



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

EVALUACIÓN DE LA EFECTIVIDAD DE PROMOCIONES DE DESCUENTO EN PRECIO CON MEDIO PREFERENTE DE PAGO EN UNA TIENDA POR DEPARTAMENTOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

MAGDALENA VICTORIA MUÑOZ SALAS

PROFESOR GUÍA:
RICARDO MONTOYA MOREIRA

MIEMBROS DE LA COMISIÓN:
ALEJANDRA PUENTE CHANDÍA
LUIS ABURTO LAFOURCADE

SANTIAGO DE CHILE
2018

RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE: Ingeniera Civil Industrial
POR: Magdalena Muñoz Salas
FECHA: 10/09/2018
PROFESOR GUÍA: Ricardo Montoya Moreira

EVALUACIÓN DE LA EFECTIVIDAD DE PROMOCIONES DE DESCUENTO EN PRECIO CON MEDIO PREFERENTE DE PAGO EN UNA TIENDA POR DEPARTAMENTOS

Las promociones han sido ampliamente estudiadas en la literatura, sin embargo, no se han encontrado investigaciones de su interacción con el medio de pago de los clientes. Este tema se vuelve interesante en un contexto como el retail chileno, donde los principales competidores cuentan con negocios financieros, y tienen entre sus prácticas comerciales el uso de promociones focalizadas al medio de pago propio del holding.

El objetivo de esta memoria es evaluar el efecto de promociones de descuento con medio preferente de pago en el comportamiento de los clientes, específicamente en tres metas del negocio, y generar recomendaciones para la configuración óptima de promociones. Para esto, se comparan diferentes tipos de regresiones y se realiza análisis de escenarios para encontrar las características óptimas de los descuentos, utilizando la información transaccional de una tienda por departamentos nacional entre 2012 y 2016.

Primero, se estudia el efecto en el volumen de ventas diarias con tarjeta, para 16 sublíneas de la categoría de productos Colchones. Los modelamientos de la cantidad de productos vendidos con tarjeta y la proporción que esta cantidad significa de las ventas totales de la sublínea, se realizan utilizando Regresiones de Poisson y Lineal, respectivamente. Los resultados muestran que si se desea maximizar las utilidades generadas por las unidades vendidas, una mayor amplitud de las promociones dentro de las sublíneas estudiadas parece ser una estrategia adecuada. Sin embargo, se debe tener cuidado con la profundidad de estos descuentos, dado que no muestran tener un efecto positivo en todas las sublíneas.

Luego, se agrupan las ventas a nivel de boleta, considerando solo las primeras compras de los clientes, y se construyen variables de la profundidad de los descuentos obtenidos, en cada oportunidad de compra, en 6 categorías de interés. En el segundo análisis se evalúa el efecto en la apertura de tarjetas de nuevos clientes, a través de una Regresión Logística. Del análisis se obtiene que si se quiere atraer a un mayor número de clientes nuevos que abran la tarjeta del retail, el nivel de las rebajas tiene siempre un efecto positivo, en mayor o menor medida, dependiendo de la categoría donde se aplican. Tercero, se estudia el efecto en el valor de los clientes (margen acumulado) atraídos al negocio, a través de una Regresión Lineal. Se obtiene que los descuentos en las categorías Televisión y Video, Computación, Línea Blanca y Telefonía, atraen a peores clientes a medida que aumenta su profundidad, por lo que se debe evitar aplicar promociones en estas categorías, si lo que se quiere es atraer a clientes más rentables.

Para la configuración óptima de las promociones, se concluye que el uso o no de descuentos, su profundidad y el tipo de productos donde se aplican, dependerá de la meta del negocio perseguida. Destaca el hecho de que los descuentos observados en la realidad están siempre por encima de los descuentos óptimos estimados para las diferentes metas. Esto estaría generando pérdidas importantes para el retail.

Como trabajo futuro se sugiere incluir información de los ingresos por parte del negocio financiero por el uso de la tarjeta propia, para poder concluir si las pérdidas estimadas en el retail se ven subsidiadas por las ganancias en el negocio financiero.

AGRADECIMIENTOS

Al finalizar esta memoria, y con ella, un largo proceso de aprendizaje y esfuerzo, me gustaría agradecer a todas las personas que me han acompañado.

Primero, a mi familia, quienes me han apoyado durante toda la vida y son un pilar fundamental para mí. Mis padres, Fátima y Alejandro, han sido un ejemplo de esfuerzo y sacrificio, y a quienes les debo ser la persona que soy hoy en día.

Al cuarteto, un grupo de maravillosas ingenieras que me acompañaron día a día en esta aventura, por todos los traspasos estudiando, carreando, hablando y resolviendo la vida. La Vale, mi compañera de viajes y en quien encontré una persona que me entiende hasta en mis divagaciones más locas. Dafne y Belén, a quienes conocí en primer año y que han estado siempre para sacarme una sonrisa cuando lo he necesitado.

A la Jose, quien me ha demostrado que no importa cuanto tiempo pase, las verdaderas amistades son incondicionales. A Gabino, quien me presentó un mundo increíble de montañismo y escalada, y tuvo la paciencia de enseñarme la mayoría de lo que sé de esos deportes. A mis amigos de sección, Felipe y Claudio, que aunque no los vea seguido siempre es un momento agradable compartir con ellos.

Al profesor Ricardo, quien a través de sus cursos me ayudo a darme cuenta que era lo que más me gustaba de la carrera, y quien me acompañó como profe guía en esta "última tarea", gracias por la paciencia y el tiempo invertido. A la profesora Alejandra, quien fue mi jefa, co-guía y amiga, muchas gracias por el apoyo y la buena onda.

TABLA DE CONTENIDO

1. Antecedentes generales	1
1.1. Introducción.....	1
1.2. Brecha del conocimiento.....	1
2. Objetivos y alcance de la memoria	3
2.1. Objetivo General	3
2.2. Objetivos Específicos.....	3
2.3. Alcances.....	3
3. Marco Conceptual.....	4
3.1. Promociones	4
3.1.1. Efectos de las promociones	4
3.1.2. Promociones focalizadas	6
3.1.3. Tarjetas de crédito y promociones	7
3.2. Valoración de un cliente.....	8
4. Metodología.....	9
4.1. Método de investigación	9
4.2. Información disponible	9
4.3. Herramientas de investigación	10
4.3.1. Estimación del efecto de las promociones en demanda	10
4.3.2. Comparación de modelos	11
4.3.3. Valorización de los clientes.....	14
4.3.4. Cálculo de la promoción óptima.....	14
5. Desarrollo Metodológico	15
5.1. Efecto de las promociones en las unidades vendidas.....	15
5.1.1. Preparación de la base de datos	15
5.1.2. Ventas por sublínea	17
5.1.3. Modelamiento	21
5.1.4. Resultados y análisis	25
5.1.5. Configuración óptima de las promociones (OE4).....	30
5.1.6. Conclusiones del Objetivo Específico 1	35
5.2. Efecto de las promociones en la apertura de tarjetas propias.....	37
5.2.1. Preparación de la base de datos	37
5.2.2. Primeras compras de los clientes	39
5.2.3. Modelamiento	44
5.2.4. Resultados y análisis	48

5.2.5.	Configuración óptima de las promociones (OE4).....	51
5.2.6.	Conclusiones del Objetivo Específico 2	53
5.3.	Efecto de las promociones en la atracción de nuevos clientes	55
5.3.1.	Preparación de la base de datos	55
5.3.2.	Valor de los clientes	56
5.3.3.	Modelamiento	59
5.3.4.	Resultados y análisis	62
5.3.5.	Conclusiones del Objetivo Específico 3	65
6.	Conclusiones	67
6.1.	Trabajo futuro.....	68
7.	Bibliografía.....	69
8.	Anexos.....	71
8.1.	ANEXO A. Información adicional Objetivo Específico 1	71
8.1.1.	Preparación de la base de datos	71
8.1.2.	Ventas por sublínea	72
8.1.3.	Modelamiento	76
8.1.4.	Resultados y análisis	78
8.2.	ANEXO B. Información adicional Objetivo Específico 2	80
8.2.1.	Primeras compras de los clientes	80
8.2.2.	Modelamiento	81
8.3.	ANEXO C. Información adicional Objetivo Específico 3.....	83
8.3.1.	Valor de los clientes	83
8.3.2.	Modelamiento	83

1. ANTECEDENTES GENERALES

1.1. Introducción

El comercio en Chile es uno de los sectores más importantes a nivel económico, representando el 24,9% de las ventas totales del país y siendo el sector con más alta participación en la economía nacional (Unidad de Estudios Ministerio de Economía, Fomento y Turismo, 2014). Dentro de este sector económico se encuentra el subsector del retail, el que engloba empresas especializadas en la comercialización masiva de productos o servicios uniformes a gran cantidad de clientes. Aquí se incluyen todas las tiendas o locales comerciales que habitualmente se encuentran en cualquier centro urbano con venta directa al público, por ejemplo, rubros como supermercados, tiendas por departamentos, casas de artículos para el hogar, ferreterías, farmacias, entre otras.

En la mayoría de estos rubros, los principales competidores no son muchos. Por lo general se encuentran encabezados por 2 o 3 empresas que concentran la mayor parte de las ventas. Ejemplo de esto se da en el rubro de tiendas por departamento, donde los 3 principales competidores reúnen el 80% de la participación de mercado (CENCOSUD, 2016).

Un elemento interesante de las grandes empresas que lideran estos rubros es que suelen ser holdings que tienen negocios en más de un sector económico. La evolución de estos holdings ha sido relativamente similar, creciendo en el mundo del retail y especializándose en cada uno de sus servicios, al mismo tiempo que aprovechan las sinergias que surgen de un holding con presencia en diversas áreas del comercio.

Otro elemento similar es la expansión horizontal que han tenido varias de estas empresas hacia los productos financieros como las tarjetas de crédito. Esta situación es inusual si se compara con la configuración del comercio en otras partes del mundo, donde los dueños de los centros comerciales, supermercados u otros, suelen ser diferentes a los emisores de los medios de pago diferentes al efectivo.

1.2. Brecha del conocimiento

La configuración empresarial del retail nacional recién descrita hace que surjan nuevas sinergias interesantes de explotar, como la transferencia de clientes desde sus negocios más consolidados a sus nuevos productos bancarios. Para ello, las técnicas utilizadas hasta ahora por los comerciantes han sido variadas, como la publicidad en sus diferentes canales, las promociones focalizadas, los programas de puntos y beneficios, entre otros.

En términos de promociones focalizadas, estas suelen diseñarse restringiéndolas solo al pago con tarjetas propias de la marca, incentivando su apertura y uso, o pueden tener diferentes condiciones (como porcentaje de descuento) según el medio de pago. Estas promociones están diseñadas bajo el supuesto de que ésta herramienta de marketing será capaz de atraer nuevos clientes a los productos financieros, a través de la apertura de nuevas tarjetas, además de poder promover la costumbre del pago con ellas, una vez abierta la cuenta.

Sin embargo, al momento de revisar literatura disponible de promociones que sustente el supuesto de diseño, ésta es acotada y casi inexistente. La investigación sobre promociones ha sido basta y extensa, abordando temas como la efectividad de esta herramienta en el aumento de ventas según sus características, su impacto en otras líneas, las diferentes fuentes del incremento en ventas y sus efectos a largo plazo, desde el punto de vista teórico y empírico, pero no se han planteado temas relacionados con el medio de pago utilizado por los clientes.

Por lo tanto, se ha detectado una brecha de conocimiento en cuanto a la utilización de promociones para incentivar la apertura de tarjetas y el cambio en el medio de pago. Esta brecha puede deberse a lo inusual de la configuración económica nacional, que solo se repite en contadas empresas en otros países latinoamericanos, por lo que no es un problema que ha llamado la atención de los investigadores hasta el momento.

La investigación aquí presentada busca contribuir a la superación de dicha brecha, analizando el desempeño de promociones focalizadas en el medio de pago en diferentes aspectos.

2. OBJETIVOS Y ALCANCE DE LA MEMORIA

2.1. Objetivo General

Evaluar el efecto de promociones de descuento con medio preferente de pago, en el comportamiento de los clientes.

2.2. Objetivos Específicos

Los objetivos específicos (OE) que se abordan en esta memoria son:

- OE1.** Evaluar el efecto de promociones de descuento con medio preferente de pago, en las unidades vendidas por subcategoría
- OE2.** Evaluar el efecto de promociones de descuento con medio preferente de pago, en la apertura de tarjetas propias de la tienda por departamento de nuevos clientes
- OE3.** Evaluar el efecto de promociones de descuento con medio preferente de pago, en la atracción de nuevos consumidores en términos del valor del cliente
- OE4.** Diseñar recomendaciones sobre las características óptimas de las promociones con medio preferente de pago

En adelante, se refiere a estos objetivos específicos bajo la sigla OE.

2.3. Alcances

En este proyecto se evalúan promociones de descuento en precio, aplicables con medio preferente de pago, es decir, con tarjeta de crédito del mismo holding de la tienda por departamentos, diferenciando su efecto según la categoría – sublínea donde se apliquen dichos descuentos. Para el primer objetivo específico, el modelamiento se realizará a nivel de subcategoría de productos, teniendo en consideración los datos transaccionales del retail desde 2014 hasta mediados del 2016. Las sublíneas consideradas corresponden a 16 sublíneas de la categoría Colchones. En el caso del segundo objetivo específico, se trabaja con la información de las primeras compras de todos los clientes registrados, producidas desde principios del 2013 hasta mediados del 2016. En este caso se analiza el impacto de las promociones de 6 categorías específicas. Por último, para el objetivo específico número tres, se trabaja con el margen acumulado de todas las compras por cliente identificado, valor del cliente, y se contrasta esta información con el contenido de sus primeras compras. Para este objetivo, se consideran las primeras compras producidas entre 2013 y 2015.

No se considera una etapa de implementación de los modelos en el diseño de nuevas promociones, debido a que este es un trabajo independiente de investigación. Es importante destacar que, por las características del proyecto y los datos disponibles, los resultados obtenidos no serán del tipo causal, pudiendo solo establecer relaciones entre las métricas estudiadas.

3. MARCO CONCEPTUAL

3.1. Promociones

La promoción de ventas es comúnmente definida como “incentivos a corto plazo para fomentar la compra o venta de un producto o servicio.” (Kotler & Armstrong, 2012). Este fomento de las compras puede tener como objetivo potenciar el desarrollo de una nueva marca y/o fortalecer una marca existente, a través de la creación de expectativas a corto plazo y la formación de relaciones a largo plazo con los clientes.

Existe una gran variedad de herramientas para la promoción de ventas, las cuales, en base a qué actor de la línea productiva están enfocadas, se pueden clasificar en tres tipos:

- Promociones para consumidores: el objetivo principal de estas promociones es motivar las ventas a corto plazo o mejorar la participación del cliente.
- Promociones comerciales: se utilizan para persuadir a los distribuidores para que vendan una marca, le otorguen espacio en los locales, la promocionen y la acerquen a los consumidores.
- Promociones para negocios: se utilizan para generar contactos de negocios, estimular las compras, recompensar a los clientes y motivar a los vendedores.

En el contexto de esta memoria la investigación se enfocará en promociones para consumidores, específicamente del tipo descuento en precio. Estas promociones pueden diferenciarse en base a quienes están enfocadas: a todos los clientes, en general para aumentar el nivel de ventas; o solo a un conjunto limitado de ellos, con objetivos más específicos como la retención, fidelización o atracción de clientes.

3.1.1. Efectos de las promociones

Los efectos de las promociones han sido altamente estudiados durante las últimas 3 décadas, concluyendo que estas pueden provocar diferentes respuestas en los consumidores, dependiendo de innumerables factores, tanto desde el punto de vista de diseño, como en donde se aplican.

Ya en 1995, Blattberg, Briesch y Fox muestran, a través de un metanálisis, que existía un amplio consenso de que las promociones provocan un aumento inmediato de las ventas (Blattberg, Briesch, & Fox, 1995). Efectos similares pueden verse reflejados en importantes trabajos posteriores, como en el de Ailawadi y Neslin, quienes establecieron que las promociones inducen a los consumidores a comprar más y consumir más rápido (Ailawadi & Neslin, 1998). Para llegar a este resultado, los autores utilizaron datos de panel de ventas de yogurt y ketchup, y generaron un modelo de incidencia, elección y cantidad, en el que el consumo de cada categoría varía con el nivel de inventario de los hogares.

Como se mencionó anteriormente estos efectos en las ventas no son siempre iguales. En 1996, se identificaron diferencias en su efectividad dependiendo del número de marcas por categoría, la penetración de la categoría donde se aplica la promoción, tiempo entre compras y la tendencia que tengan los clientes a acumular los productos.

Para ello se estudiaron 108 categorías de productos a través de análisis de regresiones (Neslin, Narasimhan, & Sen, 1996).

También se han encontrado diferencias significativas en la efectividad de las promociones a partir de la frecuencia, tanto en el corto como en el largo plazo, las características del mercado y las características de los productos (por ejemplo, en productos perecibles). Para encontrar estas diferencias, se usaron datos de promociones, su publicidad, distribución y nuevos artículos de 560 categorías de productos de consumo, durante un período de 4 años, correspondientes a datos de un supermercado holandés. Las metodologías aplicadas para testear estas diferencias fueron raíz unitaria, para determinar si la demanda evolucionaba, y modelos de vectores autorregresivos con variables exógenas, para estimar los efectos de las promociones a corto y largo plazo (Nijs, Dekimpe, Steenkamps, & Hanssens, 2001).

Pero si las promociones pueden tener efectos distintos dependiendo de sus características, ¿podrían llegar a tener efectos negativos? Con respecto a esto Kotler y Armstrong recalcan: “Lo importante es que la fijación de precios promocional pueda ser un medio eficaz para generar ventas para ciertas empresas, en ciertas circunstancias. Sin embargo, podría ser dañino para otras compañías, o si se realiza de forma constante” (Kotler & Armstrong, 2012).

Un ejemplo de esto se puede encontrar en el trabajo de Ailawadi et al. del año 2006. En este se analiza el efecto individual de las promociones ofrecidas por una farmacia estadounidense durante el año 2013 utilizando información transaccional, información de los clientes del programa de lealtad de la farmacia y datos internos de la compañía. Dentro del análisis se realizó una cuantificación del aumento de ventas por la promoción; se descompuso este aumento entre sus diferentes fuentes; se estimó el grado en que la promoción afectaba las ventas de otras categorías de productos en las tiendas; se calcularon los márgenes promocionales y no promocionales y se calculó la financiación del fabricante a las promociones para calcular el impacto de la promoción en la ganancia neta. Finalmente se realizó la examinación de cómo la promoción, la marca, la categoría y las características de la tienda influían en el impacto neto. Uno de sus principales hallazgos fue que, en promedio, el beneficio neto de las promociones era negativo, esto debido a que, a pesar del alto impacto en el volumen de ventas, los márgenes promocionales se veían tan reducidos que la utilidad final se veía altamente perjudicada. (Ailawadi, Harlem, César, & Trounce, 2006).

A pesar de que no se tratarán directamente en esta investigación, vale la pena destacar otros efectos negativos de las promociones, debido a la frecuencia de éstas y por sus consecuencias en el largo plazo. Mela et al. (1997) muestran que las promociones de larga duración pueden generar un aumento de la sensibilidad al precio por parte de los consumidores, e incluso una reducción de su percepción del valor de la marca. A esto, DelVecchio et al. (2006) añaden que, aunque es cierto que las promociones inducen compra en el corto plazo, pueden eventualmente reducir preferencia por una marca cuando ésta no se encuentra en oferta. A partir del metaanálisis de Blattberg et al. (1995), se puede concluir que existe consenso en la literatura con respecto a que las promociones recurrentes pueden cambiar el precio de referencia de los consumidores, y

que, a mayor frecuencia de las promociones, menor es el efecto que estas tienen en el aumento de ventas.

Entonces, ya habiendo establecido el aumento de ventas provocados por las promociones, el siguiente tema de interés es saber cuál es la fuente de este aumento en ventas. Normalmente, en trabajos como Gupta (1988) y Van Heerde et al. (2003), entre otros, se reconocen 3 fuentes del aumento en ventas debido a las promociones:

- Cambio de marca: Esto implica que los consumidores dejan de comprar los productos en sus marcas habituales y prefieren los que se encuentran en promoción. Esto puede tener un efecto a corto plazo, volviendo a comprar en la marca habitual pasada la promoción, o un efecto a largo plazo, cambiando definitivamente de marca, aún después de terminado el periodo. Este efecto es deseado por los proveedores cuyos productos están en promoción, pero no por los retailers, dado que no aumenta las ventas en la categoría de productos.
- Aceleración del consumo con acumulación: Implica que los clientes adelanten su compra en vista de la promoción, y guarden los productos para ser consumidos cuando la necesidad esté presente. Este efecto tiene consecuencias negativas una vez terminado el periodo promocional, debido a que las personas disminuyen su consumo posterior dado que tienen productos acumulados. Generalmente esto no es deseado ni por productores ni retailers, excepto cuando se busca reducir el inventario de algún producto, lo que podría ocurrir en términos de temporada o cuando se descontinúa una línea, entre otras ocasiones.
- Aumento del consumo: Esto significa que los consumidores compran un volumen mayor del producto en promoción sin almacenarlo. Este es un efecto positivo tanto como para retailers como para productores.

Estas fuentes del aumento del volumen pueden identificarse y cuantificarse, tal como se hace en trabajos como los de Ailawadi et al. (2006), descrito anteriormente, Gupta (1988) y Van Heerde et al. (2003).

Ya se ha establecido la importancia del estudio de las promociones como herramienta del marketing mix, por su potencial impacto en las ventas y sus diferentes efectos totales en los negocios, tanto positivos como negativos. Además, se ha establecido que las promociones podrían potencialmente provocar cambios de comportamiento en los clientes, como cambio de marca y aceleración del consumo. Un tema interesante por investigar, es si las promociones son capaces de cambiar otros comportamientos en los clientes, como, por ejemplo, el medio de pago utilizado. Para esto se debe, primero, ahondar en las promociones con un público limitado y el conocimiento disponible de estas, lo que se aborda en puntos que vienen a continuación.

3.1.2. Promociones focalizadas

Las promociones dirigidas a un público en específico pueden tener diferentes objetivos, tales como la retención de clientes, atracción de nuevos clientes, traspaso de clientes, etc. Es muy importante poder entender el poder que tienen las promociones para lograr cada uno de estos objetivos, pudiendo aprovecharlas de la mejor manera y sin provocar resultados negativos para la empresa.

Con respecto a estas promociones exclusivas, se ha mostrado en el último tiempo que son más eficientes que las promociones generales, ya que estas últimas brindan descuentos innecesarios a consumidores insensibles a los precios (Acquisti & Varian, 2005). Del mismo modo que se han mostrado efectos positivos de las promociones focalizadas, se han detectado posibles efectos negativos en los clientes, con trabajos como los realizados por Barone y Roy en 2010, donde, a partir de datos experimentales, establecen que las promociones exclusivas pueden ser percibidas de manera más, igual o menos favorable que las promociones para todos, dependiendo de diferentes características de los clientes (Barone & Roy, 2010).

En cuanto a efectos en la retención de los clientes, puede encontrarse que las promociones pueden llegar a ser armas de doble filo. Vogel y Paul estudiaron el efecto de promociones exclusivas por canal de venta de un retail en la retención de clientes, encontrando principalmente que, aunque la diferenciación de precios afectaba positivamente el valor percibido por los clientes, al mismo tiempo perjudicaba la retención por la sensación de injusticia en los precios y autodeterminación limitada (Vogel & Paul, 2015). Por lo tanto, a pesar de que las promociones exclusivas permiten optimizar los recursos utilizados y pueden tener efectos en el comportamiento de los clientes, éstas deben ser diseñadas con cuidado, evitando el despliegue de esfuerzos innecesarios y eventuales perjuicios para la empresa.

3.1.3. Tarjetas de crédito y promociones

Las tarjetas de crédito se han vuelto un importante medio de pago alternativo en la economía actual por variadas razones: comodidad, seguridad, oportunidad de liquidez, entre otras. Evidencia de esto es como algunos comerciantes comenzaron solo a aceptar pagos con tarjetas, por razones de seguridad y conveniencia. Otros, como restaurantes de servicio rápido y cafeterías ahora solo aceptan tarjetas de pago para capturar mayores ventas y aumentar la velocidad de las transacciones (Bolt & Chakravorti, 2008).

Además, se tienen beneficios directos para el comerciante por el uso de tarjeta. El llamado *efecto de la tarjeta de crédito* establece que los clientes gastan más dinero cuando pagan con ese medio que lo que gastarían si pagasen con efectivo, lo que ha recibido amplio apoyo en la literatura reciente (Chatterjee & Rose, 2012).

Por lo tanto, en un contexto donde se ha probado que el pago con tarjetas de crédito es beneficioso para el volumen de ventas y donde se tiene a las promociones como una herramienta de cambio en el comportamiento de los clientes, resulta interesante investigar si es posible que, a través de la creación de promociones focalizadas en el medio de pago, se puede incentivar el uso de las tarjetas de crédito.

Este problema se vuelve aún más interesante si el comerciante se puede beneficiar tanto del gasto directo en el comercio, como del uso de la tarjeta de crédito, como es el caso de los holdings dueños de ambas partes del proceso comercial, tanto las tiendas minoristas, como los emisores de las tarjetas de créditos (bancos). Hasta el momento no se ha encontrado literatura que se haga cargo de este tema, principalmente por el hecho de que esta situación, que emisores de tarjetas de créditos y tiendas minoristas pertenezcan al mismo holding, no es común en otras partes del mundo (EE.UU. Patente nº US 6,332,126 B1, 2001).

Sin embargo, como se describió en la sección 1.1. Introducción, esta situación es común en Chile, y ha surgido como un tema importante dentro del diseño de promociones. De esta manera, nace la interrogante de si promociones en sus negocios más consolidados, por ejemplo, en las tiendas por departamento, serían capaces de transferir clientes a las instituciones financieras con la apertura de tarjetas. Como se mencionó anteriormente, no se ha encontrado evidencia en la literatura de esto, por lo que existe un vacío de conocimiento importante para abordar e investigar. Es en ésta brecha identificada, donde hace su aporte ésta investigación.

3.2. Valoración de un cliente

La valoración de un nuevo cliente es muy importante dentro de una empresa, dado que ellos son los que la mantienen viva: “Los clientes—no los productos— son la fuente más importante de ingreso operacional” (Lehmann & Winer, 2004). Dentro del contexto de la investigación se espera poder establecer qué promociones atraen al mejor tipo de cliente, donde, el mejor tipo de cliente será el que genere más valor para la empresa, valor que no está declarado y que se requiere estimar.

El volumen de compra, frecuencia, tiempo de relación con la empresa y potencial poder comunicacional, son algunos de los factores que juegan un papel primordial en el valor que representa un nuevo cliente para la empresa. La mayoría de estos factores pueden ser cuantificados económicamente, aunque muchas veces esta tarea es altamente compleja. Para realizar esta cuantificación suelen usarse herramientas como el valor de vida del cliente (CLV), la que ha adquirido gran importancia en el último tiempo.

Para Reinartz y Kumar (2000) son 3 las principales razones de por qué CLV ha ganado tanto interés en el último tiempo. Primero se tiene el hecho de que las empresas están actualmente mayoritariamente enfocadas en el proceso de administración de los clientes, por lo que entenderlos en profundidad es extremadamente necesario. Por otro lado, se tiene que el Marketing Science Institute ha clasificado este tema como una investigación de prioridad en la creación de conocimiento. Por último, la tercera razón sería la poca evidencia empírica en el área del CLV (Reinartz & Kumar, 2000).

Otra herramienta disponible para la valorización de los clientes es el RFM, método analítico capaz de extraer conocimiento útil de los clientes a partir de datos de gran tamaño utilizando pocos criterios. Una de sus ventajas desde el punto de vista del comportamiento del consumidor según Schijns y Schroder (1996) es que es un método familiar para medir la fuerza de relación con el cliente (Schijns y Schroder, 1996).

En esta forma de valorizar al cliente se tiene en consideración tanto su frecuencia de compra, su gasto en la empresa y la distancia temporal entre sus transacciones, asumiendo que estas variables representan la importancia del cliente para la organización. Investigaciones han mostrado que mientras menor sea el tiempo desde la última compra y mayor sea la frecuencia de un cliente, es más probable que dicho cliente genere una nueva compra (Wu & Lin, 2005). También, se ha mostrado que mientras mayor sea el monto gastado por un cliente, más probable es que vuelva a comprar en la empresa (Newell, 1997).

4. METODOLOGÍA

4.1. Método de investigación

Para llevar a cabo una investigación que efectivamente llegue a resultados y a la obtención de conocimiento, es necesario mantener cierto orden lógico (metodología), que permita procesar y analizar la información de las fuentes disponibles.

En el ámbito de las investigaciones relacionadas con la minería de datos, destaca la metodología Knowledge Discovery in Database (KDD). Esta es una metodología basada en una serie de pasos para, como su nombre lo dice, obtener información relevante de una base de datos. Cada una de sus etapas se detallan a continuación.

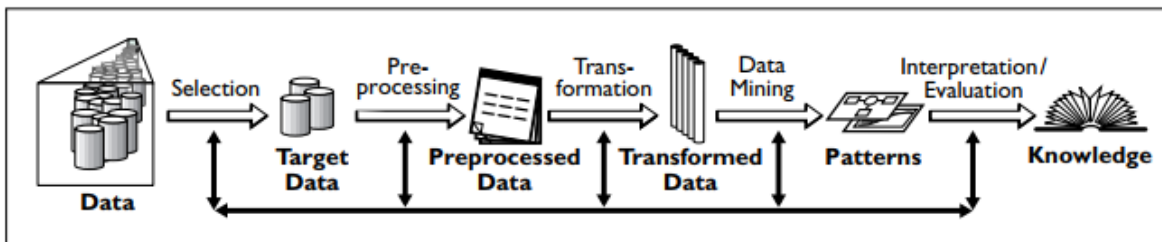


Ilustración 1: Pasos del proceso KDD
Fuente: Fayyad, Piatetsky-Shapiro, & Smyth (1996)

1. Selección de los datos: Corresponde a la recolección y extracción de los datos que se consideren relevantes para la realización de la investigación.
2. Preprocesamiento de la base: Primera exploración de los datos en búsqueda de posibles errores en ella.
3. Transformación de los datos: Se realizan las agregaciones o conversiones que se estimen convenientes en base a los objetivos perseguidos y los modelos a aplicar.
4. Minería de datos: Construcción y aplicación de los modelos escogidos.
5. Interpretación y evaluación: Análisis de los resultados obtenidos. Aquí se obtienen las conclusiones de la investigación.

Esta es la metodología seguida para este trabajo, adaptando los pasos a los requerimientos específicos del tema, y utilizando las herramientas que se detallan próximamente.

4.2. Información disponible

Para este trabajo de investigación se utiliza información de ventas de una tienda por departamentos nacional, que cuenta con una tarjeta de crédito de la misma marca, y que tiene entre sus prácticas de marketing la realización de promociones focalizadas para promover su negocio financiero.

La información disponible consta de 10 tablas de datos: una transaccional, dos con la información de los clientes, seis con información de los productos y su categorización (área, línea, sublínea, etc.) y una con la información de los puntos de venta.

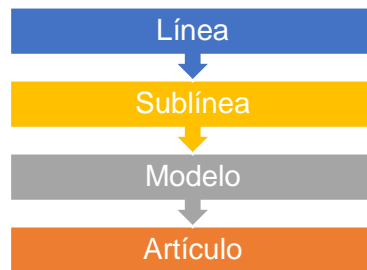


Ilustración 2: Agrupación de SKUs
Fuente: Elaboración propia

La tabla transaccional contiene la información de todas las ventas realizadas desde enero del 2012 hasta mediados de junio del 2016. Esta información se encuentra agregada a nivel de artículo vendido y consiste en 260MM de registros. Se tienen 2 tablas de información de clientes, una de ellas cuenta con 2.6MM de registros y contiene datos demográficos y de posesión de tarjeta propia, mientras la otra contiene la estimación del valor del cliente para la empresa (RFM) para una submuestra de 1.5MM de consumidores. Las tablas de información de los productos cuentan con las características de 4.8MM de SKUs diferentes, los cuales se encuentran organizados en grupos jerárquicos como se muestra en la Ilustración 2.

Con la información disponible se realiza un análisis estadístico para poder evaluar el efecto de las promociones en cada uno de los tópicos propuestos, obteniendo como resultado final las características óptimas de las promociones con medio preferente de pago, dependiendo del objetivo perseguido, y a partir de las cuales se podrán generar recomendaciones para futuras promociones.

4.3. Herramientas de investigación

En esta sección se exponen las diferentes herramientas utilizadas para abordar los objetivos propuestos. En el futuro es posible que se agreguen otras herramientas que sean necesarias en el desarrollo de la investigación.

4.3.1. Estimación del efecto de las promociones en demanda

OE1 y OE2 buscan explorar los efectos de los descuentos en la demanda de productos del retail y productos bancarios, respectivamente. A continuación, se presentan brevemente 4 opciones utilizadas para modelar la situación estudiada y estimar la demanda. Para más información sobre cada una de estas herramientas, se sugiere al lector consultar el libro *Data Analysis Using Regression and Multilevel/Hierarchical Models* de Andrew Gelman y Jennifer Hill (Gelman & Hill, 2007).

Regresión Lineal

Este modelo relaciona una variable dependiente, $Y \in \mathbb{R}$, con K variables explícitas, X_k ($k = 1, \dots, K$), o alguna transformación de estas últimas. La relación que se establece es del tipo lineal, más una perturbación aleatoria ε que recoge todos aquellos factores de la realidad no controlables u observables. Este modelo se estima a través de *mínimos cuadrados ordinarios* (MCO).

$$y = X\beta + \varepsilon$$

Regresión de Poisson

Este tipo de regresión se utiliza para modelar datos de recuento, tales como unidades compradas. En ella se asume que la variable dependiente (y) sigue una distribución de Poisson, y que las variables independientes (X) tienen una relación lineal con el logaritmo del parámetro de y .

$$y \sim \text{Poisson}(\lambda), \quad \lambda = \exp(X\beta + \varepsilon)$$

Para la estimación de este modelo se utiliza comúnmente el método de *maximización de la verosimilitud*, en adelante MLE por sus siglas en inglés. Los coeficientes obtenidos de esta estimación indican en que porcentaje cambia y (por ej. la demanda) ante la variación de una unidad de la variable independiente (por ej. la profundidad del descuento).

Regresión Binomial

En esta regresión se modela el número de éxitos y_i dentro de un número n_i de intentos. Para este modelo también se utiliza la función de enlace logit para la probabilidad de éxito, asumiendo que los éxitos siguen una distribución binomial.

$$\begin{aligned} \text{logit}(p_i) &= X\beta + \varepsilon \\ y_i &\sim \text{Bin}(n_i, p_i), \quad f_Y(y) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i} \end{aligned}$$

En este modelo también se utiliza MLE para la estimación de los parámetros.

Regresión Logística

La regresión logística se utiliza para modelar variables con resultados dicotómicos, usualmente llamados clases. En este caso, se estima la probabilidad de obtener uno de estos resultados, la cual se asume estar relacionada con la combinación lineal de los predictores escogidos. Para modelar esta relación se utiliza la función de enlace logit. Si los resultados posibles de un suceso Y son 0 y 1, y se define la clase 1 como éxito, la regresión logística queda definida como:

$$\text{logit}(\mathbb{P}(Y = 1)) = \text{logit}(p_b) = \ln\left(\frac{p_b}{1 - p_b}\right) = X\beta + \varepsilon$$

Se utiliza el método MLE para estimar los parámetros de este modelo. En este caso, la interpretación de los coeficientes no es directa, y depende del nivel de referencia que se utilice.

4.3.2. Comparación de modelos

Para realizar la comparación de los modelos anteriormente descritos, se utilizan distintas métricas de evaluación, cada una de las cuales representa una medida del desempeño del modelo, en diferentes aspectos buscados al realizar las estimaciones.

Mean absolute error y porcentaje error (MAE y MAPE)

MAE, mean absolute error, y MAPE, mean absolute percentage error, son dos medidas de exactitud de la estimación. Específicamente, para este trabajo se calculan

estas métricas dentro de la muestra para poder conocer el poder descriptivo de los modelos, y se calculan fuera de la muestra para conocer su poder predictivo.

$$MAE = \frac{1}{n} \sum_i |y_i - \hat{y}_i|$$

$$MAPE = \frac{1}{n} \sum_i \frac{|y_i - \hat{y}_i|}{y_i}$$

y_i = observaciones de la variable dependiente
 \hat{y}_i = estimaciones de la variable dependiente
 n = número de observaciones totales

Coeficiente de determinación (R^2)

R^2 es un estadístico generado en las regresiones de MCO y permite cuantificar la varianza explicada por el modelo estimado. R^2 es comparable en distintas regresiones estimadas a través de este método a pesar de que estén realizadas con modelos y datos diferentes.

Debido a la forma de este estadístico, este aumenta si se le agregan más variables a la estimación, independiente de que estas no sean significativas. Para solucionar este problema se utiliza el R^2 ajustado para las comparaciones.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

$$R^2_{adj} = 1 - \frac{N-1}{N-k-1} [1 - R^2]$$

y_i = observaciones de la variable dependiente
 \hat{y}_i = estimaciones de la variable dependiente
 \bar{y} = promedio de la variable dependiente
 N = tamaño de la muestra
 k = número de variables estimadas

Sin embargo, como varios de los modelos utilizados no se estiman a través de MCO, esta métrica no es pertinente. En cambio, se debe utilizar un pseudo R^2 para poder hacer las evaluaciones. La desventaja de esta métrica es que no es comparable cuando el, o los modelos, se aplican sobre distintas bases de datos. Existen diferentes acercamientos para calcular este nuevo estadístico, específicamente se utilizará McFadden's pseudo R^2 . Este acercamiento mantiene el objetivo esperado de R^2 de que sea una estimación de la varianza explicada por el modelo.

$$R^2_{McF} = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Null})}$$

$\hat{L}(M_{Full})$ = verosimilitud del modelo estimado
 $\hat{L}(M_{Null})$ = verosimilitud del modelo nulo (modelo solo con intercepto)

Significancia estadística

La significancia estadística de una variable permite saber si es improbable que ésta haya ocurrido como un resultado del azar. Específicamente, como métrica de comparación se testea la significancia estadística de los parámetros obtenidos de las regresiones (β_i) y se evalúa que tan probable es que el valor real de estos sea igual a cero. Para esto se utiliza un nivel de confianza del α y un test t .

Hipótesis:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Estadístico t:

$$t = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \sim t_{\alpha/2, n-2}$$

El resultado de este test es entregado de manera automática por los softwares con los que se trabaja y se reportan utilizando la siguiente notación: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Esto es, si un coeficiente se presenta con tres asteriscos (***), este es significativo con una probabilidad menor al 0.001.

Signo y rango de las variables

Como una forma de evaluar si lo entregado por las regresiones tienen sentido con la realidad, se evalúa que el signo que tengan los parámetros estimados sean los esperados. En otras palabras, si se tiene una variable independiente que se espera afecte negativamente a la variable dependiente, el signo del parámetro correspondiente a esa variable independiente debiese ser negativo y se evalúa esa condición como la correcta.

Matriz de Confusión

La matriz de confusión es una herramienta ampliamente utilizada para la visualización del desempeño de algoritmos de aprendizaje supervisado, dado que permite detectar si el modelo está clasificando o prediciendo correctamente las clases estudiadas. Suponiendo que se tienen 2 clases, Positivo y Negativo, la matriz de confusión se construye de la siguiente manera:

		Valor predicho	
		Positivo	Negativo
Valor real	Positivo	Verdadero Positivo (VP)	Falso Negativa (FN)
	Negativo	Falso Positivo (FP)	Verdadero Negativo (VN)

Tabla 1: Ejemplo Matriz de confusión
Fuente: Elaboración propia

Tanto Verdadero Positivo (VP) como Verdadero Negativo (VN), son las observaciones correctamente predichas. El objetivo de un modelamiento es reducir las magnitudes de Falso Positivo (FP) y Falso Negativo (FN) según la los objetivos y las implicancias reales de cada error.

Accuracy, Recall y Precision

Para interpretar el desempeño de un modelo como el planteado en el punto anterior, suelen utilizarse estas medidas de precisión, las que se encuentran enfocadas en representar diferentes medidas de error del modelo. Utilizando la misma notación antes precisada, se tiene:

Métrica	Definición	Fórmula
Accuracy	Tasa de acierto.	$\frac{VP + VN}{VP + VN + FN + FP}$
Recall	También llamada Sensibilidad, refleja el porcentaje de observaciones reales correctamente predichas dentro de una clase.	$\frac{VP}{VP + FN}$
Precision	Son el porcentaje de predicciones correctamente asignadas dentro de una clase	$\frac{VP}{VP + FP}$

Tabla 2: Definición de Accuracy, Recall y Precision
Fuente: Elaboración propia

Tanto Recall como Precision, pueden ser calculadas de manera homóloga para la clase Negativo.

4.3.3. Valorización de los clientes

En el OE3, se espera poder evaluar la calidad de los nuevos clientes atraídos a través de las promociones con medio preferente de pago. Esta calidad está definida por el valor que el cliente significa para la empresa desde su incorporación como cliente hasta el término de su relación. La herramienta utilizada como base para estimar el valor del cliente a partir de los datos disponibles es el Customer Lifetime Value.

Customer Lifetime Value (CLV)

El CLV de un cliente representa los beneficios que este genera a lo largo de su relación con la empresa, permitiendo diferenciar a los consumidores que son más rentables en el tiempo. Tradicionalmente se considera la tasa de retención de los clientes, sin embargo, teniendo en consideración la disponibilidad de datos, en esta investigación se utiliza una fórmula simplificada de esta métrica. Este CLV simplificado tiene una formulación muy similar a la del valor presente neto (VPN) utilizado en el área de las finanzas, y corresponde a la suma del valor presente de todas las ganancias futuras que provocará un cliente a la empresa.

$$CLV = \sum_{t=1}^T \frac{G_t}{(1+r)^t}$$

G_t = Margen de las compras del cliente en el tiempo t
 r = Tasa de interés
 T = Periodo de término de la relación con el cliente

4.3.4. Cálculo de la promoción óptima

Para encontrar la promoción óptima según el objetivo perseguido, y poder diseñar las recomendaciones establecidas en el OE4, se utilizará la herramienta de simulación de escenarios.

Simulación de escenarios (What if analysis)

Esta herramienta, utiliza modelos escogidos previamente para estimar los resultados esperados de una variable de interés, en diferentes configuraciones de la realidad. En este caso, los escenarios están compuestos por las diferentes características que definen una promoción: nivel de descuento, sublíneas aplicadas, precio final, etc. Con la formulación de los diferentes escenarios, se estima el resultado que las promociones tendrían en la realidad para la empresa. Así, se puede obtener el volumen de ventas esperado, el número de tarjetas abiertas, el valor de los clientes atraídos, etc., y se puede escoger el escenario que maximice la variable deseada.

5. DESARROLLO METODOLÓGICO

A continuación, se muestra el trabajo práctico llevado a cabo para cumplir los objetivos propuestos. Esta sección se organiza separando el desarrollo de los tres primeros objetivos específicos en tres subsecciones diferentes, incluyendo en cada uno de ellos el desarrollo del cuarto objetivo específico correspondiente a la búsqueda de las recomendaciones óptimas para cada caso.

5.1. Efecto de las promociones en las unidades vendidas

El primer objetivo específico, abordado en esta subsección, tiene como propósito evidenciar y cuantificar la relación entre descuentos con tarjeta propia del negocio y las unidades vendidas de un cierto tipo de producto. Para esto, se compararán diferentes especificaciones y formas funcionales con el fin de modelar de forma fidedigna la realidad observada, y, así, poder hacer estimaciones y encontrar los óptimos correspondientes a la maximización de las utilidades obtenidas por la empresa debido a las promociones aplicadas.

5.1.1. Preparación de la base de datos

La base utilizada en este objetivo se obtiene del procesamiento de la tabla transaccional (a nivel de SKU vendido), junto con las tablas de información de los productos. Dicho procesamiento se detalla a continuación.

Categorización y filtro de las ventas de interés

Las transacciones son categorizadas en 6 grupos de interés según el artículo comprado. Estos grupos de interés fueron definidos previamente por la empresa de retail de la cual se obtienen los datos en base al comportamiento de ventas de las diferentes líneas. Estas categorías son: Telefonía, Colchones, Electrodomésticos, Televisión y Video, Línea blanca y Computación. Se conservan sólo las transacciones pertenecientes a estas categorías y, además, se eliminan todas las observaciones que no correspondan a lo deseado, como, por ejemplo, líneas de productos no regulares, devoluciones y ventas por internet.

Agregación por día y sublínea

La información disponible se agrega inicialmente a nivel de modelo y medio de pago (con tarjeta propia u otro medio), calculando promedios simples de los precios y la suma de las ventas diarias de cada modelo a partir de los datos transaccionales, realizando las imputaciones necesarias en los precios.

Luego de haber agregado a nivel diario por modelo, se procede a agregar la información a nivel diario por sublínea. Al realizar esta agregación, los precios promedios se obtienen a partir de la ponderación de los precios diarios de cada modelo según su importancia histórica dentro de la sublínea. Para poder obtener esta importancia se utiliza el porcentaje que las unidades totales vendidas históricamente que cada modelo representan de las unidades totales vendidas en la sublínea.

Imputación de precios

Existen 3 precios diarios por sublínea que son importantes para la realización de la investigación, estos son: precio lista (P_{ts}), precio con descuento pagando con tarjeta

(P_{ts}^{DT}) y precio con descuento pagando con otro medio de pago (P_{ts}^{DO}). Debido a que la base de datos con la que se trabaja es una base del tipo transaccional, si un día no existen ventas de un modelo, no se cuenta con su información de precios para el día en cuestión. Para poder representar en el precio de la sublínea a todos sus modelos, se deben imputar los precios faltantes antes de agregar a nivel de sublínea.

Se sabe que dentro del proceso de fijación de precios del retail, el día de modificaciones es el día lunes. A partir de esto, se imputan los precios bajo la regla de que, si algún modelo no tiene alguno de los tres precios diarios necesitados, se le imputa el precio del primer día anterior que tuviese datos, excepto cuando fuese lunes, en cuyo caso se imputa el precio del primer día siguiente con datos disponibles.

Filtro temporal

Para seleccionar el periodo de tiempo a trabajar, se toman en cuenta tres objetivos: no utilizar periodos muy extensos de datos donde el comportamiento de pago de los clientes pueda cambiar radicalmente, tener datos suficientes para calibrar los modelos y priorizar los periodos más recientes. A partir de esto, se decide utilizar la información de los años 2014 y 2015 como datos de calibración, y los datos del primer semestre del 2016 como bloque de testeo.

Elección de categoría y sublíneas

Para escoger las sublíneas a estudiar, se analiza el número de imputaciones de precios realizadas durante el periodo 2014 y 2015, priorizando que existiesen datos suficientes para la calibración de los modelos en las sublíneas a utilizar. Además, se prioriza seleccionar sublíneas de una misma categoría.

La categoría con menos imputaciones es Colchones, y, dentro de esta, se tomaron todas las sublíneas que tienen menos imputaciones que el promedio general de imputaciones de la categoría. A partir de esto quedan 16 sublíneas a estudiar:

Categoría Colchones

Almohadas	Paquetes Completos Boxspring	Colchones 2 Pza.
Cómoda	Protectores De Colchón	Colchones 1 Pza.
Cama	Paquetes Completos Cama Americana	Closet/Roperos
Boxspring	Paquetes Textiles Box Spring	Colchones 1 1/2 Pza.
Veladores	Colchones Medida Especial	Cama Americana
Respaldo De Cama		

Tabla 3: Sublíneas a estudiar
Fuente: Elaboración propia

Detección de outliers e imputación

El criterio seleccionado para la detección de outliers es el método de la desviación estándar (Barnett & Lewis, 1974) con tres sigmas. Para ello se calcula el promedio de unidades vendidas por sublínea (\bar{x}_s) y su desviación estándar (σ_s). Todo valor que se encuentra fuera del intervalo $[\bar{x}_s - 3 * \sigma_s ; \bar{x}_s + 3 * \sigma_s]$ es marcado como outlier y, posteriormente, se le imputa un nuevo valor a través de una interpolación lineal.

El análisis de outliers se realiza distinguiendo las unidades vendidas según medio de pago, teniendo dos variables por sublínea: unidades diarias vendidas con tarjeta (Q_{ts}^T),

de las cuales el 1.6% son outliers, y unidades diarias vendidas con otro medio de pago (Q_{ts}^O), de las cuales el 1.81% son outliers. Como era de esperarse, la imputación generó cambios en la distribución de las cantidades, sin embargo, el cambio porcentual de sus promedios es considerablemente bajo. En la sección 8.1.1 se muestra el porcentaje de outliers de cada sublínea, además de las variaciones debido a la imputación de datos.

Posteriormente, se analizaron los modelos utilizados en el OE1 con la base de datos imputada y con la base de datos sin las observaciones outliers, encontrándose que estos tenían un desempeño ligeramente menor con la imputación de outliers.

Creación o transformación de las variables de interés

Para el modelamiento, es necesario crear y/o transformar algunas de las variables de la base de datos que son utilizadas en los desarrollos que se presentan más adelante.

- *Existe Descuento ($EDesc_{ts}$)*: Refleja la amplitud del descuento en el día t dentro de la sublínea s . Esta se puede interpretar como el porcentaje de modelos en descuento para una sublínea en un día determinado. Para su creación se siguen los siguientes pasos:
 1. A nivel de agregación por día (t) y modelo (m) se marca si existe descuento con tarjeta ese día o no, dependiendo de la cantidad vendida con descuento (Q_{tm}^{DT}):
 - i. Si $Q_{tm}^{DT} > 0 \Rightarrow EDesc_{tm} = 1$
 - ii. Si $Q_{tm}^{DT} = 0 \wedge Q_{tm}^{DO} > 0 \Rightarrow EDesc_{tm} = 0$.
 - iii. Si $Q_{tm}^{DT} = 0 \wedge Q_{tm}^{DO} = 0 \Rightarrow EDesc_{tm}$ se imputa en base a la misma regla utilizada para la imputación de precios.
 2. Al pasar a una agregación de día (t) y sublínea (s), se calcula el promedio ponderado de la variable $EDesc_{tm}$ en base a la importancia histórica de cada modelo en la sublínea. Finalmente se obtiene la variable $EDesc_{ts}$ a nivel diario por sublínea.
- *Descuento con Tarjeta (DT_{ts})*: La variable DT_{ts} es el descuento porcentual obtenido por pagar con tarjeta con respecto al precio lista de los productos.
- *Porcentaje de Unidades Vendidas con Tarjeta (QTQ_{ts})*: Esta variable se obtiene de la división de las unidades vendidas con tarjeta (Q_{ts}^T) sobre el total de unidades vendidas ($Q_{ts}^T + Q_{ts}^O + 1$). Se suma un 1 al denominador de la fracción para evitar la pérdida de observaciones debido a que no existan ventas.

5.1.2. Ventas por sublínea

Información general

Utilización	Año	# días	# observaciones
Calibración	2014	359	5.744
	2015	357	5.712
Testeo	2016	164	2.624
Total		880	14.080

Tabla 4: Distribución anual de los días estudiados
Fuente: Elaboración propia

La base de datos final de la categoría Colchones con la que se desarrolla este objetivo cuenta con un total de 14.080 observaciones, las que se dividen en calibración y testeo según la distribución especificada en la Tabla 4. Por su parte, las variables disponibles a nivel diario por sublínea se presentan en la Tabla 5. Además, se cuenta con todas las variables que se pueden construir a partir de ellas.

	Variable	Significado
<i>Precios</i>	P_{ts}	Precio lista promedio ponderado
	P_{ts}^{DT}	Precio promedio ponderado con descuento pagando con tarjeta
	P_{ts}^{DO}	Precio prom. pond. con descuento pagando con otro medio de pago
<i>Cantidades</i>	Q_{ts}^{DT}	Unidades vendidas con descuento pagadas con tarjeta
	Q_{ts}^{NT}	Unidades vendidas sin descuento pagadas con tarjeta
	Q_{ts}^{DO}	Unidades vendidas con descuento pagando con otro medio de pago
	Q_{ts}^{NO}	Unidades vendidas sin descuento pagando con otro medio de pago
<i>Variables de descuento</i>	DT_{ts}	Profundidad del descuento con tarjeta
	$EDesc_{ts}$	Amplitud del descuento con tarjeta

Tabla 5: Variables de la base de datos a trabajar
Fuente: Elaboración propia

Análisis según tipo de pago

A nivel general se tiene que las unidades más vendidas son las correspondientes a las ventas con descuento pagando con tarjeta, y que el medio de pago más popular es la tarjeta del retail, significando un 75% de las unidades totales vendidas y un 78% de los ingresos totales (Ilustración 3). Por su parte, las unidades vendidas con descuento pagando con otro medio de pago son el tipo de venta menos común. La distribución de las unidades vendidas y los ingresos de ventas son casi iguales.

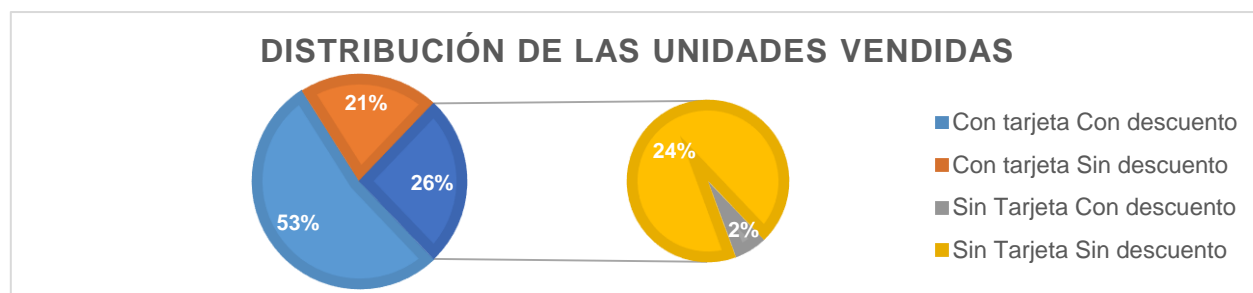


Ilustración 3: Gráfico de la distribución de unidades vendidas
Fuente: Elaboración propia

Precio	Medio de pago	Promedio	
<i>Con descuento</i>	Tarjeta	$\overline{P_{ts}^{DT}}$	\$ 139.054
	Otro medio de pago	$\overline{P_{ts}^{DO}}$	\$ 93.257
<i>Completo (lista)</i>	-	$\overline{P_{ts}}$	\$ 173.210

Tabla 6: Precios unitario promedio
Fuente: Elaboración propia

Al analizar los precios según el medio de pago (Tabla 6), se identifica una situación inusual y anti intuitiva con respecto a los precios con descuento pagando con otro medio de pago. El promedio de los precios con descuento pagando con otro medio de pago es inferior al promedio de los precios con descuento pagando con tarjeta. Esto se debe al

bajó número de ventas producidas de este tipo (2% del total de unidades), lo que implica que la imputación se realiza entre muy pocos artículos que no tienen relación real entre sí, asignando precios bajos, debido a que son precios de descuento, a SKUs que no corresponden.

Análisis según sublínea

A nivel de sublíneas, la con mayor número de unidades vendidas es Almohadas en todos los tipos de ventas (con y sin tarjeta, con y sin descuento). Existe una alta variación en las unidades totales vendidas entre las sublíneas (Ilustración 4), teniendo que 3 de estas representan más del 50% de las unidades. En la sección 8.1.2 de Anexos se pueden encontrar los estadísticos descriptivos de las unidades vendidas por sublínea.

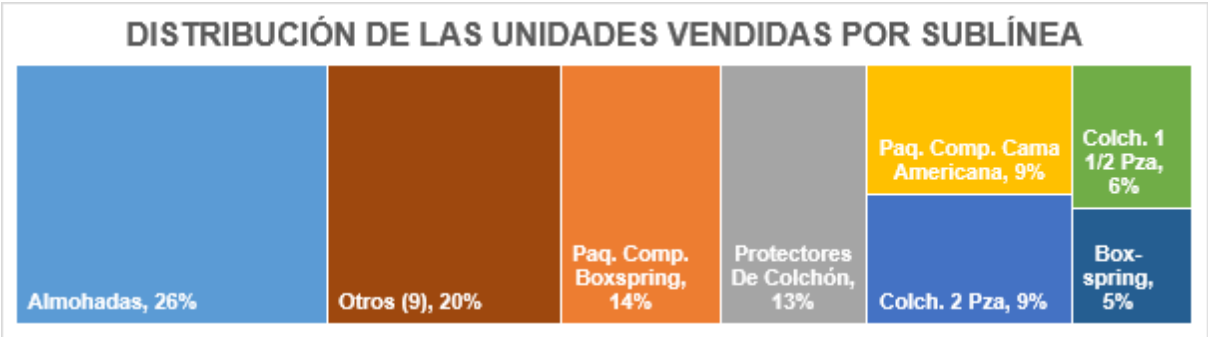


Ilustración 4: Distribución de las unidades vendidas por sublínea
Fuente: Elaboración propia

Con respecto a los promedios de unidades vendidas diariamente se da una situación similar a la descrita para las unidades vendidas totales. En general, la mayoría de las sublíneas cuenta con un coeficiente de variación bajo en términos de las unidades totales vendidas diariamente (menor o igual a 1,1), situación que se repite al analizar los diferentes tipos de ventas. Solo en las ventas con descuento sin tarjeta los coeficientes de variación de todas las líneas se salen de un rango normal (todas sobre los 2,0 puntos). Esto puede deberse al bajo número de ventas que hay de este tipo (solo el 2% del total).



Ilustración 5: Ingresos diarios promedio por sublínea
Fuente: Elaboración propia

En términos de ingresos, la sublínea más importante es Paquetes Completos Boxspring, representando el 37% de las ventas, situación que se repite si se analizan las ventas con los diferentes medios de pago y con/sin descuento. La sublínea menos importantes en términos de ingresos totales son Respaldo De Cama, lo que se repite en los ingresos por ventas con tarjetas.

En la sección 8.1.2 se puede encontrar el detalle del promedio de los precios por sublínea. Allí se puede observar que Paquetes Completos Boxspring presenta los precios promedio mayores, lo cual explica por qué esta sublínea representa los mayores ingresos históricos sin tener las mayores ventas en término de unidades. En términos generales, los precios promedio parecen no tener altas variaciones dentro de cada sublínea.

Análisis temporal

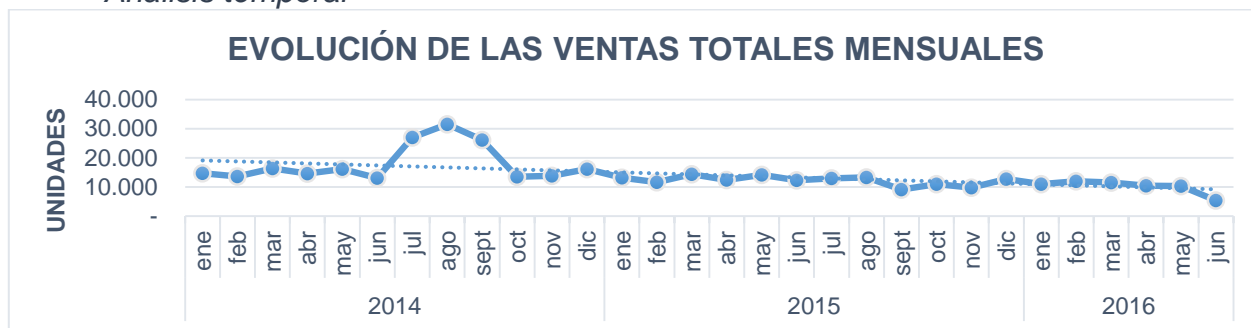


Ilustración 6: Unidades totales vendidas a históricamente
Fuente: Elaboración propia

En términos temporales, como se puede ver en la Ilustración 6, las unidades totales vendidas mensualmente han ido variando a través del tiempo, presentando una tendencia lineal a la baja. Lo mismo sucede con los ingresos promedio por año (8.1.2. Anexos).

Acorde a lo que indica la intuición, existe un alza en las ventas producidas durante el fin de semana. Por un lado, se tiene que el promedio de las ventas por sublínea un día de la semana es 26,7 unid., mientras que este número sube a 41,2 y 35,6 unid. los días sábado y domingo respectivamente. Esta diferencia es estadísticamente significativa, por lo que el efecto del día de la semana debe ser incluido en los modelos futuros.

Análisis de las variables de interés

	Promedio	Mínimo	Máximo	Desviación Estándar	Coef. de Variación
Q_{ts}^T	22	0	256	27	1,24
QTQ_{ts}	0,67	0,00	0,99	0,20	0,30
$Ln(EDesc_{ts})$	-0,72	-4,17	-0,02	0,48	-0,66
DT_{ts}	20%	0%	61%	7%	0,36
$Ln(PDT)_{ts}$	11,59	9,22	12,71	0,86	0,07

Tabla 7: Estadísticos descriptivos básicos de las variables de interés
Fuente: Elaboración propia

En la Tabla 7, se muestran los estadísticos básicos para las diferentes variables de interés que serán utilizadas en el modelamiento. La distribución de estas variables y sus estadísticos por sublínea pueden encontrarse en la sección 8.1.2. Es interesante

destacar que todas las variables presentan bajas dispersiones tanto a nivel general como dentro de las sublíneas, con coeficientes de variación menores a 1.5 puntos.

Las unidades vendidas con tarjeta (Q_{ts}^T) muestran una distribución concentrada en los valores más bajos (cerca de cero). Su evolución es muy similar a la de la variable unidades vendidas totales mostrada anteriormente, teniendo su promedio más alto el año 2014, y presentando cierta tendencia a la baja en el tiempo.

Por su parte, el porcentaje de las unidades vendidas con tarjeta (QTQ_{ts}) se caracteriza por estar mayormente concentrada en los valores altos del intervalo $[0,1]$. Las sublíneas con mayor porcentaje de ventas con tarjeta promedio son las sublíneas de Paquetes Completos Cama Americana y Paquetes Completos Boxspring. Es interesante notar que además estas dos sublíneas son las que tienen menor coeficiente de variación de esta variable. La evolución temporal de QTQ_{ts} se observa bastante estable durante el periodo estudiado, por lo cual se esperaría que las variables temporales no tengan un impacto significativo en los modelos que se estiman para este porcentaje.

De las variables de promoción, la amplitud del descuento dentro de las sublíneas ($LnEDesc_{ts}$) pertenece al intervalo $[0,1]$ y presenta una distribución concentrada en los valores altos del intervalo. De las sublíneas, Veladores es donde existe una mayor presencia de descuentos en sus modelos. Por otro lado, la profundidad del descuento pagando con tarjeta (DT_{ts}) presenta una distribución concentrada en los valores cercanos a cero. El DT_{ts} promedio por sublínea no excede el 33%, teniendo su máximo para la sublínea de Cama Americana y su mínimo en Closet/Roperos y Colchones 1 ½ Pza.

Por último, la variable del precio con descuento con tarjeta ($Ln(PDT)_{ts}$) se presenta transformada con logaritmo natural para lograr un mayor ajuste a los datos. Los promedios de ella por sublínea son en general parecidos, destacando las sublíneas Almohadas y Protectores de Colchón, los que tienen un $Ln(PDT)_{ts}$ bastante menor que el de los demás. Esto se debe a que, como se vio anteriormente, estas sublíneas tienen precios lista más bajos.

En conclusión, se puede observar variación entre las sublíneas en términos de volúmenes de ventas y promedios de las variables de interés, lo que demuestra lo importante que es realizar el modelamiento a nivel de sublíneas y no a nivel de categoría.

5.1.3. Modelamiento

Alternativas de modelamiento

Para estudiar el impacto que las promociones de descuento con medio preferente de pago tienen en las unidades vendidas con tarjeta de crédito, se decide comparar los siguientes modelos en búsqueda del más adecuado en términos de ajuste a los datos:

- 2 tipos de regresiones según la variable dependiente a modelar:
 - Tipo A, que modelan la variable dependiente Q_{ts}^T , unidades vendidas con tarjeta. El objetivo de estas regresiones es modelar el efecto directo sobre las unidades vendidas.

- Tipo B, que modelan la variable dependiente QTQ_{ts} , porcentaje de las unidades vendidas con tarjeta. El objetivo de estas regresiones es modelar el efecto de los descuentos con medio preferente de pago en la distribución del tipo de unidades vendidas.
- 3 modelos según la combinación de variables independientes incluidas:
 - Modelo 1: Incluye solo la variable $Ln(PDT)_{ts}$ más las variables de control.
 - Modelo 2: Incluye las variables $Ln(PDT)_{ts}$ y DT_{ts} más las variables de control.
 - Modelo 3: Incluye las variables $Ln(PDT)_{ts}$, DT_{ts} y $Ln(EDesc)_{ts}$ más las variables de control.
- 4 formas funcionales: Regresión de Poisson, Lineal, Binomial y Logística.
- 16 sublíneas: se estimó cada una de las regresiones con los datos de cada sublínea escogida.

Las variables de control escogidas fueron día (categórica) como el día de la semana de la observación; mes (categórica) como el mes de la observación; y año (continua) tomando como año 0 el 2014, 1 el 2015 y 2 el año 2016.

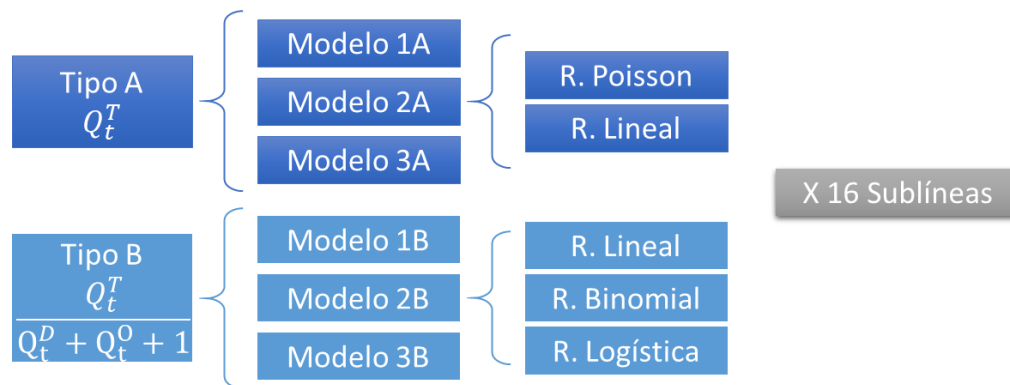


Ilustración 7: Diagrama de las regresiones realizadas
Fuente: Elaboración propia

En total se estiman 240 (15x16) regresiones distintas utilizando el software computacional Stata. Adicionalmente, se calculan las medidas de comparación antes descritas: MAE y MAPE tanto dentro como fuera de la muestra, R^2 ajustado o pseudo R^2 según correspondiese, la significancia de los betas obtenidos, y el porcentaje de parámetros significativos con signo acorde a lo esperado.

Comparación de modelos

Ante la infactibilidad técnica de exponer y analizar los resultados de las 240 regresiones realizadas, sus métricas de evaluación son agregadas a nivel de modelo. Más adelante, en base a estas agregaciones, se escoge el mejor modelo y forma funcional para continuar con la investigación.

Para realizar la agregación por modelo de R^2 , MAE y MAPE, se calculan los promedios ponderados de estas métricas otorgándole pesos a las sublíneas a partir del número de unidades vendidas históricamente de cada una¹. Los resultados de los R^2 y MAPE de los Modelos Tipo A (variable dependiente Q_{ts}^T) y tipo B (variable dependiente

¹ R^2 *: Los valores expuestos representan a R^2 ajustado o pseudo R^2 según corresponda

QTQ_{ts}) se muestran en la Tabla 8 y Tabla 9 respectivamente. Los resultados completos se pueden encontrar en la sección 8.1.3.

	Muestra	Modelo 1A		Modelos 2A		Modelos 3A	
		R ^{2*}	MAPE	R ^{2*}	MAPE	R ^{2*}	MAPE
Regresión Lineal	Dentro	40,8%	41,5%	41,8%	41,1%	47,4%	38,5%
	Fuera	-	47,4%	-	57,0%	-	54,7%
Regresión Poisson	Dentro	27,5%	40,4%	28,2%	40,0%	32,1%	36,8%
	Fuera	-	45,4%	-	50,6%	-	43,0%

Tabla 8: Métricas de comparación ponderadas de los modelos tipo A – Variable dependiente Q_t^T (1)
Fuente: Elaboración propia con uso de software stata

Para el caso de la variable dependiente Q_t^T (Tabla 8), las regresiones Lineal y Poisson que mejor explican su varianza y que tienen un mayor poder descriptivo (MAPE dentro de la muestra) son las del modelo 3. Además, entre estas dos formas funcionales, la que describe con mayor precisión la base de datos es la regresión de Poisson.

Por otro lado, el modelo 1 y 3 son quienes predicen con menor error porcentual los datos fuera de la muestra en el caso de las regresiones Lineales y de Poisson respectivamente. Entre estos dos modelos, 1A Lineal y 3A Poisson, quien predice con mayor precisión la variable Q_t^T fuera de la muestra, es el 3A de Poisson.

	Muestra	Modelo 1B		Modelo 2B		Modelo 3B	
		R ^{2*}	MAPE	R ^{2*}	MAPE	R ^{2*}	MAPE
Regresión Lineal	Dentro	9,0%	14,3%	11,0%	14,1%	20,5%	13,2%
	Fuera	-	22,3%	-	21,8%	-	19,1%
Regresión Binomial	Dentro	5,6%	14,5%	6,5%	14,3%	10,8%	13,4%
	Fuera	-	22,4%	-	22,2%	-	19,4%
Regresión Logística	Dentro	0,8%	14,3%	0,9%	14,1%	1,5%	13,2%
	Fuera	-	22,1%	-	21,9%	-	19,2%

Tabla 9: Métricas de comparación ponderadas de los modelos tipo B - Variable dependiente QTQ_{ts} (1)
Fuente: Elaboración propia con uso de software stata

Para las regresiones tipo B, variable dependiente QTQ_{ts} (Tabla 9), en todas sus formas funcionales el modelo 3 es quien explica un mayor porcentaje de la varianza de los datos, y quien, además, presenta un menor error porcentual en la estimación tanto dentro como fuera de la muestra. Al comparar entre modelos, se tiene que el modelo 3B Lineal es quien presenta una capacidad descriptiva y predictiva ligeramente mayor.

	Variable	Modelo 1A		Modelo 2A		Modelo 3A	
		Significancia	Signo Esperado	Significancia	Signo Esperado	Significancia	Signo Esperado
Regresión Lineal	$Ln(PDT)_{ts}$	66%	96%	74%	43%	63%	50%
	DT_{ts}	-	-	62%	77%	69%	66%
	$Ln(EDesc)_{ts}$	-	-	-	-	97%	100%
Regresión Poisson	$Ln(PDT)_{ts}$	92%	97%	76%	42%	86%	51%
	DT_{ts}	-	-	77%	64%	75%	63%
	$Ln(EDesc)_{ts}$	-	-	-	-	97%	100%

Tabla 10: Métricas de comparación ponderadas de los modelos tipo A - Variable dependiente Q_t^T (2)
Fuente: Elaboración propia con uso de software stata

Para comparar la significancia de las variables de los modelos, se calcula el porcentaje de las sublíneas en las que cada variable resulta significativa a un 95% (En las tablas como Significancia), ponderándolas al igual que para la agregación anterior. De la misma manera se trabaja para obtener el porcentaje de las sublíneas en las que cada variable cuenta con el signo esperado (En las tablas como Signo Esperado), tomando en consideración solo las sublíneas en las que la variable resulta significativa. Estos resultados, por tipo de regresión, se muestran en la Tabla 10 y Tabla 11. Además, en la sección 8.1.3 de Anexos se puede encontrar esta información complementada con mayores descriptivos de los β_s .

Es importante notar que la variable $Ln(PDT)_{ts}$, a pesar de tener un nivel de significancia alto en la mayoría de las regresiones, presenta un bajo porcentaje de signo esperado para los modelos 2 y 3, tanto en las regresiones tipo A (Tabla 10) como en las tipo B (Tabla 11), esto en comparación a los porcentajes alcanzados por las otras variables de interés. Por su parte, para la variable DT_{ts} , que está presente en los modelos 2 y 3, en las regresiones tipo A (Tabla 10) sobresale el modelo 2 de Poisson en términos de significancia, y el modelo 2 Lineal en términos de signo esperado. Para las regresiones tipo B (Tabla 11), el modelo 2 en todas sus formas funcionales destaca por la significancia de sus coeficientes, sin embargo, son los modelos 3 quienes presentan un mayor porcentaje de signo esperado.

	Variable	Modelo 1B		Modelo 2B		Modelo 3B	
		Significancia	Signo Esperado	Significancia	Signo Esperado	Significancia	Signo Esperado
Regresión Lineal	$Ln(PDT)_{ts}$	70%	90%	81%	32%	80%	32%
	DT_{ts}	-	-	80%	78%	76%	82%
	$Ln(EDesc)_{ts}$	-	-	-	-	97%	100%
Regresión Binomial	$Ln(PDT)_{ts}$	72%	93%	80%	32%	80%	32%
	DT_{ts}	-	-	80%	78%	76%	83%
	$Ln(EDesc)_{ts}$	-	-	-	-	97%	100%
Regresión Logística	$Ln(PDT)_{ts}$	70%	90%	81%	32%	80%	32%
	DT_{ts}	-	-	80%	78%	76%	82%
	$Ln(EDesc)_{ts}$	-	-	-	-	97%	100%

Tabla 11: Métricas de comparación ponderadas de los modelos tipo B - Variable dependiente QTQ_{ts} (2)
Fuente: Elaboración propia con uso de software stata

Finalmente, para el caso de la variable $Ln(EDesc)_{ts}$, la cual sólo está presente en el modelo 3, los resultados muestran que su desempeño en términos de significancia y signo obtenido es igual en todas sus funcionales para ambos tipos de regresiones. Es importante destacar que esta variable es la que presenta un mayor porcentaje de significancia y signo correcto en comparación a las otras variables de interés evaluadas.

Elección del modelo final

A partir del análisis recién expuesto y de la información adicional dispuesta en el punto 8.1.3 de Anexos, se puede concluir que el modelo con mejor desempeño en términos de R^2 , MAE y MAPE es el modelo 3 tanto para las regresiones tipo A y tipo B. Para el caso de la significancia, el modelo 2 parece destacar con respecto a los otros modelos, aunque esta superioridad no es generalizada. Por otro lado, ningún modelo destaca por sobre los demás en términos de signos de los parámetros.

Paralelamente, del análisis de R^2 , MAE y MAPE se distingue a la forma funcional de las regresiones de Poisson como la con mejor desempeño en la estimación de Q_{ts}^T (regresiones tipo A), y a las regresiones Lineales como las mejores estimando QTQ_{ts} (regresiones tipo B).

De la misma manera, para las regresiones tipo A, la mejor forma funcional en términos de significancia de las variables de interés nuevamente es la regresión de Poisson. El criterio de signo correcto no revela ninguna prioridad de forma funcional para este tipo de regresiones. Para las regresiones tipo B, ninguna de las formas funcionales destaca por sobre las demás en términos de significancia y signo.

En base todo a esto, se decide que el modelo a utilizar para la evaluación de las promociones en términos de unidades vendidas, y la posterior evaluación de escenarios, será el modelo 3A en su forma funcional de Poisson. Para estudiar el porcentaje que las unidades vendidas con tarjeta representan del total, se decide utilizar la forma funcional Lineal. Dado que ninguno de los modelos destaca plenamente en ninguno de los criterios para las regresiones tipo B, se decide utilizar el modelo 3 para contar con las mismas variables con las que se cuenta para el tipo A.

5.1.4. Resultados y análisis

Resultados de los modelos seleccionados

Las especificaciones de las regresiones escogidas son:

1. Modelo 3A de Poisson: Modelamiento de las unidades vendidas con tarjeta.

$$Q_{ts}^T \sim \text{Poisson}(\lambda_{ts}),$$

$$\ln(\lambda_{ts}) = \beta_0 + \beta_{\ln(PDT),s} \ln(PDT_{ts}) + \beta_{DT,s} DT_{ts} + \beta_{\ln(EDesc),s} \ln(EDesc_{ts}) + \sum_{d \in \{DOW\}} \beta_{d,s} d_{ts} + \sum_{m \in \{MOY\}} \beta_{m,s} m_{ts} + \beta_{a,s} \text{año}_{ts} + \varepsilon_{ts}$$

2. Modelo 3B Lineal: Modelamiento del porcentaje de unidades vendidas con tarjeta.

$$QTQ_{ts} = \beta_0 + \beta_{\ln(PDT),s} \ln(PDT_{ts}) + \beta_{DT,s} \ln(DT_{ts}) + \beta_{\ln(EDesc),s} \ln(EDesc_{ts}) + \sum_{d \in \{DOW\}} \beta_{d,s} d_{ts} + \sum_{m \in \{MOY\}} \beta_{m,s} m_{ts} + \beta_{a,s} \text{año}_{ts} + \varepsilon_{ts}$$

En los modelos anteriores, se utilizan DOW y MOY como los conjuntos de variables d y m indicadoras del día de la semana y mes del año, respectivamente. En ellas, se considera al día domingo y al mes de enero como el día y mes base, respectivamente.

Al resolver estas regresiones con los datos de cada una de las sublíneas se obtuvieron los parámetros reportados en la Tabla 12 y la Tabla 13. En éstas, se presentan los β_s de las variables de interés junto con la constante de cada regresión. Los resultados de las variables temporales se pueden encontrar en la sección 8.1.4.

1. Modelo 3A Poisson

A partir de los resultados (Tabla 12) se puede identificar una correlación positiva y significativa entre $\ln(EDesc_{ts})$ y las unidades vendidas con tarjeta. Esto concuerda con la intuición de que, a más amplitud del descuento con tarjeta, más incentivos a la utilización de ésta como medio de pago. Destaca el efecto que existe en la sublínea Veladores, donde una variación de $\ln(EDesc_{ts})$ se relaciona con una variación en las unidades vendidas más drástica que para las otras sublíneas.

Sublínea	$\widehat{\beta}_{cons,s}$	$\widehat{\beta}_{Ln(PDT),s}$	$\widehat{\beta}_{DT,s}$	$\widehat{\beta}_{Ln(EDesc),s}$
Almohadas	-8,18***	1,26***	3,40***	0,39***
Boxspring	130,51***	-10,36***	-4,21***	0,72***
Cama	-16,88	1,59	2,22	0,46***
Cama Americana	-22,72	2,19	2,96	0,34***
Closet/Roperos	-50,28***	4,49***	3,28***	0,10
Colchones 1 1/2 Pza.	8,72	-0,49	3,75***	0,28***
Colchones 1 Pza.	37,84***	-3,12***	-0,97	0,50***
Colchones 2 Pza.	-12,59	1,37*	3,63***	0,56***
Colchones Medida Especial	74,44	-5,85	-7,45*	0,62***
Cómoda	-17,52***	1,55***	3,95***	0,56***
Paq. Completos Boxspring	192,56***	-14,73***	-10,61***	0,42***
Paq. Completos Cama Americana	113,23***	-9,07***	-3,31***	0,44***
Paq. Textiles Box Spring	19,07	-1,37	1,54	0,65***
Protectores De Colchón	11,45***	-0,77***	-0,16	0,26***
Respaldo De Cama	110,09***	-9,13***	-13,49***	0,23**
Veladores	-20,49**	1,89**	3,03***	1,47***

Tabla 12: $\widehat{\beta}$ s de las variables de interés para Q_{ts}^T
Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Fuente: Elaboración propia con uso de software stata

En el caso de $\ln(PDT)_{ts}$ se esperaría que su parámetro tuviese signo negativo, debido a que esta variable refleja el precio pagado por el artículo en venta. Esto se da en solo 9 de las 16 sublíneas. Lo mismo ocurre con la variable DT_{ts} , la cual se esperaría tener coeficientes positivos, lo que ocurre en 9 sublíneas. Es interesante notar que para las sublíneas donde estas dos variables son significativas, ambos coeficientes tienen el mismo signo. Esto puede estar reflejando respuestas a las promociones propias de cada sublínea según el tipo de productos que incluyen.

Las variables de día para este modelo son todas significativas, excepto sábado, que resulta no significativa para cuatro sublíneas. Es este mismo día el que presenta un impacto mayor en las ventas para todas las sublíneas, seguido en segundo lugar por el día domingo. Esto concuerda con lo observado en el análisis exploratorio. Los días que tienen un menor impacto en ventas son los martes, miércoles y jueves. Las variables indicadoras de meses son en su mayoría significativas. No se observa una tendencia sostenida con respecto a qué meses presentan mayor o menor impacto en las ventas para todas las sublíneas. Sin embargo, destacan los últimos tres meses del año, donde se presenta el menor impacto para 12 de las 16 sublíneas, lo que podría indicar que en esos meses del año la gente suele estar enfocado en otro tipo de categorías por las fiestas de fin de año. La variable año es en su mayoría negativa, lo que indica una baja en las ventas a lo largo del tiempo.

2. Modelo 3B Lineal

Para el caso del porcentaje de unidades vendidas con tarjeta del total de unidades vendidas, se tienen resultados similares a los obtenidos en la regresión anterior (Tabla 13). $\ln(EDesc)_{ts}$ presenta todos sus parámetros coincidentes con el signo esperado. Esta variable solo es no significativa para la sublínea Closet/Roperos, lo que coincide con la regresión 3A. De igual manera, la variable $\ln(PDT)_{ts}$ nuevamente no tiene el signo esperado para varias de las sublíneas estudiadas. Para esta regresión, se tiene que un mayor número de sublíneas presentan el coeficiente con el signo esperado para DT_{ts} en comparación a la regresión anterior.

Sublínea	$\widehat{\beta}_{cons,s}$	$\widehat{\beta}_{\ln(PDT),s}$	$\widehat{\beta}_{DT,s}$	$\widehat{\beta}_{\ln(EDesc),s}$
Almohadas	-3,39*	0,40*	0,83***	0,09***
Boxspring	-0,74	0,12	1,40*	0,24***
Cama	-14,47*	1,24*	1,53*	0,13***
Cama Americana	3,95	-0,26	-0,23	0,09***
Closet/Roperos	-5,58***	0,55***	0,09	0,01
Colchones 1 1/2 Pza.	2,24	-0,13	0,87**	0,09***
Colchones 1 Pza.	9,43**	-0,78*	-0,16	0,15***
Colchones 2 Pza.	-8,16*	0,75*	1,33***	0,14***
Colchones Medida Especial	11,92	-0,90	-1,16	0,24***
Cómoda	-9,91***	0,86***	1,00***	0,09***
Paq. Completos Boxspring	-16,12*	1,32**	1,70**	0,16***
Paq. Completos Cama Americana	17,40***	-1,38***	0,08	0,13***
Paq. Textiles Box Spring	-8,88	0,77	1,19	0,23***
Protectores De Colchón	5,56***	-0,51***	-0,74***	0,06***
Respaldo De Cama	12,46	-0,97	-2,54*	0,09**
Veladores	-3,87	0,39	0	0,25***

Tabla 13: β s significativos de las variables de interés por sublínea para QTQ

Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Fuente: Elaboración propia con uso de software stata

Los niveles de significancia de las variables temporales para este modelo son, en general, menores en comparación al modelo anterior. Se mantiene la tendencia donde Sábado y Domingo impactan en mayor proporción el porcentaje de ventas y donde los últimos meses del año tienen el impacto más bajo. Esta vez el impacto del año es más bien variado dentro de las sublíneas, lo que podría indicar que, aunque las unidades vendidas con tarjeta han bajado, esto puede verse influenciado por una baja general de las unidades vendidas.

Métricas de evaluación

En la Tabla 14 se presentan los resultados del pseudo R^2 y R^2 ajustado, MAE y MAPE para las regresiones de Poisson y Lineales para cada una de las sublíneas. Para la regresión sobre las unidades vendidas con tarjeta, los errores porcentuales de las estimaciones dentro de la muestra van desde el 25% al 70%, con un promedio simple de 46% entre las sublíneas. Por su parte, los errores porcentuales de las estimaciones fuera de la muestra van desde el 32% al 100%, con un promedio simple del 55%. Es interesante destacar que algunas de las regresiones parecieran tener un poder predictivo mayor que un poder descriptivo, presentando MAPE fuera de la muestra menor al de dentro de la muestra.

Sublínea	Muestra	Modelo 3A Poisson			Modelo 3B Lineal		
		R ² _{pse}	MAE	MAPE	R ² _{ajd}	MAE	MAPE
Almohadas	Dentro	42%	26,4	37%	22%	0,084	14%
	Fuera		25,1	34%		0,102	16%
Boxspring	Dentro	30%	4,9	39%	28%	0,090	14%
	Fuera		4,8	52%		0,147	21%
Cama	Dentro	12%	2,8	65%	14%	0,137	25%
	Fuera		1,9	59%		0,168	26%
Cama Americana	Dentro	24%	4,9	50%	21%	0,102	17%
	Fuera		3,0	100%		0,154	40%
Closet/Roperos	Dentro	24%	3,6	46%	17%	0,089	13%
	Fuera		2,1	84%		0,153	31%
Colchones 1 1/2 Pza.	Dentro	26%	6,0	34%	22%	0,079	12%
	Fuera		4,1	32%		0,115	20%
Colchones 1 Pza.	Dentro	21%	4,5	59%	17%	0,120	23%
	Fuera		3,1	58%		0,162	40%
Colchones 2 Pza.	Dentro	30%	8,5	29%	35%	0,060	8%
	Fuera		11,7	37%		0,077	11%
Colchones Medida Especial	Dentro	12%	1,5	54%	13%	0,177	25%
	Fuera		1,4	46%		0,200	33%
CÓmoda	Dentro	19%	3,3	49%	23%	0,102	16%
	Fuera		5,4	47%		0,146	19%
Paquetes Completos Boxspring	Dentro	40%	11,4	25%	22%	0,056	7%
	Fuera		15,3	44%		0,097	13%
Paquetes Completos Cama Americana	Dentro	33%	9,0	31%	14%	0,065	9%
	Fuera		7,0	38%		0,115	17%
Paquetes Textiles Box Spring	Dentro	22%	2,2	61%	17%	0,157	26%
	Fuera		1,3	58%		0,255	28%
Protectores De Colchón	Dentro	28%	8,8	30%	12%	0,074	15%
	Fuera		7,1	37%		0,101	22%
Respaldo De Cama	Dentro	6%	1,5	52%	4%	0,188	29%
	Fuera		2,0	93%		0,206	36%
Veladores	Dentro	18%	4,0	70%	17%	0,184	22%
	Fuera		6,0	58%		0,206	27%

Tabla 14: Métricas de evaluación de las regresiones por sublínea
Fuente: Elaboración propia con uso de software stata

En el caso de las regresiones sobre el porcentaje de unidades vendidas con tarjetas, se tiene que la varianza que logra explicar el modelo en las diferentes sublíneas va desde el 4% al 35%, destacando la sublínea de Colchones 2 Pza., la cual es la única que supera el 30% de R² ajustado.

Por su parte, el MAPE tanto fuera como dentro de la muestra en este caso son bastante menores que para el tipo de regresión anterior. Sin embargo, esto se debe a que las unidades de QTQ_{ts} es porcentaje, por lo cual el error porcentual de la estimación no es comparable con el error porcentual de la estimación de Q_t^T . El MAPE dentro de la muestra para esta regresión tiene un promedio de 17% entre las sublíneas, mientras que para la evaluación fuera de la muestra el promedio del error porcentual sube a 25%. Destaca el caso de Paquetes Completos Boxspring, la cual es la sublínea con menor MAPE dentro de la muestra tanto para esta regresión como para la regresión de Poisson.

También se destaca el desempeño de las dos regresiones en el caso de la sublínea Colchones 2 Pza., donde se tiene que los MAPE fuera y dentro de la muestra son inferiores a la mayoría de las otras sublíneas.

Impacto de las variables de interés

Para poder evaluar el impacto que cada una de las variables de interés tienen sobre las métricas estimadas, es necesario recordar la interpretación que tienen los coeficientes obtenidos de cada uno de los modelos. En el caso de la regresión de Poisson, los coeficientes ($\widehat{\beta}_k$) se interpretan de tal manera que un cambio en una unidad de la variable independiente ($\Delta x_k = 1$) significa un cambio porcentual en la variable dependiente igual al coeficiente estimado ($\Delta y_k = \widehat{\beta}_k * 100\%$).

Como las variables de precio con descuento con tarjeta (*PDT*) y la amplitud del descuento dentro de la sublínea (*EDesc*) se encuentran transformadas por un logaritmo, sus coeficientes se interpretan de tal manera que un cambio porcentual en una de ellas tiene un impacto porcentual en las unidades vendidas. Para el caso del descuento con tarjeta (*DT*), esta no se encuentra transformada en el modelo, pero al ser un porcentaje, se interpreta de manera similar que el caso anterior, donde una variación de un 1 punto porcentual (1pp) del descuento implica una variación porcentual en las ventas.

Con el objetivo de ilustrar lo anterior, se toma la sublínea Cama como ejemplo. El coeficiente estimado para el precio con descuento con tarjeta en esta sublínea es $\widehat{\beta}_{Ln(PDT),CAMA} = 1,59$, esto significa que un cambio de un 1% en el precio implica un cambio de 1,59% en las unidades vendidas con tarjeta, aproximadamente. Para el descuento con tarjeta, $\widehat{\beta}_{DT,CAMA} = 2,2$, se tiene que un aumento en un 1pp de éste implica un aumento en un 2,2% en las unidades vendidas. Finalmente, el coeficiente estimado de la amplitud del descuento para Cama es $\widehat{\beta}_{Ln(EDesc),CAMA} = 0,46$, indicando que el impacto en las unidades vendidas con tarjeta de un cambio de un 1% en la amplitud del descuento dentro de la sublínea es de un 0,46%, aproximadamente.

Para el caso de la regresión Lineal, la relación entre las variables independiente y la dependiente es directa, esto es, un cambio en una unidad de una variable independiente ($\Delta x_k = 1$) implica un cambio en unidades de la variable dependiente igual al coeficiente correspondiente ($\Delta y_k = \widehat{\beta}_k$). Como el Modelo 3B tiene como variable a predecir al porcentaje de las unidades vendidas, un cambio en unidades de este es realmente un cambio en puntos porcentuales. Además, las variables predictivas se encuentran incluidas en el modelo de la misma manera que para la regresión anterior, por lo tanto, la interpretación de los impactos es casi equivalente a la descrita en la regresión anterior.

A modo de ejemplo, para la sublínea Cama, con coeficientes $\widehat{\beta}_{Ln(PDT),CAMA}$, $\widehat{\beta}_{DT,CAMA}$ y $\widehat{\beta}_{Ln(EDesc),CAMA}$ iguales a 1,24, 1,53 y 0,13 respectivamente, se tiene que: un cambio de un 1% en *PDT* implica un cambio de 1.2pp en el porcentaje de unidades vendidas con tarjeta; una variación de un 1pp en *DT* se ve relacionado con una variación de un 1,5pp en el porcentaje de las unidades vendidas con tarjeta; y que un cambio de un 1% en *EDesc* tiene un impacto de un 0,1pp en QTQ_{ts} .

Entonces, el impacto de la variación de cada una de las variables de interés, en el caso de las unidades vendidas con tarjeta, es el coeficiente estimado en porcentaje, y, en el caso del porcentaje de unidades vendidas con tarjeta, es el coeficiente estimado en puntos porcentuales. Recordar que estos coeficientes se muestran en las Tabla 12 y Tabla 13. La lectura de los impactos se realiza de la misma manera en que se ilustró para la sublínea Cama. Es importante recordar que la variable amplitud del descuento ($EDesc$) es un decimal que va desde 0 a 1, por lo que sus variaciones porcentuales representan un número muy pequeño, y es por ello que su impacto resulta relativamente bajo en comparación al impacto de las otras variables.

Destaca el alto impacto que tienen el precio y descuento con tarjeta en las unidades vendidas en las sublíneas Paquetes Completos Boxspring, Boxspring y Respaldo Cama. También, es interesante notar que el impacto de la amplitud del descuento en el porcentaje de las unidades vendidas con tarjeta es relativamente estable en las diferentes sublíneas. Cabe mencionar que el análisis del impacto del descuento y el precio con descuento con tarjeta, por separado, no es estrictamente correcto, debido a la relación existente entre estas dos variables. Un desarrollo más apropiado se presenta en la siguiente sección de análisis de escenarios.

5.1.5. Configuración óptima de las promociones (OE4)

La finalidad de la simulación de escenarios del OE1 es estimar las unidades que se hubiesen vendido con tarjeta de crédito, a partir de diferentes especificaciones de las promociones y, así, poder encontrar las configuraciones óptimas para las promociones por sublíneas a partir del modelo estimado. Para realizar esta simulación se utiliza el modelo de Poisson 3A, con variable dependiente unidades vendidas con tarjeta (Q_{ts}^T).

De las variables de interés, se decide variar la profundidad de las promociones (DT_{ts}), lo que, al mismo tiempo, varía el precio con descuento con tarjeta (PDT_{ts}). Una segunda opción es variar la amplitud de las promociones ($EDesc_{ts}$), dentro de cada sublínea, sin embargo, se descarta esta alternativa debido a las dificultades que se presentan para estimar el descuento general de cada sublínea debido a un cambio en la amplitud del descuento dentro esta.

Para poder realizar la evaluación completa del modelo, es necesario situarse en un día para la asignación de las variables de control. Debido a que este día es arbitrario dentro de la base de datos, se intenta que este sea lo más representativo posible del horizonte temporal. Para esto, se busca dentro de los datos de calibración los días donde las unidades vendidas promedio fuera lo más cercano posible al promedio general histórico diario, y, dentro de ellos, el día donde existiese menor variación entre las sublíneas. En base a esto se selecciona el día $t^* = 01/02/2015$ y las estimaciones a realizar se obtienen a partir de la siguiente ecuación:

$$\widehat{Q}_{t^*s}^T = \exp \left[\hat{\beta}_0 + \hat{\beta}_{Ln(PDT),s} \ln(PDT_s^e) + \hat{\beta}_{DT,s} \ln(DT_s^e) + \hat{\beta}_{Ln(EDesc),s} \ln(EDesc_{t^*s}) \right. \\ \left. + \sum_{d \in \{DOW\}} \hat{\beta}_{d,s} d_{t^*s} + \sum_{m \in \{MOY\}} \hat{\beta}_{m,s} m_{t^*s} + \hat{\beta}_{a,s} año_{t^*s} \right]$$

Aquí, PDT_s^e y DT_s^e corresponden al precio con descuento con tarjeta y el descuento con tarjeta que se estén evaluando. Con esta estimación, para poder conocer los beneficios obtenidos por la empresa, se calculan las utilidades esperadas de las ventas con tarjeta:

$$\widehat{U}_{t^*}^T = \widehat{Q}_{t^*}^T * \%Q_{t^*}^{DT} (PDT - C) + \widehat{Q}_{t^*}^T * (1 - \%Q_{t^*}^{DT}) (P - C)$$

$\%Q_{t^*}^{DT}$ = % de las unidades vendidas con tarjeta que tenían descuento para el día t^*

Para este cálculo de utilidades, se asume que la proporción dentro de las utilidades vendidas con tarjeta entre las con descuento y sin descuento, no cambia para el día evaluado a pesar del cambio en la profundidad del descuento. Este supuesto se basa en que el modelo ya ha sido capaz de captar la sustitución de ventas entre los distintos modelos de los productos a partir de las promociones, y que al no cambiar la amplitud de las promociones estos porcentajes deberían mantenerse. Para trabajos futuros se deja propuesto analizar estas dos formas de ventas por separado para poder revisar si este supuesto es válido o no.

Modelo Base		Sublínea	
Sublínea	\overline{DT}	Sublínea	\overline{DT}
Almohadas	26%	Colchones Medida Especial	20%
Boxspring	15%	Cómoda	20%
Cama	23%	Paq. Comp. Boxspring	19%
Cama Americana	32%	Paq. Comp. Cama Amer.	15%
Closet/Roperos	14%	Paq. Textiles Box Spring	16%
Colchones 1 1/2 Pza.	14%	Protectores De Colchón	23%
Colchones 1 Pza.	15%	Respaldo De Cama	17%
Colchones 2 Pza.	17%	Veladores	28%

Tabla 15: Promedio histórico del descuento con tarjeta por sublínea
Fuente: Elaboración propia

En primer lugar, se calcula un modelo base a partir del cual posteriormente se realizan las comparaciones con el escenario óptimo. Esta base pretende emular el comportamiento promedio de las promociones por sublínea, por lo cual se utiliza el promedio histórico de DT por sublínea (\overline{DT}_s) en el horizonte de calibración. El precio con descuento con tarjeta por sublínea se calcula a partir de este descuento promedio.

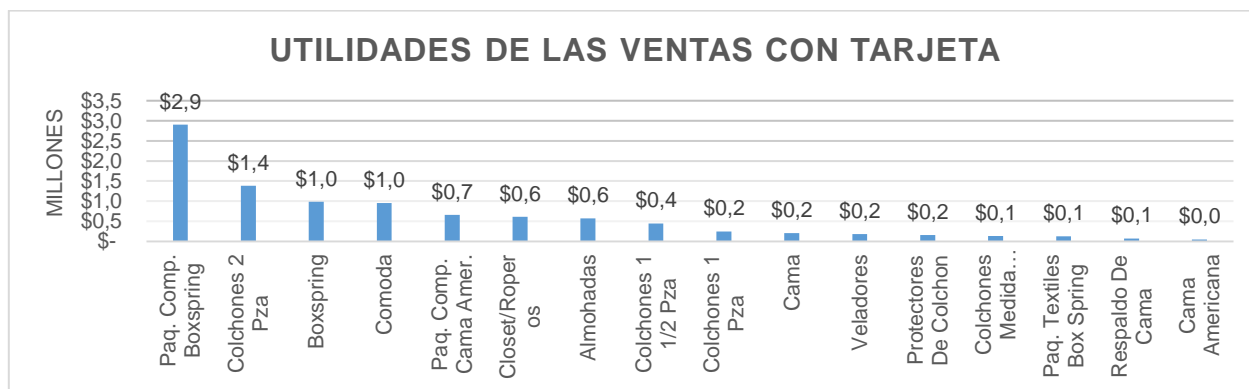


Ilustración 8: Utilidades de las ventas con tarjeta para el modelo base
Fuente: Elaboración propia

Para este día y con la configuración anteriormente especificada, se tiene que todas las sublíneas tienen utilidades positivas y bastante variables entre sí (Ilustración 8). La sublínea que entrega mayores utilidades a partir de su descuento promedio histórico es Paquetes Completos Boxspring, mientras que la que entrega menores utilidades es Cama Americana. Esto puede deberse a que Cama Americana es la sublínea que presenta un promedio histórico de profundidad del descuento mayor, lo que puede llevar a que la disminución de margen de estas ventas no logre ser sobrellevado por el aumento en unidades vendidas.

Mejor profundidad del descuento

Posterior al cálculo del modelo base, se procede a realizar las simulaciones de diferentes profundidades de descuento para cada sublínea, con el objetivo de encontrar la mejor configuración para el día 01/02/2015 en base a las utilidades obtenidas. Para esto, se utilizan las mismas formulas mostradas anteriormente, y se utilizan niveles de descuento con tarjeta múltiplos de 5%. En la Tabla 16 se muestran los resultados de la simulación por nivel de descuento en términos de las utilidades obtenidas en unidades de millones de pesos (MM), Por temas de presentación sólo se muestran los descuentos múltiplos de 10% y las utilidades menores a -\$100MM se reflejan con un signo gato.

Sublíneas	Nivel de Descuento con tarjeta										DT _s Óptimo
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	
Almohadas	0,50	0,55	0,57	0,57	0,54	0,46	0,34	0,19	0,04	-0,05	25%
Boxspring	0,62	0,86	1,12	0,91	-3,68	-44,70	#	#	#	#	25%
Cama	0,58	0,43	0,26	0,07	-0,11	-0,28	-0,40	-0,44	-0,37	-0,19	0%
Cama americana	1,05	0,78	0,45	0,10	-0,24	-0,53	-0,69	-0,68	-0,48	-0,17	0%
Closet/roperos	1,12	0,76	0,45	0,21	0,05	-0,02	-0,04	-0,03	-0,01	0,00	0%
Colch. 1 1/2 pza	0,43	0,46	0,39	0,11	-0,61	-2,21	-5,59	-12,73	-28,59	-70,75	10%
Colch. 1 pza	0,31	0,27	0,21	0,07	-0,20	-0,81	-2,37	-7,27	-29,86	#	0%
Colch. 2 pza	1,84	1,64	1,21	0,51	-0,52	-1,81	-3,23	-4,40	-4,68	-3,19	0%
Colch. medida especial	0,40	0,25	0,13	0,03	-0,09	-0,29	-0,81	-2,85	-18,40	#	0%
Cómoda	1,00	1,02	0,95	0,78	0,49	0,06	-0,44	-0,89	-1,08	-0,74	5%
Paq. Comp. Boxspring	2,75	2,97	2,81	-0,46	-26,41	#	#	#	#	#	15%
Paq. Comp. Cama Amer.	0,54	0,64	0,58	-0,44	-6,66	-45,20	#	#	#	#	15%
Paq. textiles Boxspring	0,20	0,17	0,09	-0,06	-0,36	-0,95	-2,17	-4,88	-12,22	-43,76	0%
Protectores de colchón	0,19	0,18	0,16	0,15	0,13	0,10	0,08	0,04	-0,03	-0,17	0%
Respaldo de cama	0,27	0,13	0,05	0,01	-0,03	-0,10	-0,31	-1,51	-19,94	#	0%
Veladores	0,28	0,25	0,22	0,17	0,11	0,05	-0,01	-0,05	-0,05	-0,03	0%

Tabla 16: Simulación de escenarios para 01/02/2015. Utilidades en MM.

Fuente: Elaboración propia

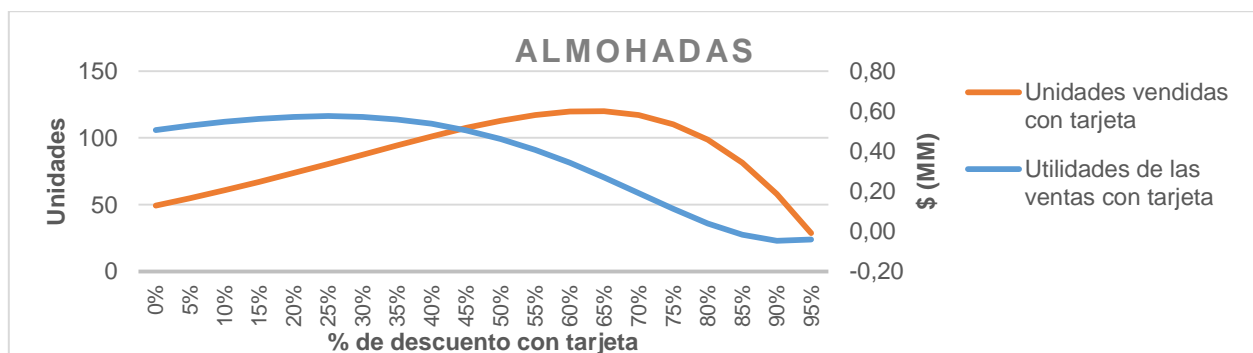


Ilustración 9: Relación unidades vendidas vs utilidades para sublínea Almohadas

Fuente: Elaboración propia

A modo de ejemplo, se presentan algunos gráficos de la relación entre las unidades estimadas totales vendidas con tarjeta, y las utilidades generadas por estos, según el nivel de descuento evaluado.

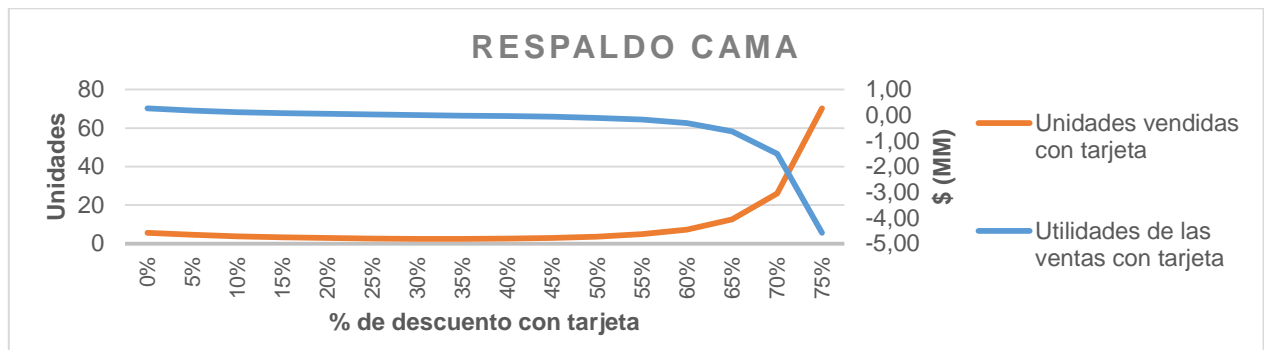


Ilustración 10: Relación unidades vendidas vs utilidades para sublínea Respaldo Cama
Fuente: Elaboración propia

Como se puede apreciar, todos los descuentos óptimos son bajos, muchos de estos llegando a ser nulos. Además, es interesante notar que las unidades estimadas no son estrictamente crecientes con el porcentaje de descuento como era de esperarse. Esto se debe a que la función utilizada en el modelo (exponencial) a pesar de ser monótona, es evaluada en dos términos que varían en dirección opuesta, el descuento con tarjeta (que va aumentando) y el precio con descuento con tarjeta (que va disminuyendo). Este juego entre las variables resulta en gráficos no estrictamente crecientes con el descuento.

Descuento óptimo versus descuento promedio

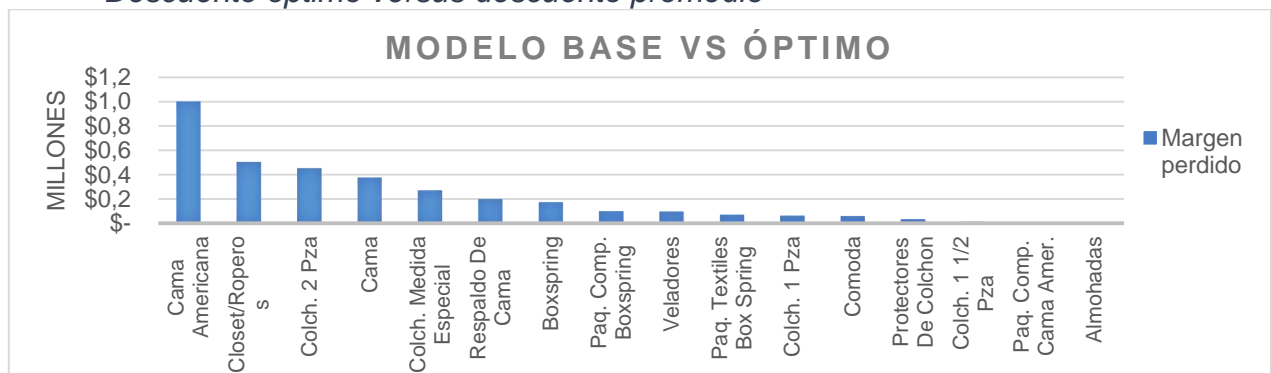


Ilustración 11: Margen perdido estimado por sublínea
Fuente: Elaboración propia

Al realizar la comparación entre el descuento óptimo por sublínea, encontrado anteriormente, con las utilidades obtenidas aplicando el descuento promedio (modelo base), se puede apreciar que, en general, existe una brecha entre lo observado en la realidad con lo ideal (Tabla 17). Esta brecha significa una pérdida de margen por sublínea, la que puede llegar a ser extrema según los modelos estimados. Por ejemplo, para el caso de la sublínea Cama Americana, el aumento en las utilidades utilizando el descuento óptimo (0%) se estima en más de un 2.000%. El alto nivel de descuento en esta sublínea y otras puede deberse a subvenciones externas a las promociones. También se tiene el caso donde el descuento aplicado en la realidad de manera promedio

pareciera ser muy similar al descuento más eficiente. Este es el caso de las Almohadas y los Paquetes Completos Cama Americana.

Sublínea	DT óptimo	\overline{DT}	Δ Utilidades óptimas y base
Almohadas	25%	26%	0%
Boxspring	25%	15%	18%
Cama	0%	23%	182%
Cama Americana	0%	32%	2.194%
Closet/Roperos	0%	14%	83%
Colchones 1 1/2 Pza.	10%	14%	3%
Colchones 1 Pza.	0%	15%	25%
Colchones 2 Pza.	0%	17%	33%
Colch. Medida Especial	0%	20%	207%
Cómoda	5%	20%	6%
Paq. Comp. Boxspring	15%	19%	3%
Paq. Comp. Cama Amer.	15%	15%	0%
Paq. Textiles Box Spring	0%	16%	54%
Protectores De Colchón	0%	23%	22%
Respaldo De Cama	0%	17%	288%
Veladores	0%	28%	55%

Tabla 17: Diferencias entre el modelo base con el escenario óptimo
Fuente: Elaboración propia

Evaluación en otros días

Al realizar la simulación de escenarios fue posible observar que los resultados son sensibles al día evaluado, esto se debe al impacto que tienen las variables de control temporal en las unidades estimadas y cuanto estas pueden suprimir el efecto de la variación de las promociones a partir de los parámetros estimados. Por esto, se buscan dos fechas que sean representativas de los extremos de la base de datos para poder dimensionar las variaciones que pueden existir en el descuento óptimo según el día que se desee evaluar dentro de cada sublínea.

Para escoger las fechas se buscan los días donde las unidades vendidas sean relativamente bajas según las variables de control (VB) y donde las ventas sean relativamente altas (VA). Para esto se seleccionan los coeficientes más bajos y más altos respectivamente para las variables de día, meses y año. En cada uno de estos días se calculan los descuentos óptimos a partir de un procedimiento análogo al anterior. Los resultados de los óptimos se resumen en la Tabla 18.

Se puede ver que efectivamente el descuento óptimo es sensible al día a evaluar, pudiendo llegar a ser muy distinto en algunos casos. El aporte de las variables temporales al modelo es alto, y puede llegar a inhibir el efecto de una promoción, haciendo la demanda más o menos elástica dependiendo del caso. Sin embargo, no es posible encontrar una regla general para el comportamiento del descuento óptimo en días de ventas relativamente altas o bajas. Esto puede deberse a que en muchas de las sublíneas la pendiente de las unidades es bastante horizontal, provocando que pequeños cambios en ella varíen drásticamente el descuento ideal, como es el ejemplo de Respaldo de Cama (Ilustración 10).

Sublíneas	Descuento Óptimo			Utilidades Óptimas (\$ MM)		
	t*	VB	VA	t*	VB	VA
Almohadas	25%	25%	5%	0,58	0,30	1,14
Boxspring	25%	15%	65%	1,16	0,49	1,23
Cama	0%	0%	0%	0,58	0,21	0,51
Cama Americana	0%	0%	0%	1,05	0,33	0,99
Closet/Roperos	0%	0%	0%	1,12	0,72	1,23
Colchones 1 1/2 Pza.	10%	15%	10%	0,46	0,19	0,73
Colchones 1 Pza.	0%	0%	95%	0,31	0,09	0,33
Colchones 2 Pza.	0%	5%	0%	1,84	0,68	1,92
Colch. Medida Especial	0%	0%	0%	0,40	0,10	0,31
Cómoda	5%	50%	15%	1,02	0,30	1,32
Paq. Comp. Boxspring	15%	15%	25%	3,00	1,53	3,68
Paq. Comp. Cama Amer.	15%	20%	15%	0,66	0,32	1,51
Paq. Textiles Boxspring	0%	0%	0%	0,20	0,07	0,41
Protectores De Colchón	0%	0%	0%	0,19	0,12	0,32
Respaldo De Cama	0%	95%	0%	0,27	218.363	0,17
Veladores	0%	0%	0%	0,28	0,20	0,89

Tabla 18: Descuentos óptimos por sublínea para las fechas extremas
Fuente: Elaboración propia

5.1.6. Conclusiones del Objetivo Específico 1

Los resultados obtenidos del trabajo fueron variados e interesantes, tanto en el área del conocimiento de las promociones y sus efectos, como en el área del negocio por sus implicancias monetarias.

De la evaluación de modelos, se evidencia que Poisson efectivamente es una buena herramienta para modelar datos de conteo, resultando ser la mejor forma funcional para el modelamiento de las unidades vendidas con tarjeta según las métricas de desempeño utilizadas. Por su parte, la regresión Lineal, a pesar de no ser acotada, tiene un buen desempeño en el modelamiento del porcentaje estudiado, esto en comparación a modelos más complejos que se limitan al intervalo [0,1]. Además, la simplicidad de interpretación, tanto del modelo de Poisson como del Lineal, entregan un beneficio extra a la hora de realizar el análisis de resultados.

En la evaluación de la inclusión de variables independientes, el modelo que entrega más información de la situación estudiada es el que cuenta con un mayor número de regresores (Modelo 3), captando diferentes factores que podrían estar influyendo en el comportamiento de los clientes.

Se pudo observar que, de acuerdo a lo esperado, la amplitud del descuento dentro de la sublínea ($EDesc$) presenta una relación positiva con las unidades vendidas con tarjeta y el porcentaje que estas representan del total. Lamentablemente, a pesar de que esta variable se muestra importante para los resultados de todas las sublíneas en términos de significancia estadística, su significancia numérica es baja debido a la coeficiente de sus coeficientes, lo que indica un bajo impacto en las métricas estudiadas.

Por otro lado, contra intuitivamente, los coeficientes del precio y descuento con tarjeta resultaron tener signo igual para varias de las sublíneas estudiadas, lo que podría

explicarse por características especiales de la demanda de cada una de las sublíneas, por el nivel de correlación de algunas variables o de efectos no observados en la base de datos.

Tanto en el análisis exploratorio como en el modelamiento posterior, las sublíneas mostraron tener características y respuestas diferentes ante las distintas configuraciones de las promociones. Los niveles de ventas, ingresos y precios difieren altamente para cada tipo de producto, y el signo del impacto de las variables de descuento y precio varía de caso a caso. Esto impide establecer una regla general con respecto a las promociones dentro de la categoría y muestra la necesidad de la evaluación diferenciada de las configuraciones óptimas de los descuentos para cada tipo de producto.

Además, los descuentos óptimos también mostraron ser altamente sensibles al día donde se evalúen, debido a su capacidad de mitigar el efecto del precio final y modificar la elasticidad de la demanda a través de sus coeficientes estimados. Por lo tanto, se hace imprescindible un monitoreo continuo en las políticas de descuento que se deben aplicar en la realidad.

Por último, los descuentos óptimos encontrados en la evaluación de escenarios son bajos, y la mayoría de ellos presentan una gran diferencia con los descuentos históricos observados. Esta situación podría estar generando una pérdida de margen potencial importante para la empresa analizada.

Si se considera solo el día evaluado en el análisis de escenarios, se puede observar una diferencia entre la utilidad del modelo base y las utilidades con el descuento óptimo de \$3,4MM para las sublíneas estudiadas de la categoría colchones. Bajo el supuesto de que esta situación fuese análoga para las otras cinco categorías de interés, ésta pérdida alcanza un valor de \$20,5MM diarios y \$1.250MM anuales aproximadamente.

A pesar de no ser éste un cálculo riguroso, permite tener una dimensión del margen potencial que ganaría al aplicar una política de descuentos eficiente en términos de unidades vendidas y utilidades obtenidas en las categorías elegidas, además de servir como cota inferior para los resultados que se podrían obtener si se ampliara el análisis a todos los productos de la empresa.

5.2. Efecto de las promociones en la apertura de tarjetas propias

El segundo objetivo específico propuesto en esta investigación es estudiar el efecto de las promociones de descuento en la apertura de tarjetas de nuevos clientes. Este objetivo se encuentra motivado por la intuición de que un cliente que comienza su relación con la empresa abriendo la tarjeta del retail, es potencialmente más valioso para ella, debido a diferencias en los niveles de gastos, además de convertirse en cliente de dos líneas de negocios distintas al mismo tiempo. En términos prácticos, se busca evidenciar el nexo entre los descuentos en las diferentes categorías de interés con la apertura de tarjeta de un cliente nuevo, lo que es equivalente a que su primera compra se haya realizado con medio preferente de pago.

5.2.1. Preparación de la base de datos

Como se verá más adelante, la variable dependiente que se quiere modelar es la variable binaria que indica si la primera compra fue pagada con tarjeta propia o no, en base a la composición de la compra y los descuentos disponibles. Para esto, se debe construir una base de datos que refleje esta información a nivel de boleta, y que permita aplicar los modelos futuros. Las bases finales se procesan como se describe a continuación, a partir de las tablas: transaccional, productos e información de los clientes.

Información de las boletas de las primeras compras

Desde la base de datos transaccional, se obtiene la primera fecha donde tienen transacciones cada uno de los compradores que aparecen en la tabla de clientes (clientes identificados). Con esta fecha, se extrae la información transaccional de las primeras compras por cliente. Además, se marcan las ventas realizadas pagando con la tarjeta propia del negocio. Se eliminan todas las transacciones que no correspondan a lo estudiado, lo que incluye líneas de productos no regulares y devoluciones.

Al igual que para el OE1, las transacciones son categorizadas en 6 grupos de interés: Telefonía, Colchones, Electrodomésticos, Televisión y Video, Línea blanca y Computación. En adelante, se hace referencia a estas categorías como categorías de interés. Se mantienen solo las observaciones de las compras de estas categorías.

Luego de generado un identificador único por boleta, se agregan las transacciones que corresponden a una misma venta, esto es, que forman parte de la misma primera compra. En este proceso se conserva información con respecto a las categorías de interés consumidas y la información de los clientes.

Creación de variables de interés

Sea c el índice de las categorías de interés, esto es, $c \in C = \{\text{Telefonía, Colchones, Electrodomésticos, Televisión y Video, Línea blanca, Computación}\}$. Además, sea a el índice de cada producto incluido en una boleta b . Para poder modelar la situación estudiada, se crean las siguientes variables a nivel de boleta b .

- *Pago Tarjeta (PT_b)*: Variable binaria que indica si el medio de pago utilizado en la boleta b fue tarjeta propia del retail.
- *Compras por categoría (IC_{cb})*: 6 variables binarias que indican si la boleta b incluye productos la categoría c .

- *Descuento por categoría (ED_{cb}):* 6 variables binarias que indican si la boleta b incluye productos con descuento de la categoría c .
- *Monto de la boleta (M_b):* Monto final pagado en la compra b , considerando los productos tanto dentro como fuera de las categorías estudiadas y los descuentos aplicados. Esta variable se encuentra dividida por \$100.000.
- *Monto del Precio Lista (MPL_b):* Suma de los precios listas de los productos de las categorías de interés incluidos en la boleta b , multiplicado por las unidades vendidas en dicha boleta. Esta variable se encuentra dividida por \$100.000.
- *Monto del Descuento (MD_b):* Suma de todos los montos descontados en los productos de las categorías de interés incluidos en la boleta b . Esta variable se encuentra dividida por \$100.000.
- *Porcentaje de descuento en productos con descuento por categoría (PD_{cb}):* 6 variables continuas que indican el descuento que tienen los productos en promoción, por categoría, incluidos en la boleta b . Para esto, se calcula la suma de los montos según precio lista ($MPLD_a$) y de los montos de descuento (MDD_a) considerando solo los productos que presentan descuento:

$$PD_{cb} = \frac{\sum_{a \in c} MDD_a}{\sum_{a \in c} MPLD_a}$$

- *Descuento histórico con tarjeta y sin tarjeta por categoría (DHT_c y DHO_c):* De la base de datos transaccional completa, se obtiene el promedio del porcentaje de descuento observado históricamente por categoría, con tarjeta (DHT_c) y con otro medio de pago (DHO_c). Estos descuentos se utilizan para estimar las diferencias de los descuentos ofrecidos en cada ocasión de compra.

En el ideal, para evaluar el efecto de los descuentos con tarjeta ofrecidos, se analizaría la decisión de cada cliente de utilizar o no tarjeta en base al descuento que podría obtener con cada medio de pago. Lamentablemente, para cada compra, solo se puede calcular de manera exacta uno de estos descuentos (con o sin tarjeta), debido a que solo se tiene el descuento aplicado para el medio de pago utilizado. Como una forma aproximada de obtener los descuentos con y sin tarjeta ofrecidos por categoría en cada ocasión de compra, se crean las siguientes métricas:

- *Descuento con tarjeta potencial:* Su objetivo es reflejar el descuento que cada cliente tenía disponible pagando con tarjeta, a partir del cual decide su medio de pago. Para esta estimación se utilizan dos métodos:
 - *Descuento con tarjeta potencial escalado ($DTPE_{cb}$):*
 - Si $PT_b = 1 \Rightarrow DTPE_{cb} = PD_{cb}$
 - Si $PT_b = 0 \Rightarrow DTPE_{cb} = PD_{cb} * \frac{DHT_c}{DHO_c}$
 - *Descuento con tarjeta potencial aumentado ($DTPA_{cb}$):*
 - Si $PT_b = 1 \Rightarrow DTPA_{cb} = PD_{cb}$
 - Si $PT_b = 0 \Rightarrow DTPA_{cb} = PD_{cb} * (DHT_c - DHO_c)$
- *Ahorro con tarjeta potencial aumentado ($ATPA_{cb}$):* Su objetivo es reflejar el ahorro que cada cliente tenía disponible si pagaba con tarjeta versus sin tarjeta, a partir

del cual decide su medio de pago. Este se obtiene del precio lista de todos los productos comprados con descuento de cada categoría:

$$ATPA_{cb} = MPL_{cb} * (DHT_c - DHO_c)/100.000$$

Creación de las bases a trabajar

Considerando que la base de datos con la que se trabaja incluye la información sobre la primera compra que hacen los clientes identificados en el horizonte temporal estudiado (2012–2016), es posible que existan clientes que hayan realizado transacciones anteriores a este periodo, y que esta no sea realmente su primera compra.

Como una manera de abordar este problema, se decide excluir una porción inicial de la base de datos, eliminando clientes que podrían haber realizados compras previas recientes fuera del intervalo observado. Esto se hace considerando que un cliente esta fugado si no realizó compras en un periodo de tiempo extenso, por lo que una nueva compra se puede considerar una primera compra, con la que reinicia su relación con la empresa. Para esto, se prueban dos largos distintos de inactividad mínima, 1 año, eliminando las primeras compras del 2012, y 2 años, eliminando los datos del 2012 y 2013. Estas serán Base 1 y Base 2 respectivamente.

En las bases creadas existe un alto porcentaje de primeras compras realizadas con tarjeta (sobre un 80%), esto implica que este tipo de compras se encuentran sobrerrepresentadas, lo que puede traer problemas al momento de estimar los modelos. Para evitar esto, se crean las SubBase 1 y SubBase 2 como submuestras de las Bases 1 y 2 respectivamente. En su construcción, se toman todas las primeras compras realizadas con otro medio de pago en cada base, y se selecciona de manera aleatoria una cantidad igual de primeras compras realizadas con tarjeta, obteniendo bases de datos balanceadas. Más adelante, se comparan estas cuatro tablas para poder evaluar cuál es mejor en términos del modelamiento y el ajuste logrado a la realidad.

5.2.2. Primeras compras de los clientes

Información general

Los datos con los que se trabaja en este objetivo corresponden a 256.009 observaciones, las cuales representan las boletas de las primeras compras de los clientes identificados ocurridas desde 2013 hasta mediados del 2016. Tal como se explicó en la sección anterior, se definieron Bases y SubBases dentro de la base de datos total, cuyos tamaños y distribución temporal se puede encontrar en la Tabla 19.

Año	Base 1	SubBase 1	Base 2	SubBase 2
2013	89.448	33.408	-	-
2014	86.731	27.409	86.731	24.124
2015	55.449	16.051	55.449	13.698
2016	24.381	5.814	24.381	4.684
Total	256.009	82.682	166.561	42.506

Tabla 19: Composición de las Bases y SubBases
Fuente: Elaboración propia

	Variable	Descripción
Variables de identificación	ID cliente	Identificador único del cliente
	ID boleta	Identificador único de la boleta b
	Fecha	Fecha de la primera compra
Variables generales	PT_b	1 si la compra de la boleta b fue con tarjeta propia, 0 si no
	M_b	Monto total de la boleta b (\$100.000)
	U_b	Unidades incluidas en la boleta b de C^*
	MPL_b	Monto total según precio lista de C^* de la boleta b (\$100.000)
	MD_b	Monto total del descuento de C^* de la boleta b (\$100.000)
Variables por categoría	PD_b	Porcentaje de descuento de la boleta b considerando C^* (%)
	IC_{cb}	1 si b incluye productos de la categoría c , 0 si no
	ED_{cb}	1 si b incluye productos en descuento de la categoría c , 0 si no
	PD_{cb}	Descuento de los productos en descuento de la categoría c en la boleta b (%)
	$DTPE_{cb}$	Descuento con tarjeta potencia escalado de la boleta b (%)
	$DTPA_{cb}$	Descuento con tarjeta potencia aumentado de la boleta b (%)
	$ATPA_{cb}$	Ahorro con tarjeta potencia aumentado de la boleta b (\$100.000)

Tabla 20: Variables disponibles

Fuente: Elaboración propia

Las variables presentadas en la Tabla 20, disponibles para todas las bases, se pueden clasificar en tres grupos según la información entregada. Las variables de identificación indican a qué cliente corresponde esa primera compra, la fecha y el medio de pago utilizado. Las variables generales entregan información de la transacción llevada a cabo, en términos de montos y unidades. Para estas, se ocupa C^* como el conjunto de los productos de las categorías de interés. Por último, las variables por categoría, entregan información sobre los productos incluidos en la boleta pertenecientes a cada una de las categorías de interés. Para mayores detalles sobre el cálculo de las variables revisar el punto 5.3.1. En lo que sigue, se realiza el análisis exploratorio a partir de la Base 1, dado que es la que contiene más información.

Distribución de las primeras compras

Las primeras compras se encuentran distribuidas temporalmente de manera dispar, al igual que el porcentaje de ellas pagadas con tarjeta. A medida que pasan los años, el número de nuevos clientes pareciera ir disminuyendo, mientras que el porcentaje de primeras compras que son pagadas con tarjeta van aumentando (Ilustración 12).

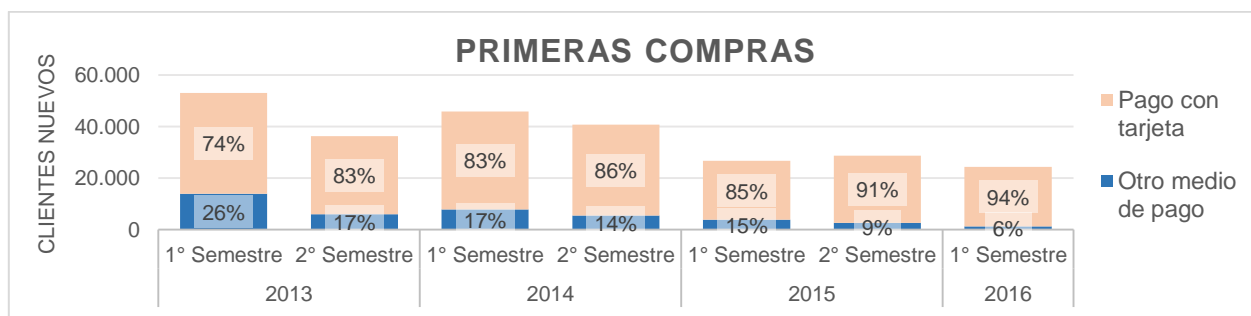


Ilustración 12: Distribución de primeras compras

Fuente: Elaboración propia

Si se desglosan ahora estas primeras compras según si incluyen algún producto de las categorías de interés en descuento, estos porcentajes son también bastante diferentes entre sí. En la Tabla 21 se puede apreciar que el 60% de las primeras compras presentan descuento en alguna de las categorías de interés, porcentaje que cambia fuertemente si se consideran solo las ventas con otro medio de pago, donde este porcentaje baja a un 5% aproximadamente.

# de boletas	Sin descuento	Con descuento	Total
Otro medio de pago	39.095 (95%)	2.261 (5%)	41.356 (100%)
Pago con tarjeta	63.377 (30%)	151.276 (70%)	214.653 (100%)
Total	102.472 (40%)	153.537 (60%)	256.009 (100%)

Tabla 21: Número de primeras compras según medio de pago y presencia de descuento
Fuente: Elaboración propia

Variables generales

En los promedios de las variables generales (Tabla 22), se pueden apreciar grandes diferencias en las boletas según el medio de pago utilizado. Un ejemplo de esto es el monto total de una boleta, cuyo promedio pagando con tarjeta es 1,7 veces mayor a una cancelada con otro medio de pago, situación muy similar a la que se presenta para el monto de la boleta según precio lista de los productos de las categorías de interés (MPL_b). La única variable que no parece tener diferencias significativas en magnitud son las unidades promedio por boleta. Otros estadísticos descriptivos de estas variables por medio de pago se pueden encontrar en la sección 8.2.1.

Variable	Otro medio de pago	Pago con Tarjeta	General
M_b	\$98.586	\$169.358	\$157.926
MPL_b	\$103.348	\$189.689	\$175.741
MD_b	\$2.085	\$22.271	\$19.010
U_b	1,21	1,29	1,28
PD_b	2%	14%	12%

Tabla 22: Promedios de las variables generales por boleta
Fuente: Elaboración propia

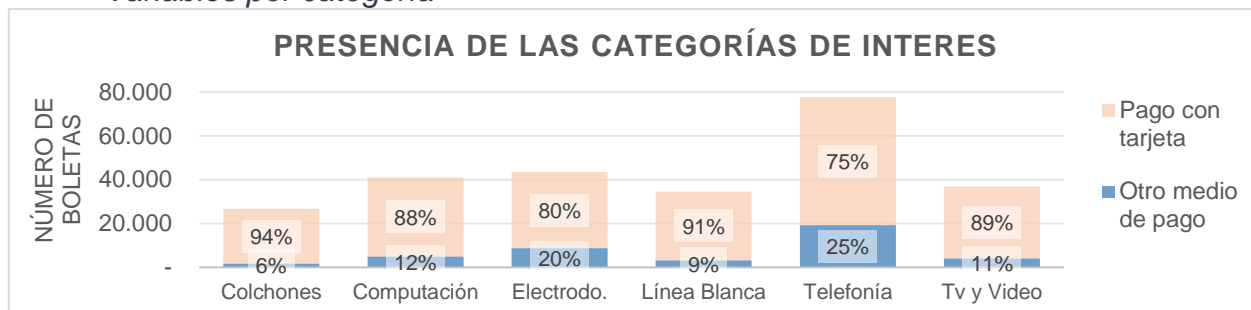
En términos temporales, los montos y unidades promedio se mantienen estables en el periodo estudiado. Sólo se observa un pequeño salto en el segundo semestre del año 2014 para las variables monto según precio lista, monto de descuento y unidades. Sin embargo, el monto total de la boleta parece no verse influenciado por estas alzas (Detalles en 8.2.1).



Ilustración 13: Descuento promedio en el tiempo
Fuente: Elaboración propia

Por otro lado, el porcentaje de descuento observado en los productos de las categorías de interés muestra un alza a través del tiempo, teniendo un promedio por boleta bajo el 10% para el primer semestre del 2013, y llegando a un nivel sobre el 15% en el segundo semestre del 2016.

Variables por categoría



*Ilustración 14: Presencia de las categorías de interés
Fuente: Elaboración propia*

En la Ilustración 14 se muestra el número de boletas que incluyen productos de cada una de las categorías de interés del total de 256.009 boletas analizadas, indicando la proporción de estas según medio de pago. A pesar de que estos porcentajes varían de categoría en categoría, se repite lo observado en la base de datos general, donde las ventas con tarjeta son considerablemente mayores al número de boletas canceladas con otro medio de pago.

La categoría de interés con mayor presencia en las boletas de las primeras compras corresponde a Telefonía. Por su parte, los productos menos recurrentes son los pertenecientes a la categoría Colchones. En contraposición, es Colchones la categoría con un mayor porcentaje de boletas canceladas con medio preferente de pago, mientras que Telefonía es la que muestra el porcentaje menor.

Continuando con el análisis, para cada categoría se toman las boletas donde tienen presencia entre los productos comprados, y se clasifican dependiendo de si alguno de los productos tiene descuento o no (Detalles en 8.2.1). De esto, se puede ver que Colchones es la categoría con un mayor porcentaje de boletas con descuento entre las boletas donde tiene presencia, seguido cercanamente por Televisión y Video. A pesar de las diferencias observadas en la Ilustración 14 en la presencia de cada categoría, el porcentaje de estas boletas donde sus productos tienen descuento no difiere altamente entre ellas, manteniéndose en un rango entre el 55% al 70% aproximadamente, siendo siempre mayor el porcentaje de boletas con descuento.

Descuentos y ahorro potencial con tarjeta

Los descuentos históricos considerando todas las ventas disponibles, por categoría, se presentan en la Tabla 23. Acorde a lo esperado, el descuento promedio con tarjeta es mayor al descuento promedio sin tarjeta para todas las categorías, más aun, es más del doble para todas ellas. El caso más extremo es el de Colchones, donde el descuento promedio con tarjeta es 6.5 veces el descuento promedio sin tarjeta.

Categoría	Descuento histórico con tarjeta (DHT_c)	Descuento histórico sin tarjeta (DHO_c)	Proporción de los descuentos (DHT_c/DHO_c)	Diferencia de los descuentos ($DHT_c - DHO_c$)
Colchones	13%	2%	6.5	11%
Computación	8%	2%	4.6	6%
Electrodomésticos	7%	2%	3.7	5%
Línea Blanca	7%	2%	5	6%
Telefonía Móvil	17%	6%	2.6	10%
Tv y Video	15%	4%	3.3	10%

Tabla 23: Descuentos históricos por categoría
Fuente: Elaboración propia

Como se explicó en la sección 5.2.1, con los descuentos históricos se crearon 3 variables por categoría para reflejar las opciones de descuento que enfrentó cada cliente al momento de decidir qué medio de pago utilizar. En la Ilustración 15, se presenta los promedios del descuento potencial con tarjeta escalado ($DTPE_c$) y aumentado ($DTPA_c$), mientras que en la Ilustración 16 se muestran los promedios del ahorro potencial de utilizar la tarjeta ($ATPA_c$).

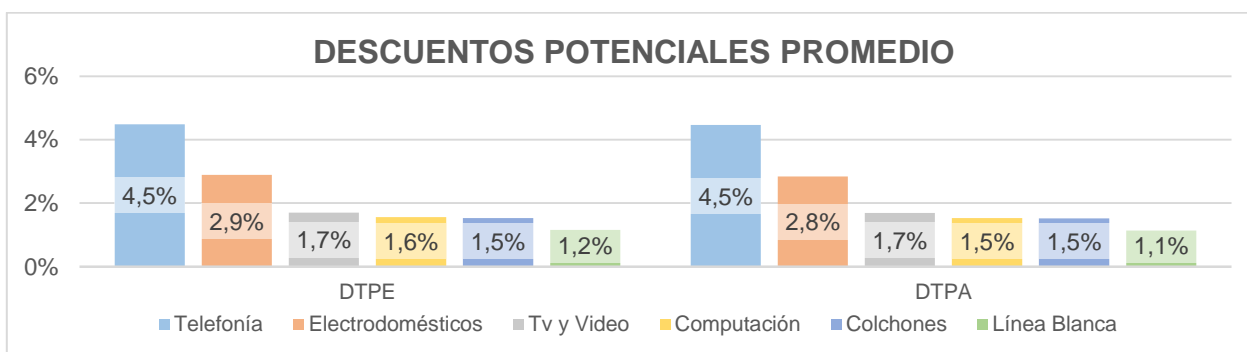


Ilustración 15: Promedio de los descuentos potenciales con tarjeta escalado y aumentado por categoría
Fuente: Elaboración propia

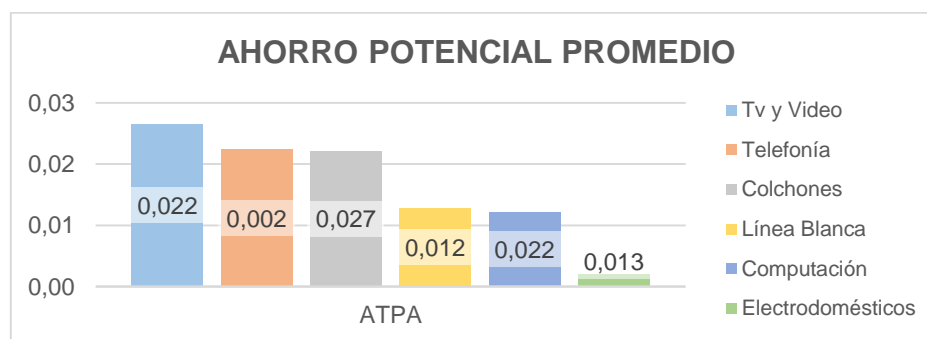


Ilustración 16: Promedio del ahorro potencial con tarjeta por categoría
Fuente: Elaboración propia

El orden de los descuentos potenciales se mantiene independiente del método utilizado (escalando o aumentando) y son bastante similares entre sí, siendo la categoría Telefonía la que presenta un mayor descuento potencial con tarjeta y Línea Blanca la con menor promedio. Para el caso del ahorro potencial el orden cambia debido a la inclusión de los niveles de precios de cada categoría. En este caso el mayor promedio lo presenta la categoría Televisión y Video y el menor la categoría Electrodomésticos. El cambio en

el orden de las categorías permitirá identificar que es más importante para los consumidores al momento de tomar una decisión de medio de pago, si la diferencia en los porcentajes de descuento o el monto de ahorro.

5.2.3. Modelamiento

Recordando que el OE2 es evaluar la efectividad de las promociones de descuento con medio preferente de pago en la apertura de nuevas tarjetas en clientes nuevos, se busca que el modelo propuesto sea capaz de responder la siguiente pregunta: ¿Cuántas aperturas se generan debido a una oferta?

En base a esto, el modelo propuesto es una regresión logística que estima la probabilidad de que la primera compra de un cliente se realice con tarjeta, lo que es equivalente a que adquiera la tarjeta, sujeto a las promociones de descuento en las categorías de interés existentes en su boleta.

$$p_b = \mathbb{P}(\text{Primera compra sea con tarjeta}|X) = \mathbb{P}(PT_b = 1|X)$$

Elección de la base para modelamiento

Como se mencionó anteriormente, se crean 4 bases de datos a partir de la base original. Primero, se crean la Base 1 y Base 2, excluyendo los datos del primer año (2012) y del primer y segundo año (2012 y 2013), respectivamente. Este corte se realiza para evitar tener primeras compras que realmente no se pueden considerar primeras, debido a que existe la posibilidad de que el cliente haya realizado una compra anterior en un periodo de tiempo acotado. En ambas bases, se encuentran sobrerrepresentadas las compras con tarjeta, por lo cual se crean las SubBases 1 y SubBase 2, balanceando las compras con y sin tarjeta de las Bases 1 y Base 2, respectivamente.

Con los primeros modelos probados se revisó el desempeño de estas 4 bases en términos de las métricas Accuracy, porcentaje de las estimaciones correctas totales, Recall, porcentaje de las observaciones de una clase correctamente clasificadas, y Precision, porcentaje de las predicciones de una clase correctas según la realidad.

En estos intentos, como era de esperarse, las medidas de Recall y Precision para la clase *otro medio de pago* calculadas tomando datos tanto dentro como fuera de la muestra fueron, en general, muy bajas para las Base 1 y 2. Esto indica que los modelos estimados sobre las bases sin balancear predicen casi siempre que las compras se pagaran con tarjeta, obteniéndose un bajo número de Verdaderos Negativos.

En relación a las medidas de desempeño de las SubBase 1 y SubBase 2, estas mostraron un alto nivel en estos primeros intentos de modelamiento, no presentando grandes diferencias entre sí, lo que indica que un corte mayor en los datos no genera ganancias a nivel de descripción y predicción de la realidad.

En conclusión, se establece que las mejores bases de datos para estimar los modelos son las SubBases 1 y 2 debido a que no ignoran una de las clases, pero de ellas no destaca ninguna sobre la otra en términos de desempeño. Finalmente, se prefiere trabajar con la SubBase 1 para privilegiar conservar un mayor número de observaciones.

Alternativas de modelamiento

Para escoger las variables independientes (X) a incluir en el modelo se tuvieron en cuenta diferentes posibilidades. En primer lugar, con el objetivo de controlar por el nivel de ventas que presenta cada compra, se desea incluir alguna de las variables generales disponibles que refleje la magnitud de la boleta estudiada.

$r(M_b, MPL_b)$	$r(M_b, MD_b)$	$r(MPL_b, MD_b)$
0,896	0,305	0,522

Tabla 24: Correlación de las variables generales consideradas
Fuente: Elaboración propia

En la Tabla 24 se puede observar que el nivel de correlación entre el monto total de la boleta (M_b) y el monto de precio lista (MPL_b) es muy elevado. Aunque en menor magnitud, lo mismo sucede entre el monto según precio lista y el monto de descuento (MD_b). Por sus niveles de correlación, no deben ser incluidas las tres variables en el modelo. Dado que M_b refleja más exactamente la magnitud de la boleta en comparación al MPL_b , y que el MD_b será incluido dentro del modelo a través de variables por categoría, se decide que la variable de nivel a incluir es solo el monto de la boleta (M_b).

En la sección 5.2.2 del análisis exploratorio, se pudo apreciar que el porcentaje de las primeras compras que tienen la tarjeta como medio de pago ha ido aumentando a través de los años, lo que hace importante incluir una variable temporal en el modelo. Para esto, se construye la variable continua año como la diferencia del año de la observación con el año 2013. También se incluye la variable transformada año².

Para reflejar el efecto que tiene cada categoría y su promoción en la probabilidad de compra con tarjeta, se decide incluir variables relacionadas al descuento por categoría. Por un lado, para incorporar si una boleta incluye productos en descuento se cuenta con la variable independiente Existe Descuento (ED_{cb}). Por el otro, para reflejar el nivel de descuento con tarjeta que presentan los productos de una categoría se cuenta con diferentes opciones: Porcentaje de Descuento (PD_{cb}), Descuento con Tarjeta Potencial Escalado ($DTPE_{cb}$), Descuento con Tarjeta Potencial Aumentado ($DTPA_{cb}$) y Ahorro con Tarjeta Potencial Aumentado ($ATPA_{cb}$). Para mayores detalles del cálculo de cada una de estas variables se recomienda revisar la sección 5.2.1.

En una etapa inicial del modelamiento se probaron configuraciones donde se incluían los dos tipos de variables de descuento, de existencia y profundidad, obteniéndose repetidamente signos contra intuitivos en los coeficientes, lo que podría deberse a un problema de correlación entre las variables.

	$r(ED_{cb}, PD_{cb})$	$r(ED_{cb}, DTPE_{cb})$	$r(ED_{cb}, DTPA_{cb})$	$r(ED_{cb}, ATPA_{cb})$
Colchones	0,807	0,803	0,807	0,715
Computación	0,767	0,749	0,766	0,790
Electrodomésticos	0,873	0,868	0,874	0,667
Línea Blanca	0,768	0,738	0,767	0,835
Telefonía	0,760	0,757	0,761	0,743
Tv y video	0,660	0,656	0,660	0,808

Tabla 25: Correlación entre la variable Existe Descuento con las variables de nivel de descuento
Fuente: Elaboración propia

En la Tabla 25 se presentan las correlaciones ($r_{x,y}$) entre la variable Existe Descuento (ED_{cb}) con los 4 tipos de variables de nivel de descuento según categoría. Entonces, por ejemplo, la correlación entre la variable ED_{cb} en la categoría Electrodomésticos con la variable PD_{cb} en la misma categoría ($r_{(ED_{Elect.b}, PD_{Elect.b})}$) es de 0,873, relación que se considera alta. Esta situación, alta correlación, se repite en todas las combinaciones de variables estudiadas, por lo que en los modelos no se deben combinar las variables ED_{cb} con las variables de nivel de descuento.

Por lo tanto, finalmente los modelos a evaluar incluyen la variable del monto de la boleta, la variable correspondiente al año y su transformación cuadrática, y un conjunto de variables de descuento por categoría:

- *Modelo 1 (M1)*: Incluye las seis variables de si existe descuento con tarjeta (ED_{cb}) para estudiar si la existencia de la promoción explica el comportamiento de pago.

$$(X'\beta)_{M1} = \beta_0 + \beta_M M_b + \beta_a \text{año}_b + \beta_{a2} \text{año}_b^2 + \sum_{c \in C} \beta_{c,ED} ED_{cb}$$

- *Modelo 2 (M2)*: Incluye las seis variables de porcentaje de descuento de los productos en descuento (PD_{cb}) para estudiar si el nivel de rebaja que obtiene un cliente es capaz de explicar el comportamiento de pago de este.

$$(X'\beta)_{M2} = \beta_0 + \beta_M M_b + \beta_a \text{año}_b + \beta_{a2} \text{año}_b^2 + \sum_{c \in C} \beta_{c,PD} PD_{cb}$$

- *Modelo 3 (M3)*: Incluye las seis variables de descuento con tarjeta potencial escalado ($DTPE_{cb}$) para estudiar si el descuento con tarjeta al que optaba el cliente en su ocasión de compra es capaz de explicar el comportamiento de pago de este.

$$(X'\beta)_{M3} = \beta_0 + \beta_M M_b + \beta_a \text{año}_b + \beta_{a2} \text{año}_b^2 + \sum_{c \in C} \beta_{c,DTPE} DTPE_{cb}$$

- *Modelo 4 (M4)*: Incluye las seis variables de descuento con tarjeta potencial aumentado ($DTPA_{cb}$) para estudiar si el descuento con tarjeta al que optaba el cliente en su ocasión de compra es capaz de explicar el comportamiento de pago.

$$(X'\beta)_{M4} = \beta_0 + \beta_M M_b + \beta_a \text{año}_b + \beta_{a2} \text{año}_b^2 + \sum_{c \in C} \beta_{c,DTPA} DTPA_{cb}$$

- *Modelo 5 (M5)*: Incluye las seis variables de ahorro con tarjeta potencial aumentado ($ATPA_{cb}$) para estudiar si el monto de rebaja al que se optaban al utilizar la tarjeta es capaz de explicar el comportamiento de pago de los clientes.

$$(X'\beta)_{M5} = \beta_0 + \beta_M M_b + \beta_a \text{año}_b + \beta_{a2} \text{año}_b^2 + \sum_{c \in C} \beta_{c,ATPA} ATPA_{cb}$$

Comparación de modelos

Recapitulando, se utiliza una regresión logística para modelar la probabilidad de que la primera compra sea con medio preferente de pago, y se comparan 5 set de variables independientes en términos de su desempeño estimativo.

$$\text{logit}(p_b) = \text{logit}(PT_b = 1) = (X'\beta)_{Mi}, \quad i \in \{1,2,3,4,5\}$$

El modelamiento se lleva a cabo utilizando el software Stata sobre la SubBase1, excluyendo todas las observaciones del año 2016 para utilizarlos posteriormente como datos de testeo. Los resultados se pueden encontrar en la sección 8.2.2 de Anexos.

En términos de significancia de las variables de interés, los 5 modelos muestran un alto desempeño, siendo los coeficientes de ED_{cb} , PD_{cb} , $DTPE_{cb}$, $DTPA_{cb}$ y $ATPA_{cb}$ significativos para todas las categorías a un nivel de confianza superior al 95%. Esto indica que las variables escogidas presentan una relación real con la probabilidad modelada y son atingentes al problema.

A priori se espera que los coeficientes estimados para todas las variables relacionadas al descuento por categoría sean de signo positivo, bajo el supuesto de que la existencia de descuentos y su mayor profundidad incentivan el uso de tarjetas, especialmente para los casos de $DTPE_{cb}$ y $DTPA_{cb}$, que representa el descuento solo disponible cuando se utiliza este medio de pago, y para el caso de $ATPA_{cb}$, que representa el monto rebajado si se usa tarjeta versus otro medio. Los modelos estimados reflejan esta intuición, siendo todos los coeficientes de las variables de descuento mayores a cero.

Finalmente, se evalúa el desempeño de cada modelo con las métricas Accuracy, Recall y Precision (Tabla 26). Los desempeños tanto descriptivos (dentro de la muestra) como predictivos (fuera de la muestra) son bastante similares entre los modelos, y se encuentran en niveles aceptables para una estimación, teniendo, por ejemplo, todos los modelos una tasa de acierto sobre el 75% dentro de la muestra y sobre el 80% fuera de la muestra. Los modelos 1 y 5 destacan en su desempeño repetidas veces, pudiendo concluir que, en términos generales, son los que mejor describen y predicen la realidad.

Medida de desempeño	Dentro de la muestra					Fuera de la muestra				
	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
<i>Accuracy</i>	82%	76%	75%	76%	82%	82%	80%	81%	80%	82%
<i>Recall</i> ($PT_b = 1$)	71%	63%	63%	63%	67%	81%	82%	86%	83%	79%
<i>Recall</i> ($PT_b = 0$)	93%	88%	85%	87%	95%	88%	72%	66%	70%	91%
<i>Precision</i> ($PT_b = 1$)	90%	83%	79%	82%	92%	95%	90%	89%	90%	96%
<i>Precision</i> ($PT_b = 0$)	78%	72%	72%	72%	76%	59%	55%	59%	56%	58%

Tabla 26: Medidas de desempeño dentro y fuera de la muestra de los modelos comparados

Fuente: Elaboración propia con uso de software stata

En resumen, ninguno de los modelos destaca en términos de significancia estadística ni concordancia de los coeficientes con el signo esperado. En términos de las métricas de desempeño, son los modelos 1 y 5 quienes muestran ser ligeramente mejores en poder descriptivo y predictivo. Recordando, el modelo 1 incluye las variables

indicadoras de si la boleta incluye productos en descuento de cada categoría, y el modelo 5 incluye las variables de ahorro potencial utilizando la tarjeta en vez de otro medio de pago. Por lo tanto, con el fin de privilegiar la posibilidad de encontrar un nivel de descuento óptimo para el incentivo de la apertura de tarjetas en nuevos clientes, se escoge el Modelo 5 para continuar con el análisis, realizar la evaluación de impacto y la simulación de escenarios.

Dado que las diferentes variables que indican el nivel de descuento por boleta no están calculadas sobre los mismos universos ni de maneras similares, esto es, no reflejan la misma información de la boleta, se evaluó la posibilidad de extender el Modelo 5, el cual cuenta originalmente con las variables $ATPA_{cb}$, añadiendo alguna de las otras variables de nivel de descuento (PD_{cb} , $DTPE_{cb}$ o $DTPA_{cb}$). Sin embargo, las correlaciones entre las variables de cada categoría (8.2.2) resultaron ser nuevamente altas, lo cual llevaría a un problema de identificación similar al descrito con anterioridad.

5.2.4. Resultados y análisis

Resultados del modelo seleccionado

$$\text{logit}(p_b) = \text{logit}(PT_b = 1) = \beta_0 + \beta_M M_b + \beta_a \text{año}_b + \beta_{a2} \text{año}_b^2 + \sum_{c \in C} \beta_{c,ATPA} ATPA_{cb}$$

El modelo final (M5) se calcula sobre la SubBase 1 y tiene la forma funcional recién expuesta. Dado que ya se ha seleccionado el modelo con el que se trabajará, y para maximizar el ajuste a los datos, los coeficientes de las variables se re-estiman utilizando todos los años disponibles en la SubBase 1, incluyendo el año 2016 (Tabla 27) .

Variable	$\hat{\beta}$
M_b	0,25***
año_b	0,02
año_b^2	0,09***
$ATPA_{\text{Colchones},b}$	32,87***
$ATPA_{\text{Computación},b}$	35,58***
$ATPA_{\text{Electrodomésticos},b}$	197,67***
$ATPA_{\text{Línea Blanca},b}$	32,10***
$ATPA_{\text{Telefonía},b}$	31,85***
$ATPA_{\text{Televisión y Video},b}$	19,04***
Constante	-1,43***
N° de observaciones	82.682
Pseudo R²	0,36

Tabla 27: Coeficientes estimados para el modelo seleccionado
Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Fuente: Elaboración propia con uso de software stata

De los resultados, se puede evidenciar que el monto total de una boleta aumenta la probabilidad de que la compra haya sido cancelada con tarjeta. También se ve reflejada la relación positiva identificada en la exploración de datos, entre los años y la probabilidad estudiada. Sin embargo, solo su componente cuadrática resulta significativa

estadísticamente, lo que indica que este crecimiento de la probabilidad con los años, va acelerando en el tiempo.

De las variables de ahorro potencial, destaca el coeficiente de la categoría Electrodomésticos, pareciendo ser el tipo de monto rebajado con un mayor impacto sobre la probabilidad de utilización de la tarjeta. No obstante, como estas variables se encuentran calculadas en términos de montos y los niveles de precios de las categorías difieren entre sí, es poco realista comparar de manera directa los coeficientes obtenidos. En el análisis de impacto, expuesto más adelante, es posible comparar de manera más precisa los resultados dado que se incluye la información del nivel de precios.

Finalmente, todas las variables consideradas en el modelo presentan una significancia estadística superior al 95%, excepto la componente lineal de los años.

Impacto de las variables

La probabilidad de que un cliente utilice tarjeta en su primera compra, en base a un conjunto de valores para las variables independientes (X), se estima de la siguiente manera:

$$\hat{p} = \exp(X' \hat{\beta}) / (1 + \exp(X' \hat{\beta}))$$

Ante un cambio en los valores definidos para las variables explicativas, esta probabilidad se verá afectada de mayor o menor manera dependiendo de la magnitud de la variación, los coeficientes estimados y el nivel inicial de la probabilidad. Es por esto que, para poder cuantificar el impacto que tienen cada una de las variables utilizadas en el modelamiento, se debe, primero, definir un nivel de referencia, con el cual comparar las probabilidades estimadas. El nivel base para esta evaluación se define como el promedio observado de las variables independientes y seleccionando el año 2013 como punto base (Tabla 28).

Variable	Formulación	Nivel de referencia (NR)
M_b	\bar{M}_b	1,579
$año_b$	0	0
$año_b^2$	0	0
$ATPA_{Colchones,b}$	$\overline{ADTP_{Colchones,b}}$	0,022
$ATPA_{Computación,b}$	$\overline{ATPA_{Computación,b}}$	0,0122
$ATPA_{Electrodomésticos,b}$	$\overline{ADTPA_{Electrodomésticos,b}}$	0,002
$ATPA_{Línea Blanca,b}$	$\overline{ADTPA_{Línea Blanca,b}}$	0,013
$ATPA_{Telefonía,b}$	$\overline{ADTPA_{Telefonía,b}}$	0,022
$ATPA_{Televisión y Video,b}$	$\overline{ADTPA_{Televisión y Video,b}}$	0,027

$$\widehat{p}_{NR} = 89.8\%$$

Tabla 28: Estimación de la probabilidad de referencia

Fuente: Elaboración propia

Con la probabilidad de referencia definida, se puede medir el impacto que tiene cada variable realizando un análisis *ceteris paribus*, esto es, variando uno de los factores mientras se mantienen los valores de las otras variables constantes. La variación utilizada es un aumento de un 10% al valor de referencia de cada variable, y se ha calculado la

probabilidad estimada de pago con tarjeta ante dicha variación (Tabla 29). Notar que, a pesar de ser un análisis donde las otras variables se mantienen constantes, para el caso de cuando cambia uno de los ahorros potenciales, también debe cambiar el monto total de la boleta por el aumento en el descuento con tarjeta. Este cambio se refleja disminuyendo el monto de la boleta en el mismo monto en el que aumenta el ahorro potencial.

Variable modificada (V)	Valor evaluado de V (\$100.000)	Probabilidad estimada (\hat{p}_V)
M_b	1,737	90,19%
$ATPA_{Colchones,b}$	0,024	90,48%
$ATPA_{Computación,b}$	0,013	90,22%
$ATPA_{Electrodomésticos,b}$	0,002	90,21%
$ATPA_{Línea Blanca,b}$	0,014	90,20%
$ATPA_{Telefonía,b}$	0,025	90,46%
$ATPA_{Televisión y Video,b}$	0,029	90,28%

Tabla 29: Análisis de impacto de las variables

Fuente: Elaboración propia

Las diferencias entre estas nuevas probabilidades y la probabilidad del nivel de referencia por categoría se muestran en la Ilustración 17. Colchones y Telefonía son las categorías que tienen un impacto mayor sobre la probabilidad estudiada a partir de un cambio del ahorro potencial obtenido en dichas categorías pagando con tarjeta de un 10% con respecto a su valor promedio. Por su parte, Electrodomésticos, a pesar de tener el coeficiente estimado más alto, tiene el segundo impacto más bajo en la probabilidad, esto debido a que el nivel de precios promedio de la categoría es considerablemente más bajo que el de las otras.

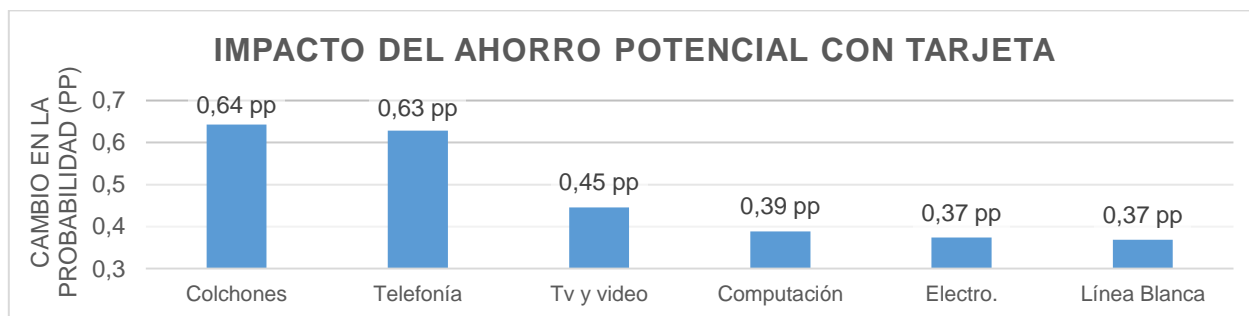


Ilustración 17: Cambio en la probabilidad por la variación de las variables de ahorro potencial

Fuente: Elaboración propia

Entonces, por ejemplo, si se aumenta en un 10% el ahorro promedio potencial de la categoría Computación, cambiando de \$1.220 a \$1.350, la probabilidad de pago con tarjeta en la primera compra aumenta de 89,8% a 90,2%. Recordando que la base de datos inicial cuenta con 256.009 primeras compras, este aumento en la probabilidad significa un aumento, en esperanza, de 994 personas que comienzan su vida como clientes del retail abriendo la tarjeta propia del negocio.

Además, es interesante analizar qué pasaría si el ahorro potencial pagando con tarjeta fuera nulo en todas las categorías, esto es, que el descuento obtenido pagando con tarjeta sea el mismo que pagando con otro medio de pago. En este caso, la

probabilidad de pago con tarjeta baja a un 26,7%, lo que se traduce a que, en esperanza, solo 68.319 de los 256.009 nuevos clientes abriría tarjeta para su primera compra, un valor considerablemente menor que lo estimado con los ahorros potenciales observados.

5.2.5. Configuración óptima de las promociones (OE4)

La hipótesis seguida en el desarrollo de este objetivo es que un cliente que comienza su vida en la empresa poseyendo la tarjeta de crédito representa un valor mayor en el largo plazo para ella. Debido a esto, se realizan promociones dirigidas al medio de pago para incentivar el uso de tarjetas. Sin embargo, al realizar estos descuentos dirigidos también existe una pérdida de margen importante en las ventas. El equilibrio entre el aumento del valor del cliente y el margen perdido establece un punto óptimo de descuento, el cual se analiza a continuación a través de una simulación de escenarios.

Para el análisis se supone que se tiene la primera compra de un cliente cuya boleta δ incluye productos de la categoría κ en descuento, y se evalúa qué pasa con la utilidad esperada de ese cliente para la empresa cuando varía el ahorro potencial con tarjeta de la categoría κ . Para ello, se decide variar el descuento histórico con tarjeta en la categoría κ , lo que repercute en el ahorro potencial de la siguiente manera:

$$ATPA_{\kappa\delta}^* = \overline{MPL}_{\kappa} (DHT_{\kappa}^* - DHO_{\kappa})$$

La probabilidad de compra con tarjeta debido a este nivel de ahorro potencial, se estima de la misma manera que para el análisis de impacto. Además, de igual manera que para el nivel de referencia, para las demás variables del modelo se utilizarán sus promedios históricos. Recordar que el cambio en el ahorro potencial de la categoría κ cambia también el monto de la boleta.

$$\begin{aligned} M_{\kappa}^* &= \bar{M} + \overline{ATPA}_{\kappa} - ATPA_{\kappa\delta}^* \\ X_{\delta}^* \hat{\beta} &= \hat{\beta}_0 + \hat{\beta}_M M_{\kappa}^* + \hat{\beta}_{\kappa, ATPA} ATPA_{\kappa\delta}^* + \sum_{c \in C \setminus \{\kappa\}} \hat{\beta}_{c, ATPA} \overline{ATPA}_c \\ \mathbb{P}(PT_{\delta} = 1)^* &= \exp(X_{\delta}^* \hat{\beta}) / (1 + \exp(X_{\delta}^* \hat{\beta})) \end{aligned}$$

Con esta probabilidad es posible calcular la utilidad esperada generada por un cliente para la empresa. Para ello, se calcula el Valor del Cliente de cada observación de la base de datos como el margen acumulado producido por el cliente para la empresa en los años disponibles, y se computa el promedio de aquellos que hayan realizado compras con descuento en la categoría κ , separando entre quienes comienzan su vida con tarjeta propia ($\overline{VC}_{\kappa T}$) y los que comienzan su vida con otro medio de pago ($\overline{VC}_{\kappa O}$). También se calculan los costos promedio de los productos incluidos en las boletas de primeras compras que incluyen productos de la categoría κ , y se computan las utilidades generadas por cada tipo de cliente según el medio de pago utilizado de la siguiente manera:

$$\begin{aligned} U_{\kappa T}^* &= \overline{VC}_{\kappa T} + MB_{\kappa T}^* - \bar{C}_{\kappa} \quad , \quad MB_{\kappa T}^* = \overline{MPL}_{\kappa} (1 - DHT_{\kappa}^*) \\ U_{\kappa O} &= \overline{VC}_{\kappa O} + MB_{\kappa O} - \bar{C}_{\kappa} \quad , \quad MB_{\kappa O} = \overline{MPL}_{\kappa} (1 - DHO_{\kappa}) \end{aligned}$$

Finalmente, la esperanza de la utilidad generada por el cliente cuya primera compra tiene la configuración de la boleta δ se calcula como:

$$\mathbb{E}(U_{\delta}) = \mathbb{P}(PT_{\delta} = 1)^* * U_T^* + (1 - \mathbb{P}(PT_{\delta} = 1)^*) * U_O$$

Los valores utilizados para la simulación se presentan en la Tabla 30. Los montos de los valores de los clientes, los montos de precio lista y los costos se encuentran calculados en UFs para poder hacer comparables los valores a través del tiempo.

Categoría (κ)	\overline{ATPA}_{κ}	$\overline{VC}_{\kappa T}$	$\overline{VC}_{\kappa O}$	DHT_{κ}	DHO_{κ}	\overline{MPL}_{κ}	\overline{C}_{κ}
Colchones	0,022	10,94	9,97	13%	2%	11,66	7,80
Computación	0,012	8,05	5,47	8%	2%	9,15	7,18
Electro.	0,002	7,28	5,60	7%	2%	1,91	1,37
Línea Blanca	0,013	8,39	5,77	7%	2%	10,74	8,56
Telefonía	0,022	6,82	4,92	17%	6%	5,23	3,66
Tv y video	0,027	7,46	7,75	15%	4%	11,00	9,04

Tabla 30: Datos por categoría utilizados para la simulación
Fuente: Elaboración propia

En la Tabla 31 se presentan los resultados obtenidos de la simulación de escenarios para cada categoría. Entonces, por ejemplo, para la categoría Computación el valor óptimo del descuento histórico con tarjeta es de 3% (Ilustración 18). No obstante, el descuento histórico con tarjeta observado en la realidad es de un 8%, perdiéndose una utilidad para la empresa de 0,3UF por cada cliente que incluyó en su primera compra productos de la categoría Computación en descuento.

Categoría (κ)	Descuento histórico con tarjeta (DHT_{κ}^*)										Óptimo
	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	
Colchones	14,29	14,39	14,38	14,31	14,21	14,10	13,99	13,87	13,76	13,64	2%
Computación	9,28	9,50	9,58	9,58	9,52	9,45	9,37	9,28	9,19	9,10	3%
Electrodo.	7,36	7,58	7,68	7,71	7,71	7,70	7,69	7,67	7,65	7,63	5%
Línea Blanca	9,92	10,10	10,14	10,09	10,02	9,92	9,82	9,71	9,61	9,50	3%
Telefonía	6,90	7,09	7,27	7,44	7,58	7,69	7,75	7,79	7,80	7,79	9%
Tv y video	9,27	9,22	9,13	9,03	8,92	8,80	8,68	8,57	8,45	8,34	1%

Tabla 31: Utilidad esperada según escenario y categoría
Fuente: Elaboración propia

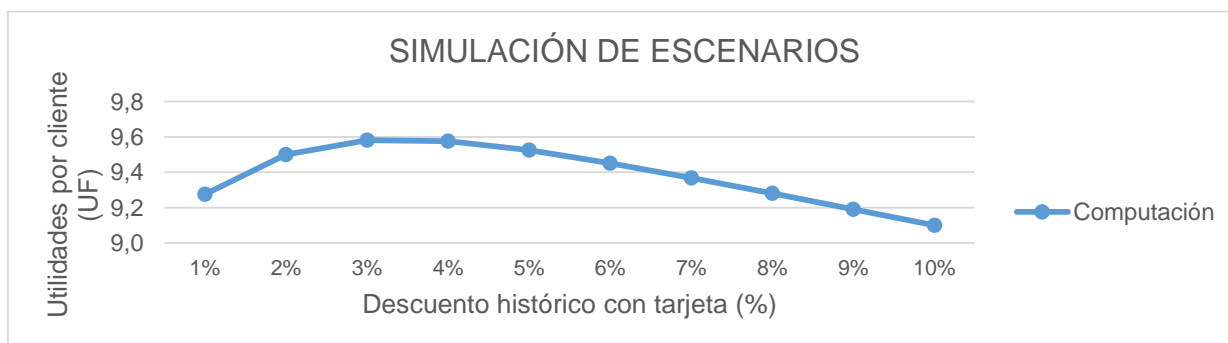


Ilustración 18: Cambio de las utilidades esperadas según escenario
Fuente: Elaboración propia

Que el descuento óptimo encontrado sea menor al descuento observado en la realidad se repite en todas las categorías, lo que puede estar generando una pérdida de margen importante para la empresa. Si se toma el número de clientes cuyas boletas de primeras compras incluyen descuento en cada categoría κ (N_{κ}) y se comparan las utilidades generadas por el descuento con tarjeta óptimo encontrados con el observado, se encuentra una pérdida de margen de 85.222UF en los 3,5 años estudiados.

Categoría (κ)	N_{κ}	$\mathbb{E}(U)$ óptimo	$\mathbb{E}(U)$ observado	$\Delta\mathbb{E}(U)$	$\Delta\mathbb{E}(U) * N_{\kappa}$
Colchones	18.781	14,39	13,29	1,10	20.627
Computación	22.365	9,58	9,28	0,30	6.748
Electrodo.	23.702	7,71	7,69	0,03	617
Línea Blanca	22.039	10,14	9,82	0,32	6.983
Telefonía	43.476	7,80	7,50	0,30	12.916
Tv y video	24.946	9,27	7,78	1,50	37.332
Total	155.309	-	-	-	85.222

Tabla 32: Margen potencial por la optimización de promociones
Fuente: Elaboración propia

5.2.6. Conclusiones del Objetivo Específico 2

Durante el desarrollo de este objetivo se pudo comprobar y cuantificar la relación existente entre las promociones de descuento y la probabilidad de que una boleta de una primera compra sea cancelada con tarjeta propia del retail. Esto se lleva a cabo bajo el supuesto de que si la primera compra de un cliente es cancelada con medio preferente de pago, implica que el nuevo cliente ha abierto recientemente la tarjeta motivado por los descuentos observados.

Un resultado interesante alcanzado en este objetivo, previo al modelamiento, fue el del promedio del descuento por boleta según medio de pago, utilizando todas las transacciones disponibles en la base original, para cada una de las categorías de interés. Como era de esperarse la diferencia entre los promedios de descuento con tarjeta y sin tarjeta es muy alta, comprobándose el hecho de que el retail estudiado utiliza consistentemente incentivos relacionado con promociones para promover el uso de su tarjeta.

En el modelamiento realizado, se comprobó el hecho de que bases de datos desbalanceadas tienen problemas para estimar las clases presentes en menor proporción, recalando la necesidad de resolver esta sub-representación previo a la estimación de los coeficientes. Además, se hizo evidente el efecto de la correlación de las variables en los modelos, las cuales generan problemas de identificación y parámetros poco interpretables en la realidad.

De los resultados del modelo se puede concluir que la probabilidad que los clientes habrán tarjeta para su primera compra va creciendo en el tiempo de manera exponencial, lo que representa una tendencia en el comportamiento del consumidor del retail a utilizar en mayor proporción este medio de pago. También se observó que a mayor gasto de los clientes (monto de la boleta), este tiende a adquirir con mayor probabilidad la tarjeta preferente. Sin embargo, estas son dos características de las compras donde la empresa

no tiene gran poder de acción (especialmente en el año de la boleta), a diferencia de los descuentos y ahorro ofrecidos a sus consumidores por la utilización de la tarjeta.

El ahorro al que puede optar un cliente al utilizar la tarjeta del retail tiene siempre un impacto positivo en el número de clientes que adquieren la tarjeta, independiente de la categoría en donde se aplique. Sin embargo, existen categorías donde los descuentos son más efectivos que en otras. De las categorías estudiadas, es Colchones la que tiene un mayor impacto sobre la probabilidad de pago con tarjeta ante variaciones alrededor de la media, y es en esta categoría donde se deben concentrar los esfuerzos si se desea generar un aumento en el número de clientes nuevos que abren la tarjeta. No obstante, es importante mencionar que para este análisis no se han considerado costos.

Por último, los niveles de descuento óptimos para la atracción de nuevos clientes con tarjeta en términos de utilidad esperada, resultaron ser siempre menores a los observados, entre un 1% a un 9%, lo que indica que el negocio está perdiendo un margen potencial al aplicar estas promociones. Del análisis de escenarios, se obtuvo que esta pérdida de utilidades alcanzaría 85.222UF aproximadamente en los 3,5 años estudiados (2013 a mediados del 2016), lo que equivale a más de \$2.000MM (Calculado en base al valor de la UF al primero de julio del 2018).

Sin embargo, esta pérdida no considera el valor que el nuevo cliente que adquiere tarjeta está generando al negocio financiero del mismo holding, por lo que en términos de empresa, los descuentos actualmente aplicados podrían no estar afectando las ganancias tan fuertemente como aquí se estima. Finalmente, se cree que son estas ganancias no observadas en los datos disponibles las que podría estar motivando que el retail establezca descuentos tan profundos con tarjeta de crédito, motivado por una subvención interna entre los diferentes negocios para las actividades promocionales estudiadas.

Cabe mencionar que el análisis realizado no es del tipo causal, por lo cual, a pesar de haber podido establecer relaciones entre las profundidades de los descuentos con la probabilidad de que un cliente comience su vida en el retail con tarjeta, no se puede concluir a ciencia cierta que una variable provoque la otra.

5.3. Efecto de las promociones en la atracción de nuevos clientes

El tercer objetivo específico propuesto en esta memoria abarca el estudio del efecto que tienen las promociones en el tipo de cliente que atrae la empresa a obtener la tarjeta del retail. En este caso, el tipo de cliente viene dado por su valor para la empresa, según las utilidades que este significa para el negocio. En la práctica, se estudia la relación entre la configuración de las promociones en la primera compra, en términos de profundidad y amplitud, con el valor del cliente poseedor de tarjeta incorporado.

5.3.1. Preparación de la base de datos

En esta sección se utiliza la tabla de transacciones, la tabla de información de los clientes y la tabla utilizada en el OE2. Esta información se trabaja previamente para poder obtener, de cada cliente identificado que posee tarjeta, la configuración de sus primeras compras, su valor de cliente en el periodo estudiado y sus características demográficas.

Lista de clientes únicos

De la base de datos creada para el OE2, con las primeras compras de los clientes que comienzan su relación con la empresa en el periodo comprendido desde inicios del 2013 a mediados del 2016, se extraen las transacciones realizadas con tarjeta propia del negocio. Con esto, se obtiene una lista de todos los clientes que abrieron tarjeta del negocio para su primera compra, junto con los datos de esta transacción.

Obtención del Valor del Cliente

De la tabla de transacciones, se extraen todas las compras realizadas por los clientes identificados en el punto anterior. Esta información es agregada a nivel de boleta, calculándose el margen para la empresa por cada una de las compras realizadas por los individuos estudiados. Posteriormente, los márgenes son transformados a Unidades de Fomento (UF), según la información obtenida del Servicio de Impuestos Internos de Chile (Servicio de Impuestos Internos, 2018) dependiendo del día de ocurrencia de la compra.

Luego de transformado los montos, estos se agregan a nivel de individuo, obteniendo el margen en UF acumulado generado por las compras de un cliente en la empresa, lo que se considera el Valor del Cliente (VC_i). Finalmente, para poder hacer comparables los valores de los clientes entre sí, estos se dividen por el número de días desde la primera compra del cliente hasta el último día del que se tiene información transaccional (16/07/2016), lo que se denomina vida del cliente. Este último valor obtenido es el Valor del Cliente Diario Promedio de cada individuo (VCD_i).

Tratamiento de datos extremos

En una exploración inicial de VCD_i , se observa que el promedio de los datos correspondientes al año 2016 es considerablemente mayor que el promedio del resto de los datos, situación generada por la corta vida de los clientes que tienen su primera compra en la última sección de la base de datos. Con el fin de evitar posibles *confounding* entre las características de las promociones y el año de ingreso de los clientes, se decide eliminar el año 2016 de la base a trabajar.

Posteriormente, al observar la distribución de frecuencias de VCD_i , se puede evidenciar que existen colas de datos importantes tanto en el lado de los números

positivos como negativos, teniéndose el caso más grave en el lado de los positivos. Esto distorsiona el promedio de los datos y aumenta considerablemente la desviación estándar de estos, lo que podría generar problemas al estimar modelos que intenten ajustarse a los valores extremos presentes.

Base	# Obs.	Promedio	Mínimo	Máximo	Desviación estándar	Coef. de Variación
<i>Sólo 2016</i>	23.034	0,114	-4,384	16,324	0,406	3,55
<i>Sin 2016</i>	191.619	0,014	-0,784	262,055	0,600	41,50
<i>Sin 2016 - Filtrada</i>	187.787	0,012	-0,002	0,076	0,012	1,03

Tabla 33: Estadísticos básicos de VCD
Fuente: Elaboración propia

Debido a esto, se decide eliminar las entradas correspondientes al 1% inferior y superior de las observaciones de VCD_i . Estos límites se determinaron a partir del análisis visual de las colas de los datos. El procedimiento realizado permite pasar de un coeficiente de variación en los datos originales (sin 2016) de 41,5 a un coeficiente de 1 en la base de datos filtrada (sin 2016).

Variables de interés

En esta oportunidad se utilizará una indexación i en las variables, para indicar los datos de cada individuo. Se mantienen las variables de interés de las primeras compras generadas para el OE2, agregando la variable binaria Descuento en Boleta (DB_i) que indica si la boleta incluye algún producto en descuento de las categorías de interés. También se aplica una transformación a VCD_i , normalizándola y generando el Valor del Cliente Diario Normalizado Promedio para cada cliente i ($VCDN_i$).

Creación de base de testeo

La base de testeo, necesaria para evaluar los posibles modelos a utilizar, se crea a partir de una muestra aleatoria correspondiente al 20% de los datos de la base sin 2016 filtrada. El otro 80% se utiliza como datos de calibración.

5.3.2. Valor de los clientes

Información general

	Variable	Descripción
<i>Variables generales</i>	ID cliente	Identificador único del cliente i
	ID boleta	Identificador único de la boleta del cliente i
	VCD_i	Valor del cliente i diario promedio (UF)
	$VCDN_i$	Valor del cliente i diario promedio normalizado (UF)
<i>Variables demográficas</i>	$EDAD_i$	Edad del cliente i
	$g_{k,i}$	Grupo socioeconómico del cliente i . Con $k \in \{ABC1,C2,C3,D,E\}$
	$s_{k,i}$	Sexo del cliente i . Con $k \in \{Mujer, Hombre, Desconocido\}$
<i>Variables generales</i>	DB_i	1 si boleta del cliente i incluye algún descuento las categorías de interés, 0 si no

Tabla 34: Variables disponibles
Fuente: Elaboración propia

La base final con la que se trabaja cuenta con 187.787 registros de clientes que comienzan su relación con el retail entre los años 2013 y 2015. Los datos incluyen tanto la información de su primera compra, como la utilidad que cada individuo genera para la empresa desde su incorporación como clientes hasta mediados del año 2016. Las variables disponibles para cada cliente se presentan en la Tabla 34, y se incluyen las variables de las boletas de las primeras compras presentadas en la Tabla 20 del OE2.

Características de los clientes

Los valores de los clientes diarios promedio (VCD_i) se encuentran principalmente concentrado alrededor de cero (Ilustración 19), teniendo un promedio de 0,012UF en el periodo estudiado (Tabla 33). A pesar de observarse una cola derecha en los datos, el coeficiente de variación de esta métrica es bajo, siendo aproximadamente 1 punto.

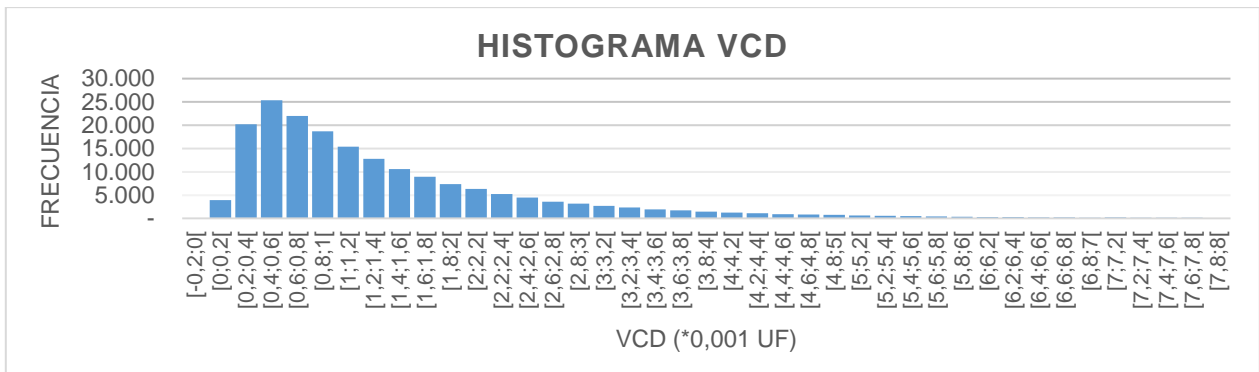


Ilustración 19: Frecuencia de los valores de los clientes diario
Fuente: Elaboración propia

Los estadísticos descriptivos básicos de VCD_i según edad, grupo socioeconómico y sexo de los clientes se pueden encontrar en la sección 8.3.1. Los clientes incluidos en la muestra se concentran etariamente entre los 30 y los 60 años, con más de un 60% de las observaciones en ese rango. En términos de grupo socioeconómico (en adelante GSE), la mayor concentración de clientes está en los grupos C2, C3 y D, con una presencia superior al 70%. El GSE que presenta un mayor valor de cliente diario promedio es el grupo ABC1, mientras que el de menor valor es el grupo D. La distribución en términos de sexo de los clientes casi igualitaria, teniendo que los hombres representan un 51%, mientras que las mujeres son un 48% de la base de datos. Las mujeres muestran ser clientes más rentables para la empresa, teniendo un valor de cliente diario promedio un 20% mayor que los hombres.

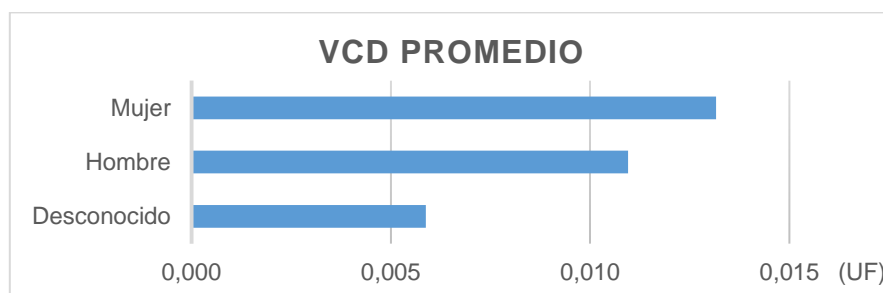


Ilustración 20: Valor del cliente diario según genero
Fuente: Elaboración propia

Características de la primera compra

La distribución temporal de las boletas de las primeras compras se muestra en la Tabla 35. Se puede observar que el número de clientes nuevos ha ido bajando en el tiempo, situación que se condice con lo observado en el OE2.

Año	2013	2014	2015	General
Número de clientes	68.630	72.225	46.932	187.787

Tabla 35: Distribución temporal de los nuevos clientes con tarjeta

Fuente: Elaboración propia

Al analizar el valor del cliente diario promedio según su fecha de primera compra, se aprecia un alza a través del tiempo (Ilustración 21), lo que podría indicar que los clientes obtenidos por el retail son cada vez más rentables en términos de márgenes reportados para la empresa. Se concluye que es importante considerar variables temporales en el modelamiento para poder capturar la tendencia observada.

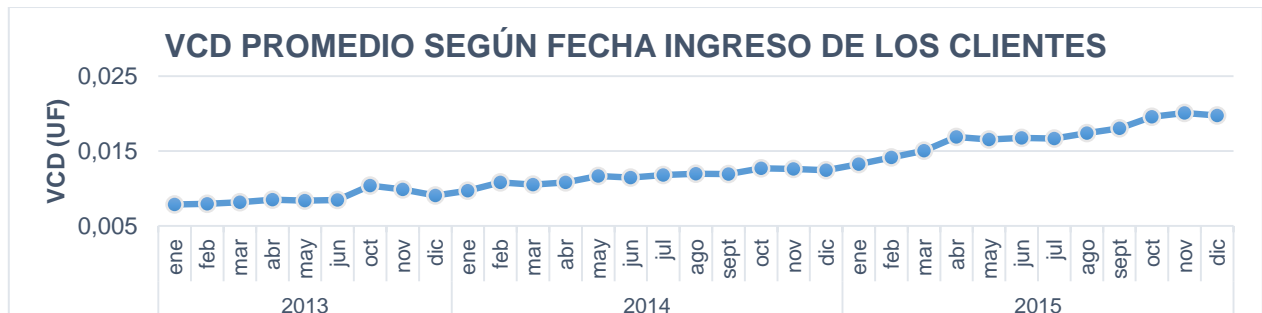


Ilustración 21: Evolución del valor del cliente según ingreso al retail

Fuente: Elaboración propia

Los promedios de las características numéricas de las primeras boletas (monto, unidades, costo, monto de descuento y monto total sin descuento) permanecen similares a los visto en el análisis exploratorio del OE2 (Tabla 22), tanto en términos numéricos como en comportamiento. Adicionalmente, las correlaciones entre dichas variables y el valor del cliente diario promedio no reflejan a priori una relación fuerte entre las medidas, siendo todas menores a 0,2.

Las boletas que incluyen algún producto en descuento en las categorías de interés representan un 69% del total, y tienen un valor del cliente promedio ligeramente mayor que las que no, siendo dicho promedio igual a 0,012 UF en el caso de las boletas para las cuales $DB_i = 1$ y 0,011 UF para las que $DB_i = 0$.

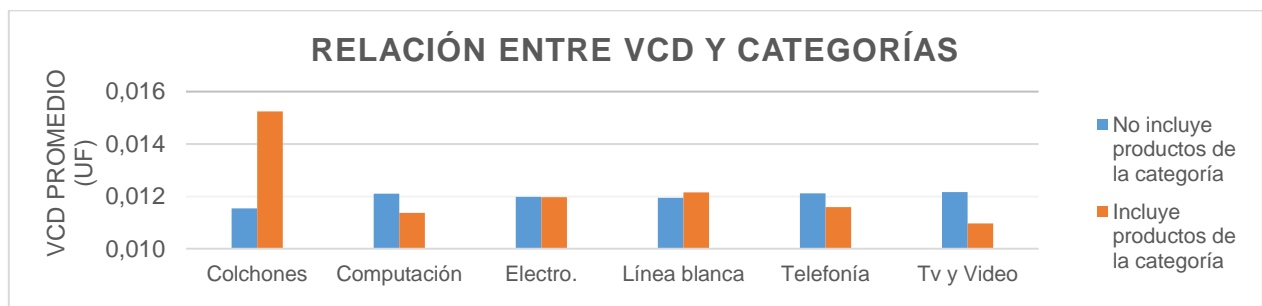


Ilustración 22: Relación entre VCD y las boletas que incluyen productos de cada categoría

Fuente: Elaboración propia

Al separar las boletas entre las que incluyen productos de las categorías de interés y las que no, se puede observar que las boletas que incluyen productos de Colchones y Línea Blanca presentan promedios mayores de VCD_i que las que no. La relación opuesta se da en las otras categorías (Ilustración 22).

Los porcentajes de boletas que incluyen productos en descuento de cada categoría se encuentran entre el 9% y el 18% (Ilustración 23). Del mismo modo que para la Ilustración 22, si se separan las boletas que incluyen productos en descuento de cada categoría con las que no, se da una relación similar a observada anteriormente, donde las boletas que incluyen productos en descuento de las categorías Colchones y Línea Blanca presentan un valor del cliente diario promedio mayor que las que no, añadiéndose en este caso también la categoría Electrodomésticos.

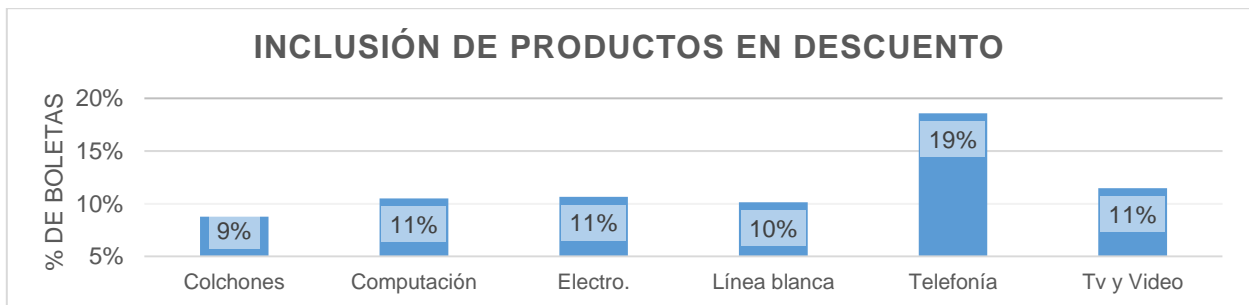


Ilustración 23: Porcentaje de boletas que incluyen productos en descuento de cada categoría
Fuente: Elaboración propia

5.3.3. Modelamiento

En este objetivo se utiliza la regresión lineal como forma funcional del modelo que busca reflejar la relación entre el tipo de cliente incorporado a la empresa, representado por su valor en el periodo estudiado, con las características de este y su primera compra. Se evalúan seis especificaciones diferentes para encontrar el grupo de variables que mejor explican la situación en análisis.

Alternativas de modelamiento

En el modelamiento se incluyen tanto variables demográficas de los clientes como variables relacionadas a su primera compra, esto con el fin de poder dimensionar qué características tienen mayor incidencia sobre el valor esperado del cliente para la empresa, y sobre cuáles de estas características se debe intentar hacer gestión.

Desde el punto de vista demográfico, se decide incorporar la edad registrada del cliente ($EDAD_i$) como una variable continua sin transformación. Se prueban otras especificaciones con esta variable transformada, tanto logarítmicamente como cuadráticamente, sin embargo, en ambos casos, el desempeño del modelo es peor que con la variable sin transformar. También se incluye la variable sexo ($s_{k,i}$) como variable categórica de tres niveles: Desconocido (nivel base), Mujer y Hombre. Por último, de las variables demográficas, se incluye el grupo socioeconómico de los clientes ($g_{k,i}$), también como variable categórica, con cinco niveles: ABC1 (nivel base), C2, C3, D y E.

Por el lado de las variables asociadas a la primera compra del cliente, al igual que para el objetivo específico anterior, de los montos asociados a la boleta sólo se incluye

el monto total pagado (M_i), para evitar correlaciones entre las variables independientes. Además, se incluyen las unidades compradas en dicha oportunidad (U_i), variable que no presenta correlación importante con el monto pagado ($r_{(M_i, U_i)} = 0,14$).

En términos temporales, se incluye el año de ingreso del cliente al retail como una variable continua que toma como año cero al 2013 ($year_i$). Esta vez no se incluye una transformación cuadrática de esta variable debido a que no mejora el desempeño del modelo. Además, se incluyen 11 variables que indican el mes de ingreso del cliente ($m_{k,i}$), tomando como mes base a enero.

De las variables asociadas a las promociones, se incluye la variable indicadora de si existió descuento o no en las categorías de interés en la primera compra (DB_i), para capturar las diferencias observadas en el análisis exploratorio. Adicionalmente, de la misma manera que en el OE2, se incluyen los grupos de variables de información de las promociones por categoría. Como la base de datos trabajada en este objetivo es una submuestra de la base de datos del objetivo anterior, el problema de las correlaciones entre los grupos de variables de promoción por categoría persiste, por lo que se prueban de manera independiente en el modelo para encontrar el conjunto de variables más adecuado para la relación estudiada.

Por lo tanto, los modelos a evaluar tienen la siguiente forma funcional:

$$VCDN_i = \beta_0 + \beta_E EDAD_i + \sum_{k \in \{SEXO\}} \beta_S s_{k,i} + \sum_{k \in \{GSE\}} \beta_G g_{k,i} + \beta_M M_i + \beta_U U_i + \beta_y year_i + \sum_{k \in \{MOY\}} m_{k,i} + \beta_D DB_i + L$$

Donde $SEXO$, GSE y MOY representan los conjuntos de las categorías de las variables de sexo, grupo socioeconómico y meses del año respectivamente. Además, L representa el grupo de variables de promociones por categoría incluidas según el modelo evaluado como se presenta a continuación. Recordar que C es el conjunto de las seis categorías de productos estudiadas.

	M1	M2	M3	M4	M5	M6
$L =$	0	$\sum_{c \in C} \beta_{c,ED} ED_{ci}$	$\sum_{c \in C} \beta_{c,PD} PD_{ci}$	$\sum_{c \in C} \beta_{c,DTPe} DTPe_{ci}$	$\sum_{c \in C} \beta_{c,DTPA} DTPA_{ci}$	$\sum_{c \in C} \beta_{c,ATPA} ATPA_{ci}$

Tabla 36: Detalle de los modelos a evaluar
Fuente: Elaboración propia

- *Modelo 1 (M1)*: No incluye variables por categoría, modelo de comparación.
- *Modelo 2 (M2)*: Incluye las seis variables de si existe descuento con tarjeta (ED_{ci}) para estudiar si la existencia de la promoción explica el valor de los clientes.
- *Modelo 3 (M3)*: Incluye las seis variables de porcentaje de descuento de los productos en descuento (PD_{ci}) para estudiar si el nivel de rebaja que obtiene un cliente es capaz de explicar el valor de este.

- *Modelo 4 (M4)*: Incluye las seis variables de descuento con tarjeta potencial escalado ($DTPE_{ci}$) para estudiar si el descuento con tarjeta al que optaba el cliente en su ocasión de compra es capaz de explicar el valor de este.
- *Modelo 5 (M5)*: Incluye las seis variables de descuento con tarjeta potencial aumentado ($DTPA_{ci}$) para estudiar si el descuento con tarjeta al que optaba el cliente en su ocasión de compra es capaz de explicar el valor de este.
- *Modelo 6 (M6)*: Incluye las seis variables de ahorro con tarjeta potencial aumentado ($ATPA_{ci}$) para estudiar si el monto de rebaja al que se optaban al utilizar la tarjeta es capaz de explicar el valor del cliente.

Comparación de modelos

A partir de los lineamientos explicados en la sección anterior, se estiman seis regresiones lineales, donde cada una incluye las variables demográficas de los clientes, las variables relacionadas a su primera compra a nivel general, y unos de los grupos de variables de promoción por categoría. Para este modelamiento se utiliza el software Stata y los resultados se pueden encontrar en la sección 8.3.2 de Anexos.

Debido a la forma funcional escogida en este objetivo, las métricas utilizadas para las comparaciones de las diferentes especificaciones son R^2 ajustado, MAE, porcentaje de variables significativas de todas las variables incluidas (Significancia general) y porcentaje de variables significativas de las variables de promociones por categoría (Sign. var. de interés). Para el cálculo de significancia se considera un nivel de confianza al 95%. En esta oportunidad no se utiliza el MAPE de las regresiones debido a que no se considera una métrica apropiada al analizar valores cercanos a cero, como es el caso de los modelos estudiados por la normalización de la variable dependiente. Los resultados de las métricas de evaluación se presentan en la Tabla 37.

<i>Métrica</i>	<i>Muestra</i>	M1	M2	M3	M4	M5	M6
R^2 ajustado	<i>Dentro</i>	15,7%	16,4%	16,1%	16,1%	16,1%	16,3%
MAE	<i>Dentro</i>	0,664	0,661	0,662	0,662	0,662	0,661
	<i>Fuera</i>	0,662	0,658	0,660	0,660	0,660	0,658
<i>Significancia general</i>	-	100%	93%	97%	97%	97%	100%
<i>Sign. var. de interés</i>	-	-	67%	100%	100%	100%	100%

Tabla 37: Métricas de desempeño de los modelos evaluados en el OE3
Fuente: Elaboración propia con uso de software stata

En términos de la varianza explicada por los modelos, reflejada en el R^2 ajustado, esta se encuentra alrededor de un 16% en todos ellos, destacando ligeramente los modelos 2 y 6. Son estos mismos dos modelos, los que muestran tener un mayor poder descriptivo y predictivo, con los menores MAE dentro y fuera de la muestra.

Los porcentajes de variables significativas de todos los modelos son altos, lo cual indica que existe una relación real entre las variables incluidas. Tanto en los modelos 1 y 6, todas sus variables son significativas al 95%. Si se miran exclusivamente las variables de las promociones por categoría, se puede ver que todos los modelos tienen todas las categorías significativas excepto el modelo 2, el cual incluye las variables indicadoras de si existe descuento en cada tipo de producto (ED_{ci}). En este caso, son Colchones y Electrodomésticos las variables que resultan no significativas.

Por lo tanto, los modelos que presentan un mejor desempeño en términos generales, según las métricas consideradas, son los modelos 2 y 6, los cuales incluyen los grupos de variables de ED_{ci} y $ATPA_{ci}$ respectivamente. De estos, el modelo 6 es el único que tiene todas sus variables de promociones por categorías significativas, por lo cual se prefiere como modelo de la situación estudiada para continuar con el análisis.

5.3.4. Resultados y análisis

Resultados del modelo seleccionado

$$VCDN_i = \beta_0 + \beta_E EDAD_i + \sum_{k \in \{SEXO\}} \beta_S s_{k,i} + \sum_{k \in \{GSE\}} \beta_G g_{k,i} + \beta_M M_i + \beta_U U_i + \beta_y year_i + \sum_{k \in \{MOY\}} m_{k,i} + \beta_D DB_i + \sum_{c \in C} \beta_{c,ATPA} ATPA_{ci}$$

El modelo seleccionado cuenta con variables demográficas y de información de la primera compra del cliente, incluyendo las variables de ahorro potencial con tarjeta aumenta ($ATPA_{ci}$) para las seis categorías de interés. Los resultados de la estimación se pueden encontrar en la Tabla 38.

Variable	$\hat{\beta}$	Variable	$\hat{\beta}$
M_i	0,15***	$SHOMBRE_i$	0,16*
U_i	0,02***	$SMUJER_i$	0,35***
$year_i$	0,31***	$g_{C2,i}$	-0,1***
$m_{Feb,i}$	0,04***	$g_{C3,i}$	-0,19***
$m_{Mar,i}$	0,06***	$g_{D,i}$	-0,24***
$m_{Abr,i}$	0,09***	$g_{E,i}$	-0,2***
$m_{May,i}$	0,12***	DB_i	0,02*
$m_{Jun,i}$	0,12***	$ATPA_{Colchones,i}$	0,34***
$m_{Jul,i}$	0,11***	$ATPA_{Computación,i}$	-0,49***
$m_{Ago,i}$	0,14***	$ATPA_{Electro.,i}$	2,06***
$m_{Sep,i}$	0,15***	$ATPA_{Línea Blanca,i}$	-0,51***
$m_{Oct,i}$	0,27***	$ATPA_{Telefonía,i}$	-0,19***
$m_{Nov,i}$	0,25***	$ATPA_{Tele.y Video,i}$	-0,57***
$m_{Dic,i}$	0,26***	Constante	-0,17*
$EDAD_i$	-0,01***	N° de observaciones	134.605
		R^2 ajustado	16%

Tabla 38: Resultados de la estimación del modelo del OE3
Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Fuente: Elaboración propia con uso de software stata

Como se mencionó anteriormente, este modelo tiene todas sus variables estadísticamente significativas y un R^2 ajustado de un 16%. Lamentablemente, aunque este nivel de varianza explicada es bajo, no se pudo encontrar alguna especificación del modelo con mejoras en esta métrica, ni en los modelos expuestos anteriormente, ni en otros modelos tanteados.

A partir de los resultados se obtiene que tanto el monto total de la primera boleta, como las unidades de productos incluidos en ella, tienen una relación positiva con el valor del cliente diario. Esto indica que clientes con primeras compras más grandes, en monto o en unidades, tienen un valor esperado diario mayor que otros clientes. Entonces, por ejemplo, una boleta \$10.000 más cara podría implicar un valor del cliente diario normalizado 0,15 mayor.

En términos temporales, se observa una relación positiva entre el año de incorporación del consumidor al retail con su valor para la empresa, mostrando que clientes más nuevos tienen un valor del cliente diario mayor. A nivel de meses, se ve una tendencia creciente entre los meses del año y el valor del cliente incorporado, teniéndose que clientes de mayor valor ingresan mayoritariamente en los últimos meses del año, siendo octubre el mes de incorporación de clientes más rentables.

En términos demográficos, la edad muestra tener una relación negativa con el valor del cliente, implicando que clientes más jóvenes que abren la tarjeta propia serían más rentables para el retail. Además, al igual como se vio en el análisis exploratorio, el grupo socioeconómico ABC1 y el género femenino serían los que tienen clientes de mayor calidad para la empresa.

La existencia de descuento en algún producto de las categorías de interés se presenta relacionado con clientes más rentables, al igual que el ahorro potencial en las categorías de Colchones y Electrodomésticos. Sin embargo, a partir de los resultados del modelo, se puede concluir que mayores ahorros potenciales con tarjeta en las categorías Computación, Línea Blanca, Telefonía y Televisión y Video se presentan en las boletas de clientes con menor valor promedio para la empresa.

Impacto de las variables

Al ser el modelo trabajado en este objetivo una regresión lineal, el impacto de las variables independientes sobre la métrica estudiada, valor del cliente diario normalizado, se puede obtener directamente del valor de los coeficientes estimados. Sin embargo, para el caso de las variables de ahorro con tarjeta potencial aumentado, el impacto también depende del nivel promedio de este ahorro según el tipo de producto, dado que no es lo mismo variar cierta cantidad de ahorro potencial en cada una de las categorías.

Variable aumentada	$\hat{\beta}$	Valor base (ATPA)	V. Aumentado (ATPA * 110%)	Variación del VCDN (* 10⁻⁴)	Variación del VCD (* 10⁻⁴)
Colchones	0,339	0,026	0,028	12,43	0,15
Computación	-0,489	0,015	0,016	-4,98	-0,06
Electro.	2,055	0,002	0,003	5,35	0,07
Línea Blanca	-0,506	0,015	0,017	-5,43	-0,07
Telefonía	-0,187	0,025	0,027	-0,97	-0,01
Tv y video	-0,570	0,031	0,034	-12,99	-0,16

Tabla 39: Valores utilizados y resultados a nivel de cliente
Fuente: Elaboración propia

Para poder dimensionar el impacto que tienen estas variables por categoría, se considera como nivel de referencia su promedio histórico, y se analiza qué pasa con el

$VCDN_i$ cuando cada uno de estos valores aumenta un 10% de manera independiente (*ceteris paribus*). Al igual que para el OE2, cuando se varían los ahorros, se hacen los ajustes correspondientes sobre el valor de la variable monto de la boleta (MB_i). Como el cambio de $VCDN_i$ no depende del nivel de las demás variables, no se escogen niveles de referencia específicos para ellas.

Para poder hacer más interpretable los resultados, se desnormaliza $VCDN_i$, utilizando la misma media y desviación estándar utilizadas anteriormente para la normalización. Con esto se obtiene, por ejemplo, que el valor del cliente diario se ve incrementado en un 0,14% si se aumenta en un 10% el ahorro potencial con tarjeta en la categoría Colchones, siendo el mayor cambio en la rentabilidad de los clientes ante una variación en el ahorro potencial en alguna de las categorías. En la Tabla 40 se puede observar el impacto generado por los otros tipos de productos.

	Colchones	Electro.	Telefonía	Compu.	L. Blanca	Tv y video
Variación de VCD	0,14%	0,06%	-0,01%	-0,06%	-0,06%	-0,15%

Tabla 40: Impacto de la variación de los ahorros potenciales en el valor del cliente diario promedio

Fuente: Elaboración propia

Además, como una forma de poder estimar el impacto total sobre la empresa, se utiliza la vida media de un cliente (750 días) para transformar el valor del cliente diario estimado (VCD_i) a valor del cliente total estimado en el periodo estudiado (VC_i), para luego multiplicarlo por el número total de clientes de la base de datos, y obtener el margen potencialmente ganado (o perdido) ante una variación de un 10% en el ahorro potencial pagando con tarjeta por categoría (Ilustración 24).

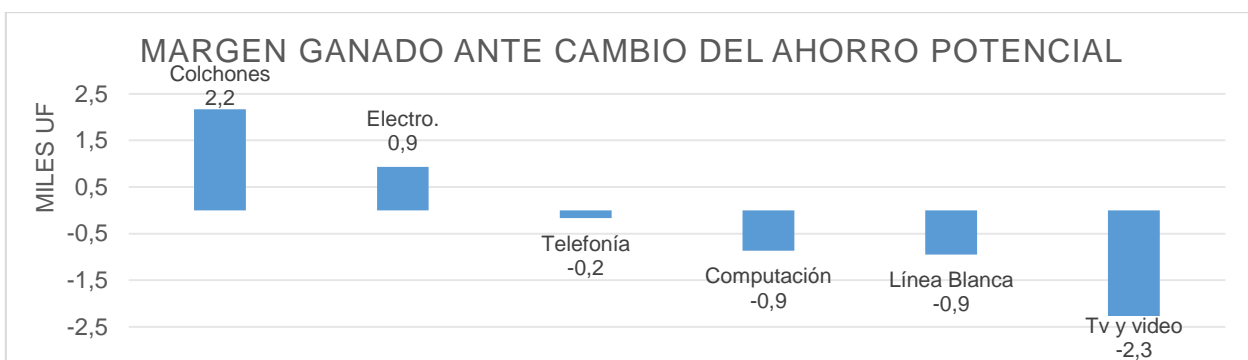


Ilustración 24: Impacto de la variación de los ahorros potenciales en el periodo estudiado

Fuente: Elaboración propia

Como se puede observar, a pesar de ser Electrodomésticos el que presenta un mayor coeficiente estimado, es Colchones quien podría generar un impacto mayor en la empresa en términos de margen acumulado ante una variación de un 10% del ahorro potencial en la categoría. Esto se debe a los niveles de ahorro promedio en cada tipo de producto. El orden de las demás categorías se condice con los valores obtenidos para los betas.

Es interesante notar que el impacto que tiene la variable $ATPA$ de las categorías Colchones y TV y Video tiene aproximadamente la misma magnitud, pero en sentidos

opuestos, pudiendo provocar resultados muy distintos en las ganancias del retail si se decide variar uno u otro descuento.

Por último, dado que la función escogida para el modelamiento es lineal, las utilidades se comportan monótonamente ante la variación en los descuentos de las categorías, haciendo imposible encontrar un máximo (o mínimo) en la rentabilidad de las promociones, y su consiguiente descuento óptimo para el retail. En consecuencia, en esta oportunidad no se desarrollará un análisis de escenarios, es decir, no se generaran recomendaciones específicas de los valores que deberían tomar las promociones (OE4).

5.3.5. Conclusiones del Objetivo Específico 3

Este tercer objetivo específico se concentró en los clientes que comienzan su vida en el retail obteniendo la tarjeta propia del negocio, y se analizó la relación entre los descuento en su primera compra, con el valor que generan estos individuos para la empresa a lo largo del periodo estudiado. De manera adicional, se evidenció el nexo entre dicho valor de los clientes con otras variables disponibles, relacionadas a su primera boleta y a características demográficas de los individuos.

Tanto los resultados del modelo, como el análisis exploratorio, muestran relaciones significativas entre las características demográficas de los clientes y su valor para la empresa. Como era de esperarse, las personas del grupo socioeconómico ABC1 son los más rentables en términos de margen acumulado, sin embargo, llama la atención que el GSE E presenta un valor del cliente diario promedio, y un beta estimado, mayor que el D. Se tiene, también, que las mujeres y los individuos más jóvenes representan mejores clientes. Estos resultados son importantes dado que indicarían a qué conjuntos de clientes es recomendable focalizar esfuerzos de atracción.

Por otro lado, los resultados sugieren que los clientes más valiosos presentan primeras compras más grandes, tanto en términos de monto como en unidades, lo que podría ser un indicio temprano para tomar decisiones relacionadas en qué individuos concentrar esfuerzos de fidelización y preventivos de fuga.

Con respecto a las promociones ofrecidas en la primera compra, se observa que el simple hecho de que exista algún descuento disponible podría atraer a mejores clientes. Sin embargo, es importante tener cuidado con respecto a donde se aplican estas promociones, pudiendo llegar a traer consecuencias completamente contrarias a las esperadas, si los descuentos son ofrecidos en algunos tipos de productos en específico.

A partir del análisis, se concluye que tanto Colchones como Electrodomésticos son tipos de productos interesantes para ofrecer promociones, si se quiere capturar a clientes de alto valor. Por el contrario, las otras categorías estudiadas muestran tener un efecto contraproducente en la rentabilidad de los clientes atraídos, a medida que crece la profundidad de los descuentos, por lo que, según los resultados obtenidos, se debería evitar hacer promociones en estas categorías, especialmente en Televisión y Video.

Vale la pena mencionar que se debe tener cuidado con la interpretación superficial de los resultados obtenidos. En una primera exploración de los datos con los que se trabaja, se ve que el número de nuevos clientes poseedores de tarjeta ha ido

disminuyendo con el paso de los años, en el periodo estudiado. Es importante no confundir este resultado con que el uso de tarjeta este disminuyendo, dado que la base de datos solo considera la primera compra en la que el cliente es identificado, y no se puede concluir del uso de tarjeta como medio de pago en términos generales.

Finalmente, es importante destacar que, en el análisis realizado, no se consideran los beneficios que los nuevos clientes podrían generar para la empresa en otros negocios del holding, como, por ejemplo, en el negocio financiero. La inclusión de estos ingresos en el cálculo del valor de los clientes podría cambiar los resultados de las categorías en las que es conveniente aplicar descuentos.

6. CONCLUSIONES

La utilización de promociones focalizadas al medio de pago utilizado, como una herramienta para maximizar las utilidades de un negocio, motiva la realización de esta investigación, especialmente ante el hecho de que no existe literatura al respecto, y no se sabe si estos esfuerzos conllevan un beneficio real para la empresa. Para el caso del retail estudiado, se pudo evidenciar que efectivamente existe una práctica generalizada de ofrecer descuentos más profundos si se utiliza el medio preferente de pago, con rebajas promedio mayores que en compras con otro medio de pago.

La utilización de las promociones exclusivas puede estar motivado por diferentes objetivos comerciales, específicamente aquí se abordaron 3 posibles metas, potenciar las unidades vendidas con tarjeta propia, promover la apertura de tarjeta del retail en clientes nuevos y capturar clientes que generen mayor valor para la empresa.

Con respecto al objetivo de maximizar las unidades vendidas con tarjeta propia, se concluye que la amplitud del descuento dentro de un tipo de productos tiene siempre un efecto positivo, sin embargo, esto no se replica en la profundidad de las promociones y depende de la sublínea donde se apliquen. En la categoría Colchones, los descuentos promedios ofrecidos en la realidad resultaron ser siempre superiores a los descuentos óptimos encontrados, lo que podría estar generando pérdidas de margen para la empresa importantes. Una aproximación realizada estima que está pérdida de margen estaría alrededor de \$1.250MM anuales.

Si lo que se busca es maximizar el número de personas que comienzan su relación con el retail obteniendo la tarjeta del holding, las promociones muestran tener un efecto siempre positivo, independiente de la categoría donde se apliquen. No obstante, si se consideran las utilidades a largo plazo de obtener un cliente con tarjeta, en comparación a un cliente que no abre cuenta bancaria con la empresa, existen categorías donde los descuentos son más efectivos que en otras, y los niveles de descuento óptimos parecen ser, nuevamente, menores a los aplicados históricamente por el retail estudiado. Se estima que esto estaría generando pérdidas para la empresa de \$2MMM en el periodo de los 3,5 años estudiados, es decir, aproximadamente \$570MM anuales.

Por último, si lo que se desea es maximizar el valor del cliente capturado, los resultados sugieren que el solo hecho de existir algún descuento en la primera compra tiene efectos positivos, sin embargo, cuando se analiza por categoría, se ve que no es conveniente realizar promociones en cualquiera de ellas. La profundidad de los descuentos aplicados en Colchones y Electrodomésticos tiene un efecto positivo en la rentabilidad de los clientes a largo plazo, no así en las otras categorías, donde las promociones muestran tener efectos contraproducentes.

Por lo tanto, se destacan 2 conclusiones del trabajo realizado. Primero, no se puede establecer una regla general de recomendación de las promociones, y éstas dependerán del objetivo buscado por el negocio. En segundo lugar, se tiene que las promociones observadas en la realidad son consistentemente mayores a los resultados de los niveles de descuento óptimo, lo que sugiere que la empresa estudiada está actualmente perdiendo márgenes importantes de utilidades en el negocio del retail.

Es importante destacar que el análisis desarrollado en esta investigación no es del tipo causal, por lo que las relaciones identificadas deben ser interpretadas con precaución, sin poder asegurar una dirección del vínculo entre las variables. Para poder asegurar causalidad entre las promociones y los resultados estudiados, es necesario realizar experimentos aleatorios sobre las características de los descuentos, lo que no estaba entre los posibles desarrollos, debido a la naturaleza de esta investigación.

Los resultados obtenidos en esta memoria son válidos para el retail estudiado, en el periodo de información disponible. La validez externa de las conclusiones dependerá de que las características del retail al que se le quieran aplicar sean parecidas a las expuestas, y que el comportamiento de los clientes no cambien drásticamente en el tiempo.

6.1. Trabajo futuro

Las conclusiones alcanzadas son válidas para el negocio considerando solo los beneficios del retail, y no se incluyen otros beneficios, a los que pueda optar el holding en alguno de sus otros negocios. En investigaciones futuras, se recomienda incluir esta información, de estar disponible, para poder entender y evaluar de manera más completa las decisiones de la empresa. Al incluir esta información se podrá calcular si los beneficios de las ventas con tarjeta propia por parte del negocio financiero son capaces de cubrir las pérdidas estimadas producidas por la alta profundidad de las promociones efectuadas.

También, se debe tener en consideración que dentro del análisis realizado podría existir un problema de endogeneidad, debido a que los niveles de descuento podrían ser establecidos como una reacción al volumen de ventas observado o a la apertura de tarjetas propias. Queda propuesto para trabajos futuros el testeo de este problema y la modificación de los modelos de ser necesario, utilizando, por ejemplo, variables instrumentales.

Por otro lado, queda propuesto la utilización de interacciones de variables en los modelos generados, para poder capturar efectos cruzados, como, por ejemplo, el efecto de las promociones en los diferentes días de la semana, en especial en modelos como los desarrollados en el objetivo específico 1, donde se evidencio que los resultados son sensibles al día de evaluación.

Finalmente, el análisis realizado se puede llevar a cabo en un nivel más desagregado en el tipo de productos, para llegar a recomendaciones de descuentos más precisas, pudiendo distinguir, por ejemplo, entre marcas de productos. Para esto, es necesario contar con información más precisa de los precios reales ofrecidos, cada día, con cada medio de pago, sin tener que realizar las imputaciones llevadas a cabo.

7. BIBLIOGRAFÍA

- Abraham, M., & Lodish, L. (1993). An Implemented System for Improving Promotion Productivity Using Store Scanner Data. *Marketing Science*, 12(3), 248-269.
- Acquisti, A., & Varian, H. (2005). Conditioning Prices on Purchase History. *Marketing Science*, 367–81.
- Ailawadi, K., & Neslin, S. (1998). The effect of promotion on consumption: Buying more and consuming it faster. *Journal of Marketing Research*, 390-398.
- Ailawadi, K., Harlem, B., César, J., & Trounce, D. (2006). Promotion profitability for a retailer: the role of promotion, brand, category, and store characteristics. *Journal of Marketing Research*, 43(4), 518-535.
- Barnett, V., & Lewis, T. (1974). *Outliers in Statistical Data* (3 ed., Vol. 37). Chichester: J. Wiley & Sons.
- Barone, M., & Roy, T. (2010). Does Exclusivity Always Pay Off? Exclusive Price Promotions and Consumer Response. *Journal of Marketing*, 74(2), 121-132.
- Blattberg, R., Briesch, R., & Fox, E. (1995). How promotions work. *Marketing science*, 14° 3_ supplement, G122-G132.
- Bolt, W., & Chakravorti, S. (2008). *Economics of Payment Cards: A Status Report*. Amsterdam: De Nederlandsche Bank NV.
- CENCOSUD. (2016). *Memoria Anual*. Santiago, Chile: Cencosud.
- Chatterjee, P., & Rose, R. (2012). Do Payment Mechanisms Change the Way Consumers Perceive Products? *Journal of Consumer Research*, 38(6), 1129–1139.
- Cheng, C.-H., & Chen, Y.-S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, 36(3), 4176-4184.
- DelVecchio, D., Henard, D., & Freling, T. (2006). The effect of sales promotion on post-promotion brand preference: A meta-analysis. *Journal of retailing*, 82(3), 203-213.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *COMMUNICATIONS OF THE ACM*, Vol. 39, No. 11. 27-34.
- Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Gupta, S. (1988). Impact of Sales Promotions on When, What, and How Much to Buy. *Journal of Marketing Research*, 25, No. 4, 342-355.
- Hughes, A. (1994). Strategic database marketing: the masterplan for starting and managing a profitable. Customer-based Marketing Program, Irwin Professional.
- Jain, D., & Singh, S. (2002). Customer Lifetime Value Research in Marketing: a Review and Future Directions. *Journal of Interactive Marketing*, 16(2), 34 - 46.
- Kotler, P., & Armstrong, G. (2012). *Marketing* (14° ed.). Ciudad de México, México: PEARSON EDUCACIÓN.
- Lehmann, D., & Winer, R. (2004). *Administración del producto* (4ta ed.). México D.F.: McGraw-Hill Interamericana Editores.
- Mela, C., Gupta, S., & Lehmann, D. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing research*, 248-261.
- Muñoz Fernández, J. F. (01 de Marzo de 2015). *TRES MÉTODOS (CON EJEMPLOS EN R) PARA IDENTIFICAR VALORES ATÍPICOS EN CONJUNTOS DE DATOS UNIVARIANTES*. Obtenido de SCRIBD: <https://www.scribd.com/document/257435992/Deteccion-de-valores-atipicos->

outliers-en-conjuntos-de-datos-univariantes-Outliers-values-detection-in-univariable-datasets

- Neslin, S., Narasimhan, C., & Sen, S. (1996). Promotional Elasticities and Category Characteristics. *Journal of Marketing*, 60, No. 2, 17-30.
- Newell, F. (1997). *The new rules of marketing: How to use one-to-one relationship marketing to be the leader in your industry*. New York: McGraw-Hills Companies Inc.
- Nijs, V., Dekimpe, M., Steenkamps, J.-B., & Hanssens, D. (2001). The Category-Demand Effects of Price Promotions. *Marketing science*, 20(1), 1-22.
- Peirce, R., Rhoads, B., Duncan, J., Loftness, S., Johnson, K., Dobbs, M., & Mayes, D. (2001). *EE.UU. Patente nº US 6,332,126 B1*.
- Reinartz, W., & Kumar, V. (2000). On the Profitability of Long Lifetime Customers: An Empirical Investigation and Implications for Marketing. *Journal of Marketing*(64), 17–35.
- Servicio de Impuestos Internos. (Mayo de 2018). *UF*. Obtenido de Servicio de Impuestos Internos: http://www.sii.cl/valores_y_fechas/uf/uf2018.htm
- Stone, B. (1975). *Successful direct marketing methods*. Crain Books.
- Unidad de Estudios Ministerio de Economía, Fomento y Turismo. (2014). *Las empresas en Chile por tamaño y sector económico desde el 2005 a la fecha*. Santiago: Ministerio de Economía, Fomento y Turismo.
- van Heerde, H., Gupta, S., & Wittin, D. (2003). Is 75% of the Sales Promotion Bump Due to Brand Switching? No, Only 33% Is. *Journal of Marketing Research*, 40(4), Journal of Marketing Research.
- Vogel, J., & Paul, M. (2015). One firm, one product, two prices: Channel-based price differentiation and customer retention. *Journal of Retailing and Consumer Services*, 26, 126-139.
- Wu, J., & Lin, Z. (2005). Research on customer segmentation model by clustering. *ACM International Conference Proceeding Series*, 113.

8. ANEXOS

8.1. ANEXO A. Información adicional Objetivo Específico 1

8.1.1. Preparación de la base de datos

Detección de outliers e imputación

Outliers detectados por sublínea, utilizando el método de 3 sigmas

Sublíneas	% de outliers en Q_{ts}^O	% de outliers en Q_{ts}^T
Almohadas	1,70%	1,70%
Boxspring	1,82%	1,48%
Cama	1,82%	1,36%
Cama Americana	2,16%	1,70%
Closet/Roperos	1,48%	2,27%
Colch. 1 1/2 Pza.	1,25%	1,59%
Colch. 1 Pza.	1,25%	1,93%
Colch. 2 Pza.	1,59%	1,93%
Colch. Medida Especial	1,48%	1,82%
Cómoda	1,25%	1,36%
Paq. Completos Boxspring	1,36%	2,05%
Paq. Completos Cama Americana	2,27%	2,05%
Paq. Textiles Box Spring	2,16%	2,05%
Protectores De Colchón	1,82%	1,48%
Respaldo De Cama	1,14%	1,82%
Veladores	1,02%	2,39%
General	1,60%	1,81%

Tabla 41: Porcentajes de outliers encontrados en cada sublínea

Fuente: Elaboración propia

Impacto de la imputación de outliers en la base de datos

Sublíneas	ΔQ^O real y Q^O imputado	ΔQ^T real y Q^T imputado
Almohadas	-1,90%	-2,44%
Boxspring	-4,61%	-3,00%
Cama	-7,86%	-3,72%
Cama Americana	-5,84%	-3,12%
Closet/Roperos	-5,34%	-3,94%
Colch. 1 1/2 Pza.	-4,17%	-1,75%
Colch. 1 Pza.	-7,16%	-5,04%
Colch. 2 Pza.	-2,47%	-2,00%
Colch. Medida Especial	-7,20%	-5,69%
Cómoda	-3,92%	-1,83%
Paq. Compl. Boxspring	-2,42%	-2,17%
Paq. Compl. Cama Americana	-5,30%	-2,64%
Paq. Textiles Box Spring	-7,15%	-5,45%
Protectores De Colchón	-2,04%	-1,71%
Respaldo De Cama	-4,78%	-4,63%
Veladores	-5,36%	-5,20%
General	-3,16%	-2,62%

Tabla 42: Diferencia provocadas por la imputación de outliers

Fuente: Elaboración propia

8.1.2. Ventas por sublínea

Análisis según sublínea

Unidades totales vendidas por sublínea (Q_{ts})

Sublíneas	Promedio	SD	CV
Almohadas	127,3	66,6	0,5
Boxspring	22,1	13,5	0,6
Cama	8,5	5,6	0,7
Cama Americana	19,1	12,2	0,6
Closet/Roperos	13,2	9,0	0,7
Colch. 1 1/2 Pza.	27,3	14,4	0,5
Colch. 1 Pza.	16,9	11,2	0,7
Colch. 2 Pza.	41,7	19,3	0,5
Colch. Medida Especial	4,0	3,3	0,8
Cómoda	12,6	6,9	0,5
Paq. Completos Boxspring	65,2	31,3	0,5
Paq. Completos Cama Americana	42,1	24,4	0,6
Paq. Textiles Box Spring	6,4	5,7	0,9
Protectores De Colchón	60,4	29,8	0,5
Respaldo De Cama	4,5	3,0	0,7
Veladores	10,4	8,2	0,8
Total	30,1	38,6	1,3

Tabla 43: Estadísticos básicos de las unidades totales vendidas diariamente por sublínea
Fuente: Elaboración propia

Precios promedio según la sublínea

Sublíneas	$\overline{P_{ts}}$	$\overline{P_{ts}^{DT}}$	$\overline{P_{ts}^{DO}}$
Almohadas	\$22.413	\$16.496	\$13.565
Boxspring	\$235.663	\$201.316	\$148.404
Cama	\$235.422	\$180.741	\$121.448
Cama Americana	\$177.421	\$119.119	\$88.825
Closet/Roperos	\$126.893	\$108.630	\$79.520
Colch. 1 1/2 Pza.	\$98.714	\$84.432	\$66.503
Colch. 1 Pza.	\$71.612	\$60.256	\$48.958
Colch. 2 Pza.	\$171.858	\$142.440	\$108.130
Colch. Medida Especial	\$242.800	\$191.859	\$123.303
Cómoda	\$233.070	\$186.867	\$139.977
Paq. Completos Boxspring	\$392.390	\$317.869	\$208.165
Paq. Completos Cama Americana	\$185.947	\$155.992	\$92.033
Paq. Textiles Box Spring	\$280.985	\$233.495	\$104.470
Protectores De Colchón	\$17.222	\$13.070	\$11.447
Respaldo De Cama	\$149.240	\$122.597	\$64.759
Veladores	\$129.713	\$89.682	\$72.607
General	\$173.210	\$139.054	\$93.257

Tabla 44: Precio unitario promedio por sublínea
Fuente: Elaboración propia

Análisis temporal

Evolución de los ingresos diarios de la categoría

	Promedio	SD	CV
2014	\$4.705.151	\$7.216.934	1,5

2015	\$3.100.909	\$4.673.121	1,5
2016	\$2.976.611	\$4.981.134	1,7
Total	\$3.732.202	\$5.948.195	1,6

Tabla 45: Ingresos totales promedio diario histórico
Fuente: Elaboración propia

Análisis de las variables de interés

Unidades vendidas con tarjeta (Q_{ts}^T)



Ilustración 25: Histograma de unidades vendidas con tarjeta
Fuente: Elaboración propia

Sublíneas	Suma	Promedio	SD	CV
Almohadas	76.057	86	48	0,56
Boxspring	14.459	16	10	0,58
Cama	5.551	6	4	0,65
Cama Americana	12.110	14	9	0,62
Closet/Roperos	9.445	11	7	0,64
Colch. 1 1/2 Pza.	18.508	21	11	0,52
Colch. 1 Pza.	10.231	12	7	0,64
Colch. 2 Pza.	29.373	33	16	0,47
Colch. Medida Especial	2.418	3	2	0,80
Cómoda	9.074	10	6	0,54
Paq. Compl. Boxspring	46.231	53	24	0,46
Paq. Compl. Cama Americana	29.707	34	17	0,51
Paq. Textiles Box Spring	3.977	5	4	0,89
Protectores De Colchón	29.249	33	16	0,49
Respaldo De Cama	2.389	3	2	0,75
Veladores	7.429	8	6	0,73

Tabla 46: Estadísticos de unidades vendidas con tarjeta por sublínea
Fuente: Elaboración propia

Porcentaje de unidades vendidas con tarjeta (QTQ_{ts})

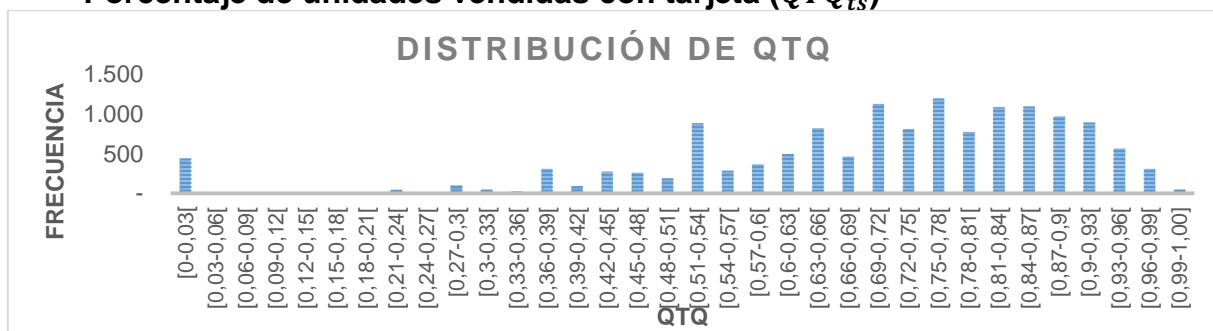


Ilustración 26: Histograma del porcentaje de unidades vendidas con tarjeta sobre las unidades totales
Fuente: Elaboración propia

Sublíneas	Promedio	SD	CV
Almohadas	0,67	0,12	0,18
Boxspring	0,72	0,14	0,20
Cama	0,66	0,20	0,31
Cama Americana	0,69	0,16	0,23
Closet/Roperos	0,74	0,16	0,21
Colch. 1 1/2 Pza.	0,75	0,12	0,16
Colch. 1 Pza.	0,66	0,18	0,27
Colch. 2 Pza.	0,78	0,11	0,13
Colch. Medida Especial	0,51	0,25	0,49
Cómoda	0,75	0,15	0,20
Paq. Completos Boxspring	0,80	0,09	0,12
Paq. Completos Cama Americana	0,80	0,10	0,12
Paq. Textiles Box Spring	0,55	0,26	0,48
Protectores De Colchón	0,55	0,11	0,19
Respaldo De Cama	0,47	0,24	0,52
Veladores	0,69	0,26	0,37

Tabla 47: Estadísticos del porcentaje de unidades vendidas con tarjeta por sublínea
Fuente: Elaboración propia

Amplitud promedio del descuento por sublínea ($Ln(EDesc_{ts})$)

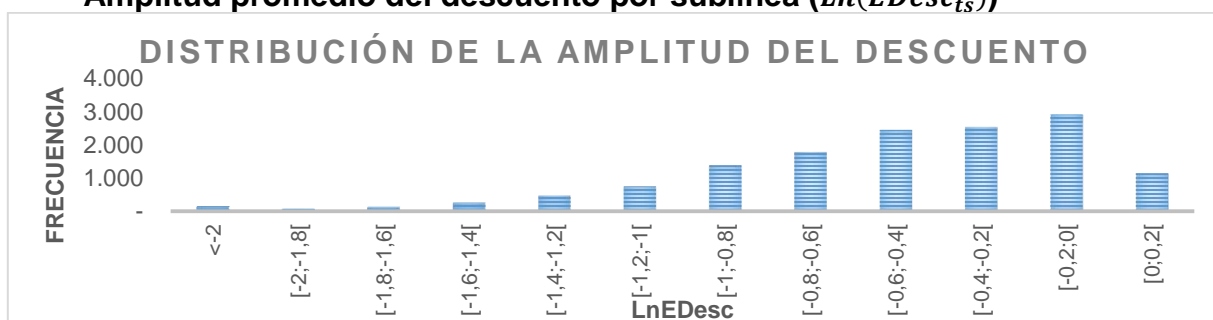


Ilustración 27: Histograma de la variable LnEDesc
Fuente: Elaboración propia

Sublíneas	Promedio	SD	CD
Almohadas	-0,54	0,493	-0,91
Boxspring	-0,62	0,294	-0,48
Cama	-0,66	0,350	-0,53
Cama Americana	-0,52	0,498	-0,96
Closet/Roperos	-0,80	0,637	-0,80
Colch. 1 1/2 Pza.	-0,78	0,477	-0,61
Colch. 1 Pza.	-0,72	0,494	-0,69
Colch. 2 Pza.	-0,51	0,381	-0,74
Colch. Medida Especial	-0,89	0,279	-0,31
Cómoda	-0,57	0,380	-0,67
Paq. Completos Boxspring	-0,62	0,222	-0,36
Paq. Completos Cama Americana	-0,77	0,207	-0,27
Paq. Textiles Box Spring	-0,94	0,193	-0,21
Protectores De Colchón	-1,04	0,593	-0,57
Respaldo De Cama	-1,33	0,437	-0,33
Veladores	-0,23	0,201	-0,87

Tabla 48: Estadísticos de LnEDesc
Fuente: Elaboración propia

Descuento con tarjeta (DT_{ts})

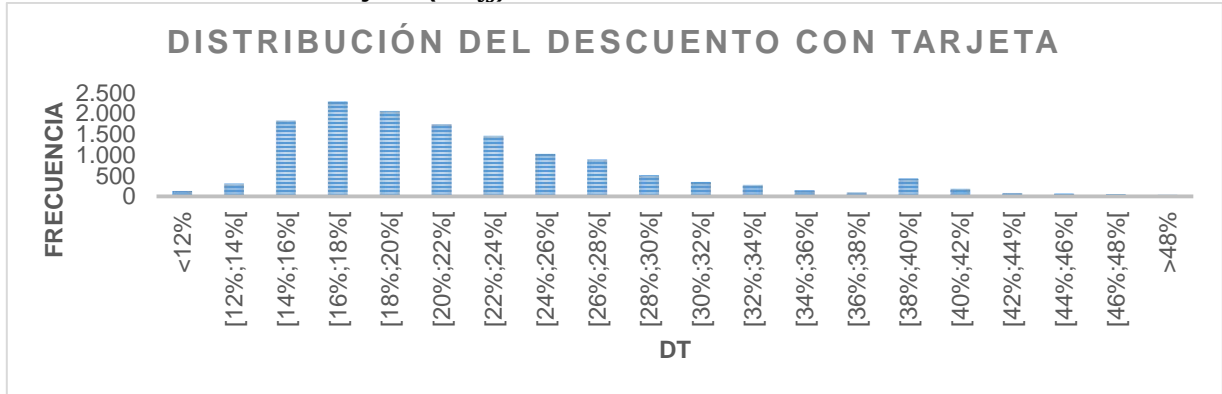


Ilustración 28: Histograma de la variable DT
Fuente: Elaboración propia

Sublíneas	Promedio	SD	CV
Almohadas	26%	6%	0,21
Boxspring	15%	2%	0,12
Cama	23%	4%	0,15
Cama Americana	33%	8%	0,24
Closet/Roperos	14%	4%	0,31
Colch. 1 1/2 Pza.	14%	2%	0,17
Colch. 1 Pza.	16%	3%	0,19
Colch. 2 Pza.	17%	3%	0,18
Colch. Medida Especial	21%	5%	0,22
Cómoda	20%	4%	0,22
Paq. Completos Boxspring	19%	3%	0,15
Paq. Completos Cama Americana	16%	3%	0,17
Paq. Textiles Box Spring	17%	2%	0,13
Protectores De Colchón	24%	6%	0,26
Respaldo De Cama	18%	3%	0,15
Veladores	31%	9%	0,29

Tabla 49: Estadísticos de DT
Fuente: Elaboración propia

Precio con descuento con tarjeta ($Ln(PDT)_{ts}$)

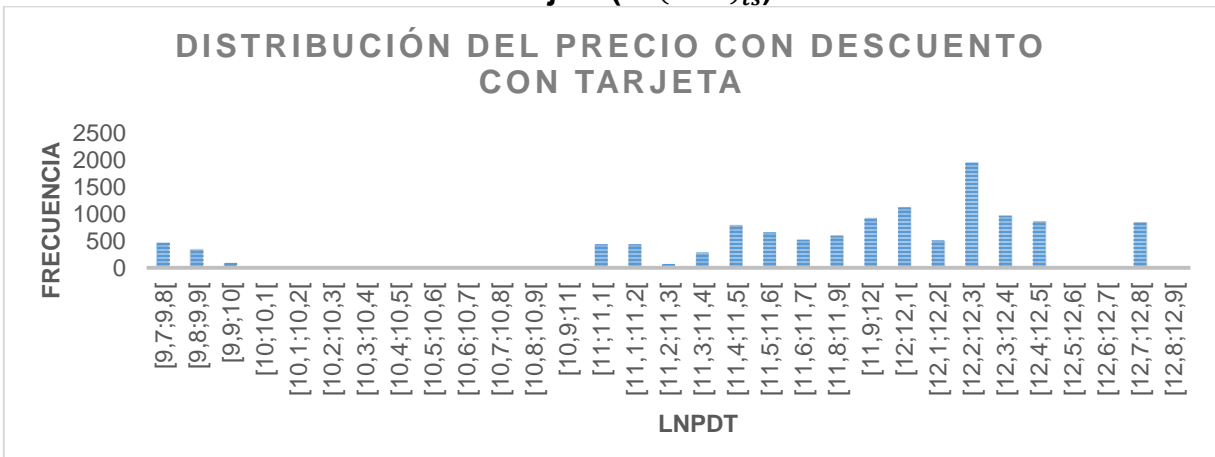


Ilustración 29: Histograma de la variable LnPDT
Fuente: Elaboración propia

Sublíneas	Promedio	DS	CV
Almohadas	9,7	0,07	0,01
Boxspring	12,2	0,04	0,00
Cama	12,1	0,05	0,00
Cama Americana	11,7	0,10	0,01
Closet/Roperos	11,6	0,08	0,01
Colch. 1 1/2 Pza.	11,3	0,05	0,00
Colch. 1 Pza.	11,0	0,04	0,00
Colch. 2 Pza.	11,9	0,04	0,00
Colch. Medida Especial	12,2	0,05	0,00
Cómoda	12,1	0,09	0,01
Paq. Completos Boxspring	12,7	0,03	0,00
Paq. Completos Cama Americana	12,0	0,02	0,00
Paq. Textiles Box Spring	12,4	0,03	0,00
Protectores De Colchón	9,5	0,09	0,01
Respaldo De Cama	11,7	0,05	0,00
Veladores	11,4	0,12	0,01

Tabla 50: Estadísticos de LnPDT

Fuente: Elaboración propia

8.1.3. Modelamiento

Comparación de modelos

Métricas de desempeño para los modelos 3A y 3B

	Muestra	Modelo 1A			Modelo 2A			Modelo 3A		
		R ^{2*}	MAE	MAPE	R ^{2*}	MAE	MAPE	R ^{2*}	MAE	MAPE
Reg.	Dentro	40,8%	13,9	41,5%	41,8%	13,8	41,1%	47,4%	13,0	38,5%
Lineal	Fuera		15,2	47,4%		17,9	57,0%		16,3	54,7%
Reg.	Dentro	27,5%	13,5	40,4%	28,2%	13,4	40,0%	32,1%	12,4	36,8%
Poisson	Fuera		13,8	45,4%		15,1	50,6%		13,2	43,0%

Tabla 51: Métricas de comparación completas ponderadas de los modelos tipo A

Fuente: Elaboración propia con uso de software stata

	Muestra	Modelo 1B			Modelo 2B			Modelo 3B		
		R ^{2*}	MAE	MAPE	R ^{2*}	MAE	MAPE	R ^{2*}	MAE	MAPE
Reg.	Dentro	9,0%	0,1	14,3%	11,0%	0,1	14,1%	20,5%	0,1	13,2%
Lineal	Fuera		0,1	22,3%		0,1	21,8%		0,1	19,1%
Reg.	Dentro	5,6%	5,0	14,5%	6,5%	5,0	14,3%	10,8%	4,6	13,4%
Binomial	Fuera		5,9	22,4%		6,2	22,2%		5,4	19,4%
Reg.	Dentro	0,8%	0,1	14,3%	0,9%	0,1	14,1%	1,5%	0,1	13,2%
Logística	Fuera		0,1	22,1%		0,1	21,9%		0,1	19,2%

Tabla 52: Métricas de comparación completas ponderadas de los modelos tipo B

Fuente: Elaboración propia con uso de software stata

Análisis de los parámetros obtenidos

Información de los coeficientes estimados:

Tipo de regresión	Variable	Significancia	Signo correcto	Máximo	Mediana	Mínimo	β_{cons}	$\frac{\beta_i^{Mediana}}{\beta_{cons}}$
Lineal	Ln(PDT)	66%	96%	2,089,3	44,2	-80,3	-404,9	-4%
Poisson	Ln(PDT)	92%	97%	37,6	3,7	-1,9	-10,0	-5%

Tabla 53: Análisis de los betas para el M1A

Fuente: Elaboración propia con uso de software stata

Tipo de regresión	Variable	Significancia	Signo correcto	Máximo	Mediana	Mínimo	β_{cons}	$\frac{\beta_i^{Mediana}}{\beta_{cons}}$
Lineal	Ln(PDT)	74%	43%	210,0	-19,3	-684,8	1.384,1	-1%
	DT	62%	77%	317,9	48,4	-306,3		4%
Poisson	Ln(PDT)	76%	42%	5,0	-1,3	-13,2	35,3	-4%
	DT	77%	64%	7,9	2,8	-15,7		8%

Tabla 54: Análisis de los betas para el M2A
Fuente: Elaboración propia con uso de software stata

Tipo de regresión	Variable	Significancia	Signo correcto	Máximo	Mediana	Mínimo	β_{cons}	$\frac{\beta_i^{Mediana}}{\beta_{cons}}$
Regresión Lineal	Ln(PDT)	63%	50%	158,4	-26,4	-759,3	2.189,3	-1%
	DT	69%	66%	351,8	40,0	-502,0		2%
	Ln(EDesc)	97%	100%	29,5	6,1	0,6		0%
Regresión Poisson	Ln(PDT)	86%	51%	4,5	-0,8	-14,7	53,2	-1%
	DT	75%	63%	4,0	3,2	-13,5		6%
	Ln(EDesc)	97%	100%	1,5	0,5	0,2		1%

Tabla 55: Análisis de los betas para el M3A
Fuente: Elaboración propia con uso de software stata

Tipo de regresión	Variable	Significancia	Signo correcto	Máximo	Mediana	Mínimo	β_{cons}	$\frac{\beta_i^{Mediana}}{\beta_{cons}}$
Regresión Lineal	Ln(PDT)	70%	90%	0,6	-0,3	-1,7	7,7	-4%
Regresión Binomial	Ln(PDT)	72%	93%	2,6	-2,5	-10,5	46,7	-5%
Regresión Logística	Ln(PDT)	70%	90%	3,0	-1,3	-10,3	43,9	-3%

Tabla 56: Análisis de los betas para el M1B
Fuente: Elaboración propia con uso de software stata

Tipo de regresión	Variable	Significancia	Signo correcto	Máximo	Mediana	Mínimo	β_{cons}	$\frac{\beta_i^{Mediana}}{\beta_{cons}}$
Regresión Lineal	Ln(PDT)	81%	32%	1,9	0,5	-1,7	-6,1	-8%
	DT	80%	78%	3,2	1,2	-3,5		-20%
Regresión Binomial	Ln(PDT)	80%	32%	12,9	2,8	-8,7	-35,3	-8%
	DT	80%	78%	20,3	6,3	-11,7		-18%
Regresión Logística	Ln(PDT)	81%	32%	11,5	2,5	-8,0	-33,1	-8%
	DT	80%	78%	19,4	6,3	-14,3		-19%

Tabla 57: Análisis de los betas para el M2B
Fuente: Elaboración propia con uso de software stata

Tipo de regresión	Variable	Significancia	Signo correcto	Máximo	Mediana	Mínimo	β_{cons}	$\frac{\beta_i^{Mediana}}{\beta_{cons}}$
Regresión Lineal	Ln(PDT)	80%	32%	1,3	0,5	-1,4	-2,2	-24%
	DT	76%	82%	1,7	1,0	-2,5		-45%
	Ln(EDesc)	97%	100%	0,2	0,1	0,1		-6%
Regresión Binomial	Ln(PDT)	80%	32%	8,2	2,8	-9,8	-13,5	-21%
	DT	76%	83%	10,1	7,0	-2,7		-52%
	Ln(EDesc)	97%	100%	1,2	0,7	0,3		-5%
Regresión Logística	Ln(PDT)	80%	32%	7,7	3,2	-9,3	-14,1	-22%
	DT	76%	82%	9,9	5,3	-10,3		-38%
	Ln(EDesc)	97%	100%	1,2	0,7	0,2		-5%

Tabla 58: Análisis de los betas para el M3B
Fuente: Elaboración propia con uso de software stata

8.1.4. Resultados y análisis

Resultados de los modelos seleccionados

Modelo 3A Poisson: Coeficientes de las variables de control

	Lun.	Mar.	Mier.	Jue.	Vier.	Sab.
Boxspring	-0,31***	-0,37***	-0,36***	-0,46***	-0,33***	0,02
Paq. completos boxspring	-0,29***	-0,34***	-0,36***	-0,50***	-0,32***	0,08***
Paq. textiles box spring	-0,25***	-0,16**	-0,31***	-0,27***	-0,25***	0,20***
Cama americana	-0,19***	-0,18***	-0,29***	-0,36***	-0,16***	0,15***
Paq. Comp. cama americana	-0,11***	-0,17***	-0,23***	-0,28***	-0,15***	0,14***
Colch. 1 1/2 pza	-0,12***	-0,15***	-0,22***	-0,31***	-0,23***	0,09***
Colch. 1 pza	-0,08*	-0,14***	-0,24***	-0,27***	-0,21***	0,11**
Colch. 2 pza	-0,12***	-0,25***	-0,26***	-0,38***	-0,27***	0,07**
Colch. medida especial	-0,28***	-0,65***	-0,31***	-0,35***	-0,32***	0,04
Almohadas	-0,38***	-0,42***	-0,47***	-0,40***	-0,26***	0,10***
Protectores de colchón	-0,03	-0,10***	-0,15***	-0,17***	-0,07**	0,16***
Cama	-0,25***	-0,23***	-0,33***	-0,39***	-0,28***	0,13**
Closet/roperos	-0,08*	-0,12**	-0,24***	-0,28***	-0,09*	0,17***
Cómoda	-0,12**	-0,28***	-0,31***	-0,37***	-0,21***	0,08*
Respaldo de cama	-0,26**	-0,38***	-0,26**	-0,41***	-0,14	0,10
Veladores	-0,18***	-0,32***	-0,13**	-0,15**	-0,18***	0,02

Tabla 59: Betas estimados para el modelo 3A de Poisson (1)

Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Fuente: Elaboración propia con uso de software stata

	Año	Feb.	Mar.	Abr.	May.	Jun.
Boxspring	0,07	0,11*	0,06	-0,01	0,08	-0,09
Paq. completos boxspring	0,34***	0,22***	0,21***	0,27***	0,32***	0,14***
Paq. textiles box spring	-0,62***	0,15	0,45***	0,31***	0,51***	0,11
Cama americana	-0,70***	-0,12**	-0,23***	-0,31***	-0,29***	-0,86***
Paq. Comp. cama americana	-0,34***	0,08*	0,21***	-0,03	0,18***	-0,17***
Colch. 1 1/2 pza	-0,39***	-0,02	-0,08*	-0,08*	0,11**	-0,23***
Colch. 1 pza	-0,42***	0,16**	0,02	-0,15**	-0,13*	-0,61***
Colch. 2 pza	-0,52***	-0,00	-0,08*	-0,05	0,04	-0,34***
Colch. medida especial	-0,23	0,36**	0,02	0,19	0,26*	0,05
Almohadas	-0,20***	-0,27***	0,08***	-0,09***	0,06**	-0,12***
Protectores de colchón	-0,37***	-0,03	-0,17***	-0,15***	-0,20***	-0,25***
Cama	-0,43***	-0,15*	-0,20**	-0,25***	-0,15	-0,40***
Closet/roperos	-0,08*	0,13*	0,15**	0,36***	0,38***	0,20**
Cómoda	0,18***	0,21**	0,20**	0,51***	0,46***	0,58***
Respaldo de cama	0,67***	-0,02	-0,25	0,20	0,34**	0,17
Veladores	-0,29***	-0,05	0,23**	0,67***	0,53***	0,61***

Tabla 60: Betas estimados para el modelo 3A de Poisson (2)

Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Fuente: Elaboración propia con uso de software stata

	Jul.	Ago.	Sept.	Oct.	Nov.	Dic.
Boxspring	0,14**	0,18***	0,05	-0,37***	-0,27***	-0,04
Paq. completos boxspring	0,36***	0,33***	0,13***	-0,03	-0,06	-0,03
Paq. textiles box spring	0,55***	0,56***	0,50***	-0,24*	-0,44***	-0,49***
Cama americana	-0,58***	-0,38***	-0,35***	-0,71***	-0,84***	-0,78***

Paq. Comp. cama americana	0,21***	0,10**	-0,08*	-0,44***	-0,38***	-0,27***
Colch. 1 1/2 pza	0,23***	0,07	0,02	-0,36***	-0,42***	-0,18***
Colch. 1 pza	0,08	0,09	-0,09	-0,66***	-0,44***	-0,32***
Colch. 2 pza	0,10**	0,19***	0,10**	-0,19***	-0,31***	-0,24***
Colch. medida especial	0,43***	0,54***	0,39***	0,03	-0,18	-0,12
Almohadas	0,33***	0,44***	0,43***	-0,14***	-0,09***	0,08***
Protectores de colchón	0,20***	0,02	0,10***	-0,35***	-0,42***	-0,33***
Cama	-0,13	-0,10	-0,34***	-0,52***	-0,42***	-0,11
Closet/roperos	0,57***	0,58***	0,72***	0,47***	0,31***	0,04
Cómoda	0,69***	0,79***	0,76***	0,58***	0,50***	0,33***
Respaldo de cama	0,62***	0,50***	0,53***	0,20	0,33**	0,28*
Veladores	1,08***	1,18***	0,94***	0,79***	0,75***	0,48***

Tabla 61: Betas estimados para el modelo 3A de Poisson (3)

Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Fuente: Elaboración propia con uso de software stata

Modelo 3B Lineal: Coeficientes de las variables de control

	Lun.	Mar.	Mier.	Jue.	Vier.	Sab.
Boxspring	-0,04**	-0,03*	-0,05**	-0,08***	-0,05**	-0,04*
Paq. completos boxspring	-0,02*	-0,03*	-0,02*	-0,04***	-0,02	0
Paq. textiles box spring	-0,05	-0,01	-0,09**	-0,06*	-0,06	0,03
Cama americana	-0,04*	-0,05*	-0,08***	-0,08***	-0,05**	0
Paq. Comp. cama americana	-0,01	-0,02*	-0,02	-0,02	-0,02	-0,01
Colch. 1 1/2 pza	-0,03*	-0,02	-0,03	-0,06***	-0,04*	-0,02
Colch. 1 pza	-0,02	-0,02	-0,05*	-0,06**	-0,04	-0,01
Colch. 2 pza	0,01	-0,02	-0,02	-0,03**	-0,01	0,01
Colch. medida especial	-0,02	-0,15***	-0,05	-0,05	-0,04	0,04
Almohadas	0	0	-0,01	0	-0,01	0
Protectores de colchón	0	0	-0,01	-0,03*	-0,01	0
Cama	-0,02	-0,04	-0,03	-0,04	-0,03	0,02
Closet/roperos	-0,01	-0,02	-0,03	-0,06***	0	0,02
Cómoda	-0,02	-0,03	-0,04*	-0,04*	-0,05**	0
Respaldo de cama	-0,05	-0,10**	-0,06	-0,08*	-0,05	0
Veladores	-0,01	-0,07*	-0,04	0	-0,03	0,02

Tabla 62: Betas estimados para el modelo 3B Lineal (1)

Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Fuente: Elaboración propia con uso de software stata

	Año	Feb.	Mar.	Abr.	May.	Jun.
Boxspring	-0,08*	0,01	0	-0,01	0,03	0,01
Paq. completos boxspring	-0,07*	0,01	-0,05***	-0,01	-0,01	-0,03
Paq. textiles box spring	-0,15*	0,02	0,12**	0,08*	0,11**	0,03
Cama americana	-0,03	-0,01	-0,05	-0,01	-0,03	-0,10***
Paq. Comp. cama americana	0,02*	0,01	0	-0,01	0,02	-0,04*
Colch. 1 1/2 pza	0,01	0,03	0	0,01	0,03	0,04
Colch. 1 pza	-0,01	0,07*	0,03	0	-0,02	-0,07*
Colch. 2 pza	-0,08***	0	-0,04*	-0,01	0	-0,03*
Colch. medida especial	-0,02	0,12**	-0,03	0,03	0,06	0,03
Almohadas	-0,02**	-0,09***	0	-0,03	-0,03	-0,03
Protectores de colchón	0,04***	0,01	0,04*	0,04*	0,02	0,03
Cama	-0,08**	-0,03	0,02	0	0,02	-0,06

Closet/roperos	-0,02	0,01	-0,03	0,04	0,02	0
Cómoda	0,05***	0,02	0,02	0,05	0,05	0,09***
Respaldo de cama	0,09	-0,05	-0,13**	-0,04	-0,01	-0,03
Veladores	0	0	0,13**	0,21***	0,16***	0,22***

Tabla 63: Betas estimados para el modelo 3B Lineal (2)
Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Fuente: Elaboración propia con uso de software stata

	Jul.	Ago.	Sept.	Oct.	Nov.	Dic.
Boxspring	0,02	-0,03	-0,06**	-0,04	-0,05*	-0,06**
Paq. completos boxspring	-0,03	-0,03	-0,05**	-0,05**	-0,05**	-0,10***
Paq. textiles box spring	0,07	0,06	0,12**	-0,01	-0,09*	-0,09
Cama americana	-0,19***	-0,07*	-0,05	-0,06	-0,10***	-0,10**
Paq. Comp. cama americana	-0,05**	-0,04*	-0,03	-0,03	-0,06**	-0,10***
Colch. 1 1/2 pza	0,02	-0,02	0	-0,02	-0,03	-0,03
Colch. 1 pza	0,02	-0,04	-0,05	-0,08*	-0,05	-0,04
Colch. 2 pza	-0,03	-0,01	0	-0,01	-0,01	-0,05**
Colch. medida especial	0,03	0,09*	0,08	0,01	-0,05	-0,04
Almohadas	-0,02	0,03	0,01	-0,03	-0,01	-0,07***
Protectores de colchón	0,04**	0	0,02	0,07***	0,05**	-0,04*
Cama	-0,02	-0,04	-0,13***	-0,04	-0,04	0
Closet/roperos	0,02	0,08**	0,10***	0,08**	0,06*	-0,03
Cómoda	0,08**	0,06*	0,09***	0,07*	0,05	0,05
Respaldo de cama	0,04	0,05	0,04	-0,02	-0,02	-0,02
Veladores	0,27***	0,24***	0,20***	0,27***	0,24***	0,18***

Tabla 64: Betas estimados para el modelo 3B Lineal (3)
Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Fuente: Elaboración propia con uso de software stata

8.2. ANEXO B. Información adicional Objetivo Específico 2

8.2.1. Primeras compras de los clientes

Variables generales					
Variable	Medio de Pago	Promedio	SD	Mínimo	Máximo
M_b	Con tarjeta	\$98.586	\$144.326	\$1	\$4.855.801
	Otro medio de pago	\$169.358	\$158.167	\$1	\$8.464.132
	General	\$157.926	\$158.173	\$1	\$8.464.132
MPL_b	Con tarjeta	\$103.348	\$155.970	\$1	\$4.999.990
	Otro medio de pago	\$189.689	\$181.146	\$1	\$7.669.610
	General	\$175.741	\$180.146	\$1	\$7.669.610
MD_b	Con tarjeta	\$2.085	\$17.598	\$-	\$819.980
	Otro medio de pago	\$22.271	\$40.442	\$-	\$2.110.000
	General	\$19.010	\$38.426	\$-	\$2.110.000
U_b	Con tarjeta	1,21	0,70	1	48
	Otro medio de pago	1,29	0,87	1	89
	General	1,28	0,85	1	89
PD_b	Con tarjeta	2%	11%	0%	100%
	Otro medio de pago	14%	15%	0%	100%
	General	12%	15%	0%	100%

Tabla 65: Estadísticos descriptivos detallados
Fuente: Elaboración propia

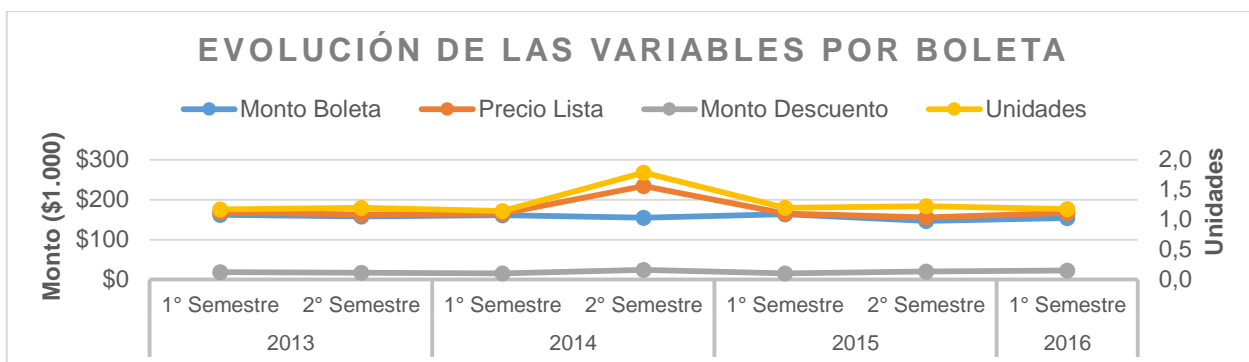


Ilustración 30: Promedios de las variables generales en el tiempo
Fuente: Elaboración propia

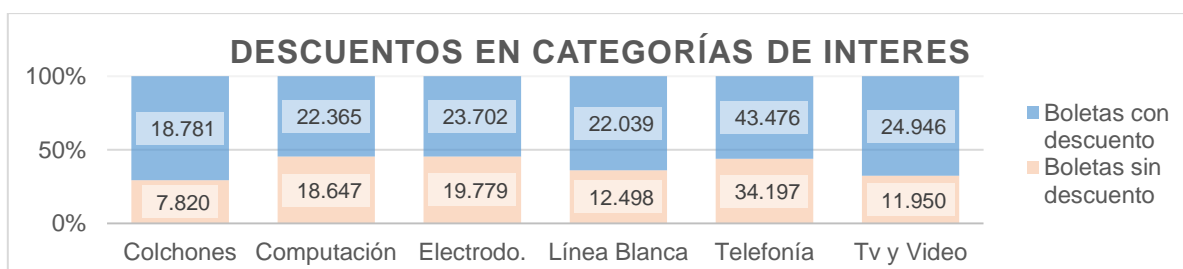


Ilustración 31: Boletas con y sin descuento en cada categoría
Fuente: Elaboración propia

8.2.2. Modelamiento

Comparación de modelos

Resultados de los modelos M1, M2, M3, M4 y M5

Variable	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{M3}$	$\hat{\beta}_{M4}$	$\hat{\beta}_{M5}$
M_b	0,32***	0,39***	0,41***	0,39***	0,24***
$year_b$	0,22***	0,48***	0,48***	0,49***	0,18***
$year_b^2$	-0,03	-0,08***	-0,08***	-0,08***	0
Colchones	5,13***	27,88***	14,54***	23,78***	33,02***
Computación	4,21***	15,72***	7,55***	13,86***	35,53***
Electrodomésticos	3,92***	12,44***	6,92***	11,36***	191,48***
Línea Blanca	4,13***	23,24***	8,73***	19,56***	32,94***
Telefonía	2,98***	2,40***	1,80***	2,27***	31,57***
Televisión y video	3,86***	7,45***	4,90***	6,70***	19,05***
Constante	-1,77***	-1,45***	-1,30***	-1,43***	-1,46***
N	76.638	76638	76638	76638	76638
Pseudo R ²	0,4	0,23	0,16	0,21	0,35

Tabla 66: Resultados Modelos 1, 2, 3, 4 y 5
N. de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Fuente: Elaboración propia con uso de software stata

Correlaciones entre $ATPA_{cb}$ y las otras variables de nivel de descuento

	$r(ATPA_{cb}, PD_{cb})$	$r(ATPA_{cb}, DTPE_{cb})$	$r(ATPA_{cb}, DTPA_{cb})$
Colchones	0.521	0.518	0.521
Computación	0.481	0.474	0.481
Electrodomésticos	0.468	0.467	0.469
Línea Blanca	0.597	0.575	0.597

Telefonía	0.460	0.458	0.460
Tv y video	0.444	0.444	0.445

Tabla 67: Correlación entre la variable ATPA con las otras variables de nivel de descuento para cada categoría
Fuente: Elaboración propia con uso

Métricas de desempeño para los modelos estimados

Modelo	Muestra	Accuracy	Recall ($PT_b = 1$)	Recall ($PT_b = 0$)	Precision ($PT_b = 1$)	Precision ($PT_b = 0$)
1	Dentro	83%	100%	3%	83%	96%
	Fuera	94%	100%	0%	94%	0%
2	Dentro	87%	95%	44%	89%	67%
	Fuera	95%	100%	2%	95%	72%
3	Dentro	86%	96%	42%	89%	67%
	Fuera	94%	100%	0%	94%	0%

Tabla 68: Métricas de desempeño, Base 1
Fuente: Elaboración propia con uso de software stata

Modelo	Muestra	Accuracy	Recall ($PT_b = 1$)	Recall ($PT_b = 0$)	Precision ($PT_b = 1$)	Precision ($PT_b = 0$)
1	Dentro	86%	100%	0%	86%	0%
	Fuera	94%	100%	0%	94%	0%
2	Dentro	86%	100%	4%	86%	73%
	Fuera	95%	100%	2%	95%	85%
3	Dentro	86%	100%	0%	86%	82%
	Fuera	94%	100%	0%	94%	0%

Tabla 69: Métricas de desempeño, Base 2
Fuente: Elaboración propia con uso de software stata

Modelo	Muestra	Accuracy	Recall ($PT_b = 1$)	Recall ($PT_b = 0$)	Precision ($PT_b = 1$)	Precision ($PT_b = 0$)
1	Dentro	74%	63%	84%	79%	71%
	Fuera	83%	89%	64%	89%	64%
2	Dentro	84%	70%	97%	95%	77%
	Fuera	83%	80%	95%	98%	59%
3	Dentro	83%	70%	95%	93%	77%
	Fuera	82%	79%	92%	97%	58%

Tabla 70: Métricas de desempeño, SubBase 1
Fuente: Elaboración propia con uso de software stata

Modelo	Muestra	Accuracy	Recall ($PT_b = 1$)	Recall ($PT_b = 0$)	Precision ($PT_b = 1$)	Precision ($PT_b = 0$)
1	Dentro	71%	55%	85%	76%	68%
	Fuera	81%	87%	66%	86%	67%
2	Dentro	86%	73%	98%	97%	80%
	Fuera	84%	79%	97%	99%	65%
3	Dentro	85%	73%	95%	94%	80%
	Fuera	83%	79%	93%	97%	64%

Tabla 71: Métricas de desempeño, SubBase 2
Fuente: Elaboración propia con uso de software stata

8.3. ANEXO C. Información adicional Objetivo Específico 3

8.3.1. Valor de los clientes

Características de los clientes

Grupo Etario	Porcentaje	VCD_i Acumulado	VCD_i Promedio
<0 o Desconocido	1%	8	0,006
0-9	0%	0	0,007
10-19	0%	0	0,012
20-29	17%	495	0,016
30-39	23%	588	0,014
40-49	21%	478	0,012
50-59	20%	392	0,011
60-69	12%	206	0,009
70-79	6%	72	0,007
80-89	1%	9	0,006
90-100	0%	1	0,007
>100	0%	1	0,006
General	-	2.249	0,012

Tabla 72: Estadísticos descriptivos por grupo etario
Fuente: Elaboración propia

GSE	Porcentaje	VCD_i Acumulado	VCD_i Promedio
ABC1	8%	196	0,014
C2	18%	417	0,013
C3	21%	457	0,012
D	35%	734	0,011
E	9%	203	0,012
Desconocido	10%	243	0,013

Tabla 73: Estadísticos descriptivos por GSE
Fuente: Elaboración propia

Sexo	Porcentaje	VCD_i Acumulado	VCD_i Promedio
Desconocido	1%	9	0,006
Hombre	51%	1.045	0,011
Mujer	48%	1.195	0,013

Tabla 74: Estadísticos descriptivos por sexo
Fuente: Elaboración propia

8.3.2. Modelamiento

Comparación de modelos

Resultados de los modelos M1, M2, M3, M4, M5 y M6

Variable	$\hat{\beta}_{M1}$	$\hat{\beta}_{M2}$	$\hat{\beta}_{M3}$	$\hat{\beta}_{M4}$	$\hat{\beta}_{M5}$	$\hat{\beta}_{M6}$
M_i	0,14***	0,14***	0,13***	0,13***	0,13***	0,15***
U_i	0,04***	0,03***	0,04***	0,04***	0,04***	0,02***
$year_i$	0,32***	0,31***	0,31***	0,31***	0,31***	0,31***
$m_{Feb,i}$	0,04***	0,04***	0,04***	0,04***	0,04***	0,04***
$m_{Mar,i}$	0,06***	0,06***	0,06***	0,06***	0,06***	0,06***
$m_{Abr,i}$	0,09***	0,09***	0,09***	0,09***	0,09***	0,09***

$m_{May,i}$	0,12***	0,12***	0,12***	0,12***	0,12***	0,12***
$m_{Jun,i}$	0,12***	0,12***	0,12***	0,12***	0,12***	0,12***
$m_{Jul,i}$	0,09***	0,1***	0,09***	0,09***	0,09***	0,11***
$m_{Ago,i}$	0,11***	0,12***	0,11***	0,11***	0,11***	0,14***
$m_{Sep,i}$	0,12***	0,14***	0,12***	0,12***	0,12***	0,15***
$m_{Oct,i}$	0,27***	0,27***	0,27***	0,27***	0,27***	0,27***
$m_{Nov,i}$	0,25***	0,25***	0,25***	0,25***	0,25***	0,25***
$m_{Dic,i}$	0,25***	0,27***	0,26***	0,26***	0,26***	0,26***
$EDAD_i$	-0,01***	-0,01***	-0,01***	-0,01***	-0,01***	-0,01***
$S_{HOMBRE,i}$	0,16*	0,17*	0,16*	0,16*	0,16*	0,16*
$S_{MUJER,i}$	0,36***	0,35***	0,35***	0,35***	0,35***	0,35***
$g_{C2,i}$	-0,1***	-0,1***	-0,1***	-0,1***	-0,1***	-0,1***
$g_{C3,i}$	-0,19***	-0,18***	-0,18***	-0,18***	-0,18***	-0,19***
$g_{D,i}$	-0,24***	-0,23***	-0,23***	-0,23***	-0,23***	-0,24***
$g_{E,i}$	-0,2***	-0,19***	-0,2***	-0,2***	-0,2***	-0,2***
DB_i	-0,02***	0,15***	0,01	0,01	0,01	0,02*
Colchones	-	0,02	0,52***	0,52***	0,52***	0,34***
Computación	-	-0,2***	-0,22***	-0,22***	-0,22***	-0,49***
Electrodomésticos	-	-0,03	0,11***	0,11***	0,11***	2,06***
Línea Blanca	-	-0,21***	-0,41***	-0,41***	-0,41***	-0,51***
Telefonía	-	-0,19***	-0,28***	-0,28***	-0,28***	-0,19***
Televisión y video	-	-0,29***	-0,38***	-0,38***	-0,38***	-0,57***
Constante	-0,18*	-0,18*	-0,17*	-0,17*	-0,17*	-0,17*
N° de observaciones	134.605	134.605	134.605	134.605	134.605	134.605
R ² ajustado	16%	16%	16%	16%	16%	16%

Tabla 75: Resultados Modelos 1, 2, 3, 4, 5 y 6
Niveles de significancia: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Fuente: Elaboración propia con uso de software stata