



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

**ESTIMACIÓN DE CUSTOMER LIFETIME VALUE A NIVEL  
DE CLIENTES USANDO VARIABLES  
SOCIO-DEMOGRÁFICAS Y TRANSACCIONALES**

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

**HA NUI BAEK**

**PROFESOR GUÍA:**  
**LUIS ABURTO LAFOURCADE**

**MIEMBROS DE LA COMISIÓN:**  
**RICHARD WEBER HAAS**  
**PABLO MARÍN VICUÑA**

**SANTIAGO DE CHILE**  
**OCTUBRE 2006**

## ÍNDICE DE CONTENIDOS

1.1. ANTECEDENTES GENERALES .....	10
1.2. JUSTIFICACIÓN Y DESCRIPCIÓN DEL PROYECTO .....	14
1.3. OBJETIVOS .....	15
1.3.1. <i>Objetivo general</i> .....	15
1.3.2. <i>Objetivos específicos</i> .....	15
1.4. ALCANCES.....	15
1.5. RESULTADOS ESPERADOS.....	16
<b>2. PLAN METODOLÓGICO.....</b>	<b>17</b>
<b>3. MARCO CONCEPTUAL.....</b>	<b>19</b>
3.1. DEFINICIONES .....	19
3.2. MODELOS PREDICTIVOS.....	20
3.2.1. <i>Media Incondicional</i> .....	21
3.2.2. <i>Media Condicional</i> .....	21
3.2.3. <i>Modelo Simple</i> .....	22
3.2.4. <i>Pareto/NBD</i> .....	23
3.2.4.1. <i>Calibración del modelo</i> .....	23
3.2.5. <i>Alternativas para el cálculo del CLV</i> .....	25
3.2.6. <i>Validación de los modelos</i> .....	26
3.3. TEST ANOVA .....	28
3.4. CLUSTERING .....	29
3.5. ÁRBOLES DE DECISIÓN .....	31
<b>4. APLICACIÓN DE LA METODOLOGÍA PARA EL CÁLCULO DEL CUSTOMER LIFETIME VALUE A NIVEL DE CLIENTES.....</b>	<b>32</b>
4.1. SELECCIÓN DE DATOS.....	32
4.2. LIMPIEZA Y PRE-PROCESAMIENTO DE DATOS.....	33
4.2.1. <i>Variables transaccionales</i> .....	33
4.2.2. <i>Variables socio-demográficas</i> .....	36
4.2.3. <i>Conclusiones de la limpieza y pre-procesamiento de datos</i> .....	39
4.3. ANÁLISIS DESCRIPTIVO .....	40
4.3.1. <i>Análisis general</i> .....	41
4.3.2. <i>Análisis por variables</i> .....	45
4.3.3. <i>Conclusiones del análisis descriptivo</i> .....	50
4.3.3.1. <i>Conclusiones del análisis general</i> .....	50
4.3.3.2. <i>Conclusiones del análisis por variables</i> .....	51
4.4. TRANSFORMACIÓN DE DATOS .....	51
4.5. SELECCIÓN DE VARIABLES.....	54

4.5.1. <i>Aplicación del test Anova</i> .....	54
4.5.2. <i>Generación de un árbol de decisión</i> .....	55
4.6. ESTIMACIÓN DEL CUSTOMER LIFETIME VALUE BASADO SÓLO EN ATRIBUTOS RFM .....	56
4.6.1. <i>Estimación del número de transacciones</i> .....	57
4.6.1.1. <i>Presencia sólo de clientes activos</i> .....	58
4.6.1.2. <i>Incorporación de clientes inactivos y aplicación de una constante</i> .....	61
4.6.1.2.1. <i>Obtención de la constante</i> .....	62
4.6.1.2.2.1. <i>Análisis por quintiles</i> .....	65
4.6.1.3. <i>Conclusiones de la estimación del número de transacciones</i> .....	67
4.6.2. <i>Estimación del monto por transacción</i> .....	68
4.6.3. <i>Cálculo del Customer Lifetime Value (CLV)</i> .....	70
4.6.3.1. <i>Error asociado al cálculo del CLV</i> .....	71
4.6.3.2. <i>Ranking de clientes</i> .....	72
4.6.4. <i>Conclusiones del uso sólo de los atributos RFM</i> .....	73
4.6.4.1. <i>Conclusiones de la estimación del número de transacciones</i> .....	73
4.6.4.2. <i>Conclusiones del cálculo del CLV</i> .....	74
4.7. ESTIMACIÓN DEL CUSTOMER LIFETIME VALUE USANDO VARIABLES DEMOGRÁFICAS Y TRANSACCIONALES .....	74
4.7.1. <i>Estimación del número de transacciones</i> .....	74
4.7.1.1. <i>Presencia sólo de clientes activos</i> .....	74
4.7.1.2. <i>Incorporación de clientes inactivos y aplicación de una constante</i> .....	78
4.7.1.2.1. <i>Obtención de la constante</i> .....	78
4.7.1.2.2. <i>Análisis por quintiles</i> .....	83
4.7.1.3. <i>Conclusiones de la estimación del número de transacciones</i> .....	87
4.7.2. <i>Estimación del monto por transacción</i> .....	87
4.7.3. <i>Estimación del monto total</i> .....	90
4.7.3.1. <i>Conclusiones de la estimación del monto total</i> .....	92
4.7.4. <i>Cálculo del CLV</i> .....	93
4.7.4.1. <i>Error asociado al cálculo del CLV</i> .....	98
4.7.4.2. <i>Ranking de clientes</i> .....	98
4.7.5. <i>Cambios esperados del comportamiento transaccional</i> .....	101
4.7.6. <i>Conclusiones</i> .....	102
4.7.6.1. <i>Conclusiones de la estimación del número de transacciones</i> .....	102
4.7.6.2. <i>Conclusiones del cálculo del CLV</i> .....	102
4.8. ANÁLISIS COMPARATIVO ENTRE LAS METODOLOGÍAS UTILIZADAS .....	103
4.8.1. <i>Estimación del número de transacciones</i> .....	103
4.8.2. <i>Cálculo del Customer Lifetime Value (CLV)</i> .....	104
<b>5. CONCLUSIONES FINALES</b> .....	<b>105</b>
5.1. <i>CONCLUSIONES DE LOS RESULTADOS OBTENIDOS</i> .....	105
5.2. <i>TRABAJOS FUTUROS</i> .....	106

<b>6. REFERENCIAS BIBLIOGRÁFICAS</b> .....	<b>107</b>
<b>7. ANEXOS</b> .....	<b>108</b>
ANEXO1: DISTRIBUCIÓN BINOMIAL NEGATIVA (NBD).....	108
ANEXO 2: FUNCIÓN DE VEROSIMILITUD.....	109
ANEXO 3: RESULTADOS DEL TEST ANOVA.....	110
ANEXO 4: PRONÓSTICO DEL NÚMERO DE TRANSACCIONES APLICANDO LA METODOLOGÍA DE ESTIMACIÓN QUE CONSIDERA TANTO LAS VARIABLES SOCIO-DEMOGRÁFICAS COMO LAS TRANSACCIONALES.....	113
ANEXO 5: MPE DEL PRONÓSTICO DE $X$ POR QUINTILES.....	117
ANEXO 6: EFECTO DE LA ESTIMACIÓN DE $X$ CONSIDERANDO EL MODELO SIMPLE MODIFICADO.....	118

## ÍNDICE DE FIGURAS

FIGURA 1: EFECTOS TEÓRICOS DEL CRM.....	11
FIGURA 2: CRECIMIENTO DE LA UTILIDAD EN EL TIEMPO GENERADA POR LOS CLIENTES .....	12
FIGURA 3: DEFINICIÓN DE RFM.....	19
FIGURA 4: RESULTADOS DEL TEST ANOVA.....	28
FIGURA 5: PROCESO DE AGRUPACIÓN K-MEANS .....	30
FIGURA 6: EJEMPLO DE UN ÁRBOL DE DECISIÓN .....	31
FIGURA 7: DISTRIBUCIÓN DEL NÚMERO DE TRANSACCIONES.....	33
FIGURA 8: OUTLIERS SEGÚN EL MONTO POR TRANSACCIÓN .....	34
FIGURA 9: DISTRIBUCIÓN DE CLIENTES SEGÚN VARIABLE ANTIGÜEDAD .....	35
FIGURA 10: DISTRIBUCIÓN DE LA VARIABLE ESTADO DE LA TARJETA .....	36
FIGURA 11: DISTRIBUCIÓN DE LA VARIABLE GÉNERO .....	36
FIGURA 12: DISTRIBUCIÓN DE LA VARIABLE ESTADO CIVIL .....	37
FIGURA 13: DISTRIBUCIÓN DE LA VARIABLE REGIÓN.....	37
FIGURA 14: DISTRIBUCIÓN DE LA VARIABLE EDAD .....	39
FIGURA 15: EVOLUCIÓN DEL NÚMERO DE CLIENTES ACTIVOS .....	41
FIGURA 16: EVOLUCIÓN DEL NÚMERO DE TRANSACCIONES TOTAL Y PROMEDIO .....	42
FIGURA 17: EVOLUCIÓN DEL MONTO PROMEDIO Y TOTAL.....	43
FIGURA 18: DISTRIBUCIÓN DEL MONTO TOTAL PROMEDIO.....	43
FIGURA 19: DISTRIBUCIÓN DEL PERÍODO DE COMPRA ( $T_x$ ).....	44
FIGURA 20: DISTRIBUCIÓN DEL PERÍODO DE OBSERVACIÓN (T).....	44
FIGURA 21: DISTRIBUCIÓN DEL NÚMERO DE TRANSACCIONES, CLIENTES Y, MONTO POR GÉNERO.....	45
FIGURA 22: DISTRIBUCIÓN DEL N° DE TRANSACCIONES Y MONTO PROMEDIO POR TRANSACCIÓN SEGÚN GÉNERO..	46
FIGURA 23: EVOLUCIÓN DEL NÚMERO DE TRANSACCIONES PROMEDIO SEGÚN GÉNERO .....	46
FIGURA 24: EVOLUCIÓN DEL MONTO PROMEDIO POR TRANSACCIÓN SEGÚN GÉNERO .....	47
FIGURA 25: DISTRIBUCIÓN DEL NÚMERO DE TRANSACCIONES, CLIENTES Y MONTO POR ESTADO CIVIL.....	47
FIGURA 26: MONTO POR TRANSACCIÓN Y N° TRANSACCIONES PROMEDIO POR EDAD .....	48
FIGURA 27: MONTO POR TRANSACCIÓN Y N° DE TRANSACCIONES PROMEDIO POR ANTIGÜEDAD .....	49
FIGURA 28: N° DE TRANSACCIONES PROMEDIO SEGÚN GÉNERO Y ESTADO CIVIL .....	50
FIGURA 29: MONTO POR TRANSACCIÓN PROMEDIO SEGÚN GÉNERO Y ESTADO CIVIL .....	50
FIGURA 30: ÁRBOL DE REGRESIÓN GENERADO .....	55
FIGURA 31: PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	59
FIGURA 32: MAPE DEL NÚMERO DE TRANSACCIONES.....	60
FIGURA 33: DIAGRAMA DE APLICAR UNA CONSTANTE .....	62
FIGURA 34: MAPE DEL NÚMERO DE TRANSACCIONES.....	64
FIGURA 35: MAPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-TX.....	65
FIGURA 36: MAPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-T.....	66
FIGURA 37: MAPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-M .....	67
FIGURA 38: PRONÓSTICO DEL MONTO POR TRANSACCIÓN POR QUINTIL DE M.....	69

FIGURA 39: MAPE POR QUINTIL DE M .....	69
FIGURA 40: BOXPLOT DEL CLV.....	71
FIGURA 41: PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	75
FIGURA 42: MPE DEL NÚMERO DE TRANSACCIONES.....	76
FIGURA 43: MAPE DEL NÚMERO DE TRANSACCIONES.....	77
FIGURA 44: PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	77
FIGURA 45: MAPE DEL NÚMERO DE TRANSACCIONES.....	78
FIGURA 46: PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	80
FIGURA 47: MPE DEL NÚMERO DE TRANSACCIONES.....	81
FIGURA 48: MAPE DEL NÚMERO DE TRANSACCIONES.....	81
FIGURA 49: PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	82
FIGURA 50: MAPE DEL NÚMERO DE TRANSACCIONES.....	83
FIGURA 50: MAPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-TX.....	84
FIGURA 51: MPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-TX.....	84
FIGURA 52: MAPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-T.....	85
FIGURA 53: MAPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-M .....	86
FIGURA 54: PRONÓSTICO DEL MONTO POR TRANSACCIÓN POR QUINTIL DE M.....	88
FIGURA 55: MPE POR QUINTIL DE M.....	89
FIGURA 56: MAPE POR QUINTIL DE M .....	89
FIGURA 57: PRONÓSTICO DEL MONTO TOTAL .....	90
FIGURA 58: MPE POR QUINTIL DEL MONTO TOTAL.....	91
FIGURA 59: MAPE POR QUINTIL DEL MONTO TOTAL.....	92
FIGURA 60: BOXPLOT DEL CLV.....	93
FIGURA 61: CLV PROMEDIO POR QUINTIL DE X .....	94
FIGURA 62: CLV PROMEDIO POR QUINTIL DE TX.....	95
FIGURA 63: CLV PROMEDIO POR QUINTIL DE T .....	95
FIGURA 64: CLV PROMEDIO POR QUINTIL DE M.....	95
FIGURA 65: CLV PROMEDIO POR GÉNERO .....	96
FIGURA 66: CLV PROMEDIO POR ESTADO CIVIL .....	96
FIGURA 67: CLV PROMEDIO POR QUINTIL DE EDAD .....	97
FIGURA 68: CLV PROMEDIO POR QUINTIL DE ANTIGÜEDAD .....	97
FIGURA 69: CLV PROMEDIO POR QUINTIL DE REGIÓN .....	97
FIGURA 70: DISTRIBUCIÓN DE CLIENTES POR VARIABLES TRANSACCIONALES .....	99
FIGURA 71: DISTRIBUCIÓN DE CLIENTES POR VARIABLES SOCIO-DEMOGRÁFICAS.....	100
FIGURA 72: CAMBIOS ESPERADOS POR QUINTIL DE X.....	101
FIGURA 73: CAMBIO ESPERADO POR QUINTIL DE MONTO TOTAL.....	101
FIGURA 74: CAMBIO ESPERADO POR QUINTIL DE CLV .....	102
FIGURA 75: MPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-T .....	117
FIGURA 76: MPE DEL NÚMERO DE TRANSACCIONES POR QUINTIL DE X-M .....	117
FIGURA 77: MAPE DEL NÚMERO DE TRANSACCIONES.....	118

## ÍNDICE DE TABLAS

TABLA 1: EJEMPLO DE VALORES DE INDICADORES .....	25
TABLA 2: RESUMEN DE DATOS.....	32
TABLA 3: VALORES DE LA VARIABLE ANTIGÜEDAD.....	35
TABLA 4: VALORES DE LA VARIABLE EDAD EXPRESADA EN AÑOS .....	38
TABLA 5: RESUMEN INDICADORES TRANSACCIONALES SEGÚN EDAD .....	38
TABLA 6: RESUMEN DE DATOS ELIMINADOS .....	39
TABLA 7: RESUMEN DE DATOS SELECCIONADOS .....	40
TABLA 8: RESUMEN DE INFORMACIÓN DISPONIBLE.....	40
TABLA 9: RESUMEN DE INDICADORES MENSUAL.....	41
TABLA 11: RESUMEN INDICADORES TRANSACCIONALES SEGÚN ESTADO CIVIL .....	48
TABLA 12: RESUMEN INDICADORES TRANSACCIONALES SEGÚN EDAD .....	48
TABLA 13: RESUMEN INDICADORES TRANSACCIONALES SEGÚN ANTIGÜEDAD.....	49
TABLA 14: VARIACIÓN DEL MONTO POR TRANSACCIÓN Y TOTAL ENTRE EL AÑO 2003 Y 2004 .....	50
TABLA 15: CATEGORÍAS SEGÚN REGIÓN .....	52
TABLA 16: CATEGORÍAS SEGÚN REGIÓN.....	52
TABLA 17: MÉTODO K-MEANS PARA LA VARIABLE EDAD.....	53
TABLA 18: MÉTODO K-MEANS PARA LA VARIABLE ANTIGÜEDAD .....	53
TABLA 19: TEST ANOVA PARA LA VARIABLE INDEPENDIENTE GÉNERO.....	54
TABLA 20: ERROR ASOCIADO AL ÁRBOL .....	56
TABLA 21: PESOS DE CADA UNA DE LAS VARIABLES .....	56
TABLA 22: PERÍODOS DE OBSERVACIÓN .....	57
TABLA 23: RESUMEN DE DATOS RFM .....	57
TABLA 24: PERÍODOS DE OBSERVACIÓN .....	58
TABLA 25: RESUMEN DE DATOS .....	58
TABLA 26: PARÁMETROS NECESARIOS PARA LA CALIBRACIÓN .....	58
TABLA 27: RESUMEN DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES.....	60
TABLA 28: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES .....	61
TABLA 29: PERÍODOS DE OBSERVACIÓN .....	61
TABLA 30: PERÍODOS DE OBSERVACIÓN .....	62
TABLA 31: ERROR DEL PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	62
TABLA 32: VALORES DE LA CONSTANTE .....	63
TABLA 33: ERROR DEL PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	63
TABLA 34: RESUMEN DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES.....	63
TABLA 35: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES .....	64
TABLA 36: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN ESTIMACIÓN DE M .....	70
TABLA 37: RESUMEN DEL CLV ANUALIZADO.....	71
TABLA 38: RESUMEN DEL MAPE .....	72
TABLA 39: ESTIMACIÓN DEL MAPE DEL CLV .....	72

TABLA 43: RESUMEN DE DATOS .....	74
TABLA 44: ERROR DEL PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	79
TABLA 45: VALORES DE LA CONSTANTE .....	79
TABLA 46: ERROR DEL PRONÓSTICO DEL NÚMERO DE TRANSACCIONES .....	79
TABLA 47: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES .....	82
TABLA 48: RESUMEN DEL MONTO POR TRANSACCIÓN.....	88
TABLA 49: RESUMEN DEL ERROR PORCENTUAL EN ESTIMACIÓN DEL MONTO TOTAL.....	91
TABLA 50: RESUMEN DE CLV ANUALIZADO .....	94
TABLA 51: RESUMEN DEL MAPE .....	98
TABLA 52: ESTIMACIÓN DEL MAPE DEL CLV .....	98
TABLA 57: TEST DE ANOVA PARA LA VARIABLE INDEPENDIENTE ESTADO CIVIL.....	110
TABLA 58: TEST DE ANOVA PARA LA VARIABLE INDEPENDIENTE EDAD.....	110
TABLA 59: TEST DE ANOVA PARA LA VARIABLE INDEPENDIENTE REGIÓN .....	111
TABLA 60: TEST DE ANOVA PARA LA VARIABLE INDEPENDIENTE ANTIGÜEDAD .....	111
TABLA 61: TEST DE ANOVA PARA LA VARIABLE INDEPENDIENTE DÍA DE PAGO .....	111
TABLA 62: TEST DE ANOVA PARA LA VARIABLE INDEPENDIENTE ESTADO DE LA TARJETA.....	112
TABLA 63: RESUMEN DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES.....	113
TABLA 64: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES .....	113
TABLA 65: RESUMEN DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES.....	114
TABLA 66: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES .....	114
TABLA 67: RESUMEN DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES.....	114
TABLA 68: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES .....	115
TABLA 69: RESUMEN DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES.....	115
TABLA 70: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES .....	115
TABLA 71: RESUMEN DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES.....	116
TABLA 72: RESUMEN DEL VALOR ABSOLUTO DEL ERROR PORCENTUAL EN EL NÚMERO DE TRANSACCIONES .....	116
TABLA 73: ERRORES DE PRONÓSTICO DEL MODELO SIMPLE .....	119
TABLA 74: ERRORES DE PRONÓSTICO DEL MODELO SIMPLE MODIFICADO.....	119



*A mis padres*  
*A mi hermana Soun*  
*A mi hermanita Andrea Soji*  
*A mi hermanita Angélica Hunhe*

## *Agradecimientos*

*Primero que todo quiero dar gracias a Dios por todo lo que me ha dado, por estar siempre conmigo en las buenas y en las malas, y por ser la Persona en quien puedo confiar plenamente sin miedos ni reproches.*

*Agradezco a mi familia, quienes con esfuerzo, me han dado todo lo necesario para alcanzar mis metas y acercarme cada vez a cumplir mi sueño.*

*Agradezco a mi hermana Soun, quien ha sido un gran apoyo este último tiempo, y he aprendido a conocerla mejor ahora que hemos madurado juntas como hermanas.*

*Agradezco a mis dos hermanitas bebés, Andrea Soji y Angélica Hunhe, quienes han crecido tanto, por darme ánimo y fuerzas para seguir adelante en los momentos de cansancio y angustia.*

*Agradezco a mis amigas y amigos por acompañarme en este camino y espero poder seguir junto a ellos por mucho tiempo más.*

# 1. INTRODUCCIÓN

## 1.1. Antecedentes generales

En el pasado, pocas empresas valoraban a sus clientes y, ante una economía de expansión y mercados de rápido crecimiento, las empresas adoptaban un marketing de *balde con fugas*, es decir, sólo se dedicaban a captar nuevos clientes, llenar y llenar el balde sin percatarse que por alguna parte de éste existía una fuga que ocasionaba el abandono de los clientes más antiguos [13].

Hoy en día, la situación es distinta, los cambios socio-demográficos, el crecimiento de la economía y la creciente innovación de la competencia implican una escasez de clientes, por lo cual las firmas están luchando por su participación de mercado haciendo elevar los costos de atraer nuevos clientes [12]. Atraer nuevos clientes sigue siendo una tarea importante, sin embargo, las empresas deben concentrarse también en conservar los clientes actuales y forjar relaciones rentables y duraderas con ellos en el tiempo. Es por esto que surge el concepto de *Customer Relationship Management (CRM)*, que se refiere a la administración de todas las interacciones que pueden tener un negocio y sus clientes, enfocándose en la optimización del ciclo de vida del cliente en su totalidad. Además, CRM es un término de la industria de la información que reúne metodologías y software para administrar de una manera eficiente y rentable las relaciones de un negocio con sus clientes [13].

Existen 3 tipos de aplicaciones de CRM [12]:

- Aplicaciones CRM operativas: éstas tienen como objetivo la automatización del contacto con el cliente mediante la coordinación entre el front-office y las funciones del back-office con el fin de mejorar la interacción con el cliente.
- Aplicaciones CRM analíticas: es una herramienta de explotación y análisis de la información acerca del cliente, que ayuda a comprender tanto lo que quieren sus clientes como su comportamiento, además facilita la adquisición de nuevos clientes y retener los presentes.
- Aplicaciones CRM cooperativas: se centra en la difusión del conocimiento en la compañía y corresponde a la gestión de los distintos canales de relación con los clientes.

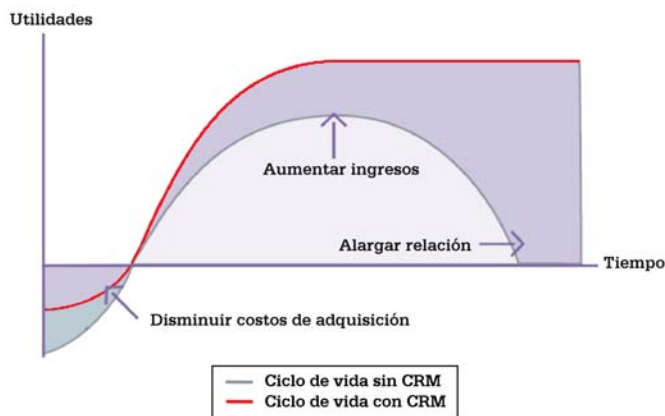
El estudio a realizarse se enmarca en el CRM analítico, ya que se pretende detectar patrones de comportamiento de compra de los clientes. Para ello, es necesario conocer las necesidades de los clientes a lo largo de su ciclo de vida, el cual puede ser visto como un proceso de 3 etapas:

- Etapa 1: Adquisición del cliente  
Se invierten recursos para que el cliente comience una relación con la firma concretándose en una primera adquisición del bien ofertado por la empresa.
- Etapa 2: Desarrollo de la relación con el cliente  
Se ofrecen productos y/o servicios adecuados a las necesidades de éste.
- Etapa 3: Retención del cliente  
Se invierten recursos para que éste continúe la relación con la empresa.

Los efectos esperados de la aplicación del CRM sobre el ciclo de vida se muestran a continuación:

- Disminuir los costos de adquisición
- Aumentar los ingresos
- Alargar la relación cliente-empresa

*Figura 1: Efectos teóricos del CRM*



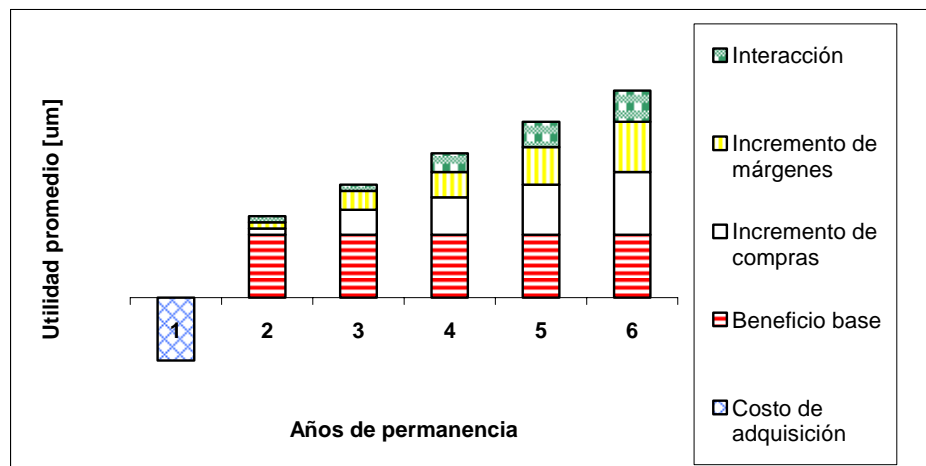
Conocer al cliente es lo más importante para una buena aplicación del CRM, ya que es fundamental conocer sus hábitos de compra, la frecuencia con que lo hace, la cantidad de productos adquiridos, sus gustos, preferencias entre otros. Todo lo anterior con el fin de identificarlos e individualizarlos y, así, hacer que se conviertan en un objetivo de gran valor para la empresa.

Es importante saber diferenciar a los clientes, ya que no todos son iguales y, es aquí donde la compañía tiene que saber reconocerlos debido a que existen clientes que llegan

a generar un mayor margen de compra y ganancia en comparación a otros. Son a estas personas a las que se les cuidará, enviándole información actualizada, invitándolos a eventos especiales y realizando un sin fin de actividades personalizadas, que generarán un nivel de confianza y lealtad por parte del cliente, haciendo que éste se sienta importante y mejore sus hábitos de compra [13].

Inicialmente, las empresas se ven obligadas a realizar inversiones para obtener clientes, pero una vez que se han realizado ventas de los primeros productos o servicios comienzan a percibir rentabilidad. A medida que transcurre el tiempo y la compañía gana experiencia con sus clientes, empieza a ser capaz de servirles con mayor eficiencia. Los clientes, al sentirse satisfechos<sup>1</sup>, darán buenas referencias de la institución a otros potenciales clientes y finalmente estarán dispuestos a pagar más por los bienes dada la confianza hacia la empresa [3].

*Figura 2: Crecimiento de la utilidad en el tiempo generada por los clientes*



*Fuente: Zero defections: Quality comes to service. Harvard Business Review, Sept - Oct 1990*

Este proceso se observa en la figura 2, donde:

- Costo de adquisición: todo negocio tiene que invertir dinero por adelantado para atraer nuevos clientes.
- Beneficio base: corresponde a la utilidad generada por los bienes adquiridos por el cliente.
- Incremento de compras: aumento en ventas.
- Incremento de márgenes: mientras más conocen a los clientes, la firma se vuelve más eficiente, siendo capaz de identificar las necesidades de éstos y, así, ofrecer

<sup>1</sup> Satisfacción del cliente se refiere al grado en que el desempeño percibido por un producto o servicio concuerda con las expectativas del comprador. Si el desempeño iguala o supera las expectativas, el cliente estará satisfecho. En caso contrario, quedará insatisfecho.

los bienes que las satisfagan.

- *Interacción*: los clientes satisfechos recomiendan el negocio a otros, se produce el *efecto boca-oreja*. Además, los clientes antiguos están más dispuestos a pagar precios más altos.

De la figura 2, se observa que a medida que aumenta la antigüedad de la relación cliente-empresa, las ganancias percibidas también aumentan. Por otro lado, cuando un cliente queda insatisfecho, éste realiza 11 malas opiniones respecto de la organización a potenciales futuros clientes [7], por lo que conocer sus necesidades es de importancia para ofrecer los bienes correctos para cada cliente.

En este entorno es fundamental revisar el grado de vinculación del cliente con la empresa y ser capaces de mantenerlo fiel. Por ello, el objetivo es transformar al cliente de valor, real o potencial, en un cliente rentable y leal. Para conseguir lo anterior se debe gestionar toda la información que se tiene del cliente, detectar el grado y las causas de su vinculación y conocer el valor que aporta para la compañía para, finalmente, tomar la decisión de fidelizarlo o eliminarlo. Es aquí donde el concepto de *customer lifetime value (CLV)* es de gran utilidad, ya que este indicador permite distinguir entre los clientes más y menos valiosos para la empresa, además de servir como referencia para dimensionar campañas de marketing, estudiar elementos de interacción y/o revisar propuestas de valor de los productos de la compañía, crear oportunidades de cross-selling<sup>2</sup>, up-selling<sup>3</sup> y abrir la posibilidad a una rápida introducción de nuevos productos o marcas.

Experiencias anteriores han obtenido que la estimación del CLV basado en el comportamiento transaccional de los clientes a través de los atributos RFM (Recency, Frecuency y Monetary Value) entrega buenos resultados<sup>4</sup> al observar segmentos de clientes pero no a nivel individual [6], lo cual indica la presencia de otros factores que afectan el comportamiento de compra de los clientes. Precisamente, uno de los objetivos de este estudio es encontrar variables que incorporen información útil para la estimación del CLV a nivel de clientes. Las alternativas de cálculo del CLV se detallan en el capítulo 3, Marco conceptual.

---

<sup>2</sup> Cross-selling se refiere a la estrategia de venta cruzada con otros productos relacionados con el principal.

<sup>3</sup> Up-selling se refiere a la venta cruzada con productos de margen superior al principal.

<sup>4</sup> El error cometido por el modelo Pareto/NBD en la estimación del número de transacciones es de 6 unidades, que representan a un 37,7% del número de transacciones promedio.

## 1.2. Justificación y descripción del proyecto

Los clientes son la fuerza dominante en la red económica de hoy. La batalla por el liderazgo del mercado, la pueden ganar aquellos que tienen un conocimiento profundo y dinámico de las necesidades de éstos, así como sus patrones de comportamiento y de lo que se puede desarrollar con él para, así, generar relaciones de largo plazo basadas en la lealtad de los clientes [13].

Dado lo anterior, las empresas están interesadas en conocer cuánto vale y cuál va a ser el comportamiento de compra futuro de cada uno de sus clientes para determinar qué clientes son valiosos y cuáles no y, así, administrar y orientar de mejor manera sus esfuerzos. Sin embargo, se desea que el valor del CLV no sea cualquiera sino que un valor confiable<sup>5</sup>. Dado que no se conoce el valor real del CLV, la confianza de éste se determina por la confianza en la estimación del número de transacciones, es decir, si la predicción del número de transacciones es confiable, se asumirá que la predicción del CLV también lo es.

Como se mencionó en el capítulo anterior, los atributos RFM no explican completamente el comportamiento futuro de los clientes ya que existen variables subyacentes no consideradas. Por ejemplo, si se tienen 2 clientes, uno doctor y el otro profesor, con el mismo comportamiento de compra, se esperaría que el valor del CLV del cliente con profesión doctor sea mayor que el del profesor, debido a las expectativas de ingreso futuro.

Experiencias anteriores indican que la predicción del número de transacciones del modelo Pareto/NBD se realiza a nivel de segmentos con un error de 38% en comparación a un 61% a nivel individual [6]. Luego, este modelo predice mejor a nivel agregado.

La cartera de clientes en estudio corresponde a los clientes de la tarjeta de crédito de una empresa minorista de retail. El presente proyecto tiene como objetivo calcular el *customer lifetime value* a nivel de clientes para el caso en estudio. Para ello, se usarán variables tanto socio-demográficas como transaccionales. Lo relevante y, a la vez, difícil consiste en la determinación de las variables a incorporar en el modelo. Luego, se construirá una metodología para la selección de variables a incluir y no incluir en el modelo y para la estimación del *customer lifetime value* a nivel de clientes.

---

<sup>5</sup> La confianza se refiere a que el valor estimado se encuentre dentro de un rango de error pequeño con respecto al valor real.

## 1.3. Objetivos

### 1.3.1. *Objetivo general*

Construir y validar una metodología para el cálculo de *customer lifetime value* a nivel de clientes usando variables socio-demográficas y transaccionales.

### 1.3.2. *Objetivos específicos*

- Calcular el CLV aplicando la metodología seguida por Hardie, Fader y Lee [1].
- Definir una metodología de evaluación de las variables socio-demográficas y, a partir de ésta, decidir qué variables se incluirán o no en el modelo.
- Modificar el modelo de estimación del CLV basado en variables transaccionales, incorporando las variables socio-demográficas seleccionadas.
- Calcular el CLV individual mediante la metodología que incorpora variables tanto socio-demográficas como transaccionales.
- Generar reglas de negocio de la información disponible de los clientes.

## 1.4. Alcances

Los alcances para el desarrollo de este trabajo son los siguientes:

- La calibración y validación de los modelos se realizarán sobre una base de datos de clientes de la tarjeta de crédito de una empresa de retail en particular.
- Las conclusiones serán válidas sólo para el caso en estudio.
- No se estudiarán otros casos o base de datos de clientes.
- No se levantarán datos para extraer nuevas variables no transaccionales. Sólo se utilizará la información disponible.
- No se realizarán propuestas de campañas de marketing.



## 1.5. Resultados esperados

Los resultados esperados del presente estudio se muestran a continuación:

- Valor del *Customer lifetime value* (CLV) a nivel de clientes utilizando la metodología seguida por Hardie, Fader y Lee [1].
- Una metodología para determinar las variables socio-demográficas que agregan y no agregan valor en la estimación del CLV.
- Una metodología para el cálculo del CLV usando información de variables socio-demográficas y transaccionales.
- Valor del *Customer lifetime value* (CLV) para cada uno de los clientes utilizando una metodología que incorpora tanto variables socio-demográficas como transaccionales.
- Documentación del efecto o impacto de la incorporación de variables no transaccionales y otros posibles factores en el cálculo del CLV.
- Reglas de negocio de los clientes de la tarjeta de crédito de la empresa de retail en estudio.

## 2. PLAN METODOLÓGICO

Los pasos a seguir para llevar a cabo este proyecto se muestran a continuación.

### 1. *Estudiar los datos del caso en estudio*

- Selección de datos.
- Pre-procesamiento y transformación de los datos.
- Análisis descriptivo de los datos y su comportamiento.
- Conclusiones generales de la base de datos.

### 2. *Determinación de variables socio-demográficas que agreguen valor a la estimación del customer lifetime value*

- Determinación de las variables socio-demográficas a incorporar y descartar en la metodología del cálculo de CLV a nivel de clientes.
- Aplicación del test Anova<sup>6</sup> sobre las posibles variables para determinar si éstas son significativas estadísticamente con respecto al número de transacciones realizadas por cada uno de los clientes.
- Una vez conocidas las variables significativas, se utiliza el método K-means<sup>7</sup> para identificar los grupos de clientes con respecto a cada una de las variables seleccionadas, es decir, categorizarlas.
- Luego, se genera un árbol de regresión<sup>8</sup> para determinar las variables más relevantes en la estimación del número de transacciones y, por lo tanto, en el cálculo del CLV.

### 3. *Correr el modelo de estimación del Customer Lifetime Value basado sólo en variables transaccionales*

- Determinación de las ventanas de tiempo a utilizar.
- Cálculo del CLV a nivel de clientes para la base de datos seleccionada.
- Análisis e interpretación de los resultados.
- Comparación de los modelos entre sí.
- Conclusiones.

---

<sup>6</sup> El test Anova se encuentra incluido en el software SPSS 11.5 utilizado para el desarrollo del presente estudio.

<sup>7</sup> El método K-means se encuentra incluido en el software SPSS 11.5 utilizado.

<sup>8</sup> La generación del árbol de regresión se realiza en Microsoft Data Engine.

**4. *Determinación de variables socio-demográficas que agreguen valor a la estimación del customer lifetime value***

- Determinación de las variables socio-demográficas a incorporar y descartar en la metodología del cálculo de CLV a nivel de clientes.
- Aplicación del test Anova<sup>9</sup> sobre las posibles variables para determinar si éstas son significativas estadísticamente con respecto al número de transacciones realizadas por cada uno de los clientes.
- Una vez conocidas las variables significativas, se utiliza el método K-means<sup>10</sup> para identificar los grupos de clientes con respecto a cada una de las variables seleccionadas, es decir, categorizarlas.
- Luego, se genera un árbol de regresión<sup>11</sup> para determinar las variables más relevantes en la estimación del número de transacciones y, por lo tanto, en el cálculo del CLV.

**5. *Construcción y validación de la metodología para el cálculo del CLV a nivel de clientes usando variables tanto categóricas como transaccionales***

- Determinación de las ventanas de tiempo a utilizar.
- Cálculo CLV a nivel individual.
- Validación.
- Análisis e interpretación de los resultados.
- Comparación de los modelos entre sí.
- Conclusiones.

**6. *Análisis comparativo entre incorporar o no variables no transaccionales***

- Comparar los errores de pronóstico de los modelos entre las dos metodologías.
- Interpretación de los resultados arrojados de la comparación.
- Conclusiones finales y futuras líneas de acción.

---

<sup>9</sup> El test Anova se encuentra incluido en el software SPSS 11.5 utilizado para el desarrollo del presente estudio.

<sup>10</sup> El método K-means se encuentra incluido en el software SPSS 11.5 utilizado.

<sup>11</sup> La generación del árbol de regresión se realiza en Microsoft Data Engine.

### 3. MARCO CONCEPTUAL

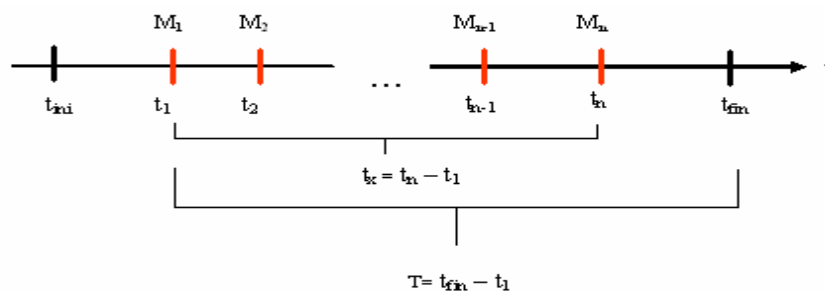
El marco conceptual se divide en cuatro partes: definiciones, modelos predictivos, test Anova, clustering y, finalmente, árboles de decisión.

#### 3.1. Definiciones

- Customer lifetime value (CLV): corresponde al valor presente de las utilidades esperadas que generará un cliente a lo largo de toda su relación con la empresa. Este indicador se calcula en base al comportamiento de compra pasada del cliente.
- Atributos RFM:
  - R (Recency) se refiere al tiempo transcurrido desde la última compra.
  - F (Frequency) corresponde a la frecuencia de compra.
  - M (Monetary value) corresponde al valor monetario de las compras.

Los atributos RFM son los más utilizados en la literatura para caracterizar el comportamiento de compra histórico del cliente. La definición de RFM<sup>12</sup> a utilizar para el desarrollo de este trabajo consta de 4 indicadores [5].

*Figura 3: Definición de RFM*



*Fuente: Elaboración propia*

Se tienen que:

$x$ : número de transacciones realizadas en el período  $t_{ini} - t_{fin}$ .

$t_x$ : período de compra del cliente, que corresponde al tiempo desde la primera compra hasta la última. Se define como  $t_x = t_n - t_1$ .

$T$ : período de observación del cliente, la cual se inicia cuando se realiza la primera

<sup>12</sup> Basado en la definición utilizada por Fader, Hardie y Lee en su modelo predictivo.

compra hasta el fin del período. Se define como  $T = t_{fin} - t_1$ .

$M$ : monto promedio por transacción durante el período  $t_{ini} - t_{fin}$ . Se define como

$$M = \frac{\sum_{i=1}^n M_i}{n}$$

El cálculo de estos indicadores es por cliente, luego cada cliente tiene asociada una tupla  $(x, t_x, T, M)$ .

- Caso contractual y no-contractual: los contractuales corresponden a casos en que existe un contrato de por medio en el cual se indica cuándo y cuánto pagar por algún servicio o bien, previamente fijado. Por ejemplo, en el caso de clientes de compañías de televisión por cable (VTR, Metrópolis entre otros). Los casos no-contractuales son lo contrario, por ejemplo, en el caso de clientes de tarjetas de crédito.

### 3.2. Modelos predictivos

Para la estimación del *Customer Lifetime Value (CLV)*, se consideran 4 modelos predictivos: Media Incondicional, Media Condicional, Modelo Simple y Pareto/NBD [3]. Estos modelos utilizan información histórica del comportamiento de compra del cliente para predecir su comportamiento futuro, principalmente, la probabilidad de seguir activo de un período a otro, el número de transacciones y monto promedio por transacción en el siguiente período, para luego estimar el valor del *customer lifetime value* para cada uno de los clientes. Esto es válido en casos no-contractuales.

El cálculo del *CLV* se realiza de la siguiente forma:

$$CLV = (\hat{M} - \hat{C}) \cdot VP(r_{anual}, \hat{x})$$

donde:

- $CLV$  = estimación de lifetime value.
- $\hat{M}$  = valor promedio de monto por transacción pronosticado.
- $\hat{C}$  = estimación del costo asociado a la transacción<sup>13</sup>.

---

<sup>13</sup> Para el desarrollo de este trabajo se considerará cero la estimación del costo asociado a la transacción, ya que no se posee esta información.

- $VP(r_{anual}, \hat{x}) = \sum_{t=1}^{\infty} \frac{\hat{x}}{(1+r_{per})^t} [P(activo)]^{t-1}$  = valor presente del número de transacciones futuras esperadas, dada una tasa de descuento anual  $r_{anual}$  y  $P(activo)$  corresponde a la probabilidad de permanecer activo de un período al siguiente.
- $\hat{x}$  = número de transacciones futuras pronosticadas.

Se requieren dos períodos: un período de calibración y otro de validación. En el de calibración se determinan los valores necesarios para el cálculo del *CLV* y en el de validación, se analiza la confiabilidad de las predicciones realizadas.

### 3.2.1. Media Incondicional

Este modelo asume que el comportamiento transaccional de un cliente es igual al promedio del total de clientes de calibración y su probabilidad de permanecer activo en el siguiente período es 1. No considera ni la fuga ni llegada de nuevos clientes.

Si el número de clientes en calibración es  $N$ , el valor del lifetime value del cliente es:

$$CLV = \bar{M}_{total} \cdot \sum_{t=1}^{\infty} \frac{\bar{x}_{total}}{(1+r_{per})^t} \cdot 1^{t-1}$$

Donde:  $\bar{M}_{total} = \frac{\sum_{j=1}^N M_j}{N}$  = valor promedio del monto promedio por transacción de los clientes en el período de calibración.

$\bar{x}_{total} = \frac{\sum_{j=1}^N x_j}{N}$  = número de transacciones promedio por cliente durante el período de calibración.

### 3.2.2. Media Condicional

Este modelo asume que el comportamiento de compra de un cliente en un período futuro será igual al promedio histórico, por lo tanto, el número de transacciones y monto promedio por transacción futuro corresponden al promedio histórico. Al igual que el modelo de Media Incondicional, no considera la "fuga" e "ingreso" de clientes.

Al considerar un análisis por segmentos que consta de  $N_s$  clientes cada grupo, el valor del *CLV* es tal que:

$$CLV = \bar{M}_s \cdot \sum_{t=1}^{\infty} \frac{\bar{x}_s}{(1 + r_{per})^t} \cdot 1^{t-1}$$

Donde:  $\bar{M}_s = \frac{\sum_{j=1}^{N_s} M_j}{N_s}$  = valor promedio del monto promedio por transacción del segmento de clientes en el período de calibración.

$\bar{x}_s = \frac{\sum_{j=1}^{N_s} x}{N_s}$  = número de transacciones promedio por cliente del segmento durante el período de calibración.

### 3.2.3. Modelo Simple

Este modelo considera los siguientes supuestos:

- El número de transacciones futuras ( $\hat{x}$ ) es proporcional a la que se observaría si se mantiene constante en el tiempo la frecuencia de compra del cliente.
- Un cliente tiene una mayor probabilidad de ser considerado activo en el período siguiente mientras más reciente haya sido su última compra.

Luego, el pronóstico para un período de largo *Teval* se tiene que:

- $\hat{x} = x \cdot \left( \frac{Teval}{T+1} \right) \cdot P(activo)$
- $P(activo) = \frac{t_x + 1}{T + 1}$

Por lo tanto el valor del cliente queda expresado como:

$$CLV = M \cdot \sum_{t=1}^{\infty} \frac{x}{(1 + r_{per})^t} \cdot \left( \frac{Teval}{T+1} \right) \left[ \frac{t_x + 1}{T+1} \right]^t$$

Este modelo considera la posibilidad de "fuga" de clientes y que un cliente "valioso" hoy no necesariamente lo será mañana.

### 3.2.4. Pareto/NBD

Se considera el modelo Pareto/NBD<sup>14</sup> adaptado por Fader, Hardie y Lee. Los supuestos de este modelo son los siguientes [1]:

- Hay independencia entre  $M$  y  $R$  y entre  $M$  y  $F$ .
- Los estados en que un cliente puede encontrarse son: activo o inactivo. Un cliente considerado inactivo en un período no puede volver al estado activo.
- Las compras se representan mediante un proceso de Poisson de tasa  $\lambda$ . Luego, el tiempo entre compras se distribuye según una exponencial de media  $1/\lambda$ .
- Se tiene que un cliente se encuentra activo por un período *no observable*  $\tau$ , la cual se distribuye según una exponencial de tasa  $\mu$ . Después de este período pasa del estado activo al inactivo.
- Las tasas de compra  $\lambda$  se distribuyen según una gamma de parámetros  $r$  (forma),  $\alpha$  (escala).
- Las tasas de fuga  $\mu$  se distribuyen según una gamma de parámetros  $s$  (forma),  $\beta$  (escala).

Se utilizan funciones gamma para modelar la distribución de las tasas de compra y fuga, ya que éstos se ajustan mejor a los datos observados debido a que el tiempo entre compras y el tiempo de inactividad de un cliente se distribuye según una función exponencial. Luego la idea es encontrar los parámetros  $r$ ,  $\alpha$ ,  $s$  y  $\beta$ , en vez de  $\lambda$  y  $\mu$ , para estimar el comportamiento individual de los clientes. Con este modelo, la probabilidad de estar activo aumenta con el momento en que se realizó la última compra considerando, además, la frecuencia de compra del cliente.

Este modelo necesita de dos períodos: un período de calibración y otro de validación.

#### 3.2.4.1. Calibración del modelo

La calibración de los parámetros necesarios para la estimación del *CLV* según el modelo Pareto/NBD,  $r$ ,  $\alpha$ ,  $s$  y  $\beta$ , se realiza por *máxima verosimilitud*. Se posee una muestra de  $N$  clientes, donde el cliente  $j$  ha realizado  $x_j$  transacciones durante el período  $(0, T_j]$ ,

siendo la última compra realizada en  $t_{x_j}$ . La función que se maximiza es el logaritmo natural de la función de verosimilitud y está dada por la siguiente expresión [1]:

---

<sup>14</sup> La definición de la distribución binomial negativa (NBD) se encuentra en el Anexo 1.



$$LL(r, \alpha, s, \beta) = \sum_{j=1}^N \ln \left[ L(r, \alpha, s, \beta / X_j = x_j, t_{x_j}, T_j) \right]$$

Donde:  $N$  = número de clientes en el período de calibración.

$x_j$  = número de transacciones realizadas por el  $j$ -ésimo cliente en el período de calibración.

$t_{x_j}$  = período de compra del  $j$ -ésimo cliente en el período de calibración.

$T_j$  = período de observación del  $j$ -ésimo cliente en el período de calibración.

Al maximizar la función de verosimilitud, se maximiza la probabilidad conjunta de observar los datos mediante una función gamma con parámetros de forma y escala determinados. La función de verosimilitud se explica detalladamente en el Anexo 2.

Se utiliza la función *fmincon* implementada en Matlab para encontrar los parámetros  $r$ ,  $\alpha$ ,  $s$  y  $\beta$  maximizando el logaritmo de la función de verosimilitud ( $LL$ ), o mejor dicho, minimizando  $-LL$ .

Luego, el valor del  $CLV$  es tal que:

$$CLV = \hat{M} \cdot DET$$

Donde:

- $\hat{M}$  = monto promedio por transacción. Es tal que:  $\hat{M} = \sum_{i=1}^N M_i$ .
- $VP(\delta, \hat{x}) = \frac{\alpha^r \beta^s \delta^{s-1} \Gamma(r+x+1) \Psi(s; s; \delta(\beta+T))}{\Gamma(r)(\alpha+T)^{r+x+1} L(r, \alpha, s, \beta / X = x, t_x, T)}$ , con  $\delta$  tasa de interés compuesta continuamente y  $\Psi(\cdot)$  es la función hipergeométrica confluyente de segundo orden.
- $\hat{x} = \frac{(r+x)(\beta+T)}{(\alpha+T)(s-1)} \left[ 1 - \left( \frac{\beta+T}{\beta+T+t} \right)^{s-1} \right] \cdot P(\text{activo})$  = número de transacciones esperadas.
- $P(\text{activo}) = \left\{ 1 + \left( \frac{s}{r+s+x} \right) (\alpha+T)^{r+x} (\beta+T)^s A_0 \right\}^{-1}$  = probabilidad de estar activo al período siguiente.

A continuación, se muestra un ejemplo con los valores que toman los indicadores para algunos clientes, bajo los siguientes parámetros de las funciones gama para las tasas de compra y las de fuga.

*Tabla 1: Ejemplo de valores de indicadores*

$r$	$\alpha$	$s$	$\beta$
7,87080	216,49010	0,66710	43,66840

<i>cliente</i>	$x$	$tx$	$T$	$M$	$\hat{x}$	$P(\text{activo})$	$DET$	$CLV$
30	2	42	69	2.403,5	3,2957	0,7406	8,2675	19.871,00
53	3	133	140	24.677	4,2484	0,9715	13,2914	327.992,95
.....								
1421	1	0	12	1.520	3,3218	0,8176	6,3851	9.705,28

De la anterior tabla, se observa que no siempre los clientes que realizan mayor número de transacciones son mejores clientes en cuanto a su frecuencia de compra. Por ejemplo, el cliente 1421 realizó una única transacción, sin embargo, el modelo Pareto/NBD pronostica mayor número de transacciones otorgándole, además, una mayor probabilidad de seguir activo durante el siguiente período en comparación al cliente 30, debido a que la compra fue realizada más recientemente en el caso del cliente 1421.

### 3.2.5. Alternativas para el cálculo del CLV

Otras alternativas para el cálculo del  $CLV$  son:

- Average lifetime value (ALV): es el valor promedio del lifetime value.

$$ALV = \frac{Monto_{total}}{N_{total}}$$

Donde:  $Monto_{total} = \sum_{j=11}^N Monto_j$  = corresponde al monto total en ventas generados

por clientes no nuevos.

$N_{total}$  = número total de clientes no nuevos.

Cada empresa determina qué tipo de clientes son considerados como nuevos.

- Monto promedio: otra alternativa es considerar que el ingreso total percibido por la empresa en cada período futuro es igual al ingreso promedio del período a considerarse. Luego se tiene que:

$$CLV = \sum_{t=1}^{\infty} \frac{\bar{M}}{(1+r_{per})^t} = \bar{M} \cdot \sum_{t=1}^{\infty} \frac{1}{(1+r_{per})^t}$$

Donde:  $\bar{M}$  = monto promedio por cliente.

$r_{per}$  = tasa de descuento compuesta anualmente.

- Otra forma es considerar que la probabilidad de que un cliente mantenga la relación con la empresa o tasa de retención del cliente podría ser estimada con la ayuda de determinantes empíricamente validados de lealtad como la satisfacción del cliente, barreras o costo de cambiarse de una empresa a otra entre otros [11], incorporando el hecho de que un cliente insatisfecho realiza 11 malas opiniones a posibles potenciales clientes.

### 3.2.6. Validación de los modelos

Como se mencionó anteriormente, se cuenta con un período para la calibración y otro para la validación del modelo. De la calibración, se obtienen los valores estimados de los parámetros  $r$ ,  $\alpha$ ,  $s$  y  $\beta$ , necesarios para el modelo Pareto/NBD en la estimación del *customer lifetime value* para cada uno de los clientes de la base de datos.

Los valores a estimar son el número de transacciones promedio y el monto promedio por transacción para cada uno de los clientes, luego para chequear que los valores estimados son confiables<sup>15</sup> se poseen datos transaccionales para el período de validación, con los cuales se realizará un análisis de errores.

Se escoge el error porcentual como medición del error del pronóstico, ya que nos indica cuánto fue el error en la estimación relativo a cada cliente, evitándose, además, el inconveniente de que frente a cualquier cambio de escala de la variable objeto de estudio afecte la magnitud de tales medidas, como ocurre con el error simple.

---

<sup>15</sup> La confianza se refiere a que el valor estimado se encuentre dentro de un rango de error pequeño con respecto al valor real.

Se utilizarán dos medidas del error de predicción, basadas en el error porcentual: MAPE<sup>16</sup> y MPE<sup>17</sup>. Por ejemplo, para el número de transacciones promedio, se tiene que:

- $EP = \frac{\hat{x} - x}{x}$  = error porcentual.
- $MAPE = promedio\left(abs\left(\frac{\hat{x} - x}{x}\right)\right)$  = promedio del valor absoluto del error porcentual.
- $MPE = promedio\left(\frac{\hat{x} - x}{x}\right)$  = promedio del error porcentual.

Donde:  $\hat{x}$  = pronóstico del número de transacciones promedio.  
 $x$  = valor real del número de transacciones.

Es importante mencionar que pueden existir casos en que un cliente haya realizado compras durante el período de calibración pero no en el de validación, luego se tiene  $x = 0$ , lo que indefiniría el error porcentual y, a su vez, el valor del MAPE y MPE. Dado lo anterior, para el análisis de errores no se consideran este tipo de casos.

La métrica *MPE* indica si el modelo tiene tendencia a sobreestimar o subestimar el valor real de la variable, mientras que el *MAPE* permite evaluar la bondad de ajuste dentro de la muestra y la exactitud o capacidad predictiva en los períodos futuros.

Los modelos que arrojen valores de indicadores de error cercanos a cero y valores pequeños de la desviación estándar serán los mejores, ya que significa que el valor estimado es cercano al real y estable en torno a dicho valor.

---

<sup>16</sup> MAPE significa, en inglés, "Mean Absolute Percentage Error".

<sup>17</sup> MPE significa, en inglés, "Mean Percentage Error".

### 3.3. Test Anova

Para saber si una variable independiente es significativa estadísticamente sobre una variable dependiente a estudiar, se realiza el test Anova. Anova significa "Analysis Of Variance" y como su nombre lo indica, compara varianzas aunque lo que contrastamos es la media.

El test Anova consiste en probar si dos o más poblaciones tienen la misma media, luego el propósito es hallar las diferencias en las medias poblacionales. Si se tienen  $n$  poblaciones, el test es el siguiente:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$$

$$H_1 : \text{No todas las medias son iguales}$$

Donde:  $\mu_n$  = es la media de la variable dependiente a testear de la población  $n$ .

El resultado del test consta de los siguientes indicadores:

- Suma de cuadrados
- Grados de libertad
- El valor del cuadrado medio
- La razón de la variación entre muestras y la variación dentro de las muestras (F).
- Nivel de significancia Sigma

*Figura 4: Resultados del test Anova*

		Sum of Squares	df	Mean Square	F	Sig.
N° de transacciones	Between Groups	2704,952	1	2704,952	684,944	,000
	Within Groups	639304,9	161884	3,949		
	Total	642009,9	161885			
Monto por transacción	Between Groups	8,52E+10	1	8,523E+10	1041,153	,000
	Within Groups	1,33E+13	161884	81865782,34		
	Total	1,33E+13	161885			
Monto total	Between Groups	1,94E+09	1	1939188829	1,979	,159
	Within Groups	1,59E+14	161884	979680464,1		
	Total	1,59E+14	161885			

Asumiendo un nivel de confianza del 95%, la regla de decisión es rechazar la hipótesis nula  $H_0$  si el nivel de significancia es menor a 0,05.

### 3.4. Clustering

Clustering o análisis de conglomerados es una técnica multivariante que permite agrupar los casos o variables de un archivo de datos en función del parecido o similitud existente entre ellos.

El programa SPSS dispone de dos tipos de análisis de conglomerados: el análisis de conglomerados *jerárquico* y el análisis de conglomerados de *K-medias*. El método jerárquico es idóneo para determinar el número óptimo de conglomerados existente en los datos, mientras que el método *K-medias* permite procesar un número ilimitado de casos pero sólo utilizando un método de aglomeración, previa determinación del número de conglomerados que se desea obtener.

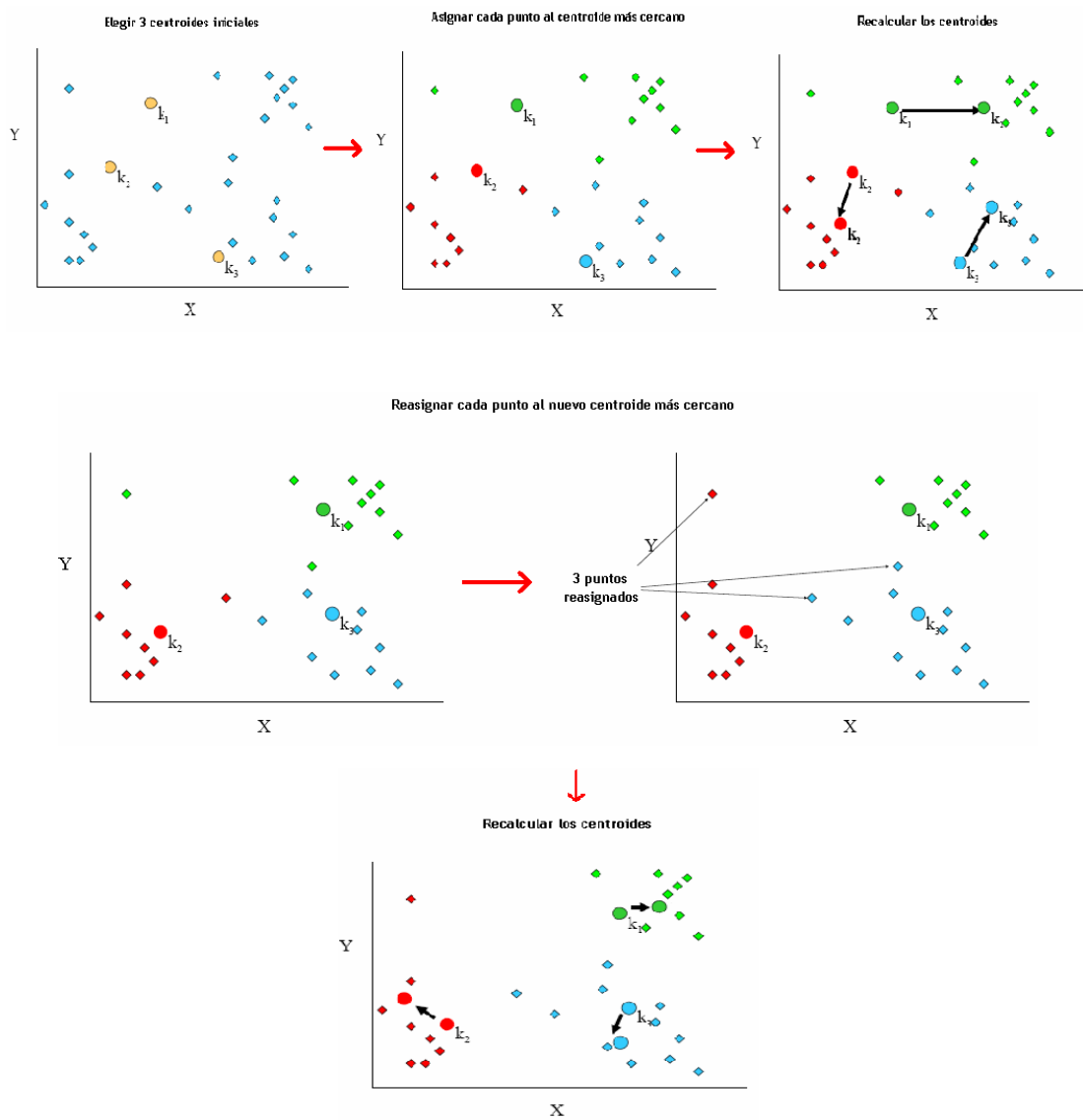
El análisis de conglomerados de *K-medias* es un método de *agrupación de casos* que se basa en las distancias existentes entre ellos en un conjunto de variables<sup>18</sup>. Versiones anteriores del procedimiento comenzaban el análisis con la asignación de los *K* primeros casos a los centros de los *K* conglomerados (centroides). En la versión actual, se comienza seleccionando los *K* casos más distantes entre sí, determinándose inicialmente el número de *K* conglomerados que se desea obtener. A continuación, se inicia la lectura secuencial del archivo de datos asignando cada caso al centro más próximo y actualizando el valor de los centros a medida que se van incorporando nuevos casos. Una vez que todos los casos han sido asignados a uno de los *K* conglomerados, se inicia un proceso iterativo para calcular los *centroides* finales de esos *K* conglomerados. El proceso descrito se muestra, gráficamente, en la figura 5 [4].

El análisis de conglomerados de *K-medias* es especialmente útil cuando se dispone de un gran número de casos. Este método se utilizará como una técnica de clasificación, clasificando los casos a partir de *centroides* ya conocidos.

---

<sup>18</sup> Este método no permite agrupar variables.

Figura 5: Proceso de agrupación K-means



Fuente: Congreso de Ciencia y Tecnología Computacional

### 3.5. Árboles de decisión

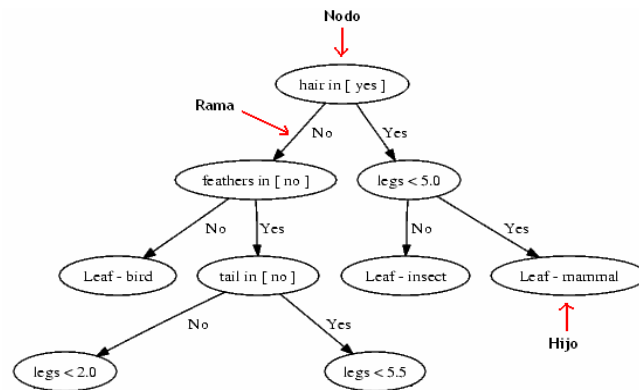
El árbol de decisión es un modelo supervisado aplicado para la segmentación de clientes y generación de reglas de clasificación entre otros. El algoritmo trata de separar los datos en conjuntos de reglas que probablemente respondan a un efecto o variable objetivo. Dado que se desea encontrar las variables no transaccionales que agregan mayor información en la estimación del "Número de transacciones", se utiliza un árbol de regresión.

Las fortalezas de este modelo son la fácil interpretación y entendimiento, genera un ranking automático de variables y tiene una rápida convergencia del algoritmo. Por otro lado, las debilidades consisten en la difícil interpretación cuando se posee mucha "profundidad" del árbol, sólo es posible tener un número discreto de ramas y no un continuo (posibilidades discretas) y la necesidad de hartos volúmenes de información.

La arquitectura del árbol está compuesta por:

- Nodo: es un punto de unión, donde se representa un lugar en el se debe tomar una decisión.
- Rama: representa un arco de conexión entre nodos.
- Hoja: es un nodo terminal (sin hijos).

*Figura 6: Ejemplo de un árbol de decisión*



Se poseen dos tipos de árboles de decisión:

- Árbol de clasificación: se trata de encontrar el grado de pertenencia de un objeto a una clase específica.
- Árbol de regresión: se trata de predecir un valor futuro de una variable en base a su comportamiento pasado.



## 4. APLICACIÓN DE LA METODOLOGÍA PARA EL CÁLCULO DEL CUSTOMER LIFETIME VALUE A NIVEL DE CLIENTES

### 4.1. Selección de datos

Este trabajo se desarrolla sobre una base de datos de clientes de la tarjeta de crédito de una empresa minorista de retail. Se posee información de todas las transacciones realizadas por los clientes a través de la tarjeta de crédito mencionada durante un período de 2 años, que comienza el 01-01-2003 y termina el 31-12-2004.

*Tabla 2: Resumen de datos*

Fecha primera transacción	02-01-2003
Fecha última transacción	31-12-2004
Número de clientes	304.574
Número de transacciones	1.663.771
Venta total	MM\$ 20.610

Cabe destacar, que el cálculo del número de transacciones se rige bajo el supuesto de que un cliente puede realizar a lo más una transacción diaria. Luego si, por ejemplo, un cliente realiza  $n$  transacciones durante el mismo día por un monto de  $M_i$  con  $i=1, \dots, n$ , éstas se agrupan en una transacción cuyo monto de compra es igual a  $M = \sum_{i=1}^n M_i$ . Dado este supuesto, cada cliente puede realizar 727<sup>19</sup> transacciones como máximo.

La información transaccional que se posee de cada uno de los clientes es la siguiente:

- $x$  = número de transacciones
- $t_1$  = fecha en que se realiza la primera transacción
- $t_n$  = fecha en que se realiza la última compra
- $M$  = monto promedio por transacción
- *Monto* = monto total de compra

El costo asociado a un cliente es desconocido, por lo tanto se estimará el ingreso futuro que generará cada cliente para la empresa y no su utilidad. Es importante mencionar que el registro de transacciones no considera devoluciones, avances ni repactaciones, sólo ventas.

Por otro lado, se posee información de las siguientes variables que caracterizan al cliente

---

<sup>19</sup> El período de 2 años considerado corresponde a 727 días.

como:

- Género
- Estado civil
- Región en la que vive el cliente
- Fecha de nacimiento
- Fecha de ingreso al sistema
- Estado de la tarjeta

## 4.2. Limpieza y Pre-procesamiento de datos

Esta sección tiene como fin eliminar los datos erróneos, inconsistentes y/o irrelevantes. A continuación, se muestra cómo se realizó la limpieza de datos para los 304.574 clientes considerando cada una de las variables.

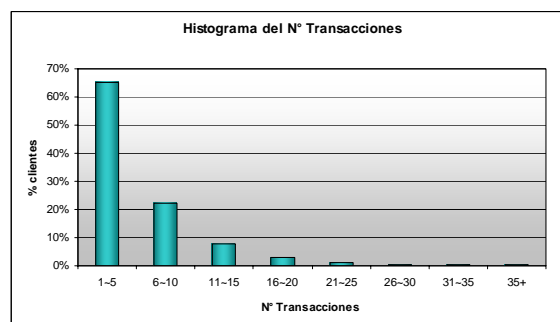
### 4.2.1. Variables transaccionales

Las variables transaccionales son: número de transacciones, monto por transacción, antigüedad del cliente y estado de la tarjeta.

- Número de transacciones

El número de transacciones se concentra en valores pequeños como se observa en la figura siguiente. Pero, también, hay existencia de pocos clientes que realizan compras con una alta frecuencia, siendo el máximo de 1 transacción semanal. Hay 45 clientes que no realizan ninguna transacción en el período de estudio ( 0,01% del total de casos).

*Figura 7: Distribución del número de transacciones*



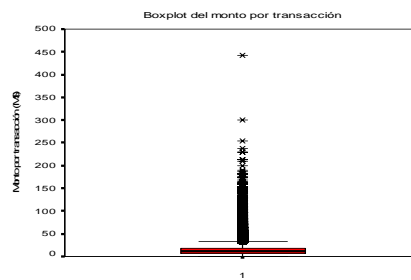
<b>Promedio</b>	3.36
<b>Desv. Estándar</b>	3.55

*Fuente: Elaboración propia*

- Monto por transacción

Se puede apreciar la existencia de clientes que desembolsan mucho dinero por cada transacción realizada. El 75% de los clientes gasta, en promedio, a lo más \$16.817, valor cercano al monto por transacción promedio del total de clientes, mientras el valor máximo alcanza los \$442.833, el cual es muy superior al promedio. Como se muestra en el boxplot del monto por transacción ( $M$ ), se observa que sí hay grandes outliers, los clientes que gastan más de \$35.000 por transacción, que corresponden al 3,47% del total de clientes, los cuales son eliminados del análisis. Hay clientes que no realizan compras durante todo el período y corresponden al 0.01% del total de clientes.

*Figura 8: Outliers según el monto por transacción*



*Fuente: Elaboración propia*

Promedio	Desv. Est.	Moda	Mínimo	1 <sup>er</sup> cuartil	Mediana	3 <sup>er</sup> cuartil	Máximo
\$13.787	\$9.932	\$3.833	\$0	\$7.749	\$11.461	\$16.817	\$442.833

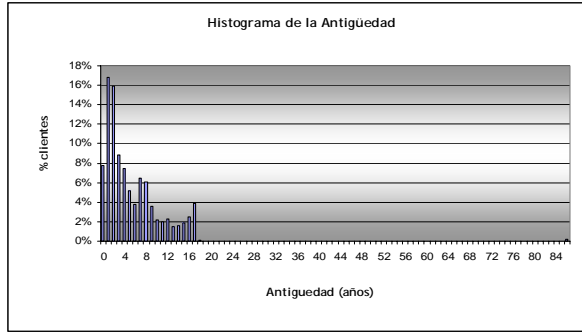
- Antigüedad

Se conoce la fecha en la cual fue ingresado el cliente al sistema y no la antigüedad, por lo tanto se transforma el dato conocido como se muestra a continuación:

$$\text{Antigüedad} = (31-12-2004) - \text{'fecha de ingreso al sistema'}$$

Como se observa en la figura 9, la relación de los clientes con la empresa se concentra en valores menores a 10 años. Además, son muy pocos los clientes que poseen una antigüedad mayor a los 25 años.

*Figura 9: Distribución de clientes según variable Antigüedad*



*Fuente: Elaboración propia*

En este caso, se deciden eliminar todos los casos que tengan una antigüedad superior a los 38 años, dado que el próximo valor de esta variable es 86 años, que corresponde a un valor erróneo<sup>20</sup> y al 0,20% del total de clientes.

*Tabla 3: Valores de la variable Antigüedad*

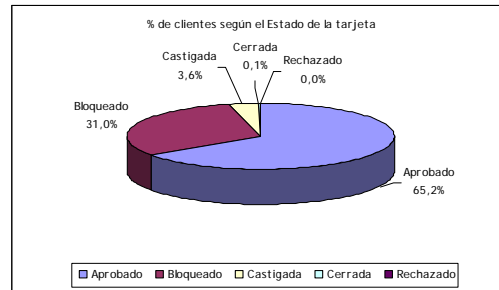
Valor	Nº casos	% del total	% acumulado
0-30 días	4.340	1,42%	1,42%
1-3 meses	12.296	4,04%	5,46%
3-6 meses	9.604	3,15%	8,61%
6-12 meses	20.466	6,72%	15,33%
1-2 años	66.336	21,78%	37,11%
2-38 años	190.935	62,69%	99,80%
Mayor que 38 años	597	0,20%	100,00%

- **Estado de la tarjeta**

Como se observa en la figura 10, todos los clientes tienen información con respecto al estado de la tarjeta, por lo tanto esta variable no genera problemas como se muestra en la siguiente figura.

<sup>20</sup> Es erróneo porque implicaría que la relación empresa-cliente fue iniciada cuando la persona tenía a lo más 3 años asumiendo que el cliente vive 90 años (caso hipotético).

*Figura 10: Distribución de la variable Estado de la tarjeta*



*Fuente: Elaboración propia*

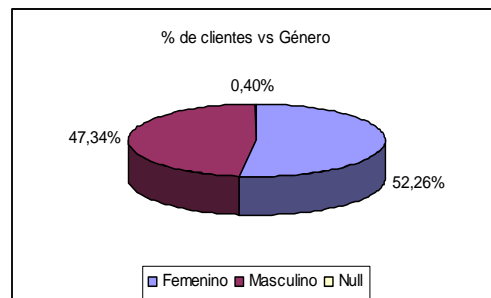
#### 4.2.2. Variables socio-demográficas

A continuación, se realiza la limpieza de datos con respecto a las variables socio-demográficas género, estado civil, región a la cual pertenece y la edad del cliente.

- Género

Se observa que existen casos en que no se posee información del género del cliente. Éstos corresponden al 0,40% del total de clientes, los cual son sacados del análisis.

*Figura 11: Distribución de la variable Género*

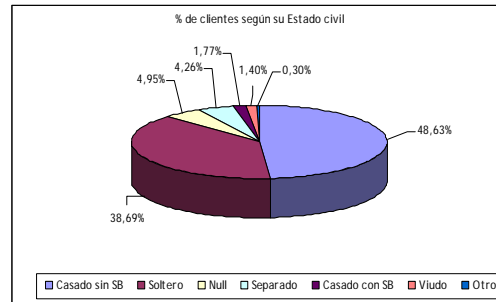


*Fuente: Elaboración propia*

- Estado civil

Como se ve en la figura 12, con respecto al estado civil, hay un 4,9% del total de casos en que no se posee información de esta variable. Estos clientes no serán considerados en el estudio.

*Figura 12: Distribución de la variable Estado civil<sup>21</sup>*

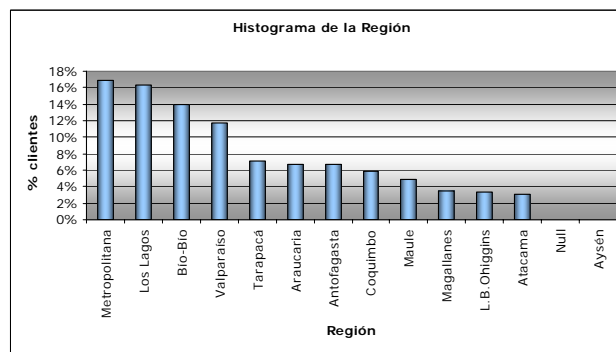


*Fuente: Elaboración propia*

- **Región**

Hay un 0,02% de los clientes, de los cuales se desconoce la región a la cual pertenece, los cuales son eliminados del análisis.

*Figura 13: Distribución de la variable Región*



*Fuente: Elaboración propia*

- **Edad**

La edad es calculada en base a la fecha de nacimiento del cliente de la siguiente manera:

$$\text{Edad} = (31-12-2004)^{22} - \text{'fecha de nacimiento'}$$

Como se aprecia en la tabla 4, existen clientes que tienen edades negativas, que corresponden a valores incorrectos, luego éstos son eliminados del estudio (0,01% del total). Los clientes menores de 18 años que poseen tarjeta de crédito, también, son eliminados y corresponden al 0,03% de los clientes.

<sup>21</sup> SB se refiere a Separación de bienes.

<sup>22</sup> El cálculo de la edad está referenciada al último día del período en estudio, correspondiente al 31-12-2004.

*Tabla 4: Valores de la variable Edad expresada en años*

Valor	Nº casos	% del total	% acumulado
Menor que cero	16	0,01%	0,01%
0-17	43	0,01%	0,02%
18-20	58	0,02%	0,04%
21-24	22.286	7,32%	7,36%
25-29	39.397	12,94%	20,30%
30-34	44.354	14,56%	34,86%
35-44	84.369	27,70%	62,56%
45-54	59.149	19,42%	81,98%
55-64	29.780	9,78%	91,76%
65-85	15.517	5,09%	96,85%
Mayor que 85	9.605	3,15%	100,00%

Los indicadores transaccionales de los clientes entre 18 y 20 años, incluyendo los extremos, toman valores muy por sobre el resto y corresponden a 58 clientes, por lo tanto, serán sacados del estudio para evitar error en el análisis posterior.

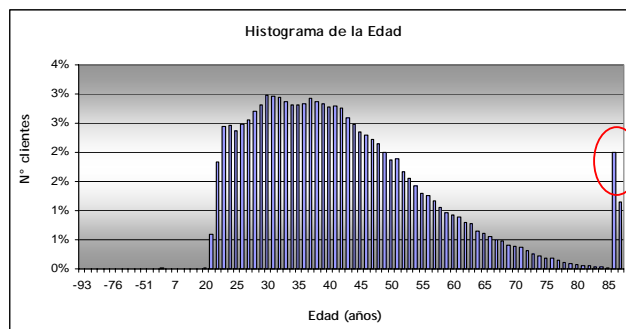
*Tabla 5: Resumen indicadores transaccionales según edad*

Edad	Nº transacciones promedio	Monto promedio por transacción (\$)	Monto total promedio (\$)
18-20	8,0	12.217	98.160
21-24	5,0	12.730	63.891
25-29	5,2	12.439	65.191
30-34	5,4	12.344	66.354
35-44	5,6	12.253	68.711
45-54	5,7	12.309	69.891
55-64	5,3	12.535	66.373
65-85	5,2	12.743	65.674
Total	5,4	12.387	67.369

En la figura 14, hay un peak en la edad de 86 años, destacado con un círculo, el cual es considerado una anomalía, ya que el 95,2% de estos clientes tienen asociadas una fecha de nacimiento igual a 1° de enero de 1919, que se creen fueron ingresadas como una fecha tipo por la misma persona para facilitar y agilizar el tipeo de la fecha de nacimiento de esos clientes. Además, existen clientes con edades sobre los 100 años, luego si se consideraran estos casos se generarían errores en el estudio, por lo tanto, se decide

eliminar los clientes que poseen una edad mayor o igual a 86 años (3,15% del total).

*Figura 14: Distribución de la variable Edad*



Fuente: Elaboración propia

#### 4.2.3. Conclusiones de la limpieza y pre-procesamiento de datos

A continuación, se muestra una tabla resumen de los casos eliminados para cada una de las condiciones determinadas en el análisis anterior.

*Tabla 6: Resumen de datos eliminados*

Condición	Número de casos	% del total
Cientes con $x=0$	45	0,01%
Outliers de $M$ ( $M > \$35.000$ )	10.558	3,47%
Edad < 21 [años] o Edad > 85 [años]	9.722	3,19%
Antigüedad > 38 [años]	596	0,20%
No posee información en la variable Género	1.217	0,40%
No posee información en la variable Estado civil	15.065	4,95%
No posee información en la variable Región	60	0,02%
<b>Total casos eliminados</b>	<b>24.051</b>	<b>7,90%</b>

Resumiendo, los datos a considerar a partir de ahora, en el análisis, cumplen las siguientes condiciones:

1. El monto por transacción ( $M$ ) es menor o igual a \$35.000.
2. El número de transacciones totales ( $x$ ) es mayor a 0.
3. La edad se debe encontrar entre los 21 y 85 años incluyendo los extremos, es decir:  $20 < \text{Edad} < 86$  [años].



4. La antigüedad no debe superar los 38 años, luego se tiene que:  
 $0 \leq \text{Antigüedad} < 39$  [años].
5. Que posea información en las variables género, estado civil y región.

Los datos seleccionados se encuentran resumidos en la tabla que se muestra a continuación.

*Tabla 7: Resumen de datos seleccionados*

	Valor	% del total
Número de clientes	280.523	92,10%
Número de transacciones	1.519.720	91,34%
Venta total	MM\$ 18.876	91,59%

### 4.3. Análisis descriptivo

El comportamiento transaccional de los clientes queda descrito por los siguientes indicadores:

- $x$  = número de transacciones realizadas durante el período de análisis
- $t_x$  = período de compra
- $T$  = período de observación
- $M$  = monto promedio por transacción
- $Monto$  = monto total

Se posee información acerca del género, estado civil, región al cual pertenece, edad, antigüedad de la relación con la empresa, día de pago y el estado de la tarjeta. Luego, para cada cliente se tiene la siguiente información:

*Tabla 8: Resumen de información disponible*

cliente	x	$t_x$ (días)	T (días)	M (\$)	Monto (\$)	Género	Estado civil	Región	Edad (años)	Antigüedad (años)	Día de pago	Estado de tarjeta
11	8	313	540	11.056	88.447	Femenino	Casado sin SB	Los Lagos	71	16	25	Bloqueado
...	...	...	...	...	...	...	...	...	..	...	...	...

El análisis se realiza desde dos perspectivas: uno a nivel general y otro por variable.

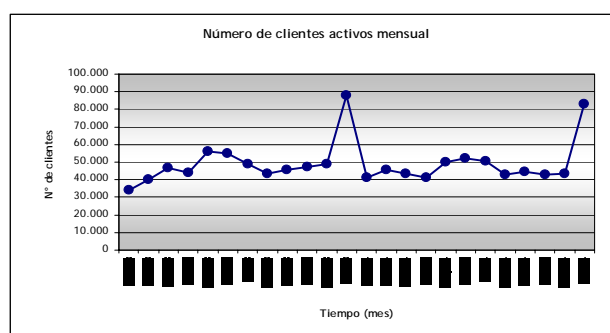
### 4.3.1. Análisis general

- Número de clientes

El número de clientes activos<sup>23</sup> mensualmente presenta una forma similar de un año a otro, por ejemplo, hay una caída del número de clientes de julio a agosto del 2003, la cual también se observa en el 2004.

Hay un aumento significativo del número de clientes que al menos realiza una transacción en el mes de diciembre de 40.000 clientes, el que se debe a la existencia de las fiestas de Navidad y Año nuevo. Por otro lado, se puede apreciar una pequeña tendencia a la baja de este indicador de un mismo mes de un año a otro.

*Figura 15: Evolución del número de clientes activos*



*Fuente: Elaboración propia*

*Tabla 9: Resumen de indicadores mensual*

Indicador	Promedio			Desv. Estándar		
	Nº transacc. por cliente	Monto por Transacción (\$)	Monto total por cliente (\$)	Nº transacciones por cliente	Monto por transacción (\$)	Monto total por cliente (\$)
Ene-03	1,22	10.460	12.655	0,54	10.594	13.283
Feb-03	1,23	14.014	16.869	0,54	13.535	16.654
Mar-03	1,30	13.451	17.183	0,64	11.619	15.915
Abr-03	1,25	12.914	15.890	0,56	10.821	14.245
May-03	1,30	12.804	16.474	0,64	12.367	18.285
Jun-03	1,28	12.259	15.410	0,60	10.646	14.060
Jul-03	1,28	13.392	16.888	0,60	11.642	15.799
Ago-03	1,25	14.547	18.010	0,57	13.622	17.970
Sept-03	1,24	13.613	16.637	0,55	11.279	14.724

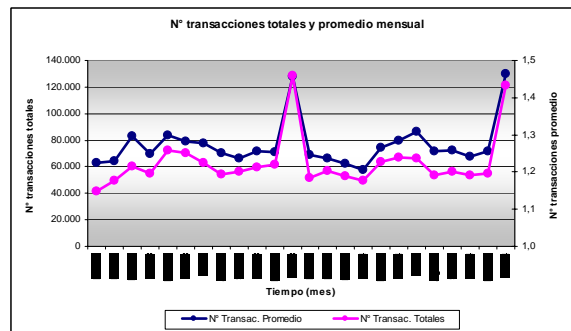
<sup>23</sup> Cliente activo corresponde a un cliente que haya realizado una compra como mínimo en el mes analizado.

Indicador	Promedio			Desv. Estándar		
	Mes	N° transacc. por cliente	Monto por Transacción (\$)	Monto total por cliente (\$)	N° transacciones por cliente	Monto por transacción (\$)
Oct-03	1,26	12.417	15.418	0,58	11.041	14.713
Nov-03	1,26	12.021	14.788	0,58	11.776	14.918
Dic-03	1,46	14.469	20.484	0,80	14.797	21.546
Ene-04	1,25	10.009	12.295	0,58	10.016	12.729
Feb-04	1,24	12.419	15.102	0,56	11.506	14.622
Mar-04	1,22	12.001	14.468	0,54	10.520	13.426
Abr-04	1,21	11.716	13.955	0,53	10.104	12.750
May-04	1,26	11.399	14.211	0,59	10.573	14.606
Jun-04	1,29	10.716	13.488	0,61	9.242	12.289
Jul-04	1,31	12.278	15.812	0,65	10.477	14.740
Ago-04	1,26	13.640	17.043	0,60	11.651	16.125
Sept-04	1,26	13.929	17.579	0,60	12.274	18.225
Oct-04	1,24	11.620	14.197	0,58	10.339	13.501
Nov-04	1,26	11.363	14.125	0,60	11.005	15.011
Dic-04	1,46	12.285	17.635	0,82	10.882	17.346

- Número de transacciones

La forma del comportamiento del número de transacciones totales a través del tiempo es explicado por el número de clientes activos en cada uno de los meses. Se puede apreciar, en la figura 16, un aumento del número de visitas promedio de los clientes durante el mes de Diciembre, debido principalmente a la Navidad y Año nuevo. En general, exceptuando el mes de Diciembre, el número de transacciones promedio se encuentra entre 1,2 y 1,3 transacciones por persona a nivel mensual.

*Figura 16: Evolución del número de transacciones total y promedio*



Fuente: Elaboración propia

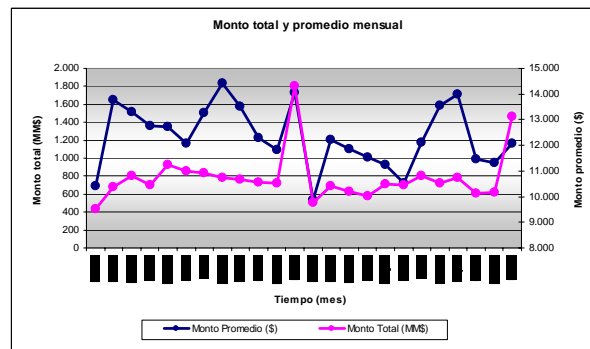
- Monto total y promedio

El menor gasto por transacción con respecto a cada año se realiza en el mes de enero, seguido del mes de junio. Por otro lado, como ya se ha visto, el mes de diciembre corresponde a un peak, sin embargo, el monto promedio por transacción en diciembre del 2004 disminuye en \$2.000 con respecto al 2003.

En general, se observa una tendencia a la baja, pero manteniendo la forma del monto promedio por transacción durante el año. Es decir, este indicador durante el 2004 se mueve en una escala menor con respecto al 2003.

El monto total mensual es estable, excepto en el caso de diciembre. El monto total percibido durante diciembre del 2004 fue de MM\$1.469, 19% menor que el monto total gastado por los clientes durante el mismo mes del 2003.

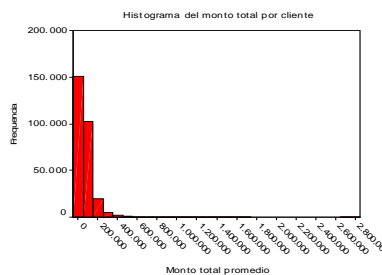
*Figura 17: Evolución del monto promedio y total*



*Fuente: Elaboración propia*

Los valores del monto total por cliente se concentran en valores pequeños.

*Figura 18: Distribución del monto total promedio*



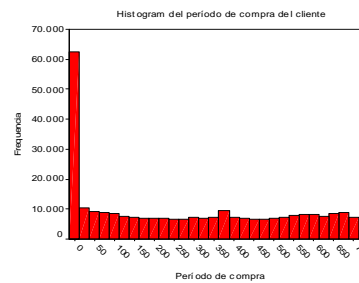
*Fuente: Elaboración propia*

Indicador Monto	Promedio	Desv. estándar	Mínimo	Máximo
Valor	67.288	72.747	0	2.824.089

- Período de compra ( $t_x$ )

El período de compra de un cliente corresponde a los días transcurridos desde la fecha de la primera compra hasta la fecha en la que se realiza la última transacción. En general, como se ve en la figura 19, la distribución de  $t_x$  es estable, excepto por la existencia de clientes que permanecen muy poco activos y que corresponden a los clientes que realizan 1 transacción. Esto indica que existen dos grupos de clientes bien definidos: los que realizan una única transacción y el resto de los clientes.

*Figura 19: Distribución del período de compra ( $t_x$ )*



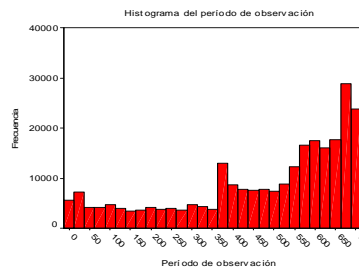
*Fuente: Elaboración propia*

Indicador $t_x$	Promedio	Desv. estándar	Mínimo	Máximo
Valor	282	240	0	729

- Período de observación (T)

El período de observación del cliente corresponde a los días transcurridos desde la fecha de la primera compra hasta la fecha de término del período (31/12/2004). Se puede ver que la mayoría de los clientes realizan su primera compra al inicio del período en estudio.

*Figura 20: Distribución del período de observación (T)*



*Fuente: Elaboración propia*

Indicador T	Promedio	Desv. estándar	Mínimo	Máximo
Valor	491	210	0	729

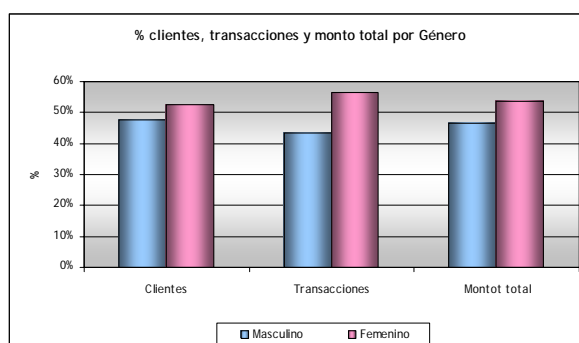
### 4.3.2. Análisis por variables

Esta sección muestra un análisis del comportamiento de los clientes distinguiendo características propias de ellos.

- Género

Dentro de la base de datos de clientes, a nivel agregado, las mujeres predominan con un 52,5%. Además, son las que realizaron mayor cantidad de transacciones durante el período (56,4% del total) y aportaron el 53,5% del monto total percibido por empresa.

*Figura 21: Distribución del número de transacciones, clientes y monto por género*



*Fuente: Elaboración propia*

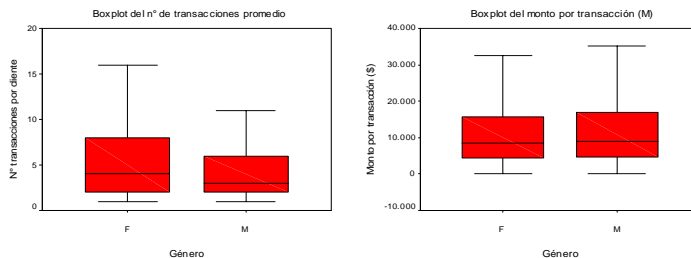
El número de transacciones promedio de las mujeres es casi 1 unidad mayor que el de los hombres y el monto total promedio gastado por el género femenino es mayor en un 4,4% con respecto al masculino. Sin embargo, no ocurre lo mismo con el monto promedio por transacción, ya que éste es superior en el caso de los hombres por \$1.500 aproximadamente.

*Tabla 10: Resumen indicadores transaccionales según género*

Género	Nº transacciones promedio	Monto promedio por transacción	Monto promedio total por cliente
Masculino	5,0	13.244	66.729
Femenino	5,8	11.786	68.701
Total	5,4	12.421	67.288

De los boxplots mostrados en la figura 22, se observa que, en cuanto al número de transacciones promedio, el 25% de las mujeres realizan 8 o más transacciones, mientras que el 25% de los hombres realiza 6 o más compras durante el período de análisis, siendo que el promedio es de 6 y 5 transacciones, respectivamente.

*Figura 22: Distribución del n° de transacciones y monto promedio por transacción según género*

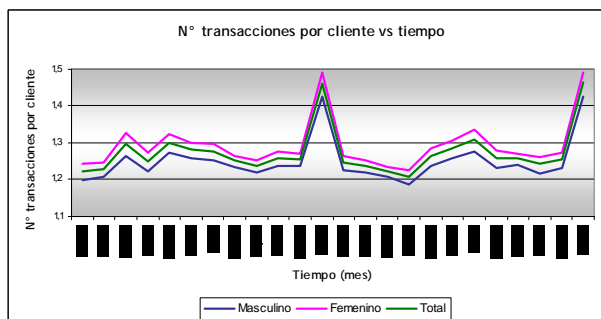


*Fuente: Elaboración propia*

Por otro lado, si analizamos las series de tiempo<sup>24</sup> (ver figuras 23 y 24), el número de transacciones y monto promedio por transacción tienen un comportamiento similar en el tiempo independiente del género del cliente, pero en distintas escalas. En cuanto al número de transacciones, las mujeres siempre se mueven por sobre los hombres, mientras que para el caso del monto promedio, tanto por transacción como por cliente, ocurre lo contrario.

Además, se observa la existencia de un peak en el mes de Diciembre, en el cual la diferencia del número de transacciones y monto promedio por transacción entre géneros se hace menor en comparación al resto de los meses. La evolución del monto promedio presenta una tendencia a la baja al comparar el gasto de los clientes del mismo mes entre años, por ejemplo, en febrero del 2003, el gasto promedio por transacción masculino fue de \$14.869, mientras que el mismo mes del siguiente año, el gasto cae a \$13.262.

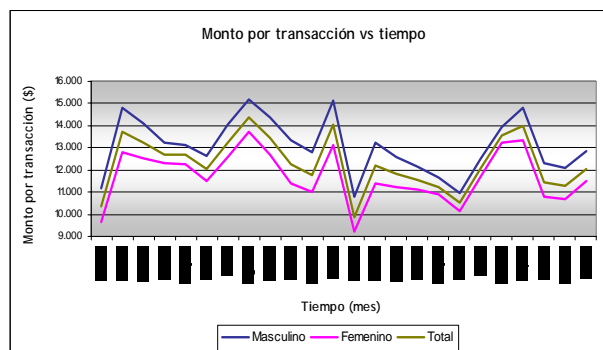
*Figura 23: Evolución del número de transacciones promedio según género*



*Fuente: Elaboración propia*

<sup>24</sup> La unidad de tiempo utilizada es el mes.

*Figura 24: Evolución del monto promedio por transacción según género*

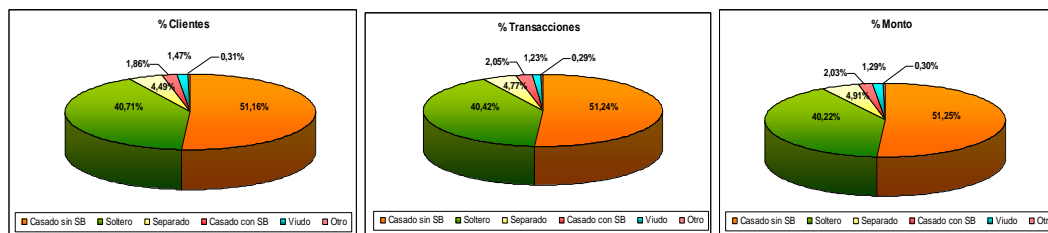


*Fuente: Elaboración propia*

- Estado civil

De la figura 25, se ve que la mayoría de los clientes son casados sin SB y solteros representando el 91,9% del total de clientes y el número de transacciones realizadas por estos dos grupos, en conjunto, corresponde al 91,7% aportando el 91,5% del monto total. Por lo tanto, estos dos grupos de clientes son muy importantes para la empresa, ya que representan más del 90% del ingreso percibido.

*Figura 25: Distribución del número de transacciones, clientes y monto por estado civil*



*Fuente: Elaboración propia*

De la tabla 11, se rescata que los clientes separados y casados con SB son los que realizan mayor cantidad de transacciones promedio, encontrándose sobre el promedio de toda la cartera de clientes.

En cuanto al monto por transacción, los viudos y separados son los que más gastan en cada una de sus compras. Como se vio anteriormente, los clientes solteros y casados sin SB son los que realizan el mayor aporte monetario y tanto el número de transacciones promedio como su monto por transacción se encuentran cerca del promedio total. Este efecto se debe a que estos dos grupos de clientes son los más numerosos.



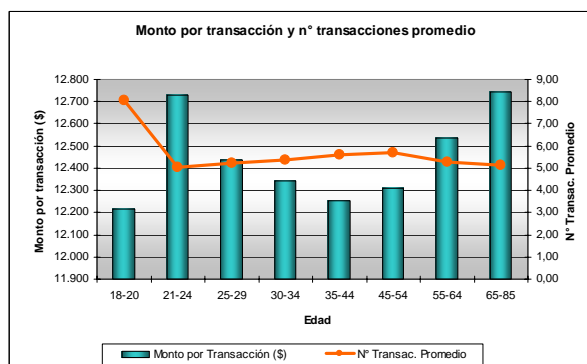
*Tabla 11: Resumen indicadores transaccionales según estado civil*

Estado civil	Nº transacciones promedio	Monto promedio por transacción	Monto total promedio
Otro	5,1	12.959	65.574
Viudo	4,6	13.019	59.300
Separado	5,8	12.743	73.592
Casado con SB	6,0	12.267	73.542
Soltero	5,4	12.357	66.311
Casado sin SB	5,4	12.427	67.504
Total	5,4	12.421	67.288

- Edad

Los clientes entre 35 y 54 años van más seguidos a realizar compras que el promedio y el monto promedio por cliente es alto.

*Figura 26: Monto por transacción y N° transacciones promedio por edad*



Fuente: Elaboración propia

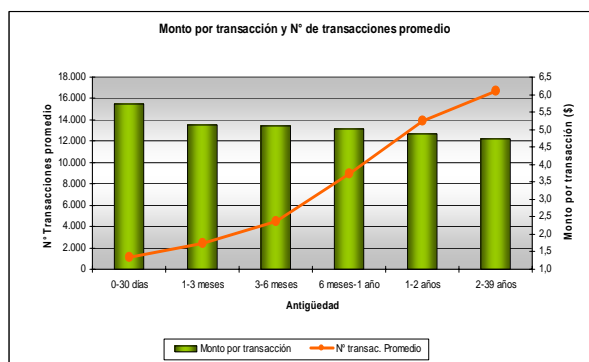
*Tabla 12: Resumen indicadores transaccionales según edad*

Edad	Nº transacciones promedio	Monto promedio por transacción	Monto promedio
21-24	5,0	12.734	63.883
25-29	5,2	12.449	65.166
30-34	5,4	12.369	66.352
35-44	5,6	12.293	68.536
45-54	5,7	12.353	69.765
55-64	5,3	12.572	66.403
65-85	5,2	12.797	66.296
Total	5,4	12.421	67.288

- Antigüedad

A medida que aumenta la antigüedad, el número de transacciones y monto promedio también aumenta, sin embargo ocurre lo contrario con el monto promedio por transacción. Es decir, los clientes antiguos para la empresa van con mayor frecuencia a comprar pero lo hacen en montos cada vez menores.

*Figura 27: Monto por transacción y N° de transacciones promedio por antigüedad*



*Fuente: Elaboración propia*

*Tabla 13: Resumen indicadores transaccionales según antigüedad*

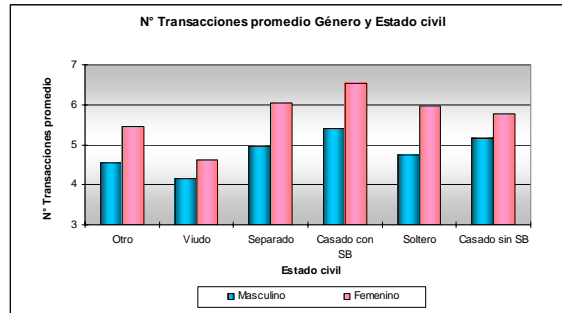
Antigüedad	N° transacciones promedio	Monto promedio por transacción	Monto promedio
0-30 días	1,3	15.482	20.593
1-3 meses	1,7	13.539	23.516
3-6 meses	2,4	13.462	31.905
6-12 meses	3,7	13.103	48.994
1-2 años	5,3	12.702	66.823
2-39 años	6,1	12.226	74.853
Total	5,4	12.421	67.288

- Género y estado civil

Ahora considerando las variables género y estado civil, de las figuras 28 y 29, se observa que los tres grupos de clientes que realizan un mayor número de transacciones promedio son, respectivamente, las mujeres casados con SB, las mujeres separadas y las mujeres solteras.

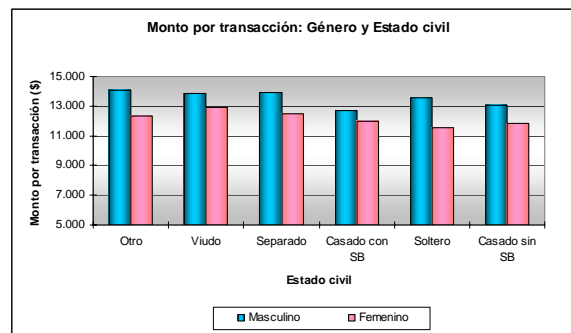
Con respecto al monto promedio por transacción, se tienen que los 3 grupos con mayor valor de este indicador son: hombres viudos, separados y otro.

*Figura 28: N° de transacciones promedio según género y estado civil*



*Fuente: Elaboración propia*

*Figura 29: Monto por transacción promedio según género y estado civil*



*Fuente: Elaboración propia*

### 4.3.3. Conclusiones del análisis descriptivo

#### 4.3.3.1. Conclusiones del análisis general

- Hay peak tanto en el número de transacciones como en la venta total en el mes de diciembre debido a las fiestas de Navidad y Año nuevo.
- Hay una tendencia a la baja del monto por transacción y total.

*Tabla 14: Variación del monto por transacción y total entre el año 2003 y 2004*

Año	Monto por transacción	Monto total
2003	\$12.991	MM\$ 10.040
2004	\$11.830	MM\$ 8.836
Variación	\$ -1.161 (- 8,94%)	\$ -1.204 (-11,99%)

- Hay 2 grupos de clientes bien definidos, aquellos que permanecen muy poco tiempo comprando ( $t_x=0$ , es decir, realizan una única transacción) y el resto de los clientes.
- La mayoría de los clientes realizan la primera compra al inicio del período del período en estudio.

#### 4.3.3.2. Conclusiones del análisis por variables

- Las mujeres son las que aportan el mayor monto total a la empresa y las realizan mayor número de transacciones totales, en comparación a los hombres. Sin embargo, los hombres son los que gastan más en promedio por transacción.
- Con respecto al estado civil del cliente, el grupo que domina en cuanto a cantidad de clientes, número de transacciones realizadas y venta generada son los clientes casados sin SB y solteros.
- A medida que aumenta la antigüedad de la relación del cliente con la empresa, mayor es el monto total aportado a ella, al igual que el número de transacciones realizadas durante todo el período. Por otro lado, los clientes antiguos gastan menos en promedio por transacción que los clientes nuevos. Más del 80% de los clientes de esta empresa tienen una antigüedad mayor a un año.
- Los grupos más importantes en cuanto al número de transacciones y monto total son los hombres casados sin SB, mujeres casadas sin SB y las mujeres solteras.

#### 4.4. Transformación de datos

Hasta ahora se han hecho 4 transformaciones que son:

- Edad = (31-12-2004) - 'fecha de nacimiento'
- Antigüedad = (31-12-2004) - 'fecha de ingreso al sistema'
- $t_x = t_n - t_1$  = período de compra.
- $T = t_{fin} - t_1$  = período de observación.

Donde:  $t_1$  = fecha de la primera compra del cliente.

$t_n$  = fecha de la última compra del cliente.

$t_{fin}$  = último día del período de observación.

A continuación, se procede a categorizar las variables para reducir los valores que toma cada una de las variables, ya que, por ejemplo, en el caso de la edad se tienen 65 valores de esta variable aproximadamente.

Para las variables edad y antigüedad se aplica el método K-medias, mientras que para las variables género, estado civil y región se categorizan según otros criterios que se detallan a continuación.

- **Género**: esta variable posee sólo 2 valores, Masculino y Femenino, y se trabajarán con ellas.
- **Estado civil**: los valores de esta variable son Casado sin SB, Soltero, Separado, Casado con SB, Viudo y Otro. Se escogen los estados que poseen la mayor cantidad de clientes y el resto se agrupa en una categoría "Otros".

*Tabla 15: Categorías según región*

Categoría	% clientes
Casado sin SB	52,1 %
Soltero	39,4 %
Separado	4,8 %
Otros	3,7 %

- **Región**: se poseen 13 valores de esta variable, correspondiente a cada una de las regiones de Chile, como su nombre lo dice.

Al igual que en el caso del estado civil, se escogen las regiones que poseen la mayor presencia de clientes y el resto se agrupa en una categoría "Otros".

*Tabla 16: Categorías según región*

Categoría	% clientes
Metropolitana	15,0 %
Valparaíso	11,7 %
Bío-Bío	13,0 %
Los Lagos	17,2 %
Otros	43,1 %

- **Edad:** se posee un rango de edad que va desde los 21 hasta los 86 años. Se aplicó el método K-means y los clusters obtenidos fueron:

*Tabla 17: Método K-means para la variable edad*

Centros de clusters iniciales				Centros de clusters finales			
	Cluster				Cluster		
	1	2	3		1	2	3
Edad (años)	85,00	21,00	53,00	Edad (años)	62,99	28,48	43,60

Rangos de los clusters finales		
Cluster	Rango Edad (años)	% clientes
1	55-85	15,7 %
2	21-34	35,2 %
3	35-54	49,1 %

*Fuente: Elaboración propia*

Finalmente, los clientes pertenecen a uno de los 3 clusters formados según la edad de ellos.

- **Antigüedad:** los valores de esta variable se encuentran entre los 0 y 39 años de antigüedad del cliente.

*Tabla 18: Método K-means para la variable antigüedad*

Centros de clusters iniciales				Centros de clusters finales			
	Cluster				Cluster		
	1	2	3		1	2	3
Antigüedad (años)	24,00	,00	12,00	Antigüedad (años)	14,94	2,53	7,98

Rangos de los clusters finales		
Cluster	Rango Antigüedad (años)	% clientes
1	12-24	14,8 %
2	0-5	57,9 %
3	6-11	27,4 %

*Fuente: Elaboración propia*

Los clientes pertenecen a uno de los 3 clusters formados según la antigüedad de los clientes en la empresa en estudio.

## 4.5. Selección de variables

La selección de las variables a incorporar se realiza en 3 etapas. Primero, se chequea que las variables sean estadísticamente significativas sobre la variable que se desea predecir, luego, se categorizan las variables continuas como la edad y antigüedad junto a aquellas que poseen muchos valores. Finalmente, se genera un árbol de decisión para identificar las variables de mayor peso en el pronóstico del número de transacciones, las cuales serán, finalmente, incorporadas en la metodología de estimación del *CLV*.

### 4.5.1. Aplicación del test Anova

Para conocer si existen diferencias estadísticas, se realizó el test Anova considerando como:

- **Variables dependientes:** número de transacciones, monto por transacción y monto total por cliente.
- **Variables independientes:** género, estado civil, edad, región, antigüedad, día de pago y estado de la tarjeta.

El test se aplica sobre los datos de la situación final estudiada, que corresponde al segundo semestre del 2003 (01/07/2003-31/12/2003).

A continuación, se muestran los resultados obtenidos considerando como variable independiente el género. El test para las otras variables se encuentra en el Anexo 3.

*Tabla 19: Test Anova para la variable independiente género*

TEST ANOVA: VARIABLE INDEPENDIENTE GÉNERO						
		Sum of Squares	df	Mean Square	F	Sig.
N° de transacciones	Between Groups	2704,952	1	2704,952	684,944	,000
	Within Groups	639304,9	161884	3,949		
	Total	642009,9	161885			
Monto por transacción	Between Groups	8,52E+10	1	8,523E+10	1041,153	,000
	Within Groups	1,33E+13	161884	81865782,34		
	Total	1,33E+13	161885			
Monto total	Between Groups	1,94E+09	1	1939188829	1,979	,159
	Within Groups	1,59E+14	161884	979680464,1		
	Total	1,59E+14	161885			

*Fuente: Elaboración propia*

Se observa que la variable género es estadísticamente significativa sobre el número de

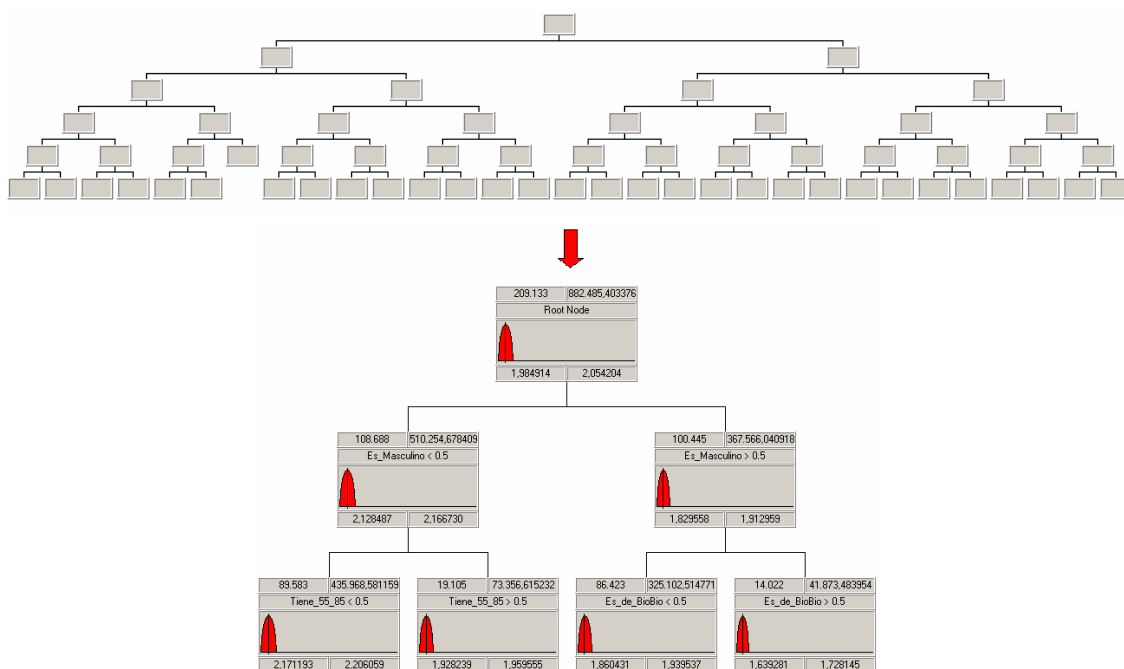
transacciones y monto por transacción, pero no sobre el monto total. Dado que lo que se estima el número de transacciones, se tiene que el género es significativo y por lo tanto es una de las variables a utilizar para diferenciar los clientes y, así, poder mejorar la estimación del valor de éste.

Finalmente, las variables género, estado civil, edad, región y antigüedad sí hacen una diferencia en el número de transacciones realizado por cada uno de los clientes, ya que el valor de sigma es menor que 0,05. Se concluye que la diferencia entre las medias es significativa, es decir, la distribución del indicador al considerar las distintas variables no es probable que se deba a una casualidad sino que a otros factores como que, tal vez, los clientes jóvenes son más derrochadores que los más viejos entre otros.

#### 4.5.2. Generación de un árbol de decisión

Se utilizan árboles de regresión, ya que éstos permiten predecir el valor de una variable objeto. En este caso, la variable a predecir es el número de transacciones y las variables predictoras corresponden al género, estado civil, edad, antigüedad y región. El árbol generado tiene una altura de 6 y consta de 61 nodos.

*Figura 30: Árbol de regresión generado*



*Fuente: Elaboración propia*

En la tabla 20, se tiene que el valor del MAPE en el período de entrenamiento asociado al árbol es de 17,0%, mucho menor que en el período de testeo. Luego, efectivamente, hay un sobreajuste en la estimación del número de transacciones.



Tabla 20: Error asociado al árbol

Período	Fecha inicio	Fecha término	MAD	MAPE
Entrenamiento	01/01/2003	31/12/2003	2,0	17,0 %
Testeo	01/01/2004	31/12/2004	2,2	72,3 %

El conocimiento que posee cada una de las variables sobre la variable a predecir se representa a través de la magnitud del peso de cada una de éstas.

Tabla 21: Pesos de cada una de las variables

Variable	Género	Edad	Región	Antigüedad	Estado civil
Peso	52,7	13,7	12,7	5,9	3,8

Dado los pesos observados, las variables seleccionadas son el género, edad y región de los clientes.

Ahora que se conocen las variables a incorporar en la estimación del número de transacciones y, por lo tanto, en el cálculo del *CLV*, se procede con el análisis de resultados considerando las variables socio-demográficas y transaccionales, para luego, finalmente, generar conclusiones en relación a la incorporación de información (género, estado civil, edad, región y antigüedad) distinta a la obtenida a través de los atributos RFM como se mostró en el capítulo anterior.

#### 4.6. Estimación del Customer Lifetime Value basado sólo en atributos RFM

Se estima el valor del *customer lifetime value (CLV)* individual usando sólo los atributos RFM, para comparar los resultados de éste con el caso en que se incorporan las variables seleccionadas (Género, Edad y Región) y concluir el aporte o no aporte de estas variables en la estimación del CLV a nivel de clientes.

Es importante conocer cómo se describe el comportamiento transaccional en cada uno de los períodos. Los atributos RFM quedan descritos por los siguientes indicadores:

- $x$  = número de transacciones realizadas durante el período de análisis
- $t_x$  = período de compra
- $T$  = período de observación
- $M$  = monto promedio por transacción

Para calcular el *CLV* es necesario estimar el número de transacciones y el monto promedio por transacción. Dado que el monto promedio por transacción no es modelado, la

confiabilidad del *CLV* depende de la precisión en la estimación de las compras futuras, luego, esto último es fundamental. A continuación, se muestran en detalle los pronósticos realizados por los modelos en estudio.

#### 4.6.1. Estimación del número de transacciones

El pronóstico final del número de transacciones considera los clientes que no realizan compras durante el período de calibración, además, de la aplicación de una constante para eliminar el error sistemático presente en los modelos Pareto/NBD y Simple. A continuación, se detalla el por qué y el efecto de los cambios realizados.

Se consideran 4 ventanas de tiempo, el primero para determinar los clientes inactivos en el período de calibración, el tercero es necesario para calcular el valor de la constante para eliminar el error sistemático de los modelos, para luego estimar el número de transacciones de cada cliente durante el segundo semestre del año 2004 (período de validación).

*Tabla 22: Períodos de observación*

Período	Fecha inicio	Fecha término
Base	02/01/2003	30/06/2003
Calibración	01/07/2003	31/12/2003
Test	02/01/2004	30/06/2004
Validación	01/07/2004	31/12/2004

De la tabla 23, se tiene que un cliente promedio realiza 2,6 transacciones en un período de 56 días siendo la primera compra hace casi 4 meses, para luego reducirlo en casi la mitad durante el período similar del año 2004.

*Tabla 23: Resumen de datos RFM*

Valor	Período de Calibración					Período de Validación		
	$x$	$t_x$ (días)	$T$ (días)	$M$ (\$)	Monto (\$)	$x$	$M$ (\$)	Monto (\$)
Promedio	2,6	56	106	13.061	32.454	1,6	6.947	19.587
Desv. Est.	2,0	60	58	9.077	31.300	2,3	8.863	32.041
Mínimo	1,0	0	0	0	0	0,0	0	0
Mediana	2,0	32	115	10.933	23.442	1,0	4.378	6.297
Máximo	37,0	183	183	137.802	619.341	39,0	116.866	940.197

A continuación, se muestran las 2 situaciones: presencia de sólo clientes activos y el caso en el que se incorporan clientes inactivos y la aplicación de la constante.

#### 4.6.1.1. Presencia sólo de clientes activos

El primer caso es cuando se tienen sólo clientes activos durante el segundo semestre del 2003 y son a éstos a quienes se les estima el comportamiento de compra futuro.

*Tabla 24: Períodos de observación*

Período	Fecha inicio	Fecha término
Calibración	01/07/2003	31/12/2003
Validación	01/07/2004	31/12/2004

*Tabla 25: Resumen de datos*

Número de clientes	161.886
Número de transacciones	679.954
Venta total	MM\$ 8.425

Se tienen 161.886 clientes activos entre el 01/07/2003 y 31/12/2003. A continuación se muestra la calibración de los parámetros necesarios del modelo Pareto/NBD para el caso en estudio.

El modelo Pareto/NBD estima el número de transacciones futuras bajo el supuesto de que las tasas de compra y fuga siguen distribuciones gamma de parámetro  $r$  y  $\alpha$  y  $s$  y  $\beta$ , respectivamente. La estimación de los parámetros se realiza por máxima verosimilitud sobre los datos del período de calibración. El método recién mencionado encuentra los parámetros que mejor se ajustan a los datos observados durante el período Enero 2003 - Diciembre 2003, en otras palabras, maximiza la probabilidad conjunta de observar los datos que se tienen dado distribuciones conocidas. Esta maximización se realiza mediante la función *fmincon* incorporada en Matlab, la cual exige un rango de búsqueda y un punto de partida para cada uno de los parámetros a estimar.

En la tabla 26, se muestran los valores para los límites superior e inferior y los puntos de partidas para los parámetros a estimar.

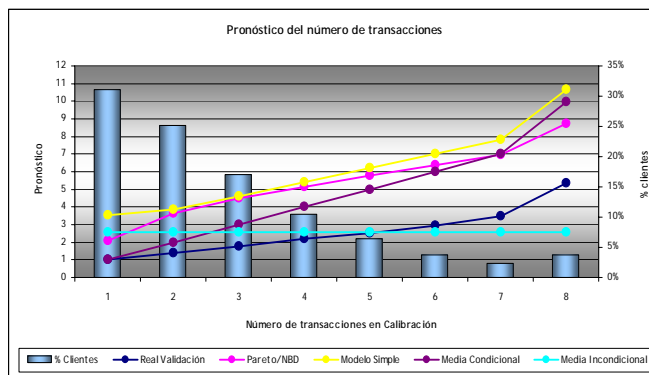
*Tabla 26: Parámetros necesarios para la calibración*

	$r$	$\alpha$	$s$	$\beta$
Punto de partida	0,01	50	0,01	10
Límite inferior	0,01	0,01	0,01	0,01
Límite superior	10	140	10	40
Valor estimado	4,0922	97,8722	0,2345	13,2554

Ahora conocidos los valores de los parámetros necesarios para el modelo Pareto/NBD, se procede a estimar el número de transacciones.

En la figura 31, se observa que los modelos, excepto por el de Media Incondicional, reflejan el hecho de que los “buenos” clientes en cuanto al número de transacciones realizados durante el período de calibración, también lo son en el período de validación, sin embargo, lo hacen por sobre lo real. Los modelos tienden a sobreestimar el número de transacciones como se muestra en la siguiente figura.

*Figura 31: Pronóstico del número de transacciones*



*Fuente: Elaboración propia*

<i>Número de transacciones</i>						
<i>Indicador</i>	<i>Real Calibración</i>	<i>Real Validación</i>	<i>Modelo Pareto/NBD</i>	<i>Modelo Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio	2,6	1,6	3,8	4,5	2,4	2,6
Desv. Est.	2,0	2,3	2,5	14,2	1,4	0,0

Para conocer la tendencia en el pronóstico de los modelos, se observa el error porcentual.

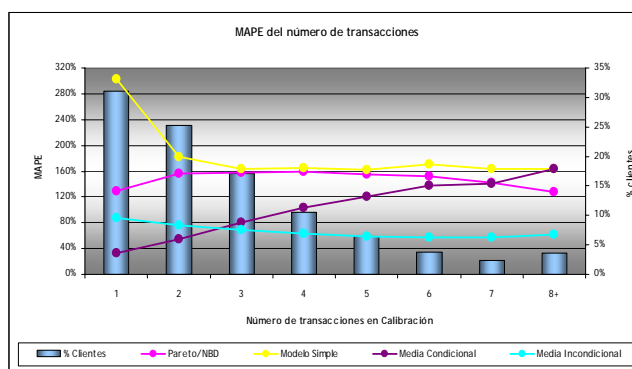
Como se aprecia en la tabla 27, los 4 modelos considerados poseen un valor positivo del MPE, lo que indica que, efectivamente, los modelos tienen tendencia a sobreestimar el número de transacciones. Según esta métrica, el modelo Media Condicional es el mejor con el menor valor del MPE, igual a 34,7% y una desviación estándar del 85,1%. En el otro extremo se encuentra el modelo Simple con un 145,6% de MPE junto con la más alta desviación estándar dentro de lo observado, a pesar de que la mediana es de 1,9%. Lo anterior, dice que el modelo Simple cuando sobreestima lo hace con una gran magnitud. El modelo Pareto/NBD es el segundo modelo que sobreestima más.

*Tabla 27: Resumen del error porcentual en el número de transacciones*

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	109,7 %	145,6 %	34,7 %	42,0 %
Desv. Est.	182,4 %	971,6 %	85,1 %	84,9 %
Mínimo	-99,8 %	-100,0 %	-90,0 %	-93,4 %
Mediana	56,6 %	1,9 %	9,6 %	28,2 %
Máximo	1.352,5 %	18.300,0 %	181,0 %	156,4 %

Como se observa en la figura 32, el modelo Pareto/NBD mantiene casi constante la curva del MAPE sin diferenciar según el tipo de cliente en cuanto al número de transacciones. También ocurre algo similar en el caso del modelo Simple, observándose este efecto a partir de los casos en que al menos realizan 2 transacciones en el período de calibración, es decir, estos 2 modelos presenta un error sistemático en el pronóstico del número de transacciones. Por otro lado, se tiene que el modelo Media Condicional predice peor el número de transacciones a medida que aumenta el número de transacciones realizados durante el segundo período del 2003, esto debido a que los clientes van disminuyendo la frecuencia de compra, siendo más drástica en el caso de los “buenos” clientes en comparación de los “malos” clientes.

*Figura 32: MAPE del número de transacciones*



*Fuente: Elaboración propia*

De la tabla 28 se observa que el modelo Media Condicional es el que posee el menor valor del MAPE, el modelo Media Incondicional es el más preciso en el pronóstico con un valor de 74,0% del MAPE y la menor desviación estándar y mediana. En el otro extremo, el modelo Simple es el peor modelo predictivo con el mayor valor del MAPE y desviación estándar. El modelo Pareto/NBD es el tercer mejor modelo.

*Tabla 28: Resumen del valor absoluto del error porcentual en el número de transacciones*

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	146,9 %	211,8 %	72,0 %	74,0 %
Desv. Est.	154,0 %	959,4 %	73,0 %	59,2 %
Mínimo	0,0 %	0,0 %	2,0 %	14,5 %
Mediana	89,0 %	91,0 %	53,3 %	35,9 %
Máximo	1.352,5 %	18.300,0 %	210,7 %	156,4 %

*Fuente: Elaboración propia*

Cabe destacar, que dentro de la información histórica en la cual se basa la estimación del número de transacciones no se consideran los clientes inactivos, es decir aquellos clientes que no realizan compras durante todo el año 2003. Dado lo anterior, las tasas de compra y fuga de los clientes del modelo Pareto/NBD se calculan sobre una muestra de datos no reales, por lo tanto, se decide incorporar este tipo de clientes en la calibración. Los clientes inactivos corresponden a personas que hayan comprado alguna vez algún producto de la empresa en estudio durante el período anterior al de calibración (es decir, que sea cliente de la compañía), pero no en éste último.

A continuación, se muestra el efecto de la incorporación de los clientes inactivos en el segundo semestre del año 2003 junto a la aplicación de una constante para eliminar el error sistemático observado.

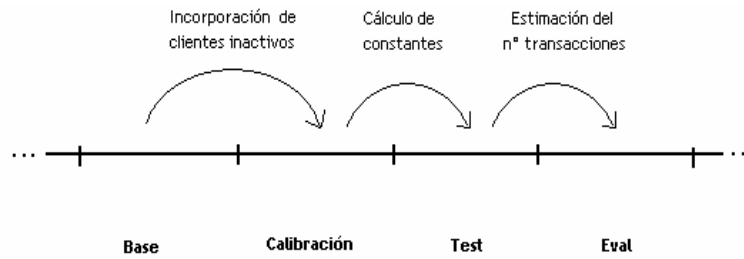
#### **4.6.1.2. Incorporación de clientes inactivos y aplicación de una constante**

Esta corresponde a la situación final (figura 33). Se agregan 47.247 casos, que corresponden a los clientes.

*Tabla 29: Períodos de observación*

Período	Fecha inicio	Fecha término
Base	02/01/2003	30/06/2003
Calibración	01/07/2003	31/12/2003
Test	02/01/2004	30/06/2004
Validación	01/07/2004	31/12/2004

*Figura 33: Diagrama de aplicar una constante*



Fuente: Elaboración propia

#### 4.6.1.2.1. Obtención de la constante

La constante a aplicar en la estimación del número de transacciones arrojadas por los modelos Pareto/NBD y Simple corresponde al valor a aplicar en las estimaciones del número de transacciones tal que los modelos en promedio posean un valor de MPE cercano a cero, es decir que no tiendan ni a sobreestimar ni subestimar. Para ello, se poseen 3 períodos.

*Tabla 30: Períodos de observación*

Período	Fecha inicio	Fecha término
Base	02/01/2003	30/06/2003
Calibración	01/07/2003	31/12/2003
Test	02/01/2004	30/06/2004

La constante se aplica de la siguiente manera:  $\hat{x}_{cc} = \hat{x}_{sc} \cdot C$

Donde:  $\hat{x}_{cc}$  = número de transacciones esperado<sup>25</sup> sin la aplicación de la constante

$\hat{x}_{sc}$  = número de transacciones esperado aplicando la constante

$C$  = valor de la constante

La estimación del número de transacciones para el período de test, sin la constante, pero sí considerando los clientes inactivos durante el período de calibración, se muestra en la siguiente tabla.

*Tabla 31: Error del pronóstico del número de transacciones*

Indicador	Modelo Pareto/NBD	Modelo Simple
MAPE	162,1 %	237,5 %
MPE	128,6 %	175,7 %

<sup>25</sup> Pero sí se considera la incorporación de clientes inactivos.

El valor de la constante por modelo se muestra en la siguiente tabla.

*Tabla 32: Valores de la constante*

Modelo	Constante
Pareto/NBD	42,8 %
Simple	31,1 %

Considerando la constante, el error arrojado por los modelos Pareto/NBD y Simple en la estimación del número de transacciones se muestra en la siguiente tabla.

*Tabla 33: Error del pronóstico del número de transacciones*

Indicador	Modelo Pareto/NBD	Modelo Simple
MAPE	65,6 %	97,6 %
MPE	-2,1 %	-4,3 %

Incorporando la constante, se logra eliminar un 96,5% del error en el caso del modelo Pareto/NBD y un 139,9% en el modelo Simple para el período de test. Ahora, dado que se conocen las constantes, se estimará el número de transacciones para el segundo semestre del 2004, período de validación.

Incorporando la constante, se observa que el modelo Pareto/NBD y Simple se ajustan más a lo que en realidad ocurre, siendo el primero el mejor como se puede apreciar en la siguiente figura.

El modelo Pareto/NBD y Simple, como se observa en la tabla 34, tienden a subestimar el número de transacciones por 11,2% y 23,6% respectivamente. Observando sólo el valor del MPE, el modelo Pareto/NBD es el mejor modelo, seguido por el modelo Simple, Media Condicional y, finalmente, el modelo Media Incondicional. Además, el modelo Pareto/NBD es el que posee la menor desviación estándar del error porcentual.

*Tabla 34: Resumen del error porcentual en el número de transacciones*

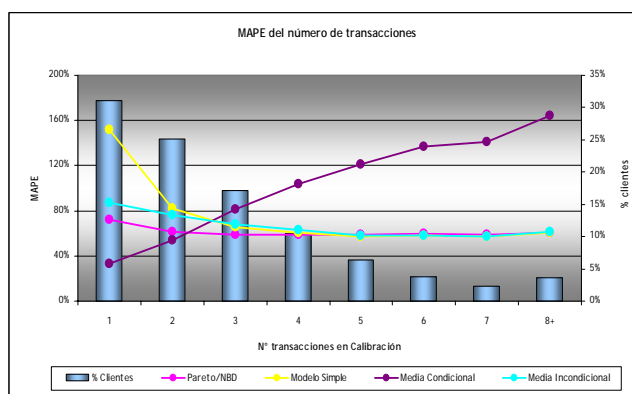
Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	-11,2 %	-23,6 %	34,7 %	42,0 %
Desv. Est.	76,5 %	302,2 %	85,1 %	84,9 %
Mínimo	-99,9 %	-100,0 %	-90,0 %	-93,4 %
Mediana	-33,3 %	-68,3 %	9,6 %	28,2 %
Máximo	502,9 %	5.622,4 %	181,0 %	156,4 %



Cabe destacar que sigue existiendo un error sistemático, sin embargo, se ha podido reducir el valor del MAPE en 83,0% en el caso del modelo Pareto/NBD y un 116,8% en el modelo Simple con respecto a la situación inicial que considera sólo los clientes activos. Como ya se explicado, el modelo Media Condicional empeora la predicción a medida que aumenta el número de transacciones realizadas por el cliente durante el período de calibración, debido a que los clientes tienen una tendencia a reducir las visitas a las sucursales de la empresa en estudio.

La métrica MAPE nos indica la precisión de pronóstico del modelo predictivo. Como se aprecia en la figura 34, se tiene que el modelo Pareto/NBD es el mejor modelo de predicción del número de transacciones a nivel de clientes dentro de los modelos estudiados.

*Figura 34: MAPE del número de transacciones*



Fuente: Elaboración propia

*Tabla 35: Resumen del valor absoluto del error porcentual en el número de transacciones*

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	63,9 %	95,0 %	72,0 %	74,0 %
Desv. Estándar	43,4 %	287,8 %	73,0 %	59,2 %
Mínimo	0,0 %	0,0 %	2,0 %	14,5 %
Mediana	59,1 %	74,7 %	53,3 %	35,9 %
Máximo	502,9 %	5.622,4 %	210,7 %	156,4 %

A continuación, se realiza un análisis por quintiles según las variables RFM con el fin de detectar las condiciones en las cuales los modelos realizan “buenos” pronósticos del número de transacciones, es decir con bajo error, y en dónde ocurre lo contrario.

#### 4.6.1.2.2.1. Análisis por quintiles

El quintil 1 corresponde al 20% de los clientes que poseen menor valor de la variable considerada, mientras que el quintil 5 representa al 20% de los clientes con mayor valor.

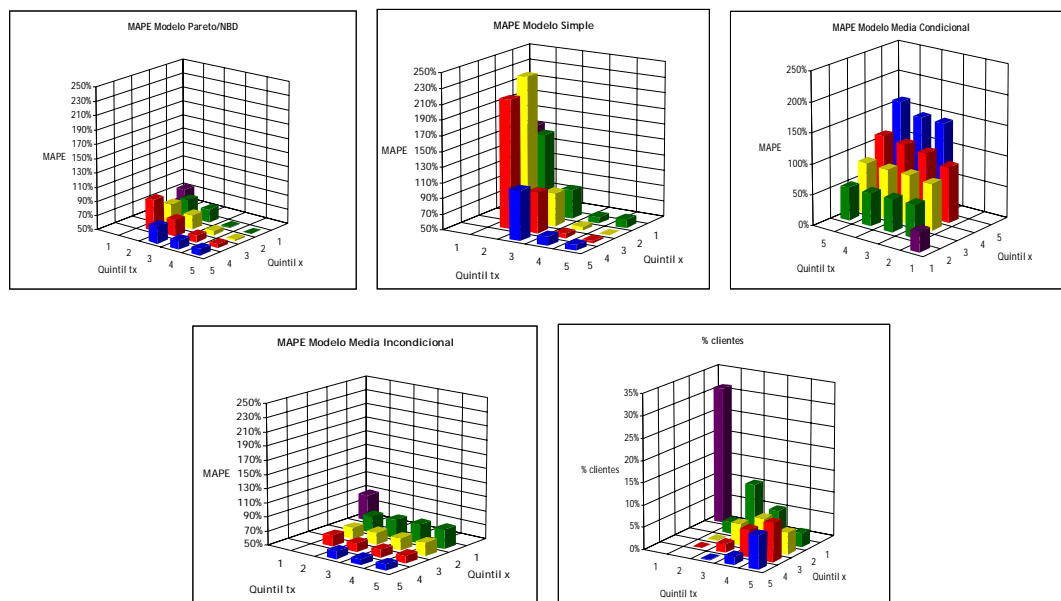
#### Quintil $x-t_x$

Las variables consideradas son la frecuencia de compra  $x$  y el período de compra del cliente  $t_x$ .

Quintil	1	2	3	4	5
Límite superior $x$	1	2	3	5	37
Límite superior $t_x$	0	5	63	124	183

El 30% de los clientes realiza 1 transacción durante los 6 meses en estudio. El modelo Pareto/NBD junto con el de Media Incondicional es el que posee los menores valores del MAPE por quintiles de  $x$  y  $t_x$ . El primero tiende a empeorar el pronóstico de  $x$  de manera moderada a medida que disminuye el período de compra del cliente. El modelo Condicional aumenta el error a medida que aumenta el número de transacciones observadas durante el segundo semestre del 2003. En cuanto al modelo Simple, el valor del MAPE se dispara cuando el período de compra del cliente es menor a 1 semana (quintil 1 y 2 de  $t_x$ ), lo que nos indica que este modelo no predice bien este indicador cuando el cliente realiza compras "puntuales".

*Figura 35: MAPE del número de transacciones por quintil de  $x-t_x$*



Fuente: Elaboración propia

## Quintil x-T

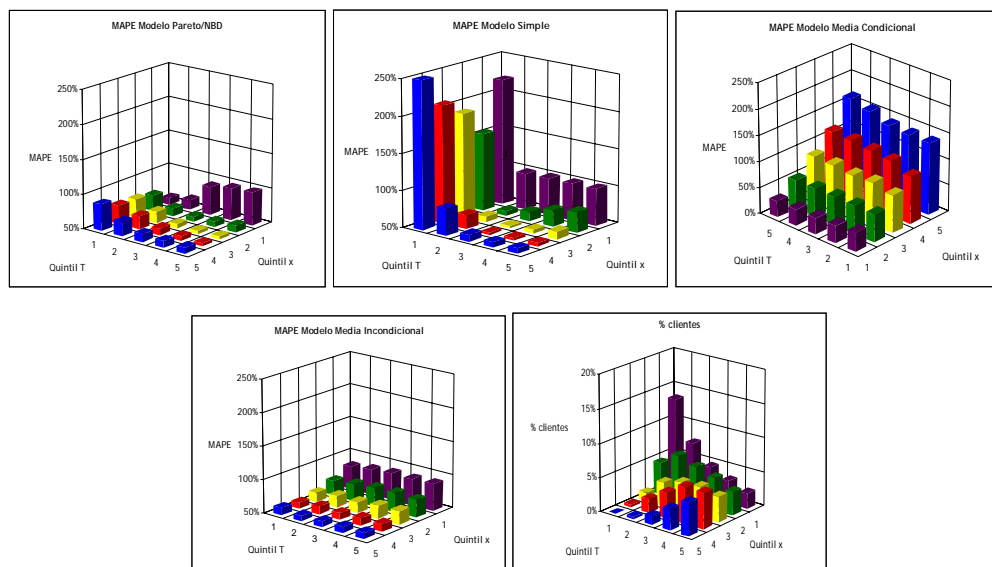
Considerando el número de transacciones  $x$  y período de observación del cliente  $T$ .

Quintil	1	2	3	4	5
Límite superior $x$	1	2	3	5	37
Límite superior $T$	40	95	136	166	183

Los modelos Pareto/NBD y Simple pronostican con bajo error el número de transacciones cuando éste es mayor que una unidad y el período de observación mayor a 40 días, sin embargo, este hecho se observa con mayor fuerza en el segundo modelo mencionado.

El modelo Media Condicional comete mayor error cuando aumenta el número de transacciones observado durante el segundo semestre del 2003, mientras que en el caso del modelo Media Incondicional ocurre lo contrario, es decir, disminuye el error a medida que aumenta  $x$  debido a que realiza un único pronóstico para todos los clientes independiente del comportamiento de compra de cada uno de ellos. No distinguen según el período de observación del cliente  $T$ .

*Figura 36: MAPE del número de transacciones por quintil de  $x-T$*



Fuente: Elaboración propia

## Quintil x-M

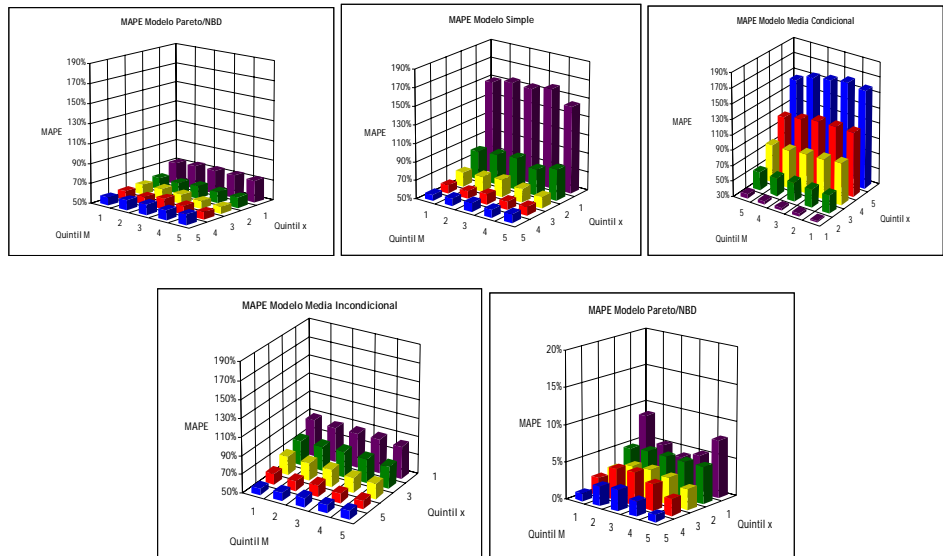
Se consideran las variables número de transacciones  $x$  y monto promedio por transacción del cliente  $M$ .

Quintil	1	2	3	4	5
Límite superior $x$	1	2	3	5	37
Límite superior $M$	\$5.896	\$9.145	\$12.980	\$18.909	\$137.802

El modelo Pareto/NBD se equivoca de manera casi indistinguible, excepto en el caso de los clientes que realizan una única transacción (quintil 1 de  $x$ ), en los cuales se observa un aumento cerca del 20% del valor del MAPE.

El modelo Simple y el de Media Incondicional realizan mejores pronósticos de  $x$  cuando se trata de “buenos” clientes, sin embargo, el cambio de magnitud del error del quintil 1 al 2 de  $x$  es mayor en el modelo Simple. Caso contrario, ocurre en el caso del modelo Media Condicional.

*Figura 37: MAPE del número de transacciones por quintil de  $x$ - $M$*



*Fuente: Elaboración propia*

#### 4.6.1.3. Conclusiones de la estimación del número de transacciones

Del análisis de la estimación del número de transacciones ( $x$ ) es necesario considerar la existencia de clientes inactivos y una constante que elimine el error sistemático presente en los modelos Pareto/NBD y Simple.

Incorporando las modificaciones mencionadas, se logra reducir el valor del MAPE asociado al pronóstico de  $x$  en 83,0% (equivale a un 56,5% del valor inicial) y en 116,8% (equivale a un 55,1% del valor inicial) respectivamente con respecto a la situación inicial en la que se consideran sólo los clientes activos.

Se observa que los clientes que, ahora, son buenos en cuanto al número de transacciones realizadas seguirán siéndolos en el período siguiente. Los modelos considerados capturan esta tendencia.

Los modelos Pareto/NBD y Simple predicen con bajo error el número de transacciones cuando se observa que el cliente visita más de una vez alguna sucursal de la empresa en estudio ( $x > 1$  transacción) y el período de compra del cliente es mayor a 1 semana. Es decir, cometen altos errores en el pronóstico del número de transacciones en los casos de clientes que realizan pocas compras y puntuales, siendo más brusco el aumento del error en el caso del modelo Simple. También, cometen un mayor error en el caso de los clientes nuevos, es decir, cuyo período de observación ( $T$ ) es menor o igual a 40 días (quintil 1 de  $T$ ).

Por otro lado, el modelo Media Incondicional mejora la predicción de  $x$  a medida que aumenta el número de transacciones observado para cada uno de los clientes.

Finalmente, el modelo Pareto/NBD es el mejor modelo de pronóstico del número de transacciones con un error de 63,9% y una desviación estándar de 43,4%.

De ahora en adelante, se consideran los resultados de este último caso, en la cual se incluyen clientes inactivos y una constante.

#### 4.6.2. Estimación del monto por transacción

Los modelos Pareto/NBD y Simple consideran que el monto promedio por transacción de cada cliente es igual al monto promedio por transacción observado ( $M$ ). Por otro lado, para el modelo de Media Incondicional, se tiene que el valor esperado de este indicador es el valor promedio del monto promedio gastado en cada compra de todos los clientes. Los períodos considerados corresponden al segundo semestre del año 2003 y 2004.

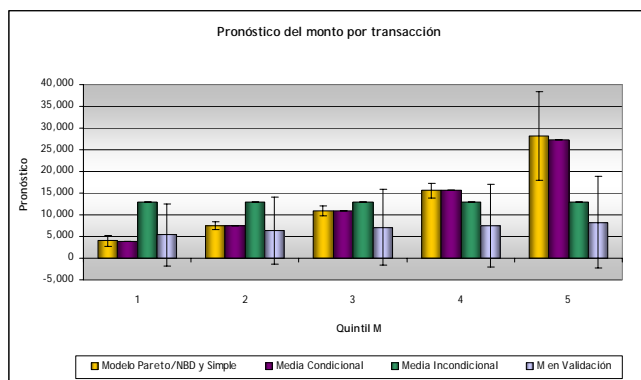
Según el modelo Pareto/NBD, la estimación de  $M$  de un cliente dado corresponde al valor del monto promedio por transacción observado entre el 01/07/2003 y el 31/12/2003. El modelo Media Condicional es considerado condicional al quintil de  $M$ , luego las estimaciones según este modelo son 5 como se muestran en la siguiente tabla.

Quintil $M$	1	2	3	4	5
$M$	3.954	7.522	10.981	15.605	27.246

Como se observa en la figura 38, se tiene que un cliente promedio reduce casi a la mitad el monto gastado por transacción de un período a otro, igual a lo que sucede con el

número de transacciones observado. En promedio, se sobreestima  $M$ , esto debido a que hay una tendencia a la baja en cuanto a este indicador.

*Figura 38: Pronóstico del monto por transacción por quintil de  $M$*

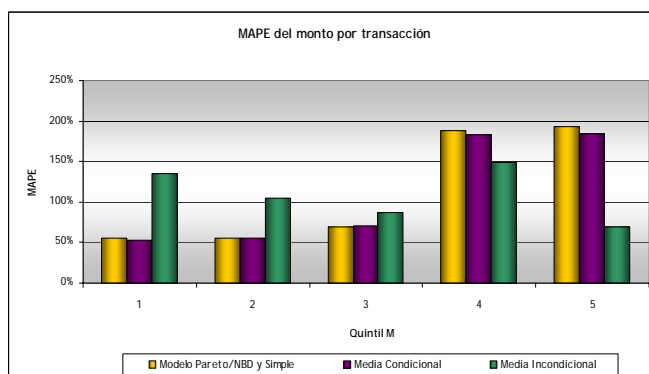


*Fuente: Elaboración propia*

<i>Monto promedio</i>	<i>Real Calibración</i>	<i>Real Validación</i>	<i>Modelo Pareto/NBD y Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio	13.061	6.947	13.061	13.061	13.061
Desv. Est.	9.077	8.863	9.077	8.070	0

El valor del MAPE de los modelos es similar, sin embargo, los modelos Pareto/NBD y Simple son los menos precisos en cuanto al monto promedio por transacción. Se puede ver que el mayor error en la predicción del monto promedio por transacción ocurre en los quintiles 4 y 5, que corresponden a los mejores clientes en cuanto a  $M$ .

*Figura 39: MAPE por quintil de  $M$*



*Fuente: Elaboración propia*

*Tabla 36: Resumen del valor absoluto del error porcentual en estimación de M*

ABS(Error Porcentual)			
<i>Indicador</i>	<i>Modelo Pareto/NBD y Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio (MAPE)	110,0 %	107,0 %	109,1 %
Desv. Estándar	5.480,4 %	2.357,5 %	4.301,4 %
Mínimo	0,0 %	0,0 %	0,0 %
Mediana	49,2 %	56,1 %	47,7 %
Máximo	1.664.000,0 %	304.653,8 %	1.305.970,2 %

El decir que un cliente que va a gastar lo mismo que el promedio de  $M$  de todos los clientes es mejor que decir que va a gastar lo mismo.

#### ***4.6.3. Cálculo del Customer Lifetime Value (CLV)***

La estimación del *CLV* se realiza a partir de la estimación del número de transacciones y del monto promedio por transacción. Sin embargo, el indicador modelado corresponde al número de transacciones promedio de cada uno de los clientes, siendo la mejor predicción generada por el modelo Pareto/NBD.

El modelo Simple es el que estima un valor promedio del *CLV* menor, siendo el de Media Incondicional el que posee el mayor valor de éste. Sin embargo, dado que el modelo Pareto/NBD es el que estima con un menor error el número de transacciones promedio a nivel individual, se puede decir que éste es el modelo que estima el valor del *CLV* a nivel de clientes con mayor confianza.

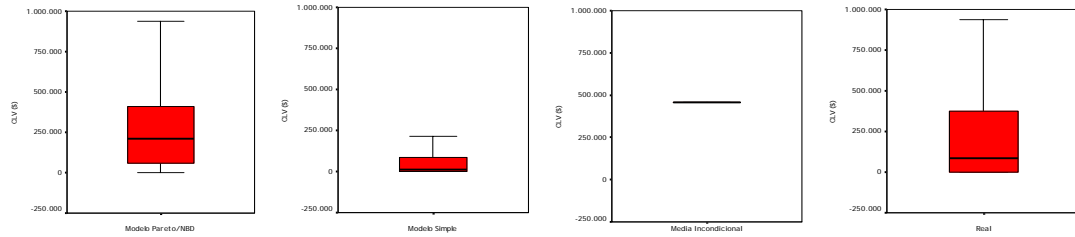
A pesar de que no se conoce el valor real del *CLV*, se puede aproximar diciendo que el cliente siempre va a tener el mismo comportamiento de compra observado durante el período de validación (segundo semestre del 2004), el cual será identificado como "Real"<sup>26</sup>.

Si consideramos que un cliente siempre va a realizar el mismo número de transacciones y monto por transacción observado en el segundo semestre del 2004, un cliente promedio vale \$267.372 para la empresa.

Cabe destacar que la mediana, \$85.957, es un valor pequeño en comparación al valor promedio, por lo tanto, existen clientes que poseen un alto valor del *CLV*, es decir, son clientes muy valiosos para la empresa (figura 40).

<sup>26</sup> No corresponde al valor real del *CLV*, es sólo un identificador.

*Figura 40: Boxplot del CLV*



*Fuente: Elaboración propia*

CLV				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Incondicional	Real
Promedio	278.778	129.963	457.156	267.372
Desv. Estándar	279.199	966.085	0	437.378
Mínimo	0	0	457.156	0
Mediana	209.831	13.657	457.156	85.957
Máximo	4.223.533	63.715.318	457.156	12.834.078

Ahora que se tienen los valores del CLV, se puede obtener el valor que aporta cada uno de los clientes, en promedio, en un período de 1 año. Se observa que el valor real y del modelo Pareto/NBD son muy similares.

*Tabla 37: Resumen del CLV anualizado*

Indicador	Modelo Pareto/NBD	Modelo Simple	Media Incondicional	Real
Promedio	6.969	3.249	11.429	6.684
Desv. Estándar	6.980	24.152	0	10.934
Mínimo	0	0	11.429	0
Mediana	5.246	341	11.429	2.149
Máximo	105.590	1.592.900	11.429	320.850

#### **4.6.3.1. Error asociado al cálculo del CLV**

Al valor estimado del *CLV* no se le puede realizar un análisis de errores, ya que no se cuenta con el valor real de éste, sin embargo, se puede estimar el error asociado a éste.

El error asociado al número de transacciones ( $x$ ) y monto promedio por transacción ( $M$ ) se muestra a continuación.



*Tabla 38: Resumen del MAPE*

Modelo	MAPE de x	MAPE de M
Pareto/NBD	63,9 %	109,1 %
Simple	95,0 %	
Media Incondicional	74,0 %	

Luego, el error asociado al *CLV* por modelo es tal que:

*Tabla 39: Estimación del MAPE del CLV*

Modelo	MAPE de <i>CLV</i>
Pareto/NBD	951,6 %
Simple	1.414,8 %
Media Incondicional	1.102,1 %

La estimación del error del valor del *CLV* de los modelos es alto dado que los errores de *x* y *M* son altos. Dada las condiciones anteriores, el modelo Pareto/NBD es el más confiable en comparación a los otros dos.

#### **4.6.3.2. Ranking de clientes**

A continuación se analizan al 10% de los mejores clientes, que corresponden a 16.190 casos, según el valor del *CLV* calculado por el modelo Pareto/NBD, Simple y Real.

De los 10% de clientes más valiosos según cada uno de los modelos, se tiene que este grupo vale \$894.819 según el modelo Pareto/NBD, \$977.059 según el modelo Simple y \$1.286.513 asumiendo que mantienen el comportamiento de compra. Esto último se debe a que no considera la posibilidad de que reduzcan las compras junto a la fuga de clientes.

*Tabla 40: Resumen de indicadores*

Modelo	Indicador	Período de Calibración					Período de Validación			CLV
		x	tx	T	M	Monto	x	M	Monto	
Pareto/NBD	Promedio	3,6	70	82	27.135	82.560	2,3	10.032	35.242	894.819
	Desv. Est.	3,1	63	63	10.613	50.925	3,2	11.054	51.695	275.383
Simple	Promedio	5,1	102	107	17.553	82.479	2,9	9.568	40.168	977.059
	Desv. Est.	3,3	61	63	9.084	51.240	3,7	10.054	55.279	2.917.528
Real	Promedio	4,1	89	117	15.477	58.155	5,9	19.095	94.247	1.286.513
	Desv. Est.	3,2	62	57	10.036	48.100	3,4	10.211	45.286	618.177

Dentro de los clientes más valiosos según el modelo Pareto/NBD, corresponden a clientes

nuevos (período de observación  $T$  es menor o igual a 40 días y antigüedad menor o igual a 5 años). Además, cerca del 45% realizaron 1 ó 2 transacciones durante el período de calibración. Cabe destacar que un poco más del 80% de los clientes valiosos son aquellos que gastan mucho dinero en cada una de sus transacciones. Por otro lado, los clientes valiosos según el modelo Simple y el comportamiento pasado (Real) se distribuyen de manera similar.

Del 10% de los clientes más valiosos para la empresa, se ve que la asignación del modelo Pareto/NBD y Simple coinciden en el 48,1% de los casos, siendo menor el porcentaje de coincidencia al compararlo con la asignación realizada al considerar que se mantiene el comportamiento de compra de los clientes observado durante el segundo semestre del 2004 ( $x$  y  $M$ ).

*Tabla 41: Asignación de los clientes según CLV*

<i>Modelo</i>	<i>Pareto/NBD</i>	<i>Simple</i>	<i>Real</i>
<i>Pareto/NBD</i>	100,0%	48,1%	23,1%
<i>Simple</i>	48,1%	100,0%	26,2%
<i>Real</i>	23,1%	26,2%	100,0%

#### **4.6.4. Conclusiones del uso sólo de los atributos RFM**

##### **4.6.4.1. Conclusiones de la estimación del número de transacciones**

Los modelos estudiados, excepto por el de Media Incondicional, siguen la tendencia real de que los buenos clientes hoy, también lo serán en el futuro. Para el caso del mejor modelo de predicción, es decir, el modelo Pareto/NBD se tienen las siguientes probabilidades de cambio de estado, como se observa en la tabla 42.

*Tabla 42: Probabilidad de cambio de estado*

<i>Modelo</i>	<i>Bueno</i>	<i>Normal</i>	<i>Malo</i>
<i>Bueno</i>	68,9 %	28,0 %	3,1 %
<i>Normal</i>	39,9 %	44,8 %	15,3 %
<i>Malo</i>	13,6 %	47,8 %	38,6 %

La incorporación de clientes inactivos y la aplicación de una constante a las estimaciones del modelo Pareto/NBD y Simple permiten mejorar el pronóstico del número de transacciones a nivel individual reduciendo el error asociado a él con respecto a la situación inicial, en la que se consideran sólo los clientes activos durante el período de calibración considerado. El valor del MAPE se reduce en 83% y 116,8% respectivamente. Considerando la situación final, se tiene que el modelo Pareto/NBD es el más preciso en el

pronóstico del número de transacciones de cada uno de los clientes cometiendo un error de 63,9%.

#### **4.6.4.2. Conclusiones del cálculo del CLV**

Los modelos Pareto/NBD y Simple aumentan considerablemente el valor del *CLV* en el caso de los clientes nuevos, es decir, cuando la primera compra fue realizada hace a lo más 40 días. Lo anterior se debe a que los modelos no pronostican bien el número de transacciones, sobreestimándolo en gran magnitud.

Se valoran más a los clientes “buenos” en la actualidad, es decir, aquellos que compran más seguido y, a su vez, gastan más por cada visita o transacción. También, el valor del *CLV* aumenta a medida que aumenta el período de compra del cliente.

El modelo Pareto/NBD es el mejor modelo, a pesar de tener asociado valores altos de error, seguido por el modelo Simple debido a que es el modelo que pronostica el número de transacciones con el menor error.

### **4.7. Estimación del Customer Lifetime Value usando variables demográficas y transaccionales**

#### ***4.7.1. Estimación del número de transacciones***

Se muestran los 2 casos: el primero corresponde al caso en que se consideran sólo los clientes activos durante el período de calibración y el segundo consiste en la incorporación de clientes inactivos junto a la aplicación de una constante

##### **4.7.1.1. Presencia sólo de clientes activos**

Esta es la situación en que sólo se consideran los clientes activos durante el segundo semestre del 2003 y son a éstos a quienes se les estima el comportamiento de compra.

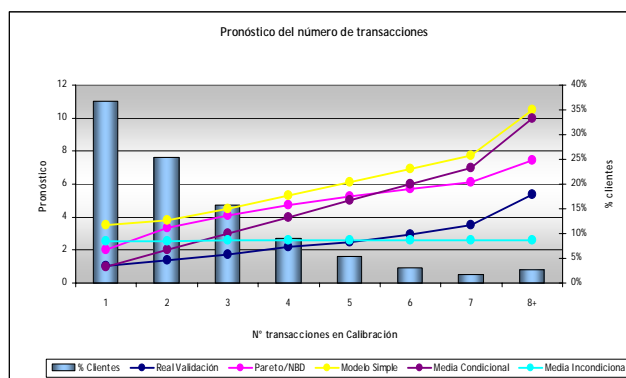
*Tabla 43: Resumen de datos*

Número de clientes	161.886
Número de transacciones	679.954
Venta total	MM\$ 8.425

De la figura 41 se observa que los modelos, excepto por el de Media Incondicional, reflejan el hecho de que los “buenos” clientes en cuanto al número de transacciones

realizados durante el período de calibración, también lo son en el período de validación. Los modelos tienden a sobreestimar el número de transacciones.

*Figura 41: Pronóstico del número de transacciones*



*Fuente: Elaboración propia*

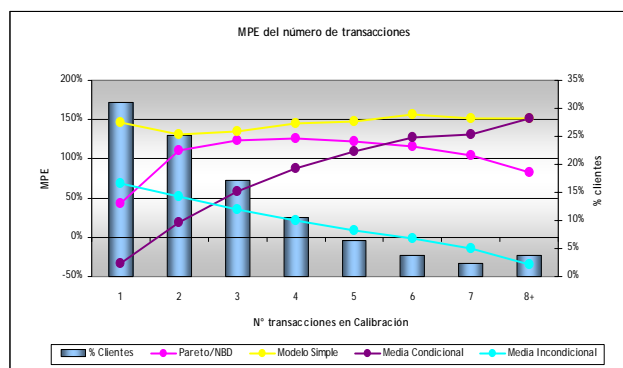
Un cliente promedio realiza 2,6 transacciones en el período de calibración y lo reduce a 1,6 transacciones en el de validación, mientras que el modelo Pareto/NBD espera que realiza 3,4 transacciones y el modelo Simple 4,4 transacciones.

Número de transacciones						
Indicador	Real Calibración	Real Validación	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio	2,6	1,6	3,4	4,4	2,4	2,6
Desv. Est.	2,0	2,3	2,2	14,0	1,4	0,2

Como se aprecia en la figura 42, los valores promedio del MPE son positivos para los modelos en estudio, luego, los modelos tienen tendencia a sobreestimar el número de transacciones.

El modelo Media Condicional es el mejor con el menor valor del MPE, igual a 34,7% y una desviación estándar del 85,1%. En el otro extremo se encuentra el modelo Simple con un 145,6% de MPE junto con la más alta desviación estándar dentro de lo observado, a pesar de que la mediana es de 1,9%. Lo anterior, dice que el modelo Simple cuando sobreestima lo hace en una gran magnitud. El modelo Pareto/NBD es el que tercer modelo que sobreestima menos.

*Figura 42: MPE del número de transacciones*



*Fuente: Elaboración propia*

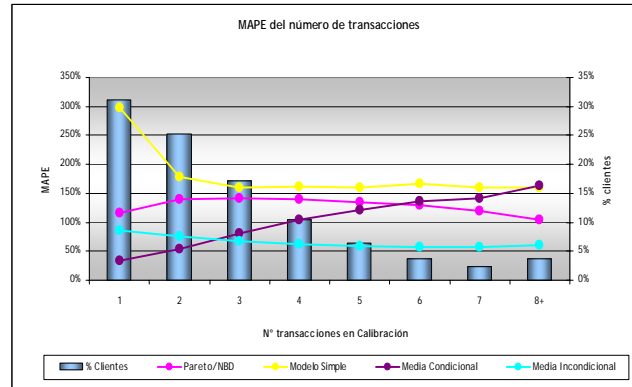
Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	93,1 %	145,6 %	34,7 %	41,6 %
Desv. Est.	161,8 %	971,6 %	85,1 %	84,9 %
Mínimo	-99,8 %	-100,0 %	-90,0 %	-92,7 %
Mediana	46,4 %	1,9 %	9,6 %	22,6 %
Máximo	1.206,6 %	18.300,0 %	181,0 %	193,1 %

Como se observa en la figura 43, la curva del MAPE del modelo Pareto/NBD es casi constante sin distinguir por el número de transacciones realizadas. En el caso del modelo Simple, también, se observa esta tendencia a partir de los casos en que al menos realizan 2 transacciones en el período de calibración. Dado lo anterior, se observa que existe un error sistemático asociado al pronóstico del número de transacciones realizados por los modelos Pareto/NBD y Simple.

Por otro lado, se tiene que el modelo Media Condicional predice peor el número de transacciones a medida que aumenta el número de transacciones realizados durante el segundo período del 2003, esto debido a que los clientes van disminuyendo la frecuencia de compra, siendo más drástica en el caso de los "buenos" clientes en comparación de los "malos" clientes.

Se tiene que el modelo Media Condicional es el que posee el menor valor del MAPE, luego éste es el más preciso en el pronóstico del número de transacciones. En el otro extremo, el modelo Simple es el peor modelo predictivo cometiendo un error de 211,8% y una desviación estándar de 959,4%. El modelo Pareto/NBD es el tercer mejor modelo.

Figura 43: MAPE del número de transacciones

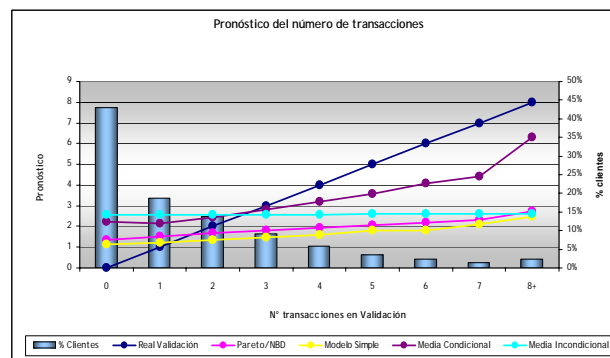


Fuente: Elaboración propia

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	130,3 %	211,8 %	72,0 %	73,2 %
Desv. Est.	133,7 %	959,4 %	73,0 %	59,8 %
Mínimo	0,0 %	0,0 %	2,0 %	2,3 %
Mediana	82,9 %	91,0 %	53,3 %	44,0 %
Máximo	1.206,6 %	18.300,0 %	210,7 %	193,1 %

Considerando el eje x como el número de transacciones durante el período de validación, se aprecia que a los clientes que van menos seguido a las tiendas de la empresa en estudio, se les sobreestima el número de transacciones, subestimando en el resto de los casos.

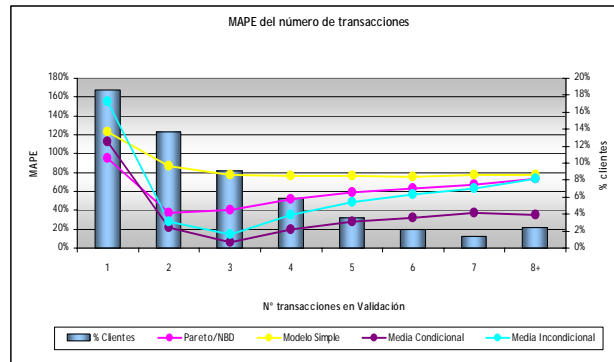
Figura 44: Pronóstico del número de transacciones



Fuente: Elaboración propia

Como se ve en la figura siguiente, los modelos son menos precisos para los clientes que realizan una única transacción.

*Figura 45: MAPE del número de transacciones*



*Fuente: Elaboración propia*

Al igual que en el capítulo anterior, se tiene que no se consideran los clientes que no realizan transacciones durante el segundo semestre del año 2003, período de calibración. Además, se observa un error sistemático en los pronósticos del modelo Pareto/NBD y Simple, la cual se intenta eliminar con la aplicación de una constante en la estimación del número de transacciones.

A continuación, se muestra el efecto de la incorporación de los clientes inactivos en el segundo semestre del año 2003 junto a la constante tanto para el modelo Pareto/NBD y Simple.

#### **4.7.1.2. Incorporación de clientes inactivos y aplicación de una constante**

Los errores en la estimación del número de transacciones incorporando cada una de las 5 variables seleccionadas de manera independiente se muestra en el Anexo 4.

Cabe destacar que se consideran los clientes inactivos durante el período de calibración y una constante. La obtención de la constante para este caso se detalla a continuación.

##### **4.7.1.2.1. Obtención de la constante**

La constante a aplicar en las estimaciones arrojadas por los modelos Pareto/NBD y Simple corresponde al valor a aplicar en las estimaciones del número de transacciones tal que los modelos en promedio posean un valor de MPE cercano a cero, es decir que no tiendan ni a sobreestimar ni subestimar.

La estimación del número de transacciones para el período 02/01/2004~30/03/2004, sin la constante, pero sí considerando los clientes inactivos durante el período de calibración, se muestra en la siguiente tabla.

*Tabla 44: Error del pronóstico del número de transacciones*

Indicador	Modelo Pareto/NBD	Modelo Simple
MAPE	146,4 %	237,5 %
MPE	113,5 %	175,7 %

Después de varias pruebas, el valor de la constante por modelo para cada uno de los grupos en estudio considerando las 3 variables seleccionadas, se muestra a continuación.

*Tabla 45: Valores de la constante*

Región	Edad	Masculino		Femenino	
		Modelo Pareto/NBD	Modelo Simple	Modelo Pareto/NBD	Modelo Simple
Metropolitana	21-34	43,6 %	25,3 %	41,2 %	25,4 %
	35-54	44,5 %	29,8 %	42,5 %	32,4 %
	55-85	46,8 %	51,6 %	44,2 %	43,9 %
Valparaíso	21-34	43,5 %	29,1 %	41,9 %	27,9 %
	35-54	45,9 %	37,2 %	45,2 %	32,6 %
	55-85	44,3 %	44,5 %	46,0 %	45,1 %
Bío-Bío	21-34	44,9 %	22,6 %	45,3 %	24,1 %
	35-54	49,0 %	32,0 %	48,7 %	42,4 %
	55-85	52,2 %	31,1 %	48,3 %	46,6 %
Los Lagos	21-34	49,9 %	19,0 %	48,4 %	29,1 %
	35-54	52,4 %	32,9 %	50,5 %	42,9 %
	55-85	51,2 %	27,8 %	51,3 %	45,9 %
Otros	21-34	42,9 %	7,2 %	44,5 %	26,4 %
	35-54	47,0 %	32,7 %	48,3 %	34,7 %
	55-85	47,5 %	26,7 %	47,7 %	40,2 %

Considerando la constante, el error arrojado por los modelos Pareto/NBD y Simple en la estimación del número de transacciones se muestra en la siguiente tabla.

*Tabla 46: Error del pronóstico del número de transacciones*

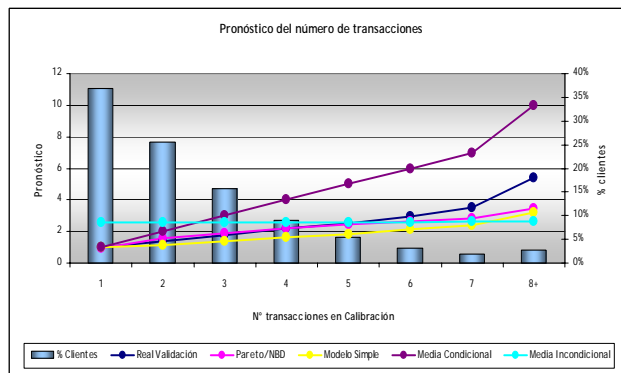
Indicador	Modelo Pareto/NBD	Modelo Simple
MAPE	64,3 %	96,8 %
MPE	-0,8 %	-7,6 %

Se logra eliminar un 82,1% del error en el caso del modelo Pareto/NBD y un 140,7% en el modelo Simple para el período de test. Ahora, dado que se conocen las constantes, se estimará el número de transacciones para el segundo semestre del 2004.



Como se ve en la figura 46, los clientes “buenos” durante el segundo semestre del 2003 siguen siendo “buenos” en el período de validación. Los modelos siguen esta tendencia, siendo el más notorio en el caso del modelo Media Condicional. Los modelos Pareto/NBD y Simple subestiman el número de transacciones. Un cliente promedio realiza 2,6 transacciones durante el período de calibración reduciéndolo en una unidad en el de validación. Un cliente promedio según el modelo Pareto/NBD es igual al real en cuanto al número de visitas realizadas a alguna sucursal de la empresa en estudio.

*Figura 46: Pronóstico del número de transacciones*



*Fuente: Elaboración propia*

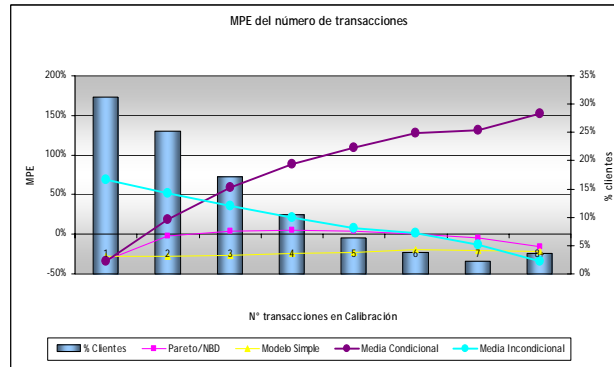
<i>Número de transacciones</i>						
<i>Indicador</i>	<i>Real Calibración</i>	<i>Real Validación</i>	<i>Modelo Pareto/NBD</i>	<i>Modelo Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio	2,6	1,6	1,6	1,3	2,6	2,6
Desv. Est.	2,0	2,3	1,0	4,2	0,2	0,2

Como se observa en la figura 47, el modelo Pareto/NBD, en general, posee un valor del MPE de 0,0%, excepto en el caso de los clientes que realizan una única transacción, en donde tiende a subestimar. El modelo Simple subestima casi en la misma proporción el número de transacciones independiente del tipo de clientes.

Por otro lado, el modelo Media Condicional sobreestima más este indicador a medida que aumenta el número de transacciones realizado por los clientes durante el segundo semestre del año 2003, ocurre lo contrario con el modelo Media Incondicional.

El modelo Pareto/NBD es el que posee el menor valor del MPE junto a la menor desviación estándar del error porcentual, seguido por el modelo Media Condicional.

*Figura 47: MPE del número de transacciones*

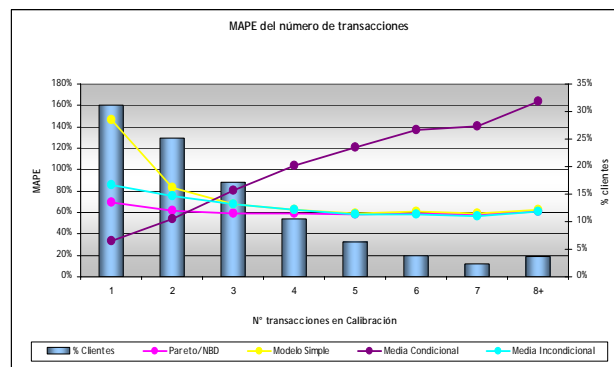


*Fuente: Elaboración propia*

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	-10,0 %	-26,4 %	34,7 %	41,6 %
Desv. Est.	75,2 %	291,4 %	85,1 %	84,9 %
Mínimo	-99,9 %	-100,0 %	-90,0 %	-92,7 %
Mediana	-31,6 %	-70,3 %	9,6 %	22,6 %
Máximo	488,7 %	8.334,6 %	181,0 %	193,1 %

La incorporación de la constante no absorbe todo el error sistemático asociado a los modelos Pareto/NBD y Simple, como se había visto en el caso en que se usan sólo los atributos *RFM*. Como se aprecia en la figura 48, para el caso en que los clientes realizan 1 transacción durante el período de calibración se tiene que el modelo Simple predice el número de visitas a alguna sucursal de la empresa en estudio con un error mucho más alto en comparación al resto.

*Figura 48: MAPE del número de transacciones*



*Fuente: Elaboración propia*

Entre los modelos en estudio, el modelo Pareto/NBD es el más preciso al momento de

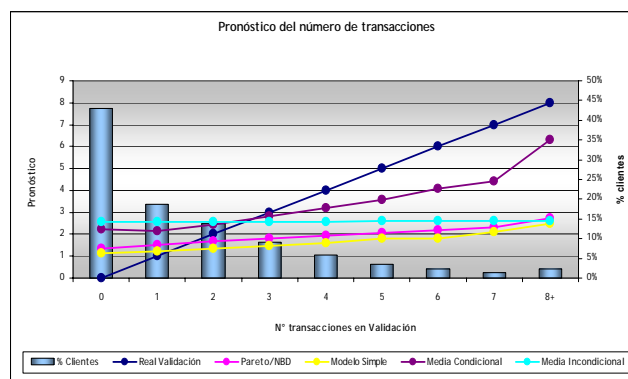
pronosticar el número de transacciones a nivel individual cometiendo un error de 63,0%. El modelo Simple es el menos exacto con un error de 94,6% y la mayor desviación estándar (tabla 47).

*Tabla 47: Resumen del valor absoluto del error porcentual en el número de transacciones*

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	63,0 %	94,6 %	72,0 %	73,2 %
Desv. Estándar	42,2 %	276,8 %	73,0 %	59,8 %
Mínimo	0,0 %	0,0 %	2,0 %	2,3 %
Mediana	58,3 %	76,9 %	53,3 %	44,0 %
Máximo	488,7 %	8.334,6 %	210,7 %	193,1 %

Ahora, considerando en el eje x el número de transacciones realizadas durante el período de validación, se observa que cerca del 45,0% de los clientes no realiza ninguna transacción. Los modelos en la mayoría de los casos, subestiman el número de transacciones futuro de un cliente.

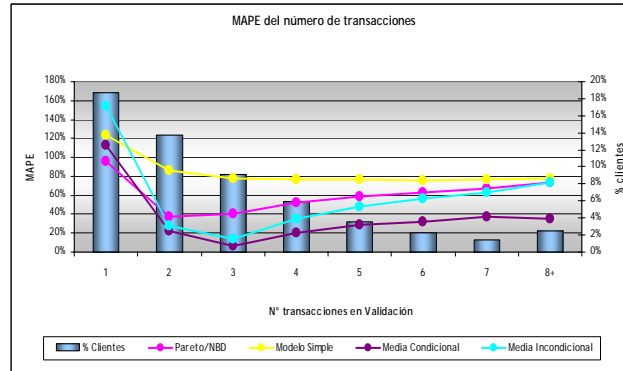
*Figura 49: Pronóstico del número de transacciones*



*Fuente: Elaboración propia*

En la siguiente figura, no se observan los clientes que no realizan transacciones, porque en estos casos, el valor del error porcentual no puede ser calculado. Gran parte del error se produce en los clientes que, efectivamente, realizan pocas transacciones.

*Figura 50: MAPE del número de transacciones*



*Fuente: Elaboración propia*

A continuación, se realiza un análisis por quintiles con el fin de identificar los grupos de clientes en los cuáles los modelos predicen el número de transacciones con mayor precisión y en cuáles no.

#### 4.7.1.2.2. Análisis por quintiles

##### Quintil $x-t_x$

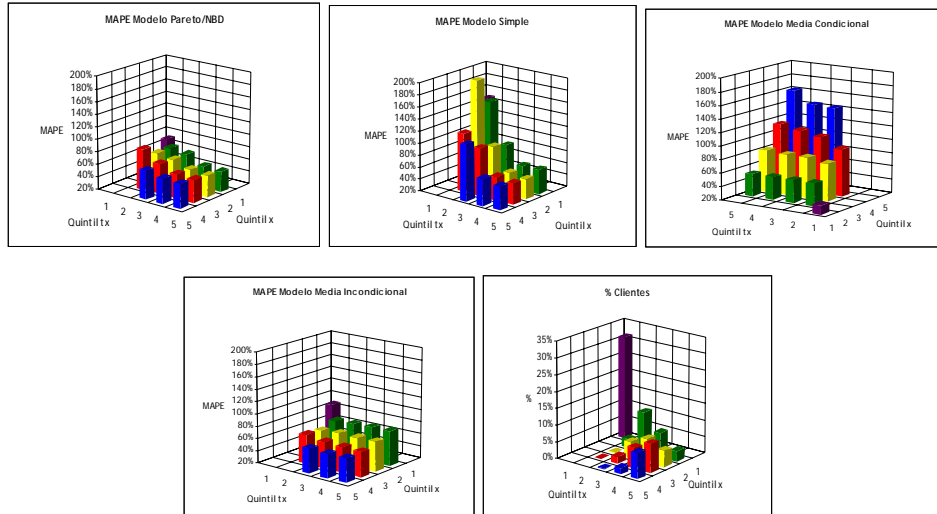
Las variables son la frecuencia de compra  $x$  y el período de compra del cliente  $t_x$ .

Quintil	1	2	3	4	5
Límite superior $x$	1	2	3	5	37
Límite superior $t_x$	0	5	63	124	183

Como se observa en los gráficos de la figura 50, un poco más del 30% de los clientes realiza una única transacción durante los 6 meses en estudio. El modelo Pareto/NBD junto con el de Media Incondicional son los que poseen valores del MAPE casi constante en cada uno de los grupos de clientes que se forman por quintiles de  $x$  y  $t_x$ .

Por otro lado, el modelo Media Condicional aumenta el error a medida que aumenta el número de transacciones observadas durante el segundo semestre del 2003. En cuanto al modelo Simple, se tiene que el error aumenta a medida que el período de compra del clientes ( $t_x$ ) es cada vez menor, luego, este último modelo no es preciso en su pronóstico del número de transacciones cuando el cliente se mantiene comprando por poco tiempo, es decir, en casos de clientes que realizan compras "puntuales".

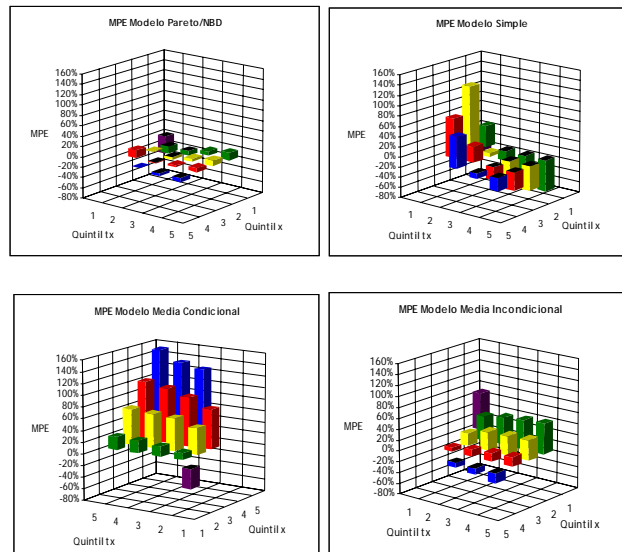
Figura 50: MAPE del número de transacciones por quintil de  $x-tx$



Fuente: Elaboración propia

El modelo Pareto/NBD es el que posee magnitudes del MPE más cercanos al 0%, lo que nos muestra que el modelo compensa la sobreestimación del número de transacciones con la subestimación y viceversa. El modelo Simple tiende a sobreestimar cuando los clientes se mantienen poco tiempo comprando y a subestimar en caso contrario. El modelo Media Condicional tiende a sobreestimar con grandes valores de éste.

Figura 51: MPE del número de transacciones por quintil de  $x-tx$



Fuente: Elaboración propia

## Quintil x-T

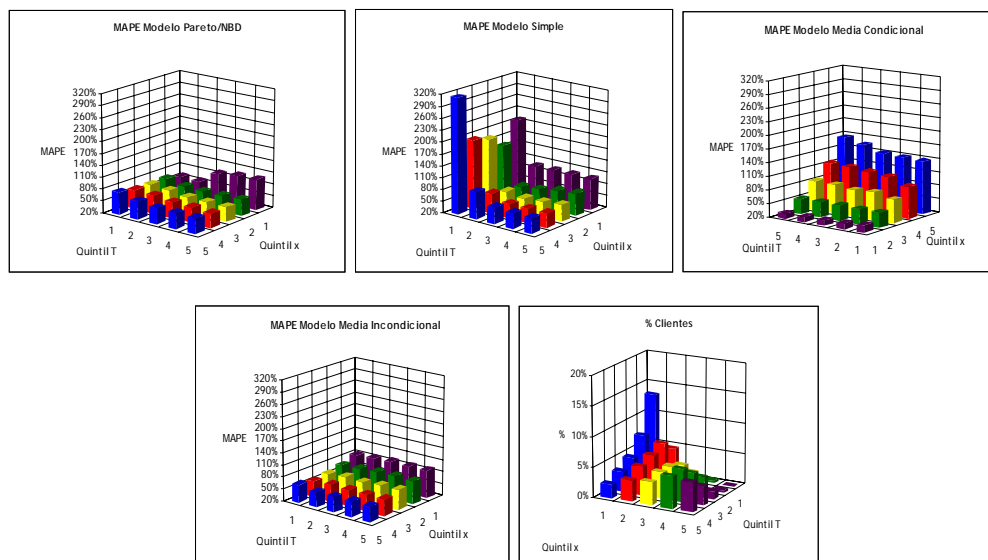
Otra variable interesante de observar es el período de observación del cliente  $T$  junto al

número de transacciones observados  $x$ .

Quintil	1	2	3	4	5
Límite superior $x$	1	2	3	5	37
Límite superior $T$	40	95	136	166	183

El modelo Pareto/NBD y el de Media Incondicional poseen un valor casi constante al observar los quintiles según  $x$  y  $T$ , sin embargo, en el caso del primer modelo se observa que el error aumenta cuando los clientes realizan una única transacción y ésta fue realizada hace mucho tiempo. Por otro lado en el caso del modelo Simple se tiene que el error es considerablemente menor en los casos en que el número de transacciones es mayor a 1 siendo la primera compra hace a lo menos 1 mes y medio. La tendencia del modelo Simple es igual al caso anterior, es decir, aumenta el error cuando aumenta el número de transacciones observado.

*Figura 52: MAPE del número de transacciones por quintil de  $x-T$*



*Fuente: Elaboración propia*

En cuanto a la tendencia de los modelos en el pronóstico del número de transacciones  $x$ , el modelo Pareto/NBD es el que posee los menores valores promedio del error porcentual, en magnitud, excepto en el caso del quintil 1 de  $x$ . Si se observa el quintil 1 de  $x$  (clientes que realizaron 1 única transacción durante el período de observación), se tiene que este modelo se equivoca más en la predicción del número de transacciones a medida que aumenta el período de observación del cliente, en otras palabras, mientras más lejos se haya realizado la transacción.

El modelo Simple sobreestima considerablemente el número de transacciones justo donde

la precisión de éste es peor. El modelo Media Condicional e Incondicional casi siempre sobreestiman  $x$ . Las figuras asociados al MPE por quintiles de  $x$  y  $T$  se encuentran en el Anexo 5.

### Quintil x-M

Ahora, se considerarán grupos de clientes según los quintiles del número de transacciones  $x$  y monto promedio por transacción del cliente  $M$ .

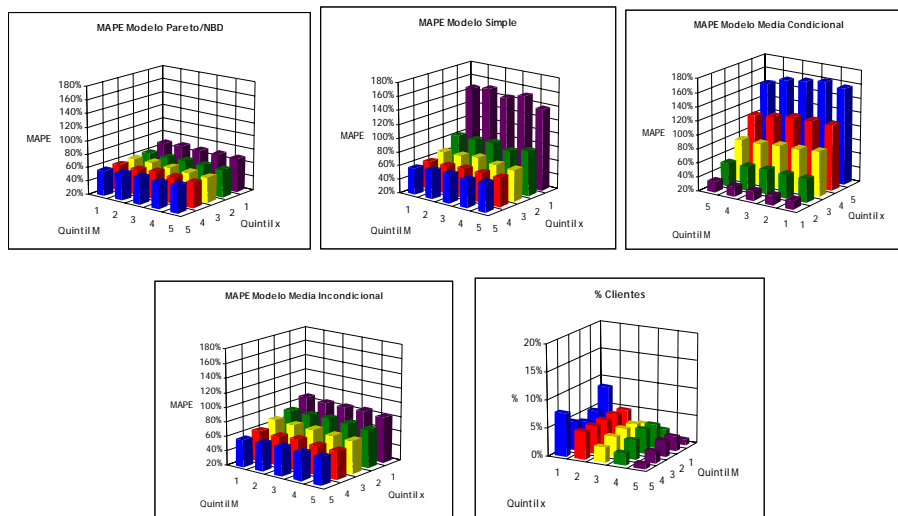
Quintil	1	2	3	4	5
Límite superior $x$	1	2	3	5	37
Límite superior $M$	\$5.896	\$9.145	\$12.980	\$18.909	\$137.802

La distribución de los clientes se diferencia más por el número de transacciones realizadas pero no por el monto promedio por transacción.

El modelo Pareto/NBD se equivoca de manera casi indistinguible entre los distintos clientes según quintiles del número de transacción y el monto promedio gastado. Por otro lado, el modelo Simple empeora notablemente la precisión en la predicción de la frecuencia de compra en los clientes del quintil 1 de  $x$ . El modelo Media Incondicional presenta una pequeña tendencia a predecir mejor cuando aumenta el número de transacciones observado, ocurre lo contrario en el caso del modelo Media Condicional.

El modelo Pareto/NBD no presenta tendencias marcadas al observar los clientes según  $x$  y  $M$ . Las figuras asociados al MPE por quintiles de  $x$  y  $M$  se encuentran en el Anexo 5.

*Figura 53: MAPE del número de transacciones por quintil de x-M*



Fuente: Elaboración propia

#### **4.7.1.3. Conclusiones de la estimación del número de transacciones**

El hecho de considerar los clientes que realizan compras durante el período de calibración y la aplicación de una constante en la estimación del número de transacciones a nivel individual de los modelos Pareto/NBD y Simple ayuda a mejorar la predicción del indicador mencionado. Se observa que el valor del MAPE se reduce en 67,3% en el caso del modelo Pareto/NBD y en un 117,2% en el caso del modelo Simple. Sin embargo, a pesar de la incorporación de la constante, sigue existiendo un error sistemático en el pronóstico del número de transacciones alrededor del 60,0%.

Los modelos captan el hecho de que los clientes buenos, en cuanto al número de visitas realizadas a alguna sucursal de la empresa, siguen siéndolos en el período siguiente.

En general, se observa que a medida que aumenta el período de compra del cliente, mejor estiman los modelos el número de transacciones de cada uno de los clientes.

El modelo Pareto/NBD realiza buenos pronósticos del número de transacciones a medida que aumenta el período de compra del cliente. Por otro lado, la predicción de  $x$  empeora en los casos de clientes que realizan un a única transacción y ésta fue realizada hace mucho tiempo.

El modelo más preciso en cuanto a la estimación del número de transacciones a nivel de clientes corresponde al Pareto/NBD con un valor de MAPE de 63,0% y una desviación estándar de 42,2%.

En el otro extremo se encuentra el modelo Simple, siendo el más inexacto, dado que pronostica  $x$  cometiendo un error de 94,6%, siendo el mayor problema de éste los clientes que realizan una única transacción siendo la primera compra hace a lo más 40 días (pertenecen al quintil 1 de  $T$ ), es decir, son clientes nuevos. En el Anexo 6 se muestra la estimación de número de transacciones realizando una modificación al modelo Simple que permita distinguir estos 2 tipos de clientes, los nuevos y el resto.

#### ***4.7.2. Estimación del monto por transacción***

La diferencia con respecto a la situación en la cual se estima el número de transacciones considerando sólo los atributos RFM es la estimación del modelo Media Incondicional, ya que en este caso, éste último diferencia los clientes según las variables género, edad y región.



*Tabla 48: Resumen del monto por transacción*

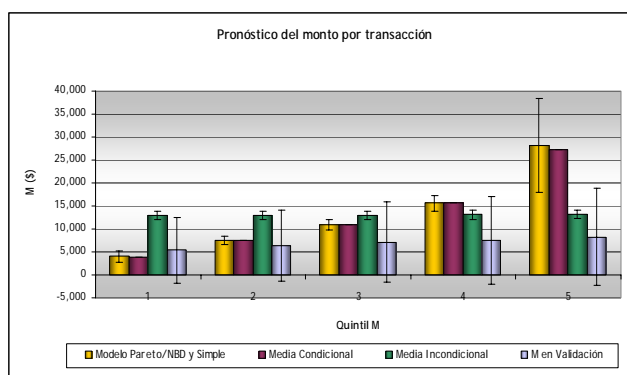
<i>Monto promedio</i>	<i>Real Calibración</i>	<i>Real Validación</i>	<i>Modelo Pareto/NBD y Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio	13.061	6.947	13.061	13.061	13.061
Desv. Est.	9.077	8.863	9.077	8.070	970

En promedio, los clientes que gastan más por transacción durante el período de calibración, también lo hacen en el de validación.

Los modelos Pareto/NBD y Media Condicional siguen esta tendencia, sin embargo no captan la magnitud en que lo hacen. El modelo Media Incondicional se mantiene casi constante.

La desviación estándar real de los pronósticos de  $M$  realizados por los modelos es muy baja en comparación a la real.

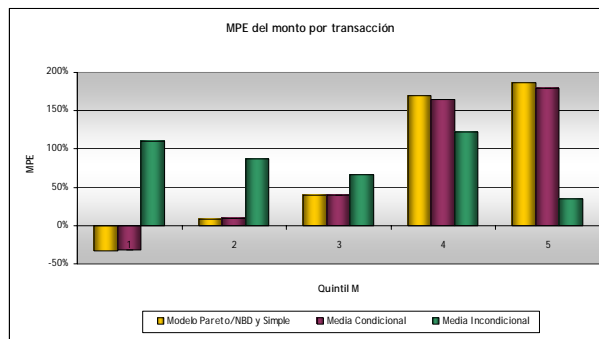
*Figura 54: Pronóstico del monto por transacción por quintil de  $M$*



*Fuente: Elaboración propia*

De la figura 55, se observa que los modelos Pareto/NBD, Simple y Media Condicional subestiman el monto promedio por transacción en los clientes que pertenecen al quintil 1 de  $M$ , es decir, quienes gastan menos de \$6.000. El modelo Media Incondicional siempre sobreestima el monto promedio por transacción, no observándose un claro patrón, con un valor de 86,2%.

*Figura 55: MPE por quintil de M*

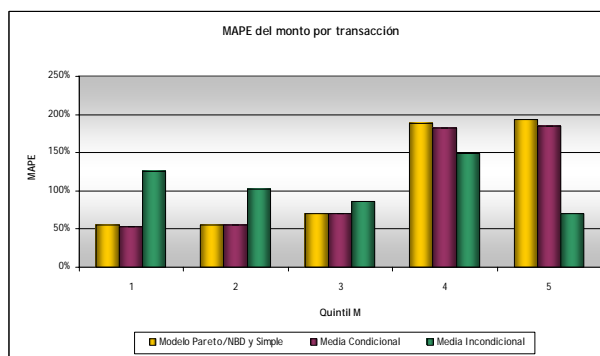


*Fuente: Elaboración propia*

Error Porcentual			
<i>Indicador</i>	<i>Modelo Pareto/NBD y Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio (MPE)	71,4 %	69,6 %	86,2 %
Desv. Est.	5.481,0 %	2.368,9 %	4.400,3 %
Mínimo	-100,0 %	-87,7 %	-88,6 %
Mediana	4,9 %	14,8 %	28,4 %
Máximo	1.664.000,0 %	304.653,8 %	1.335.900,0 %

Excepto por el modelo Media Incondicional, la precisión de los pronósticos aumenta a medida que disminuye M, es decir, los errores de pronóstico en caso de “buenos” clientes en cuanto al monto promedio gastado, es mayor.

*Figura 56: MAPE por quintil de M*



*Fuente: Elaboración propia*

ABS(Error Porcentual)			
<i>Indicador</i>	<i>Modelo Pareto/NBD y Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio (MAPE)	110,0 %	107,0 %	107,9 %

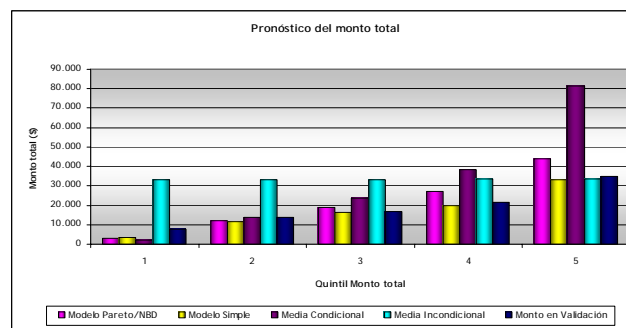
ABS(Error Porcentual)			
<i>Indicador</i>	<i>Modelo Pareto/NBD y Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Desv. Estándar	5.480,4 %	2.357,5 %	4.399,8 %
Mínimo	0,0 %	0,0 %	0,0 %
Mediana	49,2 %	56,1 %	47,0 %
Máximo	1.664.000,0 %	304.653,8 %	1.335.900,0 %

#### 4.7.3. Estimación del monto total

El monto total corresponde a la multiplicación del número de transacciones y el monto promedio por transacción, luego, el error cometido en la estimación de este indicador se debe a los errores generados en las estimaciones de los 2 componentes anteriores. Como se había mencionado, un cliente promedio que gasta \$32.454 durante el período de calibración reduce el monto total gastado a \$19.587 en el período de validación.

Los clientes del quintil 1 del monto total son los que gastan menos, mientras que los del quintil 5 son los que gastan más. Los clientes que generan mayor cantidad de ingresos para la empresa ahora, también lo serán en el futuro. En la realidad, se observa que el aumento observado de un quintil a otro no es de manera brusca, es decir, no se presentan grandes saltos, siendo el modelo Pareto/NBD y Simple quienes siguen este patrón. Por otro lado, el modelo Media Condicional tiende a diferenciar con mayor fuerza los distintos clientes en cuanto al monto total gastado.

*Figura 57: Pronóstico del monto total*

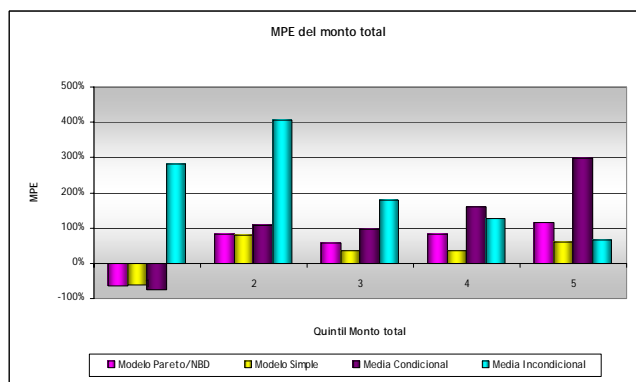


*Fuente: Elaboración propia*

<i>Monto total</i>	<i>Real Calibración</i>	<i>Real Validación</i>	<i>Modelo Pareto/NBD</i>	<i>Modelo Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio	32.454	19.587	20.518	16.561	35.078	33.437
Desv. Est.	31.300	32.041	19.975	67.804	37.993	2.913

Excepto por el modelo Media Incondicional, los modelos subestiman el monto total gastado por cada uno de los clientes pertenecientes al quintil 1, es decir, a aquellos clientes cuyo monto total gastado durante los 6 meses es menor a \$9.500, sobreestimando para el resto de los casos. El modelo Simple mantiene casi constante la magnitud de esta tendencia, excepto en el quintil 5, mientras que el modelo Pareto/NBD la va aumentando de manera moderada.

*Figura 58: MPE por quintil del Monto total*



*Fuente: Elaboración propia*

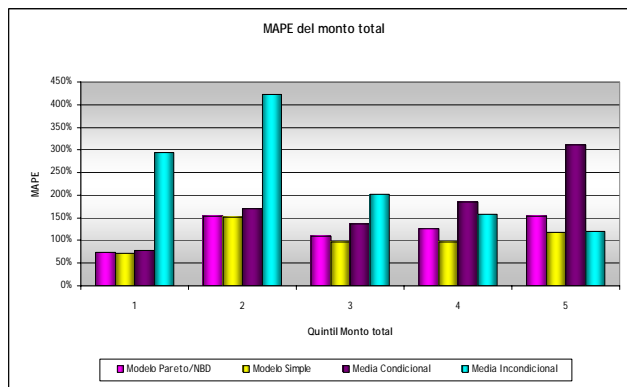
El modelo Simple, en promedio, sobreestima el monto total gastado por cada uno de los clientes en un 14,6%, seguido por el modelo Pareto/NBD con un 46,2%, siendo este último el que posee la menor desviación estándar.

*Tabla 49: Resumen del error porcentual en estimación del monto total*

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	46,2 %	14,6 %	133,5 %	200,9 %
Desv. Estándar	372,8 %	870,7 %	2.267,7 %	10.401,3 %
Mínimo	-100,0 %	-100,0 %	-94,4 %	-96,8 %
Mediana	-34,5 %	-72,0 %	15,8 %	39,9 %
Máximo	64.340,2 %	142.057,9 %	168.660,6 %	3.157.536,0 %

De la figura 59, se aprecia que el modelo Pareto/NBD es el que pronostica mejor el monto total a nivel individual con un error de 122,9% y con la menor desviación estándar, seguido por el modelo Simple con un 133,3%. La diferencia entre estos 2 modelos se debe a la brecha en la estimación del número de transacciones, ya que el pronóstico del monto promedio por transacción se realiza de la misma manera.

*Figura 59: MAPE por quintil del Monto total*



*Fuente: Elaboración propia*

ABS(Error Porcentual)				
<i>Indicador</i>	<i>Modelo Pareto/NBD</i>	<i>Modelo Simple</i>	<i>Media Condicional</i>	<i>Media Incondicional</i>
Promedio (MAPE)	122,9 %	133,3 %	180,0 %	229,8 %
Desv. Estándar	355,0 %	860,6 %	2.255,8 %	10.400,7 %
Mínimo	0,0 %	0,0 %	0,0 %	0,0 %
Mediana	68,6 %	83,0 %	68,9 %	64,0 %
Máximo	64.340,2 %	142.057,9 %	268.660,6 %	3.157.536,0 %

#### **4.7.3.1. Conclusiones de la estimación del monto total**

Los clientes buenos en cuanto al monto total gastado siguen siéndolo durante el período siguiente.

Los modelos tienden a sobreestimar el monto total gastado en la compañía durante el segundo semestre del 2004, excepto en los casos de los malos clientes en los cuales subestima. Se espera que los clientes que gastan poco gasten aún menos durante el período siguiente, mientras que en los buenos clientes se espera que aumenten el monto total gastando en la empresa.

El modelo Pareto/NBD es el mejor modelo en el pronóstico del monto total a nivel de cliente con el menor error (122,9%) y desviación estándar (555,0%), seguido por el modelo Simple.

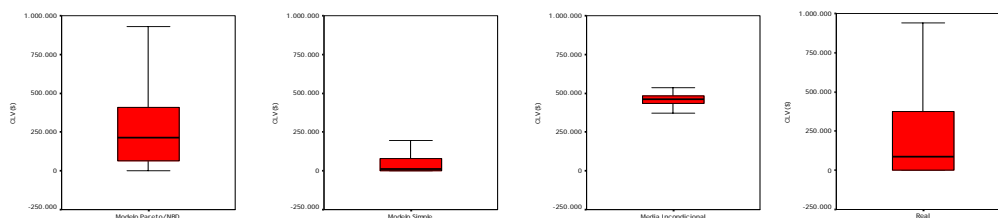
#### 4.7.4. Cálculo del CLV

La estimación del *CLV* se realiza a partir de la estimación del número de transacciones y del monto promedio por transacción. El indicador que se modela corresponde al número de transacciones promedio de cada uno de los clientes.

El modelo Simple es el que estima un valor promedio del CLV menor, siendo el de Media Incondicional el que posee el mayor valor de éste. El modelo Pareto/NBD es el que estima con un menor error el número de transacciones promedio a nivel individual, luego, se puede decir que éste es el modelo que estima el valor del CLV a nivel individual con mayor precisión, además, de ser el que posee un mayor rango de posibles valores. El modelo Media Incondicional es el que posee la menor desviación estándar dada la manera de estimación de éste.

Un cliente promedio vale \$281.076 según el modelo Pareto/NBD, \$119.100 según el modelo Simple y \$457.646 según el de Media Incondicional.

*Figura 60: Boxplot del CLV*



*Fuente: Elaboración propia*

CLV				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Incondicional	Real
Promedio	\$281.075	\$118.865	\$455.833	267.372
Dev. Estándar	\$272.666	\$877.186	\$39.314	437.378
Mínimo	\$0	\$0	\$371.430	0
Mediana	\$215.084	\$12.143	\$460.933	85.957
Máximo	\$4.000.468	\$66.995.210	\$534.390	12.834.078

Conocidos los valores del CLV para cada uno de los clientes, se puede calcular el valor gastado por el cliente anualmente. Como se muestra en la tabla 50, el modelo Pareto/NBD y el caso en que se considera que el cliente se va a comportar igual que lo observado poseen valores similares del CLV anual. Mientras que los modelos Simple y Media

Incondicional poseen valores extremos.

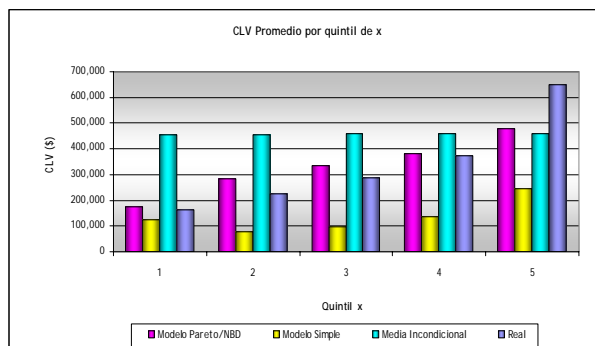
*Tabla 50: Resumen de CLV anualizado*

Valor Anual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Incondicional	Real
Promedio	7.002	2.972	11.396	6.684
Desv. Estándar	6.817	21.930	983	10.934
Mínimo	0	0	9.286	0
Mediana	5.377	304	11.523	2.149
Máximo	100.010	1.674.900	13.360	320.850

Considerando las variables tanto socio-demográficas como transaccionales, se tiene que:

- Los clientes que realizan mayor cantidad de transacciones valen más según los modelos Pareto/NBD y Simple. Cabe destacar, que el modelo Simple valora más los clientes del quintil 1 de  $x$  en comparación a los del quintil 2, debido a existencia de los clientes que realizan la compra al final del período de observación.

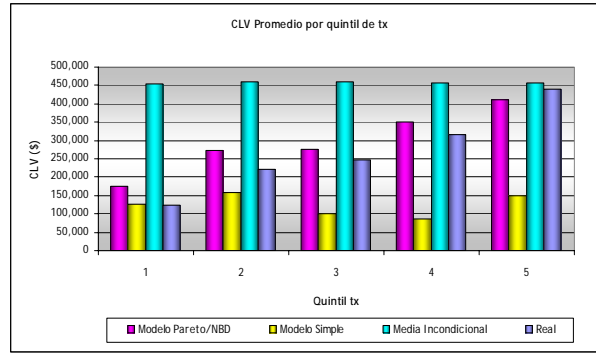
*Figura 61: CLV promedio por quintil de  $x$*



*Fuente: Elaboración propia*

- El modelo Pareto/NBD valora más a los clientes que se mantienen por más tiempo comprando. El modelo Simple otorga mayor valor a los clientes que compran por poco tiempo, dado que dentro de este grupo se encuentran los clientes que realizan 1 transacción (figura 62).

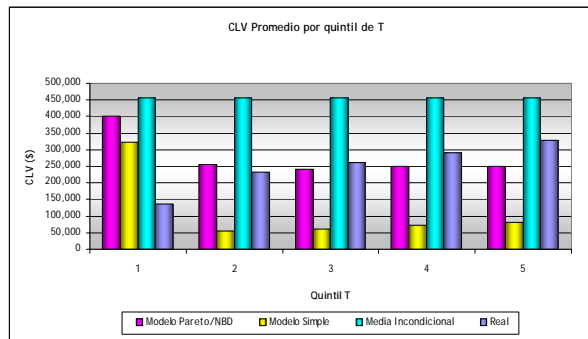
*Figura 62: CLV promedio por quintil de tx*



*Fuente: Elaboración propia*

- Los modelos valoran más a los clientes del quintil 1 de  $T$ , es decir, a aquellos clientes que realizan la primera compra hace a lo más de 40 días, siendo el impacto mayor en el caso del modelo Pareto/NBD. Para los otros quintiles se mantiene constante el valor del  $CLV$ .

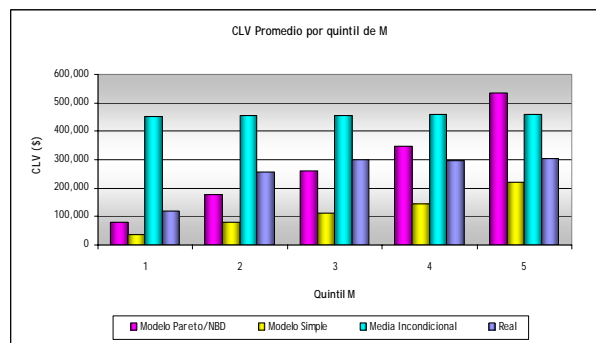
*Figura 63: CLV promedio por quintil de T*



*Fuente: Elaboración propia*

- Mientras el cliente gasta más por cada transacción realizada, se le asigna un mayor valor del  $CLV$ .

*Figura 64: CLV promedio por quintil de M*

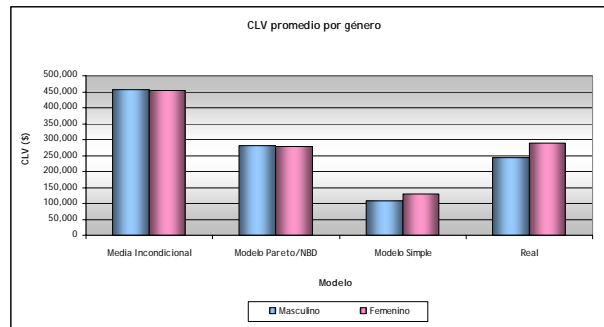


*Fuente: Elaboración propia*



- Según el modelo Pareto/NBD, los hombres valen más que las mujeres. Para los otros, se tiene que las mujeres valen más para la empresa.

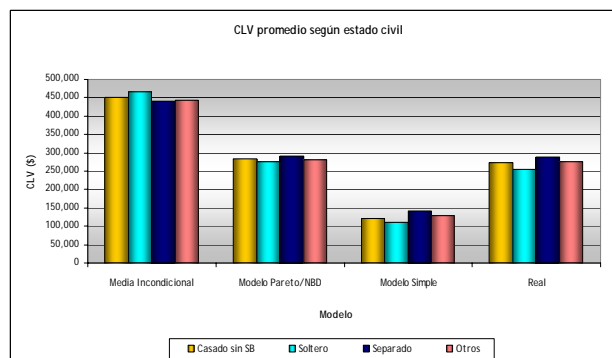
*Figura 65: CLV promedio por género*



*Fuente: Elaboración propia*

- En orden descendente según el valor promedio del CLV, se tiene que los clientes separados valen más, seguidos por los casados sin SB y, finalmente, los que valen menos corresponden a los solteros.

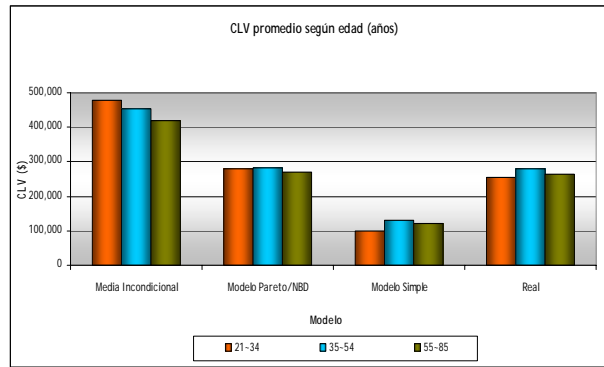
*Figura 66: CLV promedio por estado civil*



*Fuente: Elaboración propia*

- Los clientes jóvenes con edad entre los 21 y 34 años son los que valen menos para la empresa, seguidos por los clientes mayores de 54 años. Se observa que los clientes con edad entre los 35 y 54 años son los que valen más para la empresa en estudio, siendo este grupo un poco más del 47% del total de clientes (figura 67).

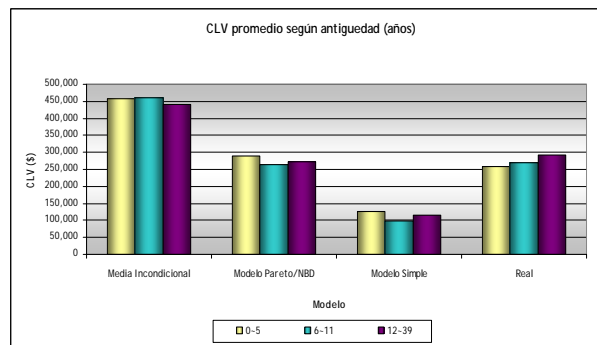
*Figura 67: CLV promedio por quintil de edad*



*Fuente: Elaboración propia*

- En cuanto a la antigüedad de los clientes en la empresa, se tiene que los clientes nuevos valen más para ella, seguido por los clientes con mayor antigüedad (mayor o igual a 12 años).

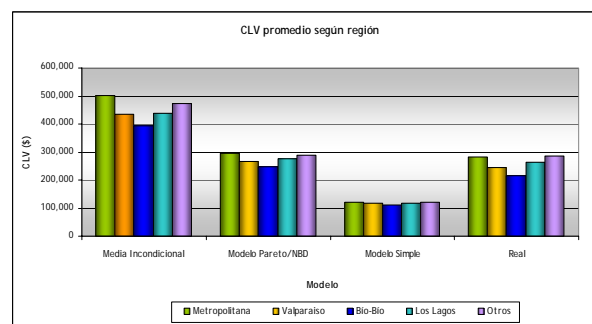
*Figura 68: CLV promedio por quintil de antigüedad*



*Fuente: Elaboración propia*

- Los clientes pertenecientes a la región de Bío-Bío son los que poseen el menor valor promedio del CLV, siendo los clientes de la región Metropolitana los que valen más.

*Figura 69: CLV promedio por quintil de región*



*Fuente: Elaboración propia*

#### 4.7.4.1. Error asociado al cálculo del CLV

A continuación, se estima el error asociado al valor calculado del indicador *CLV*. El error asociado al número de transacciones y monto promedio por transacción se muestra a continuación.

*Tabla 51: Resumen del MAPE*

Modelo	MAPE de x	MAPE de M
Pareto/NBD	63,0 %	107,9 %
Simple	94,6 %	
Media Incondicional	73,2 %	

Luego, el error asociado al *CLV* por modelo es tal que:

*Tabla 52: Estimación del MAPE del CLV*

Modelo	MAPE de <i>CLV</i>
Pareto/NBD	927,9 %
Simple	1.393,3 %
Media Incondicional	1.078,1 %

Dada las condiciones anteriores, se puede decir que el modelo Pareto/NBD es el más confiable en comparación a los otros dos.

#### 4.7.4.2. Ranking de clientes

El 10% de los clientes más valiosos corresponden a 16.190 casos. Considerando sólo el modelo Pareto/NBD, se tiene que el 76,4% del 10% de clientes más valiosos, también se encuentra dentro del 10% de clientes con mayor monto por transacción.

A continuación se analizan el 10% de los mejores clientes en cuanto al *CLV* según el modelo Pareto/NBD, Simple y, además, considerando el caso en que el cliente se comporta de la misma manera que en el segundo semestre del 2004, de ahora en adelante será considerado caso "Real".

Según el modelo Pareto/NBD, el 10% de los clientes con mayor *CLV* tienen un valor promedio de éste igual a \$879.054, \$890.826 en el caso del modelo Simple y \$1.286.513 en el "Real". Como se mencionó en la aplicación de la metodología anterior, el último valor se encuentra por sobre de los 2 anteriores debido a que no considera la fuga o disminución del comportamiento de compra del cliente.

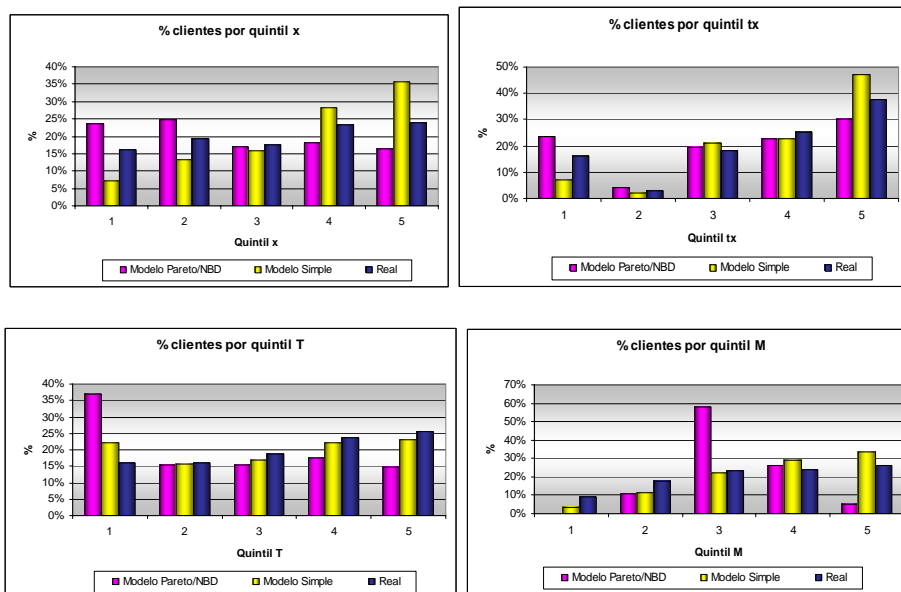
Tabla 53: Resumen de indicadores

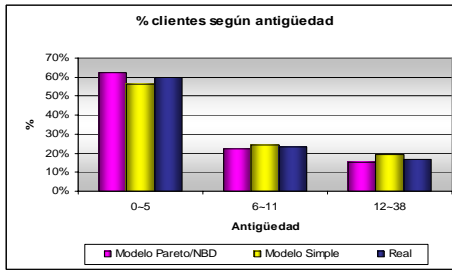
Modelo	Indicador	Período de Calibración					Período de Validación			CLV
		x	tx	T	M	Monto	x	M	Monto	
Pareto/NBD	Promedio	3,4	74	87	27.697	81.972	2,2	10.132	34.542	879.054
	Desv. Est.	2,9	63	64	10.353	50.204	3,0	11.077	50.363	262.605
Simple	Promedio	5,1	103	108	17.242	79.910	3,0	9.693	40.099	890.826
	Desv. Est.	3,3	61	62	8.991	50.563	3,7	9.894	53.923	2.647.865
Real	Promedio	4,1	86	117	15.477	58.155	5,9	19.095	94.247	1.286.513
	Desv. Est.	3,2	62	57	10.036	48.100	3,4	10.211	45.286	618.177

Cerca del 40% de estos clientes, según el modelo Pareto/NBD, corresponden a clientes nuevos, es decir, cuya primera compra fue realizada a lo más hace 40 días antes del final del período de observación. La mayoría corresponden a clientes con antigüedad menor o igual a 5 años.

Casi el 60% del 10% de los clientes más valiosos pertenecen al quintil 3 según el monto promedio por transacción. Por otro lado, los clientes valiosos según el modelo Simple y el comportamiento pasado (Real) se distribuyen de manera similar, entregando mayor prioridad a aquellos clientes que realizan un mayor número de transacciones manteniéndose comprando períodos más largos.

Figura 70: Distribución de clientes por variables transaccionales



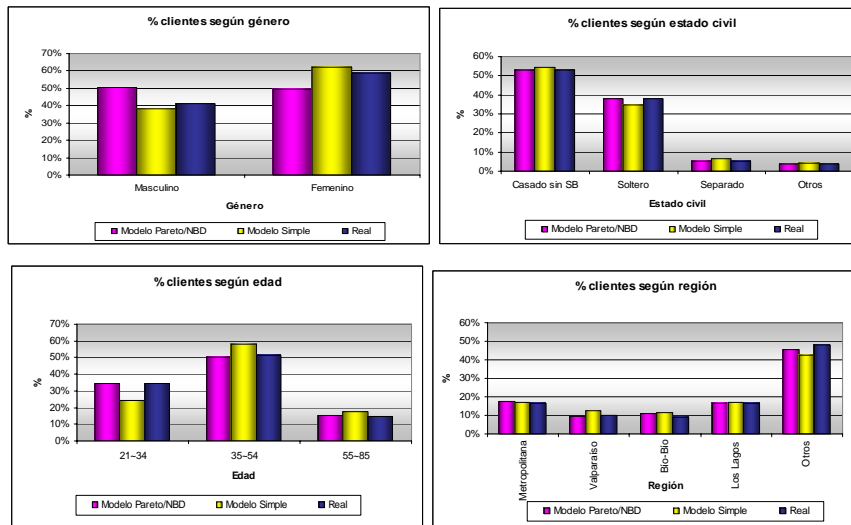


Fuente: Elaboración propia

Son mujeres que hombres las que pertenecen en este pequeño grupo de clientes. Predominan los clientes con estado civil casado sin SB y soltero.

Además, se tiene que un bajo porcentaje corresponden a los clientes viejos (edad entre los 55 y 85 años), mientras que la mayoría corresponde al grupo que tiene edad entre los 35 y 54 años.

Figura 71: Distribución de clientes por variables socio-demográficas



Fuente: Elaboración propia

La asignación de los clientes más valiosos según el modelo Pareto/NBD y Simple coinciden en el 43,5% de los casos.

Tabla 54: Asignación de los clientes según CLV

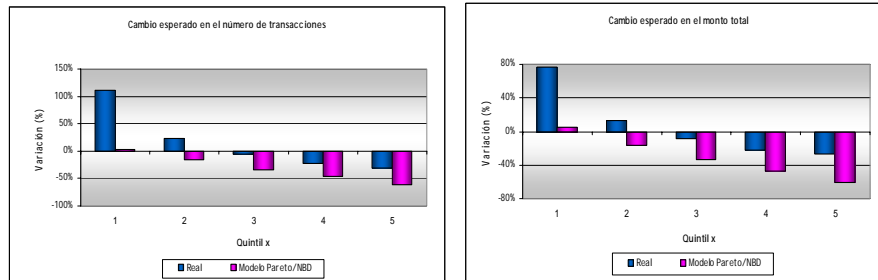
Modelo	Pareto/NBD	Simple	Real
Pareto/NBD	100,0%	43,5%	22,5%
Simple	43,5%	100,0%	26,1%
Real	22,5%	26,1%	100,0%

#### 4.7.5. Cambios esperados del comportamiento transaccional

A continuación, se muestran los cambios reales y esperados en el número de transacciones realizadas y el monto total gastado. Se considera sólo el modelo Pareto/NBD.

- A partir del quintil 3, se observa que hay una tendencia real a la baja en cuanto al número de transacciones realizadas y al monto total gastado, los modelos capturan esta tendencia en sus predicciones, sin embargo, esperan que éste sea de manera más brusca.

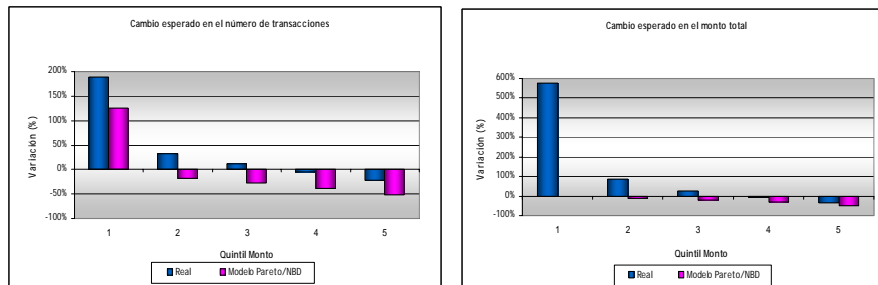
*Figura 72: Cambios esperados por quintil de x*



*Fuente: Elaboración propia*

- Los clientes que gastaban menos de \$9.500, aumentan notablemente las visitas alguna sucursal de la empresa junto al monto total gastado. Por otro lado, los clientes “buenos” presentan una pequeña tendencia a la baja en estos 2 indicadores. El modelo Pareto/NBD sigue relativamente bien la tendencia real del número de transacciones. En cuanto al monto total, no se esperan grandes cambios con respecto a la situación actual, pero, en la realidad, se observa que los clientes del quintil 1 aumentan de manera considerable el monto total gastado.

*Figura 73: Cambio esperado por quintil de Monto total*

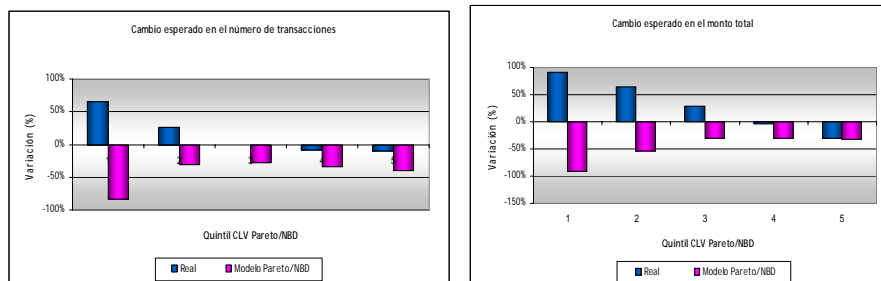


*Fuente: Elaboración propia*

- Los clientes del quintil 3 no aumentan el número de transacciones realizadas, pero, sí lo hace en cuanto al monto total, es decir, los clientes pertenecientes al quintil

3 aumentan su monto promedio. Sin embargo, independiente del quintil al cual pertenezcan los clientes, el modelo Pareto/NBD espera una caída tanto en la frecuencia de compra como en el monto total.

*Figura 74: Cambio esperado por quintil de CLV*



Fuente: Elaboración propia

#### 4.7.6. Conclusiones

##### 4.7.6.1. Conclusiones de la estimación del número de transacciones

Los modelos siguen la tendencia real de que los buenos clientes hoy, también lo serán en el futuro, en especial en el caso de los buenos y normales clientes hoy, como se muestra en la tabla 55.

*Tabla 55: Probabilidades de cambio de estado*

	Buena	Normal	Mala
Buena	66,6 %	33,0 %	0,4 %
Normal	37,6 %	53,8 %	8,6 %
Mala	13,5 %	48,1 %	38,4 %

La incorporación de clientes inactivos y la aplicación de una constante a las estimaciones del modelo Pareto/NBD y Simple permiten reducir el error en el pronóstico del número de transacciones con respecto a la situación inicial que considera sólo a los clientes activos durante el período de calibración. Finalmente, se logra reducir el valor del MAPE en 67,3% y 117,2% respectivamente. El modelo Pareto/NBD es el que pronostica el número de transacciones a nivel de clientes con mayor precisión, cometiendo un error de 63,0%.

##### 4.7.6.2. Conclusiones del cálculo del CLV

Los modelos Pareto/NBD y Simple aumentan considerablemente el valor del CLV en el caso de los clientes nuevos, es decir, cuando la primera compra fue realizada hace a lo más 40 días.

El valor del cliente es mayor a medida que aumenta el número de transacciones, gasto por transacción y mientras aumenta el período de compra de éste. También, se observa que aumenta el valor de la relación cliente-empresa en el caso del quintil 1 de  $T$ , es decir, en el caso de clientes nuevos.

Los modelos otorgan un mayor valor a las mujeres que a los hombres, observando lo mismo al considerar el 10% de clientes más valiosos. Considerando el estado civil del cliente, se tiene que el valor del  $CLV$  es mayor cuando se trata de clientes separados, seguidos por los casados sin SB.

Los clientes jóvenes son a los que se les otorga un menor valor de la relación, seguido de los viejos. Luego, los clientes más valiosos son aquellos que tienen una edad entre los 5 y 54 años.

Al observar la antigüedad de los clientes en la empresa, los clientes nuevos, es decir, con una antigüedad menor o igual a 5 años, son los más valiosos, seguidos de los más antiguos.

El modelo Pareto/NBD es el mejor modelo, a pesar de tener asociado valores altos de error, seguido por el modelo Simple debido a que es el modelo que pronostica el número de transacciones con el menor error.

#### **4.8. Análisis comparativo entre las metodologías utilizadas**

Para el cálculo del *customer lifetime value* es necesario conocer el número de transacciones y monto por transacción futura. La variable modelada es el primero de ello, luego, si la confiabilidad del cálculo del  $CLV$  depende de la precisión en la estimación del número de transacciones ( $x$ ). Del análisis realizado, se obtiene que la segunda metodología aplicada para el cálculo del  $CLV$ , es decir, la que considera tanto las variables socio-demográficas como transaccionales, permite reducir los errores asociados a la estimación de  $x$ .

##### **4.8.1. Estimación del número de transacciones**

En el caso del modelo Pareto/NBD, Se logra reducir el error de pronóstico del número de transacciones en un 16,6% en el caso inicial que considera sólo los clientes activos y en un 0,9% en la situación final que incorpora los clientes inactivos y aplica la constante. En el caso del modelo Simple, se reduce en 0% y 0,4% respectivamente. Lo anterior se debe a que en el caso inicial, no varía la estimación al aplicar las distintas metodologías.



#### *4.8.2. Cálculo del Customer Lifetime Value (CLV)*

Al aplicar las metodologías, se observa que los modelos Pareto/NBD y Simple otorgan un mayor valor del *CLV* a los clientes nuevos, cuyo período de observación es de a lo más 40 días, siendo en este grupo en donde empeora la predicción del número de transacciones.

Los clientes con estado civil separado son más valiosos, pero son pocos clientes, seguidos por los casados sin SB.

Según el modelo Pareto/NBD, los hombres poseen un mayor valor del *CLV* que los hombres, mientras que en el caso del modelo Simple, se ve que pasa de valorar más los hombres a las mujeres, dada la incorporación de información no transaccional.

Al incorporar las variables socio-demográficas se observa que los clientes más valiosos con respecto a la edad, pasa de los jóvenes a los que son un poco mayores, es decir, a aquellos que tienen entre 35 y 54 años.

## 5. CONCLUSIONES FINALES

### 5.1. Conclusiones de los resultados obtenidos

De los resultados obtenidos en este trabajo se puede decir que:

- La incorporación de las variables socio-demográficas mejora la predicción de la variable modelada necesaria para el cálculo del *CLV*, el número de transacciones, reduciendo el valor del MAPE en un 0,9% con respecto a la metodología que usa sólo las variables RFM.
- Los buenos clientes en cuanto al número de transacciones siguen siéndolos durante el período siguiente en estudio.

*Tabla 56: Probabilidades de transición de un estado a otro*

	<i>Bueno</i>	<i>Normal</i>	<i>Malo</i>
<i>Bueno</i>	66,6 %	33,0 %	0,4 %
<i>Normal</i>	37,6 %	53,8 %	8,6 %
<i>Malo</i>	13,5 %	48,1 %	38,4 %

- El modelo Pareto/NBD es el que pronostica el número de transacciones con mayor precisión, cometiendo un error de 63,0% al aplicar la metodología que considera las variables socio-demográficas y transaccionales. Como se observó en el desarrollo del trabajo, se tiene que este modelo es el que posee el menor error aproximado del cálculo del *CLV* igual a 927,9%. El alto valor de éste es debido a los altos errores en la predicción tanto del número de transacciones como del monto promedio por transacción.
- El modelo Pareto/NBD predice el número de transacciones con mayor MAPE, alrededor del 90%, cuando los clientes realizan una transacción y va aumentando a medida que aumenta el período de observación de este tipo de cliente, es decir, en los casos de clientes que realizan compras esporádicas. Además, requiere que el período de compra del cliente sea mayor para estimar  $x$  con mayor precisión, dada que la frecuencia de compra de los clientes es pequeña (promedio de 3,5 transacciones aproximadamente). Para el caso en estudio, se utilizaron ventanas de tiempo de 6 meses, el cual podría aumentarse a 1 año como mínimo.
- El modelo Simple se equivoca más en el pronóstico del número de transacciones cuando considera los clientes que realizan una única transacción y en aquellos casos en que corresponden a clientes nuevos, en este caso se define como aquellos clientes cuya primera compra fue realizada hace a lo más 40 días.

- En cuanto al cálculo del *CLV*, se tiene que la relación con los clientes nuevos es más valorada que el resto, siendo en este grupo en donde los modelos se equivocan más, sobreestimado de manera considerable el número de transacciones futuro.
- Se observa que, según el modelo Pareto/NBD, el valor del *CLV* promedio de los hombres es mayor cercano a los \$290.000. Los clientes con edad entre 34 y 54 años y/o nuevos son los que poseen un mayor valor para la empresa.

## 5.2. Trabajos futuros

- Se considera un estado estático de las variables categóricas utilizadas como el estado civil, región, edad y antigüedad de los clientes, es decir, no se considera la evolución de éste en el tiempo, por lo que sería más real considerar el paso del tiempo. El comportamiento de un cliente joven no es igual al de un adulto mayor, luego sus distribuciones de compra y de fuga también son distintas, cuyos parámetros se rescatan con la información disponible.
- Hasta ahora, se ha considerado que los clientes gastan, en promedio, igual que el promedio histórico, lo cual hace que se subestime el valor gastado por transacción de los clientes que gastan por sobre el promedio y sobreestime el de los clientes que gastan menos que el promedio de todos los clientes. Luego, se podría modelar el monto promedio por transacción de cada uno de los clientes al igual que en el caso del número de transacciones, y combinarlos para la estimación del *CLV*.
- Incorporar una tendencia de la frecuencia de compra y del monto por transacción de los clientes para poder caracterizar mejor el comportamiento de compra de éstos. Es decir, agregar información del aceleramiento (aumento) o desaceleramiento (disminución) de las variables que caracterizan el comportamiento de compra de los clientes.
- Realizar una segmentación de los clientes considerando, en conjunto, todas las variables (transaccionales y socio-demográficas) antes de aplicar la metodología, ya que así a priori se determinan grupos similares en cuanto al comportamiento de compra.
- Aplicar la metodología en otros casos para poder determinar en qué casos se predice mejor el *CLV* y así determinar los límites de la metodología.
- Incorporar el comportamiento de compra del cliente en la competencia. Por ejemplo, agregar el hecho de que el cliente también lo es de la competencia, luego se debiera considerar qué tan buen o mal cliente lo es de la competencia. Por ejemplo, en caso bancario, se posee información de la situación financiera de cada uno de los clientes en la competencia, ya sea en instituciones bancarias, casas comerciales entre otros.

## 6. REFERENCIAS BIBLIOGRÁFICAS

- [1] Fader, P., Hardie, B. 2000. Applied Probability Models in Marketing Research: Introduction. p 1-7.
- [2] Fader, P., Hardie, B., Lee, K. 2004. RFM and CLV: Using Iso-value Curves for Customer Base Analysis. p 1-38.
- [3] Herrera, C. Gestión de las relaciones con clientes (CRM).
- [4] Hurtado, C. Abril 2005. Aprendizaje en motores de búsqueda en la web. Congreso de Ciencia y Tecnología Computacional. Tulúa.
- [5] Lewis, M. 2005. Incorporating Strategic Consumer Behavior into Customer Valuation. Journal of Marketing, Vol. 69 Issue 4. p 230-138.
- [6] Marín, P. 2005. Estimación de lifetime value basado en comportamiento transaccional. Memoria Ingeniería Civil Industrial. Universidad de Chile.
- [7] Administración de las relaciones con el cliente y estrategia empresarial centrada en el cliente.
- [8] Reicheld, F y Sasser, W. 1990. Zero defections: Quality comes to service. Harvard Business Review, Septiembre - Octubre 1990.
- [9] Reinartz, W., Kumar, V. 2000. On the Profitability of Long Lifetime Customers: An Empirical Investigation and Implications for Marketing. p 1-38.
- [10] Reinartz, W., Kumar, V. 2001. The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration. Fontainebleau, France.
- [11] Valenzuela, L. Filosofía empresarial y del marketing orientada en el valor del cliente. Universidad del Bío-Bío. Concepción, Chile.
- [12] Webster, A. 2000. Estadística aplicada a los negocios y la economía. Cap. 10.
- [13] Zapata, W. 2006. Un Nuevo enfoque de la fidelización del cliente: CRM.

## 7. ANEXOS

### Anexo1: Distribución binomial negativa (NBD)

Si consideramos un experimento de Bernoulli que se efectúa repetidas veces, sin que el resultado de una repetición se vea afectado por algún otro resultado. En lugar de fijar el número de repeticiones y contar el número de éxitos obtenidos estamos interesados en obtener un determinado número de éxitos y contaremos el número de repeticiones precisas para lograrlo. Una distribución de este tipo da origen a la distribución binomial negativa.

La variable aleatoria  $X$  que representa el número de repeticiones de un experimento de Bernoulli, con probabilidad de éxito  $p$ , necesarias para alcanzar el  $k$ -ésimo éxito, exactamente en la repetición  $k+x$ , se dice que siguen una distribución binomial negativa de parámetros  $k$  y  $p$  y se representa por  $X|BN(k, p)$ . La distribución de probabilidad es:

$$P(X = x) = \binom{k+x-1}{k-1} \cdot p^k \cdot (1-p)^x, x = 0, 1, 2, \dots$$

La función de distribución acumulada de la binomial negativa es:

$$F_{BN}(k, p) = \sum_{i=0}^{[x]} \binom{k+i-1}{i-1} \cdot p^k \cdot (1-p)^i$$

Donde la media y la varianza son tal que:  $Media = \frac{k \cdot (1-p)}{p}$  y  $\sigma^2 = \frac{k \cdot (1-p)}{p^2}$

## Anexo 2: Función de verosimilitud

La función de verosimilitud es:

$$L(r, \alpha, s, \beta / X = x, t_x, T) = \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)} \left\{ \frac{1}{(\alpha+T)^{r+s} \cdot (\beta+T)^s} + \left( \frac{s}{r+s+x} \right) \cdot A_0 \right\}$$

donde:

Si  $\alpha \geq \beta$

$$A_0 = \frac{{}_2F_1\left(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+t_x}\right)}{(\alpha+t_x)^{r+s+x}} - \frac{{}_2F_1\left(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+T}\right)}{(\alpha+T)^{r+s+x}}$$

Si  $\alpha \leq \beta$

$$A_0 = \frac{{}_2F_1\left(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+t_x}\right)}{(\beta+t_x)^{r+s+x}} - \frac{{}_2F_1\left(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+T}\right)}{(\beta+T)^{r+s+x}}$$

Además, se tiene que:

$$\Gamma(.) \text{ es la función gamma, tal que: } \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

$${}_2F_1(.) \text{ es la función hipergeométrica gaussiana, tal que: } {}_2F_1(a, b; c; z) = \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{z^j}{j!}$$

con  $c \neq 0, -1, -2, \dots$  y  $(a)_j = \frac{\Gamma(a+j)}{\Gamma(a)}$ . Esta serie converge si:  $|z| < 1$  o  $|z| = 1$  y  $c-a-b > 0$

### Anexo 3: Resultados del test Anova

A continuación se muestran los resultados de la aplicación del test Anova considerando como:

- Variables dependientes: número de transacciones, monto por transacción y monto total.
- Variables independientes: estado civil, edad, región, antigüedad, día de pago y estado de la tarjeta.

*Tabla 57: Test de Anova para la variable independiente estado civil*

#### TEST ANOVA: VARIABLE INDEPENDIENTE ESTADO CIVIL

		Sum of Squares	df	Mean Square	F	Sig.
N° de transacciones	Between Groups	156,710	3	52,237	13,175	,000
	Within Groups	641853,2	161882	3,965		
	Total	642009,9	161885			
Monto por transacción	Between Groups	1,52E+08	3	50662680,67	,615	,605
	Within Groups	1,33E+13	161882	82392379,19		
	Total	1,33E+13	161885			
Monto total	Between Groups	3,03E+10	3	1,009E+10	10,298	,000
	Within Groups	1,59E+14	161882	979517616,3		
	Total	1,59E+14	161885			

*Tabla 58: Test de Anova para la variable independiente edad*

#### TEST ANOVA: VARIABLE INDEPENDIENTE EDAD

		Sum of Squares	df	Mean Square	F	Sig.
N° de transacciones	Between Groups	1313,802	2	656,901	165,977	,000
	Within Groups	640696,1	161883	3,958		
	Total	642009,9	161885			
Monto por transacción	Between Groups	5,68E+09	2	2840264129	34,487	,000
	Within Groups	1,33E+13	161883	82357718,77		
	Total	1,33E+13	161885			
Monto total	Between Groups	1,94E+11	2	9,697E+10	99,101	,000
	Within Groups	1,58E+14	161883	978500459,8		
	Total	1,59E+14	161885			

*Tabla 59: Test de Anova para la variable independiente región*

**TEST ANOVA: VARIABLE INDEPENDIENTE REGIÓN**

		Sum of Squares	df	Mean Square	F	Sig.
N° de transacciones	Between Groups	1035,021	4	258,755	65,350	,000
	Within Groups	640974,9	161881	3,960		
	Total	642009,9	161885			
Monto por transacción	Between Groups	4,89E+10	4	1,222E+10	148,918	,000
	Within Groups	1,33E+13	161881	82091755,14		
	Total	1,33E+13	161885			
Monto total	Between Groups	8,17E+11	4	2,043E+11	209,603	,000
	Within Groups	1,58E+14	161881	974662631,5		
	Total	1,59E+14	161885			

*Tabla 60: Test de Anova para la variable independiente antigüedad*

**TEST ANOVA: VARIABLE INDEPENDIENTE ANTIGÜEDAD**

		Sum of Squares	df	Mean Square	F	Sig.
N° de transacciones	Between Groups	214,670	2	107,335	27,074	,000
	Within Groups	641795,2	161883	3,965		
	Total	642009,9	161885			
Monto por transacción	Between Groups	4,62E+10	2	2,308E+10	281,059	,000
	Within Groups	1,33E+13	161883	82107700,36		
	Total	1,33E+13	161885			
Monto total	Between Groups	1,81E+10	2	9073722571	9,263	,000
	Within Groups	1,59E+14	161883	979586392,6		
	Total	1,59E+14	161885			

*Tabla 61: Test de Anova para la variable independiente día de pago*

**TEST ANOVA: VARIABLE INDEPENDIENTE DÍA DE PAGO**

		Sum of Squares	df	Mean Square	F	Sig.
N° transacciones	Between Groups	791,199	2	395,599	99,874	,100
	Within Groups	641218,7	161883	3,961		
	Total	642009,9	161885			
Monto por transacción	Between Groups	4,71E+09	2	2356623190	28,612	,431
	Within Groups	1,33E+13	161883	82363693,96		
	Total	1,33E+13	161885			
Monto total	Between Groups	8,18E+10	2	4,089E+10	41,755	,310
	Within Groups	1,59E+14	161883	979193361,3		
	Total	1,59E+14	161885			



*Tabla 62: Test de Anova para la variable independiente estado de la tarjeta*

**TEST ANOVA: VARIABLE INDEPENDIENTE ESTADO DE LA TARJETA**

		Sum of Squares	df	Mean Square	F	Sig.
N° transacciones	Between Groups	3460,248	2	1730,124	438,615	,083
	Within Groups	638549,6	161883	3,945		
	Total	642009,9	161885			
Monto por transacción	Between Groups	1,21E+11	2	6,026E+10	738,094	,156
	Within Groups	1,32E+13	161883	81648269,67		
	Total	1,33E+13	161885			
Monto total	Between Groups	1,97E+12	2	9,872E+11	1020,383	,120
	Within Groups	1,57E+14	161883	967501753,1		
	Total	1,59E+14	161885			

**Anexo 4: Pronóstico del número de transacciones aplicando la metodología de estimación que considera tanto las variables socio-demográficas como las transaccionales**

A continuación se muestran los errores en la estimación del número de transacciones considerando los atributos RFM y cada uno de las variables que se muestran. Cabe destacar que la metodología toma en cuenta la existencia de clientes inactivos y la aplicación de la constante.

**Variable Género**

A continuación se muestran los errores asociados a la estimación del número de transacciones considerando los atributos RFM y la variable género de cada uno de los clientes.

*Tabla 63: Resumen del error porcentual en el número de transacciones*

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	-10,6 %	-24,1 %	34,7 %	42,0 %
Desv. Est.	74,6 %	296,5 %	85,1 %	85,1 %
Mínimo	-99,9 %	-100,0 %	-90,0 %	-93,1 %
Mediana	-32,1 %	-68,4 %	9,6 %	21,3 %
Máximo	440,8 %	6.229,6 %	181,0 %	168,5 %

*Tabla 64: Resumen del valor absoluto del error porcentual en el número de transacciones*

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	62,8 %	94,8 %	72,0 %	73,8 %
Desv. Estándar	41,7 %	281,9 %	73,0 %	59,7 %
Mínimo	0,0 %	0,0 %	2,0 %	10,5 %
Mediana	58,4 %	74,8 %	53,3 %	39,3 %
Máximo	440,8 %	6.229,6 %	210,7 %	168,5 %

## Variable Estado civil

Los resultados obtenidos considerando el número de transacciones, el período de compra, el período de observación y la variable estado civil, se muestran en las dos tablas siguientes.

*Tabla 65: Resumen del error porcentual en el número de transacciones*

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	-10,0 %	-25,3 %	34,7 %	41,9 %
Desv. Est.	75,6 %	293,2 %	85,1 %	84,9 %
Mínimo	-99,9 %	-100,0 %	-90,0 %	-93,3 %
Mediana	-31,7 %	-68,9 %	9,6 %	26,9 %
Máximo	458,9 %	6.615,1 %	181,0 %	162,0 %

*Tabla 66: Resumen del valor absoluto del error porcentual en el número de transacciones*

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	63,3 %	93,7 %	72,0 %	73,0 %
Desv. Estándar	42,5 %	282,3 %	73,0 %	58,0 %
Mínimo	0,0 %	0,0 %	2,0 %	15,4 %
Mediana	58,5 %	73,7 %	53,3 %	36,6 %
Máximo	458,9 %	5.981,6 %	210,7 %	153,8 %

## Variable Edad

Considerando los atributos RFM y la variable edad de cada uno de los clientes, se obtienen los siguientes errores en el pronóstico del número de transacciones a nivel individual.

*Tabla 67: Resumen del error porcentual en el número de transacciones*

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	-10,0 %	25,5 %	34,7 %	41,6 %
Desv. Est.	75,2 %	296,9 %	85,1 %	84,8 %
Mínimo	-99,8 %	-100,0 %	-90,0 %	-93,4 %

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Mediana	-31,7 %	-69,2 %	9,6 %	28,3 %
Máximo	438,4 %	6.977,1 %	181,0 %	164,8 %

*Tabla 68: Resumen del valor absoluto del error porcentual en el número de transacciones*

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	63,1 %	94,5 %	72,0 %	73,7 %
Desv. Estándar	42,2 %	282,6 %	73,0 %	59,1 %
Mínimo	0,0 %	0,0 %	2,0 %	11,7 %
Mediana	58,4 %	75,6 %	53,3 %	40,7 %
Máximo	438,4 %	6.977,1 %	210,7 %	164,8 %

### Variable Región

Los resultados incorporando la región a la cual pertenece el cliente son:

*Tabla 69: Resumen del error porcentual en el número de transacciones*

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	-9,9 %	-25,3 %	34,7 %	41,9 %
Desv. Est.	75,1 %	294,6 %	85,1 %	84,9 %
Mínimo	-99,9 %	-100,0 %	-90,0 %	-93,2 %
Mediana	-31,6 %	-69,0 %	9,6 %	24,8 %
Máximo	456,0 %	6.126,4 %	181,0 %	166,6 %

*Tabla 70: Resumen del valor absoluto del error porcentual en el número de transacciones*

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	63,0 %	94,5 %	72,0 %	73,9 %
Desv. Estándar	42,1 %	280,2 %	73,0 %	59,3 %
Mínimo	0,0 %	0,0 %	2,0 %	11,1 %

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Mediana	58,3 %	75,0 %	53,3 %	40,0 %
Máximo	456,0 %	6.126,4 %	210,7 %	166,6 %

### Variable Antigüedad

Los errores asociados a la estimación del número de transacciones considerando los atributos RFM y la antigüedad de la relación de cada uno de los clientes con la empresa se muestran a continuación.

*Tabla 71: Resumen del error porcentual en el número de transacciones*

Error Porcentual				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MPE)	-9,8 %	-25,5 %	34,7 %	42,1 %
Desv. Est.	75,2 %	296,1 %	85,1 %	85,1 %
Mínimo	-99,9 %	-100,0 %	-90,0 %	-93,5 %
Mediana	-31,4 %	-69,3 %	9,6 %	27,2 %
Máximo	439,0 %	7.248,6 %	181,0 %	165,8 %

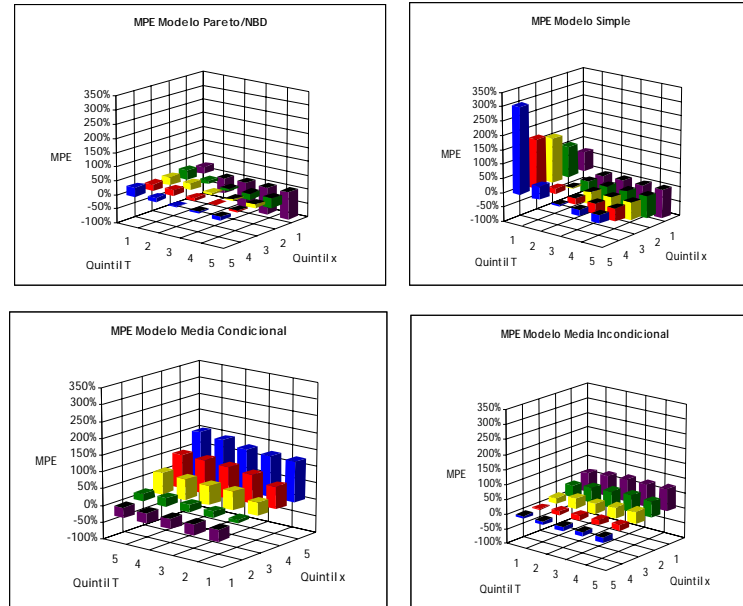
*Tabla 72: Resumen del valor absoluto del error porcentual en el número de transacciones*

ABS(Error Porcentual)				
Indicador	Modelo Pareto/NBD	Modelo Simple	Media Condicional	Media Incondicional
Promedio (MAPE)	63,0 %	94,4 %	72,0 %	74,1 %
Desv. Estándar	42,1 %	281,8 %	73,0 %	59,3 %
Mínimo	0,0 %	0,0 %	2,0 %	11,4 %
Mediana	58,3 %	75,9 %	53,3 %	36,4 %
Máximo	439,0 %	7.24,6 %	210,7 %	165,8 %

## Anexo 5: MPE del pronóstico de $x$ por quintiles

**Quintiles  $x$ -T:** Se muestran los valores de MPE del pronóstico del número de transacciones por quintiles de  $x$  y T.

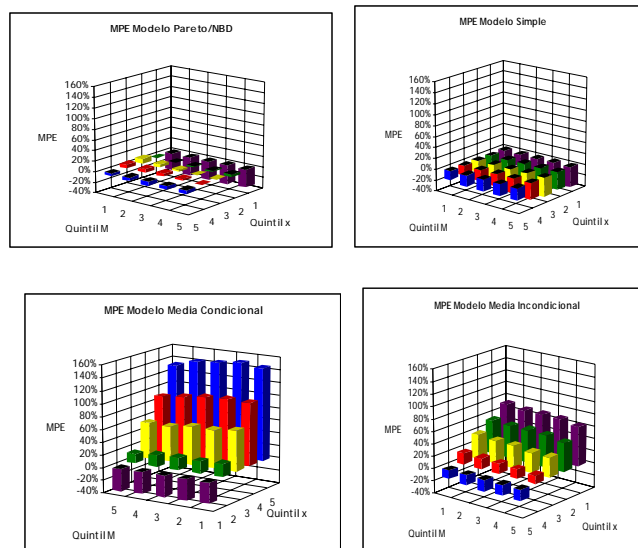
*Figura 75: MPE del número de transacciones por quintil de  $x$ -T*



*Fuente: Elaboración propia*

**Quintiles  $x$ -M:** Se muestran los valores de MPE del pronóstico del número de transacciones por quintiles de  $x$  y M.

*Figura 76: MPE del número de transacciones por quintil de  $x$ -M*



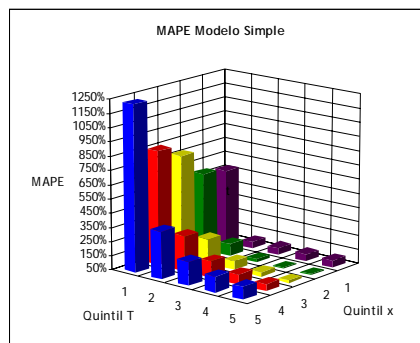
*Fuente: Elaboración propia*

**Anexo 6:** Efecto de la estimación de  $x$  considerando el modelo Simple modificado

Los errores en la estimación del número de transacciones se realiza considerando los clientes inactivos, pero no la constante, debido a que se desea ver el efecto real en los pronósticos de este indicador.

Como se ha mencionado, el modelo Simple genera problemas cuando se tratan de clientes nuevos, es decir pertenecen al quintil 1 de  $T$ .

*Figura 77: MAPE del número de transacciones*



*Fuente: Elaboración propia*

Una manera de corregir este hecho es considerar las variables socio-demográficas seleccionadas para estimar el número de transacciones para este grupo de clientes, las variables género, edad y región. Luego, el modelo Simple pronostica el número de transacciones de la siguiente manera:

- Cientes que pertenecen al quintil 1 de  $T$   $\rightarrow \hat{x} = \bar{x}_{ijk}$

Donde:  $\bar{x}_{ijk}$  = promedio del número de transacciones de los clientes con género  $i$ ,

edad  $j$  y región  $k$ .

$i$  =valores de la variable género

$j$  =valores de la variable edad

$k$  =valores de la variable región

- El resto de los clientes

$$\hat{x} = x \cdot \left( \frac{T_{test}}{T+1} \right) \cdot P(activo), \text{ con } P(activo) = \frac{t_x + 1}{T + 1}$$

A continuación se muestran los errores en el pronóstico del número de transacciones sin la modificación del cálculo para los clientes del quintil 1 de T.

*Tabla 73: Errores de pronóstico del modelo Simple*

Modelo Simple		
Indicador	ABS(Error Porcentual)	Error Porcentual
Promedio	211,8 %	145,6 %
Desv. Estándar	959,4 %	971,6 %
Mínimo	0,0 %	-100,0 %
Mediana	91,0 %	1,9 %
Máximo	18.300,0 %	18.300,0 %

En la siguiente tabla, se muestra considerando el modelo Simple modificado.

*Tabla 74: Errores de pronóstico del modelo Simple modificado*

Modelo Simple modificado		
Indicador	ABS(Error Porcentual)	Error Porcentual
Promedio	105,2	25,4
Desv. Estándar	125,2	161,6
Mínimo	0,0	-100,0
Mediana	79,6	-33,2
Máximo	3.190,9	3.190,9

El hecho de distinguir los clientes que realizan la primera compra hace a lo más 40 días de los demás, permite reducir el error de pronóstico del número de transacciones en 106,6% y la desviación estándar en 834,2%.