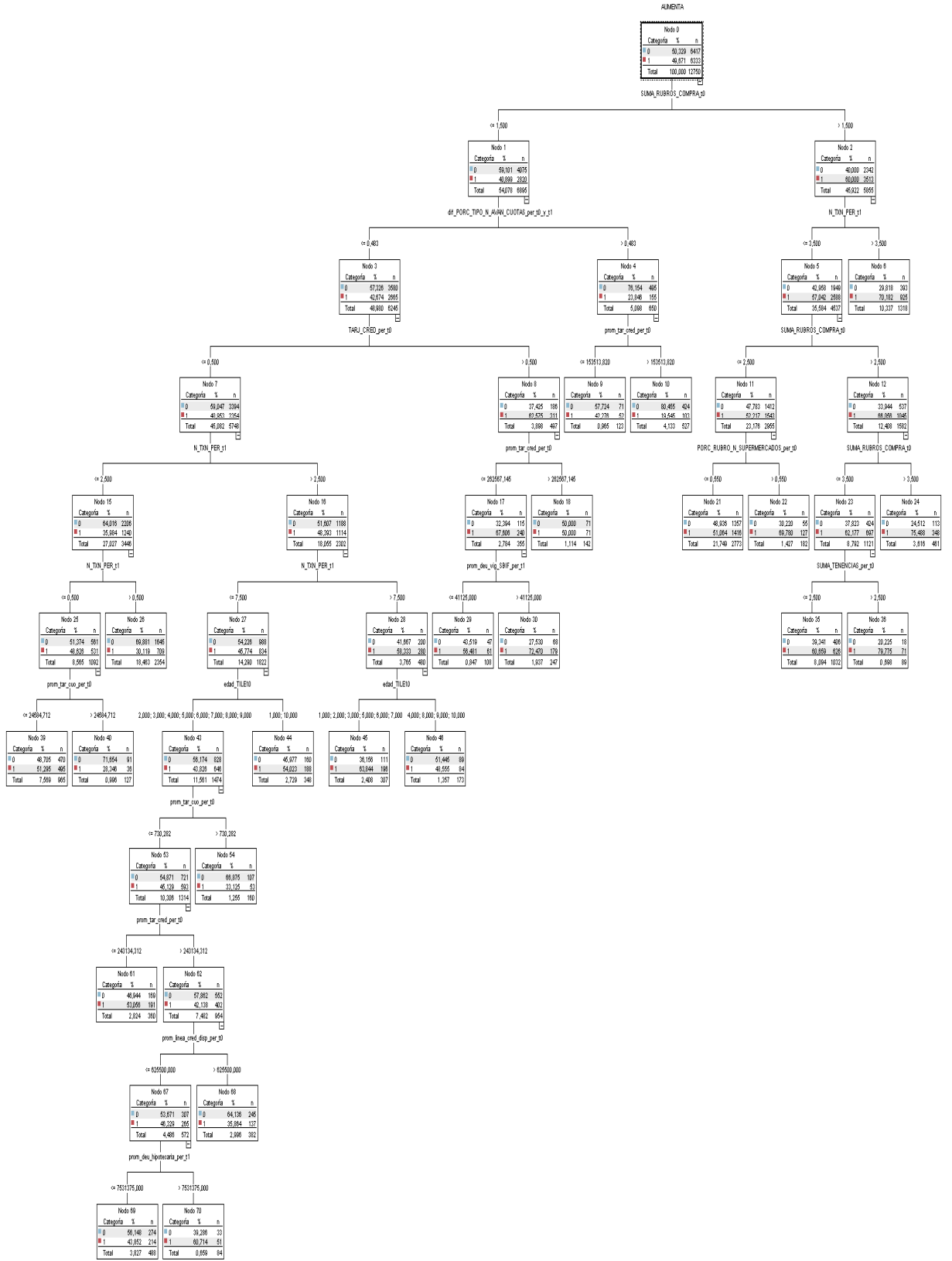


Figura 42: Esquema modelo de Árbol de decisión para el segmento Intermedio

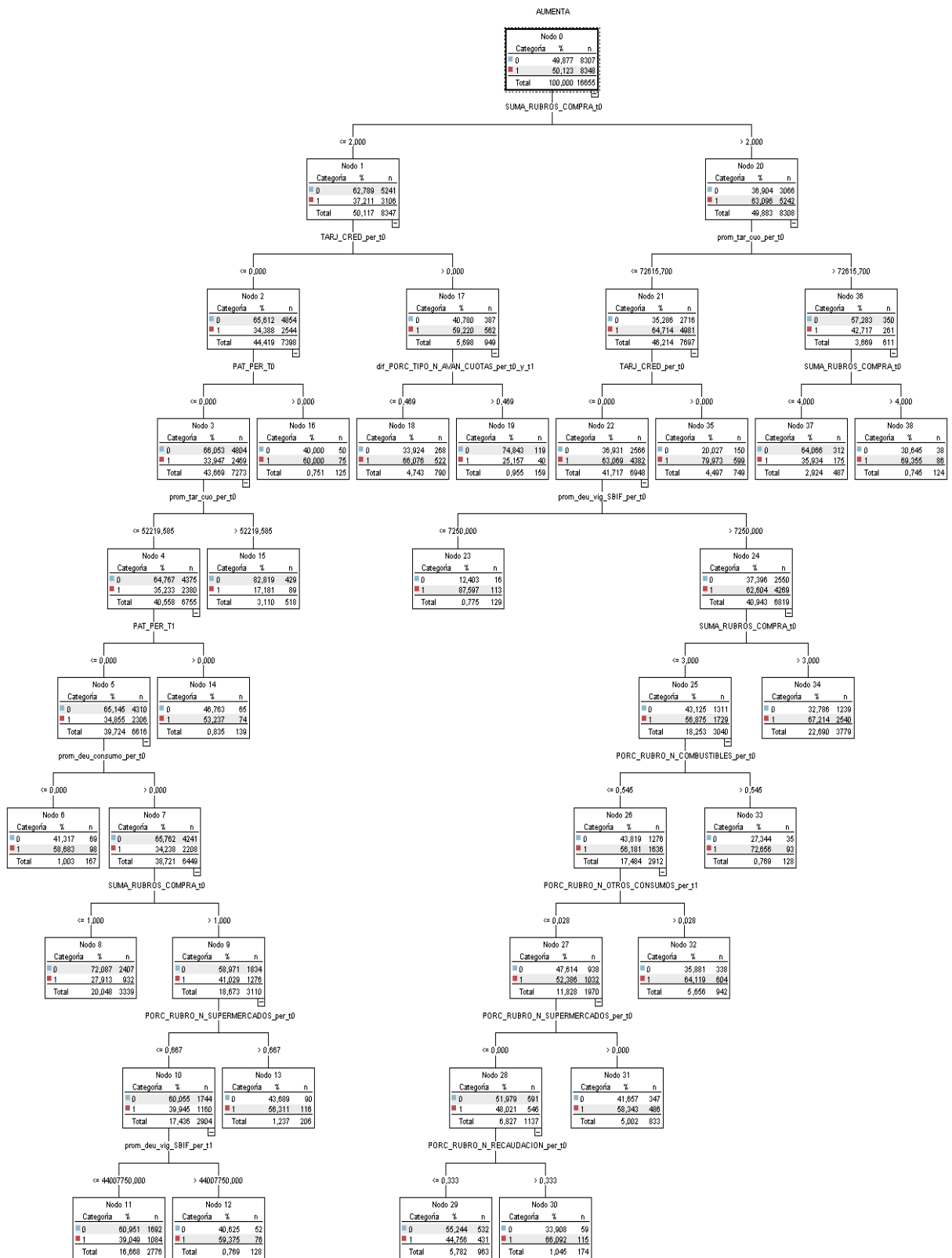
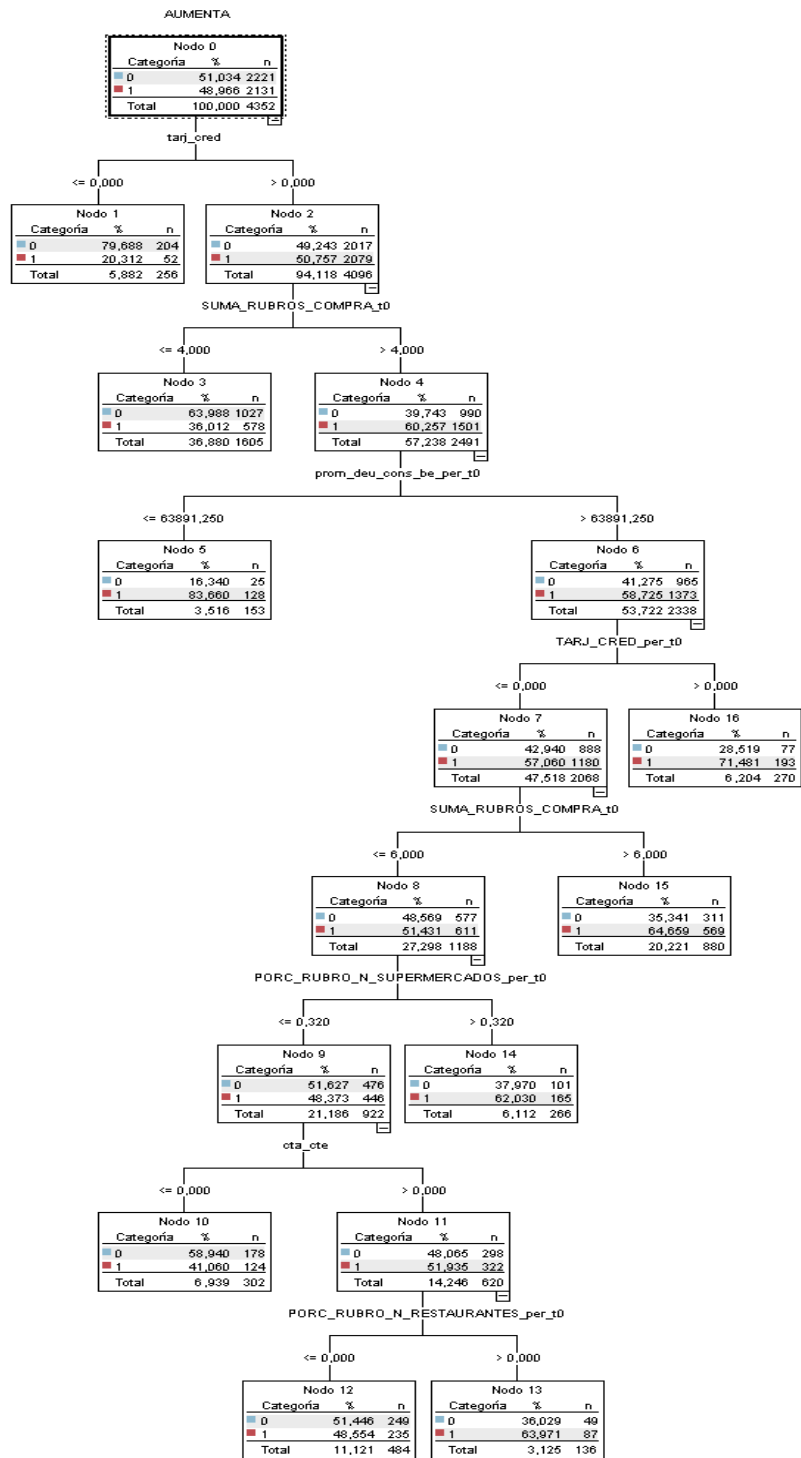


Figura 435: Esquema modelo de Árbol de decisión para el segmento de Alto consumo



## ANEXO H: Diccionario de Variables Árboles de Decisión.

A continuación se describen las variables ocupadas en los distintos árboles de decisión, separadas por segmento y ordenadas por nombre.

### Modelo Segmento Inactivos

Tabla 93: Diccionario variables segmento Inactivos

Variable	Significado
<b>camp_per_t0</b>	Número de Campañas realizadas a un cliente en el período t0
<b>camp_per_t1</b>	Número de Campañas realizadas a un cliente en el período t1 (t1 es anterior a t0)
<b>CTA_CTE_Transferencia_n_txn_per_t0_Dummy</b>	Número de transferencias de cuenta corriente en el período t0 (Dummy)
<b>Deposito_n_txn_per_t0_Dummy</b>	Número de depósitos con otros medios de pago en el período t0 (Dummy)
<b>dif_prom_tar_cred_per_t0_y_t1</b>	Diferencia entre el cupo promedio utilizado de tarjeta de crédito entre el período t0 y el anterior (t1)
<b>Giro_n_txn_per_t0</b>	Número de giros con otros medios de pago en el período t0
<b>Giro_n_txn_per_t0_Dummy</b>	Número de giros con otros medios de pago en el período t0 (Dummy)
<b>Giro_n_txn_per_t0_quint5_B</b>	Número de giros con otros medios de pago en el período t0 (Quintiles con correlación con variable dependiente: 5)
<b>n_prods</b>	Número de productos contratados

<b>Pago_n_txn_per_t0_Dummy</b>	Si la persona registra o no pagos en otros medios de pago (Dummy)
<b>Pago_n_txn_per_t0_quint5</b>	Si la persona registra o no pagos en otros medios de pago (Quintiles con correlación con variable dependiente: 5)
<b>prom_cred_cons_per_t0</b>	Promedio de los créditos de consumo contratados en t0
<b>prom_deu_cons_be_per_t0</b>	Promedio de deuda de consumo en BancoEstado en el período t0 (t1 es anterior a t0)
<b>prom_deu_cons_be_per_t1</b>	Promedio de deuda de consumo en BancoEstado en el período t1 (t1 es anterior a t0)
<b>prom_share_consumo_per_t0</b>	Promedio share deuda de consumo en BancoEstado en el período t0
<b>prom_share_consumo_per_t1</b>	Promedio share deuda de consumo en BancoEstado en el período t1 (t1 es anterior a t0)
<b>prom_tar_cred_per_t0</b>	Promedio cupo tarjeta de crédito ocupado en el período t0
<b>prom_tar_cred_per_t1</b>	Promedio cupo tarjeta de crédito ocupado en el período t1 (t1 es anterior a t0)
<b>seg_salud</b>	Si tiene seguro de salud contratado (Dummy)
<b>sum_prom_creditos_cualquiera_per_t0</b>	Suma de los promedios de deuda de créditos (de cualquier tipo) vigentes en t0
<b>sum_prom_medios_pago_per_t0</b>	Suma de los promedios de cupo utilizado en los otros medios de pago
<b>tan_tmo_ang_id</b>	Antigüedad del cliente (en años)

<b>tarj_cred</b>	Si la tarjeta de crédito es normal (Dummy)
<b>tarj_cuota_fija</b>	Si la tarjeta de crédito es de cuota fija (Dummy)
<b>tenencias_med_de_pago</b>	Número de medios de pago contratados por el cliente
<b>Transferencia_n_txn_per_t0_Dummy</b>	Número de transacciones con otros medios de pago (Dummy)
<b>Transferencia_n_txn_per_t0_quint5_B</b>	Número de transacciones con otros medios de pago (Quintiles con correlación con variable dependiente: 3, 4 y 5)

### Modelo Segmento Muy bajo consumo

Tabla 94 : Diccionario variables segmento Muy bajo consumo

<b>Variable</b>	<b>Significado</b>
<b>cred_cons</b>	Si tiene contratado un crédito de consumo (Dummy)
<b>cta_cte</b>	Si tiene contratada una cuenta corriente (Dummy)
<b>dif_PORC_TIPO_N_AVAN_CUOTAS_per_t0_y_t1</b>	(Porcentaje del total de transacciones hechas en avance en varias cuotas en t0) - (Porcentaje del total de transacciones hechas en avance en varias cuotas en t1)
<b>edad_TILE10</b>	Edad (Quintiles)
<b>N_TXN_PER_t1</b>	Número de transacciones en el período



	t1 (t1 es anterior a t0)
<b>PORC_RUBRO_N_OTROS_CONSUMOS_per_t1</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro OTROS CONSUMOS en el período t1 (t1 es anterior a t0)
<b>PORC_RUBRO_N_SUPERMERCADOS_per_t0</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro SUPERMERCADOS en el período t0
<b>PORC_RUBRO_N_VEST_CALZ_per_t0</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro vestuario y calzado en el período t0
<b>prom_cred_cons_per_t0</b>	Promedio de los créditos de consumo contratados en t0
<b>prom_cred_univ_per_t1</b>	Promedio de los créditos de consumo contratados en t0
<b>prom_deu_vig_SBIF_per_t0</b>	Promedio de deuda vigente el sistema financiero (SBIF) en el período t0
<b>prom_share_consumo_per_t0</b>	Promedio share deuda de consumo en BancoEstado en el período t0
<b>prom_tar_cred_per_t0</b>	Promedio cupo tarjeta de crédito ocupado en el período t0
<b>sum_prom_ahorro_per_t0</b>	Suma de los promedios de dinero depositado en productos de ahorro en t0
<b>sum_prom_creditos_cualquiera_per_t0</b>	Suma de los promedios de deuda de créditos (de cualquier tipo) vigentes en t0
<b>sum_prom_lineas_per_t0</b>	Suma de los promedios utilizados en distintas líneas de crédito en t0
<b>sum_prom_medios_pago_per_t0</b>	Suma de los promedios de cupo utilizado en los otros medios de pago
<b>SUMA_RUBROS_COMPRA_t0</b>	Suma de los distintos rubros donde la

	persona realizó transacciones en el período t0
<b>SUMA_TENENCIAS_DUMMY_per_t0</b>	Número productos contratados en el período t0 (Dummy)
<b>SUMA_TENENCIAS_DUMMY_per_t1</b>	Número productos contratados en el período t1 (Dummy)
<b>SUMA_TENENCIAS_per_t1</b>	Número productos contratados en el período t1 (t1 es anterior a t0)
<b>TARJ_CRED_per_t0</b>	Si contrató tarjeta de crédito en el período t0
<b>TARJ_CRED_per_t1</b>	Si contrató tarjeta de crédito en el período t1 (t1 es anterior a t0)
<b>tenencia_creditos_cualquiera</b>	Número de créditos que posee el cliente (de cualquier tipo)
<b>tenencia_lineas</b>	Número de líneas de crédito que posee el cliente (de cualquier tipo)
<b>tenencia_med_de_pago</b>	Número de medios de pago que posee el cliente

## Modelo Segmento Bajo Consumo

Tabla 95: Diccionario variables segmento Bajo consumo

Variable	Significado
<b>dif_PORC_TIPO_N_AVAN_CUOTAS_per_t0_y_t1</b>	(Porcentaje del total de transacciones hechas en avance en varias cuotas en t0) - (Porcentaje del total de transacciones hechas en avance en varias cuotas en t1)
<b>edad_TILE10</b>	Edad (Quintiles)

<b>N_TXN_PER_t1</b>	Número de transacciones en el período t1 (t1 es anterior a t0)
<b>PORC_RUBRO_N_SUPERMERCADOS_per_t0</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro SUPERMERCADOS en el período t0
<b>prom_deu_vig_SBIF_per_t1</b>	Promedio de deuda vigente en el sistema financiero (SBIF) en el período t1 (t1 es anterior a t0)
<b>prom_deuda_hipotecaria_per_t1</b>	Promedio de deuda hipotecaria en el período t1 (t1 es anterior a t0)
<b>prom_lin_cred_disp_per_t0</b>	Promedio de línea de crédito disponible en t0
<b>prom_tar_cred_per_t0</b>	Promedio cupo tarjeta cuota fija ocupado en el período t0
<b>prom_tar_cuo_per_t0</b>	Promedio cupo tarjeta cuota fija ocupado en el período t0
<b>SUMA_RUBROS_COMPRA_t0</b>	Suma de los distintos rubros donde la persona realizó transacciones en el período t0
<b>SUMA_TENENCIAS_per_t0</b>	Número productos contratados en el período t0
<b>TARJ_CRED_per_t0</b>	Si contrató tarjeta de crédito en el período t0

## Modelo Segmento Intermedio

Tabla 96: Diccionario variables segmento Intermedio

<b>Variable</b>	<b>Significado</b>
<b>dif_PORC_TIPO_N_AVAN_CUOTAS_per_t0_y_t1</b>	(Porcentaje del total de transacciones hechas en avance en varias cuotas en t0) - (Porcentaje del total de

	transacciones hechas en avance en varias cuotas en t1)
<b>PAT_PER_t0</b>	Número de Pagos Automáticos de Tarjeta(PAT) contratados en el período t0
<b>PAT_PER_t1</b>	Número de Pagos Automáticos de Tarjeta(PAT) contratados en el período t1 (t1 es anterior a t0)
<b>PORC_RUBRO_N_COMBUSTIBLES_per_t0</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro COMBUSTIBLES en el período t0
<b>PORC_RUBRO_N_OTROS_CONSUMOS_per_t1</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro OTROS CONSUMOS en el período t1 (t1 es anterior a t0)
<b>PORC_RUBRO_N_RECAUDACION_per_t0</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro RECAUDACION en el período t0
<b>PORC_RUBRO_N_SUPERMERCADOS_per_t0</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro SUPERMERCADOS en el período t0
<b>prom_deu_vig_SBIF_per_t0</b>	Promedio de deuda vigente en el sistema financiero (SBIF) en el período t0
<b>prom_deu_vig_SBIF_per_t1</b>	Promedio de deuda vigente en el sistema financiero (SBIF) en el período t1 (t1 es anterior a t0)
<b>prom_deuda_consumo_per_t0</b>	Promedio deuda consumo contratada en t0
<b>prom_tar_cuo_per_t0</b>	Promedio cupo tarjeta cuota fija ocupado en el período t0
<b>SUMA_RUBROS_COMPRA_t0</b>	Suma de los distintos rubros donde la persona realizó transacciones en el

	período t0
<b>TARJ_CRED_per_t0</b>	Si contrató tarjeta de crédito en el período t0

## Modelo Segmento Alto consumo

Tabla 97: Diccionario variables segmento Alto consumo

<b>Variable</b>	<b>Significado</b>
<b>cta_cte</b>	Si tiene contratada una cuenta corriente (Dummy)
<b>PORC_RUBRO_N_RESTAURANTES_per_t0</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro RECAUDACION en el período t0
<b>PORC_RUBRO_N_SUPERMERCADOS_per_t0</b>	Porcentaje de las transacciones de un cliente que fueron realizadas en el rubro SUPERMERCADOS en el período t0
<b>prom_deu_cons_be_per_t0</b>	Promedio de deuda de consumo en BancoEstado en el período t0
<b>SUMA_RUBROS_COMPRA_t0</b>	Suma de los distintos rubros donde la persona realizó transacciones en el período t0
<b>tarj_cred</b>	Si la tarjeta de crédito es normal (Dummy)
<b>TARJ_CRED_per_t0</b>	Si contrató tarjeta de crédito en el período t0