



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DISEÑO DE UN MODELO PARA LA ACTIVACIÓN DE CLIENTES EN LOS RUBROS DE LA TARJETA DE CRÉDITO DE UN RETAIL FINANCIERO

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

DIEGO EXEQUIEL LÓPEZ HERNÁNDEZ

PROFESOR GUÍA:

ANDRÉS MUSALEM SAID

MIEMBROS DE LA COMISIÓN:

LUIS ABURTO LAFOURCADE

JUAN ROMERO GODOY

SANTIAGO DE CHILE

2018

DISEÑO DE UN MODELO PARA LA ACTIVACIÓN DE CLIENTES EN LOS
RUBROS DE LA TARJETA DE CRÉDITO DE UN RETAIL FINANCIERO

El presente trabajo de título se lleva a cabo en una empresa de Retail Financiero nacional que se caracteriza por ser la administradora de la tarjeta de crédito de un reconocido grupo a nivel latinoamericano. La estrategia de la empresa es conseguir la principalidad, siendo una manera de medirlo la cantidad de necesidades que cubren los clientes con su tarjeta, a través de calcular los rubros distintos en que es utilizada. Es por ello, que el objetivo general de este proyecto es diseñar un modelo de asignación de ofertas para la activación de los clientes de un Retail Financiero en los distintos rubros de gestión de su tarjeta de crédito, balanceando las preferencias del cliente y las ganancias para la empresa.

Con respecto al desarrollo de la solución, se plantean tres objetivos específicos: En primer lugar, determinar el efecto causal que posee sobre el gasto y la rentabilidad (de las compras con la tarjeta) del cliente el hecho de que se active en un nuevo rubro, a través de una técnica de *matching*. Concluyéndose, que existe un efecto positivo y significativo de activarse en 15 de los 16 rubros sobre el gasto y la rentabilidad del uso de la tarjeta. De estos, se destaca Viajes, que genera el mayor aumento del gasto con \$78.000 mensuales (un 117% más que el promedio de los rubros) y un aumento promedio de la rentabilidad de las compras en \$1.100 mensuales (el doble que el promedio de los rubros).

El segundo objetivo, es diseñar un modelo para predecir el próximo rubro en el cual se activará cada cliente, según sus características y comportamiento pasado. El modelo escogido fue un probit multivariado, que estipula la definición de un orden general de ofrecimiento de los rubros debido a su probabilidad de compra y efecto sobre la compra de otros rubros. Luego, el tercer objetivo, es establecer un criterio de asignación de rubros a ofertar a cada cliente que aumente el gasto de estos, recomendándose a la empresa la utilización del criterio del lift, debido a que el gasto percibido aumenta en un 18% por sobre la utilización de la clásica asignación por mayor propensión. También, de los primeros objetivos, se logra identificar que el 14% de la cartera tiene alta probabilidad de activarse en al menos un rubro, lo que repercutiría en un aumento de su gasto en un 20%. Por lo anterior, se recomienda a la empresa potenciar a través de ofertas los rubros en que los clientes no se activan naturalmente, pero generan una diferencia grande en el gasto de los clientes, como Educación y Recaudación.

Finalmente, se estipulan las principales limitaciones del proyecto, que dan paso a dos líneas de trabajos propuestos: el estudio de la activación de rubros en clientes nuevos, debido a que estos fueron eliminados del análisis, y la incorporación del efecto del marketing directo sobre las probabilidades de compra.

TABLA DE CONTENIDO

1.	Antecedentes generales	1
1.1	La empresa	1
1.1.1	Historia.....	1
1.1.2	Declaraciones estratégicas	1
1.1.3	Modelo de negocios y estructura organizacional	2
1.1.4	Tamaño de la empresa	3
1.2	Mercado	3
1.2.1	Retail en Chile.....	3
1.2.2	Retail Financiero	4
1.3	Desempeño organizacional	5
2.	Descripción del proyecto	6
2.1	Áreas involucradas.....	6
2.2	Identificación de la oportunidad.....	6
2.2.1	Crecimiento de la cartera	7
2.2.2	Principalidad	7
2.2.3	Gasto y rentabilidad de la cartera	10
2.2.4	Tamaño de la oportunidad	12
3.	Objetivos.....	13
3.1	Objetivo general	13
3.2	Objetivos específicos.....	13
4.	Marco Conceptual.....	14
4.1	Mejor próxima oferta.....	14
4.1.1	Logit	15
4.1.2	Probit multivariado	15
4.1.3	Modelamiento jerárquico bayesiano.....	17
4.2	Inferencia causal	17
4.2.1	Métodos de emparejamiento.....	18
4.3	Modelos de respuesta	21
4.4	Evaluación de desempeño	22
4.4.1	Matriz de confusión	22
4.4.2	Análisis COR, AUC y Ganancia	23

5.	Metodología	24
5.1	Efecto activación de los rubros.....	25
5.1.1	Estimación del <i>propensity score</i>	25
5.1.2	Algoritmo de emparejamiento y soporte común	26
5.2	Mejor próximo rubro	27
5.2.1	Modelo logit.....	27
5.2.2	Modelo probit multivariado	28
5.3	Asignación de ofertas	29
5.4	Análisis de campañas.....	30
6.	Alcances	31
7.	Comprensión de los datos y análisis descriptivo	32
7.1	Información sociodemográfica y del negocio.....	32
7.2	Transacciones y rubros de gestión.....	33
7.3	Rentabilidad	36
8.	Preparación de los datos	38
9.	Efecto de los rubros sobre la rentabilidad.....	40
9.1	Cálculo del propensity score	41
9.2	Emparejamiento y soporte común.....	42
9.3	Calidad del emparejamiento y estimación del efecto	44
10.	Mejor próximo rubro.....	48
10.1	Modelo logit.....	48
10.2	Modelo probit multivariado	51
10.2.1	Modelo de efectos específicos	51
10.2.2	Modelo de ordenamiento	53
10.2.3	Elección de modelo predictivo	57
11.	Criterios de asignación y usos posibles.....	58
12.	Conclusiones	63
13.	Limitaciones y trabajos propuestos.....	65
14.	Bibliografía.....	66
15.	Anexos.....	69

TABLA DE FIGURAS

Figura 1: Estructura gerencial de la empresa.	2
Figura 2: Evolución histórica de la utilidad de la empresa.	3
Figura 3: Porcentaje de tarjetas de crédito vigentes en el mercado por emisor.	4
Figura 4: Colocaciones de las principales emisoras de tarjetas de crédito no bancarias.	5
Figura 5: Evolución histórica de la cantidad de cuentas abiertas de la empresa.	7
Figura 6: Histograma de la cantidad de rubros en los que utilizan la tarjeta de crédito los clientes.	8
Figura 7: Porcentaje de venta y porcentaje de venta acumulada durante el año 2016 de los rubros de gestión.	9
Figura 8: Gasto promedio de los clientes según la cantidad de rubros en los que compra.	10
Figura 9: Rentabilidad promedio de los clientes según la diversidad de rubros en que compran, respecto del grupo con mayor rentabilidad.	11
Figura 10: Distribución de los clientes de la empresa según su diversidad de rubros y gasto con la tarjeta de crédito durante el año 2016.	12
Figura 11: Ejemplo matriz de confusión binaria.	22
Figura 12: Ejemplo de curvas de ganancia y comparación con selección aleatoria.	23
Figura 13: Ilustración de los pasos del ciclo de la metodología CRISP-DM.	24
Figura 14: Metodología para la estimación de efectos causales basadas en propensity score.	25
Figura 15: Distribución de la cartera de clientes según su edad y tipo de tarjeta.	33
Figura 16: Distribución mensual de los clientes con compra y su gasto con la tarjeta de crédito.	34
Figura 17: Cantidad de clientes que compran y su gasto total por rubro durante el periodo de observación.	35
Figura 18: Evolución del gasto total por rubro de gestión con la tarjeta de crédito.	36
Figura 19: Rentabilidad anual promedio de los clientes que compraron en cada rubro de gestión durante el 2016.	37
Figura 20: Distribución de la cantidad de transacciones y gasto de los clientes para detección de valores extremos.	39
Figura 21: Línea de tiempo de los periodos definidos para la aplicación del método de emparejamiento.	40
Figura 22: Cantidad de clientes en la base analítica de cada rubro y porcentaje de activación en mes de estudio.	42
Figura 23: Verificación gráfica del soporte común de propensity score para los clientes que se activan en farmacias y los que no.	43

Figura 24: Variación de la cantidad de rubros para las cuáles están balanceadas el propensity score y gasto pasado según cantidad de vecinos utilizados.	44
Figura 25: Verificación gráfica del balance de la compra pasada en salud y el gasto pre-activación a distintas probabilidades de activación en farmacias.....	45
Figura 26: Estimación final del efecto de activarse en el rubro farmacias sobre el gasto del cliente.....	46
Figura 27: Efecto estimado de la activación de cada rubro de gestión sobre el gasto y rentabilidad de los clientes.	47
Figura 28: Capacidad predictiva de los modelos logits estimados para cada rubro en la muestra de testeo.....	48
Figura 29: Comparación de curvas de ganancia de los modelos de efectos específicos sobre las activaciones.	52
Figura 30: Comparación de curvas de ganancia de los modelos de ordenamiento sobre las activaciones.	55
Figura 31: Escala de adquisición de rubros estimado según el modelo probit multivariado de ordenamiento.	56
Figura 32: Comparación de las curvas de ganancia de los modelos predictivos seleccionados.....	57
Figura 33: Evaluación de ganancia sobre las activaciones en el rubro ofrecido según regla de selección.....	59

ÍNDICE DE TABLAS

Tabla 1: Descripción de las variables relevantes a nivel de clientes.	32
Tabla 2: Campos relevantes de la base de datos de transacciones y su descripción.....	33
Tabla 3: Jerarquía de rubros, sub rubros y categorías para el rubro de gestión Automotriz.	34
Tabla 4: Gasto y transacciones de los clientes de la base analítica entre julio de 2016 y junio de 2017.	38
Tabla 5: Gasto y transacciones durante el periodo de estudio de los clientes eliminados y de los que se mantienen en la base.	39
Tabla 6: Verificación del balance de las covariables relevantes en la base emparejada para el rubro farmacias.....	45
Tabla 7: Resultados estimación del efecto causal de la activación en farmacias sobre el gasto del cliente.	46
Tabla 8: Evaluación del WAIC y comparación de las distintas especificaciones del modelo logit de ordenamiento.	55
Tabla 9: Composición de las campañas vía e-mail para la activación de los comercios asociados del rubro Restaurant.	61

1. ANTECEDENTES GENERALES

El presente trabajo de memoria se desarrolla en una empresa de Retail Financiero (desde ahora “la empresa”), por lo que en esta sección se presentará una breve caracterización de la organización en conjunto a la descripción del mercado en que se desarrolla.

1.1 LA EMPRESA

En los siguientes subcapítulos se observan las principales características de la empresa en que se desarrolla el proyecto.

1.1.1 Historia

El origen de la empresa se remonta a fines del siglo XIX, como una pequeña sastrería en Santiago de Chile. Luego, la tienda se fortalecería al introducir nuevos productos relacionados con el vestuario y el hogar, hasta transformarse en una tienda por departamentos. Es precisamente en este desarrollo, que en 1980 implementan su propia tarjeta de crédito, que en el año 1981 daría forma a la empresa en que se desarrolla el presente trabajo. Posteriormente, se agregarían distintas líneas de negocio hasta convertirse en uno de los grupos más grandes de Latinoamérica.

En particular, una de las líneas desarrolladas fueron los Servicios Financieros que aportaron más del 50% de las ganancias del grupo en el año 2016. Es aquí donde la organización que maneja la tarjeta de crédito se sitúa como pilar fundamental de la organización, jugando un rol de conector entre los distintos negocios a través de la propia tarjeta de crédito y su programa de fidelización.

1.1.2 Declaraciones estratégicas

Con respecto a las declaraciones estratégicas, la empresa define como misión *hacer posible las aspiraciones de las personas, mejorar su calidad de vida y superar sus expectativas a través de una oferta integrada de servicios, potenciada por los beneficios del grupo. A su vez, estipula que su visión es ser la compañía preferida por las personas, generando relaciones de largo plazo, a partir de: Ser líderes por su transparencia, simplicidad y conveniencia; Atraer, desarrollar y motivar un equipo de excelencia, comprometido, colaborativo y apasionado por los clientes; finalmente, ser valorados por su aporte a las comunidades donde trabajan.*

1.1.3 Modelo de negocios y estructura organizacional

Como toda tarjeta de crédito, el modelo de negocios básico que la sustenta es el brindar financiamiento con su consecuente cobro de intereses y comisiones por el uso de su servicio. En este sentido, la empresa es una fuente de financiamiento de carácter individual y no institucional, por lo que su principal público objetivo son las personas individuales entre 21 y 65 años, sin poseer una mayor discriminación, a pesar de que la cartera se compone en un 53% por personas del grupo socioeconómico D, de un total de aproximadamente 3 millones de clientes activos que posee la empresa (en el Anexo A se puede apreciar la distribución de la cartera de clientes según grupo sociodemográfico).

Con respecto a las formas de financiamiento, se pueden considerar dos escenarios: en primer lugar, la simple utilización de la tarjeta de crédito con sus inherentes intereses y comisiones; o, en segundo lugar, la prestación de servicios financieros como lo son los créditos de consumo.

Además de lo que respecta directamente al otorgamiento de financiamiento, la empresa posee otras líneas de productos que complementan la utilización de su tarjeta, por ejemplo, la asignación de seguros asociados a la tarjeta, pago automático de cuentas y ofrecimiento de otros productos del grupo. El objetivo de todo este ecosistema es el posicionar a su tarjeta como el principal medio de pago de sus clientes, denominándole a este concepto principalidad.

Acompañando al modelo principal, existen dos importantes herramientas que otorgan una gran ventaja competitiva: el club de fidelización (basado en el típico sistema de puntos) y las alianzas comerciales. Asimismo, otra característica diferenciadora es su posicionamiento, alcanzando, por ejemplo, la mayor recordación espontánea entre las tarjetas de crédito de casas comerciales, con un 74% en 2016. Para llevar a cabo este modelo, la estructura que utiliza la organización es evidentemente funcional, presidida por un Gerente General bajo el cual se encuentra la línea gerencial como se puede apreciar en la Figura 1.



Figura 1: Estructura gerencial de la empresa.
Fuente: Elaboración propia.

1.1.4 Tamaño de la empresa

En cuanto a su tamaño, en la actualidad la empresa posee más de 2.100 trabajadores, situándose como una gran empresa. La mayoría de sus colaboradores son mujeres, con un 76% del total, además es una empresa de carácter joven en que el promedio de edad de sus trabajadores es de 36 años.

Con respecto a los canales que utiliza para la comunicación con sus clientes, destacan los de carácter no presencial como su Call Center, aplicación móvil y página web, mientras que su principal canal físico son sus 100 sucursales, con 60 de ellas en la Región Metropolitana.

En el ámbito financiero, a diciembre de 2016, la utilidad de la empresa alcanzó MM\$122.678, con un crecimiento del 7,3% con respecto al año 2015, como se puede ver en la Figura 2.

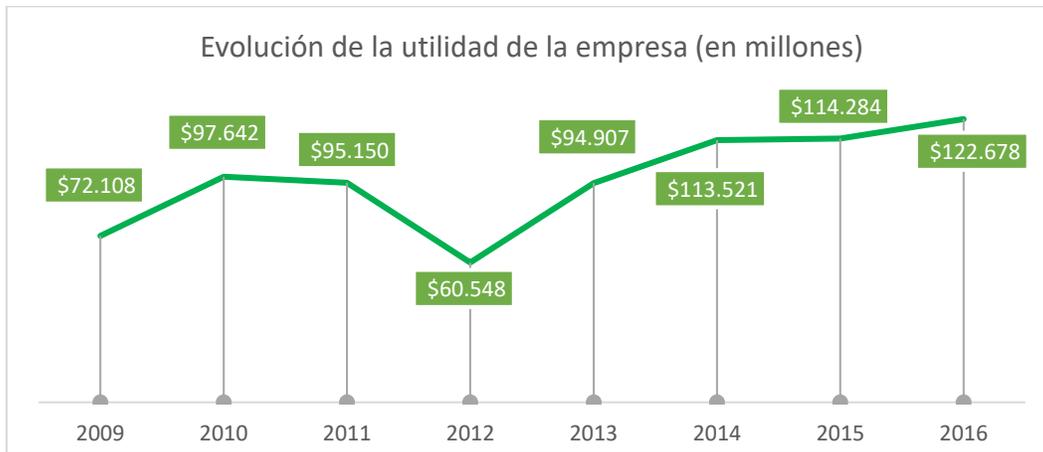


Figura 2: Evolución histórica de la utilidad de la empresa.
Fuente: Elaboración propia en base a memoria anual 2016 de la empresa.

1.2 MERCADO

La empresa se inserta en el mundo del Retail Financiero que se deriva del negocio original del Retail, por ello en el siguiente capítulo se estudian estos mercados.

1.2.1 Retail en Chile

El Retail o comercio al por menor, es una industria bastante arraigada en Chile, presentando un crecimiento sostenido a través de los años. Al mes de marzo del 2017, la actividad del comercio al por menor registró un crecimiento real de 4,9% durante el último año, igualado al promedio del comercio que creció en un 4,9% en el mismo periodo, según el Instituto Nacional de Estadísticas (INE, 2017).

1.2.2 Retail Financiero

El Retail Financiero hace referencia a una parte del negocio financiero que no es manejada por entes bancarios, sino que por empresas de Retail. Esta asociación entre el Retail y el negocio financiero se realiza fundamentalmente a través de las tarjetas de crédito de casas comerciales, dentro de las cuáles la empresa y la tienda por departamentos del grupo fueron pioneras en el país.

Es por lo anterior, que la competencia de la empresa es híbrida, teniendo por una parte las tarjetas de crédito de casas comerciales como Presto, La Polar o ABCDIN y, también, las tarjetas de crédito de bancos como las del Banco de Chile, Banco Santander, entre otros, también, se pueden agregar como competencia indirecta otras instituciones que se dedican al financiamiento de las personas naturales como la Casa de Crédito Prendario o las cooperativas de ahorro y crédito. En la actualidad la organización en cuestión es el segundo emisor con más tarjetas de crédito bajo Banco Ripley (ver Figura 3), con un total de MM 2,8 de tarjetas activas y con operaciones durante el último año.

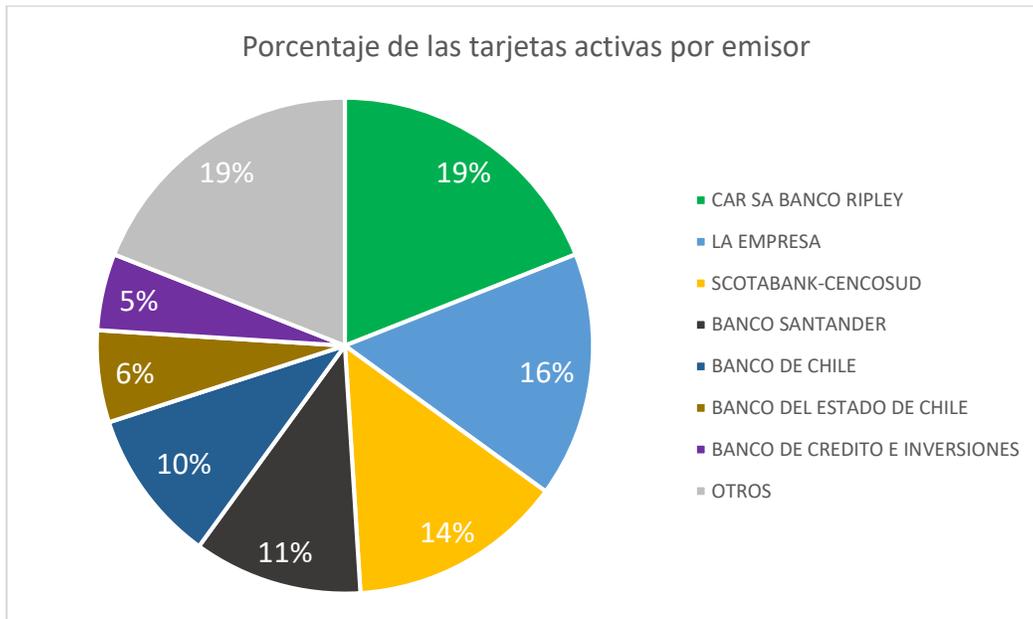


Figura 3: Porcentaje de tarjetas de crédito vigentes en el mercado por emisor.
Fuente: Superintendencia de Bancos e Instituciones Financiera (SBIF, 2016).

Adentrándose específicamente en el mercado de las tarjetas de crédito pertenecientes a emisores no bancarios, la empresa se ubica en el primer lugar en colocaciones con MM\$1.422.884, siendo su principal competidor la tarjeta PRESTO con MM\$467.747 de colocaciones. En la Figura 4, se pueden apreciar las colocaciones de las principales emisoras de tarjetas de crédito no bancarias.

Un fuerte cambio que vivió la industria del Retail Financiero en Chile durante el último tiempo fue la bancarización de dos grandes exponentes, la tarjeta CENCOSUD y la tarjeta Ripley, por lo que el número total de tarjetas de crédito administradas por empresas de Retail ha caído duramente de 5,9 millones en 2012 a 4 millones en 2016. A pesar de ello, las colocaciones totales del Retail Financiero han aumentado los últimos tres trimestres a tasas mayores al 10% según el Comité de Retail Financiero (Comité de Retail Financiero, 2016).

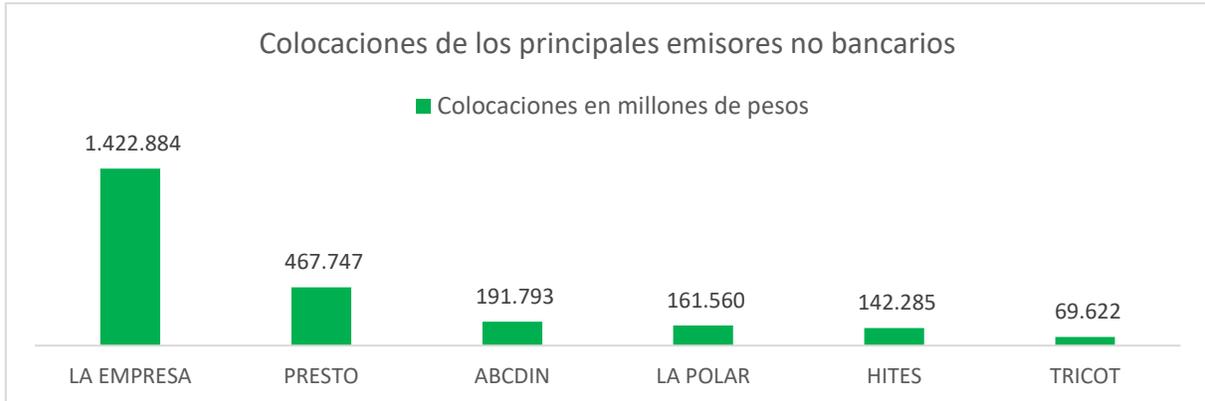


Figura 4: Colocaciones de las principales emisoras de tarjetas de crédito no bancarias.
Fuente: Superintendencia de Bancos e Instituciones Financiera (SBIF, 2016).

Asimismo, la misma entidad reporta que la deuda promedio de los clientes subió en un 12,9% durante el último año, llegando a un promedio de \$508.000, otorgándole gran importancia a la inclusión de las tarjetas de tipo Visa y MasterCard que ampliaron el rango de utilización de las antiguas tarjetas cerradas de las casas comerciales.

1.3 DESEMPEÑO ORGANIZACIONAL

Al igual que la industria del Retail Financiero en general, la organización ha continuado creciendo en lo que respecta a cantidad de colocaciones a tasas del 9% en el periodo 2010 a 2016. Del mismo modo, también han aumentado las cuentas activas desde 2,4 millones en el año 2010 hasta 2,9 millones a fines del año 2016.

Otro punto relevante, ha sido el cambio en la composición de las tarjetas de crédito de la empresa, puesto que en el año 2010 sólo un tercio de las tarjetas permitía la compra en negocios fuera del grupo, mientras que a fines del año 2016 el porcentaje de este tipo de tarjetas aumentó hasta un 68%, lo que repercute en un cambio total de la estrategia de la empresa, ahora enfocada en ser el medio de pago principal de sus clientes.

2. DESCRIPCIÓN DEL PROYECTO

En este capítulo se darán a conocer las características del área de trabajo específica en la que se desarrolla el proyecto, además de la oportunidad que lo sustenta.

2.1 ÁREAS INVOLUCRADAS

Una de las grandes ventajas de la empresa en que se realiza el trabajo es el ser la tarjeta de crédito que une a su grupo, por lo que como empresa posee un acceso a una gran cantidad de información transaccional, económica y sociodemográfica de los clientes. Esta condición, por una parte, hace difícil el manejo de tal cantidad de datos, pero, por contrapartida abre la posibilidad de extraer conocimiento sobre los clientes. En este ámbito, la Gerencia de Inteligencia de Clientes tiene el objetivo de permear a la empresa de una visión integral de los clientes, que permita a las áreas estratégicas y de gestión el poder tomar mejores decisiones.

Para la realización de este objetivo, la Gerencia de Inteligencia de Clientes, se divide en tres subgerencias: Data Mining y Reportería, Gestión de Clientes y Research.

En específico, el trabajo de título se realizará en la Subgerencia de Data Mining, que dentro del objetivo general de la gerencia de Inteligencia de Clientes tiene el rol fundamental de la entrega de información para la toma de decisiones de la empresa y el conocimiento de los clientes, basados principalmente en el análisis de datos y fuentes cuantitativas, siendo entonces los beneficiados por el trabajo de esta subgerencia la totalidad de áreas que necesiten permearse de información y análisis de los clientes, en especial, dos áreas que tiene contacto directo con el cliente: la Subgerencia de Gestión de Clientes y la Gerencia Comercial.

2.2 IDENTIFICACIÓN DE LA OPORTUNIDAD

El proyecto, se basa en la identificación de los mejores rubros en los cuáles se debería activar a los clientes que permitan aumentar el gasto y la rentabilidad de estos, de manera de permitir una mejor gestión del crecimiento de la cartera.

2.2.1 Crecimiento de la cartera

A pesar del crecimiento histórico de la cantidad de tarjetas, este no se ha mantenido en el último tramo, sobre todo en la cantidad de cuentas abiertas¹ apreciable en la Figura 5.

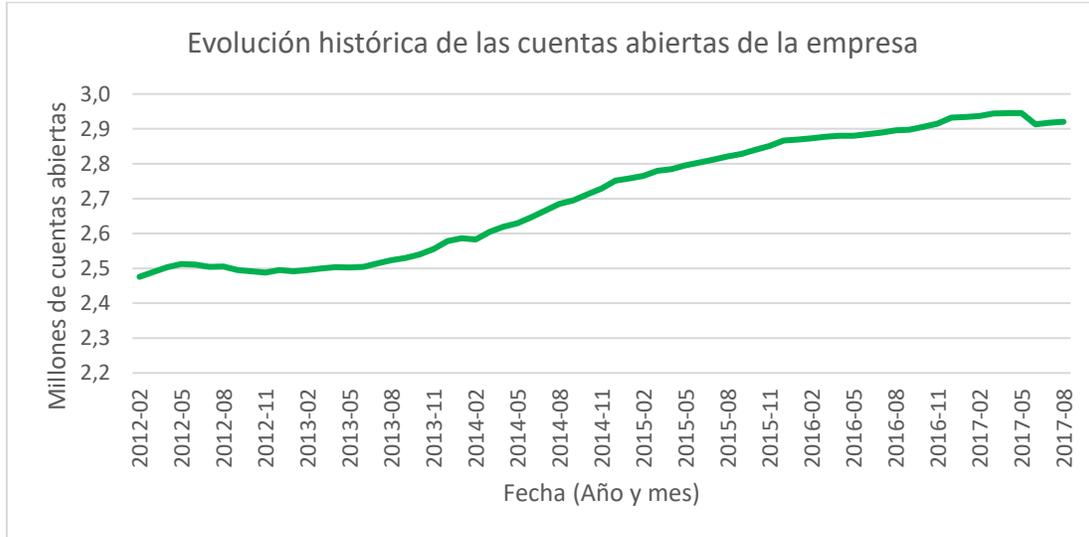


Figura 5: Evolución histórica de la cantidad de cuentas abiertas de la empresa.
Fuente: Elaboración propia.

En los últimos años, la empresa había apalancado su crecimiento económico en su creciente cartera de clientes, que se elevó hasta tasas del 6% durante el año 2014. Pero, a partir de abril del año 2016 se logra apreciar que la cantidad total de cliente sólo ha aumentado en un 1%, poniendo en problemas la estrategia de crecimiento a través del aumento de tarjetahabientes. Por ello, se inicia una nueva estrategia con enfoque en la capitalización del potencial de los clientes actuales de la compañía a través de la principalidad.

2.2.2 Principalidad

La principalidad se refiere al hecho de ser el principal medio de pago de los clientes, es decir, situarse como la primera opción de estos al momento de realizar cualquier transacción. En particular, para la empresa, la principalidad es medida en tres aspectos: la diversidad del uso de la tarjeta como medida de las necesidades que cubre, la tenencia de productos de la empresa y la cantidad de transacciones, siendo punto principal de la estrategia el aumento de la diversidad.

¹ Una cuenta abierta se refiere al contrato de un cliente que en la actualidad posee la capacidad de realizar transacciones con su tarjeta inmediatamente o previo pago de su deuda, debido a que esta se encuentra con una mora no superior a 90 días.

En este aspecto, la medida que utiliza la empresa para medir la diversidad del uso de la tarjeta de sus clientes es la cantidad de rubros de gestión² diferentes en los que estos la utilizan (la lista completa de rubros de gestión se puede ver en el Anexo B), siendo claramente una mayor cantidad de rubros lo deseable para la empresa. Pero, a pesar de la posibilidad de realizar compras en 16 rubros distintos, la moda de rubros en que compraron los clientes fue de sólo 2 durante el 2016, como se aprecia en la Figura 6.

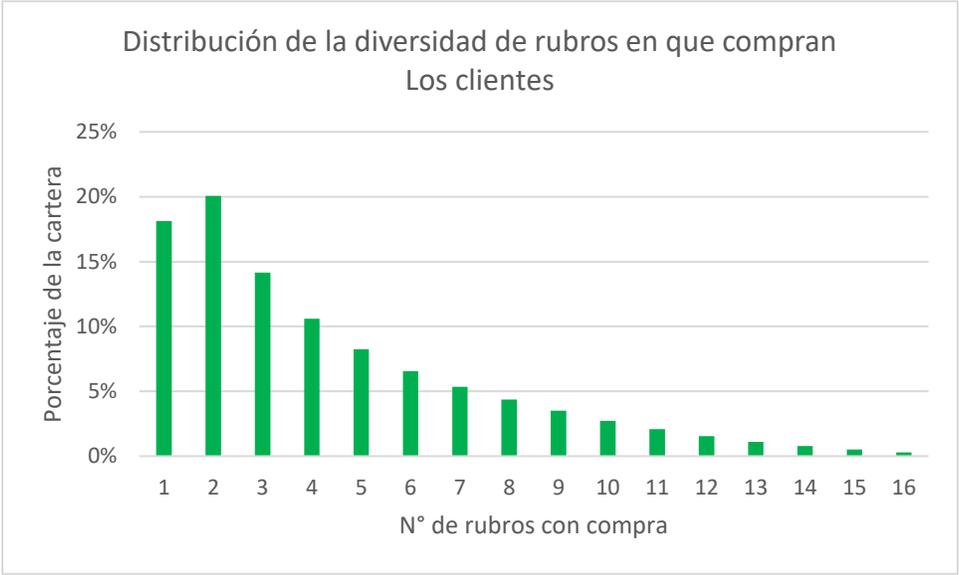


Figura 6: Histograma de la cantidad de rubros en los que utilizan la tarjeta de crédito los clientes.
Fuente: Elaboración propia.

Añadiéndose a esta situación no deseada por la empresa, los montos de uso de la tarjeta de crédito se concentran también en una baja variedad de rubros. En la Figura 7, se aprecia que los cuatro rubros de gestión con mayor venta acumulan más del 70% de esta, conjeturándose que, a pesar de la posibilidad de utilizar la tarjeta en una amplia gama de comercios, los clientes no aprecian su utilización en muchos de estos, lo que atenta contra el objetivo de la principalidad.

Es así, que se sustenta la oportunidad en la estrategia corporativa que apunta a situar la tarjeta de crédito como principal de los clientes, estableciéndose como una de las líneas de acción de la empresa la activación de los tarjetahabientes en rubros de gestión en los que actualmente no compran.

² Un rubro de gestión es una unificación de distintos comercios en los que es utilizada la tarjeta de crédito de la empresa, en base a su similitud en el negocio al que pertenecen, forma de pago u otros. Esta agrupación se realiza en 19 grupos observables en el Anexo B (en el análisis se unifican los de menor relevancia como Otros, en el desarrollo del informe se detalla este hecho).

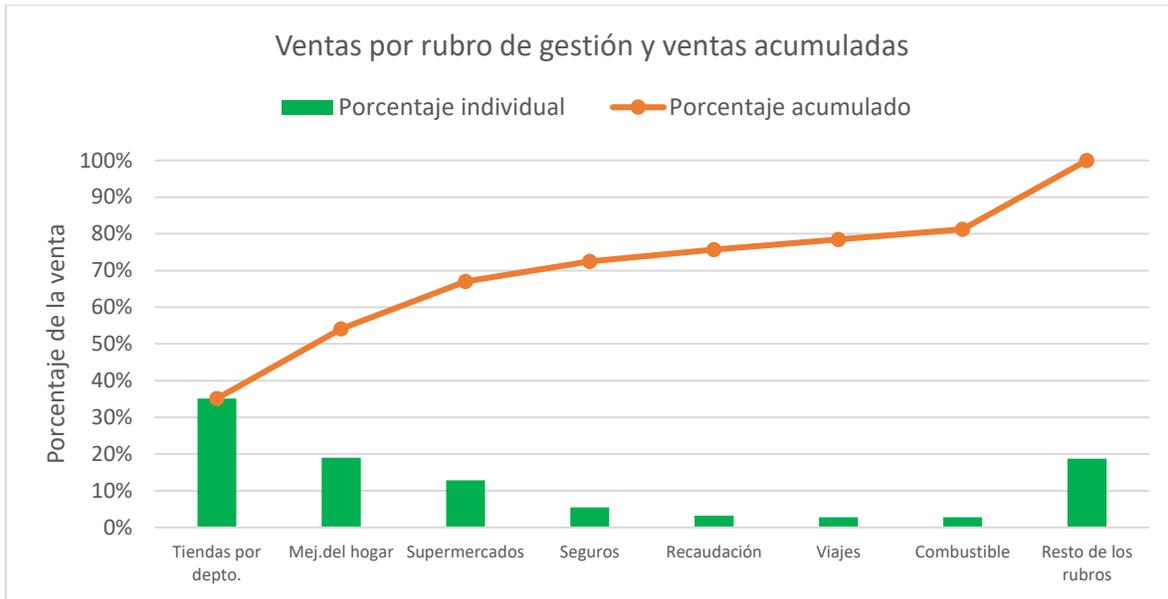


Figura 7: Porcentaje de venta y porcentaje de venta acumulada durante el año 2016 de los rubros de gestión.
Fuente: Elaboración propia.

Por lo anterior, en búsqueda de la principalidad, el área de Gestión de Clientes como centralizadores de la comunicación con el cliente por parte de la gerencia de Inteligencia de Clientes ha optado por la realización de campañas masivas. Con esto, han logrado comunicar la posibilidad del uso de la tarjeta en distintos rubros, pero también se provoca una pérdida de eficiencia al destinar recursos de las campañas a clientes que tienen muy pocas posibilidades de responder de manera positiva a estas. Para la solución de este inconveniente, se ha optado en la gerencia por establecer una selección de los clientes que tengan más posibilidades de utilizar la campaña recibida, pero hasta el momento esta selección es dada básicamente por el conocimiento experto de los trabajadores del área o por algunas hipótesis lógicas³ que se pudiesen definir y el uso de métodos o herramientas más complejas ha sido acotado a ciertos rubros o campañas específicas.

³ Un ejemplo de este tipo de hipótesis es suponer que para que un cliente compre en el rubro de gestión Combustibles debería poseer un automóvil.

2.2.3 Gasto y rentabilidad de la cartera

El aspecto principal de cualquier negocio es la generación de ganancias, que en el caso de la empresa vienen principalmente de tres fuentes: ingresos por intereses y comisiones, servicios financieros y *merchant*⁴.

En lo que refiere al gasto de los clientes, este sigue una relación lógica con respecto a la diversidad con la cual el cliente usa su tarjeta, teniéndose que en promedio mientras en más rubros de gestión haya utilizado la tarjeta el cliente mayor es su gasto, como se aprecia en la Figura 8.

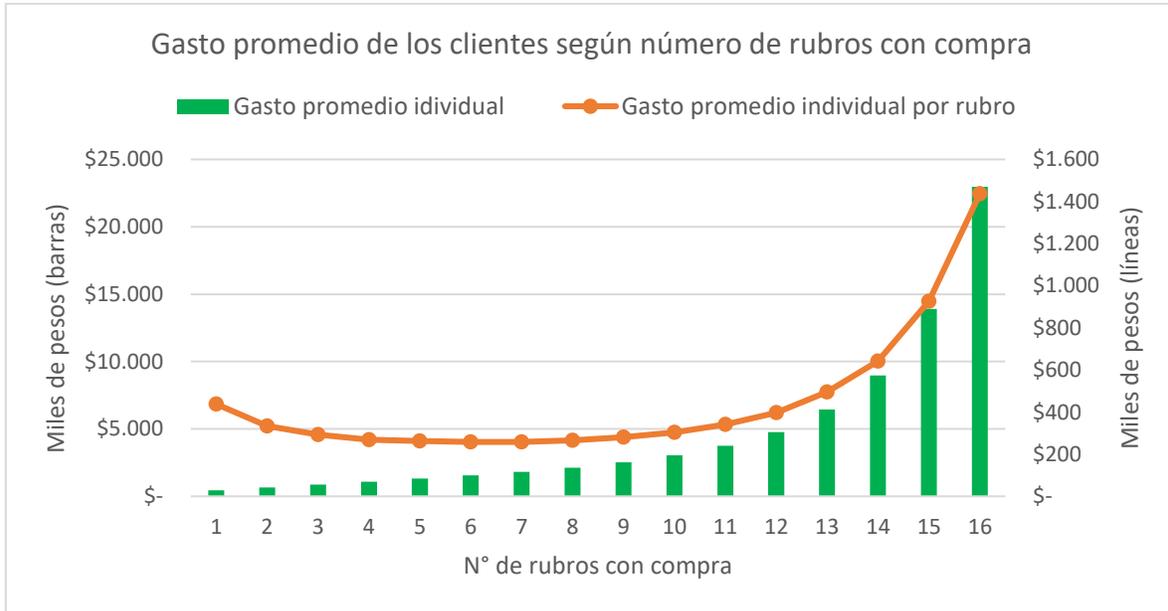


Figura 8: Gasto promedio de los clientes según la cantidad de rubros en los que compra.
Fuente: Elaboración propia.

Asimismo, se puede apreciar que el gasto promedio de los clientes por rubro se comporta de manera levemente decreciente hasta una diversidad de 6 rubros, pero a partir de 8 rubros sigue una tendencia creciente fuerte. Esto, expresa que en los clientes con baja diversidad un nuevo rubro agrega menos gasto en promedio que los que ya tenía, mientras que para los con alta diversidad hay evidencia para decir que cada rubro agregado por el cliente aumenta de mayor manera su gasto que los que ya compraba, lo que recalca las grandes oportunidades de ganancias de llevar a los clientes hasta una alta cantidad de rubros con compra.

⁴ Corresponde a los ingresos recibidos como un porcentaje del monto gastado por los clientes con su tarjeta de crédito en los distintos comercios, cabe señalar que este porcentaje no es igual en todos los comercios.

En cambio, a nivel individual no existe una relación directa entre los clientes con mayor diversidad de uso de su tarjeta y la rentabilidad que le otorgan a la empresa. En efecto, en la Figura 9 se puede apreciar como la mayor rentabilidad promedio se observa en los clientes con una diversidad de 13 rubros. A su vez, el caso más especial sucede con los clientes más principales en términos de diversidad, con 16 rubros activos, que presentan en promedio sólo un 18% de la rentabilidad que entrega el grupo más rentable.

Es por ello, que se hace aún más indispensable, la correcta elección de los rubros que deben ser ofertados a los distintos clientes. Por lo tanto, este proyecto busca la definición del próximo mejor rubro a ofertar, buscando establecer un orden o ranking de activación de estos, para los distintos clientes con el objetivo de aumentar su gasto y una parte de la rentabilidad.

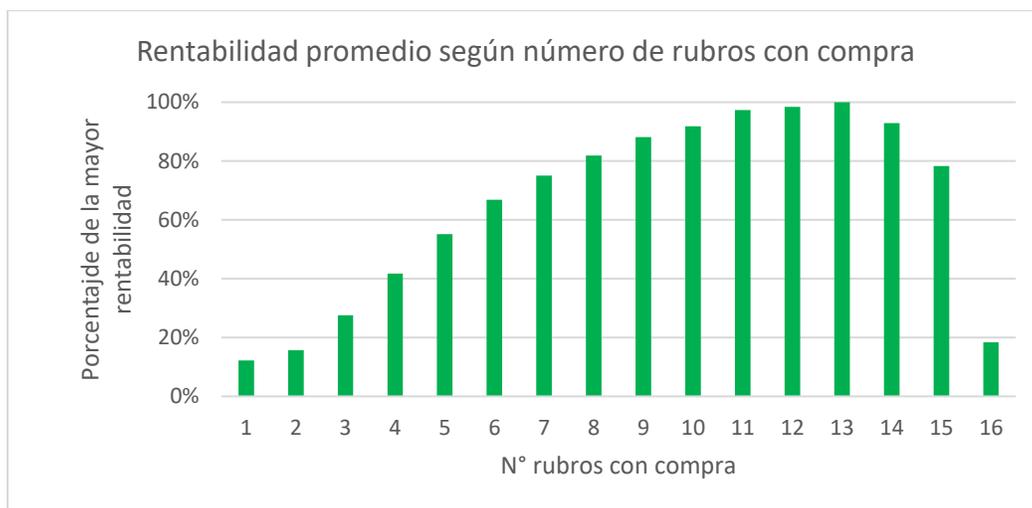


Figura 9: Rentabilidad promedio de los clientes según la diversidad de rubros en que compran, respecto del grupo con mayor rentabilidad.
Fuente: Elaboración propia.

Respecto a ello, la mayor hipótesis que sustenta la solución propuesta es el hecho de que la realización de un ofrecimiento en el rubro indicado mediante marketing directo (principalmente e-mail) basado en las características del cliente y sus similares, aumenta la probabilidad de que este cliente se active en este.

A su vez, dadas las características del negocio, otra hipótesis importante es que la compra de los clientes en un nuevo rubro no canibaliza sus compras en los demás, sino que sólo marca un cambio en su medio de pago. Por último, se piensa que existirá una relación entre los rubros en los que el cliente se encuentra activo y el gasto y la rentabilidad que el cliente presenta, debido a las características propias que pueda poseer el rubro o las personas que acceden a estos.

2.2.4 Tamaño de la oportunidad

Es importante también, definir los posibles beneficios de aumentar la diversidad del uso de la tarjeta de los clientes. Particularmente, una de las razones que sustentan el proyecto es que los clientes con una mayor diversidad poseen mayores gastos lo que se corroboró en la Figura 9. En particular, el promedio de diferencia en gasto entre clientes con diferencia de un rubro de diversidad es de \$374.614.

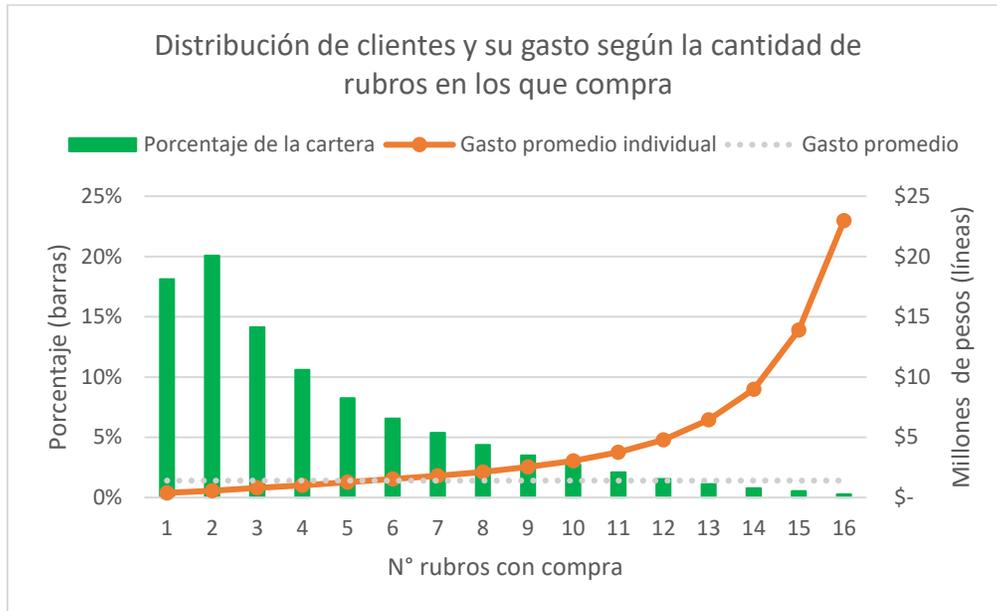


Figura 10: Distribución de los clientes de la empresa según su diversidad de rubros y gasto con la tarjeta de crédito durante el año 2016.
Fuente: Elaboración propia.

Asimismo, se aprecia en la Figura 10, que como es de esperar la cantidad de clientes es menor a medida que se sube en la diversidad, lo que deja un gran espacio de crecimiento para desarrollar la cartera.

Es así como, realizando un análisis *ceteris paribus*, en el caso de que se logre con este proyecto que el 1% de la cartera de clientes aumente su diversidad en un rubro y asumiendo que con esto aumentarían su gasto hasta el promedio del nivel de diversidad al cual se están cambiando, se obtiene que el gasto incremental alcanzable es de MM\$9.788 anualmente, cercano al 0,3% del gasto anual de la cartera durante el último año. A pesar de ello, como se plantea en el problema, este aumento de gasto no se transfiera de forma directa en ingresos para la empresa, ya que realizando un análisis similar al del gasto, el aumento incremental de rentabilidad sería sólo de un 0,08%.

3. OBJETIVOS

Dado el contexto en el que se desempeña la empresa y la oportunidad descrita con anterioridad, a continuación, se plantean los objetivos de este trabajo.

3.1 OBJETIVO GENERAL

Diseñar un modelo de asignación de ofertas para la activación de los clientes de un Retail Financiero en los distintos rubros de gestión de su tarjeta de crédito, balanceando las preferencias del cliente y las ganancias para la empresa.

3.2 OBJETIVOS ESPECÍFICOS

1. Estimar el efecto causal individual que posee sobre el gasto y la rentabilidad (de las compras) de un cliente su activación en los distintos rubros de gestión.
2. Diseñar un modelo para predecir el próximo rubro en el cual se activará cada cliente según sus características y comportamiento pasado.
3. Establecer un criterio de asignación de rubros a ofertar a cada cliente para aumentar el gasto de estos, basado en sus probabilidades de activación y la ganancia que le significa a la empresa.

4. MARCO CONCEPTUAL

Como es recurrente en un proyecto, se ven intervenidas bastantes disciplinas para su ejecución. Por ello, en el siguiente capítulo se plantea el marco conceptual que sustenta la elaboración de esta memoria de título.

4.1 MEJOR PRÓXIMA OFERTA

Existen variados métodos para permitir la selección de una próxima o las próximas mejores ofertas que se le pueden realizar a un cliente, en particular, se acogen las denominadas *Next Best Offer* (NBO).

El modelo general utilizado para abordar la selección de la mejor próxima oferta realizable a un cliente es el *Next Best Offer* visto en (Knott, Hayes, & Neslin, 2002), para ello existen variados modelos específicos que toman la decisión desde distintas aristas. En primer lugar, bastamente utilizados son los modelos de elección binaria, en particular el logit, mediante el cual se estipulan las probabilidades de los clientes a escoger determinados productos y mediante estas establecer reglas para la asignación de las ofertas.

Con mayor sofisticación, se cuenta con modelos como el de (Kamakura, Kossar, & Wedel, 2004) que permite diferenciar a los clientes según su propensión a ser innovadores en el consumo de ciertos productos y bajo esta clasificación permitir dirigir las ofertas a los clientes. También, se encuentran modelos como el de (Liu & Cai, 2007) que se basa en clasificar a los clientes según su madurez en el negocio y con ello predecir sus siguientes movimientos. El corazón de esta metodología está en la creación de estados de madurez de un cliente de manera tal de modelar estos estados y sus transiciones mediante redes bayesianas.

Por último, se puede tomar una metodología de ofrecimiento ordenado de productos. En este tipo de modelos, se busca como principal objetivo descubrir la mejor oferta para el cliente en el siguiente periodo dado un orden lógico de obtención de productos, como lo es el modelo probit multivariado visto en (Li, Sun, & Wilcox, 2005). El principal producto obtenido en estas metodologías es la determinación de una probabilidad de compra hacia los distintos productos, considerando un orden en su adquisición y la coocurrencia⁵ de compra entre estos.

⁵ Se utiliza coocurrencia en los términos especificados por (Manchanda, Ansari, & Gupta, 1999) como la existencia de correlación en la compra de dos categorías o productos en un mismo periodo u ocasión de compra.

Debido a las características del problema, se propone la utilización en primera instancia de modelos logit para cada uno de los rubros de gestión a estudiar, debido a su probada utilidad como línea base para la comparación de modelos más sofisticados. Asimismo, luego se dispondrá de un modelo probit multivariado como el utilizado por (Li, Sun, & Wilcox, 2005) principalmente por tres motivos: la capacidad de captar relaciones entre los distintos rubros en un mismo periodo, establecer un ordenamiento de estos y, por último, su flexibilidad para incorporar variables explicativas. A continuación, se muestran las especificaciones de estos modelos.

4.1.1 Logit

El modelo de regresión logística (logit) es un modelo estructural de elección discreta que nace de la necesidad de analizar variables dependientes no continuas. En particular, se busca modelar la probabilidad con que un individuo n elija la alternativa i dentro de un set de opciones, asumiendo que escogerá la opción que maximiza su utilidad (Montoya, Modelos de Elección Discreta, 2016), la que viene dada por una parte determinística v_{ni} y otra estocástica ε_{ni} como se muestra en (1).

$$u_{ni} = v_{ni} + \varepsilon_{ni} \quad (1)$$

Específicamente, el modelo logit aparece al asumir que la componente estocástica de la utilidad distribuye valor extremo tipo I o Gumbel (Montoya, Logit + Probit, 2016), con lo que se deriva la probabilidad de elección de n por el individuo i que se muestra en (2).

$$P_{ni} = \frac{e^{v_{ni}}}{\sum_j e^{v_{nj}}} \quad (2)$$

4.1.2 Probit multivariado

El probit multivariado es una ampliación del modelo probit, en el que se generaliza la elección binaria para estimar de manera conjunta un grupo de variables dependientes binarias correlacionadas entre sí (Vargas, 2003).

En el modelo probit, se asume que existe una relación entre la decisión observada del individuo i (y_i) y una respectiva contraparte no observada (y_i^*), símil a la utilidad, como se muestra en (3).

$$y_i = \begin{cases} 1 & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases} \quad (3)$$

Con respecto a la variable no observada, se supone al igual que en el modelo logit que esta depende una componente determinística (x_i) y otra aleatoria (ε_i) como se muestra en (4), quedando expresa la probabilidad de elección del individuo, dadas sus características observables como en (5).

$$y_i^* = x_i\beta + \varepsilon_i \quad (4)$$

$$\Pr(y_i = 1|x_i, \beta) = \Phi(y_i^*) = \Phi(x_i\beta + \varepsilon_i), \text{ y } \varepsilon_i \sim N(0, 1) \quad (5)$$

Dónde Φ es la función de densidad acumulada de la distribución normal estándar, x_i el vector de variables determinísticas, β el vector de parámetros a estimar y ε_i la componente no observable de la utilidad.

De esta manera, el probit multivariado es la extensión del modelo probit a un set de elecciones que se traduce en un sistema de ecuaciones como el que se observa en (6).

Lamentablemente, para el caso general en que la cantidad de elecciones es mayor a 2, no existe solución cerrada para las integrales incluidas en la función de verosimilitud, por lo que una de las opciones largamente utilizadas para flexibilizar el modelo es el análisis jerárquico bayesiano (Vargas, 2003).

$$\begin{bmatrix} y_{i1}^* \\ \vdots \\ y_{iJ}^* \end{bmatrix} = \begin{bmatrix} z_{i1}\beta_1 \\ \vdots \\ z_{iJ}\beta_J \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iJ} \end{bmatrix} \quad (6)$$

$\varepsilon_i \sim N_J(0, \Omega)$

$$y_{ij} = \begin{cases} 1 & \text{si } y_{ij}^* > 0 \\ 0 & \text{si } y_{ij}^* \leq 0 \end{cases}$$

4.1.3 Modelamiento jerárquico bayesiano

El modelamiento jerárquico bayesiano, es un tipo de modelo estadístico que permite establecer jerarquía en la dependencia de las variables que explican un determinado fenómeno que estima los distintos parámetros utilizando métodos bayesianos. En específico, utilizado sobre la estructura de un modelo probit establece distintos niveles de relaciones entre variables para predecir la realización de elecciones binarias (Manchanda, Ansari, & Gupta, 1999).

En primer lugar, se define la utilidad para el individuo i de escoger n , al igual que en los modelos anteriores, dependiendo de una componente determinística y otra estocástica, donde esta última sigue una distribución normal multivariada como se muestra en (7).

$$u_{ni} = v_{ni} + \varepsilon_{ni} \tag{7}$$
$$\varepsilon_{ni} \sim MVN(0, \Sigma)$$

En segundo lugar, de forma necesaria para la estimación de los parámetros por el método bayesiano, se debe incluir la definición de los *priors* de los parámetros definidos. Estos elementos capturan las creencias previas del analista con respecto a la distribución de las características estudiadas y sus relaciones (Manchanda, Ansari, & Gupta, 1999).

Finalmente, la estimación de los parámetros estipulados en el modelo jerárquico bayesiano, se utiliza el algoritmo de iteración de *Gibbs Sampler*. Este método, recurre a la simulación de Montecarlo a través de cadenas de Markov (MCMC) de forma cíclica que se obtiene a partir de una serie de distribuciones condicionales sobre una distribución inicial (Rossi, Allenby, & McCulloch, 2012).

4.2 INFERENCIA CAUSAL

La inferencia causal se refiere a la estimación de efectos causales, ya sea mediante experimentos o estudios. En particular, en este trabajo nos enfocaremos en el efecto causal de estar expuesto a un tratamiento (τ_i), lo que se expresa como la comparación de los resultados de un individuo i que está en tratamiento ($Y_i(1)$) versus los resultados de un individuo i no tratado ($Y_i(0)$), como se aprecia en 8.

$$\tau_i = Y_i(1) - Y_i(0) \tag{8}$$

El principal problema de la inferencia causal es que sólo podemos observar uno de los potenciales resultados para un individuo i , dado que estará o no en tratamiento. Una forma eficiente de atacar este inconveniente es la comparación de un grupo tratado y un grupo que no lo haya sido (grupo de control) tan similares como sea posible, para ello los experimentos aleatorios utilizan una asignación que asegura que las distribuciones de estos grupos sean balanceadas en sus características más relevantes antes de entrar en tratamiento o no (Stuart, 2010).

Entonces, para la estimación de los efectos del tratamiento se han elaborado distintas métricas, siendo la más relevante para el caso el ATT (del inglés average treatment effect on the treated), que se enfoca explícitamente en la diferencia en los resultados de los individuos tratados, como se aprecia en (9).

$$\tau_{ATT} = E[\tau|D = 1] = E[Y(1)|D = 1] - E[Y(0)|D = 1] \quad (9)$$

Dónde D igual a 1 expresa el hecho de que hubo tratamiento y D igual a 0 que no.

El problema se mantiene en el hecho de que el factor $E[Y(0)|D = 1]$ no es observable, por lo que se debe escoger un estimador de este. Particularmente, se asume que los resultados para ambos grupos (tratamiento y control) deberían sólo variar debido al sesgo de selección. Concluyéndose, que para que τ_{ATT} sea identificable, el efecto de no ser tratado sobre el grupo de control es un buen estimador para el resultado de no ser tratado dado que si se fue intervenido (Stuart, 2010).

Lamentablemente, la asignación aleatoria no es practicable a posteriori del tratamiento, para lo que han surgido métodos que permiten construir estos grupos de control y así poder analizar los efectos del tratamiento, siendo unos de los más utilizados los métodos de emparejamiento o *matching* que se estudian a continuación.

4.2.1 Métodos de emparejamiento

Los métodos de emparejamiento se hacen cargo de la inferencia causal básicamente en 4 etapas:

Definir cercanía: se estipula la medida que se utilizará para determinar cuándo un individuo es un buen emparejamiento o control para otro.

Implementación: se realiza el emparejamiento de individuos con su respectiva contraparte utilizando la medida de cercanía definida.

Evaluación: evaluar la calidad de los emparejamientos obtenidos, dependiendo de estos se puede volver de manera iterativa a los pasos anteriores.

Análisis: estudio de los resultados anteriores y estimación del efecto del tratamiento.

4.2.1.1 Medidas de distancia

En lo que respecta a las medidas de cercanía, la más utilizada es el denominado *propensity score* definida por (Rosenbaum & Rubin, 1983) que define la distancia entre los individuos i y j como se muestra en (10).

$$D_{ij} = |e_i - e_j| \quad (10)$$

En este caso, e_i es el *propensity score* del individuo i , que resume todas las covariables a estudiar (X) en un escalor: la probabilidad de ser tratado condicional a las características del individuo, como se aprecia en (11).

$$e_i = P[D = 1|X] \quad (11)$$

El *propensity score* posee entonces una propiedad clave, dado a que un valor definido de este la distribución de las variables independientes entre el grupo de control y el tratado son iguales, el agrupar individuos con similar puntaje se replica una selección aleatoria (por lo menos para las covariables observadas). Por esto, la diferencia en promedios de los efectos vistos entre los grupos emparejados con *propensity score* es un estimador imparcial del efecto del tratamiento (Stuart, 2010).

Otra medida de distancia comúnmente utilizadas son la distancia exacta, que define la distancia entre dos elementos como 0 cuando son iguales e infinita entre elementos distintos. También, se utiliza para casos en que se quiere calcular distancias entre variables aleatorias en más de una dimensión la distancia de Mahalanobis, siendo muy similar a la conocida distancia euclídea con la diferencia de que esta toma en cuenta la correlación entre las variables aleatorias (X) como se muestra en (12) a través de la matriz de varianzas y covarianza Σ (Salas & Portillo, 2008).

$$D_{ij} = (X_i - X_j)' \cdot \Sigma^{-1} \cdot (X_i - X_j) \quad (12)$$

4.2.1.2 Algoritmos de emparejamiento

En el caso en que se desean obtener resultados generales de la estimación del efecto del tratamiento, son dos los algoritmos más comunes, los que se explican a continuación en el caso de que la distancia escogida es el *propensity score*:

- Nearest Neighbor Matching

El método más sencillo de emparejamiento es el del vecino más cercano, el cual consiste en buscar para cada individuo tratado, el o los individuos no tratados con el valor del *propensity score* más similar.

Este método posee dos variantes, con o sin reemplazo, en caso de aplicarse el reemplazo un individuo del grupo de control puede ser emparejado a más de un individuo tratado, elección más común debido a que reduce el sesgo en la estimación de los resultados y no es sensible al orden en que se realizan los emparejamientos (Dehejia & Wahba, 1999) Por lo que el conjunto de individuos emparejados con el individuo i estará dado por $C(i)$ como se observa en (13).

$$C(i) = \min_j ||e_i - e_j|| \quad (13)$$

Una vez que cada individuo que fue expuesto al tratamiento fue emparejado con un individuo del grupo de control, el ATT es obtenido simplemente promediando las diferencias entre grupos (Becker & Ichino, 2002).

- Radius Matching

Una de las desventajas que puede tener el método del vecino más cercano es que este puede ser muy distinto al individuo con el que se empareja. La solución más directa es definir una distancia máxima entre sus *propensity scores* para poder ser aceptado el emparejamiento. A esta distancia máxima se le llama *caliper*.

Al aplicar este método, cada individuo del grupo de control que es elegido pareja del individuo tratado debe estar a una distancia menor al *caliper*, por lo cual, las contrapartes escogidas son más parecidas que los encontrados con el método del vecino más cercano, aunque esto también implica que hay clientes tratados que podrían no tener pareja (Becker & Ichino, 2002). Una modificación a lo anterior nace de aceptar más de una pareja para cada individuo tratado, el *radius matching*, dónde se construye un radio con el valor del *caliper*, por lo que todos los individuos dentro de este radio son emparejados con el individuo tratado, este conjunto se representa como $C(i)$ en (14).

$$C(i) = \{e_j \mid \|e_i - e_j\| < r\} \quad (14)$$

Un tema relevante al analizar la calidad del emparejamiento obtenido con las técnicas de *matching* es el balance entre variables importantes del grupo tratado y del grupo de control. Entonces, surgieron variadas técnicas que buscan asegurar el balance en algunas variables deseadas o en todas ellas.

Por ejemplo, en el *genetic matching* (Diamond & Sekhon, 2013), se asigna una matriz de pesos a las variables utilizadas para la definición de la distancia entre los individuos para, a través de la optimización de ellos, definir con cuáles se alcanza el balance en las variables. Por otro lado, cuando sólo se desea obtener balance en unas pocas variables relevantes y estas son continuas se encuentran métodos como el Mahalanobis matching dentro de propensity score calipers (Rubin & Thomas, 2000).

En este último método, se busca mantener las propiedades del emparejamiento de *propensity score matching*, pero además establecer una serie de covariables claves en la cuáles se desea el balance. Para ello, se define un *caliper* (c) sobre la diferencia lineal de los *propensity score* (e) y dentro de los clientes que quedan dentro de este radio se establece un emparejamiento por distancia de Mahalanobis entre las variables claves (Z), quedando la distancia de este método especificada como en (15).

$$D_{ij} = \begin{cases} (Z_i - Z_j)' \Sigma^{-1} (Z_i - Z_j) & \text{if } |\text{logit}(e_i) - \text{logit}(e_j)| \leq c \\ \infty & \text{if } |\text{logit}(e_i) - \text{logit}(e_j)| > c \end{cases} \quad (15)$$

Por lo general, entre las variables claves se utiliza alguna medida del resultado a medir antes del tratamiento. Luego, la estimación del ATT sigue representada como la diferencia entre grupos, con el cuidado de no contabilizar de manera repetida los individuos que son control de más de un tratado.

4.3 MODELOS DE RESPUESTA

Otra de las grandes incógnitas en el mundo del marketing es determinar la efectividad de las campañas de marketing directo al momento de cambiar el comportamiento del cliente. Para atacar este problema se utilizan los modelos de respuesta que buscan determinar los clientes más prospectos a ser influenciados por la acción de marketing directo.

El modelamiento básico de estos problemas es la generación de estimaciones para los grupos de control y los grupos tratados mediante las acciones de marketing directo y a partir de estas concluir las características que influyen la respuesta, mediante herramientas como los modelos de elección binaria o algoritmos de clasificación (Lo, 2002).

4.4 EVALUACIÓN DE DESEMPEÑO

4.4.1 Matriz de confusión

La matriz de confusión es un método gráfico de representación de la capacidad de asertividad de un modelo (comúnmente binario) contrastado con sus errores. Es así como se construye una matriz como la observada en la Figura 11.

		Valor predicho	
		Negativo	Positivo
Valor real	Negativo	Verdaderos negativos	Falsos positivos
	Positivo	Falsos negativos	Verdaderos positivos

Figura 11: Ejemplo matriz de confusión binaria.
Fuente: Elaboración propia.

Las métricas que se desprenden desde la matriz de confusión son variadas, encontrándose entre las más utilizadas para la comparación de modelos cuando la proporción de casos positivos es pequeña la sensibilidad y la especificidad (Altman & Bland, 1994).

- *Sensibilidad (tasa de verdaderos positivos)*: Corresponde a la proporción de casos definidos como positivos que fueron bien clasificados (16).

$$Sensibilidad = \frac{verdaderos\ positivos}{verdaderos\ positivos + falsos\ negativos} \quad (16)$$

- *Especificidad (tasa de falsos positivos)*: Indica el error de clasificación del modelo, correspondiendo a los registros mal calificados como casos positivos (17).

$$Especificidad = \frac{falsos\ positivos}{falsos\ positivos + verdaderos\ negativos} \quad (17)$$

4.4.2 Análisis COR, AUC y Ganancia

El análisis COR (Característica Operativa Relativa) se utiliza para la evaluación de modelos de predicción, basándose en la matriz de confusión. En especial, se toma la sensibilidad y la especificidad con las que se construye la curva COR. A continuación, se calcula el área bajo la curva (AUC), la cual indica un mayor desempeño al ser más cercana a un 100%.

Otro método bastante utilizado para clasificadores binarios es la curva de ganancias acumuladas, en la cual se grafica la sensibilidad del modelo sobre el porcentaje de los casos mejor evaluados según el clasificador, lo que genera el gráfico de ganancias como se muestra en la Figura 12. En este caso, al igual que con la curva COR, un mejor modelo será el que cuya curva de ganancia se ubique por sobre la de los modelos a comparar.

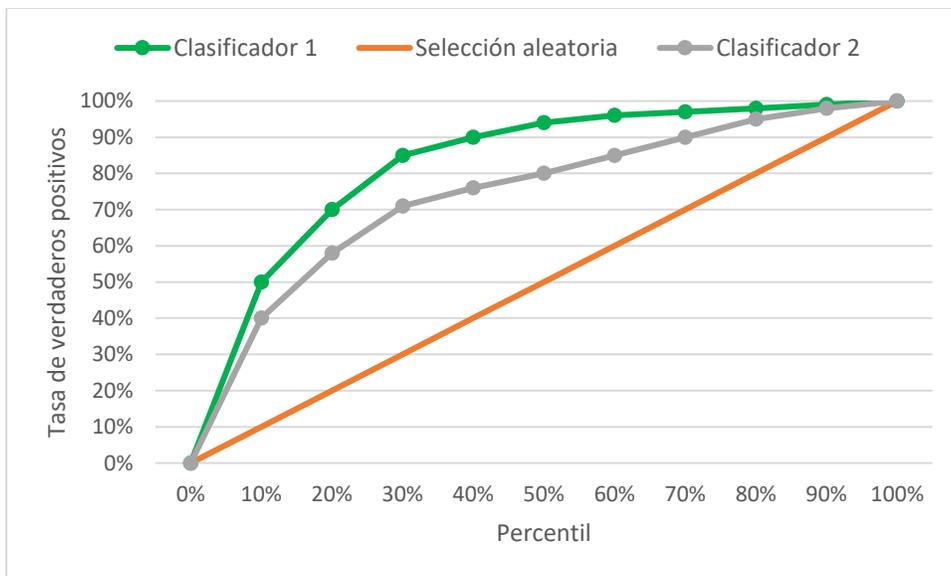


Figura 12: Ejemplo de curvas de ganancia y comparación con selección aleatoria.
Fuente: Elaboración propia.

5. METODOLOGÍA

Para hacerse cargo de los objetivos del proyecto, la metodología seguida es CRISP-DM (Shearer, 2000), la cual está marcada por un ciclo de 6 pasos fundamentales que se ilustran en la Figura 13.

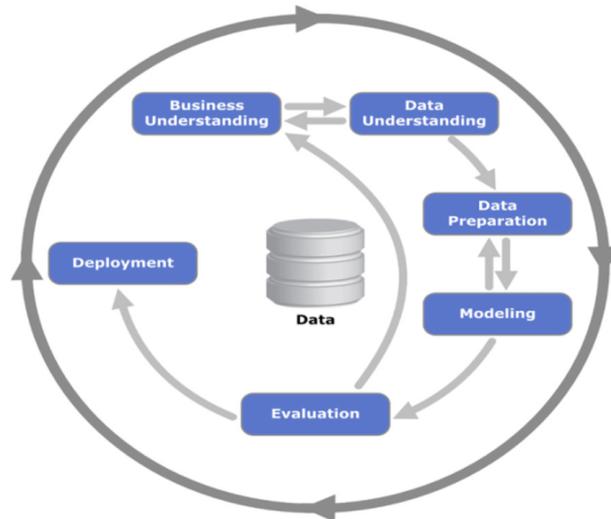


Figura 13: Ilustración de los pasos del ciclo de la metodología CRISP-DM.
Fuente: (Shearer, 2000)

1. Comprensión del negocio: Fase inicial enfocada en la definición de los objetivos y requerimientos del proyecto, ya realizado en los capítulos anteriores del informe.
2. Comprensión de los datos: Comienza con la recolección de los datos y procede con actividades para detectar posibles problemas de calidad en los datos, tener una primera aproximación con el comportamiento de estos y descubrir información relevante. Este proceso se realiza básicamente con herramientas de gestión SQL para extraer información de las bases de datos de la empresa.
3. Preparación de los datos: Contempla todas las actividades necesarias para construir los conjuntos de datos finales a utilizar a partir de los datos recolectados.
4. Modelamiento: Selección y aplicación de distintas técnicas de modelado. La herramienta principal para esta etapa será el software R (R Core Team, 2007).
5. Evaluación: Se evalúan los modelos seleccionados en el paso anterior y se revisan los procedimientos realizados con objetivos en revisar el cumplimiento de los objetivos.
6. Despliegue: Fase de recopilación de los resultados obtenidos, en la cual se organizan y son presentados de una manera utilizable para el cliente del proyecto.

A continuación, se detalla la metodología específica que se seguirá en la etapa de modelamiento y la de evaluación para los cuatro sub etapas de modelamiento del proyecto: Emparejamiento para la estimación del efecto de la activación de los rubros sobre el gasto y la rentabilidad, los modelos de propensión a la compra en los distintos rubros, determinación de un criterio de asignación y el análisis de respuesta de las campañas.

5.1 EFECTO ACTIVACIÓN DE LOS RUBROS

Para estimar el efecto de activarse en un rubro en el gasto y rentabilidad del cliente se utiliza el método de Mahalanobis matching dentro de propensity score calipers, para lo cual se seguirán los cuatro grandes pasos metodológicos que se observan en la Figura 14 propuestos por (Caliendo & Kopeinig, 2008). Es importante aclarar, que este proceso será llevado a cabo para los 16 rubros de gestión incluidos en el análisis, siendo el tratamiento definido como la activación de un cliente en uno de estos rubros.



Figura 14: Metodología para la estimación de efectos causales basadas en propensity score.
Fuente: (Caliendo & Kopeinig, 2008).

5.1.1 Estimación del *propensity score*

El valor individual del *propensity score* será estimado mediante regresiones logísticas para cada rubro. La variable dependiente por predecir será la realización de una activación en el rubro en cuestión durante un mes en específico. En tanto, las variables independientes estarán agrupadas en tres categorías: transaccionales, donde se incluirá la compra o no en los demás rubros durante el mes anterior, montos de compra y otros; segundo, características sociodemográficas como edad, sexo o grupo socioeconómico; y tercero, características del cliente asociadas al negocio, como el tipo de tarjeta que posee, su cupo o tenencia de productos de la empresa.

Para evitar problemas de correlación en las variables independientes de los modelos, se verificará la correlación entre las variables continuas mediante el coeficiente de correlación de Pearson y para las variables de tipo categóricas se utiliza el coeficiente V de Cramer.

Luego, se dará paso a la estimación de los parámetros correspondientes de los distintos modelos generados por el método de máxima verosimilitud a través del software R. Cabe señalar que para prevenir problemas de sobreajuste se dividirá de forma aleatoria la base de clientes en un 80% para entrenamiento de los modelos y un 20% para validación. De este modo, se generarán distintas especificaciones de modelos a través de un proceso de *stepwise*⁶, las que se compararán a través del criterio de información de Bayes (18) definido en (Schwarz, 1978). El BIC balancea el deseo de obtener un modelo con mejor ajuste (máxima verosimilitud, LL) en contra del deseo de modelos con una menor cantidad de parámetros (p) dependiendo del tamaño de la muestra (n). El modelo con menor BIC es escogido, que tiene aproximadamente el mismo resultado que escoger el modelo con la mayor probabilidad posterior (Raftery, 1985).

$$BIC = -2LL + p \log n \quad (18)$$

Cabe señalar, que los modelos generados en este punto para el *propensity score* son los mismos que serán utilizados como base de comparación para los modelos NBO, dejando en la sección de mejor próximo rubro el análisis predictivo de estos modelos y otros detalles.

5.1.2 Algoritmo de emparejamiento y soporte común

En esta etapa, también denominada de implementación, se realiza el emparejamiento de individuos, en este caso este procedimiento se realiza en dos pasos: Primero, el cálculo de la distancia definida por el método de Mahalanobis matching entre calipers de propensity score, vista en 15, entre los distintos individuos de las bases analíticas de cada rubro. Para esta distancia, se tomarán como covariables claves el *propensity score*, por su capacidad de resumir las covariables relevantes y su importancia para predecir la activación en un rubro, y la medida a evaluar (el gasto del cliente) en un periodo anterior.

Segundo, se utiliza el algoritmo de emparejamiento de *distancia de Mahalanobis con callipers en los propensity scores lineales*, con reemplazo, con una cantidad de vecinos que se determinará a través de un análisis de sensibilidad sobre el balance de las variables relevantes. Todo este proceso es implementando en R a través del paquete *MatchIt* (Ho, Imai, King, & Stuart, 2011).

⁶ El *stepwise* hace referencia a un proceso sistematizado en que se agregan y quitan individualmente variables independientes al modelo para evaluar el aporte particular de utilizar cada una de ellas.

Para evaluar la calidad del emparejamiento se usarán un t-test y un test chi cuadrado como lo proponen en (Rosenbaum & Rubin, 1985) para asegurarse que no haya diferencias significativas en el promedio de las variables relevantes del grupo que se activa en un rubro y el que no.

A continuación, se pasa a la estimación del ATT comparando los grupos resultantes de la verificación de la calidad del emparejamiento. Para finalizar con el análisis de los resultados.

5.2 MEJOR PRÓXIMO RUBRO

El centro de este trabajo de memoria es la proposición del mejor próximo rubro dónde se debe realizar una oferta para cada uno de los clientes. Para ello se estipulan dos niveles de análisis: primero, uno referido al entendimiento de las características y comportamiento que hacen propenso a un cliente a activarse en un rubro en particular, a través de la utilización de modelos logit por cada rubro. Y, en segundo lugar, la utilización de un modelo probit multivariado para incorporar la coocurrencia de compra entre rubros, además de determinar un ordenamiento o ranking en que se adquieren estos. A continuación, se detalla la metodología para este análisis.

5.2.1 Modelo logit

Como punto inicial, para entender las variables que poseen mayor importancia al momento de predecir la activación de un cliente en un rubro, se llevan a cabo regresiones logísticas para cada uno de ellos (cuya estimación ya fue tratada en el punto 5.1.1).

En este punto, se elaborará un análisis descriptivo de los parámetros estimados para la activación de cada rubro, de manera de sistematizar las principales características que hace a un cliente propenso a escogerlo. Además, de determinar cuáles son las variables independientes que poseen relevancia en una mayor cantidad de rubros a través de su significancia individual, para luego ser usadas en los siguientes modelos.

Asimismo, para adentrarse en los efectos de las covariables utilizadas sobre cada modelo se incorpora un análisis de los efectos marginales promedio de estas (o AME por su nombre en inglés: average marginals effects).

Para, finalmente, evaluar la calidad de la predicción de los modelos a través de la curva COR y métricas asociadas.

5.2.2 Modelo probit multivariado

Como ya fue mencionado, para la incorporación de efectos de coocurrencia y orden en la adquisición de los distintos rubros se utilizará un modelo probit multivariado similar al de (Li, Sun, & Wilcox, 2005), siguiendo las tres etapas principales del análisis jerárquico bayesiano estipuladas en el marco teórico de este informe.

Primero, se pasará a definir la utilidad que le entrega al cliente i la elección del rubro j en el periodo de tiempo t , como siempre estipulando una componente determinística y otra aleatoria. En segundo lugar, se pasará a la definición de los *priors* de cada uno de los parámetros definidos, para finalizar con la utilización del método *Gibb Sampler* para la estimación de estos. Cabe señalar que para la implementación del análisis bayesiano se utilizara el lenguaje Stan (Carpenter, y otros, 2017) aplicado en el software R.

Debido a la complejidad computacional de estimar el modelo probit multivariado, para poder realizar comparaciones sobre distintas posibles especificaciones del modelo se trabajará con una simplificación de este y una muestra de los clientes de tamaño trabajable. En particular, se estimarán regresiones logísticas, que pueden compartir o no parámetros, a través de un enfoque bayesiano, siendo la gran diferencia con el modelo probit el hecho de que no se considera la correlación temporal que pueda existir entre las compras en distintos rubros en un mismo periodo (coocurrencia).

A través de estos modelos logits estimados con un enfoque bayesiano, se seguirán dos líneas de modelamiento: por una parte, estimar una especificación similar a la realizada con los logits para el *propensity score* en que cada modelo tendrá una estructura lineal en sus parámetros. Y por otro lado, se trabajará en una especificación que pueda entregar información sobre un posible orden en la activación de los distintos rubros de gestión, permitiendo en este caso modelos no lineales que comparten parámetros.

Un paso común para todos los modelos estimados de forma bayesiana, será la verificación de la recuperación de parámetros simulados, de forma tal de corroborar que el modelo es capaz de encontrar el mecanismo subyacente que genera los datos. Para ello se procederá en 5 pasos: primero, generar las variables independientes de forma aleatoria. Segundo, generar los parámetros de forma aleatoria dentro de sus posibles valores. Tercero, construir la variable independiente según las variables y parámetros generados siguiendo la especificación del modelo. Cuarto, establecer las variables dependientes como independientes para el siguiente periodo y repetir los pasos anteriores. Quinto y final, estimar los parámetros y corroborar que el valor real de los parámetros se encuentra en el intervalo de confianza a un 95% de estos.

Con respecto a la evaluación de los modelos estimados de forma bayesiana, existe una amplia gama de métricas y metodologías utilizables. Un enfoque son los clásicos criterios de información como el AIC y el BIC (ya revisado para el logit) o simplemente la verosimilitud, los cuáles han sido modificados para el caso de modelos bayesianos con el fin de utilizar totalmente la distribución posterior estimada y no sólo basada en un punto estimado como las antes mencionadas. Uno de estos criterios es el WAIC (Widely Aplicable Information Criterion) que posee la ventaja de ser asintóticamente igual a realizar validación cruzada bayesiana y es invariante ante distintas parametrizaciones de un modelo (Watanabe, 2010), cuya especificación se ve en el Anexo C. Finalmente, se estima la diferencia en WAIC entre los modelos y su respectiva desviación estándar para determinar la preferencia por uno.

Por último, al igual que para los modelo logit se evaluarán y compararán estos modelos con los anteriores en términos de las curvas de ganancia.

5.3 ASIGNACIÓN DE OFERTAS

El siguiente paso es la definición de criterios para la asignación de la próxima oferta a cada cliente balanceando su propensión y el gasto esperado que le entrega a la empresa.

En particular, se propone la utilización de una métrica que permita comparar de manera unidimensional la oferta que debe ser asignada a cada cliente. Esta debe considerar dos puntos relevantes: la probabilidad con que el cliente se activará en un rubro y el gasto esperado del rubro en cuestión. Entonces, se define el gasto esperado por la activación como se muestra en (19).

$$R_{ni} = P_{ni} * r_n \quad (19)$$

Dónde R_{ni} representa una estimación de la ganancia de que el cliente i se active en el rubro n , que es el resultado de la ponderación del gasto esperado al activarse en el rubro n (r_n) con la probabilidad de que el cliente i se active en el rubro n (P_{ni}). Claramente, el criterio de selección será seleccionar la oferta de mayor R_{ni} para cada cliente.

Por otra parte, esta métrica se comparará con otros métodos de asignación clásicos en los modelos NBO como lo son la imputación de ofertas según el decil de propensión en que se encuentra el cliente o la asignación del rubro en el que posee mayor ganancia o *lift* con respecto del promedio de propensión de la cartera.

5.4 ANÁLISIS DE CAMPAÑAS

Para finalizar, se realizará un análisis de las campañas vía email del negocio en un rubro definido según el análisis anterior y los datos disponibles de campañas. El principal objetivo será definir si estos rubros logran tener una mayor tasa de activación al ser expuestos a este tipo de marketing y si hay características de los clientes que los hagan más sensibles a estas campañas.

Para esto se hará un análisis descriptivo de la campaña y del rubro al que pertenece en los meses que se realizó la campaña, para luego profundizar el análisis con un modelo de respuesta basado en regresiones logísticas binarias, en los que la variable dependiente será la utilización de la oferta propuesta en la campaña o la compra en el rubro asociado a esta, dejando tres tipos de variables independientes: primero, si el cliente recibió el correo o pertenecía al grupo de control; segundo, características y comportamiento previo del cliente; y tercero, el efecto cruzado entre los dos niveles anteriores.

6. ALCANCES

Tomando en cuenta las distintas limitaciones del negocio y de la empresa, en conjunto a los intereses del trabajo y tiempos de ejecución, se proponen los siguientes alcances:

- Se utiliza sólo la información de los clientes que posee actualmente la empresa, no se consideró la ejecución de encuestas o experimentos para obtener información de los clientes.
- Sólo son objeto de estudios los clientes identificables por la empresa, más en específico sólo se contemplan las compras en las que estos clientes utilizaron la tarjeta de crédito de la empresa.
- El análisis se realizó sobre una base de clientes con cierto grado de antigüedad, debido a que por conocimiento de la compañía se sabe que los clientes tienen cierto periodo de maduración con el uso de la tarjeta de crédito y otros productos, por lo que un análisis del tipo que se propone en este trabajo para clientes nuevos puede ser una continuación de este.
- El trabajo de memoria no incorpora la creación de nuevas promociones, sino sólo contempla la asignación de ofertas existentes o la selección de rubros en particular en el que se deberían realizar las ofertas.
- Dado el tiempo disponible, no se contempla en este trabajo la realización de un experimento de campo como forma de validación de las ofertas propuestas. Sino sólo se utilizará la historia transaccional de los clientes y el histórico de campañas ya realizadas por la empresa.
- Los rubros utilizados para la proposición de próximas ofertas son los que actualmente maneja la empresa, por lo que serán un insumo y este trabajo no se hace cargo de la generación de nuevos rubros más allá de nuevas divisiones o agrupaciones necesarias.

7. COMPRENSIÓN DE LOS DATOS Y ANÁLISIS DESCRIPTIVO

Para el desarrollo de este trabajo de memoria se consideran fundamentalmente cuatro fuentes de información: transacciones a nivel de boletas por cliente, información sociodemográfica de los clientes, información del contrato de los clientes y la rentabilidad anual de estos.

7.1 INFORMACIÓN SOCIODEMOGRÁFICA Y DEL NEGOCIO

En lo que respecta a la información a nivel de clientes, se tienen datos sociodemográficos y del contrato con la empresa, de los cuáles se seleccionan los estipulados en la Tabla 1.

Tabla 1: Descripción de las variables relevantes a nivel de clientes.

Información sociodemográfica		Información del negocio	
Variable	Descripción	Variable	Descripción
ID_CLIENTE	Identificador único del cliente	ID_CLIENTE	Identificador único del cliente
FECHA_NAC	Fecha de nacimiento del cliente	ID_CONTRATO	Identificador único de los contratos
GSE	Grupo socioeconómico del cliente	MONTO_CUPO	Cupo asignado al cliente
ID_SEXO	Sexo del cliente	FECHA_APER_CTO	Fecha en que el cliente inició el contrato
		ID_TARJETA	Tipo de tarjeta asociada al contrato
		ID_MORA	Indica situación de mora de la cuenta del cliente

De esta información, se puede capturar que la empresa posee registro de 2.862.612 clientes con cuenta abierta, de los cuales la mayoría corresponde a mujeres con un 52% de la cartera y con respecto al grupo socioeconómico, el predominante es el D con un 53% (el detalle de la distribución por GSE en Anexo A). También, construyendo la edad de los clientes a partir de su fecha de nacimiento, se puede observar que el rango etario va desde los 18 hasta los 91 años, encontrándose su promedio en 48 años.

Por parte de la información del negocio una característica importante es el tipo de tarjeta del cliente, las cuáles se pueden dividir en dos grandes grupos: tarjetas abiertas, que se pueden utilizar en cualquier comercio que posea sistema Transbank, y tarjetas cerradas, sólo utilizables en los negocios del grupo y algunos

comercios aliados particulares. Mayoritariamente, los clientes poseen tarjetas abiertas (68%), aunque en el rango etario de los 18 a los 25 años es mayor la proporción de tarjetas cerradas que se eleva hasta un 59% como se aprecia en la Figura 15.

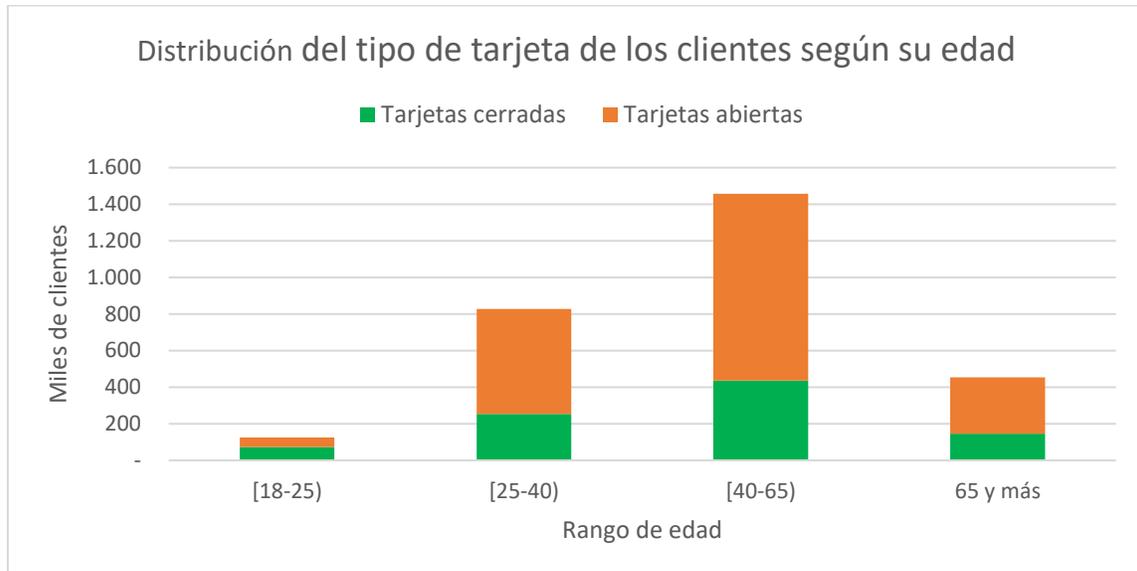


Figura 15: Distribución de la cartera de clientes según su edad y tipo de tarjeta.
Fuente: Elaboración propia.

7.2 TRANSACCIONES Y RUBROS DE GESTIÓN

En primera parte, se posee la información de todas las transacciones realizadas con la tarjeta de la empresa desde el 1 de junio de 2015 hasta el 31 de agosto de 2017. En particular, las variables más relevantes para el proyecto se resumen en la Tabla 2.

Tabla 2: Campos relevantes de la base de datos de transacciones y su descripción.

CAMPO	DESCRIPCIÓN
NUMERO_TRN	Identificador único de la transacción
ID_TIPO_TRANSACCION	Identificador del tipo de transacción (en particular para el análisis se utiliza sólo las que representan ventas)
ID_CLIENTE	Identificador único del cliente
MONTO_LIQUIDO	Monto líquido de la transacción
ID_RUBRO	Identificador de la categoría del comercio
ID_COMERCIO	Identificador del comercio en el que se realizó la transacción
FECHA_COMPRA	Fecha en la que se realizó la transacción
ID_TARJETA	Identificador del tipo de tarjeta utilizado en la transacción (Abierta o cerrada)
ID_CUOTAS	Cantidad de cuotas en las que se realiza la compra

De manera exploratoria, se logra observar que la cantidad promedio de clientes que poseen transacciones en un mes en particular es de 2,2 millones. Asimismo, se rescata que existe clara estacionalidad en diciembre en el monto de gasto de los clientes, además de mantenerse una tendencia al alza de este, como se aprecia en la Figura 16.

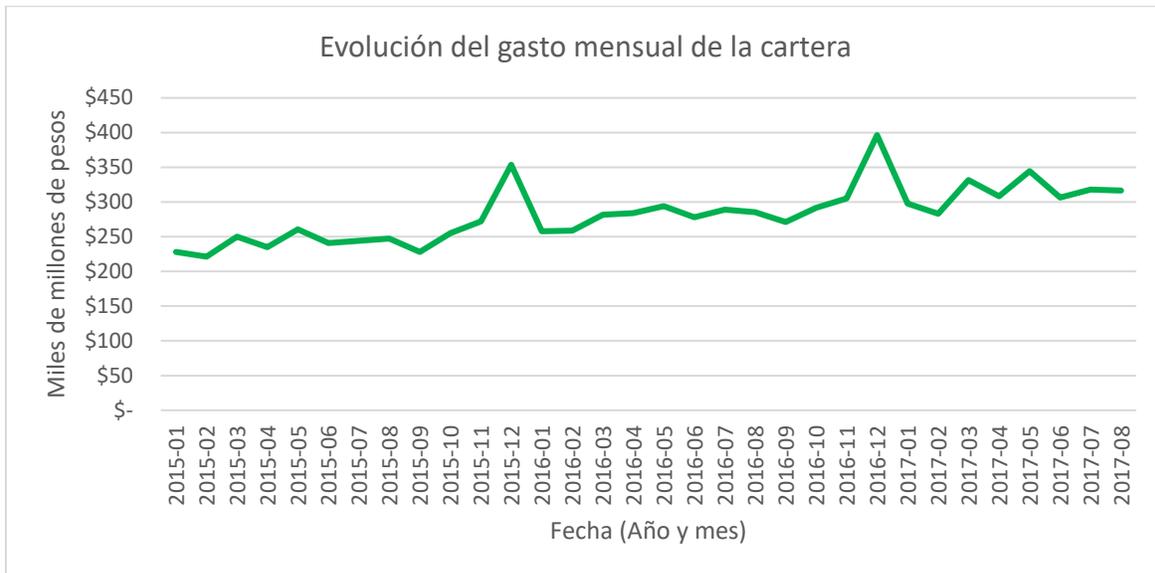


Figura 16: Distribución mensual de los clientes con compra y su gasto con la tarjeta de crédito.
Fuente: Elaboración propia.

De esta misma fuente, se puede verificar que durante el periodo de estudio existieron 814.033 comercios distintos con transacciones, por lo que para su agregación se cuenta con tres niveles de jerarquías definidas por la empresa: categoría, sub rubro de gestión y rubro de gestión. En la Tabla 3 se puede apreciar la jerarquía para el rubro de gestión Automotriz.

Tabla 3: Jerarquía de rubros, sub rubros y categorías para el rubro de gestión Automotriz.

RUBRO DE GESTIÓN	SUB RUBRO DE GESTIÓN	CATEGORÍA
AUTOMOTRIZ	VENTA AUTOMOTRIZ	VENTA MOTOCICLETAS
		VENTA AUTOS Y CAMIONES NUEVOS/USADOS
		AGENCIA ARRIENDO AUTOMÓVILES
		TIENDAS DE NEUMÁTICOS
	ACCESORIO AUTOMOTRIZ	TIENDAS DE AUTO SERVICIO PARA VEHÍCULOS
		TIENDAS DE PIEZAS DE AUTOMÓVIL

Dada la pertenencia de la empresa a un grupo, en la Figura 17 se puede apreciar cómo es de esperar que el mayor porcentaje del gasto con la tarjeta de crédito se produce en los tres principales rubros en los que el grupo tiene un comercio importante: Tiendas por Departamento, Mejoramiento del Hogar y Supermercados.

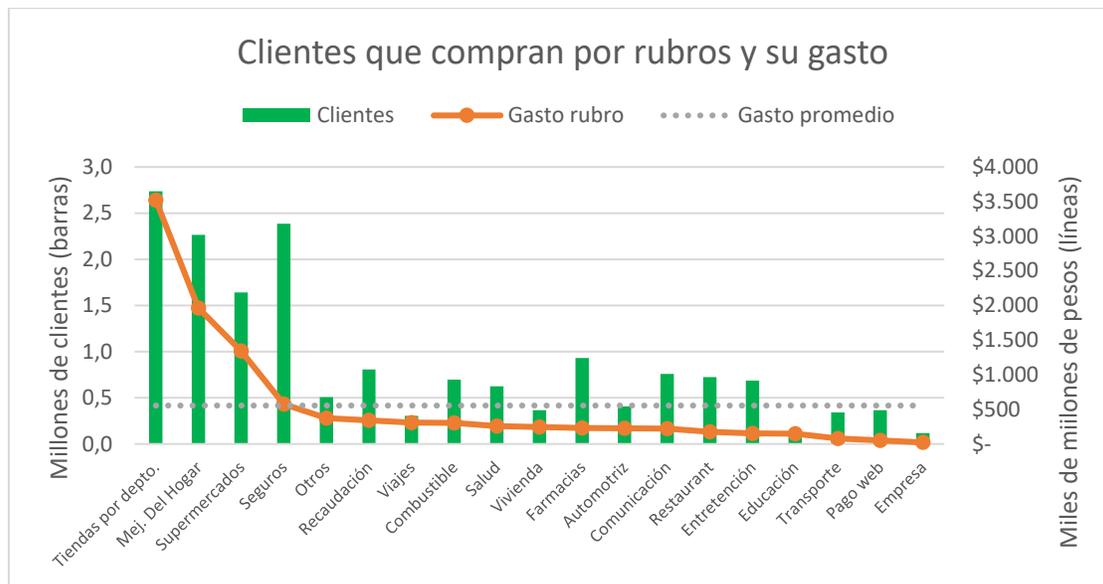


Figura 17: Cantidad de clientes que compran y su gasto total por rubro durante el periodo de observación.
Fuente: Elaboración propia.

Además, debido a la baja relevancia en monto que poseen los rubros de gestión Empresa y Pago web y a sus características particulares que los hacen poco gestionables por parte de la empresa, estos se agruparán de ahora en adelante junto al rubro Otros y no formarán parte del análisis a nivel de rubros de gestión.

En la misma línea, se puede apreciar cómo varían los montos de venta en los distintos rubros de gestión a través de los distintos meses. En específico, en la Figura 18 se muestra la serie temporal del gasto para los 4 rubros que acumulan mayor gasto de los clientes y cuyo valor no posee grandes variaciones a través del año, exceptuando a Tiendas por Departamento que posee dos *peaks* de gasto en diciembre y mayo, dados por las festividades de navidad y el día de la madre.

Respecto del comportamiento del gasto del resto de los rubros, se puede ver en el Anexo D que esta se ubica siempre bajo los \$10 miles de millones, exceptuando las grandes alzas de gasto que posee el rubro Recaudación durante marzo debido a variados servicios que se pagan en este mes, por ejemplo, los permisos de circulación vehicular en las municipalidades.

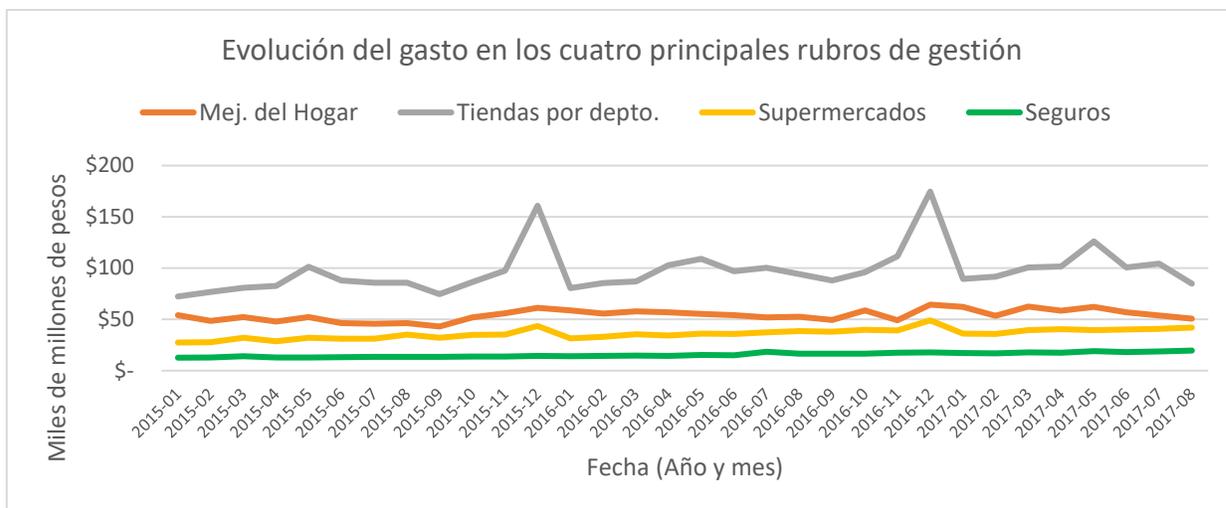


Figura 18: Evolución del gasto total por rubro de gestión con la tarjeta de crédito.
Fuente: Elaboración propia.

Por otra parte, a nivel de clientes, se puede apreciar cómo se distribuyen según la cantidad de rubros en los que compran y cuáles son sus montos de gasto. En la Figura 10, ya expuesta en la descripción del proyecto, se observa que a medida que crece el número de rubros con compra aumenta el gasto que poseen los clientes con la tarjeta, pero a su vez son menos los clientes que se ubican en el segmento. En particular, los clientes compraron en promedio en 4,3 rubros de gestión durante el año 2016.

Por último, tomando en cuenta el tipo de tarjeta con el que se realizan las transacciones, primero, se observa que a nivel de gasto los de tarjeta abierta gastan en promedio MM\$1,43 más que los de tarjeta cerrada anualmente (basado en la información del año 2016). Además, mientras los clientes de tarjeta cerrada sólo realizaron compras en un promedio de 2,6 rubros durante el 2016, los de tarjeta abierta lo hicieron en un promedio de 5,1, por lo que puede significar una característica relevante en la probabilidad de compra de ciertos rubros.

7.3 RENTABILIDAD

Por parte de la rentabilidad anual, en el año 2016 se tiene que el promedio de la cartera ronda los \$48.000, teniendo un gran rango de valores que van desde clientes con rentabilidad negativa de cerca de -MM\$30 y otros de alta rentabilidad cercana a los MM\$6. Además, con respecto al tipo de tarjeta de los clientes, se observa que la rentabilidad de los clientes con tarjeta abierta es en promedio 11 veces la rentabilidad de los con tarjeta cerrada.

Ahora, enlazando la información transaccional de los clientes y la rentabilidad, podemos observar cómo se comporta esta última según la actividad de los clientes en los distintos rubros de gestión. En la Figura 19 se visualiza que los clientes con compras durante el último año en los rubros Vivienda y Automotriz son los que poseen en promedio mayor rentabilidad, cercana a los \$145.000, mientras que entre los de menor rentabilidad promedio están los tres principales con participación del grupo (Tiendas por Departamento, Supermercados y Mejoramiento del Hogar).

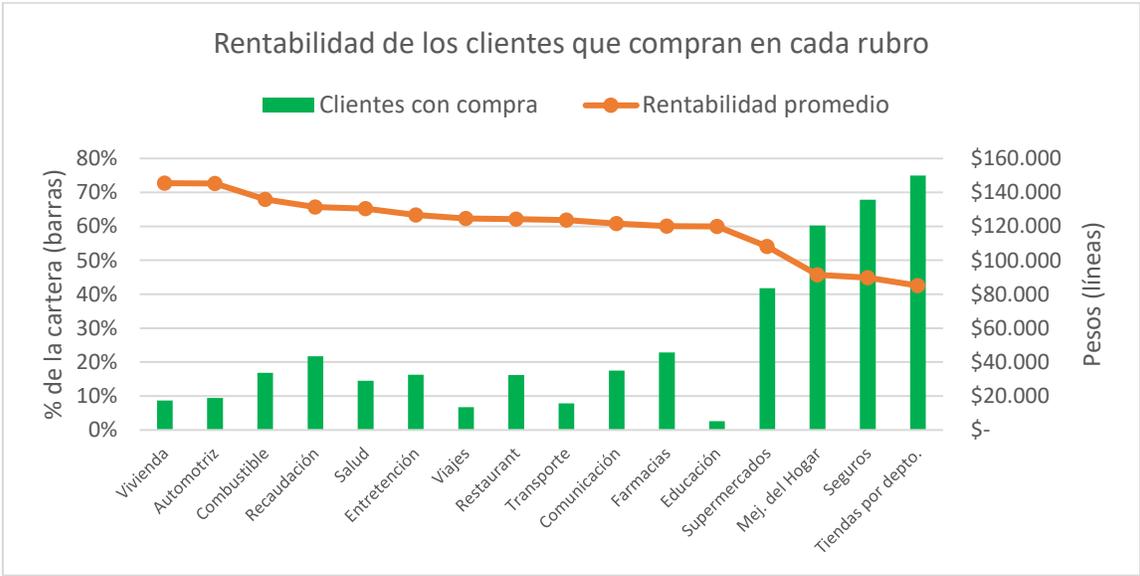


Figura 19: Rentabilidad anual promedio de los clientes que compraron en cada rubro de gestión durante el 2016.
Fuente: Elaboración propia.

La rentabilidad anual recién mostrada se conforma por muchos elementos, no todos ligados al uso de la tarjeta de crédito, como el canje de puntos del club de fidelización o la tenencia de otros productos de la empresa. Por ello, el análisis se desarrollará sobre los ingresos de rentabilidad mensual que tienen directa relación sobre el uso de la tarjeta por el cliente. En particular, estos ingresos vienen de dos fuentes: ingresos por intereses asociados a la compra en cuotas y *merchant*.

Con esta primera mirada, se puede determinar la relevancia de aislar los efectos que produce la actividad en cada rubro de gestión sobre el gasto y la rentabilidad debido a las diferencias que puede generar para la empresa, a pesar de que claramente existe una gran cantidad de características que pueden estar explicando este fenómeno.

8. PREPARACIÓN DE LOS DATOS

En lo que respecta al preprocesamiento de los datos seleccionados, se presenta un procedimiento principal, la eliminación de clientes con comportamientos extremos.

Los clientes tomados en consideración para el análisis, desde este punto en adelante, son aquellos que realizaron la apertura de su tarjeta de crédito hasta el 30 de junio de 2015, debido a que los análisis se realizarán sobre su comportamiento entre julio de 2016 y junio de 2017 (no se toman en cuenta clientes con antigüedad menor a 1 año por ser considerados clientes nuevos y tener un comportamiento especial) y no se permitirá el ingreso de nuevos clientes para el estudio. Además, se les exige tener su cuenta abierta durante todo el periodo de estudio y que hayan realizado al menos una transacción durante este.

Luego de estos filtros, la base está compuesta por 1.227.080 clientes, de los cuáles se eliminan adicionalmente 13 clientes por tener una edad mayor a 200 años, lo que se considera error de la fuente de datos. Con esto, se queda con 1.227.067 clientes de los que se observan datos transaccionales básicos en la Tabla 4.

Tabla 4: Gasto y transacciones de los clientes de la base analítica entre julio de 2016 y junio de 2017.

Cantidad de clientes	Promedio Transacciones	Gasto total	Gasto promedio
1.227.067	71.5	\$ 2.414.231.008.227	\$ 1.967.481

El objetivo de esta etapa es apartar del análisis a los clientes que poseen gastos y/o transacciones en cantidades anómalas, por ejemplo, debido a la existencia de tarjetahabientes que utilizan la tarjeta sin fines personales. Para la eliminación, se utiliza como criterio de anormalidad la distancia de Mahalanobis (Matsumoto, Kamei, Monden, & Matsumoto, 2007) calculada sobre el total de transacciones y monto gastado por el cliente. Una de las limitantes de este método es que asume distribuciones normales en sus variables, lo que no cumplen ni el gasto ni las transacciones por persona como se aprecia en el Anexo E y el Anexo F, respectivamente.

Para atacar este problema, se realiza la transformación logarítmica de ambas variables, de las cuáles se puede observar su densidad en el Anexo G y el Anexo H, respectivamente, además de sus gráficos de cuartiles sobre cuartiles en el Anexo I. De estos gráficos, se concluye que la transformación logarítmica del gasto se comporta de forma prácticamente normal, mientras que la transformación logarítmica de las transacciones lo hace en algunos tramos, por lo que se permite la utilización del método.

Luego, se define una distancia de 13,8 como el criterio de corte para valores extremos, que representa el nivel al cuál se encuentra el 99,9% de la varianza de los datos asumiendo la distribución normal. Como resultado, se obtiene que 14.401 (1,17%) clientes se excluyen de la muestra (*outliers*), mostrándose en la Figura 20 la distribución de sus gastos y transacciones respecto de los clientes que se mantienen en esta.

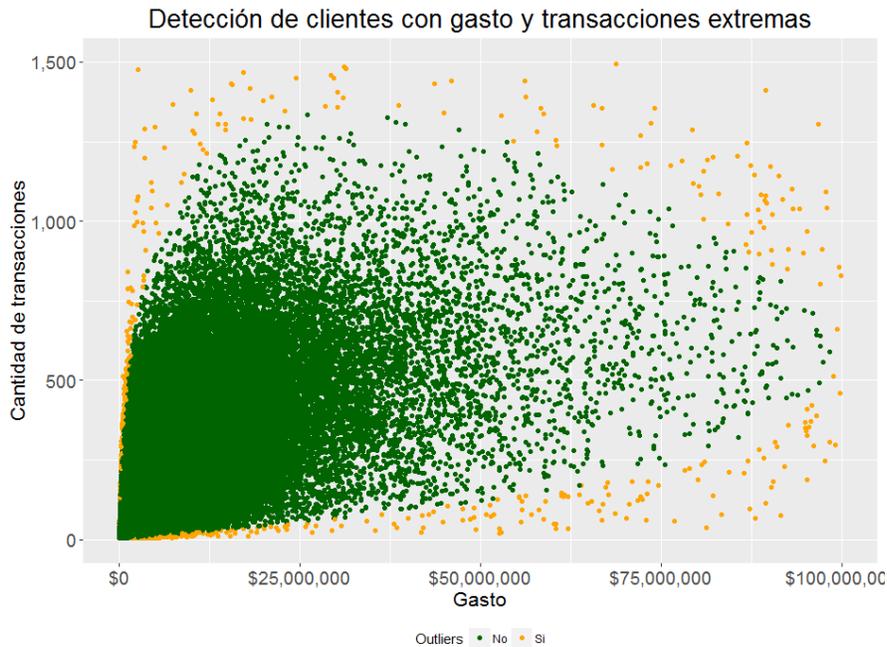


Figura 20: Distribución de la cantidad de transacciones y gasto de los clientes para detección de valores extremos.

Fuente: Elaboración propia.

En la Tabla 5, se pueden observar las transacciones y gasto promedio de los clientes que se mantiene en la base y de los que no, corroborando que el grupo eliminado más importantes es el de clientes con compras puntuales con valores extremadamente altos.

Tabla 5: Gasto y transacciones durante el periodo de estudio de los clientes eliminados y de los que se mantienen en la base.

Outliers	Cantidad de clientes	Gasto promedio	Transacciones promedio
No	1.212.666	\$ 1.939.372	71,8
Si	14.401	\$ 4.334.498	49,4

9. EFECTO DE LOS RUBROS SOBRE LA RENTABILIDAD

En este capítulo se presenta el desarrollo de la metodología de emparejamiento por distancia de Mahalanobis dentro de calipers por propensity score y sus principales resultados.

En primer lugar, se define el tratamiento, que corresponde al hecho de activarse en un rubro en particular, siendo considerada como activación en este caso la realización de al menos una transacción en el rubro correspondiente cuando no se había comprado el mismo en los tres meses anteriores⁷.

A continuación, dado el objetivo de la definición de efectos causales por parte de cada uno de los rubros sobre el gasto y la rentabilidad del cliente, se estipulan tres periodos relevantes: periodo pre-activación, periodo de activación y evaluación, observándose en la Figura 21 los rangos de fechas seleccionados para cada uno de estos periodos (en meses).

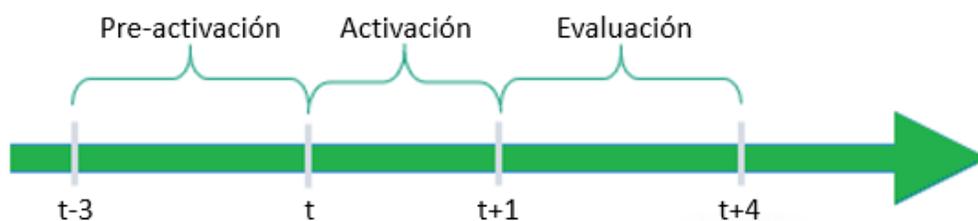


Figura 21: Línea de tiempo de los periodos definidos para la aplicación del método de emparejamiento.
Fuente: Elaboración propia.

En particular, del periodo de pre-activación se toman las características necesarias para el emparejamiento, en la activación se establece si el cliente se activa en un rubro o no, y en el periodo de evaluación se tiene como objetivo final el análisis del gasto y la rentabilidad de estos.

Es así como, del total de registros de la muestra⁸, para el análisis de cada rubro sólo se consideran los pares cliente-mes en los que no se han realizado transacciones en el rubro en particular durante el periodo de pre-activación, pero que tengan al menos una transacción en otro rubro durante el mismo intervalo, para asegurarse de no tomar en cuenta clientes que se puedan estar fugando.

⁷ Esta definición de activación se toma principalmente por convenciones tomadas por las áreas comerciales de la empresa.

⁸ De la base analítica de clientes se seleccionó una muestra aleatoria del 10% para realizar el emparejamiento, que corresponden a 121.266 clientes.

9.1 CÁLCULO DEL PROPENSITY SCORE

Para la asignación de clientes comparables por el método escogido, el primer paso es la estimación de la propensión a que el cliente se active en un rubro dado su comportamiento y características en el periodo de pre-activación.

La estimación del *propensity score* se realiza mediante un modelo logit binario para cada rubro de gestión, en los que la variable dependiente corresponde a la activación en el rubro (el tratamiento). Por parte de las variables explicativas, se consideran tres aspectos fundamentales del cliente: características sociodemográficas, características asociadas al negocio y comportamiento en los distintos rubros en el mes anterior. El set inicial de variables puede ser observado en el Anexo J.

Luego, se pasa a descartar las distintas covariables según sus correlaciones, que se pueden observar en el Anexo K y el Anexo L para las variables continuas y categóricas, respectivamente, dejando como criterio de corte para la eliminación de las variables un valor de 0,39. Cabe destacar que no se considera en este punto la eliminación de las variables de actividad en los distintos rubros, ya que un punto importante de la investigación es corroborar los efectos cruzados entre estos.

También, cabe mencionar que para captar las estacionalidades de los distintos rubros se utilizan variables del tipo dummy que indican el mes que se está prediciendo. Además, para evitar que la estimación de las estacionalidades estuviese sesgada por lo sucedido en sólo un año, se agregan a la base las compras de los mismos clientes de la muestra durante el año anterior.

Es con lo anterior, que se conforma la formulación de la componente observada de la utilidad para el cliente i de activarse en el rubro j en el periodo t (V_{ijt}) general de los modelos logits estimados como se muestra en (20). Donde D es un vector con las dummies que indican la compra o no de los distintos rubros, X el vector de características del cliente y las ligadas al negocio y Z el vector de dummies con el periodo que se está prediciendo, cada uno de estos acompañados con sus respectivos coeficientes.

$$V_{ijt} = \alpha_j + \sum_{k \neq j} \beta_k \cdot D_{ikt-1} + \bar{\gamma}_j \cdot X_{it-1} + \bar{\theta}_j \cdot Z_t \quad (20)$$

A continuación, para cada uno de los rubros, se especifica su base analítica correspondiente a los clientes que no poseen transacciones en aquel rubro durante los tres meses anteriores, con lo cual la cantidad de clientes en cada base analítica y el porcentaje de activaciones se ve resumido en la Figura 22.

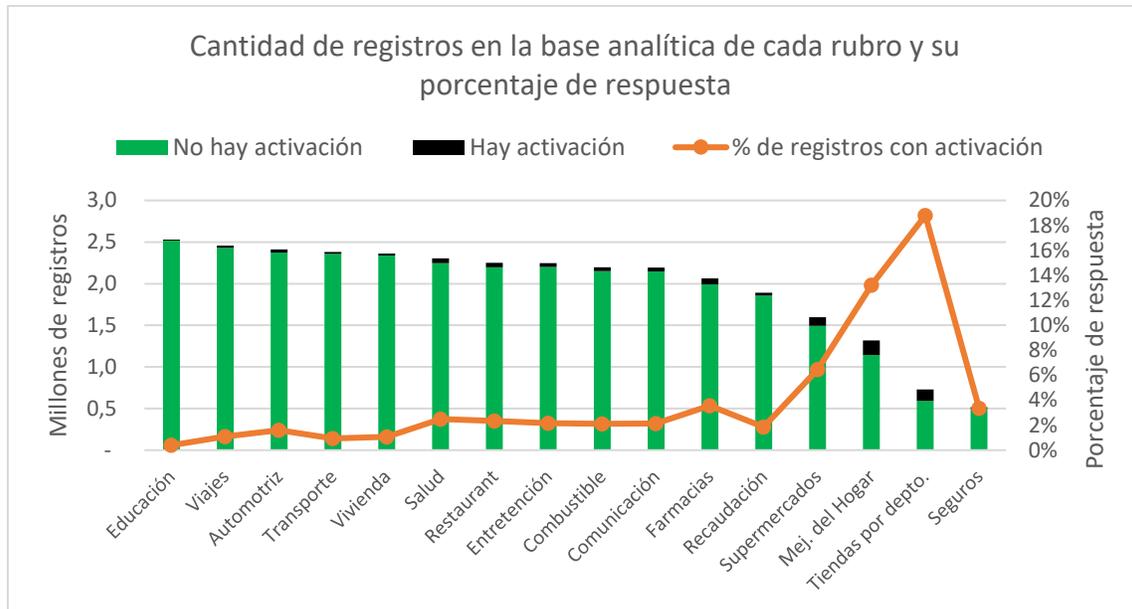


Figura 22: Cantidad de clientes en la base analítica de cada rubro y porcentaje de activación en mes de estudio. Fuente: Elaboración propia.

Para continuar, se separa la camada de clientes en un 80% para entrenamiento y un 20% para testeo. Para, posteriormente, estimar los parámetros de los modelos en la muestra de entrenamiento, aplicando el proceso de selección de menor BIC con *stepwise* (se pueden revisar las variables seleccionadas para cada modelo en el Anexo M y el BIC de estos en el Anexo N). Con respecto a las especificaciones seleccionadas, se puede verificar que el BIC varía desde 62.901 hasta 728.866, mostrando que existen grandes diferencias en el ajuste de los modelos. Asimismo, en ningún caso se mantienen todas las variables incorporadas en el proceso de *stepwise*, por lo que queda claro que no se han seleccionado las especificaciones con mayor verosimilitud.

Desde este punto en adelante, se ejemplificarán por simplicidad los distintos pasos de la metodología con el rubro farmacias, dejando los resultados del resto de los rubros especificados en su respectivo anexo.

9.2 EMPAREJAMIENTO Y SOPORTE COMÚN

Antes de la ejecución del emparejamiento, se debe verificar el soporte común de *propensity score* entre los clientes que se activan y los que no.

En el ejemplo del rubro farmacias, corroboramos que el valor mínimo de *propensity score* es de 0,8% y el máximo de 76%, lo que muestra el gran rango de probabilidad que existe de activarse en el rubro (hecho similar se aprecia en los demás rubros en el Anexo O) y en específico el rango de *propensity score* en que

se encuentran clientes tanto en el grupo que se activa como en el que no es entre 0,8% y 75,4%, como se aprecia en la Figura 23. En la misma línea, es interesante apreciar que, como es de esperar si el clasificador hace un buen trabajo, el promedio del *propensity score* en el grupo que se activa es bastante mayor (10,8%) que el de los que no se activaron (4,9%).

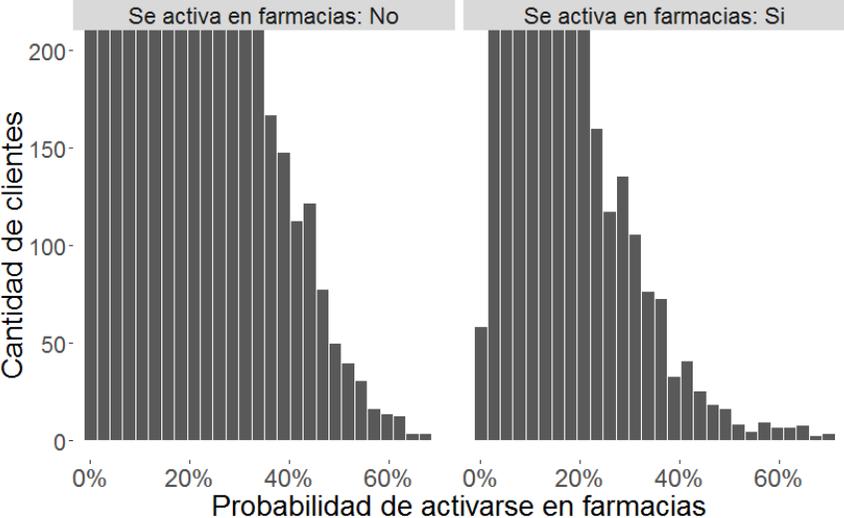


Figura 23: Verificación gráfica del soporte común de propensity score para los clientes que se activan en farmacias y los que no.
Fuente: Elaboración propia.

Como se explicó en la metodología, el emparejamiento de clientes se realiza mediante la distancia de Mahalanobis considerando calipers en el propensity score lineal. Con respecto al *caliper*, se define de manera diferenciada para cada rubro como el 25% de la desviación estándar de sus propensity score lineales, como se recomienda generalmente (Rosenbaum & Rubin, 1985).

Mientras que, para determinar la cantidad de vecinos a utilizar para el *radius matching*, con reemplazo, existe un trade-off entre varianza y sesgo al agregar una mayor cantidad de vecinos, debido a que se reduce la varianza al construir el grupo de control para cada cliente con una mayor cantidad de información, pero aumenta el sesgo debido a la utilización de clientes más lejanos que el más similar (Smith, 1997).

Es por lo anterior que, dado que utilizar una mayor cantidad de vecinos hace la estimación más robusta, se realiza un análisis de sensibilidad del balance de las variables en las que se desea asegurar el balance (*propensity score* y gasto) ante la cantidad de vecinos utilizados. Para ello, se toma una muestra del 5% de los clientes de la camada en cada mes y se aplica el emparejamiento, para luego comparar en la Figura 24 el balance dados los vecinos utilizados.

Con esta información, se da un rango de tolerancia del 10% al balance de cada variable (es decir, sólo se permite que cada variable no esté balanceada en un máximo de 2 rubros), con lo que se determina que la cantidad de vecinos a utilizar será de 4.

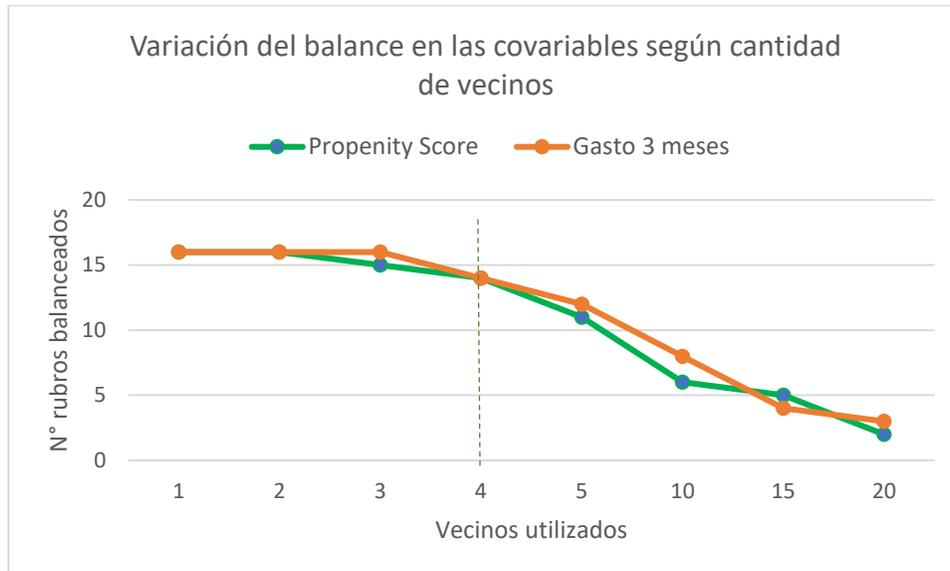


Figura 24: Variación de la cantidad de rubros para las cuáles están balanceadas el propensity score y gasto pasado según cantidad de vecinos utilizados.
Fuente: Elaboración propia.

Una vez realizado el emparejamiento, se puede verificar que se encuentra al menos un emparejamiento para el 100% de los casos en que hubo activación. Aunque, no en todos los casos se encuentran 4 vecinos dentro del *caliper*, teniendo que el promedio de vecinos para cada registro es de 3,2.

9.3 CALIDAD DEL EMPAREJAMIENTO Y ESTIMACIÓN DEL EFECTO

En este punto, se procede a verificar el balance entre las variables relevantes para el emparejamiento (gasto en el periodo pre-activación y *propensity score*) a través de un test chi cuadrado para el primero y uno de medias para el segundo.

Para el ejemplo del rubro farmacias, se pueden observar en la Tabla 6 los promedios de estas variables junto al p-valor de su correspondiente test, con lo que se verifica que no existen diferencias significativas con un 95% de confianza entre el grupo que se activa y el que no en ninguna de las variables.

Tabla 6: Verificación del balance de las covariables relevantes en la base emparejada para el rubro farmacias.

Variable	Tratados farmacias	No tratados farmacias	Diferencia	p-valor
PS	10,8%	10,6%	0,2%	0,07
Gasto en pretratamiento	\$588.945	\$560.784	\$12.523	0,07

En el Anexo P, se muestran los resultados del test de proporciones y test de medias para cada uno de los rubros, con lo que se comprueba la calidad del emparejamiento concluyéndose que se poseen grupos comparables para el análisis post-activación, exceptuando cuatro rubros que poseen diferencias en el periodo pre-activación, pero se mantiene la cantidad de vecinos para que la evaluación sea estándar.

Asimismo, podemos corroborar de manera gráfica que este método de emparejamiento mantiene la propiedad del emparejamiento por *propensity score* en que ha distintos valores de probabilidad de activación se mantiene balance entre las variables estudiadas. En la Figura 25, se observa este comportamiento en la estimación para el rubro farmacias para el gasto en el periodo de pre-activación y la compra en salud (variable independiente interesante debido a que marca una relación temática entre rubros).

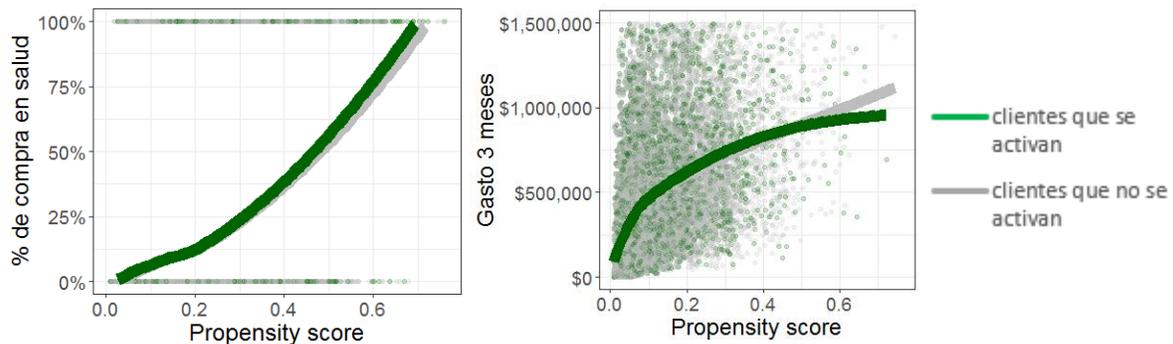


Figura 25: Verificación gráfica del balance de la compra pasada en salud y el gasto pre-activación a distintas probabilidades de activación en farmacias.

Fuente: Elaboración propia.

A partir de esta verificación, se da paso a la estimación del efecto sobre el gasto y la rentabilidad (del uso de la tarjeta) de la activación en cada rubro. Para el caso del rubro farmacias, verificamos que la diferencia estimada en gasto entre los clientes que se activaron y los que no es de \$193.263 (diferencia significativa con un 95% de confianza), este resultado es complementado con un enfoque de diferencia en diferencias (dado que en el periodo pre-activación ya existía una diferencia de \$70.381), concluyendo que la diferencia neta en gasto por activarse en farmacias es de \$122.882 como se observa en la Figura 26.

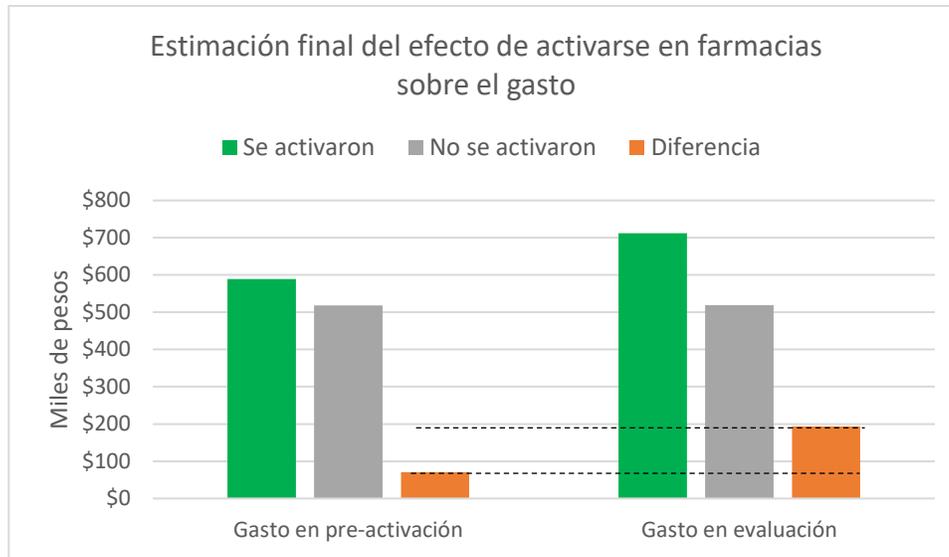


Figura 26: Estimación final del efecto de activarse en el rubro farmacias sobre el gasto del cliente.
Fuente: Elaboración propia.

Para reafirmar el hecho del aislamiento del efecto de activar en farmacias, se comparan en la Tabla 7 los gastos en el periodo previo a la activación de los clientes antes y luego del emparejamiento, corroborando que en los clientes emparejados la diferencia es bastante menor que antes de realizar el procedimiento. En el Anexo P, se encuentran los valores de los efectos para el resto de los rubros.

Tabla 7: Resultados estimación del efecto causal de la activación en farmacias sobre el gasto del cliente.

Variables	Estado	Promedio en los que se activan	Promedio en los que no se activan	Diferencia
Gasto en periodo pre-activación	Antes del emparejamiento	\$588.945	\$309.697	\$279.248
	Después del emparejamiento	\$588.945	\$518.564	\$70.381

Una vez realizado este procedimiento para los distintos rubros de gestión, se construye la Figura 27, donde se aprecia que los rubros que al ser activados provocan un mayor aumento en el gasto son Viajes y Recaudación con una diferencia de \$233.868 y \$143.115, respectivamente.

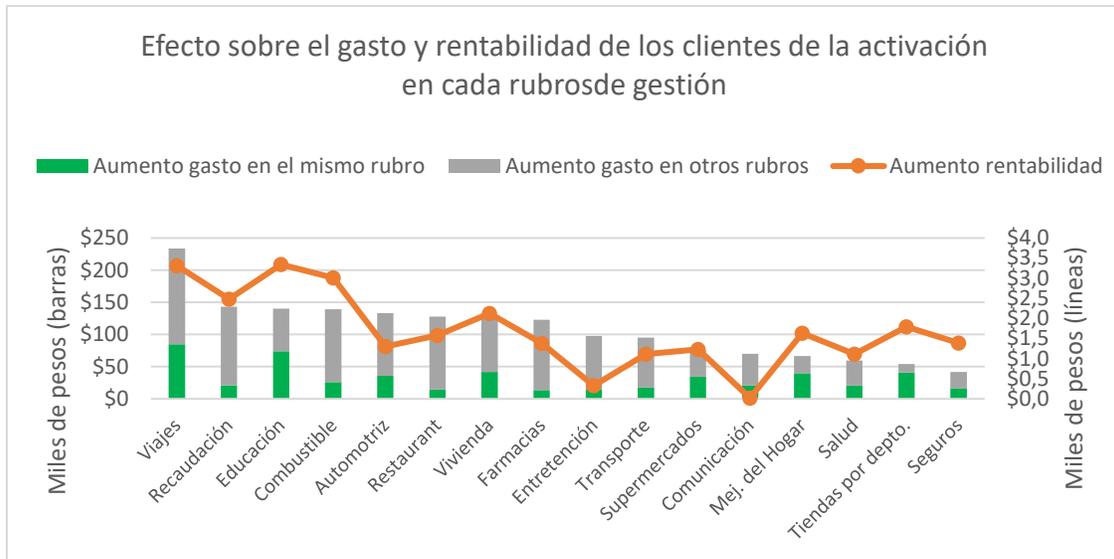


Figura 27: Efecto estimado de la activación de cada rubro de gestión sobre el gasto y rentabilidad de los clientes.

Fuente: Elaboración propia.

Por contrapartida, los rubros que al activarse generan una menor diferencia son tres de los que poseen una mayor cantidad de gasto de los clientes y con fuerte presencia del grupo (Tiendas por departamento, Mejoramiento del hogar y Seguros), junto con Salud, todo ellos con un aumento del gasto en los 3 meses posteriores a la activación menor a \$60.000.

Es importante resaltar, que los efectos mostrados, además de ser positivos son significativos para todos los rubros (con una confianza del 95%), exceptuando a Seguros. Con lo que se concluye la existencia de un efecto causal de activarse en un rubro sobre el gasto futuro y la rentabilidad por el uso de la tarjeta.

También, se ve en la Figura 27, que del gasto que se aumenta al activarse en un rubro, en la mayoría de los casos este aumento no viene dado mayoritariamente por gastos en el mismo rubro activado, sino que en una mezcla con los demás, lo que recalca existencia de efectos cruzados estudiados con los logits.

Con respecto a la rentabilidad, esta no sigue el mismo comportamiento del gasto, de hecho, los 3 rubros con presencia del grupo recién mencionados por su bajo efecto sobre el gasto de los clientes dejan de ser los de menor diferencia en rentabilidad, debido principalmente a la mayor utilización de compras con cuotas en y que el *merchant* es mayor en ellos. A pesar de lo anterior, los cuatro rubros que provocan mayor aumento en el gasto también lo son para el aumento de rentabilidad (aunque no en el mismo orden), por lo que se declaran rubros relevantes y que deberían ser potenciados para el beneficio de la empresa.

10. MEJOR PRÓXIMO RUBRO

En el siguiente capítulo, se muestran los resultados de los dos modelos de propensión escogidos para la asignación del mejor próximo rubro a ofertar según las características y comportamiento histórico de los clientes: modelo logit y modelo probit multivariado.

10.1 MODELO LOGIT

Los modelos logits corresponden a los mismo estimados para el cálculo del *propensity score*, por lo que el detalle de su estimación se puede ver en el punto 8.1.

Retomando el ejemplo del rubro farmacias, en el que el valor del AUC en la muestra de entrenamiento (77%) y en la muestra de testeo (77%) son cercanos (apreciables en la Figura 28 y junto a otras métricas predictivas en el Anexo N), lo que ocurre para todos los rubros, se verifica que no hay problemas de sobre ajuste.

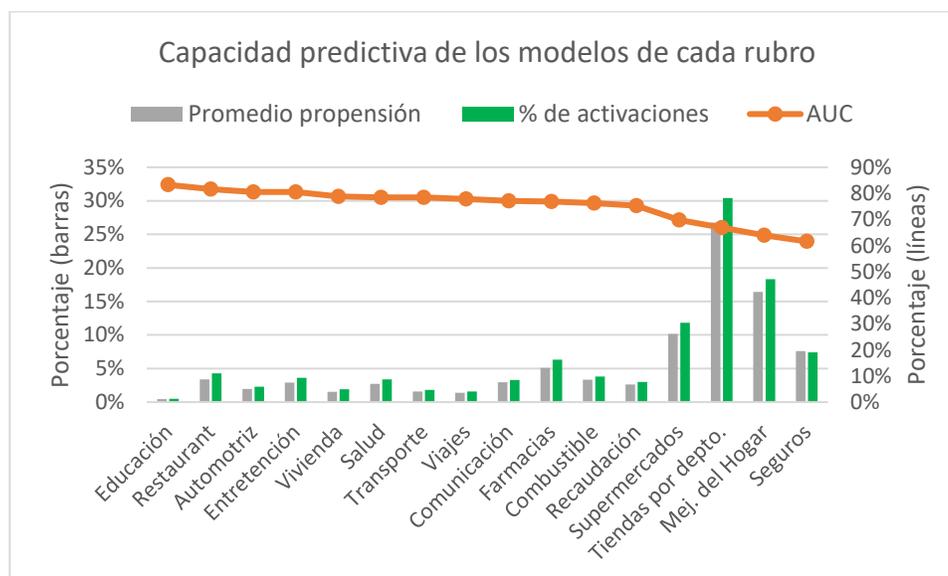


Figura 28: Capacidad predictiva de los modelos logit estimados para cada rubro en la muestra de testeo. Fuente: Elaboración propia.

De la misma manera, en la Figura 28, se observa que el AUC se ubica por lo general sobre un 70%, teniendo sus menores valores en los cuatro rubros con mayor porcentaje de activación (Supermercados, Tiendas por Departamento, Mejoramiento del Hogar y Seguros), lo que se verifica a través de la diferencia entre el promedio de la propensión y el real porcentaje de activaciones, que en todos los casos está siendo subestimado.

Por otra parte, respecto de los coeficientes estimados de cada variable para los distintos modelos (ver en Anexo M), se puede destacar que el intercepto es el coeficiente de mayor módulo y es negativo, lo que indica que es bastante improbable activarse en el rubro farmacias dado que no he realizado compras en el pasado reciente (este hecho se repite en los distintos rubros, con menor fuerza en los rubros que poseen una gran cantidad de compras como Tiendas por Departamentos).

Para analizar los efectos de las distintas variables sobre la activación en los distintos rubros se utilizan sus efectos marginales promedios, apreciables en su totalidad en el Anexo M.

Tomando una vez más como ejemplo el rubro Farmacias, se observa en primera instancia con las variables no ligadas a efectos de rubros, que el tipo de tarjeta y el enganche por puntos poseen coeficientes similares, siendo el más importante la tenencia de una tarjeta abierta premium (correspondiente a un tipo especial de las tarjetas abiertas con mayores beneficios y asociadas a clientes con mayores gastos) que aumenta marginalmente la probabilidad de compra en un 3,24%, aunque no es tan grande como en los rubros Restaurant (5,51%) y Salud (5,31%).

Siguiendo, con respecto a la actividad durante el último mes en otros rubros, se observa que todos tienen un efecto positivo en la probabilidad de activación. Esto se condice con lo esperado de que los clientes que mientras más rubros esté utilizando un cliente, mayor principalidad de la tarjeta, por ende, es más fácil que la pueda utilizar en un nuevo rubro.

Asimismo, se observa que los rubros más potentes en efecto activación son Supermercados (4,33%) y Salud (2,93%), de los cuáles el primero se repite a través de los distintos rubros (se detalla este hecho más adelante) y Salud está indicando una asociación temática de rubros relacionados intrínsecamente.

Por último, pasando al análisis de la estacionalidad en el rubro, se observa que consistente con el periodo de mayor gasto en Farmacias, agosto es el mes que más aumento en la probabilidad de compra genera (1,56%).

Con estos resultados se divide el análisis en tres puntos principales: poder explicativo de las variables, sensibilidad de los rubros y asociaciones temáticas.

Primero, comenzando con las variables no ligadas a la actividad en rubros, se observa que hay dos coeficientes de igual magnitud al intercepto en la mayoría de los casos: tarjeta abierta clásica y la tarjeta abierta premium, aumentando siempre la probabilidad de activarse en los rubros, siendo su efecto más potente en los rubros sin presencia del grupo.

En primera parte, se concluye que la tenencia de una tarjeta no cerrada eleva transversalmente la probabilidad de activarse en un rubro, teniéndose como causa de este resultado que claramente existe una limitación para los clientes con la tarjeta cerrada y que por contrapartida los clientes que poseen tarjetas abiertas premium son, en general, clientes que le dan mayor uso que se transforma en mayores probabilidades de activación.

En la misma línea, pero con menor poder, se encuentra el hecho de que el cliente sea enganchado por los puntos del club de fidelización⁹ y el haber cambiado la tarjeta en el último mes. La primera variable va fuertemente relacionada con un comportamiento conocido en el negocio de que los clientes con una alta tasa de canje de puntos amplían el uso de la tarjeta para la acumulación de estos.

También, con respecto a la actividad pasada en otros rubros, se evidencia que existen cuatro rubros que apalancan el uso de la tarjeta en nuevos rubros de manera general, estos son Restaurant, Farmacias, Combustible y Supermercados con un aumento en la probabilidad de compra promedio de 1,8%, 1,7%, 1,6% y 1,5%, respectivamente. Mientras que, por contrapartida, se observa que rubros como Recaudación o Seguros no marcan una tendencia hacia la activación de clientes, siendo sus efectos promedio en la probabilidad de activarse en otros rubros estadísticamente nulos.

En segundo punto, se establece que existen rubros sensibles a la actividad en otros rubros verificados a través de los coeficientes estimados, siendo los mayores Supermercados (con un promedio 0,39) y Restaurant (0,33), que aumentan de gran valor su probabilidad de activación a medida que más rubros activos se tuvo en el pasado. Contrario a lo anterior, se vislumbra que mejoramiento del hogar (0,17) no ve fuertemente afectada su probabilidad de activación por la actividad pasada en los demás rubros. Este comportamiento, da indicios de la existencia de distintos niveles de rubros, en los que algunos para ser activados necesitarán de un mayor nivel de madurez respecto de su uso de la tarjeta de crédito.

Finalmente, se observa que existen asociaciones temáticas en la activación de rubros relacionado directamente a que existe un comportamiento natural en el hecho de realizar compras en ambos en periodos consecutivos. Por ejemplo, ya se estipuló la relación de compra en Salud (compuesto por clínicas, hospitales y otros servicios de salud) y en Farmacias, asimismo se observa ese comportamiento con el rubro Automotriz y el de Combustibles o Restaurantes con Entretención.

⁹ Se refiere al hecho de que el cliente ha gastado al menos el 90% de los puntos acumulados en el club de fidelización durante el último año.

10.2 MODELO PROBIT MULTIVARIADO

Los modelos probit multivariados que se desarrollan en este capítulo tienen dos grandes objetivos: primero, establecer un orden o ranking en que se realiza la activación en los rubros de gestión; y segundo, incorporar en la estimación el hecho de que existe coocurrencia en las compras de los rubros, es decir, que existe relación entre los rubros que se compran o no en un mismo periodo.

Para lograr lo anterior, se seguirán dos líneas de modelamiento. Primero, estimar un modelo que capture comportamientos similares a los modelos logits anteriores (modelo de efectos específicos), en el cual cada rubro afecta de distintas maneras la compra en los diferentes rubros. Y, en segundo lugar, establecer un modelo que permita describir la activación de los distintos rubros en un orden (modelo de ordenamiento).

Es importante mencionar que, dadas las restricciones de capacidad del software y hardware utilizado, junto a la complejidad computacional de este tipo de modelamiento, se debió seleccionar una muestra aleatoria de 1.500 clientes y sus compras durante un año, separándose igual que para los logits en un 80% para estimación y un 20% para validación.

10.2.1 Modelo de efectos específicos

El primer paso para la estimación de los modelos es la definición de la utilidad latente. En este caso, se busca reflejar una hipótesis de comportamiento en que cada rubro posee una probabilidad intrínseca de ser comprado en cada periodo y que cada rubro afecta de manera diferente la compra futura de los demás, sin haber una relación clara entre estos dos aspectos.

Además, se agregan como covariables adicionales el tipo de tarjeta del cliente y si está enganchado por puntos debido a la probada relevancia de estas características a través de los modelos logits anteriores, permitiendo que estas afecten a todos los rubros de manera diferenciada. Con esto la utilidad observada que presenta la compra del rubro j para el cliente i en el periodo t queda definida como en (21).

$$V_{ijt} = \alpha_j + \sum_{k=1}^j \beta_{kj} \cdot D_{ikt-1} + \theta_{1j} \cdot T_{visa\ classica} + \theta_{1j} \cdot T_{visa\ premium} + \theta_{1j} \cdot EngPuntos \quad (21)$$

Esta forma funcional de la utilidad se estima en primer lugar de la forma descrita en la metodología como logit bayesiano, para luego ser estimada en forma de probit multivariado con el objetivo de verificar el efecto agregado por el modelamiento de la correlación en la compra en distintos rubros en el mismo

periodo. En específico, los *priors* definidos para los parámetros de la utilidad observable son poco informativos, básicamente normales centradas en 0 con su correspondiente desviación estándar. En tanto, para la matriz de correlaciones (para que el modelo sea identificable) de la normal multivariada que rige la variable dependiente del modelo probit, se escoge una distribución LKJ sobre la descomposición de Cholesky de esta, con parámetro 1 (también poco informativo).

Lamentablemente, estas dos especificaciones no son comparables por WAIC debido a que poseen una cantidad distinta de puntos estimados. Por lo que se verifica el beneficio de incluir la coocurrencia de compra a través de las curvas de ganancia que se muestran en la Figura 29. En ellas se verifica que el modelo probit multivariado tiene un mejor, aunque no bastante superior, rendimiento predictivo, por lo que ahora nos centramos en el estudio de sus resultados.

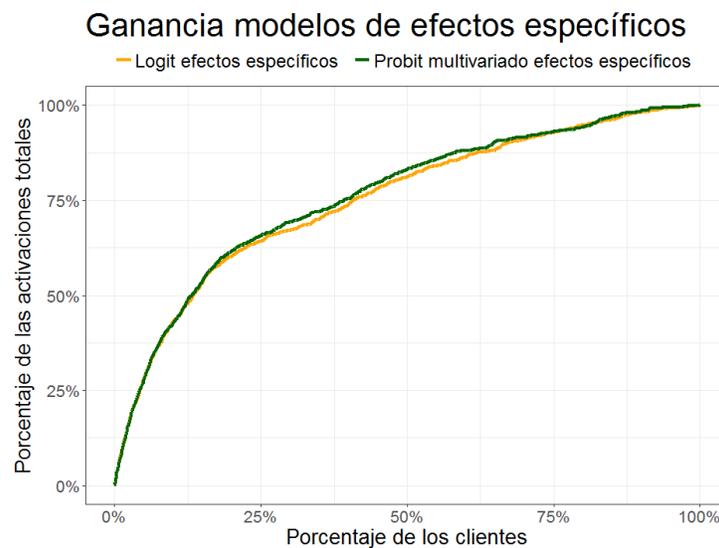


Figura 29: Comparación de curvas de ganancia de los modelos de efectos específicos sobre las activaciones. Fuente: Elaboración propia.

A grandes rasgos, los coeficientes estimados siguen comportamientos similares a los vistos con los logits anteriores. Por ejemplo, los interceptos (α_j) son en general los coeficientes de mayor módulo y negativos, también, se mantiene la relevancia del tipo de tarjeta que tiene el cliente siendo más fuerte en los rubros menos comprados como Viajes o Automotriz (ver detalle en Anexo Q).

Un hecho diferente, es que en este caso se estiman los coeficientes asociados a la repetición de compra en los rubros cuyos valores son largamente mayores a los efectos cruzados entre rubros, lo que corrobora la importancia de incluir la tenencia actual de productos como predictor de la que hablan (Manchanda, Ansari, & Gupta, 1999), a pesar de que estos no tendrán incidencia en la predicción de activaciones es importante aislar su efecto.

El aspecto diferenciador de este modelo respecto de los modelos logits ya estimados es la incorporación de la coocurrencia de compra entre rubros, la que se ve representada en la estimación de la matriz de correlaciones de las variables dependientes (sus coeficientes estimados se ven en el Anexo R), en la cual se aprecian correlaciones de hasta 0,36 (obviando la correlación de 1 se las variables con sí mismas).

De la coocurrencia de compra se destacan tres resultados. En primera parte, se corrobora que algunas de las relaciones en la compra de rubros ya revisados no sólo ocurren entre periodos, sino que también en el mismo mes, lo que aplica para lo visto entre Combustible y Automotriz (correlación de 0,25) y la relación entre Salud y Farmacias (correlación de 0,3).

En segundo lugar, las mayores correlaciones se ven en Supermercados con Combustible (0,36), Farmacias (0,32) y Restaurant (0,33). Estos tres últimos rubros tienen la similitud de tener alianzas importantes con la empresa y ser objeto frecuente de campañas. Aunque, se podría haber esperado una mayor correlación entre los rubros íconos del grupo (Tiendas por Departamento y Mejoramiento del Hogar) y Supermercados o las alianzas, lo que no ocurre, por lo que se agregan evidencias de que Supermercados en particular es un rubro que acarrea compra en los demás.

En tercer lugar, se observan rubros que no parecen tener coocurrencia de compra con los demás, ejemplos de esto son Recaudación o Comunicación. Es particular el caso de Recaudación, del cual se había estimado que su efecto sobre el gasto futuro de los clientes era importante, por lo que se puede concluir que su efecto es principalmente hacia meses futuros y no en el actual.

10.2.2 Modelo de ordenamiento

Este modelamiento, busca representar la existencia de un orden o jerarquía en el cual los clientes van adquiriendo rubros. Para ello, se decanta por un modelo que tiene como hipótesis que los clientes realizan una especie de acumulación de rubros en la cual existe una mayor probabilidad de agregar un nuevo rubro cuando ya se compraron los que están por debajo de este en la escala.

Además, se asume que cuando el cliente comienza a comprar los rubros que se ubican en los últimos lugares de la escala de adquisición de rubros es reflejo de que este alcanza una mayor madurez, por lo tanto, estos rubros elevan de mayor manera la probabilidad de comprar los demás rubros.

El comportamiento anterior, se refleja en la ecuación (22) a través de los coeficientes β , los cuáles no poseen restricciones en sus posibles valores, ni tampoco están restringidos a condiciones especiales entre ellos como, por ejemplo, especificar a priori que alguno de los rubros tiene el menor coeficiente. Asimismo,

se incorpora el parámetro de pendiente γ del que más adelante se detallará su utilidad y decisión de inclusión. También, debido a la corroborada persistencia de compra verificada en el modelo de efectos específicos y la importancia de aislar los efectos propios de los de activar otros rubros se incorpora el parámetro δ . Finalmente, se agrega nuevamente el efecto del tipo de la tarjeta del cliente y su enganche con el sistema de puntos. Con respecto a los *priors* se utilizan los mismos mencionados para el modelo anterior.

En particular, el coeficiente β captura el efecto deseado en caso de ser positivo, debido que en un hipotético caso negativo describiría un rubro cuya probabilidad de compra es bastante alta de forma natural y que disminuye la probabilidad de que en el futuro se compren los demás rubros, lo que resulta contra intuitivo para el caso de la empresa. Para solucionar esto se realizaron distintas modificaciones, las que perjudicaban la convergencia del modelo, así que dado que de la estimación todos los coeficientes resultaron positivos con un 95% de confianza se decidió no modificarlo.

$$V_{ijt} = -\beta_j + \gamma_j \sum_{k \neq j} \beta_k \cdot D_{ikt-1} + \delta_j \cdot D_{ijt-1} + \theta_{1j} \cdot T_{visa\ classica} + \theta_{1j} \cdot T_{visa\ premium} + \theta_{1j} \cdot EngPuntos \quad (22)$$

De forma general, el parámetro γ captura la fuerza con que la actividad en otros rubros afecta la probabilidad de compra en el rubro estudiado. Pero, debido a que no es claro aún si existe el comportamiento declarado, se realiza un análisis de sensibilidad respecto a este coeficiente utilizando el formato de logits bayesianos. Se comparan cuatro escenarios para el coeficiente γ , primero que su valor es 0 lo que significa que no existe influencia sobre la compra de un rubro el hecho de haber comprado otros en el pasado. En segundo lugar, se asume que existe el efecto descrito anteriormente, lo que se refleja en un γ igual a 1. Tercero, se sigue el caso anterior en el que existe influencia entre rubros, pero se permite que el poder de este sea definido por el modelo. Finalmente, se permite que el parámetro γ sea diferente para cada rubro, lo que lograría capturar el hecho de que no todos los rubros se ven afectados con la misma fuerza por la compra pasada en los demás rubros.

Las cuatro especificaciones de γ se ven en la Tabla 8, junto al WAIC asociado a cada modelo y la comparación de este respecto a la cuarta especificación (debido a que esta es la de menor WAIC), verificándose debido a la diferencia positiva calculada que la formulación 4 es la preferida utilizando este criterio de información, siendo también la de mayor verosimilitud.

Tabla 8: Evaluación del WAIC y comparación de las distintas especificaciones del modelo logit de ordenamiento.

Especificación de γ_j	WAIC	Diferencia epld-WAIC (s.e.)	Log-verosimilitud
1) $\gamma_j = \mathbf{0}$	43.052,8	392,2 (31,5)	-17.187.458
2) $\gamma_j = \mathbf{1}$	51.138,6	4435,1 (82,2)	-20.418.571
3) $\gamma_j = \gamma$	42.374,4	53,1 (13,1)	-16.915.245
4) $\gamma_j = \gamma_j$	42.268,3	0 (0)	-16.865.166

A continuación, se estima esta especificación en el formato de probit multivariado para incorporar la coocurrencia de compra. Verificando, a través de las curvas de ganancia de la Figura 30, que este modelo posee un mejor poder predictivo (al igual que para los modelos de efectos específicos la diferencia no es sustancial) sobre las activaciones de los clientes en los distintos rubros, por lo que se detallan sus resultados.

Ganancia modelos de ordenamiento

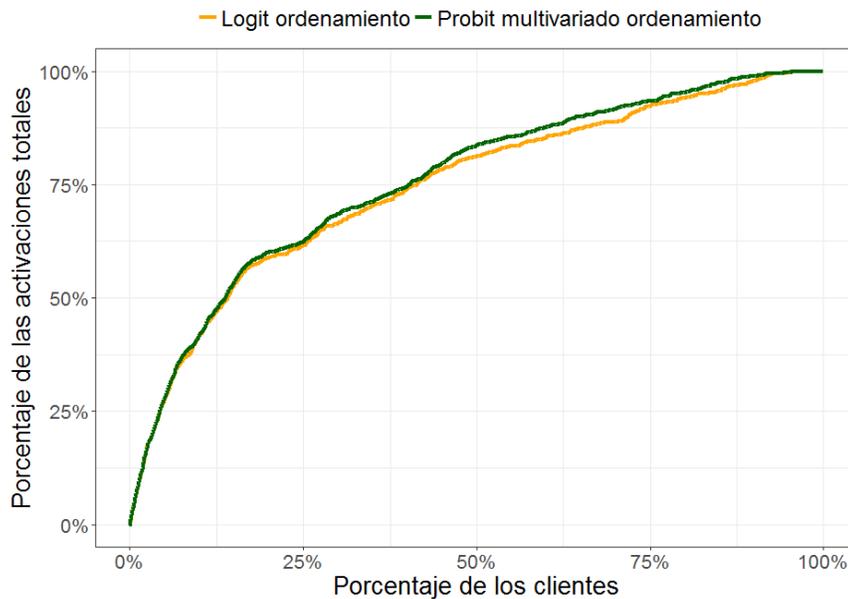


Figura 30: Comparación de curvas de ganancia de los modelos de ordenamiento sobre las activaciones.
Fuente: Elaboración propia.

Lo primero analizable de este modelo es su matriz de correlaciones entre variables independientes (ver en Anexo S) es consistente y sigue un comportamiento similar al observado en la matriz de correlaciones del probit multivariado de efectos específicos.

Ahora bien, con respecto a los coeficientes β estimados, estos sugieren que existe un orden estocástico (puesto que no se estableció ningún orden *a priori*) en que los rubros se van comprando o acumulando. También, vemos que bajo esta especificación se observa de forma general un orden lógico de los rubros en que respeta las proporciones en que se compra cada uno, situando en su inicio los rubros más comprados y hacia el final los que lo hacen en menor medida.

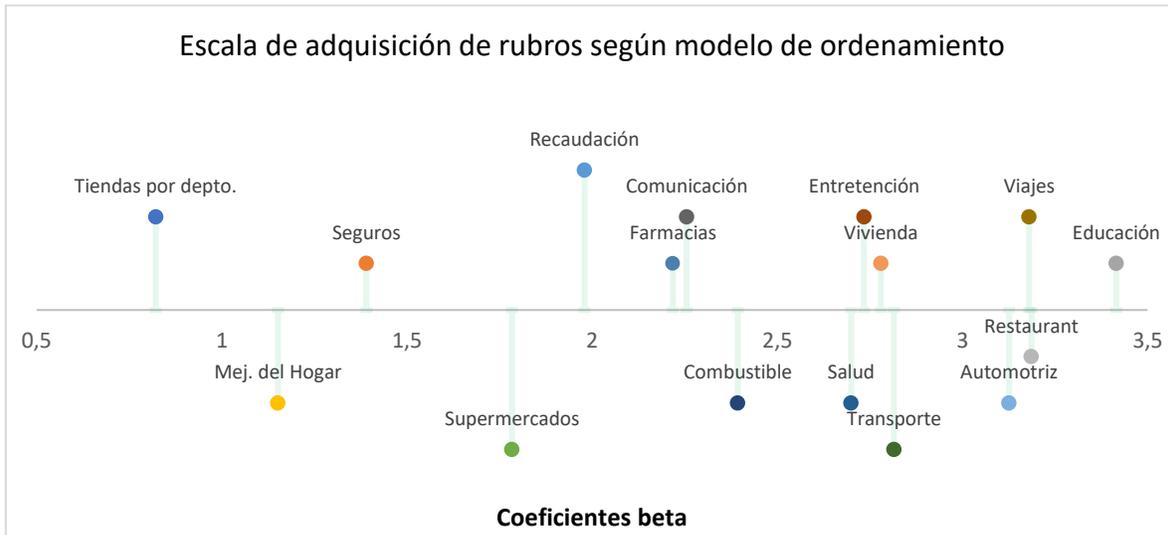


Figura 31: Escala de adquisición de rubros estimado según el modelo probit multivariado de ordenamiento. Fuente: Elaboración propia.

A partir de los valores estimados para los coeficientes de orden β_j que se muestran en la Figura 31, se infiere que la forma en que se acumularían o adquirirían rubros comienza con los rubros íconos del grupo, siendo el que más aporta a la madurez de compra del cliente Supermercados, en concordancia con lo anteriormente analizado.

A su vez, los parámetros de pendiente γ_j muestran que el efecto de la acumulación de rubros está a una escala mucho menor que la dificultad de activación, con un promedio de 0,05. Pero, marcando que hay rubros que se ven mucho más impactados con la compra de los demás como lo son Restaurant (0,1) y Supermercados (0,08), con coeficientes hasta 9 veces superiores a los de Recaudación (0,01) o Seguros (0,01).

Por último, en cuánto a los parámetros que captan la persistencia en la compra, se observa una especie de relación negativa con los coeficientes anteriores en la que, para los rubros que se ven menos afectados por la acumulación de otras compras (Recaudación y Seguros) se aprecian mayores valores de repetición de compra en meses consecutivos (con coeficientes 3,1 y 3,3 respectivamente). Por otra parte, los rubros que presentan menor coeficiente de repetición de compra son

Tiendas por Departamento y Mejoramiento del Hogar, influenciados por su gran cantidad de compras puntuales (los coeficientes están en el Anexo T).

10.2.3 Elección de modelo predictivo

Finalmente, se comparan las distintas formas de modelamiento vistas según su poder predictivo sobre las compras que representan activaciones en la muestra de evaluación.

En la Figura 32, se verifica que los modelos logits tradicionales¹⁰ se encuentran levemente por debajo de los modelos probit multivariados, los cuáles se encuentran prácticamente empatados en términos de ganancia.

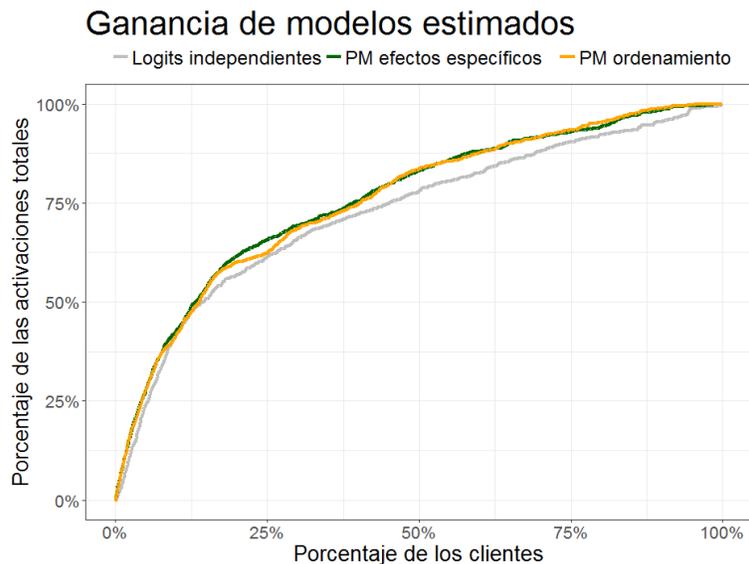


Figura 32: Comparación de las curvas de ganancia de los modelos predictivos seleccionados.
Fuente: Elaboración propia.

Para escoger entre los modelos probit, utilizamos el WAIC, inclinándonos por el modelo de ordenamiento (1.005.192) por sobre el de efectos específicos (938.851), cuya diferencia es negativa en el elpd-WAIC de -33.170 (con un error estándar de 198.7) a favor de la primera especificación. También, verificando esta elección, se encuentra que el logaritmo de la verosimilitud es mayor para el modelo de ordenamiento (con un valor de -365.745.994 en comparación al -381.269.884 del modelo de efectos específicos).

¹⁰ Esta vez los modelos logits tradicionales (los utilizados para la estimación del *propensity score*) fueron estimados sobre la misma base que los modelos bayesianos para poder compararlos en iguales condiciones.

11. CRITERIOS DE ASIGNACIÓN Y USOS POSIBLES

Debido a la cantidad de resultados, tanto finales como intermedios, de este trabajo, existen diversos usos posibles de ellos por parte de la empresa, de los cuáles se abordarán dos: la selección del próximo mejor rubro para un cliente y la selección de clientes para campañas.

En primer lugar, nos enfocamos en el caso en el que se desea decidir para cada cliente, ya sea toda la cartera o un grupo de esta, cuál es el mejor rubro en el que no está activo que se le podría ofrecer, cuestionamiento principal de este trabajo.

Para definir eso, dado el análisis anterior, se selecciona el modelo probit multivariado de ordenamiento como el más acorde para realizar predicciones de activaciones. De calcular las probabilidades de activación para todos los clientes de la cartera para el mes de octubre de 2017 podemos extraer, por ejemplo, que existen rubros en los que el modelo no indica un alto potencial de activación de forma natural, es decir sin una intervención oportuna como una oferta por parte de la empresa, dado que no existe ningún cliente con una probabilidad mayor a un 50%. Estos rubros son Automotriz, Comunicación, Educación, Mejoramiento del Hogar, Recaudación, Seguros y Vivienda, a los que la empresa debería asignar actividades de marketing para poder lograr una cantidad de activaciones considerables. Por contrapartida, de los rubros en que se esperan activaciones, destacan Tiendas por Departamento y Entretención, rubros en los cuáles se espera que de forma natural se activen un 12% y un 1% de la cartera total de clientes, respectivamente. Es así como, en un mes en particular, se estima que un 14% de los clientes tiene altas probabilidades de activarse en algún rubro, lo que repercutiría en un aumento del gasto de estos clientes en aproximadamente 23 mil millones de pesos durante los próximos tres meses, que se traducen en 685 millones de pesos de ingresos (rentabilidad de las compras con la tarjeta) para la empresa, siendo estos valores un 20% y 36% superior a los de tres meses antes de la activación por parte de estos clientes, respectivamente.

Entonces, lo que sigue es encontrar una regla que nos permita equilibrar la propensión de los clientes a activarse y el aumento de gasto que esto genera en los clientes para incrementar las ganancias de la empresa. Para evaluar esto, se realiza la estimación de las probabilidades a activarse de los clientes durante el mes de agosto de 2017 para todos los clientes de la muestra utilizada. Para luego, definir a través de 3 reglas (mayor propensión, gasto esperado y lift) cuál dentro de los rubros no activos de los clientes es el mejor a ofrecer. Terminando, con el cálculo de la ganancia esperada que se obtiene al utilizar cada una de estas reglas, asumiendo que la persona que se activa en el rubro ofrecido incrementará su gasto según lo estimado en el punto 9.

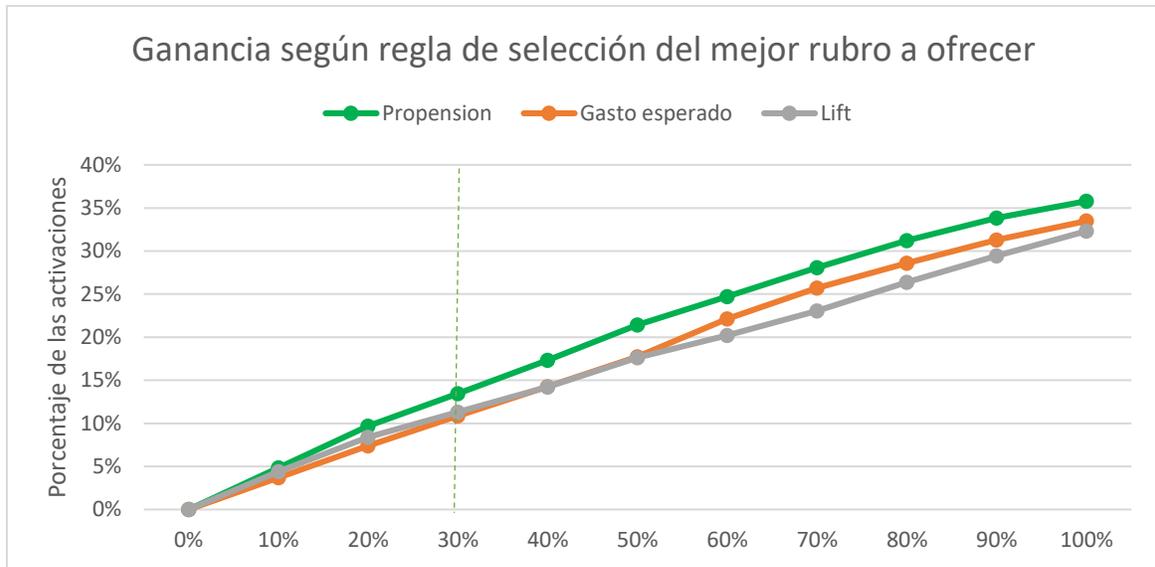


Figura 33: Evaluación de ganancia sobre las activaciones en el rubro ofrecido según regla de selección. Fuente: Elaboración propia.

Así, se observa en la Figura 33, una adaptación de las curvas de ganancia para este caso, en la que se ordenan los clientes junto a la regla que define su mejor rubro a activar de forma descendente, para evaluar la cantidad de activaciones que se capturan al tomar cada percentil según el criterio utilizado.

En un primer caso, en el que se decide ofrecer el determinado como mejor próximo rubro a todos los clientes de la muestra, corroboramos que en promedio las reglas estudiadas cubren un tercio de las activaciones de los clientes. En cuanto a la precisión, el criterio ganador es la propensión (36%), seguidos del gasto esperado (33%) y el lift (32%). En cuanto a los retornos percibidos se da en el orden contrario, con el lift en primer lugar (MM\$75), luego el gasto esperado (MM\$68) y la propensión (MM\$63). Ahora, si se decide sólo tocar, por ejemplo, al 30% de los clientes mejor evaluados según la regla por temas de capacidad de comunicación u otras restricciones de la empresa, se mantiene el comportamiento anterior con su respectivo aumento en precisión 45%, 38% y 36% de la propensión, el gasto esperado y el lift, respectivamente.

Este comportamiento se explica debido a que, a pesar de la pérdida de precisión que se tiene al elegir métricas distintas a la propensión, las otras permiten en mayor medida la asignación de rubros más complicados de activar y que también tienen retornos mayores, como Viajes o Recaudación, a diferencia de la propensión que privilegia los rubros más fáciles de activar como Tiendas por Departamento o Supermercados que son de los que menor gasto futuro implican.

Por ello, se concluye que cuando el único objetivo es mejorar el retorno para la empresa la mejor opción sería utilizar como criterio el lift, pero cuando existen costos importantes de contactar a una gran cantidad de clientes se puede pasar a utilizar el gasto esperado (y en peores casos la propensión) debido a su mayor precisión.

Continuando con el criterio del lift, se estipula que a pesar de que el objetivo del modelamiento NBO es determinar la mejor oferta según las características específicas del cliente, los grupos de clientes propensos a un mismo rubro comparten características comunes. Para ejemplificar, los clientes a los que se les asigna el rubro Educación (el último según el ordenamiento estimado por el modelo) son los que gastaron más durante el último año con su tarjeta de crédito (con un promedio de MM\$14 sobre el promedio de la cartera de MM\$1,3) y tienen mayores cupos (176% mayor al promedio de la cartera), además de tener una mayor proporción de clientes con correo registrado válidos (85%), una alta tasa de canjeadores de puntos del club de fidelización (81%) e ingresos a la página web de la empresa (82%) durante el último año, configurando un perfil maduro en el uso de su tarjeta. Mientras que, los clientes a los que se les asignan los primeros rubros de la secuencia (Tiendas por Departamento, Mejoramiento del Hogar y Seguros) cumplen con las características esperables de clientes con un menor grado de relación con la empresa (siendo también los grupos más populosos), con menores gastos (menores a MM\$1) y cupos (bajo MM\$1,2), una baja tasa de clientes con canjes (6%) e ingresos a la web (13%), además de poseer una última compra más distante en el tiempo (el detalle de estos valores se puede apreciar en el Anexo U).

El segundo uso, como herramientas de selección de los mejores clientes para cierta campaña o rubro, sigue el enfoque más utilizado en la empresa, por eso se compara contra una campaña vía e-mail que se realiza mes a mes para activar a los clientes en ciertos comercios del rubro Restaurant, que ofrece un descuento sobre el total de la boleta un día en específico de la semana, escogida por su objetivo activador, la disponibilidad de datos y la sensibilidad del rubro respecto de compra en los demás vista en los modelos de propensión estimados.

En particular se evalúan las campañas de los meses junio, julio y agosto de 2017, en las cuáles el grupo objetivo fueron los clientes que no habían comprado antes en los comercios asociados a esta misma campaña, de las que podemos ver la composición del grupo contactado y de control y sus tasas de respuesta en la Tabla 9.

Tabla 9: Composición de las campañas vía e-mail para la activación de los comercios asociados del rubro Restaurant.

Mes campaña	Grupo	Clientes	% respuesta	p-valor test de proporciones
Junio	Mail	60.535	0,13%	0,4846
	Control	6.764	0,09%	
Julio	Mail	162.644	0,18%	0,0392
	Control	18.214	0,11%	
Agosto	Mail	142.873	0,13%	0,0028
	Control	16.026	0,04%	
Total	Mail	366.052	0,15%	0,0005
	Control	41.004	0,08%	

En promedio, la tasa de respuesta incremental de estas campañas fue de un 0,07%, un porcentaje bastante bajo pero significativo, explicado básicamente por su carácter activador. A pesar de este objetivo, al estudiar la compra pasada de los clientes que se activan en estos restaurantes, se corrobora que un 48,9% ya había comprado en restaurantes por lo que en gran medida la campaña provocó un efecto de incorporación de nuevos restaurantes o cambio de ellos y no una activación en el rubro.

Para verificar lo observado, se realiza un modelo logit de respuesta a la campaña, con interacciones de primer orden entre haber recibido el mail y las demás variables independientes. Este modelo se construye mediante un proceso de *stepwise* dónde sólo se mantienen las variables o interacciones que tienen efectos significativos, terminando con la configuración y resultados del Anexo V.

El resultado más relevante de los modelos de respuesta estimados es que por sí sólo el haber recibido el correo de la campaña tiene un efecto positivo sobre la activación en los restaurantes de la campaña. Pero, al agregar la interacción de pertenecer al grupo mail con el haber comprado en otros restaurantes durante los últimos tres meses (única interacción que sobrevive al proceso de *stepwise*), el recibir el correo pasa a no tener un efecto significativo y sólo se mantiene su efecto significativo en los clientes ya activos en restaurantes.

Asimismo, se verifica que un cupo sobre \$700.000 que es un criterio de selección que se mantiene a través del tiempo para esta campaña tiene un efecto positivo significativo en la respuesta, al igual que tener tarjeta abierta premium, que está altamente relacionado con ser un cliente premium, que es el otro criterio que se utilizó en las tres campañas¹¹.

Dado lo anterior, se pasa a verificar cómo variaría la activación en el rubro Restaurant de pasar del criterio de selección utilizado por la empresa para el envío del correo con seleccionar la misma cantidad de clientes de mayor propensión a activarse en restaurantes según el modelo probit multivariado de ordenamiento.

En esta comparación, se logra verificar que mientras el criterio de selección de la empresa, basado en el cupo del cliente, su categoría y la tasa de lectura a correos anteriores tiene una precisión de un 6,3%, mientras que utilizando la asignación del modelo se obtiene una precisión del 10%.

Esta diferencia de 3,7% en la precisión sobre los 302.160 clientes seleccionados, repercute en que los clientes seleccionados por el modelo de propensión gastaron MM\$201 más en restaurantes que los seleccionados por el criterio de la empresa en el siguiente mes del que fueron escogidos.

¹¹ Cabe señalar que, dado que la oferta no es exclusiva del formato e-mail, se incluyó en la estimación como grupo de control un segmento de clientes que no cumplían con los criterios de selección para la campaña.

12. CONCLUSIONES

La situación de la empresa, respecto de su crecimiento de clientes requería un cambio de estrategia que fue enfocada en el concepto de principalidad, lograr que el cliente utilice su tarjeta de crédito como medio primario de pago.

El uso de la tarjeta en distintos rubros de comercios, la diversidad como medida de la principalidad del cliente, marca un objetivo claro, sustentado en que los clientes que utilizan su tarjeta en un rubro más que otro gastan en promedio \$374.614 más durante un año y mientras más rubros usan es mayor este efecto.

Es de allí que nace la pregunta de cuáles rubros conviene primero activar, debido a que no todos los clientes están dispuestos o propensos a utilizar su tarjeta en ellos, ni todos los rubros generan el mismo retorno para la empresa. Por ello, el objetivo general de esta memoria de título era elaborar un modelo predictivo y un criterio de asignación que permitiesen atacar esta decisión para cada cliente.

El primer objetivo para resolver este problema fue la estimación del efecto que tiene sobre el gasto y la rentabilidad de las compras de los clientes el activarse en cada rubro de gestión, a través del método de matching por distancia de Mahalanobis con calipers en sus propensity scores, que aseguró que la comparación del efecto se realizara sobre grupos de clientes con igual distribución en su propensión a activarse y el gasto que venían realizando.

Este método definió que existe un efecto causal positivo y significativo sobre el gasto y rentabilidad de los clientes en 15 de los 16 rubros, siendo Seguros la excepción. En particular, los rubros que provocaban un mayor aumento en el gasto en los tres meses posteriores a la activación son Viajes y Recaudación con \$233.868 y \$143.115, respectivamente. En tanto, analizando la rentabilidad de estos gastos, se mantienen como rubro relevante Viajes, añadiéndose Combustible como los dos rubros que aumentan la rentabilidad en más de \$1.000 mensuales. Por esto, se propone este trio de rubros como buenos objetivos finales para capitalizar los gastos de los clientes.

A continuación, el segundo objetivo fue construir un modelo predictivo para establecer las probabilidades de los clientes a activarse en cada uno de los rubros. Para este propósito, se evaluaron distintos modelos, imponiéndose el probit multivariado de ordenamiento, debido a su mejor curva de ganancia. Este modelo, posee dos elementos diferenciadores: Primero, a través de la especificación multivariada determinar coocurrencia de compra entre rubros, hecho que no significó un gran aumento en su poder predictivo, pero que fue escogido debido a la inferencia realizable de su estimación, corroborando la existencia de rubros que se compran en conjunto en mayor medida que los demás, dejando como claro ejemplo los rubros Combustible y Automotriz.

A pesar de lo anterior, también existen evidencias de que la activación de un rubro pueda afectar de manera negativa la activación de otro rubro, siendo dos casos relevantes los efectos negativos apreciados entre Transporte y Combustible y entre Tiendas por Departamento y Recaudación, corroborados con coeficientes negativos (y significativos al 95%) en los modelos que mostraban interacciones directas entre rubros.

Segundo, en base a su WAIC (un 7% menor al del modelo competidor) se corroboró que de los modelos estimados, el probit multivariado de ordenamiento es el que mejor representaba el proceso de compra de los clientes, por lo que se pudo determinar un orden de adquisición de los rubros que parte por los con fuerte presencia del grupo (Tiendas por Departamento y Supermercados) culminando con el rubro Educación.

A partir de los resultados anteriores, se verificó que el 14% de la cartera activa posee altas probabilidades de activar en al menos un rubro durante el próximo mes, lo que dados los efectos causales estudiados, repercutiría en un aumento de su gasto en un 20% y de la rentabilidad ligada al uso de la tarjeta en un 36%.

El tercer objetivo, destinado a definir con qué criterio sería mejor equilibrar las posibilidades de que el cliente active un rubro y el beneficio estimado para la empresa. De este punto, se concluyó que en el caso de que la empresa desee enfocarse en aumentar el retorno de la selección debería utilizar el lift, debido a que puede aportar hasta un 18% más de gasto por parte de los clientes que seleccionar sólo por la propensión del modelo predictivo. A su vez, se sitúa que la métrica de gasto esperado es la que da un mejor balance entre precisión y ganancia si es que no se quiere sacrificar en demasía la precisión de las campañas.

También, se evaluó la diferencia de utilizar el modelo propuesto en este trabajo en contra de los criterios de selección de la empresa en una campaña de restaurantes. Mediante esta, se corroboró que el modelo predictivo tiene un 3,7% más de precisión, lo que se traduce en capturar cerca de MM\$67 más por campaña debido a este cambio en la eficiencia de la selección.

En definitiva, se concluye que la metodología seguida en este trabajo logra dar respuesta a los objetivos propuestos, definiendo la importancia que tiene la selección precisa y personalizada de los rubros a ofrecer para un cliente.

13. LIMITACIONES Y TRABAJOS PROPUESTOS

A pesar de lo anterior, existe una serie de limitaciones en este trabajo, que dan paso a líneas de estudio que pueden perfeccionar los resultados obtenidos.

Primero, el trabajo fue realizado sobre clientes con una antigüedad de al menos un año en la empresa y utiliza su comportamiento pasado para definir las posibles futuras acciones a tomar sobre ellos. Por lo tanto, es baja la utilidad de los resultados obtenidos en el caso de los clientes nuevos que no poseen comportamiento histórico, por lo que se plantea como un trabajo complementario el poder determinar los drivers que desencadenan las primeras compras y el rubro en que se realiza.

Una segunda limitación, fue la no inclusión ni estimación de los efectos producidos por las acciones de marketing directo en las probabilidades de que los clientes se activen en un rubro. Esta limitante viene principalmente de la inexistencia, al momento de iniciarse este trabajo, de campañas aplicadas sobre rubros en general de modo de poder hacer inferencia sobre la historia de contactos con el cliente. De esto, se propone, por una parte, aprovechar la flexibilidad de los modelos probit multivariados propuestos para incorporar variables ligadas al marketing directo a medida que estas se van generando con las nuevas campañas asociadas a rubros. Por otra parte, se puede continuar una línea totalmente distinta del análisis del marketing directo a través de la realización de experimentos de campo o modelos del tipo *uplift*.

En tercer lugar, este trabajo tomó como insumo la clasificación a nivel de rubros utilizada por la empresa, lo que da paso a posibilidades de indagar en posibles relaciones a nivel menos agregados, por ejemplo, realizando un análisis similar a nivel de subrubros o de categorías.

Finalmente, un punto relevante, es que los resultados son reflejo de la generalidad de la cartera, tanto para los modelos de propensión como para la estimación del efecto sobre gasto y rentabilidad de activarse en los rubros, siendo que en variadas ocasiones será relevante conocer este comportamiento en grupos específicos de clientes, por ejemplo, un segmento de bajo o alto valor. Al igual que para la limitante anterior, un enfoque de mejora es utilizar la posibilidad de incluir jerarquías en los modelos bayesianos para permitir que los distintos efectos varíen según grupos de clientes o incluso para cada persona. Por otra parte, con respecto al efecto de activación de rubros se propone que la mejor línea es aplicar el análisis sobre los grupos requeridos.

14. BIBLIOGRAFÍA

- Altman, D., & Bland, M. (1994). Statistics Notes: Diagnostic tests 1: sensitivity and specificity. *BMJ*, 1308-1552.
- Becker, S., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The stata journal*, 358-377.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 31-72.
- Carpenter, B., Gelman, G., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*.
- Comité de Retail Financiero. (diciembre de 2016). *Tablero de indicadores estadísticos*.
- Dehejia, R., & Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 1053-1062.
- Diamond, A., & Sekhon, J. (2013). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Review of Economics and Statistics*, 932-945.
- Ho, D., Imai, K., King, G., & Stuart, E. (18 de abril de 2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 1-28. Obtenido de <https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf>
- INE. (3 de mayo de 2017). *Índice de actividad del comercio (IAC)*. Obtenido de [http://www.ine.cl/docs/default-source/boletines/%C3%ADndice-de-actividad-del-comercio/2017/bolet%C3%ADn-sectores-econ%C3%B3micos-%C3%ADndice-de-la-actividad-del-comercio-\(iac\)-marzo-2017.pdf?sfvrsn=4](http://www.ine.cl/docs/default-source/boletines/%C3%ADndice-de-actividad-del-comercio/2017/bolet%C3%ADn-sectores-econ%C3%B3micos-%C3%ADndice-de-la-actividad-del-comercio-(iac)-marzo-2017.pdf?sfvrsn=4)
- Kamakura, W., Kossar, B., & Wedel, M. (2004). Identifying innovators for the cross-selling of new products. *Management Science*, 1120-1133.
- Knott, A., Hayes, A., & Neslin, S. (2002). Next-product-to-buy models for cross-selling applications. *Journal of Interactive Marketing*, 59-75.
- Li, S., Sun, B., & Wilcox, R. (2005). Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research*, 233-239.

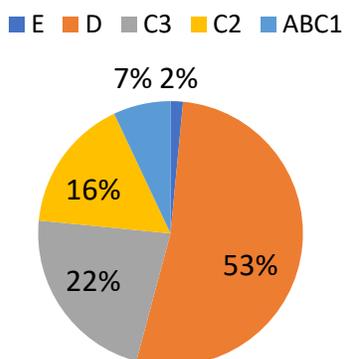
- Liu, C., & Cai, S. (2007). Customer cross-selling model based on customer maturity and product grade. *International Management Review*, 33-50.
- Lo, V. (2002). The true lift model: a novel data mining approach to response modeling in database marketing. *ACM SIGKDD Explorations Newsletter*, 78-86.
- Manchanda, P., Ansari, A., & Gupta, S. (1999). The “shopping basket”: A model for multicategory purchase incidence decisions. *Marketing science*, 95-114.
- Matsumoto, S., Kamei, Y., Monden, A., & Matsumoto, K. (2007). Comparison of outlier detection methods in fault-proneness models. *Empirical Software Engineering and Measurement* (págs. 461- 463). ESEM.
- Montoya, R. (2016). Logit + Probit. Santiago.
- Montoya, R. (2016). Modelos de Elección Discreta. Santiago.
- R Core Team. (2007). *R: A Language and Environment for Statistical Computing*. Obtenido de <https://www.R-project.org/>
- Raftery, A. (1985). A model for high-order Markov chains. *Journal of the Royal Statistical Society*, 528-539.
- Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 41-55.
- Rosenbaum, P., & Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 33–38.
- Rossi, P., Allenby, G., & McCulloch, R. (2012). *Bayesian statistics and marketing*. John Wiley & Sons.
- Rubin, D., & Thomas, N. (2000). Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates. *Journal of the American Statistical Association*, 573-585.
- Salas, J., & Portillo, M. (2008). P. Ch. Mahalanobis y las aplicaciones de su distancia estadística. *CULCyT: Cultura Científica y Tecnológica*, 13-20.
- SBIF. (junio de 2016). Obtenido de <http://www.sbif.cl/sbifweb/servlet/InfoFinanciera?indice=4.1&idCategoria=564&tipocont=568>
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 12-22.
- Smith, H. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology*, 325–353.
- Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 1-22.
- Vargas, M. (2003). *Estimación del Modelo Probit Multivariante (Tesis para obtención del grado académico de Licenciatura en Economía)*. La Paz.
- Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research*, 3571-3594.

15. ANEXOS

Anexo A: Porcentaje de la cartera de clientes de la empresa según GSE.

Porcentaje de clientes según GSE



Anexo B: Rubros de gestión definidos por la empresa.

Posición	Rubro de Gestión
1	TIENDAS POR DEPARTAMENTO
2	SUPERMERCADOS
3	MEJORAMIENTO HOGAR
4	COMBUSTIBLE
5	SALUD
6	VIVIENDA
7	VIAJES
8	FARMACIAS
9	AUTOMOTRIZ
10	COMUNICACION
11	RESTAURANT
12	EDUCACION
13	ENTRETENCION
14	SEGUROS
15	TRANSPORTE
16	RECAUDACIÓN
17	OTROS
18	PAGO WEB
19	EMPRESA

Anexo C: Especificación del criterio WAIC.

El WAIC se define como:

$$\text{WAIC} = -2 \widehat{\text{elpd}}_{\text{waic}}.$$

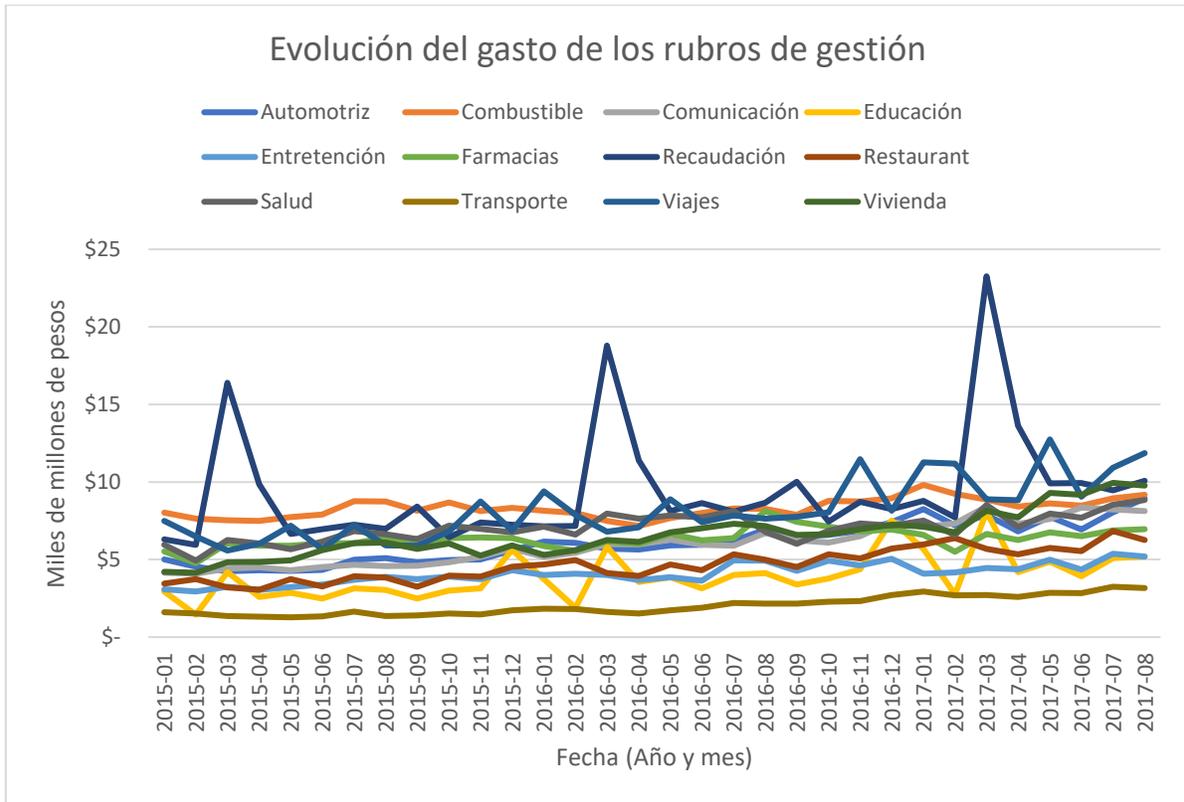
Dónde cada una de las componentes queda determinada por las siguientes ecuaciones, siendo la comparación aplicada sobre el $\widehat{\text{elpd}}_{\text{waic}}$.

$$\widehat{\text{elpd}}_{\text{waic}} = \widehat{\text{lpd}} - \widehat{p}_{\text{waic}}.$$

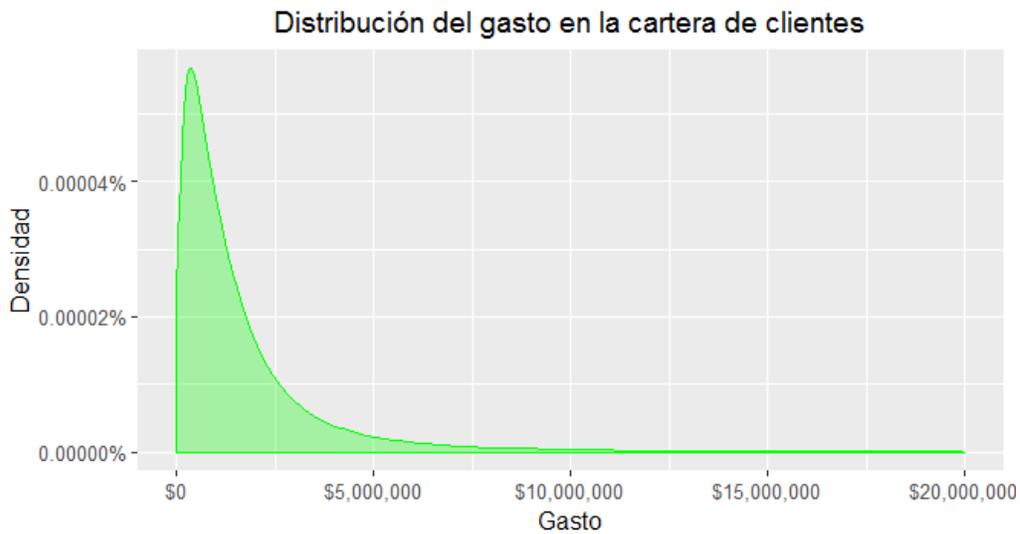
$$p_{\text{waic}} = \sum_{i=1}^n \text{var}_{\text{post}}(\log p(y_i|\theta)).$$

$$\widehat{p}_{\text{waic}} = \sum_{i=1}^n V_{s=1}^S (\log p(y_i|\theta^s))$$

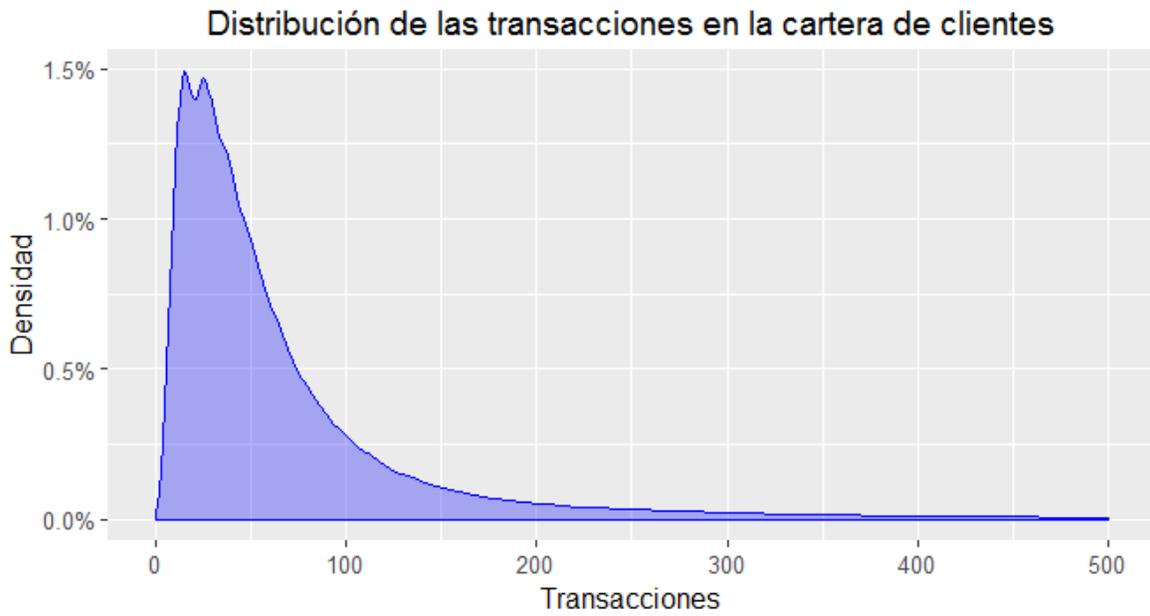
Anexo D: Gasto mensual por los principales rubros de gestión durante el último año móvil.



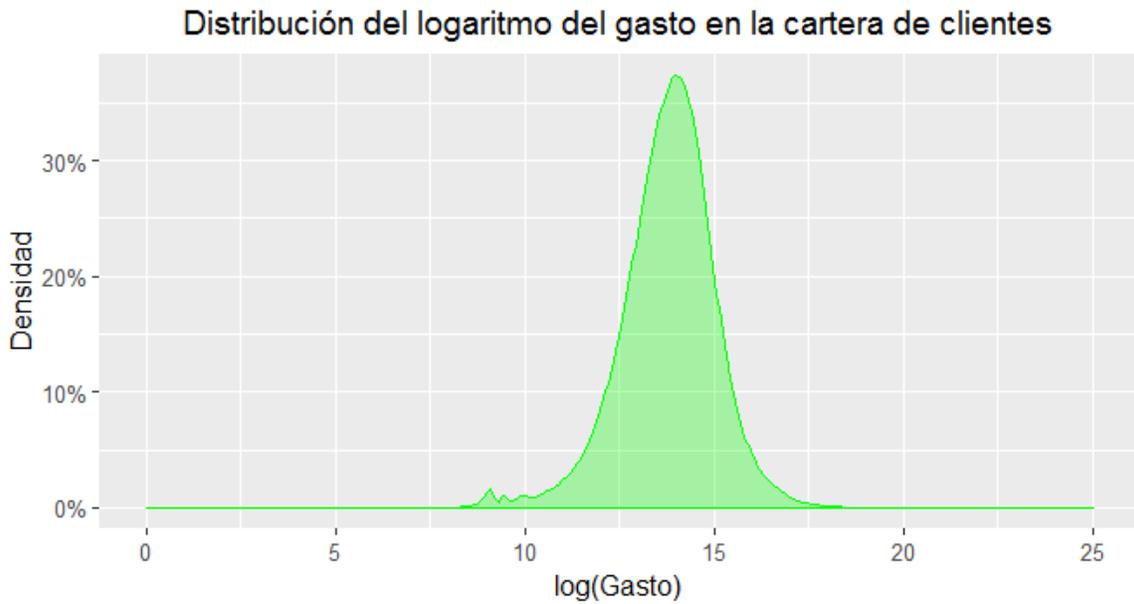
Anexo E: Distribución del gasto en los clientes de la base analítica.



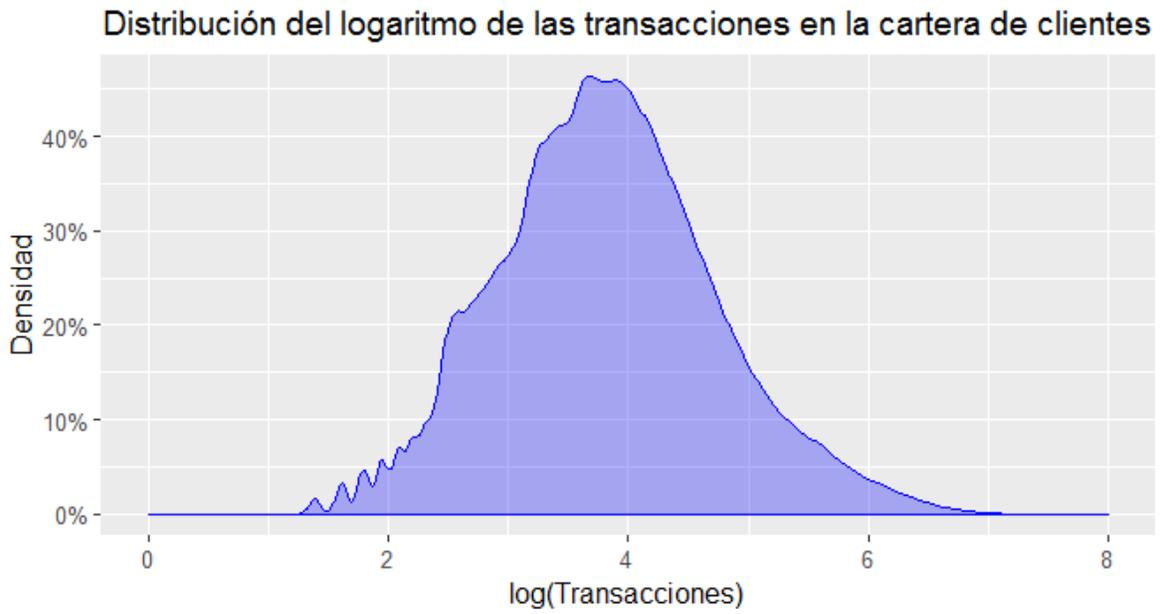
Anexo F: Distribución de la cantidad de transacciones en los clientes de la base analítica.



Anexo G: Distribución del logaritmo del gasto en los clientes de la base analítica.



Anexo H: Distribución del logaritmo de la cantidad de transacciones en los clientes de la base analítica.



Anexo I: Gráficos cuantil-cuantil para las transformaciones logarítmicas del gasto y la cantidad de transacciones de los clientes.

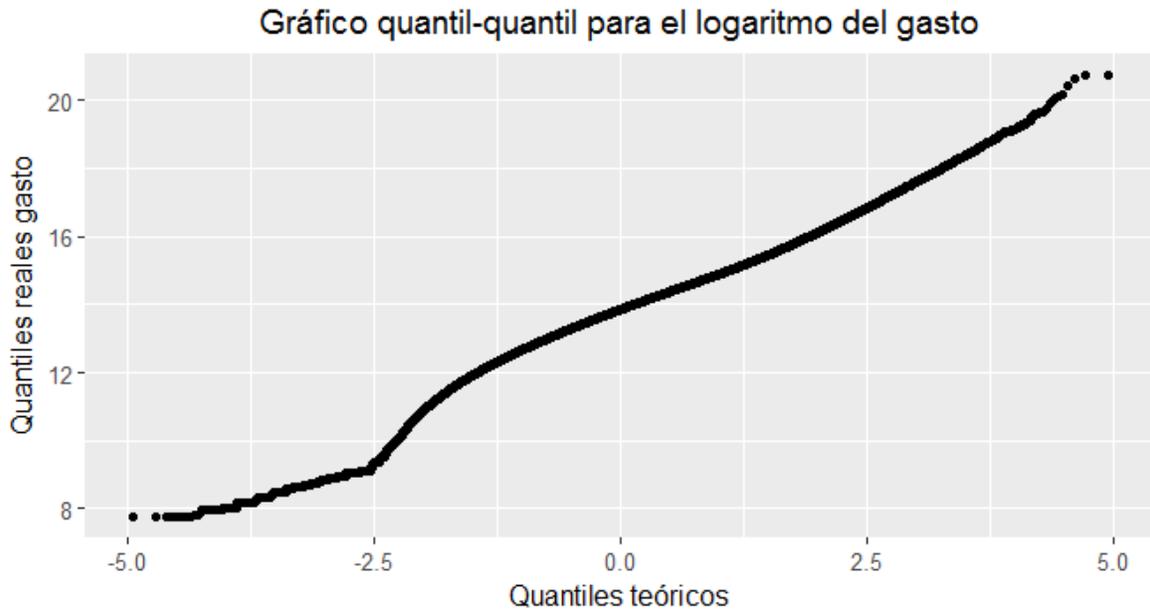
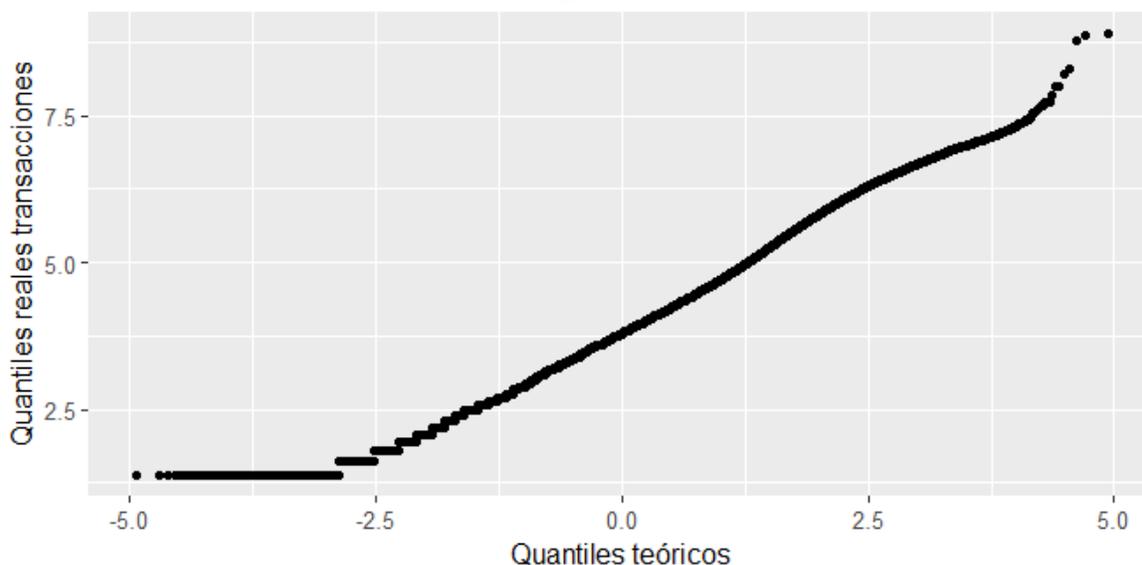


Gráfico quantil-quantil para el logaritmo de la cantidad de transacciones



Anexo J: Set inicial de variables para la realización de los modelos logit.

Variable	Tipo	Descripción
DISPONIBLE	Continua	Monto disponible al final del periodo
CUPO	Continua	Monto del cupo del cliente
INGRESOS_WEB	Continua	Ingresos a la página web del negocio
NLINEAS_RETAIL	Continua	Número de líneas distintas del Retail del grupo en las que compró el cliente en el periodo
EDAD	Continua	Edad del cliente
VAR_CUPO	Continua	Variación porcentual del cupo del cliente respecto del mes anterior
NRUBROS	Continua	Número de rubros activos durante el mes
HOMBRE	Binaria	Indica si cliente es hombre
T_ABIERTA_CLASICA	Binaria	Indica si la tarjeta del cliente es abierta y clásica (normal)
T_ABIERTA_PS	Binaria	Indica si la tarjeta del cliente es abierta y de tipo Premium.
ING_ADICIONAL	Binaria	Indica si ingresó cuentas adicionales en el mes

ENGANCHADO_PUNTOS	Binaria	Indica si el cliente ha gastado más del 90% de los puntos acumulados en el club de fidelización durante el último año
ENGANCHADO_FINANCIAMIENTO	Binaria	Marca del negocio que indica si el cliente utiliza la tarjeta como medio de financiamiento.
ENGANCHADO_PROMOCION	Binaria	Marca del negocio que indica si el cliente utiliza la tarjeta en promociones.
ID_PREMIUM	Catórica	Indica la categoría a la cual pertenece el cliente (normal o premium)
TRASPASO	Binaria	Indica si el cliente cambió su tarjeta en el mes
CAMBIO_CATEGORIA	Binaria	Indica si el cliente cambió su categoría en el mes
AUTOMOTRIZ	Binaria	Indican si el cliente realizó o no transacciones durante el mes en cada rubro en específico
COMBUSTIBLE	Binaria	
COMUNICACIÓN	Binaria	
EDUCACION	Binaria	
ENTRETENCION	Binaria	
FARMACIAS	Binaria	
MEJORAMIENTO_HOGAR	Binaria	
OTROS	Binaria	
RECAUDACION	Binaria	
RESTAURANT	Binaria	
SALUD	Binaria	
SEGUROS	Binaria	
SUPERMERCADOS	Binaria	
TIENDAS_POR_DEPARTAMENTO	Binaria	
TRANSPORTE	Binaria	
VIAJES	Binaria	
VIVIENDA	Binaria	

Anexo K: Coeficiente de correlación de Pearson para las variables continuas.

Coef. Pearson	DISPONIBLE	MONTO_CUPO	INGRESOS_WEB	EDAD	VAR_CUPO	NRUBROS_L1
DISPONIBLE	1	0,91	0	0,15	0,01	0,04
MONTO_CUPO	0,91	1	0,02	0,13	0	0,23
INGRESOS_WEB	0	0,02	1	-0,05	0	0,06
EDAD	0,15	0,13	-0,05	1	-0,05	-0,13
VAR_CUPO	0,01	0	0	-0,05	1	0,01
NRUBROS_L1	0,04	0,23	0,06	-0,13	0,01	1

Anexo L: Coeficiente V de Cramer para las variables categóricas y binarias.

V de Cramer	ENGANCHADO_PUNTOS	ENGANCHADO_PROMOCION	ENGANCHADO_FINANCIAMIENTO	ID_PREMIUM	HOMBRE	T_VISA_CLASICA	T_VISA_PS	TRASPASO	CAMBIO_CATEGORIA
ENGANCHADO_PUNTOS	1	0,14	0,11	0,44	0,02	0,05	0,22	0,02	0,07
ENGANCHADO_PROMOCION	0,14	1	0	0,08	0,02	0,01	0,05	0,01	0,01
ENGANCHADO_FINANCIAMIENTO	0,11	0	1	0,1	0,04	0,03	0,04	0	0,02
ID_PREMIUM	0,44	0,08	0,1	1	0,07	0,28	0,65	0,04	0,13
HOMBRE	0,02	0,02	0,04	0,07	1	0,04	0,06	0,02	0,01
T_VISA_CLASICA	0,05	0,01	0,03	0,28	0,04	1	0,5	0	0,01
T_VISA_PS	0,22	0,05	0,04	0,65	0,06	0,5	1	0,08	0,05
TRASPASO	0,02	0,01	0	0,04	0,02	0	0,08	1	0,13
CAMBIO_CATEGORIA	0,07	0,01	0,02	0,13	0,01	0,01	0,05	0,13	1

Anexo M: Detalle de las variables independientes, coeficientes estimados y p-valor de estos para los modelos logits de activación en cada rubro.

Variable independiente	Automotriz	p-valor	AME	Combustible	p-valor	AME	Comunicación	p-valor	AME
INTERCEPTO	-7,00	0,00		-4,28	0,00		-3,45	0,00	
HOMBRE	0,84	0,00	1,60%	0,57	0,00	1,81%	0,12	0,00	0,34%
EDAD	-0,01	0,00	-0,02%	-0,02	0,00	-0,06%	-0,02	0,00	-0,06%
VAR_CUPO	0,11	0,00	0,20%			0,00%	0,09	0,00	0,25%
T_ABIERTA_CLASICA	2,44	0,00	4,64%	0,99	0,00	3,14%	1,19	0,00	3,30%
T_ABIERTA_PREMIUM	2,69	0,00	5,12%	1,30	0,00	4,14%	1,40	0,00	3,88%
CAMBIO_CATEGORIA			0,00%			0,00%	0,07	0,13	0,20%
ENGANCHADO_PUNTOS	0,50	0,00	0,96%	0,51	0,00	1,62%	0,38	0,00	1,06%
AUTOMOTRIZ_L1			0,00%	0,91	0,00	2,89%			0,00%
COMBUSTIBLE_L1	1,13	0,00	2,15%			0,00%	0,22	0,00	0,62%
COMUNICACION_L1	0,08	0,00	0,16%	0,15	0,00	0,48%			0,00%
EDUCACION_L1			0,00%	0,40	0,00	1,27%			0,00%
ENTRETENCION_L1			0,00%			0,00%	1,48	0,00	4,08%
FARMACIAS_L1	0,16	0,00	0,30%	0,40	0,00	1,27%	0,15	0,00	0,40%
MEJ_HOGAR_L1	0,34	0,00	0,64%	0,28	0,00	0,88%	0,04	0,00	0,12%
RECAUDACION_L1	0,15	0,00	0,29%	0,11	0,00	0,35%	0,09	0,00	0,25%
RESTAURANT_L1	0,13	0,00	0,25%	0,42	0,00	1,34%			0,00%
SALUD_L1	0,24	0,00	0,46%	0,22	0,00	0,70%			0,00%
SEGUROS_L1			0,00%			0,00%	0,08	0,00	0,21%
SUPERMERCADOS_L1	0,40	0,00	0,76%	0,80	0,00	2,54%	0,29	0,00	0,81%
TIENDAS_POR_DEPTO_L1	0,14	0,00	0,27%			0,00%	0,17	0,00	0,48%
TRANSPORTE_L1	-0,03	0,19	-0,05%	-0,04	0,02	-0,14%	0,26	0,00	0,72%
VIAJES_L1	0,22	0,00	0,41%			0,00%			0,00%
VIVIENDA_L1	0,17	0,00	0,33%	0,27	0,00	0,87%	0,76	0,00	2,10%
ABRIL	-0,19	0,00	-0,36%	-0,11	0,00	-0,36%	-1,10	0,00	-3,04%
AGOSTO			0,00%			0,00%	-1,13	0,00	-3,13%
DICIEMBRE	-0,12	0,00	-0,23%	-0,08	0,00	-0,25%	-0,87	0,00	-2,41%
FEBRERO			0,00%			0,00%	-0,90	0,00	-2,48%
JULIO			0,00%			0,00%	-1,13	0,00	-3,12%
JUNIO	-0,12	0,00	-0,23%	-0,13	0,00	-0,42%	-0,93	0,00	-2,59%
MARZO			0,00%			0,00%	-0,91	0,00	-2,53%
MAYO			0,00%	-0,09	0,00	-0,29%	-0,93	0,00	-2,56%
NOVIEMBRE	-0,09	0,00	-0,18%			0,00%	-1,46	0,00	-4,05%
OCTUBRE			0,00%			0,00%	-1,49	0,00	-4,12%
SEPTIEMBRE	-0,08	0,00	-0,15%	-0,19	0,00	-0,60%	-1,54	0,00	-4,26%

Variable independiente	Educación	p-valor	AME	Entretención	p-valor	AME	Farmacias	p-valor	AME
INTERCEPTO	-7,00	0,00		-4,18	0,00		-4,19	0,00	
HOMBRE			0,00%	-0,18	0,00	-0,49%	-0,55	0,00	-2,56%
EDAD	-0,02	0,00	-0,01%	-0,04	0,00	-0,10%			0,00%
VAR_CUPO	0,08	0,09	0,04%			0,00%			0,00%
T_ABIERTA_CLASICA	1,91	0,00	0,88%	1,48	0,00	4,06%	0,58	0,00	2,72%
T_ABIERTA_PREMIUM	2,28	0,00	1,06%	1,86	0,00	5,10%	0,69	0,00	3,24%
CAMBIO_CATEGORIA			0,00%			0,00%			0,00%
ENGANCHADO_PUNTOS	0,56	0,00	0,26%	0,35	0,00	0,95%	0,55	0,00	2,58%
AUTOMOTRIZ_L1	0,21	0,00	0,10%			0,00%	0,19	0,00	0,90%
COMBUSTIBLE_L1	0,35	0,00	0,16%	0,20	0,00	0,56%	0,56	0,00	2,63%
COMUNICACION_L1	0,38	0,00	0,18%	0,40	0,00	1,10%	0,13	0,00	0,58%
EDUCACION_L1			0,00%	0,45	0,00	1,25%	0,26	0,00	1,21%
ENTRETENCION_L1	0,44	0,00	0,20%			0,00%	0,19	0,00	0,91%
FARMACIAS_L1	0,25	0,00	0,11%	0,34	0,00	0,92%			0,00%
MEJ_HOGAR_L1			0,00%			0,00%	0,15	0,00	0,68%
RECAUDACION_L1			0,00%			0,00%			0,00%
RESTAURANT_L1	0,39	0,00	0,18%	0,72	0,00	1,98%	0,58	0,00	2,73%
SALUD_L1	0,39	0,00	0,18%	0,30	0,00	0,82%	0,63	0,00	2,93%
SEGUROS_L1	-0,19	0,00	-0,09%	-0,12	0,00	-0,32%	-0,07	0,00	-0,32%
SUPERMERCADOS_L1	0,42	0,00	0,19%	0,34	0,00	0,93%	0,93	0,00	4,33%
TIENDAS_POR_DEPTO_L1	0,45	0,00	0,21%	0,48	0,00	1,32%	0,44	0,00	2,06%
TRANSPORTE_L1	0,26	0,00	0,12%	0,46	0,00	1,27%	0,15	0,00	0,70%
VIAJES_L1	0,39	0,00	0,18%	0,43	0,00	1,19%			0,00%
VIVIENDA_L1	0,41	0,00	0,19%	0,11	0,00	0,30%	0,29	0,00	1,37%
ABRIL	-0,71	0,00	-0,33%	0,24	0,00	0,65%			0,00%
AGOSTO	-0,27	0,00	-0,13%	1,22	0,00	3,35%	0,33	0,00	1,56%
DICIEMBRE			0,00%	0,33	0,00	0,91%	0,16	0,00	0,74%
FEBRERO	-0,68	0,00	-0,31%	0,17	0,00	0,46%			0,00%
JULIO	-0,30	0,00	-0,14%	0,58	0,00	1,58%	0,10	0,00	0,49%
JUNIO	-0,88	0,00	-0,41%			0,00%	0,07	0,00	0,31%
MARZO	0,27	0,00	0,12%			0,00%	0,23	0,00	1,05%
MAYO	-0,65	0,00	-0,30%	0,50	0,00	1,38%	0,15	0,00	0,70%
NOVIEMBRE	-0,55	0,00	-0,25%	0,17	0,00	0,47%	0,16	0,00	0,76%
OCTUBRE	-0,60	0,00	-0,28%	0,40	0,00	1,10%	0,24	0,00	1,10%
SEPTIEMBRE	-0,58	0,00	-0,27%	0,31	0,00	0,84%	0,16	0,00	0,76%

Variable independiente	Mej. Hogar	p-valor	AME	Recaudación	p-valor	AME	Restaurant	p-valor	AME
INTERCEPTO	-2,19	0,00		-6,05	0,00		-3,97	0,00	
HOMBRE	0,15	0,00	2,00%	0,36	0,00	0,90%			0,00%
EDAD	0,00	0,00	0,02%	0,01	0,00	0,03%	-0,03	0,00	-0,10%
VAR_CUPO	0,06	0,00	0,86%	0,21	0,00	0,52%	0,07	0,00	0,23%
T_ABIERTA_CLASICA	0,10	0,00	1,39%	0,32	0,00	0,79%	1,25	0,00	3,97%
T_ABIERTA_PREMIUM	0,29	0,00	3,94%	0,69	0,00	1,69%	1,73	0,00	5,51%
CAMBIO_CATEGORIA	0,19	0,00	2,50%			0,00%	0,15	0,00	0,49%
ENGANCHADO_PUNTOS	0,58	0,00	7,78%	0,57	0,00	1,40%	0,34	0,00	1,07%
AUTOMOTRIZ_L1	0,31	0,00	4,18%	0,44	0,00	1,08%	0,20	0,00	0,62%
COMBUSTIBLE_L1	0,35	0,00	4,77%	0,59	0,00	1,44%	0,57	0,00	1,81%
COMUNICACION_L1			0,00%	0,26	0,00	0,64%	0,14	0,00	0,45%
EDUCACION_L1			0,00%	0,34	0,00	0,84%	0,48	0,00	1,53%
ENTRETENCION_L1			0,00%	0,14	0,00	0,34%	0,52	0,00	1,66%
FARMACIAS_L1	0,20	0,00	2,71%	0,25	0,00	0,60%	0,60	0,00	1,89%
MEJ_HOGAR_L1			0,00%	0,09	0,00	0,23%	0,07	0,00	0,21%
RECAUDACION_L1	-0,04	0,00	-0,59%			0,00%			0,00%
RESTAURANT_L1	0,09	0,00	1,19%	0,25	0,00	0,62%			0,00%
SALUD_L1	0,10	0,00	1,29%	0,28	0,00	0,70%	0,38	0,00	1,22%
SEGUROS_L1			0,00%	0,08	0,00	0,20%			0,00%
SUPERMERCADOS_L1	0,30	0,00	3,99%	0,25	0,00	0,61%	0,94	0,00	2,99%
TIENDAS_POR_DEPTO_L1	0,23	0,00	3,06%	0,08	0,00	0,20%	0,40	0,00	1,26%
TRANSPORTE_L1			0,00%	0,44	0,00	1,07%	0,54	0,00	1,71%
VIAJES_L1	0,09	0,00	1,25%	0,28	0,00	0,68%	0,48	0,00	1,52%
VIVIENDA_L1	0,12	0,00	1,68%	0,55	0,00	1,36%	0,26	0,00	0,84%
ABRIL	-0,07	0,00	-0,98%	0,41	0,00	1,00%	-0,11	0,00	-0,36%
AGOSTO	-0,07	0,00	-0,99%	0,55	0,00	1,34%			0,00%
DICIEMBRE	0,18	0,00	2,46%	0,71	0,00	1,74%			0,00%
FEBRERO	-0,11	0,00	-1,49%	0,23	0,00	0,57%	0,15	0,00	0,48%
JULIO	-0,13	0,00	-1,81%	0,26	0,00	0,65%	0,21	0,00	0,65%
JUNIO	-0,07	0,00	-0,92%	0,34	0,00	0,83%			0,00%
MARZO	0,04	0,00	0,59%	1,79	0,00	4,39%	0,28	0,00	0,90%
MAYO			0,00%	0,50	0,00	1,24%			0,00%
NOVIEMBRE	-0,26	0,00	-3,51%	0,35	0,00	0,86%	-0,07	0,00	-0,22%
OCTUBRE	0,04	0,00	0,56%	0,34	0,00	0,82%			0,00%
SEPTIEMBRE	-0,14	0,00	-1,85%	0,53	0,00	1,31%	-0,24	0,00	-0,76%

Variable independiente	Salud	p-valor	AME	Seguros	p-valor	AME	Supermercados	p-valor	AME
INTERCEPTO	-5,74	0,00		-1,88	0,00		-2,75	0,00	
HOMBRE	-0,37	0,00	-0,97%	0,09	0,00	0,60%			0,00%
EDAD	-0,01	0,00	-0,01%	-0,01	0,00	-0,07%	-0,01	0,00	-0,11%
VAR_CUPO			0,00%	0,16	0,00	1,06%			0,00%
T_ABIERTA_CLASICA	1,74	0,00	4,61%			0,00%	0,52	0,00	4,63%
T_ABIERTA_PREMIUM	2,00	0,00	5,31%	0,34	0,00	2,31%	0,74	0,00	6,56%
CAMBIO_CATEGORIA			0,00%	-0,15	0,09	-1,04%			0,00%
ENGANCHADO_PUNTOS	0,48	0,00	1,27%	-0,57	0,00	-3,83%	0,58	0,00	5,13%
AUTOMOTRIZ_L1	0,23	0,00	0,60%	0,29	0,00	1,94%	0,36	0,00	3,22%
COMBUSTIBLE_L1	0,33	0,00	0,88%			0,00%	0,76	0,00	6,76%
COMUNICACION_L1	0,11	0,00	0,28%	-0,14	0,00	-0,94%	0,10	0,00	0,92%
EDUCACION_L1	0,32	0,00	0,85%	0,24	0,00	1,61%	0,38	0,00	3,37%
ENTRETENCION_L1	0,28	0,00	0,74%			0,00%	0,15	0,00	1,35%
FARMACIAS_L1	0,69	0,00	1,82%			0,00%	0,72	0,00	6,44%
MEJ_HOGAR_L1	0,13	0,00	0,34%	0,11	0,00	0,77%	0,23	0,00	2,02%
RECAUDACION_L1			0,00%	-0,26	0,00	-1,77%	-0,03	0,00	-0,23%
RESTAURANT_L1	0,46	0,00	1,22%	0,23	0,00	1,56%	0,90	0,00	7,96%
SALUD_L1			0,00%	0,24	0,00	1,61%	0,32	0,00	2,86%
SEGUROS_L1			0,00%			0,00%	0,03	0,00	0,29%
SUPERMERCADOS_L1	0,54	0,00	1,44%	-0,22	0,00	-1,49%			0,00%
TIENDAS_POR_DEPTO_L1	0,39	0,00	1,04%	-0,03	0,06	-0,22%	0,27	0,00	2,39%
TRANSPORTE_L1	0,27	0,00	0,73%			0,00%	0,23	0,00	2,01%
VIAJES_L1	0,17	0,00	0,46%	0,26	0,00	1,75%	0,08	0,00	0,71%
VIVIENDA_L1	0,21	0,00	0,55%	0,33	0,00	2,22%	0,35	0,00	3,11%
ABRIL			0,00%			0,00%	0,22	0,00	1,92%
AGOSTO	0,08	0,00	0,20%	0,13	0,00	0,88%	0,31	0,00	2,77%
DICIEMBRE			0,00%	0,10	0,00	0,67%	0,48	0,00	4,29%
FEBRERO	-0,06	0,01	-0,16%			0,00%	0,15	0,00	1,31%
JULIO	0,13	0,00	0,33%			0,00%	0,18	0,00	1,56%
JUNIO			0,00%	0,12	0,00	0,81%	0,22	0,00	1,99%
MARZO	0,20	0,00	0,54%	0,09	0,01	0,59%	0,31	0,00	2,80%
MAYO	0,10	0,00	0,27%	0,23	0,00	1,59%	0,20	0,00	1,82%
NOVIEMBRE	0,12	0,00	0,32%	0,14	0,00	0,95%	0,21	0,00	1,89%
OCTUBRE	0,22	0,00	0,57%			0,00%	0,26	0,00	2,30%
SEPTIEMBRE	-0,10	0,00	-0,27%			0,00%	0,09	0,00	0,80%

Variable independiente	Tiendas depto.	p-valor	AME	Transporte	p-valor	AME	Viajes	p-valor	AME	Vivienda	p-valor	AME
INTERCEPTO	-0,49	0,00		-5,59	0,00		-6,11	0,00		-7,43	0,00	
HOMBRE	-0,54	0,00	-9,99%			0,00%	-0,05	0,00	-0,06%	0,06	0,00	0,10%
EDAD	-0,02	0,00	-0,34%	-0,02	0,00	-0,03%	-0,02	0,00	-0,03%	0,00	0,00	0,01%
VAR_CUPO			0,00%	0,05	0,12	0,07%	0,04	0,24	0,05%			0,00%
T_ABIERTA_CLASICA	0,27	0,00	4,94%	1,59	0,00	2,52%	2,25	0,00	3,10%	1,95	0,00	3,07%
T_ABIERTA_PREMIUM	0,48	0,00	8,88%	1,95	0,00	3,08%	2,95	0,00	4,06%	2,19	0,00	3,46%
CAMBIO_CATEGORIA			0,00%			0,00%	0,30	0,00	0,41%			0,00%
ENGANCHADO_PUNTOS	0,55	0,00	10,25%	0,43	0,00	0,68%	0,40	0,00	0,54%	0,38	0,00	0,61%
AUTOMOTRIZ_L1	0,20	0,00	3,72%	0,17	0,00	0,26%	0,12	0,00	0,17%	0,26	0,00	0,41%
COMBUSTIBLE_L1	0,16	0,00	2,93%	0,16	0,00	0,26%	0,26	0,00	0,36%	0,56	0,00	0,89%
COMUNICACION_L1	0,13	0,00	2,46%	0,45	0,00	0,70%	0,26	0,00	0,36%	0,57	0,00	0,89%
EDUCACION_L1	0,37	0,00	6,86%	0,29	0,00	0,46%	0,40	0,00	0,55%	0,31	0,00	0,50%
ENTRETENCION_L1	0,27	0,00	5,05%	0,47	0,00	0,74%	0,51	0,00	0,71%	0,13	0,00	0,20%
FARMACIAS_L1	0,49	0,00	9,11%	0,20	0,00	0,32%	0,11	0,00	0,15%	0,36	0,00	0,56%
MEJ_HOGAR_L1			0,00%			0,00%	0,12	0,00	0,17%	0,13	0,00	0,21%
RECAUDACION_L1	-0,06	0,00	-1,17%	0,30	0,00	0,47%			0,00%	0,13	0,00	0,21%
RESTAURANT_L1	0,41	0,00	7,54%	0,48	0,00	0,75%	0,51	0,00	0,70%	0,29	0,00	0,46%
SALUD_L1	0,26	0,00	4,76%	0,20	0,00	0,32%	0,25	0,00	0,34%	0,13	0,00	0,21%
SEGUROS_L1	-0,04	0,00	-0,72%			0,00%	-0,14	0,00	-0,19%	0,08	0,00	0,12%
SUPERMERCADOS_L1	0,29	0,00	5,34%	0,38	0,00	0,60%	0,05	0,02	0,06%	0,87	0,00	1,38%
TIENDAS_POR_DEPTO_L1			0,00%	0,30	0,00	0,47%	0,37	0,00	0,51%	0,13	0,00	0,20%
TRANSPORTE_L1	0,06	0,00	1,11%			0,00%	0,41	0,00	0,57%	0,34	0,00	0,54%
VIAJES_L1	0,29	0,00	5,44%	0,70	0,00	1,10%			0,00%			0,00%
VIVIENDA_L1			0,00%	0,40	0,00	0,63%			0,00%			0,00%
ABRIL	0,26	0,00	4,80%			0,00%	-0,21	0,00	-0,29%	0,17	0,00	0,27%
AGOSTO	0,28	0,00	5,14%	-0,09	0,00	-0,14%			0,00%	0,31	0,00	0,49%
DICIEMBRE	0,98	0,00	18,14%			0,00%	-0,24	0,00	-0,33%	0,13	0,00	0,21%
FEBRERO	0,25	0,00	4,62%			0,00%	0,11	0,00	0,15%			0,00%
JULIO	0,33	0,00	6,17%	0,15	0,00	0,24%			0,00%	0,41	0,00	0,64%
JUNIO	0,28	0,00	5,11%			0,00%	-0,28	0,00	-0,39%	0,40	0,00	0,62%
MARZO	0,29	0,00	5,31%	0,11	0,00	0,17%	-0,17	0,00	-0,24%	0,29	0,00	0,46%
MAYO	0,47	0,00	8,74%			0,00%	0,19	0,00	0,26%	0,54	0,00	0,85%
NOVIEMBRE	0,37	0,00	6,78%	-0,08	0,01	-0,12%			0,00%	0,25	0,00	0,39%
OCTUBRE	0,30	0,00	5,56%			0,00%	-0,16	0,00	-0,22%	0,29	0,00	0,45%
SEPTIEMBRE	0,21	0,00	3,84%	-0,13	0,00	-0,21%	-0,09	0,00	-0,13%	0,16	0,00	0,26%

Anexo N: Métricas de ajuste y poder predictivo de los modelos logits de activación por rubro.

Rubro (modelo)	Sensibilidad	Especificidad	AUC entrenamiento	Precisión	AUC validación	BIC
Automotriz	73%	73%	81%	6%	81%	200.961
Combustible	70%	69%	76%	8%	76%	286.755
Comunicación	69%	70%	77%	7%	77%	233.482
Educación	76%	75%	83%	2%	83%	62.901
Entretención	74%	72%	80%	9%	80%	265.059
Farmacias	70%	70%	77%	14%	77%	397.285
Mej. Hogar	59%	62%	64%	26%	64%	728.866
Recaudación	66%	71%	75%	7%	75%	197.455
Restaurant	74%	74%	82%	11%	82%	296.861
Salud	71%	71%	78%	8%	79%	267.798
Seguros	61%	56%	62%	10%	61%	114.359
Supermercados	65%	64%	70%	20%	69%	538.674
Tiendas depto.	62%	62%	67%	42%	67%	632.645
Transporte	72%	70%	78%	4%	78%	166.472
Viajes	72%	69%	78%	4%	78%	159.781
Vivienda	71%	72%	79%	5%	79%	168.558

Anexo O: Verificación de zonas de soporte común para el propensity score de cada rubro.

Rubro	Activación	Cantidad de registros	Menor ps	Promedio ps	Máximo ps
Automotriz	No	237.400	0%	2%	99%
	Si	3.856	0%	6%	40%
Combustible	No	214.984	0%	3%	68%
	Si	4.624	0%	7%	63%
Comunicación	No	214.800	0%	3%	98%
	Si	4.706	0%	10%	77%
Educación	No	251.874	0%	0%	98%
	Si	1.002	0%	2%	25%
Entretención	No	219.954	0%	3%	76%
	Si	4.879	0%	8%	70%
Farmacias	No	198.981	1%	5%	75%
	Si	7.350	1%	11%	76%
Mej. del Hogar	No	114.273	8%	16%	65%
	Si	17.408	8%	19%	62%
Recaudación	No	185.763	0%	3%	77%
	Si	3.506	0%	8%	81%
Restaurant	No	219.817	0%	3%	94%
	Si	5.273	0%	10%	79%

Salud	No	224.537	0%	3%	62%
	Si	5.741	0%	7%	61%
Seguros	No	49.953	2%	8%	47%
	Si	1.737	2%	9%	50%
Supermercados	No	149.342	2%	10%	88%
	Si	10.347	2%	14%	80%
Tiendas por depto.	No	59.162	6%	26%	87%
	Si	13.698	8%	31%	91%
Transporte	No	235.925	0%	2%	100%
	Si	2.249	0%	5%	59%
Viajes	No	243.171	0%	1%	49%
	Si	2.667	0%	4%	42%
Vivienda	No	233.810	0%	2%	63%
	Si	2.510	0%	4%	42%

Anexo P: Detalle verificación del balance y estimación del efecto causal sobre el gasto de los emparejamientos de los distintos rubros de gestión.

Rubro	Variable	Promedio/proporción en grupo que se activa	Promedio/proporción en grupo que no se activa	Diferencia	P-valor del test
Automotriz	Propensity Score	6%	5%	0%	71%
	Gasto pre-activación	\$1.025.546	\$887.859	\$137.687	8%
	Rentabilidad pre-activación	\$19.022	\$16.523	\$2.499	0%
	Gasto evaluación	\$1.166.840	\$896.080	\$270.760	0%
	Rentabilidad evaluación	\$22.598	\$18.798	\$3.800	0%
Combustible	Propensity Score	7%	7%	0%	89%
	Gasto pre-activación	\$584.354	\$548.056	\$36.297	10%
	Rentabilidad pre-activación	\$13.267	\$13.132	\$135	73%
	Gasto evaluación	\$717.232	\$541.456	\$175.777	0%
	Rentabilidad evaluación	\$17.990	\$14.844	\$3.146	0%
Comunicación	Propensity Score	10%	8%	2%	0%
	Gasto pre-activación	\$728.828	\$687.551	\$41.277	3%
	Rentabilidad pre-activación	\$16.568	\$13.425	\$3.143	0%
	Gasto evaluación	\$795.778	\$684.368	\$111.411	0%
	Rentabilidad evaluación	\$19.971	\$16.805	\$3.166	0%
Educación	Propensity Score	2%	2%	0%	31%
	Gasto pre-activación	\$1.393.934	\$1.269.269	\$124.666	25%
	Rentabilidad pre-activación	\$22.410	\$20.366	\$2.045	18%
	Gasto evaluación	\$1.568.149	\$1.303.430	\$264.720	1%
	Rentabilidad evaluación	\$29.815	\$24.430	\$5.385	0%
Entretención	Propensity Score	8%	8%	0%	28%
	Gasto pre-activación	\$798.238	\$698.287	\$99.952	5%
	Rentabilidad pre-activación	\$16.469	\$14.063	\$2.405	0%
	Gasto evaluación	\$887.715	\$689.931	\$197.784	0%
	Rentabilidad evaluación	\$18.197	\$15.465	\$2.733	0%
Farmacias	Propensity Score	11%	11%	0%	7%
	Gasto pre-activación	\$588.945	\$518.564	\$70.381	7%
	Rentabilidad pre-activación	\$13.200	\$12.136	\$1.064	7%
	Gasto evaluación	\$712.362	\$519.099	\$193.263	0%
	Rentabilidad evaluación	\$16.163	\$13.714	\$2.449	0%
Mej. del Hogar	Propensity Score	19%	18%	1%	0%
	Gasto pre-activación	\$327.192	\$274.938	\$52.254	0%
	Rentabilidad pre-activación	\$9.493	\$8.559	\$934	0%
	Gasto evaluación	\$435.416	\$316.369	\$119.047	0%
	Rentabilidad evaluación	\$12.143	\$9.571	\$2.572	0%
Recaudación	Propensity Score	8%	7%	1%	0%
	Gasto pre-activación	\$842.885	\$705.315	\$137.570	1%
	Rentabilidad pre-activación	\$15.747	\$14.314	\$1.434	13%
	Gasto evaluación	\$985.568	\$704.882	\$280.685	0%
	Rentabilidad evaluación	\$20.265	\$16.351	\$3.914	0%

Rubro	Variable	Promedio/proporción en grupo que se activa	Promedio/proporción en grupo que no se activa	Diferencia	P-valor del test
Restaurant	Propensity Score	10%	10%	0%	58%
	Gasto pre-activación	\$674.089	\$614.432	\$59.656	15%
	Rentabilidad pre-activación	\$14.422	\$14.339	\$83	81%
	Gasto evaluación	\$788.888	\$601.583	\$187.304	0%
	Rentabilidad evaluación	\$18.105	\$16.443	\$1.662	0%
Salud	Propensity Score	7%	7%	0%	59%
	Gasto pre-activación	\$798.836	\$709.098	\$89.739	11%
	Rentabilidad pre-activación	\$16.043	\$14.851	\$1.193	19%
	Gasto evaluación	\$909.385	\$716.379	\$193.006	0%
	Rentabilidad evaluación	\$19.997	\$16.804	\$3.193	0%
Seguros	Propensity Score	9%	8%	0%	31%
	Gasto pre-activación	\$489.751	\$476.038	\$13.713	67%
	Rentabilidad pre-activación	\$7.950	\$8.647	-\$697	15%
	Gasto evaluación	\$558.909	\$503.127	\$55.782	11%
	Rentabilidad evaluación	\$10.948	\$10.254	\$694	22%
Supermercados	Propensity Score	14%	14%	0%	10%
	Gasto pre-activación	\$363.724	\$332.008	\$31.716	5%
	Rentabilidad pre-activación	\$10.287	\$9.417	\$870	5%
	Gasto evaluación	\$458.825	\$351.575	\$107.249	0%
	Rentabilidad evaluación	\$13.021	\$10.923	\$2.098	0%
Tiendas por depto.	Propensity Score	31%	29%	3%	0%
	Gasto pre-activación	\$198.073	\$171.144	\$26.930	0%
	Rentabilidad pre-activación	\$7.023	\$6.012	\$1.011	0%
	Gasto evaluación	\$313.387	\$232.400	\$80.987	0%
	Rentabilidad evaluación	\$9.328	\$6.524	\$2.805	0%
Transporte	Propensity Score	5%	4%	0%	93%
	Gasto pre-activación	\$1.019.390	\$953.032	\$66.358	26%
	Rentabilidad pre-activación	\$18.776	\$17.449	\$1.326	6%
	Gasto evaluación	\$1.118.122	\$956.467	\$161.654	0%
	Rentabilidad evaluación	\$22.202	\$19.764	\$2.438	0%
Viajes	Propensity Score	4%	4%	0%	97%
	Gasto pre-activación	\$1.067.736	\$966.469	\$101.267	20%
	Rentabilidad pre-activación	\$18.843	\$16.617	\$1.226	20%
	Gasto evaluación	\$1.286.820	\$951.685	\$335.135	0%
	Rentabilidad evaluación	\$23.742	\$19.205	\$4.537	0%
Vivienda	Propensity Score	4%	4%	0%	98%
	Gasto pre-activación	\$816.848	\$746.399	\$70.449	13%
	Rentabilidad pre-activación	\$16.366	\$15.091	\$1.275	10%
	Gasto evaluación	\$947.269	\$751.061	\$196.209	0%
	Rentabilidad evaluación	\$19.776	\$16.337	\$3.400	0%

Anexo Q: Coeficientes estimados mediante el modelo probit multivariado de efectos específicos.

COVARIABLES	AUTOMOTRIZ	COMBUSTIBLE	COMUNICACION	EDUCACION
INTERCEPTO	-3,387111194	-2,386543054	-2,42319216	-3,575331659
AUTOMOTRIZ	0,847320116	0,478269372	0,027350522	-0,417004988
COMBUSTIBLE	0,540891941	1,757018332	0,067584124	0,118151554
COMUNICACION	0,1872987	0,111769318	2,810831904	-0,045603424
EDUCACION	0,101460213	0,256286955	0,172095037	2,264836449
ENTRETENCION	0,072329044	-0,007354002	0,624693112	0,310943757
FARMACIAS	0,00836687	0,063190773	-0,146399399	0,054735596
MEJ_HOGAR	0,206800237	0,20349533	0,047139635	0,031824606
RECAUDACION	0,115649732	0,084254648	0,056425947	0,163486918
RESTAURANT	0,13353662	0,188985175	-0,055036622	0,33115706
SALUD	0,061436258	0,25548539	0,150320196	0,377381669
SEGUROS	0,089972872	0,141134186	0,130518507	-0,170054607
SUPERMERCADOS	0,214340196	0,296276403	-0,043578043	-0,243079518
TIENDAS_DPTO	0,011802666	0,035040652	0,125632846	0,225064668
TRANSPORTE	-0,204486623	0,031460226	0,128786331	0,212477195
VIAJES	0,302623644	0,126916453	-0,036849449	-0,078309276
VIVIENDA	-0,081480105	0,115459644	0,252957727	0,119217221
T_AB_CLASICA	0,987290294	0,289011568	0,428686533	0,562141143
T_AB_PREMIUM	0,996284211	0,128675463	0,229630935	0,608144521
ENG_PUNTOS	0,108953777	0,272038178	0,161787626	0,342251061

COVARIABLES	ENTRETENCION	FARMACIAS	MEJ_HOGAR	RECAUDACION
INTERCEPTO	-2,688933847	-2,096003705	-1,090147463	-2,092981937
AUTOMOTRIZ	-0,042690404	-0,01153903	0,153980106	-0,050747629
COMBUSTIBLE	0,06283258	0,137409536	0,197603954	0,072897926
COMUNICACION	0,206193631	0,049591976	0,050376314	0,037594844
EDUCACION	0,097843493	0,011545214	0,114956846	0,010207856
ENTRETENCION	1,871185758	0,083004372	-0,058379548	-0,194914434
FARMACIAS	0,155994765	1,120498069	0,11115694	-0,0246031
MEJ_HOGAR	-0,048371101	0,073207606	0,663682497	-0,057076608
RECAUDACION	-0,102862632	0,02311193	-0,026149796	3,116119648
RESTAURANT	0,334296569	0,258821281	-0,010689986	0,121455311
SALUD	0,044659104	0,272378142	-0,029430459	-0,004893681
SEGUROS	0,061681157	-0,110825284	-0,054768536	0,063195659
SUPERMERCADOS	0,068606723	0,422568745	0,178196619	0,066608768
TIENDAS_DPTO	0,139200818	0,191366248	0,034745809	-0,055994658
TRANSPORTE	0,277601438	0,061017007	-0,095066982	0,056747487
VIAJES	0,181639957	0,051753121	0,171155861	0,120169553
VIVIENDA	-0,061429177	0,151076672	0,057098372	0,234537581
T_AB_CLASICA	0,691342244	0,318876795	0,273752551	0,055644357
T_AB_PREMIUM	0,563490536	0,215611911	0,035451296	0,017667344
ENG_PUNTOS	0,112720762	0,263999345	0,352555829	0,291463859

COVARIABLES	RESTAURANT	SALUD	SEGUROS	SUPERMERCADOS
INTERCEPTO	-3,047062614	-2,590451257	-1,502485936	-1,626847948
AUTOMOTRIZ	0,101045064	0,059317429	-0,206761534	0,189371804
COMBUSTIBLE	0,280174572	0,218674586	0,122190019	0,38795521
COMUNICACION	0,122318108	-0,018711785	0,100882014	0,006573048
EDUCACION	0,407707991	0,233774758	0,287092068	0,012270604
ENTRETENCION	0,422286733	0,088961411	0,119881921	0,11535757
FARMACIAS	0,194187197	0,317386131	-0,098772457	0,37019624
MEJ_HOGAR	0,017142977	0,011510399	0,094353461	0,102136008
RECAUDACION	-0,031101737	0,021061052	0,026400749	-0,003604846
RESTAURANT	1,043287691	0,187854053	-0,100577227	0,330044348
SALUD	0,23020329	1,087461408	0,045850171	0,186552898
SEGUROS	-0,025955098	-0,092100088	3,30782132	-0,038038759
SUPERMERCADOS	0,433606199	0,178067585	0,07174834	1,181167083
TIENDAS_DPTO	0,326577294	0,172381325	0,045916953	0,133601967
TRANSPORTE	0,354512956	0,195390287	0,009278528	0,226353772
VIAJES	0,076316292	-0,065934785	-0,005083959	-0,105712588
VIVIENDA	0,041958711	0,122507795	0,029710449	0,216447961
T_AB_CLASICA	1,00596561	0,492498148	0,090877417	0,422522621
T_AB_PREMIUM	0,77500775	0,489150379	0,078389354	0,166648791
ENG_PUNTOS	0,115916735	0,184521741	-0,028474576	0,378189111

COVARIABLES	TIENDAS_DPTO	TRANSPORTE	VIAJES	VIVIENDA
INTERCEPTO	-0,675789898	-2,800640522	-3,591078563	-2,77289083
AUTOMOTRIZ	0,021061652	-0,1143441	0,105040293	-0,05434021
COMBUSTIBLE	0,026523794	-0,047283375	0,084083173	0,189942067
COMUNICACION	0,159919152	0,206033123	0,277577673	0,345022619
EDUCACION	0,39231789	0,454085355	0,011491795	-0,046021607
ENTRETENCION	0,277314051	0,216017576	0,274345451	0,014987136
FARMACIAS	0,215550245	0,097505996	0,000977719	0,149243778
MEJ_HOGAR	0,037862312	-0,059121198	0,18797744	0,006831497
RECAUDACION	-0,07291539	0,095748915	-0,111142324	0,145582986
RESTAURANT	0,256782586	0,116835012	0,157038289	0,11926815
SALUD	0,146773036	0,164539251	0,130864819	0,005284836
SEGUROS	-0,095145574	-0,001935467	0,089732773	-0,062146236
SUPERMERCADOS	0,172596277	0,178180942	-0,051743114	0,172074578
TIENDAS_DPTO	0,691380185	0,019388123	0,193375346	-0,10560252
TRANSPORTE	0,030703153	2,800165155	0,183358949	0,172305762
VIAJES	0,11648415	0,12077421	1,207265571	-0,122525942
VIVIENDA	-0,149434071	0,133253996	-0,158249492	2,402092389
T_AB_CLASICA	0,247157609	0,583614733	1,340164144	0,637244501
T_AB_PREMIUM	0,094602524	0,402888987	1,04940983	0,467376113
ENG_PUNTOS	0,303051916	0,134260342	0,14479948	0,169775018

Anexo R: Coeficientes de la matriz de correlación del modelo probit multivariado de efectos específicos.

	AUTOMOTRIZ	COMBUSTIBLE	COMUNICACION	EDUCACION	ENTRETENCION	FARMACIAS	MEJ_HOGAR	RECAUDACION	RESTAURANT	SALUD	SEGUROS	SUPERMERCADOS	TIENDAS_DPTO	TRANSPORTE	VIAJES	VIVIENDA
AUTOMOTRIZ	1	0,250324749	-0,008560169	0,046371335	-0,00206961	0,084768233	0,138954583	0,083848256	0,11934684	0,019517294	0,146129959	0,159478608	0,027083669	0,041324011	0,038182513	0,13684534
COMBUSTIBLE	0,250324749	1	0,063757919	-0,047585182	0,116746182	0,239966947	0,125315974	0,043484331	0,275353949	0,143795394	0,008914462	0,356234126	0,044541482	0,070042314	0,037008178	0,157527441
COMUNICACION	-0,008560169	0,063757919	1	0,076050667	-0,154633754	0,03270862	0,011265771	0,04947592	0,062370613	0,035693729	-0,006101761	0,012369174	0,059149677	0,042172191	0,089093917	0,176001898
EDUCACION	0,046371335	-0,047585182	0,076050667	1	-0,005204175	-0,003238561	0,062904853	0,100644012	0,155686875	0,014532182	-0,042272446	-0,008440078	0,064281074	-0,060011009	0,070769236	0,123474336
ENTRETENCION	-0,00206961	0,116746182	-0,154633754	-0,005204175	1	0,077588771	0,03731249	0,001973925	0,22650867	0,122598525	-0,074696593	0,118732938	0,12522309	0,112783018	0,11160556	0,086844048
FARMACIAS	0,084768233	0,239966947	0,03270862	-0,003238561	0,077588771	1	0,107827979	0,039709647	0,242413671	0,299285448	0,056759129	0,324912511	0,165078068	0,04963286	0,075329762	0,147270344
MEJ_HOGAR	0,138954583	0,125315974	0,011265771	0,062904853	0,03731249	0,107827979	1	0,045523437	0,089902156	0,037959705	0,070590166	0,158753139	0,077928715	0,045384041	0,098869098	0,067889547
RECAUDACION	0,083848256	0,043484331	0,04947592	0,100644012	0,001973925	0,039709647	0,045523437	1	-0,042817863	0,054089989	0,048824764	0,002980118	0,007173051	-0,040022905	-0,025186539	0,012687349
RESTAURANT	0,11934684	0,275353949	0,062370613	0,155686875	0,22650867	0,242413671	0,089902156	-0,042817863	1	0,130243067	0,088949064	0,326233348	0,200493162	0,199447411	0,144028211	0,115573011
SALUD	0,019517294	0,143795394	0,035693729	0,014532182	0,122598525	0,299285448	0,037959705	0,054089989	0,130243067	1	-0,001297235	0,164312828	0,132737218	0,148213546	0,020442971	0,094698128
SEGUROS	0,146129959	0,008914462	-0,006101761	-0,042272446	-0,074696593	0,056759129	0,070590166	0,048824764	0,088949064	-0,001297235	1	0,03200155	0,169054373	0,033516554	0,001006064	-0,009982394
SUPERMERCADOS	0,159478608	0,356234126	0,012369174	-0,008440078	0,118732938	0,324912511	0,158753139	0,002980118	0,326233348	0,164312828	0,03200155	1	0,169054373	0,065244275	0,032181543	0,158425283
TIENDAS_DPTO	0,027083669	0,044541482	0,059149677	0,064281074	0,12522309	0,165078068	0,077928715	0,007173051	0,200493162	0,132737218	0,098861471	0,169054373	1	0,102699128	0,095435941	0,015495591
TRANSPORTE	0,041324011	0,070042314	0,042172191	-0,060011009	0,112783018	0,04963286	0,045384041	-0,040022905	0,199447411	0,148213546	0,033516554	0,065244275	0,102699128	1	0,068572718	0,084782015
VIAJES	0,038182513	0,037008178	0,089093917	0,070769236	0,11160556	0,075329762	0,098869098	-0,025186539	0,144028211	0,020442971	0,001006064	0,032181543	0,095435941	0,068572718	1	-0,006199989
VIVIENDA	0,13684534	0,157527441	0,176001898	0,123474336	0,086844048	0,147270344	0,067889547	0,012687349	0,115573011	0,094698128	-0,009982394	0,158425283	0,015495591	0,084782015	-0,006199989	1

Anexo S: Coeficientes de la matriz de correlación del modelo probit multivariado de ordenamiento.

	FARMACIAS	EDUCACION	TIENDAS_DPTO	SUPERMERCADOS	MEJ_HOGAR	RECAUDACION	ENTRETENCION	COMUNICACION	COMBUSTIBLE	RESTAURANT	SALUD	AUTOMOTRIZ	VIAJES	TRANSPORTE	VIVIENDA	SEGUROS
FARMACIAS	1,000	-0,001	0,177	0,330	0,116	0,039	0,090	0,020	0,236	0,264	0,313	0,097	0,074	0,053	0,148	0,031
EDUCACION	-0,001	1,000	0,034	0,042	0,044	0,105	0,000	0,104	-0,063	0,125	-0,024	0,092	0,068	-0,029	0,099	-0,062
TIENDAS_DPTO	0,177	0,034	1,000	0,174	0,081	-0,004	0,140	0,060	0,046	0,208	0,143	0,026	0,097	0,094	0,021	0,094
SUPERMERCADOS	0,330	0,042	0,174	1,000	0,163	0,000	0,125	0,013	0,350	0,330	0,186	0,167	0,028	0,056	0,166	0,030
MEJ_HOGAR	0,116	0,044	0,081	0,163	1,000	0,051	0,036	0,000	0,132	0,092	0,048	0,152	0,088	0,032	0,064	0,066
RECAUDACION	0,039	0,105	-0,004	0,000	0,051	1,000	0,016	0,051	0,029	-0,046	0,058	0,090	-0,040	-0,045	0,010	0,063
ENTRETENCION	0,090	0,000	0,140	0,125	0,036	0,016	1,000	-0,170	0,109	0,227	0,129	-0,005	0,122	0,114	0,099	-0,062
COMUNICACION	0,020	0,104	0,060	0,013	0,000	0,051	-0,170	1,000	0,049	0,071	0,024	-0,026	0,108	0,032	0,169	0,000
COMBUSTIBLE	0,236	-0,063	0,046	0,350	0,132	0,029	0,109	0,049	1,000	0,258	0,134	0,248	0,038	0,071	0,149	0,018
RESTAURANT	0,264	0,125	0,208	0,330	0,092	-0,046	0,227	0,071	0,258	1,000	0,157	0,124	0,151	0,187	0,144	0,078
SALUD	0,313	-0,024	0,143	0,186	0,048	0,058	0,129	0,024	0,134	0,157	1,000	0,034	0,033	0,146	0,068	-0,002
AUTOMOTRIZ	0,097	0,092	0,026	0,167	0,152	0,090	-0,005	-0,026	0,248	0,124	0,034	1,000	0,031	0,051	0,137	0,144
VIAJES	0,074	0,068	0,097	0,028	0,088	-0,040	0,122	0,108	0,038	0,151	0,033	0,031	1,000	0,035	0,008	0,001
TRANSPORTE	0,053	-0,029	0,094	0,056	0,032	-0,045	0,114	0,032	0,071	0,187	0,146	0,051	0,035	1,000	0,092	0,023
VIVIENDA	0,148	0,099	0,021	0,166	0,064	0,010	0,099	0,169	0,149	0,144	0,068	0,137	0,008	0,092	1,000	-0,023
SEGUROS	0,031	-0,062	0,094	0,030	0,066	0,063	-0,062	0,000	0,018	0,078	-0,002	0,144	0,001	0,023	-0,023	1,000

Anexo T: Coeficientes estimados mediante el modelo probit multivariado de ordenamiento.

PARÁMETROS	RESTAURANT	SUPERMERCADOS	COMBUSTIBLE	FARMACIAS
GAMA	0,095381059	0,079233225	0,066729342	0,066662472
BETA	3,186114194	1,783637166	2,393258848	2,217657251
DELTA	1,114388616	1,164516243	1,728591914	1,150187198
T_AB_CLASICA	1,149096902	0,444542847	0,338281158	0,363883815
T_AB_PS	0,909259648	0,211862586	0,205244602	0,266401278
ENG_PUNTOS	0,16904643	0,391176313	0,307634817	0,318632303

PARÁMETROS	SALUD	AUTOMOTRIZ	ENTRETENCION	EDUCACION
GAMA	0,059454585	0,057047872	0,046127845	0,044981977
BETA	2,699335365	3,126073069	2,734238874	3,415617
DELTA	1,131093249	0,947489639	1,92160919	2,275707385
T_AB_CLASICA	0,502250582	0,759068466	0,763459031	0,350061914
T_AB_PS	0,502943406	0,800799425	0,619530522	0,402714612
ENG_PUNTOS	0,215699179	0,162303583	0,101890998	0,2856892

PARÁMETROS	VIVIENDA	TIENDAS_DPTO	TRANSPORTE	VIAJES
GAMA	0,043719088	0,042525217	0,041742754	0,039294615
BETA	2,779830282	0,821552964	2,814979045	3,18010935
DELTA	2,457253151	0,703231669	2,795970754	1,25238177
T_AB_CLASICA	0,545783932	0,282787731	0,590304455	1,096578864
T_AB_PS	0,378820672	0,118858472	0,400172507	0,770303631
ENG_PUNTOS	0,184315545	0,294194186	0,126598254	0,128692205

PARÁMETROS	COMUNICACION	MEJ_HOGAR	SEGUROS	RECAUDACION
GAMA	0,035214375	0,024591076	0,012553552	0,011227444
BETA	2,255318419	1,151493909	1,390236616	1,979491074
DELTA	2,777307121	0,678828134	3,281682786	3,086935372
T_AB_CLASICA	0,377589728	0,269763789	0,027281729	0,003542458
T_AB_PS	0,193061774	0,037264019	0,010689332	-0,066258531
ENG_PUNTOS	0,14660696	0,369996036	-0,007246742	0,265983797

Anexo U: Características de los clientes a los que se les asigna cada rubro según el criterio del lift.

Rubro	Cantidad de clientes	Gasto último año móvil	% con mail	Edad	Cupo	Recency (meses)	% con canje en el último año	% con ingresos a la web en el último año
Automotriz	81.633	\$ 4.128.544	72%	46	\$ 1.948.543	0,1	60%	47%
Combustible	103.789	\$ 2.380.032	66%	47	\$ 1.689.830	0,2	55%	39%
Comunicación	69.752	\$ 2.787.745	74%	44	\$ 1.574.100	0,3	47%	49%
Educación	16.737	\$ 14.946.643	85%	43	\$ 3.550.881	0,1	81%	82%
Entretención	35.826	\$ 1.807.984	77%	41	\$ 1.173.109	0,2	40%	53%
Farmacias	132.096	\$ 2.108.436	69%	45	\$ 1.291.979	0,3	56%	43%
Mejoramiento del Hogar	267.013	\$ 865.734	63%	43	\$ 939.675	0,8	25%	30%
Recaudación	49.053	\$ 3.774.093	78%	42	\$ 1.781.807	0,4	67%	57%
Restaurant	82.419	\$ 2.928.146	72%	46	\$ 1.588.075	0,3	53%	47%
Salud	36.665	\$ 3.897.395	73%	46	\$ 1.708.421	0,2	69%	51%
Seguros	954.042	\$ 300.217	49%	51	\$ 1.124.690	13,8	6%	13%
Supermercados	361.986	\$ 1.345.753	63%	48	\$ 1.370.318	0,4	40%	35%
Tiendas por departamento	652.265	\$ 691.295	56%	49	\$ 1.158.669	3,4	19%	28%
Transporte	39.106	\$ 5.062.551	79%	45	\$ 2.120.545	0,1	66%	59%
Viajes	32.027	\$ 4.379.636	83%	42	\$ 2.052.030	0,1	62%	59%
Vivienda	46.968	\$ 4.944.457	80%	42	\$ 1.957.731	0,2	66%	61%

Anexo V: Resultados de la estimación del modelo logit de respuesta sobre la campaña de restaurantes.

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-6.950675	0.335887	-20.693	< 2e-16	***
MAIL	0.346737	0.221586	1.565	0.1176	
CUPO_700	0.469399	0.230949	2.032	0.0421	*
EDAD	-0.022032	0.003713	-5.933	2.97e-09	***
T_ABIERTA_PREMIUM	0.969397	0.091387	10.608	< 2e-16	***
ENGANCHADO_PUNTOS	-0.448371	0.092057	-4.871	1.11e-06	***
COMUNICACION_L1	-0.405600	0.100086	-4.053	5.07e-05	***
RESTAURANT_L1	0.663698	0.372389	1.782	0.0747	.
TRANSPORTE_L1	0.681016	0.097392	6.993	2.70e-12	***
MAIL:RESTAURANT_L1	0.657012	0.379522	1.731	0.0834	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					