



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

**METODOLOGÍA PARA ESTIMAR LA PRINCIPALIDAD DE LA TARJETA DE CRÉDITO
DE UNA TIENDA POR DEPARTAMENTOS**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

ISMAEL ALEJANDRO EGUIA JACOB

**PROFESOR GUÍA:
RICARDO MONTOYA MOREIRA**

**MIEMBROS DE LA COMISIÓN:
MARCEL GOIC FIGUEROA
NANCY LARA RODRÍGUEZ**

**SANTIAGO DE CHILE
ENERO 2013**

**METODOLOGÍA PARA ESTIMAR LA PRINCIPALIDAD DE LA TARJETA DE CRÉDITO
DE UNA TIENDA POR DEPARTAMENTOS**

En ambientes competitivos, los consumidores pueden adquirir sus requerimientos de productos y servicios en diversas empresas. En este contexto, las compañías desarrollan estrategias de gestión de clientes, como programas de fidelización, estrategias de cross-selling o up-selling y muchos otros esfuerzos de marketing focalizados, con el objetivo de que los clientes adquieran la mayoría de sus requerimientos con ellos. Sin embargo, estas estrategias se realizan con información incompleta, puesto que desconocen el total de gasto que los consumidores poseen en cada una de las categorías.

El objetivo principal de esta memoria es desarrollar una metodología que permita estimar cuánto del gasto mensual de una persona es capturado por medio de la tarjeta de crédito perteneciente a una tienda por departamentos, lo cual se denomina principalidad.

Para lograr el objetivo propuesto, se trabaja con una muestra de 161.005 clientes que han realizado alguna compra en el periodo comprendido entre junio de 2010 a mayo de 2012. La estimación de la principalidad se logra utilizando el modelo binomial generalizado (GBM), el cual, en ausencia de datos del comportamiento del cliente fuera de la compañía, supone que el número de transacciones totales que éste realiza se distribuye Poisson. En primera instancia, se aplica el GBM utilizando sólo datos transaccionales, para luego incluir variables demográficas y de negocio, modelo llamado GBM extendido.

La estimación realizada con el GBM extendido tiene un pseudo ρ^2 de 66% y un MAPE tanto de ajuste como de validación del orden del 12,5%, mientras que el modelo actual que posee la empresa tiene un error de un 25%. Si bien, tanto el modelo de la empresa como el propuesto en este trabajo entregan una principalidad promedio de un 20%, la metodología propuesta permite una mayor variabilidad en el potencial de gasto junto con asegurar que el gasto total es siempre mayor o igual al realizado con la tarjeta de crédito.

Una vez realizada la estimación se realiza una segmentación basada en la principalidad y en el gasto total, donde se observa que sólo un 1,8% de la muestra total, representa a los clientes que cuentan con un alto gasto y con un gran porcentaje de éste realizado con la tarjeta. Por otro lado, al evaluar a que categoría está más vinculado cada cliente, se observa que aquellos asignados a las unidades de consumo que no son cubiertas por el holding al cual pertenece la empresa, como lo son educación, salud y recreación, presentan la mayor principalidad, donde los asignados a educación tienen el mayor valor con un promedio de 47% pero representan sólo el 0.43% de la muestra.

Como trabajo futuro se propone obtener información para una muestra de clientes respecto de cuanto gastan mensualmente, con el objeto de evaluar de mejor manera los resultados entregados por la metodología propuesta y poder desarrollar modelos supervisados, para luego comparar los niveles de ajuste. Además, se propone evaluar de manera dinámica las estrategias que serán implementadas para aumentar la principalidad.

AGRADECIMIENTOS

A mis padres por su apoyo, cariño, preocupación y soporte en los periodos de estrés.

A mis hermanos, en específico a Ricardo por haberme enseñado a ser responsable y serio en las cosas que uno hace, por su apoyo y sustento desde que nací.

A mi abuelita Olga y tíos maternos, por brindarme cariño, apoyo y consejos durante mi carrera en la universidad.

A mi polola Karina, quien siempre estuvo conmigo, brindándome apoyo y cariño en los buenos y malos momentos, junto con soportar mi peor carácter. También a su familia, por la hospitalidad ofrecida en todos estos años.

A mis amigos de la universidad, por los buenos momentos vividos en estos años, donde hubo largas noches de estudios y trabajos, como muchas otras de fiestas, salidas y juntas.

A las personas del área de la empresa, por su simpatía y momentos agradables que me hicieron pasar en el transcurso de la memoria. En específico a Juan Pablo y Daniel, por su paciencia, amabilidad, buena disposición y ayuda.

A los profesores Ricardo y Marcel, por su paciencia, compromiso y enseñanzas, junto con las oportunidades brindadas estos dos últimos años.

A todos los que colaboraron de alguna u otra forma aconsejando, criticando y ayudando durante mi estancia en la universidad.

ÍNDICE DE CONTENIDOS

1.	ANTECEDENTES GENERALES	9
2.	DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN	10
3.	OBJETIVOS	11
3.1.	Objetivo General	11
3.2.	Objetivos Específicos	11
4.	ALCANCES Y RESULTADOS ESPERADOS	12
4.1.	Alcances.....	12
4.2.	Resultados esperados.....	13
5.	MODELO ACTUAL UTILIZADO POR LA EMPRESA	13
5.1.	Estimación del potencial de compra de cada cliente.....	13
5.2.	Unidades de consumo a cubrir.....	14
5.3.	Cálculo de la principalidad	15
5.4.	Ajuste del modelo.....	15
5.5.	Críticas al modelo.....	17
6.	METODOLOGÍA	17
6.1.	Metodología para el tratamiento de los datos	18
6.2.	Metodología para el desarrollo del modelo	19
6.2.1.	Modelo Binomial Generalizado	19
6.2.2.	Regresión Log-Lin	22
6.3.	Metodología para evaluar la calidad del ajuste	23
6.4.	Metodología para estimar la principalidad.....	23
6.5.	Metodología para segmentar a los clientes.....	24
6.6.	Propuestas de acción.....	24
7.	TRATAMIENTO DE LA BASE DE DATOS	25
7.1.	Selección de los datos	25
7.2.	Procesamiento de los datos	26
7.3.	Transformación de los datos	28
7.4.	Análisis descriptivo y evaluación de las variables seleccionadas	29
8.	DESARROLLO DEL MODELO	33
8.1.	Simulación.....	34
8.1.1.	Resultados de la simulación	34
8.2.	Modelo Log-Lin	36
8.3.	Modelo Binomial Generalizado transaccional	38

8.4.	Modelo Binomial Generalizado extendido	42
8.5.	Aplicación de la metodología a las categorías vestuario y hogar	46
9.	RESULTADOS DEL GBM EXTENDIDO	48
9.1.	Resultados generales del modelo GBM extendido	48
9.1.1.	Comparación con el modelo actual de la empresa	51
9.2.	Principalidad y valor del cliente	53
9.3.	Segmentos en función de la principalidad y <i>size of wallet</i>	56
9.4.	Principalidad y segmentación por categorías de consumo	57
9.4.1.	Segmentación por categorías de consumo	59
9.5.	Segmentación utilizando <i>K-mean</i>	63
9.6.	Propuestas de acción.....	65
10.	CONCLUSIONES.....	68
10.1.	Conclusiones del trabajo.....	68
10.2.	Limitaciones del modelo.....	69
10.3.	Recomendaciones y Trabajos futuros.....	70
11.	BIBLIOGRAFÍA.....	71
12.	ANEXOS	72

ÍNDICE DE GRÁFICOS

Gráfico 1: Ajuste del modelo actual de la empresa	17
Gráfico 2: Distribución de clientes por zona	30
Gráfico 3: Gasto promedio por zona con la tarjeta durante el último año.....	30
Gráfico 4: Distribución de la base de datos según rango de edad	31
Gráfico 5: Gasto promedio durante el último año según rango de edad	31
Gráfico 6: Comparativa del gasto promedio por cliente con la tarjeta en ambos periodos de estudio.....	32
Gráfico 7: Monto promedio mensual gastado con la tarjeta vs recencia según tipo de cliente.....	33
Gráfico 8: Monto promedio mensual gastado con la tarjeta Visa vs recencia según tipo de cliente.....	33
Gráfico 9: MAPE como función del número de transacciones promedio con la tarjeta en cada repetición	35
Gráfico 10: MAPE como función del número de transacciones promedio con cualquier medio de pago en cada repetición	35
Gráfico 11: Ajuste del GBM transaccional	40
Gráfico 12: Número de transacciones con la tarjeta vs número de transacciones estimados con cualquier medio de pago utilizando GBM transaccional.....	41
Gráfico 13: Ajuste del GBM extendido	44
Gráfico 14: Número de transacciones con la tarjeta vs número de transacciones estimados con cualquier medio de pago utilizando GBM extendido	46
Gráfico 15: Clientes ordenados según <i>Share of wallet</i>	48
Gráfico 16: Histograma de clientes según <i>share of wallet</i>	49
Gráfico 17: Recencia y antigüedad según quintil de principalidad	50
Gráfico 18: Porcentaje de tipos de clientes según principalidad	51
Gráfico 19: Número de transacciones vs gasto total estimado por la empresa.....	52
Gráfico 20: Histograma de clientes según <i>share of wallet</i> estimado por la empresa	53
Gráfico 21: Concentración de la rentabilidad del cliente para el negocio	54
Gráfico 22: Concentración de la rentabilidad del cliente para la empresa de la tarjeta de crédito	55
Gráfico 23: Grupos de clientes según principalidad de la tarjeta y gasto total con cualquier medio de pago	57
Gráfico 24: Principalidad por categoría	59
Gráfico 25: Porcentaje de clientes vinculados a cada unidad de consumo	60
Gráfico 26: Principalidad global según segmento asignado	61

ÍNDICE DE TABLAS

Tabla 1: Descripción del grupo superior e inferior en gasto con la tarjeta en el primer año de estudio.....	27
Tabla 2: Descripción en el segundo año de estudio.....	27
Tabla 3: Descripción de los nuevos grupos una vez eliminados los registros fuera de rango en el primer periodo	27
Tabla 4: Descripción de los nuevos grupos en el segundo periodo	28
Tabla 5: Segmentación de clientes según puntos acumulados.....	32
Tabla 6: Estadísticas para regresión Log-Lin utilizando <i>cross validation</i>	36
Tabla 7: Estadísticas para regresión Log-Lin con base de muestra global	36
Tabla 8 Resumen coeficientes estimados para modelo Log-Lin	37
Tabla 9: Resumen coeficientes estimados para GBM transaccional.....	39
Tabla 10: Estadísticas para GBM transaccional utilizando <i>cross validation</i>	39
Tabla 11: Estadísticas para GBM transaccional con base de muestra global.....	40
Tabla 12: Error por quintil de gasto promedio mensual con la tarjeta en el periodo 2 del GBM transaccional	41
Tabla 13: Resumen coeficientes estimados para GBM extendido	43
Tabla 14: Estadísticas para GBM extendido utilizando <i>cross validation</i>	44
Tabla 15: Estadísticas para GBM extendido con base de muestra global	44
Tabla 16: Error por quintil de gasto promedio mensual en el segundo periodo para el GBM extendido.....	45
Tabla 17: Resumen de los tres modelos	46
Tabla 18: Análisis descriptivo de las categorías vestuario y hogar según gasto con la tarjeta	47
Tabla 19: Comparación gasto estimado por la empresa vs GBM extendido.....	51
Tabla 20: Cantidad de clientes según terciles de principalidad vs terciles de rentabilidad de clientes para el negocio.....	56
Tabla 21: Cantidad de clientes según terciles de principalidad vs terciles de rentabilidad de clientes para la empresa de la tarjeta de crédito	56
Tabla 22: Distribución del gasto mensual por quintil de ingreso.....	58
Tabla 23: Características demográficas de los segmentos	61
Tabla 24: Caracterización transaccional de los segmentos	62
Tabla 25: Total de clientes por tipo en la muestra	62
Tabla 26: Principalidad por unidad de consumo según segmento	63
Tabla 27: Centros de los segmentos para variables de principalidad	64
Tabla 28: Centros de los segmentos para variables de negocio	65

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Etapas de la metodología KDD	18
---	----

1. ANTECEDENTES GENERALES

En muchos de los sectores económicos del país los consumidores tienen la posibilidad de elegir más de un lugar para comprar el mismo producto o servicio en un periodo de tiempo determinado. Lo anterior se puede evidenciar en sectores como el retail, donde pueden realizar compras de cualquier categoría, como por ejemplo vestuario, en una gran cantidad de tiendas y formatos, como supermercados, tiendas por departamentos, etc.

Actualmente, la gran mayoría de las compañías sólo poseen información acerca de la interacción que sus clientes realizan con ellos, siendo escasa o no existente la que tiene la competencia con éstos. Por lo tanto, un consumidor que se ve muy rentable de potenciar tal vez no lo sea, puesto que podría estar gastando todo lo que tiene en la empresa focal. Lo anterior la compañía lo desconoce, por lo que destina gastos de marketing en potenciar a un cliente que no le generará un retorno sobre la inversión. Este problema es transversal en muchos sectores de la economía, tanto para productos como servicios. Si se lograra conocer la fracción del gasto del cliente que se está capturando, se podrían realizar mejores estrategias de gestión de clientes, las cuales estarían mejor fundadas.

Dada la parcialidad o no existencia de esta información, las compañías no tienen otra opción más que realizar estrategias de *CRM (Customer Relationship Management)* desde un punto de vista limitado, puesto que éstas se basan solamente en información de registros internos que la firma posee del cliente. Lo anterior, genera lo siguiente [1]:

- Basarse solamente en las transacciones que el cliente realiza con la empresa, podría entregar resultados equivocados acerca del potencial de compra que posee el cliente, pudiendo sobrestimarlo o subestimarlos.
- La falta de información a nivel individual de las transacciones que el cliente hace en la industria completa, es una barrera importante al momento de realizar un buen *CRM*.
- Siguiendo la línea del punto anterior, esta falta de información dificulta el diseño de buenos programas de lealtad de clientes, aplicaciones de *up-selling*, promociones focalizadas, y muchos otros esfuerzos de marketing con el objetivo de lograr el mejor retorno sobre la inversión.

La empresa en la cual se desarrolla el presente trabajo, perteneciente a un holding de retail, presenta el problema al igual que la gran mayoría de las compañías de la industria, de no conocer el porcentaje del gasto de sus clientes que captura, en este caso, por medio de su tarjeta de crédito (desde ahora, "*Share of wallet (ShoW)*" o *Principidad*¹), por consiguiente, tampoco conoce el volumen de gasto de los clientes

¹ Representa una medida de cuanto del total de gasto de un cliente es realizado con la tarjeta de crédito de la empresa. Este indicador da luces de clientes que poseen un potencial de crecimiento, al contrastarlo con el *size of wallet*, como también de clientes que están abandonando a la empresa, lo que se vislumbraría con una baja de este indicador, dando señales para realizar campañas de retención.

con cualquier medio de pago en cada categoría (desde ahora, “*Size of wallet (SoW)*”²). Conociendo ambas variables, es posible determinar el potencial de gasto que queda por adquirir de cada cliente (desde ahora, “*Potencial of Wallet (PoW)*”), el cual representa la diferencia entre el SoW y el ShoW.

Lo mencionado anteriormente es aún más importante para esta empresa, ya que en Septiembre de 2009 amplió el campo de acción de su tarjeta, luego de haber comenzado una alianza con Visa. La tarjeta, que antiguamente funcionaba solo al interior del holding, ahora permite que sus clientes la utilicen en todos los comercios locales e internacionales, extendiendo el número de categorías para realizar las compras.

En este trabajo, la tarjeta de la compañía resulta ser el producto ofrecido, y los otros medios de pago la competencia, por lo que el consumidor en cada categoría puede elegir cuanto gastar con la tarjeta y cuanto con otro medio de pago.

Los modelos existentes en la literatura que permiten determinar el SoW, requieren datos acerca de cuanto gastan en total con la competencia una muestra de los clientes [2], información obtenida por medio de una encuesta (ver [1] y [3]). Esta información les resulta necesaria para poder ajustar los parámetros en los modelos, que ayudarán a explicar el comportamiento del resto de los clientes que no poseen información adicional, de la manera más precisa posible.

En este estudio, se presentará el modelo Binomial generalizado (GBM), el cual no necesita de datos obtenidos por una encuesta para estimar los parámetros. En primera instancia, éste utilizará sólo información transaccional (al interior de la empresa), el cual será denominado como modelo binomial generalizado transaccional. En una segunda iteración, el modelo incluirá además, variables demográficas y de negocio (cupó, tipo de cliente, etc.) para poder predecir el SoW, denominándolo como modelo binomial generalizado extendido. Es importante mencionar que, para una mejor evaluación del modelo, es importante contar con datos provenientes de una muestra de los clientes que reporten su SoW, de manera de poder evaluar con mayor precisión el ajuste del modelo, el cual es construido sólo con datos internos de la empresa.

2. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN

Estimar la *principalidad* (ShoW) de la tarjeta de la empresa para cada uno de sus clientes, es fundamental para el área de Inteligencia de Negocios de la firma, puesto que en la actualidad se desconocen variables muy importantes al momento de desarrollar estrategias de CRM, como lo son el potencial de gasto que cada cliente presenta en cada categoría (PoW), y cuánto están adquiriendo del gasto de ellos en cada una de las categorías por medio de la tarjeta de crédito.

² La “Billetera” de un cliente (Size of wallet), es definida como la cantidad total que puede gastar en una categoría de producto específica.

En orden de mejorar la rentabilidad de cada uno de los clientes dentro de una empresa, el ShoW puede servir para cumplir este objetivo de dos maneras, permitiendo identificar a los clientes que [2]:

- Presentan un mayor grado de lealtad, para los cuales se deben desarrollar estrategias que impidan su deserción.
- Tengan un potencial de crecimiento. Lo anterior se refiere a que si un cliente tiene un bajo porcentaje de principalidad, se deben generar políticas para que éste realice una mayor cantidad de compras con la tarjeta, lo que provocará un aumento en las ganancias de la empresa.

A partir de lo anterior, el problema de negocio que el área de *Customer & Business Intelligence* de la empresa quiere solucionar, es el de aumentar el conocimiento de sus clientes, en relación a cuánto de su gasto se está realizando por medio de la tarjeta. Conocer esto, les permitirá gestionar a los segmentos de clientes según el grado de principalidad que posean. Para un grupo con alto grado de principalidad, la empresa puede generar políticas que mantengan el nivel actual, mientras que para uno con un bajo nivel de principalidad, la empresa puede diseñar acciones que potencien el uso de la tarjeta, para obtener una mayor participación de ésta. Por lo tanto, el indicador que será desarrollado en la memoria, será el indicador principal para la estrategia de gestión de los clientes, lo cual les permitirá realizar gastos más fundados para lograr este objetivo sin incurrir en sobre costos.

Estimar la principalidad resulta ser muy complejo, ya que a pesar de que hoy en día se cuenta con herramientas tecnológicas avanzadas que permiten procesar grandes cantidades de información y descubrir patrones de comportamiento a nivel individual al interior de las bases de datos, tanto la empresa en la que se desarrolla el proyecto como muchas otras, poseen el problema de la información incompleta. Lo anterior, se refiere a que las bases de datos están conformadas sólo por las transacciones que los clientes han realizado con la compañía. En el caso de esta empresa, sólo se conoce cuánto y qué tipo de transacciones hacen los clientes con la tarjeta de crédito, desconociendo esta información para cualquier otro medio de pago. Debido a la limitación de información que existe, resulta difícil estimar el nivel de gasto de un cliente agregando todos los medios de pago, lo que resulta ser información exógena a la empresa, por lo que para poder lograrlo, se deben hacer supuestos sobre la información inexistente, utilizando sólo datos transaccionales de la tarjeta de crédito y variables que describan aspectos propios del consumidor para hacer la estimación

3. OBJETIVOS

3.1. Objetivo General

Desarrollar una metodología para estimar el potencial de compra de los clientes.

3.2. Objetivos Específicos

- Identificar las principales variables que permiten estimar el *size of wallet*.

- Desarrollar los modelos que permitan estimar el *size of wallet*.
- Desarrollar el indicador de principalidad utilizando el mejor modelo.
- Generar segmentos de clientes según principalidad.
- Proponer líneas de acción, con el objeto de aumentar o mantener la principalidad en las distintas categorías utilizando la información proveniente de los segmentos encontrados.

4. ALCANCES Y RESULTADOS ESPERADOS

4.1. Alcances

Esta memoria pretende generar una metodología que permita identificar cuánto representa la tarjeta de crédito de una tienda por departamentos en el gasto total que realizan cada uno de sus clientes con cualquier medio de pago. Debido a las restricciones que imponen los datos existentes y al tiempo limitado que se posee para desarrollar la memoria, el alcance del proyecto se limitará a lo siguiente:

- **Se estimará la principalidad para una muestra de clientes que actualmente utilicen la tarjeta de crédito:** Para un mejor manejo de los datos, y debido a capacidad computacional, sólo se trabajará con una muestra de 161.005 clientes activos, utilizando un periodo de dos años de historia transaccional.
- **Se estimará el potencial de compra total y en dos categorías:** Junto con estimar el SoW de cada cliente de la muestra, considerando de manera agregada las unidades de consumo del Instituto Nacional de Estadísticas, INE (vestuario, vivienda, hogar, salud, transporte, recreación, educación, otros), también se calculará el SoW que posee en dos de estas unidades de manera independiente: Vestuario y Hogar. La elección de sólo dos categorías a ser tratadas de manera individual viene de la gran cantidad de personas que no presentan compras registradas en el periodo en el resto de las categorías, pudiendo generar una gran cantidad de error al realizar una estimación del potencial en ellas. Además, según lo mencionado por la empresa, Vestuario y Hogar representan las unidades con un mayor número de clientes activos, haciendo posible una estimación razonable del potencial de gasto en esas dos unidades individuales. Dado que para efectos estratégicos es importante evaluar en qué categorías cada cliente es más y/o menos principal, para generar acciones por unidad de consumo la estimación en éstas se basará en el procedimiento actual de la compañía, de manera de facilitar la comparación con los métodos actuales.
- **Se realizará una segmentación de los clientes en base al indicador generado:** Una vez que se posea el indicador de principalidad, se generarán segmentos de clientes en base a éste, con el objetivo de identificar información relevante para la compañía en torno a la conducta de los poseedores de la tarjeta, la cual resulta útil para proponer lineamientos generales de acción.

- **No se generará información adicional de la que posee la empresa, como lo es el gasto de las personas con otros medios de pago en cada categoría:** Dado que la empresa no posee información acerca de cuánto gasta mensualmente una muestra de clientes considerando todos los medios de pagos, en este estudio no se levantará esa información, por lo que para evaluar el modelo se utilizará el método que actualmente usa la compañía para evaluar su metodología.
- **No se implementarán las líneas de acción.**
- **No se implementará un sistema automatizado de cálculo de principalidad:** Debido a que el indicador de principalidad irá variando en cada periodo de tiempo, éste debe ser actualizado. En este trabajo, no se implementará un sistema que realice la automatización del modelo al interior de la compañía, limitándose únicamente a generar la metodología para lograr obtener el indicador.
- **No se realizará un análisis de fuga de clientes en los grupos que posean baja principalidad:** Debido a que el indicador de principalidad representa una medida de lealtad de los clientes, y es posible identificar según diversos criterios que clientes se podrían estar fugando, en esta memoria no se abordará el análisis relacionado a fuga de clientes.

4.2. Resultados esperados

Al finalizar la memoria se espera obtener:

- Variables demográficas, transaccionales y de negocio que influyen en el *size of wallet*.
- Indicador de principalidad.
- Segmentos de clientes según la principalidad de la tarjeta de crédito y su perfil.
- Líneas de acción para aumentar el *share of wallet*.

5. MODELO ACTUAL UTILIZADO POR LA EMPRESA

En este capítulo se explica la solución que la empresa posee en estos momentos para estimar el SoW de su cartera de clientes. Además, se muestra el nivel de ajuste que posee y algunas críticas a su metodología.

5.1. Estimación del potencial de compra de cada cliente

Actualmente la empresa, para poder estimar el gasto total que realiza cada cliente, considera que este gasto depende del poder adquisitivo que posee cada uno de ellos. El problema de esto, es que no se posee el dato de renta para todos los clientes, sino que sólo para una muestra.

Para solucionar lo anterior, suponen que el ingreso depende de las variables demográficas del cliente, para lo cual consideran las siguientes:

- **Sexo:** Usan esta variable ya que según diversos estudios, los hombres ganan en promedio más que las mujeres. Esta variable posee dos categorías.
- **Edad:** Utilizan esta variable puesto que los jóvenes y adultos mayores, en promedio, ganan menos que los adultos jóvenes. Según tramos etarios generados por la empresa, ésta contiene cinco categorías:
 - 18 años a 25 años.
 - 26 años a 35 años.
 - 36 años a 45 años.
 - 45 años a 60 años.
 - Mayores a 60 años.
- **Zona de residencia:** Dado que según la zona de residencia la renta varía, se utiliza esta variable para lograr discriminar el ingreso de los segmentos. El país se dividió en ocho zonas, donde seis de éstas corresponden a Santiago, la séptima categoría corresponde a residentes de zonas de la región metropolitana no consideradas parte de Santiago, y la octava categoría son los habitantes de regiones.
- **Grupo socioeconómico:** Si bien la zona de residencia puede dar señales del grupo socioeconómico, se incluye esta variable para que la segmentación sea aun más exhaustiva. Sin embargo, el GSE no está disponible en toda la base de datos, por lo que para los datos faltantes, se imputa según la moda en cada uno de los grupos asignados. Esta variable posee cinco categorías.

A partir del número de categorías existentes en cada una de las cuatro variables mencionadas anteriormente, se generan 400 grupos de clientes.

Una vez realizado lo anterior, se toma el siguiente supuesto: “Cada cliente tendrá el mismo poder adquisitivo que el resto de los integrantes de su grupo”. Como no se tiene el poder adquisitivo, suponen que el Top 5% de cada grupo que gasta más con la tarjeta de crédito, son 100% principales, es decir, realizan todas sus compras con la tarjeta, por lo que el valor promedio de este 5% de clientes será el potencial de compra para el resto de los integrantes del grupo.

5.2. Unidades de consumo a cubrir

Junto con estimar el potencial de compra, se debe decidir que ítems/unidades de consumo cubrir para estimar la principalidad en cada una de éstas, ya que resulta información útil para realizar acciones de marketing a modo de aumentar este indicador en cada una de las categorías. Para ello, utilizan las unidades de consumo que maneja el INE, el cual contempla las siguientes 9 categorías: Alimentación, Vestuario, Vivienda, Hogar, Salud, Transporte, Recreación, Educación y Otros. Por lo tanto, cada una de las

compras que realiza el cliente en un periodo de tiempo dado con la tarjeta, son asignadas a cada una de estas 9 categorías.

Además, dado que el nivel de gasto que cada persona asigna mensualmente a estas categorías varía a nivel de cliente, se debe evaluar cuánto pesa en el consumo mensual cada una de ellas. Para esto, nuevamente se basan en el INE, el cual entrega los datos de gastos por ítem según quintil de ingreso. A modo de ejemplo, según estos datos, el quintil más rico gasta el 26% de su ingreso en transporte, versus el más pobre que gasta solamente un 15%. El problema de utilizar esta medida es que no se posee el dato de renta para todos los clientes, por lo que no pueden evaluar a qué quintil pertenece cada uno de ellos. Para solucionar este problema, realizan quintiles según el potencial estimado, por lo que el grupo de mayor potencial es homologado al quintil de mayor ingreso según el INE, realizando lo mismo para el resto de los segmentos.

5.3. Cálculo de la principalidad

Una vez estimado tanto el potencial de compra que posee cada cliente, como la importancia relativa de cada unidad de negocio para cada uno de ellos, y utilizando el nivel de gasto realizado con la tarjeta en cada una de éstas, es posible definir una fórmula que permite estimar la principalidad para cada uno de los clientes en cada categoría de consumo. La fórmula generada en el área de Inteligencia de negocios es la siguiente:

$$Principalidad_{i,j} = \min \left\{ \frac{Gasto\ con\ la\ tarjeta_{i,j}}{Potencial_i * Peso_{i,j}}, 1 \right\} \quad (1)$$

donde i representa a un cliente y j la categoría en la que se está calculando la principalidad. A partir de lo anterior, la principalidad total para cada cliente es la siguiente:

$$Principalidad_{total,i} = \sum_{Item\ j} Principalidad_{i,j} * Peso_{i,j} \quad (2)$$

Cabe mencionar que el peso obtenido desde el INE es recalculado en algunos casos, dado que algunos ítems de gasto fueron considerados parte de otras unidades de consumo distinta a lo asignado por la institución. Como ejemplo de esto, se puede mencionar el ítem comunicación, donde el INE lo tiene asignado en la unidad de consumo Transporte y comunicaciones, y la empresa lo reasignó a la unidad gastos de la vivienda.

5.4. Ajuste del modelo

Para evaluar el ajuste del modelo, la empresa utilizó una muestra de 1.000 clientes que poseen la variable renta, descartando a los que presentan un ingreso inferior a los \$182.000 y superior a los \$1.500.000, dado que éstos representan menos

del 7% de la base de clientes que poseen la variable, por lo que son considerados como valores fuera de rango.

Una vez seleccionada la muestra, y basándose en los resultados entregados por la VI encuesta de presupuestos familiares que realiza el INE³, el potencial de gasto estimado de las personas fue dividido por un factor de gasto sobre la renta entregado en este estudio, diferenciado por quintil de ingreso⁴. Como este factor es un promedio por quintil, la empresa lo modificó y linealizó a medida que el ingreso al interior de la muestra aumentaba, donde el factor mínimo es de un 29% (para el quintil con mayor ingreso luego de ser modificado) y el máximo es de un 160% (para el quintil con menor renta). La linealidad del factor se realiza con la siguiente fórmula:

$$\mathbf{Factor}_i = \mathbf{160\%} - \left(\frac{\mathbf{160\%} - \mathbf{29\%}}{\mathbf{n}} \right) * \mathbf{i} \quad (3)$$

donde i corresponde al número correlativo para cada cliente, el cual es ordenado de manera ascendente según el nivel de ingresos y n el número de individuos en la muestra.

Luego, para recuperar el ingreso a partir de la estimación del potencial y poder compararlo con la renta real, la cual se tiene sólo para un grupo menor de clientes, se utilizó lo siguiente:

$$\mathbf{Renta\ estimada}_i = \frac{\mathbf{Potencial\ estimado}_i}{\mathbf{Factor}_i} \quad (4)$$

Finalmente, para la evaluación del ajuste se utilizó el error absoluto promedio porcentual (MAPE), arrojando un error de un 25%. El Gráfico 1 muestra el nivel de ajuste, donde el área coloreada en naranja, representa las bandas tomando una desviación de un 50% respecto a la renta real, indicando que la gran mayoría de los datos se encuentran en ese intervalo de error.

³http://www.ine.cl/canales/chile_estadistico/encuestas_presupuestos_familiares/VI_ENCUESTA/pdf/EPF%20tomo%20III.pdf

⁴ La lógica detrás de esto es que las personas pueden no gastar todo su ingreso, ya que pueden gastar un porcentaje inferior a modo de ahorro, o gastar más que su ingreso ya que pueden gastar un porcentaje superior por medio del endeudamiento.

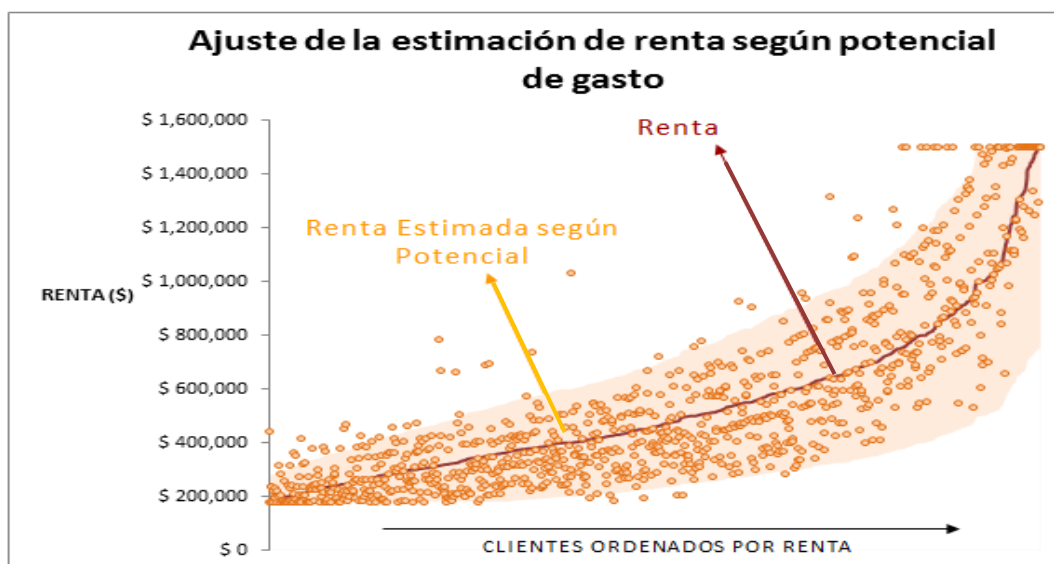


Gráfico 1: Ajuste del modelo actual de la empresa
Fuente: Área Customer & Business Intelligence de la empresa

5.5. Críticas al modelo

A partir de lo anterior, se observa que la manera de estimar el *size of wallet* para cada cliente está basado en muchos supuestos, lo cual genera errores en la estimación que son perjudiciales para realizar correctas estrategias de marketing. Uno de estos supuestos, es el hecho de considerar que los clientes pertenecientes a un mismo grupo de similares características, presentan el mismo gasto total mensual, no considerando la posible heterogeneidad al interior de cada uno de los grupos en relación al nivel de gasto que éstos poseen. El segundo supuesto que se realiza, el cual es el más fuerte en este modelo, es el hecho de considerar que los clientes que más gastan al interior de cada grupo utilizando la tarjeta, son clientes principales, es decir, el gasto con la tarjeta representa el total de consumo que realizan en un periodo de tiempo dado. Lo anterior no necesariamente es cierto, puesto que estos mismos clientes pueden tener incluso un gasto mayor, que es realizado con otro medio de pago. Es por esto que, el potencial que fijan para el resto de los integrantes de ese grupo puede tanto sobrestimar como subestimar el potencial de compra para los clientes pertenecientes a él. Sin embargo, para relajar estas restricciones y supuestos es necesario desarrollar un modelo que logre estimar el potencial de gasto de cada uno de los poseedores de la tarjeta, basándose únicamente en su propio comportamiento transaccional y características demográficas, de manera de obtener variabilidad dentro de cada uno de los grupos.

6. METODOLOGÍA

En este capítulo se aborda la metodología utilizada en cada una de las etapas de la memoria. Se describen las herramientas que se emplean y la forma en las que son usadas para lograr los objetivos propuestos.

6.1. Metodología para el tratamiento de los datos

La manera de abordar el tratamiento de los datos es por medio de la metodología denominada KDD (Knowledge Discovery in Databases). Esta metodología a lo largo de sus etapas, busca identificar patrones comprensibles, útiles, válidos y novedosos dentro de un conjunto de datos [4]. Las etapas con las que cuenta son las siguientes:

- Selección de los datos
- Pre procesamiento
- Transformación
- Aplicación de modelos
- Evaluación e interpretación de los resultados generados

Los pasos anteriormente expuestos son resumidos en la siguiente ilustración [4]:

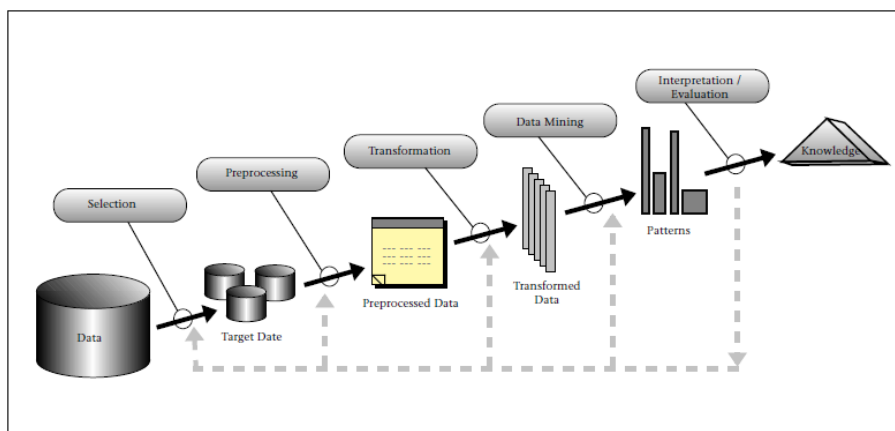


Ilustración 1: Etapas de la metodología KDD

Las etapas de selección, pre procesamiento y transformación son realizadas en el software estadístico SPSS Statistics versión 17.

La selección de los datos busca responder correctamente a los objetivos propuestos, es decir, encontrar las variables que logran describir el potencial de compra de los clientes con el menor error posible. Para ello, se utilizan las variables descritas en la literatura donde fue aplicado el modelo (ver [2]), el cual es descrito en la siguiente sección, junto con variables que de manera preliminar parecen razonables para explicar el nivel de gasto de una persona. Además, estas variables son evaluadas utilizando el test anova de un factor, con el objeto de ver si existen diferencias significativas entre los grupos de clientes en torno al gasto total realizado con la tarjeta, en el periodo comprendido entre junio de 2011 y mayo de 2012. Las variables con las que se cuenta se pueden agrupar en cuatro grandes grupos: Variables transaccionales, variables sociodemográficas, variables de comportamiento del cliente en la empresa y variables de negocio.

Por otro lado, el tamaño de la muestra es escogido de tal manera que sea un número suficiente para poder realizar estimaciones con un bajo nivel de sesgo,

considerando que deben ser clientes activos que tengan la variable renta conocida con el objeto de evaluar la capacidad predictiva del modelo.

Para evitar errores en la estimación, se realiza una limpieza de los datos que pueden causar ruido, como lo son los datos faltantes, valores fuera de rango, inconsistencias en las distintas variables o registros que no aportan información al modelo. Junto con lo anterior, se realiza una transformación en los datos para corregir inconsistencias identificables y corregibles, actualizar variables, y crear otras que no se encuentran explícitas en la base de datos pero que son utilizadas en el modelo.

6.2. Metodología para el desarrollo del modelo

Para estimar la principalidad, se deben hacer supuestos acerca del comportamiento de las personas en torno al nivel de gasto que realizan, debido a que no se tiene conocimiento acerca de lo que realmente está ocurriendo. Uno de los modelos presentados en la sección siguiente, el más importante en esta memoria, asume que las transacciones totales que hace un cliente siguen una distribución poisson, lo que implica que el tiempo entre cada transacción se distribuye exponencial.

El otro modelo corresponde a una regresión log-lin, el cual es utilizado para estimar la renta bajo el supuesto de que el factor por el cual se pondera el ingreso para obtener el gasto total de una persona, es correcto. Por lo que, dado que es la renta la variable utilizada por la empresa para evaluar el ajuste del modelo, si esta regresión tuviese un mejor ajuste que el modelo binomial (cuyos resultados son ponderados por la inversa del factor utilizado para obtener la renta) se tendría dos maneras de obtener el gasto total de las personas. Sin embargo, esta visión es más limitada que el modelo GBM, ya que el gasto de las personas está condicionado a un factor que puede no ser correcto, en contraste al GBM, el cual estima directamente el nivel de gasto de las personas.

Ambos modelos son implementados en la herramienta de programación R 2.14.2.

6.2.1. Modelo Binomial Generalizado

El modelo binomial generalizado mostrado en [2], describe el número de transacciones que un cliente realiza en un periodo de tiempo dado. Para este valor, denotado por n_i , se asume que se distribuye como una variable aleatoria poisson de parámetro λ_i ,

$$f_1(n_i|\lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{n_i}}{n_i!} \quad (5)$$

En este modelo, n_i representa el *size of wallet* de un cliente para una categoría de producto dada. Si bien n_i cuenta el número de transacciones, éste representa una cantidad de dinero del gasto total, ya que como se muestra más adelante, cada transacción es un monto de gasto.

Dado que los clientes no son iguales, para incluir heterogeneidad el parámetro λ_i es modelado utilizando variables observables de la siguiente manera,

$$\lambda_i = \exp(\beta'V_i) \quad (6)$$

donde $V_i = (v_{i,1}, \dots, v_{i,L})$ son los regresores que influyen en el gasto total del individuo i .

En segundo lugar, se asume que el número de transacciones que el cliente i realiza en la empresa, denotado por x_i , sigue una distribución binomial,

$$f_2(x_i|\pi_i, n_i) = \binom{n_i}{x_i} \pi_i^{x_i} (1 - \pi_i)^{n_i - x_i} \quad (7)$$

En contraste a n_i , x_i es una variable observada por la empresa. La variable π_i corresponde a la probabilidad de que un cliente al hacer una transacción, lo realice utilizando la tarjeta de la tienda. Para estimar la probabilidad anterior, se utiliza una regresión logística, la cual se define de la siguiente manera,

$$\text{Ln}\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha'W_i \quad (8)$$

donde $W_i = (w_{i,1}, \dots, w_{i,J})$ son los J regresores que influyen en la probabilidad de que el individuo i utilice la tarjeta de crédito. Una vez que el parámetro α es estimado, la probabilidad de elección de la tarjeta en cada transacción se estima como sigue,

$$\hat{\pi}_i = \frac{1}{1 + e^{-\hat{\alpha}'w_i}} \quad (9)$$

Los parámetros α y β son estimados por máxima verosimilitud, donde la probabilidad de una observación individual es igual a

$$f(x_i|\alpha, \beta) = \int f_2(x_i|n, \alpha) f_1(n|\beta) dn \quad (10)$$

lo cual puede ser escrito de la siguiente manera

$$f(x_i|\alpha, \beta) = \sum_{n=x_i}^{\infty} \binom{n}{x_i} \pi_i^{x_i} (1 - \pi_i)^{n-x_i} \frac{e^{-\lambda_i} \lambda_i^n}{n!} \quad (11)$$

donde π_i es una función de α , y λ_i es una función de β . Sustituyendo $m_i = n - x_i$, se tiene lo siguiente

$$f(x_i|\alpha, \beta) = \frac{(\lambda_i \pi_i)^{x_i} e^{-\lambda_i \pi_i}}{x_i!} \quad (12)$$

Los valores estimados, $\hat{\alpha}$ y $\hat{\beta}$, se obtienen minimizando la siguiente función de verosimilitud,

$$-\log L(x_1, \dots, x_n|\alpha, \beta) \propto \sum_{i=1}^N [\lambda_i \pi_i - x_i \log(\lambda_i \pi_i)] \quad (13)$$

donde N es el número de clientes en la muestra.

Para predecir el *size of wallet* o gasto total en cada categoría, se sabe que $n_i \geq x_i$, dado que el número total de transacciones que un consumidor realiza en una categoría, es mayor o igual a lo que realiza sólo con la tarjeta de crédito. De esta manera, se tiene lo siguiente:

$$\hat{n}_i = E\langle n_i | n_i \geq x_i \rangle = \lambda_i \frac{1 - F(x_i - 2, \lambda_i)}{1 - F(x_i - 1, \lambda_i)} \quad (14)$$

donde $F(x, \lambda)$ es la función acumulada de una distribución poisson con parámetro λ en el valor x .

Para discretizar los montos promedios mensuales gastados con la tarjeta durante el periodo junio de 2011 a mayo de 2012, a modo de obtener x , se debe tener en cuenta que la estimación del *size of wallet* mediante la ecuación 14, queda indefinida con valores muy grandes de éste. Dicho lo anterior, en este estudio se opta por escoger la moda de estos montos promedios mensuales para discretizar, la cual representa el valor de cada transacción.

Debido a que el modelo utilizado es relativamente nuevo, siendo aplicado sólo en un trabajo de investigación en la industria bancaria, es necesario evaluar en una primera etapa su comportamiento. Para realizar esto, se desarrolla una simulación generando los datos a partir de las distribuciones mencionadas anteriormente y evaluando la capacidad predictiva del modelo. Esta simulación se repite mil veces, variando los valores simulados, de manera de identificar correctamente en qué casos el modelo se comporta de buena manera y en qué casos no tiene una buena capacidad predictiva.

Cabe mencionar que esta simulación sólo se realiza considerando el modelo binomial más simple (GBM transaccional), dado que el objetivo principal es ver que efectivamente el modelo tiene capacidad de predicción.

Una vez hecho lo anterior, se realiza la estimación utilizando los datos reales. El modelo binomial es dividido en dos: binomial generalizado transaccional y binomial

generalizado extendido. El GBM transaccional utiliza sólo variables transaccionales, donde tanto el vector que da cuenta del gasto total de un individuo (V_i), como el referido a la probabilidad de utilizar la tarjeta (W_i), están en función del tiempo que el cliente posee ésta y del gasto promedio realizado con ella en el periodo anterior al evaluado. Para el caso del modelo binomial generalizado extendido, el vector (W_i), será una función de variables transaccionales y de negocio, mientras que (V_i) estará en función de variables demográficas y transaccionales.

En primera instancia, se estimará el *share of wallet* de la tarjeta para cada uno de los clientes de la muestra, considerando el gasto agregado de todas las categorías. Luego, se estimará el *size of wallet* sólo para las categorías vestuario y hogar, debido a que son las categorías que más importancia tienen en el holding, y debido al hecho que el modelo no es exacto, estimar el potencial en cada categoría utilizando el modelo conllevaría a una gran cantidad de error, por lo que para obtener la principalidad en cada categoría se utilizará el procedimiento actual de la empresa, utilizando como base el potencial de gasto estimado de manera global con el modelo.

Cabe mencionar, que la decisión de utilizar los montos promedios mensuales en el periodo en el que se desea calcular la principalidad, se debe a que existen limitaciones de cupo que no permite a las personas comprar todo lo que deseen, por lo que al evaluar un horizonte de tiempo mayor se soluciona en gran medida el problema que se ocasionaría si se evaluase la principalidad mes a mes. Por lo que, al momento de calcular la principalidad, ésta corresponde a la principalidad promedio durante el periodo de tiempo evaluado.

6.2.2. Regresión Log-Lin

Este modelo es idéntico a una regresión lineal común en cuanto a estimación de parámetros y supuestos que deben cumplirse. La diferencia radica en que se quiere predecir el logaritmo de la variable dependiente. En este modelo la interpretación de sus coeficientes también es diferente, los cuales miden el porcentaje en que cambia la variable dependiente en caso de una variación absoluta de una variable independiente.

La forma funcional es la siguiente:

$$\text{Log}(y) = \alpha_0 + \alpha_1 * x_1 + \alpha_2 * x_2 + \dots + \alpha_n * x_n + \varepsilon \quad (15)$$

Se eligió este modelo en lugar de una regresión lineal simple debido al hecho de que se debe asegurar que la variable dependiente, en este caso la renta, sea siempre positiva.

El objetivo de este modelo es tener una comparación para el modelo GBM, donde la diferencia radica en que el GBM estima directamente el gasto total de las personas y la regresión estima la renta, la cual debe ser ponderada para obtener el gasto total de las personas bajo el supuesto de que estos factores son los correctos.

6.3. Metodología para evaluar la calidad del ajuste

Para evaluar la calidad de los modelos, se utiliza como indicador de error al MAPE de la estimación del *size of wallet*. Este indicador se define como sigue:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i} \quad (16)$$

donde Y_i corresponde al benchmark utilizado por la empresa, es decir, el ingreso real de los clientes presentes en la muestra, e \hat{Y}_i corresponde al potencial de gasto estimado de cada cliente de la muestra dividido por el porcentaje de gasto sobre el sueldo entregado por el INE . Se reportará el MAPE de ajuste y de validación.

Una segunda medida para evaluar la calidad del modelo, es el estadístico denominado *the likelihood ratio index* (ρ), el cual mide qué tan bien se ajusta el modelo a los datos con los parámetros estimados, en comparación a un modelo cuyos parámetros son todos ceros [5]. El ratio se define como sigue:

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(\mathbf{0})} \quad (17)$$

donde $LL(\hat{\beta})$ es el valor de la función verosimilitud evaluada en los parámetros estimados y $LL(\mathbf{0})$, representa la verosimilitud cuando todos los parámetros son fijados en cero. Los valores extremos para este indicador son 0 y 1, tomando el valor cero cuando el modelo genera los mismos resultados que sí no existiese un modelo asociado ($LL(\hat{\beta}) = LL(\mathbf{0})$) y toma el valor uno cuando la función de verosimilitud es máxima, es decir, $LL(\hat{\beta}) = 0$ [5].

Los indicadores anteriores se calculan por medio de las estimaciones que se realizan utilizando el 80% de la base de datos y también, como el promedio de diez valores provenientes de estimaciones realizadas en diez bases de datos más pequeñas, dado que para evaluar la calidad de los modelos se dividirá la base original en diez grupos donde cada uno de ellos representa el 10% de la base. El objetivo de lo anterior es utilizar la metodología *k-fold cross validation*, con k igual diez [6]. Se usa un *ten-fold cross validation* ya que el modelo GBM resulta ser sensible a los datos que se le entregan como input.

6.4. Metodología para estimar la principalidad

La principalidad en cada categoría será estimada como la razón entre el gasto realizado con la tarjeta en esa categoría y el potencial de gasto mensual en ella, debido a que de otra forma el indicador contendría mucho error por el hecho de que el resto de las categorías tendrían una estimación errónea.

De esta manera, la principalidad se obtiene utilizando el potencial de compra estimado por el modelo binomial generalizado, pero para obtener el potencial dentro de cada categoría, éste será ponderado por los porcentajes INE, el cual es asignado de acuerdo al quintil de pertenencia según el potencial de compra de cada uno de los clientes.

6.5. Metodología para segmentar a los clientes

Para generar información útil de gestionar, es importante realizar una segmentación de los clientes en base a la principalidad de cada uno de ellos con el objetivo de obtener información para proponer líneas de acción y formar grupos con igual principalidad, esto realizado por categoría y en torno a la tarjeta de manera global. Por medio de esta segmentación, es posible caracterizar a los clientes que presentan un gran potencial de crecimiento en el uso de la tarjeta, junto a los que no se les puede extraer más y sólo resta realizar políticas que mantengan su uso actual.

Para realizar lo anterior, en primer lugar se generan segmentos según quintiles de *share of wallet* y de *size of wallet*. Luego, para identificar a qué unidades de consumo se encuentran más vinculado cada uno de los clientes, se generan segmentos de acuerdo a la principalidad relativa en las categorías, donde cada usuario de la tarjeta es asignado a una unidad si su *share of wallet* en ella es mayor que su *share of wallet* en el resto de las categorías.

Finalmente, se realiza una segmentación utilizando el *k-means*, el cual se encuentra implementado en el programa SPSS. El objetivo de esto, es evaluar si por medio de esta herramienta estadística se puede encontrar aún más información que sea valiosa para generar propuestas de acción.

En específico, es interesante evaluar qué clientes tienen un potencial de rentabilidad alto para poder generar estrategias que permitan capturar más de su gasto por medio de la tarjeta en cada una de las categorías.

6.6. Propuestas de acción

Las directrices que guían las propuestas se basan en la información de los segmentos encontrados anteriormente. Estas propuestas tienen como objetivo aumentar el gasto de los clientes con la tarjeta de crédito, por lo que se pueden distinguir dos acciones principalmente:

- Cobertura: Esta acción busca traspasar clientes a segmentos más principales, lo cual se logra incitándolos a comprar en categorías que aún no lo hacen.
- Alcance: Esta acción busca aumentar el *reach* en las categorías menos principales para cada segmento de clientes, es decir, provocar que los clientes vinculados a una unidad de consumo, utilicen aún más la tarjeta en las categorías en las cuales son menos principales.

Lo anterior, es posible de lograr por medio de promociones focalizadas en cada categoría para cada segmento, incentivando así el uso de la tarjeta.

7. TRATAMIENTO DE LA BASE DE DATOS

En esta etapa se describen las diferentes acciones que se realizaron en la base de datos antes de comenzar a desarrollar los modelos.

7.1. Selección de los datos

Actualmente, la empresa posee alrededor de 2,7 millones de clientes titulares activos. Globalmente, se tienen dos bases de datos, una que registra las transacciones de los clientes en cada uno de los negocios (Holding, Alianzas y comercios asociados a Visa) y otra con información de cada uno de sus clientes, como lo son datos demográficos, de comportamiento y de negocio. La muestra inicial de clientes otorgada por la empresa, consta de 220.994 clientes, los cuales fueron seleccionados de manera aleatoria, cuidando que sólo fueran clientes que actualmente pueden realizar transacciones con la tarjeta y que tuvieran la variable renta asociada.

El periodo de estudio para este grupo de clientes abarca las transacciones realizadas durante 24 meses, desde junio de 2010 a mayo de 2012. Como etapa preliminar, cada una de las transacciones realizadas son asignadas a alguna de las 9 categorías INE (alimentación, vestuario, vivienda, hogar, salud, transporte, recreación, educación, otros), donde la asignación de cada compra a una de estas categorías, es realizada según criterio de la empresa de acuerdo a la sub línea de compra.

Dentro de las transacciones que hizo cada uno de los clientes de la muestra en cada mes, se seleccionaron a aquellas que tenían asociada alguna de las categorías INE, descartando aquellas que presentaban un monto negativo o estaban asociadas a promociones y donaciones.

Una vez hecho lo anterior, se procede a agregar las bases provenientes tanto de los distintos negocios del holding como fuera de éstos, logrando obtener montos de gasto mensual para cada cliente en cada una de las 9 unidades de consumo. Posterior a ello, se cruza la base de datos transaccional con la que presenta información demográfica y de comportamiento del cliente, base de datos que será procesada como se describe en la sección 7.2.

Dentro de la totalidad de las variables, se deben elegir aquellas que puedan tener influencia tanto en el total de compras que realiza una persona en un periodo de tiempo dado, como en las que sólo hizo al interior del holding. Las variables transaccionales escogidas que podrían entregar información, utilizadas para realizar los análisis, son las siguientes:

- **Monto_{it}**: Cantidad de dinero gastado usando la tarjeta en la categoría *i* (categoría INE o gasto total con la tarjeta) en el periodo de tiempo *t*.
- **Fecha última compra_{it}**: Periodo en que realizo la última compra en categoría *i* con la tarjeta.

Dentro de las que se relacionan al negocio y al comportamiento del cliente, se mencionan las siguientes:

- **Cupo:** Cupo que posee la tarjeta.
- **Tipo de tarjeta:** Tipo de tarjeta que posee el cliente a mayo de 2012.
- **Fecha de apertura:** Fecha en la que abrió la cuenta.
- **Tipo de cliente:** Según segmentación de la empresa, clientes Elite, Premium y Normal.

En cuanto a las variables demográficas, se optó por incluir todas, excepto el grupo socioeconómico⁵, ya que según [7], los consumidores con diferentes características poseen distintos patrones de comportamiento de compra. Las variables incluidas son las siguientes:

- **Sexo.**
- **Edad.**
- **Zona de residencia.**

Como proxy para el gasto total y a modo de evaluar si las variables escogidas aportan información relevante, se escogió para cumplir con esta función al gasto total realizado con la tarjeta durante el periodo junio de 2011 a mayo de 2012. Debido a que sólo se posee información de comportamiento del cliente con la empresa, parece razonable suponer que las variables que influyen en el gasto total también deberían hacerlo en el gasto total realizado con la tarjeta.

7.2. Procesamiento de los datos

Dentro de la base generada, la cual en un principio contenía 216.965 clientes activos, existen errores en los datos producto de valores faltantes en otras variables relacionadas, o datos que se encuentran fuera de un rango aceptable. Además, existen registros que no son útiles para la realización del modelo, los cuales deben ser eliminados por no aportar información. Para realizar las estimaciones, es necesario hacer una limpieza y corrección de manera de impedir que algunos valores generen ruido en los resultados.

Como punto de partida, se identifican clientes que no tienen asociada una fecha de nacimiento, provocando que en su edad aparezca el valor "2012". Estos registros fueron eliminados, los cuales representan un 0,015% del total de la muestra.

Debido a que se estimará la principalidad para el periodo junio de 2011 a mayo de 2012 de manera anual, y dado que para poder realizar una estimación acerca del gasto total que realiza un cliente en un periodo de tiempo dado se requiere historia de su comportamiento, los clientes que abrieron la tarjeta durante el año en que se realizará la estimación fueron eliminados del análisis. La cantidad de este tipo de clientes en la muestra son 20.468, que representan un 9% de la muestra inicial.

⁵ El GSE no se considera debido a que existe una gran cantidad de clientes que no poseen esta información. Además esta variable fue sometida a correcciones, por lo que no resulta ser muy confiable.

Por otro lado, según lo conversado con la empresa, los clientes que no hayan realizado compras durante los dos años de estudio (junio 2010 a mayo 2012), no se les calculará la principalidad dado que, si bien aún tienen la tarjeta habilitada para realizar compras, son clientes ya fugados. El número de registros eliminados debido a esta razón corresponden a 7.856, que representan un 3,6% de la base original.

Para la detección de los valores fuera de rango que pueden generar distorsiones en la estimación, se evaluó el monto total gastado con la tarjeta en el periodo junio 2010 a mayo 2011 y de junio 2011 a mayo 2012, donde la primera será utilizada como uno de los regresores para la estimación del *size* y *share of wallet*, y la segunda es utilizada para obtener el número de transacciones que realiza cada cliente. Para realizar esta evaluación, se consideró sólo a las personas que realizaron algún gasto con la tarjeta durante el periodo en evaluación, generando 200 grupos con igual cantidad de personas en cada uno de ellos, ordenados de mayor a menor gasto. La situación inicial para el grupo de mayor y de menor gasto para el primer periodo, presenta las siguientes características:

Grupo	Media	Mínimo	Máximo
Grupo con mayor gasto	\$ 15.225.000	\$ 8.182.386	\$ 199.497.000
Grupo con menor gasto	\$ 4.309	\$ 110	\$ 6.960

Tabla 1: Descripción del grupo superior e inferior en gasto con la tarjeta en el primer año de estudio

Fuente: Elaboración propia en base a datos de la empresa

La situación durante el segundo periodo es presentada a continuación:

Grupo	Media	Mínimo	Máximo
Grupo con mayor gasto	\$ 18.934.000	\$ 10.443.605	\$ 310.000.000
Grupo con menor gasto	\$ 4.440	\$ 1	\$ 7.990

Tabla 2: Descripción en el segundo año de estudio

Fuente: Elaboración propia en base a datos de la empresa

Según lo conversado con la empresa, en ambas tablas los grupos se escapan de lo que se puede considerar normal en el uso de la tarjeta, razón por la cual los cuatro grupos, donde cada uno de estos representa el 0,5% de los clientes con montos positivos, fueron eliminados. El total de registros eliminados por este motivo asciende a 4.596, lo que representa un 2% de la muestra inicial.

Los grupos de mayor y menor gasto luego de realizar lo anterior, tienen la siguiente descripción:

Grupo	Media	Mínimo	Máximo
Grupo con mayor gasto	\$ 6.906.000	\$ 6.018.612	\$ 8.178.103
Grupo con menor gasto	\$ 9.367	\$ 6.970	\$ 11.970

Tabla 3: Descripción de los nuevos grupos una vez eliminados los registros fuera de rango en el primer periodo

Fuente: Elaboración propia en base a datos de la empresa

Grupo	Media	Mínimo	Máximo
Grupo con mayor gasto	\$ 8.679.800	\$ 7.556.422	\$ 10.442.753
Grupo con menor gasto	\$ 10.397	\$ 7.990	\$ 12.000

Tabla 4: Descripción de los nuevos grupos en el segundo periodo

Fuente: Elaboración propia en base a datos de la empresa

Se puede observar que los rangos para cada uno de los nuevos grupos resultan ser más aceptables que lo que se tenía anteriormente, logrando obtener una gran varianza en esta variable pero dentro de rangos aceptados como normales.

Para obtener el número de transacciones que realiza un cliente en el segundo periodo, se trabaja con el monto promedio mensual, por lo que se debe extraer de la muestra a aquellos clientes que presenten un monto promedio muy alto o muy bajo. El valor máximo que es utilizado para poder calibrar los modelos es de \$700.000, por lo que montos superiores a éste son considerados valores fuera de rango. De manera similar, el mínimo aceptable es de \$5.000. De acuerdo a este criterio, el número de registros eliminados fue de 7.740, que representa un 3,6% de los datos iniciales.

Finalmente, el último filtro realizado en la base de datos, cuyo objetivo es evaluar la capacidad predictiva del modelo, fue eliminar a clientes con una renta mayor a \$1.500.000 y menor a \$180.000, los cuales representan un 7% de la base de datos inicial, es decir, 15.268 registros eliminados. Este rango de renta fue el utilizado por la empresa para evaluar la capacidad predictiva de su modelo, y dado que se usa el mismo benchmark para evaluar la capacidad predictiva del modelo desarrollado en este proyecto, se debe utilizar el mismo criterio. Además, el rango aceptable elegido por la empresa se debe a que los clientes cuya renta cumple con el criterio, representan el 93% del total de clientes que poseen esta variable asociada.

Una vez realizado este procesamiento de datos, la muestra final con la cual se trabajará corresponde a 161.005 clientes, que representa un 74,2% de la muestra inicial.

7.3. Transformación de los datos

Luego de ser realizada la selección y limpieza de los datos, es importante generar, a partir de los datos disponibles, nueva información que no se encuentra de manera explícita en la base de datos y que es útil para desarrollar la metodología que será propuesta. Como primera transformación, se eliminaron las transacciones que aparecían con montos positivos en periodos anteriores a la fecha de apertura. Estos casos ocurren debido a que existen clientes que cerraron la tarjeta y luego la volvieron a abrir, por lo que se tomó como primera fecha de apertura la que actualmente registra el sistema.

Una segunda etapa, fue la de crear variables que no se encuentran explícitas en la base de datos y que serán utilizadas en los modelos. Dentro de éstas, se encuentran los montos gastados de manera agregada en cada una de las categorías de consumo y de manera global con la tarjeta en cada año de estudio (junio 2010-mayo 2011 y junio 2011-mayo 2012) para cada uno de los clientes de la muestra. Además, debido a que el

tipo de tarjeta indicado en la base de datos es la que a mayo de 2012 poseía el cliente, se debe generar para cada mes y categoría el tipo de tarjeta que el cliente tenía en ese periodo, generando luego el monto total que gastó en cada mes según el tipo de tarjeta que poseía.

Debido a que existen desactualizaciones en datos como antigüedad y fecha de última compra del cliente, se genera una fecha de última compra para cada categoría de consumo y para la tarjeta como gasto total. Además, se crea nuevamente la variable antigüedad del cliente, para poder generar variables transaccionales relacionadas al gasto promedio en un periodo de tiempo determinado.

Por otro lado, se generó una variable de recencia, la cual indica cuánto tiempo ha transcurrido desde la última compra del cliente a mayo de 2011 y otra para mayo de 2012.

En esta etapa, también se generaron variables dummy para las diferentes zonas de residencia y para el sexo. Finalmente, otra variable dummy creada, tiene que ver con el tipo de cliente, hecha según una segmentación realizada por la empresa.

7.4. Análisis descriptivo y evaluación de las variables seleccionadas

Para comprender cómo está constituida la base con la que se trabaja, se realiza un análisis descriptivo de las variables más relevantes. Además, para decidir acerca de qué variables introducir en el modelo, se realizan test anova de un factor entre las medias de los diferentes grupos en cuanto al gasto total realizado con la tarjeta en el año 2, ya que como se mencionó en la sección 7.1, éste será el proxy al gasto total realizado con cualquier medio de pago por un cliente para poder evaluar las variables.

Al estudiar la distribución de clientes por género al interior de la base de datos, se observa que un 60% son hombres y un 40% mujeres. Luego, al observar como fue la distribución del gasto total realizado con la tarjeta durante el periodo junio 2011-mayo 2012, se puede ver que los hombres representan el 59% sobre un total de \$152.859.000.000, con un valor igual a \$90.536.000.000. Al evaluar si las medias entre estos dos grupos de clientes presentan diferencias significativas mediante un análisis anova de un factor, se observa que efectivamente existen diferencias entre las medias de éstos con un p-valor de 0,01, por lo que la variable sexo será incluida como una de las variables para predecir el *size of wallet*.

Como segunda variable para describir los datos y evaluar si la variable influye en el gasto total de las personas, se estudia la zona de residencia. El lugar de residencia fue agrupado en 8 zonas distintas:

- Santiago Oriente.
- Santiago Sur-Oriente.
- Santiago Centro.
- Santiago Norte.
- Santiago Poniente.

- Santiago Sur.
- Región Metropolitana fuera de Santiago.
- Regiones.

La distribución de los clientes según esta variable es la siguiente:

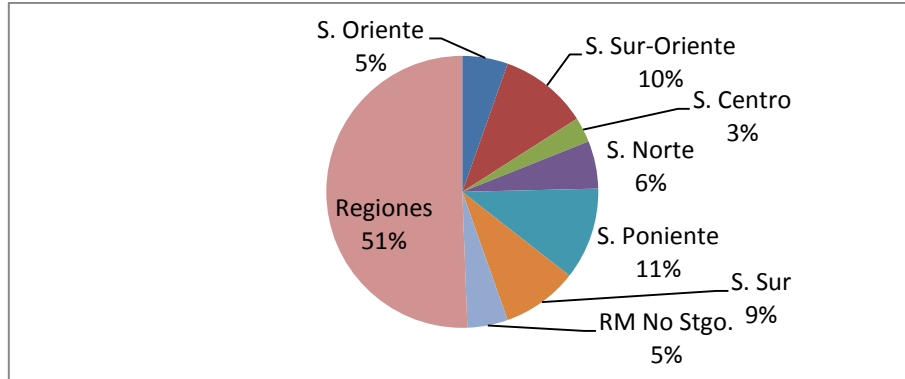


Gráfico 2: Distribución de clientes por zona

Fuente: Elaboración propia en base a datos de la empresa

Al estudiar cómo se comporta el gasto promedio durante el último año en cada una de las zonas descritas anteriormente, se observa que éste es muy similar entre zonas, siendo levemente superior la zona Santiago Centro.

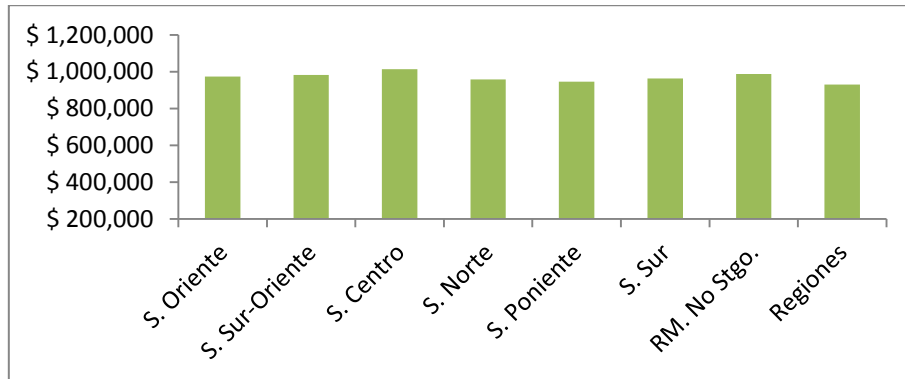


Gráfico 3: Gasto promedio por zona con la tarjeta durante el último año

Fuente: Elaboración propia en base a datos de la empresa

Al evaluar si existen diferencias significativas entre las zonas de residencia en torno al gasto realizado el último año, se desprende que estadísticamente son diferentes de acuerdo a un test anova de un factor, obteniendo un p-valor de cero, razón por la cual también será incluida como una de las variables que explican el gasto total realizado.

Para evaluar la composición en la base de datos según el rango edad, se generaron 9 grupos etarios, cuya distribución en la base de datos es la siguiente:

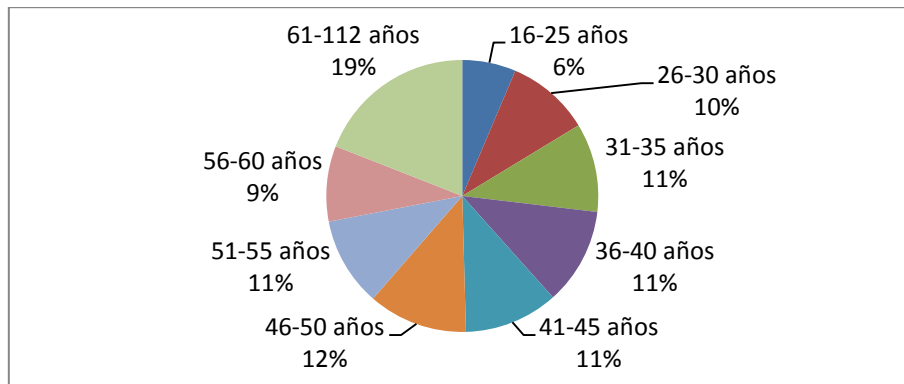


Gráfico 4: Distribución de la base de datos según rango de edad
 Fuente: Elaboración propia en base a datos de la empresa

Al observar el gasto promedio durante el último año en cada uno de estos grupos, se observa que a medida que el cliente envejece, gasta más, volviendo a decrecer su gasto pasado los 50 años.

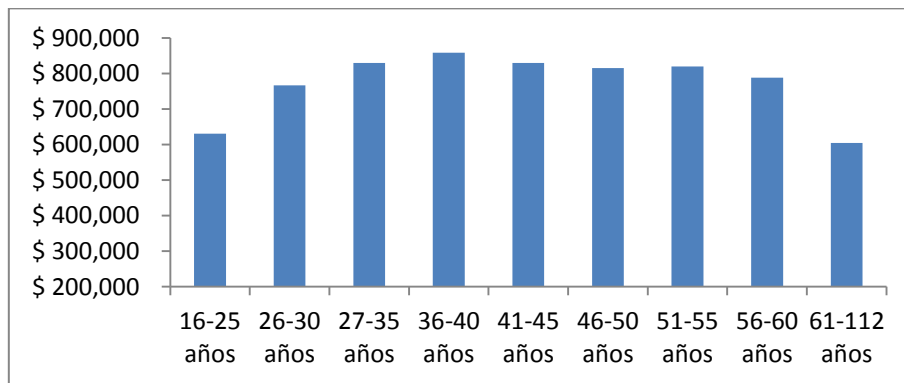


Gráfico 5: Gasto promedio durante el último año según rango de edad
 Fuente: Elaboración propia en base a datos de la empresa

Al realizar el test anova de un factor, se evidencia que efectivamente existen diferencias, arrojando un p-valor igual a cero, razón por la cual esta variable también será utilizada para predecir el gasto total de las personas.

Es interesante evaluar cómo ha sido el gasto con la tarjeta durante los dos últimos años, evidenciando que si bien el comportamiento es similar en ambos años, el gasto realizado durante el segundo año ha sido superior.

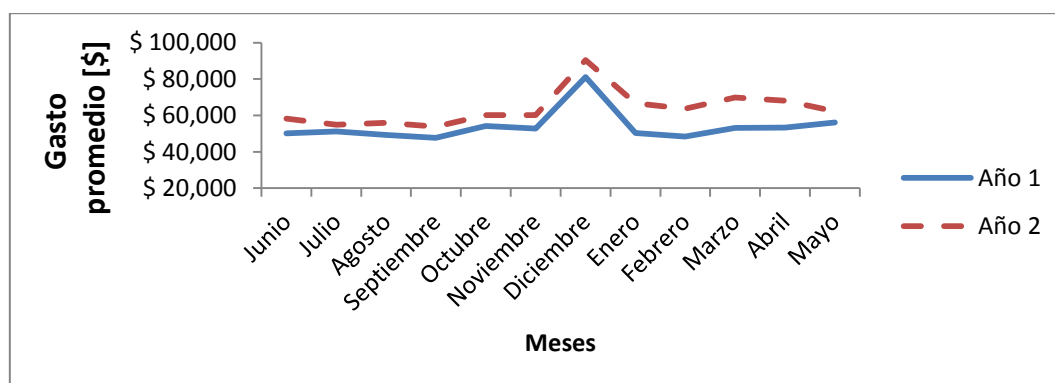


Gráfico 6: Comparativa del gasto promedio por cliente con la tarjeta en ambos periodos de estudio
Fuente: Elaboración propia en base a datos de la empresa

Lo anterior se puede deber principalmente a que más clientes han migrado a la tarjeta de crédito que incorpora Visa, lo cual les da la oportunidad de realizar compras en una mayor cantidad de lugares y categorías de consumo.

Debido a que el comportamiento en el uso de la tarjeta entre un año y otro es similar, el gasto en el periodo anterior al de estimación de la principalidad será utilizado tanto para la predicción del *size of wallet* como del *share of wallet*, ya que esta variable describe como es la interacción del cliente con la empresa.

Actualmente, la empresa tiene una segmentación para sus clientes según la cantidad de puntos acumulados. La segmentación es la siguiente:

Tipo de cliente	Rango de puntos acumulados
Elite	Mayor a 50.000 puntos
Premium	Entre 15.000 y 50.000 puntos
Normal	Menor a 15.000 puntos

Tabla 5: Segmentación de clientes según puntos acumulados

Fuente: Elaboración propia en base a datos extraídos de la pagina de la empresa

Además, en la categoría normal, la empresa los divide en 3 grupos según su antigüedad, distinguiéndose así los que llevan más de 3 años, los que llevan entre 1 y 3 años y los que llevan menos de un año.

Al estudiar cómo ha sido el gasto promedio mensual dentro de cada uno de estos grupos en el segundo año en función de la recencia, se evidencia una gran diferencia de los Elite en comparación a los otros dos grupos. En primer lugar, se hará la comparación sólo con los clientes que poseen la tarjeta normal. Al observar el Gráfico 7 mostrado a continuación, se observa que si bien el número de clientes Elite que utilizan la tarjeta son muy pocos, son los que tienen un gasto promedio mucho mayor al resto, con una muy baja recencia. Es por la razón anterior que, para el holding, estos clientes resultan ser los más valiosos, puesto que son los que más gastos realizan, pero para la empresa de la tarjeta son los clientes más caros, dado que si bien compran mucho con la tarjeta, generalmente es al contado, por lo que acumulan muchos puntos y no generan intereses.

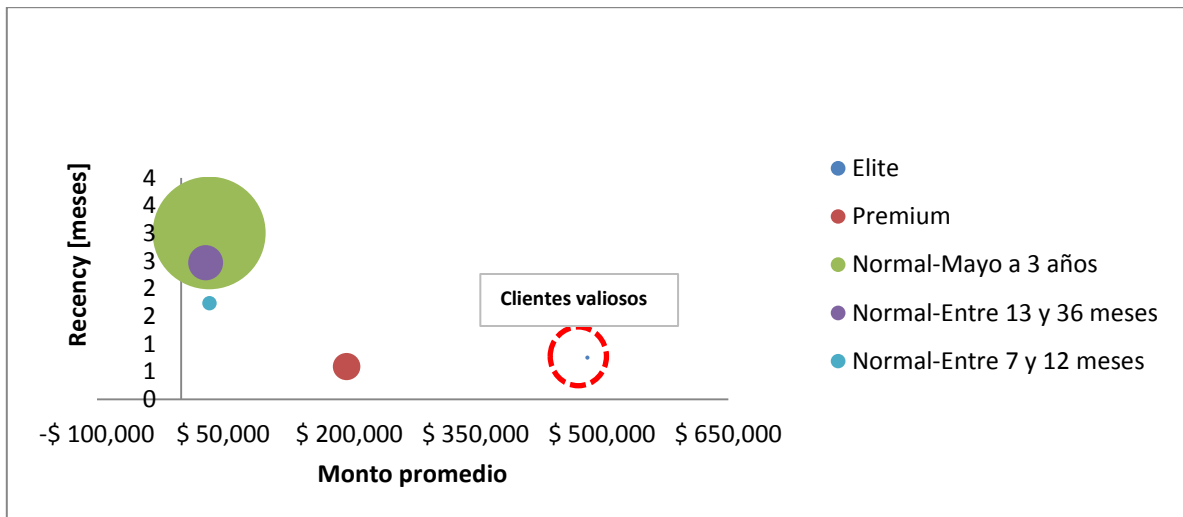


Gráfico 7: Monto promedio mensual gastado con la tarjeta vs recencia según tipo de cliente
 Fuente: Elaboración propia en base a datos de la empresa

Algo similar ocurre al hacer la comparación entre estos grupos, considerando sólo a aquellos que poseen la tarjeta Visa, con la salvedad de que los montos promedios son superiores a los que utilizan la tarjeta normal, como fue mostrado anteriormente.

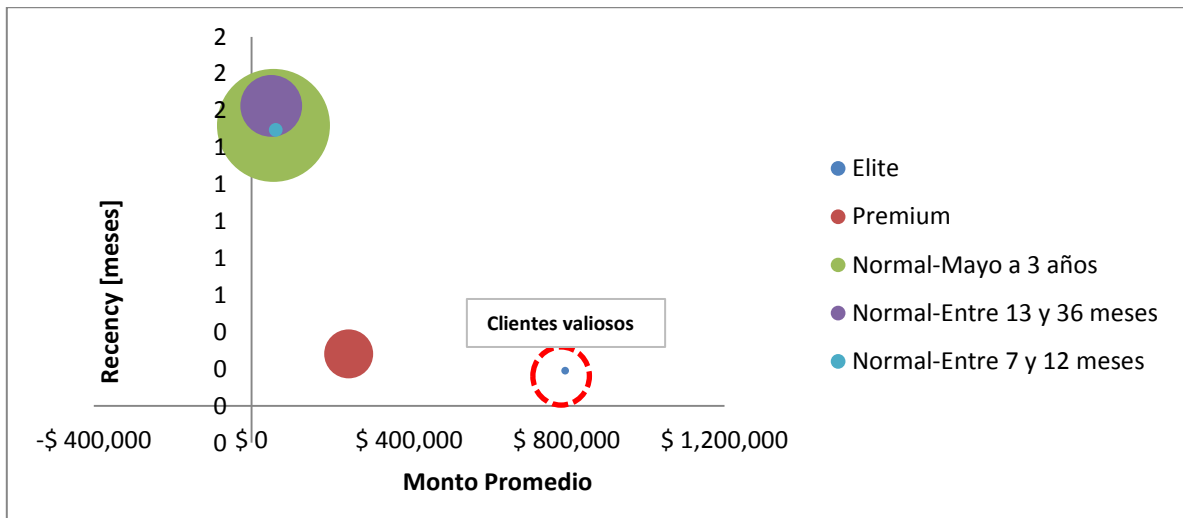


Gráfico 8: Monto promedio mensual gastado con la tarjeta Visa vs recencia según tipo de cliente
 Fuente: Elaboración propia en base a datos de la empresa

Debido a que claramente existen diferencias al momento de utilizar la tarjeta en cada uno de estos grupos, esta segmentación será utilizada como variable para predecir el *share of wallet*.

8. DESARROLLO DEL MODELO

En este capítulo se muestra la estimación del potencial de compra de cada uno de los clientes presentes en la muestra utilizando el modelo GBM transaccional, GBM extendido y la regresión Log-Lin.

Con el objetivo de evaluar el comportamiento del modelo GBM, en un comienzo se realiza una simulación de los datos para evaluar la capacidad de éste para recobrar los valores simulados, además de determinar bajo qué condiciones tiene una mejor capacidad predictiva. Esto sólo se realiza con el GBM transaccional, de manera de demostrar que tiene capacidad predictiva.

Posterior a ello, se realiza una estimación utilizando la regresión Log-Lin, el GBM transaccional y luego el GBM extendido, evaluando el ajuste y comparándolo con el rendimiento del actual modelo que utiliza la empresa. Luego, se elige el que presenta un mejor rendimiento.

8.1. Simulación

Debido a que el modelo GBM es un modelo nuevo en la estimación del *share of wallet*, es necesario evaluar la capacidad de predicción que posee éste mediante una simulación.

Para el desarrollo de la misma, se trabaja con mil clientes ficticios, donde las variables de cada uno de ellos son simuladas bajo distribuciones uniformes, considerando los rangos reales presentes en la muestra de datos. Una vez obtenidas las variables para cada uno de los mil clientes, es posible simular a partir de las distribuciones correspondientes los valores de n_i y x_i (poisson y binomial respectivamente).

Una vez que se tienen los valores de n_i y x_i , estos pueden ser considerados como el número de transacciones reales que realizan cada uno de los clientes. Por lo tanto, lo que sigue es aplicar el modelo utilizando las variables para cada uno de los clientes junto a x_i , para luego evaluar la capacidad del modelo para recobrar n_i .

Los pasos anteriores se repiten mil veces variando los valores simulados, por lo tanto, existen mil grupos donde cada uno de ellos contiene mil clientes. Esto, se realiza para poder evaluar bajo qué condiciones el modelo se comporta de mejor manera.

La implementación de la simulación se realiza en la herramienta de programación R.

8.1.1. Resultados de la simulación

Al evaluar la capacidad predictiva del modelo en el contexto de la tarjeta de crédito por medio de la simulación, se logra evidenciar que este modelo tiene una mejor capacidad predictiva cuando tanto el número de transacciones realizadas con la tarjeta como el número de transacciones realizadas con cualquier medio de pago van aumentando. Se puede observar lo anterior en el Gráfico 9, donde se muestra el x_i promedio de cada una de las 1000 repeticiones, evidenciando que a medida que el número de transacciones disminuye, el MAPE aumenta.

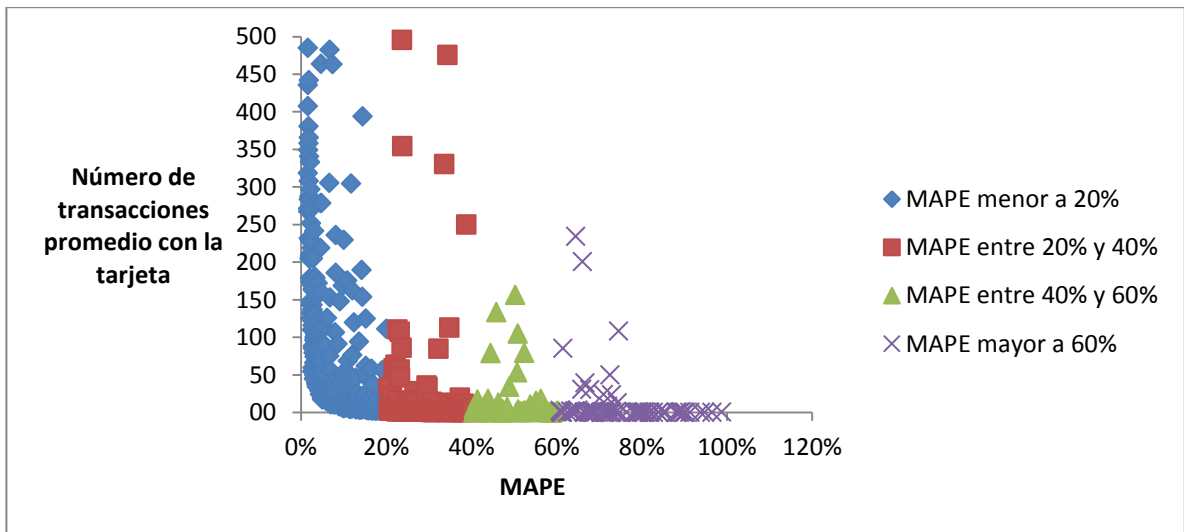


Gráfico 9: Número de transacciones promedio con la tarjeta como función del MAPE en cada repetición

Fuente: Elaboración propia

En el Gráfico 10 se evidencia la misma situación anterior, pero considerando el número de transacciones que realiza una persona utilizando cualquier medio de pago. En el gráfico se muestra el n_i promedio de cada repetición en función del MAPE asociado. Cabe mencionar que el n_i mostrado es el obtenido a partir de la distribución y no el estimado por el modelo.

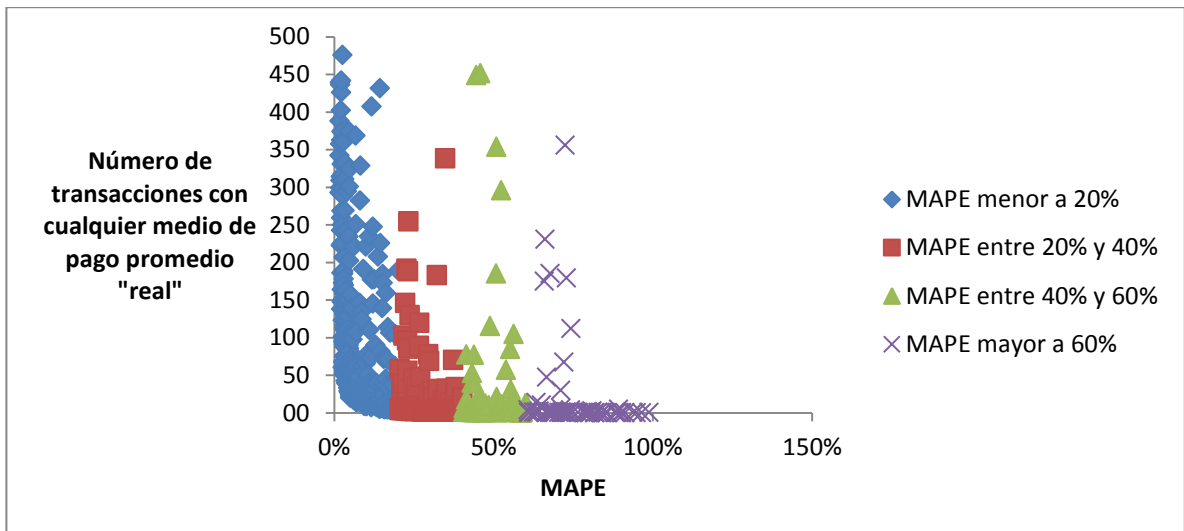


Gráfico 10: Número de transacciones promedio con cualquier medio de pago como función del MAPE en cada repetición

Fuente: Elaboración propia

El gráfico que muestra a n_i en función de x_i según el MAPE obtenido, es mostrado en el Anexo A. En aquel gráfico es posible observar no sólo lo mencionado anteriormente, sino que además, el modelo GBM transaccional no logra predecir de buena manera cuando los clientes realizan un número de compras mucho mayor con cualquier medio de pago que los realizados con la tarjeta, es decir, a aquellos que poseen un bajo *share of wallet*.

8.2. Modelo Log-Lin

El primer modelo que se evalúa, corresponde al modelo log-lin. En este modelo se incorporaron todas las variables que podrían influir, tanto el *share of wallet* como en el *size of wallet*, de manera de encontrar la mejor especificación.

Este modelo, al igual que los que se presentarán en las siguientes secciones, fue desarrollado utilizando la muestra de 161.005 clientes, los cuales fueron divididos en diez grupos donde cada uno de estos representan el 10% de la muestra, con el objeto de realizar un *ten-fold cross validation*.

A continuación se muestra la especificación del modelo:

$$\ln(Renta_i) = \alpha_0 + \alpha_1 * MP1 + \alpha_2 * Te + \alpha_3 * A + \alpha_4 * Cupo + \alpha_5 * A^2 + \alpha_6 * R + \alpha_7 * D_{genero} + \alpha_8 * D_{zona\ i} \quad (18)$$

donde {MP1} corresponde al monto promedio gastado con la tarjeta el periodo 1, {Te} corresponde a la antigüedad, {A} corresponde a la edad, {R} a la recencia, {D} a variable dummy, donde **zona_i** hace referencia a que corresponde a 7 dummies de zona, con i desde 1 a 7. La incorporación de la variable **edad²** se debe a que como se mostró en el Gráfico 5, existe una forma casi cuadrática en el gasto en función de la edad.

El resumen del *ten-fold cross validation* es mostrado en la siguiente tabla, donde el detalle se puede ver en el Anexo B.

Estadística	Valor
MAPE ajuste	36%
MAPE validación	36%
R ² ajustado	20,5%

Tabla 6: Estadísticas para regresión Log-Lin utilizando *cross validation*

Fuente: Elaboración propia

Se puede observar que el ajuste de este modelo es 11 puntos peor que el actual procedimiento que utiliza la empresa a nivel de renta, como fue mostrado en la sección 5.4. Esto se debe a que la variabilidad existente en la renta de los clientes no es posible modelarlo con una regresión de este tipo, ya que no captura heterogeneidad con la forma funcional que posee. Además, la cantidad de varianza explicada por el modelo es solo de un 20,5%, lo que confirma lo mencionado anteriormente.

Finalmente, al aplicar este modelo utilizando la totalidad de los clientes de la muestra, divididos en ajuste (80%) y validación (20%), se observa que se corrobora lo arrojado por la metodología *ten-fold cross validation*, como es mostrado en la Tabla 7.

Estadística	Valor
MAPE ajuste	36,2%
MAPE validación	36,1%
R ² ajustado	20,4%

Tabla 7: Estadísticas para regresión Log-Lin con base de muestra global

Fuente: Elaboración propia

En la Tabla 8 se muestra la estimación de los coeficientes para las 15 variables utilizadas en el modelo.

Coeficiente	Valor	Desviación estándar	Significancia al 95 % (p-valor<0.05)
α_0	12,2000000	0,01280000	0,00
α_1	0,00000005	0,00000002	0,00
α_2	-0,0003690	0,00001790	0,00
α_3	0,0264000	0,00057700	0,00
α_4	0,00000002	0,00000000	0,00
α_5	-0,0002670	0,00000603	0,00
α_6	0,0033700	0,00037200	0,00
α_7	0,0886000	0,00247000	0,00
α_8	0,1850000	0,00549000	0,00
α_9	0,0008540	0,00406000	0,83
α_{10}	0,1030000	0,00712000	0,00
α_{11}	-0,0131000	0,00528000	0,01
α_{12}	0,0002510	0,00399000	0,95
α_{13}	-0,0221000	0,00432000	0,00
α_{14}	-0,0342000	0,00571000	0,00

Tabla 8 Resumen coeficientes estimados para modelo Log-Lin
Fuente: Elaboración propia

En la tabla anterior, se puede observar que dos de las quince variables resultan no ser significativas, las cuales corresponden a dos zonas de residencia (zona Sur-Oriente y zona Santiago-Poniente). Al analizar los signos de las variables que a priori se conoce el efecto sobre la renta, se observa que en su mayoría son consistentes. Por ejemplo, al analizar la variable cupo (coeficiente α_4), se observa que si éste aumenta en \$100.000, la renta aumenta en un 2%, por lo que la relación es correcta, ya que el cupo que una persona posee en su tarjeta de crédito está vinculado positivamente con la renta que ésta percibe. Al mirar los coeficientes asociados a la edad (α_3 y α_5), se observa que a medida que aumenta ésta, la renta también lo hace, pero dado que el coeficiente de la edad al cuadrado es negativo (coeficiente α_5), se tiene que existe una relación cóncava, es decir, a partir de cierto punto, a medida que la persona envejece, la renta percibida disminuye, lo cual también es cierto ya que en general las personas que se encuentran en un rango de edad intermedio perciben más ingresos en promedio que los adultos mayores y los más jóvenes.

Otra variable que se conoce su efecto en la renta promedio, resulta ser el género (coeficiente α_7), donde se confirma que en promedio los hombres ganan más que las mujeres, y el modelo arroja que los hombres ganan un 8% más que el sexo femenino.

Si bien el modelo es consistente con lo esperado en cuanto a los signos de las variables, su poder predictivo y de ajuste es más bajo en comparación al modelo de la empresa, y dado que su nivel de varianza explicado también es bajo, no se gana información respecto a la situación actual de la compañía.

8.3. Modelo Binomial Generalizado transaccional

El segundo modelo a evaluar, corresponde al GBM transaccional. En éste, como se mencionó en el marco metodológico, el monto promedio mensual gastado por un cliente con la tarjeta durante el periodo junio de 2011 a mayo de 2012, debe ser discretizado para poder obtener el valor de x_i . El monto utilizado en la calibración de este modelo es de \$30.000, ya que éste es uno de los montos mayor a cero con frecuencia alta dentro de la muestra de clientes, por lo que es considerado como uno de los montos promedios “típico” al interior de la muestra.

Las variables explicativas utilizadas en este modelo son:

- El tiempo que cada cliente llevaba en la compañía hasta mayo de 2011 (*Antigüedad*)
- El monto promedio mensual de gasto durante el periodo anterior al que se quiere calcular la principalidad (junio de 2010 a mayo de 2011).

La forma funcional explícita para el caso de la tasa de gasto es la siguiente:

$$\lambda_i = \exp \left(\beta_0 + \beta_1 * \log(MP1_i + 1) + \beta_2 * \frac{((Antigüedad) - \min(Antigüedad))}{(\max(Antigüedad) - \min(Antigüedad))} \right) \quad (19)$$

Para el caso de la probabilidad de elección de la tarjeta de crédito, la especificación es la siguiente:

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \alpha_0 + \alpha_1 * \frac{((Antigüedad) - \min(Antigüedad))}{(\max(Antigüedad) - \min(Antigüedad))} + \alpha_2 * \frac{((MP1) - \min(MP1))}{(\max(MP1) - \min(MP1))} \quad (20)$$

La decisión acerca de normalizar la variable *Antigüedad* (variable que se encuentra en meses), se debe a que el modelo es mucho más estable en cada uno de los grupos de ajuste y calibración. Lo mismo sucede con la variable monto promedio en la ecuación 20.

En la Tabla 9 se muestran los valores estimados para los coeficientes de cada una de las cuatro variables:

Coeficiente	Valor	Desviación estándar	Significancia al 95 % (p-valor<0.05)
β_0	2,260432	0,01459	0,00
β_1	-0,021154	0,00103	0,00
β_2	0,752739	0,01731	0,00
α_0	-1,374063	0,01303	0,00
α_1	-1,053775	0,02751	0,00
α_2	9,150317	0,07217	0,00

Tabla 9: Resumen coeficientes estimados para GBM transaccional

Fuente: Elaboración propia

Al observar esta tabla, se ve que todas las variables resultan ser significativas con un 95% de confiabilidad. Ahora, al analizar los signos de las variables, se puede observar que la gran mayoría resulta acorde a lo esperado. En el caso de las variables asociadas a la tasa de compra, se esperaría que tanto la constante como la asociada al monto gastado aporten positivamente, ya que de acuerdo al monto promedio que gasta una persona de forma mensual, ésta debe aportar positivamente en la tasa de transacciones que esa persona realiza, y la constante también debería ser positiva, ya que es de esperar que la tasa de compra sea al menos, mayor a uno mensualmente. Es por lo anterior que el parámetro β_1 tiene un signo contrario a lo esperado, ya que el modelo indica que a medida que el monto gastado en el periodo uno con la tarjeta es mayor, la tasa de compra global con cualquier medio de pago disminuye. Con respecto a la probabilidad de elegir la tarjeta en cada transacción, se observa que mientras mayor tiempo se posea la tarjeta, es menos probable que ésta sea utilizada (α_1 es menor a cero). Si bien lo anterior no es intuitivo, esto podría ser explicado por el hecho de que al momento de sacar una tarjeta, existe un motivo detrás, lo que genera que haya un uso intensivo, y en el mediano y largo plazo, el uso de ésta se estabilice. Finalmente, se observa que el monto promedio gastado en el periodo uno influye positivamente en la probabilidad de uso de la tarjeta, lo cual es consistente con el hecho de que las personas con un monto de gasto mayor, tienen mayor tendencia a utilizarla.

El resumen de las estadísticas arrojado por el *ten-fold cross validation* es mostrado en la Tabla 10. El detalle de esto se encuentra en el Anexo C.

Estadística	Valor
<i>MAPE ajuste</i>	26,57%
<i>MAPE validación</i>	26,55%
AIC	580.002
BIC	580.012
ρ^2	63,96%

Tabla 10: Estadísticas para GBM transaccional utilizando *cross validation*

Fuente: Elaboración propia

Nuevamente, al utilizar completamente la base de clientes utilizada en esta memoria, dividida en base de ajuste y de validación, se corroboran los resultados entregados en el *ten-fold cross validation*, como se puede ver en la Tabla 11. Cabe

mencionar que los coeficientes mostrados en la Tabla 9 fueron estimados utilizando la base de ajuste (80% de la base total de la muestra).

Estadística	Valor
<i>MAPE ajuste</i>	26,79%
<i>MAPE validación</i>	26,71%
AIC	514.556
BIC	514.566
ρ^2	63,85%

Tabla 11: Estadísticas para GBM transaccional con base de muestra global
Fuente: Elaboración propia

Como se muestra en la tabla anterior, el MAPE obtenido al utilizar el mismo benchmark que el de la empresa es de un 26%, lo cual es 1 punto peor que el modelo actual que posee la compañía. El nivel de ajuste es mostrado en el Gráfico 11⁶.

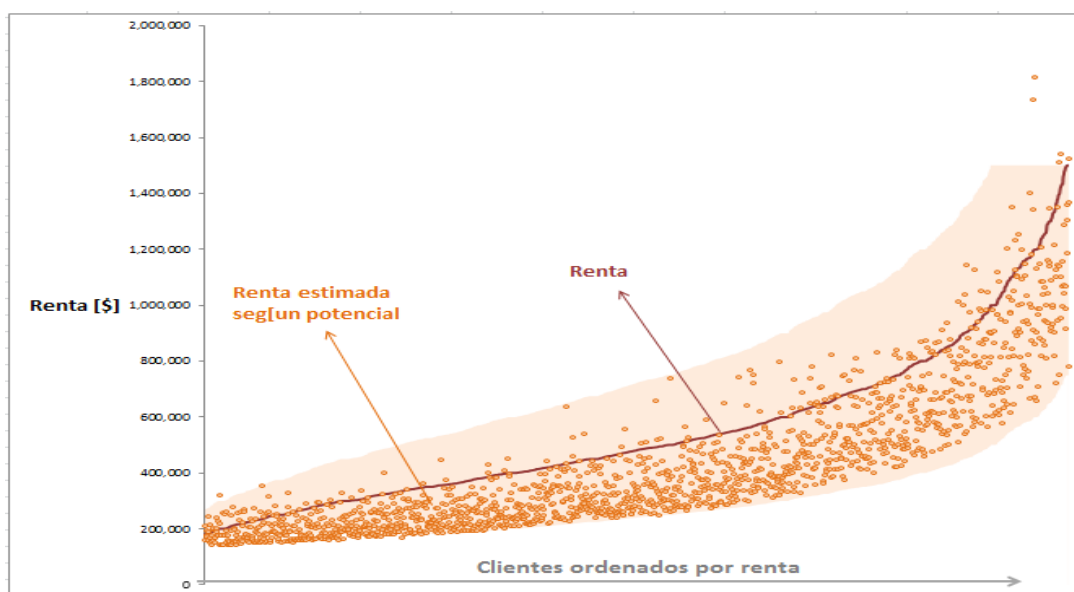


Gráfico 11: Ajuste del GBM transaccional
Fuente: Elaboración propia

En el gráfico anterior se puede apreciar que la gran mayoría de los clientes se encuentran dentro del intervalo de $\pm 50\%$ en relación a la renta real, lo cual es mostrado por el área anaranjada. El error máximo en la muestra (Max MAPE) corresponde a un 110,9%.

Al evaluar el error por quintiles de clientes según el gasto promedio mensual realizado con la tarjeta en el periodo 2⁷, se observa que el error del modelo va en aumento según incrementa el quintil de gasto, a excepción del quinto quintil, el cual tiene el error más bajo con un 22,4%. Esto es mostrado en la Tabla 12.

⁶ Para el gráfico, solo se utilizó una muestra aleatoria de 1000 clientes, donde el MAPE es de un 26%

⁷ El V quintil es el que más gasto realiza con la tarjeta

Quintil	MAPE
I	23,9%
II	28,4%
III	29,2%
IV	29,9%
V	22,4%

Tabla 12: Error por quintil de gasto promedio mensual con la tarjeta en el periodo 2 del GBM transaccional

Fuente: Elaboración propia

Los histogramas de n_i y π_i estimados son mostrados en el Anexo D. En particular, en el histograma de π_i se observa que una gran cantidad de clientes posee una probabilidad mayor al 40% de utilizar la tarjeta en cada transacción (alrededor de un 25% de la muestra), una probabilidad bastante alta si se considera la gran cantidad de medios de pagos y clubes de fidelización existentes hoy en día.

En el Gráfico 12 es posible observar que existe una relación positiva entre el número de transacciones realizadas con la tarjeta de crédito y el número de transacciones realizadas con cualquier medio de pago, lo cual es lógico ya que el número de compras con la tarjeta están contenidas en el número de transacciones global. La baja dispersión observada cuando el número de transacciones con la tarjeta aumenta, se explica ya que para los clientes que gastan una gran cantidad utilizando este medio de pago, es muy probable que sea cerca de todo su gasto.

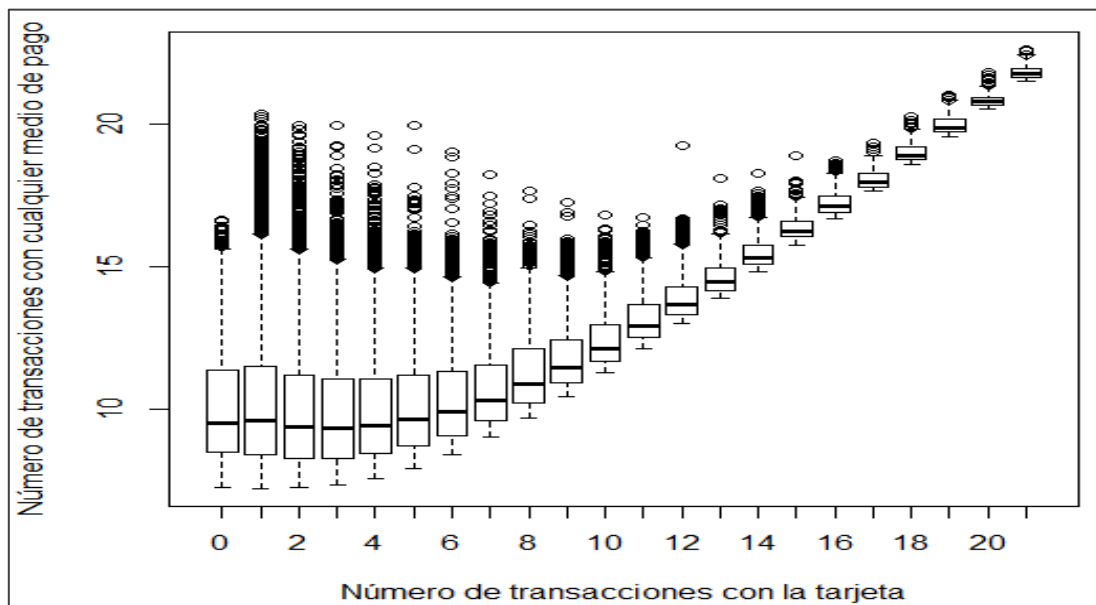


Gráfico 12: Número de transacciones con la tarjeta vs número de transacciones estimados con cualquier medio de pago utilizando GBM transaccional

Fuente: Elaboración propia

8.4. Modelo Binomial Generalizado extendido

El tercer modelo a evaluar, corresponde al GBM extendido. En éste, al igual que en el anterior, el monto promedio mensual gastado por un cliente con la tarjeta durante el periodo junio de 2011 a mayo de 2012, debe ser discretizado para poder obtener el número de transacciones realizadas con ésta. El monto utilizado en la calibración de este modelo nuevamente es de \$30.000.

Las variables explicativas utilizadas en este modelo, repartidas entre *size of wallet* y *share of wallet*, son:

- El tiempo que cada cliente llevaba en la compañía hasta mayo de 2011 (tenure): Variable normalizada, utilizada tanto en *share of wallet* como en el *size of wallet*.
- El monto promedio mensual de gasto durante el periodo anterior al que se quiere calcular la principalidad (junio de 2010 a mayo de 2011).
- La edad del cliente.
- El género del cliente.
- La zona donde vive.
- El cupo total de su tarjeta.
- El tipo de cliente según la clasificación de la empresa, la cual fue mostrada en la Tabla 5.

La forma funcional explícita para el caso de la tasa de gasto es la siguiente:

$$\lambda_i = \exp(\beta_0 + \beta_1 * \log(MP1_i + 1) + \beta_2 * \text{Norm}(\text{Tenure}) + \beta_3 * \log(\text{Age} + 1) + \beta_4 * D_{\text{genero}} + \beta_i * D_{\text{zonai}}) \quad (21)$$

donde {Norm()} indica que la variable está normalizada y {D} a una variable dummy, donde $zoni$ hace referencia a que corresponde a 7 dummies de zona, con i desde 1 a 7.

Para el caso de la probabilidad de elección del medio de pago, la definición es la siguiente:

$$\text{Ln}\left(\frac{\pi_i}{1 - \pi_i}\right) = \alpha_0 + \alpha_1 * \text{Norm}(MP1) + \alpha_2 * \text{Norm}(\text{Tenure}) + \alpha_3 * \log(\text{Cupo} + 1) + \alpha_i * D_{\text{tipo de cliente}} \quad (22)$$

donde la variable dummy tipo de cliente son dos: Normal tipo I (más de 3 años en la empresa) y Normal tipo II (entre 1 y 3 años en la empresa).

En la Tabla 13 se muestran los valores estimados para los coeficientes de cada una de las dieciocho variables. El código utilizado para la estimación se encuentra en el Anexo E:

Coefficiente	Valor	Desviación estándar	Significancia al 95 % (p-valor<0.05)
β_0	3,535	0,029	0,00
β_1	-0,003	0,001	0,01
β_2	0,465	0,022	0,00
β_3	-0,304	0,006	0,00
β_4	-0,007	0,003	0,02
β_5	-0,027	0,007	0,00
β_6	0,008	0,005	0,14
β_7	0,008	0,009	0,35
β_8	0,014	0,007	0,04
β_9	0,014	0,005	0,01
β_{10}	0,031	0,006	0,00
β_{11}	0,019	0,007	0,01
α_0	-4,097	0,053	0,00
α_1	3,586	0,061	0,00
α_2	-0,948	0,032	0,00
α_3	0,286	0,004	0,00
α_4	-1,166	0,015	0,00
α_5	-1,209	0,017	0,00

Tabla 13: Resumen coeficientes estimados para GBM extendido
Fuente: Elaboración propia

Al observar las variables, se tiene que 16 de ellas son significativas. Las únicas dos variables que resultan ser no significativas corresponden a las dummies asociadas a la zona Santiago Sur-Oriente (al igual que en el modelo log-lin) y Santiago Centro, a un 95% de confianza.

Al analizar el signo de las variables, se observa que el de la constante asociada a la tasa de compra es consistente con lo mencionado en la sección 8.3, donde se obtuvo que ésta es positiva, señalando que existe un número de compras promedio al menos mayor a uno. En cuanto a la variable asociada al monto promedio gastado en el periodo uno con la tarjeta en la estimación de la probabilidad, ésta tiene un signo positivo, lo que es lógico debido a que es de esperar que mientras mayor sea el gasto promedio de un cliente, mayor debiese ser la probabilidad de utilizar la tarjeta en cada transacción, como fue mostrado en la sección anterior. Al observar la variable asociada al sexo, se ve que las mujeres tienen una tendencia a gastar más que los hombres (β_4 es menor a cero). Finalmente, al observar las variables asociadas al tipo de cliente (α_4 y α_5), se observa que ambas son negativas, lo que deja en evidencia que la probabilidad de que este tipo de clientes use la tarjeta, es menor a un cliente Premium o Elite.

Los resultados del *ten-fold cross validation* para este modelo, son mostrados en la Tabla 14. El detalle de esto se encuentra en el Anexo F.

Estadística	Valor
<i>MAPE ajuste</i>	12,4%
<i>MAPE validación</i>	12,3%
AIC	545.467
BIC	545.477
ρ^2	66,1%

Tabla 14: Estadísticas para GBM extendido utilizando *cross validation*
Fuente: Elaboración propia

De igual forma que en los modelos anteriores, al utilizar completamente la base de clientes, dividida en base de ajuste y de validación, se corroboran de manera similar los resultados entregados en el *ten-fold cross validation*, como se puede ver en la Tabla 15.

Estadística	Valor
<i>MAPE ajuste</i>	12,5%
<i>MAPE validación</i>	12,6%
AIC	483.948
BIC	483.958
ρ^2	66%

Tabla 15: Estadísticas para GBM extendido con base de muestra global
Fuente: Elaboración propia

Al observar los resultados anteriores, se puede ver que el MAPE obtenido al utilizar el mismo benchmark que el de la empresa es de un 12,5%, lo cual es 12,5 puntos mejor que el modelo actual que posee la compañía. El nivel de ajuste es mostrado en el Gráfico 13⁸.

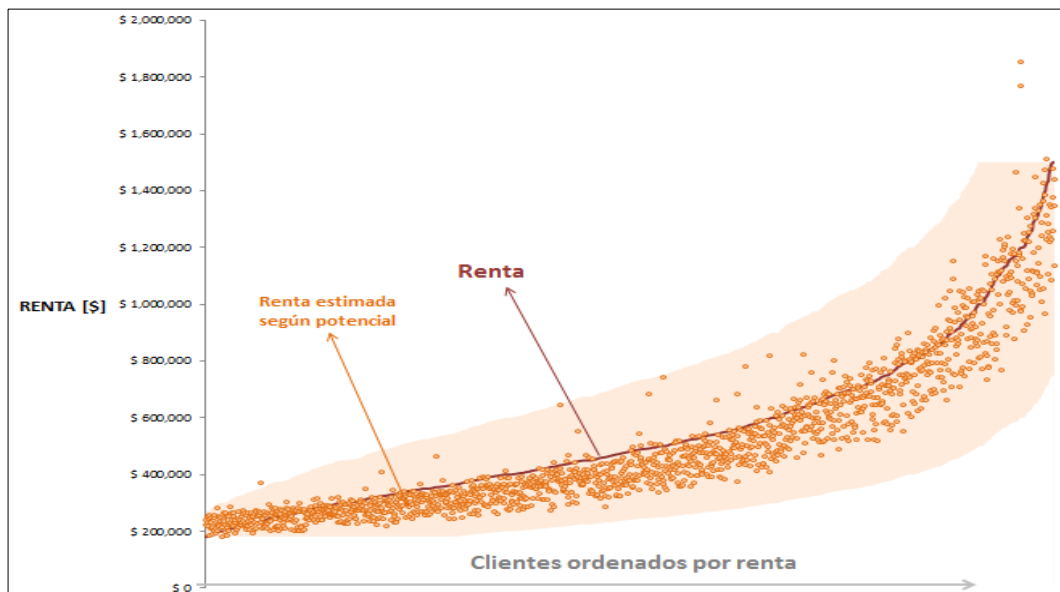


Gráfico 13: Ajuste del GBM extendido
Fuente: Elaboración propia

⁸ Para el gráfico, solo se utilizó una muestra aleatoria de 1000 clientes, donde el MAPE es de un 12,3%

Como se puede ver en el gráfico anterior, existe un gran nivel de ajuste, donde la gran mayoría de los clientes se encuentran al interior de un intervalo de error de $\pm 50\%$ en relación a la renta real, con un error máximo de 114,5% (Max MAPE).

Al evaluar el error por quintiles de clientes según el gasto promedio mensual realizado con la tarjeta en el segundo periodo, se observa que el error es casi constante a lo largo de los quintiles, en donde cada uno de estos presenta un MAPE aceptable del orden del 12%. Al igual que en el modelo GBM transaccional, el que presenta el menor error es el quinto quintil. Esto se muestra en la Tabla 16.

Quintil	MAPE
I	13,8%
II	12,1%
III	13,0%
IV	12,4%
V	11,5%

Tabla 16: Error por quintil de gasto promedio mensual en el segundo periodo para el GBM extendido

Fuente: Elaboración propia

Los histogramas de n_i y π_i estimados para este modelo son mostrados en el Anexo G. En particular, resulta interesante compara el histograma de π_i de este modelo en contraste al GBM transaccional, ya que se observa que en este caso existe una menor cantidad de individuos con probabilidad mayor al 40% de utilizar la tarjeta en cada transacción (14%).

Al comparar el Gráfico 12 con el Gráfico 14 que es presentado a continuación, es posible evidenciar que este modelo entrega resultados muy parecidos al anterior en cuanto a la tendencia positiva del nivel de gasto en función del número de transacciones realizado con la tarjeta de crédito. Sin embargo, este modelo tiene un desempeño superior, ya que al incluir más variables, obtiene un valor más bajo tanto en los estadísticos AIC y BIC, es decir, aun cuando el modelo es castigado por el número de variables, logra un mejor ajuste a los datos. Además, al observar el ρ^2 se evidencia que el GBM extendido (ρ^2 es de un 66%) funciona mejor que el GBM transaccional (ρ^2 es de 64%) con los parámetros estimados, en comparación a un modelo en que todos los parámetros son cero.

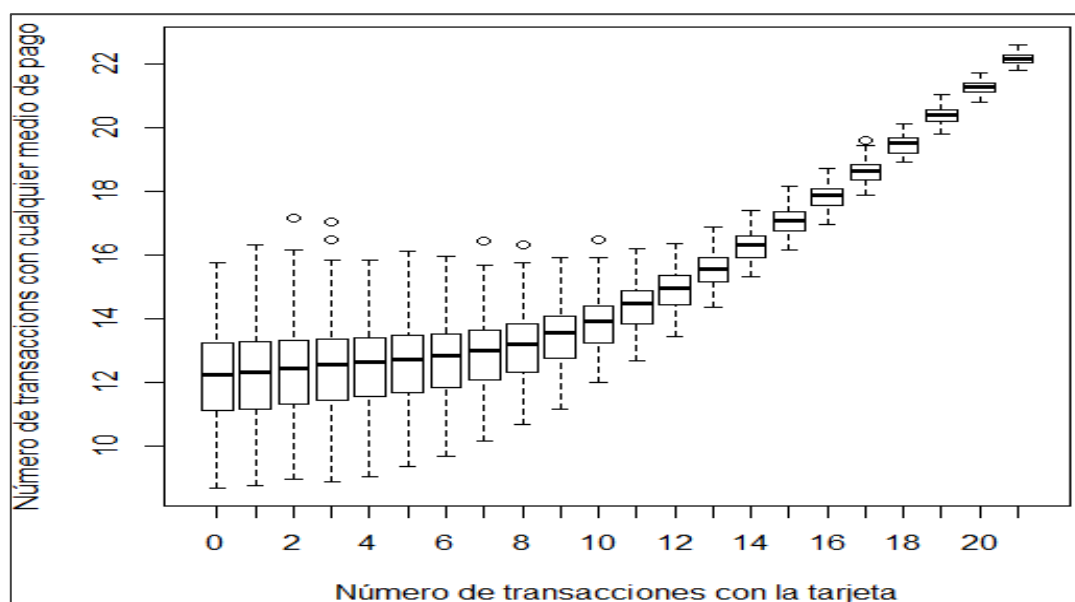


Gráfico 14: Número de transacciones con la tarjeta vs número de transacciones estimados con cualquier medio de pago utilizando GBM extendido

Fuente: Elaboración propia

Debido a todo lo mencionado anteriormente, será este modelo el que se utilizará en lo que resta del trabajo, debido a que es el que presenta un mejor comportamiento. En la Tabla 17 se muestra un resumen de los tres modelos y el ajuste de la empresa.

Estadística	Modelo de la empresa	Log-Lin	GBM transaccional	GBM extendido
<i>MAPE ajuste</i>	25%	36,2%	26,79%	12,5%
<i>MAPE validación</i>	-	36,1%	26,71%	12,6%
AIC	-	-	514.556	483.948
BIC	-	-	514.566	483.958
ρ^2	-	20,4%	63,85%	66%

Tabla 17: Resumen de los tres modelos

Fuente: Elaboración propia

Como se puede ver en la tabla anterior, el GBM extendido es superior a los otros dos modelos tanto en el MAPE de ajuste como en el MAPE de validación, además su especificación es mejor que el GBM transaccional al comparar tanto AIC y BIC, como fue mencionado anteriormente.

8.5. Aplicación de la metodología a las categorías vestuario y hogar

Para aplicar la metodología a las dos categorías más relevantes para el holding (hogar y vestuario), es necesario eliminar a los clientes que tienen un promedio de gasto mensual, en cada una de estas categorías, fuera del rango normal tanto en el periodo uno como en el dos, con el objeto de que no causen ruido en la estimación. Debido a lo anterior, el monto promedio mensual máximo gastado con la tarjeta en el periodo dos en la categoría vestuario y hogar, luego de eliminar a estos registros, es de

\$59.998 y \$164.764 respectivamente. Las estadísticas descriptivas son mostradas en la Tabla 18.

Categoría	Monto promedio máximo periodo I	Monto promedio máximo periodo II	Media del monto promedio del periodo II en la categoría
<i>Vestuario</i>	\$59.990	\$59.998	\$11.744
<i>Hogar</i>	\$134.950	\$164.764	\$18.975

Tabla 18: Análisis descriptivo de las categorías vestuario y hogar según gasto con la tarjeta
Fuente: Elaboración propia

Al realizar la estimación en la categoría vestuario, se utilizó un monto de \$1.800 para obtener el número de transacciones realizadas con la tarjeta durante el periodo dos (x_i). La utilización de este monto, se debe a que es uno de los valores que se encuentra en el rango del valor típico de gasto promedio al interior de la muestra, y además es el valor más pequeño que logra un ρ^2 alto (76,1%) sin que se indefina la estimación de los errores estándar, junto con poder estimar el *size of wallet* sin problemas según la ecuación 14.

A partir de los resultados obtenidos, la principalidad promedio en esta categoría es de un 33,1%. Como se verá más adelante, este valor es 17 puntos más bajo que lo que se obtiene al utilizar los porcentajes definidos por el INE para estimar el potencial de compra en la categoría. Uno de los factores que influye en la variación anterior, es el hecho de que al calcular cuánto representa el gasto en vestuario del total de compras mensuales que realiza una persona con cualquier medio de pago, éste en promedio es de un 8,2%, en contraste al 5,3% entregado por el INE.

Para la estimación realizada en la categoría hogar, se utilizó un monto de \$2.500 para obtener el valor de x_i . Las razones de escoger este valor son las mismas que las mencionadas para el caso de la categoría vestuario, obteniendo un ρ^2 de 75%. La principalidad promedio en la muestra en la unidad de consumo hogar es de 29%, lo cual es 19 puntos más bajo de lo que se obtiene utilizando los porcentajes INE. Nuevamente, se evidencian diferencias con respecto a los porcentajes entregados en la VI encuesta de presupuestos familiares del INE, donde el gasto promedio en la categoría hogar representa en promedio un 7,1% del gasto total mensual que realiza una persona, en contraste del 15,6% obtenido por el modelo.

A partir de lo anterior, se observa de que si bien el modelo logra un buen rendimiento en ambas categorías en cuanto al ajuste a los datos, en comparación a un modelo con todos los parámetros en cero (ρ^2 de 75%), existen diferencias importantes al estimar el *share of wallet* utilizando los porcentajes entregados por el Instituto Nacional de Estadísticas. Lo anterior podría indicar que el modelo sobrestima el potencial de gasto al aplicarlo en categorías por separado, o bien el comportamiento de compra de los clientes que utilizan la tarjeta de crédito de la empresa es muy diferente a lo indicado en el estudio del INE en cuanto al porcentaje del gasto destinado a cada una de las unidades de consumo, estudio que considera una muestra representativa de la población de Chile. Lo anterior es muy probable debido a que la muestra de clientes seleccionados son clientes con cuenta corriente, lo que provoca que tengan un comportamiento de compra diferente a la población promedio.

Como se mencionó en la sección de alcances, existen categorías con una gran cantidad de clientes inactivos, por lo que no es posible realizar una estimación sin generar una gran cantidad de error, motivo por el cual se utilizarán los porcentajes entregados por el INE para obtener el potencial y principalidad en cada una de las unidades de consumo.

9. RESULTADOS DEL GBM EXTENDIDO

En este capítulo se muestra en primer lugar, los resultados obtenidos con el modelo GBM extendido, para luego compararlos con los del modelo actual de la empresa. Luego, se realizará un análisis del valor del cliente contrastándolo con la principalidad. Posterior a ello, se hace una segmentación de clientes utilizando solamente el *share* y *size of wallet* global, para luego mostrar el *share of wallet* en cada una de las categorías usando como base el porcentaje entregado por el estudio del INE junto con una segmentación basado en esto. Inmediatamente después, se realiza una segmentación utilizando la herramienta estadística *k-means*.

Una vez que se tienen definidos los segmentos de clientes, se proponen líneas de acción, con el objeto de aumentar el *share of wallet* de la cartera de clientes.

9.1. Resultados generales del modelo GBM extendido

Al aplicar el modelo GBM extendido a la muestra de clientes, se obtiene que la principalidad promedio en ésta es de un 20,19% y el nivel de gasto promedio mensual es de \$377.000. En el Gráfico 15 se muestra a los clientes ordenados según el *share of wallet*, donde se puede evidenciar que la gran mayoría de los clientes presentan una principalidad inferior a un 50%, mostrando que existe una oportunidad para capturar una mayor cantidad de compras por medio de la tarjeta de crédito.

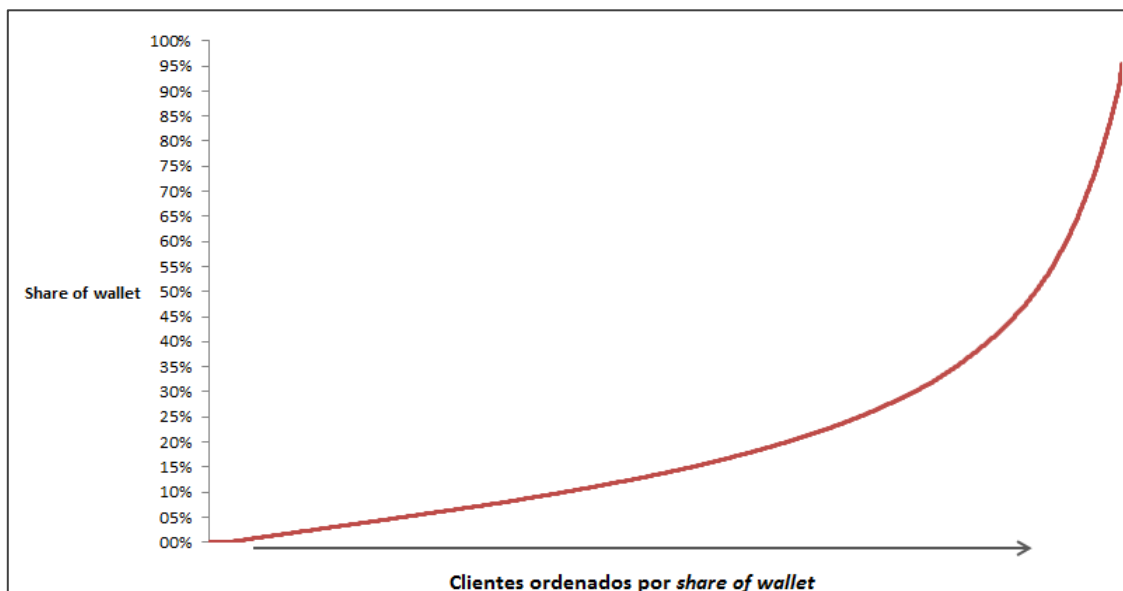


Gráfico 15: Clientes ordenados según *Share of wallet*
Fuente: Elaboración propia

En particular, un 63% de los clientes de la muestra presentan una principalidad entre 0% y 20%, y sólo un 9,4% de los clientes presentan una principalidad mayor al 50%. En el Gráfico 16 se muestra un histograma de los clientes según la principalidad estimada, donde se puede ver que a priori, los clientes que se encuentran en el extremo izquierdo del gráfico son aquellos que tienen un mayor potencial para ser capturado, sin embargo, como se verá en la sección 9.3, no es suficiente fijarse solamente en el *share of wallet*, sino que además, uno debiera hacer el cruce con el *size of wallet*.

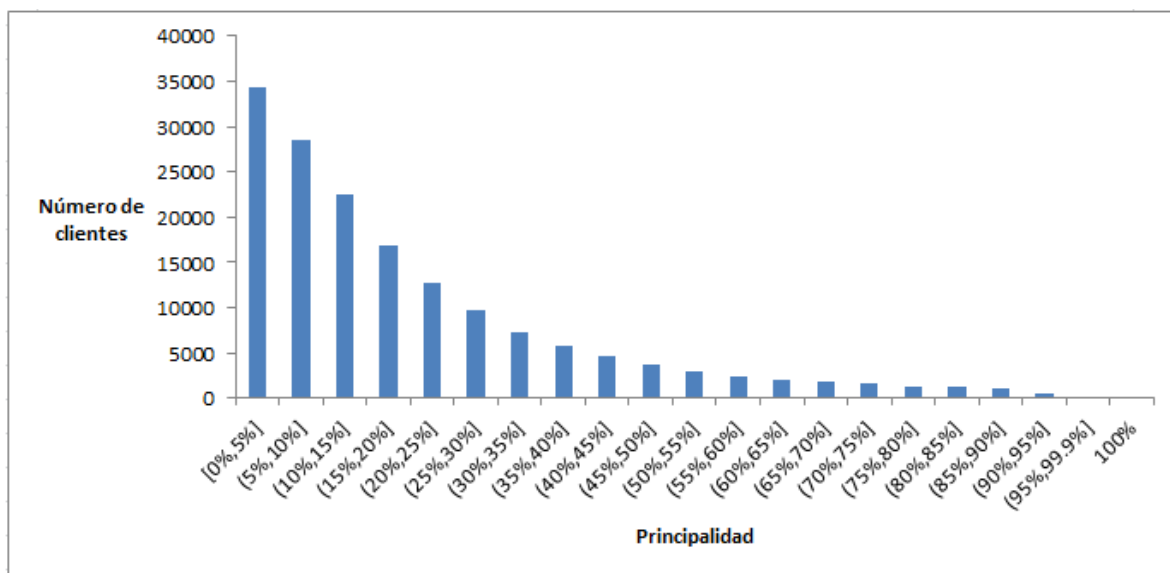


Gráfico 16: Histograma de clientes según *share of wallet*
Fuente: Elaboración propia

Al hacer la distinción entre aquellos que presentan la tarjeta Visa y los que no, se observa que la principalidad promedio para los primeros es de un 23,4%, mientras que para los segundos es de un 14,24%. Lo anterior, se debe a que los clientes que presentan la tarjeta Visa tienen la posibilidad de realizar compras no sólo al interior del holding, sino que además en cualquier comercio asociado, lo que les da la posibilidad de poder abarcar un mayor número de ítems al interior de cada unidad de consumo utilizando la tarjeta de crédito.

Luego, al observar la recencia y la antigüedad de los clientes según quintil de *share of wallet*, se puede ver que aquellos pertenecientes al quinto quintil presentan en promedio una recencia de 0,09 meses (alrededor de 3 días) y son clientes con una antigüedad promedio de 12 años. En el extremo opuesto, los clientes del primer quintil presentan una recencia de 2 meses y medio y una antigüedad promedio de 11 años. Lo anterior se puede observar en el Gráfico 17.

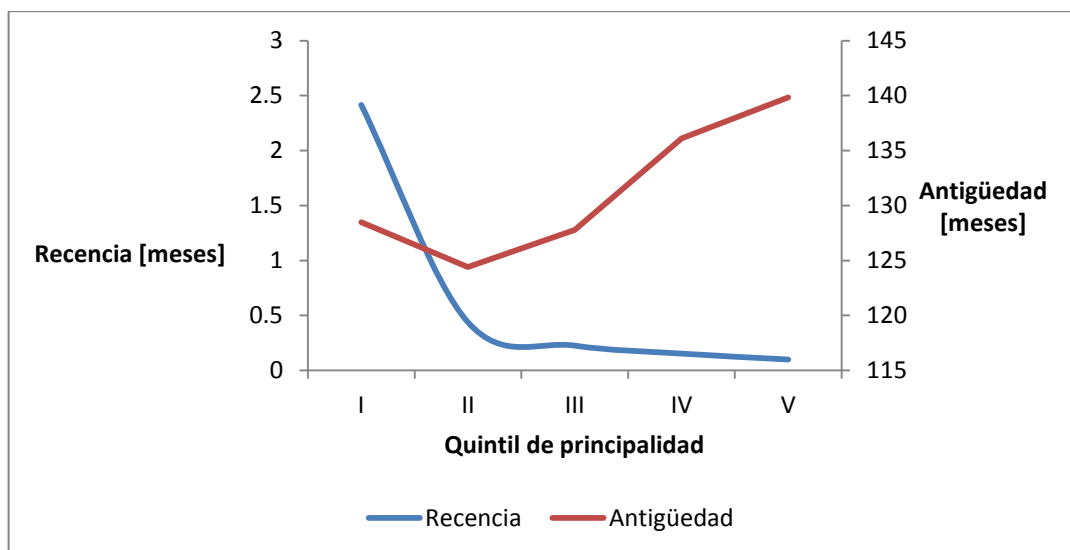


Gráfico 17: Recencia y antigüedad según quintil de principalidad
Fuente: Elaboración propia

Lo anterior es esperable ya que los clientes pertenecientes al quinto quintil de *share of wallet*, son los que utilizan la tarjeta para la gran mayoría de sus compras, por lo que es común que tengan un bajo número de días sin que la usen. Por el contrario, los clientes con un bajo nivel de principalidad, son aquellos que utilizan la tarjeta para compras esporádicas y particulares, razón por la cual presentan una recencia mayor.

Al utilizar la clasificación que tiene la empresa para sus clientes, esto es, según la cantidad de puntos acumulados como fue mostrado en la Tabla 5, y contrastarla con quintiles de principalidad, se puede observar que cerca de un 70% de los clientes Elite se encuentran en el quinto quintil de principalidad, lo cual es consistente con el hecho de que estos clientes son los que poseen un uso frecuente de la tarjeta o tienen un ticket promedio elevado. Por otro lado, de manera completamente opuesta, los clientes clasificados como normales no acumulan una gran cantidad de puntos en el plazo de un año, por lo que son clientes que poseen una baja frecuencia de uso o tienen un ticket promedio muy bajo, razón por la cual la gran mayoría de ellos (72,7%) se encuentran en el primer quintil de principalidad. En cuanto a los clientes Premium, cerca del 68% de éstos se encuentran entre el tercer y quinto quintil. Lo anterior es mostrado en el Gráfico 18.

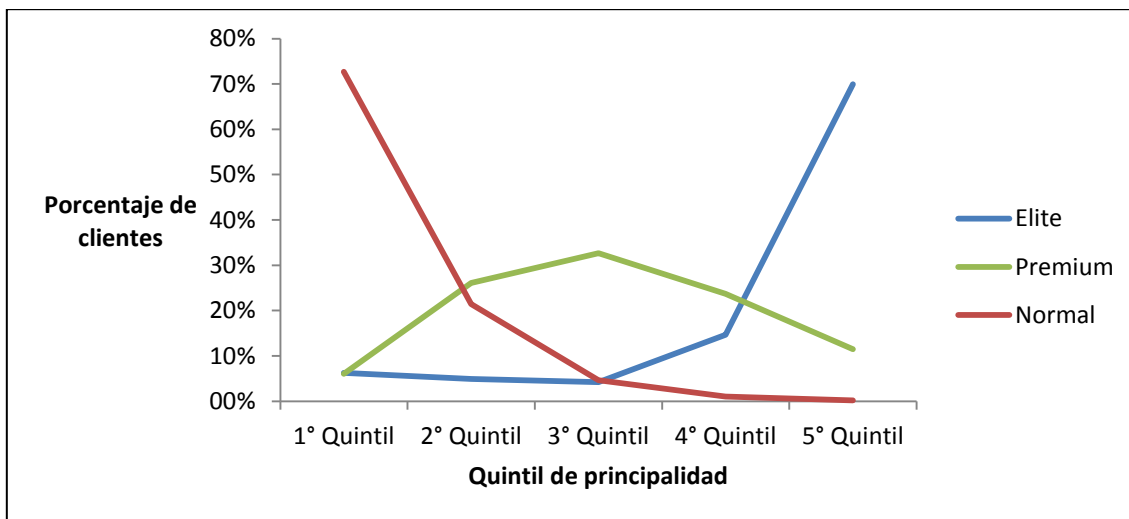


Gráfico 18: Porcentaje de tipos de clientes según principalidad
Fuente: Elaboración propia

9.1.1. Comparación con el modelo actual de la empresa

Al realizar un análisis general del modelo actual de la empresa y compararlo con los resultados de la metodología propuesta, se pueden vislumbrar varios resultados interesantes. En primer lugar, al mostrar la dispersión del gasto total estimado por la empresa (ver Gráfico 19), se observa que no existe una gran heterogeneidad al interior de la muestra, dada la forma en que se estimó el gasto total, en contraste a lo mostrado en el Gráfico 14. Además, esta metodología puede tanto subestimar como sobrestimar el gasto total para muchos clientes, ya que por ejemplo, si uno observa a los clientes que tienen un número de transacciones igual a 20, se ve que muchos tienen un potencial estimado inferior a los \$600.000⁹, lo cual es incoherente con el hecho de que el gasto con la tarjeta de crédito está contenido en el gasto total que realiza una persona.

Luego, al comparar el gasto total, se puede ver que el promedio estimado por la empresa en la muestra es de \$414.571, versus los \$377.000 obtenidos por la metodología. En la Tabla 19 se muestran estadísticas descriptivas de lo anterior.

	Media	Máximo	Mínimo	Desviación estándar
Empresa	\$414.571	\$1.246.646	\$187.063	\$131.938
GBM	\$376.715	\$677.375	\$261.463	\$46.508

Tabla 19: Comparación gasto estimado por la empresa vs GBM extendido
Fuente: Elaboración propia

⁹ Recordar que el valor utilizado para discretizar los gastos con la tarjeta es de \$30.000.

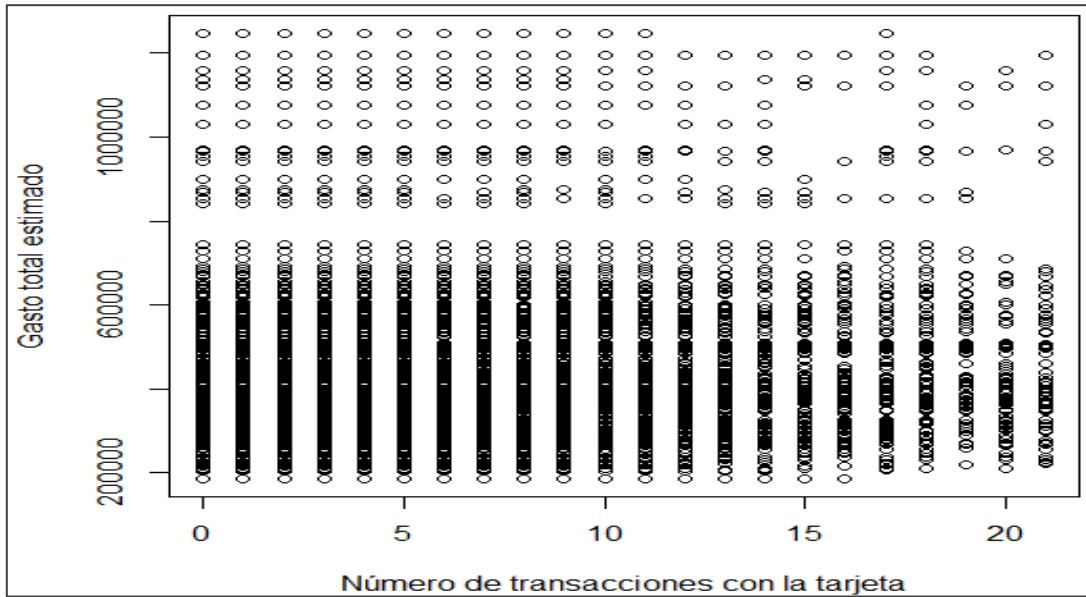


Gráfico 19: Número de transacciones vs gasto total estimado por la empresa
Fuente: Elaboración propia en base a información entregada por la empresa

Finalmente, al comparar la principalidad estimada se observa que en promedio ambas metodologías son parecidas, ya que por un lado el GBM extendido entrega un *share of wallet* promedio de 20,19% y el modelo de la empresa arroja un 20,06%. El problema que presenta este último, como se mencionó anteriormente, es que existen clientes en los que se sobrestima el potencial de gasto y a otros en los que se subestima, sin embargo, el GBM también presenta este problema, dado que como se verá más adelante, existen clientes que no compran más con la tarjeta porque su cupo no lo permite, lo que genera que se subestime su gasto total.

Como se puede observar en el Gráfico 20, existe una gran cantidad de clientes que presentan principalidad uno, los cuales tienen un potencial de gasto subestimado, lo que genera que su *share of wallet* este sobrestimado¹⁰. Es por lo anterior que, dado que existen clientes con principalidad tanto sobrestimada como subestimada, el efecto final en el promedio se compensa, logrando la misma media que lo arrojado por el GBM extendido. De hecho, en el Gráfico 20 se ve que la cantidad de clientes en cada grupo de principalidad es muy similar a lo arrojado por el GBM extendido, pero la diferencia principal radica en que el GBM extendido provoca una mayor variabilidad en el potencial de gasto y siempre este potencial es mayor o igual al gasto con la tarjeta.

¹⁰ Lo anterior se debe a que la principalidad fue calculada como el $\min(1, \text{share of wallet})$

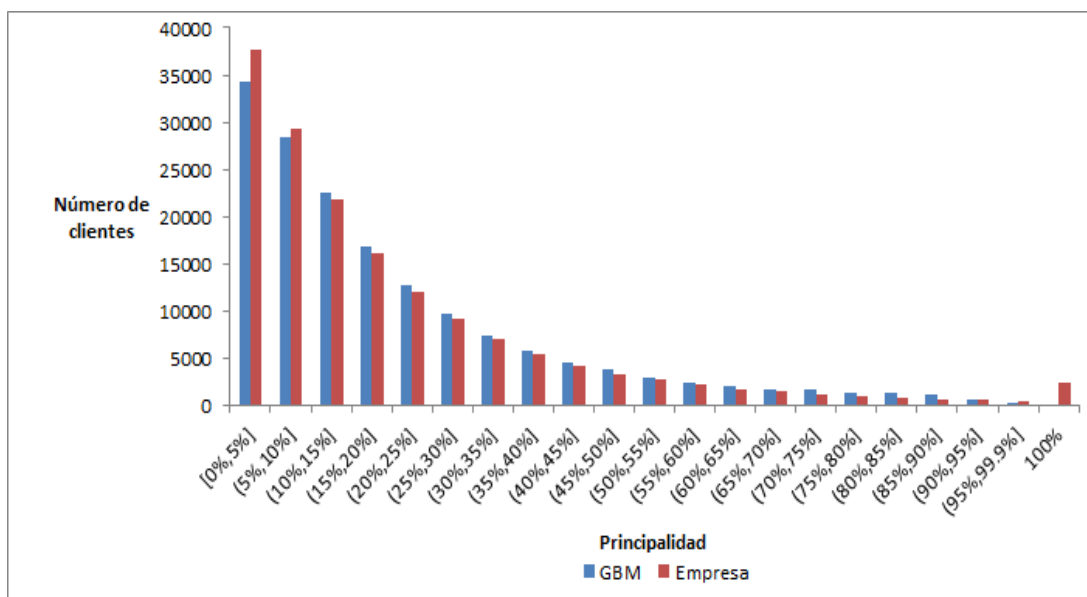


Gráfico 20: Histograma de clientes según share of wallet estimado por GBM vs estimado por la empresa

Fuente: Elaboración propia en base a información entregada por la empresa

9.2. Principalidad y valor del cliente

Actualmente, en la compañía poseen dos métricas respecto al valor del cliente en un periodo determinado. La primera métrica consiste en evaluar el valor que tuvo el cliente para el negocio (holding), indicador diseñado según el monto gastado, el margen de contribución de la línea en la que compró y los costos asociados (costos operacionales, gastos de administración y ventas, y otros costos asociados). La segunda métrica resulta de evaluar la rentabilidad de los clientes, pero para la empresa de la tarjeta de crédito, donde sus ganancias y costos por cada uno de los consumidores son diferentes a lo considerado por el negocio. Los ingresos en este indicador se construyen utilizando la comisión por venta con la tarjeta¹¹, los intereses que paga el cliente debido al endeudamiento y los intereses provocados por la mora en el pago. Los costos asociados son de las mismas características a los del primer indicador, sólo que son los costos propios de la empresa de la tarjeta de crédito¹². Es así como un comprador puede ser muy rentable en ambas métricas, o sólo serlo en términos del negocio y tener una baja rentabilidad para la empresa de la tarjeta de crédito, o viceversa. Lo anterior podría ocurrir en caso de que un cliente compre frecuentemente o compre montos muy altos utilizando la tarjeta de crédito, pero nunca endeudarse.

Al estudiar la rentabilidad en términos del negocio en el periodo 2011, es interesante mostrar como al interior de la muestra, el 15% de los clientes más rentables en esta métrica explican el 80% de la rentabilidad total de la cartera considerada en este trabajo (calculada como la suma de las rentabilidades individuales). Es posible

¹¹ Los diferentes negocios al interior del holding, las alianzas con la empresa y los comercios asociados, le pagan una comisión de la venta a la compañía cuando utilizan la tarjeta de crédito.

¹² Estos indicadores los construye el área de riesgo de la empresa. No fue posible obtener la metodología específica utilizada para la construcción.

visualizar lo anterior en el Gráfico 21, donde en el eje de las abscisas los consumidores se encuentran *rankeados* según el X% top de la muestra (el 100% top considera a todos los clientes). En el eje de las ordenadas, se muestra el porcentaje acumulado de la rentabilidad total de la muestra (\$3.528.672.025 en el año 2011) por cada X% top de clientes. Se observa que en algunos puntos este valor supera el 100%, lo cual se debe a la existencia de clientes con rentabilidad negativa.

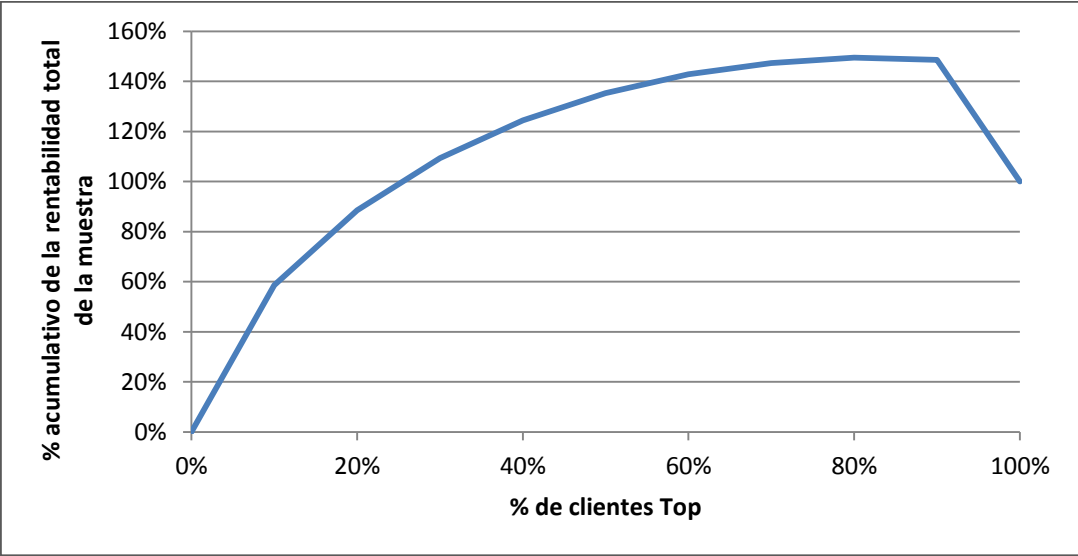


Gráfico 21: Concentración de la rentabilidad del cliente para el negocio
Fuente: Elaboración propia en base a información entregada por la empresa

Algo similar ocurre al analizar la rentabilidad de los clientes según la empresa de la tarjeta de crédito, donde se observa que el top 30% de los clientes en esta métrica, explica el 80% de la rentabilidad total de la muestra. Nuevamente, se ve que el porcentaje acumulado de rentabilidad supera el 100%, lo anterior entre el top 80% y el top 100%, debido nuevamente a los clientes con rentabilidad negativa. Todo esto es mostrado en el Gráfico 22. La rentabilidad total de la muestra en el 2011 considerando este indicador fue de \$20.380.825.539.

Resulta interesante mencionar el hecho de que en ambas métricas, ocurre un hecho cercano a lo conocido como *Principio de Pareto*, donde se dice que el 20% de los clientes explican el 80% de la rentabilidad del negocio, siendo en este caso la rentabilidad total de los clientes.

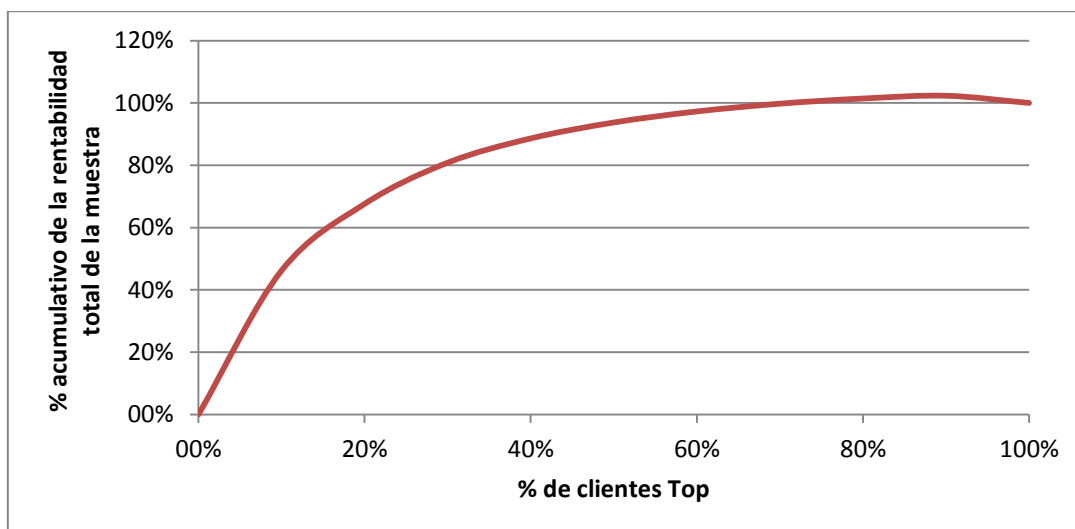


Gráfico 22: Concentración de la rentabilidad del cliente para la empresa de la tarjeta de crédito
Fuente: Elaboración propia en base a información entregada por la empresa

Si bien es posible argumentar que la empresa debe enfocarse en todo el grupo de clientes que explican la mayor parte de la rentabilidad de la cartera, realizando las mismas acciones para mantener su uso actual, esto no es completamente cierto, ya que existe un grupo de clientes que poseen una rentabilidad alta y además, tienen una principalidad baja, lo que les da la posibilidad de aumentar el valor de ese cliente aún más, en contraposición de aquellos clientes que poseen una rentabilidad alta y una principalidad alta también, por lo que las acciones de marketing para estos clientes apuntan a mantener su uso actual, a diferencia del grupo con bajo *share of wallet*, en que las acciones deberían ir enfocadas en capturar un mayor gasto con la tarjeta de crédito.

Al separar por tercil la rentabilidad de los clientes en términos del negocio, considerando por un lado a los que tienen rentabilidad negativa y por otro a los que tienen rentabilidad positiva, y contrastarlo con terciles de principalidad, se observa que un 22,36% de los clientes de la muestra se encuentran en el tercer tercil de rentabilidad positiva y en el primer tercil de *share of wallet*. Estos clientes, a diferencia del 2,71% que se encuentran tanto en el tercer tercil de principalidad como de rentabilidad positiva, tienen la posibilidad de aumentar su rentabilidad capturando una mayor porción de su gasto por medio de la tarjeta de crédito. Lo anterior se muestra en la Tabla 20. Por otro lado, existen aquellos clientes que presentan una alta principalidad pero rentabilidad negativa, para los cuales la empresa debiese tratar de mantener el gasto actual que poseen, pero trasladándolo a categorías más rentables para el negocio.

Terciles de principalidad	Rentabilidad Negocio negativa			Rentabilidad Negocio positiva		
	I	II	III	I	II	III
I	3,16%	3,94%	4,33%	17,13%	29,41%	22,36%
II	1,33%	0,88%	0,55%	0,95%	3,28%	8,26%
III	0,53%	0,21%	0,14%	0,19%	0,65%	2,71%

Tabla 20: Cantidad de clientes según terciles de principalidad vs terciles de rentabilidad de clientes para el negocio

Fuente: Elaboración propia

Dado que el proyecto es realizado en la empresa de la tarjeta de crédito, es más interesante analizar lo que ocurre al observar la segunda métrica de rentabilidad. Al observar la Tabla 21, se observa que sólo un 2,39% de los clientes de la muestra se encuentran tanto en el tercer tercil de principalidad como de rentabilidad positiva. Por otro lado, un 23,42% de los clientes se encuentran en el primer tercil de principalidad y en el tercer tercil de rentabilidad positiva, clientes que tienen un potencial para aumentar aún más su valor por medio de la utilización de la tarjeta de crédito. En general, el 76% de los clientes se encuentran en el primer tercil de *share of wallet* y tienen una rentabilidad positiva, lo que da luces de que el valor de la cartera de clientes se puede aumentar aún más si se traslada a éstos a mayores terciles de principalidad.

Terciles de principalidad	Rentabilidad tarjeta de crédito negativa			Rentabilidad tarjeta de crédito positiva		
	I	II	III	I	II	III
I	1,36%	1,08%	1,26%	25,74%	27,47%	23,42%
II	0,08%	0,26%	0,23%	2,30%	4,85%	7,53%
III	0,13%	0,23%	0,09%	0,57%	1,02%	2,39%

Tabla 21: Cantidad de clientes según terciles de principalidad vs terciles de rentabilidad de clientes para la empresa de la tarjeta de crédito

Fuente: Elaboración propia

A los clientes que se encuentran en el tercer tercil de principalidad, pero presentan una rentabilidad negativa, es complicado llevarlos al lado derecho de la tabla, ya que dado que están gastando una gran cantidad de dinero utilizando la tarjeta, la única manera de hacerlo es generando un cambio en su conducta de compra, incentivándolos a comprar en cuotas, ya que son clientes que por lo general compran al contado, razón de por qué no son rentables para la empresa.

9.3. Segmentos en función de la principalidad y *size of wallet*

Utilizando los resultados arrojados por el GBM extendido, es posible realizar una primera segmentación de los clientes basado solamente en la principalidad de la tarjeta y el nivel de gasto realizado con cualquier medio de pago. Como se muestra en el Gráfico 23, se formaron 25 grupos de clientes, donde fueron segmentados según quintil de principalidad y de *size of wallet*.

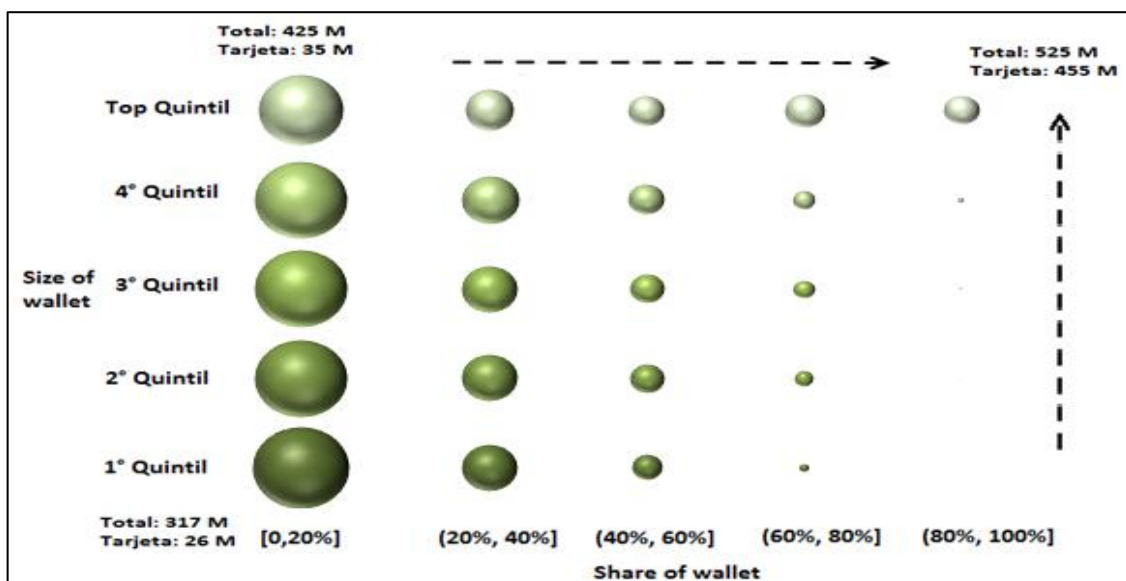


Gráfico 23: Grupos de clientes según principalidad de la tarjeta y gasto total con cualquier medio de pago

Fuente: Elaboración propia

En el gráfico anterior, es posible observar que los clientes que se encuentran en el extremo superior izquierdo, quienes representan un 10,5% de la muestra total de clientes, gastan en promedio \$425.000 mensual con cualquier medio de pago, y de este total, en promedio sólo \$35.000 es realizado con la tarjeta de la tienda. Son estos clientes los que tienen un mayor potencial a ser capturado con la tarjeta de crédito, por lo tanto la compañía debería enfocarse en trasladarlos al extremo superior derecho, de manera de capturar más de su gasto por medio de la tarjeta. Por otro lado, en el extremo superior derecho se encuentran los clientes ideales para el holding, los cuales gastan una gran cantidad de plata con cualquier medio de pago y de este total, entre un 80% y un 100% es realizado con la tarjeta de crédito de la tienda. Estos clientes representan un 1,8% de la muestra total, a los cuales se les deben hacer campañas para retenerlos y para que mantengan su uso actual de la tarjeta. En el extremo inferior izquierdo, se encuentran los clientes que no resultan interesantes para la compañía (14,2% de la muestra), por lo que no se debería destinar una gran cantidad de gastos de marketing en ellos, ya que además de comprar poco con la tarjeta, son los que gastan la menor cantidad de dinero con cualquier medio de pago. Las tablas relacionadas al gráfico anterior son mostradas en el Anexo H.

Debido a que la compañía no tiene influencia directa en el gasto total que realiza una persona utilizando cualquier medio de pago, la empresa a través de campañas sólo puede influir en la dimensión del *share of wallet*, donde el objetivo debiese ser el de trasladar a la gran mayoría de los clientes pertenecientes a los diferentes segmentos, al extremo derecho.

9.4. Principalidad y segmentación por categorías de consumo

Como se mencionó en las secciones anteriores, para calcular el gasto total en cada una de las unidades de consumo, se utilizó el porcentaje sobre el gasto total destinado a cada una de estas categorías entregado por el INE, el cual está dividido

según quintil de ingreso. Dado que no se posee el ingreso para todos los clientes de la compañía, se generaron quintiles según el gasto estimado con el GBM extendido. Al considerar estos porcentajes, se debe tener en consideración el posible sesgo que conlleva, debido a que el comportamiento de gasto de los clientes de la empresa son diferentes al de la población nacional. Es de esperar que los porcentajes de gasto al interior de la empresa sean superiores a los presentados por el instituto de estadísticas, ya que este estudio considera a una muestra representativa de la población chilena, existiendo tanto poseedores y no poseedores de tarjetas de crédito, por lo que el comportamiento de compra promedio es diferente al grupo de clientes de la compañía. En la Tabla 22 se muestran estos porcentajes, donde se observa por ejemplo que en la categoría alimentos, el quintil más rico destina un 16% de su gasto total en esta unidad, versus el quintil más pobre que destina un 36,2%. Cabe mencionar que en la categoría hogar y transporte, hay una diferencia en cuanto a lo entregado por el INE, ya que el ítem de comunicaciones, perteneciente originalmente a la categoría transporte y comunicación, fue cambiada a la unidad de gastos en vivienda. Los ítems de cada categoría se encuentran en el Anexo I.

Unidad de consumo	I	II	III	IV	V
ALIMENTOS	36,2%	31,4%	28,3%	23,8%	16,0%
VESTUARIO	5,0%	5,6%	5,6%	5,4%	5,0%
VIVIENDA	22,7%	20,7%	20,1%	19,9%	18,0%
HOGAR	6,7%	6,5%	6,6%	7,3%	8,5%
SALUD	4,3%	4,3%	5,4%	5,3%	5,8%
TRANSPORTE	10,6%	14,4%	15,0%	17,4%	21,3%
RECREACION	3,2%	3,9%	3,9%	4,2%	4,3%
EDUCACION	3,2%	4,4%	5,1%	5,6%	7,0%
OTROS	8,1%	8,8%	10,1%	11,0%	14,1%
Total	100%	100%	100%	100%	100%

Tabla 22: Distribución del gasto mensual por quintil de ingreso
Fuente: VI Encuesta de presupuestos familiares. INE

Al ponderar estos porcentajes por el gasto total estimado, se obtiene el gasto total de cada cliente de la muestra en cada unidad de consumo, obteniendo luego el *share of wallet* en cada una de estas categorías. En el Gráfico 24 se muestra la principalidad promedio en cada una de estas unidades.

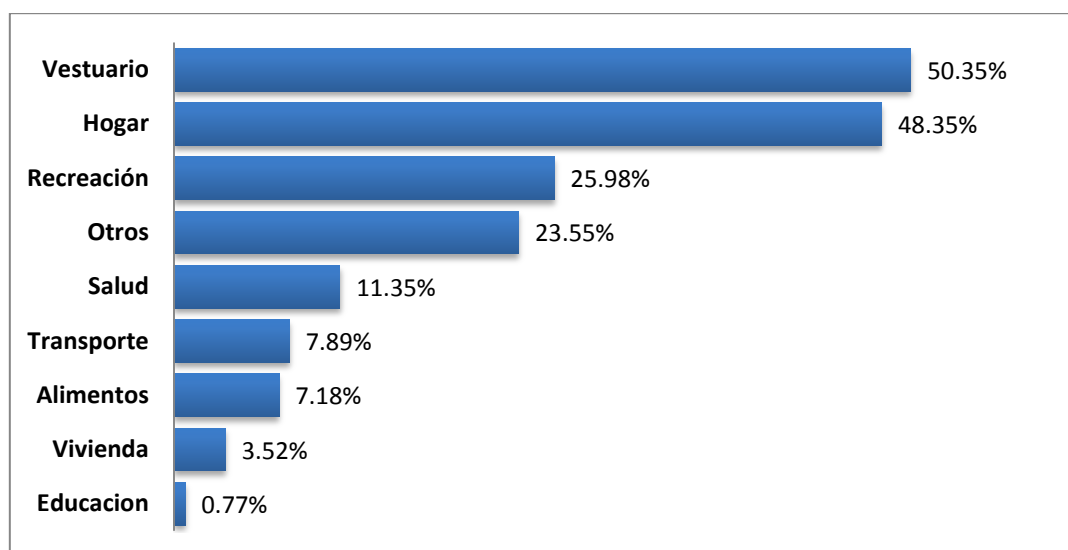


Gráfico 24: Principalidad por categoría

Fuente: Elaboración propia

Como se puede observar, los clientes son en promedio más principales en la categoría vestuario, es decir, alrededor del 50% de sus compras en esta unidad la realizan con la tarjeta de crédito de la empresa. Dos puntos más abajo se encuentra la categoría hogar, con un 48% de principalidad promedio. Es de esperar que estas dos categorías sean las que presentan el mayor *share of wallet*, ya que dos de los negocios centrales del holding cubren por completo estas unidades de consumo, por lo que sus clientes al realizar compras en estas categorías tienden a utilizar la tarjeta más que en cualquier otra.

En cuanto a la categoría alimentación, donde uno de los negocios del holding cubre en gran medida sus ítems (a excepción de gastos en restaurants), se observa que sólo posee un 7,18% de principalidad. Lo anterior se podría explicar por el bajo o nulo posicionamiento que tiene en este negocio el holding respecto a sus competidores, donde dos de ellos dominan la industria, por lo que no resulta natural que los clientes de la tarjeta de crédito la utilicen en esta categoría, ya que existen otros clubes de fidelización propios de los competidores que permiten realizar compras en sus locales. Sin embargo, este bajo porcentaje resulta ser una oportunidad para el holding, dado que existe un gran potencial para realizar *cross-selling* y otorgar beneficios en caso de utilizar la tarjeta en esta categoría. Finalmente, se observa que la categoría educación resulta ser aquélla donde los clientes utilizan en menor medida la tarjeta de crédito de la empresa.

En el Anexo J se muestra la matriz de correlación entre las principalidades de cada categoría.

9.4.1. Segmentación por categorías de consumo

Con el objetivo de evaluar con qué categoría está más vinculado un cliente por medio de la tarjeta de crédito, se realizó una segmentación basado en el siguiente criterio: “Un cliente estará asociado a una unidad de consumo si en ella presenta la

mayor principalidad respecto de las otras categorías". Debido a lo anterior, un consumidor puede ser parte de una o más categorías (si presenta igual principalidad en alguna de las unidades), e incluso a ninguna en caso de que su principalidad global sea cero en el periodo de estudio. A partir de esto, se obtiene que la mayor cantidad de clientes se encuentran vinculados a la categoría vestuario, con un 45,6% de la muestra, mientras que la unidad de consumo que posee la menor cantidad, es la de educación, con sólo un 0,43% de los clientes. En el Gráfico 25 se observa lo anterior, donde además se puede ver que las categorías siguen el mismo orden que en el Gráfico 24.

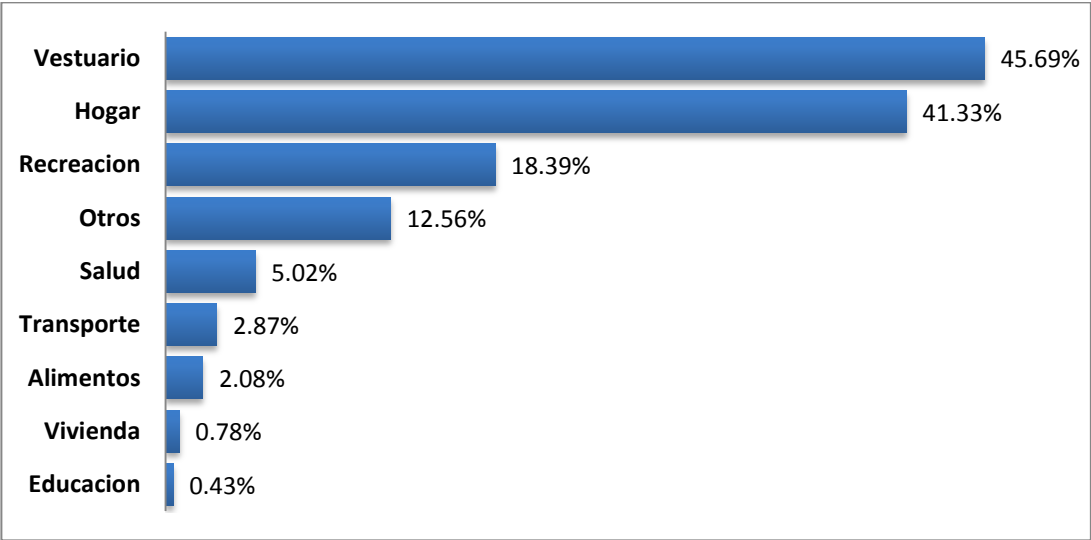


Gráfico 25: Porcentaje de clientes vinculados a cada unidad de consumo
Fuente: Elaboración propia

Al evaluar la principalidad global promedio de cada uno de estos segmentos, se observa que aquellos clientes que están vinculados con la categoría educación, son los que presentan la mayor principalidad global, con un 47%. Lo anterior indica que una vez que un cliente utiliza su tarjeta para gastos en educación, es muy probable que la esté usando en gran medida en el resto de las unidades de consumo. Lo contrario ocurre con aquellos clientes que son vinculados a la categoría vestuario, la cual contiene a la mayor cantidad de clientes como se observó en el Gráfico 25, presentando una principalidad promedio de un 25%. En general, como se puede ver en el Gráfico 26, se observa que los clientes que están vinculados por medio de la tarjeta de crédito en las categorías que no tienen relación con lo que ofrece el holding (educación, salud, recreación, etc.), son aquellos que utilizan su tarjeta en mayor medida en cualquier categoría. Una excepción a lo anterior son los clientes vinculados a la categoría de alimentos, ya que como fue mencionado, esta es una categoría en la cual la mayor parte de sus ítems son ofrecidos por uno de los negocios del holding, pero presenta una baja principalidad promedio considerando la muestra de clientes. Sin embargo, aquellos que están vinculados a esta categoría por medio de la tarjeta, tienen un *share of wallet* global de 46%, es decir, dado que es muy probable que estos clientes estén realizando gran parte de las compras de esta categoría en el negocio del holding, es probable que sean clientes muy leales tanto a los negocios de la compañía, como a la tarjeta de crédito.

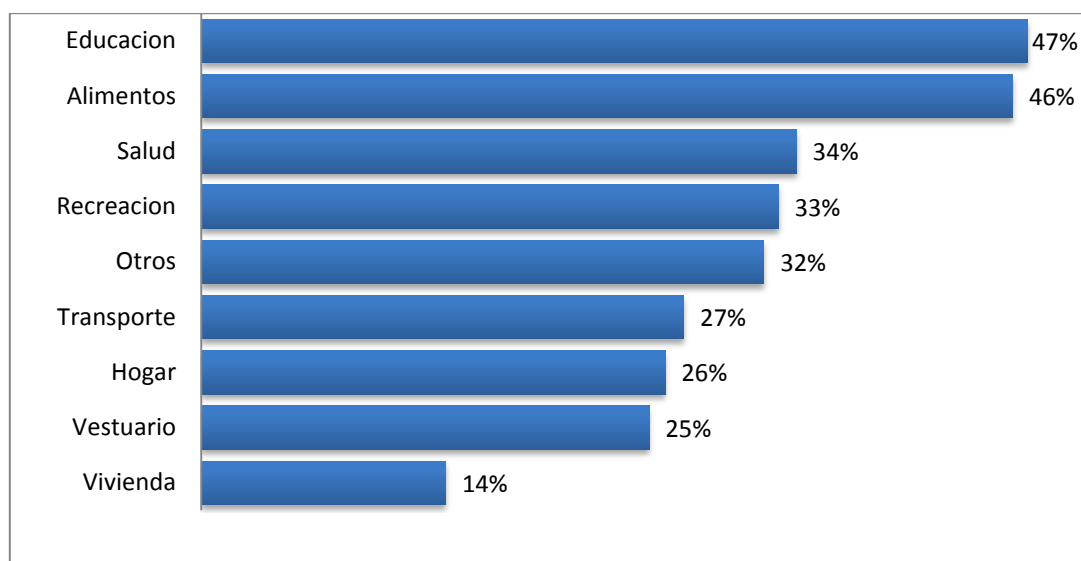


Gráfico 26: Principalidad global según segmento asignado

Fuente: Elaboración propia

Es interesante conocer cómo está compuesto cada uno de esos segmentos en cuanto al tipo de cliente y variables transaccionales, ya que una descripción demográfica no tiene sentido en este caso, puesto que al tomar como ejemplo a la variable edad o sexo, se puede evidenciar que no existe una gran diferencia entre un segmento y otro, además de no agregar información a la compañía. Lo anterior se puede ver en la Tabla 23. Sin embargo, este tipo de características serán tomadas en cuenta al momento de proponer líneas de acción.

Segmento	% Hombres	% Mujeres	Edad promedio
Educación	58,4%	41,6%	42
Vivienda	68,6%	31,4%	45
Alimentos	65,8%	34,2%	46
Transporte	81,0%	19,0%	48
Salud	46,7%	53,3%	49
Otros	61,2%	38,8%	44
Recreación	64,5%	35,5%	43
Hogar	63,0%	37,0%	46
Vestuario	52,2%	47,8%	42

Tabla 23: Características demográficas de los segmentos

Fuente: Elaboración propia

Al caracterizar los segmentos anteriores de manera transaccional, se puede observar en la Tabla 24 que aquellos que presentan una mayor principalidad global promedio, contienen un mayor porcentaje de clientes Elite y Premium en relación al resto de las categorías. Si bien es cierto que estos mismos segmentos también presentan un gran porcentaje de clientes normales, la cantidad de éstos en esos grupos es pequeña en relación al total de clientes del tipo Normal presentes en la muestra, lo cual es mostrado en la Tabla 25. Lo anterior, es consistente con lo que fue mostrado en

el Gráfico 18, donde se mostró que un bajo porcentaje de clientes del tipo Normal se encuentran en el cuarto y quinto quintil de principalidad.

Es interesante notar como la tendencia en relación al porcentaje de clientes en que su tarjeta funciona como Visa, aumenta con los segmentos más principales, donde el grupo con un mayor porcentaje de éstos, resulta ser el segmento educación, con un 96,4%. Algo similar ocurre con la media del gasto promedio en el periodo dos, el cual también aumenta en los grupos más principales. Lo anterior podría dar indicios de que al momento de poseer la tarjeta Visa, las personas comienzan a utilizarla en gran parte de sus gastos mensuales, por lo que sería interesante que la compañía evaluase una mayor migración de clientes a este tipo de tarjeta, con el objeto de aumentar la principalidad promedio de su cartera.

Segmento	Número de clientes	Monto promedio Tarjeta/ persona	% Clientes Elite y Premium	% Clientes Normal	% Clientes con Visa
Alimentos	3352	\$ 211.546	53,7%	46,3%	77,4%
Educación	695	\$ 203.155	47,3%	52,7%	96,4%
Salud	8075	\$ 139.532	36,1%	63,9%	73,0%
Recreación	29605	\$ 131.826	29,2%	70,8%	73,3%
Transporte	4620	\$ 109.877	27,7%	72,3%	70,9%
Otros	20214	\$ 128.279	23,2%	76,8%	68,0%
Hogar	66542	\$ 103.528	21,0%	79,0%	63,6%
Vestuario	73562	\$ 101.034	21,0%	79,0%	72,6%
Vivienda	1243	\$ 59.016	13,0%	87,0%	59,6%

Tabla 24: Caracterización transaccional de los segmentos

Fuente: Elaboración propia

Total clientes Elite	143
Total clientes Normal	138430
Total clientes Premium	22411

Tabla 25: Total de clientes por tipo en la muestra

Fuente: Elaboración propia

Finalmente, se debe analizar cómo se comportan los clientes pertenecientes a cada segmento, en el resto de las categorías en torno al *share of wallet*. Como se puede ver en la Tabla 26, los segmentos que contienen a los clientes más principales de manera global (educación, alimentos, salud), presentan los *share of wallet* más altos en el resto de las categorías. Por ejemplo, los clientes asignados al segmento de educación, presentan su menor principalidad en la categoría vivienda, con un 10,3%, sin embargo es la más alta en comparación al resto de los grupos en la misma categoría (sin considerar, lógicamente, al segmento vivienda). Por otro lado, se observa que aquellos segmentos en que la principalidad global es baja (vivienda, vestuario y hogar), presentan bajas principalidades en el resto de las categorías, a excepción de las unidades de consumo vestuario y hogar, donde presentan un *share of wallet* en torno al 50%, y en su propia categoría, en torno al 80%.

	Hogar	Vestuario	Otros	Recreación	Transporte	Vivienda	Alimentos	Salud	Educación
Hogar	82,9%	46,3%	23,5%	25,3%	8,7%	3,7%	8,1%	10,5%	0,7%
Vestuario	45,0%	81,9%	26,7%	28,1%	8,6%	4,3%	9,9%	15,7%	1,1%
Otros	44,4%	48,8%	78,7%	27,7%	8,3%	4,7%	7,6%	11,3%	0,9%
Recreación	56,0%	59,3%	31,5%	92,9%	9,8%	5,1%	10,8%	14,5%	1,3%
Transporte	33,1%	32,2%	18,9%	16,8%	61%	3,7%	9,8%	13,1%	1,0%
Vivienda	14,1%	13,6%	10,6%	7,7%	4,0%	31,2%	4,7%	5,4%	0,7%
Alimentos	50,9%	54,1%	28,0%	36,4%	17,6%	7,4%	69,2%	27,8%	2,2%
Salud	47,6%	59,5%	29,5%	30,7%	13,4%	5,8%	17,6%	82,1%	2,0%
Educación	55,8%	73,6%	37,6%	44,5%	19,3%	10,3%	20,5%	29,0%	96,1%

Tabla 26: Principalidad por unidad de consumo según segmento

Fuente: Elaboración propia

9.5. Segmentación utilizando *K-means*

Para realizar una segmentación utilizando la herramienta estadística de *k-means*, se debe establecer a priori el número de clúster a formar. Debido a la gran cantidad de registros, no es posible realizar un dendograma que permita visualizar el número de segmentos más adecuados, ni tampoco es posible utilizar la “regla del codo”. Es por esto, que la elección del número de grupos es realizada luego de hacer ocho iteraciones, donde en cada una de éstas se utiliza una cantidad distinta de segmentos, los cuales van desde dos hasta nueve.

Las variables utilizadas en la segmentación son previamente estandarizadas para manejar una misma escala de medida y poder compararlas entre sí. Lo anterior es realizado según la siguiente fórmula:

$$\bar{\omega} = \frac{\omega - MIN(rango)}{MAX(rango) - MIN(rango)} \quad (23)$$

donde ω corresponde al valor original y $\bar{\omega}$ al valor normalizado en una escala común de 0 a 1. Las variables fueron las siguientes:

- *Share of wallet* global.
- Gasto total estimado.
- *Share of wallet* en cada una de las nueve unidades de consumo.
- Antigüedad.
- Cupo.
- Recencia.
- Tipo de cliente (dummy).
- Edad.
- Género (dummy).
- Zona de residencia (dummy).

Luego de analizar cada una de las iteraciones, se decide trabajar con ocho segmentos, ya que en éstos se pueden distinguir claramente distintos perfiles de

clientes. Los centros de cada segmento son mostrados en tres tablas, una correspondiente a los centros según principalidad, otra enfocada a variables del negocio y finalmente otra de carácter demográfico.

Como se puede ver en la Tabla 27, los segmentos dos y cuatro son los que presentan los valores más altos en todas las variables. Dado que son los grupos que presentan la principalidad más alta, es de esperar que éstos estén conformado principalmente por clientes Elite y Premium, lo que es corroborado en la Tabla 28. Estos dos grupos son los que contienen a los clientes más leales, dado que tienen un gran uso de la tarjeta de crédito en todas las unidades de consumo. El grupo uno representa a aquellos clientes que utilizan la tarjeta de crédito en mayor medida sólo al interior del holding, ocurriendo lo mismo con el segmento tres, los cuales se diferencian del primer conglomerado en que contiene a clientes que gastan montos mayores con la tarjeta en relación a su gasto total, razón por la que presentan una mayor principalidad. El grupo cinco corresponde a los clientes menos leales a la empresa, ya que poseen un bajo nivel de utilización de la tarjeta en todas las unidades de consumo, pero presentan un gran potencial a ser capturado. En tanto el grupo siete, puede ser vinculado directamente al segmento inferior izquierdo mostrado en el Gráfico 23, ya que son clientes que presentan un bajo potencial de gasto junto con un bajo nivel de principalidad. En cuanto a los grupos seis y ocho, se observa que son similares al grupo uno, pero entre ellos se diferencian en las variables descritas en la Tabla 28 y Tabla K-2.

	Segmento							
	I	II	III	IV	V	VI	VII	VIII
Potencial	372.053	410.651	376.143	405.591	374.564	380.275	342.661	382.672
Show	14,84%	53,72%	25,58%	49,60%	7,36%	12,85%	16,66%	14,31%
Show Alimentos	8,35%	25,93%	6,32%	20,32%	2,09%	3,14%	6,18%	4,41%
Show Educación	0,43%	2,65%	0,76%	2,36%	0,19%	0,36%	0,67%	0,52%
Show Vestuario	41,67%	80,09%	71,51%	90,45%	16,14%	47,59%	53,69%	39,10%
Show Vivienda	2,79%	9,87%	3,86%	7,44%	1,84%	2,05%	3,22%	2,63%
Show Salud	6,16%	25,67%	10,90%	34,60%	3,70%	9,52%	9,11%	8,37%
Show Hogar	42,80%	82,36%	68,09%	80,40%	28,80%	39,47%	40,57%	43,59%
Show Otros	18,63%	48,44%	31,78%	51,88%	10,42%	16,98%	20,44%	19,26%
Show Recreación	20,71%	56,62%	38,86%	59,74%	9,02%	14,82%	26,04%	19,58%
Show Transporte	5,36%	24,87%	10,03%	12,91%	4,91%	3,06%	8,23%	5,23%

Tabla 27: Centros de los segmentos para variables de principalidad

Fuente: Elaboración propia

Como se puede ver en la Tabla 28, los segmentos que presentan una mayor principalidad (II y IV) son aquellos que presentan el mayor cupo, además, como se mencionó anteriormente, están conformados principalmente por clientes Elite y Premium. Junto con lo anterior, al observar la recencia de estos dos segmentos, se corrobora lo mostrado en el Gráfico 17, donde se observa que presentan el menor tiempo promedio desde la última compra.

	Segmento							
	I	II	III	IV	V	VI	VII	VIII
Antigüedad	99,85	153,384	132,895	150,976	154,508	153,713	19,989	159,114
Cupo	\$718.027,14	\$1.629.149,3	\$1.112.166,97	1.385.237,5	\$1.034.310,9	\$954.474,8	\$388.123,7	\$1.079.791,6
Recencia	1,688	.332	.475	.260	2,934	1,768	2,174	1,727
Elite	.000	.002	.002	.004	.000	.000	.000	.001
Normal II	.254	.000	.000	.004	.000	.000	1,000	.023
Premium	.016	.998	.000	.789	.006	.005	.000	.018

Tabla 28: Centros de los segmentos para variables de negocio

Fuente: Elaboración propia

Por otro lado, al observar la recencia del grupo cinco, se reafirma que son los clientes menos leales, ya que no sólo presentan el *share of wallet* más pequeño, sino que además presentan en promedio el valor más alto en este indicador en relación a los demás segmentos. Este grupo está compuesto principalmente por clientes Normales tipo III. En cuanto al grupo siete, éstos corresponden a los clientes más nuevos de la muestra, con el cupo más bajo y el segundo con el mayor tiempo desde la última compra (2 meses aproximadamente). Éstos están conformados en su totalidad por los clientes Normales tipo II. Tanto el grupo uno como el tres están compuestos principalmente por clientes Normales tipo III, los cuales como se mencionó, utilizan mayormente su tarjeta en vestuario y hogar. Ambos grupos se diferencian en el *share of wallet*, lo cual se reafirma al observar que la recencia del tres es mucho menor al del grupo uno (15 días versus dos meses).

El número de individuos en cada uno de los grupos es mostrado en el Anexo K, junto con los centros para las variables demográficas, los cuales no serán analizados en profundidad dado que no aportan mucha información. Sólo cabe decir respecto a estos centros que los segmentos dos, tres y cinco están compuestos sólo por hombres mientras que los grupos cuatro y seis sólo por mujeres, y el segmento siete presenta a los clientes más jóvenes con un promedio de edad de 37 años.

9.6. Propuestas de acción

A partir de los análisis realizados anteriormente, específicamente en las secciones 9.3 y 9.4, es posible proponer lineamientos generales que permitan aumentar el *share of wallet*. Estas propuestas se dividen básicamente en dos: cobertura y alcance. La primera consiste en aumentar el número de categorías en las que están comprando, de manera de aumentar el gasto con la tarjeta de crédito, mientras que la segunda hace

referencia a “alcanzar” el máximo porcentaje del gasto total en cada categoría en la que ya se encuentran activos.

El primer lineamiento se crea a partir de los segmentos mostrados en el Gráfico 23, donde se mostró que la empresa debiese trasladar a la gran mayoría de sus clientes al extremo derecho, enfocándose principalmente en aquellos que están entre el cuarto y quinto quintil de *size of wallet* y en el primer y segundo quintil de *share of wallet*. Para poder aumentar el número de categorías en las que están comprando con la tarjeta, se debe primero encontrar cuáles son las posibles razones que generan la limitante. En general se identifican tres causas:

- Límite de cupo

Al observar en la muestra cuanto es el cupo disponible promedio en el periodo de estudio después de pagar las deudas asociadas a la tarjeta, se observa que este es de un 60,5% del cupo total. En particular, un 35,8% de los clientes de la muestra presentan un cupo disponible después del pago, inferior a un 50% del cupo real, lo que provoca que su posibilidad de gasto con la tarjeta disminuya aún más.

En particular, es interesante evaluar la cantidad de clientes que tienen asociado un cupo inferior a su potencial de gasto, donde se ve que un 9,3% de la muestra tiene asociado un cupo inferior al 60% de su gasto total mensual. Lo anterior genera que estos clientes tengan siempre una principalidad baja, nunca pudiendo alcanzar su potencial.

Debido a lo mencionado anteriormente, el área debe evaluar en conjunto con riesgo la posibilidad de aumentar el cupo a los clientes que presentan cupo insuficiente (inferior al potencial) y a los que presentan un cupo saturado (inferior al 50% del cupo real). Un aumento de cupo generaría un cambio en la conducta de estos clientes, ya que les daría la posibilidad de comprar montos mayores con la tarjeta.

- No poseer la tarjeta Visa.

Este punto hace referencia a aquellos clientes que tienen una principalidad alta en las unidades de consumo que cubre el holding, pero no pueden acceder al resto de las categorías debido a que en ausencia de alianzas, no pueden utilizar la tarjeta de crédito en éstas. Por el motivo anterior, es que la empresa debe evaluar nuevamente en conjunto con el área de riesgo, una migración de estos clientes a la tarjeta Visa. Esta acción, permitiría que estos clientes comiencen a utilizar la tarjeta en el resto de las unidades de consumo, lo que provocaría que su gasto con Visa aumente, capturando un mayor porcentaje de su gasto total.

- No tener conocimiento de lo que es Visa.

En cuanto a esto, la solución consiste en proveer información a las personas que tienen Visa, un bajo *share of wallet*, un alto potencial de gasto y compras principalmente al interior del holding, ya que a partir de lo conversado con la empresa, existen clientes que poseen tarjeta Visa, pero desconocen que ésta puede ser utilizada en comercios

fuera del holding, por lo que su principalidad es baja. Muchos de estos clientes son potenciales usuarios de la tarjeta en el resto de las categorías, lo que generaría un aumento del *share of wallet* de cada uno de ellos si tienen un mayor conocimiento de su uso.

El segundo lineamiento, referido al alcance, nace a partir de los segmentos presentados en la Tabla 26, donde se observa que existen clientes vinculados en mayor medida a ciertas unidades de consumo pero presentan baja principalidad en el resto de las categorías. El objetivo final de este lineamiento es trasladar a los clientes a los segmentos que presentan un *share of wallet* global mayor (educación, alimentos, salud), ya que una vez que son principales en estas categorías, lo son en la mayoría de las unidades de consumo.

Para lograr lo anterior, se proponen dos herramientas, las cuales consideran las características demográficas de los clientes, puesto que el ciclo de vida de una persona, su género y lugar de origen, influyen en el nivel de propensión a gastar en ciertas categorías. Un ejemplo de esto son los jóvenes, los cuales tienen baja propensión a gastar en educación en relación a personas de un rango etario adulto joven, quienes tienen mayor probabilidad de tener hijos. Las dos metodologías son las siguientes:

- Modelos de propensión

Por medio de modelos de propensión, que incluyan variables demográficas, de negocio, comportamiento y principalmente de *share of wallet*, es posible identificar grupos de clientes que tienen una mayor probabilidad de aumentar su gasto con la tarjeta en las categorías que poseen baja principalidad por medio de campañas de marketing. Una de las campañas que sirve para este propósito, consiste en ofrecer una mayor cantidad de puntos a los segmentos objetivos en caso de utilizar su tarjeta en las unidades de consumo que se quieren potenciar. Si bien este tipo de campañas, como cualquier otra que tenga por objetivo generar clientes más leales, tiene efectos positivos pequeños al corto plazo, podrían tener efectos satisfactorios mucho mayores en el largo plazo en cuanto al *share of wallet* de los clientes [8]. Lo anterior, combinado con los primeros lineamientos mencionados, los cuales apuntan a cambiar la conducta del consumidor en el largo plazo, ayudan a aumentar la principalidad de la tarjeta de crédito en cada una de las unidades de consumo, logrando aumentar el número de categorías en que las personas gastan con el medio de pago de la empresa.

- *Cross-selling*

Al igual que los modelos de propensión, las ventas cruzadas tienen por objeto aumentar el gasto en las categorías en que los clientes son menos principales. A partir de la información entregada por la Tabla 26, se pueden generar estrategias de *cross-selling*, como por ejemplo ofrecer descuentos en las categorías en las que los clientes se encuentran más vinculados, en caso de que gasten en otra unidad de consumo en la que son menos principales y exista una propensión a comprar en ella, razón por la cual esta herramienta está vinculada a la presentada anteriormente.

Por medio de las dos herramientas presentadas, se pretende traspasar a los clientes vinculados a las categorías con menor principalidad global, a aquellas que presentan un mayor *share of wallet*, ya que estas categorías en general presentan un alto valor de este indicador en la mayoría de las unidades de consumo.

A partir de los dos lineamientos presentados, se busca que la tarjeta de crédito sea transversal en las categorías para todos los clientes, en primer lugar incentivándolos a comprar en ellas (primer lineamiento), para luego provocar un aumento en el gasto en cada una de ellas utilizando el medio de pago de la empresa (segundo lineamiento).

10. CONCLUSIONES

10.1. Conclusiones del trabajo

Esta memoria tuvo como objetivo general estimar el gasto mensual de los clientes considerando todos los medios de pago, utilizando sólo el comportamiento transaccional con la tarjeta de crédito de la empresa y variables demográficas y de comportamiento. Para lograr lo anterior, se utilizó una metodología nueva, la cual no requiere que una muestra de clientes reporten su potencial de gasto para poder calibrar el modelo. Dado lo anterior, este trabajo resulta ser complejo, ya que se tiene el problema de información incompleta, es decir, se desconoce cuánto gastan mensualmente una muestra de los clientes considerando cualquier medio de pago, lo cual es información exógena a la empresa. Es por lo anterior que este modelo requiere hacer supuestos acerca del potencial que se quiere estimar, donde supone que el número de transacciones totales que hace una persona se distribuye como una variable aleatoria de poisson.

El proyecto surge de la necesidad de una empresa que quiere estimar cuánto del gasto mensual de sus clientes están capturando por medio de la tarjeta de crédito que ellos ofrecen, información relevante para poder diseñar campañas de marketing mejor fundadas, teniendo en cuenta que clientes presentan potencial de gasto para adquirir con la tarjeta y quienes están gastando todo lo que pueden con ésta.

Dado que la metodología propuesta en este proyecto sólo ha sido utilizada en un proyecto de investigación, fue necesario realizar una simulación previa para evaluar que el modelo tenía capacidad de predicción, donde se obtuvo que su error de predicción era menor cuando el *share of wallet* de los clientes era mayor. Sin embargo, esta simulación fue realizada sólo considerando variables tipo transaccionales, ya que el objetivo principal de realizar esto era el de tener certeza que el modelo si lograba recuperar información, por lo que no es directo inferir lo anterior al aplicar el modelo que considera variables demográficas.

El modelo utilizado finalmente, corresponde al GBM extendido, el cual incluye variables demográficas, transaccionales y de negocio para poder predecir el gasto total de cada cliente. Este modelo, al utilizar el mismo benchmark que usa la empresa para evaluar la metodología actual que ellos poseen, arroja un error de ajuste y validación de un 12,5%, lo cual resulta ser 12,5 puntos menor al que actualmente utiliza la compañía.

Al comparar los resultados arrojados por la metodología propuesta en comparación a los que posee la empresa, se observa que la principalidad promedio de ambas metodologías resulta ser similar (20%), sin embargo, el modelo de la empresa tiene el problema de que muchos de sus clientes presentan un *share of wallet* subestimado y otros sobrestimados dada la forma en que fueron calculados. En particular, una gran cantidad de clientes presentan principalidad uno con el modelo de la empresa, ya que tienen un potencial de gasto estimado inferior al que hacen con la tarjeta de crédito, lo cual no tiene sentido. Por otro lado, el GBM genera una mayor heterogeneidad en el gasto estimado de los clientes, en contraste a la metodología actual, donde el gasto depende del grupo de clientes al que pertenece, los cuales son formados según variables demográficas.

Un resultado interesante, se obtiene al contrastar el valor del cliente con su nivel de principalidad. A partir de esto, se obtiene que existe una gran cantidad de clientes que se encuentran en los segmentos con mayor rentabilidad, los cuales presentan potencial para aumentar aún más ésta, dado que muchos de ellos se encuentran con una principalidad baja.

A partir de la segmentación realizada, se obtuvo información útil para proponer líneas de acción que aumenten la principalidad de la cartera de clientes. Dentro de ésta, se encuentra el hecho de que los clientes vinculados a unidades de consumo que no son satisfechas por el holding, son los que presentan una mayor principalidad, donde el grupo más principal resulta ser aquellos que tienen el mayor *share of wallet* en educación. Además, estos grupos son los que poseen un mayor porcentaje de clientes con Visa. Con esta información, se propusieron dos líneas de acción con el objeto de aumentar la principalidad de la cartera, donde la primera esta enfocada en cubrir una mayor cantidad de categorías con la tarjeta, lo cual se realiza por medio de aumentos de cupo y migración de clientes a la tarjeta Visa, y el segundo curso de acción consiste en alcanzar un mayor gasto con la tarjeta de crédito en cada una de las unidades de consumo, para lo cual se utilizan técnicas de *cross-selling* y modelos de propensión, los cuales apoyan a definir los segmentos objetivos para las campañas de marketing.

10.2. Limitaciones del modelo

Dado que para la aplicación del modelo se requiere que los montos gastados en el periodo de evaluación estén discretizados, se necesita escoger un monto para aquella función tal que el número de transacciones no sea tan grande, ya que la estimación mediante la ecuación 14, quedaría indefinida. Es importante mencionar que si se pudiera utilizar los montos gastados sin ninguna modificación, la estimación sería mucho mejor en cuanto a verosimilitud, pero el problema recae en que el *size of wallet* quedaría indefinido. Es por lo anterior que, si se aplica la metodología propuesta a clientes con gastos muy superiores al máximo presente en la muestra, no es posible estimar su gasto total. Una solución a lo anterior consiste en utilizar un monto mayor al propuesto para discretizar¹³, de manera de que el número de transacciones para este tipo de clientes sea menor. El problema de esto, es que se pierde información y

¹³ Se debe destacar que el valor de \$30.000 se adapta a la muestra utilizada, sin embargo este valor podría variar si se considera otra muestra de clientes.

exactitud para la gran mayoría de los clientes, los cuales se encuentran en un rango de gasto adecuado. Es por lo anterior de que existe un *trade off* entre poder estimar para todos los clientes su gasto con una menor precisión, versus la opción de estimarlo para la gran mayoría de los clientes, dejando de lado los casos outliers, con mayor exactitud.

Una segunda limitación del modelo tiene que ver con la distribución impuesta al número de transacciones, donde se asumió que se distribuye como poisson, lo que implica que el tiempo entre cada transacción es exponencial. Lo anterior genera que, dada las propiedades de la exponencial, exista pérdida de memoria, por lo que la probabilidad de realizar una compra después de un periodo de tiempo corto, es la misma que si ha pasado un periodo de tiempo más largo, lo cual no es cierto para la mayoría de las categorías. Un ejemplo de esto se aprecia en la categoría vestuario, donde es de esperar que la probabilidad de realizar una compra sea más alta cuando ha pasado un tiempo mayor desde la última compra.

Finalmente, debido a la existencia tanto de clientes que presentan un cupo inferior a su potencial, como otros que presentan un cupo saturado, sus compras con la tarjeta están condicionadas a estas limitantes, lo que genera que no se pueda capturar toda la información de su comportamiento de gasto, por lo que modelo tiende a subestimar el potencial de gasto de una gran masa de clientes.

10.3. Recomendaciones y Trabajos futuros

A modo de recomendación y trabajo futuro, se sugiere levantar información para una muestra de clientes respecto de cuanto gastan mensualmente, lo cual apoyará dos objetivos. El primero de ellos, responde a evaluar con mayor precisión la calidad de predicción de la metodología propuesta en esta memoria, mientras que el segundo tiene relación con la aplicación de la metodología propuesta en [1], el cual utiliza *list augmentation*, es decir, agregar datos provenientes de encuestas o fuentes secundarias a una muestra de clientes en la base de datos existente, para luego aplicar tres modelos supervisados.

Por otro lado, también se propone como trabajo futuro aplicar la metodología propuesta en [9], la cual, al igual que la metodología propuesta en este trabajo, no utiliza datos externos para calibrar el modelo. Esta metodología modela el tiempo entre compras, asumiendo una distribución *Erlang 2* (en la metodología de éste trabajo, dado que el número de transacciones se distribuye como poisson, el tiempo entre compras es exponencial, es decir, *Erlang 1*).

El objetivo de los trabajos propuestos anteriormente, es evaluar cuán diferentes resultan ser las estimaciones y en que grupos de clientes difieren más y/o menos, además de analizar cuál de las tres metodologías posee un mayor grado de predicción para los datos con los que se está trabajando.

Finalmente, se propone analizar de manera dinámica la evolución del *share of wallet*, incorporando variables de la empresa, como lo son campañas o acciones de marketing.

11. BIBLIOGRAFÍA

- [1] R. Yuxing Du, W. Kamakura and C. Mela, "Size and Share of Customer Wallet," *Journal of Marketing*, vol. LXXI, pp. 94-113, 2007.
- [2] C. C. Nicolas Glady, "Predicting customer wallet without survey data," *Journal of service research*, vol. XI, pp. 219-231, 2009.
- [3] C. Perlich, S. Rosset, B. Zadrozny, S. Merugo, S. Weiss and R. Lawrence, "Wallet estimation models," IBM Research Report, NY, 2006.
- [4] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," 1996.
- [5] K. Train, *Discrete Choice Methods with Simulation*, Cambridge University Press, 2003.
- [6] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *International Joint Conference on Artificial Intelligence*, 1995.
- [7] W. Kamakura and V. Mittal, "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the moderating effect of customer characteristics," *Journal of marketing research*, vol. XXXVIII, pp. 131-142, 2001.
- [8] P. C. Verhoef, "understanding the effect of customer relationship management efforts on customer retention and customer share development," *Journal of Marketing*, vol. LXVII, pp. 30-45, 2003.
- [9] Y. Chen and J. Steckel, "Modeling Credit Card Share of wallet: Solving the incomplete information problem," *Journal of marketing research*, vol. XLIX, pp. 655-669, 2012.

12. ANEXOS

ANEXO A

En el siguiente gráfico, el cual es construido utilizando los mil registros de cada repetición (es decir, 1.000.000 puntos en el gráfico), es posible concluir lo mismo que en la sección 8.1.1, donde se evidencia de que la gran mayoría de puntos que se encuentran en repeticiones con bajo MAPE, corresponde a aquellos que tienen un alto valor tanto en n_i como en x_i . Es importante observar que el modelo tiene un MAPE alto en el caso en que n_i es muy superior a x_i .

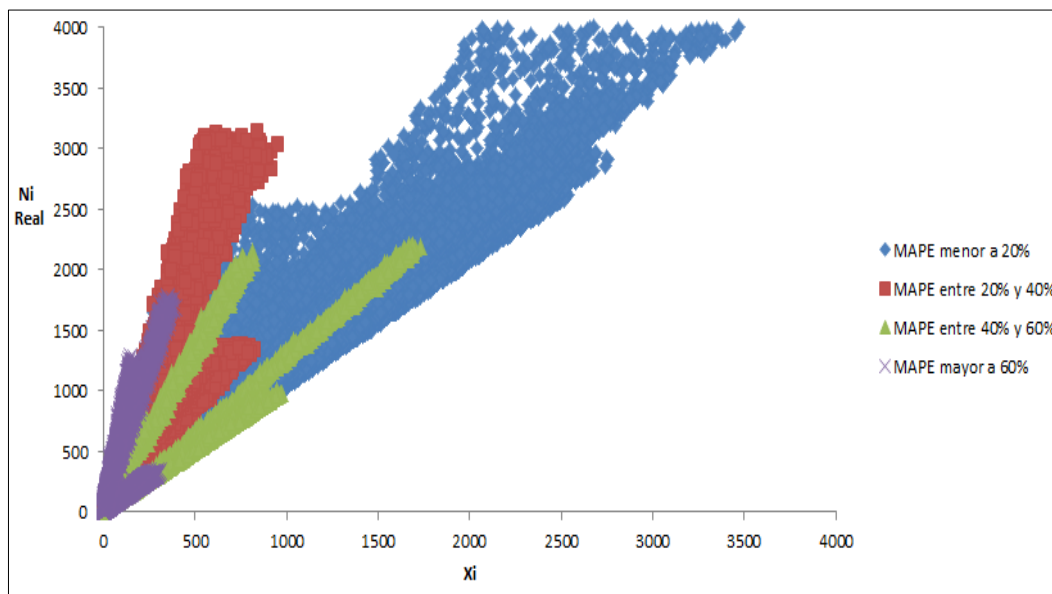


Gráfico A- 1: MAPE obtenido en función de ni y xi
Fuente: Elaboración propia

ANEXO B: *Ten-fold cross validation* para el modelo log-lin

Iteración	MAPE AJUSTE	MAPE PREDICCIÓN	R^2 Ajustado
1	36.2%	36.1%	0.205
2	36.2%	36.2%	0.206
3	36.2%	36.1%	0.207
4	36.2%	36.0%	0.204
5	36.2%	36.4%	0.206
6	36.2%	35.9%	0.205
7	36.2%	36.2%	0.206
8	36.1%	36.8%	0.206
9	36.2%	36.0%	0.205
10	36.2%	36.0%	0.205
Promedio	36.2%	36.2%	0.2055

Tabla B-1: Fold cross validation
Fuente: Elaboración propia

ANEXO C: *Ten-fold cross validation* para el GBM Transaccional

Iteración	MAPE AJUSTE	MAPE PREDICCIÓN	AIC	BIC	ρ^2
1	26,60%	26,38%	571507	571517	63,9%
2	26,42%	26,36%	578229	578239	63,9%
3	26,64%	26,87%	579671	579681	64,0%
4	26,41%	26,50%	580935	580945	63,9%
5	26,57%	26,60%	582329	582339	64,0%
6	26,69%	26,75%	581922	581932	64,0%
7	26,42%	26,54%	581257	581267	64,0%
8	26,72%	26,41%	583048	583058	63,9%
9	26,79%	26,63%	580778	580787	64,0%
10	26,48%	26,47%	580341	580351	64,0%
<i>Promedio</i>	26,57%	26,55%	580001.7	580011.6	64,0%

Tabla C- 1: Fold cross validation para GBM transaccional

Fuente: Elaboración propia

ANEXO D: Histogramas del modelo GBM transaccional

En el siguiente gráfico se muestra el histograma de los n_i estimados con el GBM transaccional. En él se puede apreciar que el rango de transacciones promedio mensual utilizando cualquier medio de pago va desde aproximadamente 7 a 22.

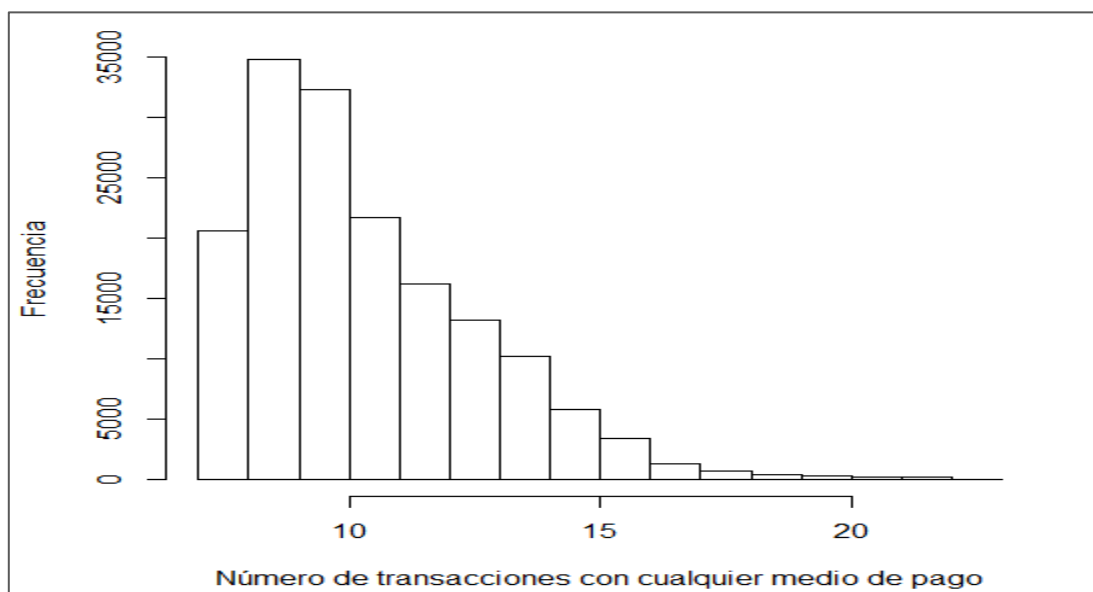


Gráfico D-1: Histograma de los n_i estimados con GBM transaccional

Fuente: Elaboración propia

En el Gráfico D-2 se muestra el histograma de los π_i estimados con el GBM transaccional. En este histograma, como se mencionó en la sección 8.3, se puede apreciar que existe una gran masa de clientes con probabilidades muy altas de utilizar la tarjeta en cada transacción.

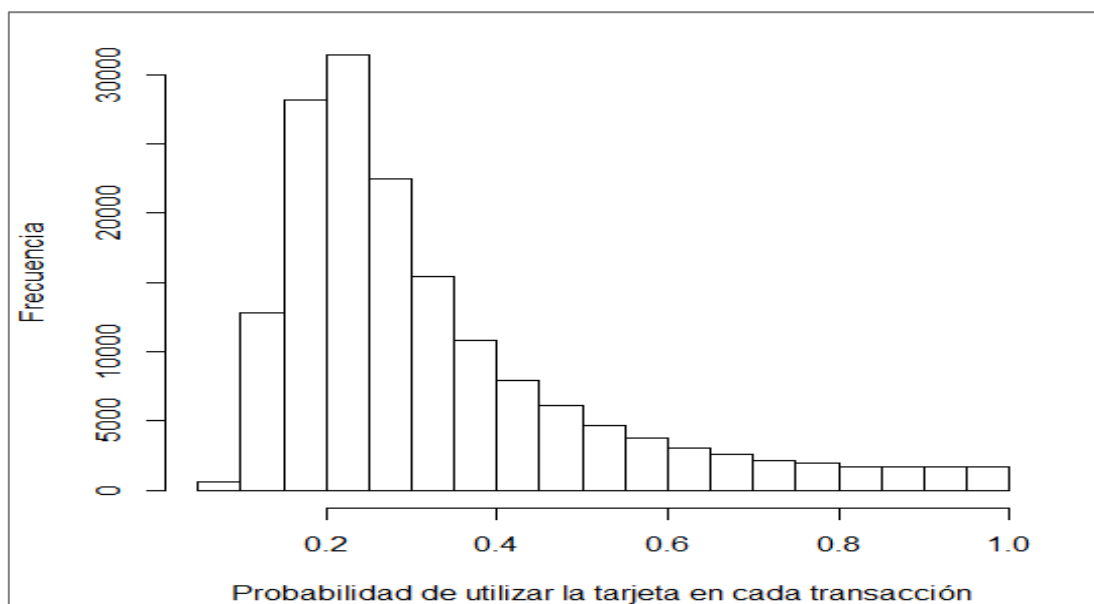


Gráfico D- 2: Histograma de la probabilidad estimada con el GBM transaccional
Fuente: Elaboración propia

ANEXO E: Código para la estimación del GBM Extendido

El siguiente código corresponde a la estimación de los parámetros utilizando el modelo GBM extendido.

```
rm(list=ls())      # clear the list of objects
graphics.off()    # clear the list of graphs
options(digits = 5)  # number of digits to display
### READ AND FORMAT THE DATA###
library(foreign) #Paquete para leer archivos desde SPSS

paneldata = read.spss("baseajuste_80.sav", to.data.frame=T)
n = nrow(paneldata) #número de clientes
idcliente=paneldata[,1]
age = paneldata[,2]
monto1=paneldata[,3]
monto2=paneldata[,4]
recency=paneldata[,5]
tipo_cliente=paneldata[,6]
gse=paneldata[,7]
rango_edad=paneldata[,8]
sexo=paneldata[,9]
zona=paneldata[,10]
tenure=paneldata[,11]
```

```

potencial_actual=paneldata[,12]
renta=paneldata[,13]
añoapertura=paneldata[,15]
mesapertura=paneldata[,16]
cupo=paneldata[,26]
factori=paneldata[,27]
dummy_sexo=matrix(0,n,1)
for(j in 1:n){
if (sexo[j]==1) dummy_sexo[j]=0
else dummy_sexo[j]=1
}
dummy_zona1=matrix(0,n,1)
dummy_zona2=matrix(0,n,1)
dummy_zona3=matrix(0,n,1)
dummy_zona4=matrix(0,n,1)
dummy_zona5=matrix(0,n,1)
dummy_zona6=matrix(0,n,1)
dummy_zona7=matrix(0,n,1)
dummy_zona8=matrix(0,n,1)

for(j in 1:n){
if (zona[j]==1) dummy_zona1[j]=1
else dummy_zona1[j]=0
}
for(j in 1:n){
if (zona[j]==2) dummy_zona2[j]=1
else dummy_zona2[j]=0
}
for(j in 1:n){
if (zona[j]==3) dummy_zona3[j]=1
else dummy_zona3[j]=0
}
for(j in 1:n){
if (zona[j]==4) dummy_zona4[j]=1
else dummy_zona4[j]=0
}
for(j in 1:n){
if (zona[j]==5) dummy_zona5[j]=1

```

```

else dummy_zona5[j]=0
}
for(j in 1:n){
if (zona[j]==6) dummy_zona6[j]=1
else dummy_zona6[j]=0
}
for(j in 1:n){
if (zona[j]==7) dummy_zona7[j]=1
else dummy_zona7[j]=0
}
for(j in 1:n){
if (zona[j]==8) dummy_zona8[j]=1
else dummy_zona8[j]=0
}
dummy_elite=matrix(0,n,1)
dummy_premium=matrix(0,n,1)
dummy_normal1=matrix(0,n,1)
dummy_normal2=matrix(0,n,1)

for(j in 1:n){
if (tipo_cliente[j]==1) dummy_elite[j]=1
else dummy_elite[j]=0
}
for(j in 1:n){
if (tipo_cliente[j]==4) dummy_premium[j]=1
else dummy_premium[j]=0
}
for(j in 1:n){
if (tipo_cliente[j]==3) dummy_normal1[j]=1
else dummy_normal1[j]=0
}
for(j in 1:n){
if (tipo_cliente[j]==2) dummy_normal2[j]=1
else dummy_normal2[j]=0
}

tenureaño1=(2011-añoapertura)*12+(5-mesapertura)
avermonto2=matrix(0,n,1)

```

```

for(j in 1:n){
if (tenure[j]>=12) avermonto2[j]=monto2[j]/12
else avermonto2[j]=monto2[j]/tenure[j]
}
avermonto1=matrix(0,n,1)
for(j in 1:n){
if (tenureaño1[j]>=12) avermonto1[j]=monto1[j]/12
else avermonto1[j]=monto1[j]/tenureaño1[j]
}
norm=30000 #Valor que discretiza
x=matrix(0,n,1) #Número de transacciones con la tarjeta
for(j in 1:n) {
x[j]=ceiling(avermonto2[j]/norm)
}
ones = rep(1,n)

V=cbind(ones,log(avermonto1+1),((tenureaño1)-min(tenureaño1))/(max(tenureaño1)-
min(tenureaño1)),log(age+1),dummy_sexo,dummy_zona1,dummy_zona2,dummy_zona3,dummy_zona4,
dummy_zona5,dummy_zona6,dummy_zona7)
W=cbind(ones,((avermonto1)-min(avermonto1))/(max(avermonto1)-min(avermonto1)),((tenureaño1)-
min(tenureaño1))/(max(tenureaño1)-min(tenureaño1)),log(cupo+1),dummy_normal1,dummy_normal2)

##DRAW OF THE REALISED VALUES
loglikel<- function(beta){
lambdai=exp(V%%beta[c(1:12)])
logpi=W%%beta[c(13:18)]
phi=1/(1+exp(-logpi))
lli=lambdai*phi-x*log(lambdai*phi)
return(sum(lli))
}

### LIKELIHOOD MINIMIZATION
beta.start = 0.1*rep(1,18)
mymleS = optim(par=beta.start, fn=loglikel, hessian=TRUE,method="SANN", control = list(maxit=500,
trace=TRUE, REPORT=10))
mymleB = optim(par=mymleS$par, fn=loglikel, hessian=TRUE,method="BFGS", control = list(maxit=3000,
trace=TRUE, REPORT=10))

```

#Reportes

```
mle = mymleB$par  
se = sqrt(diag(solve(mymleB$hessian)))  
cbind(mle,se)
```

#Reescribir parámetros estimados

```
lambdai=exp(V%%mle[1:12]) #beta:1-5.....betha  
logpi=W%%mle[13:18] #beta:6-8.....alpha  
phi=1/(1+exp(-logpi))  
lli2=lambdai*phi-x*log(lambdai*phi)+log(factorial(x))  
verosimilitud=sum(lli2)  
npar = length(mle)  
rho = 1 - ( -(verosimilitud) / -(loglikel(beta=rep(0,npar))+sum(log(factorial(x))))));  
aic = -2*-(verosimilitud) + 2*length(npar) ;  
bic = -2*-(verosimilitud) + length(npar)*log(n) ;
```

#Estimación del Size of wallet

```
niestim = (lambdai*(1-ppois(x-2,lambdai,log=F))/(1-ppois(x-1,lambdai,log=F)))  
nrent=niestim*norm/factori  
MAPE= (1/n)*sum(abs((nrent-renta)/renta))
```

ANEXO F: *Ten-fold cross validation* para el GBM Extendido

Iteración	MAPE AJUSTE	MAPE PREDICCIÓN	AIC	BIC	ρ^2
1	12,4%	12,4%	537530	537540	66,06%
2	11,9%	11,9%	543818	543827	66,09%
3	12,7%	12,8%	545065	545075	66,11%
4	12,3%	12,3%	546427	546437	66,09%
5	12,1%	12,1%	547417	547427	66,14%
6	12,4%	12,4%	547314	547323	66,10%
7	12,6%	12,7%	546875	546885	66,12%
8	12,6%	12,5%	548323	548333	66,09%
9	12,7%	12,6%	546224	546234	66,10%
10	12,3%	12,2%	545677	545687	66,14%
<i>Promedio</i>	12,4%	12,4%	545467	545477	66,10%

Tabla F-1: Fold cross validation para GBM extendido

Fuente: Elaboración propia

ANEXO G: Histogramas del modelo GBM extendido

En el Gráfico G- 1, se observa que el modelo extendido identifica dos grande grupos de clientes en torno a la probabilidad de utilizar la tarjeta de crédito. En el primero de ellos existe una gran cantidad de clientes con una probabilidad en torno al 20%, los cuales tienen una baja principalidad, mientras que el segundo segmento, el cual está conformado por una pequeña cantidad de clientes, son aquellos que tienen un *share of wallet* superior al 50%, ya que la probabilidad de utilizar la tarjeta en ese grupo es del orden del 60%.

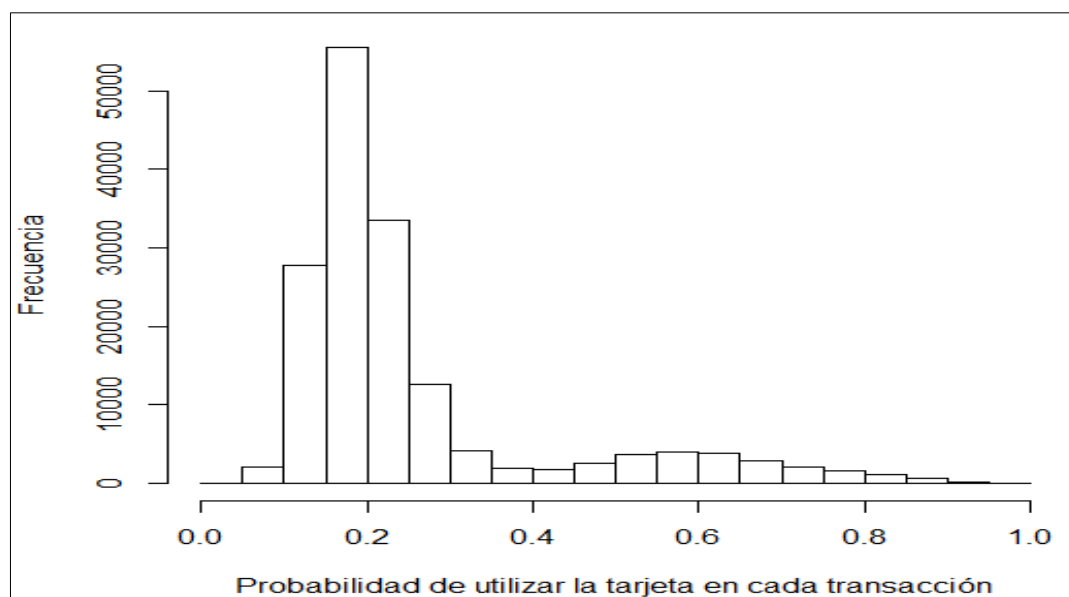


Gráfico G- 1: Histograma de la probabilidad estimada con el GBM extendido
Fuente: Elaboración propia

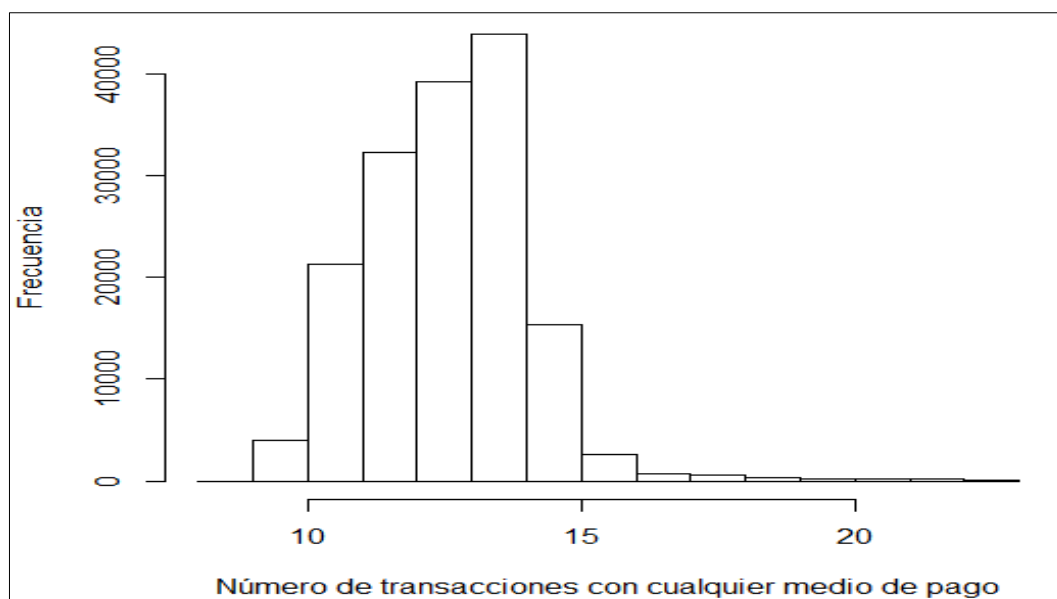


Gráfico G- 2: Histograma de los ni estimados con GBM extendido
Fuente: Elaboración propia

ANEXO H: Descripción de los segmentos según principalidad y *share of wallet*

Size/Share of wallet	[0,20%]	(20%,40%]	(40%,60%]	(60%,80%]	(80%,100%]
Quintil Bajo	\$ 316.816,60	\$ 318.249,72	\$ 320.851,88	\$ 327.641,95	-
2° quintil	\$ 351.528,49	\$ 351.590,06	\$ 351.729,07	\$ 353.827,62	\$ 364.520,58
3° quintil	\$ 378.485,74	\$ 378.541,52	\$ 378.325,11	\$ 378.268,10	\$ 384.089,18
4° quintil	\$ 399.868,34	\$ 399.907,02	\$ 400.268,90	\$ 400.496,85	\$ 401.494,79
Top quintil	\$ 425.045,72	\$ 424.821,70	\$ 425.722,01	\$ 443.308,95	\$ 525.310,50

Tabla H- 1: Gasto promedio con cualquier medio de pago

Fuente: Elaboración propia

Size/Share of wallet	[0,20%]	(20%,40%]	(40%,60%]	(60%,80%]	(80%,100%]
Quintil Bajo	\$ 26.162,61	\$ 89.418,76	\$ 152.946,34	\$ 210.713,85	-
2° quintil	\$ 29.775,06	\$ 99.738,23	\$ 170.346,68	\$ 233.018,09	\$ 294.223,17
3° quintil	\$ 32.273,10	\$ 106.594,34	\$ 183.135,39	\$ 256.371,18	\$ 317.793,96
4° quintil	\$ 34.371,93	\$ 112.746,64	\$ 193.288,61	\$ 273.703,76	\$ 327.598,41
Top quintil	\$ 34.972,34	\$ 121.253,42	\$ 211.966,58	\$ 313.496,58	\$ 455.247,85

Tabla H- 2: Gasto promedio con la tarjeta de crédito

Fuente: Elaboración propia

Size/Share of wallet	[0,20%]	(20%,40%]	(40%,60%]	(60%,80%]	(80%,100%]
Quintil Bajo	14,2033%	4,4284%	1,2566%	0,1118%	0,0000%
2° quintil	13,0777%	4,6415%	1,7728%	0,5056%	0,0025%
3° quintil	12,9361%	4,6557%	1,7666%	0,6324%	0,0093%
4° quintil	12,6093%	4,8334%	1,8468%	0,6659%	0,0447%
Top quintil	10,5296%	3,5090%	1,9021%	2,2779%	1,7809%

Tabla H- 3: Porcentaje de clientes en cada segmento

Fuente: Elaboración propia

ANEXO I: Ítems en unidades de consumo

ALIMENTOS Y BEBIDAS
PAN, CEREALES, PASTAS
CARNE
PESCADO Y MARISCO
LECHE, QUESO Y HUEVOS
ACEITES MANTEQUILLA Y MARGARINA
FRUTA
VERDURAS Y TUBERCULOS
AZUCAR, CAFE, TE, DULCES Y CONDIMENTOS
BEBIDAS Y GASTOS EN RESTAURANTS

Tabla I- 1: Unidad de consumo Alimentos

Fuente: VI Encuesta de presupuestos familiares. INE

VESTUARIO Y CALZADO

VESTUARIO PARA HOMBRE Y NIÑO

VESTUARIO PARA MUJER Y NIÑA

VESTUARIO Y ACCESORIOS PARA GUAGUAS

ACCESORIO DE VESTUARIO PARA HOMBRE

ACCESORIO DE VESTUARIO PARA MUJER

MATERIAL PARA CONFECCION Y ARREGLO DE ROPA

HECHURA Y ARREGLO DE PRENDAS DE VESTIR

CALZADO

VESTUARIO SIN ESPECIFICAR

Tabla I- 2: Unidad de consumo Vestuario

Fuente: VI Encuesta de presupuestos familiares. INE

GASTO DE LA VIVIENDA, AGUA, COMBUSTIBLE Y ELECTRICIDAD

ARRIENDO, DIVIDENDO, CONTRIBUCIONES Y OTROS GASTOS DE REPARACIÓN Y
MANTENCIÓN

AGUA POTABLE

COMBUSTIBLE Y ELECTRICIDAD

COMUNICACIONES

Tabla I- 3: Unidad de consumo Vivienda

Fuente: VI Encuesta de presupuestos familiares. INE

EQUIPAMIENTO Y CUIDADOS DE LA CASA

MUEBLES

ACCESORIOS Y DECORACIONES

TEXTILES PARA EL HOGAR

ARTEFACTOS PARA EL HOGAR

VAJILLA, CRISTALERIA, UTENSILIOS DOMESTICOS

PRODUCTOS Y UTENSILIOS PARA MANTENIMIENTO DEL HOGAR

SERVICIO DOMESTICO

Tabla I- 4: Unidad de consumo Hogar

Fuente: VI Encuesta de presupuestos familiares. INE

GASTO EN SERVICIOS MEDICOS

MEDICAMENTOS

ARTICULOS AUXILIARES

APARATOS Y EQUIPOS AUXILIARES

SERVICIOS DE PROFESIONALES DE LA SALUD

ATENCION HOSPITALARIA

SEGURO DE SALUD

Tabla I- 5: Unidad de consumo Salud

Fuente: VI Encuesta de presupuestos familiares. INE

TRANSPORTE

TRANSPORTE PRIVADO

TRANSPORTE PUBLICO

Tabla I- 6: Unidad de consumo Transporte
Fuente: VI Encuesta de presupuestos familiares. INE

ENSEÑANZAENSEÑANZA

Tabla I- 7: Unidad de consumo Educación
Fuente: VI Encuesta de presupuestos familiares. INE

RECREACION

EQUIPOS Y ACCESORIOS

REPARACION DE ARTICULOS RECREATIVOS

SERVICIOS DE RECREACION Y ESPARCIMIENTO

LIBROS, DIARIOS Y REVISTAS

Tabla I- 8: Unidad de consumo Recreación
Fuente: VI Encuesta de presupuestos familiares. INE

OTROS BIENES Y SERVICIOS

CUIDADO Y EFECTOS PERSONALES

RELOJES, JOYAS Y FANTASIAS

OTROS ARTICULOS PERSONALES

TABACO

COMPUTADOR PERSONAL, ARTICULOS Y SUMINISTROS DE ESCRITORIO

GASTO EN LUGARES PARA ALOJAMIENTO

SERVICIOS FINANCIEROS

OTROS SERVICIOS

Tabla I- 9: Unidad de consumo Otros
Fuente: VI Encuesta de presupuestos familiares. INE

ANEXO J: Matriz de correlación entre principalidad por categorías

	Alimentos	Vestuario	Hogar	Otros	Recreación	Educación	Vivienda	Transporte	Salud
Alimentos	1.000	0.259	0.188	0.180	0.216	0.096	0.183	0.254	0.330
Vestuario		1.000	0.263	0.301	0.277	0.076	0.163	0.151	0.296
Hogar			1.000	0.219	0.217	0.029	0.120	0.157	0.127
Otros				1.000	0.231	0.056	0.170	0.145	0.174
Recreación					1.000	0.061	0.138	0.120	0.162
Educación						1.000	0.074	0.068	0.084
Vivienda							1.000	0.132	0.148
Transporte								1.000	0.195
Salud									1.000

Tabla J- 1: Correlación entre principalidad por categorías
Fuente: Elaboración propia

ANEXO K: Clientes en cada segmento del *k-means* y centros para variables demográficas

Segmento	Número de clientes
I	6594
II	12215
III	23147
IV	12049
V	32755
VI	36878
VII	25046
VIII	12300

Tabla K- 1: Número de clientes en cada clúster
Fuente: Elaboración propia

	Grupo							
	I	II	III	IV	V	VI	VII	VIII
Edad	41,19	45,27	43,53	44,16	47,83	46,25	36,82	46,37
Dummy Hombre	.6330	1,0000	1,0000	.0000	1,0000	.0000	.6751	.5590
Stgo. Oriente	.0000	.0431	.0418	.0856	.0507	.0990	.0354	.0000
Stgo. Sur Oriente	.0000	.1044	.0047	.0953	.0000	.0000	.0851	1,0000
Stgo. Centro	.0000	.0303	.0288	.0432	.0263	.0425	.0335	.0000
Stgo. Norte	.0000	.0527	.0663	.0631	.0631	.0655	.0678	.0000
Stgo. Poniente	.0000	.1059	.1171	.1105	.1363	.1315	.1182	.0000
Stgo. Sur	.0000	.0928	.0914	.0922	.1031	.1038	.1185	.0000
RM No Stgo.	1,0000	.0493	.0014	.0435	.0000	.0000	.0000	.0000

Tabla K- 2: Centros de los segmentos para variables demográficas
Fuente: Elaboración propia