



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

EVALUACIÓN DE EFECTOS DE ACTIVACIÓN Y EFECTOS CRUZADOS
ENTRE CATEGORÍAS EN CLIENTES CON CATEGORÍAS INACTIVAS
DE UNA TIENDA POR DEPARTAMENTOS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

IGNACIO ANDRÉS MARÍN CASAS-CORDERO

PROFESOR GUÍA:
PABLO MARÍN VICUÑA

MIEMBROS DE LA COMISIÓN:
CAROLINA SEGOVIA RIQUELME
RUBÉN DAZA BARRA

SANTIAGO DE CHILE
2020

EVALUACIÓN DE EFECTOS DE ACTIVACIÓN Y EFECTOS CRUZADOS ENTRE CATEGORÍAS EN CLIENTES CON CATEGORÍAS INACTIVAS DE UNA TIENDA POR DEPARTAMENTOS

Durante las últimas décadas, el avance de las tecnologías de la información y el dinamismo en el comportamiento de compra de los clientes, han aumentado la necesidad de comprender en profundidad las acciones futuras que se deberían tomar para aumentar la participación del cliente en la empresa.

La presente memoria, llevada a cabo en una empresa de retail de tienda por departamentos, perteneciente a un reconocido grupo a nivel latinoamericano, tiene por objetivo principal evaluar los efectos de activación en clientes con categorías de producto inactivas y la relación que existe entre la compra en otras categorías y la activación. Asimismo, el propósito y su objetivo se originan en la necesidad de aumentar la frecuencia de compras del cliente y con ello, mejorar la personalización de las ofertas de email marketing para incentivar la compra en categorías que se encuentren inactivas.

Con lo anterior, para el desarrollo de la solución se plantean cuatro objetivos específicos. En primer lugar, evaluar el efecto causal que posee la activación en diferentes categorías de productos y el aumento en el gasto de los clientes, a través del uso de una técnica de emparejamiento. Con respecto a los resultados, se concluye que existe un efecto positivo y significativo en las 11 categorías de producto analizadas, de las cuales se destacan Electro hogar y Decoración que generan un aumento en el gasto de \$16.633 (111% sobre el promedio) y \$14.840 (88% sobre el promedio), respectivamente. Luego, como segundo objetivo se tiene el diseñar y construir un modelo para predecir la próxima categoría en la que se activará un cliente, dado su comportamiento y características tanto sociodemográficas como propias del negocio. El modelo escogido es el Logit binario, dada su facilidad de procesamiento e interpretación.

El tercer objetivo consiste en definir un criterio que permita balancear la propensión de activarse del cliente y su gasto posterior, del que se obtiene que el mejor criterio es el de gasto esperado, que aumenta un 40% el gasto posterior y presenta una leve pérdida de precisión con respecto al uso clásico de la propensión como tasa de respuesta. Además, se obtiene que, en promedio, un 5% de los clientes tiene alta probabilidad de activación natural, lo que aumentaría su gasto posterior en un 21%. Por ello, se propone a la empresa potenciar acciones de marketing a clientes en categorías que no posean alta probabilidad a activarse, pero si un alto gasto posterior, como lo son Calzado mujer y Decoración.

En cuarto lugar, se pretende elaborar un orden de asignación de ofertas dada la relación que existe entre la activación de una categoría y la compra en otras. De los resultados se concluye que, para todas las categorías inactivas, existe una relación positiva de al menos una categoría distinta en su activación. En particular, Moda hombre con Calzado hombre (9,3%), Deportes (6,8%) e Infantil (6,6%); y para Moda mujer, con Calzado mujer (9,3%), Accesorios (8,5%) y Ropa interior (8,3%).

Finalmente, se exponen las principales limitaciones del trabajo, que dan origen a propuestas de proyectos futuros. Entre ellos, el estudio de la relación que existe en la compra conjunta de dos o más categorías en la misma ocasión de compra y la incorporación del efecto de herramientas de marketing directo sobre la propensión.

*Para mi madre, mi novia y mi familia
a quienes debo todo lo que soy...*

TABLA DE CONTENIDO

1. ANTECEDENTES GENERALES	1
1.1 ANTECEDENTES DE LA INDUSTRIA DEL RETAIL	1
1.2 ANTECEDENTES DE LA ORGANIZACIÓN	3
1.2.1 RUBRO	3
1.2.2 MISIÓN, VISIÓN Y VALORES	3
1.2.3 ESTRUCTURA ORGANIZACIONAL	4
1.2.4 PRODUCTO Y CLIENTES	4
1.2.5 DIMENSIONAMIENTO DE ACTIVIDAD	5
1.2.6 VENTAJAS COMPETITIVAS	6
1.3 MERCADO Y MARCO INSTITUCIONAL	6
1.4 DESEMPEÑO ORGANIZACIONAL	8
2. PLANTEAMIENTO Y DESCRIPCIÓN DEL PROYECTO	10
2.1 INFORMACIÓN DEL ÁREA DE LA ORGANIZACIÓN	10
2.2 CONTEXTO DEL TRABAJO EN EL ÁREA	11
2.3 IDENTIFICACIÓN DEL PROBLEMA U OPORTUNIDAD	12
2.4 ESTADO DEL ARTE	12
2.5 INDICADORES DE RELEVANCIA, CAUSAS Y EFECTOS	13
2.5.1 INDICADORES DE RELEVANCIA	13
2.5.2 CAUSAS	14
2.5.3 EFECTOS	14
2.6 HIPÓTESIS Y SOLUCIÓN PROPUESTA	14
2.7 TRABAJOS PREVIOS	15
2.8 OBJETIVOS	16
2.8.1 OBJETIVO GENERAL	16
2.8.2 OBJETIVOS ESPECÍFICOS	16
2.9 RESULTADOS ESPERADOS	16
2.10 ALCANCES	17

3. MARCO CONCEPTUAL	18
3.1 MEJOR PRÓXIMA OFERTA	18
3.1.1 LOGIT	19
3.2 INFERENCIA CAUSAL	19
3.2.1 MÉTODOS DE EMPAREJAMIENTO	20
3.2.1.1 MEDIDAS DE CERCANÍA	21
3.2.1.2 ALGORITMOS PARA ESTIMACIÓN DE EMPAREJAMIENTO	21
3.3 MÉTRICAS DE EVALUACIÓN	23
3.3.1 MATRIZ DE CONFUSIÓN	23
3.3.2 CURVAS COR – AUC Y DE GANANCIA	23
4. METODOLOGÍA	25
4.1 EFECTO DE LA ACTIVACIÓN DE CATEGORÍAS DE PRODUCTO	27
4.1.1 ETAPA 1: ESTIMACIÓN DEL PROPENSITY SCORE	28
4.1.2 ETAPA 2: ELECCIÓN DE ALGORITMO DE EMPAREJAMIENTO	29
4.1.3 ETAPA 3: VERIFICACIÓN DE SOPORTE COMÚN	29
4.1.4 ETAPA 4: CALIDAD DE EMPAREJAMIENTO Y ESTIMACIÓN DEL EFECTO	29
4.2 MEJOR PRÓXIMA CATEGORÍA	30
4.2.1 MODELO DE REGRESIÓN LOGÍSTICA (LOGIT)	30
4.3 ELECCIÓN DE PRÓXIMA CATEGORÍA PARA CORREOS DE MARKETING	30
5. DESARROLLO METODOLÓGICO	32
5.1 COMPRESIÓN DE LOS DATOS	32
5.1.1 INFORMACIÓN TRANSACCIONAL	32
5.1.2 INFORMACIÓN DE CLIENTES	34
5.1.3 ANÁLISIS EXPLORATORIO DEL CONJUNTO DE DATOS	35
5.2 PREPARACIÓN DE LOS DATOS	41
5.3 MODELAMIENTO Y EVALUACIÓN DE RESULTADOS	43
5.3.1 ANÁLISIS DE EFECTOS CAUSALES DE ACTIVACIÓN EN CATEGORÍAS	43
5.3.1.1 CÁLCULO DEL PROPENSITY SCORE	44

5.3.1.2	SOPORTE COMÚN Y EMPAREJAMIENTO	47
5.3.1.3	CALIDAD DEL EMPAREJAMIENTO	48
5.3.1.4	ESTIMACIÓN DEL EFECTO CAUSAL	50
5.3.2	MEJOR PRÓXIMA CATEGORÍA	53
5.3.2.1	RELEVANCIA EXPLICATIVA NO LIGADA A ACTIVIDAD EN CATEGORÍAS	55
5.3.2.2	SENSIBILIDAD A LA COMPRA EN OTRAS CATEGORÍAS	55
5.3.2.3	ASOCIACIONES CRUZADAS Y TEMÁTICAS ENTRE CATEGORÍAS	56
6.	<u>CRITERIOS DE ASIGNACIÓN DE CATEGORÍAS Y USOS POSIBLES</u>	57
7.	<u>CONCLUSIONES</u>	62
8.	<u>LIMITACIONES Y TRABAJOS FUTUROS</u>	64
9.	<u>BIBLIOGRAFÍA</u>	66
10.	<u>ANEXOS</u>	70

INDICE DE FIGURAS

FIGURA 1: PARTICIPACIÓN EN TIENDAS POR DEPARTAMENTOS.	1
FIGURA 2: VARIACIÓN DE CANTIDAD DE EMPRESAS PARTICIPANTES EN EVENTOS CYBER.	2
FIGURA 3: VARIACIÓN EN LA CANTIDAD DE TRANSACCIONES POR EVENTO CYBER.	2
FIGURA 4: DIAGRAMA GERENCIAL DE LA EMPRESA.	4
FIGURA 5: DIAGRAMA ORGANIZACIONAL ÁREA INTELIGENCIA DE CLIENTES.	10
FIGURA 6: REPRESENTACIÓN DE COMPORTAMIENTO DE CURVAS ROC Y CLASIFICACIÓN ALEATORIA.	24
FIGURA 7: CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM).	25
FIGURA 8: ETAPAS DE IMPLEMENTACIÓN DE EMPAREJAMIENTO POR PROPENSITY SCORE.	28
FIGURA 9: EVOLUCIÓN DE CANTIDAD DE TRANSACCIONES MENSUALES (ÚLTIMOS 2 AÑOS).	34
FIGURA 10: EVOLUCIÓN DE MONTO DE VENTAS MENSUAL DE (ÚLTIMOS 2 AÑOS).	34
FIGURA 11: DISTRIBUCIÓN ETARIA POR GÉNERO ENTRE EL CONJUNTO DE CLIENTES.	35
FIGURA 12: DISTRIBUCIÓN DE ESTADO DE ACTIVIDAD POR TIPO DE CLIENTES.	37
FIGURA 13: DISTRIBUCIÓN ETARIA DE CLIENTES ANTIGUOS.	37
FIGURA 14: DISTRIBUCIÓN ETARIA DE CLIENTES NUEVOS.	37
FIGURA 15: ZONA DE RESIDENCIA DE CLIENTES ANTIGUOS.	37
FIGURA 16: ZONA DE RESIDENCIA DE CLIENTES NUEVOS.	37
FIGURA 17: DISTRIBUCIÓN DE GÉNERO CLIENTES ANTIGUOS.	38
FIGURA 18: DISTRIBUCIÓN DE GÉNERO CLIENTES NUEVOS.	38
FIGURA 19: COMPORTAMIENTO DE COMPRA EN DISTINTAS AGRUPACIONES DE PRODUCTOS EN CLIENTES ACTIVOS E INACTIVOS.	39
FIGURA 20: DISTRIBUCIÓN PORCENTUAL DE LA VENTA PROMEDIO ANUAL POR CATEGORÍA DE PRODUCTO POR TIPO DE CANAL.	39

FIGURA 21: NAVEGACIÓN PROMEDIO ANUAL EN CATEGORÍAS PARA CLIENTE ACTIVOS E INACTIVOS.	40
FIGURA 22: FRECUENCIA DE COMPRA EN DISTINTAS LÍNEAS DE PRODUCTO Y GASTO PROMEDIO ANUAL.	41
FIGURA 23: CRITERIO DE EXCLUSIÓN PARA EL PAR 1: MONTOS DE VENTA ANUAL Y CANTIDAD DE PRODUCTOS COMPRADOS ANUAL.	42
FIGURA 24: VENTANA DE TIEMPO PARA ESTUDIAR EFECTOS CAUSALES DE ACTIVACIÓN EN CATEGORÍAS.	44
FIGURA 25: CANTIDAD DE CLIENTES EN LAS BASES ANALÍTICAS DE CADA CATEGORÍA Y SU PORCENTAJE DE ACTIVACIÓN.	44
FIGURA 26: CURVA DE RECEPCIÓN OPERATIVA DEL RECEPTOR QUE INDICA UN ÁREA BAJO LA CURVA (AUC) DE 69%.	47
FIGURA 27: VERIFICACIÓN GRAFICA DEL SOPORTE COMÚN DEL PROPENSITY SCORE PARA LA CATEGORÍA DECORACIÓN.	48
FIGURA 28: BALANCE ENTRE VARIABLES RELEVANTES PARA EL EMPAREJAMIENTO POR PROPENSITY SCORE Y GASTO PRE-ACTIVACIÓN.	49
FIGURA 29: EFECTO CAUSAL DE ACTIVACIÓN EN LA CATEGORÍA DECORACIÓN SOBRE EL GASTO DEL CLIENTE EN POST-ACTIVACION.	51
FIGURA 30: EFECTO ESTIMADO DE LA ACTIVACIÓN DE CLIENTES SOBRE EL GASTO EN EL PERIODO POST-ACTIVACIÓN EN DECORACIÓN DIFERENCIADO POR EL TIPO DE COMPRA.	51
FIGURA 31: EFECTO ESTIMADO DE LA ACTIVACIÓN DE CLIENTES SOBRE EL GASTO EN EL PERIODO POST-ACTIVACIÓN EN CADA CATEGORÍA ESTUDIADA.	52
FIGURA 32: RESUMEN DE MÉTRICA DE PREDICCIÓN AUC PARA TODAS LAS CATEGORÍAS.	53
FIGURA 33: GANANCIA RESPECTO A LA PRECISIÓN DE CADA CRITERIO DE ELECCIÓN DE MEJOR PRÓXIMA CATEGORÍA.	58

INDICE DE TABLAS

TABLA 1: INGRESOS ANUALES 2017 Y 2018 PARA RETAIL EN PAÍSES DE PRESENCIA.	6
TABLA 2: ESTADO GENERAL DE CLIENTES EN TIENDA POR DEPARTAMENTOS.	11
TABLA 3: DETALLE DE CARACTERÍSTICAS DE CORREOS DE EMAIL MARKETING.	13
TABLA 4: MATRIZ DE CONFUSIÓN.	23
TABLA 5: INFORMACIÓN PRINCIPAL DE BASE DE DATOS TRANSACCIONAL.	32
TABLA 6: DESCRIPCIÓN DE JERARQUÍA TIPO LÍNEA ELECTRO HOGAR (AGRUPACIÓN DE PRODUCTOS).	33
TABLA 7: INFORMACIÓN PRINCIPAL DE BASE DE DATOS SOCIODEMOGRÁFICA.	34
TABLA 8: INFORMACIÓN PRINCIPAL DE BASE DE DATOS DE NAVEGACIÓN.	35
TABLA 9: COMBINACIÓN DE VARIABLES PARA ESTUDIO DE OUTLIERS.	42
TABLA 10: COMPORTAMIENTO ANUAL DE GASTO, VISITAS CON COMPRA Y CANTIDAD DE PRODUCTOS PARA CLIENTES EXCLUIDOS Y NO EXCLUIDOS.	43
TABLA 11: BALANCE DE VARIABLES CLAVES ESCOGIDAS PARA EL EMPAREJAMIENTO.	49
TABLA 12: DIFERENCIA EN EL GASTO PRE-ACTIVACIÓN PARA AMBOS GRUPOS ANTES Y DESPUES DEL EMPAREJAMIENTO.	50
TABLA 13: RESUMEN DEL EFECTO CAUSAL DE ACTIVACIÓN EN LA CATEGORÍA DECORACIÓN SOBRE EL GASTO POST-ACTIVACIÓN.	50
TABLA 14: TIPO DE DESCUENTO USADOS PARA LA ACTIVACIÓN DE CLIENTES.	60
TABLA 15: RELEVANCIA DE CATEGORÍAS QUE SEGÚN SU ACTIVIDAD DE COMPRA PREVIA AUMENTAN LA PROBABILIDAD DE ACTIVARSE EN MODA HOMBRE.	60

1. ANTECEDENTES GENERALES

En esta primera sección, se presentan los principales antecedentes del proceso de trabajo de memoria de título, realizado en una importante empresa de retail de Latinoamérica, y en particular, de Chile (desde ahora en adelante, La empresa) [1].

1.1 Antecedentes de la industria del retail

La industria del retail, o, dicho en otras palabras, el comercio que se realiza al detalle directamente al consumidor lleva muchos años en una posición consolidada en Chile a través de distintos actores del mercado que han logrado posicionar sus empresas a nivel nacional mediante diversas cadenas, tales como, supermercados, farmacias, tiendas por departamentos, entre otras. Entre las principales cadenas que lograron consolidación en el país están: Cencosud, D&S, La empresa, Ripley, La Polar, SMU, Mall Plaza, Parque Arauco e Hites.

Durante los últimos años, esta industria ha tenido que adaptar su modelo de negocio a las nuevas tendencias tecnológicas, entre ellos el comercio electrónico y móvil, creando nuevos desafíos que se enfocan en desarrollar nuevas herramientas que permitan dar respuesta a la exigencia de productos y servicios cada vez más personalizados, que mejoren la relación con el cliente [2].

En particular, con respecto a tiendas por departamentos (rubro en el que se llevará a cabo la memoria de título) se tiene que existen tres actores principales, los cuales concentran casi la totalidad del mercado. En la Figura 1 se observa la participación de mercado de diferentes empresas que poseen tiendas por departamentos [3].

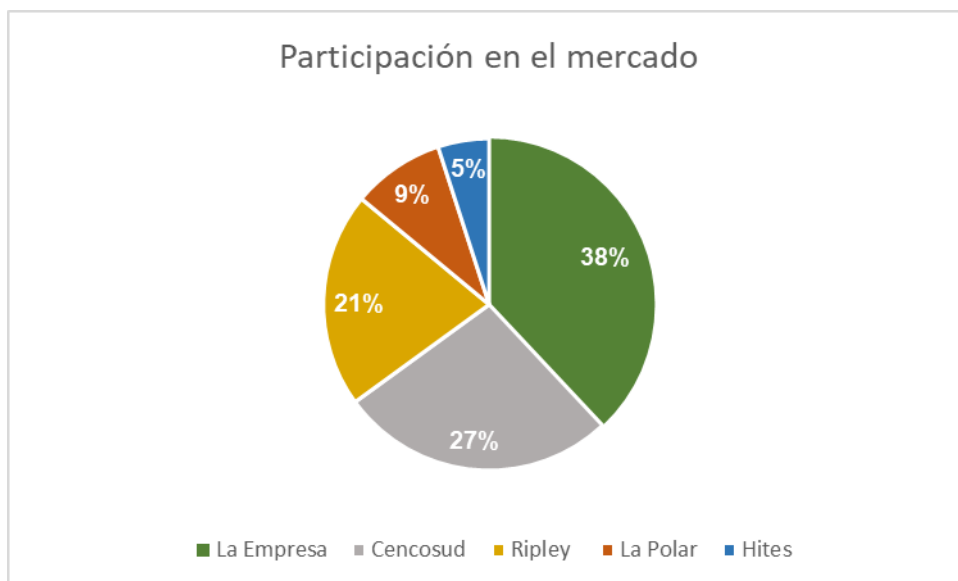


Figura 1: Participación en tiendas por departamentos.
Fuente: Elaboración propia.

Asimismo, es importante destacar el crecimiento del comercio electrónico en el retail, que, gracias a la mayor penetración y democratización del acceso a internet y al uso de nuevas tecnologías, ha permitido fortalecer y ampliar una nueva fuerza de venta

para la industria del retail. Lo anterior es posible evidenciarlo de acuerdo al crecimiento de eventos especiales que concentran la totalidad de la oferta a través del comercio electrónico, como lo son eventos del tipo Cyber. Esta tendencia al alza es posible observarla tanto por el lado de la oferta como por el de la demanda, como se muestra en las Figuras 2 y 3.

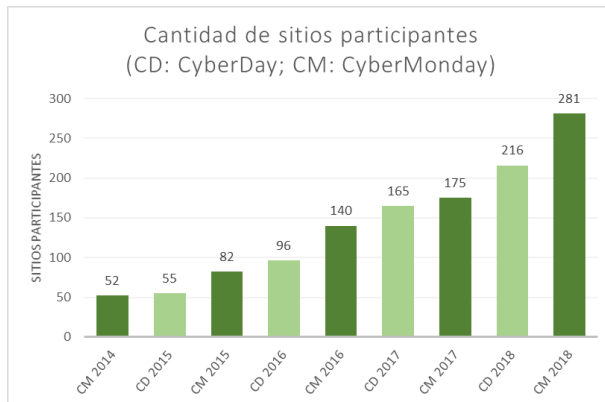


Figura 2: Variación de cantidad de empresas participantes en eventos Cyber.

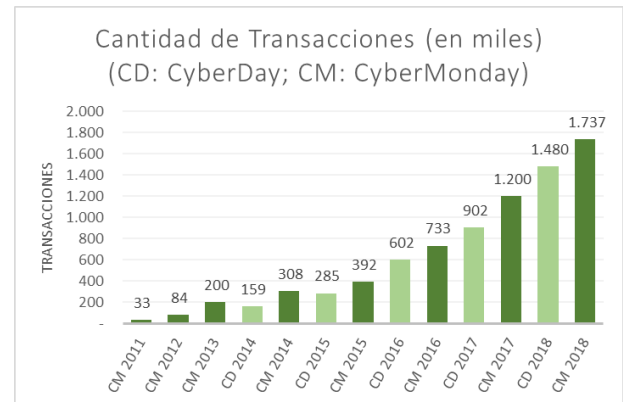


Figura 3: Variación en la cantidad de transacciones por evento Cyber.

Fuente: Elaboración propia extraída de CCS.

Con respecto al contexto competitivo del mercado nacional, como se mencionó anteriormente, existen tres actores principales en la industria del retail – entre los cuales se encuentra La empresa – que, en concordancia con los cambios en transformación y migración digital, han implementado diversas herramientas que mejoren el entendimiento sobre el comportamiento de compra del cliente.

Así pues, los desafíos de estos actores se alinean en lo que se refiere a la *Omnicanalidad*¹, es decir, que el cliente perciba y experimente la misma calidad en los distintos canales de compra a su disposición. Para ello, cada empresa ha realizado esfuerzos en inversión que contemplan diversas estrategias que apuntan hacia el mismo objetivo, como lo son, la migración masiva de datos a servicios de almacenamiento en la nube, creación de gerencias de analítica avanzada de datos, ciencia de los datos y gobierno de datos que permitan aprovechar la disponibilidad de información que se recopila a través de los distintos canales de interacción que poseen las empresas con sus clientes. Estos cambios también se perciben en la implementación de diversas herramientas de automatización de procesos a través del uso de inteligencia artificial y aprendizaje de máquinas, como, por ejemplo, los *chat-bots* de apoyo y ayuda tanto al colaborador como al cliente y los servicios de venta remota.

En el ámbito internacional, el aumento en el uso de canales digitales y el aprovechamiento de la información también se encuentra en alza, donde, entre la industria que prioriza en mayor medida el uso de aprendizaje de máquinas e inteligencia artificial² se encuentra la industria del marketing y ventas (de la que forma parte el retail) y que son claves para los propósitos propuestos. De forma particular se tienen empresas

¹ Extraído de memoria anuales 2018 de empresas Ripley, Cencosud y La empresa.

² Aprendizaje de máquinas e inteligencia artificial son herramientas tecnológicas que tienen la capacidad de poder enseñar a una máquina a encontrar patrones no evidentes haciendo uso de grandes volúmenes de datos.

internacionales de gran relevancia que actualmente pretenden integrar diversos mecanismos haciendo uso de inteligencia artificial, entre ellos se encuentran Alibaba, Amazon, Apple, entre otras. Estas empresas poseen vasta experiencia en transformación y migración digital a través del *e-commerce*, ciencia de datos, manejo de datos masivos y lenguajes de programación lo que les ha permitido generar herramientas robustas de aprendizaje de máquinas e inteligencia artificial para la automatización de procesos (Forbes, 2019) [4] [5].

Cabe destacar, por último, que lo que caracteriza y tienen en común estas empresas en cuanto al uso de tecnología en la ciencia de los datos, es la constante búsqueda de la predicción del comportamiento del consumidor, en otras palabras, cual podría ser con mayor probabilidad la siguiente compra o acción del cliente.

1.2 Antecedentes de la organización

Chile, en virtud del crecimiento de la población, la economía y la industrialización, ha experimentado un importante crecimiento a nivel nacional a través del último siglo [6]. Ante este panorama no han estado exentas las empresas de origen nacional, entre ellas La empresa, lugar donde será desarrollada la memoria de título. Asimismo, y en virtud del aumento en el acceso a información masiva (*Big Data*), tanto la industria como la sociedad, han cambiado su comportamiento y necesidades relacionadas con su mayor involucramiento al mundo digital [7].

Es de acuerdo a lo anterior, que resulta de suma relevancia, comprender el estado y evolución de las personas a través de su ciclo de vida comercial, que permita optimizar la relación con sus necesidades y requerimientos. Durante los últimos años, el estudio del comportamiento del consumidor a través de la ciencia de los datos ha cobrado un alto valor para la industria, ya que su propósito permite identificar diversos patrones no evidentes insertos en un universo de datos desagregados. La empresa, siguiendo a su lema, de situar al cliente en primera posición, no ha escatimado esfuerzos para apostar por la innovación y la tecnología para aumentar el conocimiento y entendimiento de sus clientes a través de distintas metodologías, herramientas y mecanismos de vanguardia que le permitan mantener y consolidar su posición de líder de mercado a nivel país y latinoamericano.

1.2.1 Rubro

La empresa forma parte de un *holding* mayor que posee actividades comerciales en diferentes rubros de acuerdo con el sector económico al cual pertenecen. En particular, retail desarrolla actividades en el comercio al por menor, mediante tiendas por departamento, a través de múltiples canales de compra.

1.2.2 Misión, visión y valores

La organización define su misión como “El motor y centro de nuestro negocio, es la pasión por nuestros clientes, buscamos conocerlos y servirlos en todas sus necesidades de consumo, lo que nos lleva a reinventarnos e innovar para entregarles la mejor experiencia de compra, a través de una estrategia de formatos y canales diferenciada, con marcas fuertemente posicionadas en nuestros mercados” y su visión

como “Queremos ser la plataforma de retail online y offline líder en América Latina”. De lo anterior se desprende el foco tanto en el entendimiento del cliente como la transferencia al mundo digital. Los valores que afirma defender la empresa son “somos un equipo”, “superamos las expectativas de los clientes”, “actuamos con sentido”, “crecemos por nuestros logros” y “hacemos que las cosas pasen”. Por tanto, se infiere que existen valores que guardan estrecha relación con fortalecer la relación con el cliente a través de una labor impecable y en continuo mejoramiento.

1.2.3 Estructura organizacional

La distribución organizacional que utiliza La empresa en su esquema corporativo corresponde a una estructura matricial en esencia y funcionamiento, no obstante, para efectos prácticos, se representa mediante una estructura funcional clásica. Lo anterior se observa en la Figura 4, correspondiente a la planilla corporativa, que tiene como puntal inicial a su Gerente General, para luego dar paso a la planilla gerencial asociada a las funciones fundamentales de la empresa.

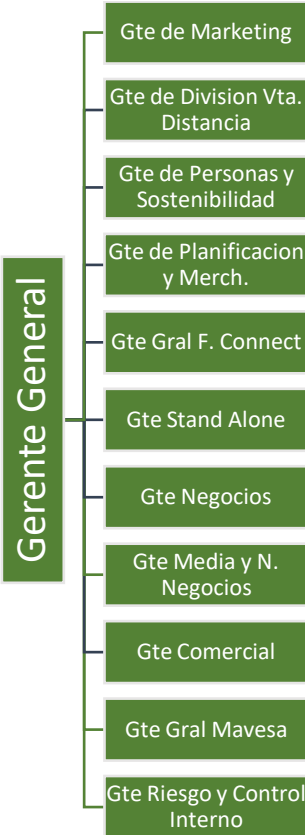


Figura 4: Diagrama gerencial de la empresa. Fuente: Elaboración propia.

1.2.4 Producto y clientes

La empresa, desde su origen, centra sus operaciones en el formato de tiendas por departamento, a través del cual ofrece un vasto catálogo de productos tanto de uso personal como para mejoramiento del hogar. El catálogo de productos se divide en una serie de categorías que contemplan tipos de productos como vestuario, calzado, artículos

de belleza, artículos electrónicos, electrodomésticos, muebles y accesorios de decoración, entre otros.

La empresa posee un modelo de negocios enfocado en la mujer como cliente principal, por tanto, desarrolla su propuesta estratégica y comercial en su satisfacción. Para atender a sus clientes, la empresa promueve la omnicanalidad, es decir, que todos los servicios que interactúan con el cliente sean percibidos de la misma manera indistintamente y con el mismo estándar de calidad. Para ello, La empresa en Chile, dispone tanto de canales presenciales, compuesto por tiendas físicas, como virtuales, compuesto por el portal web de la empresa y la aplicación para teléfonos móviles.

1.2.5 Dimensionamiento de actividad

La unidad de negocios de tiendas por departamento posee presencia en diversos países de Latinoamérica, tales como Chile, Perú, Colombia y Argentina a través de 114 tiendas, sumado a su presencia digital mediante la plataforma online y aplicación móvil. Asimismo, posee diferentes porcentajes de participación de mercado por cada país, siendo en Chile un 23%, en Perú un 18%, en Colombia un 7% y en Argentina un 2%.

De acuerdo con los datos entregados durante el último cuarto del 2018, los ingresos de la unidad de retail ascendieron a un monto de \$2.559.540 millones, un aumento de 2,6% con respecto al último cuarto del año precedente. Además, el canal online de negocios experimentó un crecimiento de 24,8% en comparación al 4Q 2017 por un total de \$203.028 millones, ingreso que corresponde a un 9,7% del total de los ingresos del retail. Sin embargo, el crecimiento relativo de los ingresos de la unidad de retail estuvo parcialmente contrarrestado por un decrecimiento de los ingresos en tiendas por departamento, explicado en parte por una dinámica de consumo débil y un decaimiento en la participación de compras por parte de extranjeros. Los ingresos anuales del periodo 2018 corresponden a \$1.525.908 millones, lo que corresponde a una disminución del 0,6% en comparación al 2017. Los resultados de ingresos anuales por retail para cada país se resumen en la Tabla 1.

Las ventas totales durante el 2018 tuvieron un decrecimiento correspondiente al 3,7%, muy por debajo del crecimiento experimentado durante el año 2017 (4,8%). Además, el resultado operacional de tiendas por departamento reportó un EBITDA de \$71.558 millones, una disminución del 35,9% con respecto al EBITDA reportado el 2017. Finalmente, la utilidad neta consolidada para el último trimestre del 2018 reportó un monto de \$160.638 millones, una disminución de un 11,1% con respecto al mismo cuarto del año anterior.

Ingresos retail periodo 2017 - 2018 (MM\$)				
	Anual 2017	Anual 2018	Var %	Var % (Moneda local)
Chile				
Tiendas por departamento	1.534.916	1.525.908	-0,6	-0,6
Mejoramiento del hogar	2.058.298	2.169.506	5,4	5,4
Supermercados	738.661	776.142	5,1	5,1
Perú				
Tiendas por departamento	580.860	605.025	4,2	5,8
Mejoramiento del hogar	601.184	615.901	2,4	4,1
Supermercados	702.487	773.460	10,1	13,3
Colombia				
Tiendas por departamento	356.894	402.257	12,7	8,5
Mejoramiento del hogar	751.542	783.372	4,2	3,8
Argentina				
Tiendas por departamento	319.043	241.362	-24,3	34,3
Mejoramiento del hogar	170.958	125.248	-26,7	27,5
Brasil				
Mejoramiento del hogar	194.937	182.790	-6,2	10,8
Ventas online	473.189	598.091	26,4	

Tabla 1: Ingresos anuales 2017 y 2018 para retail en países de presencia.
Fuente: Elaboración propia (extraída de Memoria anual Empresa 2018).

En el ámbito nacional, para tiendas por departamento a diciembre del año 2018, la empresa posee 47 locales que comprenden una superficie total de venta de 338.698 m².

1.2.6 Ventajas competitivas

La empresa, desde su inauguración en Chile, se ha mantenido como líder en su categoría y durante las últimas décadas ha transferido este modelo y metodología a diversos países de Latinoamérica. Este liderazgo, se ha perpetuado en virtud de ventajas competitivas que le han permitido seguir a la vanguardia de la dinámica comercial de los últimos años. Entre las ventajas competitivas que se desprenden de las afirmaciones de sus propios dueños, se encuentran, la sinergia creada a través de la omnicanalidad para sus formatos físico y digital. Una fuerte inversión en digitalización y logística, potenciado por un conocimiento y entendimiento de los clientes, que se recopila a través del registro de interacciones y transacciones que realizan los clientes en cada una de las plataformas a su disposición.

1.3 Mercado y marco institucional

La empresa se encuentra inserta en la industria del comercio al por menor a través de su tiendas físicas y digitales, donde vende una amplia gama de productos tanto de uso personal como para mejoramiento del hogar.

La empresa en general apunta a múltiples clientes, tanto hombres como mujeres, de diversa edad, necesidad y preferencias. No obstante, y en particular para la unidad de tiendas por departamento, centra su mayor atención en atraer como cliente predilecto a mujeres, pues se tiene el supuesto que es la integrante de la familia que posee la última palabra en la decisión de compra para todo el grupo familiar.

El compromiso de la empresa para con el cliente refiere a aspectos claves como la privacidad, la seguridad, el manejo de la información y la transparencia, que son parte fundamental de las guías y normas que rigen a la organización. Para ello, y cumpliendo con la normativa y regulación de la SVS a través de la Norma de Carácter General N°385 acordada el 2015, se realizan capacitaciones anuales para el 100% de los colaboradores. Asimismo, la empresa busca la retroalimentación del cliente a través de mecanismos como el Net Promoter Score (NPS), entre otros.

El objetivo principal es entregar la mejor experiencia de compra, procurando asegurar la calidad y seguridad de los productos, servicios e instalaciones de la compañía. Por otro lado, siguiendo el dinamismo y digitalización de sus clientes, ha apostado por una estrategia de omnicanalidad y desarrollo tecnológico continuo complementando tanto los canales físicos como digitales. Este último experimentó un alza de 36% durante el 2017, alcanzando los \$472.534 millones en venta.

Es importante mencionar, de acuerdo con la creciente migración digital que ha experimentado la empresa y el retail en general, las nuevas políticas de privacidad en cuanto al uso y tratamiento de datos personales. Leyes como el Reglamento General de Protección de Datos (GDPR en inglés) publicada en 2016 en la Unión Europea o la posible modificación a la ley 19.628 (vigente desde 1999) en Chile durante 2019 en respuesta a la promulgación de la Ley 21.096 en la Constitución durante el 2018 que instaura el derecho a la protección de datos personales como derecho fundamental. Estas leyes regulan el derecho al uso y explotación de la información personal y que, en ámbitos de este proyecto, podrían constituir riesgos futuros para la implementación de herramientas y/o medidas que pretendan predecir el comportamiento de clientes haciendo uso de datos personales. Lo anterior, puesto que, instaura restricciones a su almacenamiento, explotación, transferencia y uso, como también entrega derecho a la persona, para pedir antecedentes acerca de información que pudiera estar recopilando una empresa sin su expreso consentimiento y que obligaría a las empresas a ser mucho más rigurosas en este sentido [8] [9].

Dentro de su modelo de negocios, la empresa posee dos formas de atraer a sus clientes. En primer lugar, se tiene la opción de adquirir los derechos de venta de marcas internacionales exclusivas de alta relevancia y contingentes con respecto a las tendencias. Además, la empresa posee derechos compartidos (entendiendo esta denominación cuando existe otro competidor en el mercado que posee la misma marca a la venta) para otros. Por último, existe otra alternativa, que durante los últimos años se ha vuelto más relevante, la que corresponde al desarrollo de diversas marcas propias que apuntan a diferentes segmentos de clientes y que son manufacturadas en empresas propias.

Algunas cifras adicionales relevantes son que el negocio del retail registró más de 300 millones de transacciones y cerca de 400 millones de visitas al año (2018) en sus centros comerciales.

Con respecto a los proveedores, se tiene un trabajo mutuo constante a través de la confianza, la colaboración, y la comunicación abierta a través de múltiples canales. Los

datos entregados para el periodo 2018 muestran un total de 27.855 proveedores activos, un 11% menos que el 2017, de los cuales ninguno supera por sí sólo el 10% de las ventas totales realizadas. Por otro lado, los negocios del retail trabajan con Sedex, la mayor plataforma mundial en temas de Responsabilidad Social en la Cadena de Valor, y, además, un 68% y 30% de los contratos posee cláusula de Responsabilidad Social y medioambiental respectivamente, donde 700 fueron auditados durante el 2018 en prácticas de RSE (Responsabilidad Social Empresarial). Estas medidas permiten de alguna u otra forma garantizar que los proveedores también sigan los lineamientos de la empresa, en sus aspectos ético, legal y socialmente responsable.

Cabe destacar, la importancia que ha cobrado el medioambiente durante la última década, La empresa, en su compromiso de contribuir más allá de su negocio ha instaurado herramientas y buenas prácticas que han sido reconocidas en reiteradas ocasiones por diferentes organizaciones internacionales. Asimismo, ha impulsado iniciativas que apuntan a la integración de nuevas regulaciones, como la Ley REP (Responsabilidad Extendida del Productor), y tendencias, como la Economía Circular, a través de la recuperación de materiales antes considerados y tratados como desechos.

Finalmente, La empresa, en su unidad de negocios de tiendas por departamento, se ha caracterizado desde su origen por estar a la vanguardia en muchos aspectos del mercado, lo que le ha permitido mantener una posición de liderazgo en la región durante los últimos años. Su posicionamiento de mercado en Chile, Perú, Colombia y Argentina en este aspecto es de liderazgo. Cabe destacar que también posee posición de liderazgo en otras unidades de negocio, como, por ejemplo, Mejoramiento del Hogar en Chile, Perú y Colombia.

1.4 Desempeño organizacional

La empresa actualmente se encuentra en una etapa de crecimiento, a través de diferentes iniciativas de inversión y robustecimiento que pretenden consolidar su posicionamiento de mercado en la región, en vista de las nuevas tendencias y el dinamismo en el comportamiento de compra de los clientes. Estas iniciativas encuentran su justificación en el crecimiento y transferencia digital, el uso de inteligencia artificial y la automatización, así como el mejoramiento en los procesos logísticos y de la cadena de suministro.

De acuerdo a lo anterior, las principales inversiones están orientadas al desarrollo en tecnologías de la información y logística que busca seguir avanzando en el desarrollo de un ecosistema digital de servicios. Se contempla para este propósito un plan de inversiones de US\$ 4.200 millones para el periodo 2019-2022, donde un 37% corresponde a Tecnologías de la información (TI) y logística, un 34% a la apertura de nuevas tiendas y centros comerciales y un 29% a remodelación y ampliación de tiendas y centros existentes.

Durante el primer trimestre del 2019, La empresa reportó utilidades por US\$ 104 millones, un 38% menos que el mismo periodo del año anterior y un EBITDA de US\$ 412 millones, disminuyendo un 6% con respecto al 2018. Por otro lado, los ingresos experimentaron un alza relativa de un 3% con respecto al año anterior, alcanzando los

US\$ 3.319 millones. Lo anterior se debe a la contracción de la demanda por parte de clientes extranjeros comprando en el país y a las inversiones realizadas para potenciar el ecosistema físico y digital de la compañía. Se suma a lo anterior la adquisición de Linio, la mejora en las aplicaciones móviles, la expansión del click & collect y las facilidades de entrega al cliente [10].

En palabras de su Gerente General Gaston Bottazzini, esta inversión en crecimiento apuesta a que “La empresa está haciendo hoy las inversiones que sustentarán el crecimiento futuro de la compañía. Estamos construyendo un ecosistema de servicios integrados, que nos permita servir a nuestros clientes en toda la región cuándo quieran y donde quieran.” Se desprende entonces, la marcada dirección y sentido por la omnicanalidad, entregando al cliente un sistema integrado que se adapte a los cambios en su comportamiento y dinamismo en el seguimiento de tendencias, así como también a los cambios y dinanismos globales.

2. PLANTEAMIENTO Y DESCRIPCIÓN DEL PROYECTO

2.1 Información del área de la organización

El área en la cual será desarrollada la memoria de trabajo de título corresponde a la Gerencia de Inteligencia de Clientes, extensión de la Gerencia de Negocios que responde ante la Gerencia General y que, a la vez, forma parte del *holding*. Esta área se encarga de generar oportunidades de negocio a través de la administración, análisis y tratamiento de datos para la toma de decisiones a través de la segmentación de los clientes de la empresa.

Cabe destacar el trabajo coordinado con el área de Inteligencia Web para la incorporación de nuevas funcionalidades en el sitio web, la automatización del envío de emails, entrega de información a clientes, gestión de campañas de marketing y la mejora de los sistemas de soporte tecnológico. Asimismo, el área también posee injerencia en el diseño de las piezas publicitarias de las campañas y provee información que sustenta la toma de decisiones comerciales.

El área está compuesta por: (1) Gerente de Inteligencia de Negocios, (1) Ingeniero Jefe de Análisis de segmentos, (1) Ingeniero Jefe de Campañas, (4) Ingenieros de Inteligencia de Negocios y (1) Ingeniero Analista de Inteligencia de Negocios. El detalle de la representación gráfica de la jerarquía organizacional del área se observa en la Figura 5.

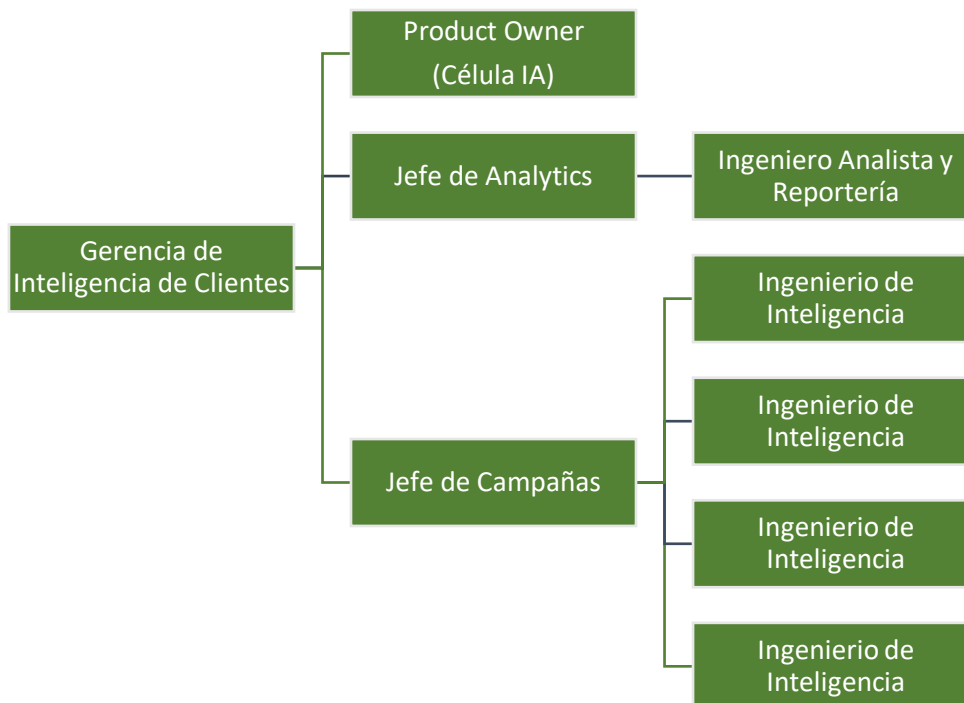


Figura 5: Diagrama organizacional área Inteligencia de clientes.
Fuente: Elaboración propia.

Entre las principales actividades de los integrantes del área se encuentra la administración de cuatro a cinco categorías de los productos comercializados por La empresa, de modo que deben realizar estudios para la identificación de oportunidades de negocio, segmentación de clientes y el apoyo a la planificación comercial en función de las directrices de la industria y las preferencias de los consumidores. Por otro lado, existe una extensión del área, que consiste en una célula que se encarga de generar alternativas a través del uso de aprendizaje de máquinas e inteligencia artificial para la segmentación de clientes, predicción de compra y la generación de mecanismos de retención de clientes. Esta última se encuentra a cargo de un (1) Product Owner.

La persona que demanda este trabajo de título es la Gerenta de Inteligencia de Clientes. Por último, es importante mencionar que el impacto de este trabajo de memoria de título irá en directo beneficio del área anteriormente mencionada, puesto que consiste en un proyecto integral impulsado por esta misma área, no obstante, podrían existir beneficios indirectos y relativos con relación a la utilidad que pueda tener para otras áreas, como Inteligencia Web o Marketing.

2.2 Contexto del trabajo en el área

Tal como fue mencionado en la sección anterior, dentro de las labores que se llevan a cabo en el área de Inteligencia de clientes existe el análisis y segmentación de clientes en virtud de su comportamiento histórico transaccional e información adicional recopilada por la empresa. Lo anterior responde al requerimiento específico de la gerencia comercial de realizar diversos envíos diarios, a través de email marketing, tanto de información de eventos y contenido en categorías de productos, como también, de ofertas y descuentos.

Sigue que, para cumplir el objetivo de incentivar la compra de los productos ofertados, la búsqueda de clientes óptimos se lleva a cabo siguiendo la regla de seleccionar clientes con buen comportamiento de compra o haber comprado productos similares en periodos recientes. Sin embargo, el uso de esta metodología implica que la gran mayoría de los clientes elegidos, son clientes que evidencian alta actividad de compra en la empresa y que generalmente poseen mayor valor. Asimismo, para gestionar distintos tipos de clientes, se realizan envíos genéricos de categorías que por naturaleza poseen alto volumen de compra o mayor margen entre todos los clientes.

En la empresa, el cliente pasa por diversos estados dependiendo de la actividad de compra que posea. En primer lugar, se tiene la clasificación de mayor volumen, que clasifica a los clientes en tres estados: activo, inactivo y fugado. El detalle de este tipo de estados se observa en la Tabla 2.

Estado de cliente	Descripción
Activo	Cliente que ha comprado al menos una categoría durante los últimos 3 meses
Inactivo	Cliente que ha comprado en al menos una categoría entre el periodo de 3 meses y 1 año
Fugado	Cliente que no ha comprado en más de 1 año

Tabla 2: Estado general de clientes en tienda por departamentos.

Fuente: Elaboración propia.

En particular, dentro de los clientes activos, se tienen clientes que poseen actividad en múltiples categorías de productos y que son los clientes que en su mayoría son elegidos para incentivar la compra de un tipo de producto o marca específica, pues se espera que su respuesta al incentivo genere una compra.

Por otro lado, se tienen clientes que son sólo intensivos en una categoría y que son clientes a los que se pretende incentivar una compra en otras categorías que no haya comprado (objetivo que se identifica como compra cruzada de categorías).

Por último, se tienen los clientes inactivos, de los cuales se desprenden dos tipos. El primero corresponde a clientes que son completamente inactivos, es decir, no compran ninguna categoría en el periodo de tres meses y un año. Y, el segundo, que corresponde a clientes que son inactivos en una categoría, y que pueden o no, haber comprado en otras categorías distintas a la categoría inactiva. El tipo de cliente que se describe en este último párrafo corresponde, en efecto, al cliente del cual se desprende e identifica la oportunidad que será descrita en la siguiente sección y que deriva en este trabajo de memoria de título.

2.3 Identificación del problema u oportunidad

La oportunidad identificada pretende apoyar, sustentar y/o refutar las decisiones comerciales acerca de a qué clientes es mejor contactar para las campañas u ofertas que requieran la activación de compra de un cliente en una categoría de productos de la empresa. Para ello, se propone como herramienta útil la construcción de un modelo numérico que permita seleccionar la mejor próxima categoría de producto en la que se activará un cliente (inactivo en categorías) y la relación que existe en el comportamiento de compra previo de otras categorías en la activación. Cabe mencionar que no abordar y resolver este problema tendría como consecuencia no llegar a conocer las diferencias que existen entre este tipo de clientes y los otros de la empresa, y que las acciones de marketing sigan ejecutándose sin considerar las características específicas de cada tipo de cliente.

Este problema u oportunidad tiene su origen en la falta de un sistema que administre el ciclo de vida de un cliente mediante un único flujo de gestión de proceso, que distinga las distintas campañas que son utilizadas para aumentar la frecuencia de compra y retención en clientes. En particular, el presente estudio sólo centra su atención en mejorar la interacción en clientes con categorías inactivas y su activación, no obstante, en el flujo como un todo, existen otros tipos de clientes en diferentes estados, como lo son: los clientes fugados, propensos a fuga, clientes de alto valor, entre otros.

2.4 Estado del arte

La gestión de cliente actual se administra en base al criterio de experto, criterio comercial y métricas de frecuencia de compra. Lo anterior, de acuerdo con las decisiones que se toman en cuanto a requisitos y exigencias de la empresa, permite filtrar la base

total de clientes para quedarse con clientes que presentan un buen comportamiento de compra (alta frecuencia) o un patrón³ de compra evidente.

Estas bases de clientes son extraídas día a día, dependiendo de la necesidad de despacho de correos de email marketing con un propósito particular (comunicación de información, campaña u oferta).

Los correos de email marketing que comúnmente son enviados, la política de email marketing y su criterio de despliegue se detallan en la Tabla 3 a continuación:

Email marketing	Detalle
Tipos de correo	<ul style="list-style-type: none"> • Trigger: Tiene por propósito aumentar la frecuencia de compra a través de sugerencias de productos que el cliente suele comprar. • Contenido: Para comunicar información de marcas o eventos especiales (Cyber, navidad, HOT sales, etc.). • Cupón: Se ofrecen diversos niveles de descuento en categorías de producto. • Monetizado: Información/descuento a petición de marcas, que pagan por el servicio de envío de correos.
Política de envíos	<ul style="list-style-type: none"> • Aproximadamente 1 millón de envíos diarios. • Sólo un envío por persona diario. • Estrategias específicas para clientes fugados, inactivos o de alto valor tienen restricción de 1 envío por semana.

Tabla 3: Detalle de características de correos de email marketing.
Fuente: Elaboración propia.

2.5 Indicadores de relevancia, causas y efectos

2.5.1 Indicadores de relevancia

Viene de la observación del comportamiento de diversos indicadores de rendimiento (KPI's) que consideran fidelización, valor del tiempo de vida del cliente, la venta incremental (que consiste en cuan útil o rentable fue la intervención de La empresa en el comportamiento de compra del cliente), gasto esperado del cliente, entre otros. Estos indicadores se observan para medir la efectividad de las campañas y se evalúan para cada una de ellas.

De acuerdo a lo anterior y al contexto que componen las labores llevadas a cabo en el área, con el objetivo de promover e incentivar la propensión y comportamiento de compra de los clientes, es posible identificar tanto potenciales causas como efectos que se desprenden de este problema.

³ Un patrón de compra evidente refiere a que sea posible identificar de manera simple (exploración de datos) la mejor oferta para un tipo de cliente sin necesidad de aplicar métodos de mayor complejidad (modelos de predicción).

2.5.2 Causas

- Dinamismo en el comportamiento de compra de clientes, haciendo casi inevitable un aumento en el volumen y tipo de interacción que incentiven compras.
- Integración de múltiples canales de compra, haciendo imprescindible desarrollar diversos mecanismos de interacción con el cliente.
- Aumento del flujo flotante en la plataforma online y aumento en el volumen de transacciones, tanto del canal físico como el virtual.
- Inexistencia de lógica sistémica en el orden de envío de ofertas que responda a distintas categorías de productos y distintos tipos de clientes.

2.5.3 Efectos

- Comprensión más detallada de los segmentos de clientes y su ciclo de compra.
- Mejora en la administración de clientes, entendiendo su comportamiento de compra histórica y necesidades.
- Optimización de envío de ofertas de categorías de producto, que estén estrechamente relacionadas con el comportamiento de compra previo del cliente.

2.6 Hipótesis y solución propuesta

De acuerdo con los antecedentes proporcionados, se identifican las siguientes hipótesis:

- Hipótesis 1: La gestión de la cartera de clientes se ve afectada, aumentando la fuga y disminuyendo la fidelización, porque no existe una priorización de ofertas que guarde relación con su comportamiento de compra previo.
- Hipótesis 2: El aumento en la activación de distintas categorías de productos, aumenta el gasto de cliente.
- Hipótesis 3: El estado en que se encuentra el ciclo de vida del cliente (activo, inactivo, fugado) influyen en el comportamiento de compra. Asimismo, lo hace su antigüedad con la empresa y el canal por el cual lleva a cabo sus compras, entre otras características relevantes.

Solución propuesta: Evaluar el efecto que tiene la activación de una categoría en el gasto del cliente y generar un modelo predictivo que permita predecir la próxima categoría de producto en que se activará, su relación con la compra en otras categorías de productos y la importancia de las características propias de cada cliente.

2.7 Trabajos previos

Con el propósito de justificar el objetivo de la memoria, se detallan los trabajos previos realizados en la Gerencia que lo demanda y además se incluye el trabajo de memoria previo realizado a nivel corporativo del *holding*.

Evaluación de la efectividad de promociones personalizadas en la retención y aumento de los clientes de alto valor para el año siguiente.

(Riquelme, Francisca; 2017)

Evaluación de promociones en la retención de un segmento de clientes para el año siguiente.

(Cardoza, Luis; 2018)

Con respecto a los trabajos mencionados anteriormente se extrae que estos modelos y resultados corresponden a especificaciones que agrupan la totalidad de los datos y que pretenden entregar información para el siguiente año.

En cuanto a los hallazgos más importantes del primer trabajo mencionado se tiene que el envío de correos de marketing tiene un efecto significativo en la propensión a ser cliente de alto valor y que esta propensión posee alta variación entre diferentes tipos de clientes y su actividad en compras. En relación con el segundo, entre los principales hallazgos se encuentran variables importantes en la predicción de fuga, como lo son, las compras durante el último año, la antigüedad del cliente y el *recency*. Y, por otro lado, que las ofertas asociadas a montos de descuento poseen mayor significancia que los descuentos porcentuales.

Dado el requerimiento de acortar los periodos de análisis, es necesario un enfoque más específico para tratar el comportamiento de los clientes, ya sea evaluando mes a mes o tomar en consideración como se comportan entre sí distintos tipos de productos para distintos tipos de clientes.

Evaluación de la efectividad de un sistema de recomendación de productos.

(Rubilar, Guillermo; 2019)

Con respecto al trabajo anterior, es posible afirmar que se hace cargo del requerimiento de agilizar los periodos de predicción de manera eficiente, donde, dada su especificación, se entrega una recomendación basada en la frecuencia de compra anterior del cliente y a la estacionalidad del mes en el cual está prediciendo.

Además, se encuentra que, el modelo seleccionado entrega mejores resultados que el criterio de juicio experto, donde, se proponen diversos canales de interacción con el cliente que pudiesen tener un efecto significativo en la mejora del comportamiento de compra.

Cabe destacar que, este trabajo se limitó a sólo un par de conjuntos de categorías (Vestuario y Calzado) y no se considera el efecto que la compra de la una afecta a otras.

2.8 Objetivos

Tomando en consideración lo expuesto en el sub capítulo anterior, se plantean los siguientes objetivos para el trabajo de memoria.

2.8.1 Objetivo General

Evaluar el efecto de la activación de clientes en diferentes categorías de producto inactivas en su gasto posterior y el efecto cruzado que existe entre la compra en otras categorías y la activación, de tal forma que permita predecir la mejor próxima categoría a ofrecer y mejorar el *targeting* de clientes para acciones de email marketing.

2.8.2 Objetivos Específicos

- Determinar la relación (inferencia causal) entre la activación de distintas categorías de productos y el aumento/disminución en el gasto del cliente.
- Diseñar y construir un modelo que permita predecir la próxima categoría de producto en la que se activará un cliente.
- Definir un criterio de asignación de categoría para cada cliente, que permita balancear su probabilidad de activación y el gasto posterior (3 meses) a la activación.
- Determinar un criterio de asignación de relevancia de otras categorías en la activación de la categoría analizada, que permita priorizar la segmentación de clientes para ofertas de email marketing.

2.9 Resultados esperados

En relación con los requisitos y estándares que se esperan del trabajo realizado, expuesto en esta memoria de título, se desprenden del desarrollo ordenado de los objetivos antes planteados, los resultados esperados que se exponen a continuación de manera detallada.

- Cuantificar la inferencia causal entre la activación de una categoría específica y el aumento en el gasto del cliente.
- Encontrar, a través del modelo estadístico seleccionado, la mejor próxima categoría de producto, basado en la propensión de compra y gasto esperado, que permita gatillar la activación de clientes de la empresa.
- Determinar explícitamente las categorías de productos que tienen mayor importancia en la activación de ciertas categorías de producto.

- Cuantificar la validez y utilidad del modelo planteado de cara al uso potencial para el cliente del proyecto (la empresa).

2.10 Alcances

Considerando las características de este trabajo de título, los requerimientos y limitaciones pretendidos por la empresa y el negocio, las capacidades y tiempo de ejecución disponible para llevarlo a cabo, se definen los siguientes alcances que determinarán la profundidad de los resultados del proyecto.

- Se utilizarán los datos de clientes disponibles e identificables actualmente en la empresa y que cuenten con al menos, una transacción en la organización durante el periodo de tiempo analizado. Por tanto, no se contemplará para ello crear una nueva base de datos con clientes potenciales o realizar experimentos o encuestas para recopilar mayor información.
- El trabajo de título no contemplará la creación de nuevas campañas para los clientes, sino más bien permitirá optimizar el proceso de segmentación de clientes para acciones de email marketing.
- El desarrollo de este proyecto no contemplará un pilotaje o implementación de la solución propuesta, sino más bien sentará las bases y la metodología para su aplicación. La implementación y proceso de experimentación se deja como una posible continuación de este.
- Esta memoria no se hará cargo de la pertinencia de la segmentación de categorías existente, ni el criterio que se utiliza para estipular la ventana de tiempo en que una categoría pasa de estar activa a inactiva, sino que serán utilizadas en su estado actual como un insumo para cumplir el objetivo del proyecto.

3. MARCO CONCEPTUAL

El siguiente capítulo tiene por objeto declarar y detallar los diversos conceptos y disciplinas que intervienen en el desarrollo de la solución propuesta y que sustentan la ejecución de la memoria de título.

3.1 Mejor próxima oferta

El asunto principal que da forma al tema de memoria y que representa una oportunidad de aumentar el gasto del cliente y por consecuencia el posible aumento en la venta incremental para la empresa es saber qué tipo de productos podrían ser ofrecidos al cliente que gatillen su próxima compra.

Lo anterior refiere al sistema “Next Best Action/Offer/Product” [11] [12] que viene a responder a la pregunta planteada, ya que mide la propensión de compra del cliente en base a un agrupado de alternativas con diferente probabilidad de ocurrencia, donde esta ocurrencia, se relaciona con la información histórica del segmento de cliente al cual pertenece.

Para ello existen múltiples modelos de diversa complejidad que abordan el problema desde distintos enfoques. Entre los más utilizados se encuentran los modelos de elección binaria, como el Logit, utilizado por (Knott, Hayes & Neslin, 2002) [13] que asigna la elección de determinados productos mediante la probabilidad que posee cada cliente a comprarlo.

Por otro lado, existen modelos de mayor complejidad y sofisticación, como lo es el modelo Probit multivariado utilizado por (Manchanda, Ansari & Gupta, 1999) [31] y (Vargas, 2003) [14] con el objetivo de estimar variables dependientes binarias que estén correlacionadas entre sí.

El sistema planteado considera el estudio de datos históricos de los consumidores a través de modelos estadísticos/económicos que permitan pronosticar el comportamiento de compra de un cliente cualquiera. Para ello existen diversas formas de abordar el problema según las características o especificaciones que se requieran.

Dentro del análisis basado en minería de datos se tienen diversas posibilidades de elaboración de modelos, cada uno con diferentes capacidades de procesamiento, alcance y profundidad, que dependen de las características propias del problema subyacente. Por tanto, se especifican las diferentes herramientas que se contemplan para abordar el problema y que fueron seleccionadas en virtud de las características de los datos, los recursos de procesamiento y la complejidad intrínseca de su elaboración y atinencia al problema.

3.1.1 Logit

El modelo Logit, es un modelo de elección del tipo binaria que asigna probabilidad a los clientes de escoger ciertos productos, y, en virtud de estos resultados, crear reglas de asignación para las ofertas de mayor pertinencia que pudiesen incentivar una compra. Sus usos son variados, como es el caso visto por (Knott, Hayes & Neslin, 2002) que apunta a predecir la siguiente compra del cliente con el objetivo de resolver problemas de venta cruzada.

Por otro lado, se tienen Logits más sofisticados para resultados de una variable objetivo categórica con más de dos niveles, como el utilizado por (Prinzie & Van den Poel, 2008) [15] sobre el uso de Logit multinomial mediante un algoritmo de clasificación generado a través del uso de Random Forest, (Adams & Wilson, 1996) sobre el uso de Logit multinomial con coeficientes aleatorios o (Adams; Wilson & Wen-chung, 1997) [16] y sobre el uso de Logit multinomial mediante la creación de coeficientes aleatorios a través del modelo de Rasch multidimensional [17].

Formalmente, se tiene para un modelo Logit de elección discreta, donde se desea explicar el comportamiento de una variable dependiente no continua. Lo que se intenta modelar es la probabilidad que un individuo cualquiera elija una alternativa específica dentro de un grupo de opciones, suponiendo que un cliente racional elegirá la opción que maximice su utilidad. Por tanto, la utilidad (u_{ni}) del cliente se define en (2), como:

$$u_{ni} = v_{ni} + \varepsilon_{ni} \quad (2)$$

Donde (v_{ni}) corresponde a la componente determinística y (ε_{ni}) a la componente estocástica. De lo cual el modelo Logit (Goic, 2015) [18], asume que la componente estocástica distribuye valor extremo tipo I o Gumbel, y que tiene una probabilidad de un cliente i sobre una elección n , definida en (3), como sigue:

$$P_{ni} = \frac{e^{v_{ni}}}{\sum_j e^{v_{nj}}} \quad (3)$$

3.2 Inferencia Causal

La inferencia causal refiere a una forma de estimar los efectos causales de experimentos o, para efectos de este estudio, a través de datos observacionales. Para ello lo que se pretende es replicar un experimento aleatorio, donde, tanto el grupo tratado como el grupo control posean distribuciones similares en sus variables explicativas.

Entonces, se tiene que el efecto causal de ser expuesto a un tratamiento (τ_i) se expresa en (4), como la diferencia entre un individuo en tratamiento contra un individuo no tratado, sigue que:

$$\tau_i = Y_i(1) - Y_i(0) \quad (4)$$

Donde, $Y_i(1)$ corresponde a un individuo i que está expuesto al tratamiento (grupo tratado) y $Y_i(0)$ corresponde a un individuo que no está expuesto al tratamiento (grupo control).

De lo expuesto anteriormente surge un inconveniente, y es que, dada la expresión para la inferencia causal, sólo es posible observar uno de los resultados para un individuo, dado que es o no es tratado. Sin embargo, una posible solución corresponde a comparar un grupo tratado y un grupo control que sean lo más similares posibles. Para ello, los experimentos aleatorios aseguran que las distribuciones para ambos grupos se encuentren balanceadas previo a encontrarse o no, en tratamiento.

Con el objetivo de comparar los efectos de un tratamiento específico existen métricas que permiten evaluar dicho tratamiento en los tratados, donde la más relevante corresponde al ATT (Average Treatment effect on the Treated) como se muestra en (5):

$$\tau_{ATT} = E[\tau|D = 1] = E[Y(1)|D = 1] - E[Y(0)|D = 1] \quad (5)$$

Donde, $D = 1$ refiere a que hubo tratamiento. Al igual que antes, se tiene que la expresión dada por $E[Y(0)|D = 1]$ no es observable, por lo que se debe escoger un estimador para este. No obstante, dada las características de balance para cada grupo, sólo deberían existir variaciones dado el sesgo de selección. Por tanto, es posible asumir que el efecto de no ser sometido a tratamiento para el grupo de control es un buen estimador de $E[Y(0)|D = 1]$, es decir, de no ser tratado dado que si fue intervenido (Stuart, 2010) [19].

Luego, dado que la asignación aleatoria no puede ser aplicada posterior al tratamiento, se han elaborado diversos métodos de emparejamiento que facilitan la construcción de estos grupos de control y permiten evaluar los efectos de un tratamiento.

3.2.1 Métodos de emparejamiento

Con el objetivo de elaborar el emparejamiento se sigue la metodología propuesta en (Stuart, 2010), que consta de cuatros pasos clave:

- i. **Definir “cercanía”:** determina la distancia que se usa para determinar si un individuo es un buen emparejamiento para otro.
- ii. **Implementación:** se lleva a cabo el emparejamiento de acuerdo con la distancia determinada en el paso anterior.
- iii. **Evaluación:** se evalúa la calidad de los emparejamientos realizados, desde aquí se puede volver de manera iterativa a los puntos anteriores hasta lograr un buen emparejamiento.
- iv. **Análisis:** se realiza un análisis de los resultados obtenidos y se estima el efecto del tratamiento, dado el emparejamiento en el paso (iii).

3.2.1.1 Medidas de cercanía

En lo que refiere a una medida de distancia D_{ij} que capture las similitudes entre un individuo i y otro individuo j , se tienen diversas formas, tales como:

- Distancia exacta: se define como cero la distancia para dos elementos iguales e infinito para elementos distintos.

$$D_{ij} = \begin{cases} 0 & \text{si } X_i = X_j \\ \infty & \text{si } X_i \neq X_j \end{cases} \quad (6)$$

- Distancia de Mahalanobis: se usa para casos en que se quiere calcular la distancia entre variables aleatorias multidimensionales, esto pues, toma en consideración la correlación de las variables aleatorias X_i y X_j mediante su matriz de varianza covarianza Σ , como se muestra en (7).

$$D_{ij} = (X_i - X_j)' \cdot \Sigma^{-1} \cdot (X_i - X_j) \quad (7)$$

- Propensity score: Definido por (Rosenbaum & Rubin, 1983) [20] el propensity es una medida de distancia que captura totalmente las características de las variables X_i aleatorias para un individuo cualquiera i , en un único escalar e_i , donde $e_i(X_i) = P[D = 1|X_i]$, que refiere a la probabilidad de ser tratado sujeto a las variables aleatorias en estudio. Por tanto, se define la distancia como sigue:

$$D_{ij} = |e_i - e_j| \quad (8)$$

Asimismo, se tiene que, existe una extensión del propensity score que pretende reducir el sesgo de selección a través de una definición lineal del mismo mediante regresiones logísticas, como se muestra en (9):

$$D_{ij} = |\text{logit}(e_i) - \text{logit}(e_j)| \quad (9)$$

3.2.1.2 Algoritmos para estimación de emparejamiento

Para el caso en que el propensity score es la medida de distancia elegida, se tienen diversas formas de estimar el emparejamiento, entre ellas, el Stratification Matching, Kernel Matching, Nearest-Neighbor Matching y Radius Matching. Entre los más comunes y de mayor facilidad de implementación se encuentran los dos últimos mencionados, por lo que se procede a describir ambos en mayor detalle.

- Nearest-Neighbor Matching

Este mecanismo consiste en encontrar para cada individuo tratado, el o los individuos no tratados con el propensity score más semejante, es decir, la menor distancia. Este mecanismo se puede abordar de dos formas, con y sin reemplazo, donde

el reemplazo consiste en dar la libertad para que un individuo perteneciente al grupo control pueda ser emparejado a más de un individuo del grupo tratado. La ventaja que presenta esto último es que reduce el sesgo en los resultados y, además, el orden en que un individuo tratado es emparejado no importa (Stuart, 2010). Se tiene entonces de la especificación dada por (Becker & Ichino, 2002) [21] que para $C(i)$ el conjunto de individuos del grupo control emparejados con el individuo i del grupo tratado dado un propensity score de e_i se escribe como.

$$C(i) = \min_j \|e_i - e_j\| \quad (10)$$

- Radius Matching

Otro método que permite realizar el emparejamiento usando propensity score es el Radius Matching, que define un radio máximo para el propensity score para el cual los individuos del grupo control puedan ser emparejados al individuo tratado. Lo anterior implica que, para que un individuo del grupo control sea emparejado este debe tener una distancia menor a la definida por el radio máximo, por lo que se debe poner atención a la distancia que se fija para este, pues puede existir que individuos tratados no tengan ningún emparejamiento, no obstante, el uso de un radio menor implica una mayor calidad del emparejamiento (Becker & Ichino, 2002). Se tiene entonces para el conjunto de individuos no tratados dentro del radio máximo r cumplen:

$$C(i) = \{e_i \mid \|e_i - e_j\| < r\} \quad (11)$$

Sigue que, para el caso en que se quiera comparar la similitud entre el grupo tratado y grupo control, un punto importante es que las distribuciones de las variables explicativas estén balanceadas. Para ello, las distintas métricas de cercanía expuestas anteriormente pueden ser combinadas. En particular, para el caso en que se desea obtener balance en unas pocas covariables claves, y además, estas son continuas, se define la distancia de Mahalanobis dentro de *calipers* por *propensity score* (Rubin & Thomas, 2000) [22] entre los individuos i y j , como sigue:

$$D_{ij} = \begin{cases} (Z_i - Z_j)' \cdot \Sigma^{-1} \cdot (Z_i - Z_j) & \text{si } |\text{logit}(e_i) - \text{logit}(e_j)| \leq c \\ \infty & \text{si } |\text{logit}(e_i) - \text{logit}(e_j)| > c \end{cases} \quad (12)$$

Que corresponde a un emparejamiento por distancia de Mahalanobis para los individuos dentro del *caliper*. Donde c representa el *caliper* - que se define dentro de la diferencia lineal entre los propensity score -, Z es el vector de covariables claves y Σ es la matriz de varianza covarianza de Z .

La estimación del ATT tanto para el método de Nearest-Neighbor como el de Radius corresponde al promedio de las diferencias entre grupos (Becker & Ichino, 2002).

3.3 Métricas de evaluación

3.3.1 Matriz de confusión

La matriz de confusión es una representación gráfica de un tipo de métrica de evaluación de la asertividad de modelos o algoritmos de clasificación, comúnmente clasificadores del tipo binario. Una matriz de confusión genérica se observa en la siguiente Tabla 4:

		Valor de predicción	
		Positivo	Negativo
Valor Real	Positivo	Verdaderos positivos	Falsos negativos
	Negativo	Falsos positivos	Verdaderos negativos

Tabla 4: Matriz de confusión.
Fuente: Elaboración propia.

De acuerdo con lo expuesto, se desprenden una serie de métricas basadas en la combinación de los valores que se muestran en la tabla, tales como la precisión, exactitud, la sensibilidad y la especificidad. En particular, para el caso de estudio, donde se tiene una pequeña proporción de casos positivos se usa comúnmente según (Altman & Bland, 1994) [23], la sensibilidad y la especificidad con el objetivo de comparar modelos.

Sensibilidad (tasa de verdaderos positivos): Proporción de registros definidos como positivos que fueron efectivamente clasificados.

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos}}{\text{verdaderos positivos} + \text{falsos negativos}} \quad (13)$$

Especificidad (tasa de falsos positivos): Proporción de registros que son mal clasificados por el modelo como casos positivos.

$$\text{Especificidad} = \frac{\text{falsos positivos}}{\text{falsos positivos} + \text{verdaderos negativos}} \quad (14)$$

3.3.2 Curvas COR – AUC y de ganancia

La curva COR (Característica Operativa del Receptor) es una representación gráfica de la sensibilidad frente a la especificidad, entregada por la matriz de confusión, para evaluar un modelo de predicción del tipo binario. Por tanto, con ambas métricas se construye la curva COR para luego calcular el área bajo la curva (AUC en inglés), resultado que al obtener valores cada vez más cercanos a 100%, indican un aumento en el desempeño del modelo analizado. En la figura 6 se muestran diferentes comportamientos de clasificadores, donde, se define el área entre las curvas y la línea punteada inferior como el área bajo la curva (AUC) que define a la métrica con el mismo nombre. Para cada curva se tiene que, $AUC_{COR2} > AUC_{COR1} > AUC_{Aleatorio}$.

Asimismo, se tiene que una forma de comportamiento gráfico similar al COR – AUC es la curva de ganancias acumuladas, que, al representar buenos clasificadores, posee un comienzo con aumento brusco para luego descender su crecimiento en valores cercanos al 100%, como se observa en la Figura 6. Cabe destacar que, para una asignación aleatoria o un modelo que no entrega mayor información, la curva tendrá un comportamiento lineal como una diagonal que se traza del 0 al 100% como se muestra en color dorado en la Figura 6.

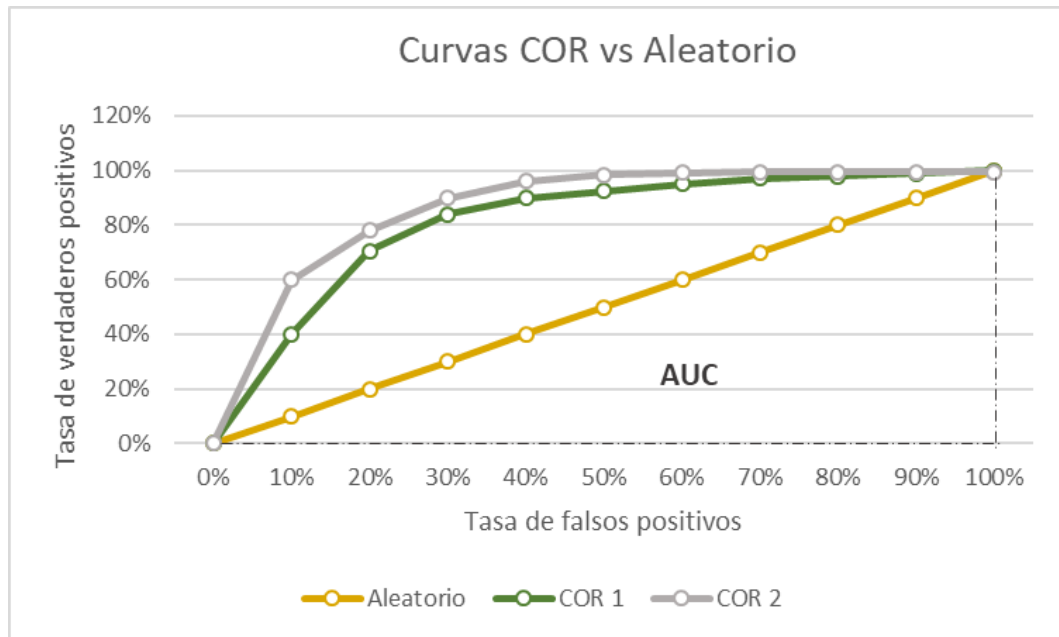


Figura 6: Representación de comportamiento de curvas ROC y clasificación aleatoria.
Fuente: Elaboración propia.

4. METODOLOGÍA

Para abordar este tipo de problemas que implica el análisis de grandes volúmenes de datos, sobre el comportamiento de compra o transaccional de un cliente con dependencia en el tiempo, se destacan metodologías basadas en la exploración de datos y el desarrollo de modelos. A partir de ello se contemplan tres metodologías que cumplen los requerimientos del proyecto, estos son, el método CRISP-DM [24], SEMMA y KDD [25]. Los trabajos realizados por (Shafique & Qaiser, 2014) y (Azevedo & Santos, 2008) [26] que comparan estas metodologías entre ellas, muestran que, a pesar que no existen diferencias tan significativas entre ellos, las metodologías más completas son CRISP-DM y KDD pues permitan desarrollar con mayor precisión y orden el procedimiento para resolver el problema.

Con el objetivo de llevar a cabo este trabajo de título de forma que su metodología propuesta sea abordada de la manera más rigurosa posible, se propone el uso del modelo *CRISP-DM* debido a su extenso y reiterado uso en proyectos que implican la utilización de *Data Mining* y a la característica que tiene de abordar o comprender “el negocio” como forma preliminar de abordar el problema planteado, cosa que no se presenta en las metodologías como SEMMA o KDD.

El modelo consta de 6 etapas: Comprensión del negocio y/o problema, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y finalmente, Implementación. A modo de ejemplo ilustrativo, el diagrama que se observa en la Figura 7, muestra el proceso iterativo de las etapas mencionadas anteriormente.

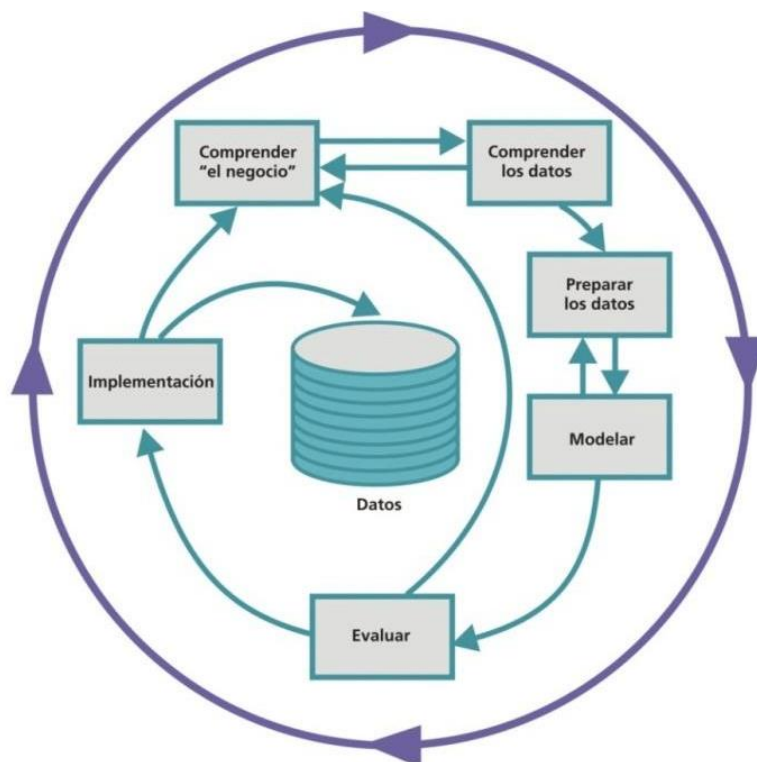


Figura 7: Cross Industry Standard Process for Data Mining (CRISP-DM).
Fuente: Medium.com.

Para efectos del desarrollo del trabajo se presenta entonces la secuencia de seguimiento de etapas mediante la metodología propuesta, como sigue.

- 1. Comprensión del negocio y/o problema:** Esta etapa inicial consiste en comprender la esencia del problema o proyecto abordado, su justificación, propósito y objetivos, desde la perspectiva del negocio. Esta etapa es abordada en profundidad en los primeros capítulos de la presente memoria. Así entonces, es posible fijar los objetivos del minado de los datos, cuáles son las variables, datos y herramientas que se tienen en consideración actualmente y de qué forma este problema podría ser abordado desde la Ingeniería Civil Industrial, tomando en consideración las particularidades y requerimientos que posea la empresa, para finalmente poder elaborar el plan de trabajo.

Se tiene entonces que, en relación con la información disponible, se aborda el presente capítulo mediante el uso de las Memorias Anuales de la empresa, la información pública disponible en medios digitales y la información propia recopilada durante el tiempo de duración del proyecto.

- 2. Comprensión de los datos:** El propósito de esta fase es explorar genéricamente la base de datos que permita extraer información relevante acerca de la misma. Comienza con una recopilación de los datos, para luego describir, explorar, y analizar la base de modo tal que permita identificar problemas en su calidad o comportamiento evidente que sea útil a priori para tener un acercamiento al conocimiento del registro transaccional de los clientes. Esta etapa se desarrolla mediante consultas a través de la herramienta de gestión y administración de datos PL SQL que permite la extracción de registros desde el *Data Warehouse*.

La empresa tiene a su disposición una gran cantidad de bases de datos que recopilan y consolidan todas las transacciones e interacciones realizadas tanto en las tiendas físicas como en la tienda online. En particular, se utilizará información transaccional, sociodemográfica y de navegación en su página web a nivel de cada cliente en el intervalo correspondiente a dos años móviles (2017 – 2019). Asimismo, se agregará la información corporativa, correspondiente a datos del negocio y uso de la tarjeta de crédito perteneciente al *holding* para realizar las transacciones.

- 3. Preparación de los datos:** El propósito de esta fase es elaborar una base de datos manejable y de buen estándar que permita desarrollar las etapas posteriores. Su desarrollo consiste en procedimientos como la selección, limpieza, construcción, integración y formato final de la base de datos que permitan utilizar la información contenida en la misma, de acuerdo con los modelos propuestos.

Para ello, se considera que la base puede contener registros que afecten la distribución de los datos para ciertas variables importantes por lo que se excluirán valores extremos o atípicos.

Por otro lado, por la forma en que se realizarán los distintos análisis se prevé existirá el problema de registros nulos (ceros), dado principalmente, por la diferenciación que se hará en la compra de diferentes categorías de productos y por

las características en el comportamiento de compra que posee el cliente escogido para el estudio.

- 4. Modelamiento:** Construcción y aplicación de los modelos previamente estudiados y seleccionados, que permitan evaluar su desempeño y revisar el ajuste de los parámetros. Para este propósito se hará uso la herramienta computacional de programación R.

Esta etapa consiste en una serie de pasos que concluyen con el resultado esperado final para el proyecto y sus implicaciones. Lo anterior se aborda en profundidad en la sección 4.1, 4.2 y 4.3 que se expone de manera posterior en la sección.

- 5. Evaluación:** Se evalúan los modelos aplicados anteriormente, a través de sus resultados, con el propósito de revisar el cumplimiento de los objetivos planteados y fijar acciones futuras para su ajuste, ya sean modificaciones a los datos o el mismo modelo aplicado.

La evaluación se llevará a cabo en conjunto con la fase de modelamiento de manera iterada y que pretende obtener un modelo óptimo que permita resolver el problema. Para ello se evaluarán en detalle los resultados y parámetros entregados por el modelo final con el objetivo de elaborar un entregable que permita formular acciones de marketing directo. Finalmente, serán evaluados los criterios de elección propuestos posteriormente en 4.3 con el fin de entregar diversas métricas a la empresa para mejorar la interacción con el cliente.

- 6. Despliegue:** Etapa final de recopilación y ordenamiento de los resultados obtenidos que permitan generar un reporte final y facilitar la revisión del trabajo, de manera sencilla y útil para el cliente. Lo anterior se llevará a cabo utilizando las facilidades que entrega el Software R para generar reportes tabulados.

Cabe destacar que, para el desarrollo de la memoria, en la etapa detallada como modelamiento se sigue una metodología específica que se muestra en la sección precedente.

4.1 Efecto de la activación de categorías de producto

Con el objetivo de cuantificar el efecto que tiene la activación de diferentes categorías de productos en el gasto del cliente se utiliza la técnica conocida como emparejamiento por distancia de Mahalanobis dentro de *calipers* por *propensity score*.

Para lograr lo anterior se sigue la secuencia de etapas que se pueden observar en la Figura 8 y que son propuestas por (Caliendo y Kopeinig, 2008) [27]. Esta secuencia será ejecutada para cada una de las 11 categorías de producto analizadas, siendo la activación de un cliente en una categoría, la variable objetivo definida como tratamiento.

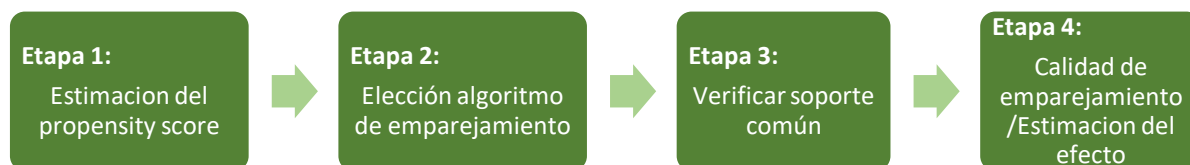


Figura 8: Etapas de implementación de emparejamiento por propensity score.
Fuente: extraído de (Caliendo & Kopeinig, 2008).

4.1.1 Etapa 1: Estimación del propensity score

Para la estimación del valor individual del propensity score se llevarán a cabo regresiones logísticas para cada una de las categorías de producto analizadas. La variable objetivo se define como la realización de una activación (compra) en la categoría durante un mes específico.

Asimismo, las variables explicativas corresponderán a conglomerados de datos que son posibles de segmentar en tres grupos principales: información transaccional y de interacción, que incluye la compra o no en las demás categorías de productos durante el mes anterior, la navegación en la categoría analizada, entre otras. Además, características sociodemográficas como género, edad, estado civil, monto de renta, u otros. Finalmente, características del cliente propias del negocio, como el tipo de tarjeta (de propiedad de la empresa) que posee, cupo o uso de cuotas.

Previo a la ejecución del modelo, como tal, para evitar problemas de correlación entre las variables explicativas, se verificará la correlación, mediante el coeficiente de correlación de Pearson, para variables de tipo continua y el coeficiente V de Cramer, para las del tipo categóricas.

Luego, se dará paso al cálculo de los modelos y la estimación de sus parámetros correspondientes generados mediante el método de máxima verosimilitud mediante la utilización del Software R. Por tanto, se generarán diferentes especificaciones para cada modelo a través de un proceso sistemático de *Stepwise*⁴, las que serán comparadas mediante el criterio de información de Bayes (BIC) sugerido por (Schwarz, 1978) [28]. Este criterio, como se observa en (15), tiene como propósito equilibrar la intención de obtener un modelo con mejor ajuste en máxima verosimilitud ($\ln(\hat{L})$) en contra de la intención de obtener modelos con una menor cantidad de parámetros (p) dependiendo del tamaño muestral (n). El resultado pretende escoger el modelo con la mayor probabilidad posterior, mediante la búsqueda del modelo con menor BIC, tal como expone (Raftery, 1985) [29]. Cabe mencionar que, para evitar inconvenientes de sobre ajuste de los modelos, se separa la muestra de clientes de manera aleatoria y balanceada, en un 80% para entrenamiento y un 20% para testeo.

$$\text{BIC} = -2 \cdot \ln(\hat{L}) + p \cdot \ln(n) \quad (15)$$

La especificación de los modelos de regresión logística que se estiman para llevar a cabo la técnica de emparejamiento son los mismos que serán utilizados como base de

⁴ El *Stepwise* refiere al proceso sistematizado de incluir y excluir variables explicativas de manera individual en el modelo para estimar y cuantificar el aporte específico de utilizar cada una de ellas.

comparación para la elección de la mejor próxima categoría (Next Best Product to Buy), por lo que su análisis predictivo y otras menciones, se detallan en la sección 4.2 de elección de mejor próxima categoría.

4.1.2 Etapa 2: Elección de algoritmo de emparejamiento

Para esta etapa de implementación, se lleva a cabo el emparejamiento de clientes a través de dos pasos secuenciales. En primer lugar, se calcula la distancia por el método de Mahalanobis dentro de *calipers* por *propensity score*, como se expuso en (12), entre los diferentes clientes de cada base analítica para cada categoría analizada. Para el cálculo de esta distancia, se utilizarán como variables independientes claves el *propensity score*, por su atingencia y pertinencia para resumir las variables explicativas importantes de los modelos de regresión logística y su relevancia en la predicción de la activación de la categoría, y, la medida a evaluar (el gasto de cada cliente) en un periodo de previo que se detallará en secciones posteriores.

En segundo lugar, se utiliza el algoritmo de emparejamiento por distancia de Mahalanobis dentro de *calipers* por *propensity score lineales*, con reemplazo, con una cantidad de vecinos que serán determinados mediante un análisis de mejora en el balance del resultado del emparejamiento.

Ambos pasos expuestos previamente serán implementados en el Software R mediante la utilización del paquete MatchIt, elaborado por (Ho, Imai, King & Stuart, 2011) [30].

4.1.3 Etapa 3: Verificación de soporte común

Esta etapa pretende verificar la existencia de solapamiento en la región de soporte común que permita encontrar individuos comparables en ambos grupos de la muestra de clientes (grupo tratamiento y control), pues el propósito de la última etapa es cuantificar el efecto entre estos, como lo indican (Dehejia y Wahba, 1999) [31].

Para su desarrollo existen dos formas de comprobar lo expuesto con anterioridad. Por un lado, de forma sencilla, se tiene, el realizar un análisis visual en la densidad de la distribución del *propensity score* en ambos grupos, como indica (Lechner, 2001b) [32]. Por otro lado, también es posible añadir en el análisis la observación de estadísticos importantes, tales como el mínimo, máximo y promedio del *propensity score* en ambos grupos.

4.1.4 Etapa 4: Calidad de emparejamiento y estimación del efecto

Para la evaluación de la calidad del emparejamiento, al igual que la verificación de soporte común se lleva a cabo de dos formas. Por una parte, se realizará un análisis visual del balance de variables claves mediante representación gráfica de su distribución en ambos grupos analizados (tratamiento y control). Por otra, se realizará un test de chi-cuadrado y un t-test para comprobar que no existan diferencias significativas en el promedio de las variables relevantes de ambos grupos.

Finalmente, se llevará a cabo la estimación *del Average treatment effect on the treated* (ATT) o efecto de tratamiento comparando en los grupos resultantes del emparejamiento, la medida a evaluar posterior al tratamiento, que, en este caso, corresponderá al gasto del cliente en el periodo post-activación, que será definido en secciones posteriores.

4.2 Mejor próxima categoría

Dentro de los objetivos principales del trabajo de memoria corresponde la proposición de la mejor próxima categoría a ofertar a clientes que presentan categorías inactivas. Para ello se propone el uso de modelos de regresión logística (Logit) para cada categoría, que permita entender la características y comportamiento que hace propenso a un cliente de activarse en una categoría inactiva específica.

4.2.1 Modelo de regresión logística (Logit)

Para entender a cabalidad las características que hacen propenso de activarse a un cliente, se analizarán, mediante regresiones logísticas, las variables que poseen mayor relevancia en el modelo y cuya especificación fue referida en el punto 4.1.1.

En este paso, para los modelos de activación de cada categoría, se realizará un análisis descriptivo de los parámetros estimados con el fin de sistematizar las principales características que hacen a un cliente ser más o menos propenso. Además, para cuantificar de mejor manera los efectos aislados de las variables explicativas en cada uno de los modelos, se incorpora un análisis de los efectos marginales promedio (AME, por sus siglas en inglés).

Asimismo, para evaluar la calidad y poder de predicción de cada uno de los modelos estimados se calculan y exhiben una serie de métricas como la curva COR, el AUC, entre otras.

Cabe destacar que los AME serán utilizados también para elaborar un criterio de asignación de categorías para distintos tipos de clientes con categorías inactivas, donde, se escogerán las categorías que presenten efectos marginales de mayor relevancia. Lo anterior con el fin de apoyar la segmentación de clientes para envío de ofertas e información en categorías inactivas.

4.3 Elección de próxima categoría para correos de marketing

El último paso de la etapa de modelamiento corresponde a la definición de criterios para la asignación de la próxima categoría a comunicar a cada cliente, para ello se pretende equilibrar la propensión estimada y el gasto esperado que entrega.

En particular, se propone la utilización de una métrica que permita comparar de manera sencilla la categoría que debe ofrecerse a cada cliente. Por tanto, se considera el efecto conjunto de la probabilidad de activación de un cliente a la categoría y el gasto

esperado en esta misma. Así entonces, se define el criterio de asignación como el gasto esperado por la activación, como sigue en (16).

$$G_{ij} = P_{ij} \cdot g_j \quad (16)$$

Donde, G_{ji} representa una estimación de la ganancia de que el cliente i se active en la categoría j , y que corresponde al resultado de ponderar la probabilidad P del cliente i a activarse en la categoría j con el gasto esperado g de activarse en la categoría j .

Queda que, el criterio de asignación de la categoría para cada cliente será elegir la categoría con mayor G_{ij} .

Finalmente, esta métrica específica se comparará con otros métodos de asignación estándar para los modelos de Mejor próxima oferta como lo son la elección de categorías según el decil de mayor propensión y la asignación mediante el *Lift* o mayor ganancia con respecto al promedio de propensión de la muestra de clientes. Así entonces, las métricas mencionadas anteriormente se construyen de la siguiente forma.

Propensión: Probabilidad P del cliente i de activarse en la categoría j en el periodo t , como se muestra en (17).

$$P_{ijt} \quad (17)$$

Lift de propensión: Probabilidad P del cliente i de activarse en la categoría j en el periodo t sobre la probabilidad promedio \bar{P} de la categoría j en el periodo t , como se muestra en (18).

$$\frac{P_{ijt}}{\bar{P}_j} \quad (18)$$

Lift de gasto esperado: El producto entre la probabilidad P del cliente i de activarse en la categoría j en el periodo t y el criterio de gasto esperado (G_{ij}) definido en (16) sobre la probabilidad promedio \bar{P} de la categoría j en el periodo t , como se muestra en (19).

$$\frac{P_{ijt} \cdot G_{ij}}{\bar{P}_j} \quad (19)$$

5. DESARROLLO METODOLÓGICO

5.1 Compresión de los datos

Para el desarrollo de esta sección se consideran diversas fuentes de información: información transaccional a nivel de boleta, información sociodemográfica y del negocio, e información de navegación en el portal web de la empresa.

5.1.1 Información transaccional

En primer lugar, se tiene información de todas las transacciones realizadas a nivel de boleta desde el 01-10-2017 al 30-09-2019 (correspondiente a 2 años móviles desde la fecha de extracción de la base, lo anterior responde a que, para motivos del estudio de clientes y su gestión de vida, se consideran sólo los clientes con al menos una transacción durante los últimos 2 años). Asimismo, se observa que en esta base de datos transaccional existen 7.561.362 de clientes, de los cuales no todos serán considerados para el estudio, por razones que serán detalladas en la sección posterior.

Para el caso del proyecto, se tienen las siguientes variables de mayor relevancia en la Tabla 5.

Información transaccional	
Campo	Descripción
Id_dia	Fecha en que fue realizada la compra
Num_rut_comprador	Identificador único del cliente
Monto_liquid	Monto liquidado de la transacción
Id_linea & id_sublinea	Identificador único del tipo de producto adquirido
Id_local	Identificador único del local donde se hizo la transacción, sea tienda o canal web
Cant_venta	Cantidad de productos comprados en la transacción

Tabla 5: Información principal de base de datos transaccional.
Fuente: Elaboración propia.

De esta base se puede observar que existe – para el periodo de 2 años móviles – un universo de poco más de 6 millones de clientes únicos, con transacciones positivas y correctamente identificados, lo anterior, pues, se excluyen a los clientes del tipo:

- Ruts atípicos o inválidos, tales como los menores a 100.000 y nullos o del tipo valores genéricos, tales como 1, 111, 123, 99999, entre otros.
- Transacciones que no corresponden a compras, sino que a devoluciones, recambios o cancelaciones.
- Transacciones correspondientes a tiendas islas que se encuentran fuera de la tienda misma de la empresa, las denominadas “*standalone*”, tales como Tienda isla Americanino, Mango, entre otras.

Asimismo, se tiene que en el periodo mencionado se realizaron transacciones en 1.504.834 sku's distintos, los que corresponden a 2.436 Sub-clases⁵ de productos. De acuerdo a lo anterior, se decide trabajar con la categorización a nivel de líneas (jerarquías) que corresponden a 19 agrupaciones de productos distintos. En la Tabla 6, a modo de ejemplo, se observa la descripción de jerarquías correspondiente a la línea Electro hogar. Un mayor detalle de las jerarquías de gestión de la empresa se muestra en el Anexo A.

Línea	Sub-línea	Clase	Sub-clase
Electro hogar	Computación	Notebooks	Gamer Tradicional Nueva tecnología
		Accesorios	Teclado Pendrive Mouse
		Video juegos	Consolas Juegos Accesorios
	Lavado	Centros de lavado	Lavaseca Centro de lavado
		Secadoras	Eléctricas A gas
		Lavavajillas	Lavavajilla

Tabla 6: Descripción de jerarquía tipo línea electro hogar (agrupación de productos).
Fuente: Elaboración propia.

Además, con el propósito de facilitar el procesamiento de los datos y utilizar los criterios de agrupación comercial de jerarquías para el envío de email marketing, se decide crear 11 agrupaciones diferentes de productos, utilizando la definición de línea y sub-línea expuesto con anterioridad. Los detalles de esta diferenciación y agrupación de jerarquías se observan en la sección Anexos B.

Por otro lado, de las Figuras 9 y 10 se observa que tanto para la cantidad de transacciones como montos líquidos mensuales existe una leve, pero perceptible, tendencia a la baja, lo que podría significar una disminución en la frecuencia de compra de los clientes o una pérdida o fuga de clientes. Además, se identifican claras fechas estacionales, las cuales corresponden a los meses de Cyber Day, Cyber Monday y Navidad, siendo importante recoger el efecto que significan estas fechas en particular, a la hora de estructurar el modelo.

⁵ Sub-clase corresponde a la descripción categórica más pequeña dentro del sistema de jerarquías que administra la empresa, donde, si Sub-clase es fragancias mujer jerarquía de línea entonces es Perfumería.

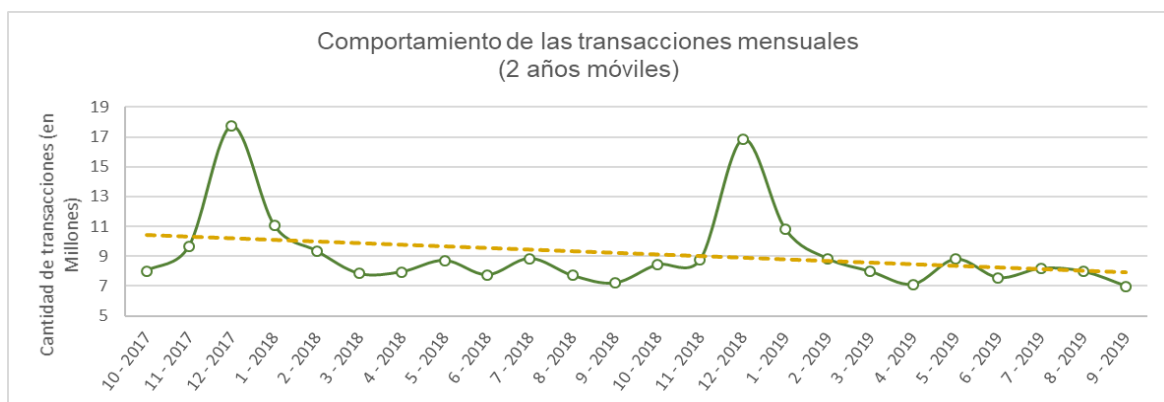


Figura 9: Evolución de cantidad de transacciones mensuales (últimos 2 años).
Fuente: Elaboración propia.

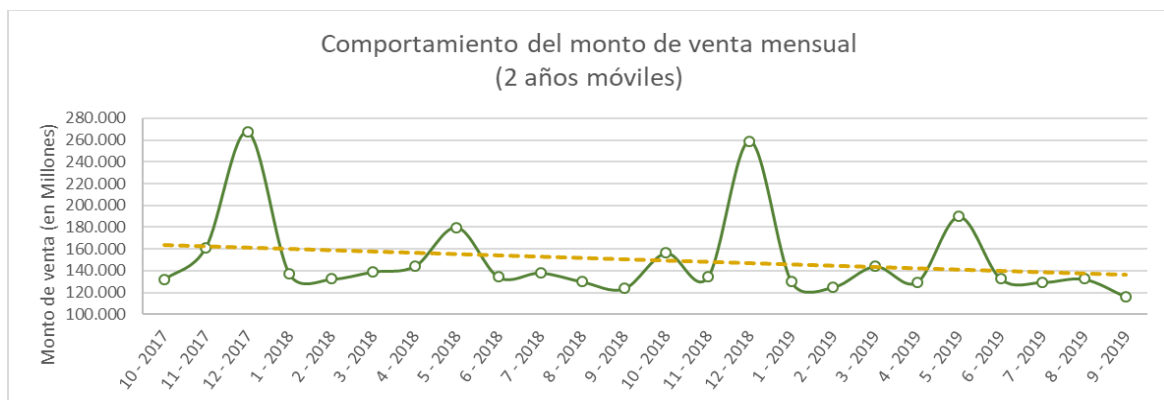


Figura 10: Evolución de monto de ventas mensual de (últimos 2 años).
Fuente: Elaboración propia.

5.1.2 Información de clientes

Con relación a la información almacenada por cliente, se tienen datos tanto sociodemográficos y del negocio como también de navegación en página web. Entre los campos más relevantes a considerar se tienen los correspondientes a la Tabla 7.

Información sociodemográfica y del negocio	
Campo	Descripción
Num_rut_Clte	Identificador único del cliente
Edad_def	Edad actual del cliente
Fecha_Nacimiento	Fecha de nacimiento del cliente
Comuna_def	Comuna/provincia de residencia del cliente
Sexo_def	Género del cliente
Estado_civil	Indica el estado civil del cliente (Soltero(a) o Casado(a))
Monto_renta	Monto de renta que percibe el cliente
Cupo	Cupo disponible en la tarjeta de crédito de la empresa
Cuotas	Indica si el cliente utiliza cuotas para realizar la compra
Hijos	Indica la existencia de hijos en el cliente
Tipo_tarjeta	Indica el tipo de tarjeta que posee el cliente

Tabla 7: Información principal de base de datos sociodemográfica.
Fuente: Elaboración propia.

A través del análisis de esta información, es posible afirmar que, del total de clientes para el estudio, se tiene que la mayoría corresponde a mujeres con un 57% del total. Asimismo, la mayor concentración de clientes se encuentra en el rango [25-35] años con un 26% del total de clientes, donde, un 41% del total de clientes se encuentra casado(a).

Además, de la Figura 11, se observa de acuerdo a ciertos rangos de edad, que existen diferencias claras en la prevalencia del género femenino para todos los rangos, con mayor relevancia para rangos superiores a los 35 años. De lo anterior se extrae que, La empresa, al focalizar su atención y objetivo comunicacional en las mujeres, así como también su volumen de oferta, posee una prevalencia evidente con respecto a los hombres.

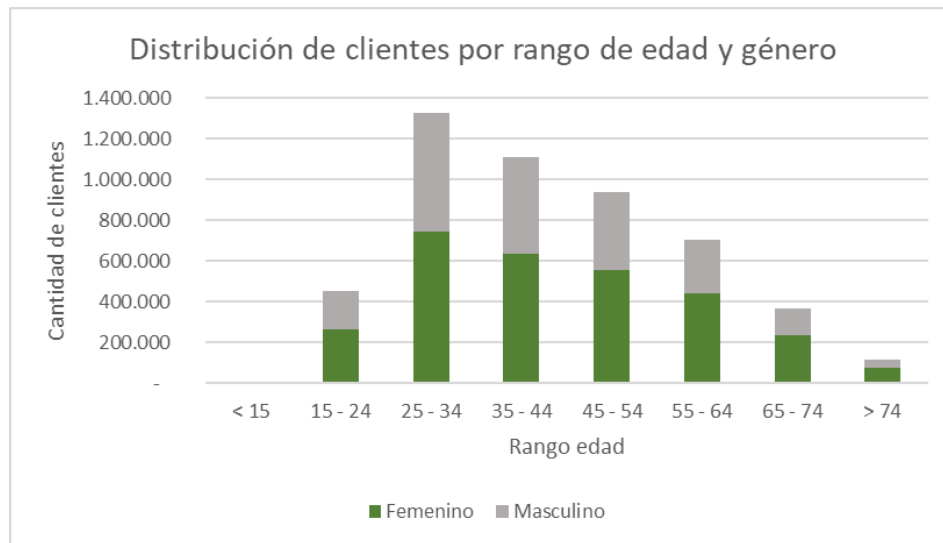


Figura 11: Distribución etaria por género entre el conjunto de clientes.
Fuente: Elaboración propia.

Por otro lado, con el objetivo de considerar la actividad que mantiene el cliente con la empresa y la relación comunicacional, se consideran datos de navegación de los clientes en la página web. La información de mayor relevancia – en consideración con el objetivo del trabajo – se observa en la Tabla 8.

Información de navegación web	
Campo	Descripción
Num_rut_clte	Identificador único del cliente
Id_linea & sub_linea	Identificador único del tipo de producto navegado
Fecha_nav	Fecha de navegación
Cant_vista	Cantidad de veces que se navegó el producto

Tabla 8: Información principal de base de datos de navegación.
Fuente: Elaboración propia.

5.1.3 Análisis exploratorio del conjunto de datos

Dada las características del problema abordado y los objetivos del modelamiento, se toman en consideración desde este punto en adelante todas las transacciones que hayan ocurrido entre julio del 2018 a junio de 2019. Esto pues, se pretende dejar los

meses de julio, agosto y septiembre de 2019 para estimar el gasto posterior de los clientes, como será detallado en secciones posteriores.

Asimismo, se tiene que este análisis exploratorio pretende encontrar ciertas diferencias observables en variables claves que se desean incluir en el análisis y modelo posterior. Por un lado, si existen diferencias evidentes en el comportamiento de clientes activos e inactivos, tanto en general como entre categorías. Para ello, se consideran clientes activos e inactivos a la fecha de extracción de la base, es decir, clientes con estas características al mes de junio de 2019 y su comportamiento en el último año móvil.

Por otro lado, se considera también diferenciar las transacciones de clientes antiguos y nuevos, puesto que es de interés conocer la diferencia de comportamiento entre ambos y si ello entrega información útil que pueda ser incluida en el estudio. Lo anterior se debe a que se tiene la hipótesis de que el comportamiento de clientes activos⁶ contra inactivos es distinto y que, a su vez, este también difiere entre clientes antiguos⁷ contra nuevos.

Por último, se hace diferencia dentro del análisis de las transacciones durante el último año mencionado, a si estas fueron realizadas mediante el portal web de la empresa, o, por el contrario, fueron realizadas en tiendas físicas. Esto pues se pretende encontrar diferencias evidentes en el canal de compra para algunas de las categorías incluidas y que esta información pueda ser utilizada como insumo para el estudio.

En primer lugar, se analizan los datos siguiendo una diferenciación por información sociodemográfica para edad, zona de residencia y género.

Como se observa de la Figura 12, se tiene que existen un mayor volumen de clientes antiguos con respecto a los clientes nuevos, asimismo, se observa una mayor presencia de clientes inactivos en el grupo de clientes nuevos en comparación con el de clientes antiguos. Luego, de las Figuras 13 y 14, que muestra la distribución etaria para clientes antiguos y nuevos, se observa que para ambos grupos la mayor concentración de clientes se encuentra en el rango etario de 25 a 44 años.

Lo anterior se puede explicar si se asume que los clientes antiguos poseen un estado mayor de maduración en la empresa y que ésta ha logrado una mayor retención entre ellos y, por lo tanto, un mejor comportamiento de compra con respecto a clientes que recién comienzan una relación con la empresa.

⁶ Se define, por criterio comercial, un cliente activo como un cliente que ha comprado al menos un producto en los últimos 3 meses e inactivo a uno que ha realizado alguna compra entre 3 meses y 1 año, pero no ha comprado en los últimos 3 meses. El resto se considera como cliente fugado. Para el caso de este estudio, este criterio también se utiliza para definir a un cliente activo o inactivo en cada categoría.

⁷ Se define como cliente nuevo en la unidad de negocio a un cliente que ha comprado sólo durante el último año y un cliente antiguo a un cliente que ha comprado este último año y además compró en el año inmediatamente anterior.

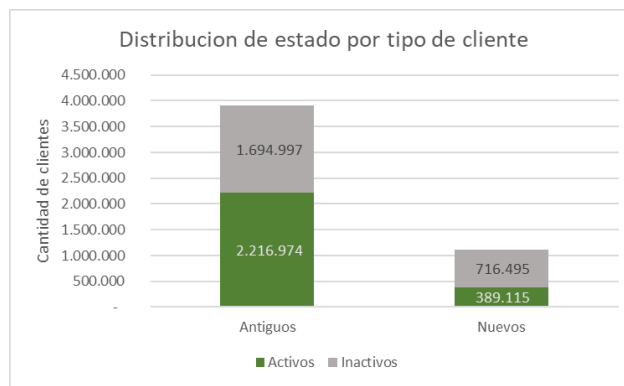


Figura 12: Distribución de estado de actividad por tipo de clientes.
Fuente: Elaboración propia.

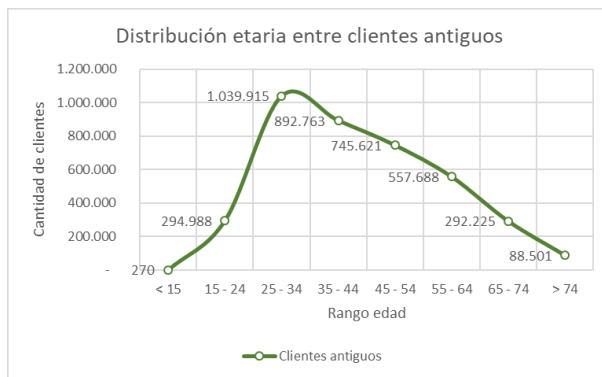


Figura 13: Distribución etaria de clientes antiguos.
Fuente: Elaboración propia.

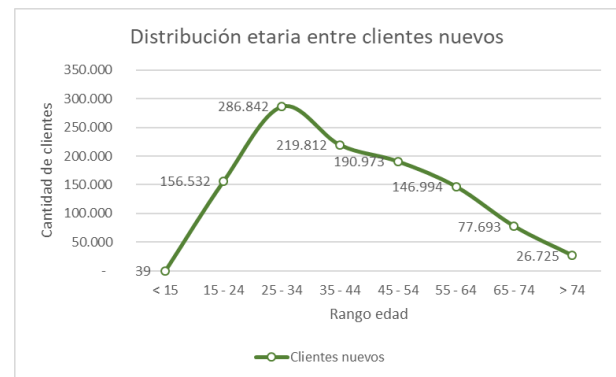


Figura 14: Distribución etaria de clientes nuevos.
Fuente: Elaboración propia.

Con respecto a la información residencial, se tiene que, como se observa en las Figuras 15 y 16, para ambos tipos de clientes, la concentración es mayor en la Región Metropolitana, en particular para la zona poniente, esto pues, la zona oriente incluye un menor número de comunas.

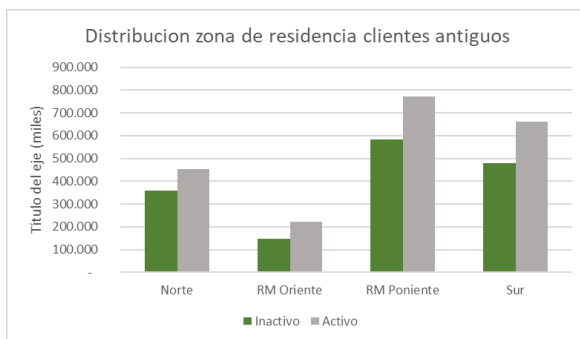


Figura 15: Zona de residencia de clientes antiguos.
Fuente: Elaboración propia.

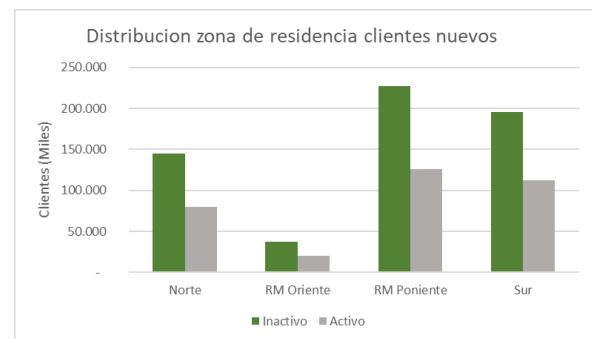


Figura 16: Zona de residencia de clientes nuevos.
Fuente: Elaboración propia.

Finalmente, para el análisis de género, se tiene que para clientes antiguos existe un mayor volumen de clientes del género femenino. Por otro lado, para el caso de los clientes nuevos, esa diferencia no es observable, siendo similar la distribución para ambos segmentos. El estado de los clientes (activo e inactivo) se observa similar tanto para clientes antiguos como para nuevos, siendo el estado de inactividad de clientes, el de mayor presencia. Lo anterior se muestra en las Figuras 17 y 18.

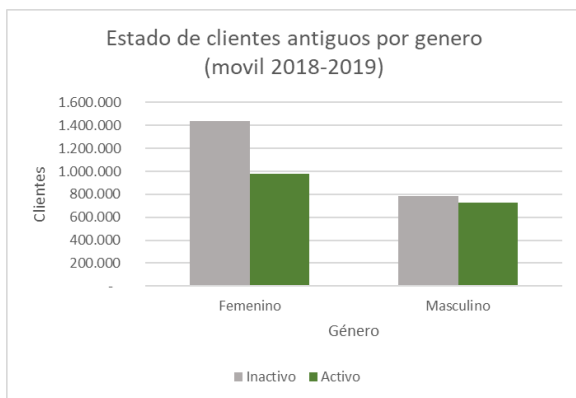


Figura 17: Distribución de género clientes antiguos.

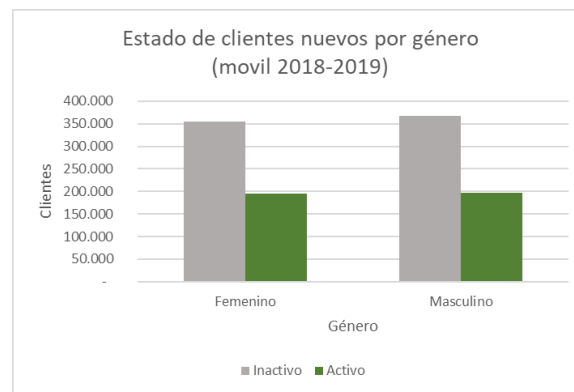


Figura 18: Distribución de género clientes nuevos.

Fuente: Elaboración propia.

Ahora bien, para las Figuras 19, 20 y 21, se utiliza la agrupación de categorías de producto definida en la sección Anexos B. Esta diferenciación tiene por objeto observar posibles diferencias evidentes en el comportamiento de compra y navegación web de clientes (activos e inactivos en categorías) en las agrupaciones que se utilizarán en el estudio.

Por tanto, sigue que, con el análisis de comportamiento de compra, en la Figura 19, se aprecia que la categoría con mayor venta promedio anual corresponde a la categoría Electro, seguida por Deco, resultado esperable pues la diferencia de precios entre estas categorías con respecto a las otras es alta. Lo anterior se comprueba si se observa la cantidad de clientes que han comprado en cada categoría, donde categorías como Moda hombre, Moda Mujer e Infantil, poseen mayor volumen de clientes, pero su gasto promedio es menor a las mencionadas.

Por otro lado, es posible identificar una diferencia clara de comportamiento entre clientes activos contra inactivos, donde, estos últimos, poseen una menor venta promedio en todas las categorías de estudio. Cabe mencionar que, un análisis similar se llevó a cabo para otras variables, como cantidad de productos comprados y visitas a tiendas (física y virtual) con compra, los que se muestran en los Anexos C y D respectivamente.

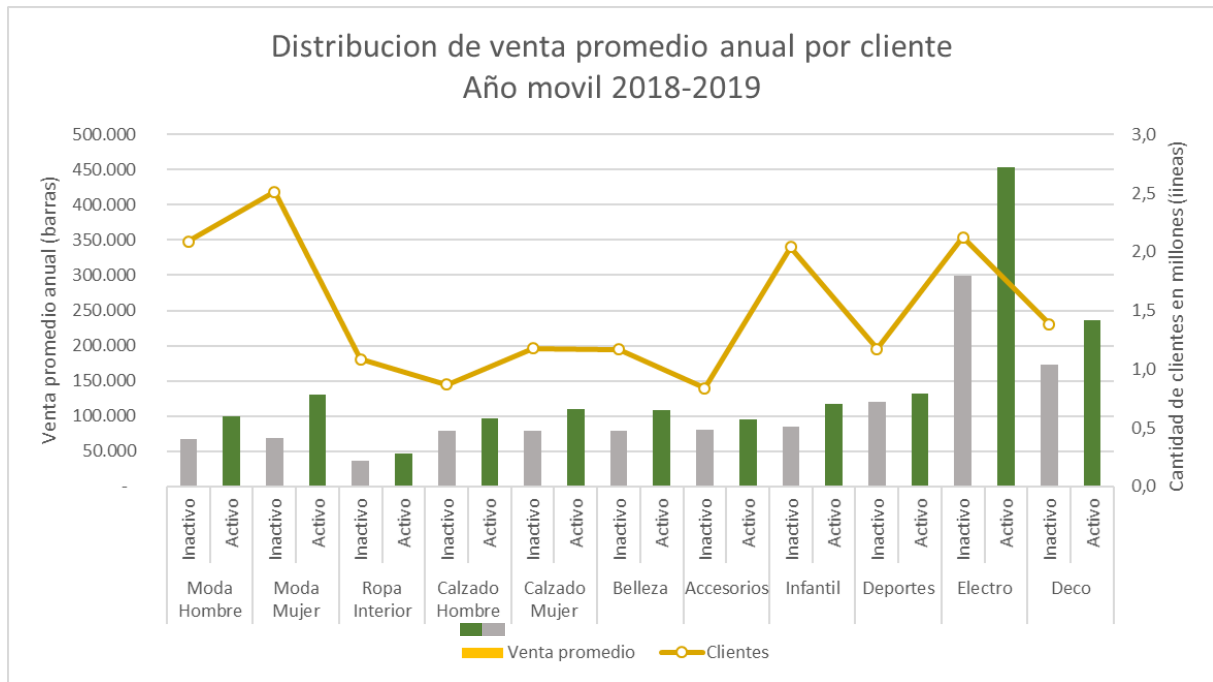


Figura 19: Comportamiento de compra en distintas agrupaciones de productos en clientes activos e inactivos. Fuente: Elaboración propia.

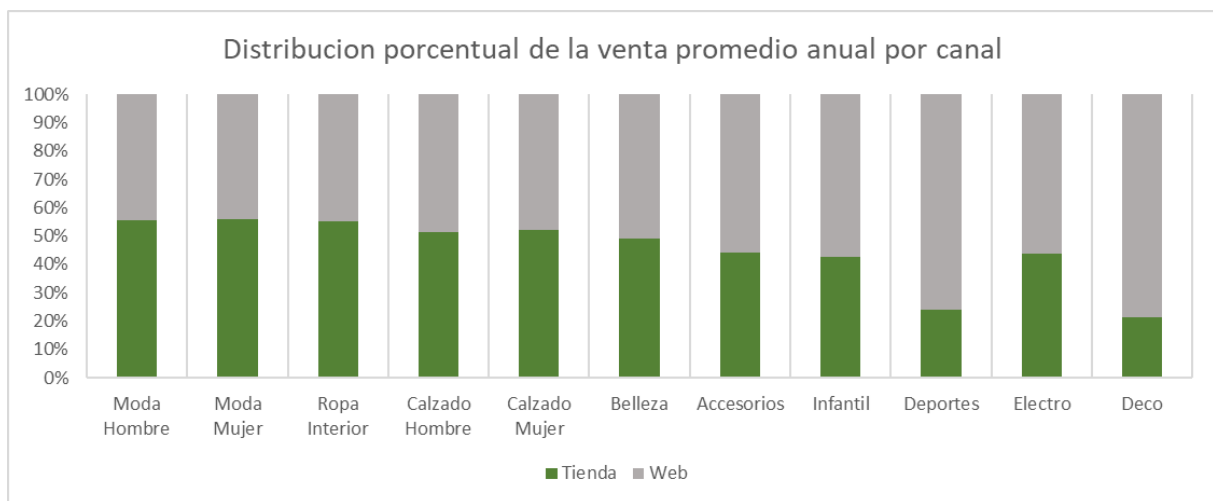


Figura 20: Distribución porcentual de la venta promedio anual por categoría de producto por tipo de canal. Fuente: Elaboración propia.

Por último, se llevó a cabo una diferenciación por el tipo de canal que fue utilizado para la compra, donde, como se observa en de la Figura 20, categorías de producto como Accesorios, Infantil, Deportes, Electro y Decoración poseen un mayor porcentaje de venta efectuado mediante el canal web de la empresa. Asimismo, se observa una leve relevancia del canal físico por sobre el canal web en categorías como Moda Hombre, Moda Mujer y Ropa Interior.

Como fue mencionado con anterioridad, una información importante que será considerada para el estudio es la influencia de la navegación en la página web sobre las distintas agrupaciones de productos. Por lo anterior, se estudia el comportamiento de navegación de los clientes tanto activos como inactivos para cada categoría, como se

muestra en la Figura 21. En esta se observa que el promedio anual de navegación para clientes activos es mayor en todas las categorías estudiadas, con mayor presencia en categorías blandas⁸ de Moda Mujer y Calzado Mujer, seguidas por Electro y Deco.

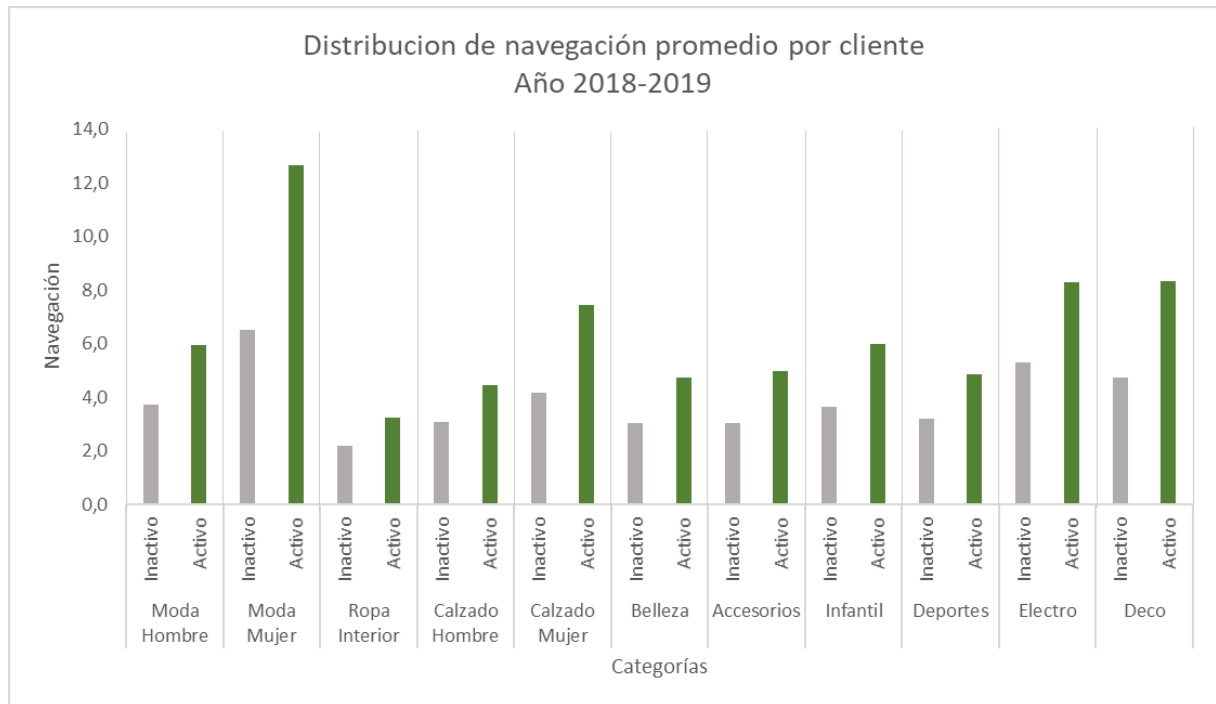


Figura 21: Navegación promedio anual en categorías para cliente activos e inactivos.
Fuente: Elaboración propia.

Finalmente, otra hipótesis que se desea responder es que un cliente multi-categorico posee un mejor comportamiento de compra y, por lo tanto, un mayor gasto en la unidad de negocios que un cliente mono-categorico. Para ello, se analiza la frecuencia de compra anual promedio de distintas categorías sobre todos los clientes del estudio.

Se encuentra que, en lo expuesto en la Figura 22, la gran mayoría de los clientes son mono-categoricos, es decir, concentran mayor volumen de compra en 1-3 líneas de producto y desciende de manera paulatina. Asimismo, se analiza el promedio anual de gasto, donde se encuentra que el mayor promedio de gasto anual se encuentra en los clientes que son multi-categoricos, es decir, compran en un gran número de líneas [10,19]. En virtud de lo anterior, se concluye que existe un valor relevante en convertir a un cliente mono-categorico en uno multi-categorico

⁸ Categoría blanda refiere al término acuñado por la empresa para diferenciar a los productos que no son Electro hogar o Decoración.

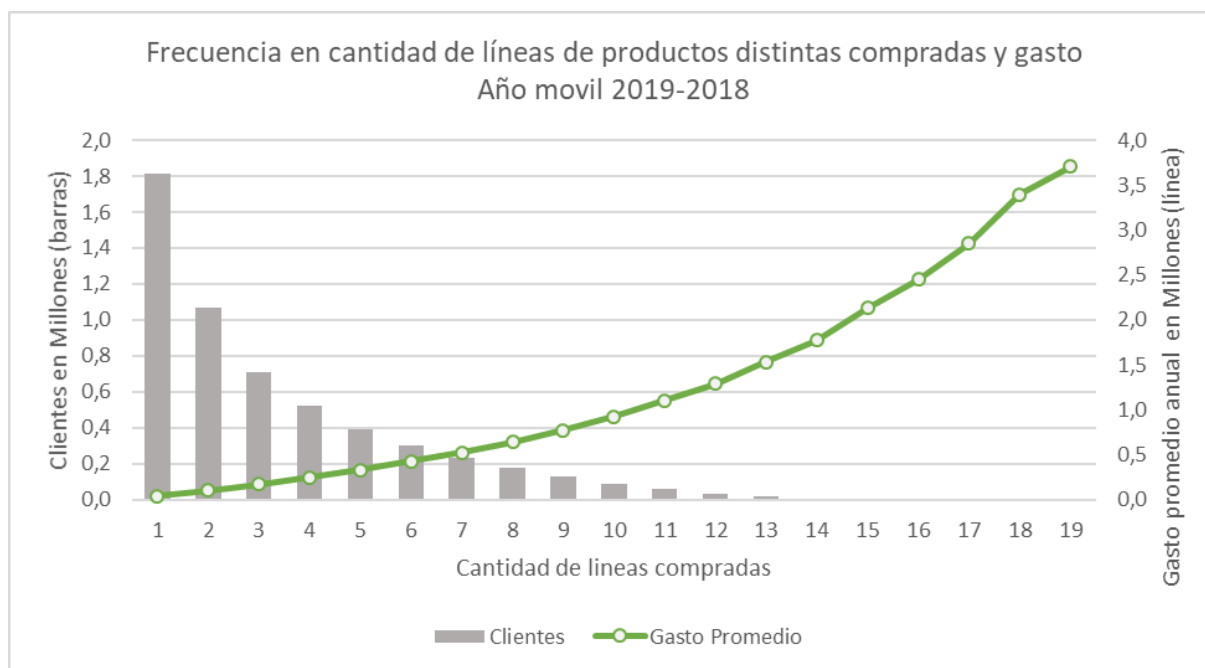


Figura 22: Frecuencia de compra en distintas líneas de producto y gasto promedio anual.
Fuente: Elaboración propia.

5.2 Preparación de los datos

Como puntal inicial de esta etapa, correspondiente al tratamiento de los datos, que permite el posterior desarrollo de los modelos seleccionados, se realiza un estudio e identificación de los clientes con comportamientos anómalos y/o extremos.

Con lo anterior en consideración, la base total de clientes para el periodo de julio de 2018 a junio 2019 corresponde a 5.755.215 clientes con al menos una compra durante ese periodo. Como fue mencionado, el objetivo es excluir de la base de estudio a clientes que presenten un comportamiento extremo, tales como una inmensa cantidad de compras asociado a un monto total muy pequeño, o a la inversa, un enorme monto de compra asociado a un número muy pequeño de compras. En otras palabras, se desea excluir a clientes que aumentan de manera considerable la varianza entre variables importantes.

Para llevar a cabo lo anterior se utiliza como criterio de exclusión, la distancia de Mahalanobis planteada por (Matsumoto, Kamei, Monden, & Matsumoto, 2007) para tres especificaciones distintas. La utilidad de esta técnica, por sobre la ejecutada mediante box-plot, es la posibilidad de analizar comportamientos extremos de forma multivariada. Así entonces, las tres variables seleccionadas para el análisis de *outliers* corresponden a: montos de compra, visitas con compra y cantidad de compras, anuales por cliente, como se observa en la Tabla 9. Cabe destacar que, la técnica puede ser ejecutada utilizando las tres variables de manera conjunta, no obstante, se decide realizar combinaciones en pares para facilitar la observación del comportamiento de exclusión de clientes a través de representaciones gráficas.

Combinaciones para análisis de <i>outliers</i> por cliente en el año móvil		
	Variable 1	Variable 2
Par 1	Monto de venta	Cantidad de productos comprados
Par 2	Monto de venta	Visitas con compra
Par 3	Cantidad de productos comprados	Visitas con compra

Tabla 9: Combinación de variables para estudio de *outliers*.

Fuente: Elaboración propia.

Como se observa de los anexos E, F y G, tanto el monto de venta, visitas con compra y la cantidad de productos comprados, no poseen distribuciones normales, requisito que se exige para utilizar la distancia de Mahalanobis. Por lo tanto, se procede a realizar la transformación logarítmica de estas variables. En virtud de lo anterior, se observa en los anexos H, I y J que el monto de compra, ahora, posee una distribución aproximadamente normal en todo su dominio y las variables de cantidad de productos comprados y visitas con compra anuales, poseen una distribución normal en algunos tramos, no obstante, permite utilizar el método.

Sigue que, como se muestra en la Figura 23 para el *par 1*, de acuerdo con el criterio de Mahalanobis, la distancia de corte se establece en 11,34, que representa la distancia en que se encuentra el 99,9% de la varianza, siguiendo una métrica de chi-cuadrado, asumiendo distribución normal. La representación gráfica del criterio de exclusión para los pares 2 y 3, se exponen en la sección Anexos K y L.



Figura 23: Criterio de exclusión para el par 1: montos de venta anual y cantidad de productos comprados anual.

Fuente: Elaboración propia.

Finalmente, se tiene que, la cantidad de clientes excluidos corresponde a 47.290 (0,82%) del total de la muestra.

De los resultados obtenidos, es posible extraer que, los clientes que fueron excluidos de la muestra efectivamente corresponden a comportamientos de compra

anómalos, como se muestra en la Tabla 10, donde el promedio de montos de venta, visitas con compra y cantidad anual de productos comprados son muy diferentes para cada grupo.

Outlier	Clientes	Monto venta	Visitas con compra	Cant. Productos
No	5.707.925	\$ 244.092	4	11
Si	47.290	\$ 953.946	6	32

Tabla 10: Comportamiento anual de gasto, visitas con compra y cantidad de productos para clientes excluidos y no excluidos.

Fuente: Elaboración propia.

5.3 Modelamiento y evaluación de resultados

Con el objetivo de simplificar el tiempo de procesamiento y entregar mayor información con respecto a los efectos que se quieren estimar, desde aquí en adelante las 19 líneas de productos que administra la empresa serán agrupadas en 11 categorías diferentes, siguiendo la misma estructura que fue mencionada con anterioridad para realizar el análisis exploratorio. Las agrupaciones son Moda Hombre, Moda Mujer, Calzado Hombre, Calzado Mujer, Deportes, Ropa Interior, Belleza, Accesorios, Infantil, Electro y Deco. El detalle de qué líneas de productos contiene cada una de ellas se detalla en la sección Anexos B.

5.3.1 Análisis de efectos causales de activación en categorías

Para esta primera etapa de modelación, se pretende dar cuenta del efecto que tiene sobre el gasto de los clientes el activarse o no en una categoría específica. Este procedimiento se concibe con el fin de replicar un experimento aleatorio utilizando grupos de tratamiento y control balanceados y comparables. Para ello, el procedimiento metodológico corresponde al descrito en 4.1.

Con el objetivo de llevar a cabo el análisis, se utiliza el método de emparejamiento por distancia de Mahalanobis dentro de *calipers* por *propensity score*. Para ello, se define como variable objetivo el comprar o no en una categoría específica cuando no se ha efectuado compra alguna en la categoría en los tres meses anteriores.

Se propone para el análisis tres periodos relevantes, pre-activación, activación y post-activación para cada categoría, siendo los periodos de análisis estipulados en 3 meses, 1 mes y 3 meses respectivamente, como se observa en la Figura 24.

Del periodo de pre-activación (3 meses) se rescata la información que construye el criterio para la activación, esto es, no tener compras en la categoría analizada.

El periodo de activación (1 mes) corresponde a la ventana de tiempo en que el cliente se activó o no en la categoría analizada.

Finalmente, en la ventana de post-activación se analiza la diferencia en gasto promedio para cada categoría luego de su activación en la categoría analizada.

Cabe destacar que la ventana de activación se hizo mes a mes para el último año móvil, es decir, activación de cada cliente en cada mes.

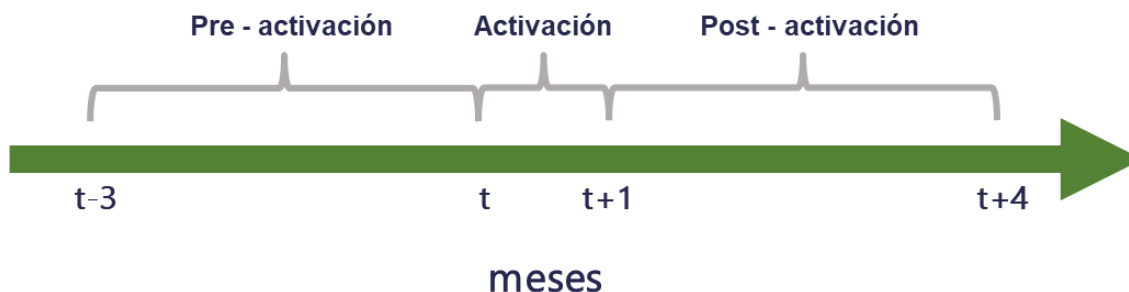


Figura 24: Ventana de tiempo para estudiar efectos causales de activación en categorías.
Fuente: Elaboración propia.

A continuación, para cada una de las categorías de producto estudiadas, se detalla su base analítica, que corresponde a los clientes que no han realizado compra en aquella categoría durante los tres meses anteriores (pre-activación). Se tiene entonces que la cantidad de registros en cada base para cada categoría y el porcentaje de activaciones en el año de estudio se resume como sigue en la Figura 25.

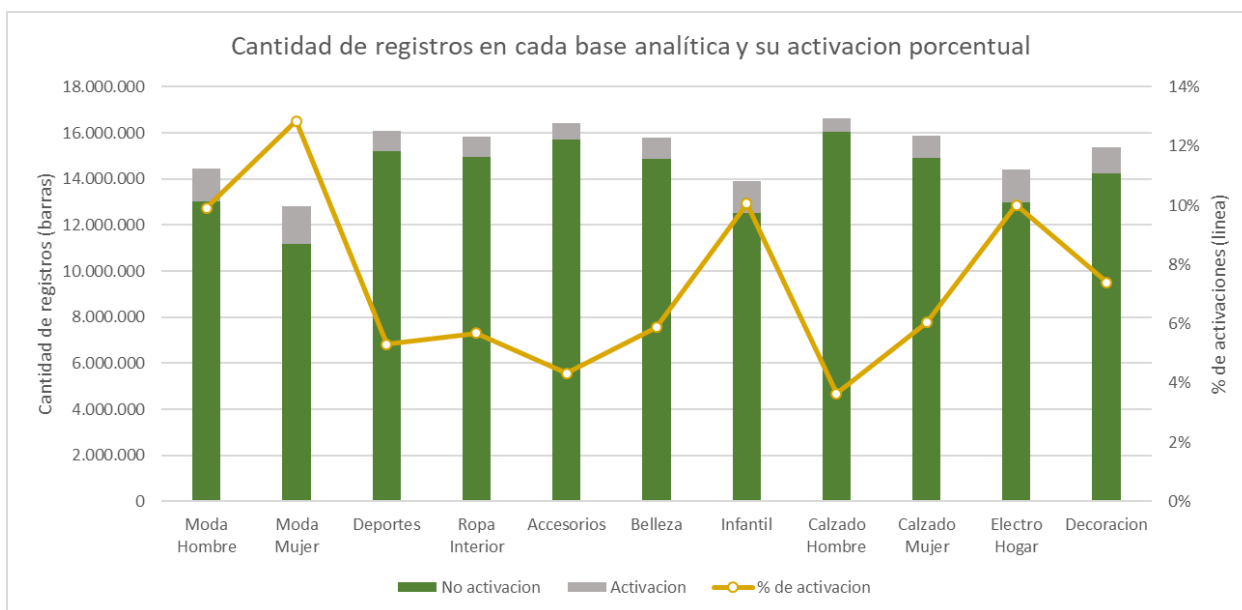


Figura 25: Cantidad de clientes en las bases analíticas de cada categoría y su porcentaje de activación.
Fuente: Elaboración propia.

5.3.1.1 Cálculo del propensity score

Para realizar el análisis comparativo del gasto del cliente después de activarse o no en una categoría específica, lo primero que se debe hacer es calcular la propensión a que el cliente efectivamente se active dado su comportamiento de compra y características de emparejamiento exigidas en la ventana de Pre-activación.

Para ello, se decide calcular el *propensity score* mediante un modelo de regresión logística (Logit) de tipo binario para cada categoría de producto analizada, donde, la variable objetivo corresponde a la activación en un mes en particular. Dentro de las variables explicativas se considera información del tipo transaccional, sociodemográfica, del negocio y de navegación web, el detalle de las variables consideradas en primera instancia se observa en Anexos M.

Es importante mencionar que, las variables explicativas de actividad o comportamiento de compra previo del cliente se construyen observando un mes atrás con respecto a su activación en un mes específico. Lo anterior se realiza de manera iterativa para los 12 meses del año móvil analizado con el objetivo de captar información de actividad de compra para todos los meses del año de estudio. Además, con el motivo de simplificar el tiempo de procesamiento (pues existían restricciones fuertes de la capacidad de procesamiento del computador a disposición), se seleccionan de manera aleatoria y balanceada, muestras representativas correspondientes al 1% de cada una de las bases analíticas por categoría. Lo anterior se llevó a cabo verificando la conservación de la distribución en variables importantes (estacionalidad, género, edad, estado de cliente, entre otras).

Luego, se procede a eliminar variables explicativas mediante el análisis de su matriz de correlaciones, tanto para variables del tipo continua (numéricas) como para las variables del tipo categóricas. En virtud de los resultados obtenidos, se descartan las variables Visitas con compra, Cantidad de líneas compradas, Gasto mes anterior, dado un criterio de corte mayor a 0,55 de coeficiente correlación, como se observa en Anexos N y O. Para esta etapa, no se considera la eliminación de las variables del tipo actividad en diferentes categorías, pues, el propósito central de la memoria es conocer el efecto cruzado y causal que ello implica sobre la activación.

Por otro lado, con el objetivo de captar el efecto estacional de las activaciones en distintos meses durante el año móvil considerado, se añaden variables de tipo binaria que indican el mes para el que se está realizando la predicción.

Entonces, la formalización de la componente observada de la utilidad para el cliente i de activarse en la categoría de productos k en el mes t (U_{ikt}) para la estimación de modelos Logit binarios se formula en (20) como:

$$U_{ikt} = \alpha_k + \sum_{k \neq j} \vec{\beta}_j \cdot X_{ij(t-1)} + \vec{\gamma}_k \cdot Z_{i(t-1)} + \vec{\delta}_k \cdot Y_t \quad (20)$$

Donde, $X_{ij(t-1)}$ corresponde a la activación (binaria) de las distintas categorías de productos j , con $k \neq j$, durante el mes anterior para el cliente i en el mes $t - 1$. $Z_{i(t-1)}$ corresponde a las variables de las características sociodemográficas, de navegación y de la empresa del cliente i en el mes $t-1$. Y_t corresponde a la variable binaria correspondiente al mes en que se está prediciendo la activación. Finalmente, el parámetro α_k corresponde al porcentaje o, en este caso, a la probabilidad de activarse de un cliente cuando se encuentra inactivo en la categoría; β_j corresponde al coeficiente que indica la probabilidad de activarse en la categoría analizada k cuando existe o no

una activación en una categoría j con $k \neq j$; γ_k corresponde al coeficiente que indica la probabilidad de activarse en la categoría analizada k que depende las características sociodemográficas del cliente; y δ_k corresponde al coeficiente que indica la probabilidad de activarse en la categoría analizada k asociada a la activación de un cliente en un mes específico.

Sigue que, se separa de forma balanceada cada muestra de análisis en un 80% para entrenamiento y un 20% para testeo. Así, para las bases de entrenamiento, se estiman los parámetros de los modelos para cada categoría de productos aplicando el proceso sistemático de eliminación de variables explicativas o *Stepwise*, mediante la selección de especificación de modelos con menor BIC. De lo anterior, es posible observar que existen diferentes especificaciones para cada categoría, lo que queda demostrado en la variación del BIC en la sección Anexos R. Cabe destacar que en ningún caso se mantienen todas las variables inicialmente incluidas para evaluar cada modelo mediante *Stepwise*, y que, por tanto, no se escogen los modelos con mayor verosimilitud.

Las especificaciones finales de los modelos Logit por categoría y sus coeficientes estimados, se observan en Anexos P.

Desde este punto en adelante, para facilitar el análisis y la simplicidad de los resultados, se detallarán los pasos correspondientes a la metodología para la categoría Decoración. No obstante, la totalidad de los resultados para las demás categorías pueden ser consultadas en la sección Anexos.

Para la categoría Decoración, de la Figura 26, es posible observar el comportamiento del ajuste del modelo que posee un AUC de 69%, lo que indica que el modelo se encuentra por encima de una selección aleatoria (línea discontinua), pero como es de esperar, no describe el comportamiento del cliente a la perfección (sobre ajuste). En la sección Anexos R, se muestran métricas de evaluación asociadas para cada categoría. Un mayor detalle acerca de las especificaciones de cada modelo, la relevancia de sus coeficientes y su interpretación será abordado en la sección 5.3.2.

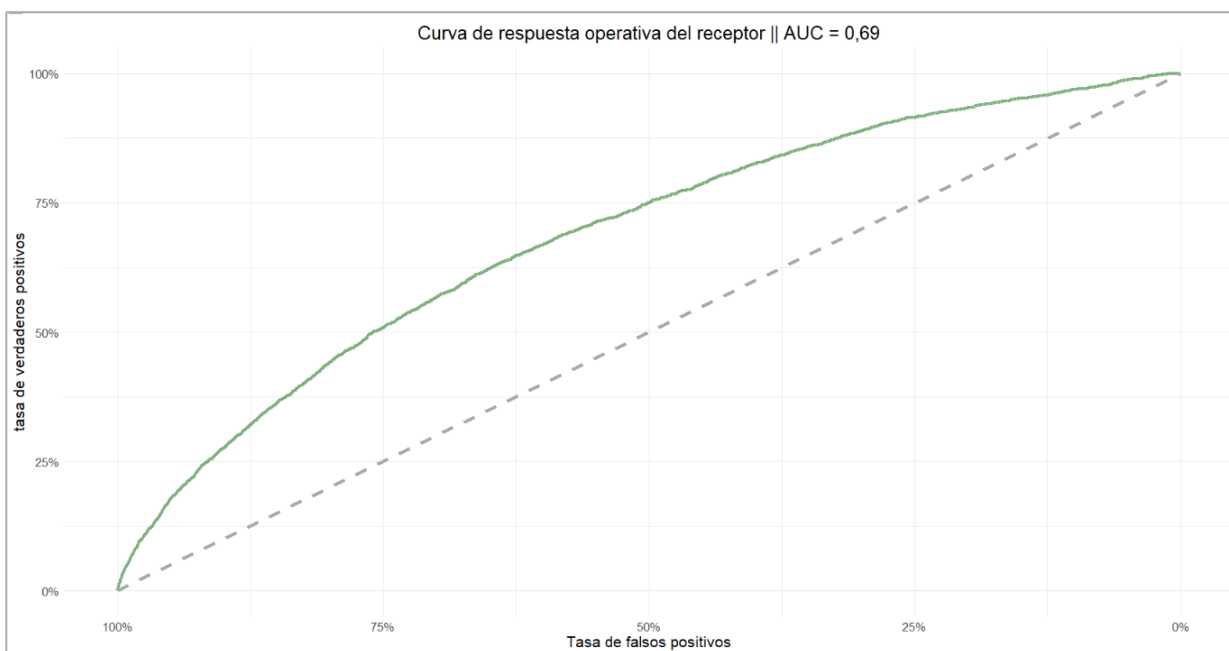


Figura 26: Curva de recepción operativa del receptor que indica un área bajo la curva (AUC) de 69%.
Fuente: Elaboración propia.

5.3.1.2 Soporte común y emparejamiento

Para llevar a cabo el emparejamiento en la categoría Decoración, primero que todo, se debe verificar el soporte común del *propensity score* entre ambos grupos de clientes, los que se activan y los que no.

Se observa, de los principales estadísticos para la categoría, que existe un gran rango de probabilidad para la activación de clientes, siendo 2,6% y 99% sus valores mínimo y máximo para el *propensity score*, respectivamente. Además, se tiene que, el rango de solapamiento del *propensity score* – requisito para el emparejamiento – se encuentra entre 2,9% y 98,7%.

Asimismo, se comprueba que el clasificador de propensión realiza un buen trabajo, pues el promedio para los clientes que no se activan corresponde a 10,4% y 15,6% para los que se activan, como se muestra en la Figura 27. Donde, se decidió acotar el eje correspondiente a la cantidad de clientes para observar de manera más clara la diferencia en la distribución del *propensity score* para grupo. El resumen de los estadísticos y representación gráfica del soporte común para ésta y las demás categorías se muestran en Anexos R y T, respectivamente.

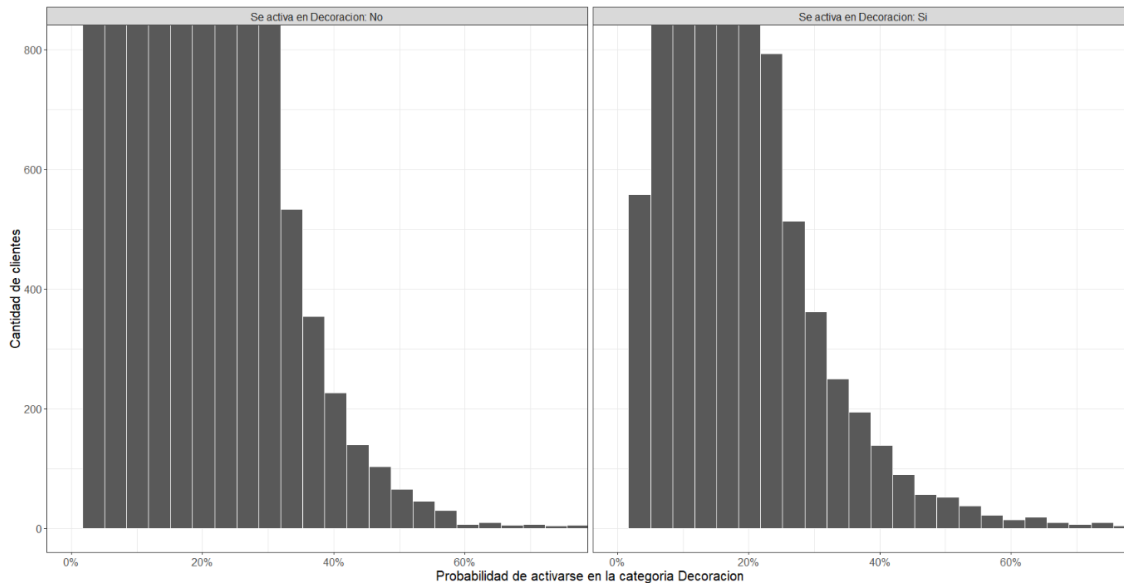


Figura 27: Verificación gráfica del soporte común del *propensity score* para la categoría Decoración.
Fuente: Elaboración propia.

Sigue que - siguiendo las etapas propuestas para el modelamiento en la metodología - se ejecuta el emparejamiento luego de la verificación del soporte común y la elección del algoritmo de emparejamiento, que como fue descrito en el punto 4.1.2 corresponde a la distancia de Mahalanobis dentro de *calipers* por *propensity score lineal*. Cabe mencionar, que el *caliper* se define para cada categoría como el 25% de la desviación estándar de los *propensity score lineales*, como se recomienda en (Rosenbaum & Rubin, 1985) [33].

Para determinar la cantidad de vecinos a utilizar mediante el emparejamiento por radio, con reemplazo, se debe considerar la compensación que existe entre varianza y sesgo al aumentar la cantidad de vecinos. Lo anterior, se debe a que al escoger una mayor cantidad de vecinos se reduce la varianza en el *caliper*, no obstante, se eligen, cada vez, vecinos menos comparables y más lejanos al de mayor similitud (Smith, 1997) [34].

Por lo tanto, se escoge la cantidad de vecinos que genera mayor mejora en el balance de las variables claves, de tal forma que se obtiene que la cantidad de vecinos a utilizar corresponde a 3 para cada una de las categorías.

Así entonces, realizado el emparejamiento con la cantidad de vecinos mencionados, se observa que existe al menos un emparejamiento para el 100% de los casos de la muestra de clientes en que efectivamente hubo activación. Sin embargo, se observa que no se encuentran 3 vecinos para todos los casos, siendo 2,3 el promedio de vecinos para cada registro.

5.3.1.3 Calidad del emparejamiento

A continuación, corresponde verificar el balance del emparejamiento entre las variables claves que se utilizaron para ello - *propensity score* y gasto en periodo de pre-activación - mediante un test de medias y un test de chi-cuadrado, respectivamente.

Como se observa de la Tabla 11, la diferencia en el promedio de las variables mencionadas no muestra diferencias significativas con un 95% de confianza entre ambos grupos (el que se activa y el que no). Esta información se observa en Anexos S para las demás categorías, donde, para los resultados del balance del emparejamiento, se comprueba para 9 de las 11 categorías no existen diferencias significativas y, por tanto, se poseen grupos comparables para analizar el gasto post-activación. Cabe mencionar que, para las categorías con desbalance en el *propensity score*, se decide seguir la elección de 3 vecinos para mantener lo más semejante posible las características del emparejamiento para todas las categorías estudiadas.

Decoración				
Variable	Activación	No activación	Diferencia	p-value test
Propensity score	15,69%	15,54%	0,15%	0,91
Gasto Pre-activación	\$ 105.238	\$ 103.865	\$ 1.373	0,62

Tabla 11: Balance de variables claves escogidas para el emparejamiento.
Fuente: Elaboración propia.

Por otro lado, lo anterior también es posible compararlo de manera gráfica, donde, como se observa en la Figura 28, el método de emparejamiento mantiene las propiedades de comportamiento del *propensity score* en ambos grupos, en que a valores diferentes de probabilidad de activación se mantiene el balance en ambas de las variables analizadas (*propensity score* y gasto pre-activación). De forma adicional se muestra el balance para tres variables de comportamiento de compra importantes en la activación de la categoría, como lo son las categorías de Electro hogar (variable interesante pues estas categorías son las que presentan mayor similitud en cuanto a comportamiento de compra de los clientes), Moda Hombre e Infantil.

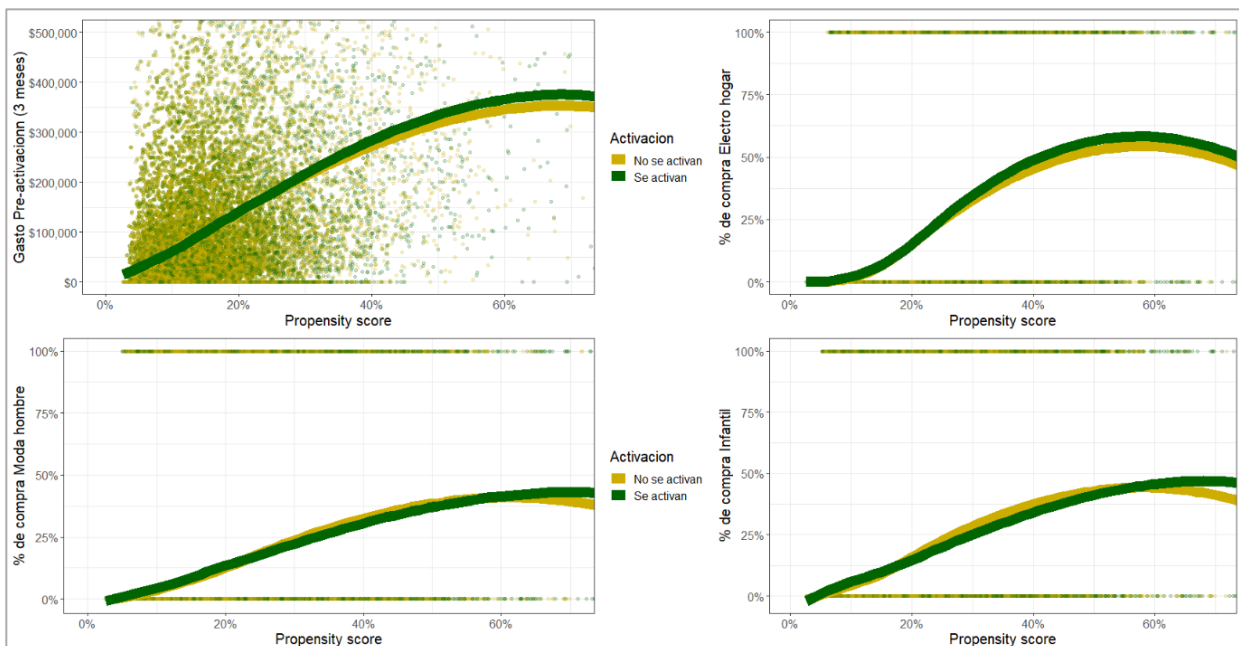


Figura 28: Balance entre variables relevantes para el emparejamiento por propensity score y gasto pre-activación.
Fuente: Elaboración propia.

Los resultados gráficos para las categorías restantes del estudio se muestran en la sección de Anexos U.

5.3.1.4 Estimación del efecto causal

En primer lugar, en la Tabla 12 se muestran los resultados de la diferencia en gasto pre-activación (variable utilizada para el emparejamiento) en Decoración, para antes y después de aplicar el emparejamiento, donde, la diferencia en el gasto pre-activación es bastante menor para los clientes emparejados.

Por lo tanto, la disminución en la brecha del gasto pre-activación, verifica la utilidad de llevar a cabo el emparejamiento con el objetivo de obtener muestras de clientes comparables.

Variable	Situación	Promedio en grupo que se activa	Promedio en grupo que no se activa	Diferencia
Gasto en periodo pre-activación	Antes del emparejamiento	\$105.238	\$63.065	\$42.173
	Después del emparejamiento	\$105.238	\$103.865	\$1.373

Tabla 12: Diferencia en el gasto pre-activación para ambos grupos antes y después del emparejamiento.
Fuente: Elaboración propia.

Finalmente, se procede a la estimación del efecto de la activación en cada categoría sobre el gasto del cliente en el periodo post-activación.

Para el caso de la categoría analizada (Decoración), se comprueba que la diferencia estimada del gasto promedio, en el periodo post-activación, de los clientes que se activan y los que no corresponde a \$16.213 (diferencia significativa al 95% de confianza). Asimismo, la estimación del efecto se complementa con un enfoque de diferencia en diferencias, pues, la estimación de la diferencia en el periodo pre-activación entrega un resultado de \$1.373.

Variable	Periodo	Promedio en grupo que se activa	Promedio en grupo que no se activa	Diferencia	Diff in Diff
Gasto	Pre - activación	\$105.238	\$103.865	\$1.373	\$14.480
	Post - activación	\$136.739	\$120.526	\$16.213	

Tabla 13: Resumen del efecto causal de activación en la categoría Decoración sobre el gasto post-activación.
Fuente: Elaboración propia.

Por lo tanto, como se observa en la Tabla 13, se concluye que la diferencia neta en el gasto post-activación, de un cliente que se activa en la categoría Decoración, es de \$14.480, una variación del 10,6% con respecto al gasto post-activación en la categoría. Asimismo, esta diferencia puede ser observada en la Figura 29 que representa el efecto que se obtiene de aplicar el método de diferencia en diferencias.

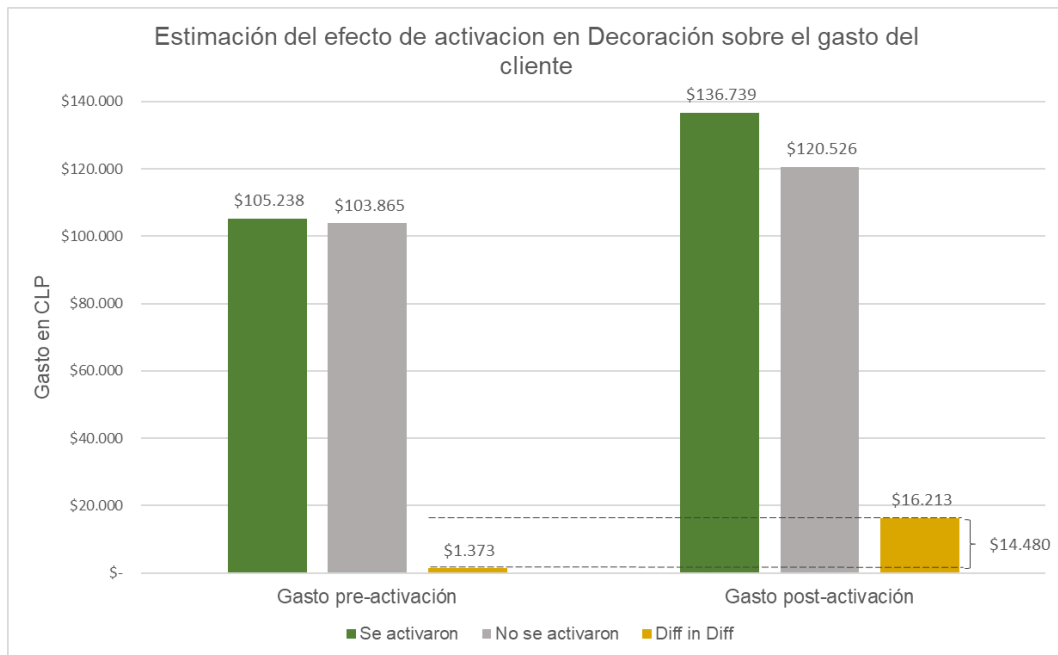


Figura 29: Efecto causal de activación en la categoría Decoración sobre el gasto del cliente en post-activación. Fuente: Elaboración propia.

El procedimiento anterior, detallado paso a paso, sólo para la categoría de Decoración en esta sección se lleva a cabo para todas las categorías de manera diferenciada y se expone en la sección Anexos S.

Por último, para la categoría analizada (Decoración), se lleva a cabo la diferenciación entre si este efecto en el gasto post-activación (\$14.480) corresponde a compras realizadas en la misma categoría que se encontraba antes inactiva o, por otro lado, si se lleva a cabo en otra diferente. De lo anterior, se concluye que, como muestra la Figura 30, este efecto corresponde a un aumento en el gasto en la misma categoría, es decir, el cliente, luego de activarse en la categoría decoración, durante los 3 meses posteriores vuelve a realizar compras en esta misma categoría.

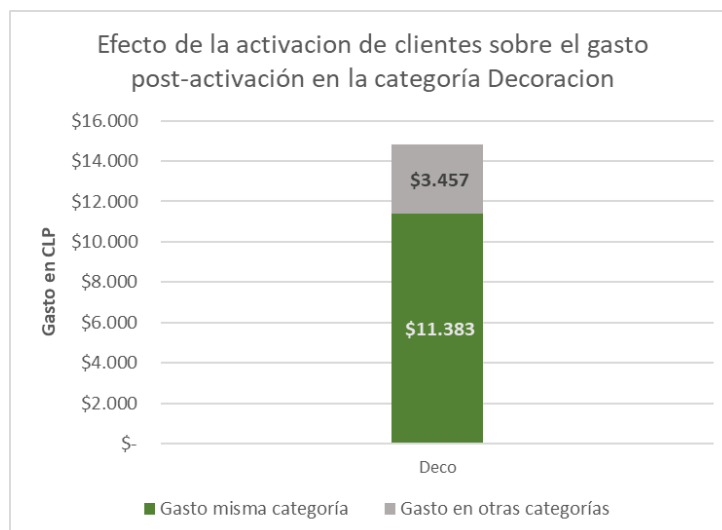


Figura 30: Efecto estimado de la activación de clientes sobre el gasto en el periodo post-activación en Decoración diferenciado por el tipo de compra. Fuente: Elaboración propia.

El procedimiento realizado para la categoría Decoración, acerca del efecto diferenciado del gasto post-activación se realiza para cada una de las otras categorías analizadas, lo que permite construir la Figura 31, que permite identificar si el gasto posterior se lleva a cabo en la misma categoría o en otras, y que resume el efecto causal de activación sobre el gasto promedio post-activación.

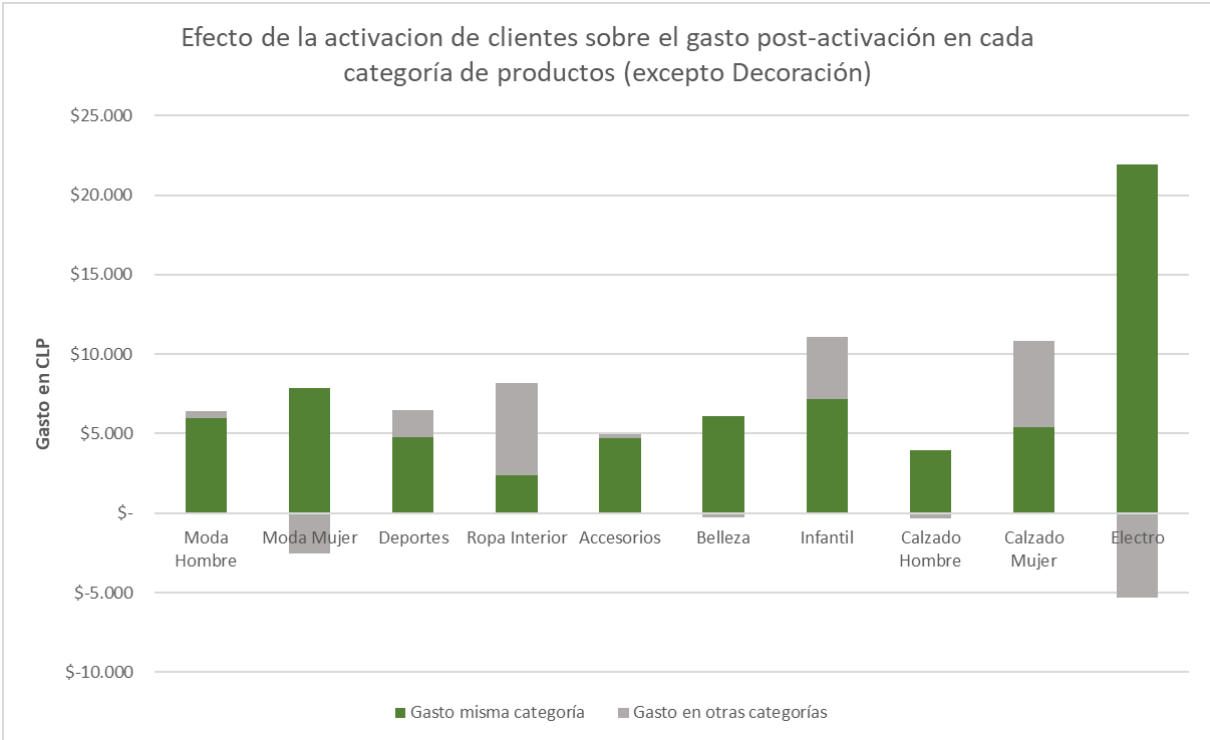


Figura 31: Efecto estimado de la activación de clientes sobre el gasto en el periodo post-activación en cada categoría estudiada.

Fuente: Elaboración propia.

De lo anterior es posible extrapolar que, la diferencia en el aumento del gasto en el periodo post-activación se muestra con mayor efecto en categorías como Electro hogar, Decoración, Calzado Mujer e Infantil que presentan un aumento mayor a \$10.000 luego de activarse. Por el contrario, las categorías de Calzado Hombre y Accesorios son las que presentan la diferencia en gasto más baja, con un efecto menor a \$5.000.

Cabe destacar que 2 de las 4 categorías que presentan un mayor efecto, son categorías que presentan gran presencia web y mayores montos de compra promedio, como lo son Electro hogar y Decoración, pues, gran parte de su oferta consiste en productos de gran volumen y mayor valor monetario.

Por otro lado, se observa que para 9 de las 11 categorías analizadas el aumento del gasto al activarse viene dado principalmente por un aumento en gasto en su misma categoría, lo que podría inferir que los clientes añaden una categoría adicional a su canasta de compra usual en la tienda y existe un efecto directo sobre la misma categoría. Por lo tanto, existe una oportunidad clara de aumentar los efectos cruzados entre categorías de producto en la mayoría de ellos que incremente el gasto posterior en otras categorías.

Finalmente, se destaca que, los efectos expuestos en las Figuras 29, 30 y 31 son positivos y significativos - a un 95% de confianza – por lo que se concluye que, existe un efecto causal observable de activarse en distintas categorías sobre el gasto post-activación de los clientes.

5.3.2 Mejor próxima categoría

En la presente sección se muestran los resultados para la asignación de la mejor próxima categoría a comprar según el comportamiento histórico de los clientes y sus características propias en virtud de los resultados obtenidos a través de la estimación de la propensión mediante los modelos Logit.

Como fue mencionado con anterioridad, los modelos Logit utilizados para esta sección corresponden a los mismos que fueron usados para el desarrollo del emparejamiento por *propensity score*, por lo que los detalles de su especificación pueden ser observados en la sección 5.3.1.1.

De la Figura 32 se observa la principal métrica predictiva para todas las categorías (AUC), donde el valor mayor se observa para la categoría Moda Mujer (72%) – categoría principal de la empresa – y el valor más bajo se observa en la categoría Accesorios (65%).

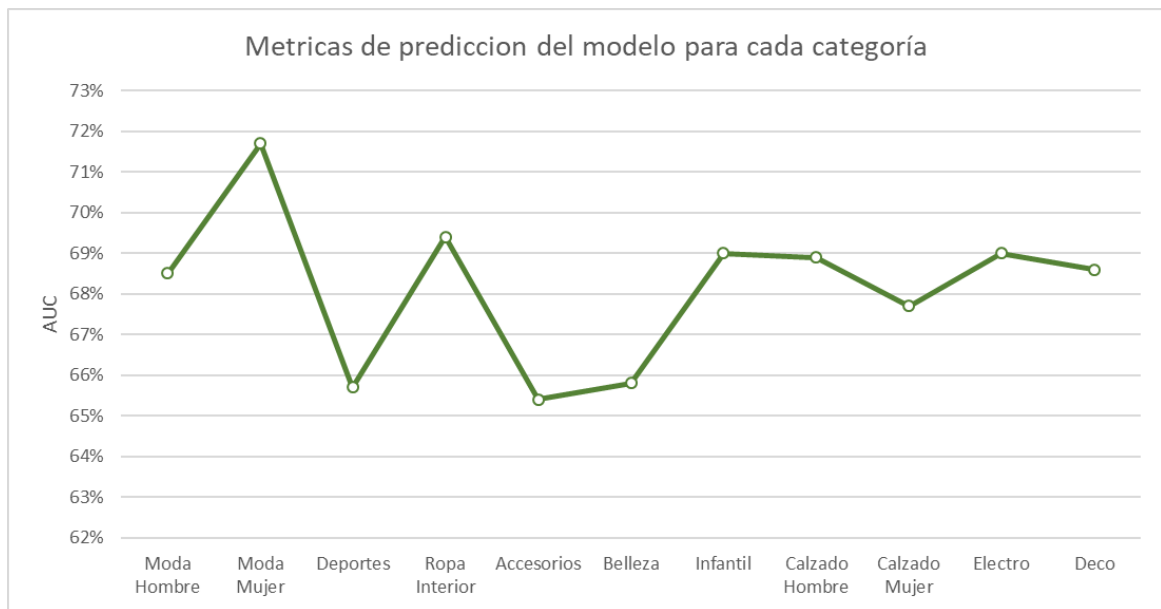


Figura 32: Resumen de métrica de predicción AUC para todas las categorías.
Fuente: Elaboración propia.

Sigue que, sin perder generalidad, nuevamente para esta etapa, el análisis en detalle de los coeficientes estimados y sus conclusiones se realizan para la categoría de Decoración. No obstante, se mencionan los principales hallazgos para las demás categorías. Los resultados que serán mencionados se observan en su totalidad en la sección Anexos P.

De los resultados estimados para las variables de cada modelo, se destaca que el coeficiente de mayor modulo corresponde al intercepto y para todas las categorías, es

negativo, lo que indica que la probabilidad de activarse, cuando no se ha realizado compra en la categoría durante los 3 meses anteriores, es muy baja. Este hecho se repite de manera similar para todas las categorías, siendo mayor su efecto en categorías como Moda Hombre, Belleza y Ropa Interior; y menor para categorías como Electro Hogar e Infantil. Lo anterior podría tener su explicación en los periodos de recompra de cada categoría, siendo estas últimas, categorías que pudiesen presentar una menor frecuencia de compra.

Con el objetivo de analizar de mejor forma los efectos de las distintas variables en la activación de las categorías se calculan sus efectos marginales promedio (AME, en inglés), los que pueden ser observados en detalle en Anexos P.

Para la categoría Decoración, se tiene que, para las variables ligadas a características del cliente (no ligadas a actividad en otras categorías ni a estacionalidad), como *Omnicana*⁹ y Solo tienda, que corresponden al tipo de canal de compra que utilizó el cliente en el último año, presentan un efecto negativo y significativo con un efecto marginal de disminución en la probabilidad de compra de un (5,11%) y un (10,29%) respectivamente. Lo anterior, se debe principalmente a que esta categoría y Electro Hogar presentan alta presencia web y los principales mecanismos de atracción e interacción con el cliente también pretenden intensificar constantemente el canal digital para estas categorías.

En la misma línea, con respecto a la zona de residencia del cliente, se observa que existe un mayor efecto marginal para clientes pertenecientes a la Zona Oriente metropolitana de un (2,96%) en el aumento de la probabilidad a activarse.

Cabe destacar que, con respecto a la navegación web en la misma categoría durante el mes anterior, esta presenta un efecto positivo y significativo, no obstante, su efecto marginal promedio es sólo de (0,25%).

Sigue que, con respecto a la actividad de compra en otras categorías durante el último mes previo a la activación, se observa que la gran mayoría tiene un efecto positivo y significativo (a excepción de Calzado Hombre) en la probabilidad de activarse. Esto guarda estrecha relación con la acumulación de categorías de compra, donde, a mayor cantidad de categorías distintas que el cliente compre, mayor es la probabilidad de activarse en la categoría estudiada.

Asimismo, se observa, para Decoración, que los efectos marginales promedio más grandes se encuentran en categorías como Electro Hogar (5,67%), Infantil (4,14%) y Moda Hombre (3,70%). Cabe destacar que para el primero existe una relación cruzada similar con respecto al efecto marginal en Decoración, por lo que existe una clara influencia entre ellos.

Al mismo tiempo, el efecto en Infantil podría verse explicado, por un lado, a través de la situación familiar en la que se encuentre el cliente, pues se podría inferir que los clientes que se activan en Decoración son clientes que en su mayoría poseen hijos(as).

⁹ Un cliente que presenta un comportamiento Omnicanal en el canal de compra quiere decir que utiliza tanto el canal físico como digital de la empresa durante el periodo de tiempo analizado (1 año)

Y, por otro lado, el efecto en Moda Hombre podría explicarse si se toma como supuesto que clientes del género masculino que se activan en Decoración, aprovechan la ocasión de compra también para adquirir otros productos para uso personal.

Por último, con respecto al análisis de la estacionalidad en la categoría, se observa que existe una gran influencia en las fechas de eventos especiales, tales como eventos Cyber en mayo y octubre con efectos marginales promedio de (3,36%) y (2,80%), y navidad en diciembre, con un efecto de (3,93%).

Así entonces, con estos resultados, es posible dividir el análisis en tres principales aristas de manera transversal para todas las categorías: relevancia explicativa de variables no ligadas a actividad de compra en otras categorías, sensibilidad a la compra en otras categorías y, por último, asociaciones cruzadas naturales y temáticas entre categorías.

5.3.2.1 Relevancia explicativa no ligada a actividad en categorías

En primer lugar, para las variables no ligadas a actividad, se tiene que, el canal de compra del cliente posee gran relevancia en categorías que por su naturaleza presentan comportamientos opuestos con respecto al uso de un canal u otro. En particular, existe un efecto marginal promedio positivo de (15%) y (17%) en clientes que son *omnicanales* y sólo tienda para categorías como Moda Hombre, Moda Mujer, Ropa Interior y Accesorios. Asimismo, en categorías como Electro Hogar y Decoración presenta un comportamiento opuesto, con un efecto promedio negativo de (7%) y (13%). Lo anterior se debe a que los primeros corresponden a productos que por naturaleza el cliente prefiere adquirir en el canal presencial por sobre el canal web y en los segundos, este último canal por naturaleza es el predilecto.

Por otro lado, se observa que la tenencia de una tarjeta abierta perteneciente a la empresa no presenta efectos mayormente relevantes en la activación de clientes a través de gran parte de las categorías, lo que podría significar una oportunidad de aumentar el uso de la misma, que incentive la activación de categorías en este tipo de clientes.

Asimismo, se observa un efecto transversal positivo en la navegación en la categoría analizada y en todas las demás categorías. Sin embargo, su efecto marginal promedio para todas es menor al 1%. Lo anterior se debe principalmente a que no es posible identificar la identidad de todos los clientes que navegan en la página web, por lo que existe un espacio de trabajo para mejorar el proceso de captura de esta información.

Finalmente, se observa que ser cliente antiguo entrega valores diversos para las diferentes categorías, por lo que no se observa una evidencia clara y consistente en la existencia de una brecha con respecto a los clientes nuevos, que permita concluir acerca de un comportamiento de compra diferenciable entre ellos.

5.3.2.2 Sensibilidad a la compra en otras categorías

Con relación a la actividad previa en otras categorías, se observa que existen 5 categorías de producto que aumentan de manera relevante y transversal la activación de

las demás categorías, estas son Moda Hombre (3,2%), Moda Mujer (3,5%), Infantil (3,9%), Electro Hogar (4,1%) y Decoración (3,6%). Por contraparte, categorías como Ropa Interior, Accesorios y Calzado Hombre sólo presentan efectos significativos en algunas categorías puntuales.

De lo anterior, se extrae que existen categorías que presentan un efecto general de apalancamiento en la activación de clientes en menor o mayor medida. Y, por otro lado, existen categorías que por su naturaleza u ocasión de compra incrementan la probabilidad de activación de manera particular en otra categoría y que podría ser utilizado como incentivo para generar efectos cruzados específicos.

5.3.2.3 Asociaciones cruzadas y temáticas entre categorías

Se observa de los resultados que existen efectos cruzados naturales entre categorías que afectan directamente en la activación de clientes. En particular, se tiene el efecto conjunto entre Moda Hombre y Moda Mujer, que pueden ser asociados a la misma ocasión de compra. Asimismo, se tiene el efecto conjunto entre 3 categorías, como lo son Electro Hogar, Decoración e Infantil.

Por otro lado, es posible destacar efectos temáticos como podrían ser Moda Hombre y Calzado Hombre; Moda Mujer y Calzado Mujer, Moda Mujer y Belleza, o bien, Moda Mujer y Accesorios.

El detalle de los efectos cruzados mencionados en esta sección se puede observar de forma ordenada en la sección de Anexos V.

6. CRITERIOS DE ASIGNACIÓN DE CATEGORÍAS Y USOS POSIBLES

De acuerdo con los resultados finales del trabajo desarrollado, existen diversos usos posibles que podrían ser útiles para la empresa. En particular, serán detallados dos: la elección de la próxima mejor categoría de activación para un cliente y la segmentación de clientes inactivos para envío de correos de marketing que provoquen efectos cruzados entre categorías.

Para abordar lo primero, se toma el caso en que el objetivo es tomar una decisión a nivel de un cliente particular, ya sea para todos los clientes o un grupo de ellos, que cumplan las características de inactivación, sobre cuál es la mejor próxima categoría en la que se encuentra inactivo y que se le podría comunicar.

Entonces, sigue que, se procede a realizar una predicción de mejor categoría a comprar para todos los clientes con categorías inactivas para el mes de octubre de 2019. Cabe mencionar que la elección de este mes se realiza con el objetivo de ejemplificar la capacidad de predicción del modelo para un caso particular (mes de octubre), sin perder generalidad, y que no representa una limitación del modelo, pues este procedimiento puede llevarse a cabo para cualquier mes del año que se quiera predecir. Siguiendo entonces, se tiene que, existen 1.258.102 clientes únicos con al menos una categoría inactiva para el mes de la predicción. De los resultados de las probabilidades de activación de los clientes analizados se extrae, por ejemplo, que para todas las categorías estudiadas existe un bajo nivel de activación natural para todos los clientes, es decir, existen pocos clientes con probabilidades mayores a 50% de activarse en al menos una categoría inactiva.

En particular, entre las categorías con mayor probabilidad de activación promedio para el mes de octubre, se encuentran Moda Mujer (3,5%) y Electro hogar (2,2%). Por un lado, esto se puede explicar pues Moda Mujer es la categoría que posee mayor volumen de compra anual y que no se observa estacionalidad evidente a lo largo de todos los meses. Con respecto al segundo, se tiene que, como se vio en secciones precedentes, tiene al canal web por canal de compra predilecto y dado que durante este mes se suelen dar eventos que incentivan este canal de compra, su probabilidad aumenta.

Por el contrario, categorías como Ropa Interior, Accesorios, Calzado Hombre, Deportes y Belleza presentan una cantidad de clientes con probabilidades de activación natural muy bajas, menos al 0,1% del total, por lo que se recomienda que la empresa realice acciones de marketing que permitan aumentar la cantidad de activaciones en estas categorías.

Es así como, para un mes específico, se estima que cerca de un 5,3 % de los clientes inactivos tiene alta probabilidad de activación (mayor a 50%) en alguna categoría, por lo que si efectivamente llegaran a activarse lo que repercutiría en un aumento del gasto de estos clientes en cerca de 380 millones de pesos (\$380.664.135) durante los tres meses siguientes (periodo post-activación), un 21,1% superior a los 3 meses previos a la activación.

Ahora bien, con el objetivo de proponer a la empresa un criterio que permita balancear tanto la propensión a la activación de cada cliente y el aumento del gasto estimado en la sección 5.3.1.4, se construyen cuatro reglas de elección que permitan escoger la mejor próxima categoría para cada cliente. Para desarrollar esto, a modo de ejemplo y con el objetivo particular de comparar los distintos criterios de asignación propuestos, se lleva a cabo una estimación de la probabilidad de activarse para el mes de julio de 2019 para todos los clientes que presenten inactivación en categorías. Luego, se define para los cuatro criterios a comparar (propensión, gasto esperado, *Lift* de propensión y *Lift* de gasto esperado), cuál dentro de las categorías no activas de los clientes es el mejor a comunicar. Por último, se concluye con el cálculo del gasto posterior a obtener (siguiendo el gasto post-activación estimado en la sección ya mencionada) y la precisión que se obtiene usando cada uno de estos criterios. Cabe mencionar que el criterio de gasto esperado, propensión y *Lift*'s se construyen siguiendo las métricas propuestas en la sección 4.3.

Con lo anterior, en la Figura 33, se observa una modificación de una curva de ganancia para definir el mejor criterio, donde, la base de clientes se divide en deciles y se ordenan de manera decreciente con respecto a su propensión a la mejor próxima categoría. Lo anterior se lleva a cabo para evaluar el porcentaje de activaciones que captura el criterio en cada decil.

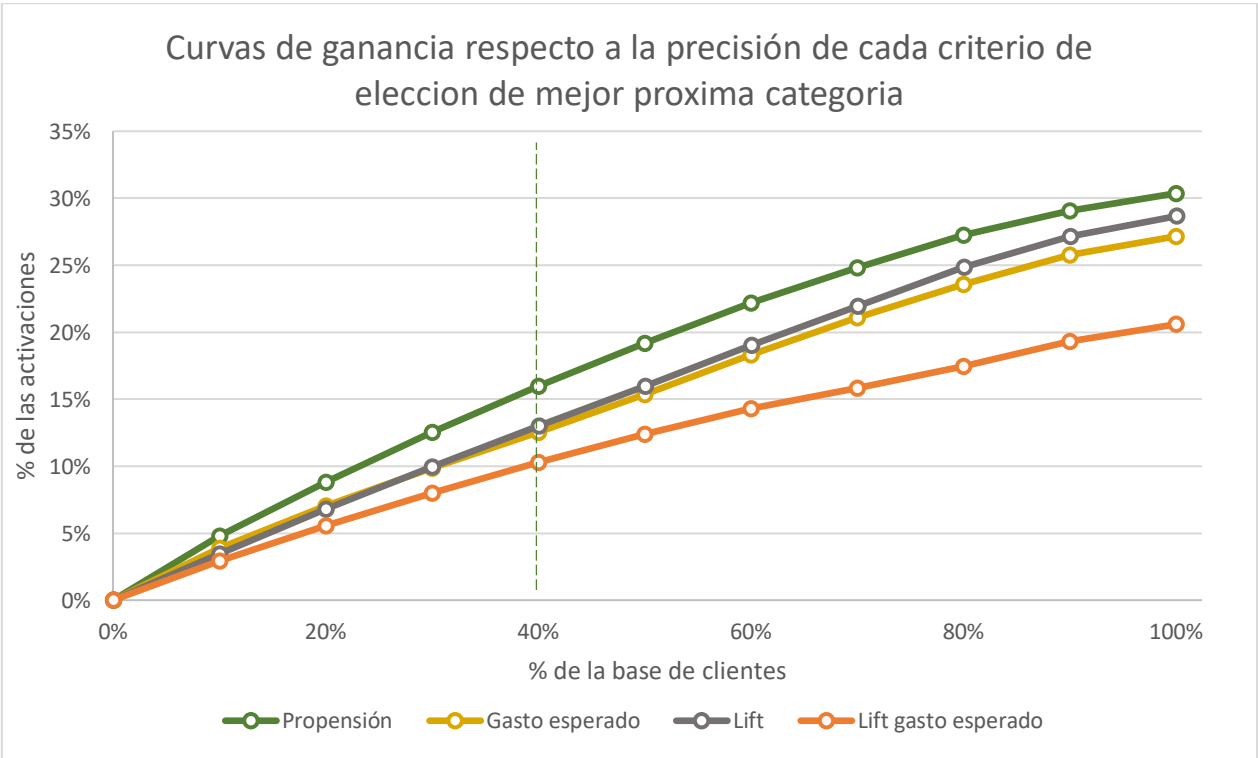


Figura 33: Ganancia respecto a la precisión de cada criterio de elección de mejor próxima categoría. Fuente: Elaboración propia.

Con respecto a los resultados de la comparación de los criterios, de la Figura 33 se identifica que, si se decide contactar al 100% de la base de clientes inactivos, en promedio, los criterios capturan cerca de un 30% de las activaciones efectivas, a excepción del ultimo criterio (*Lift* del gasto esperado). Si tomamos en cuenta sólo la

precisión, el mejor criterio corresponde a la propensión (30%), seguido por el *Lift* de la propensión (29%), el gasto esperado (27%) y el *Lift* del gasto esperado (21%). Por otro lado, con respecto al gasto estimado en post-activación que resulta de elegir la mejor próxima categoría para cada cliente, se obtiene el orden inverso, donde, en primer lugar, se encuentra el *Lift* del gasto esperado (\$16.220.170.392), luego el gasto esperado (\$13.246.671.348), el *Lift* de la propensión (\$10.187.626.410) y la propensión (\$9.460.701.335).

Asimismo, si por decisiones comerciales y/o por costos comunicacionales, se decide contactar sólo a una parte (40%) de la base de clientes mejor evaluados, se mantiene el orden mencionado con anterioridad con una mejora en la precisión de 40% para la propensión, 31% para el gasto esperado, 32% para el *Lift* y un 26% para el *Lift* del gasto esperado.

Lo que se obtiene guarda relación con que, si bien al elegir los criterios de gasto esperado o ambos *Lift* se pierde precisión, estos permiten activar en mayor medida categorías que son más difíciles de activar, como Calzado Mujer o Ropa Interior, pero que poseen mayores gastos posteriores estimados. Por el contrario, la propensión concede mayor peso a categorías más fáciles de activar, como lo son Moda Mujer y Moda Hombre que son los que se estimó, poseen un menor gasto post-activación. Lo anterior se observa en la tabla expuesta en la sección Anexos V, que muestra la distribución de clientes en la elección de la mejor próxima categoría para cada uno de los criterios analizados durante el mes de julio de 2019.

Por lo tanto, se concluye que, si la empresa sólo desea aumentar el gasto posterior del cliente la mejor decisión es usar el criterio de gasto esperado. Esto pues, a pesar de que el *Lift* del gasto esperado entrega mejores resultados en gasto posterior, pierde un 9% en precisión con respecto a la propensión. No obstante, de existir costos relevantes para contactar a toda la base de clientes se propone utilizar el *Lift* y en peores casos la propensión, debido a su mayor precisión.

Como fue mencionado, se detalla un segundo uso, el cual corresponde al apoyo al *targeting* de clientes (en este caso, para clientes inactivos) para su elección en ciertas campañas, o bien, comunicación con respecto a la categoría misma u otras. Cabe mencionar que, no existe un mecanismo actual que permita seleccionar la mejor categoría a ofrecer o comunicar para este tipo de clientes, por lo que este posible uso representa una primera forma de mejorar la interacción que tiene la empresa para con este tipo clientes.

En particular, se detalla el enfoque más utilizado por la empresa para la activación de clientes, que corresponde, como muestra la Tabla 14, al envío de un correo con una oferta de descuento que guarda relación con la frecuencia histórica de compra o con la última interacción que tuvo el cliente con la empresa. Sin embargo, este criterio no mide la sensibilidad que posee el cliente a ciertas categorías específicas.

Campaña	Tipo	Detalle
Frecuencia Omnicanal	Oferta 1	X% Dscto. en categoría más comprada
	Oferta 2	X% Dscto. última categoría comprada en Fcom
	Oferta 3	X% Dscto. última categoría comprada en Tienda

Tabla 14: Tipo de descuento usados para la activación de clientes.

Fuente: Elaboración propia.

De acuerdo a lo anterior, para el caso de este tipo de campañas, se proponen dos líneas de acción para el envío de correos siguiendo los resultados obtenidos mediante el modelo de propensión. Por un lado, si lo que se desea es activar al cliente en una categoría cualquiera, se propone segmentar a los clientes siguiendo el criterio de mayor propensión, similar a lo utilizado para elegir la mejor categoría a ofrecer. Por otro lado, si lo que se desea es comunicar una categoría o marca en particular para un conjunto de clientes, se propone utilizar el segmento con la mayor propensión de la categoría específica siguiendo los resultados de propensión entregados por el modelo Logit.

Finalmente, de acuerdo los resultados encontrados mediante los modelos Logit, se propone un criterio de *targeting* para provocar efectos de compra cruzada de categorías. Lo que se busca obtener con esto es que clientes que hayan comprado una categoría cualquiera, pero posean una categoría inactiva, se les comunique una oferta en la categoría que se encuentra inactiva para incentivar su activación y/o efectos cruzados.

En particular, por ejemplo, para la categoría Moda Hombre, como se muestra en la Tabla 15, las categorías que más relevancia poseen en el aumento de la probabilidad de activación de la categoría son Calzado Hombre (9,28%) y Deportes (6,82%). El detalle del ordenamiento, en relación con el cálculo de los efectos marginales promedio, se muestra para todas las categorías en la sección Anexos W.

Moda Hombre	
Categoría	AME
Calzado hombre	9,28%
Deportes	6,82%
Infantil	6,60%
Electro	6,06%
Moda mujer	6,05%
Deco	5,16%
Belleza	5,12%
Ropa interior	3,56%
Calzado mujer	3,17%

Tabla 15: Relevancia de categorías que según su actividad de compra previa aumentan la probabilidad de activarse en moda hombre.

Fuente: Elaboración propia.

Por lo tanto, para la segmentación de clientes en los que se quiera activar y comunicar sólo la categoría específica que se encuentre inactiva (como en Anexos X), se propone seleccionar a clientes que durante su última compra hayan adquirido algunas de las categorías que más efecto posean sobre la activación. Por ejemplo, para activar Moda Hombre, seleccionar clientes que hayan comprado en primera instancia Calzado Hombre y/o Deportes, para luego sucesivamente, dependiendo de la cantidad de clientes que se requieran para la campaña, se añadan los clientes que hayan adquirido las categorías siguientes en el ordenamiento expuesto en la Tabla 15. Como ejemplo, se elabora la gráfica de envío, como se expone en la sección Anexos Z (a) para activación de Moda Hombre.

Por otro lado, si lo que se quiere comunicar en el correo es más de una categoría, como se muestra en el correo genérico enviado por la empresa en Anexos Y, se propone que si se quiere incentivar la recompra en la categoría Calzado Hombre o Deportes, también se añadan al envío, comunicación de contenido u ofertas de la categoría Moda Hombre, que incentive tanto la compra cruzada como su activación, como se muestra en la gráfica expuesta en la sección Anexos Z (b).

Asimismo, si la comunicación sigue siendo para incentivar la compra cruzada y, por consecuencia, se muestra más de una categoría, se tiene que, si el deseo es activar más de una categoría, se comuniquen las categorías a activar dependiendo de la relevancia del efecto que se posea en una categoría en que ambos coincidan. En particular, por ejemplo, si se desea comunicar Moda mujer y Accesorios o Moda Mujer, Accesorios e Infantil para clientes con estas categorías inactivas, se filtre por clientes que hayan comprado Belleza durante el último mes. Cabe destacar que en ningún caso esta regla pretende reemplazar en su totalidad a las demás, pues, como se observa en Anexos W para la categoría Infantil, una variable explicativa clave corresponde a la existencia de hijos, por tanto, el primer filtro que se debiese aplicar para este tipo de activaciones corresponde a la existencia de hijos. A modo de ejemplo, se expone en Anexos Z (c) el caso de incentivos de Moda Mujer y Accesorios, y, asimismo, incentivos para Moda Mujer, Accesorios e infantil en el caso que se desease añadir la variable existencia de hijos(as) en Anexos Z (d).

Es importante mencionar por último que, para asignar las categorías que presentan mayor relevancia en la activación, se considera aquellos efectos marginales que se encuentran por encima del promedio general de efecto sobre todas las categorías analizadas, pues se quiere evitar asumir que la significancia entregada por el modelo de predicción represente en su totalidad el efecto cruzado que exista entre categorías. Por lo tanto, se consideran relevantes los efectos marginales mayores a 3,72%.

7. CONCLUSIONES

El contexto de la industria del retail, respecto al aumento de la oferta de competidores, y la situación de la empresa, respecto al cambio en el comportamiento de los clientes, requería que se llevaran a cabo diversos cambios de estrategia en el concepto de la personalización de ofertas que incentiven la recompra o aumento de frecuencia de compra en el cliente.

La identificación y análisis de la diversidad de categorías que adquiere un cliente, su estado (activo, inactivo o fugado) y su frecuencia de compra son conceptos de interés para la empresa, puesto que, una mejora en los resultados implicaría, de una u otra forma, un mayor nivel de retención y fidelidad, medidas consideradas de alto valor para la organización. Lo anterior marca un objetivo claro, que se sustenta en que un cliente con una categoría activa adicional gasta en promedio \$203.388 más durante un año y que este efecto aumenta a medida que más categorías permanecen activas durante el año.

Es de acuerdo a lo expuesto que se articula la interrogante de cuál o cuáles categorías debiesen ser activadas, debido a que no todos los clientes poseen la misma propensión de compra a todas las categorías por igual, ni todas las categorías representan el mismo gasto y retorno para la empresa. Por lo tanto, el objetivo general de esta memoria de título consistía en evaluar el efecto de la activación en categorías inactivas y el efecto cruzado entre ellas, mediante el diseño y construcción de un modelo de predicción.

Para resolver la oportunidad identificada, se trazaron cuatro objetivos específicos. En primer lugar, la estimación del efecto causal que posee activarse en una categoría de producto específica sobre el gasto post-activación del cliente, mediante la técnica de emparejamiento por distancia de Mahalanobis dentro de *calipers* por *propensity score*, que garantizaba que la comparación del efecto de gasto posterior se ejecutara sobre conglomerados de clientes con similar distribución en su propensión a la activación y el gasto que habían realizado durante los 3 meses previos (periodo pre-activación).

La técnica utilizada definió que existe un efecto causal positivo y significativo sobre el gasto post-activación de los clientes en todas las categorías analizadas. En particular, las categorías de producto que evidenciaron un mayor aumento en el gasto posterior a la activación fueron Electro hogar, Decoración, Infantil y Calzado mujer con gastos de \$16.633, \$14.840, \$11.053 y \$10.838 respectivamente. Por contrapartida, Calzado Hombre y Accesorios generaron un menor aumento en el gasto con \$3.588 y \$4.963, respectivamente. Así entonces, se propone a los cuatro primeros como objetivos finales acordados si se desea capitalizar el gasto post-activación de los clientes.

Sigue que, el segundo objetivo específico consistió en el diseño y construcción de un modelo de predicción que permitiese estimar las probabilidades de activación de los clientes en cada una de las categorías analizadas. Con este propósito, se elige el Logit del tipo binario, debido a sus menores costos de procesamiento de la información y a su facilidad interpretativa.

De lo anterior, se obtuvo que existen diversos comportamientos en la probabilidad de activación para cada categoría y al mismo tiempo, a la sensibilidad que existe en la compra previa de otras categorías en la activación de la categoría analizada. Por un lado, se observa que existe una alta relevancia del canal de compra – sea tienda o web -, género del cliente y estacionalidad – como los eventos de *Cyber's* y Navidad - en algunas categorías. Por otro lado, se observa que existen categorías que aumentan la probabilidad de activación en categorías específicas, y que, por tanto, las estrategias para incentivar la activación debiesen ser diferenciadas, y otras, que apalancan de manera transversal la activación de otras categorías como Electro hogar e Infantil, y que, por lo tanto, se propone incentivar siempre la compra en estas últimas.

A partir de los resultados que se obtuvieron de la propensión, se corroboró que existe, en promedio, un 5% de los clientes con categorías inactivas que posee altas probabilidades de activarse de forma natural en al menos una categoría durante el próximo mes, lo que tendría como efecto observable, un aumento promedio del gasto de un 21% en los próximos 3 meses con respecto a su gasto en los 3 meses previos a la activación y un aumento promedio estimado de un 7% con respecto a los que no se activaron.

Luego, como tercer objetivo se tenía definir un criterio de asignación de categorías que permitiese balancear para cada cliente, la propensión a activar una categoría y el gasto post-activación estimado. De ello, se concluye que, si la empresa sólo desea aumentar el gasto posterior del cliente se utilice el criterio de gasto esperado, pues conlleva un 30% y 40% más de gasto que el *Lift* de propensión y la propensión, respectivamente, y un 6% más de precisión que el *Lift* de gasto esperado. Se destaca que, si se desea lograr un equilibrio entre ambas decisiones (precisión y gasto posterior) la opción más adecuada corresponde al *Lift* de propensión.

El objetivo específico cuarto y final, consistía en definir un criterio, similar a lo realizado en el objetivo anterior, con el fin de mejorar la segmentación de clientes para el envío de correos de marketing que permitiese incentivar efectos de compra cruzada entre categorías. Se concluye que existen efectos cruzados relevantes para 9 de las 11 categorías analizadas y se propone seguir una metodología de comunicación a través del envío de correos que permita generar tanto recompra como activación de categorías para un cliente seleccionado.

Finalmente, se concluye que el método propuesto y seguido en este trabajo de memoria de título logra dar solución a la totalidad de los objetivos propuestos, lo que determina en sus resultados, tanto intermedios como finales, la relevancia que posee la selección precisa y personalizada de las categorías que se ofrecen a un cliente para su activación.

8. LIMITACIONES Y TRABAJOS FUTUROS

A pesar de los resultados obtenidos del trabajo anterior y la consecución de los objetivos planteados, existe una serie de limitaciones que dan paso a diversas líneas de trabajos futuros que pudiesen explorar tanto nuevas interrogantes que nacen de este estudio como otros que pudiesen perfeccionar las técnicas empleadas y los resultados obtenidos.

En primer lugar, el trabajo se llevó a cabo mediante el análisis de clientes con categorías inactivas, donde, no se separaron de la muestra a clientes antiguos y nuevos, ni tampoco, a clientes que sólo utilizaran un canal de compra, como canal de predilección, sino que sólo fueron incluidas como variables explicativas para el modelo.

De los resultados con respecto a la antigüedad del cliente, no se encontraron diferencias concluyentes entre categorías, por lo que un estudio adicional que permita encontrar fehacientemente si existe evidencia cuantificable que permita diferenciar tratamientos hacia estos clientes queda propuesto.

Por otro lado, con respecto al canal de compra, se encontró que existen diferencias relevantes en el canal de compra para algunas categorías, por lo que se plantea como un trabajo complementario determinar cómo afecta este mismo enfoque diferenciando a cada cliente por su respectivo canal de compra predilecto y los mecanismos que pudiesen desencadenar activaciones diferentes a la encontradas en este trabajo.

Asimismo, una segunda limitación fue la no inclusión de la coocurrencia de compra entre categorías o cómo afecta la activación conjunta de categorías, es decir, la relación subyacente que existe en que dos categorías distintas se adquieran en la misma ocasión de compra y que efectos pudiesen tener sobre las activaciones. Por lo tanto, se propone como trabajo futuro, un perfeccionamiento a esta solución, el incluir la coocurrencia en el desarrollo de un modelo de activación con un enfoque igual o similar a este a través del uso del Probit multivariado mediante estimación bayesiana (Manchanda, Ansari & Gupta, 1999) [35].

Otra limitante se encuentra en que este trabajo utiliza como insumo principal el comportamiento histórico del cliente para definir las posibles acciones futuras que se pudiesen tomar sobre ellos. Por tanto, se propone estudiar el caso para clientes que no poseen registro histórico en la empresa y las conductas que pudiesen desencadenar las activaciones en diferentes categorías.

Como cuarta limitación, el presente trabajo dispuso como insumo la clasificación de categorías a nivel de líneas de gestión de productos utilizada por la empresa mediante criterio comercial. Lo anterior, otorga la posibilidad de investigar posibles nuevas agrupaciones no evidentes dentro de cada categoría, o bien, robustecer los resultados obtenidos mediante el estudio a nivel de líneas, sub-líneas o marcas.

En quinto lugar, se tiene la no inclusión ni estimación (por temas de procesamiento y disponibilidad de datos) de los efectos de marketing directo a través de envíos de correo

electrónico o mensajes de texto en la probabilidad de activación de categorías. Esta limitante se originó principalmente, pues, al momento de desarrollar este estudio no existía la información consolidada de estos resultados que permitiera hacer inferencia con respecto a las interacciones históricas de la empresa con estos clientes antes de la activación.

Finalmente, un punto importante a destacar es que los resultados finales que se obtuvieron reflejan como una generalidad el comportamiento de los clientes inactivos en categorías, tanto para la estimación del efecto causal de activación como también para la estimación de los modelos de mejor próxima categoría, no obstante, sería relevante conocer como podría ser este comportamiento en diferentes conglomerados de clientes, en particular, segmentos de alto valor (Clientes F y Super F¹⁰), entre otros. De acuerdo a lo anterior, un enfoque de mejora sería considerar la posibilidad de realizar estimaciones diferenciadas para este tipo de clientes, o, por otro lado, incluir esta diferenciación en los modelos desarrollados mediante el uso de estimadores bayesianos.

¹⁰ Clientes F y Super F corresponden a clientes que agrupan el mayor volumen de compra de la empresa, en general, gastos anuales superiores a los \$500.000 y más de 7 visitas con compra anual.

9. BIBLIOGRAFÍA

- [1] «S.A.C.I. FALABELLA - Memorias Anuales - SVS». [En línea]. Disponible en: <http://www.cmfchile.cl/institucional/mercados/entidad.php?mercado=V&rut=90749000&grupo=&tipoentidad=RVEMI&row=AAAwy2ACTAAABBy8AAG&vig=VI&control=svs&pestanía=49> [accedido may 17, 2019].
- [2] «Industria del Retail», *CERET*. [En línea]. Disponible en: <http://www.ceret.cl/industria-del-retail/>. [accedido: sep. 30, 2019].
- [3] A. Guerrero, «Falabella, Cencosud y Ripley concentran el 86% del negocio de las multitiendas en Chile», *América Retail*, dic. 14, 2015. [En línea]. Disponible en: <https://www.america-retail.com/weekly/falabella-cencosud-y-ripley-concentran-el-86-del-negocio-de-las-multitiendas-en-chile/> [accedido sep. 30, 2019].
- [4] B. Marr, «The 10 Best Examples Of How Companies Use Artificial Intelligence In Practice», *Forbes*. [En línea]. Disponible en: <https://www.forbes.com/sites/bernardmarr/2019/12/09/the-10-best-examples-of-how-companies-use-artificial-intelligence-in-practice/> [accedido may 01, 2020].
- [5] L. Columbus, «State Of AI And Machine Learning In 2019», *Forbes*. [En línea]. Disponible en: <https://www.forbes.com/sites/louiscolumbus/2019/09/08/state-of-ai-and-machine-learning-in-2019/> [accedido may 01, 2020].
- [6] J. D. Gregorio, «Crecimiento económico en Chile: Evidencia, fuentes y perspectivas». Centro de Estudios Públicos. 2016. [En línea]. Disponible en: https://www.cepchile.cl/cep/site/docs/20160304/20160304093619/r98_degregorio_crecimi-economic.pdf. [Accedido: dic. 07, 2019].
- [7] Willem F. Schol, «Chile: El mundo del retail está cambiando y cambiará más, ¿estamos listos para ello? - América Retail». [En línea]. Disponible en: <https://www.america-retail.com/chile/chile-el-mundo-del-retail-esta-cambiando-y-cambiara-mas-estamos-listos-para-ello/> [accedido sep. 23, 2019].
- [8] R. Arrieta, «El nuevo entorno regulatorio de la protección de datos personales en Chile», 2019. [En línea]. Disponible en: <https://iapp.org/news/a/el-nuevo-entorno-regulatorio-de-la-proteccion-de-datos-personales-en-chile/> [accedido may 03, 2020].
- [9] P. P. Gutierrez, «Protección de Datos Personales en Chile y GDPR», *La Tercera*, abr. 27, 2018. [En línea]. Disponible en: <https://www.latercera.com/opinion/noticia/proteccion-datos-personales-chile-gdpr/146187/> [accedido may 03, 2020].
- [10] «Falabella reduce sus utilidades un 38% en el primer trimestre de 2019, pero sigue creciendo en e-commerce», *La Tercera*, may 14, 2019. [En línea]. Disponible en: <https://www.latercera.com/pulso/noticia/falabella-reduce-utilidades-38-primer-trimestre-2019-sigue-creciendo-e-commerce/656009/> [accedido may 17, 2019].

- [11] «NGDATA | What is Next Best Offer? Meaning of and Best Practices», *NGDATA*, ene. 05, 2017. [En línea]. Disponible en: <https://www.ngdata.com/what-is-next-best-offer/> [accedido dic. 27, 2019].
- [12] G. Kihlstrom, «Council Post: Getting Started With Next-Best-Action Marketing», 2019. [En línea]. Disponible en: <https://www.forbes.com/sites/forbesagencycouncil/2019/09/25/getting-started-with-next-best-action-marketing/#6021fbc9174b> [accedido dic. 27, 2019].
- [13] A. Knott, A. Hayes, y S. Neslin, «Next-product-to-buy models for cross-selling applications - Knott - 2002 - Journal of Interactive Marketing - Wiley Online Library», 2002. [En línea]. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1002/dir.10038> [accedido dic. 23, 2019].
- [14] M. Vargas, «Estimación del Modelo Probit Multivariante: Una Mejora», *University Library of Munich, Germany, MPRA Paper*, ene. 2003.
- [15] A. Prinzie y D. Van den Poel, «Random Forests for multiclass classification: Random MultiNomial Logit», *Expert Systems with Applications*, vol. 34, n.º 3, pp. 1721-1732, abr. 2008, doi: 10.1016/j.eswa.2007.01.029.
- [16] R. J. Adams, M. Wilson, y W. Wang, «The Multidimensional Random Coefficients Multinomial Logit Model», *Applied Psychological Measurement*, vol. 21, n.º 1, pp. 1-23, mar. 1997, doi: 10.1177/0146621697211001.
- [17] G. Engelhard, *Objective Measurement: Theory Into Practice*. Greenwood Publishing Group, 1992.
- [18] M. Goic, «Apunte.pdf». Apunte preliminar para el ramo de Marketing II. Facultad de ciencias físicas y matemáticas, Universidad de Chile, 2015. Disponible en: https://www.ucursos.cl/usuario/1428d1e5203d7988008b967e79a8b468/mi_blog/r/Apunte.pdf. [Accedido: dic 27, 2019]
- [19] E. A. Stuart, «Matching methods for causal inference: A review and a look forward», *Stat Sci*, vol. 25, n.º 1, pp. 1-21, feb. 2010, doi: 10.1214/09-STS313.
- [20] P. R. Rosenbaum y D. B. Rubin, «The central role of the propensity score in observational studies for causal effects», *Biometrika*, vol. 70, n.º 1, pp. 41-55, abr. 1983, doi: 10.1093/biomet/70.1.41.
- [21] Sascha O. Becker, Andrea Ichino, «Estimation of Average Treatment Effects Based on Propensity Scores - 2002». Accedido: oct. 16, 2019. [En línea]. Disponible en: <https://journals.sagepub.com/doi/abs/10.1177/1536867X0200200403>.

- [22] D. B. Rubin y N. Thomas, «Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates», *Journal of the American Statistical Association*, vol. 95, n.º 450, pp. 573-585, jun. 2000, doi: 10.1080/01621459.2000.10474233.
- [23] D. G. Altman y J. M. Bland, «Diagnostic tests. 1: Sensitivity and specificity», *BMJ*, vol. 308, n.º 6943, p. 1552, jun. 1994, doi: 10.1136/bmj.308.6943.1552.
- [24] C. Shearer, «The CRISP-DM model: the new blueprint for data mining», *J Data Warehouse*, vol. 5, pp. 13-22, ene. 2000.
- [25] U. Fayyad, G. Piatetsky-Shapiro, y P. Smyth, «From Data Mining to Knowledge Discovery in Databases», 1, vol. 17, n.º 3, pp. 37-37, mar. 1996, doi: 10.1609/aimag.v17i3.1230.
- [26] A. I. R. L. Azevedo y M. F. Santos, «KDD, SEMMA and CRISP-DM: a parallel overview», *IADS - DM*, 2008, [En línea]. Disponible en: <https://recipp.ipp.pt/handle/10400.22/136>. [accedido: jun. 28, 2019].
- [27] M. Caliendo y S. Kopeinig, «Some Practical Guidance for the Implementation of Propensity Score Matching», *Journal of Economic Surveys*, vol. 22, n.º 1, pp. 31-72, 2008, doi: 10.1111/j.1467-6419.2007.00527.x.
- [28] G. Schwarz, «Estimating the Dimension of a Model», *Ann. Statist.*, vol. 6, n.º 2, pp. 461-464, mar. 1978, doi: 10.1214/aos/1176344136.
- [29] A. E. Raftery, «A Model for High-Order Markov Chains», *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 47, n.º 3, pp. 528-539, 1985, doi: 10.1111/j.2517-6161.1985.tb01383.x.
- [30] D. E. Ho, K. Imai, G. King, y E. A. Stuart, «MatchIt : Nonparametric Preprocessing for Parametric Causal Inference», *J. Stat. Soft.*, vol. 42, n.º 8, 2011, doi: 10.18637/jss.v042.i08.
- [31] R. H. Dehejia y S. Wahba, «Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs», *Journal of the American Statistical Association*, vol. 94, n.º 448, pp. 1053-1062, dic. 1999. doi: 10.1080/01621459.1999.10473858.
- [32] M. Lechner, «A Note on the Common Support Problem in Applied Evaluation Studies», Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 259239, nov. 2000. [En línea]. Disponible en: <https://papers.ssrn.com/abstract=259239>. [accedido: dic. 24, 2019].

- [33] P. R. Rosenbaum y D. B. Rubin, «Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score: The American Statistician: Vol 39, No 1», 1985. [En línea]. Disponible en: <https://amstat.tandfonline.com/doi/abs/10.1080/00031305.1985.10479383> [accedido dic. 27, 2019].
- [34] H. L. Smith, «Matching With Multiple Controls to Estimate Treatment Effects in Observational Studies», 1997. [En línea]. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9531.271030> [accedido dic. 27, 2019].
- [35] P. Manchanda, A. Ansari, y S. Gupta, «The “Shopping Basket”: A Model for Multicategory Purchase Incidence Decisions», *Marketing Science*, vol. 18, n.º 2, pp. 95-114, may 1999, doi: 10.1287/mksc.18.2.95.

10. ANEXOS

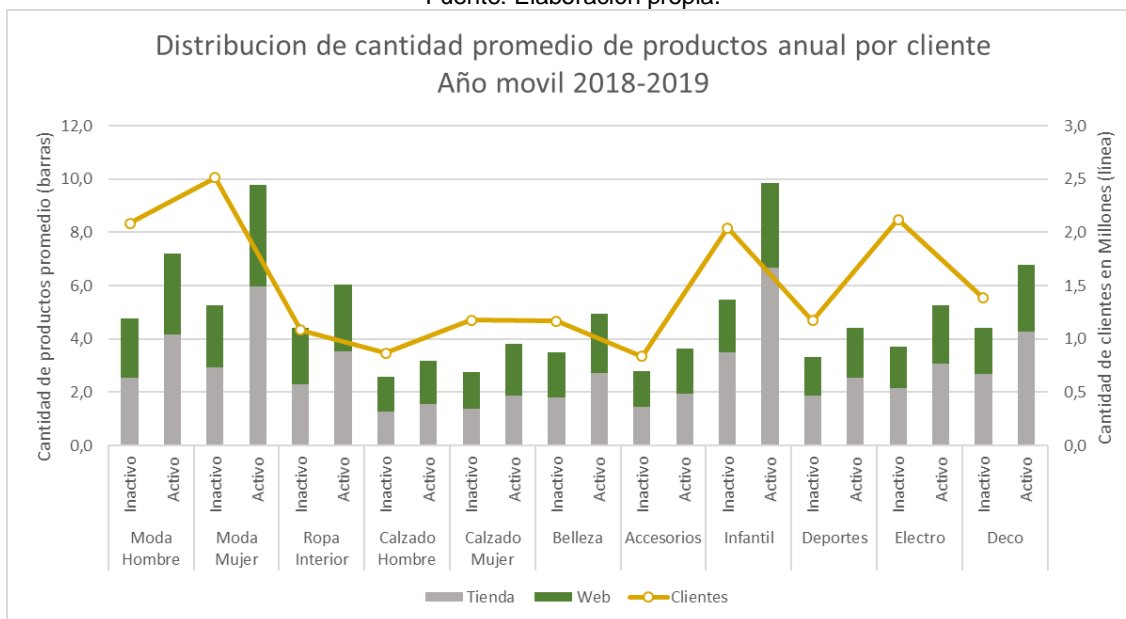
Anexos A: Jerarquías de gestión de productos.
Fuente: Elaboración propia.

Linea - Sublinea	Descripcion
J01	Hombres
J02	Rincon juvenil hombres
J03	Deportes
J04	Damas
J05	Rincon juvenil Damas
J06	Ropa interior
J07	Accesorios mujer
J08	Perfumeria
J09	Niños
J10	Calzado
J11	Electro hogar
J12	Blanco
J13	Muebles
J14	Dormitorio
J15	Menaje
J16	Decoracion
J17	Regalos
J18	Gourmet y concesiones
J32	Crater & Barrel

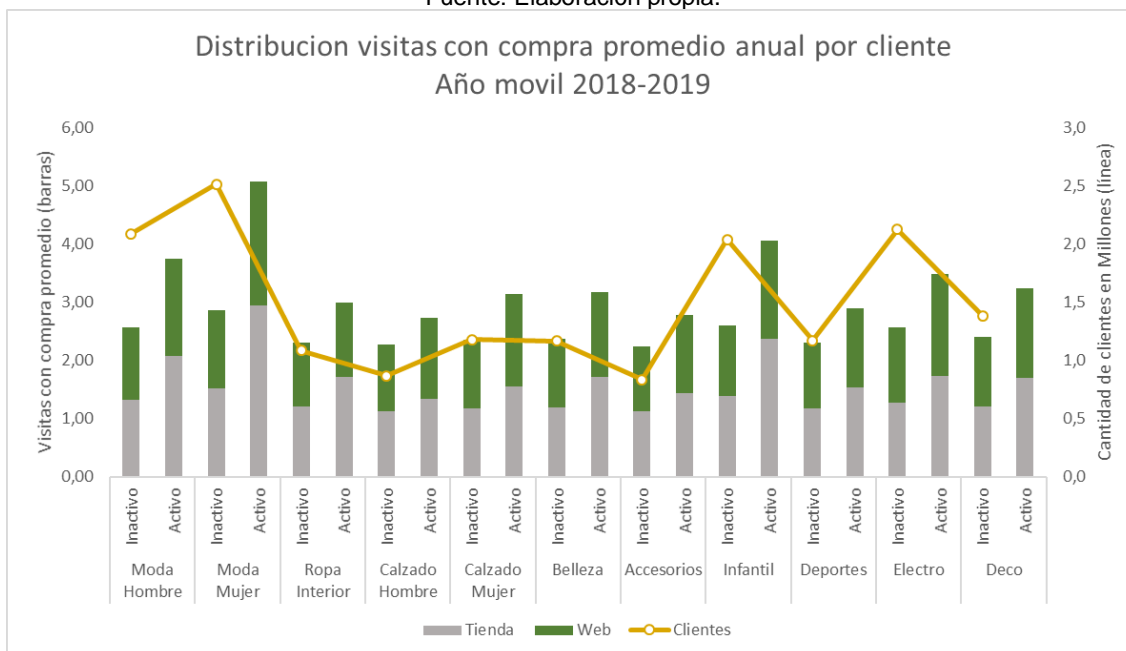
Anexos B: Agrupación de jerarquías de gestión de productos.
Fuente: Elaboración propia.

Categorías	Jerarquías (líneas de productos)	Identificador de línea
Moda hombre	Hombre – Juvenil hombre	J01-J02
Moda mujer	Mujer – Juvenil mujer	J04-J05
Calzado hombre	Calzado	J1001-J1003
Calzado mujer	Calzado	J1002- J1004
Deportes	Deportes	J03
Ropa Interior	Ropa interior	J06
Accesorios	Accesorios mujer	J07
Belleza	Perfumería	J08
Infantil	Infantil	J09
Electro	Electro hogar	J11
Deco	Blanco – Muebles – Dormitorio – Menaje – Decoración – Regalos – Gourmet – Crate & Barrel	J12-J13-J14-J15-J16-J17- J18-J32

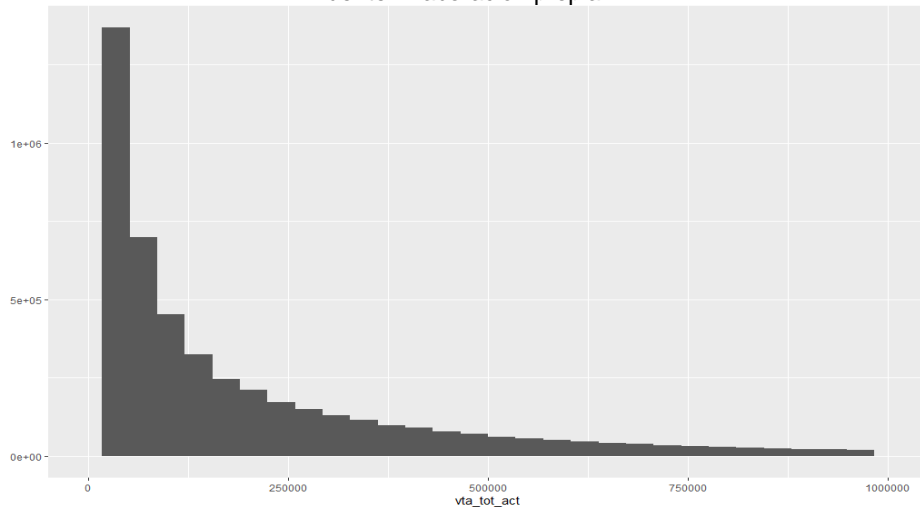
Anexos C: Cantidad de productos comprados promedio en 1 año en cada categoría para clientes activos e inactivos.
Fuente: Elaboración propia.



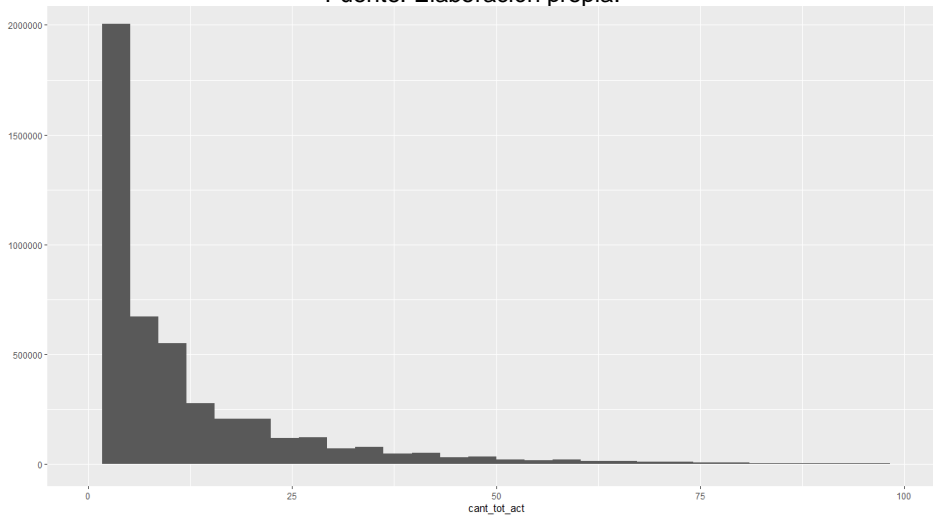
Anexos D: Visitas con compra promedio anual en cada categoría para clientes activos e inactivos.
Fuente: Elaboración propia.



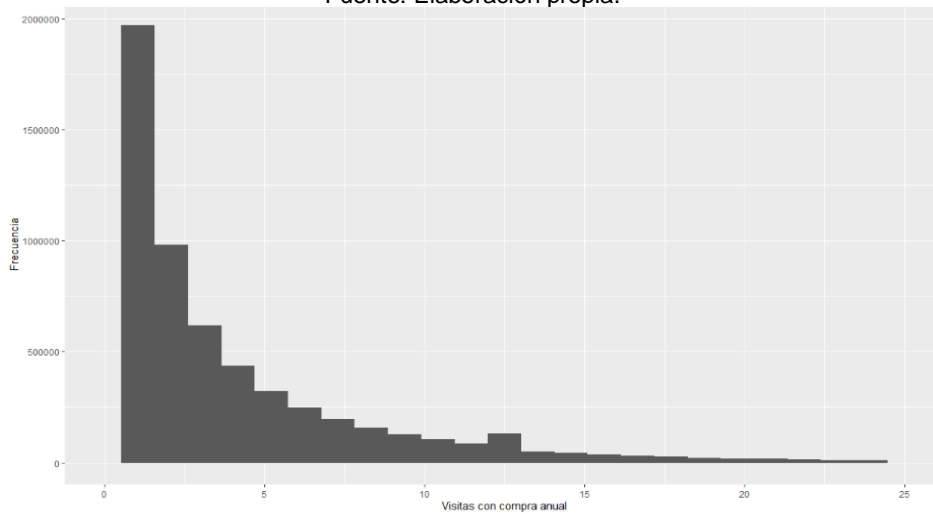
Anexos E: Distribución de frecuencia de montos líquidos de venta anuales por cliente.
Fuente: Elaboración propia.



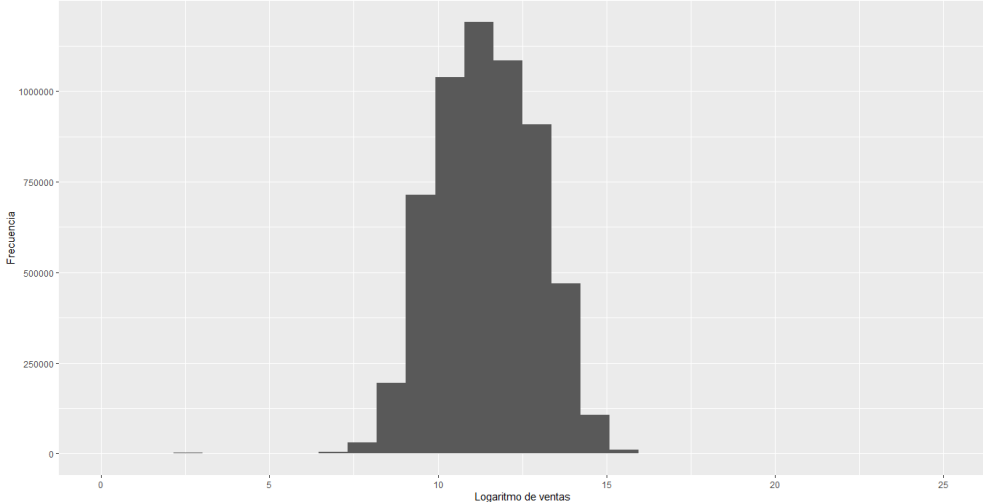
Anexos F: Distribución de frecuencia de cantidad de productos comprados anuales.
Fuente: Elaboración propia.



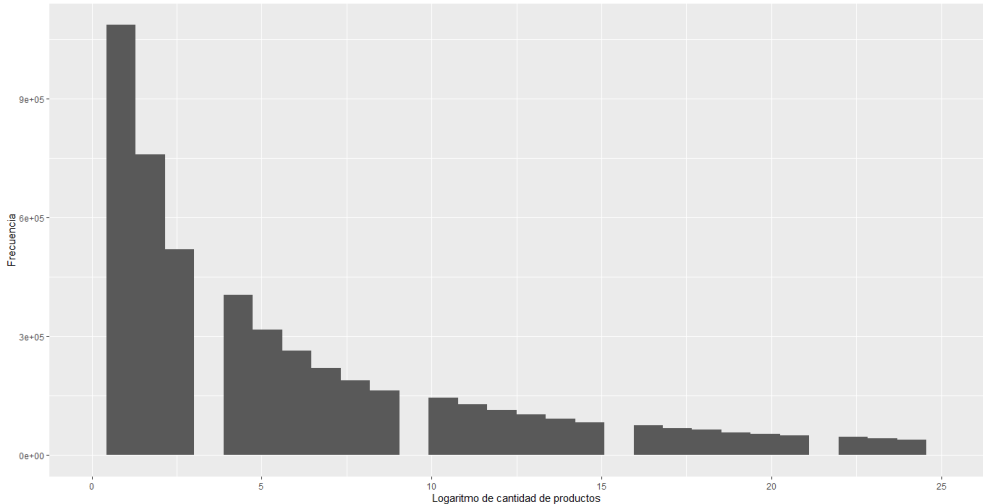
Anexos G: Distribución de visitas con compra anual por cliente.
Fuente: Elaboración propia.



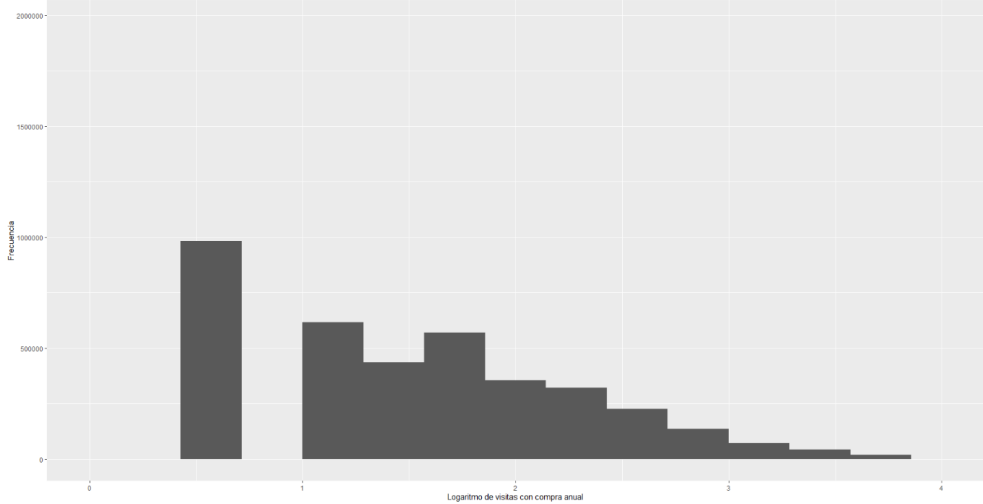
Anexos H: Transformación logarítmica de la frecuencia de montos líquidos de venta anuales por cliente.
Fuente: Elaboración propia.



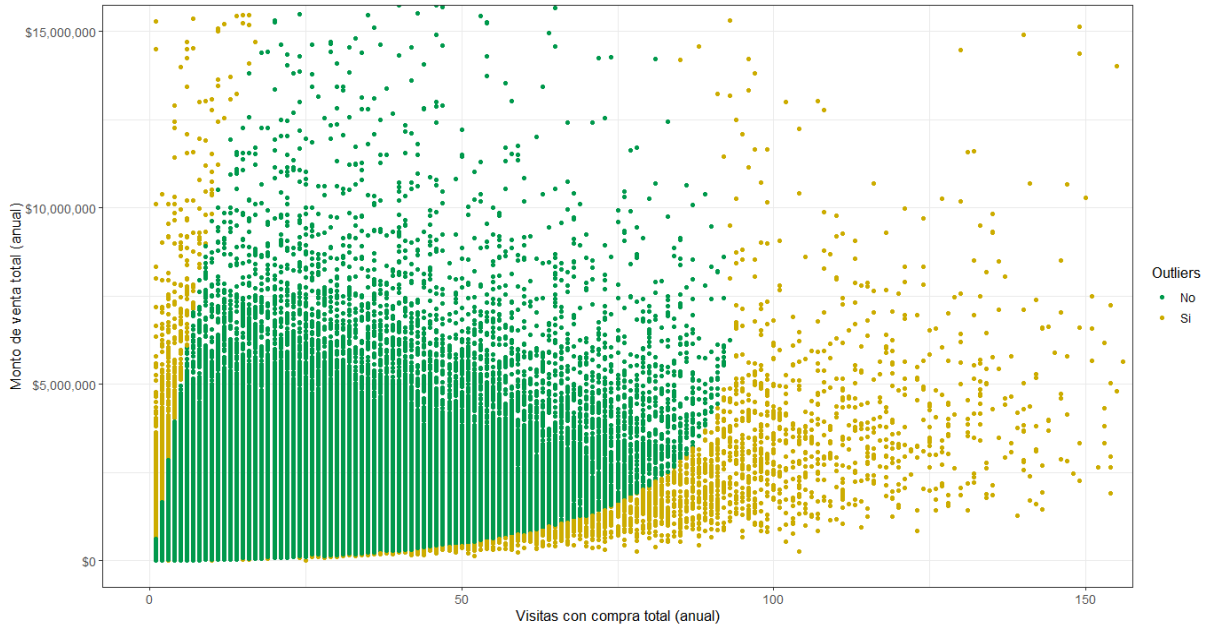
Anexos I: Transformación logarítmica de la cantidad de productos comprados anualmente por cliente.
Fuente: Elaboración propia.



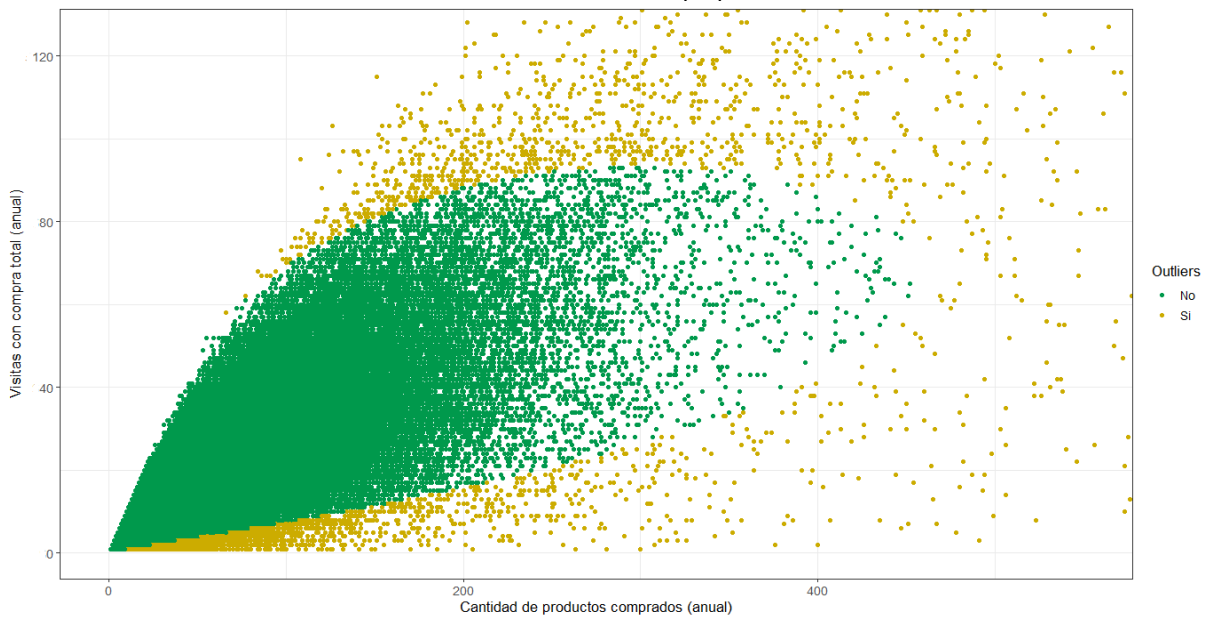
Anexos J: Transformación logarítmica de visitas con compra anual por cliente.
Fuente: Elaboración propia.



Anexos K: Criterio de exclusión para el par 2: Montos de venta anual y visitas con compra anual.
Fuente: Elaboración propia.



Anexos L: Criterio de exclusión para el par 3: Cantidad de productos comprados y visitas con compra anual.
Fuente: Elaboración propia.



Anexos M: Detalle de variables preliminares para el propensity score mediante Logit.
Fuente: Elaboración propia.

Set de variables explicativas		Detalle	Tipología
VAR_GASTO_L1	Variación gasto último mes	Variación del gasto mensual del cliente con respecto al mes anterior	Numérica continua
VAR_GASTO_L2	Variación gasto penúltimo mes		Numérica continua
EDAD_DEF	Edad	Indica la edad actual del cliente	Numérica discreta
SEXO_DEF	Genero	Indica 1 si es hombre	Binaria
NAV_CAT_L1	Navegación último mes	Indica 1 si navegó en la categoría	Binaria
N_LINEAS	Nº líneas compradas los últimos 3 meses	Indica la cantidad de líneas diferentes compradas	Numérica discreta
DIR_GSE	Grupo socioeconómico	Indica a que estrato socioeconómico pertenece el cliente (ABC1, C2, C3, D, E)	Categórica
ZONA_RESIDENCIA	Zona de residencia	Indica la zona de residencia del cliente (Oriente, Poniente, Norte o Sur)	Categórica
MONTO_RENTA	Monto renta	Indicador de renta percibida por el cliente	Numérica
CUPO	Cupo	Indicador del cupo tarjeta de crédito cliente	Numérico
ANTIGÜEDAD	Antigüedad	1 si el cliente es antiguo (compras en año anterior y año actual)	Binaria
ESTADO_CIVIL	Estado civil	1 si el cliente está casado(a)	Binaria
CMR_OPEN_NORMAL	CMR abierta normal	1 si el cliente posee cuenta abierta y t. crédito Normal	Binaria
CMR_OPEN_PREMIUM	CMR abierta Premium	1 si el cliente posee cuenta abierta y t. crédito Premium	Binaria
OMNICANAL	Canal de compra omnicanal del cliente	1 si el cliente compra en ambos canales de la empresa (tanto en web como en tienda)	Binaria
SOLO TIENDA	Canal de compra solo tienda del cliente	1 si el cliente compra solo en el canal físico (tienda) de la empresa	Binaria
MODAH	Moda Hombre	1 si compro la categoría en el último mes	Binaria
MODAM	Moda Mujer		Binaria
DEPORTES	Deportes		Binaria
ROPAINT	Ropa Interior		Binaria
ACCESORIOS	Accesorios		Binaria
BELLEZA	Belleza		Binaria
INFANTIL	Infantil		Binaria
CALZADOH	Calzado Hombre		Binaria
CALZADOM	Calzado Mujer		Binaria
ELECTRO	Electro hogar		Binaria
DECO	Decoración		Binaria

Anexos N: Matriz de coeficientes de correlación de Pearson para variables continuas.

Fuente: Elaboración propia.

Pearson Coef	VISITAS	N_LINEAS	GASTO_L1	VAR_GASTO_L1	VAR_GASTO_L2	NAV_CAT_L1	EDAD_DEF	MONTO_RENTA	CUPO	N_HIJOS
VISITAS	1,00	0,88	0,85	0,22	0,02	0,01	0,01	0,03	0,11	0,00
N_LINEAS	0,88	1,00	0,84	0,22	0,03	0,01	0,01	0,04	0,11	0,01
GASTO_L1	0,85	0,84	1,00	0,17	0,02	0,01	0,01	0,04	0,12	0,00
VAR_GASTO_L1	0,22	0,22	0,17	1,00	-0,01	0,00	0,00	0,01	0,02	0,00
VAR_GASTO_L2	0,02	0,03	0,02	-0,01	1,00	0,00	0,00	0,00	0,01	0,00
NAV_CAT_L1	0,01	0,01	0,01	0,00	0,00	1,00	0,00	0,01	0,01	0,00
EDAD_DEF	0,01	0,01	0,01	0,00	0,00	0,00	1,00	-0,01	0,20	0,54
MONTO_RENTA	0,03	0,04	0,04	0,01	0,00	0,01	-0,01	1,00	0,24	-0,01
CUPO	0,11	0,11	0,12	0,02	0,01	0,01	0,20	0,24	1,00	0,03
N_HIJOS	0,00	0,01	0,00	0,00	0,00	0,00	0,54	-0,01	0,03	1,00

Anexos O: Matriz de coeficientes V de Cramer para variables categóricas.

Fuente: Elaboración propia.

V Cramer Coef.	ANTIGÜEDAD	CANAL_CLIENTE	CMR_OPEN_ELITE	CMR_OPEN_NORMAL	CMR_OPEN_PREMIUM	CUOTERO	DIR_GSE	ESTADO_CIVIL	SEXO_DEF	USA_CMR	ZONA_RESIDENCIA
ANTIGÜEDAD	1,00	0,20	0,09	0,13	0,09	0,01	0,05	0,09	0,11	0,13	0,05
CANAL_CLIENTE	0,20	1,00	0,15	0,10	0,15	0,08	0,05	0,04	0,07	0,17	0,09
CMR_OPEN_ELITE	0,09	0,15	1,00	0,24	1,00	0,12	0,02	0,04	0,07	0,14	0,06
CMR_OPEN_NORMAL	0,13	0,10	0,24	1,00	0,24	0,11	0,05	0,06	0,02	0,14	0,08
CMR_OPEN_PREMIUM	0,09	0,15	0,24	0,24	1,00	0,12	0,02	0,04	0,07	0,14	0,06
CUOTERO	0,01	0,08	0,12	0,11	0,12	1,00	0,12	0,07	0,09	0,11	0,13
DIR_GSE	0,05	0,05	0,02	0,05	0,02	0,12	1,00	0,13	0,06	0,02	0,30
ESTADO_CIVIL	0,09	0,04	0,04	0,06	0,04	0,07	0,13	1,00	0,02	0,07	0,06
SEXO_DEF	0,11	0,07	0,07	0,02	0,07	0,09	0,06	0,02	1,00	0,06	0,04
USA_CMR	0,13	0,17	0,14	0,14	0,14	0,11	0,02	0,07	0,06	1,00	0,02
ZONA_RESIDENCIA	0,05	0,09	0,06	0,08	0,06	0,13	0,30	0,06	0,04	0,02	1,00

Anexos P: Regresiones logísticas para cada categoría, la estimación de sus parámetros y los efectos marginales promedio calculados para cada parámetro estimado.

Fuente: Elaboración propia.

Variable explicativa	Moda Hombre	p-value	AME	Moda Mujer	p-value	AME	Deportes	p-value	AME	Ropa Interior	p-value	AME
INTERCEPTO	-3,85	0,00		-2,38	0,00		-2,20	0,00		-4,90	0,00	
MODA HOMBRE				0,43	0,00	7,10%	0,45	0,00	3,23%	0,22	0,00	1,56%
MODA MUJER	0,44	0,00	6,05%				0,37	0,00	2,69%	0,61	0,00	4,33%
DEPORTES	0,50	0,00	6,82%	0,26	0,00	4,38%				0,32	0,00	2,26%
ROPA INTERIOR	0,26	0,00	3,56%	0,50	0,00	8,33%						
ACCESORIOS				0,51	0,00	8,53%				0,35	0,00	2,52%
BELLEZA	0,38	0,00	5,12%	0,38	0,00	6,30%	0,37	0,00	2,71%	0,50	0,00	3,53%
INFANTIL	0,48	0,00	6,60%	0,43	0,00	7,19%	0,47	0,00	3,40%	0,43	0,00	3,06%
CALZADO HOMBRE	0,68	0,00	9,28%				0,67	0,00	4,86%			
CALZADO MUJER	0,23	0,00	3,17%	0,55	0,00	9,25%				0,41	0,00	2,92%
DECO	0,38	0,00	5,16%	0,36	0,00	5,59%	0,41	0,00	2,94%	0,33	0,00	2,37%
ELECTRO	0,44	0,00	6,06%	0,34	0,00	6,00%	0,50	0,00	3,65%	0,41	0,00	2,92%
FEBRERO	-0,24	0,00	3,29%	0,37	0,00	6,24%				-0,17	0,00	-1,21%
MARZO	-0,13	0,00	1,76%	0,28	0,00	4,62%						
ABRIL							-0,17	0,00	-1,22%			
MAYO										0,13	0,00	0,94%
JUNIO	0,43	0,00	5,84%	-0,21	0,00	-3,49%				0,11	0,00	-1,30%
JULIO	0,18	0,00	2,40%							-0,18	0,00	
AGOSTO				-0,11	0,00	-1,88%						
SEPTIEMBRE												
OCTUBRE												
NOVIEMBRE												
DICIEMBRE	0,51	0,00	6,92%	0,13	0,00	2,12%	0,45	0,00	3,25%			0,80%
NAV_CAT_L1	0,04	0,00	0,58%	0,02	0,00	0,40%	0,04	0,00	0,30%	0,05	0,00	0,33%
CUPO	0,00	0,00	0,07%	0,01	0,00	0,21%						
CLTE ANTIGUO	0,10	0,00	1,35%	0,17	0,00	2,85%				0,27	0,00	1,95%
OMNICANAL	1,43	0,00	19,49%	1,32	0,00	22,01%				1,95	0,00	13,87%
SOLO TIENDA	1,74	0,00	23,77%	1,64	0,00	27,27%	-0,08	0,00	-0,61%	2,12	0,00	15,08%
CMR_OPEN_NORMAL												
EDAD							-0,01	0,00	-0,09%	0,01	0,00	0,04%
MASCULINO	1,11	0,00	15,11%	-1,42	0,00	-23,75%	0,53	0,00	3,83%	-1,11	0,00	-7,92%
CASADO(A)				0,08	0,00	1,34%						
HIJOS	-0,13	0,00	1,83%	0,29	0,00	4,83%				-0,12	0,00	-0,85%
MONTO RENTA				0,02	0,00	0,34%				0,03	0,00	0,20%
CUOTAS CMR	0,06	0,00	0,83%			1,11%	0,12	0,00	0,84%			
ZONA NORTE				0,13	0,00	2,20%	-0,10	0,00	-0,70%	0,09	0,00	0,63%
ZONA RM ORIENTE							-0,49	0,00	-3,54%			
ZONA RM PONIENTE				0,09	0,00	1,50%	-0,23	0,00	-1,68%			

Variable explicativa	Accesorios	p-value	AME	Belleza	p-value	AME	Infantil	p-value	AME	Calzado hombre	p-value	AME
INTERCEPTO	-3,64	0,00		-3,85	0,00		-1,34	0,00		-3,27	0,00	
MODA HOMBRE	0,34	0,00	1,84%	0,41	0,00	3,06%	0,32	0,00	4,53%	0,52	0,00	2,82%
MODA MUJER	0,61	0,00	3,33%	0,41	0,00	3,05%	0,24	0,00	3,37%			
DEPORTES	0,29	0,00	1,60%	0,21	0,00	1,55%	0,26	0,00	3,70%	0,43	0,00	2,32%
ROPA INTERIOR	0,26	0,00	1,43%	0,39	0,00	2,94%						
ACCESORIOS				0,53	0,00	4,01%	0,26	0,00	3,61%	0,37	0,00	1,97%
BELLEZA	0,60	0,00	3,28%				0,32	0,00	4,55%			
INFANTIL	0,35	0,00	1,90%	0,35	0,00	2,64%				0,45	0,00	2,44%
CALZADO HOMBRE							0,34	0,00	4,86%			
CALZADO MUJER	0,48	0,00	2,59%	0,30	0,00	2,24%	0,30	0,00	4,17%	0,44	0,00	2,38%
DECO	0,44	0,00	2,36%	0,40	0,00	2,99%	0,36	0,00	5,09%	0,33	0,00	1,79%
ELECTRO	0,39	0,00	2,09%	0,50	0,00	3,74%	0,40	0,00	5,69%	0,44	0,00	2,37%
FEBRERO							0,59	0,00	8,35%			
MARZO				0,37	0,00	2,81%	0,28	0,00	3,98%	0,20	0,00	1,08%
ABRIL				0,31	0,00	2,36%						
MAYO	0,26	0,00	1,39%	0,80	0,00	5,99%						
JUNIO				0,32	0,00	2,43%	-0,12	0,00	-1,70%			
JULIO				0,20	0,00	1,51%	0,22	0,00	3,16%			
AGOSTO				0,35	0,00	2,66%	0,23	0,00	3,25%			
SEPTIEMBRE				0,35	0,00	2,59%						
OCTUBRE	0,20	0,00	1,09%	0,45	0,00	3,38%						
NOVIEMBRE	0,27	0,00	1,45%	0,39	0,00	2,96%	0,11	0,00	1,59%			
DICIEMBRE	0,62	0,00	3,38%	1,19	0,00	8,97%	0,59	0,00	8,32%			
NAV_CAT_L1	0,04	0,00	0,20%	0,06	0,00	0,46%	0,03	0,00	0,48%	0,05	0,00	0,29%
CUPO				0,02	0,00	0,12%						
CLTE ANTIGUO	0,22	0,00	1,17%	0,25	0,00	1,91%	-0,13	0,00	-1,88%	-0,24	0,00	-1,30%
OMNICANAL	0,59	0,00	3,23%	0,29	0,00	2,17%						
SOLO TIENDA	0,58	0,00	3,17%	0,28	0,00	2,13%	0,12	0,00	1,62%			
CMR_OPEN_NORMAL										0,11	0,00	0,57%
EDAD	0,00	0,00	-0,02%	0,01	0,00	0,04%	-0,03	0,00	-0,44%			
MASCULINO	-0,48	0,00	-2,60%	-0,16	0,00	-1,17%	-0,23	0,00	-3,31%	1,16	0,00	6,23%
CASADO(A)							-0,11	0,00	-1,61%			
HIJOS	-0,18	0,00	-0,97%	-0,20	0,00	-1,53%	1,31	0,00	18,51%			
MONTO RENTA	0,04	0,00	0,19%	0,03	0,00	0,19%				-0,04	0,00	-0,19%
CUOTAS CMR							0,19	0,00	2,69%	0,14	0,00	0,78%
ZONA NORTE							0,14	0,00	2,03%			
ZONA RM ORIENTE							-0,12	0,00	-1,75%	-0,20	0,00	-1,07%
ZONA RM PONIENTE							0,08	0,00	1,15%			

Variable explicativa	Calzado mujer	p-value	AME	Electro	p-value	AME	Deco	p-value	AME
INTERCEPTO	-2,76	0,00		-1,39	0,00		-2,32	0,00	
MODA HOMBRE	0,18	0,00	1,39%	0,23	0,00	3,09%	0,40	0,00	3,70%
MODA MUJER	0,59	0,00	4,59%	0,10	0,00	1,34%	0,30	0,00	2,73%
DEPORTES	0,35	0,00	2,75%	0,29	0,00	4,00%	0,22	0,00	2,04%
ROPA INTERIOR							0,32	0,00	2,93%
ACCESORIOS	0,41	0,00	3,16%				0,29	0,00	2,68%
BELLEZA	0,28	0,00	2,20%	0,25	0,00	3,42%	0,34	0,00	3,09%
INFANTIL	0,34	0,00	2,66%	0,33	0,00	4,50%	0,45	0,00	4,14%
CALZADO HOMBRE	0,55	0,00	4,30%						
CALZADO MUJER				0,18	0,00	2,51%	0,29	0,00	2,69%
DECO	0,28	0,00	2,16%	0,41	0,00	5,63%			
ELECTRO	0,40	0,00	3,07%				0,62	0,00	5,67%
FEBRERO	0,17	0,00	1,34%						
MARZO	0,14	0,00	1,09%	0,32	0,00	4,42%	0,23	0,00	2,16%
ABRIL				0,16	0,00	2,22%	0,24	0,00	2,23%
MAYO	0,23	0,00	1,77%	0,41	0,00	5,54%	0,36	0,00	3,36%
JUNIO				0,23	0,00	3,08%			
JULIO				0,23	0,00	3,08%			
AGOSTO				0,21	0,00	2,84%	0,14	0,00	1,31%
SEPTIEMBRE	-0,17	0,00	-1,30%	0,18	0,00	2,44%			
OCTUBRE				0,28	0,00	3,82%	0,30	0,00	2,80%
NOVIEMBRE	0,13	0,00	1,02%				0,30	0,00	2,78%
DICIEMBRE				0,30	0,00	4,11%	0,43	0,00	3,93%
NAV_CAT_L1	0,03	0,00	0,24%	0,04	0,00	0,50%	0,03	0,00	0,25%
CUPO				0,05	0,00	0,63%	0,02	0,00	0,21%
CLTE ANTIGUO				-0,27	0,00	-3,65%			
OMNICANAL	0,12	0,00	0,90%	-0,62	0,00	-8,43%	-0,55	0,00	-5,11%
SOLO TIENDA				-1,20	0,00	-16,44%	-1,12	0,00	-10,29%
CMR_OPEN_NORMAL				-0,14	0,00	-1,97%	-0,12	0,00	-1,10%
EDAD	0,01	0,00	0,04%	0,01	0,00	0,10%	0,01	0,00	0,13%
MASCULINO	-0,77	0,00	-5,98%	0,68	0,00	9,25%			
CASADO(A)									
HIJOS									
MONTO RENTA							0,03	0,00	0,28%
CUOTAS CMR	0,13	0,00	0,98%	-0,06	0,00	-0,79%	-0,12	0,00	-1,07%
ZONA NORTE	0,12	0,00	0,91%	-0,14	0,00	-1,93%	-0,12	0,00	-1,06%
ZONA RM ORIENTE							0,32	0,00	2,96%
ZONA RM PONIENTE							0,09	0,00	0,84%

Anexos Q: Principales métricas se evaluación para el modelo de regresión logística para cada categoría.

Fuente: Elaboración propia

Categoría	AUC	Sensitivity	Especificity	Precision	Recall	Accuracy	BIC
Moda Hombre	69%	58%	71%	29%	58%	68%	100.832
Moda Mujer	72%	75%	58%	38%	75%	62%	133.975
Deportes	66%	68%	56%	12%	68%	57%	68.970
Ropa Interior	69%	77%	50%	12%	77%	52%	66.544
Accesorios	65%	58%	65%	9%	58%	65%	56.451
Belleza	66%	55%	69%	15%	55%	68%	70.108
Infantil	69%	60%	70%	31%	60%	68%	99.678
Calzado Hombre	69%	58%	58%	10%	72%	58%	60.020
Calzado Mujer	68%	56%	70%	15%	56%	69%	72.253
Electro	69%	66%	63%	28%	66%	63%	100.769
Deco	69%	61%	66%	18%	61%	66%	79.418

Anexos R: Resumen de los principales estadísticos de soporte común para la estimación del propensity score de cada categoría.

Fuente: Elaboración propia.

Categoría	Activación	Cantidad de registros	Propensity score		
			Minimo	Promedio	Maximo
Moda Hombre	No	94.932	1,4%	16,6%	99,7%
	Si	20.594	1,7%	23,4%	99,7%
Moda Mujer	No	84.809	1,0%	22,2%	98,5%
	Si	27.971	1,7%	32,8%	100,0%
Deportes	No	116.902	1,6%	7,9%	99,5%
	Si	10.408	2,1%	11,0%	99,0%
Ropa Interior	No	116.583	0,2%	7,8%	88,7%
	Si	10.245	0,3%	11,8%	83,3%
Accesorios	No	123.516	1,0%	5,8%	98,7%
	Si	7.830	1,4%	8,8%	97,7%
Belleza	No	115.476	1,8%	8,2%	99,0%
	Si	10.789	2,3%	12,0%	99,0%
Infantil	No	90.590	1,1%	17,3%	99,8%
	Si	20.694	1,6%	24,4%	100,0%
Calzado Hombre	No	125.318	1,8%	5,7%	100,0%
	Si	7.910	1,8%	9,1%	99,8%
Calzado Mujer	No	115.668	2,7%	8,5%	99,9%
	Si	11.222	2,8%	12,0%	99,5%
Electro	No	94.532	4,6%	16,8%	88,9%
	Si	20.710	4,8%	23,5%	100,0%
Deco	No	109.553	2,6%	10,4%	99,0%
	Si	13.468	2,9%	15,6%	98,7%

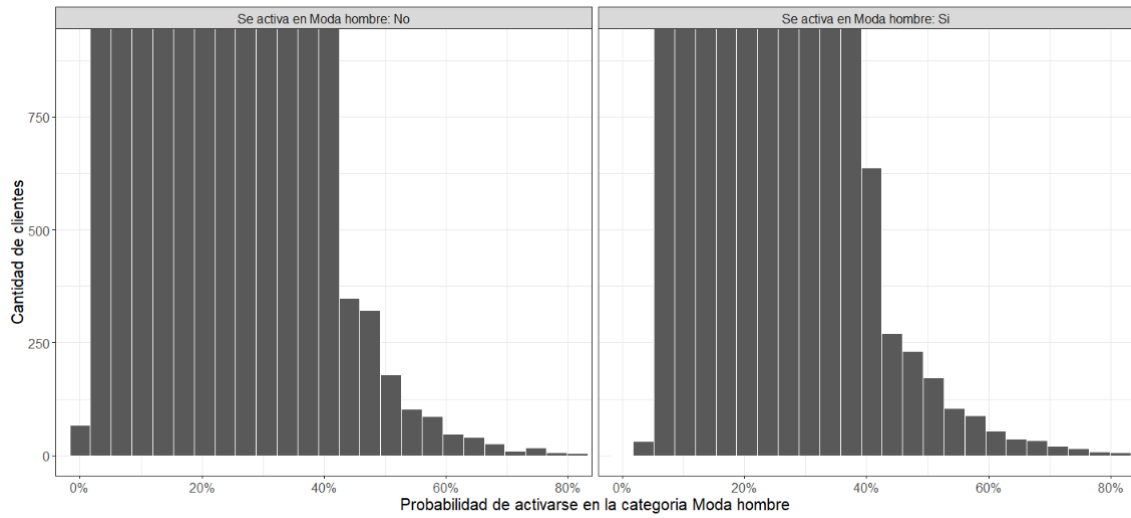
Anexos S: Resumen de los resultados del efecto causal en gasto post activación después del emparejamiento para cada categoría.

Fuente: Elaboración propia.

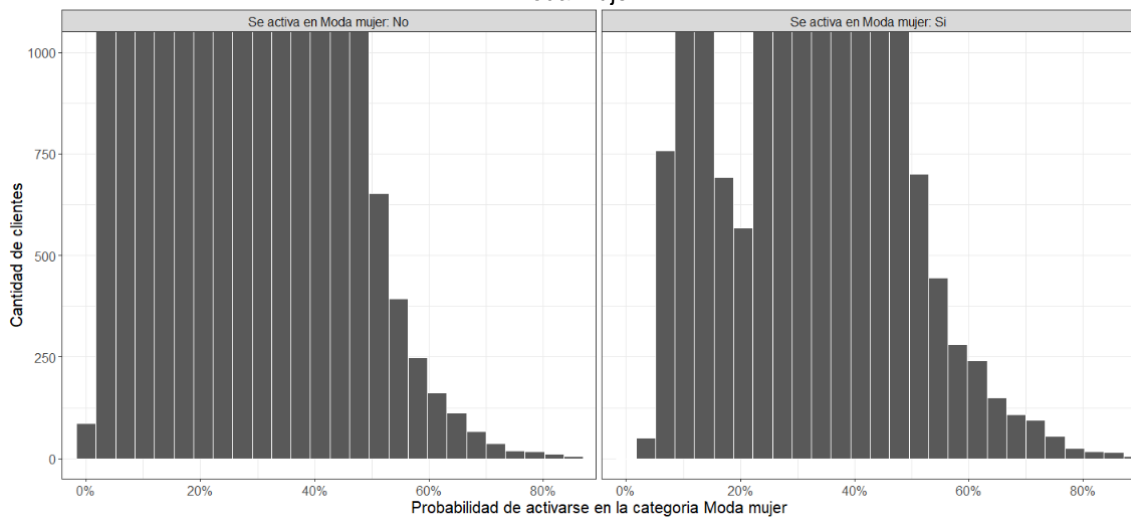
Categoría	Variable	Promedio en grupo de activacion	Promedio en grupo de no activacion	Diferencia	P-value test	Diff in Diff	
						Monetaria	Porcentual
Moda Hombre	Propensity score	23%	23%	0%	0,98		
	Gasto en pre-activacion	\$ 79.729	\$ 79.268	\$ 461	0,93	\$ 6.395	6,6%
	Gasto en post-activacion	\$ 95.684	\$ 88.828	\$ 6.856	0,00		
Moda Mujer	Propensity score	33%	33%	0%	0,83		
	Gasto en pre-activacion	\$ 63.061	\$ 62.751	\$ 310	0,95	\$ 5.059	6,0%
	Gasto en post-activacion	\$ 82.872	\$ 77.503	\$ 5.369	0,00		
Deportes	Propensity score	11%	11%	0%	0,90		
	Gasto en pre-activacion	\$ 106.988	\$ 105.228	\$ 1.760	0,44	\$ 6.493	5,5%
	Gasto en post-activacion	\$ 115.567	\$ 107.314	\$ 8.253	0,00		
Ropa Interior	Propensity score	12%	12%	0%	0,80		
	Gasto en pre-activacion	\$ 107.767	\$ 107.139	\$ 628	0,83	\$ 8.163	6,8%
	Gasto en post-activacion	\$ 118.830	\$ 110.039	\$ 8.791	0,00		
Accesorios	Propensity score	9%	7%	1%	0,00		
	Gasto en pre-activacion	\$ 123.876	\$ 122.420	\$ 1.456	0,54	\$ 4.963	3,8%
	Gasto en post-activacion	\$ 128.813	\$ 122.394	\$ 6.419	0,00		
Belleza	Propensity score	12%	12%	0%	0,27		
	Gasto en pre-activacion	\$ 108.584	\$ 107.243	\$ 1.341	0,60	\$ 5.821	4,6%
	Gasto en post-activacion	\$ 121.796	\$ 114.634	\$ 7.162	0,00		
Infantil	Propensity score	24%	24%	0%	0,77		
	Gasto en pre-activacion	\$ 73.864	\$ 72.738	\$ 1.126	0,64	\$ 11.053	10,2%
	Gasto en post-activacion	\$ 104.149	\$ 91.970	\$ 12.179	0,00		
Calzado Hombre	Propensity score	9%	9%	1%	0,03		
	Gasto en pre-activacion	\$ 101.776	\$ 101.098	\$ 678	0,90	\$ 3.588	3,2%
	Gasto en post-activacion	\$ 110.390	\$ 106.124	\$ 4.266	0,00		
Calzado Mujer	Propensity score	12%	12%	0%	0,55		
	Gasto en pre-activacion	\$ 102.271	\$ 101.862	\$ 409	0,92	\$ 10.838	9,0%
	Gasto en post-activacion	\$ 119.502	\$ 108.255	\$ 11.247	0,00		
Electro	Propensity score	24%	24%	0%	0,98		
	Gasto en pre-activacion	\$ 59.379	\$ 58.822	\$ 557	0,88	\$ 16.633	13,9%
	Gasto en post-activacion	\$ 115.619	\$ 98.429	\$ 17.190	0,00		
Deco	Propensity score	16%	16%	0%	0,91		
	Gasto en pre-activacion	\$ 105.238	\$ 103.865	\$ 1.373	0,62	\$ 14.840	10,6%
	Gasto en post-activacion	\$ 136.739	\$ 120.526	\$ 16.213	0,00		

Anexos T: Verificación del soporte común para las categorías de producto estudiadas.
Fuente: Elaboración propia.

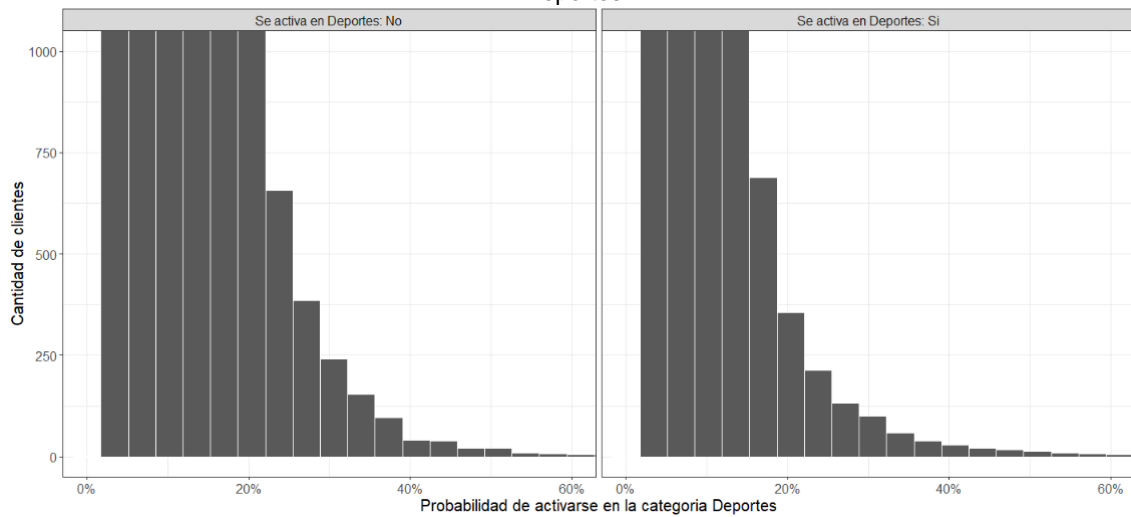
Moda Hombre



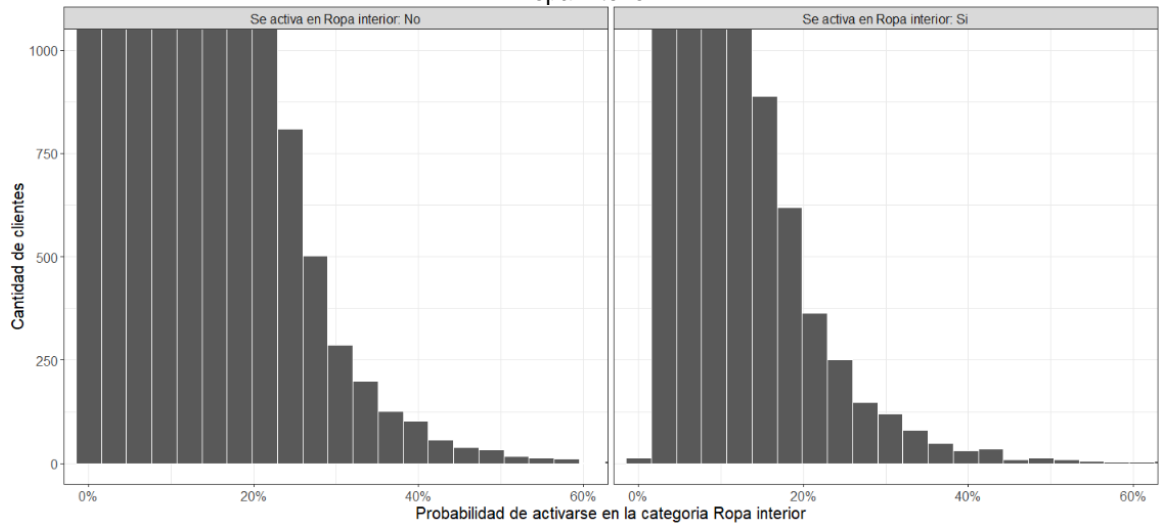
Moda Mujer



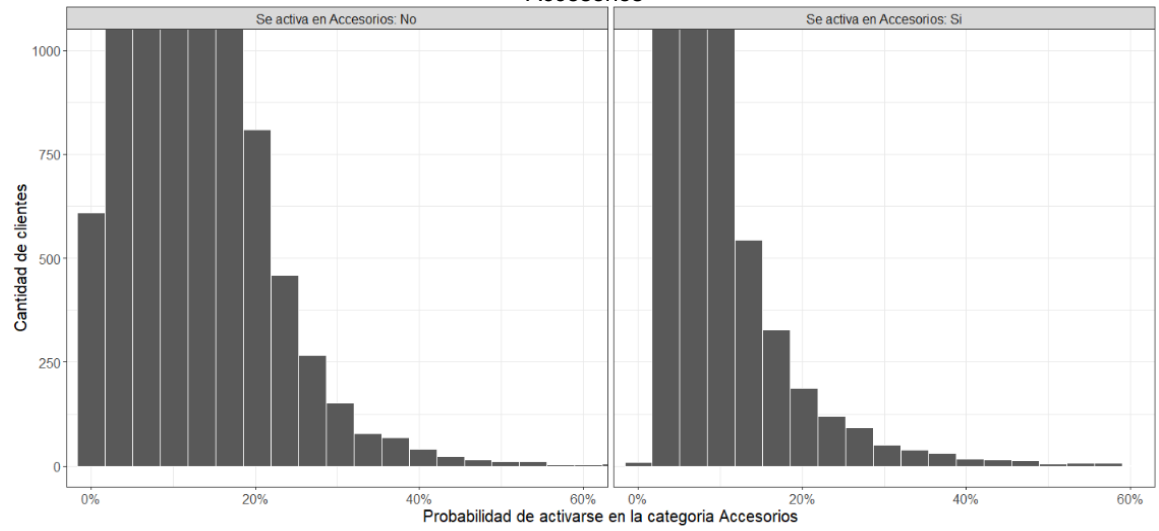
Deportes



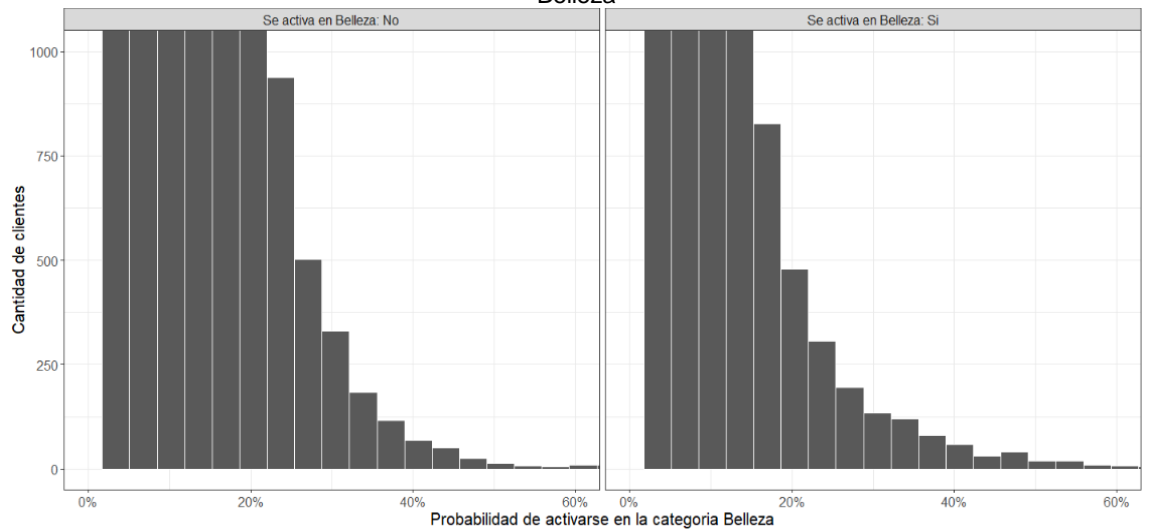
Ropa Interior



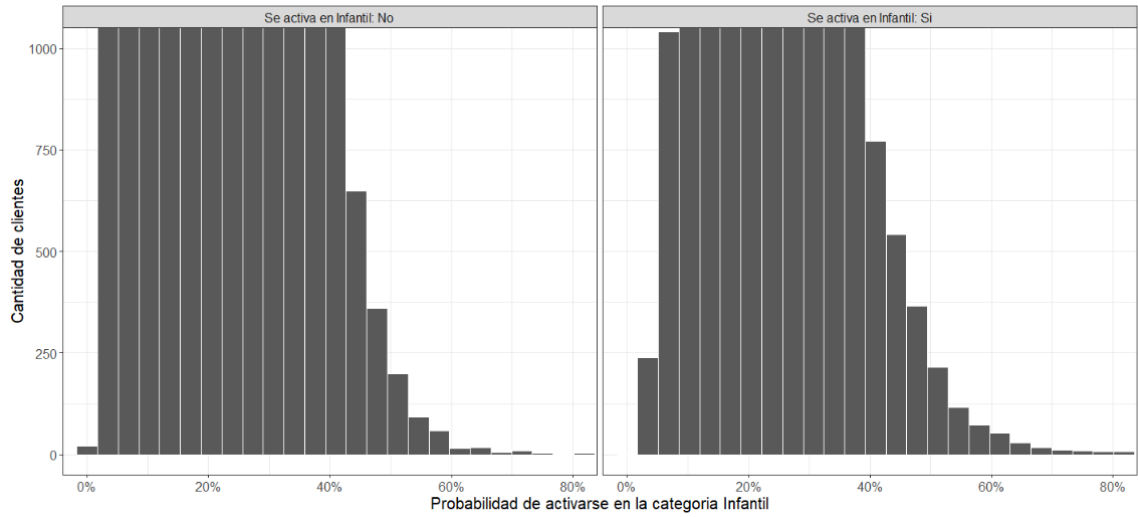
Accesorios



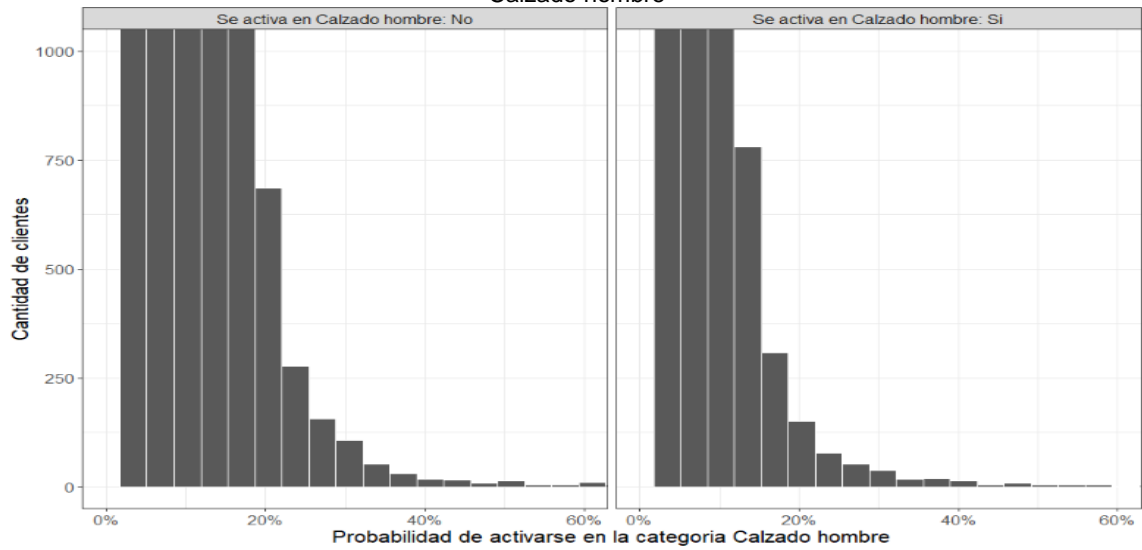
Belleza



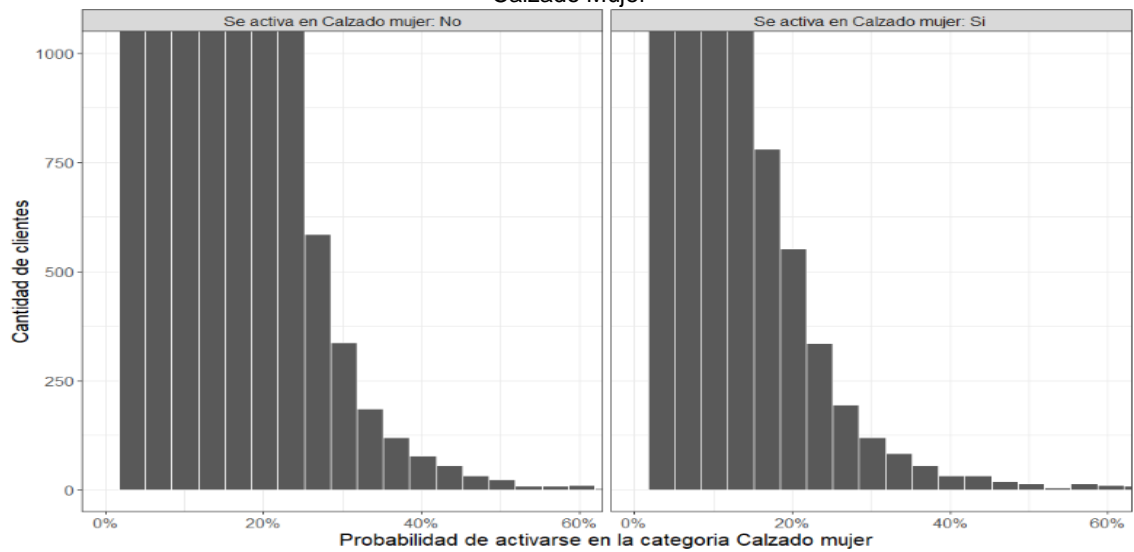
Infantil



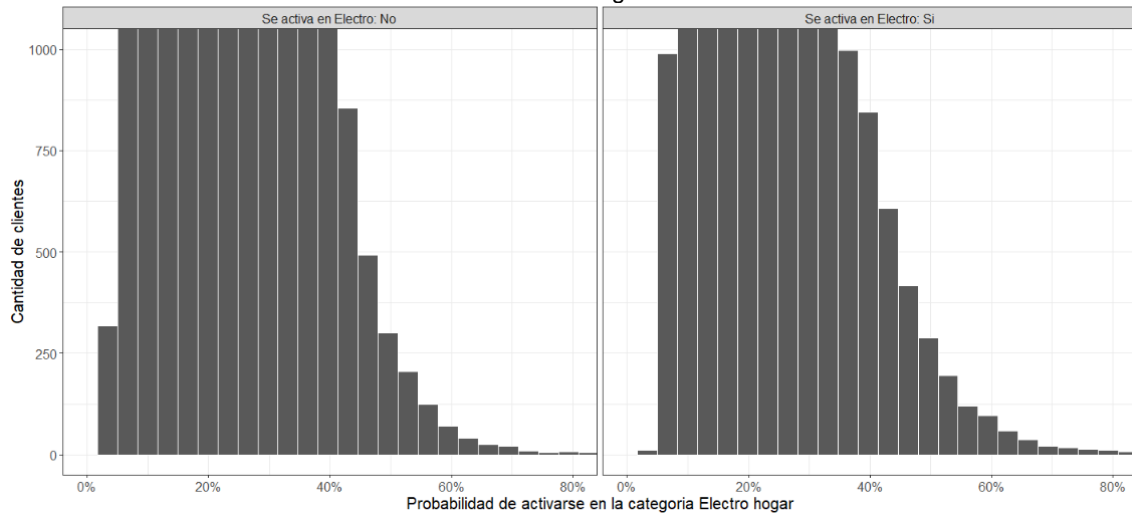
Calzado hombre



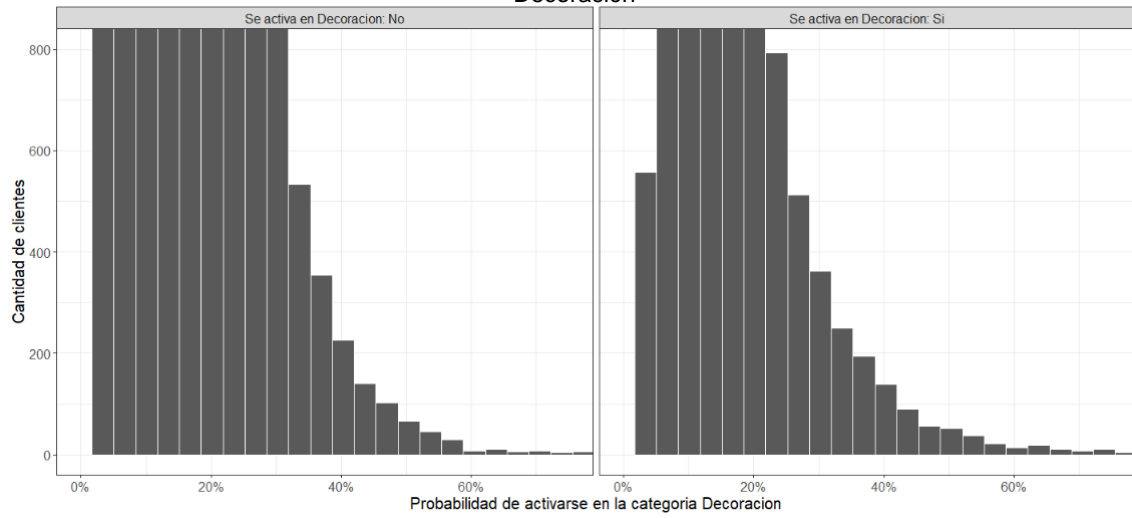
Calzado Mujer



Electro hogar

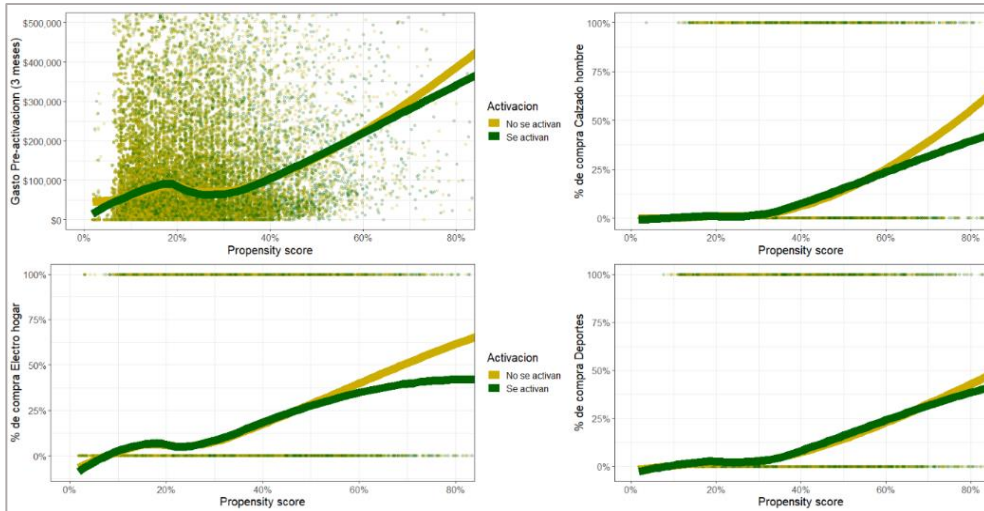


Decoración

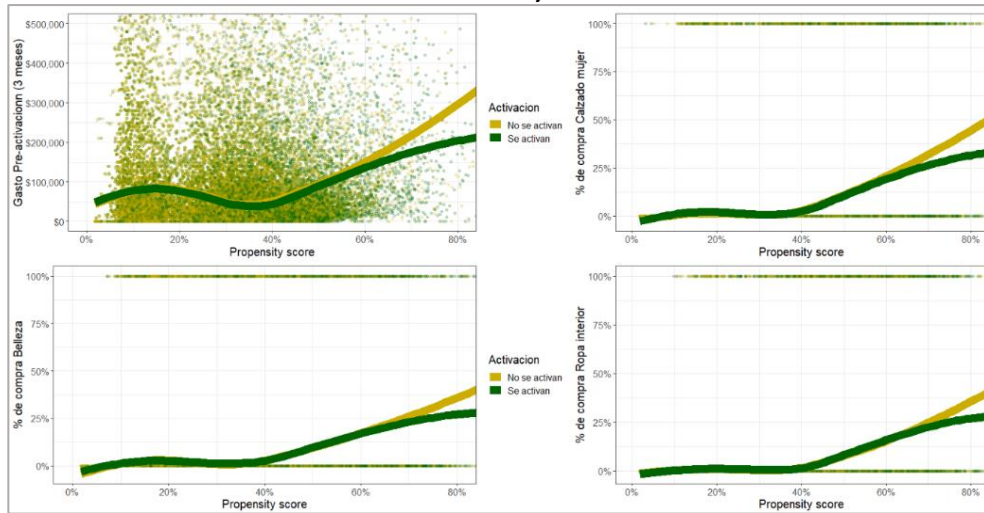


Anexos U: Balance de variables claves de emparejamiento y variables de comportamiento de compra relevantes para cada categoría.
 Fuente: Elaboración propia.

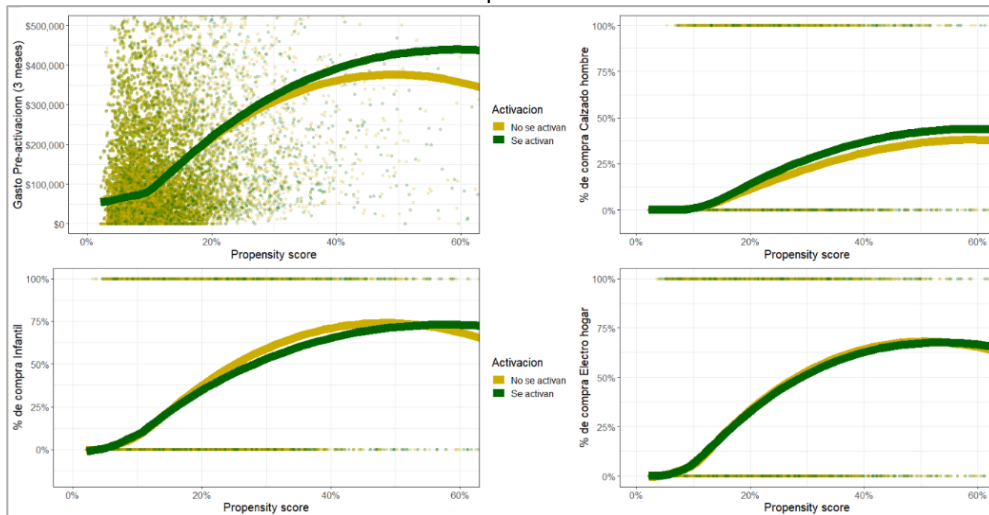
Moda Hombre



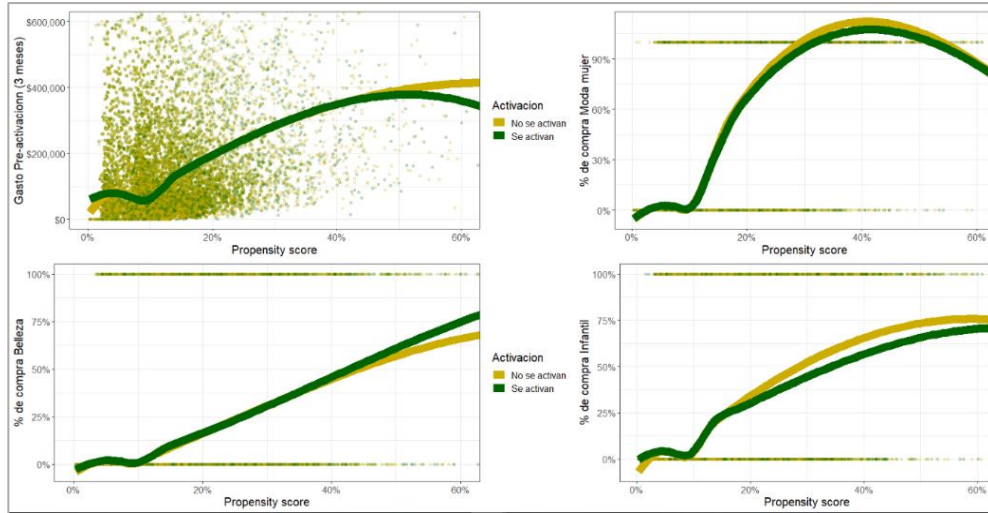
Moda Mujer



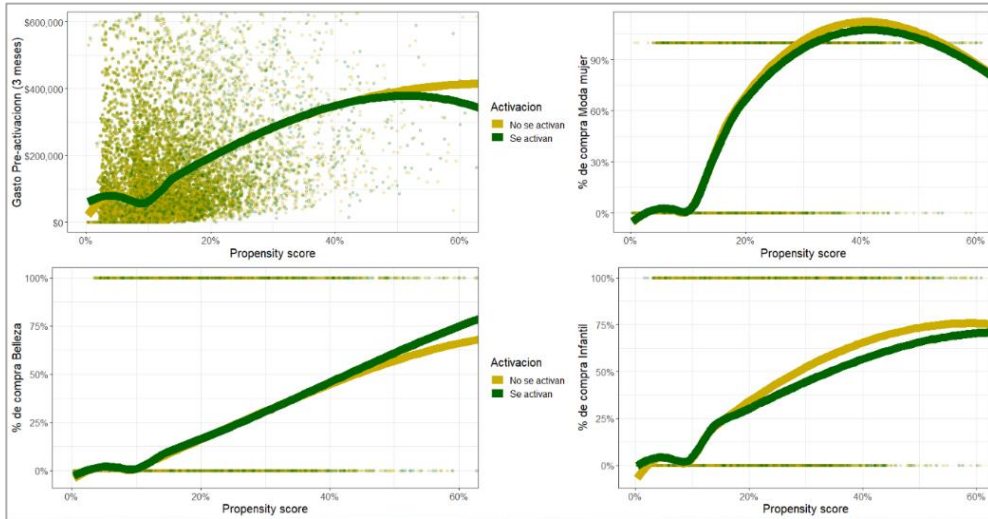
Deportes



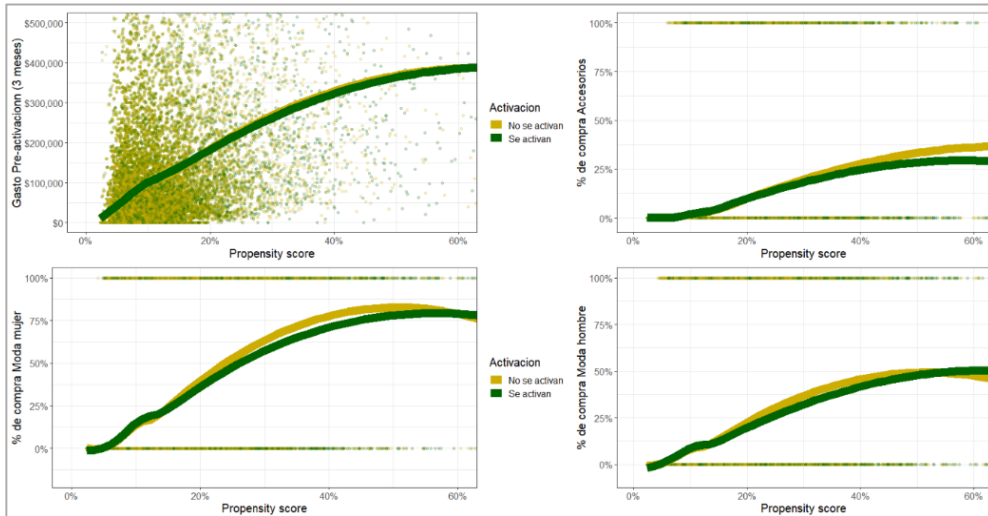
Ropa Interior



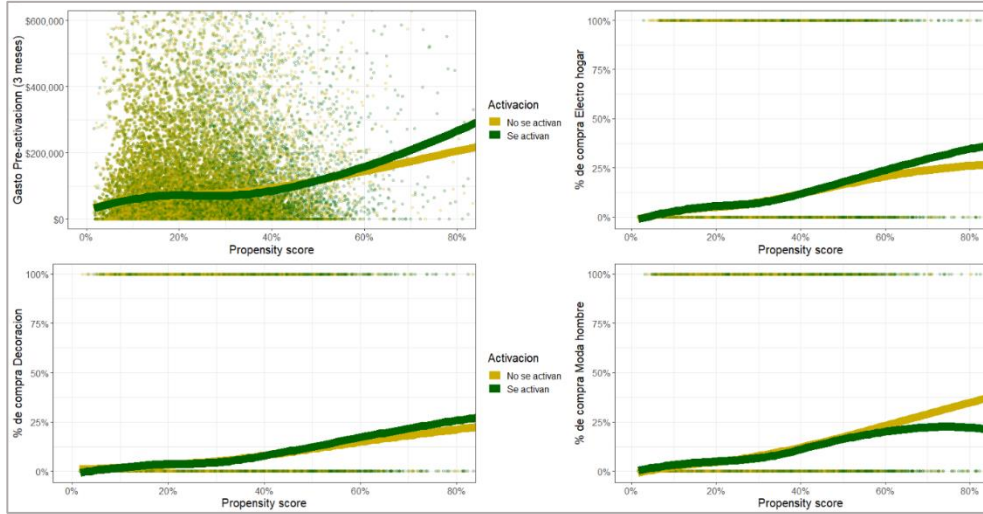
Accesorios



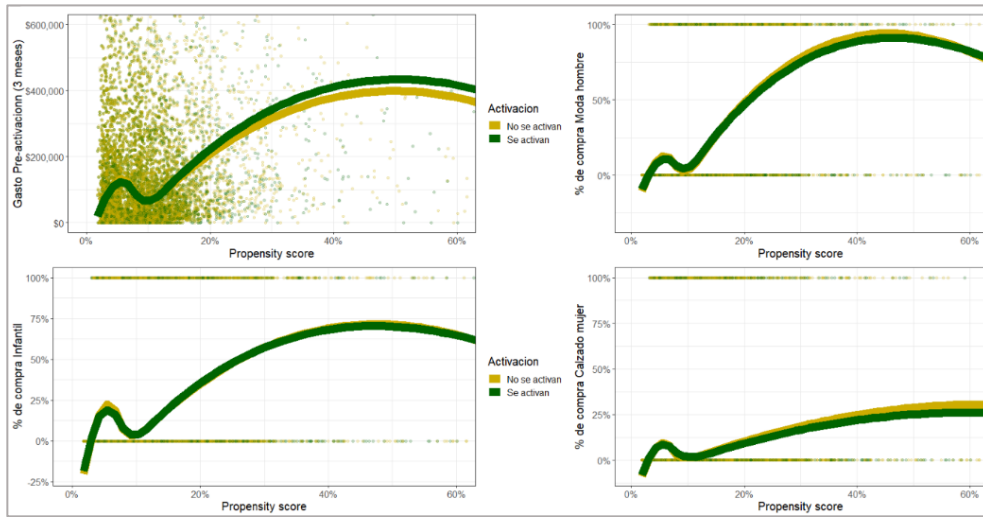
Belleza



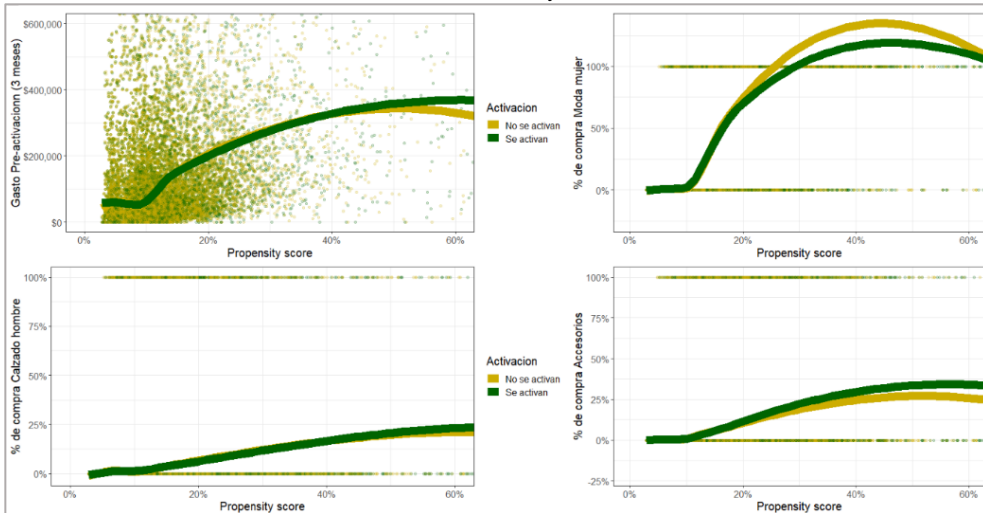
Infantil



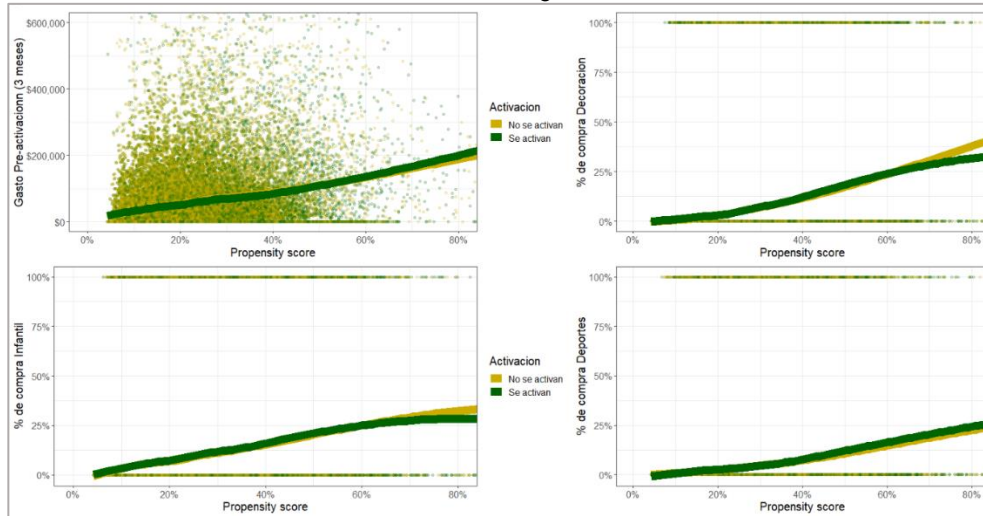
Calzado Hombre



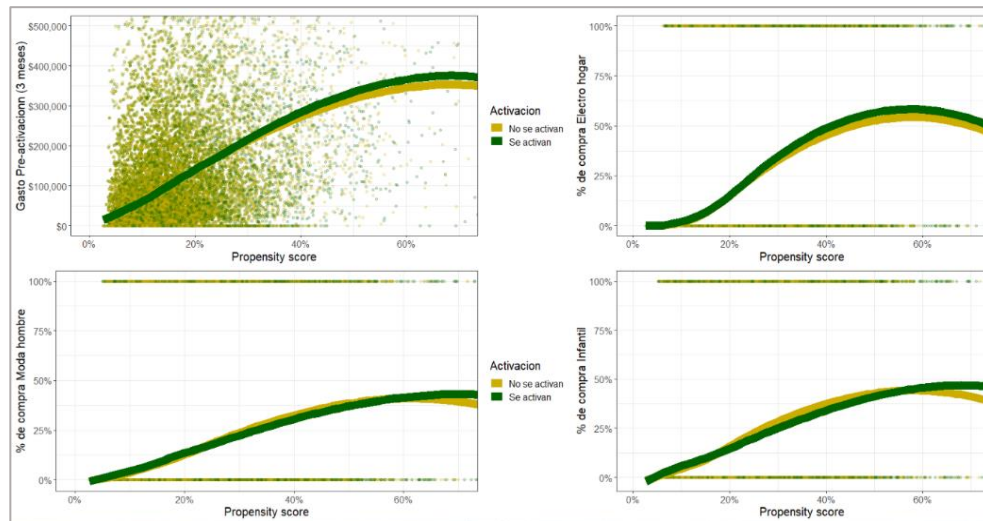
Calzado Mujer



Electro hogar



Decoración



Anexos V: Distribución de la elección de mejor próxima oferta para los diferentes criterios estudiados para el mes de julio 2019.

Fuente: Elaboración propia.

Categorías	Propension	Gasto esperado	Lift	Lift gasto esperado
Moda Hombre	306.145	179.160	148.158	99.966
Moda Mujer	511.434	242.051	279.610	3.554
Deportes	9.230	4.318	41.057	30.076
Ropa Interior	17.569	10.219	143.431	74.963
Accesorios	2.176	556	67.870	4.967
Belleza	3.625	1.493	28.992	6.868
Infantil	234.420	584.080	170.216	329.325
Calzado Hombre	4.699	440	178.494	3.339
Calzado Mujer	14.139	47.965	74.469	271.328
Electro Hogar	242.070	90.717	124.232	16.089
Decoracion	34.372	218.880	123.350	539.404

Anexos W: Ordenamiento de efecto de las compras en otras categorías durante el mes anterior sobre la categoría analizada.

Fuente: Elaboración propia.

Moda Hombre		Moda Mujer		Deportes		Ropa interior		Accesorios		Belleza	
Categoría	AME	Categoría	AME	Categoría	AME	Categoría	AME	Categoría	AME	Categoría	AME
Calzado hombre	9,28%	Calzado mujer	9,25%	Calzado hombre	4,86%	Moda mujer	4,33%	Moda mujer	3,33%	Accesorios	4,01%
Deportes	6,82%	Accesorios	8,53%	Electro	3,65%	Belleza	3,53%	Belleza	3,28%	Electro	3,74%
Infantil	6,60%	Ropa interior	8,33%	Infantil	3,40%	Infantil	3,06%	Calzado mujer	2,59%	Moda hombre	3,06%
Electro	6,06%	Infantil	7,19%	Moda hombre	3,23%	Calzado mujer	2,92%	Deco	2,36%	Moda mujer	3,05%
Moda mujer	6,05%	Moda hombre	7,10%	Deco	2,94%	Electro	2,92%	Electro	2,09%	Deco	2,99%
Deco	5,16%	Belleza	6,30%	Belleza	2,71%	Accesorios	2,52%	Infantil	1,90%	Ropa interior	2,94%
Belleza	5,12%	Electro	6,00%	Moda mujer	2,69%	Deco	2,37%	Moda hombre	1,84%	Infantil	2,64%
Ropa interior	3,56%	Deco	5,59%			Deportes	2,26%	Deportes	1,60%	Calzado mujer	2,24%
Calzado mujer	3,17%	Deportes	4,38%			Moda hombre	1,56%	Ropa interior	1,43%	Deportes	1,55%

Otras variables	AME	Otras variables	AME	Otras variables	AME	Otras variables	AME	Otras variables	AME	Otras variables	AME
Solo tienda	23,8%	Solo tienda	27,27%	Masculino	3,83%	Solo tienda	15,08%	Diciembre	3,38%	Diciembre	8,97%
Omnicanal	19,5%	Omnicanal	22,01%	Diciembre	3,25%	Omnicanal	13,87%	Omnicanal	3,23%	Mayo	5,99%
Masculino	15,1%	Masculino	-23,75%			Masculino	-7,92%	Solo tienda	3,17%	Octubre	3,38%

Infantil		Calzado hombre		Calzado mujer		Electro hogar		Decoracion	
Categoría	AME	Categoría	AME	Categoría	AME	Categoría	AME	Categoría	AME
Electro	5,69%	Moda hombre	2,82%	Moda mujer	4,59%	Deco	5,63%	Electro	5,67%
Deco	5,09%	Infantil	2,44%	Calzado hombre	4,30%	Infantil	4,50%	Infantil	4,14%
Calzado hombre	4,86%	Calzado mujer	2,38%	Accesorios	3,16%	Deportes	4,00%	Moda hombre	3,70%
Belleza	4,55%	Electro	2,37%	Electro	3,07%	Belleza	3,42%	Belleza	3,09%
Moda hombre	4,53%	Deportes	2,32%	Deportes	2,75%	Moda hombre	3,09%	Ropa interior	2,93%
Calzado mujer	4,17%	Accesorios	1,97%	Infantil	2,66%	Calzado mujer	2,51%	Moda mujer	2,73%
Deportes	3,70%	Deco	1,79%	Belleza	2,20%	Moda mujer	1,34%	Calzado mujer	2,69%
Accesorios	3,61%			Deco	2,16%			Accesorios	2,68%
Moda mujer	3,37%			Moda hombre	1,39%			Deportes	2,04%

Otras variables	AME	Otras variables	AME	Otras variables	AME	Otras variables	AME	Otras variables	AME
Hijos	18,51%	Masculino	6,23%	Masculino	-5,98%	Solo tienda	-16,44%	Solo tienda	-10,29%
Febrero	8,35%					Omnicanal	-8,43%	Omnicanal	-5,11%
Diciembre	8,32%					Masculino	9,25%	Diciembre	3,93%

Anexos X: Ejemplo de formato de grafica de envío de correos de marketing mono-categoricos
Fuente: Graficas de envío Falabella Retail.

f **LIQUIDACIÓN**

Relojes
40% DCTO
Productos Seleccionados.

Apúrate, encuentra tu estilo y disfruta precios imperdibles

VER TODO →

ROSSIO
ACCESSORIES COLLECTION
VER TODO >

LACOSTE
VER TODO >

f | | | |

Este e-mail es un anuncio de Falabella.com. Te enviamos este e-mail ya que nos escribiste tu dirección de correo en nuestros registros para recibir novedades, ofertas, promociones y más. Por favor, no respondas este correo, si necesitas más información llámanos al 600 390 6500.

¡PROTEGERNOS DE FRAUDES ES RESPONSABILIDAD DE TODOS!

Con estos 3 simples consejos, podrás evitar inconvenientes:

- 1.- Falabella no envía e-mails desde una extensión distinta a @falabella.com
- 2.- Nunca te pediremos por e-mail contraseñas o claves de acceso
- 3.- Ingresa a Falabella.com directamente desde tu navegador de Internet

f Despacho 24 HRS. | **Región Metropolitana**

HOT SALE

¡Aprovecha y renuévate!
No dejes pasar las mejores ofertas en la mejor Tecnología.

VER TODO

Despacho 24 HRS. | **Región Metropolitana**

SAMSUNG LED NU7095 4K Ultra HD Smart TV **\$239.990**
NORMAL: \$319.990

LG LED 55UM7100 4K Ultra HD Smart TV **\$279.990**
NORMAL: \$349.990

Electrodomésticos
VER TODO

f | | | |

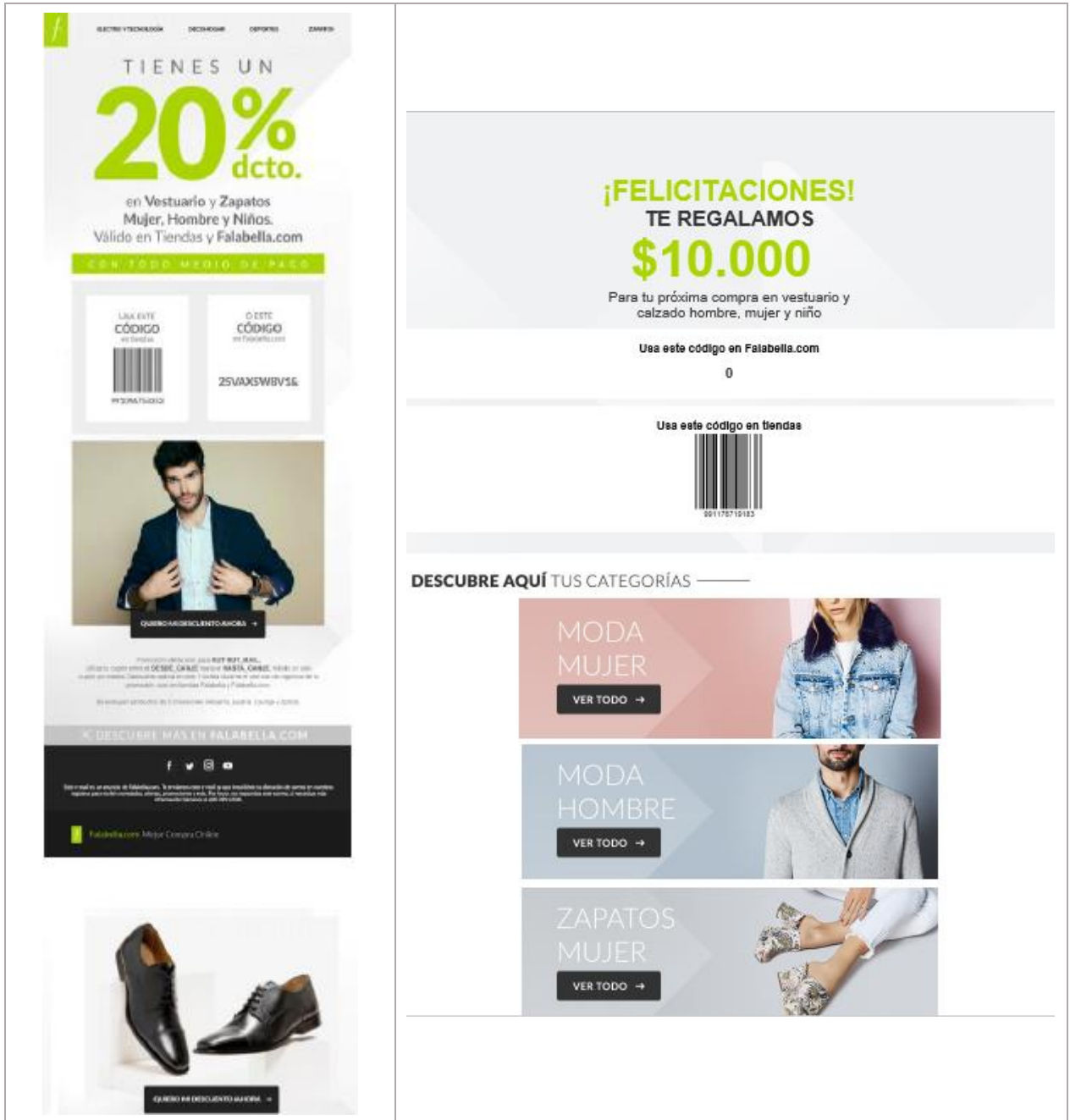
Este e-mail es un anuncio de Falabella.com. Te enviamos este e-mail ya que nos escribiste tu dirección de correo en nuestros registros para recibir novedades, ofertas, promociones y más. Por favor, no respondas este correo, si necesitas más información llámanos al 600 390 6500.

¡PROTEGERNOS DE FRAUDES ES RESPONSABILIDAD DE TODOS!

Con estos 3 simples consejos, podrás evitar inconvenientes:

- 1.- Falabella no envía e-mails desde una extensión distinta a @falabella.com
- 2.- Nunca te pediremos por e-mail contraseñas o claves de acceso
- 3.- Ingresa a Falabella.com directamente desde tu navegador de Internet

Anexos Y: Ejemplo de formato de grafica de envío de correos de marketing multi-categoricos.
Fuente: Graficas de envío Falabella Retail.



Anexos Z: Ejemplo de piezas publicitarias sugeridas para incentivar activación específica (a), activación con compra cruzada (b), activación de múltiples categorías (c) y activación de múltiples categorías asociada a variables sociodemográficas (d) como existencia de hijos(as).

Fuente: Diseño de graficas: Falabella Retail – Disposición espacial: Elaboración propia.

PANTALONES & JEANS
Encuentra tus favoritos a un click

[VER TODO >](#)

PANTALONES
JEANS [VER TODO >](#)

<p>BASEMENT \$14.990 <small>NORMAL: \$19.990</small> VER PRODUCTO ></p>	<p>BEARCLIFF \$12.990 <small>NORMAL: \$16.990</small> VER PRODUCTO ></p>	<p>AMERICANINO \$34.990 VER PRODUCTO ></p>
---	--	--

CAMISAS & CHAQUETAS
Renuévate con los diseños, cortes y estampados que tus marcas favoritas tienen para ti.

<p>\$39.990</p> <p>BASEMENT VER TODO ></p>	<p>\$17.990</p>
--	-----------------

LIQUIDACION 2x1
EN TIENDAS FALABELLA

VESTUARIO HOMBRE

¡COMBINALOS COMO QUIERAS! TEMPORADA VERANO PRODUCTOS SELECCIONADOS

TODO MEDIO DE PAGO
Sobre precio normal

EN FALABELLA.COM
Vestuario Hombre
Temporada Verano
Productos Seleccionados.

50% DCTO [LO QUIERO! >](#)

LIQUIDACION

Apúrate y encuentra las mejores Zapatillas de Mujer y Hombre a precios únicos

[VER TODO ZAPATILLAS >](#)

¡Si este 2020 te propusiste entrenar, ESTE DESCUENTO ES PERFECTO!


VESTUARIO DEPORTIVO MEJORES MARCAS

30% DCTO
Productos Seleccionados.

[VER TODO >](#)

(a)

(b)




Vestuario Mujer

H A S T A

60% DCTO

VER TODO >



PANDORA

40%

Productos seleccionados

¡Encuentra tu producto Pandora favorito y llévalo a un increíble precio! Aplica en Falabella y Falabella.com


VER TODO

Charms

VER TODO CHARMS

Pulseras


VER TODO PULSERAS



Resplandor de la Naturaleza

\$55.200


P. normal \$92.000



Moments Broche Flor

\$55.200


P. normal \$92.000



Corazones Flotantes

\$66.600


P. normal \$111.000



Aros

Pendientes Shine Triángulos

VER TODO AROS



¡LO MEJOR DEL 2019!

El año ya se acaba, pero aún estás a tiempo de descubrirlos.

(c)



brand fashion market

Descubre las mejores marcas en Vestuario

VER TODO VER ESPECIAL



RAPSODIA

VER TODO



PRUNE

VER TODO



LIQUIDACIÓN

Accesorios

40% DCTO

Productos Seleccionados.

Descubre tus Joyas, Carteras o Relojes favoritos a los mejores precios

VER TODO CARTERAS > VER TODO JOYAS >



ALDO

\$24.990

Normal: \$34.990

VER PRODUCTO >



AMPHORA

\$21.990

Normal: \$31.990

VER PRODUCTO >



CON TU TARJETA CMR PUEDES OBTENER DESCUENTOS ÚNICOS

Encuentra grandes ofertas en juegos para niños, Cunas, Coches y Sillas de Auto

- Mundo juguetería [VER TODO >](#)
- Mundo Bebé [VER TODO >](#)

(d)