



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

**VALORIZACIÓN DE LOS USUARIOS DE TARJETA DE CRÉDITO DE UNA
TIENDA POR DEPARTAMENTOS INCORPORANDO EL RIESGO FINANCIERO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

RICARDO ANDRES EGUIA JACOB

**PROFESOR GUÍA:
RICARDO MONTOYA M.**

**MIEMBROS DE LA COMISIÓN:
LUIS ABURTO
MANUEL REYES**

**SANTIAGO DE CHILE
ABRIL 2009**

RESUMEN DE LA MEMORIA
PARA OPTAR AL TITULO DE
INGENIERO CIVIL INDUSTRIAL
POR: RICARDO EGUIA JACOB
FECHA: 03/04/2009
PROF. GUIA: SR. RICARDO MONTOYA

VALORIZACIÓN DE LOS USUARIOS DE TARJETA DE CRÉDITO DE UNA TIENDA POR DEPARTAMENTOS INCORPORANDO EL RIESGO FINANCIERO

Las grandes tiendas por departamento han ampliado su gama de productos ofrecidos, incluyendo los servicios financieros desarrollando sus propias tarjetas de crédito, que permiten mayor flexibilidad de pago y mayor alcance de productos. Sin embargo, existe una diversa cartera de clientes, como algunos que no compran regularmente a cuotas y otros que no cumplen con sus pagos, introduciendo así el riesgo. Entonces surge la importancia de identificar cuáles son los clientes más valiosos para la empresa.

El presente trabajo de título se desarrolla en una empresa de retail la cual cuenta con su propia tarjeta de crédito. La memoria tiene como objetivo estimar el valor de los clientes incorporando el riesgo financiero. Para cumplirlo, se hace énfasis al estudio de dos variables fundamentales: El interés generado por el uso de la tarjeta y el riesgo. El interés depende de muchos factores como lo son el holding, el número de cuotas, el monto de la compra, entre otros. Esto convierte el valor del interés en una variable de alta volatilidad que no permite predecir un valor exacto de este. Por lo tanto se decide ordenar a los clientes según su valor, dividiéndolos en 4 grupos.

Para determinar el valor de los clientes se utilizará la fórmula de Lifetime Value, considerando como margen el interés que pagan los clientes al pagar por cuotas y por otra parte el riesgo financiero. Para estimar el interés, se utilizan 3 conjuntos de variables distintos combinados con 4 tipos de modelos de demanda: Lineal, Log-Normal, Poisson y Binomial Negativo; para así, identificar que combinación de estos permite valorizar y ordenar a los clientes con mayor nivel de precisión utilizando la matriz de confusión como herramienta de medición. Los dos primeros conjuntos de variables son referentes a las variables RFM mientras que el otro utiliza además variables demográficas y otras asociadas a la tarjeta.

Mediante un análisis por cuartiles, se obtiene que el cuartil superior entrega el 80% del margen interés a la firma mientras que el resto entrega 15%, 4% y 1%. El valor de cada grupo es de \$256Mil, \$41Mil, \$13Mil y \$1,6Mil respectivamente, donde el cuartil superior e inferior muestran mayor estabilidad que los cuartiles centrales al mantenerse un mayor número de clientes en el mismo grupo de un periodo a otro. Se identifica que el modelo de poisson presenta los mejores resultados de distinción de clientes para cada conjunto de variables mostrando una leve superioridad con el set RFM, llegando así a ubicar aproximadamente el 49% de los clientes en su grupo correspondiente, el 43% en la vecindad y tan solo el 8% en los grupos más apartados. Destacando su mayor nivel de acierto para el mejor y peor cuartil con un 58,5% y 57% respectivamente.

La metodología propuesta permitiría a la empresa identificar a sus mejores clientes tomando en consideración el riesgo asociado.

Índice

1. Introducción.....	1
2. Justificación y Planteamiento del Proyecto	3
3. Objetivos	5
3.1 Objetivo General.....	5
3.2 Objetivos Específicos.....	5
4. Alcances.....	5
5. Metodología.....	6
Comprensión del Problema.....	6
Preparación de los Datos.....	6
Estudio Situación Actual y Elementos de la Empresa.....	6
Creación y Evaluación Modelos.....	6
Conclusiones.....	7
6. Marco Conceptual	8
6.1 Buenos y Malos Clientes	8
6.2 Determinación del valor del cliente	8
6.2.1 Limitación de CLV.....	11
6.3 Modelos de Comportamiento.....	11
6.3.1 Regresión lineal	12
6.3.2 Regresión Log- Log.....	13
6.3.3 Modelo Semi- Log.....	13
6.3.4 Modelo de Poisson.....	14
6.3.5 Modelo Binomial Negativo.....	15
6.4 Métodos de Estimación.....	15
6.4.1 Estimación mediante mínimos cuadrados ordinarios.....	16
6.4.2 Estimación mediante método de Máxima Verosimilitud	17
6.5 Medición de error y bondad de ajuste.....	18
6.5.1 R^2 y R^2 Ajustado.....	18
6.5.2 Indicadores AIC y BIC	19
6.5.3 Indicador de Sobredispersión (Overdispersion)	20
6.5.4 Desviación Estándar	20
6.5.5 Correlación de Spearman	20
6.5.6 Matriz de Confusión	21
7. Aplicación de la Metodología.....	22

7.1 Comprensión del Problema	22
7.2 Preparación de los Datos.....	23
7.2.1 Base de Datos Principal	23
7.2.2 Base de Ventas	24
7.3 Descripción de Elementos del Sistema Crediticio de la Empresa.....	26
7.3.1 Análisis agregado de los clientes	26
7.3.2 Riesgo	29
7.3.3 Pago de Interés	34
7.4 LTV	40
7.4.1 Lifetime Value Histórico	40
7.4.1.1 Riesgo por LTV	42
7.4.1.2 Dinámica del Valor del Cliente	42
7.4.1.3 Caracterización Muestra y Variables.....	44
7.4.2 Estimación Monto Interés y LTV	48
7.4.2 Modelos RFM simple	49
7.4.2.2 Modelos RFM segregados	60
7.4.2.3 Modelos con variables transaccionales, demográficas y relacionadas a la tarjeta	71
7.4.3 Mejores modelos por conjuntos de variables.....	84
7.5 Conclusiones	89
7.5.1 Conclusiones Generales	89
7.5.2 Conclusiones respecto a los modelos.....	90
7.5.3 Estudios futuros y propuestas de mejora	92
8. Bibliografía	94
9. Anexos	95
Anexo 1: Rango Score Clientes.....	95
Anexo 2: Tendencia de Interés.....	95
Anexo 3: Matriz de Transición LTV.....	96
Anexo 4: Ajuste por R^2	96
Anexo5: Ordenamiento de clientes a nivel individual y por segmento	97

Índice de Tablas

Tabla 1: Ejemplo Matriz de Confusión.....	21
Tabla 2: Rango situación actual	26
Tabla 3: Distribución de clientes con actividad según Rango Score	29
Tabla 4: Descriptivo de variables por nivel de riesgo	30
Tabla 5: Matriz de transición para el riesgo.....	31
Tabla 6: Descriptivos para variables asociadas a la tarjeta de crédito por Holding...34	
Tabla 7: Descriptivos de variables por segmentos de valor	40
Tabla 8: Valor promedio anual de clientes por segmento	41
Tabla 9: Segmentos de valor v/s segmento riesgo.....	42
Tabla 10: Dinamismo LTV para 12 meses	42
Tabla 11: Dinamismo LTV 8.....	43
Tabla 12: Distribución por sexo	44
Tabla 13: Detalle años cuenta.....	44
Tabla 14: Detalle edad	45
Tabla 15: Detalle monto cupo.....	46
Tabla 16: Coeficientes de modelos para conjuntos de variables RFM simple	50
Tabla 17: Bondad de Ajuste para RFM simple	51
Tabla 18: Ordenamiento individual fase uno	52
Tabla 19: Ordenamiento por segmento fase uno	52
Tabla 20: Porcentaje de acierto para modelos fase uno	53
Tabla 21: Modelos Poisson fase uno	55
Tabla 22: Ordenamiento individual fase dos	55
Tabla 23: Ordenamiento por segmento fase dos	55
Tabla 24: Porcentaje de aciertos fase dos	56
Tabla 25: Ordenamiento individual fase tres	58
Tabla 26: Ordenamiento por segmento fase tres	58
Tabla 27: Porcentaje de acierto para modelos fase tres	58
Tabla 28: Coeficientes de modelos para conjuntos de variables RFM segregados ..61	
Tabla 29: Bondad de Ajuste para RFM segregados.....	62
Tabla 30: Ordenamiento individual fase uno	63
Tabla 31: Ordenamiento por segmento fase uno	63
Tabla 32: Porcentaje de aciertos fase uno	64
Tabla 33: Ordenamiento individual fase dos	66
Tabla 34: Ordenamiento por segmento fase dos	66
Tabla 35: Porcentaje de aciertos fase dos	66
Tabla 36: Ordenamiento individual fase tres	68
Tabla 37: Ranking por segmento fase tres.....	68
Tabla 38: Porcentaje de aciertos fase tres	69
Tabla 39: Coeficientes de modelos para conjuntos de variables RFM full	72
Tabla 40: Bondad de ajuste para RFM full	73
Tabla 41: Ordenamiento individual fase uno	74
Tabla 42: Ordenamiento por segmento fase uno	75
Tabla 43: Porcentaje de aciertos fase uno	75

Tabla 44: Ordenamiento individual fase dos	77
Tabla 45: Ordenamiento por segmento fase dos	78
Tabla 46: Porcentaje de aciertos fase dos	78
Tabla 47: Ordenamiento individual fase tres	80
Tabla 48: Ordenamiento por segmento fase tres	80
Tabla 49: Porcentaje de aciertos fase tres	81
Tabla 50: Rango Scoring.....	95
Tabla 51: Bondad de Ajuste Modelo Lineal.....	96
Tabla 52: Bondad de Ajuste Modelo Log-Normal.....	97
Tabla 53: Bondad de Ajuste Modelo Lineal.....	97
Tabla 54: Bondad de Ajuste Modelo Log-Normal.....	97
Tabla 55: Bondad de Ajuste Modelo Lineal.....	97
Tabla 56: Bondad de Ajuste Modelo Log-Normal.....	97

Índice de Gráficos

Gráfico 1: Ejemplo Diagrama de dispersión y recta	12
Gráfico 2: Ejemplo Configuración Log Normal	13
Gráfico 3: Ejemplo método máxima Verosimilitud	18
Gráfico 4: Evolución situación actual.....	26
Gráfico 5: Evolución Monto deuda	27
Gráfico 6: Evolución monto mora	28
Gráfico 7: Evolución Puntaje Score.....	30
Gráfico 8: Interés Promedio por Scoring condicionado a tener actividad	32
Gráfico 9: Interés promedio por rango score condicionado a interés	33
Gráfico 10: Interés total por segmento riesgo.....	33
Gráfico 11: Transacciones por Holding	35
Gráfico 12: Proporción de transacciones con interés por Holding.....	36
Gráfico 13: Contribución de interés	37
Gráfico 14: Ejemplo Cliente A	39
Gráfico 15: Ejemplo Cliente B	39
Gráfico 16: Interés promedio por sexo	44
Gráfico 18: Interés promedio según años de la cuenta	45
Gráfico 17: Distribución años cuenta.....	44
Gráfico 20: Interés promedio según edad	46
Gráfico 19: Distribución edad.....	45
Gráfico 22: Interés promedio según monto cupo.....	47
Gráfico 21: Distribución monto cupo	46
Gráfico 23: Modelo Poisson RFM simple	88
Gráfico 24: Evolución Interés para Holding 1	95
Gráfico 25: Evolución Interés para Holding 2	95
Gráfico 26: Evolución Interés para Holding 3	96

1. Introducción

Las grandes tiendas de retail se caracterizan por su gran tamaño físico, su variada diversidad de productos y servicios ofrecidos a sus clientes como también por sus ubicaciones en los principales lugares del país y la gran competitividad presente en esta industria.

En el marco de este negocio, las tiendas de departamento han extendido las ofertas de productos traspasando la frontera del retail llegando a ofrecer servicios financieros como además el desarrollo de sus propias tarjetas de crédito (Estudio realizado por Cepal en 2007). Estas tarjetas de crédito pueden ser utilizadas dentro de las mismas tiendas o en alianzas, lo cual permite a los consumidores comprar productos que no podrían al contado gracias al pago mediante cuotas asociado a la tarjeta. Además, permiten obtener avances y otros beneficios como seguros, asimilando de esta manera los servicios entregados por un banco.

El impacto de estas tarjetas de retail es tal, que hoy en día existen en Chile sobre 13 millones de tarjetas emitidas triplicando de esta manera en el último tiempo el número de las tarjetas bancarias aprobadas y además representa cerca del 65% de las ventas de una tienda de departamento¹.

Haciendo referencia a Sunil Gupta (Managing Customers as Investments, 2008) “Los clientes son el elemento vital de cualquier compañía”. Sin clientes una firma no tiene ingresos ni ganancias y por lo tanto tampoco valor de mercado. Es por eso que las compañías deben enfocarse no sólo en atraer nuevos clientes, si no más importante aún, conservar a los clientes actuales mediante la creación de relaciones rentables y duraderas a lo largo del tiempo para así asegurar el éxito de la empresa en el largo plazo. Customer Relationship Management (CRM) tiene su origen en lo antes mencionado, establecer una relación duradera con los clientes. CRM basa su estrategia en el cliente para así aumentar la rentabilidad de la firma a lo largo del tiempo. Esto permite entre otras cosas entender y diferenciar a los clientes mediante acciones de marketing, donde el valor y el riesgo asociado a estos pueden variar substancialmente de uno a otro. De hecho según el principio de Pareto (regla 80-20) se postula que el 80% del valor es aportado por el 20% de los clientes. Es por lo tanto muy importante distinguir e identificar a los clientes, que dada su condición pueden ser tratados como un “portafolio de clientes” compuesta tanto por buenos como malos clientes.

Lo mismo ocurre con los usuarios de las tarjetas de crédito, como ya se ha mencionado en todo negocio hay diferentes tipos de clientes, unos más valiosos que otros. Una manera de reconocer y seleccionar a los clientes más rentables en términos de su valor para la empresa, es el enfoque Customer Lifetime Value (CLV o LTV).

¹ Fuente: Compañía de retail.

LTV es definido como el valor presente de todos los ingresos futuros generados por el cliente, en otras palabras, se obtiene el valor del cliente con un enfoque a largo plazo fundamentando sus principios en que el valor del dinero vale más hoy que en el mañana. Sin embargo, hoy en día la gran mayoría de las empresas para determinar cuáles son los mejores clientes de la firma se basan en resultados del año anterior, como el volumen de compra total o cuantas veces compró el cliente (número de transacciones), es decir, una mirada a corto plazo.

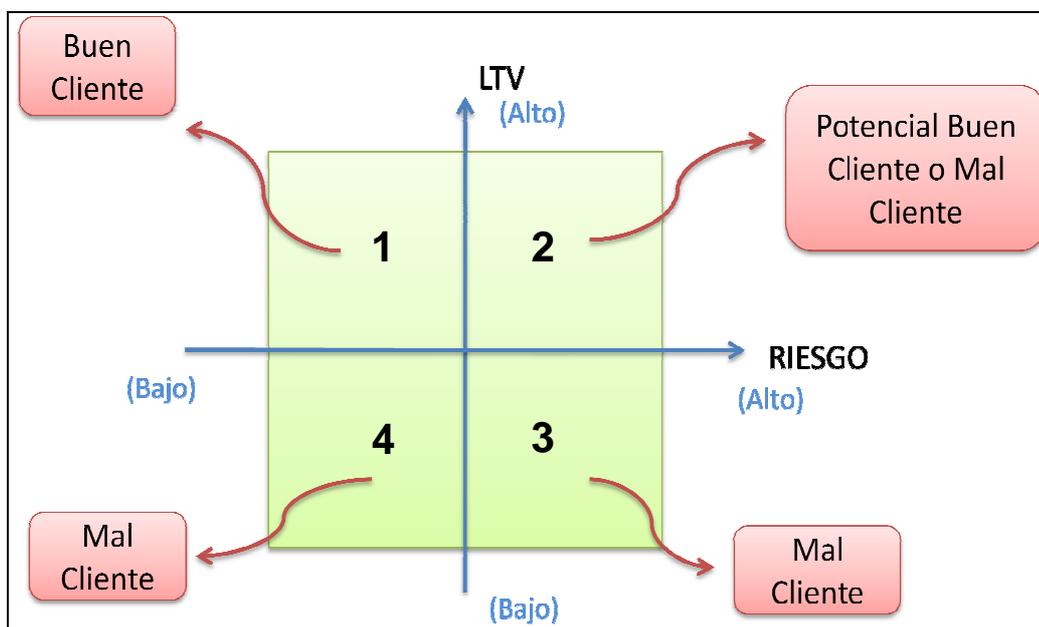
Por otra parte el negocio de las tarjetas de crédito tiene un riesgo asociado puesto que al pertenecer a un área financiera cabe la probabilidad que al comprar a cuotas o pedir un crédito, no se cumpla con la deuda en el futuro, lo que produce que se convierta en un negocio riesgoso. Por una parte el interés que genera el uso de la tarjeta puede llegar a ser una fuente de grandes ingresos para la compañía, de hecho, este representa cerca del 12% de los ingresos de una empresa de retail chilena. Por otro lado si el nivel de morosidad crece demasiado y los consumidores se endeudan y no cumplen con sus pagos, esto puede llegar a ser perjudicial para el éxito de la compañía en el largo plazo. De aquí radica la importancia de valorizar y jerarquizar y distinguir a los clientes respecto al interés que pagarán a mediano o largo plazo, considerando el riesgo asociado que tiene cada uno.

2. Justificación y Planteamiento del Proyecto

La enorme expansión de las grandes compañías de retail hacia nuevas áreas como la financiera, en la cual entrega servicios como avances, seguros y opciones de pago, han ocasionado que éstas no sólo se preocupen de adquirir más clientes y mantenerlos sino que además deben estar conscientes del riesgo asociado a estos. El hecho de dar estas alternativas financieras a sus clientes, puede tener efectos tanto positivos como negativos. Positivo, puesto que en cierto grado estrecha la relación con estos, ya que los retiene cuando estos deben pagar en varias cuotas, como asemejando a un servicio con contrato y la compañía además gana dado los intereses obtenidos por este servicio. Por otra parte negativo, dado que si no se conoce que tan buen pagador es el cliente, este puede obtener una cuantiosa cantidad en término monetarios y nunca devolverlos, lo cual ocasionaría una gran pérdida a la empresa. Esto ocurriría si el cliente es riesgoso.

Para el área de la empresa en el cual se lleva a cabo esta memoria (Área financiera encargada de todo lo referente a la tarjeta de crédito de la empresa), es sumamente importante reconocer a los clientes que paguen mayor interés puesto que son una fuente de ingreso relevante para la firma y distinguirlos además por su nivel de riesgo financiero.

Figura 1: Combinación modelos de Lifetime Value y de Riesgo²



La idea entonces es llegar a un resultado como el mostrado en la Figura 1, donde se pueden identificar distintos tipos de clientes:

- Buenos Clientes (Cuadrante 1): Son los que presentan un LTV alto y un riesgo asociado bajo. Por ende son clientes en general confiables que entregan mucho valor a la empresa.

² Fuente: Elaboración propia.

- Potenciales clientes buenos (Cuadrante 2): Clientes que a pesar de tener un alto valor, tienen alta probabilidad de no cumplir con sus pagos y convertirse así en clientes malos.
- Malos clientes (cuadrantes 3 y 4): Caracterizados por ser clientes que entregan poco valor a la firma. Cabe destacar que los clientes pertenecientes al cuadrante 3 son los peores clientes, ya que no solo entregan poco valor sino que además son riesgosos.

Al clasificar a los clientes según estos resultados, la empresa podrá enfocar sus esfuerzos de marketing a clientes que lo ameriten y que presenten las condiciones necesarias para aumentar la rentabilidad de la compañía. En cifras se gasta aproximadamente entre \$12 y \$15 millones anualmente en campañas publicitarias directas³. Con el proyecto la empresa pretende enfocar sus esfuerzo de marketing apuntando a los clientes más valiosos para la empresa que valen la inversión de marketing directo y de esta manera obtener mayores ganancias al dirigirse a los clientes más rentables para la empresa. Por lo tanto, el proyecto vale en términos monetarios un retorno del gasto en campañas publicitarias directas más un delta lo cual proviene del incentivo por parte de las campañas a comprar en la compañía. En término cualitativos, vale un delta por lograr retener y fidelizar a los clientes más valiosos en término de generación de márgenes financieros.

El proyecto, es difícil llevar a cabo por diferentes razones. Por un lado se necesita una base de datos fidedigna con las variables necesarias para realizar el trabajo las cuales deben ser analizadas para decidir cuáles son relevantes en la implementación de los modelos. Segundo, la gran complejidad de pronosticar el valor futuro de las distintas variables a utilizar (véase capítulo 7.3), especialmente el interés. Esta variable como se verá en el transcurso de la memoria, presenta gran volatilidad entre una compra y otra, pues depende de muchos factores que son analizados. Y finalmente que las compras realizadas por los clientes no son regulares y tampoco así el pago de las deudas.

El objetivo de esta memoria consiste en identificar a los clientes más valiosos para la empresa a lo largo del tiempo incorporando el riesgo financiero (El Riesgo será explicado en el capítulo 7.3.2).

³ Gerente área financiera encargada de las tarjetas de crédito.

3. Objetivos

3.1 Objetivo General

Desarrollar una metodología que permita valorizar a los usuarios de una tarjeta de crédito de una empresa de retail.

3.2 Objetivos Específicos

- Valorizar a los clientes mediante el modelo de lifetime value escogido.
- Encontrar el modelo que obtenga los mejores resultados referentes a su capacidad de distinción entre clientes.
- Ordenar a los clientes tanto a nivel individual como por segmento respecto a la valoración de los clientes en un plazo de tiempo determinado.
- Identificar el conjunto de variables que permitan valorizar y distinguir a los clientes de forma más acertada.

4. Alcances

- Se utilizarán para el estudio solamente la información entregada por una empresa en particular.
- Se utilizará una muestra representativa de la población para llevar a cabo el trabajo. La calibración y validación de los modelos se hará sobre esta muestra.
- No se incluirá en el trabajo clientes que no hayan utilizado la tarjeta de la empresa.
- El modelo sólo será válido para los casos que presenten las variables requeridas para llevar a cabo la predicción.

5. Metodología

En el presente capítulo se expondrá la manera de proceder para cumplir los objetivos planteados en esta memoria. A continuación se exponen en detalle cada uno de los pasos metodológicos a considerar, el cual fue creado utilizando como base la metodología.

Comprensión del Problema

En esta etapa, se procederá a comprender en su totalidad el problema a enfrentar desde la perspectiva de la empresa y los requerimientos que esta quiere. A partir de lo anterior se expondrá una propuesta de solución al problema planteado.

Preparación de los Datos

En esta fase, la idea principal es realizar actividades para familiarizarse con los datos, conocerlos y entenderlos. En otras palabras, mediante el uso de estos es posible comprender su calidad y características, lo cual será de suma importancia para reconocer las variables que pueden ser de utilidad para el proyecto.

Finalmente, por una parte se limpiarán los datos que presenten errores como por ejemplo datos fuera de rango o registros incompletos. Por el otro, se corregirán los datos que sirven para el análisis, pero que sin embargo, están mal ingresados al sistema.

Estudio Situación Actual y Elementos del Sistema Financiero de la Empresa

En este paso, se caracterizará la situación de los clientes referente a su uso de la tarjeta, para diversas variables relacionadas, se ilustrarán tanto las tendencias de estos como su proyección histórica. Además, se determinarán las variables que pueden llegar a entregar valor para la estimación del customer lifetime value y se estudiarán las relaciones existentes entre estas. Por otra parte se describirán distintos elementos relacionados con el ámbito financiero utilizados por parte de la firma, con el objetivo de obtener información relevante para la resolución el problema a resolver en esta memoria.

Creación y Evaluación Modelos

En este punto se estudiará el Lifetime Value histórico de la compañía para así obtener información relevante como cuánto es el valor de un cliente perteneciente al grupo más rentable en comparación al grupo de menos valor al dividir el análisis en cuartiles, conociendo además la contribución que tiene cada uno de ellos respecto al margen interés entregado a la compañía. Luego, se procederá a construir y calibrar variados modelos, con distintas variables explicativas que sean capaces de predecir el valor de la variable dependiente, que en este caso particular será el monto interés pagado por el cliente. Para esto se utilizarán modelos como la regresión lineal, el modelo log-log, el modelo de Poisson y finalmente el modelo binomial negativo. Por otra parte, los resultados de los modelos antes descritos serán en una primera instancia comparados y evaluados para conocer cuál de ellos se acomoda en mejor

medida a lo requerido. Para aquello, se analizará el error de pronóstico asociado a cada uno de los modelos ya mencionados.

Conclusiones

En este apartado, se concluirá respecto a los modelos encontrados y los objetivos planteados en la memoria. A su vez, se entregará sugerencias para desarrollar y trabajos futuros que entreguen aportes tanto como complemento de esta memoria como para la mejora de la compañía. Finalmente, se cruzará este modelo con el modelo de riesgo para determinar cuáles son los mejores clientes para la compañía y se dejará en conciso las acciones que debe seguir la empresa para que esta pueda utilizar la información de una forma efectiva y rentable, donde los resultados entregables son:

- Una metodología para la estimación del lifetime value y el ordenamiento de sus clientes por nivel de valor.
- Metodología para la limpieza de la base de datos.
- Documentación con la caracterización de las variables relevantes.

6. Marco Conceptual

Dentro de este capítulo se expondrán los distintos métodos, modelos y definiciones que se utilizan para el desarrollo de la memoria.

6.1 Buenos y Malos Clientes

La definición de buen o mal cliente puede variar dependiendo de las necesidades de cada empresa o área de esta en particular. En general se supone que alguien que compra frecuentemente y sus montos promedios por transacción son altos es un buen cliente. Sin embargo, el enfoque de esta memoria no toma a los buenos clientes de la manera tradicional. A continuación se dará a conocer las definiciones de “buen” y “mal” cliente utilizadas en el presente estudio para distinguir a los clientes por su valor y así cumplir tanto con los objetivos de la memoria como con los requerimientos del área donde se desarrolla este trabajo.

Buenos clientes: Son las personas que utilizan frecuentemente la tarjeta de la compañía, compran a muchas cuotas, cumplen con sus pagos y paguen lo mínimo que deben pagar. Es decir, generan mucho interés para la firma.

Malos clientes: En contraste con la definición anterior, son las personas que no pagan a cuotas o a muy pocas. Por otra parte también son malos clientes los clientes que no cumplen con sus pagos. Por lo tanto personas que compren frecuentemente con montos promedios considerables pero no utilizan la tarjeta para pagar, o en otras palabras no generan interés al comprar, entonces serán definidos como malos clientes dentro de este contexto.

6.2 Determinación del valor del cliente

Customer Lifetime Value (CLV) ha ganado gran importancia como una métrica de marketing, tanto académicamente como en la práctica. Compañías como IBM, Harrah's, ING y Capital One utilizan CLV como una herramienta para gestionar y medir el éxito de sus negocios. Ahora ¿Por qué existe creciente interés en este concepto?

Gupta, Hanssens y Hardie (Modelin Customer Lifetime Value, 2006) explican varios factores del porqué, los cuales son señalados a continuación.

Primero, hay una presión constante por parte de las compañías en hacer un marketing responsable. Puesto que las actuales métricas utilizadas como conciencia de marca, ventas, entre otras, no son suficientes para mostrar un retorno sobre la inversión de marketing. De hecho, las acciones de marketing pueden llegar a dañar la rentabilidad de un negocio a largo plazo. (Yoo y Hanssen, 2005).

Segundo, métricas financieras como el precio de la acción tampoco resuelven el problema de conocer el valor de sus clientes. Aunque estas métricas son útiles, tienen una capacidad limitada de diagnóstico. Esto pues, estudios han demostrado

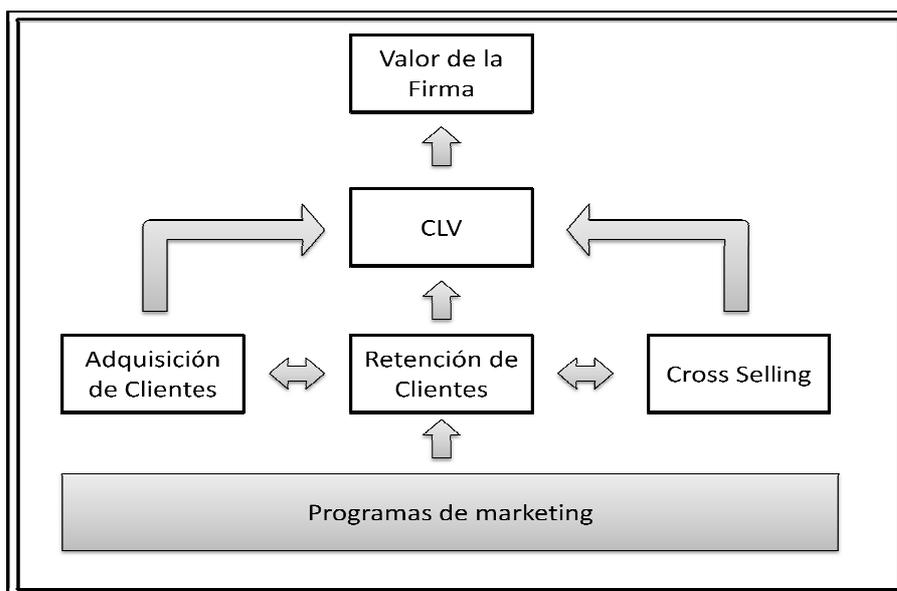
que no todos los clientes son igualmente valiosos (Gupta y Lehman 2005; Lemon y Zeithaml 2004). Por lo tanto, sería deseable poder enfocar los recursos de marketing a diferentes grupos de clientes, específicamente a los más valiosos. Estas respuestas no pueden ser obtenidas mediante métricas financieras. En contraste, el Lifetime value es una métrica capaz de identificar a los clientes más rentables para la compañía y así distribuir los recursos de marketing de una manera eficiente (Kumar y Reinartz 2006).

Como último factor se puede mencionar la gran mejora de la tecnología de la información que permite recolectar enorme cantidad de información transaccional, la cual permite a la firma revelar preferencias más que intenciones. Esta información es fundamental para poder reconocer y diferenciar entre los distintos clientes.

La Figura 2 usada por investigadores como Gupta, Lehman y Petersen en sus estudios de LTV, muestra que lo que la firma hace referente a sus acciones de marketing, influencia el comportamiento del cliente como por ejemplo la retención de estos, las ventas cruzadas, etc. Por lo cual, en efecto, afecta el CLV de los clientes o en otras palabras su rentabilidad para la empresa (véase gráfico). Además esta ilustración destaca los distintos estudios que se pueden modelar. Algunos de estos son el impacto de los programas de marketing para la adquisición, ventas cruzadas, retención de clientes. Otros consisten en conocer la fuga de clientes y el porqué de esto.

La presente memoria consistirá en modelar el valor presente que tendrá el cliente referente al interés que pagará a la empresa en el futuro y así diferenciar entre “buenos” y “malos” clientes.

Figura 2: Efecto Lifetime Value⁴



⁴ Fuente: Modeling Customer Lifetime Value. 2006. S. Gupta, D. Hanssens, B. Hardie.

Customer Lifetime Value (CLV) es generalmente definido como el valor presente de todas las ganancias futuras obtenidas de un cliente a través de su vida de relación con una empresa.

Cabe agregar que CLV es similar a un flujo de caja con tasa de descuento usado en finanzas. Sin embargo, hay dos diferencias claves, primero que todo CLV está definido y estimado típicamente a nivel individual o nivel segmento. Esto permite diferenciar entre clientes que son más rentables que otros. Por otra parte hay modelos CLV que incorporan la posibilidad de que los clientes se vayan con la competencia en el futuro. Se representa el CLV definido por Gupta, Lehman y Stuart como⁵:

$$CLV = \sum_{t=1}^T \left(\frac{(P_t - C_t) * R_t}{(1 + i)^t} \right) - AC$$

Fórmula 1

Donde P_t = Precio pagado por el consumidor en el tiempo t .

C_t = Costo directo por el servicio en el período t .

i = Tasa de descuento de la firma.

R_t = Probabilidad de que el cliente vuelva a comprar.

AC = Costo de adquisición.

T = Tiempo horizonte para estimar CLV.

Algunas consideraciones para la presente memoria:

- P_t será definido como el interés pagado por el cliente en el período t .
- No se tomará en cuenta ningún tipo de costo, puesto que el costo asociado al cliente es un dato desconocido en este caso.
- LTV será sacado para un período de 12 meses.
- Se utilizarán sólo a clientes que hayan presentado actividad durante el período en estudio.
- Por otro lado, la probabilidad de que el cliente vuelva a comprar será tomada como 1, ya que el LTV será sacado para un período de 12 meses utilizando sólo a clientes que hayan presentado actividad durante el período en estudio y por ende se supondrá que tendrá alguna actividad durante el próximo período.
- La tasa de descuento es de un 15% anual. No existe alguna teoría referente al valor que este deba tomar, sin embargo, se recomienda en la literatura utilizar una tasa de descuento cercana al 15% para horizontes de tiempo de corto plazo.

Por lo tanto se estará estimando el ingreso futuro del cliente dado por el interés que pagará a la firma durante los próximos 12 meses siguientes.

⁵ Gupta, Lehmann y Stuart 2004 como también Reinartz y Kumar 2003

Entonces la fórmula para obtener el LTV quedaría definida de la siguiente manera:

$$LTV = \sum_{t=1}^T \frac{FF_t}{(1+i)^t}$$

Fórmula 2

Con FF = Flujos Futuros de Interés por parte del consumidor en el período t.

i = Tasa de descuento de un 15%.

T = Horizonte de tiempo igual a un año, es decir, T =12.

6.2.1 Limitación de CLV

Dado que la mayoría de los modelos CLV están diseñados con una tasa de descuento, se dice que éstos están muy conectados tanto con las métricas del área de marketing como las de finanzas. Sin embargo, hay un concepto financiero que ha sido ampliamente descuidado, el riesgo. Hasta el momento las investigaciones realizadas respecto al CLV se han enfocado en predecir o comprender los niveles futuros de las transacciones respecto a los clientes. Por ende, si entramos a una industria la cual ofrece servicios financieros, en la cual el riesgo es una variable importantísima para conocer y caracterizar a los clientes, los modelos actuales de CLV pierden validez, puesto que no consideran el riesgo asociado a los clientes.

A continuación se expondrán los métodos que se utilizarán para estimar la variable dependiente, necesaria para el desarrollo de la memoria.

6.3 Modelos de Comportamiento

Para enfrentar el problema de estimación de la variable dependiente, en este caso el margen interés, se han desarrollado una gran cantidad de metodologías o modelos, como por ejemplo los modelos de regresión simple, métodos de suavización exponencial, series de tiempo, redes neuronales, etc. El utilizar uno u otro modelo depende del problema que se esté enfrentado y la disponibilidad de datos con la que se cuente. En este estudio se utilizarán los modelos regresivos.

La regresión es una técnica estadística utilizada para simular la relación existente entre dos o más variables (Gujarati, 2004). Por lo tanto se puede emplear para construir un modelo que permita predecir el comportamiento de una variable dada. La regresión es muy utilizada para interpretar situaciones reales, pero comúnmente se hace de mala forma, por lo cual es necesario realizar una selección adecuada de las variables que van a construir las ecuaciones de la regresión, ya que tomar variables que no tengan relación en la práctica, nos arrojará un modelo carente de sentido, es decir ilógico.

Según sea la dispersión de los datos (nube de puntos) en el plano cartesiano, pueden darse alguna de las siguientes relaciones, Lineal, Logarítmica, Exponencial,

Cuadrática, entre otras. En este estudio en particular se estudiarán los modelos predictivos de regresión múltiple: Lineal, Log- Log, Poisson y Binomial Negativo.

6.3.1 Regresión lineal

La regresión lineal estima los coeficientes de la ecuación lineal, con una o más variables independientes que mejor prediga el valor de la variable dependiente. Hay que tener en claro que se llama lineal cuando los valores son lineales en los coeficientes y no necesariamente en las variables predictivas. Por ejemplo, para el caso en estudio, se quiere predecir el total del interés a obtener por el cliente en un plazo determinado a partir de variables independientes como la frecuencia de compra, tiempo pasado desde la última compra (recency), el monto de la compra, número de cuotas, etc.

Tomemos una representación de la regresión lineal generalizada:

$$Y = \alpha + \beta * X_1 + \delta * X_2 + \dots + \varphi * X_n + \varepsilon$$

Fórmula 3

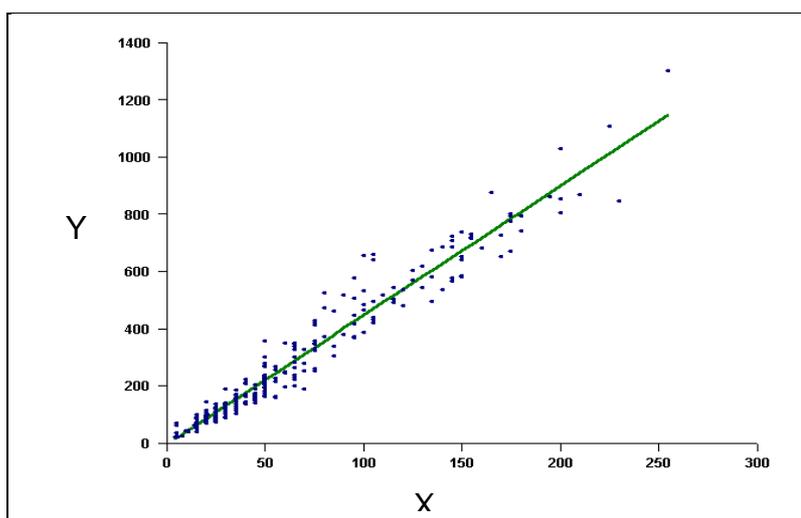
En la cual: Y es la variable dependiente.

$X_{1\dots n}$ son las variables independientes.

α, β, δ y φ , valores a ser estimados.

A continuación se muestra la configuración que debieran seguir los datos para realizar una regresión lineal significativa.

Gráfico 1: Ejemplo Diagrama de dispersión y recta⁶



Este modelo presenta la ventaja de que es fácil de implementar y se puede utilizar en la mayoría de los casos. La desventaja es que no asegura valores positivos (El interés debe ser siempre mayor o igual a 0).

⁶ Fuente: Reproducción gráfico libro Gujarati. 2004.

6.3.2 Regresión Log- Log

También conocido como modelo de regresión log-lineal, esta también se puede correr mediante una regresión lineal puesto que utiliza los mismos supuestos. Sin embargo, la interpretación de sus coeficientes es distinta al enfoque anterior. El modelo Log-Log se representa de la siguiente manera:

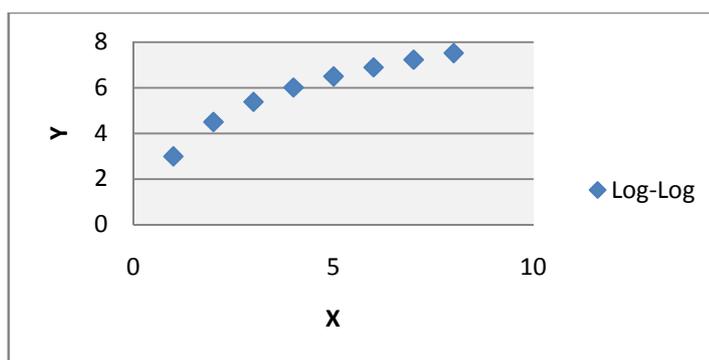
$$\text{Log}(Y) = \alpha + \beta * \text{Log}(X_1) + \delta * \text{Log}(X_2) + \dots + \varphi * \text{Log}(X_n) + \varepsilon$$

Fórmula 4

En este caso los coeficientes no representan un cambio de valor “total” sobre la variable endógena sino que representa un cambio porcentual. En otras β mide la elasticidad de Y con respecto a X.

Además este modelo es utilizado en los casos en los cuales la variable dependiente toma solo valores positivos. Lo cual se ve reflejado en la estructura de la ecuación anterior. A continuación un ejemplo de una configuración de dispersión Logarítmica:

Gráfico 2: Ejemplo Configuración Log Normal



La ventaja de este modelo es que asegura valores positivos. La desventaja es que los valores entregados al modelo deben ser mayores a 0 o sino quedan casos indefinidos.

6.3.3 Modelo Semi- Log

Este modelo es también conocido como modelo de crecimiento(o Log-Lin). Sus coeficientes miden el cambio proporcional constante o relativo en Y para un cambio absoluto dado en el valor del regresor X. Este se puede representar de la siguiente manera:

$$\text{Log}(Y) = \alpha + \beta * X_1 + \delta * X_2 + \dots + \varphi * X_n + \varepsilon$$

Fórmula 5

Generalmente es usado, tal como dice su nombre, para medir la tasa de crecimiento de la variable dependiente. Además compartiendo cierta bondad con el modelo anterior, posibilita la posibilidad de que la variable dependiente solo tome valores positivos.

6.3.4 Modelo de Poisson

La regresión de Poisson se puede utilizar para modelar una variada matriz de sucesos. Este modelo se puede representar de forma sencilla en la siguiente manera:

$$\ln(\lambda) = \alpha + b * X$$

Fórmula 6

Al observar la ecuación anterior, se puede apreciar que es muy similar al modelo Log-Lin, sin embargo su principal diferencia radica en su distribución. Este modelo asume que la variable dependiente Y sigue una distribución de Poisson, es decir:

$$f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!},$$

Fórmula 7

Donde: e es una exponencial (equivalente a 2,718 aproximado)

K es el número de éxitos dentro de un evento

λ es un número positivo, igual al número esperado de ocurrencias que ocurren durante algún intervalo dado.

La distribución de Poisson es aplicado a sistemas con un gran número de eventos posibles, cada uno de la cuales es raro, en otras palabras, de baja probabilidad de ocurrencia.

Por ejemplo, la probabilidad que un teléfono seleccionado al azar esté envuelta en una llamada es muy baja, pero sin embargo, el número de teléfonos existentes es muy grande de tal modo que hay un conjunto finitos de número que tienen una llamada activa en cualquier momento. O la probabilidad de que un automóvil esté envuelto en un accidente es muy baja, pero hay tantos automóviles que de todas maneras hay accidentes cada día.

En el caso particular de la memoria, la probabilidad que un cliente realice una compra utilizando la tarjeta de crédito particular de la compañía en un día particular pagando con interés es muy baja, sin embargo, la cantidad de personas que tienen la tarjeta es muy elevado, por lo cual existe un número finito de clientes que en ese día sí realizarán una transacción con su tarjeta pagando interés.

Por lo tanto este modelo tiene la ventaja de adaptarse a la realidad de los datos tal como se explica en el ejemplo anterior. Además asegura valores positivos para la variable dependiente. Las desventajas es que su método de estimación de coeficientes es más complejo, sin embargo, dada las tecnologías de hoy en día esto no es un gran problema.

6.3.5 Modelo Binomial Negativo

Este modelo permite modelar situaciones iguales o muy similares a las utilizadas con el modelo de Poisson. La forma general del modelo es de la siguiente manera:

$$\ln(\lambda) = \alpha + b * X$$

Fórmula 8

Lo cual permite que los valores entregados sean número positivos. Ahora, en este caso el modelo utiliza una distribución binomial negativa dada por:

$$f(k; r, p) = \binom{k+r-1}{k} \cdot p^r \cdot (1-p)^k$$

Fórmula 9

Donde

$$\binom{k+r-1}{k} = \frac{\Gamma(k+r)}{k! \cdot \Gamma(r)} = (-1)^k \cdot \binom{-r}{k}$$

Fórmula 10

y

$$\Gamma(r) = (r-1)!$$

Fórmula 11

Este método se usa típicamente cuando algunos estadísticos que vienen con el reporte de resultados al correr el modelo de Poisson muestran que existe sobredispersión. La distribución de probabilidad usada por el modelo Binomial Negativo permite mayor explicación para la sobredispersión de los datos. Estos estadísticos serán descritos en la sección 6.5. Cabe agregar que este modelo presenta las mismas ventajas y desventajas que el modelo de Poisson.

Como a priori no se puede conocer qué modelo tendrá los mejores resultados no se puede descartar ninguno. Por esto, en la presente memoria se analizan cada uno de ellos para así escoger el mejor modelo según los resultados obtenidos.

La pregunta ahora es: ¿Pero cómo se estiman los coeficientes?

6.4 Métodos de Estimación

Existen principalmente dos maneras de hacer esto. La primera es mediante mínimos cuadrados ordinarios y la segunda mediante el método de máxima verosimilitud. Cabe agregar que la primera está limitada a ecuaciones con distribución normal, por lo cual no es aplicable a los modelos de Poisson y binomial negativo. Por otro lado el método de máxima verosimilitud se puede aplicar a este tipo de modelos. Cabe destacar además que para muestras grandes los resultados de las estimaciones son prácticamente iguales al otro método ya mencionado.

6.4.1 Estimación mediante mínimos cuadrados ordinarios

Tomemos en cuenta la siguiente ecuación derivada de la fórmula anterior:

$$f(\chi) = \alpha + \beta * \chi + \varepsilon$$

Fórmula 12

En que: $f(x)$ Es el valor predicho por el modelo.

ε El valor del error asociado al modelo. De esta manera $\varepsilon = Y - \hat{Y}$

Ahora, para obtener los coeficientes se utiliza el método MCO, es decir, mínimos cuadrados ordinarios. El cual consiste básicamente en minimizar la suma de los errores elevados al cuadrado que se tendrían (véase fórmula 13), suponiendo distintos valores posibles para los parámetros, al estimar los valores de la variable endógena a partir de los de las variables exógenas en cada una de las observaciones muestrales, usando el modelo propuesto, y comparar esos valores con los que realmente tomó la variable endógena.

$$E_c(f) = \frac{\sum_{k=1}^n (\varepsilon_k)^2}{n}$$

Fórmula 13

Los parámetros que lograrán ese mínimo, el de la suma de los errores cuadráticos, se acepta que son los que estamos buscando de acuerdo con criterios estadísticos. El método MCO también proporciona información (en forma de ciertos valores estadísticos adicionales, que se obtienen además de los de los parámetros) para ver en qué medida los valores de los parámetros que obtenidos resultan fiables, por ejemplo, para hacer contrastes de hipótesis, esto es, ver si ciertas suposiciones que se habían hecho acerca del modelo resultan o no ciertas. Se puede usar también esta información adicional para comprobar si se pueden prescindir de algunas de esas variables, para ver si es posible que los valores de los parámetros hayan cambiado con el tiempo (o si los valores de los parámetros son diferentes en una zona económica de los de otra, por ejemplo), o para ver en qué grado son válidas predicciones a cerca del futuro valor de la variable endógena si se supone que las variables exógenas adoptan nuevos valores.

La estimación de los coeficientes de los modelos que vienen a más adelante, son estimados mediante el método de máxima verosimilitud, dado que el método MCO como ya se ha explicado no es aplicable para estos casos.

6.4.2 Estimación mediante método de Máxima Verosimilitud

De acuerdo con Gujarati (2001), el método de de Máxima Verosimilitud, presenta algunas propiedades teóricamente más fuertes que las del método de mínimos cuadrados ordinarios, este método también es conocido como método de “razón de mínima varianza.”

A diferencia del método MCO, el método de máxima verosimilitud nos dice que escogeremos como valor estimado del parámetro aquél que tiene mayor probabilidad de ocurrir según lo que hemos observado, es decir aquél que es más compatible con los datos observados.

Formalmente el método queda definido de la siguiente manera:

Sea X una variable aleatoria con función de probabilidad $f(X|\theta)$, donde θ es un parámetro desconocido. Sean X_1, \dots, X_n los valores observados en una muestra aleatoria de tamaño n . La función de verosimilitud de la muestra es:

$$L(X, \theta) = f(X_1, X_2, \dots, X_n)$$

Fórmula 14

Ahora bien, como las variables son independientes entonces la función de verosimilitud puede expresarse como:

$$L(X, \theta) = f_1(X_1, \theta) * f_2(X_2, \theta) \cdots f_n(X_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

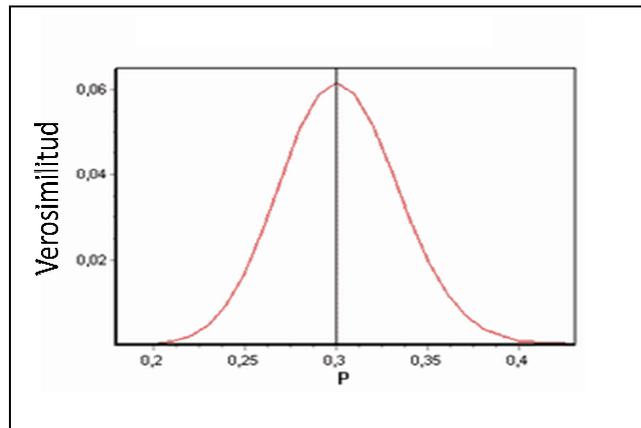
Fórmula 15

Por lo tanto al ser la función de verosimilitud una función de densidad conjunta de la muestra aleatoria, entonces el objetivo que se pretende con el método de estimación tal como ya se había comentado, es encontrar aquellos valores de los parámetros que maximicen la probabilidad de obtener los valores que se dieron en la muestra, esto es conocido como el método de máxima verosimilitud..

Por consiguiente, para encontrar estos estimativos se debe derivar la función de verosimilitud con respecto a cada uno de los parámetros a estimar, igualar a cero y despejar el respectivo valor. En otras palabras, el estimador de máxima verosimilitud de θ es el valor de θ que maximiza la función de verosimilitud $L(X, \theta)$.

A modo de ejemplo a lo antes mencionado, se puede apreciar en el gráfico 4 en donde la maximización de la función de verosimilitud de un solo parámetro p . En este caso $p = 0,3$ maximiza la función de verosimilitud.

Gráfico 3: Ejemplo método máxima Verosimilitud



Cabe agregar que por simplicidad en general se maximiza la función logarítmica de esta función conocida como Log-Likelihood, por lo cual la función a maximizar es la siguiente:

$$\text{Log } L(X, \theta) = \sum_{i=1}^n \log f_i(x_i, \theta).$$

Fórmula 16

6.5 Medición de error y bondad de ajuste

Una vez calibrado los modelos en el cual se obtienen los valores de los coeficientes, se necesita conocer cual modelo se adecua mejor a los datos y cuán bien lo hace. Para esto, se debe medir el error asociado a los modelos. Existen varios criterios que dependerán del modelo a utilizar. Para predicciones mediante modelos regresivos se puede usar R². A continuación se dará una breve descripción de cada uno de ellos.

6.5.1 R² y R² Ajustado

R² está descrito de la siguiente manera:

$$R^2 = \frac{RSS}{TSS} = \frac{\text{Suma de cuadrados de la regresión}}{\text{Suma total de los cuadrados}}$$

Esto significa que R² mide la bondad del modelo, cuán bien se ajusta el modelo a los datos, en otras palabras el porcentaje de variación explicada por la regresión. Entonces siguiendo la ecuación anterior, se puede ver que:

R² = 1 => El modelo de ajusta perfectamente a los datos.

R² = 0 => El modelo de regresión no explica en lo absoluto los datos.

R² = 0,63 => El modelo explica el 63% de la varianza.

Un problema de este indicador, es que es una función creciente. Por lo tanto mientras se le inserten más variables al modelo, más incrementará el valor de R^2 , por lo que no se reconocerá en su totalidad si una variable predictiva es significativa o no para el modelo, en cuanto a su impacto hacia este.

Es por esto que se introduce R^2 ajustado, el cual mide la varianza explicada ajustada por los grados de libertad del modelo.

6.5.2 Indicadores AIC y BIC

Los índices AIC y BIC, criterio de información de Akaike y criterio de información bayesiano respectivamente, son dos medidas utilizadas para determinar que tan bien los modelos que usan el método de máxima verosimilitud se ajustan a una base de datos. Donde el mejor modelo es aquel que posea menor valor de AIC o BIC dependiendo el criterio a utilizar.

Los índices AIC y BIC se definen de la siguiente forma:

AIC:

$$AIC = 2 \cdot k - 2 \cdot (\text{Log likelihood}_{max})$$

Fórmula 17

BIC:

$$BIC = k \cdot \ln(n) - 2 \cdot (\text{Log likelihood}_{max})$$

Fórmula 18

Donde,

k = grados de libertad del modelo (o número de parámetros estimados)

n = número de observaciones

$\text{Log likelihood}_{max}$ = valor máximo del Log likelihood obtenido por el modelo

Estos dos índices toman en consideración tanto el ajuste del modelo como el número de parámetros utilizados en el ajuste. El ajuste del modelo es medido negativamente por el término $(-2 \cdot \text{Log likelihood})$ en que mientras más grande sea este término, peor es el ajuste. En tanto la complejidad es medida positivamente en el término $2 \cdot k$, en el caso del AIC, y por $k \cdot \ln(n)$ para el caso del BIC. Estos son una penalización proporcional al número de parámetros que incorpora el modelo. Se puede observar que BIC penaliza de forma más estricta, respecto a AIC.

Ahora bien, ¿Cuál es la diferencia entre AIC y BIC?

Para esto hay que tener claro que es lo que se quiere, por una parte si lo principal es minimizar el error, entonces se recomienda el uso de AIC, por otro lado si se quiere dar mayor consideración a la interpretación de los datos, se sugiere el uso del indicador BIC.

6.5.3 Indicador de Sobredispersión (Overdispersion)

Los estadísticos llamados “Deviance” y “Pearson Chi-Square” entregados en el reporte de resultados al correr la regresión mediante el método de máxima verosimilitud, son utilizados para detectar sobredispersión. Esto se hace al dividir los estos indicadores por los grados de libertad (número de casos menos el número de parámetros en el modelo). Si su valor es mayor que 1 entonces es una señal de que existe sobredispersión. (Mayor detalle en capítulo 7.4)

Por lo tanto, si al correr el modelo de Poisson este obtiene valores mayores que 1 al realizar lo antes descrito, entonces se recomienda el uso del modelo regresivo binomial negativo.

Anticipadamente esto sí ocurre y es el motivo principal en conjunto con que este entrega valores positivos y tiene comportamientos muy similares al modelo de Poisson y por eso su utilización dentro del estudio.

6.5.4 Desviación Estándar

Otra métrica, la cual es muy conocida pero sumamente importante, es la desviación estándar. Supongamos que un modelo obtiene un valor promedio muy parecido al real, pero sin embargo su desviación estándar es alta. Esto implica que a nivel global el modelo puede ser consistente pero a nivel individual su nivel de predicción es malo. Por lo tanto es una métrica que se debe tener en cuenta, sobre todo si se quiere obtener una valorización del cliente a nivel individual o de segmento.

6.5.5 Correlación de Spearman

La correlación de Spearman o por rangos utiliza valores de escala discretos. A diferencia de la correlación de Pearson que mide la relación de magnitudes entre dos variables, este tipo de correlación por rangos permite determinar si existe relación entre dos variables con valores nominales o de jerarquía.

La correlación de Spearman es obtenida de la siguiente manera:

$$\rho = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2} \sqrt{n(\sum y_i^2) - (\sum y_i)^2}}$$

Fórmula 19

6.5.6 Matriz de Confusión

Una matriz de confusión es una herramienta de visualización donde cada columna de la matriz representa el número de predicciones de cada grupo, mientras que cada fila representa a las instancias en el grupo real. Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos o más grupos.

A continuación se muestra una matriz de ejemplo en la cual A y B representan el número de casos que el modelo predice correctamente para los grupos “Adultos” y “Ancianos” respectivamente. Mientras que C y D los desajustes del modelo ubicando a las personas del grupo “Adulto” y “Anciano” en el grupo contrario al cual pertenecen.

Tabla 1: Ejemplo Matriz de Confusión

		Predicho	
		Adulto	Anciano
Real	Adulto	A	B
	Anciano	C	D

En conclusión:

Los primeros indicadores R^2 , AIC y BIC son utilizados principalmente para conocer que tan bien los modelos se ajustan a los datos. Si bien esto es importante, no será el indicador para determinar al mejor modelo puesto que lo que se busca en esta memoria es distinguir entre los “buenos” y “malos” clientes y este indicador no proporciona esa respuesta. Por otro lado, la correlación de Spearman y la matriz de confusión si entregan esta respuesta, por lo cual serán las métricas principales para determinar el mejor modelo para estimar el Lifetime Value y el ordenamiento de los clientes.

7. Aplicación de la Metodología

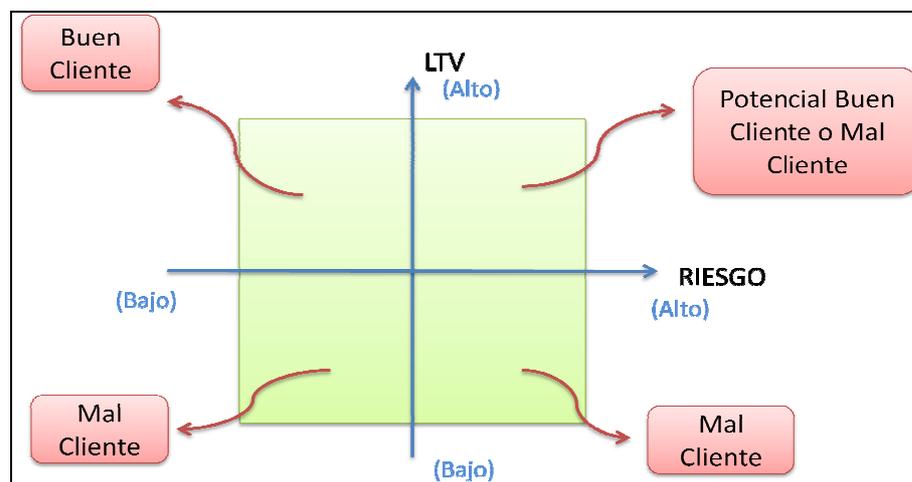
En este capítulo se muestra el desarrollo de la metodología antes mencionada y como se abordan cada una de las etapas.

7.1 Comprensión del Problema

Esta etapa está caracterizada principalmente por reuniones. Las cuales consistieron básicamente en la comunicación de las necesidades y requerimientos de la empresa al memorista respecto al trabajo. A partir de esto se llevó a cabo el análisis y preparación de la manera de enfrentar el problema y finalmente se organizó la estructura de cómo se iba a resolver este.

A continuación se muestra una Figura con el modo en que se aborda el problema:

Figura 3: Combinación modelos de Lifetime Value y de Riesgo



Esto implica crear un modelo de Lifetime Value que valore al cliente no en la forma tradicional conocida en la literatura de marketing, sino respecto al interés pagado por este al momento de realizar compras o utilización de otros servicios entregados por la compañía con el uso de la tarjeta de crédito propia de la empresa.

Una vez obtenido esto se querrá llegar a algo similar a la Figura 3, en la cual se combina tanto el modelo creado de LTV con el modelo de riesgo entregado por la empresa. Y dado el cuadrante en el que estén insertos los clientes, se podrá decir que tan valiosos son para la compañía.

7.2 Preparación de los Datos

Para esta etapa la preparación y limpieza de datos es analizada en dos partes, base de datos principal y base de datos de ventas.

7.2.1 Base de Datos Principal

Esta base representa la base de datos principal del cliente, con una gran cantidad de variables asociadas. Esta base ha ido aumentando su número de variables llegando hoy a tener 250 variables. Por lo tanto su entendimiento era esencial para conocer que variables tendrán utilidad para el proyecto y no descartarlas a priori puesto que pueden llegar a tener un valor fundamental.

Finalmente se escogieron 15 variables de las 250, que fueron seleccionadas por su relevancia con respecto al estudio a llevar a cabo, al tener relación con: El ámbito financiero o más específicamente el uso de las tarjetas de crédito; variables demográficas y variables de compra.

Esta es una base en la cual las variables están compuestas mensualmente, las cuales están enumeradas de acuerdo a la cantidad de meses. De esta manera se facilitó el estudio de la situación actual de los clientes. Esta base consta de 189.816 casos.

Una vez finalizado el punto anterior se realizó un proceso de limpieza de datos. Este consistió primero que todo en:

- Detección de *outliers* (valores fuera de rango)
- Detección de *missing value* (valores perdidos)

Una vez detectados estos valores que “ensucian” el proceso analítico posterior, se debe decidir qué se hará con ellos. Algunas posibilidades que se barajaron:

- Eliminar los casos con valores perdidos de la agrupación
- Reemplazar los valores “extraños” por el promedio de la variable
- Reemplazar los valores perdidos por “0” (sólo si corresponde).

Se optó por la eliminación de estos casos, puesto que la gran mayoría de los casos detectados, eran casos perdidos en las variables relevantes del proyecto.

Errores Detectados en la Base de Datos

Inmediatamente después de la limpieza antes mencionada se realizó un segundo análisis más exhaustivo, detectando lo siguiente:

- Casos con variable “Monto Disponible” tienen valores fuera de lo normal.
- 30.000 clientes (15% de la muestra) en que monto deuda actual es mayor que monto cupo. Por esto se deduce que existen datos no actualizados en el monto cupo.
- 33.000 tienen días mora mayor que 700 llegando hasta 9.000 (17% de la muestra).
- Años cuenta igual a 0 no tienen información completa de meses anteriores

Luego, se tomó la decisión de eliminar los datos de clientes relativamente nuevos que no presentaban un año completo con la cuenta abierta puesto no tenían la información necesaria para el análisis. Además se eliminó la variable “Monto Disponible”. La variable “Monto Cupo” a pesar de estar desactualizada, se dejó en la base puesto que solo será utilizada para la descripción de la situación actual.

Finalmente se eliminaron los casos que presentaban Situación Actual: Cuenta Cerrada. Ya que son personas que no tienen actividad alguna durante el período en estudio, pero sin embargo tienen registro en sus variables, los cuales afectan los resultados.

Hasta este punto, la muestra quedó compuesta de 158.343 clientes.

7.2.2 Base de Ventas

Estas bases presentan muchos errores de distintas dimensiones, tanto para los datos de venta de los holdings de la empresa como de sus aliados respecto a compras u otros servicios utilizados efectuados con la tarjeta.

Las bases se presentan de la siguiente manera:

ID_CLIENTE	ID_AGNO_ MES	DESC_AGNO_MES	ID_CUOTAS	MONTO_DE_CO MPRAS_CMR	N_BOLETAS_ CMR	MONTO_DE_ ABONO	VALOR_CUOTA
------------	-----------------	---------------	-----------	--------------------------	-------------------	--------------------	-------------

A continuación se enumera una lista con los errores encontrados respecto a la base:

1. Valores Cuota Negativos
2. Montos Compras Negativos
3. Montos Abono Negativos
4. Interés Negativo(Monto Compra - Monto Abono) > Cuotas* Valor Cuota
5. Interés > 0 y Cuota = 1
6. Monto Compra = 0
7. Monto Compra > 10.000.000
8. Valor Cuota > 5.000.000
9. Valor Cuota = 0

De hecho era tal la incongruencia de datos que para un holding de la empresa su media en cuanto al interés calculado por los clientes era negativo, ya que se detectaban casos que tenían intereses negativos de -5.000.000 millones. Lo cual alteraba considerablemente los resultados.

7.2.2.1 Metodología para solucionar y limpiar la base de datos:

Acorde a lo anterior el procedimiento tomado fue el siguiente⁷:

1. Crear nuevas variables con el valor absoluto de cada variable.
2. Eliminar los casos fuera de rango, esto es valor cuota mayor que 5 millones. Puesto que el máximo valor dado a un monto de tarjeta es de 5 millones.
3. Eliminar casos los cuales presente Monto Abono mayor que Monto Compra.
4. Descartar casos en que Monto Compra es mayor que Monto_CMV (Cuotas*Valor Cuota).
5. Eliminar Casos con compra con tarjeta menores a \$500 (Casos muy atípicos).
6. Descartar casos en los cuales Monto Interés es mayor que cero si el Cuotas es igual a 0⁸.
7. Eliminar casos con valor cuota igual a 0.
8. Eliminar casos que no presenten actividad durante el período en estudio.

Cabe agregar que los intereses fueron calculados de la siguiente manera:

$$\text{Interés} = (\text{Valor Cuota} * \text{Cuotas}) - (\text{Monto Venta} - \text{Monto Abono}).$$

Una vez solucionado los problemas se cruzó cada base de ventas de los holdings entre sí para obtener el monto de interés mensual pagado por cada cliente. Luego esta se cruzó con la base principal creada siguiendo la misma estructura antes descrita. Quedo finalmente una muestra con 98.886 casos.

Respecto a la base anterior, se tomó como supuesto que el interés solo se asociará a la compra respectiva. Por una parte por simplicidad del trabajo y por otra, para no perder el interés pagado por el cliente. Por ejemplo, si el cliente compra a 36 cuotas en uno de los últimos períodos de estudio y se dividiera el interés generado por la compra en ese número de períodos, se perdería gran parte del interés generado por el cliente.

⁷ Este procedimiento fue llevado a cabo con la aprobación del área donde se realiza el trabajo.

⁸ Compras efectuadas a Una Cuota no tienen interés.

7.3 Descripción de Elementos del Sistema Crediticio de la Empresa

7.3.1 Análisis agregado de los clientes

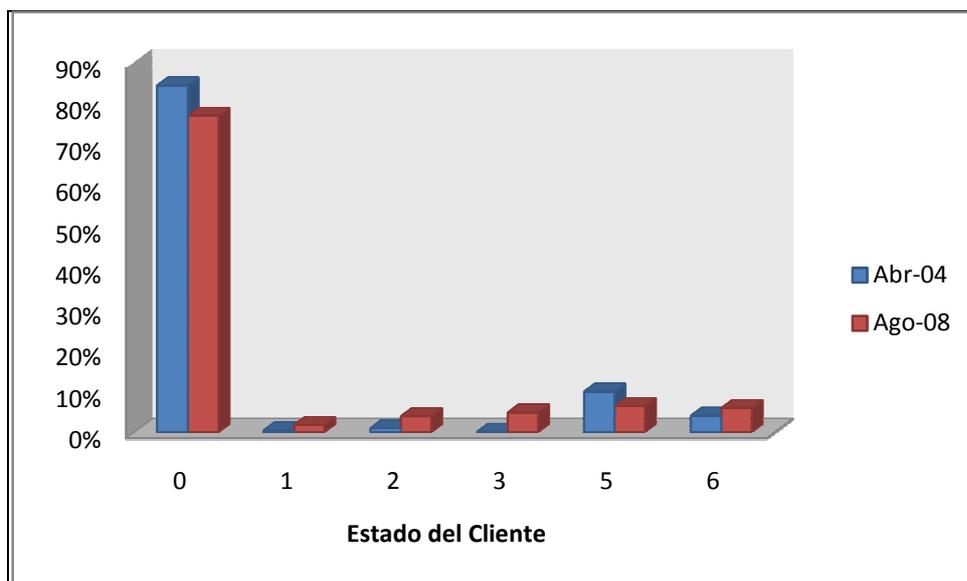
En este punto se describirá cómo evoluciona el sistema actual de los clientes durante el período en estudio. Esto es, su capacidad para comprar o pagar. Para entender esto se define:

Tabla 2: Rango situación actual

Estado del cliente	
0	Sin Restricción
1	Cuenta Nueva
2	Cuenta con Observaciones
3	Cuenta Castigada(Mora mayor a 180 días)
5	Debe Cancelar(Hasta 90 días mora)
6	Pasa a Relaciones Comerciales(90 a 180 días mora)

Luego, en el siguiente gráfico se puede ver la evolución antes mencionada:

Gráfico 4: Evolución situación actual



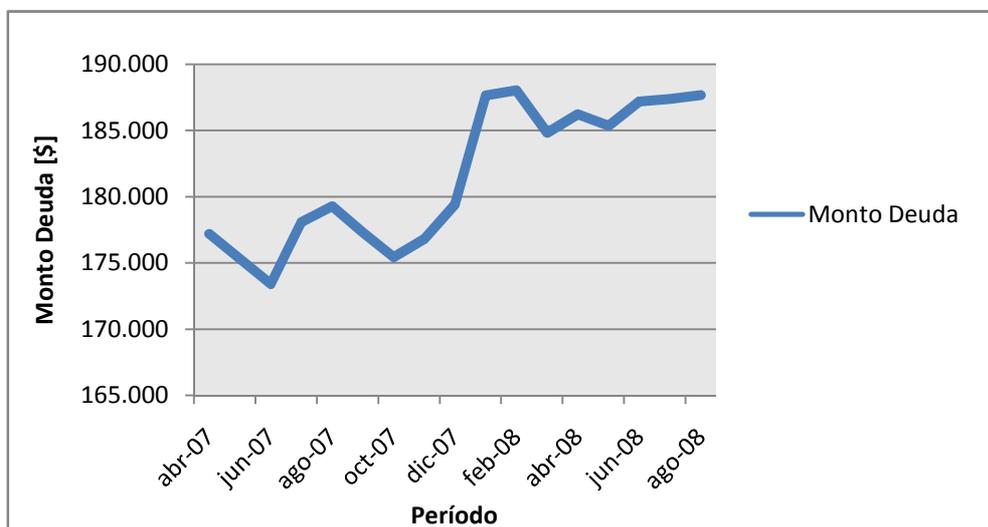
Como se puede apreciar en el gráfico anterior, hay una disminución notoria de clientes que al principio del estudio no tenían restricción. Esto se ve claramente al aumentar “cuentas con observaciones”, “cuentas castigadas” y “pasar a relaciones comerciales”.

Además se puede notar que existe una disminución de clientes que estaban en el estado 5 “Deben Cancelar” esto pues, pasan al estado “Relaciones Comerciales” o “Cuenta Castigada”. Por lo tanto, todo esto implica que la población tiende a empeorar su situación acorde pasa el tiempo, lo cual es perjudicial para la empresa.

Se debe agregar que los clientes que pasan a los estados 3, 5 o 6 quedan identificados por lo general como clientes riesgosos, dado que pasa mucho tiempo antes de que paguen su deuda o en el peor de los casos no lo hagan.

A continuación se describirá la evolución del Monto Deuda promedio a nivel global.

Gráfico 5: Evolución Monto deuda



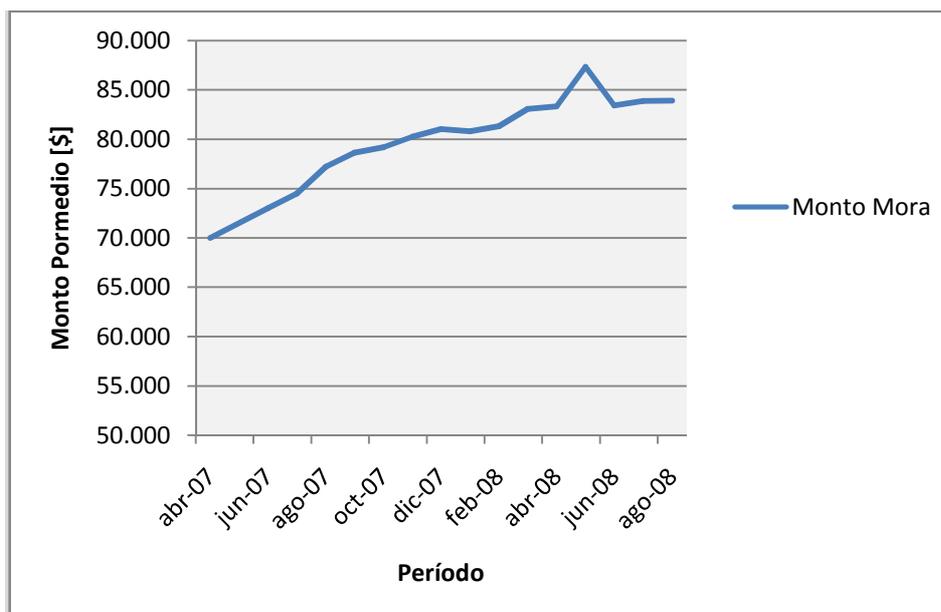
De este gráfico se desprende que a medida que pasa el tiempo la gente se endeuda cada vez más, es decir, utiliza en mayor medida la tarjeta de crédito proporcionada por la firma. A pesar que la variación desde el primer período hacia último (Agosto-2008) no es muy considerable, su tendencia va en aumento lo cual de todas maneras hay que tener en cuenta puesto que una cifra de estas es muy considerable al combinarla con la cantidad de usuarios que tienen y utilizan la tarjeta de crédito de la compañía.

Lo anterior, se puede deber principalmente a la gran cantidad de opciones de uso que se le puede dar a la tarjeta. Por ejemplo, se puede utilizar en farmacias, estaciones de servicio, préstamos, viajes, entre muchas otras.

En complemento con lo anterior, se debe ver cuánto de ese dinero no se paga a tiempo. Puesto que si la gente paga todo lo que debe, significaría que no existe riesgo asociado al prestar el servicio para ya sea comprar productos o pedir avances de dinero, porque todos se considerarían clientes “confiables” que mantienen sus pagos al día.

Como en la realidad se sabe que no pasa esto y que siempre existe gente que por distintos motivos no puede cancelar sus deudas se espera que haya asociado un monto mora, es decir, que no paga en la fecha pactada.

Gráfico 6: Evolución monto mora



Tal como ya se había supuesto, del gráfico anterior se puede observar que no solo la deuda aumenta sino que también el monto mora, o sea, el monto que no ha sido cancelado cuando se debe pasado 30 días. Lo cual demuestra que existe un grupo de personas que no cumplen con los pagos de su deuda pactada que pueden considerarse riesgosas para la compañía, dado que si llegan al punto de no pagar la firma pierde dinero al invertir en este tipo de personas.

Por lo tanto, al contrastar este gráfico con el resultado anterior se puede inferir que la gente no solo está utilizando más la tarjeta, sino que existe a la vez una creciente deuda que no se está cancelando.

Para corroborar lo antes dicho, se analizará en el capítulo 7.3.2 el modelo que ocupa actualmente la compañía para ubicar y distribuir a sus clientes mediante el nivel de riesgo estimado. Esto es una medida creada por la empresa justamente para conocer el perfil de sus clientes, sin embargo por confidencialidad, el funcionamiento de este modelo no fue entregado. Finalmente, se estudiarán los casos por segmentos creados a partir del riesgo impuesto para cada uno.

7.3.2 Riesgo

El modelo de riesgo de la empresa asigna un puntaje a sus clientes llamado “Puntaje Score”. Este puntaje fluctúa entre 0 y 999. Mientras menos puntaje sea asignado, implica que más riesgoso será el cliente.

La base de clientes es dividida principalmente en 4 categorías de riesgo:

- Categoría Score 1: Los clientes dentro de este grupo son los considerados los menos riesgosos. Estos clientes se caracterizan por mantener sus cuentas al día, cumplir con sus fechas de pago y el monto correspondiente. Estos clientes son premiados en general por su conducta con un aumento de cupo.
- Categoría Score 2: Son clientes caracterizados por ser clientes neutros, es decir, cumplen en general con sus pagos sin embargo puede que lleguen a presentar retrasos en sus pagos.
- Categoría Score 3: Clientes considerados de gran riesgo. Estos clientes en general han tenido en alguna ocasión una deuda sin cancelar por mucho tiempo de un monto considerable. Además se da que este tipo de clientes no paga frecuentemente en la fecha que debe pagar.
- Categoría score 4: Esta última categoría condensa a los clientes potencialmente más peligrosos. Estos caen a esta categoría principalmente por dos razones. La primera porque han pasado 9 meses desde su última compra y por otra parte que han pasado 9 meses desde su último pago, siendo esta última la razón predominante.

La distribución de clientes con actividad según su nivel de riesgo⁹ es mostrada a continuación:

Tabla 3: Distribución de clientes con actividad según Rango Score

Scoring	
1	69%
2	19%
3	6%
4	6%

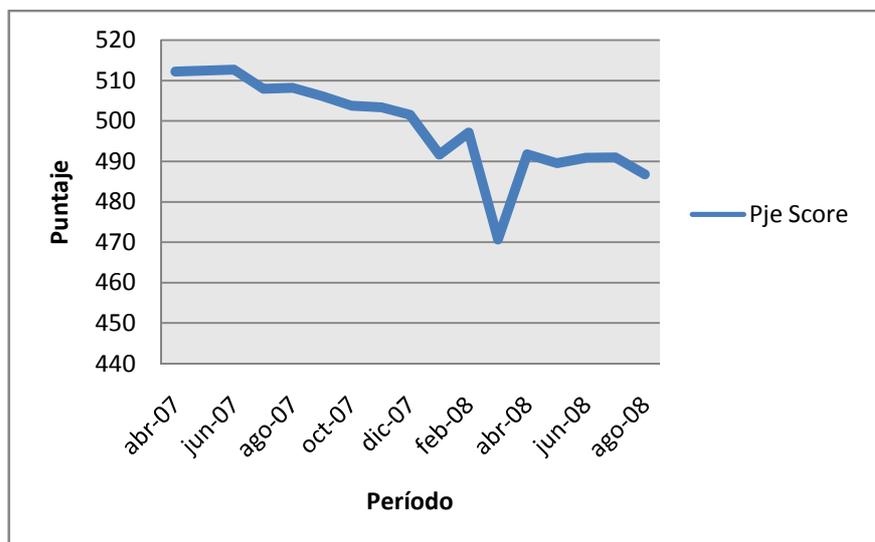
Dentro de esta distribución se ve que claramente existe una predominancia por parte de los clientes menos riesgosos, en el cual un 70% de estos sobre el total global tiene actividad. Por otro lado la gran mayoría de los clientes de rango score 3 y 4 han sido castigados por no pago de sus deudas lo cual no les permite realizar compras y permanecen inactivos. Es reflejado en que tan solo el 12% de los clientes con actividad pertenecen a estos grupos. Finalmente un 19% de los clientes con transacciones durante el período de estudio pertenece al grupo scoring 2.

⁹ Véase Anexo 1 para mayor detalle

Los resultados anteriores muestran que existe relación entre el porcentaje de clientes con actividad con su nivel de riesgo asociado. Esto se debe principalmente como ya se ha mencionado a que la gran mayoría de los clientes clasificados como riesgosos han sido castigados por deuda.

Ahora, ¿cómo evoluciona el puntaje score en el tiempo? Para responder esta pregunta se utiliza el siguiente gráfico:

Gráfico 7: Evolución Puntaje Score Promedio



Como se puede apreciar, a nivel global los clientes se hacen más riesgosos a medida que pasa el tiempo. Esto significa que cada vez existe mayor probabilidad de que los clientes no paguen lo que deben pagar o cuando deben hacerlo. Esto también se puede interpretar que existe un filtro muy bajo para otorgar la tarjeta. A corto plazo puede verse como una manera de captar mayor número de clientes con la oportunidad de utilizar la tarjeta, la que supuestamente aumentaría la rentabilidad de esta, sin embargo a largo plazo esto puede llegar a ser dañino, puesto que si captan mucha gente que al final no cumple con sus pagos la empresa se puede ver afectada.

7.3.2.1 Variables Financieras por Riesgo

En esta sección se analizarán las diferencias para un conjunto de variables financieras respecto al uso de la tarjeta según los segmentos de riesgo antes definidos.

Tabla 4: Descriptivo de variables por nivel de riesgo

Scoring	Monto Cupo	% Uso Cupo	Deuda	Monto Mora	Proporción Mora	Días Mora
4	330.000	60%	180.000	110.000	61,1%	111
3	280.000	95%	270.000	30.000	11,1%	14
2	430.000	75%	320.000	23.000	7,2%	8
1	880.000	30%	225.000	2.000	0,9%	1

Se ve que existe una clara correlación entre nivel de riesgo y monto cupo excluyendo el grupo scoring 4 que tiene un cupo promedio mayor que el segmento 3. Por otro lado el segmento menos riesgoso tiene un cupo promedio de \$880.000 el cual es el mucho mayor que los demás siendo el doble que el grupo 2 y casi el triple que el del segmento 3.

En cuanto a las deudas, se observa que el segmento 2 es el presenta una deuda mayor que el resto siendo esta de \$320.000 promedio, mientras que la menor es del grupo más riesgoso. Ahora bien, el monto mora más alto (monto que se debió haber pagado y aún no se hace) está asociado al grupo más riesgoso, con un monto de \$110.000 siendo esto un 61% de la deuda total y además tiene asociado más de 100 días de atraso o sin pagar. Por otra parte, los clientes menos riesgosos presentan los índices más bajos, mostrando tan solo un 0,9% de deuda atrasada respecto a la total y 1 día de mora promedio. Para el resto de los segmentos a medida que el cliente es considerado más riesgoso, más aumenta su monto mora y días mora. Por lo tanto, en este último análisis sí existe una correlación total, esto es, entre el nivel de riesgo y la morosidad.

El Rango Score, en términos generales se mantiene estable. Por ende los clientes que están clasificados dentro de un grupo scoring se mantienen en el tiempo. Los clientes ubicados en el rango 2 y 3 presentan mayor movimiento de un estado a otro, manteniéndose el 75% de estos en el mismo grupo. Por otra parte el 94% y el 98% de los grupos scoring 1 y 4 se mantienen dentro de su mismo segmento. Esto se puede ver en más detalle en la siguiente tabla:

Tabla 5: Matriz de transición para el riesgo

Matriz Transición Riesgo				
	1	2	3	4
1	94%	4%	0%	2%
2	6%	75%	15%	4%
3	1%	17%	75%	7%
4	0%	1%	1%	98%

Cabe destacar, que los clientes que se mueven o son clasificados distintos de su nivel de riesgo anterior, son movidos en su gran mayoría dentro de su vecindad (segmentos más cercanos) y tan solo la minoría a un grupo más alejado.

7.3.2.2 Interés y Número de Cuotas por Riesgo

En este apartado, se describirá cómo evoluciona el interés para cada nivel de riesgo, para así saber si existen diferencias significativas entre lo que paga cada uno ellos.

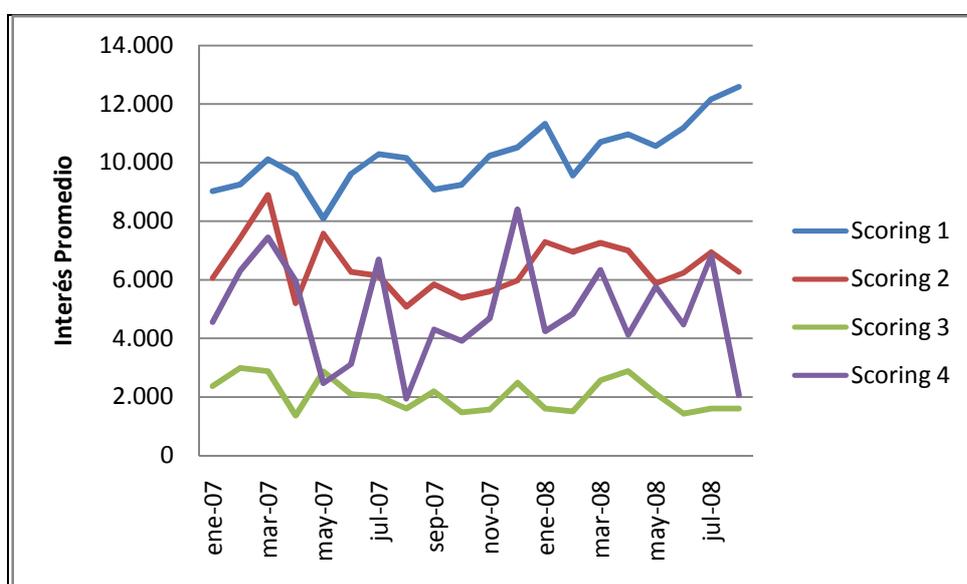
El interés pagado ha tenido un incremento a lo largo del tiempo, lo cual es beneficioso para la empresa. Esto se puede deber a tres razones endógenas y una exógena. Las endógenas vienen dadas por el número de cuotas, puesto que a

mayor número de cuotas, mayor es el interés pagado; por el monto de compra ya que a mayor monto de compra mayor pago de interés y también respecto al holding donde se realizan las transacciones. Estas son analizadas en el capítulo 7.3.3 “Sistema Interés”. Finalmente la razón exógena viene dada por la tasa de interés la cual se ajusta por el valor del mercado.

El interés pagado por segmento scoring será estudiado en dos partes. La primera condicionada a que el cliente tenga actividad con la tarjeta independiente si paga interés o no y la segunda solo transacciones con interés.

A continuación se muestra el resultado de la primera parte:

Gráfico 8: Interés Promedio por Scoring condicionado a que cliente presenta actividad

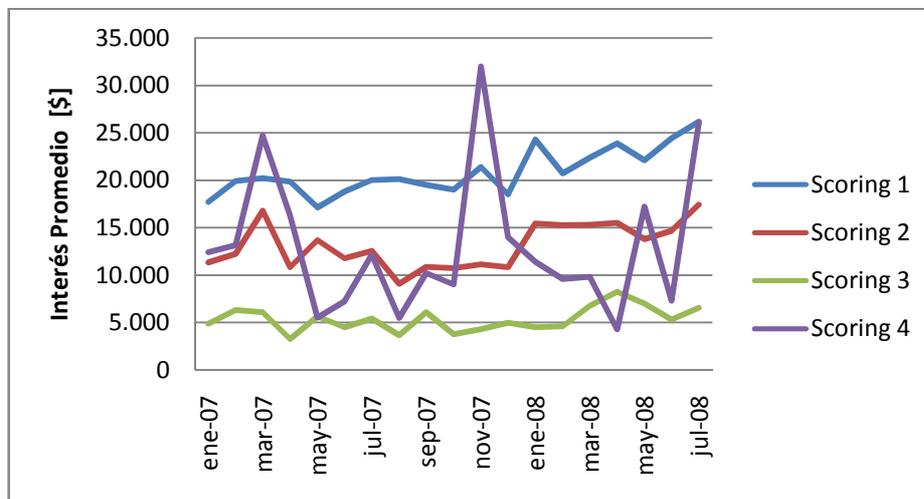


Tal como se puede apreciar del gráfico anterior, el segmento menos riesgoso tiene un monto interés promedio mayor que el resto, alcanzando su valor máximo en el período final sobre los \$12.000. Esto se debe a que el interés promedio ha ido en aumento acorde pasa el tiempo. Por otra parte, el grupo scoring 3 presenta el interés promedio más bajo, esto puede estar dado principalmente a que este grupo tiene menos cupo para comprar y la mayoría de las compras que realizan deben ser a 1 o 2 cuotas precio contado, lo que no tiene interés asociado¹⁰. Además se nota que la tendencia se mantiene al mismo nivel a lo largo del tiempo. El grupo scoring 4 es el que presenta mayor volatilidad asociada. A lo antes dicho, estos clientes presentan muchos meses sin comprar o pagar, los que están aquí deben ser del primer caso puesto que el resto no puede realizar transacciones con la tarjeta y por ende se ve que los peak se dan en fechas especiales como navidad o período escolar. Finalmente el segundo grupo menos riesgoso se ubica en segundo lugar con valores cercanos a los \$7.000 y manteniendo una tendencia aunque con una leve disminución acorde pasa el tiempo.

¹⁰ Más detalle de esto en el capítulo Sistema de Interés

Al analizar el interés promedio tomando en cuenta solo las transacciones con interés, se mantiene el mismo patrón que antes, sin embargo como era de esperar el interés promedio aumenta puesto que se han sacado todas las transacciones que no tenían un interés asociado (véase gráfico 9).

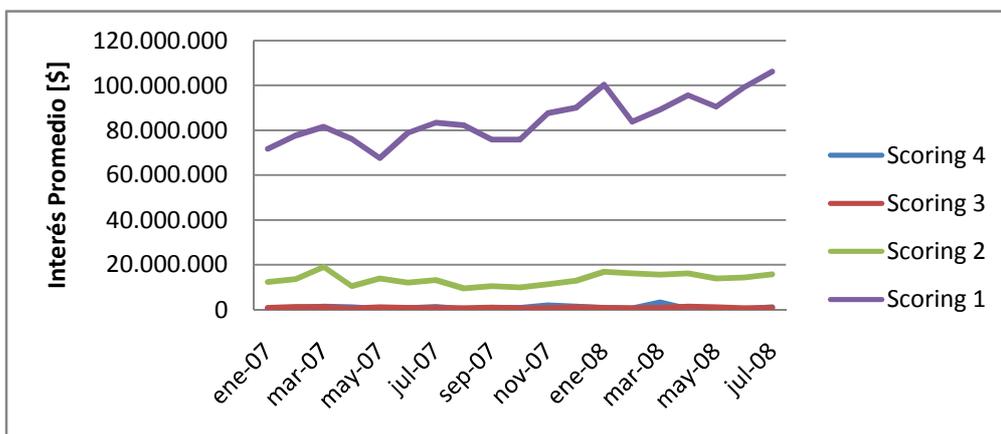
Gráfico 9: Interés promedio por rango score condicionado a transacciones con interés



En este caso se verifican los supuestos anteriores referentes al segmento scoring 4, donde la volatilidad se acrecienta aún más, notándose con mayor claridad las diferencias de pago de interés hechas por este grupo dentro de fechas especiales como navidad y período escolar en comparación de meses regulares, en la cual el interés pagado baja considerablemente. Por otra parte el grupo menos riesgoso presenta un interés promedio de \$20.000 aumentando a medida que pasa el tiempo. Finalmente los grupos 2 y 3 muestran tendencias regulares en el tiempo siendo este último el grupo que paga menos interés en promedio.

Al tomar en cuenta el análisis antes realizado en el cual la gran mayoría de los clientes que presentan actividad pertenecen al grupo de menos riesgo y que este es el grupo que mayor interés promedio paga, se puede ver como es la suma de interés por grupo entregado a la firma en el siguiente gráfico:

Gráfico 10: Interés total por segmento riesgo



7.3.3 Pago de Interés

En este capítulo se analizan los distintos factores que influyen en el monto de interés a pagar al realizar una transacción con la tarjeta de la compañía.

Los clientes al tener la posibilidad de utilizar la tarjeta de crédito tanto en holdings propios de la empresa para servicios como avances o préstamos de dinero y en alianzas de esta (Las cuales incluyen negocios como bencineras o farmacias), pueden utilizar su tarjeta para comprar productos o servicios de variados rangos de precios. Por ejemplo, en alianzas se puede comprar con la tarjeta productos de muy bajos precios y en los holdings propios a productos con costos superiores al millón de pesos. Esto permite que el cliente pueda comprar productos o servicios completamente diferentes de un mes a otro dependiendo las necesidades del momento que tenga este.

Este sistema de endeudamiento permite utilizar la tarjeta tanto como su monto cupo autorizado lo permita, inclusive hay casos especiales que CMR permite ampliar su cupo para comprar cosas puntuales con hasta un 10% extra del cupo original. Este monto cupo tiene un límite de \$5.000.000 y variará dependiendo del cumplimiento del cliente con sus pagos, el nivel de riesgo que le sea asociado y la actividad que tenga respecto al uso de la tarjeta.

7.3.3.1 Holdings

En esta sección se analizarán las diferencias entre un holding y otro según variables financieras y transacciones realizadas.

La siguiente tabla muestra las diferencias de montos de compra, número de compras e interés promedio pagado en cada uno de los holdings asociados a la firma.

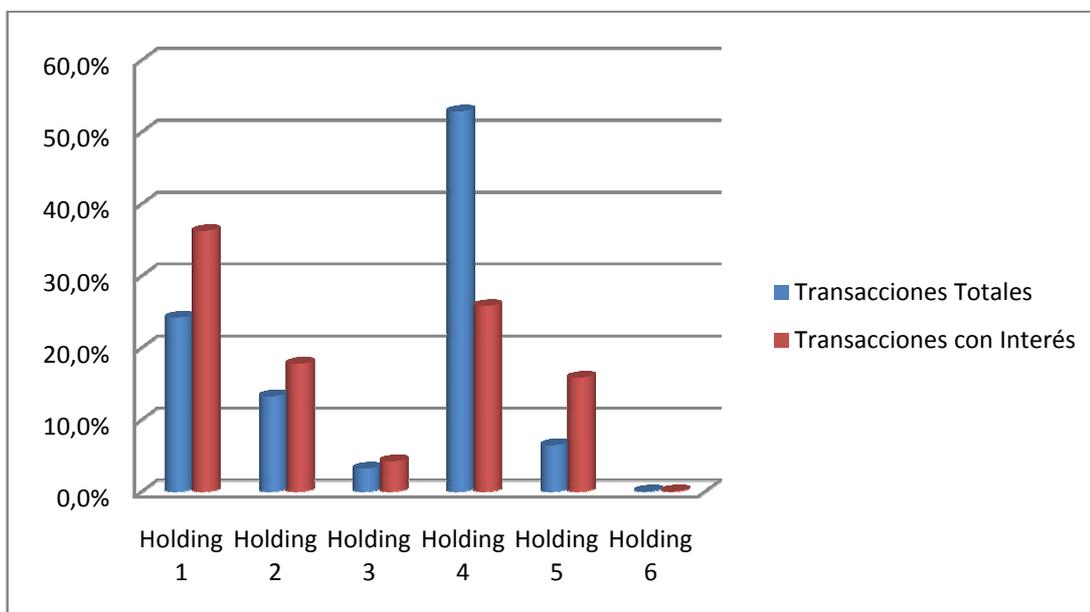
Tabla 6: Descriptivos para variables asociadas a la tarjeta de crédito por Holding

<i>Holding</i>	<i>Interés [\$]</i>	<i>Monto Compra [\$]</i>	<i>Nº Cuotas</i>
Holding 1	5.189	59.104	3,0
(Desv. Estand.)	19.857	106.101	2,7
Holding 2	4.530	66.964	2,8
(Desv. Estand.)	15.168	149.203	2,5
Holding 3	1.701	42.393	2,2
(Desv. Estand.)	4.117	51.993	1,8
Holding 4	390	15.325	1,3
(Desv. Estand.)	2.247	37.744	0,7
Holding 5	44.886	184.171	8,8
(Desv. Estand.)	112.396	351.923	7,7
Holding 6	36.689	481.356	5,8
(Desv. Estand.)	74.794	516.813	6,3

El interés promedio que se paga por transacción en el holding 5 es el más elevado siendo este de aproximadamente \$45.000, en el cual su monto promedio de compra a cancelar con la tarjeta es de \$184.000 y una cuota promedio de 8,8 por transacción. En contra parte el holding 4 tiene asociado un interés promedio de \$390 con un número de cuotas promedio de 1,3 y un monto compra de \$15.000. Esto significa que tanto el número de cuotas como el monto de compra asociado al holding 1 es 7 veces mayor que el holding 4 y por eso su diferencia de interés. Por otro lado el holding 6 tiene el segundo interés asociado más alto con 37.000 mientras que los holdings 1 y 2 presentan un interés promedio entre \$4.000 y \$5.000 lo cual no es una cifra menor con un número de cuotas asociado cercana a 3 cuotas promedio por compra. (véase anexo para ver evolución interés por holding) Ahora bien hay que destacar que el holding 5 tiene 3 veces el número de cuotas y monto compra en comparación a los anteriores y por eso la gran diferencia de interés. Pero ¿Cuántas transacciones son hechas por cada holding?

Como se ve del análisis anterior, el interés promedio varía mucho de un holding a otro como también lo hacen el monto de compra y el número de cuotas. Las alianzas (Holding 4) presentan el interés promedio más bajo al igual que su monto de compra y número de cuotas. Sin embargo, este tipo de prestaciones es el más utilizado en cuanto a transacciones con la tarjeta (ver gráfico N°11). Por otra parte los holdings 5 y 6 son los que presentan mayores intereses asociados pero el número de transacciones hechas en estos holdings son las más bajas en conjunto con las del holding 3, especialmente el holding 6 que sus transacciones son prácticamente el 0% del total.

Gráfico 11: Transacciones por Holding

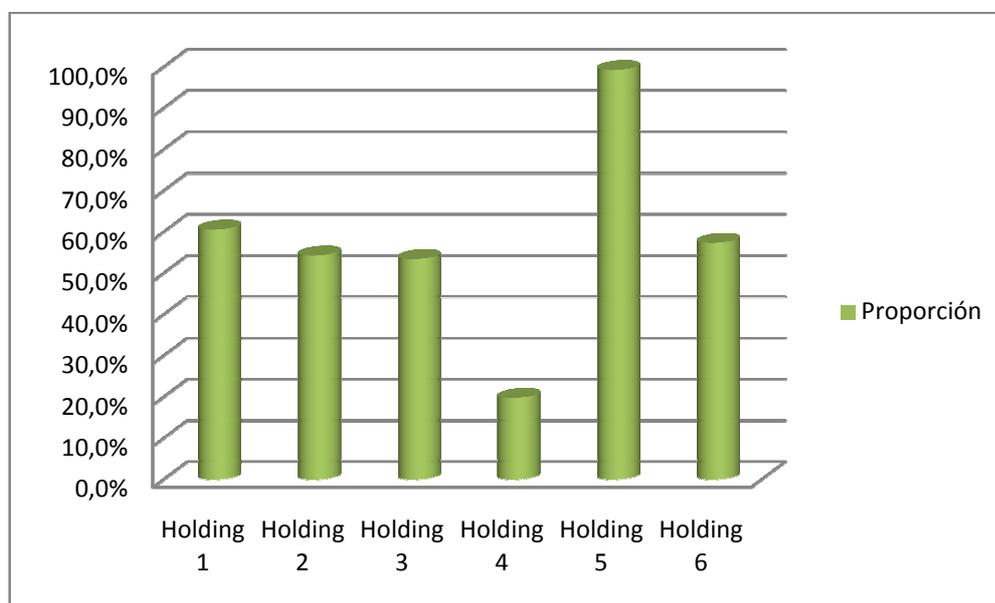


A nivel global, el 40% de las transacciones totales realizadas con la tarjeta tienen un interés asociado. Tal como se había explicitado, las alianzas presentan el mayor número de transacciones totales realizadas con la tarjeta de crédito, llegando a ser esta el 55% del total y el holding 1 con un 23 en el segundo lugar. Sin embargo el

holding 4 solo representa el 25% de las transacciones con interés, mientras que el holding 1 presenta el mayor porcentaje de transacciones con interés, siendo este el 38% del total. Por otro lado el porcentaje de las transacciones con interés que se realizan en el holding 2 y 5 ("Préstamo de dinero") respecto al total, son cercanas al 20% cada una. Cabe recordar que este último está entre los 3 holdings con menos transacciones con la tarjeta pero con el interés promedio asociado más alto por compra.

Los resultados expuestos eran respecto al total de transacciones con interés y sin interés a nivel global, es decir, con la suma de todas las transacciones realizadas en cada uno de los holdings. Ahora la pregunta es ¿Cuál es la proporción de transacciones que se realizan con interés en cada uno de los holdings?

Gráfico 12: Proporción de transacciones con interés por Holding

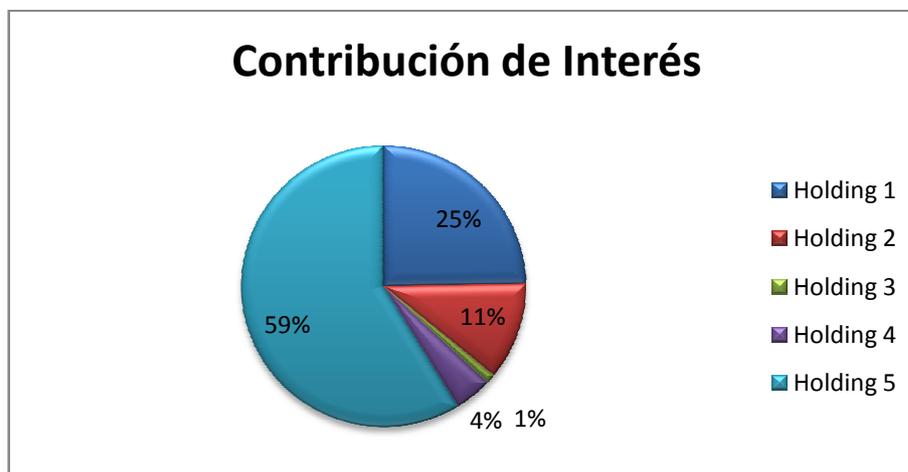


Es interesante ver que el 99% de las transacciones realizadas en el Holding 5 son transacciones con interés. Esto es un indicador muy fuerte si se tiene en cuenta que este holding entrega en promedio \$44.886 de interés por transacción. Ahora para el caso del Holding 4 el cual tiene la mayor cantidad de transacciones con la tarjeta, dentro de las transacciones hechas en este solo el 19% son con interés, siendo el más bajo de todos y cabe recordar que además entrega el menor valor de interés por transacción (\$390). En cuanto a los demás holdings, estos se presentan de manera más pareja ubicándose todo ellos entre un 50 y 60% de transacciones con interés respecto del total para cada uno de ellos. Sin embargo el Holding 1 es el que presenta el mayor número de transacciones a nivel global con un monto de interés promedio de \$5.189, un monto considerable para la cantidad de transacciones que se realiza con interés dentro de este y lo mismo para el holding 2 que tiene un interés de \$4.500 promedio por transacción. Para el caso del holding 6, si bien más del 50% de sus transacciones son con interés y presenta un elevado monto interés

por transacción, el número de estas es prácticamente nula en comparación con los otros holdings y por ende su contribución a la firma.

Dado el análisis de los resultados anteriores, de ante mano ya se sabe que el interés proporcionado por cada uno de estos holdings a la firma es completamente diferente el uno del otro, lo cual se puede ver a continuación:

Gráfico 13: Contribución de interés



El holding que presenta las contribuciones más significativas para la firma son los entregados por el Holding N°5, el cual según los datos anteriores solo representaba el 6,7% de las transacciones totales y un 16,6% de las transacciones con interés a nivel global, pero que dentro de las transacciones hechas en el holding, el 99% eran con interés. Por lo tanto el interés no dependerá solo de la cantidad de transacciones ni del monto de la compra sino que una combinación entre las dos antes mencionadas en conjunto con el número de cuotas con que se realizan las compras y del holding en el cual se hace la transacción (véase Tabla N°7).

Cabe notar que el Holding “Alianza” que presentaba el mayor número de transacciones totales superando el 50%, solo contribuye a la firma con un 5% del interés total, lo cual es deducible puesto que su interés promedio no supera los \$390 y su porcentaje de transacciones con interés es solo del 27% del total.

Por otra parte el Holding N°1, es el segundo mayor contribuidor de interés de la compañía siendo este el 26% del total, esto se a que es el que presenta el mayor número de transacciones con interés respecto del total y un interés asociado promedio dado por \$5.190 y un número de cuotas sobre la mayoría de las demás holdings de la empresa. Hay que destacar que entre el holding N°1 y N°5 se encuentra acumulado el 75% del interés total entregado a la firma. Además el Holding 2 a pesar de estar en 3ra posición respecto a las transacciones con interés a nivel global, entrega el 9% de contribución de interés a la firma, el tercero con mayor interés con \$4.500 de monto promedio por transacción realizada con interés. Por lo tanto se sospecha que los clientes que presenten mayor actividad dentro de estos tres holdings serán los clientes más valiosos según lo ya definido.

7.3.3.2 Sistema de Cuotas

El sistema de la tarjeta de crédito permite comprar en cuotas, las cuales varían desde 1 cuota hasta 36 cuotas en el siguiente rango: 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 15, 24, 36.

Cada una de estas tiene una tasa de interés diferente. Si el cliente paga a 1 cuota o 2 cuotas precio contado no pagará interés. Para el resto sí, las cuales dependen del número de cuotas (a mayor número de cuotas, mayor es el interés asociado a la compra), del tipo de producto o servicio que se quiere obtener (Avance, Electrónica, Vestuario, etc.) y del mercado (IPC). Por esto la tasa en sí es volátil y puede variar con el tiempo. A medida que el cliente compra con un mayor número de cuotas, mayor será el interés a pagar. De hecho existe una amplia diferencia entre el interés generado al pagar con 3 cuotas y 36 cuotas. Este último genera el mayor interés que se podría alcanzar en una compra, lo que queda sujeto al monto de la compra en sí, realizada en la transacción. Sin embargo los clientes que compran a 36 o 24 cuotas, esperan un intervalo de tiempo considerable para volver a realizar otra. Esto se puede deber a varias situaciones. Una de ellas es que su monto cupo no se lo permite, otra razón viene dado por que su valor cuota a cancelar tiene un valor alto y esta debe ser pagada por 3 años. Y finalmente porque solo utiliza la tarjeta para realizar compras de productos puntuales de amplio costo y ya no desea volver a utilizar la tarjeta hasta alguna nueva oferta, es decir, sería un cliente esporádico. El número de cuotas con el cual se realiza la compra puede variar desde \$0 (Caso 1 y 2 cuotas precio contado) hasta una cifra superior al millón de pesos, el cual puede ser dado por una compra de alta envergadura asociada a un número de cuotas mayor que 24.

Todo lo anterior demuestra que existe una alta volatilidad asociada al interés generado por una compra, puesto que no solo depende del monto de compra sino que además en que holding se realiza la transacción, que tipo de producto desea obtener y finalmente a cuantas cuotas desea pagar. Lo que es incierto entre un período y otro.

Por esto se pueden encontrar clientes que un período pueden comprar artículos de gran costo a un número alto de cuotas lo que genera un interés elevado y otros períodos en que no pagará nada con interés o un monto muy pequeño. Otros clientes por ejemplo pueden ser activos durante todo el período en estudio con compras de monto tanto grandes como pequeñas pero que no pagarán nunca interés.

Al mismo número de cuotas, varía mucho el monto compra. Y para el mismo monto compra puede variar mucho el n^o cuotas. La correlación entre un mes y otro en la mayoría de las variables es positiva pero muy baja (Monto Compra, Interés, Cuotas). Esto se ve reflejado para el caso del interés el cual presenta una correlación de 0,061 entre el interés pagado en el período t con el período siguiente, lo que indica que prácticamente no existe relación en cuanto al pago de interés de un período y otro.

Algunos ejemplos de lo antes mencionado se pueden ver a continuación, en el cual se muestran dos clientes A y B obtenidos aleatoriamente dentro del grupo de los clientes más valiosos. En ambos casos el interés varía considerablemente de un período a otro, llegando a variar hasta en un monto mayor a \$400.000 para el cliente A con una desviación estándar de \$147.000 y de \$210.063 para el cliente B.

Gráfico 14: Ejemplo Cliente A

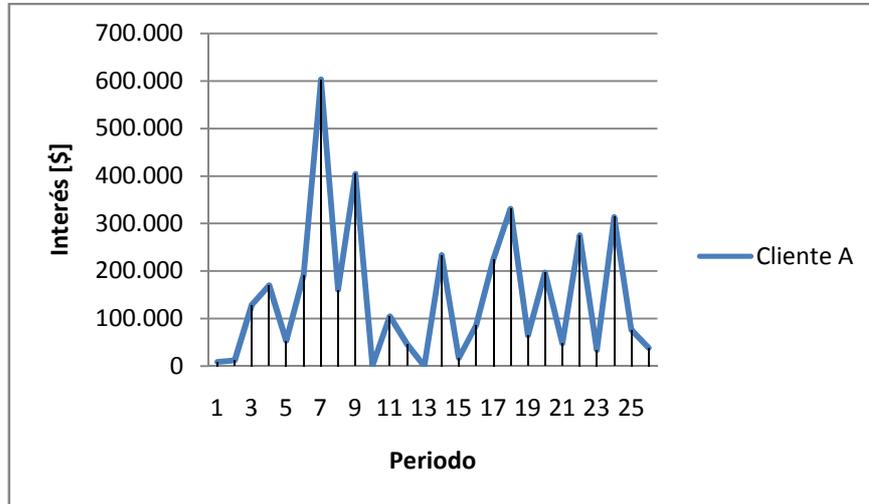
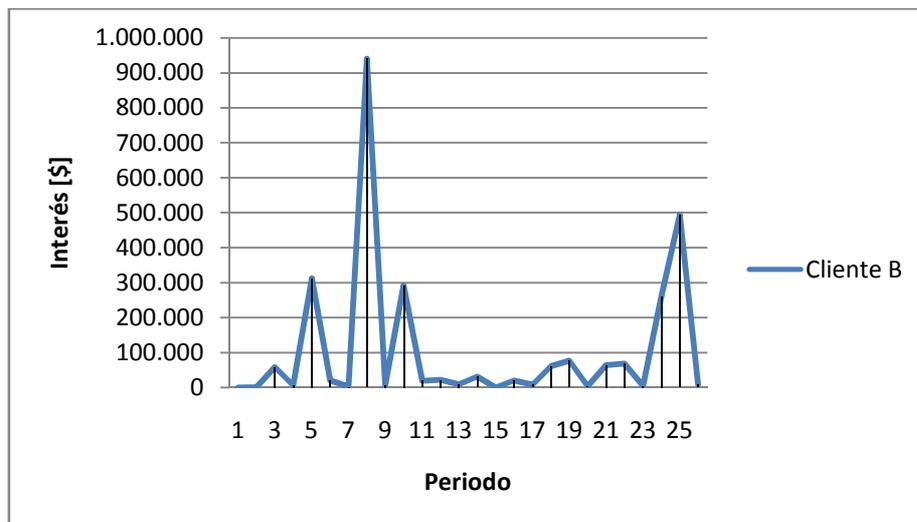


Gráfico 15: Ejemplo Cliente B



Ahora, dado que no se puede calcular el monto de interés exacto o cercano que el cliente pagará en el futuro por lo explicitado en este capítulo. ¿Se puede distinguir entre buenos y malos clientes?

7.4 LTV

7.4.1 Lifetime Value Histórico

Con el fin de conocer las diferencias existentes entre los clientes más valiosos y los menos valiosos se obtuvo el Lifetime Value (Gupta, Lehmann y Stuart, 2006) de hace dos años (2006) utilizando 24 meses de datos históricos de la empresa, es decir, reconocer cuales eran los clientes más valiosos para la compañía en Julio del 2006. Una vez obtenido el LTV de cada cliente, estos fueron ordenados acorde a su valor obtenido. Luego para realizar el análisis fueron divididos en 4 segmentos proporcionales con un 25% de la población cada uno. En el cual el segmento "Ranking1" contiene a los más valiosos, mientras que el segmento "Ranking 4" los de menos valor.

Tabla 7: Descriptivos de variables por segmentos de valor

	Ranking 1	Ranking 2	Ranking 3	Ranking 4
Monto Compra	252.479	118.361	80.770	65.271
Nº CUOTAS	8,2	5,0	4,0	3,0
Scoring (Riesgo)	1,2	1,3	1,5	1,6
Edad	47	45	44	44
SEXO	0,55	0,50	0,48	0,47
Años Cuenta	10,5	9,0	7,8	8,2
Monto Cupo	1.038.015	689.281	531.863	612.537
Contribución Interés	80%	15%	4%	1%
Transacciones Holding 1	7,2	6,2	4,4	2,1
Transacciones Holding 2	3,0	2,2	1,5	0,6
Transacciones Holding 3	0,9	0,8	0,5	0,1
Transacciones Holding 4	5,3	4,4	3,0	1,4
Transacciones Holding 5	4,7	2,3	1,2	0,4
Transacciones Holding 6	0,03	0,01	0,00	0,00

De la tabla anterior se puede ver que el segmento más valioso representa para la compañía el 78% de los ingresos dados por el interés, seguido por el segmento 2 el cual contribuye con el 16% del total. Por lo tanto los primeros dos segmentos más valiosos entregan en total el 94% del interés lo cual indica que tan solo estos dos segmentos tienen prácticamente el valor total dado a la compañía lo cual demuestra una notoria diferencia entre un grupo de compradores y otro en cuanto al uso de la tarjeta.

El monto de compra promedio y el Número de Cuotas promedio del segmento más valioso es claramente mayor a los demás, siendo estos 4 y 3 veces mayores que los entregados por el segmento 4 respectivamente. En cuanto a la edad y los años

transcurridos desde que obtuvo su tarjeta, no existe una gran diferencia entre grupos pero hay un leve aumento de estas por parte del segmento más valioso.

Se observa del monto cupo entregado que los clientes del Ranking 1 tienen un cupo mayor que los demás sobre el millón de pesos mientras que los otros no superan los 700.000 de cupo promedio, lo que los posibilita a tener una mayor opción de endeudamiento y por consiguiente de mayor pago de interés.

Al analizar la cantidad de compras hechas con intereses en cada uno de los holdings en los cuales se puede utilizar la tarjeta de crédito, se puede apreciar que el segmento más valioso presenta mayor número de compras para cada uno de los holdings mientras que el segmento menos valioso tiene los menores niveles de compra con interés asociados. En mayor detalle el número de compras hechas por el segmento Ranking 1 en el Holding N°5 es 11 veces mayor que las realizadas por el segmento 4 y 4 veces mayor que las del segmento 3. Recordando del capítulo anterior, este holding entrega en promedio \$45.000 por transacción, lo cual se refleja en cuanto a las contribuciones de interés dadas por cada segmento. Esto mismo ocurre para todos los demás holdings, en especial el N°1 y el N° 2, los cuales son los siguientes holdings que más interés entregan a la firma luego del holding N°5. En resumen, los segmentos más valiosos realizan compras de mayor monto que las demás, compran a mayor número de cuotas, realizan más transacciones en los distintos holdings y tienen un mayor cupo para realizar compras, lo que se traduce en su contribución de interés casi total para la firma.

Tabla 8: Valor promedio anual de clientes por segmento

Valor	Ranking 1	Ranking 2	Ranking 3	Ranking 4
Promedio	\$ 256.039	\$ 41.661	\$ 13.050	\$ 1.646
Desv std.	\$ 31.951	\$ 6.440	\$ 1.890	\$ 45

Lo antes descrito se traduce a los valores de la tabla 8, donde un cliente perteneciente al segmento de mayor valor vale \$256.000. Un cliente de este grupo vale más de 100 veces lo que vale un cliente del segmento ranking 4 y más de 20 veces lo que vale un cliente de ranking 3, lo cual demuestra una vez más la diferencia existente entre un cliente de un grupo y otro. Finalmente un cliente del segundo grupo más valioso vale en promedio \$41.600, tres veces mayor que el segmento 3 y 25 veces el valor de un cliente del segmento de menos valor.

Una vez conocidos los clientes más valiosos y las diferencias entre los 4 grupos segmentados proporcionalmente por su LTV, la pregunta ahora es como se distribuyen estos segmentos explicitados anteriormente por el modelo riesgo. A continuación se responderá esa pregunta.

7.4.1.1 Riesgo por LTV

Al cruzar los resultados del modelo de riesgo con los segmentos antes descritos se obtienen los siguientes resultados: los clientes pertenecientes al grupo Scoring 1, es decir, los clientes menos riesgosos están repartidos de una manera proporcional entre los segmentos de LTV con un leve predominio para los grupos de LTV más valiosos. Para los clientes de Scoring 2, este se encuentra predominantemente en los segmentos Ranking 2 y 3, sin embargo la proporción no es muy superior respecto de los otros grupos. En cuanto a los dos grupos más riesgosos, se puede ver que estos pertenecen en gran medida a los dos grupos menos valiosos, es decir, los segmentos Ranking 3 y 4 con el 80% y el 88% respectivamente.

Tabla 9: Segmentos de valor v/s segmento riesgo

	Scoring 1	Scoring 2	Scoring 3	Scoring 4
Ranking 1	28,5%	20,7%	5%	3,6%
Ranking 2	25,4%	28,5%	15,3%	8,6%
Ranking 3	22,4%	32,1%	37,9%	23,2%
Ranking 4	23,7%	18,7%	41,8%	64,6%

En conclusión, los clientes más riesgosos se encuentran en mayor medida en los grupos menos valiosos de LTV, lo que implica que existe una fuerte correlación entre estos. Por otra parte los clientes de Scoring 1 y 2 se encuentran distribuidos de forma no tan desigual entre los segmentos de LTV, por lo tanto no se puede hablar de algún tipo de correlación esperada en estos casos.

7.4.1.2 Dinámica del Valor del Cliente

Ya se ha descrito a los grupos por Lifetime Value y por riesgo, ahora la pregunta a responder es: ¿Qué tan dinámico es un cliente? Con esto se pretende responder si un cliente de un período a otro continúa siendo del mismo segmento, sigue siendo valioso o cambia a lo largo del tiempo. Para hacer esto se calculará el Lifetime Value con los datos históricos de la misma manera que antes pero en esta ocasión serán calculados con intervalos de tiempo de 12, 8 y 4 meses¹¹.

- 12 Meses

Tabla 10: Dinamismo LTV para 12 meses

	Matriz Transición LTV 12			
	Ranking 1	Ranking 2	Ranking 3	Ranking 4
Ranking 1	58%	23%	10%	10%
Ranking 2	24%	38%	23%	15%
Ranking 3	10%	25%	36%	29%
Ranking 4	9%	14%	31%	47%

¹¹ Véase Anexo para otros períodos.

En este caso el LTV fue calculado en 2 períodos de 12 meses cada uno y se compararon entre ellos para conocer el dinamismo de los clientes. Tal como indica la tabla N°9, el 58% de los clientes que pertenecían al segmento de los clientes más valiosos Ranking 1 se mantienen en el mismo rango durante el siguiente período, mientras que el 23% de estos pasa al segmento 2. Por lo tanto aproximadamente el 81% de los clientes que en el primer período pertenecían al segmento más valioso se mantiene dentro de los clientes más rentables para la empresa. Para los segmentos Ranking 2 y 3, el mayor porcentaje de ellos se mantiene en ese segmento y otro gran porcentaje dentro de la vecindad, siendo los mínimos los que pasan a un segmento más alejado de la vecindad. Finalmente los pertenecientes al grupo Ranking 4, el 78% se mantiene dentro de los 2 segmentos menos valiosos. En resumen, a pesar de que las diagonales no son del 100%, es decir, los clientes no están siempre fijos en el mismo grupo de un período a otro, si la gran mayoría se mantiene ahí o en la vecindad.

Por lo tanto, predecir el monto de interés exacto de un cliente no se puede dado los distintos factores antes explicado que producen gran volatilidad entre el interés pagado por cada ocasión, si es más probable estimar en que grupo pertenecerá el cliente.

- 8 Meses

Tabla 11: Dinamismo LTV 8

Matriz Transición LTV 8				
	Ranking 1	Ranking 2	Ranking 3	Ranking 4
Ranking 1	55%	23%	11%	11%
Ranking 2	24%	37%	24%	16%
Ranking 3	11%	25%	37%	27%
Ranking 4	10%	16%	28%	46%

Para el cálculo de LTV en períodos de 8 meses se ve un mayor cambio entre los clientes de un segmento a otro en comparación con el caso anterior, sin embargo, este cambio es pequeño. Este cambio producto de la disminución de meses (De 12 a 8), representa que la gente no siempre se mantiene comprando con intereses constantemente y que hay clientes que solo en algunos meses compra con un monto de interés alto.

En conclusión existe un grupo de valiosos clientes que se mantiene comprando constantemente con intereses lo cual le permite mantenerse siempre en el mismo segmento más valioso, mientras tanto hay otros clientes valiosos que no lo hacen tan esporádicamente, pero sin embargo cuando lo hacen, realizan una compra con un monto alto de interés asociado. Además la gran mayoría de los clientes se mantiene dentro de su segmento, y otra gran mayoría cerca de su vecindad, es decir, dentro de los segmentos más cercanos. Y la minoría es la que pasa hacia un segmento más alejado de su vecindad.

Por lo tanto, si se puede estimar el valor del cliente con el objetivo de distinguir entre los “buenos” y “malos” clientes, lo cual es una de los objetivos de esta memoria.

7.4.1.3 Caracterización Muestra y Variables

En este capítulo se describe la muestra de clientes utilizada, que es necesaria para poder obtener los valores en los modelos, que servirán finalmente para proyectar los mejores resultados a la población total, es decir, a todos los usuarios de tarjeta de crédito de la empresa. Los clientes, tienen como condición tener alguna actividad durante el período de estudio. Las variables a caracterizar son sexo, años cuenta, edad y monto cupo, es decir, variables tanto demográficas como variables propias de la tarjeta de la empresa.

La muestra utilizada está compuesta por una muestra aleatoria de 13.250 clientes que tienen como condición tener actividad durante el período en estudio. A continuación se describen las distintas variables.

Sexo

Tabla 12: Distribución por sexo

Sexo	Frecuencia	Porcentaje
Mujeres	195244	49,8%
Hombres	197169	50,2%
Total	392413	100%

Se puede observar de la tabla que la muestra está dividida proporcionalmente entre hombres y mujeres. Al estudiar la relación entre el interés y el sexo (véase gráfico 17), es interesante destacar que existe un interés promedio mayor por parte de los hombres (1) con una diferencia de aproximadamente \$1.000 en comparación a las mujeres.

Años Cuenta

Tabla 13: Detalle años cuenta

Años Cuenta	
Mínimo	1
Máximo	27
Media	8,15
Mediana	6
Moda	3
Desv. Estándar	6,06

Gráfico 16: Interés promedio por sexo

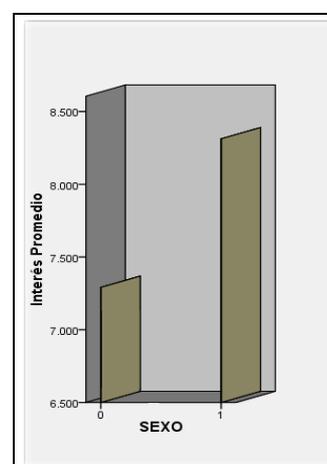
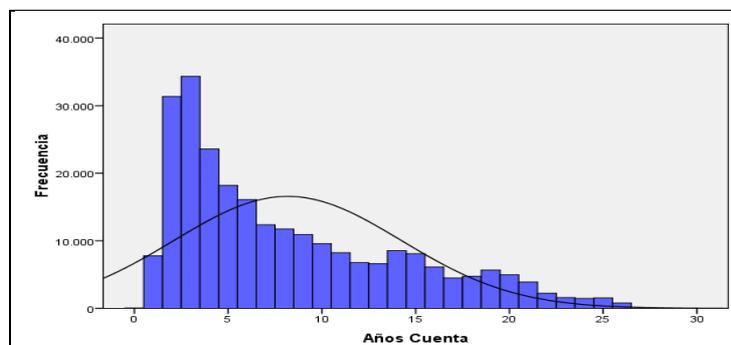
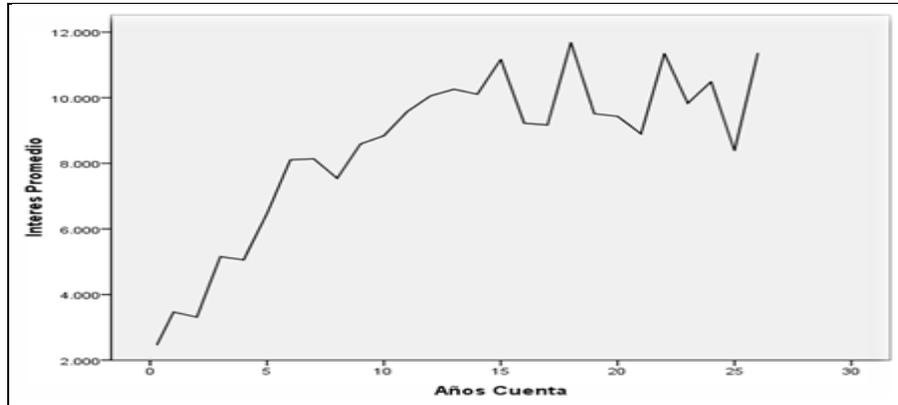


Gráfico 17: Distribución años cuenta



La cantidad de años que un cliente tiene la tarjeta está dominado principalmente por un período de corto tiempo (gráfico 18) en que la moda tiene un valor de 3 años y la mediana de 6. La media por otra parte está es de 8 años en el que el máximo son 27 y el mínimo 1. En el siguiente gráfico se analiza la relación de esta variable con el interés.

Gráfico 18: Interés promedio según años de la cuenta



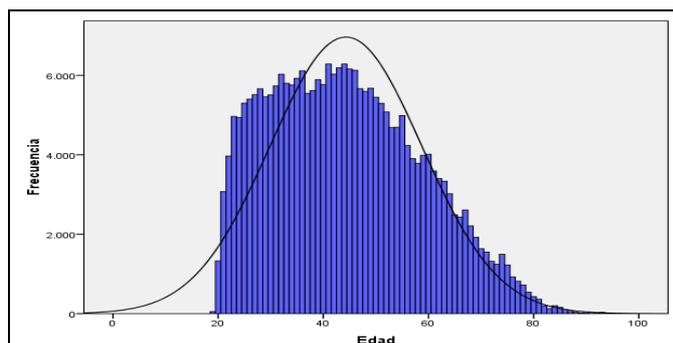
Como se puede apreciar en el gráfico, dentro de los primeros años con la tarjeta se puede ver que a medida que aumenta la cantidad de años que se tiene la tarjeta de la compañía, mayor es el interés promedio asociado. En que ya a partir de los 15 años existe mayor turbulencia o volatilidad y los valores asociados ya no aumentan significativamente.

Edad

Tabla 14: Detalle edad

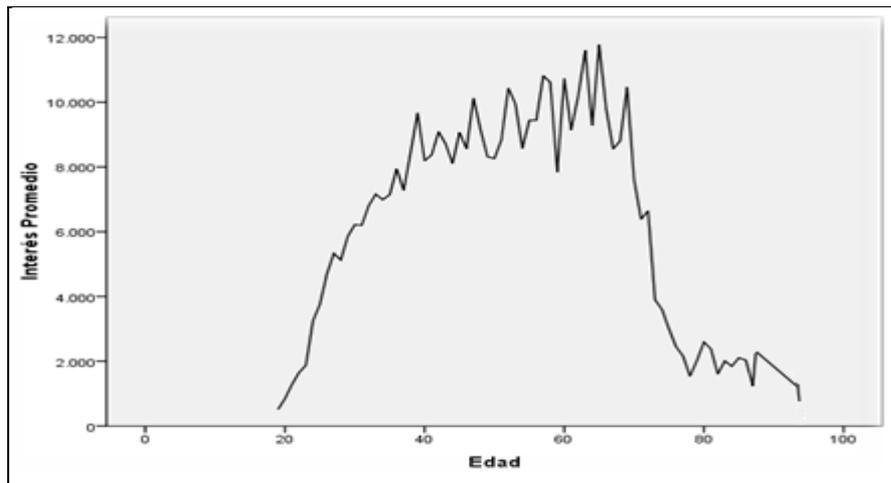
Edad	
Mínimo	19
Máximo	94
Media	44,34
Mediana	43
Moda	44
Desv. Estándar	14,36

Gráfico 19: Distribución edad



Al observar el gráfico anterior, se observa que la muestra está bien distribuida respecto a la edad, en la cual la media, la mediana y la moda son todas muy similares, lo que queda claro al ver cierta distribución en “campana”. La mínima edad está dada por 19 años, puesto que no se permite sacar la tarjeta estrictamente bajo los 1 años, el máximo está dado por 94 años. Cabe destacar cierta disminución ya al partir de los 50 años reduciéndose los clientes con tarjeta que presenten actividad para esta edad.

Gráfico 20: Interés promedio según edad



Al analizar la relación existente entre el interés y esta variable, se observa que a medida que el cliente “envejece” mayor interés va pagando. Además, el máximo interés promedio alcanzado se obtiene dentro del rango de edad de 40 a 70 años, es decir, ya gente adulta con niveles de ingreso más alto y consolidado. Sin embargo, se ve que existe una tendencia fuerte a la disminución del pago de interés al pasar esta edad.

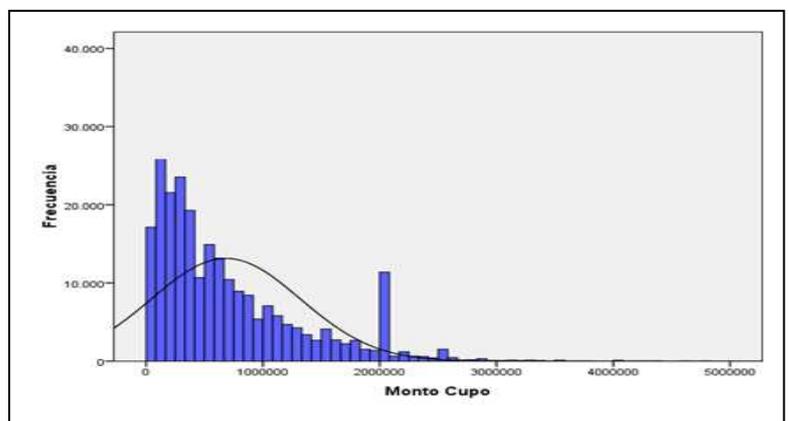
Monto Cupo

Dentro de los datos se ve que la gran mayoría de los clientes tiene un monto cupo menor al \$1.000.000, en la cual la media es de \$682.000 y la mediana de 460.000. A pesar de esto la moda viene dada por un valor de \$2.000.000, los cuales deben ser grupo compuesto de clientes poco riesgosos según el modelo de riesgo de la firma. El mínimo viene dado por el valor más bajo que se le puede otorgar a alguien al sacar la tarjeta que es de un monto de \$80.000, mientras que el máximo el de \$5.000.0000, el cual es el tope máximo de cupo entregado a un cliente.

Tabla 15: Detalle monto cupo

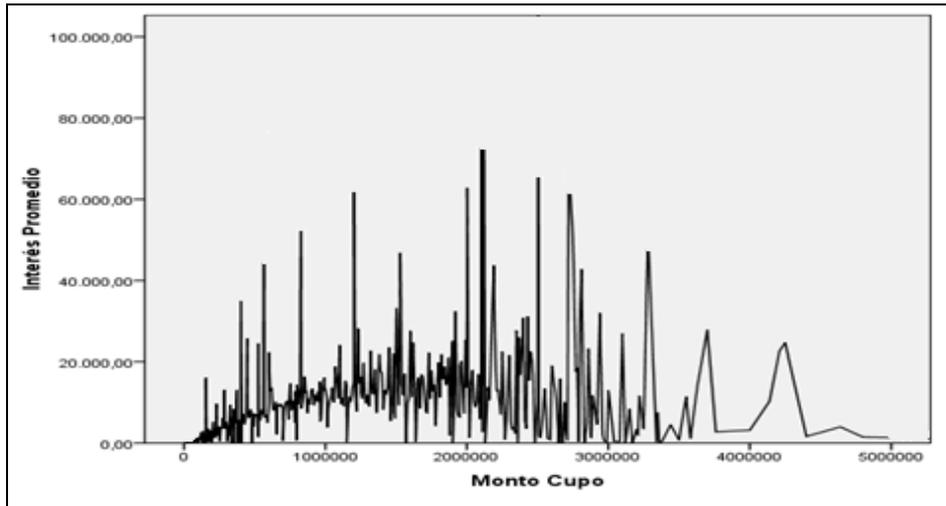
Monto Cupo	
Mínimo	80.000
Máximo	5.000.000
Media	682.491
Mediana	460.000
Moda	2.000.000
Desv. Estándar	636.540

Gráfico 21: Distribución monto cupo



Ahora al analizar la relación dada por el interés y monto cupo se observa lo siguiente:

Gráfico 22: Interés promedio según monto cupo



Existe gran volatilidad asociada al monto cupo, sin embargo existe una tendencia a pagar mayor interés a medida que aumenta el cupo, disminuyendo este ya aproximadamente al llegar a un monto cupo de \$2.000.000, en el cual ya cercano a los \$3.000.000 se pierde la tendencia y no existe algún patrón determinado para los cupos sobre este valor.

7.4.2 Estimación Monto Interés y LTV

Para obtener el Lifetime Value con respecto al interés que pagará el cliente en el futuro dada la utilización de la tarjeta de crédito y así identificar a los “buenos” y “malos” clientes, se necesita estimar el valor del monto “interés” en el período $t+1$, en función de un conjunto de variables independientes medidas en el período t .

Para estimar el valor del interés, se utilizaron 4 tipos de modelos regresivos: Modelos Lineales, Modelos Log Normales, Modelos de Poisson y por último Modelos Binomiales Negativos. Cada uno de ellos fue estudiado por una parte solo con variables RFM y por otra con variables demográficas y variables referentes a la tarjeta como por ejemplo años cuenta, número promedio de cuotas, entre otras. Además en cada caso fueron insertadas dummies de mes con el propósito de encontrar estacionalidades en el tiempo. Luego de obtener estos valores se calculó el LTV con la fórmula planteada en el marco conceptual (Fórmula 2).

Se cuenta con 19 meses de datos, para lo cual se ha decidido que el período de calibración costará de 12 meses para así contar con todos los meses y poder ver el efecto estacionalidad de estos. Por otra parte el período de validación constará de los siguientes 7 meses.

Los resultados entregados y analizados respecto al LTV obtenido por cada modelo serán los siguientes:

Correlación de Spearman¹² a nivel Individual. Este tipo de correlación permite conocer que tan bien predicen los rankings los modelos (Greene, Edición 1999). En este caso los clientes más valiosos están dentro de los primeros lugares del ranking y los menos valiosos se ubican en los últimos lugares (para mayor información, véase anexo 4 con tabla de ranking).

Correlación de Spearman a nivel de Segmento. Los grupos serán ordenados de menor a mayor dado su valorización. Luego serán divididos en cuatro grupos proporcionales y así obtener a los clientes más valiosos y los menos valiosos. Se volverá a utilizar la correlación de Spearman para conocer la capacidad predictiva de los modelos respecto de los rankings a nivel de segmento.

Matriz de Confusión. Finalmente se creará una matriz de confusión que permitirá conocer el porcentaje de acierto del modelo para cada segmento creado anteriormente, es decir, cuantos clientes en porcentaje fueron colocados dentro de su segmento real y cuantos fueron colocados dentro de otros.

A continuación se detallará el estudio de cada uno de ellos en los que se analizarán sus resultados en tres fases para así conocer el error asociado al pasar de una fase a otra.

¹² Revítese anexo 4 para mayor información

- a) Primera Fase: Utilizando los coeficientes obtenidos en el período de calibración, se analizarán los resultados y errores para el mismo período inserto con el objetivo de conocer la capacidad de predicción del modelo.
- b) Segunda Fase: Se utilizarán los coeficientes obtenidos en el período de calibración para analizar y validar la capacidad predictiva del modelo dentro del período de validación solo evaluando los coeficientes y conocer de esta manera la consistencia de estos coeficientes para períodos futuros. Para esto, las variables independientes se mantendrán con su valor real.
- c) Fase Final: En esta etapa del período de validación, se agregará a lo ya realizado en la fase dos, las variables independientes obtenidas a través de media móviles con una ventana de tiempo de 12 meses. Esto con el objetivo de validar las variables independientes y conocer su consistencia con respecto al caso anterior, de esta manera obtener cuánto varían los resultados de los modelos al utilizar este método de medias móviles respecto de los valores obtenidos en la fase dos.

Finalmente se expondrán las conclusiones respectivas para cada caso.

Los modelos expuestos más adelante cumplen los siguientes requerimientos:

- Las variables deben ser significativas. Puesto que en la teoría las variables no significativas no influyen en el resultado del modelo.
- Las variables deben tener un coeficiente interpretable. Esto es sumamente importante, ya que los coeficientes deben tener un significado consistente con el modelo. En caso contrario, si bien puede darse que el ajuste del modelo aumente, no se le puede dar una interpretación correcta a los variables y por ende saber el real motivo de la fortaleza (coeficiente alto) de la variable en el modelo.

Para este tipo de modelos se utilizaron las variables RFM (Recency, Frecuency y Monto Interés), las que fueron analizadas de dos maneras diferentes definidas como: RFM Simple y RFM Segregado.

7.4.2 Modelos RFM simple

Este consiste en tomar la frecuencia de compra y recency condicionado a transacciones con interés a nivel agregado, en otras palabras, indiferente en que holding se realizaba la compra. Acorde a esto el modelo queda descrito como:

$$\text{Interés } (t+1) = C_1 + C_2 * \text{Frecuencia } (t) + C_3 * \text{Recency } (t) + C_4 * \text{Interés Promedio } (t) + \sum C_x * \text{Mes } X (t+1)$$

Fórmula 20

Donde:

Interés (t+1): Es la variable dependiente que se quiere predecir. Esta representa el monto interés total que se pagará en el período t+1.

Frecuencia (t): Es la frecuencia de compra con interés, independiente del holding en que se realice.

Recency (t): Entrega cuantos meses han pasado desde la última compra con interés, independiente del holding en que fue realizada la transacción hasta el período t.

Interés Promedio (t): Lleva el promedio del interés pagado por el cliente hasta el período t.

Mes X (t+1): Dummy que será 1 si el período t+1 es el mes X y 0 en otro caso. $X \in [1,12]$

Los coeficientes obtenidos para cada variable en cada uno de los modelos se presentan a continuación:

Tabla 16: Coeficientes de modelos para conjuntos de variables RFM simple

	Modelos y Coeficientes			
	Lineal	Log Normal	Poisson	Binomial Negativo
Constante	1195,317	0,698	5,071	6,536
Frecuencia	7986,954	5,987	0,588	0,976
Recency		-0,025		0,047
Interés Promedio	0,443	0,044	0,435	0,216
Mes 3	749,885	0,137		0,280
Mes 12		0,747		

Los cuadros que no presentan coeficientes son dados así puesto que en esos modelos las variables no tenían significancia por lo cual fueron sacadas. La variable es considerada significativa si su p-valor es menor o igual a 0,05.

Las dummies de meses fueron retiradas exceptuando diciembre y marzo puesto que al colocarlas todas dentro del modelo se pierde interpretabilidad de los coeficientes incluyendo las mismas. Las dummies de diciembre y marzo fueron dejadas por experticia del gerente del área y apoyado por los estudios realizados de las repercusiones de estas en cuanto al interés pagado en los distintos meses. Utilizando solo estas dummies no se pierde sentido de interpretación de las variables dentro del modelo.

En la tabla siguiente se muestran los valores referentes al ajuste entregados por los modelos.

Tabla 17: Bondad de Ajuste para RFM simple¹³

	Bondad de Ajuste			
	Lineal	Log Normal	Poisson	Binomial Negativo
Deviance/df	1645307926,4	14,35	30652,8	11,72
Pearson Chi-Square/df	1645307926,4	14,35	474036,1	72,10
AIC	3825687	874851	4874736988	3059861
BIC	3825737	874891	4874737018	3059920

Si bien los ajustes de los modelos al ver los valores de los estadísticos AIC y BIC no son buenos (valores muy altos¹⁴), cabe recordar que lo que se está buscando en esta memoria no es calcular el valor exacto del cliente si no que distinguir entre los “buenos” y “malos” clientes. Cabe notar que los valores de las dos primeras métricas, (Deviance/df y Pearson Chi-Square/df) específicamente los dados para el modelo de Poisson, son muy superiores a 1 lo cual es un indicador de que existe sobredispersión en los datos.

A continuación se realiza un análisis referente a los coeficientes obtenidos. Para esto hay que tener en cuenta que el valor de los parámetros del modelo lineal no es comparable con los demás modelos puesto que usan configuraciones diferentes.¹⁵

Análisis Coeficientes RFM Simple

Frecuencia: Esta variable dice que el cliente mientras más compre con interés mayor será su interés pagado lo cual es intuitivo. De gran peso para todos los modelos en general, siendo este más significativo en el caso del modelo log-normal que le atribuye gran cantidad de la explicación del modelo a esa variable. Cabe notar además que el modelo de Poisson asigna aproximadamente la mitad que el peso en comparación al binomial negativo, por lo tanto es un valor más importante para este último modelo.

Recency: Esta variable tan solo es significativa para los modelos log-normal y binomial negativo, presentando este poco peso en cada uno de sus modelos para explicar la variable dependiente. Ahora bien, cada modelo le asigna un signo diferente al parámetro por lo que su interpretación es más confusa. Para el primer caso se puede interpretar que mientras más tiempo pasa menos valor se pagará al período siguiente y lo contrario pasa para el modelo B.N.

Interés Promedio: Coeficiente signficante para todos los modelos, el cual tiene su mayor peso en el modelo lineal y de Poisson, siendo el valor de este último el doble que el modelo binomial negativo y la variable que explica gran parte del modelo. Por otro lado el modelo log-normal le otorga el peso más bajo en comparación al resto de los otros modelos.

¹³ Con la intención de comprar entre modelos, se han dejado los indicadores AIC y BIC. Véase anexo 5 para ver valores de R² obtenidos para los modelos lineales y log-normales.

¹⁴ Recordar que en estos indicadores mientras más bajo es su valor, mejor es el nivel de ajuste.

¹⁵ Véase Marco Conceptual

Meses 3 y 12: Mes 3 no es significativo solo para el modelo de Poisson y tiene el mayor peso en el modelo lineal y binomial negativo. Por otra parte el mes 12 tan solo es significativo para el modelo log-normal otorgándole un peso de 0,747.

Fase Uno: Calibración

Correlación de Spearman a Nivel Individual

Tabla 18: Ordenamiento individual fase uno

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,726	0,593	0,770	0,696
Significancia	0	0	0	0

De la tabla anterior se puede apreciar que el modelo de Poisson presenta mayor nivel de correlación respecto al ranking real con un 77%, mientras que el peor modelo viene dado por el log-normal el cual tiene una correlación de tan solo el 59%. Cabe destacar que el modelo lineal y el binomial negativo se encuentran entre el 73% y 70% respectivamente, lo cual demuestra que presentan un buen nivel de ajuste.

Correlación de Spearman a Nivel de Segmento

Tabla 19: Ordenamiento por segmento fase uno

	Ordenamiento por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,675	0,549	0,726	0,657
Significancia	0	0	0	0

Como en el punto anterior, Poisson se mantuvo como el modelo que mejor distingue entre los buenos y malos clientes ahora a nivel de segmento. Esto pues presenta una correlación de 73% lo cual es un valor bueno si consideramos la gran volatilidad existente entre las compras y pagos de interés de cada transacción. El modelo lineal tiene una correlación de 67,5% respecto del ranking real referente al valor del cliente, lo cual indica que sigue siendo un buen modelo predictivo utilizando estas variables al igual que el binomial negativo. Por último el modelo log-normal presenta el nivel de relación más bajo de todos los modelos.

Tabla 20: Porcentaje de acierto para modelos fase uno

Ranking Real	Ranking Predicho	Porcentaje Acierto Modelos			
		Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	67%	48%	71%	63%
	2	23%	30%	23%	26%
	3	8%	17%	5%	7%
	4	2%	5%	1%	3%
2	1	24%	34%	22%	27%
	2	39%	32%	47%	38%
	3	27%	25%	25%	26%
	4	10%	10%	6%	9%
3	1	6%	15%	5%	7%
	2	29%	28%	22%	25%
	3	39%	33%	47%	40%
	4	26%	23%	26%	28%
4	1	2%	3%	3%	3%
	2	8%	10%	8%	10%
	3	26%	25%	23%	26%
	4	63%	61%	67%	61%

Los resultados serán analizados por segmentos.

Segmento Ranking 1:

Para los clientes más valiosos, es decir, los pertenecientes al Ranking 1, el modelo Poisson predice el 71% de estos de forma correcta. Si bien no es el 100%, cabe destacar que del resto de los clientes pertenecientes al segmento 1, el 23% los ubica en el Ranking 2 y tan solo el 1% dentro del segmento menos valioso. Además en forma global el 94% de los clientes más valiosos los ubica dentro de los 2 segmentos más valiosos, los cuales al recordar el capítulo 8, el 80% y el 15% de las contribuciones entregadas a la empresa son dadas por el segmento Ranking 1 y Ranking 2 respectivamente. El modelo lineal y el binomial negativo, presentan valores muy similares acorde a su nivel de acierto para este segmento. Ambos superan el 60% de acierto, en el cual el modelo lineal es levemente superior con un 67% de efectividad. Además en los dos modelos el 90% de los clientes son ubicados dentro de los 2 segmentos más valiosos, por ende, para este segmento al igual que el modelo de Poisson, ambos distinguen con un nivel muy bueno entre los buenos y malos clientes. Para el caso del modelo log-normal, este presenta los valores más bajos de acierto, puesto que solo el 48% de los clientes es ubicado correctamente.

¹⁶ Matrices para los Modelos Lineales y Binomial Negativo se encuentran en anexo.

Por otra parte, ubica el 22% de clientes valiosos dentro de los segmentos menos valiosos.

Segmento Ranking 2:

Para los clientes ubicados dentro de estos 2 segmentos, el modelo de Poisson vuelve a ser el que con mayor precisión ubica a los clientes con un 47% de exactitud. Un 22% es ubicado en el segmento más valioso, lo cual el 69% es de los clientes es ubicado dentro de los dos segmentos más valiosos y tan solo el 6% es ubicado en el segmento 4. Para el modelo lineal y binomial, ambos nuevamente presentan valores muy similares colocando el 39% aproximadamente de los clientes de forma correcta y el 10% en el segmento menos valioso. El modelo log-normal presenta los peores resultados acertando el 32% de los clientes.

Segmento Ranking 3:

El modelo de Poisson posiciona el 47% de los clientes bien, por otra parte el lineal y el binomial un 39% y 40% respectivamente. Más abajo se encuentra el modelo log-normal que solo acierta el 32% de los clientes.

Segmento Ranking 4:

Dentro de los clientes menos valiosos Poisson vuelve a obtener los mejores resultados, el cual estima correctamente el 67% de los clientes del segmento 4, mientras que los otros modelos se encuentran cercanos al 60%, donde el modelo lineal presente una leve superioridad con un 63%. Cabe destacar que dentro de este segmento el 85% de todos los clientes son posicionados dentro de los 2 segmentos menos valiosos, lo cual indica que todos tienen un buen nivel de acierto para este caso.

En resumen:

El modelo de Poisson presenta un mayor nivel de predicción para todos los segmentos. El cual presentó un acierto de un 71% para el segmento 1, un 67% para el segmento 4 y finalmente de un 47% para el segmento 2 y 3. Mientras que el modelo log-normal presentó los peores resultados para todos los segmentos.

Los modelos en general presentaron mejores resultados para los clientes del segmento 1 y 4, mientras que para los clientes pertenecientes a los segmentos del centro 2 y 3, se pierde precisión. Esto se debe principalmente a la poca diferencia existente entre la contribución de interés entregada a la compañía por parte de estos grupos (véase capítulo 9.1 "Holdings), incluyendo también al segmento 4, lo que provoca que se distribuyan de manera más proporcional al momento de ubicar a los clientes pertenecientes a estos segmentos.

A continuación se presenta en forma más detallada los resultados obtenidos para el mejor modelo encontrado anteriormente.

Tabla 21: Modelos Poisson fase uno

Modelo Poisson				
Ranking Real	Predicción Ranking	Frecuencia	Porcentaje Acierto	Porcentaje Acumulado
1	1	2340	71%	71%
	2	759	23%	94%
	3	172	5%	99%
	4	41	1%	100%
2	1	713	22%	22%
	2	1567	47%	69%
	3	835	25%	94%
	4	197	6%	100%
3	1	157	5%	5%
	2	731	22%	27%
	3	1558	47%	74%
	4	866	26%	100%
4	1	102	3%	3%
	2	255	8%	11%
	3	747	23%	33%
	4	2211	67%	100%

Fase dos: Validación Coeficientes

Correlación de Spearman a Nivel Individual

Tabla 22: Ordenamiento individual fase dos

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,684	0,575	0,715	0,648
Significancia	0	0	0	0

Tal como en el período de calibración, el modelo de Poisson sigue manteniendo la correlación más alta para rankings individuales con un 71,5%, como también el modelo log-normal se mantiene como el menos eficaz con un 57,5%. Cabe anotar además que todos los modelos bajaron sus correlaciones en un 4% promedio, una cifra baja, la cual no es muy significativa si se toma en cuenta el cambio en las tendencias de los consumidores a lo largo del tiempo.

Correlación de Spearman a Nivel de Segmento

Tabla 23: Ordenamiento por segmento fase dos

	Ordenamiento por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,639	0,532	0,659	0,579
Significancia	0	0	0	0

Las correlaciones de ranking a nivel de segmento mantienen el mismo patrón que para el caso anterior, mostrándose una correlación más baja respecto al ordenamiento individual. Este fenómeno es debido a que los ordenamientos por ranking son más castigados en comparación con los individuales puesto que el ranking de los segmentos puede tomar solo valores entre 1 y 4 en cambio a nivel individual se toman valores entre 1 y 13.251.

Matriz de confusión respecto a rankings por segmento

Tabla 24: Porcentaje de aciertos fase dos

Ranking Real	Ranking Predicho	Porcentaje Acierto Modelos			
		Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	63%	47%	65%	58%
	2	25%	29%	27%	29%
	3	10%	18%	7%	11%
	4	2%	6%	1%	2%
2	1	24%	33%	22%	27%
	2	39%	32%	41%	35%
	3	28%	25%	31%	29%
	4	9%	10%	6%	9%
3	1	9%	16%	9%	10%
	2	27%	28%	22%	24%
	3	38%	33%	41%	36%
	4	26%	23%	29%	31%
4	1	4%	4%	5%	5%
	2	9%	11%	9%	12%
	3	24%	24%	21%	25%
	4	63%	61%	65%	58%

Segmento Ranking 1:

Dentro del segmento más valioso, el modelo de Poisson presenta los mejores resultados acorde a lo visto en las correlaciones analizadas, ubicando correctamente al 65% de los clientes mientras que al 27% del mismo grupo los ubica en el segmento 2. Además solo el 1% de los clientes son ubicados en el segmento menos valioso. En otras palabras, el modelo sigue distinguiendo bien entre los “buenos” y “malos” clientes, donde el 92% de los clientes más valiosos son ubicados dentro de los 2 segmentos de mayor valor y el resto en los otros dos grupos. El modelo lineal se presenta con buenos resultados de distinción acertando el 63% de los clientes y el 88% entre los dos mejores grupos. Tan solo coloca el 2% del grupo uno en el segmento menos valioso. El modelo log-normal presenta los más bajos resultados ubicando correctamente solo al 47% de los clientes y al 24% de los mejores clientes los ubica entre los dos grupos menos valiosos.

Segmento Ranking 2:

Al igual que para el segmento anterior, el modelo de Poisson presenta el mejor nivel de precisión. Además cabe destacar que nuevamente el modelo lineal se ubica en segundo lugar con valores levemente distintos a los de Poisson, sin embargo, 9% de estos clientes los ubica dentro del grupo menos valioso al igual que el modelo binomial. Por otra parte el modelo log-normal solo posiciona bien al 32% de los clientes de este segmento y el 10% los ubica en el peor grupo.

Segmento Ranking 3:

Tal como en el caso anterior, el modelo de Poisson ubica correctamente al 41% de los clientes del grupo 3, seguido por el modelo lineal con un 38%, luego sigue el binomial con un 36% y finalmente el log-normal con un 33%.

Segmento Ranking 4:

Para el grupo 4, Poisson tienen el nivel de acierto más alto con un 65% de los clientes, además cabe destacar que dentro de este grupo de los menos valiosos, el 86% de estos son ubicados dentro de los 2 segmentos de menos valor y el 5% de estos clientes son puestos dentro del grupo de los más valiosos. El modelo lineal ubica correctamente el 63% de este segmento menos valioso y tan solo el 4% dentro del grupo 1.

En resumen:

Dentro del período de validación se mantienen más menos los mismos resultados que en el período de calibración con una disminución respecto de ese período de un 4% promedio para los rankings y de una variación promedio de un 1,8% dentro de los valores de la matriz de confusión, donde el modelo que más varía es el modelo de Poisson con un 2,8%.

El modelo de Poisson continúa siendo el modelo que ubica de manera más correcta a los clientes, donde se destaca el acierto a los grupos 1 y 4 con un 65% de efectividad para ambos casos seguido por el modelo lineal.

Continúa la tendencia para los segmentos 2 y 3 los que nuevamente presentan menos porcentaje de acierto para todos los modelos en general en cuanto a los grupos por lo ya explicado en el punto anterior.

Fase tres: Validación Total (Variables Independientes obtenidas por medias móviles con ventana de tiempo de 12 meses)

En este caso los resultados fueron analizados usando los coeficientes obtenidos del período de calibración y la inclusión de las variables independientes sacadas mediante medias móviles.

Correlación de Spearman a Nivel Individual

Tabla 25: Ordenamiento individual fase tres

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,569	0,480	0,603	0,559
Significancia	0	0	0	0

Los resultados siguen manteniéndose como en las fases anteriores en el cual la correlación más alta viene dada por el modelo de Poisson y la más baja por el modelo log-normal. Sin embargo las correlaciones bajan en un 10% promedio con respecto a la fase dos. Cabe destacar además que el modelo binomial negativo reduce su diferencia con respecto al modelo lineal, en que ambos alcanzan correlaciones muy similares para el ranking a nivel individual. Esto se debe a que la diferencia de correlación del modelo binomial con la fase anterior es de un 8,7% en cambio la del lineal alcanza un 11,6%.

Correlación de Spearman a Nivel de Segmento

Tabla 26: Ordenamiento por segmento fase tres

	Ordenamiento por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,534	0,457	0,566	0,523
Significancia	0	0	0	0

Para los valores por segmentos, Poisson sigue manteniendo su diferencia con los demás modelos aunque para este caso la diferencia se ve disminuida levemente. Con respecto a la fase anterior, el ranking por segmento al igual que el individual disminuye su correlación en un 10% aproximadamente, producto de la predicción de las variables independientes mediante medias móviles.

Matriz de confusión respecto a rankings por segmento

Tabla 27: Porcentaje de acierto para modelos fase tres

		Porcentaje Acierto Modelos			
Ranking Real	Ranking Predicho	Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	56%	45%	58%	55%
	2	24%	28%	25%	26%
	3	13%	18%	11%	13%
	4	7%	9%	5%	6%
2	1	27%	32%	25%	28%
	2	36%	31%	38%	34%
	3	25%	24%	27%	26%
	4	13%	14%	10%	12%

3	1	11%	17%	11%	11%
	2	28%	27%	25%	26%
	3	34%	31%	36%	34%
	4	26%	25%	28%	30%
4	1	5%	6%	6%	6%
	2	12%	13%	12%	15%
	3	28%	28%	25%	27%
	4	55%	53%	57%	52%

Los resultados analizados por segmento se analizan a continuación:

Segmento Ranking 1:

Se observa a modo general que el modelo de Poisson presenta los valores más altos de predicción lo que rectifica los resultados de la correlaciones, acertando el 58% de los clientes más valiosos para la compañía y además el 83% de estos es ubicado dentro de los 2 segmentos de mayor valor. Tan solo el 5% es ubicado dentro del segmento menos valioso. Estos valores, comparados con los valores obtenidos en la fase dos disminuyen su porcentaje de predicción en un 7% y aumenta su error en un 4% respecto de la ubicación de estos clientes en el segmento menos valioso. Si bien baja su porcentaje de aciertos, el modelo aún mantiene una buena distinción entre los “buenos” y “malos” clientes.

Los modelos binomial y lineal presentan valores muy similares en el cual el modelo lineal predice con un 1% mejor a los clientes del segmento 1 pero presenta mayor error al ubicar al 7% de estos clientes dentro de los menos valiosos en comparación del 6% del modelo binomial.

Como en todo el proceso, el modelo log-normal presenta los resultados más bajos en todos los aspectos, en el cual solo acierta al 45% de los clientes más valiosos y al 27% los ubica dentro de los segmentos de menor valor siendo el 9% de estos ubicados en el segmento 4.

Segmento Ranking 2:

Al igual que para el segmento anterior, los niveles de predicción para cada modelo disminuyen con respecto a la fase dos. Para el caso de Poisson, disminuye su porcentaje en un 3%, pasando de un 41% a un 38%, sin embargo aún continúa siendo el modelo con los mejores resultados. De los clientes pertenecientes a este grupo, el 10% los asigna dentro del segmento menos valioso. Para los demás modelos también se ve una baja en comparación a la fase anterior.

Segmento Ranking 3:

Los resultados de este segmento, mantienen los mismos patrones del segmento anterior, en que Poisson predice correctamente el 36% de los clientes y coloca dentro de la vecindad a un 53% de estos. Un 11% de los clientes de este segmento3 son puestos dentro del grupo de los más valiosos. El modelo binomial y el lineal presentan valores muy similares y finalmente el modelo log-normal tiene los peores resultados.

Segmento Ranking 4:

El modelo de Poisson predice correctamente el 57% de estos clientes, seguido por el modelo lineal con un 55%, presentando valores muy similares al anterior. De forma excepcional el modelo binomial registra el nivel de acierto más bajo para este segmento con un 52%, sin embargo, el modelo log-normal tiene valores muy similares a este.

En conclusión:

El aumento de error ocasionado por la introducción al modelo de las variables independientes obtenidas mediante medias móviles fue de un 10% promedio para las correlaciones a nivel global en comparación a la fase anterior.

El modelo que obtuvo los índices de aciertos más elevados fue Poisson, presentándose como el modelo con mejor consistencia a lo largo de las 3 fases, seguido por el modelo lineal. Respecto al nivel de predicción estudiado en la matriz de confusión, Poisson disminuyó su porcentaje de acierto en un 5% promedio, siendo la disminución más notable la producida en el segmento de los clientes menos valiosos, en el cual paso de un acierto de un 65% a un 57%. Si bien aumenta el error, el modelo aún presenta gran capacidad de distinción entre los “buenos” y “malos” clientes.

Para esto se puede ver que el modelo de Poisson ubicó el 83% de los mejores clientes dentro de los 2 segmentos más valiosos y el 82% de los peores clientes dentro de los 2 grupos menos valiosos.

Aunque para los segmentos 2 y 3 se ve de la tabla 11 que los clientes pertenecientes a dichos grupos han sido distribuidos más uniformemente que para los otros segmentos, esto se debe a como ya se ha explicado anteriormente a la poca diferenciación en cuanto a la distribución por parte de estos grupos. Lo que le produce una complicación al modelo distinguir más claramente entre estos.

Estas variables RFM en modo simple han mostrado tener un buen nivel de diferenciación y de predicción para los segmentos de clientes, en el cual ha destacado como se ha dicho ya, el modelo de Poisson.

Ahora se analizarán los resultados con variables RFM segregadas por holding particularmente las variables de frecuencia y recency.

7.4.2.2 Modelos RFM segregados

En este caso la frecuencia y el recency fueron incorporados por holding. La principal razón de esto fue por los resultados obtenidos en el análisis de transacciones e interés por holding, el cual demuestra claramente las diferencias en cuanto al interés generado por cada uno de ellos. El modelo tiene la siguiente forma:

$$\text{Interés } (t+1) = C_1 + C_2 * \text{Frecuencia_H1 } (t) + C_3 * \text{Recency_H1 } (t) + C_4 * \text{Frecuencia_H2 } (t) + C_5 * \text{Recency_H2 } (t) + C_6 * \text{Frecuencia_H3 } (t) + C_7 * \text{Recency_H3 } (t) + C_8 * \text{Frecuencia_H4 } (t) + C_9 * \text{Recency_H4 } (t) + C_{10} * \text{Frecuencia_H5 } (t) + C_{11} * \text{Recency_H5 } (t) + C_{12} * \text{Frecuencia_H6 } (t) + C_{13} * \text{Recency_H6 } (t) + C_{14} * \text{Interés Promedio } (t) + \sum C_x * \text{Mes X } (t+1)$$

Fórmula 21

Donde:

-Interés (t+1): Es la variable dependiente que se quiere predecir. Esta representa el monto interés total que se pagará en el período t+1.

-Frecuencia_Hx(t): Frecuencia de compra con interés en el Holding X hasta el período t.

-Recency_Hx (t): Contabiliza cuantos meses han pasado desde la última compra con interés en el Holding X hasta el período t.

-Interés Promedio (t): Entrega el promedio del interés pagado por el cliente hasta el período t.

-Mes X (t+1): Dummy que será 1 si el período t+1 es el mes X y 0 en otro caso.

Los coeficientes obtenidos para cada variable en cada uno de los modelos se presentan a continuación:

Tabla 28: Coeficientes de modelos para conjuntos de variables RFM segregados

	Modelos y Coeficientes			
	Lineal	Log Normal	Poisson	Binomial Negativo
Constante		1,533	5,583	7,253
Frecuencia Holding 1	5016,228	3,390	0,239	0,482
Frecuencia Holding 2	10065,144	2,566	0,670	0,675
Frecuencia Holding 3	3754,983	4,206		
Frecuencia Holding 4		2,593		
Frecuencia Holding 5	16861,637	3,358	1,160	2,154
Recency Holding 1		-0,023		
Recency Holding 2		-0,033		-0,020
Recency Holding 3	93,251	0,037		
Recency Holding 4		-0,037		
Recency Holding 5		-0,014		
Interés Promedio	0,374	0,051	0,370	0,153
Mes 3	1051,535			0,209
Mes 12		0,851		0,202

Los cuadros sin valores representan las variables que no presentaban significancia y por ende fueron retiradas del modelo¹⁷.

A continuación se muestra la tabla con la bondad de ajuste de cada modelo.

Tabla 29: Bondad de Ajuste para RFM segregados

	Bondad de Ajuste			
	Lineal	Log Normal	Poisson	Binomial Negativo
Deviance/df	1640825602,67	14,09	31138,96	0,77
Pearson Chi-Square/df	1640825602,67	14,09	216119,19	2,58
AIC	2,60899E+14	2532584,634	4951685130	1596888,928
BIC	2,60899E+14	2532714,331	4951685270	1597038,579

Tal como con el conjunto de variables antes visto los valores de AIC y BIC son altos, siendo obteniendo el modelo log-normal el mejor nivel de ajuste. Además, tal como era de esperarse, la sobredispersión se hace presente mostrando en el caso del modelo lineal y de Poisson valores muy elevados.

Análisis Coeficientes:

Frecuencia por Holding: Para el caso del modelo lineal, como era de esperarse dado el estudio realizado referente a los holdings, el coeficiente de las frecuencias 1, 2 y 5 son los más altos, siendo este último el que presenta el valor mayor con 16.861, seguido por la frecuencia del holding 2. Por lo tanto mientras más se compre en esos holdings, mayor será el valor estimado del interés en el siguiente período. El holding 4 que tiene el menor interés asociado no es significativo para ningún modelo excepto el log-normal. Para el caso de los modelos de Poisson y binomial negativo, siguen la misma estructura que el modelo lineal en el cual la frecuencia del holding 5 es la más alta. Se destaca también que los valores de los coeficientes del modelo B.N. son más altos o tienen más pesos que los dados por Poisson. En estos casos la variable frecuencia holding 3 no es significativa. Cabe recordar que este holding se posicionaba penúltimo en contribución con interés.

Recency por Holding: El recency no es significativo para el caso del modelo de Poisson y tan solo el recency del holding 2 es significativo para el modelo B.N., sin embargo, su peso dentro del modelo es muy bajo. Para el caso del modelo lineal ocurre algo similar en que solo es significativo el recency del holding 3 y ocurre el mismo fenómeno que para los coeficientes del conjunto RFM Simple, en el cual tienen signos distintos y se atribuye la misma explicación a la ya descrita en ese caso. Por otro lado los recency de cada holding se presentan significativos dentro del modelo log-normal, pero en cual sus pesos son bajos.

¹⁷ Véase anexo para ver los resultados con las variables no significativas.

Interés Promedio: El coeficiente de esta variable tiene gran peso tanto en el modelo lineal como el de Poisson, siendo el valor de este último más del doble que el otorgado en el modelo binomial negativo. Por último el log-normal es el modelo que menos peso le da a esta variable en comparación con el resto.

Mes 3 y 12: Tienen valores significativos en algunos casos. El mes 12 tiene un peso alto en el modelo log-normal. El único modelo en el cual ambos son significativos es el modelo B.N., sin embargo sus valores asociados no son muy altos.

Fase Uno: Resultados Período Calibración

Correlación de Spearman a Nivel Individual

Tabla 30: Ordenamiento individual fase uno

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,728	0,628	0,763	0,725
Significancia	0	0	0	0

Utilizando este conjunto de variables, Poisson muestra los resultados más favorables a nivel de ranking individual con un 76,3% de relación con respecto a los ordenamientos reales. Por otra parte el modelo lineal y el binomial negativo también presentan un alto porcentaje de correlación en relación al ranking real estando estos cercanos al 73%. Finalmente el modelo log-normal tiene los resultados más bajos con un 62,8% de relación.

Correlación de Spearman a Nivel de Segmento

Tabla 31: Ordenamiento por segmento fase uno

	Ordenamiento por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,677	0,581	0,720	0,681
Significancia	0	0	0	0

Los patrones de correlación se mantienen similares al ranking individual, exceptuando que en este caso el modelo binomial negativo tiene una correlación levemente mayor que la del modelo lineal, esto significa que sus valores entregados en el ranking individual pueden estar un tanto más alejados con respecto al lineal, pero distingue un poco mejor por segmento.

Matriz de confusión respecto a rankings por segmento

Tabla 32: Porcentaje de aciertos fase uno

Ranking Real	Ranking Predicho	Porcentaje Acierto Modelos			
		Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	66%	51%	70%	63%
	2	25%	28%	24%	26%
	3	7%	16%	5%	8%
	4	2%	5%	1%	2%
2	1	25%	35%	22%	28%
	2	40%	32%	46%	40%
	3	26%	24%	26%	25%
	4	9%	10%	6%	7%
3	1	6%	12%	5%	7%
	2	27%	31%	22%	25%
	3	40%	34%	47%	42%
	4	27%	22%	26%	26%
4	1	3%	2%	3%	2%
	2	8%	9%	8%	8%
	3	27%	26%	22%	25%
	4	62%	63%	66%	64%

Segmento Ranking 1:

Dentro del segmento más valioso, Poisson presenta los mejores resultados ubicando correctamente el 70% de los clientes y el 94% de este segmento los coloca dentro de los dos grupos más valiosos. Además tan solo el 1% de los mejores clientes los ubica en el segmento 4. En segundo lugar se encuentra el modelo lineal con un 66% de acierto seguido por el modelo binomial con un 63%. Ambos ubican el 2% de este segmento dentro del grupo menos valioso. Finalmente se encuentra el modelo log-normal que muestra un mal ajuste con respecto a la distinción de los buenos clientes, ya que tan solo acierta un 51% de estos.

Segmento Ranking 2:

Para el segmento 2, el modelo de Poisson nuevamente tiene un mayor número de aciertos respecto a los demás modelos, siendo este de un 46% y el 48% de los clientes de este segmento es ubicado dentro de la vecindad (segmentos más cercanos). Solo el 6% de este grupo es ubicado en el segmento de menos valor. En este caso el modelo lineal y el binomial obtiene valores muy similares, sin embargo, el modelo lineal presenta mayor error al ubicar al 9% de estos clientes dentro del grupo menos valioso en comparación del 7% del modelo binomial. Los valores entregados por el modelo log-normal son los peores y cabe destacar además que el mayor porcentaje de estos clientes los ubica en un segmento que no es el correcto.

Segmento Ranking 3:

El modelo de Poisson coloca correctamente el 47% de los clientes y ubica solo el 5% de estos dentro del segmento de mayor valor, lo cual indica buena distinción entre los “buenos” y “malos” clientes. A diferencia del segmento 1, aquí el modelo binomial si bien tiene valores cercanos al lineal, presenta un nivel de acierto levemente mayor con un 42%, sin embargo ubica el 7% de estos en el segmento de mayor valor en comparación del 6% del modelo lineal. Como en todos los caso el modelo log-normal solo ubica bien el 34 de los clientes

Segmento Ranking 4:

Para este segmento que concentra a los clientes que entregan menos valor, los modelos en general presentan un buen nivel de distinción, siendo los aciertos mayores al 60%, en el cual destaca el modelo de Poisson con un 66% seguido por el modelo binomial con un 64%. Además todos los modelos ubican entre un 2% y 3% de estos clientes en el segmento 1. Además en forma general, más del 88% de estos clientes son ubicados dentro de los dos segmentos menos valiosos.

En resumen:

Los modelos en general mostraron un buen nivel de distinción entre los “buenos” y “malos” clientes. Mostrándose una mayor distinción para los clientes de los segmentos ranking 1 y 4.

El modelo de Poisson presentó los mejores resultados tanto a nivel de ranking como en distinción de clientes dados por la matriz de confusión.

Además el modelo lineal obtuvo unos valores de correlación de ranking individual un poco mayores que el binomial y este un poco mejores en el ordenamiento de segmento.

El modelo lineal mostró mejor ajuste para predecir al segmento más valioso en comparación con el binomial, sin embargo este último presentó mejor ajuste para predecir a los segmentos menos valiosos.

Los modelos tienen menos nivel de acierto para los segmentos del centro (grupo 2 y 3) dada la poca diferencia entre sus niveles de contribución de interés lo que ocasiona que los clientes que se encuentra con valores muy cercanos y son ubicados en distintos rankings sean confundidos y posicionados en el segmento cercano.

En comparación con el conjunto de variables antes estudiados en el período de calibración, el modelo de Poisson mostró una leve disminución y el modelo lineal en general se mantuvo igual en el cual disminuyó sus ajustes para los segmentos 1 y 4 pero mejoró para los segmento 2 y 3. Por otra parte el modelo binomial se vio beneficiado aumentando su nivel de acierto para prácticamente todos los segmentos como también sus correlaciones de ordenamiento, mostrando por lo tanto mejores resultados al segregar las variables de frecuencia y recency por holding.

Fase dos: Validación Coeficientes

Correlación de Spearman a Nivel Individual

Tabla 33: Ordenamiento individual fase dos

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,673	0,593	0,712	0,681
Significancia	0	0	0	0

Al utilizar los coeficientes obtenidos en el período de calibración en el de validación, se ve una baja en la correlación individual de un 4,6% promedio para los distintos modelos. Si bien hay una disminución, los modelos aún mantienen un nivel alto de correlación especialmente el modelo de Poisson con un 71,2% el cual es seguido por el modelo binomial negativo con un 68,1%.

Correlación de Spearman a Nivel de Segmento

Tabla 34: Ordenamiento por segmento fase dos

	Ordenamiento por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,629	0,548	0,666	0,634
Significancia	0	0	0	0

Siguiendo los patrones del ordenamiento individual, Poisson vuelve a tener la correlación más alta de todos los modelos seguido por el modelo binomial con un 63,4%. El modelo log-normal presenta valores bastante bajos en relación a los demás modelos con una correlación de tan solo el 54,8%. Además se presenta una baja en cuanto a la correlación de la fase anterior de un 4,6%.

Matriz de confusión respecto a rankings por segmento

Tabla 35: Porcentaje de aciertos fase dos

		Porcentaje Acierto Modelos			
Ranking Real	Ranking Predicho	Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	62%	49%	64%	59%
	2	26%	28%	28%	27%
	3	10%	17%	7%	12%
	4	2%	6%	1%	2%
2	1	25%	34%	23%	28%
	2	38%	32%	41%	37%
	3	29%	25%	31%	28%
	4	9%	10%	6%	8%

3	1	9%	14%	8%	9%
	2	25%	30%	22%	26%
	3	37%	33%	41%	37%
	4	28%	23%	29%	27%
4	1	4%	3%	4%	3%
	2	11%	11%	11%	10%
	3	24%	24%	21%	24%
	4	61%	62%	64%	63%

Los resultados por segmento son los siguientes:

Segmento Ranking 1:

Acorde a los resultados de las correlaciones, el modelo de Poisson presenta el mayor nivel de aciertos con un 64% de los clientes pertenecientes a este segmento. Además se puede apreciar que posiciona el 92% de estos clientes dentro de los dos grupos de mayor valor y tan solo el 1% en el segmento 4. En contraste, el modelo log-normal tan solo ubica bien al 49% de estos clientes y el 6% dentro de los peores clientes. Por otra parte el modelo lineal y binomial aciertan un 62% y 59% respectivamente.

Segmento Ranking 2:

Para este grupo, el modelo de Poisson acierta el 41% y el 54% está distribuido dentro de la vecindad. Lo cual como se ha dicho anteriormente es normal, puesto que deben ser clientes que se encuentran al límite de cada segmento y dado que no existe tanta diferencia entre sus valores provoca estos resultados. Cabe destacar además que Poisson solo ubica el 6% de este grupo en el segmento 4. El modelo lineal y binomial para este caso presenta valores similares con un 38% de certeza aproximadamente para cada uno.

Segmento Ranking 3:

Al igual que para el grupo anterior, Poisson ubica correctamente el 41% del grupo 3, seguido por el lineal y binomial con un 37% y finalmente el log-log con un 33% sin embargo este último coloca un 14% de estos clientes en el grupo de mayor valor, en cambio el resto tan solo un 9% en promedio.

Segmento Ranking 4:

Los modelos en forma global presentan los mejores ajustes, los cuales tiene un porcentaje de acierto mínimo de 61% dado por el lineal y el máximo de un 64% entregado por el modelo Poisson. Además en promedio, los modelos ubicaron el 86% de los clientes de este grupo dentro de los dos segmentos de menos valor. Cabe agregar que los modelos muestran buena distinción entre los “buenos” y “malos” clientes colocando tan solo el 3,5% de estos dentro del grupo más valioso.

En resumen:

Al validar los coeficientes, se denotó una disminución de las correlaciones de los rankings en un 4,6% promedio respecto a la fase anterior, es decir hubo un leve aumento en el error. A pesar de esto los modelos, presentaron buenos niveles de distinción entre “buenos” y “malos” clientes.

El modelo que obtuvo los mejores resultados fue el de Poisson con un 64% de acierto para los segmento 1 y 4 de un 41% para los grupos 2 y 3. Ahora con respecto a la fase anterior, este disminuye su porcentaje de acierto en un 4,9% promedio para cada segmento siendo este el valor más alto en comparación con otros modelos.

Los modelos en general tienen mejor nivel de diferenciación de clientes para los menos valiosos ubicando correctamente en promedio el 62,5% de este grupo.

Dentro de los valores obtenidos para los segmentos más valiosos el modelo log-normal se diferencia mucho de los demás modelos, mostrando un mal nivel de distinción para estos grupos, especialmente el más valioso.

Fase tres: Validación Total (Variables Independientes obtenidas por medias móviles con ventana de tiempo de 12 meses)

Correlación de Spearman a Nivel Individual

Tabla 36: Ordenamiento individual fase tres

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,567	0,507	0,601	0,586
Significancia	0	0	0	0

En esta fase final, se nota una baja en la correlación de un 10% respecto a la fase anterior, sin embargo aún se mantienen los patrones en que el modelo de Poisson obtiene los mejores resultados manteniendo aún un buen nivel con un 60,1% seguido por el binomial con un 58,6% y en último lugar el modelo log-normal con un 50,7%

Correlación de Spearman a Nivel de Segmento

Tabla 37: Ranking por segmento fase tres

	Ranking por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,530	0,474	0,563	0,549
Significancia	0	0	0	0

Se continúa con la misma tendencia que en el caso anterior, manteniéndose como era de esperarse el modelo de Poisson con la mayor correlación con respecto al ranking real. Además la disminución de los ordenamientos por segmentos en comparación con la fase anterior también es de un 10%.

Matriz de confusión respecto a rankings por segmento

Tabla 38: Porcentaje de aciertos fase tres

Ranking Real	Ranking Predicho	Porcentaje Acierto Modelos			
		Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	56%	47%	58%	55%
	2	25%	27%	26%	27%
	3	13%	18%	11%	13%
	4	6%	8%	5%	5%
2	1	27%	32%	25%	28%
	2	36%	31%	38%	35%
	3	26%	24%	27%	26%
	4	12%	13%	10%	11%
3	1	12%	16%	11%	12%
	2	26%	29%	24%	26%
	3	34%	30%	36%	34%
	4	28%	25%	29%	28%
4	1	6%	5%	6%	5%
	2	13%	13%	12%	12%
	3	27%	28%	26%	27%
	4	54%	54%	56%	56%

Los resultados por segmento son los siguientes:

Segmento Ranking 1:

En relación a los valores de las correlaciones por segmento, el modelo de Poisson continúa manteniendo los mejores resultados de este grupo con un 58% de acierto sin embargo muestra una disminución de un 6% respecto a la fase dos y el 84% es ubicado dentro de los dos segmentos más valiosos, lo cual indica que aún hay un buen nivel de distinción entre “buenos” y “malos” clientes. El 5% es ubicado dentro del segmento de menos valor. Por otro lado, el modelo lineal y binomial obtienen valores de predicción muy similares, en que el modelo lineal acierta el 56% mientras que el binomial negativo el 55% sin embargo el modelo lineal coloca al 6% de estos clientes dentro del grupo de menos valor en cambio el modelo B.N. el 5%. Finalmente, log-normal mantiene los resultados más bajos con un 47% de clientes posicionados correctamente.

Segmento Ranking 2:

El modelo de Poisson presenta los mejores niveles de predicción para este grupo, con un 38% de acierto seguido por el modelo lineal con un 36%. Además el modelo de Poisson coloca el 10% de estos clientes dentro del segmento 4 mientras que el lineal el 12%, mostrando menos poder de distinción. Por otro lado el modelo B.N. acierta el 35% de los clientes y un 10% los ubica en el segmento menos valioso. Como era de esperar el modelo log-normal presenta los peores resultados de distinción.

Segmento Ranking 3:

Se puede apreciar que los aciertos han disminuido levemente respecto a los del segmento anterior, prediciendo el modelo de Poisson correctamente el 36% de clientes disminuyendo así en un 5% respecto de la fase anterior. Además ubica un 52% de este grupo dentro de la vecindad, por las mismas razones ya antes descritas. Por otra parte el modelo lineal y el B.N. obtienen los mismos porcentajes acertando el 34% y un 54% son ubicados en la vecindad. Finalmente se encuentra el modelo log-normal con un 30%, presentando los peores resultados.

Segmento Ranking 4:

Para el grupo de clientes de menor valor, excepcionalmente el modelo binomial negativo obtiene los mejores resultados con un 56% de estos predichos correctamente y el 83% de estos ubicados dentro de los dos segmentos menos valiosos. En comparación con el modelo Poisson que acierta el 56% de los clientes el cual muestra una baja de un 8% de precisión en comparación con la fase anterior. Además ubica el 6% de estos en el grupo 1 mientras que el B.N. solo el 5%. Cabe agregar además que en este caso tanto el modelo lineal como el log-normal presentan valores muy similares a los modelos ya analizados, prediciendo correctamente el 54% de los clientes pertenecientes a este grupo.

En conclusión:

Los modelos bajan en un 10% promedio su nivel de correlación respecto a la fase dos, producto del reemplazo de los valores reales de las variables independientes por su predicción mediante medias móviles. Por otro lado, acorde a los valores de la matriz de confusión, en promedio disminuyen su nivel de acierto en un 4,4%. Si bien son valores significativos, era de esperarse tener un error asociado al predecir las variables independientes, sin embargo es pequeño si se toma en cuenta que el método usado fue el de medias móviles, mostrando por ende buen nivel de adecuación a los valores reales.

El mejor modelo presentado dado este conjunto de variables fue el de Poisson, el cual mostró mejores niveles de predicción durante las 3 fases para los 4 segmentos exceptuando el grupo 4 en la última fase, sin embargo para este segmento obtuvo valores muy similares al modelo que presentaba los mejores resultados. Cabe agregar que este modelo presentó la disminución más alta respecto al nivel de

aciertos, disminuyendo en un 5,4% promedio en comparación a la fase dos, reflejándose su mayor disminución en el segmento 4 con un 8%. Si bien aumenta el error, el modelo aún presenta gran capacidad de distinción entre los “buenos” y “malos” clientes y se mantiene como el modelo con los mejores resultados.

Particularmente, el modelo de Poisson ubicó el 84% de los mejores clientes dentro de los 2 segmentos más valiosos y el 81% de los peores clientes dentro de los 2 grupos menos valiosos.

En comparación con el conjunto de variables definidas como RFM Simple, los modelos a nivel agregado presentan mejorías, especialmente el modelo binomial negativo el cual al analizar sus valores obtenidos en el ordenamiento por segmento de la fase tres, aumenta su correlación de ranking en un 2,5%, seguido por el log-normal con 1,7%. Sin embargo el mejor modelo antes encontrado, el cual se mantiene también con los mejores resultados para este set de variables, el modelo de Poisson, disminuye su correlación en un 0,34%.

Por lo tanto no se encontraron mejoras significativas al utilizar la frecuencia y el recency segregado por holding, puesto que lo importante era encontrar un modelo que obtuviera mejores resultados con respecto al análisis anterior. Sin embargo el modelo que presenta mejores resultados con las variables RFM segregado no superó al modelo de Poisson con las variables RFM simple.

A continuación se procederá a realizar el análisis incluyendo variables demográficas y referentes a la tarjeta.

7.4.2.3 Modelos con variables transaccionales, demográficas y relacionadas a la tarjeta¹⁸

En esta sección se incluirán en el análisis, variables transaccionales, demográficas y referentes a la tarjeta en el mejor modelo RFM antes encontrado, con el objetivo de conocer si estas variables mejoran la predicción de “buenos” y “malos” clientes.

Dado que en la parte anterior los modelos mejoraron en forma agregada utilizando las frecuencias y los recency por holding, a este se le incluirán las siguientes variables:

-Edad (t+1): Años que tiene la persona.

-Años Cuenta (t+1): Número de años que han pasado desde que el cliente obtuvo la tarjeta de crédito de la compañía.

-Número Cuotas Promedio (t): Número de cuotas promedio con el cual se cancelan las compras dentro de los holdings de la empresa.

-Monto Compra Promedio (t): Monto promedio de la compra que se paga mediante la tarjeta, el cual se obtiene descontando del monto de la compra total el abono dado por el cliente.

¹⁸ Por simplicidad el modelo será definido como RFM Full

-Rango Scoring (t+1): Variable que viene dada por el modelo de riesgo de la compañía. Como se ha descrito anteriormente 1 implica al cliente más riesgo mientras que 4 al cliente menos riesgoso.

-Monto Cupo (t+1): Monto máximo en dinero permitido para hacer compras con la tarjeta de la firma.

-Porcentaje Uso Tarjeta (t): Entrega el porcentaje de uso de la tarjeta respecto al límite dado por el monto cupo. Esta variable fue obtenida de la siguiente manera:

$$\frac{\text{Monto Deuda}}{\text{Monto cupo}}$$

Fórmula 22

Para elegir la combinación óptima de variables para cada modelo, se hicieron una serie de pruebas en la cual se usaron distintos set de variables y finalmente se escogieron acorde a sus resultados de correlación tanto para nivel individual como de segmento los mejores.

A continuación, se muestran los resultados de estos modelos:

Tabla 39: Coeficientes de modelos para conjuntos de variables RFM full

	Modelos y Coeficientes			
	Lineal	Log Normal	Poisson	Binomial Negativo
Constante	-924	0,638	6,511	7,212
Frecuencia Holding 1	1.983	3,123	0,167	0,271
Frecuencia Holding 2	8.284	2,396	0,559	0,587
Frecuencia Holding 3		3,736		
Frecuencia Holding 4		2,480		
Frecuencia Holding 5	9.580	2,907	0,550	0,509
Recency Holding 1		-0,016		
Recency Holding 2		-0,027		-0,012
Recency Holding 3		0,030		
Recency Holding 4		-0,033		
Recency Holding 5		-0,008	-0,010	-0,016
Interés Promedio	0,251		0,255	0,083
Años Cuenta	160	0,026	0,018	
Edad		0,173		0,009
Sexo	574		0,084	
Monto Cupo				
Rango % Uso		0,249		
N° Cuotas Promedio	1.668	0,233	0,077	0,196
Monto Promedio		0,046	0,024	0,024
Scoring	-1014,972	-0,429	-0,333	-0,237
Mes 3	726			0,190
Mes 12		0,834		0,263

Los ajustes de los modelos se pueden ver en la tabla 39, en que al igual que en los casos anteriores existe sobredispersión en los datos y acorde a los valores entregados por AIC y BIC, el mejor ajuste a los datos los presenta el modelo log-normal.

Tabla 40: Bondad de ajuste para RFM full

	Bondad de Ajuste			
	Lineal	Log Normal	Poisson	Binomial Negativo
Deviance/df	1636511342	14,0349034	29828,8282	11,53067754
Pearson Chi-Square/df	1636511342	14,0349034	273970,697	64,56137622
AIC	2,60209E+14	2523770,01	4743495426	3029245,899
BIC	2,60209E+14	2523929,64	4743495526	3029375,597

Análisis Coeficientes:

Frecuencia por Holding: Para todos los modelos excepto el log-normal, las frecuencias de los holdings que contribuyen con más interés a la compañía tienen significancia, en que los mayores pesos los tienen el holding 2 y 5. En que se destaca que el peso de la frecuencia del holding 1 para el modelo B.N. es casi el doble que el de Poisson. Por lo tanto a mayor compra dentro de esos holdings, mayor es el interés asociado al próximo período (puesto que estos coeficientes son positivos). Por otro lado las frecuencias dentro del modelo log-normal son todas significativas, las cuales presentan valores muy similares unas con otras.

Recency por Holding: En general los recency independiente del holding, no tienen pesos muy significativos. Cabe destacar que en todos los modelos exceptuando el log-normal, la mayoría de los recency no son significativos. Para Poisson solo el recency del holding 5 tiene peso afectando este negativamente. En el caso del modelo B.N. el recency del holding 5 y 2 tienen significancia, siendo ambos de bajo peso y negativos. En otras palabras, mientras más tiempo pase menor es el interés que se obtendrá en el próximo período. Por último dentro del modelo log-normal son todos significativos sin embargo sus coeficientes son muy bajos.

Interés Promedio: Con respecto a esta variable se puede ver que tiene un peso relevante para el modelo lineal y de Poisson. Para el caso del modelo binomial negativo, esta tiene asociado un coeficiente 3 veces menor que el de Poisson. Y finalmente no es significativa para el modelo log-normal.

Años Cuenta: Tiene asociado un signo positivo como era de esperarse (véase Figura 18), esto es, mientras más años el cliente tenga la tarjeta mayor es el valor esperado del interés. Cabe agregar que sus pesos no son elevados en ningún modelo y para el caso del B.N. no es significativo.

Edad: Al igual que en el caso anterior este tiene asociado un signo positivo, puesto que a mayor edad mayor interés esperado. Tiene significancia para los modelo log-normal y binomial negativo, sin embargo este presenta un peso relevante para el primero siendo el valor del binomial negativo muy bajo.

Sexo: Tiene signo positivo en los modelos que presenta significancia, pues los hombres pagan mayor interés en promedio que las mujeres, o cual produce este efecto. Ahora sus valores son bajos y poco significantes.

Rango % Uso: Tan solo tiene significancia para el modelo log-normal, el cual se interpreta de la siguiente manera: Mientras más proporción de la tarjeta se use, mayor es el interés esperado en el siguiente período.

Cuotas Promedio: Con respecto a esta variable todos los coeficientes son significativos, la cual tiene un alto valor para todos los modelos excluyendo Poisson que tan solo tiene asociado un 0,07. Es decir, tiene un gran peso positivo referente a la variable dependiente, en que mientras más grande es el número de cuotas promedio mayor es el interés asociado al próximo período.

Monto Promedio: Coeficiente con signo positivo el cual se espera que a mayor valor de este, el interés estimado sea mayor. Sus coeficientes son bajos pero los valores que puede tomar esta variable son altos. El mayor peso lo tiene el modelo log-normal con un 0,047. En el caso del modelo lineal, este parámetro no es significativo.

Scoring: Al contrario de la gran mayoría de las variables, esta presenta signo negativo y es significativa para todos los modelos. Esto ocurre por lo visto en el gráfico 09, en el cual se puede apreciar que a nivel global, mientras más riesgoso es un cliente menos es el interés promedio pagado. Cabe recordar que mientras el scoring sea más alto el cliente serpa considerado más riesgoso.

Mes 3 y 12: Ambas variables son significativas solo en el modelo B.N las cuales son positivas dado que se espera un aumento para esas fechas como época escolar y navidad, siendo en este caso el coeficiente mes 12 mayor. Para el modelo de Poisson no ninguna tiene significancia pero por otro lado el mes 3 es significativa para el modelo lineal y el mes 12 para el modelo log-normal, siendo el peso de estos coeficientes algo considerables.

Fase Uno: Resultados Período Calibración

Correlación de Spearman a Nivel Individual

Tabla 41: Ordenamiento individual fase uno

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,714	0,651	0,757	0,738
Significancia	0	0	0	0

Para este conjunto de variables la correlación de todos los modelos con respecto al ordenamiento individual real es alta, sobre todo para los modelos de Poisson y binomial negativo con un 75,7% y 73,8% respectivamente. La relación más baja está dada por el modelo log-normal con un 65%. Cabe destacar además que existe una diferencia de un 10% entre el nivel de correlación del modelo log-normal con el de Poisson, lo que demuestra que este último tiene mejor nivel de distinción entre clientes.

Correlación de Spearman a Nivel de Segmento

Tabla 42: Ordenamiento por segmento fase uno

	Ordenamiento por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,664	0,606	0,711	0,692
Significancia	0	0	0	0

Continuando el patrón dado en las correlaciones anteriores, Poisson muestra los mejores resultados con una relación a nivel de segmento de 71,1%, seguido por el binomial negativo con un 69,2% y en último lugar el modelo log-normal. A continuación se verán estos resultados reflejados en la matriz de confusión.

Matriz de confusión respecto a rankings por segmento

Tabla 43: Porcentaje de aciertos fase uno

		Porcentaje Acierto Modelos			
Ranking Real	Ranking Predicho	Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	68%	53%	69%	67%
	2	24%	28%	25%	24%
	3	7%	15%	5%	7%
	4	2%	4%	1%	2%
2	1	25%	34%	23%	27%
	2	40%	33%	44%	41%
	3	25%	24%	26%	24%
	4	10%	9%	7%	9%
3	1	5%	12%	5%	5%
	2	26%	30%	23%	27%
	3	38%	35%	44%	40%
	4	32%	23%	28%	29%

4	1	2%	1%	2%	2%
	2	10%	9%	9%	8%
	3	31%	26%	25%	29%
	4	56%	64%	63%	61%

Los resultados al analizar por segmento son los siguientes:

Segmento Ranking 1:

Como era de esperar dado los resultados de las correlaciones, el modelo de Poisson obtiene el nivel de acierto más alto con un 69%, colocando además el 94% de estos clientes de ranking 1 dentro de los dos segmentos de mayor valor y tan solo el 1% en el segmento 4, mostrando así un muy buen nivel de distinción. Por su parte el modelo lineal y binomial negativo también muestran un buen nivel, colocando correctamente el 68% y 67% respectivamente y tan solo el 2% de este grupo más valioso dentro del segmento 4.

Segmento Ranking 2:

Al igual que en el grupo anterior, Poisson obtiene los mejores resultados obteniendo un 44% de acierto y otro 49% de este grupo es ubicado dentro de la vecindad. Ubica el 7% de estos clientes en el segmento de menos valor. Log-normal presenta malos resultados, sin embargo, el modelo lineal y binomial muestran mejores resultados, en que nuevamente obtienen valores muy similares con un 40% y 41% de acierto respectivamente y posicionan aproximadamente un 9% de estos dentro del grupo de menor valor.

Segmento Ranking 3:

Para este segmento ya dentro de los dos menos valiosos, el modelo de Poisson tiene el mayor nivel de distinción con un 44% y tan solo ubica el 5% de estos dentro del grupo de mayor valor. Hay que destacar que en este grupo el modelo B.N. se diferencia del lineal presentando mejores resultados con un 40% de acierto contra un 38% del lineal. En último lugar está el modelo log-normal con un 35% de eficacia, sin embargo ubica el 12% de estos clientes dentro del segmento ranking 1 lo cual es un valor considerable.

Segmento Ranking 4:

En general todos los modelos presentan buenos resultados colocando correctamente más del 61% de los clientes exceptuando modelo lineal que tan solo acierta el 56%. El modelo log-normal obtiene el mayor nivel de acierto con un 64%, seguido por Poisson con un 63%. Alrededor del 2% de los clientes de este grupo son puestos en el segmento ranking 1 mostrando un muy buen nivel de distinción en que destaca el modelo log-normal con un 1%.

En conclusión:

Los modelos mostraron una alta capacidad de ajuste respecto a los ordenamientos de clientes, considerando los valores obtenidos dentro de sus correlaciones, tanto a nivel individual como de segmento. Además con los resultados de la matriz de confusión, denotaron un buen nivel de distinción entre “buenos” y “malos” clientes, destacando su alto porcentaje para los segmentos 1 y 4.

El modelo de Poisson presenta los mejores resultados a nivel global siendo su nivel de acierto el mejor en 3 de los 4 grupos de clientes y el segundo mejor del segmento restante.

El modelo log-normal presenta los peores resultados en general sobre todo para los grupos más valiosos, sin embargo, obtuvo mayor eficiencia para el segmento de menor valor.

El mejor modelo de este set de variables (Poisson) no supera los resultados obtenidos por parte de los mejores valores de los grupos anteriores, viéndose su correlación de ranking tanto a nivel individual como de segmento disminuida en un 1,5% aproximado con respecto al mejor modelo de los casos anteriores. Por lo tanto la inclusión de estas variables dentro de los modelos no tuvo un efecto significativo para aumentar el nivel de distinción. Cabe destacar sí, que el modelo B.N alcanza sus valores más altos, reflejándose esto en el aumento de su correlación de ranking de segmento en un 1,1% en comparación con el valor más alto obtenido en los casos anteriores.

Fase dos: Validación Coeficientes

Correlación de Spearman a Nivel Individual

Tabla 44: Ordenamiento individual fase dos

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,666	0,608	0,699	0,685
Significancia	0	0	0	0

Al validar los coeficientes dentro del período de validación, se ve una disminución cercana al 5% en relación a la fase anterior para todos los modelos, lo cual es un valor bajo. Además como era de suponer dado los resultados anteriores, el modelo de Poisson tiene la relación más alta con un 70%, seguido muy de cerca por el modelo lineal y binomial con un 66.6% y un 68,5% respectivamente.

Correlación de Spearman a Nivel de Segmento

Tabla 45: Ordenamiento por segmento fase dos

	Ordenamiento por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,617	0,565	0,651	0,636
Significancia	0	0	0	0

Las correlaciones bajan en un 5,5% promedio para todos los modelos. La mayor baja la tiene Poisson con un 6%. Por otra parte, los patrones se mantienen iguales que el ordenamiento anterior en que Poisson presenta los valores de correlación más altos con un 65% seguido por el modelo binomial con un 63,6%.

Matriz de confusión respecto a rankings por segmento

Tabla 46: Porcentaje de aciertos fase dos

		Porcentaje Acierto Modelos			
Ranking Real	Ranking Predicho	Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	62%	51%	64%	61%
	2	27%	28%	27%	27%
	3	9%	17%	8%	10%
	4	3%	5%	1%	2%
2	1	25%	34%	24%	26%
	2	38%	32%	39%	38%
	3	27%	25%	30%	27%
	4	10%	10%	7%	9%
3	1	9%	13%	8%	9%
	2	24%	29%	23%	25%
	3	36%	34%	38%	36%
	4	31%	23%	30%	30%
4	1	4%	3%	4%	4%
	2	11%	11%	11%	10%
	3	28%	24%	24%	27%
	4	56%	62%	61%	60%

Segmento Ranking 1:

Dentro del segmento de mayor valor, se aprecia que el nivel de acierto a disminuido en un 5% promedio con un peak de 6% dado por Poisson. De todas maneras el modelo de Poisson presenta los mejores resultados con un 63% de acierto, mostrando valores muy similares el modelo lineal y B.N. con un 62% y un 61% correspondientemente. Por otra parte, es relevante percatar que estos 3 modelos ubican más del 89% de estos clientes dentro de los dos segmentos de mayor valor y el modelo de Poisson mantiene un porcentaje de 1% de ubicación de este grupo en el segmento 4, por lo cual muestran un muy buen nivel de distinción de los clientes más valiosos. En último lugar está el modelo log-normal con solo un 51% de acierto.

Segmento Ranking 2:

Poisson obtiene los mejores resultados acertando el 39% de los clientes de este grupo. Además el modelo lineal y B.N. presentan valores muy similares. La disminución respecto a la fase anterior es de un 2,8% promedio en este caso con la mayor disminución dada por Poisson con un 5% de diferencia. Finalmente, el modelo log-normal mantiene los resultados más bajos.

Segmento Ranking 3:

Los patrones del segmento anterior siguen igual, en el cual Poisson tiene los mejores resultados a pesar de que su diferencia de eficacia disminuye en un 6% en comparación a la fase anterior. Los demás modelos tan solo disminuyen en un 2% promedio, lo que produce que tanto el lineal como el B.N. tengan valores similares a Poisson. Estos 3 modelos ubican un 54% aproximado de estos clientes cerca de la vecindad y un 8% dentro del grupo más valioso.

Segmento Ranking 4:

Excepcionalmente a los grupos anteriores el modelo log-normal obtiene los mejores resultados (62% acierto), seguido por Poisson y el binomial con valores muy similares 61% y 60% respectivamente. En último lugar está el modelo lineal con un 56% de acierto. La diferencia de eficacia en relación a la fase anterior es de tan solo un 1,3%. En el cual el máximo viene dado por Poisson con un 2,4%, un valor bajo si se compara con las disminuciones obtenidas anteriormente.

En resumen:

Los modelos excepto el log-normal, presentan una correlación en el ordenamiento individual y por segmento sobre el 60% respecto al ranking real lo que muestra buen nivel de distinción entre clientes, destacando el modelo de Poisson con los valores más altos en ambos casos. Esto queda confirmado al obtener los porcentajes de acierto más elevado mostrando así gran poder de diferenciación especialmente para los grupos 1 y 4.

Este conjunto de variables presenta una disminución de un 5% en cuanto a las correlaciones de ranking de la fase anterior, siendo esta la más elevada en comparación con las anteriores. La baja más notoria viene dada por Poisson con un 6% aproximado. En cuanto al nivel de acierto de grupos, la disminución más notoria viene dada en el segmento más valioso con un 5% promedio. Cabe destacar sí, que a pesar de esta disminución el porcentaje que traspasa de este grupo al segmento menos valioso es de tan solo un 0,4%.

El modelo lineal y B.N. tienen valores de acierto similares para los primeros segmentos, sin embargo, dentro de los segmentos de menor valor el modelo lineal pierde capacidad predictiva de ranking, mostrando por ende mayor error en comparación al modelo binomial negativo, lo cual produce que este se convierta en el segundo modelo con mejores resultados.

Nuevamente el modelo log-normal presenta los peores resultados a nivel general, pero muestra gran nivel de acierto para el grupo menos valioso obteniendo los valores más altos para este caso.

Fase tres: Validación Total (Variables Independientes obtenidas por medias móviles con ventana de tiempo de 12 meses)

Correlación de Spearman a Nivel Individual

Tabla 47: Ordenamiento individual fase tres

	Ordenamiento Individual			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,570	0,525	0,600	0,586
Significancia	0	0	0	0

El modelo de Poisson presenta la correlación más alta con un 60% acorde al ordenamiento real. Por otra parte tanto el modelo lineal como el B.N presentan valores muy similares. La disminución de la correlación en comparación a la fase anterior es de un 9,4% promedio mostrando la mayor disminución el modelo B.N con un 9,9%, sin embargo mantiene la segunda correlación más alta. Finalmente el modelo log-normal tiene el nivel de relación más bajo a pesar de que su disminución fue de un 8,5% respecto a fase dos.

Correlación de Spearman a Nivel de Segmento

Tabla 48: Ordenamiento por segmento fase tres

	Ordenamiento por Segmento			
	Ranking Lineal	Ranking Log-Normal	Ranking Poisson	Ranking Binomial Negativo
Ranking Real	0,523	0,490	0,563	0,549
Significancia	0	0	0	0

Los lugares dados por el nivel de correlación se mantienen, en otras palabras, el modelo Poisson obtiene los valores más altos (56%) seguido por el binomial y en último lugar el modelo log normal con un 49% de correlación. La disminución para el ordenamiento a nivel de segmento baja en un 8,6% promedio siendo el más afectado el lineal con un 9,4%. Además el modelo log-normal obtiene sus valores más altos en comparación con los niveles de correlación obtenidos con los otros conjuntos de variables, sin embargo el modelo con mejores resultados (Poisson) no supera su nivel de relación con respecto al caso RFM Simple, teniendo un disminución de un 0,34% y obtiene el mismo nivel que para el caso RFM Segregado.

A continuación se verán estos resultados reflejados en la matriz de confusión.

Matriz de confusión respecto a rankings por segmento

Tabla 49: Porcentaje de aciertos fase tres

Ranking Real	Ranking Predicho	Porcentaje Acierto Modelos			
		Lineal	Log-Normal	Poisson	Binomial Negativo
1	1	56%	48%	57%	55%
	2	26%	27%	27%	26%
	3	12%	17%	12%	13%
	4	6%	8%	4%	5%
2	1	27%	32%	26%	28%
	2	34%	31%	35%	34%
	3	26%	24%	28%	26%
	4	13%	13%	11%	12%
3	1	11%	15%	12%	11%
	2	25%	29%	24%	26%
	3	33%	31%	33%	33%
	4	31%	25%	31%	30%
4	1	6%	5%	6%	5%
	2	15%	13%	14%	13%
	3	29%	28%	27%	28%
	4	50%	55%	54%	53%

Segmento Ranking 1:

El modelo de Poisson presenta los mejores resultados para este grupo, tanto por nivel de acierto con un 57% como por el menor porcentaje de estos clientes ubicados el segmento menos valioso. Cabe destacar además que el 84% de estos clientes ranking 1 son ubicados dentro de los dos grupos de mayor valor. El modelo lineal presenta un número de acierto similar a Poisson (56%) sin embargo ubica un 6% de estos clientes en segmento 4. Por otra parte el modelo log-normal tiene los peores resultados de lo ya antes descrito. Ahora en comparación con la fase dos, hay una disminución en el acierto de un 5% siendo el mayor valor dado por

Poisson con un 6,5%. Por otra parte tan solo existe un aumento de error de un 3% al colocar estos clientes de mayor valor dentro del grupo 4.

Segmento Ranking 2:

Tal como en el segmento anterior, Poisson tiene el nivel predictivo más alto y ubica el 11% de estos en el grupo menos valioso. Cabe mencionar que el 54% de estos clientes Poisson los ubica dentro de la vecindad, esto como ya se ha explicado viene dado principalmente por la cercanía de los valores en los límites de un segmento y otro. El model lineal y B.N. presentan un nivel de distribución muy similar con un 34% de eficacia y ubicando el 54% de este grupo dentro de la vecindad. La disminución del acierto es de un 3% respecto a la fase anterior, destacando que el modelo log-normal tan solo disminuye un 0,3%

Segmento Ranking 3:

Para este grupo, no existe mucha diferencia entre los distintos modelos exceptuando el log-normal que presenta los más bajos resultados. Existe una leve diferenciación entre el Poisson y el restante puesto que los últimos ubican un 1% menos de este grupo en el segmento de mayor valor.

Segmento Ranking 4:

En general todos los modelos presentan un muy buen nivel de distinción prediciendo correctamente más del 50% de este grupo y el 80% aproximadamente es ubicado dentro del grupo de menor valor. En promedio el 5,5% de este grupo es colocado en el segmento 1, mostrando así buen nivel de distinción. Además, tal como se ha dado en las fases anteriores referentes a este grupo, el modelo log-normal obtiene los mejores resultados con un 55% de acierto seguido por Poisson con un 54%. La disminución de efectividad es de un 6,7% respecto a la fase dos, mostrando la mayor disminución el modelo log-normal con un 7,5%. Cabe destacar que el aumento de error al ubicar estos clientes de menor valor dentro del grupo ranking 1 es de un 1,6% promedio.

En conclusión:

Las correlaciones bajan en un 9,4% para el caso del ordenamiento individual y un 8,6% respecto al ordenamiento de segmento. Sin embargo, los niveles de relación aún son altos en que Poisson presenta la correlación más alta en ambos casos con un 60% en el orden individual.

En comparación con los resultados de los otros conjuntos de variables, el log-normal presenta sus valores más altos. Por otra parte el modelo lineal presenta un baja de un 1,1% respecto al mejor valor antes alcanzado en la correlación de segmento y los modelos de Poisson y B.N. no superan los niveles de relación antes alcanzados aunque si igualan sus mejores resultados. Por ende no existe mejora con este set de variables para estos últimos modelos señalados.

En cuanto a los niveles de predicción, si bien existe una baja de acierto significativa respecto a la fase anterior con un 6,7% en promedio dado en el segmento 4 y un 5% en el grupo 1 producto de la obtención de los valores de las variables independientes mediante medias móviles, los modelos en general aún mantiene un buen nivel de distinción entre “buenos” y “malos” clientes, destacando a Poisson con los valores de predicción más altos sobre todo para los grupos 1 y 4 con un 57% y 54% respectivamente. Por otra parte el aumento dado en los grupos antagónicos a estos es de tan solo un 2,4%.

Durante todo el proceso el modelo lineal y el B.N. presentaron valores muy similares para los grupos de mayor valor, sin embargo comenzaban a diferenciarse a medida que avanzaban a los segmentos de menor valor, mostrando mejores resultados por parte del B.N. que se ve reflejado en su mayor nivel de correlación por segmento. Por su lado, el modelo log-normal presenta los peores resultados a nivel general, mostrando mayor eficacia en el grupo de menos valor.

El nivel de predicción respecto a los otros conjuntos de variables, es muy similar siendo este un tanto menor para el grupo más valioso pero mayor para el segmento de menos valor, es decir, existe cierto trade off entre este conjunto de variables y los antes analizados.

Conclusiones generales de los resultados:

Para cada uno de los conjuntos de variables estudiados el modelo de Poisson obtuvo los resultados más sobresalientes, tanto para las correlaciones de ordenamiento individual y de segmento como porcentaje de acierto de ubicación correcta de los clientes. Por lo tanto este modelo mostró la mejor distinción entre “buenos” y “malos” clientes.

El error asociado al pasar de la fase uno a la fase dos en el caso de las correlaciones (en cuanto disminuye el nivel de correlación) es de un 4,6% promedio, mostrándose la mayor baja para el conjunto de variables RFM Full con un 5,1%. En el caso particular el modelo de Poisson en comparación con el resto de los modelos, muestra la mayor baja de correlación con un 5,4% en el cual la mayor disminución fue dada por el conjunto RFM Full con un 5,8%.

El error asociado referente a la disminución del porcentaje de acierto al pasar de la fase uno a la siguiente, es de un 4,1% para el segmento ranking 1, 2,4% para el grupo 2, 3,2% segmento 3 y finalmente un 1,1% para el grupo ranking 4. Específicamente el error dado en el modelo de Poisson viene dado por una disminución en el acierto de un 5,9% para el segmento 1, 5,7% grupo 2, 7,8% grupo 3 y finalmente un 2,2% en segmento ranking 4. Hay que destacar que la baja de aciertos en los segmentos 1 y 4 no se refleja en un aumento de sus grupos

opuestos¹⁹ puesto que el aumento de estas es mínima con un 0,2% y 1,6% respectivamente.

Al reemplazar los valores reales de las variables independientes por medias móviles, es decir, pasar de la fase dos a la tres, se ve un aumento del error, al disminuir en el caso de las correlaciones individuales en un 9,8% y en un 9% para el ordenamiento por segmento. El conjunto de variables que obtuvo la mayor baja fue el RFM Simple con un 10,3% y 9,2% correspondientemente. Con el mismo patrón, el modelo de Poisson muestra su mayor disminución con el conjunto de variables RFM Simple, con un 11,2% en el caso de ordenamiento individual y de un 10,3% en el ranking por segmento y los menores valores alcanzados están dados por el set RFM Full con un 9,8% y un 8,8% respectivamente. Por lo tanto las variables incluidas tanto demográficas como de la tarjeta permiten que el error disminuya al pasar a la última fase de validación.

Durante el período de la fase de calibración, los mejores resultados fueron dados por el modelo de Poisson en el conjunto RFM, sin embargo este disminuye en mayor medida al utilizar medias móviles en la fase tres. Por otra parte el conjunto RFM Full, es el que más disminuye sus valores al pasar a la fase dos, producto de utilizar los datos de validación con los coeficientes obtenidos en la etapa de calibración, pero hay que tener en cuenta que era el que menor valor de correlación tiene en la fase uno. Lo que finalmente produce que no haya mayor diferenciación en cuanto a los valores entregados por cada uno de los modelos de Poisson en la fase final.

A continuación se muestra una comparación de los resultados obtenidos para los ordenamientos de cada fase referente a los modelos de Poisson y luego una comparación más detallada de estos modelos con los valores obtenidos en la matriz de confusión con los distintos conjuntos de variables estudiados.

7.4.3 Mejores modelos por conjuntos de variables

A continuación se comparan los resultados obtenidos para los mejores modelos obtenidos para cada conjunto de variable, es decir, los modelos de Poisson.

Resultados obtenidos para los ordenamientos individuales y por segmento:

Tabla 49: Resumen ranking individual

Correlaciones	Ordenamiento Individual		
	Fase 1	Fase 2	Fase 3
RFM Simple	0,770	0,715	0,603
RFM Segregado	0,763	0,712	0,600
RFM Full	0,757	0,699	0,600

Tabla 50: Disminución correlación ranking individual

Diferencia	Ordenamiento Individual	
	Fase 1 a Fase 2	Fase 2 a Fase 3
RFM Simple	5,5%	11,2%
RFM Segregado	5,1%	11,2%
RFM Full	5,8%	9,9%

¹⁹ El grupo opuesto del segmento 1 es el grupo 4 y el del 4 es el segmento 1

Tal como ya se ha comentado, el modelo RFM Simple comienza con los mejores resultados, sin embargo ya llegando a la fase 3, este se equipara con el resto al tener la segunda disminución más alta al pasar a la fase 2 y la más alta al llegar a la 3, mostrando si una mínima ventaja de un 0,3%. Por otro lado el conjunto RFM Full tiene la baja más alta al pasar a la fase dos, pero sin embargo presenta la disminución menor al continuar a la fase siguiente.

Tabla 51: Resumen ranking por segmento

Correlaciones	Ordenamiento por Segmento		
	Fase 1	Fase 2	Fase 3
RFM Simple	0,726	0,672	0,566
RFM Segregado	0,720	0,666	0,563
RFM Full	0,711	0,651	0,563

Tabla 52: Disminución correlación ranking por segmento

Diferencia	Ordenamiento por Segmento	
	Fase 1 a Fase 2	Fase 2 a Fase 3
RFM Simple	5,4%	10,6%
RFM Segregado	5,4%	10,3%
RFM Full	6,0%	8,8%

En cuanto al ordenamiento por segmento pasa lo mismo que en el caso anterior, aunque el conjunto RFM Simple presenta la mayor baja al pasar a la fase 3, sin embargo disminuye lo mismo que el grupo RFM Segregado en la fase 2. A pesar de lo anterior sigue obteniendo una correlación levemente mayor que el resto, siendo esta diferencia de un 0,3%.

Comparación resultados mejores modelos obtenidos incluyendo método utilizado por la empresa:

A continuación se analizan las distribuciones entregadas por los modelos de Poisson por cada conjunto de variable a la cual se incluyen los valores obtenidos al utilizar el método del área de la empresa. Para esto se explicará brevemente en qué consiste.

Para estimar que clientes serán los más valiosos en el período siguiente, se basan en el número de transacciones realizadas por cada uno de ellos con la tarjeta de crédito de la firma en el período anterior. Entonces para comparar los valores obtenidos con este método con los mejores modelos por cada set de variables, se ordenó a los clientes por número de transacciones respecto al período "T" (período de calibración) y se obtuvo un ranking donde el cliente más valioso para el período siguiente (en este caso el de validación) sería el que tuvo mayor número de transacciones el período anterior.

Ahora al analizar las distribuciones entregadas por los modelos de Poisson por cada set de variables y el método usado actualmente se obtiene lo siguiente:

Tabla 36: Resultados modelos de Poisson

		Modelos Poisson			
Ranking Real	Ranking Predicho	RFM Simple	RFM Segregado	RFM Full	Transacciones
1	1	58,40%	57,80%	56,90%	46,00%
	2	25,30%	25,80%	27,20%	26,98%
	3	11,20%	11,40%	11,50%	19,59%
	4	5,10%	5,00%	4,30%	7,43%
2	1	25,00%	24,90%	25,70%	30,53%
	2	37,50%	37,80%	34,80%	32,08%
	3	27,10%	27,00%	28,40%	25,75%
	4	10,40%	10,30%	11,10%	11,64%
3	1	10,80%	10,80%	11,60%	16,79%
	2	24,70%	24,30%	24,50%	30,45%
	3	36,20%	36,20%	33,30%	33,31%
	4	28,30%	28,60%	30,60%	19,46%
4	1	5,80%	6,40%	5,80%	11,60%
	2	12,50%	12,10%	13,50%	19,55%
	3	25,50%	25,40%	26,80%	38,80%
	4	57,00%	56,10%	54,00%	30,05%

A continuación se analizan los resultados de la tabla anterior por segmento.

Segmento Ranking 1:

Dentro del grupo de los clientes más valiosos, se destaca el mayor porcentaje de acierto dado por el conjunto RFM Simple con un 58,4%, sin embargo ubica el 5,1% de estos clientes en el segmento menos valioso. El segundo mejor resultado es entregado por el set RFM Segregado con un 57,8% y ubica un 5% de este grupo en el segmento 4. El grupo RFM Full tiene el menor porcentaje de acierto aunque hay que destacar que ubica el menor número de clientes en el grupo 4 siendo este un 4,3%. Finalmente el método usado actualmente “transacciones” presenta tan solo un 46% de acierto, lo cual es un 12,5% menos que el mejor valor encontrado. Además ubica el 25% de este de mayor valor dentro de los dos segmentos menos valiosos, presentando así la peor distinción para este segmento.

Segmento Ranking 2:

Los conjuntos RFM Simple y Segregado obtienen los mejores resultados para este segmento con un 37,5% aproximado de aciertos y un 52% de este grupo dentro de la vecindad. Además ubican el menor porcentaje de este grupo en el segmento ranking 4. A diferencia del grupo anterior, el set RFM Full coloca un mayor número

de estos clientes dentro del segmento de menos valor. Por otro lado el método “transacciones” presenta valores similares al resto de los modelos, sin embargo, tiene un 6% promedio menos de nivel de acierto en comparación a los otros modelos.

Segmento Ranking 3:

Se vuelven a repetir los resultados anteriores, siendo los conjuntos RFM Simple y Agregado los que presentan el mayor número de acierto y menor porcentaje de colocación dentro del grupo más alejado (en este caso segmento 1). Por otra parte, el “transacciones” muestra un bajo nivel de distinción para este segmento puesto que ubica el 47% de este grupo dentro de los dos segmentos más valiosos donde el 17% de estos pertenecen al grupo ranking 1.

Segmento Ranking 4:

Para este grupo se nota una diferencia, el cual favorece al conjunto RFM Simple con el mayor número de aciertos con un 57% en contra de RFM Segregado que obtiene un 56% mientras que RFM Full un 54%. Además RFM Simple ubica el menor porcentaje de este grupo dentro del segmento más lejano (Ranking 1) con un 5,8% al igual que RFM Full y por otro lado RFM Segregado coloca un 6,4% de estos clientes menos valioso en el grupo de mayor valor. Finalmente “transacciones” tan solo tiene un 30% de acierto y además ubica el mayor porcentaje de este grupo dentro del segmento de mayor valor. Por lo tanto su nivel de distinción es el peor en comparación a los otro modelos

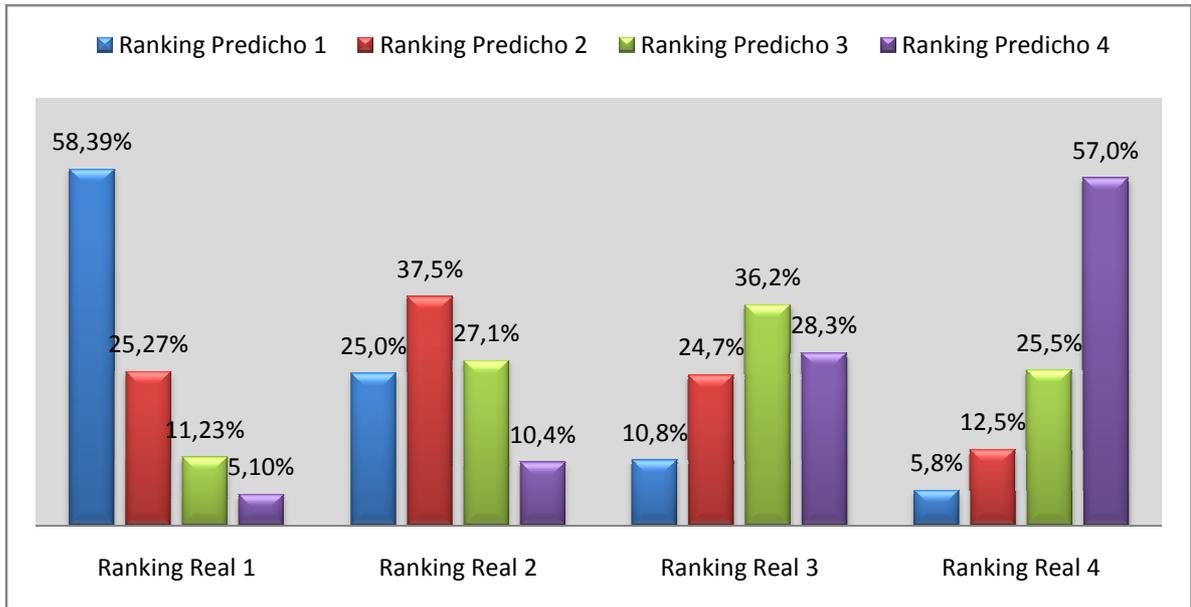
En conclusión:

Al analizar en mayor detalle los mejores resultados encontrados para los distintos grupos de variables, se concluye que el modelo de Poisson con el conjunto RFM Simple muestra los mejores resultados tanto para las correlaciones individuales y por segmento como para la distribución de clientes analizada en la matriz de confusión, en que a modo general presenta el mayor número de acierto y el menor error al colocar a los clientes de un grupo dentro del segmento más alejado. Por lo tanto la división de la frecuencia y recency a nivel de holding o la inclusión de variables demográficas o referentes a la tarjeta de crédito no tuvieron mayor efecto.

Hay que destacar además que los modelos de Poisson para cada conjunto de variables presentaron mejor nivel predictivo que el actualmente utilizado por la empresa, mostrando este último un bajo nivel de distinción especialmente para los segmentos de menor valor.

A continuación se muestra gráficamente los resultados de la fase final del mejor modelo encontrado.

Gráfico 23: Modelo Poisson RFM simple



7.5 Conclusiones

En el desarrollo del presente trabajo se han analizado distintos tipos de modelos regresivos con una combinación de conjuntos de variables diferentes con el objetivo de encontrar el mejor modelo para estimar el lifetime value de cada cliente respecto al interés que le pagaría a la empresa en el futuro teniendo en cuenta el riesgo asociado a este obtenido por el modelo de riesgo utilizado por la compañía y así poder distinguir entre los “buenos” y “malos” clientes.

7.5.1 Conclusiones Generales

Al terminar el trabajo y analizar los distintos resultados obtenidos, queda la satisfacción de que estos modelos de LTV pueden llegar a ser una poderosa herramienta para la toma de decisiones, principalmente estratégicas. Ya que le da a la empresa estimaciones que le permitirá reconocer a gran parte de sus “mejores” clientes apoyándose en una base sólida, permitiendo de esta forma a la empresa llegar a realizar esfuerzos de marketing más efectivos lo que entregaría finalmente mejores resultados económicos.

Se ha sido capaz de encontrar un modelo con la capacidad de estimar el valor futuro del cliente referente al interés que le pagará a la compañía producto de sus transacciones con la tarjeta de crédito propia de la firma. La empresa actualmente no contaba con un modelo de estas características por lo que ahora puede enfocar sus esfuerzos de marketing a clientes que sí entregan valor a la empresa, lo cual puede llegar a afianzar aún más sus lazos en el largo plazo y al mismo tiempo reducir sus costos de marketing directo al no gastar parte de estos recursos en clientes que finalmente no entregan valor consecuente con este enfoque.

Los modelos de LTV estudiados utilizando diversos conjuntos de variables si bien muestran un bajo nivel de estimación del valor real del cliente especialmente a nivel individual, no tienen el fin de estimar este valor sino más bien de distinguir a los “mejores” clientes de los menos valiosos y así ayudar a la toma de decisiones. Lo anterior no hubiera sido posible sin antes aplicar un paso previo: La limpieza de datos. Las bases de datos entregadas por la empresa con las cuales se realizó esta memoria contaban con un gran número de errores. Por ejemplo existían variables que estaban dentro de la base y eran usadas para diversos propósitos, pero sin embargo no tenían datos correctos (datos incongruentes) o la variable ya no era actualizada a los valores actuales. Además existían clientes con información incompleta dentro de variables en que su valor era necesario. Por lo tanto la selección de variables y clientes para el trabajo fue mediante un proceso muy minucioso el cual fue realizado a través de un análisis exhaustivo de esta. Es relevante recalcar que la nobleza de los resultados depende mucho de la calidad de los datos utilizados y por lo tanto la explicación del porqué del esfuerzo y tiempo en esta sección.

Dentro de los clientes hay un grupo que claramente destaca sobre el resto siendo su valor significativamente superior. Más específicamente el 25% de los clientes más valiosos entregan el 80% del valor total a la firma en que su valor promedio anual por cliente trasciende a los \$256.000 mientras que en contraste el 25% de los clientes de menor valor tan solo alcanzan un valor promedio anual de \$ 1.640. Cabe destacar además que el 58% del segmento de clientes más valiosos se mantienen dentro de los más valiosos en el próximo período, mientras que el 23% de estos pasan al segundo segmento de mayor valor. Si bien no se mantiene el 100% de los clientes, existe de todas maneras cierta estabilidad al permanecer la gran parte de los clientes dentro del grupo de los más valiosos en el siguiente período.

Existe una tendencia general a que los clientes clasificados como más riesgosos entreguen poco valor a la compañía y por ende estar dentro de los segmentos de menos valor. Sin embargo esto no es así para los clientes de menor riesgo (Scoring 1 y 2) ya que estos se encuentran mayormente distribuidos en los distintos segmentos de valor, en otras palabras, no existe mayor relación entre los clientes más valiosos y su riesgo asociado.

Finalmente respecto a los objetivos propuestos en esta memoria, existe gran satisfacción con el trabajo realizado, ya que a pesar de todas las dificultades y complicaciones por las cuales se tuvo que pasar, se logró cumplir con cada uno de los objetivos planteados.

A continuación se presentan las conclusiones referentes a los resultados obtenidos por los distintos modelos y conjuntos de variables estudiados en sus distintas etapas de análisis.

7.5.2 Conclusiones respecto a los modelos

El modelo de Poisson se mantuvo como el modelo con los mejores resultados para todo conjunto de variables y para cada etapa. Esto fue dado tanto para ordenamiento individual y de segmento, como de distinción entre los “buenos” y “malos” clientes al ver los valores entregados dentro de la matriz de confusión. Si bien sus valores no fueron de un 100% de acierto, sus porcentajes fueron muy altos particularmente para el caso de los segmentos 1 y 4. Sobre todo si se considera la gran volatilidad asociada entre una compra y otra para el pago del interés, puesto que esta depende de una gran cantidad de factores como lo son el holding en que se compra, el número de cuotas con el cual se realiza, el monto a pagar con la tarjeta, el departamento en que se realiza la transacción y los valores de mercado. Por lo tanto el modelo de Poisson mostró muy buenos resultados al predecir gran parte de los clientes dentro de su segmento correspondiente en cada conjunto de variables.

El modelo que obtuvo los peores resultados fue el log-normal para cada conjunto de variables, sin embargo muestra buen nivel de distinción para los segmentos de menos valor.

Al comparar los valores obtenidos por el modelo de Poisson respecto a los distintos conjuntos de variables utilizados no hubo gran diferencia entre uno y otro, sin embargo hay una leve diferencia entre estos, en que se destaca el mayor porcentaje de acierto de los clientes más valiosos sin aumentar la colocación incorrecta de éstos dentro de los segmentos de menor valor usando el conjunto de variables RFM Simple. Por lo tanto, el hecho de haber dividido las frecuencias y los recency de compra por holding no tiene significancia para los valores entregados para el modelo de Poisson. De hecho la correlación de rankings disminuye levemente y así también el porcentaje de acierto en cada etapa de estudio. Luego al incorporar variables demográficas y otras referentes a la tarjeta, tampoco tuvo efecto en comparación con el uso de las variables RFM Simple, de hecho, con la inclusión de estas variables se alcanzó el mayor nivel de desacierto aunque se recalca nuevamente que este fue una disminución muy baja frente a las demás y su capacidad de ranking tuvo el mismo nivel que RFM Segregada. Sin embargo hay que destacar que al utilizar estos conjuntos de variables, algunos modelos alcanzan valores más altos respecto al conjunto RFM Simple. Específicamente, el modelo lineal y binomial negativo alcanzan sus valores más altos con el conjunto RFM Segregado mientras que el modelo log-normal con el conjunto RFM Full. Por lo tanto, el modelo de Poisson utilizando el conjunto de variables RFM Simple se considera el mejor modelo al obtener los mejores resultados.

A nivel global el aumento del error al pasar de la fase uno a la dos es relativamente bajo mientras que al pasar de la fase dos a la tres este aumenta al doble. Particularmente el modelo de Poisson el que obtiene los mejores resultados en cada modelo y cada etapa, presenta una disminución de la corrección de ordenamiento individual de un 5,4% promedio al pasar de la fase uno a las dos y de un 5,6% para el caso del ordenamiento de segmento alcanzando sus valores más altos con el conjunto RFM Full. Por otra parte esta, disminución es de un 10,8% al pasar de la fase dos a la tres en el ordenamiento individual y de un 9,9% para el ranking por segmento, alcanzando sus valores más altos para el set RFM Simple y Segregado con un 11,2%. Por ende, se ve una disminución en la última transición de casi el doble que la entregada por parte de la primera transición (fase uno a dos), producto del reemplazo de los valores reales de las variables independientes por sus valores obtenidos mediante medias móviles. Esto se ve reflejado en la disminución de aciertos en los clientes de cada segmento y la mayor colocación de clientes de un grupo dentro del segmento más alejado a ese (fuera de la vecindad). Si bien lo anterior era de esperarse hay que destacar que la disminución no es elevada considerando que son valores a predecir en el futuro, en otras palabras, la existencia de incertidumbre sobre su valor real. Mostrando así resultados satisfactorios.

7.5.3 Estudios futuros y propuestas de mejora

En esta sección se plantearán diversas propuestas y estudios encontrados a lo largo de la memoria.

Primero, se detectó al realizar el trabajo que existe un gran número de clientes que ya no realiza transacciones dentro de la empresa, o en otras palabras, clientes fugados. Sería interesante para la empresa examinar el porqué de esta situación y qué hacer para revertirlo. Para esto podría realizarse un estudio cualitativo, por cual habría que salir a terreno y aplicar encuestas, entrevistas o focus groups.

Otro punto importante es referente a la base de datos. Es de suma relevancia que se realice un levantamiento en el data warehouse y limpiar las bases completamente ya que durante el trabajo se encontraron una serie de errores que dificultaron la realización del trabajo. Por ejemplo se deberían sacar los clientes con la cuenta ya cerrada por largo tiempo, además revisar cuidadosamente las variables y eliminar las que no son usadas, ya que se hallaron en el estudio variables desactualizadas con datos erróneos que la compañía seguía utilizando sin saber de esto. Además revisar de porqué hay tantos datos outlayers dentro de las bases de ventas lo cual no es normal siendo que es un proceso que debiera mantener las variables con valores reales y no fuera de rango o con signos cambiados, puesto que son variables muy importantes para conocer gran cantidad de cosas relacionadas a las ventas. Finalmente referente a las bases de datos esta el tema de la coordinación entre áreas. Esto pues, las distintas áreas van incorporando variables dentro de la base principal sin saber el significado de estas por parte de las otras áreas las cuales pueden llegar a ser de gran utilidad.

Con respecto al tema desarrollado existen varios estudios que se pueden realizar para completar y sacar mayores beneficios para la firma los cuales son descritos a continuación.

Para la mejora del modelo se podría estudiar el mismo conjunto de variables pero con el uso de información agregada para cuando se cuente con mayor cantidad de períodos en la base de datos. Dado que al analizar los resultados con distintos conjuntos de variables se ha concluido que el uso de variables tanto demográficas como de la tarjeta de crédito no mejoran los resultados o el nivel de distinción de clientes. Sin embargo, al utilizar información agregada como por ejemplo en vez de utilizar un período de 1 mes se utilice un período de 3 meses puede que el nivel de ajuste de los modelos mejore y por lo tanto los resultados ya que se nota un aumento en la correlación del interés a mayor número de períodos.

Para complementar el trabajo, se puede estudiar qué tipo de productos son los más demandados o que más le interesan a los clientes más valiosos. Para así no solo focalizar los esfuerzos de marketing en estos clientes con productos y ofertas irrelevantes para ellos sino que ofrecerles cosas que para ellos sean de interés lo cual produciría un mayor incentivo en estos clientes a gastar más con el uso de la

tarjeta en los productos ofrecidos por la compañía, aumentando de esta manera la satisfacción del cliente y las utilidades de la empresa.

Finalmente recordando la motivación de esta memoria donde estaba el gráfico de LTV vs Riesgo, es estudiar cómo mantener a los clientes más valiosos dentro de su cuadrante. Mantener a los clientes más valiosos dentro de la compañía es una de las principales bases para que una empresa sea rentable y crezca en el tiempo, por lo tanto motivar y crear lazos a largo plazo con estos clientes es fundamental. Por otro lado es de suma importancia estudiar también como hacer que los clientes del cuadrante superior izquierdo (tienen un LTV alto pero son riesgosos) pasen al otro cuadrante. Ya que si bien son clientes que pagan mucho interés a la compañía, existe la probabilidad de que finalmente no cumplan con sus pagos lo cual sería perjudicial para la firma, sin embargo, si se halla la manera de que aseguren o disminuyan la probabilidad de que estos clientes no paguen el efecto sería el contrario y la empresa lograría aumentar su rentabilidad a lo largo del tiempo.

8. Bibliografía

- [1] M. Haenlein, A. Kaplan, A. Beeser. A model to determine customer lifetime value in retail banking context. *European Management Journal* Vol. 25, 221-234. 2007.
- [2] S.Gupta, D. Hanssens B. Hadie. Modeling Customer Lifetime Value. *Journal of Service Research*. Vol. 9,134-155. 2006.
- [3] R. Colombo, W. Jiang. A stochastic RFM model. *Journal of Interactive Marketing*, Vol. 13. 1999.
- [4] P. Berger, N. Eechambadi, M. George. From Customer Lifetime Value to Shareholder Value. *Journal of Service Research*, Volume 9, 156-167. 2006
- [5] F. Wangenheim, P. Lentz. Applying Financial Risk and Volatility Measures to Customer Segmentation and Risk-Adjusted Lifetime Value Determination. 2005.
- [6] D. Gujarati. *Principios de Econometría*. Cuarta Edición, 2004
- [7] H. Greene. *Análisis Econométrico*. Tercera Edición, 1999.
- [8] W. Reinartz, V. Kumar. The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration. 2001. jcmc.indiana.edu/vol9/issue4/lawson_body.html
- [9] G. Dallal. Negative Binomial and Poisson regression; Testing For Overdispersion in Poisson regression. 2008.
- [10] H. Bauer, M. Hammerschmidt, M. Braehler. The customer lifetime value concept and its contribution to corporate valuation.
- [11] T. Panda. *Creating Customer Life Time Value through Effective CRM in Financial Services Industry*. 2004.
- [12] S. Gupta, D. Lehmann. *Managing Customers as Investments*. Cuarta Edición, 2008.

9. Anexos

Anexo 1: Rango Score Clientes

Descripción de los puntajes asociados a cada nivel de rango.

Tabla 50: Rango Scoring

Rango Score	Puntaje
4	0
3	1-299
2	300-449
1	450-999

Anexo 2: Tendencia de Interés

Tal como en el caso del Holding 1, se puede apreciar la tendencia en aumento para el Interés aunque en distintas magnitudes dependiendo del Holding en que se realizan las transacciones.

Gráfico 24: Evolución Interés para Holding 1

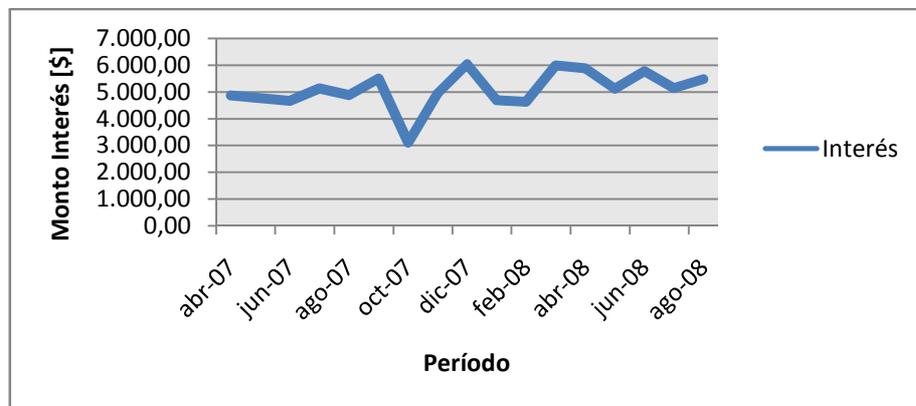


Gráfico 25: Evolución Interés para Holding 2

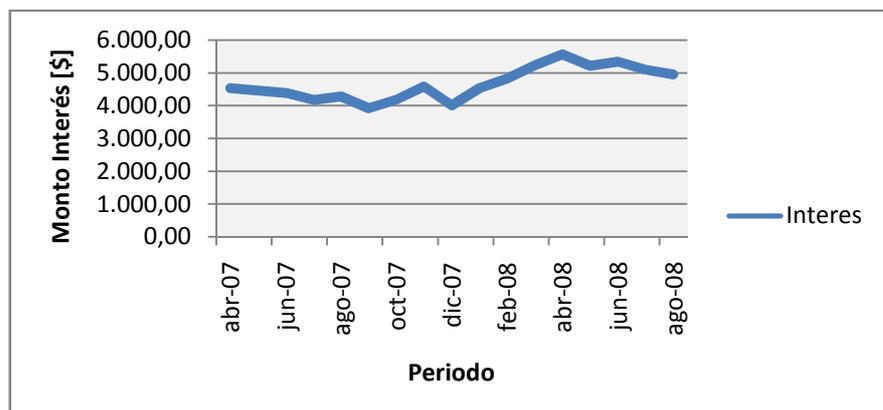
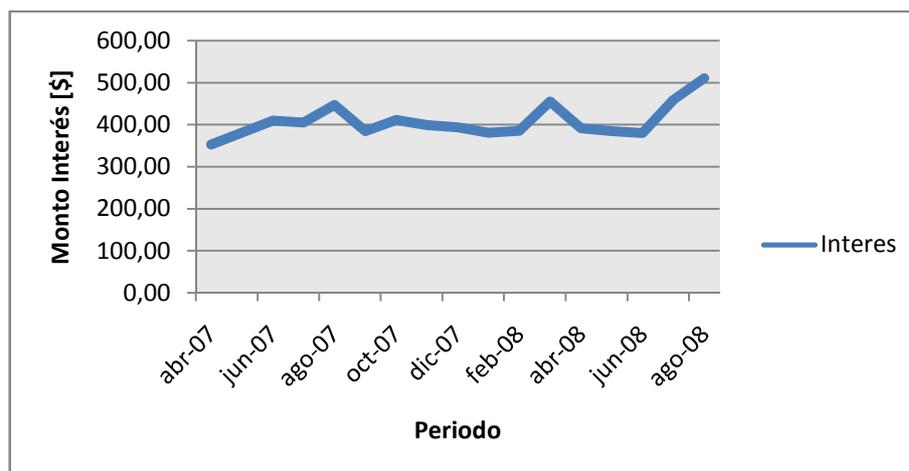


Gráfico 26: Evolución Interés para Holding 3



Anexo 3: Matriz de Transición LTV

Transición cada 4 Meses

Tabla 54: Dinamismo LTV para 4 Meses

Matriz Transición LTV 4				
	Ranking 1	Ranking 2	Ranking 3	Ranking 4
Ranking 1	49%	23%	14%	14%
Ranking 2	25%	35%	25%	16%
Ranking 3	14%	25%	38%	22%
Ranking 4	12%	17%	23%	48%

El número de clientes que se mantiene en el mismo segmento disminuye dado que en este caso un período equivale a 4 meses.

Anexo 4: Ajuste por R²

A continuación se señalan los valores de R² obtenidos para el modelo lineal y log-normal para cada conjunto de variables.

RFM Simple:

Tabla 51: Bondad de Ajuste Modelo Lineal

Model Summary				
Modelo	R	R Square	Adjusted R Square	Std. Error of the Estimate
	0,173	0,035	0,035	40576,259

Tabla 52: Bondad de Ajuste Modelo Log-Normal

Model Summary				
Modelo	R	R Square	Adjusted R Square	Std. Error of the Estimate
	0,431	0,186	0,186	3,781

RFM Segregado:**Tabla 53: Bondad de Ajuste Modelo Lineal**

Model Summary				
Modelo	R	R Square	Adjusted R Square	Std. Error of the Estimate
	0,249	0,062	0,062	40507,106

Tabla 54: Bondad de Ajuste Modelo Log-Normal

Model Summary				
Modelo	R	R Square	Adjusted R Square	Std. Error of the Estimate
	0,445	0,198	0,198	3,754

RFM Full:**Tabla 55: Bondad de Ajuste Modelo Lineal**

Model Summary				
Modelo	R	R Square	Adjusted R Square	Std. Error of the Estimate
	0,254	0,065	0,064	40455,229

Tabla 56: Bondad de Ajuste Modelo Log-Normal

Model Summary				
Modelo	R	R Square	Adjusted R Square	Std. Error of the Estimate
	0,450	0,203	0,203	3,742

Anexo5: Ordenamiento de clientes a nivel individual y por segmento

A continuación se muestra un ejemplo de cómo es ordenado cada cliente dado su valor obtenido por nivel individual y de segmento. Luego dado estos resultados se saca la correlación existente entre los valores reales y predichos mediante la correlación de spearman.

Tabla 55: Ejemplo Tabla Rankings por Valor

ID CLIENTE	LTV Real	LTV Predicho	Ranking Individual Real	Ranking Individual Predicho	Ranking Segmento Real	Ranking Segmento Predicho
79151191	851197	564529	178	46	1	1
60284343	845948	498403	181	68	1	1
32853212	591402	359355	336	180	1	1
36930922	591091	76305	337	5095	1	2
87807181	498307	100226	431	3457	1	2
46266451	457974	192278	499	1004	1	1
73026121	397858	179324	601	1156	1	1
62195063	370584	327633	657	240	1	1
39648651	295511	99601	878	3491	1	2
38146332	289423	124324	895	2390	1	1
53809448	188420	174076	1433	1241	1	1
46547207	187049	171483	1441	1291	1	1
51110101	186528	107476	1449	3099	1	1
95355820	125614	97606	2136	3619	1	2
90677965	109729	157010	2423	1553	1	1
87143210	92643	81417	2799	4682	1	2
21065051	86688	211803	2937	793	1	1
24766770	83877	151344	3016	1660	1	1
96631031	82456	94628	3077	3793	1	2
47465110	81423	68846	3104	5744	1	2
77565369	77033	142639	3264	1858	1	1
76979763	75158	84936	3324	4411	2	2
42052381	66612	173651	3643	1250	2	1
38749844	66460	61404	3649	6512	2	2
65242058	66075	78422	3664	4927	2	2
99208172	61191	55040	3855	7244	2	3
89354153	57353	83990	4025	4485	2	2
43193780	48249	87047	4547	4253	2	2
74832170	48214	66872	4552	5927	2	2
88888016	47415	101403	4603	3395	2	2
95537463	32072	43026	5753	8775	2	3
42301336	31947	71141	5771	5531	2	2
37587221	29124	50950	6028	7752	2	3
76753600	28390	85733	6090	4354	2	2
110647250	25405	65789	6439	6024	2	2
39481897	24530	62875	6551	6342	2	2
76790112	23791	54714	6651	7276	3	3
49759115	23598	54756	6673	7269	3	3
107151373	22508	52300	6813	7588	3	3
53422390	21329	41366	6987	9002	3	3
110599286	19888	43232	7170	8758	3	3
23401567	19805	78061	7182	4952	3	2

44699508	19269	28732	7251	10619	3	4
61264651	19112	52504	7275	7555	3	3
105482825	17954	36044	7445	9694	3	3
41919633	17798	23059	7466	11318	3	4
108677758	17737	51524	7474	7686	3	3
48598396	17337	65876	7537	6012	3	2
85687847	17313	64506	7544	6163	3	2
80433717	17060	4227	7588	12954	3	4
109655238	17035	32417	7596	10166	3	4
36370740	15579	69028	7851	5731	3	2
68890017	14415	71157	8054	5529	3	2
75992679	13353	43527	8225	8710	3	3
95826500	13298	168914	8236	1339	3	1
88815811	9914	52716	8858	7529	3	3
78836037	9835	40470	8874	9127	3	3
50230040	9747	37153	8895	9547	3	3
54450487	9358	67924	8980	5838	3	2
87807006	9251	31094	9003	10314	3	4
19649637	6646	28921	9720	10598	3	4
88010797	6444	41789	9780	8946	3	3
81284420	5887	13644	9952	12279	4	4
37677248	5482	35154	10085	9820	4	3
56559671	4923	6812	10273	12773	4	4
42556381	4343	22460	10451	11383	4	4
24339602	4169	26714	10519	10868	4	4
97222349	3724	37954	10682	9455	4	3
90905052	3577	15284	10744	12120	4	4
49013433	3304	24563	10857	11144	4	4
112521255	2980	34202	10991	9943	4	4
41794781	2603	18178	11101	11837	4	4
61103357	2073	38999	11317	9315	4	3
85388555	1631	28836	11500	10604	4	4
76511729	1291	42462	11663	8863	4	3
25681691	1095	31310	11759	10288	4	4
36151486	871	11466	11869	12455	4	4
56028686	682	22899	11970	11337	4	4

Anexo 5: Obtención coeficientes con todas las variables por junto de variables

Conjunto de Variables RFM Simple

- Modelo lineal

Coeficientes						
Modelo		Coeficientes		Coeficientes Estandarizados	t	Sig.
		B	Estad. Error	Beta		
	Constante	2887,327	442,574		6,524	0
	Frecuencia	7897,265	463,035	0,054	17,055	0
	Recency	-11,3	39,043	0	-0,289	0,772
	Interés Promedio	0,443	0,008	0,149	57,434	0
	Mes2	-1573,709	499,191	-0,011	-3,153	0,002
	Mes3	-877,72	498,892	-0,006	-1,759	0,079
	Mes4	-2105,779	498,678	-0,014	-4,223	0
	Mes5	-2507,019	498,5	-0,017	-5,029	0
	Mes6	-1805,338	498,395	-0,012	-3,622	0
	Mes7	-1353,782	498,344	-0,009	-2,717	0,007
	Mes8	-1848,177	498,316	-0,012	-3,709	0
	Mes9	-2223,055	498,285	-0,015	-4,461	0
	Mes10	-2281,506	498,288	-0,015	-4,579	0
	Mes11	-1205,415	498,318	-0,008	-2,419	0,016
	Mes12	-942,799	498,332	-0,006	-1,892	0,059
Variable Dependiente: Interés T+1						

- Modelo Lineal solo con dummies de meses 3 y 12

Coeficientes						
Modelo		Coeficientes		Coeficientes Estandarizados	t	Sig.
		B	Estad. Error	Beta		
	Constante	1159,3	281,7		4,115	0,000
	Frecuencia	7948,8	462,3	0,054	17,194	0,000
	Recency	-5,6	38,9	0,000	-0,143	0,886
	Interés Promedio	0,4	0,0	0,150	57,440	0,000
	Mes3	815,1	369,9	0,005	2,203	0,028
	Mes12	745,3	369,9	0,005	2,015	0,044
Variable Dependiente: Interés T+1						

- Modelo Log-Normal

Coeficientes				
Modelo	Coeficientes	Coeficientes Estandarizados	t	Sig.

	B	Estd. Error	Beta		
Constante	0,70	0,04		15,981	0,000
Frecuencia	5,99	0,05	0,400	131,783	0,000
Recency	-0,02	0,00	-0,019	-6,324	0,000
Interés Promedio	0,04	0,01	0,026	8,432	0,000
Mes3	0,14	0,03	0,009	3,950	0,000
Mes12	0,75	0,03	0,049	21,619	0,000
Variable Dependiente: Log(Interés T+1)					

- Modelo de Poisson

Estimación Parámetros				
Parámetros		Estd. Error		
			Wald Chi-Square	Sig.
Constante	8,22194652	0,04	34722,86	0,000
Frecuencia	1,24228179	0,06	442,35	0,000
Recency	-0,05650258	0,01	50,22	0,000
Interés Promedio	1,7711E-05	0,00	2041,83	0,000
Mes3	0,05276284	0,05	1,14	0,286
Mes12	0,13920281	0,05	8,08	0,004
(Scale)	227105,099			
Variable Dependiente: Interés T+1				

- Modelo Binomial Negativo

Estimación Parámetros				
Parámetros		Estd. Error		
			Wald Chi-Square	Sig.
Constante	8,01171144	0,05	24055,33	0,000
Frecuencia	0,98813305	0,09	125,94	0,000
Recency	-0,03879213	0,01	31,50	0,000
Interés Promedio	4,2939E-05	0,00	388,02	0,000
Mes3	0,20995789	0,07	9,89	0,002
Mes12	0,1960375	0,07	8,53	0,003
(Scale)	53,5028087			
Variable Dependiente: Interés T+1				

Conjunto de Variables RFM Segrado

- Modelo Lineal

Coeficientes					
Modelo	Coeficientes		Coeficientes Estandarizados	t	Sig.
	B	Estad. Error	Beta		
Constante	1010,5	505,8		1,998	0,046
Recency Holding 1	-21,3	27,8	-0,003	-0,768	0,443
Recency Holding 2	-45,4	27,7	-0,006	-1,638	0,101
Recency Holding 3	76,9	34,8	0,008	2,209	0,027
Recency Holding 4	-52,0	27,4	-0,007	-1,893	0,058
Recency Holding 5	38,5	28,3	0,005	1,361	0,174
Frecuencia Holding 1	4404,2	666,6	0,023	6,607	0,000
Frecuencia Holding 2	8757,9	1001,5	0,031	8,745	0,000
Frecuencia Holding 3	3118,4	1446,7	0,007	2,155	0,031
Frecuencia Holding 4	-476,0	641,6	-0,003	-0,742	0,458
Frecuencia Holding 5	17526,0	1045,0	0,063	16,771	0,000
Interés Promedio	0,4	0,0	0,127	44,521	0,000
Mes3	851,2	390,9	0,006	2,177	0,029
Mes12	655,1	389,9	0,004	1,680	0,093
Variable Dependiente: Interés T+1					

- Modelo Log-Normal

Coeficientes					
Modelo	Coeficientes		Coeficientes Estandarizados	t	Sig.
	B	Estad. Error	Beta		
Constante	1,5	0,1		27,624	0,000
Recency Holding 1	0,0	0,0	-0,028	-8,773	0,000
Recency Holding 2	0,0	0,0	-0,043	-12,919	0,000
Recency Holding 3	0,0	0,0	0,038	11,161	0,000
Recency Holding 4	0,0	0,0	-0,050	-14,377	0,000
Recency Holding 5	0,0	0,0	-0,018	-5,137	0,000
Frecuencia Holding 1	3,4	0,1	0,173	54,202	0,000
Frecuencia Holding 2	2,6	0,1	0,090	27,445	0,000
Frecuencia Holding 3	4,2	0,1	0,088	30,971	0,000
Frecuencia Holding 4	2,6	0,1	0,146	43,563	0,000
Frecuencia Holding 5	3,4	0,1	0,118	35,535	0,000
Interés Promedio	0,1	0,0	0,030	10,116	0,000
Mes3	0,0	0,0	0,000	0,168	0,866
Mes12	0,9	0,0	0,056	23,529	0,000
Variable Dependiente: Interés T+1					

- Modelo de Poisson

Estimación Parámetros				
Parámetros		Estad. Error	Wald Chi-Square	Sig.
Constante	8,450	0,070	14565,973	0,000
Recency Holding 1	-0,013	0,004	10,177	0,001
Recency Holding 2	-0,018	0,004	24,047	0,000
Recency Holding 3	0,027	0,004	38,195	0,000
Recency Holding 4	-0,004	0,004	1,124	0,289
Recency Holding 5	-0,032	0,004	83,561	0,000
Frecuencia Holding 1	0,553	0,077	51,610	0,000
Frecuencia Holding 2	0,801	0,100	64,505	0,000
Frecuencia Holding 3	0,751	0,156	23,295	0,000
Frecuencia Holding 4	-0,008	0,074	0,012	0,914
Frecuencia Holding 5	1,169	0,091	163,385	0,000
Interés Promedio	0,000	0,000	866,076	0,000
Mes3	0,052	0,051	1,054	0,305
Mes12	0,126	0,051	6,189	0,013
(Scale)	216119,191			
Variable Dependiente: Interés T+1				

- Modelo Binomial Negativo

Estimación Parámetros				
Parámetro		Estad. Error	Wald Chi-Square	Sig.
Constante	8,219	0,088	8635,423	0,000
Recency Holding 1	-0,010	0,005	4,453	0,035
Recency Holding 2	-0,022	0,005	22,241	0,000
Recency Holding 3	0,022	0,006	14,871	0,000
Recency Holding 4	-0,006	0,005	1,491	0,222
Recency Holding 5	-0,025	0,005	26,524	0,000
Frecuencia Holding 1	0,719	0,119	36,505	0,000
Frecuencia Holding 2	0,754	0,177	18,063	0,000
Frecuencia Holding 3	0,732	0,271	7,295	0,007
Frecuencia Holding 4	0,011	0,111	0,010	0,918
Frecuencia Holding 5	1,103	0,196	31,626	0,000
Interés Promedio	0,000	0,000	243,915	0,000
Mes3	0,164	0,069	5,713	0,017
Mes12	0,238	0,068	12,154	0,000
(Scale)	2,58			
Variable Dependiente: Interés T+1				

Conjunto de Variables RFM Full

- Modelo Lineal

Modelo	Coeficientes		Coeficientes Estandarizados	t	Sig.
	B	Estd. Error	Beta		
Constante	-2515,89	692,08		-3,635	0,000
Recency Holding 1	-10,70	27,90	-0,001	-0,383	0,701
Recency Holding 2	-15,44	27,81	-0,002	-0,555	0,579
Recency Holding 3	51,69	34,82	0,005	1,485	0,138
Recency Holding 4	-65,55	27,44	-0,009	-2,389	0,017
Recency Holding 5	3,80	28,36	0,001	0,134	0,893
Frecuencia Holding 1	388,99	703,97	0,002	0,553	0,581
Frecuencia Holding 2	6676,60	1007,22	0,024	6,629	0,000
Frecuencia Holding 3	503,76	1461,86	0,001	0,345	0,730
Frecuencia Holding 4	-1377,99	647,50	-0,008	-2,128	0,033
Frecuencia Holding 5	9163,97	1152,95	0,033	7,948	0,000
Interés Promedio	0,15	0,01	0,050	10,100	0,000
Años Cuenta	-50,46	21,27	-0,007	-2,373	0,018
Edad	-8,08	7,90	-0,003	-1,022	0,307
Sexo	55,85	209,50	0,001	0,267	0,790
Monto Cupo	0,00	0,00	0,062	17,295	0,000
Rango % Uso	227,26	79,77	0,009	2,849	0,004
Cuotas Promedio	1910,51	116,73	0,080	16,367	0,000
Monto Promedio	0,01	0,00	0,023	5,305	0,000
Scoring	-540,14	146,79	-0,010	-3,680	0,000
Mes 3	939,51	389,94	0,006	2,409	0,016
Mes 12	671,98	389,00	0,005	1,727	0,084
Variable Dependiente: Interés T+1					

- Modelo Log-Normal

Modelo	Coeficientes		Coeficientes Estandarizados	t	Sig.
	B	Estd. Error	Beta		
Constante	-46392,44	2013,49		-23,04	0,00
Recency Holding 1	25,75	28,44	0,00	0,91	0,37
Recency Holding 2	20,89	27,90	0,00	0,75	0,45
Recency Holding 3	-6,07	35,42	0,00	-0,17	0,86
Recency Holding 4	-51,90	27,51	-0,01	-1,89	0,06
Recency Holding 5	80,03	28,72	0,01	2,79	0,01
Frecuencia Holding 1	-36,28	709,33	0,00	-0,05	0,96
Frecuencia Holding 2	8407,21	1018,58	0,03	8,25	0,00
Frecuencia Holding 3	-1180,20	1477,27	0,00	-0,80	0,42
Frecuencia Holding 4	-1904,73	649,64	-0,01	-2,93	0,00

Frecuencia Holding 5	16958,10	1096,25	0,06	15,47	0,00
Interés Promedio	-43,71	74,28	0,00	-0,59	0,56
Años Cuenta	-570,67	156,58	-0,01	-3,64	0,00
Edad	-895,11	343,23	-0,01	-2,61	0,01
Sexo	-76,09	210,62	0,00	-0,36	0,72
Monto Cupo	4189,67	152,27	0,10	27,51	0,00
Rango % Uso	2034,20	206,16	0,03	9,87	0,00
Cuotas Promedio	7529,53	378,80	0,10	19,88	0,00
Monto Promedio	-562,15	85,24	-0,03	-6,60	0,00
Scoring	-1428,87	298,12	-0,01	-4,79	0,00
Mes 3	1089,47	391,45	0,01	2,78	0,01
Mes 12	578,41	390,17	0,00	1,48	0,14
Variable Dependiente: Interés T+1					

- Modelo de Poisson

Estimación Parámetros				
Parámetros		Estd. Error	Wald Chi-Square	Sig.
Constante	6,892	0,132	2729,969	0,000
Frecuencia Holding 1	0,162	0,078	4,332	0,037
Frecuencia Holding 2	0,205	0,100	4,155	0,042
Frecuencia Holding 3	0,309	0,155	3,962	0,047
Frecuencia Holding 4	-0,077	0,072	1,129	0,288
Frecuencia Holding 5	0,506	0,103	24,157	0,000
Recency Holding 1	0,000	0,004	0,002	0,968
Recency Holding 2	-0,011	0,004	9,786	0,002
Recency Holding 3	0,013	0,004	9,217	0,002
Recency Holding 4	-0,008	0,004	5,090	0,024
Recency Holding 5	-0,020	0,004	30,341	0,000
Interés Promedio	0,201	0,012	265,819	0,000
Años Cuenta	-0,005	0,003	3,409	0,065
Edad	0,000	0,001	0,173	0,678
Sexo	0,011	0,028	0,147	0,701
Monto Cupo	0,000	0,000	394,283	0,000
Rango % Uso	0,071	0,011	44,424	0,000
Cuotas Promedio	0,067	0,007	86,296	0,000
Monto Promedio	0,000	0,000	0,047	0,829
Scoring	-0,325	0,032	100,377	0,000
Mes 3	0,104	0,050	4,302	0,038
Mes 12	0,099	0,049	4,070	0,044
(Scale)	207480,4			
Variable Dependiente: Interés T+1				

- Modelo Binomial Negativo

Estimación Parámetros				
Parámetros		Estd. Error	Wald Chi-Square	Sig.
Constante	7,006	0,116	3634,897	0,000
Frecuencia Holding 1	0,319	0,108	8,777	0,003
Frecuencia Holding 2	0,336	0,155	4,690	0,030
Frecuencia Holding 3	0,332	0,235	1,992	0,158
Frecuencia Holding 4	0,008	0,099	0,007	0,935
Frecuencia Holding 5	0,516	0,182	8,048	0,005
Recency Holding 1	0,001	0,004	0,095	0,757
Recency Holding 2	-0,012	0,004	8,512	0,004
Recency Holding 3	0,011	0,005	4,436	0,035
Recency Holding 4	-0,005	0,004	1,446	0,229
Recency Holding 5	-0,021	0,004	22,850	0,000
Interés Promedio	0,083	0,009	91,691	0,000
Años Cuenta	-0,002	0,003	0,353	0,552
Edad	0,001	0,001	0,879	0,348
Sexo	0,076	0,032	5,516	0,019
Monto Cupo	0,000	0,000	350,397	0,000
Rango % Uso	0,062	0,012	24,968	0,000
Cuotas Promedio	0,146	0,016	83,451	0,000
Monto Promedio	0,000	0,000	0,716	0,397
Scoring	-0,205	0,022	88,489	0,000
Mes 3	0,103	0,061	2,895	0,089
Mes 12	0,270	0,060	20,337	0,000
(Scale)	38,582			
Variable Dependiente: Interés T+1				