



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

DISEÑO Y EVALUACIÓN EXPERIMENTAL DE UN SISTEMA DE RECOMENDACIÓN  
DE PRODUCTOS PARA UN SITIO DE COMERCIO ELECTRÓNICO

MEMORIA PARA OPTAR AL TÍTULO DE  
INGENIERO CIVIL INDUSTRIAL

RODRIGO IGNACIO LÓPEZ PÉREZ

PROFESORA GUIA:  
CAROLINA SEGOVIA RIQUELME

MIEMBROS DE LA COMISIÓN:  
ALEJANDRA PUENTE CHANDÍA  
PEDRO URZÚA SALINAS

SANTIAGO DE CHILE  
2020

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL  
POR: RODRIGO IGNACIO LÓPEZ PÉREZ  
FECHA: 2020  
PROF. GUÍA: CAROLINA SEGOVIA R.

## DISEÑO Y EVALUACIÓN EXPERIMENTAL DE UN SISTEMA DE RECOMENDACIÓN DE PRODUCTOS PARA UN SITIO DE COMERCIO ELECTRÓNICO

El comercio electrónico plantea oportunidades únicas a las distintas empresas que participan de este rubro. La vasta cantidad de información y el gran número de productos que pueden ser promocionados son factores en ello. Pero también plantean como desafío el hacer que la experiencia de compra de los visitantes sea lo más significativa posible, con tal de convertir a la mayor cantidad de ellos en consumidores.

Un mecanismo que se suele encontrar en la literatura y en el día a día para lograr un mayor interés de los usuarios son los sistemas de recomendación. En particular, el trabajo de título realizado se propone como una oportunidad de desarrollar un primer sistema de recomendación de productos para una empresa chilena de ventas electrónicas minoristas. El trabajo realizado consistió en una primera instancia en la implementación de dos métodos de clasificación de productos (en base al número de compras y el número de visualizaciones) para mostrar recomendaciones a todos los visitantes del sitio. En una etapa posterior, se incorporó un nivel de personalización de los artículos a recomendar a través de una agrupación de clientes por afinidad utilizando el algoritmo de K-Means sobre su información de compra y navegación. En ambas instancias, los efectos de las recomendaciones fueron probados a través de experimentos controlados en línea (o A/B test).

Los resultados muestran que la incorporación de un sistema de recomendaciones constituido por un carrusel de lo más visto en categorías de electrónica y hogar, y uno de lo más visto en vestuario y calzado, generarán incrementales significativos en el ingreso del sitio. Esta amplitud en el tipo de recomendaciones a mostrar permitió interesar a un número mayor de visitantes. A pesar de la baja exposición del espacio en donde las recomendaciones fueron implementadas, un importante número de visitantes expuestos terminó comprando gracias a las recomendaciones, evidenciando el potencial que este elemento puede tener. Además, la metodología de personalización utilizada permitió lograr que ciertos visitantes se vieran más interesados en las recomendaciones al lograr enfocarlas en sus intereses, a pesar de que esta focalización aún es mejorable. Este mayor interés en categorías variadas resultó en un mayor gasto promedio del grupo expuesto a recomendaciones personalizadas.

Se concluye que la incorporación de sistemas de recomendación, incluso sencillos como los evaluados, mejorarán el desempeño del sitio. Se plantea el uso de estos como punto de comparación para desarrollar sistemas más sofisticados. Se propone a la empresa evaluar la implementación de un mayor número de carruseles de recomendación, complementando esta mayor amplitud de productos con personalización, ya sea por cliente o grupo de ellos, para maximizar la pertinencia de los artículos sugeridos.

*A mi familia,  
por su amor y apoyo incondicional.*

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
<b>2. Descripción del Proyecto</b>	<b>3</b>
2.1. Oportunidad Identificada . . . . .	3
2.2. Objetivos . . . . .	5
2.2.1. Objetivo General . . . . .	5
2.2.2. Objetivos Específicos . . . . .	5
2.3. Alcances . . . . .	5
2.3.1. Punto de partida para sistemas más complejos . . . . .	5
2.3.2. Número de recomendaciones personalizadas y funcionalidad . . . . .	6
2.3.3. Ubicación y elemento de despliegue . . . . .	6
<b>3. Marco Teórico</b>	<b>7</b>
3.1. Sistemas de Recomendación . . . . .	7
3.2. Algoritmos de Segmentación . . . . .	9
3.2.1. Clasificación de algoritmos . . . . .	9
3.2.2. Métricas de validez interna en segmentación . . . . .	10
3.3. Experimento Controlado en línea . . . . .	11
3.4. Test de Hipótesis . . . . .	12
3.5. Indicadores del Negocio . . . . .	13
<b>4. Metodología</b>	<b>15</b>
4.1. Fase Exploratoria . . . . .	16
4.2. Primera Fase: Recomendaciones No Personalizadas . . . . .	17
4.3. Segunda Fase: Recomendaciones Personalizadas usando Segmentos . . . . .	18
4.4. Implementación de Experimentos . . . . .	19
4.4.1. Diseño Experimental para recomendaciones no personalizadas . . . . .	19
4.4.2. Diseño Experimental para recomendaciones personalizadas . . . . .	21
<b>5. Desarrollo Metodológico</b>	<b>23</b>
5.1. Fase Exploratoria . . . . .	23
5.1.1. Descripción de productos y categorías . . . . .	23
5.1.2. Descripción de clientes . . . . .	25
5.1.3. Descripción de navegación . . . . .	27
5.1.4. Experimentación exploratoria . . . . .	29
5.2. Primera Fase: Recomendaciones No Personalizadas . . . . .	30
5.2.1. Generación de recomendaciones . . . . .	30

5.2.2.	Resultados experimentales . . . . .	33
5.2.2.1.	Resultados generales de conversión y gasto . . . . .	34
5.2.2.2.	Resultados de uso y post-click . . . . .	36
5.2.2.3.	Resultados por dispositivo . . . . .	39
5.3.	Segunda Fase: Recomendaciones Personalizadas usando Segmentos . . . . .	42
5.3.1.	Generación de segmentos por afinidad . . . . .	43
5.3.1.1.	Afinidad según compras . . . . .	43
5.3.1.2.	Afinidad según navegación . . . . .	47
5.3.1.3.	Asignación de segmentos para recomendaciones . . . . .	50
5.3.2.	Generación de recomendaciones por segmento . . . . .	51
5.3.3.	Resultados experimentales . . . . .	52
5.3.3.1.	Resultados generales de conversión y gasto . . . . .	53
5.3.3.2.	Resultados de uso y post-click . . . . .	54
<b>6.</b>	<b>Conclusiones</b>	<b>57</b>
<b>7.</b>	<b>Trabajo Futuro</b>	<b>60</b>
	<b>Bibliografía</b>	<b>60</b>
	<b>Anexo A. Visualización de las recomendaciones</b>	<b>62</b>
	<b>Anexo B. Análisis Exploratorio</b>	<b>64</b>
	<b>Anexo C. Resultados Segmentación</b>	<b>66</b>

# Índice de Tablas

4.1.	Distinción entre categorías en base al monto y frecuencia de compra. . . . .	18
4.2.	Nomenclatura de grupos para primera experimentación. . . . .	20
4.3.	Tamaño muestral para alcanzar poder estadístico a priori en base a una estimación del valor de la métrica y el lift a encontrar. . . . .	21
5.1.	Agrupaciones jerárquicas de productos en la empresa. . . . .	24
5.2.	Porcentaje de clientes del total que realizan alguna compra en cada categoría.	26
5.3.	Productos únicos incluidos en cada versión de las recomendaciones no personalizadas. . . . .	31
5.4.	Estadística de precios de las recomendaciones no personalizadas generadas. . .	32
5.5.	Tasa de conversión de los productos recomendados (no personalizado). . . . .	33
5.6.	Resultados de conversión y gasto tras primera fase de experimentación. . . . .	34
5.7.	Tasa de uso por carrusel para los distintos grupos no personalizados. . . . .	37
5.8.	Comparación tasa de uso entre lo más comprado y lo más visto. . . . .	37
5.9.	Comparación tasa de uso entre recomendadores de categorías duras y blandas.	38
5.10.	Resultados de métricas posteriores al click en recomendaciones no personalizadas.	38
5.11.	Estimación del ingreso adicional esperado por sesión aportado directamente por las recomendaciones no personalizadas. . . . .	39
5.12.	Resultados generales de conversión y gasto para dispositivos de escritorio (no personalizadas). . . . .	40
5.13.	Resultados generales de conversión y gasto para dispositivos móviles (no personalizadas). . . . .	40
5.14.	Comparación de tasas de uso por dispositivo para recomendaciones no personalizadas. . . . .	41
5.15.	Estadísticas de los productos recomendados por categoría. . . . .	53
5.16.	Resultados de conversión y gasto tras segunda fase de experimentación. . . . .	54
5.17.	Comparación de la tasa de uso entre grupo expuesto a recomendaciones de categorías blandas y personalizadas por segmento. . . . .	55
5.18.	Resultados de métricas posteriores al click en recomendaciones personalizadas.	56
5.19.	Estimación del ingreso adicional esperado por sesión aportado directamente por las recomendaciones personalizadas. . . . .	56
C.1.	Cuantiles para la distribución del gasto total por cliente tras eliminación de outliers. . . . .	66

# Índice de Ilustraciones

1.1.	Evolución de los ingresos de la empresa. . . . .	2
4.1.	Caracterización de las líneas de productos según precio promedio y unidades vendidas. . . . .	15
5.1.	Caracterización de las líneas de productos según precio promedio y unidades vendidas. . . . .	24
5.2.	Distribución de frecuencia del número de líneas distintas compradas por cliente.	27
5.3.	Esquema de las secciones de la página de inicio y su tasa de utilización aproximada. . . . .	28
5.4.	Distribución de clicks por slot en carrusel de fase exploratoria. . . . .	29
5.5.	Ejemplo de recomendaciones no personalizadas desplegadas en el sitio. . . . .	31
5.6.	Posición de los clicks en los carruseles de recomendaciones no personalizadas para dispositivos de escritorio. . . . .	41
5.7.	Posición de los clicks en los carruseles de recomendaciones no personalizadas para dispositivos móviles. . . . .	42
5.8.	Frecuencia de compra en días por cliente. . . . .	44
5.9.	Resultados para K-Means utilizando variables FM totales . . . . .	45
5.10.	Segmentación de clientes en base a FM utilizando K-Means con 9 grupos . . . . .	45
5.11.	Resultados para el puntaje de Sum of Squares en segmentación por compras por categoría. . . . .	46
5.12.	Distribución del número de sesiones con visita a PDP por usuario. . . . .	48
5.13.	Resultados para el puntaje de Sum of Squares en segmentación por navegación por categoría. . . . .	49
5.14.	Ejemplo de recomendaciones personalizadas desplegadas en el sitio. . . . .	51
A.1.	Despliegue del sitio web en dispositivos móviles y distribución de las vistas. . . . .	62
A.2.	Despliegue del sitio web en dispositivos de escritorio y distribución de las vistas. . . . .	63
B.1.	Pirámide demográfica de los clientes activos. . . . .	64
B.2.	Precio promedio y unidades vendidas por producto para las distintas agrupaciones jerárquicas. . . . .	65
B.3.	Distribución de las unidades y el monto vendido dentro de cada clase. . . . .	65
C.1.	Distribución del gasto total por cliente tras eliminación de outliers. . . . .	66
C.2.	Distribución de la recencia de compra al realizar segmentación en base a RFM. . . . .	67
C.3.	Resultados para los centros de K-Means utilizando el gasto porcentual por categoría. . . . .	68
C.4.	Resultados para los centros de K-Means utilizando el gasto porcentual por categoría y muestreo por segmentación. . . . .	68
C.5.	Distribución del número de hits por usuario. . . . .	68

C.6.	Segmentación de clientes en base a al número de hits y sesiones totales utilizando KMeans. . . . .	69
C.7.	Resultados para los centros de K-Means utilizando intensidad de navegación por categoría. . . . .	69
C.8.	Resultados para los centros de K-Means utilizando intensidad de navegación por categoría y muestreo por segmentación. . . . .	70
C.9.	Reglas de asignación de categorías utilizando los resultados de las etapas de segmentación. . . . .	71

# Capítulo 1

## Introducción

En el siguiente informe se describe el proyecto de título realizado en una importante empresa de tiendas por departamento nacional. Esta empresa ha sido durante varios años la número uno en ventas en el país en el mercado de las ventas minoristas, gracias a sus distintos canales de venta como son las tiendas físicas, las ventas telefónicas y el comercio electrónico.

Además de su larga historia atendiendo las necesidades de sus clientes en Chile, la empresa cuenta también hace varios años con operaciones en Perú, Colombia y Argentina, tanto en formato de tiendas físicas como ventas online, por lo que también es considerado un importante competidor dentro de la región.

Una de las características de esta empresa es la gran diversidad de productos que ofrece, que incluyen vestuario, artículos del hogar, electrónica, decoración, entre otros. En la actualidad la empresa comercializa más de 300.000 productos (SKU's) diferentes, agrupados en cuatro niveles de jerarquía denominados como Líneas, Sub líneas, Clases, y Sub Clases. Es en esta situación de alta diversidad en donde el comercio electrónico de la firma juega un rol muy importante, pues a pesar de contar en la actualidad con 47 grandes tiendas a lo largo de todo el país, la tarea de mostrar a los consumidores el gran surtido de productos y marcas de manera ordenada es algo realizable solamente a través de la plataforma web.

En cuanto al volumen de ventas anuales, los reportes institucionales de la firma indican ingresos en Chile superiores a los 4.000 millones de dólares en 2019. En este mismo año, la empresa informa que alrededor de un 21 % de estos ingresos corresponden a ventas online, lo que significa un crecimiento del 22 % con respecto al año anterior, mostrando el dinamismo de este canal y su rápida adopción por parte de los clientes. Estos ingresos provienen anualmente de alrededor de 5,5 millones de clientes en todo Chile, hombres y mujeres, de todas las edades y situaciones socioeconómicas, lo que describe la diversidad de perfiles a los que la empresa debe atender. Dentro de estos 5,5 millones de clientes, alrededor de un tercio de ellos utiliza el sitio web de la empresa para realizar alguna de sus compras cada año, existiendo categorías de productos en que la preferencia por los medios digitales es mucho mayor (como es, por ejemplo, el caso de computación y electrodomésticos).

De acuerdo a los datos reportados en los últimos años por la empresa, sus ingresos se han mantenido en un constante aumento (en términos nominales), aunque la tasa de crecimiento ha ido disminuyendo, como se puede observar en los datos de la Figura 1.1. Estas cifras

indican que la empresa se encuentra actualmente en una etapa de madurez, y que a pesar de su estado de consolidación existe la posibilidad de que su crecimiento se estanque en el futuro o incluso que comience un decrecimiento de sus ingresos.

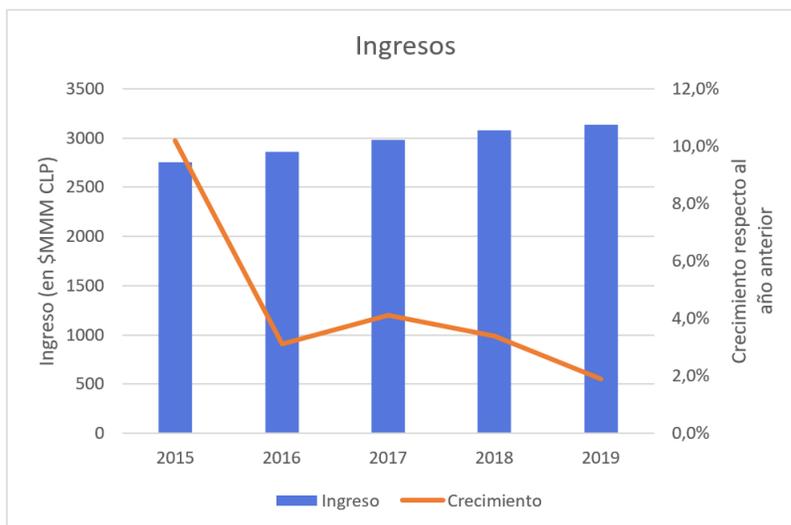


Figura 1.1: Evolución de los ingresos de la empresa.

A pesar de los datos anteriores, la empresa continúa expandiéndose en lo relacionado con el canal digital. Esto pues en los últimos años el crecimiento de las ventas por internet ha sido explosivo, más allá de que el crecimiento general haya disminuido. Según los datos informados por la empresa en los últimos años, las ventas online aumentaron un 36 % en el año 2017, un 26 % en el año 2018 y un 22 % en el año 2019. Esta tasa de crecimiento mayor en el canal online en comparación al crecimiento total de las ventas de la empresa indica que las ventas a distancia componen cada año una porción más grande de los ingresos totales. Aún así, la tendencia sugiere que si no se potencia este canal de la manera correcta, su crecimiento también corre el riesgo de estancarse.

Según los antecedentes antes señalados, se ve como los medios digitales se presentan como una oportunidad para que el negocio siga creciendo adaptándose a las tendencias actuales de los consumidores, presentando una oportunidad para reducir costos operacionales y para aumentar sus ventas al llegar a un espectro más amplio de clientes con una mayor oferta de productos. Es por esto que la estrategia actual de la empresa consiste en potenciar una experiencia omnicanal (integrando medios físicos y digitales), intentando generar una experiencia de compra significativa de cara al cliente teniendo en consideración que la apertura ofrecida por el canal digital para llegar a más clientes también favorece la mayor competitividad.

En la actualidad la empresa se encuentra desarrollando y utilizando diversos mecanismos de atracción y activación de clientes a través del medio online, tales como campañas de fidelización, email marketing, utilización de triggers, personalización de elementos del sitio, entre otras. Es en este contexto en donde se posicionan los sistemas de recomendación de productos como un mecanismo adicional para potenciar la experiencia de compra de los usuarios, haciendo uso de la gran cantidad de información que la empresa registra cada día sobre ellos.

# Capítulo 2

## Descripción del Proyecto

### 2.1. Oportunidad Identificada

Los datos anteriormente descritos señalan que una importante oportunidad para la empresa se encuentra en su comercio electrónico, que cada año crece en cuanto al volumen de ventas, productos y clientes. Así mismo, la creciente popularidad de los comercios electrónicos en la última década en Chile y el mundo ha significado un aumento en el número de competidores, pues a través de este medio se reducen los costos de entrada y promoción para empresas nuevas y existentes, y se reducen los costos de acceso para los consumidores. Por esto, para que la empresa siga creciendo y liderando el mercado de las ventas minoristas, debe ofrecer un valor agregado a la experiencia de navegación y compra que enfrentan los clientes cada día.

Las características únicas del comercio electrónico presentan oportunidades como el bajo costo para exhibir productos o la capacidad de recopilar gran cantidad de información de los usuarios. A pesar de esto, dado el gran volumen de productos que son comercializados por la empresa es un desafío constante lograr que cada cliente que entra al sitio pueda encontrar algún producto de su interés, resultando finalmente en una venta. La heterogeneidad de los más de 2 millones de visitantes que el sitio web de la empresa tiene indica que el interés de ellos frente a los más de 300 mil productos ofrecidos en la plataforma será muy diferente, existiendo sólo un número acotado de productos que serán verdaderamente relevantes para cada usuario. Es bajo esta misma premisa que sitios web en todo el mundo, de distintos rubros, han logrado agregar valor a la experiencia de los usuarios ofreciéndoles recomendaciones que reduzcan sus costos de búsqueda (un aumento del *product finding*), fomenten el descubrimiento de nuevos productos (un aumento del *product discovery*) y guíen su navegación. Cuando los datos y la infraestructura tecnológica lo permitan, estas recomendaciones pueden también incluir diversos niveles de personalización para lograr que lo mostrado sea lo más afín posible a los intereses de cada visitante.

Estudios académicos han analizado empíricamente el impacto de los sistemas de recomendación en sitios web, encontrando consistentemente beneficios tales como el aumento en la conversión de visitantes a compradores, el aumento de las ventas cruzadas o el incremento en el descubrimiento de productos [1]. En términos más prácticos, una rápida exploración por los sitios web de los retailers más importantes del mundo, como es el caso de Amazon, muestran la implementación de recomendaciones de productos en diversos formatos y a través de distintas secciones del sitio. Según lo analizado por la consultora McKinsey, se estima que

un 35 % de las ventas generadas por la empresa norteamericana provienen de alguno de sus sistemas de recomendación [2]. En este mismo estudio también se destaca el mayor impacto que tienen las recomendaciones realizadas por otros consumidores, por sobre recomendaciones provenientes desde áreas comerciales.

A pesar de los beneficios antes señalados, en la actualidad la empresa no incluye en ninguna de sus páginas algún tipo de sistema de recomendación, desaprovechando la gran cantidad de información que se almacena sobre los usuarios y sus comportamientos dentro del sitio. Es así como el proyecto se plantea como una oportunidad para diseñar un primer sistema de recomendación de productos que sea funcional e implementable, buscando replicar los efectos positivos obtenidos por otros sitios web. Dado que este proyecto será uno de los primeros acercamientos de la empresa a la generación automática de un sistema de este estilo, se propone dar una mirada amplia sobre las recomendaciones que se pueden llegar a generar basadas en la popularidad de los artículos disponibles en el sitio, utilizando como medida la cantidad de compras y visualizaciones registradas. La empresa ha decidido que un espacio del sitio que potencialmente puede verse beneficiado de un sistema de recomendaciones es la página de inicio, utilizando un elemento denominado como “carrusel” de productos para desplegar los artículos sugeridos. Este elemento consiste en un listado horizontal de productos, deslizable para acceder a distintas vistas, y que es utilizado por gran cantidad de comercios electrónicos.

Se espera además, según los antecedentes descritos, que la efectividad de las recomendaciones mejore al añadir un grado de personalización sobre ellas. Dados los plazos y las capacidades técnicas con las que cuenta actualmente la empresa, se propone como una segunda etapa en el proyecto la generación de recomendaciones diferenciadas para distintos segmentos de clientes, construidos según sus preferencias sobre las distintas categorías de productos comercializados en el sitio web, en base a la información que se dispone sobre sus historiales de compra y de navegación. Se busca con esto diseñar un método que permita focalizar aún más las recomendaciones y aumentar la pertinencia de los artículos a sugerir. Para evaluar el impacto de los sistemas de recomendación a generar se propone el diseño y la ejecución de experimentos controlados en línea, con el objetivo de determinar diferencias estadísticamente significativas en las métricas claves de conversión, venta y uso del sitio.

El valor de la implementación de estos sistemas radica en el bajo costo de implementación y mantenimiento que ellos tienen, que consiste principalmente en el tiempo necesario para diseñar y automatizar los procesos de generación de las recomendaciones. Dado este bajo costo, incluso incrementos pequeños en el ingreso deben ser considerados dado el gran volumen de visitantes que el sitio tiene cada mes. En este sentido, incluso si un porcentaje pequeño de los visitantes se ve afectado positivamente por los carruseles de recomendación, se puede justificar monetariamente la inclusión de uno o más elementos de este estilo.

Más allá de los incrementos en venta que los sistemas puedan generar, se espera que la metodología permita obtener aprendizajes que sirvan como base para el conocimiento que la empresa tiene acerca de los mecanismos de recomendación y guíen las iteraciones futuras a realizar en esta materia.

## **2.2. Objetivos**

### **2.2.1. Objetivo General**

Implementar un sistema de recomendación de productos en el sitio de comercio electrónico de la empresa, utilizando distintos niveles de personalización, buscando impactar positivamente métricas de conversión y uso.

### **2.2.2. Objetivos Específicos**

Para cumplir con el objetivo general del proyecto, se definen los siguientes objetivos específicos:

- Diseñar recomendaciones no personalizadas de productos e implementarlas dentro del sitio web.
- Generar segmentos de clientes en base a sus preferencias de compra y navegación por categoría.
- Diseñar recomendaciones para cada segmento de clientes e implementarlas dentro del sitio web.
- Diseñar y conducir experimentos controlados en línea que permitan evaluar el impacto de los sistemas de recomendación generados.

## **2.3. Alcances**

### **2.3.1. Punto de partida para sistemas más complejos**

Se debe tener en consideración que el proyecto no busca encontrar un sistema de recomendación de productos que optimice la conversión o el gasto de los visitantes del sitio web (en el sentido de encontrar un máximo absoluto), dado que existirán una gran cantidad de factores a tener en cuenta al momento de diseñar cualquier sistema de recomendación, y en los plazos definidos será imposible evaluar todas las configuraciones posibles. Por lo mismo, el trabajo desarrollado se concibe como un análisis exploratorio de algunas opciones de sistemas de recomendación implementables en el sitio de comercio electrónico, esperando que los sistemas diseñados y evaluados generen aprendizajes y sirvan como un primer nivel de comparación para la empresa, así como una base para incorporar mejoras o diseñar sistemas más sofisticados en el futuro.

Dado el carácter exploratorio ya señalado del proyecto, se intentará diseñar y exponer los sistemas generados a la mayor cantidad de usuarios posibles, no habiendo restricciones a priori acerca de clientes objetivos o productos a incluir en las recomendaciones. De esta forma se busca que los sistemas de recomendación a generar funcionen de la forma más automatizada posible, cumpliendo con el requerimiento de la empresa de que estos no necesiten grandes esfuerzos para ser mantenidos.

### **2.3.2. Número de recomendaciones personalizadas y funcionalidad**

En relación a lo anterior, también se ha considerado para el diseño de los sistemas de recomendación detallados en este informe la capacidad tecnológica existente al momento de desarrollar el proyecto, que comprende el período entre enero y julio del 2020. Esta limitante consiste principalmente en que para personalizar las recomendaciones de los usuarios identificados esto debe hacerse a nivel de grupo y con un número reducido de ellos, debido a ciertas configuraciones manuales que deben hacerse. Si bien pueden existir otros modelos más sofisticados que permitan generar recomendaciones más personalizadas para cada usuario y con una mayor tasa de actualización, la implementación de sistemas de este estilo requiere de un mayor desarrollo previo, lo cual escapa del alcance de este trabajo.

Además, el trabajo de memoria no busca generar solamente prototipos conceptuales de sistemas de recomendación, si no que se plantea con la finalidad de generar versiones funcionales de ellos, es decir, que puedan ser implementados de manera inmediata en el sitio web de la empresa y se puedan mantener operativos una vez concluido el proyecto.

### **2.3.3. Ubicación y elemento de despliegue**

Por parte de la empresa se dispuso como espacio de interés la página de inicio del sitio web. En particular se restringió el trabajo a un espacio fijo bajo otras secciones principales de la página, como se observa en la Figura A.2 del anexo. La restricción a este espacio será tomada en cuenta al momento de decidir la información a utilizar, que consistirá principalmente en el historial de compra y navegación de los clientes, excluyendo información contextual. Además, el espacio señalado tendrá en general una baja exposición comparada con otras secciones de la página principal, dado que los usuarios deberán desplazarse hacia abajo en el sitio para llegar al espacio dispuesto para la experimentación. Esto seleccionará automáticamente al tipo de clientes que serán efectivamente expuestos a las recomendaciones.

En relación a lo anterior, también se debe considerar que las recomendaciones de productos serán entregadas a los usuarios en formato de un carrusel de productos, elemento del sitio cuyas configuraciones gráficas y de usabilidad han sido predefinidas por otras áreas de la empresa y no serán configurables dentro del trabajo más allá de los artículos que lo componen, aunque no se debe dejar de señalar que también corresponde a un elemento en desarrollo que puede ser mejorado en el futuro.

# Capítulo 3

## Marco Teórico

### 3.1. Sistemas de Recomendación

Los comercios electrónicos han permitido a las compañías durante varios años ofrecer mayores opciones de compra para cubrir un rango más amplio de posibles consumidores. Esto también ha significado una mayor cantidad de información que los usuarios deben procesar para encontrar productos que satisfagan sus necesidades. Estudios exploratorios acerca del tema han mostrado como los sistemas de recomendación han permitido a las compañías de comercio electrónico filtrar esta información y facilitar así la experiencia de navegación de los usuarios, convirtiéndose en un mecanismo común en la actualidad para mejorar la experiencia de compra en los sitios web [3]. En base a estos estudios, se puede evidenciar que los sistemas de recomendación benefician las ventas de tres formas distintas:

- Convierten usuarios que navegan, en usuarios que compran. Esto ayudando a los visitantes a encontrar productos que, consciente o inconscientemente, desean comprar.
- Incrementan las ventas cruzadas, pudiendo sugerir productos adicionales o complementarios a los visitantes del sitio.
- Mejoran la lealtad de los clientes. En el caso de generar un valor agregado de cara a los consumidores, estos se verán incentivados a regresar a aquellos sitios en donde se les ayude a encontrar productos interesantes.

Si bien los sistemas de recomendación son muy populares en la actualidad en los comercios electrónicos, no son exclusivos de estas plataformas. También se puede evidenciar su uso al momento de recibir recomendaciones de películas, música o incluso personas. Desde un punto de vista abstracto las nociones que aplican en cada uno de estos contextos son similares y en principio pueden ser aplicadas sobre cualquier entidad. En particular los comercios electrónico suelen usar como entidad principal los productos, pero estas también pueden corresponder a marcas o categorías.

Se debe mencionar que los sistemas de recomendación pueden presentarse en un sitio web con variadas interfaces, utilizando distintas lógicas o métodos de generación, pudiendo ser o no personalizadas para cada usuario en base a la información con la que se disponga. Para

el caso de este proyecto, se centrará el análisis en la utilización de una interfaz de recomendación denominada comúnmente como carrusel de productos, permitiendo agrupar una gran cantidad de artículos en una franja reducida de espacio.

Según lo señalado en la literatura, los sistemas de recomendación pueden ser clasificados según la información que se utilice para su construcción, así como los procedimientos que harán uso de ella para determinar el conjunto de artículos (en este caso, productos) a sugerir [4]. Así se definen seis técnicas de recomendación:

1. Colaborativas: consistentes principalmente en la asociación de grupos de clientes con cierta similitud, medida en base a sus preferencias explícitas o implícitas (evaluaciones de productos, compras, etc.). Los productos a recomendar para cada usuario son extrapolados a partir de las evaluaciones y decisiones de compra de estos clientes similares.
2. Basadas en contenido: estas técnicas hacen uso de las atributos que poseen los distintos productos, generando asociaciones entre aquellos que comparten gran cantidad de atributos similares.
3. Demográficas: utilizan información de los clientes para encontrar grupos demográficos similares y analizar las preferencias de productos que cada uno de estos grupos manifiesta.
4. Basadas en utilidad: define una función de utilidad en base a los atributos de que posee cada producto, buscando estimar la probabilidad de preferencia de cada cliente o grupo de clientes. Las recomendaciones serán entonces los productos con mayor probabilidad.
5. Basadas en conocimiento: corresponde a técnicas que infieren productos que calzan con las necesidades de un grupo de usuarios en base al conocimiento del negocio y el estudio de comportamientos anteriores.
6. Híbridas: como su nombre lo indica, son técnicas de recomendación que integran 2 o más de las técnicas mencionadas anteriormente, ya sea de manera integrada o secuencial.

Cada una de estas técnicas de recomendación tiene sus puntos a favor y en contra, y su elección dependerá de los alcances buscados con la intervención, la cantidad de datos con los que se disponga y los recursos tecnológicos con los que se cuente. En general, todos los sistemas de recomendación deberán responder a las interrogantes de qué artículos recomendar y a quienes. Un método que ha ganado popularidad y ha mostrado buenos resultados en los sistemas de recomendación actuales consiste en la utilización de filtros colaborativos [5]. Estos filtros tienen como idea fundamental que los usuarios que han compartido los mismos intereses en el pasado también lo harán en el futuro. Tienen como ventaja que al mismo tiempo que se encuentran relaciones entre clientes, se decide que artículos mostrar a cada uno.

Para el caso de este proyecto los filtros colaborativos generarán una gran cantidad de recomendaciones diferentes, que serán difíciles de implementar con las tecnologías actuales de la empresa. Por esto se estima conveniente utilizar un modelo híbrido de recomendación que considere tanto asociaciones entre clientes (similar a lo realizado por un filtro colaborativo), pero obteniendo grupos de personas similares en base a sus intereses de compra y navegación. Esto permitirá tener un número acotado y controlado de recomendaciones a generar, lo que permitirá manejar de mejor forma la calidad y la coherencia de ellas, sin dejar de considerar

que los modelos generados servirán como una primera iteración de sistemas mas sofisticados en el futuro. Así la probabilidad de preferencia de un cliente hacia un grupo de productos será estimado de manera independiente al proceso de elección de los productos a sugerir.

## 3.2. Algoritmos de Segmentación

La utilización de algoritmos de segmentación será necesaria para la asociación automática y objetiva de clientes con similares preferencias, medidas en base a sus comportamientos de compra y navegación. Al utilizar modelos de este estilo, se podrán identificar patrones en grandes volúmenes de datos, pudiendo además controlar por el nivel de especificidad que se requiera para la generación de las recomendaciones, por ejemplo al modificar el número de grupos a generar o utilizando modelos de segmentación en forma de ensamble.

### 3.2.1. Clasificación de algoritmos

Existe una gran diversidad de algoritmos de segmentación que utilizan distintas reglas para la construcción de grupos. Por esto es posible obtener una clasificación de los algoritmos más populares en la actualidad [6], generando cinco grupos en base a los procedimientos de agrupamiento que realizan:

1. Basados en Particiones: corresponde a algoritmos que se basan en dividir los datos en un número definido de grupos donde, en general, cada grupo debe estar compuesto de al menos un punto y cada punto debe pertenecer a exactamente un grupo. Algoritmos populares en este grupo son K-Means, Fuzzy C-Means, CLARA, entre otros.
2. Basados en Jerarquías: en este caso los datos son organizados a modo de árbol y de manera secuencial en base a su proximidad hasta alcanzar el número requerido de grupos. Entre estos modelos se incluyen BIRCH, CURE, entre otros.
3. Basados en Densidad: en este tipo de algoritmos, los grupos son armados de manera progresiva en base a los vecinos más cercanos manteniendo cierto nivel de densidad a través de cada iteración. Modelos que utilizan algoritmos de este estilo son DBSCAN o DENCLUE.
4. Basados en Grilla: este tipo de algoritmos divide los datos en grillas, permitiendo acumular los datos en un nivel superior para luego realizar agrupaciones a nivel de grillas, permitiendo menores tiempos de procesamiento. Algoritmos que utilizan estas técnicas son STING u OptiGrid.
5. Basados en Modelo: en este caso los grupos son construidos teniendo en consideración la optimización de un modelo matemático predefinido, basándose en el supuesto de que los datos están construidos a partir de distribuciones de probabilidad subyacentes. Ejemplos de estos algoritmos son MCLUST o Expectation Maximization (EM).

En el caso del proyecto que se ha planteado, se considera que una buena opción será la utilización de algoritmos basados en particiones, dado que el objetivo de la segmentación

será poder encontrar un número establecido de grupos de clientes con preferencias similares sobre un número de categorías de productos. Por ejemplo, el modelo de K-Means permitirá agrupar a los clientes en base a distancias en variables como el gasto por categoría y navegación por categoría. Además, a través de los centroides calculados para los K segmentos, se puede interpretar la naturaleza de los segmentos, lo que puede ser utilizado para la posterior generación de recomendaciones dirigidas a ese segmento en específico.

### 3.2.2. Métricas de validez interna en segmentación

Los modelos de segmentación pertenecen a la categoría de algoritmos no supervisados, es decir, no es posible conocer los segmentos *reales* a los que pertenecen los datos, y por lo tanto las estimaciones generadas a través de los modelos no pueden ser contrastadas a través de un nivel de precisión o exactitud. Esto propone una dificultad al momento de decidir el número de segmentos indicados para obtener un buen desempeño.

A pesar de lo anterior, se puede hacer uso de métricas de validez interna para tener cierta noción de la naturaleza de la segmentación realizada. Estas métricas de validez interna se basan en la noción de que una buena segmentación permitirá crear grupos muy similares en su interior (*compactos*) y muy diferentes entre ellos (*separados*). Según estudios realizados al respecto [7] existirán tres principales métricas que presentan buenos resultados al momento de reflejar la calidad de los grupos a generar:

- Calinski-Harabasz Score (CH):

$$CH = \frac{BSS}{WSS} * \frac{N - k}{k - 1} \quad (3.1)$$

Donde  $BSS$  es la suma de cuadrados de las distancias entre el centro de cada cluster y el centroide de los datos (midiendo separación),  $WSS$  es la suma de cuadrados de las distancias entre cada punto y el centro de su cluster (midiendo compacidad),  $N$  es la cantidad de puntos y  $k$  el número de segmentos. La factor de normalización hará que el puntaje disminuya a medida que  $k$  aumenta. Un mayor puntaje de CH indicará una segmentación de mejor calidad.

- Sum-of-Squares (SS):

$$SS = \frac{WSS}{BSS} * k \quad (3.2)$$

Consiste en una adaptación del CH score, invirtiendo la relación entre compacidad y separación, con un impacto mucho mayor del número de segmentos  $k$ . Un menor puntaje de SS indicará una segmentación de mejor calidad.

- Silhoutte:

El método de silueta también establece una relación entre compacidad y separación, pero calcula un puntaje para cada elemento en el conjunto de datos, promediando el puntaje de silueta de la totalidad de datos. Si  $x_i$  pertenece al cluster  $p$  entonces:

$$\frac{1}{N} \sum_{i=0}^N s_{x_i} \text{ con } s_{x_i} = \frac{b_{q_i} - a_{p_i}}{\max\{a_{p_i}, b_{p_i}\}} \text{ y } b_{q_i} = \min d_{q_i} \quad (3.3)$$

Donde  $d_{q_i}$  es la distancia promedio entre el punto  $x_i$  y cada otro punto en el cluster  $q$ ,  $a_{p_i}$  es la distancia promedio entre el punto  $x_i$  y todos los demás puntos de cluster  $p$ . Un número más cercano a 1 indicará una segmentación de mejor calidad.

### 3.3. Experimento Controlado en línea

Una vez contruidos los sistemas de recomendación será necesario evaluar su desempeño en base a métricas relevantes para el negocio. A pesar de que se pueden consultar datos históricos para realizar la evaluación, una de las mejores formas de probar el sistema es en una situación real. Para ello, existe la posibilidad ejecutar un diseño experimental que permita acercamientos a las respuestas en base a las hipótesis planteadas en primer lugar. La relevancia fundamental de un experimento bien diseñado radica en poder establecer una causalidad entre cierta intervención (en este caso, la incorporación de un sistema de recomendación de productos) y el efecto observado, que puede ser cuantificado permitiendo concluir acerca de la calidad de la intervención [8].

Los diseños experimentales no son únicos, existen muchos caminos para responder las preguntas investigativas que se puedan plantear, con distintas selecciones de muestra y asignación de tratamientos. En particular, se hará uso de un estudio experimental controlado, en el cual se definen previamente los tratamientos a implementar y cómo responderán la pregunta de investigación, así como se estudiarán y determinarán los riesgos estadísticos asociados al diseño planteado.

En el caso particular de los sitios de comercio electrónico, será más pertinente hablar de experimentos controlados en línea, que se diferencian de otros experimentos controlados principalmente por el volumen de visitantes que pueden ser expuestos a los tratamientos, presentando un desafío para asegurar la calidad de las mediciones, así como también tienen la particularidad de realizar un asignación aleatoria a los tratamientos a medida que los visitantes llegan al sitio en contraste a un muestro aleatorio previo que es utilizado en otros diseños experimentales.

De la literatura basada en este tipo de experimentos [9] se logran identificar los siguientes puntos críticos que será necesario analizar y tener en consideración para poder conducir de manera correcta experimentos controlados en línea, asegurando la calidad de los resultados a obtener:

1. Criterio de Evaluación Global (CEG): correspondiente a la métrica que determinará si el experimento fue o no exitoso. A pesar de que en cada experimento pueden haber distintas variables a medir con el objetivo de complementar el análisis, se recomienda predefinir un único CEG, en línea con los objetivos de la organización.
2. Factor y Variación: factor se entiende como una variable de experimentación que se cree puede impactar el CEG. Un factor puede tener dos o más variaciones. Un A/B

test tradicional corresponderá a un prueba de un factor con dos variaciones.

3. Unidad Experimental: entidad sobre la cual se calcularán las métricas y que se asume son independientes entre si. Por ejemplo una unidad puede ser usuarios, visitas, compras, etc.
4. Hipótesis Nula: será la hipótesis de que el CEG no sufrirá ningún cambio producto de la intervención en el factor experimental a evaluar.
5. Nivel de Significancia ( $\alpha$ ): corresponde a la probabilidad de cometer un error tipo I (falso positivo), es decir, rechazar la hipótesis nula incorrectamente.
6. Poder Estadístico ( $1 - \beta$ ) y Tamaño Muestral ( $N$ ): el poder estadístico consiste en la probabilidad de no cometer un error tipo II en el análisis (falso negativo), es decir, no rechazar incorrectamente la hipótesis nula. En otras palabras, el poder estadístico se entiende como la probabilidad de evidenciar un efecto significativo en caso de existir. El cálculo del poder estadístico será importante pues un experimento con bajo poder no permitirá obtener conclusiones a partir de los resultados. El poder estadístico dependerá del nivel de significancia ( $\alpha$ ), del efecto esperado en la métrica de evaluación y del tamaño muestral alcanzado. Esto hace necesario realizar un análisis de poder a priori que permita calcular el tamaño muestral objetivo al momento diseñar un experimento con el nivel de significancia y el poder estadístico que se requiera.
7. Test Nulo o Test A/A: en el caso particular de la experimentación en línea, será muy importante revisar la calidad de los datos recopilados por los sistemas tecnológicos que funcionan dentro del sitio, dado que no es extraño encontrarse con errores de medición en este tipo de pruebas debido dado el gran volumen de datos y el gran número de tecnologías que interactúan entre sí [10]. Una forma de comprobar la salud de las métricas que serán obtenidas antes y durante el transcurso de los experimentos es a través de la realización de tests nulos, que consisten en dividir el tráfico de visitantes tal como lo hará el sistema en los experimentos, pero sin ofrecer ningún cambio entre los dos (o más) grupos. Lo esperable es que al calcular métricas sobre ambos grupos no se evidencie una diferencia estadísticamente significativa.

### 3.4. Test de Hipótesis

Para determinar si las métricas tuvieron mejoras entre los grupos experimentales, no bastará solo con determinar si el valor calculado es mayor o menor que otro. Lo tradicional para realizar estas comparaciones será plantear un test de hipótesis con respecto a la métrica de interés para concluir con cierta significancia estadística, basado en ciertos supuestos de sus distribuciones de probabilidad e incorporando la variabilidad de los datos. Estos test pueden definirse tanto de una como de dos colas (dependiendo del sentido de la diferencia a encontrar). En el caso de un test de una cola (derecha) se plantean las hipótesis nula y alternativa de la siguiente forma:

$$H_0 : \hat{\theta}_1 \leq \hat{\theta}_2$$

$$H_a : \hat{\theta}_1 \geq \hat{\theta}_2$$

La forma de concluir sobre un test de hipótesis tradicional dependerá del tipo de métrica que se estará comparando. En el caso de métricas que representan proporciones en base a opciones binarias (como pueden ser el número de visitantes que compran dividido en el total, o el número de clicks partido en el número de visualizaciones) el análisis se realizará a través de la construcción de un estadístico  $Z$  que tendrá una distribución  $N(0, 1)$ :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}; \text{ con } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \quad (3.4)$$

En el caso de comparar métricas que correspondan a promedios de números (como el gasto promedio por cliente o por sesión) el análisis debe realizarse a través de la construcción de un estadístico  $t$  que distribuirá según una  $t$ -student con  $n_1 + n_2 - 2$  grados de libertad, donde  $n_1$  y  $n_2$  corresponden al tamaño de las muestras respectivas de ambos grupos. El estadístico estará entonces dado por la fórmula:

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}; \text{ con } S_j^2 = \frac{\sum_i (x_{ij} - \hat{X}_j)^2}{n_j - 1} \quad (3.5)$$

En ambos casos el estadístico calculado será contrastado con un valor de comparación que será determinado según el nivel de confianza (usualmente 95 %) que se espera del test de hipótesis planteado.

### 3.5. Indicadores del Negocio

Para poder determinar el impacto de las recomendaciones en el sitio, ya sea este negativo, positivo o nulo, se deben definir un conjunto de métricas que serán comparadas entre grupos para evaluar dicho impacto. En el desarrollo del proyecto, se tomó en general como unidad de medida una “sesión”, que estará definida como una navegación realizada de manera continua por un visitante durante un plazo máximo de 12 horas, o en caso de existir un período de inactividad de más de 30 minutos. Esta definición de sesión suele ser estándar en distintos sistemas de analítica web.

Para entender el efecto de las recomendaciones a implementar y aprovechando la gran cantidad de información disponible, se propone analizar tres conjuntos de métricas que explicarán comportamientos diferentes de los visitantes. Estos son:

1. **Métricas de conversión e ingreso:** permitirán medir de manera global la proporción de visitantes que realizan compras y el monto gastado por ellos, considerando a todas las sesiones registradas dentro del grupo. Esto permitirá medir efectos que serán muy difíciles de medir a través de un análisis exhaustivo. Entre estas métricas están:

- Tasa de conversión (TC):

$$TC = \frac{N \text{ sesiones con compra}}{N \text{ sesiones totales}}$$

- Gasto promedio por sesión (ARPS):

$$ARPS = \frac{\text{Ingreso total del grupo}}{N \text{ sesiones totales}}$$

- Gasto promedio de sesiones con compra(ARPPS):

$$ARPPS = \frac{\text{Ingreso total del grupo}}{N \text{ sesiones con compra}}$$

2. **Métricas de uso:** relativa a la cantidad de visitantes que realizan algún click sobre algún elemento de los carruseles de recomendación en comparación con el total asignado al grupo.

- Tasa de uso:

$$TU = \frac{N \text{ sesiones con click en carrusel}}{N \text{ sesiones totales}}$$

3. **Métricas post-click:** permitirán hacer un seguimiento de los visitantes luego de haber hecho uso del carrusel, para entender los efectos directos que esto tendrá en sus comportamientos, como saber si tras el click se aumentó la propensión a comprar del grupo o si los visitante terminaron comprando el producto de su interés. Entre estas métricas están:

- Tasa de Conversión Post-click:

$$TC_{postclick} = \frac{N \text{ sesiones con compra y click}}{N \text{ sesiones con click en carrusel}}$$

- Tasa de Conversión Post-click (sobre el mismo producto):

$$TC_{postclick}^{mismo SKU} = \frac{N \text{ sesiones que compraron producto clickeado}}{N \text{ sesiones con click en carrusel}}$$

- Tasa de Agregar al Carro:

$$TC_{carro} = \frac{N \text{ de sesiones que agregan click al carro}}{N \text{ sesiones con click en carrusel}}$$

- Gasto en productos clickeados:

$$ARPPS_{postclick}^{mismo SKU} = \frac{\text{Gasto en productos clickeados}}{N \text{ sesiones que compraron producto clickeado}}$$

# Capítulo 4

## Metodología

El proyecto será realizado a través de un proceso iterativo de trabajo basado en la metodología CRISP-DM [11]. Esto quiere decir que el trabajo se realizará en ciclos flexibles en donde se consideran las distintas etapas de entendimiento de la oportunidad de negocio, el entendimiento de los datos, su preparación y limpieza, el modelamiento y la evaluación de los sistemas generados. En particular, la metodología propuesta para este trabajo se compone de tres ciclos principales que consisten en una fase preliminar exploratoria, con foco en el entendimiento del problema y la exploración y limpieza de datos; una fase para la construcción y evaluación de sistemas de recomendación no personalizados, con foco en la construcción de rankings de productos; y una última fase para la construcción y evaluación de sistemas personalizados basados en segmentos, con foco en la agrupación de clientes para ofrecerles las mejores recomendaciones según su perfil. Al finalizar cada ciclo, que se puede observar gráficamente en la Figura 4.1, se propone una evaluación de los sistemas de recomendación generados a través de un experimento controlado en línea, aplicado en el sitio de comercio electrónico, para generar aprendizajes al pasar de un ciclo a otro.

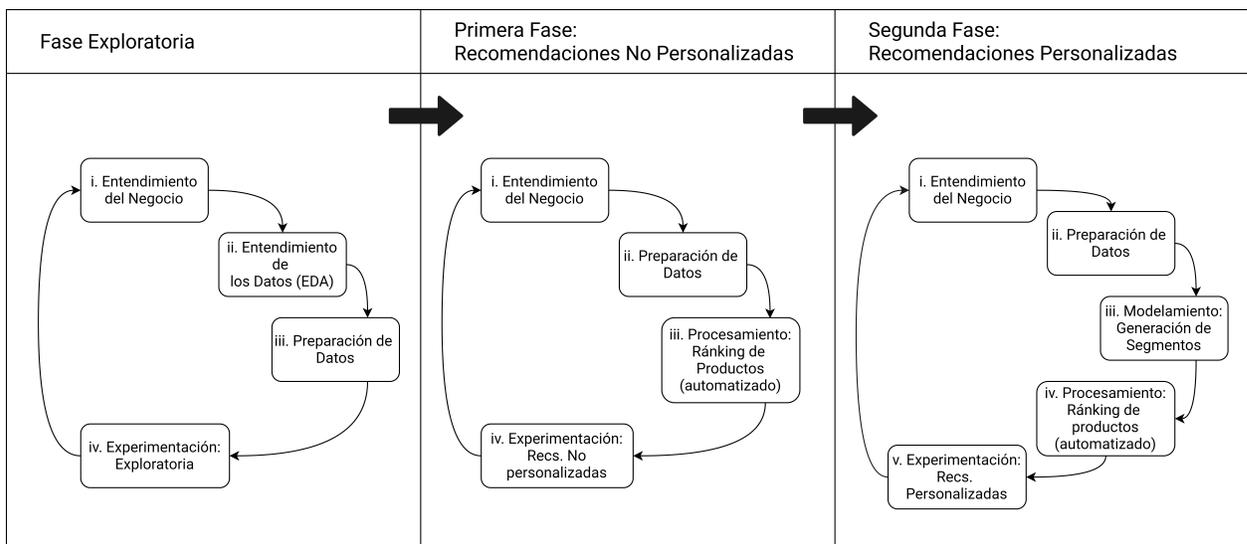


Figura 4.1: Caracterización de las líneas de productos según precio promedio y unidades vendidas.

## 4.1. Fase Exploratoria

La etapa de análisis exploratorio y descriptivo de los datos será relevante para identificar la información que será más o menos importante en el proceso siguiente de formulación de recomendaciones, segmentos y diseño experimental. Para esto se manejarán distintas fuentes de información disponibles en la empresa, que se agrupan a grandes rasgos en cuatro grupos:

1. **Clientes:** sobre los que se incluye información sociodemográfica como es el sexo, la edad, la comuna y región de residencia, y la existencia de hijos. Además se incluyen otras variables relevantes para la empresa como son las que indican si un cliente posee la tarjeta de crédito de la tienda o si se trata de un cliente frecuente y con alto gasto.
2. **Productos:** sobre los que se incluye información del precio, marca, categoría a la que pertenecen, calificación de los usuarios (en algunos casos), si está en oferta o no y otros atributos particulares para cada tipo de producto. Se considera la categorización de los productos en la empresa consiste de cuatro niveles jerárquicos: líneas, sublíneas, clases y subclases.
3. **Compras:** los datos transaccionales de compra incluyen las realizadas tanto en el sitio web como en tiendas físicas. Estos datos entregan información sobre el producto comprado, el monto, la cantidad, la boleta a la que pertenecen, el cliente que compró (en la mayoría de los casos), la fecha, la tienda de compra y el medio de pago. La información está disponible a partir de 2015 para compras en tiendas físicas y a partir de 2016 para compras en el sitio web.
4. **Navegación:** los datos transaccionales de navegación registran las secciones del sitio por las cuales pasan los usuarios cada día. En algunos casos también se tiene información sobre las interacciones de los usuarios con el sitio, como son los clicks sobre distintos elementos de la página. Estos datos permiten hacer seguimiento dentro del sitio a nivel de sesión, aunque en algunos casos permite extender este seguimiento hacia múltiples sesiones o a nivel de usuario en caso de una navegación identificada a través del rut o el mail.

Para lograr descubrir información relevante en los datos se propone analizar patrones y tendencias dentro de cada uno de los cuatro grupos antes mencionados a través de estadística descriptiva y visualizaciones. Ya entendiendo de forma general los datos se procederá a investigar relaciones entre grupos, por ejemplo como se asocian las distintas categorías de productos con los distintos tipos de cliente o las compras con la navegación de los usuarios.

El objetivo final de la etapa de análisis descriptivo será la de determinar aquellas variables que serán utilizadas para construir las recomendaciones de productos y posteriormente definir segmentos, así como definir qué atributos deberán ser creados a partir de los datos originales con los que se cuenta. Complementando lo anterior, se propone esta etapa como un método para entender el comportamiento de datos anómalos, la variabilidad de las métricas y definir la forma de tratar con estos casos.

## 4.2. Primera Fase: Recomendaciones No Personalizadas

Para la primera instancia de generación de recomendaciones, se crearán mecanismos no personalizados, es decir, que sean desplegados a todos los visitantes por igual. Esto será útil cuando los clientes no logran ser individualizados por las herramientas del sitio, lo cual sucede con la mayoría de los visitantes. Para determinar aquellos productos que serán buenos candidatos a ser recomendados (dado que presentarán distintas características atractivas para los clientes) se decide buscar mecanismos que puedan ser totalmente automatizables y estén basados en estadísticas extraídas del sitio. La ventaja de esto es que permite escalar la solución y reducir el esfuerzo necesario para mantenerla.

Con el objetivo de generar reglas automáticas para elegir los productos a recomendar, se consideraron los datos con los que se dispone y una exploración de soluciones implementadas por otras empresas y plataformas, proponiendo dos maneras sencillas de ordenar los productos intentando reflejar el interés general de los consumidores por explorarlos o comprarlos:

- **Productos más comprados:** se propone ordenar los productos según la cantidad de unidades compradas en el sitio durante un intervalo móvil de tiempo, que permita equilibrar la cantidad de información disponible y la actualidad de la misma. En base a la cantidad de datos se propone utilizar una ventana móvil de 15 días con respecto al cálculo del ranking. Además, se considerará el ordenamiento a partir de las unidades vendidas y no el monto total, para poder comparar de mejor forma productos con precios bajos con otros de un valor más elevado.
- **Productos más navegados:** de manera similar a la regla anterior, se propone ordenar los productos según la cantidad de visualizaciones que se registraron en sus páginas de producto (PDP) durante un intervalo móvil de tiempo de 5 días, permitiendo equilibrar cantidad y actualidad de la información. Al ser mayor el volumen de datos de navegación que el de compra, se podrá utilizar una ventana de tiempo más pequeña y reciente.

Otra posibilidad considerada, en base a otras implementaciones de sistemas de recomendación, corresponde a generar rankings de productos en base a la calificación entregada por los usuarios. A pesar de que en general estos sistemas suelen tener buenos resultados pues muestran de manera explícita el grado de satisfacción, en el caso de la empresa la cantidad de artículos que cuentan con esta información son muy reducidos, por lo tanto se descarta esta metodología para buscar otras que permitan abordar un universo mayor de productos.

Además de la elaboración de los dos métodos anteriormente mencionados, la fase exploratoria del trabajo permite identificar que habrán dos grandes categorías de productos que evidencian comportamientos diferentes frente a los consumidores, tanto por el monto comprado, como por el número de transacciones realizadas. Estos grupos serán denominados como categorías “duras” y categorías “blandas”, y en la Tabla 4.1 se ejemplifican los tipos de productos que integrarán cada una. Para aprovechar de mejor forma las diferencias presentes entre ellas, se propone generar los dos rankings antes descritos para cada una de estas agrupaciones, resultando en 4 sistemas de recomendación diferentes.

Tabla 4.1: Distinción entre categorías en base al monto y frecuencia de compra.

Categorías Duras	Categorías Blandas
Electro-Hogar	Vestuario Mujer
Muebles	Vestuario Hombre
Dormitorio	Vestuario Deportivo
Decoración	Accesorios Mujer
	Belleza
	Calzado

### 4.3. Segunda Fase: Recomendaciones Personalizadas usando Segmentos

Dado que un objetivo del proyecto es buscar una mejora de los sistemas de recomendación generales incluyendo un grado de personalización, permitiendo aprovechar los datos recopilados sobre los clientes, se incorporará un perfilamiento de clientes a través de segmentos, con el objetivo de generar un número reducido y abordable de recomendaciones diferentes. En este sentido, la segmentación buscará utilizar los datos explícitos que se tienen sobre los clientes para agrupar perfiles similares, siendo así objetivos de una misma recomendación.

Para la generación de los segmentos se dispone de información sobre el comportamiento de compra y de navegación de los clientes, sobre las distintas categorías de productos que comercializa la empresa. Estas categorías agrupan productos similares (en características, funcionalidad, etc) a distintos niveles de agregación. Evidentemente mientras mayor sea el nivel de agregación menos precisas podrán ser las recomendaciones, y al mismo tiempo mientras menor sea el nivel de agregación se podrá crear una mayor particularidad de segmentos.

Dado que se requiere generar un número reducido de grupos de gran tamaño, se plantea como primera iteración una segmentación de clientes en base a la categoría más agregada que maneja la empresa: línea. A manera de simplificar el análisis, se propone utilizar en primera instancia sólo la información de compra de los clientes para realizar la clasificación. Por lo tanto la primera clasificación será en base a la frecuencia y el monto comprado por línea lo que ya permitirá determinar preferencias de los clientes.

En una segunda iteración se propone ampliar el análisis considerando además la información de navegación de los clientes, medida a través del número de visualizaciones sobre las distintas categorías, que permitirá utilizar información más reciente y masiva acerca de los usuarios del sitio. Una vez concluidas ambas segmentaciones, se propone utilizar un conjunto de reglas de asignación que consideren los resultados anteriores para decidir cual será la categoría más indicada para recomendar a cada uno de los clientes.

Para la creación de los segmentos se propone utilizar la técnica de K-means, que se basa en encontrar centroides que describan un número K de segmentos, de manera tal que se minimice la distancia total de los puntos (clientes) a su centroide respectivo. El número de segmentos será determinado en primera instancia según la ganancia de información que represente el

agregar un segmento más a la clasificación, aunque también se propone considerar el juicio experto para decidir el número de segmentos en cuanto a la interpretabilidad y funcionalidad de los grupos generados.

Una vez determinados los segmentos de cada iteración, se generarán recomendaciones en base a los productos que mejor representen a las categorías con mayor afinidad por parte de cada segmento. Estos productos podrán ser elegidos mediante un puntaje u ordenamiento en base a compras y navegación, similar a la generación de recomendaciones no personalizadas, pero esta vez acotado a una categoría o grupo de categorías específicas.

## 4.4. Implementación de Experimentos

Una vez generadas las recomendaciones para cada iteración (no personalizadas y personalizadas) se propone probar su efectividad mediante diseños experimentales controlados. La mejor forma de hacer esto es a través de pruebas A/B/n. Esto quiere decir que una población objetivo será separada en entre 2 hasta n grupos o variaciones diferentes, incluyendo un control que será el grupo “por defecto” a vencer, y en donde la única diferencia entre los grupos será la variación sobre la cual se quiere realizar la prueba. La idea es poder controlar por la mayor cantidad de factores posibles para que las diferencias encontradas entre grupos puedan ser atribuibles exclusivamente a la variación asignada a cada uno. En el caso de este proyecto, una variación será un tipo de sistema de recomendación, que se le implementará a cada grupo en su experiencia de navegación en el sitio.

Para lograr lo anterior, los grupos serán poblados con muestras aleatorias del universo total de usuarios. Realizando esta aleatorización para un número suficientemente grande de datos, se puede afirmar con mayor seguridad que cualquier otra característica observable y no observable de los usuarios se asignará de manera uniforme a lo largo de todos los grupos, asegurando que la única diferencia entre un grupo y otro sea la variación o el tratamiento que se está intentando medir. Con el objetivo de comprobar o refutar las hipótesis en las distintas fases del proyecto, y así encontrar lógicas de recomendación que sean factibles de implementar, se detallan a continuación los experimentos diseñados.

### 4.4.1. Diseño Experimental para recomendaciones no personalizadas

Con el objetivo de determinar la eficacia de un sistema de recomendación de forma general, se plantea el primer diseño experimental. Como se pudo ver en parte en el análisis descriptivo, el comportamiento de compra de los clientes hacia ciertas categorías de productos puede ser diferente por un tema de precio o por la naturaleza propia de los productos. Es por esta razón que se decidió para esta primera instancia probar un sistema de recomendación que separe productos pertenecientes a categorías duras de productos pertenecientes a categorías blandas. Es así como se definen las siguientes variaciones de recomendaciones a probar:

- **Categorías Duras:** categorías que tengan un alto precio promedio y estén compuestos por productos que los clientes compran con baja frecuencia.

- Grupo de Control: grupo de usuarios al que no se le inyectará ninguna recomendación de productos duros.
  - Más Comprados (duros): los primeros 28 productos más comprados en las categorías duras en los últimos 15 días.
  - Más Navegados (duros): los primeros 28 productos más navegados en las categorías duras en los últimos 5 días.
- **Categorías Blandas:** categorías compuestas por productos con un bajo precio promedio y que los clientes compren con frecuencia media y alta.
- Grupo de Control: grupo de usuarios al que no se le inyectará ninguna recomendación de productos blandos.
  - Más Comprados (blandos): los primeros 28 productos más comprados en las categorías blandas en los últimos 15 días.
  - Más Navegados (blandos): los primeros 28 productos más navegados en las categorías blandas en los últimos 5 días.

La principal hipótesis a probar es que la incorporación de alguno de estos sistemas de recomendación generará incrementales de conversión y venta, con respecto a la ausencia de carruseles de recomendación. Además, se propone que una mayor diversificación de recomendaciones (al incluir ambas agrupaciones de productos) generará mayor incrementales. Por último, se propone que la metodología de recomendaciones basadas en lo más comprado aumentará más la conversión que la metodología basada en lo más visto.

Para probar estas hipótesis se decide generar grupos con las interacciones de las variaciones antes mencionadas, para cada uno de los carruseles de productos, con el objetivo de poder dar versatilidad al análisis realizado posteriormente con un diseño más fraccionado. La nomenclatura de los grupos para este primer experimento se muestran en la Tabla 4.2. Con esta configuración se obtienen 9 grupos experimentales.

Tabla 4.2: Nomenclatura de grupos para primera experimentación.

		Recomendaciones Blandas		
		Control (O)	Más Comprados (C)	Más Navegados (N)
Recomendaciones Duras	Control (O)	OO	OC	ON
	Más Comprados (C)	CO	CC	CN
	Más Navegados (N)	NO	NC	NN

Para estimar el tamaño de la muestra que será necesario alcanzar, se utilizará un análisis de poder estadístico a priori. Para esto, se generó la Tabla 4.3 que indica el tamaño que debe tener cada grupo para un test de proporciones de una cola. Esta tabla asume un nivel de significancia  $\alpha = 0,05$ , un poder estadístico  $(1 - \beta) = 0,8$  y tamaños iguales entre grupos experimentales.

Tabla 4.3: Tamaño muestral para alcanzar poder estadístico a priori en base a una estimación del valor de la métrica y el lift a encontrar.

Lift	Valor actual de la métrica									
	2,0%	2,5%	3,0%	3,5%	4,0%	4,5%	5,0%	5,5%	6,0%	6,5%
1%	6.088.579	4.845.885	4.017.422	3.425.663	2.981.844	2.636.651	2.360.497	2.134.552	1.946.265	1.786.946
2%	1.529.561	1.217.342	1.009.196	860.520	749.013	662.286	592.904	536.136	488.831	448.802
3%	683.100	543.649	450.682	384.277	334.473	295.736	264.747	239.392	218.264	200.385
4%	386.097	307.269	254.718	217.181	189.028	167.131	149.614	135.282	123.338	113.232
5%	248.288	197.591	163.793	139.651	121.545	107.463	96.197	86.979	79.298	72.798
6%	173.245	137.867	114.282	97.435	84.800	74.973	67.111	60.679	55.319	50.783
7%	127.887	101.769	84.357	71.920	62.592	55.337	49.533	44.784	40.826	37.478
8%	98.376	78.283	64.887	55.319	48.143	42.562	38.096	34.443	31.399	28.823
9%	78.095	62.143	51.507	43.911	38.214	33.782	30.237	27.337	24.920	22.875
10%	63.553	50.570	41.914	35.732	31.095	27.488	24.603	22.242	20.275	18.611

Asumiendo un tasa de conversión cercana al 3,5%, para evidenciar un lift cercano al 2% se necesitarán 860.520 muestras por grupo. Es claro que la tasa de conversión será la métrica más difícil de mover, y en general será suficiente con determinar un aumento en la tasa de uso (que en teoría de debería ser más fácil de mover y describirá el interés de los visitantes en las recomendaciones), pero las métricas de uso no pueden ser comparadas con el grupo control, dado que no existirá ningún elemento sobre el cual contrastar. Dada la estimación anterior, para completar los 9 grupos se necesitarán alrededor de 7.7 millones de sesiones. Según los datos de navegación históricos, esta cantidad de sesiones podrá alcanzarse en cerca de 17 días, por lo que será el tiempo de experimentación propuesto.

Una vez finalizado el tiempo de experimentación, se procederá a analizar el conjunto de métricas detallados en la metodología entre los grupos experimentales. Para eso, se realizarán las siguientes comparaciones entre grupos:

- OO <-> CC <-> NN <-> NC <-> CN: permitirá conocer en primer lugar, si los sistemas de recomendación lograron métricas superiores que ante la ausencia de una recomendación. Al mismo tiempo, permitirá probar entre reglas cuál de todos los sistemas (de forma general) fue el más efectivo.
- OO <-> CO <-> NO: permitirá saber que regla fue más efectiva para generar recomendaciones dentro de las categorías duras.
- OO <-> OC <-> ON: permitirá saber que regla fue más efectiva para generar recomendaciones dentro de las categorías blandas.
- OO <-> CO <-> OC: permitirá saber cuál de los dos sistemas se desempeñó mejor (duros o blandos) al ser creado con una regla en base a compras.
- OO <-> NO <-> ON: permitirá saber cuál de los dos sistemas se desempeñó mejor (duros o blandos) al ser creado con una regla en base a navegación.

#### 4.4.2. Diseño Experimental para recomendaciones personalizadas

Ya con los resultados de la primera fase de experimentación, en esta segunda fase se plantea poder determinar si el método de segmentación utilizado para agregar un grado de

personalización a los sistemas de recomendación obtiene mejores resultados que con sistemas de recomendación no personalizados. El objetivo entonces será lograr dividir a los clientes en un número  $K$  (determinado en la segmentación anterior) de grupos en base a sus compras y navegación por líneas de producto y elegir las mejores recomendaciones para cada una de estas asociaciones de clientes. Para probar empíricamente el sistema personalizado, cada segmento generado será dividido en tres grupos:

- Grupo de Control: grupo de usuarios pertenecientes al segmento  $K$  que no se les inyectará ninguna recomendación.
- Mejor recomendador no personalizado: en base a los resultados de la etapa previa, seleccionar un recomendador no personalizado para mostrar a una fracción de clientes del grupo  $K$ .
- Recomendación para segmento  $K$ : en base a las líneas que definan de mejor forma al segmento  $K$ , realizar recomendaciones de productos dentro de esas líneas en base a reglas de compra y navegación.

Una vez que los segmentos sean formados, los visitantes serán aleatorizados de manera uniforme en estos tres grupos. En este caso el número posible de muestra por intervalo de tiempo se reduce, dado que sólo un tercio de las visitas podrán ser identificadas y, por lo tanto, asignadas a un segmento en específico. Además, se espera que la segmentación arroje grupos que no tengan exactamente el mismo tamaño. Se debe considerar que la empresa ha puesto como plazo máximo de experimentación 21 días. En este sentido, el número esperado de sesiones que se espera registrar será de cerca de 3 millones, lo que significaría alrededor de 1 millón de unidades experimentales por grupo. Volviendo a la Tabla 4.3, se observa que con esta cantidad de muestra (y asumiendo una tasa cercana al 4%) el efecto mínimo a encontrar se encontrará entre un 1 y 2%, pero se estima que la diferencia entre las tasas de la conversión de los grupos no será tan grande.

Considerando lo anterior, se tomará como métrica de comparación principal la tasa de uso de los carruseles de recomendación implementados, ya que en este caso se espera que la diferencia pueda ser mucho mayor. De esta forma, al realizar una experimentación durante 21 días, se podrán tener resultados que permitan concluir si las recomendaciones personalizadas generadas a través del método de segmentación propuesto efectivamente capturarán un mayor interés por parte de los visitantes.

Por último, se debe considerar que dada la gran cantidad de segmentos a analizar (un número cercano a 20), si se analizan los resultados realizando esta desagregación, el número de unidades experimentales por grupo podrá ser muy reducido, provocando que los resultados no sean significativos en muchos casos. Por lo tanto este análisis será tomado en consideración sólo a modo descriptivo y en caso de que entregue información relevante acerca del funcionamiento del sistema.

# Capítulo 5

## Desarrollo Metodológico

### 5.1. Fase Exploratoria

De manera previa al desarrollo e implementación de los sistemas de recomendación de productos y de las pruebas en línea, se prepararon y analizaron las fuentes de datos disponibles con el objetivo de entender patrones en los datos a tener en consideración para las fases posteriores, así como para encontrar posibles fuentes de errores a ser prevenidas durante las fases de experimentación. Es así como en esta fase se trabajó en paralelo tanto en el procesamiento de datos históricos, como en el análisis de los resultados arrojados tanto por la herramienta de experimentación-personalización como por la herramienta de analítica de la empresa. En las siguientes secciones se describen algunos de los hallazgos mas relevantes para el desarrollo del proyecto.

#### 5.1.1. Descripción de productos y categorías

Una de las primeras consideraciones a tener es en lo relativo al número de productos y la categorización que ellos tienen dentro de la empresa. Al momento de la realización de este proyecto, el sitio web de la empresa contaba con alrededor de 350 mil productos publicados considerando sus distintas variaciones (colores, tallas, etc.). Este número se reduce a aproximadamente 170 mil al no considerar estas variaciones, pero ciertamente sigue siendo una gran cantidad de productos, además de consistir en una gran cantidad de categorías diferentes.

En específico, la empresa ordena sus productos en 4 niveles de categorías que tienen un sentido jerárquico. Estas se muestran en la Tabla 5.1, indicando además ejemplos para cada uno de estos niveles y el número de categorías para cada uno de ellos. En principio todos los niveles tienen un sentido funcional que permite separar productos muy distintos entre si y agrupar a los similares. A pesar de que al bajar desde el nivel “Línea” al nivel “Clase” los productos son mucho más parecidos entre si (en características, utilidad, precio, etc.), también el número de categorías se vuelve más numerosa. Esto se debe tener en cuenta al momento de generar agrupaciones, pues incrementar el nivel de detalle podría mejorar los resultados de la personalización al ser más específico en el tipo de producto, pero aumentaría la complejidad de generación e interpretación de estos grupos.

Tabla 5.1: Agrupaciones jerárquicas de productos en la empresa.

Nivel	Ejemplo	Cantidad
Línea	Dormitorio - Moda Mujer - Electrohogar	20
Sub Línea	Colchones - Bottoms - Telefonía	150
Clase	Colch. King - Jeans - Prepagos y Libres	783
Sub Clase	Con Figuras - Jeans Básicos - Libres	2.402

También se debe tener en consideración acerca de los productos que habrán otras diferencias que no se presentan de forma explícita en la categorización. Una de estas características es el precio promedio de los productos en cada categoría, donde es claro que habrán grupos que presentarán montos mas de diez veces mayores que las categorías de precios más reducidos, lo que determinará cierto comportamiento en las compras de los productos. En general, los productos que pertenecen a categorías con precio promedio reducido serán comprados en mayores cantidades y con mayor frecuencia (por un mismo cliente) que los productos pertenecientes a categorías más caras. Esta predisposición y frecuencia de compra tendrá un impacto en la probabilidad de conversión de los visitantes del sitio ante las distintas categorías. Estas características permiten hacer una distinción entre categorías “Duras”, de un precio más elevado y en general con una menor tasa de conversión debido a que se trata de productos que cada cliente comprará una sola vez en varios años (por ejemplo: televisores, lavadoras o camas); y categorías “Blandas”, que tendrán un precio menor y serán adquiridos con mayor facilidad y frecuencia por parte de los usuarios (por ejemplo: vestuario, calzado).

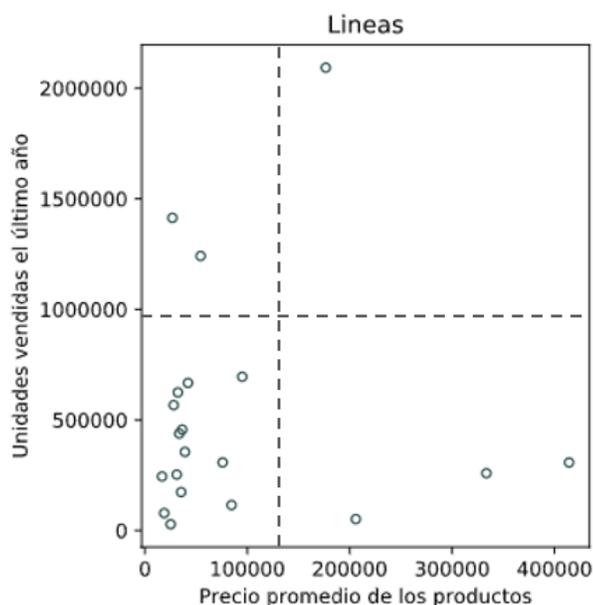


Figura 5.1: Caracterización de las líneas de productos según precio promedio y unidades vendidas.

En la Figura 5.1 se observa esta diferenciación al nivel de agrupación de Líneas. El precio promedio de esta categorización de productos se complementa con la cantidad de unidades

vendidas en el último año. El gráfico muestra en los cuadrantes de la izquierda aquellas categorías que pueden ser consideradas como blandas y a la derecha aquellas que pueden ser consideradas como duras. En general el mayor número de unidades vendidas se obtiene en las categorías blandas, aunque el gráfico también sugiere tres categorías que se escapan de las tendencias del resto, indicando que son categorías muy compradas más allá de su precio. Lo anterior muestra que aunque la caracterización entre duras y blandas es correcta en general, habrán categorías particulares que serán muy populares (o poco populares) más allá de su precio. El mismo análisis se replica para el resto de niveles en la Figura B.2, donde en el nivel de “clase” se puede observar como la acumulación de puntos hacia la izquierda muestra que en general se mantiene la tendencia sobre que las categorías de menor valor suelen ser compradas con mayor frecuencia, aunque no son tan pocos los casos en que esa regla no se cumple del todo.

Por lo tanto, más allá de que una distinción entre categorías duras y blandas será conveniente (dadas sus diferencias en características y en el tipo de cliente que muestra interés en ellas), no se debe dejar de considerar que habrán categorías que serán más importantes para los ingresos de la empresa, definido por la relación entre un alto precio y un gran número de unidades vendidas. Las categorías que cumplirán con esta relación estarán definidas por los objetivos de compra de los visitantes que suelen utilizar el sitio web.

### 5.1.2. Descripción de clientes

La base de clientes de la empresa es extensa y diversa. Si se añaden las compras realizadas en sus tiendas físicas a las realizadas en su sitio web, se contabilizan más de 5.5 millones de clientes activos (con compras en el último año). La distribución del sexo y la edad de ellos se puede observar gráficamente en la Figura B.1 en anexos, donde se puede apreciar la heterogeneidad de los mismos. A pesar de que los clientes de la empresa suelen ser tanto hombres como mujeres, se puede observar en el gráfico que un mayor porcentaje de ellos corresponden a mujeres. Con respecto a la edad, si bien se tienen clientes para todos los rangos, la mayoría suele concentrarse cerca de los 30 años, que también es el rango de edad en donde mayor es evidente al aporte de las compras online, dando cuenta de la mayor penetración que este canal tiene en los clientes más jóvenes.

Además del sexo y su edad, se tiene que los clientes de la empresa suelen provenir de distintos grupos socioeconómicos, evaluados según su nivel de gasto y lugar de residencia. De los clientes que realizan su compra por el sitio web de la empresa más del 59% de ellos tienen hijos y más del 57% de ellos cuentan con la tarjeta de crédito de la marca. Todas estas características describen la heterogeneidad de los clientes y sus necesidades, que la empresa debe ser capaz de atender.

Considerando sólo las ventas realizadas mediante el sitio web, se contabilizan alrededor de 1.4 millones de clientes por año. En base a las compras realizadas durante el último año, se puede observar que a través de este canal existen ciertas tendencias sobre las categorías más compradas por los usuarios. Considerando una categorización a nivel de línea, en la Tabla 5.2 se muestra el número de clientes que realizaron alguna de sus compras en cada categoría. Esta estadística muestra que un 67% de los clientes que realizaron alguna compra en el último

año, lo hicieron dentro de la categoría denominada como “J11”, distanciándose bastante de la que se encuentra en segundo lugar (“J10” con un 44%). Una forma de interpretar esto es que ambas categorías (sobre todo la primera) son categorías importantes y de interés para la mayoría de los clientes. Así mismo, habrán categorías en que el interés estará dado por un grupo muy específico. A pesar de lo anterior, las demás categorías suelen tener un interés distribuido de forma relativamente pareja, indicando que existe un universo de clientes para cada categoría. Además, el hecho de que una persona no haya comprado un producto en cierta línea no significa necesariamente que dicha línea no sea de su interés, puesto que puede haberse visto negado a comprar por las características propias de los productos de esa categoría, porque la línea no está bien visibilizada en el sitio, porque no hay buenas ofertas u otra razón.

Tabla 5.2: Porcentaje de clientes del total que realizan alguna compra en cada categoría.

<b>Linea</b>	<b>% de clientes con compra</b>
J01	18 %
J02	16 %
J03	36 %
J04	23 %
J05	22 %
J06	12 %
J07	22 %
J08	30 %
J09	38 %
J10	44 %
J11	67 %
J12	24 %
J13	26 %
J14	28 %
J15	26 %
J16	17 %
J17	16 %
J18	8 %
J21	12 %
J32	5 %

Esta hipótesis acerca de la importancia de ciertos productos para los clientes (y por lo tanto para la empresa) se complementa al analizar el ingreso absoluto aportado por cada categoría. Por ejemplo, al analizar el ingreso aportado a nivel de “clase” (que se observa gráficamente en la Figura B.3 en anexos), se observa como un gran porcentaje del ingreso total fue realizado por un número reducido de categorías, cumpliéndose el principio de Pareto. En particular, se tiene que alrededor del 80% del ingreso del último año a través del canal online fue aportado por el 9% de las categorías con más ventas, recordando que el tamaño del aporte está influenciado tanto por el valor promedio de los productos como por el nivel de interés que los visitantes tienen de adquirirlos.

Otra comportamiento interesante a analizar acerca de los clientes corresponde a deter-

minar cuántos de ellos suelen ser compradores de una única categoría, y cuántos de ellos resultan ser multi-categoría. La Figura 5.2 muestra la cantidad de clientes por número de categorías compradas utilizando el nivel de línea. El gráfico muestra que alrededor del 33 % de los clientes del último año realizaron sus compras en una única línea, aunque esto también está relacionado con el número de clientes que realizaron una única compra (en especial, de un único producto). A pesar de que el gráfico muestra que la mayoría de los clientes compra una única línea, no es menor el porcentaje de clientes que suele comprar 3 o más líneas diferentes, indicando un gran número de clientes multi-categoría. De todas formas, la mayoría de los clientes comprarán un número limitado de categorías que mejor representarán sus intereses.

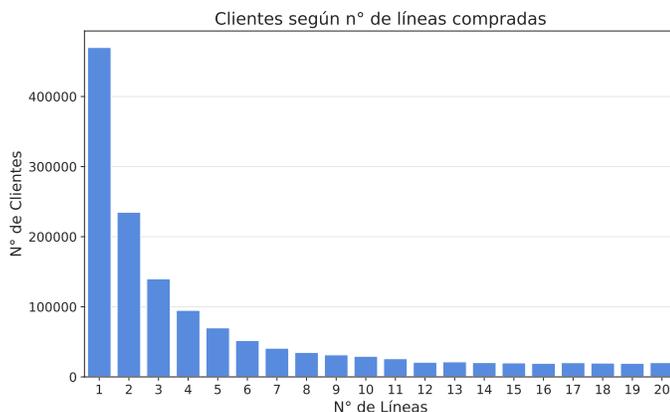


Figura 5.2: Distribución de frecuencia del número de líneas distintas compradas por cliente.

### 5.1.3. Descripción de navegación

Antes de comenzar con el diseño de las recomendaciones de productos, también será importante considerar el comportamiento de navegación dentro del sitio, específicamente dentro de la página de inicio que será el lugar donde se implementarán los carruseles de recomendación. En particular, se analizaron las métricas registradas por la herramienta de analítica de la empresa durante el mes de abril del 2020, previo a la implementación de los experimentos. Durante este mes se registraron alrededor de 10,8 millones de sesiones en la página de inicio, de las cuales cerca del 38 % fueron realizadas a través de la versión para dispositivos móviles del sitio, y un 62 % se realizaron a través de la versión para PC y Mac. Además, de este total de sesiones registradas el 34 % correspondían a sesiones de usuarios identificados (por rut o mail). Estas últimas sesiones fueron realizadas por alrededor de 900 mil clientes que pudieron ser individualizados. El número de usuarios que corresponden al 66 % de sesiones no identificadas no podrá determinarse con exactitud, pero se estima que serán al menos 1.7 millones.

Será importante señalar que habrá una diferencia importante entre las sesiones identificadas y no identificadas en cuanto a la probabilidad de conversión y el gasto promedio realizado. En el caso de las sesiones identificadas que pasaron por la página de inicio, el 6,5 % de ellas terminaron realizando alguna compra, mientras que en el caso de las sesiones no identificadas las que terminaron realizando una compra fueron el 2,1 %. Además el gasto promedio de las sesiones identificadas que compraron fue de \$109.117, mientras que las no identificadas que

compraron gastaron en promedio \$99.014. Dada esta mayor probabilidad de conversión y mayor gasto, las sesiones identificadas aportarán más de tres veces lo que las no identificadas al ingreso, por lo que será recomendable poner especial atención a ellas.

Otro análisis importante a destacar es hacia qué secciones de la página de inicio suelen dirigirse los visitantes. Durante el período en que se realizó el trabajo, la empresa no contaba con una herramienta funcional para conocer el porcentaje de usuarios que visualiza cada sección de sitio, pero una aproximación similar se puede hacer con el número de clicks registrados por la herramienta de analítica en la página de inicio. En la Figura 5.3 se presenta un esquema del sitio en su versión de escritorio, con una aproximación del tamaño y la posición de sus secciones para ejemplificar la atención que recibe cada elemento. Los porcentajes mostrados corresponden al número de sesiones que hicieron click en dicha sección, del total de sesiones registradas que pasaron por la página. Se debe considerar que existirán secciones de la página que no son registradas (por errores no solucionados con la herramienta de analítica).

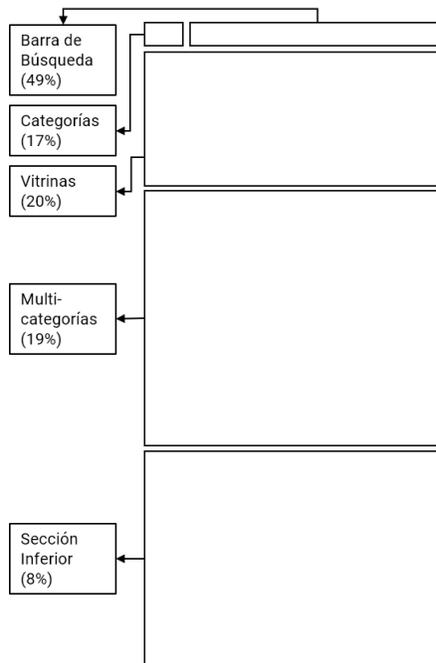


Figura 5.3: Esquema de las secciones de la página de inicio y su tasa de utilización aproximada.

La figura anterior muestra como un gran porcentaje de los visitantes hacen uso principalmente de la barra de búsqueda, mientras que sólo el 8 % utiliza algún elemento de la sección inferior del sitio, que será el lugar en donde serán ubicados los carruseles de recomendación. En análisis de los datos muestra que de las sesiones que pasan por la página de inicio del sitio sólo un 30 % de ellas retorna a la página de inicio para realizar un nuevo click. Independiente de lo anterior, sólo el 13 % de las sesiones totales retorna a la página de inicio para utilizar un elemento diferente del sitio, pudiendo o no estar dentro de la misma sección bosquejada en la figura superior. En resumen, la mayoría de las sesiones no suele volver atrás ni utilizar más de una sección del sitio.

Será imposible saber el número exacto de sesiones que visualizarán la sección inferior del sitio, y aunque sólo el 8% de los visitantes haga uso de algún elemento en este espacio no se debe dejar de considerar que habrán ciertos visitantes que lo visualizarán sin hacer click en él. Por lo tanto, aunque no es posible saber cuántos (y sobre todo cuales) visitantes serán realmente expuestos a las recomendaciones, sí se debe considerar que sólo una fracción del total de ellos sabrán de su existencia.

#### 5.1.4. Experimentación exploratoria

Para finalizar la fase exploratoria, se aprovecharon los resultados generados a raíz de una primera experimentación piloto realizada por el área de la empresa, con relación a los carruseles de recomendación. En este período de experimentación se implementó un carrusel de recomendaciones de forma manual, integrado por una selección de productos con una buena calificación por parte de los clientes. A pesar de que parte de los resultados no son concluyentes, sí se pueden analizar algunas métricas como la distribución de los clicks.

Una de las particularidades del carrusel al que se hace referencia en esta fase exploratoria es que el orden de sus productos se generaba de manera aleatoria, por lo que no existe ninguna relación entre esto y la calidad de la recomendación. Aún así, como se observa en la Figura 5.4, una gran cantidad de clicks se concentran en las primeras dos y seis posiciones, que corresponden al número de artículos mostrados en una primera vista del carrusel tanto en dispositivos móviles como de escritorio, respectivamente. Aunque lo anterior no indica necesariamente que los elementos efectivamente visualizados se encuentran en la misma proporción que los clicks, sí sugiere que la importancia de las primeras posiciones en el efecto de los carruseles de recomendación es muy alta y definirá en gran parte la percepción que los visitantes tienen del sistema.

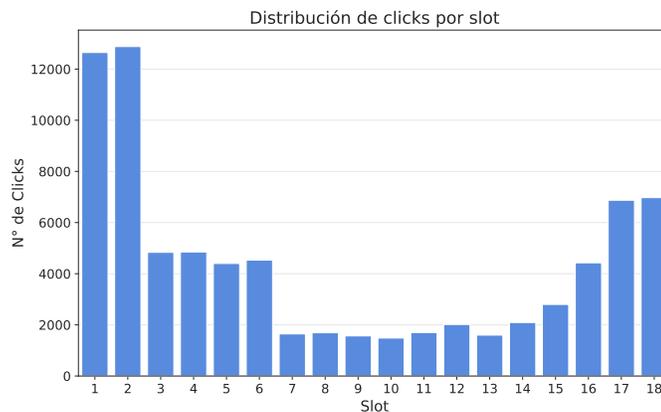


Figura 5.4: Distribución de clicks por slot en carrusel de fase exploratoria.

Además de este análisis de uso del carrusel, se aprovechó esta instancia para entender la variabilidad de las métricas que serían utilizadas en las fases de experimentación posteriores. En particular, se analizó el gasto promedio por sesión (ARPS) dentro de la página de inicio. Para esto se simuló en varias instancias un muestreo aleatorio de dos grupos (con 50% de los

datos cada uno) entre las sesiones del mes. Ambos grupos no debieran representar ninguna diferencia real, más allá de la variación normal de los datos. Los resultados de la simulación muestran que al armar dos grupos de 2.5 millones de sesiones cada uno, el 95 % de las veces se obtuvo una diferencia entre los ARPS promedio de \$103 o más, mientras que al utilizar 1.2 millones por grupo este valor subió a \$131 y al utilizar sólo 600 mil por grupo subió a \$186. Lo que quiere decir esto es que con tamaños muestrales muy pequeños, las diferencias “normales” que pueden ocurrir entre los ARPS de los grupos son muy altas, dada la gran variabilidad de los montos de compra de los visitantes. Es por esto que para asegurar que una diferencia evidenciada de alrededor de \$100 pueda ser concluida con significancia estadística, se necesitarán cerca de 2.5 millones de muestras. Esto dificulta mucho el poder concluir sobre diferencias menores en esta métrica (por ejemplo, de \$20), aún cuando esta diferencia, de existir, tendría una significancia práctica dado el gran volumen de sesiones que se registran cada semana en el sitio. Esto debe tenerse en consideración al momento de planificar los plazos de futuras experimentaciones.

## **5.2. Primera Fase: Recomendaciones No Personalizadas**

Concluida la fase exploratoria, se pudo observar que una gran cantidad de visitantes del sitio navegan sin poder ser identificados por parte de la empresa. Por lo tanto, las primeras recomendaciones a desarrollar deberán ser implementables tratando a todos los visitantes como iguales. Para cumplir este objetivo, se detalla a continuación el proceso automatizado de generación de los artículos a recomendar en base a compra y navegación, tanto para categorías duras como para categorías blandas, concluyendo con los resultados del período de experimentación ejecutado.

### **5.2.1. Generación de recomendaciones**

Para la generación de productos recomendados para categorías duras y blandas se implementó un proceso automatizado de extracción, transformación y carga de datos utilizando las bases de datos de la empresa, las plataformas de procesamiento en la nube con las que se trabaja y la plataforma de experimentación-personalización que se utiliza para realizar cambios modulares dentro del sitio. El proceso fue automatizado para que todos los días y cada dos horas se actualizaran los rankings de los productos más comprados y más vistos, generados a partir de las bases de datos actualizadas de ventas y de navegación respectivamente. Así la plataforma de experimentación-personalización se encargó cada vez de renderizar en el sitio un carrusel de productos con las recomendaciones generadas. Una vista de las recomendaciones desplegadas en la versión de escritorio en un día específico se muestra en la Figura 5.5 para ejemplificar los elementos vistos por los visitantes.

Para la generación de estos ranking se etiquetaron todos los productos según las categorías del negocio en las macro categorías de duras y blandas, utilizando los resultados obtenidos en la fase exploratoria y las recomendaciones comerciales de la empresa. Se debe mencionar que

todos los productos listados y publicados al momento de la generación de los ranking fueron considerados. Además, para contabilizar las unidades compradas y el número de visitas a la página de producto, se consideraron las distintas variaciones del producto (color, talla, etc.) como un único producto, identificado a través de su SKU padre.

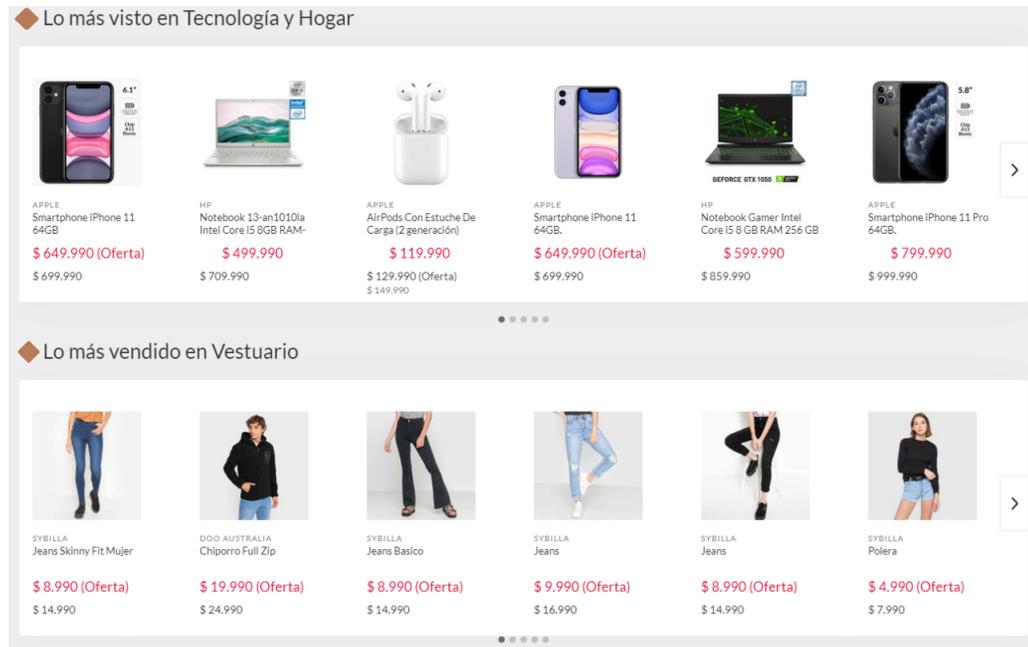


Figura 5.5: Ejemplo de recomendaciones no personalizadas desplegadas en el sitio.

Una vez comenzado el período de experimentación, las recomendaciones de productos se comenzaron a generar automáticamente y durante este período se almacenó la información de los ranking de productos generados para categorías blandas y categorías duras, en su versión de los más comprados y los más vistos. Un análisis de esta información muestra las diferentes características de los cuatro rankings generados. En la Tabla 5.3 se puede observar la cantidad de productos únicos que se incluyeron en cada versión de la recomendación, considerando todos los productos mostrados (28 por cada iteración) y los mostrados en las primeras 6 posiciones, que como ya se pudo observar en la fase exploratoria son los que más influencia tienen en el uso del carrusel de recomendaciones.

Tabla 5.3: Productos únicos incluidos en cada versión de las recomendaciones no personalizadas.

Ranking	Duras		Blandas	
	Más Comprados	Más Vistos	Más Comprados	Más Vistos
En el Top 28	58	103	70	115
En el Top 6	15	31	16	29

La Tabla 5.3 muestra que al realizar el ranking en base a los productos más vistos del sitio, se mostraron entre un 64-107 % más recomendaciones que generándolo a través de los

productos con mayor cantidad de unidades compradas. Esto se debe a que utilizando información sobre el número de visitas se puede utilizar un intervalo de días menor gracias a la mayor disponibilidad de datos. Así los ranking generados en base a visitas se ven más afectados por las tendencias recientes. En cambio, al necesitar más días para recopilar una cantidad de información similar en ventas, los ranking en base a compras varían menos entre una fecha y la otra, y las recomendaciones se vuelven más estáticas. Evidentemente los resultados serán diferentes si se utiliza un intervalo de días recientes más pequeño, pero se debe tener en cuenta que los datos de navegación tienen un volumen alrededor de 20 veces mayor que los de compra.

También será importante caracterizar los productos que fueron incluidos en las recomendaciones en el período de implementación experimental. En la Tabla 5.4 se observa el precio promedio, mínimo y máximo de los productos recomendados para las distintas categorías. Estos datos muestran como las recomendaciones en categorías blandas desplegaron productos con un precio promedio cerca de quince veces menor que los desplegados en categorías duras. Además, al comparar los productos más comprados con los más vistos dentro de ambas categorías, se tiene que los primeros poseen un precio promedio inferior. Esto muestra la relación inversa que tiene el precio de los productos con la probabilidad que tiene el visitante de convertir. Lo mismo se puede decir de los productos que estuvieron entre los 6 más comprados o navegados, donde los precios promedios son en general menores. A pesar de que habrán otros factores que determinarán la probabilidad de compra de un visitante, sí es cierto que entre dos productos que satisfacen similares requerimientos el general de los clientes terminará comprando el más barato.

Tabla 5.4: Estadística de precios de las recomendaciones no personalizadas generadas.

Ranking	Precio	Duras		Blandas	
		Más Comprados	Más Vistos	Más Comprados	Más Vistos
En el Top 28	Mínimo	\$ 12.167	\$ 16.990	\$ 2.502	\$ 4.990
	Promedio	\$ 299.150	\$ 453.504	\$ 19.894	\$ 32.353
	Máximo	\$ 1.299.990	\$ 1.999.990	\$ 98.742	\$ 380.000
En el Top 6	Mínimo	\$ 15.353	\$ 79.990	\$ 4.942	\$ 6.990
	Promedio	\$ 246.849	\$ 532.322	\$ 12.026	\$ 25.109
	Máximo	\$ 916.056	\$ 1.149.990	\$ 67.256	\$ 64.990

Una conclusión similar se puede obtener al revisar las tasas de conversión promedio de los productos recomendados mostrados en la Tabla 5.5. Estas tasas de conversión fueron calculadas en el intervalo de tres semanas previas al proceso de experimentación, y se definen como el porcentaje de sesiones que compraron uno de los productos recomendados, dividido por el número total de sesiones que visitaron las páginas de producto de alguna de las recomendaciones. La Tabla 5.5 muestra que los productos de categorías blandas poseen una tasa de conversión mayor que los productos recomendados en categorías duras. Además, como era de esperarse, la tasa de conversión de los productos más comprados es mayor que la de los productos más vistos, en ambas categorías. Mientras que los productos más comprados tendrán características que los hacen muy adquiridos por los clientes (como el precio), los productos más vistos pueden destacar también por otras características de interés para ellos

(como la novedad). A pesar de lo anterior, se puede notar como la tasa de conversión de los productos de categorías duras no es tanto menor si se tiene en consideración que su precio es alrededor de 15 veces el precio de un producto de categorías blandas. Lo mismo se debe considerar entre lo más comprado y lo más navegado, ya que a pesar de que la primera lógica entrega productos con una mayor probabilidad de que el visitante convierta, la segunda lógica puede entregar productos más interesantes y con mayor ingreso esperado.

Tabla 5.5: Tasa de conversión de los productos recomendados (no personalizado).

Ranking	Duras		Blandas	
	Más Comprados	Más Vistos	Más Comprados	Más Vistos
En el Top 28	3,06 %	1,99 %	3,83 %	2,69 %
En el Top 6	3,28 %	1,96 %	4,06 %	2,51 %

Finalmente, se debe considerar que a pesar de las diferencias descritas anteriormente entre los recomendaciones realizadas en base a lo más comprado y lo más visto, ciertamente habrá una relación entre ambos ranking de productos. En el caso de categorías duras, alrededor del 15 % de los productos entraron al top 6 en las recomendaciones de ambas lógicas, mientras que en el caso de categorías blandas estos productos representan el 18 %. Esto indica que sí habrá cierta relación entre los productos más comprados y los más vistos, pero en general ambas lógicas entregarán recomendaciones diferentes.

## 5.2.2. Resultados experimentales

Una vez concluido el período de experimentación para la primera fase se recopilaron y analizaron los resultados capturados por la plataforma de analítica de la empresa. Durante este período se recopiló la navegación de alrededor de 7.200.000 sesiones, equivalentes a cerca de 3.250.000 visitantes que pudieron ser individualizados por la herramienta de analítica (estos no representan necesariamente la cantidad de usuarios únicos que visitaron el sitio).

De las sesiones y visitantes para los cuales se registró navegación a través de la página de inicio del sitio, se eliminaron aquellos que por error la herramienta asignó con más de un grupo experimental (situación que ocurrió menos del 0.5 % de los casos), así como aquellos que registraron una única página vista (alrededor del 10 % de las sesiones cumplen esta condición), los cuales pueden ser identificados como un rebote y no como visitantes que realmente navegaron el sitio, y no serán de interés en esta ocasión.

Además se debe considerar que de las sesiones que fueron registradas durante el transcurso del experimento, un 40 % de ellas fueron realizadas a través de la versión para navegadores de escritorio (PC o Mac) mientras que el restante 60 % fueron realizadas a través de la versión web móvil para celulares. Se debe considerar que en el experimento se excluyeron los registros de la aplicación para smartphones, al no tener la misma implementación de recomendaciones que las versiones de escritorio y móvil.

Por otra parte, de las sesiones registradas en el experimento, en cerca del 33 % de ellas se pude identificar el usuario que se encontraba navegando el sitio. Al separar por tipo de dispositivo esta proporción cambia, ya que en el caso de escritorio el 44 % de las sesiones corresponde a usuarios identificables, mientras que en móvil sólo el 26 % de las sesiones son identificables.

### 5.2.2.1. Resultados generales de conversión y gasto

Las primeras métricas a analizar tendrán relación con la conversión y el gasto general de los grupos experimentales, con el objetivo de determinar si la colocación de las recomendaciones tuvo un impacto positivo en comparación al grupo que no se le mostró ningún carrusel con recomendaciones. En la Tabla 5.6 se muestran las tasas de conversión (TC) y el gasto promedio por sesión (ARPS) de los grupos experimentales. Los incrementales (en paréntesis) y los p-valores calculados son con respecto al grupo control (OO) a un 95 % de confianza.

Tabla 5.6: Resultados de conversión y gasto tras primera fase de experimentación.

Grupo	TC		Z	P-valor	ARPS		t	P-valor
CC	3.47 %	(+0.74 %)	0.8614	0.1945	\$3,758	(+\$168)	1.5222	0.0640
CN	3.45 %	(+0.22 %)	0.2514	0.4007	\$3,574	(-\$16)	0.1907	0.4244
CO	3.45 %	(+0.28 %)	0.3298	0.3708	\$3,637	(+\$47)	0.5689	0.2847
NC	3.50 %	(+1.47 %)	1.7013	0.0444	\$3,721	(+\$131)	1.4223	0.0775
NN	3.46 %	(+0.39 %)	0.4461	0.3278	\$3,786	(+\$196)	1.7832	0.0373
NO	3.47 %	(+0.78 %)	0.9127	0.1807	\$3,662	(+\$72)	0.8894	0.1869
OC	3.50 %	(+1.52 %)	1.7736	0.0381	\$3,525	(-\$65)	0.8257	0.2045
ON	3.45 %	(+0.11 %)	0.1235	0.4508	\$3,610	(+\$20)	0.2550	0.3994
OO	3.44 %				\$3,590			

Recordando la nomenclatura de los grupos, se observa que la tasa de conversión aumenta en mayor medida al incluir recomendaciones de productos de categorías blandas, sobre todo al basar éstas recomendaciones en los productos más comprados del sitio. Tomando en cuenta un nivel de confianza del 95 % se tiene que el único incremental significativo en la tasa de conversión lo obtuvo el grupo OC (es decir, sin recomendaciones de categorías duras y lo más comprado en blandas). A pesar de que el resto de los grupos no obtuvieron un incremental estadísticamente significativo, se observa que al ser todos los incrementales positivos con respecto al control, existe una tendencia que indica que la inclusión de los carruseles de recomendaciones probados aumentaría la cantidad de visitas en el sitio que convierten.

Ahora bien, analizando la sección derecha de la tabla relativa al gasto por sesión, se tiene que no todos los grupos lograron un incremental positivo con respecto al control. Más aún, los mayores incrementales se obtuvieron al momento de incluir alguna recomendación de productos de categorías duras, que poseen un valor promedio mucho más elevado, como ya se mencionó en la sección anterior. Además, el grupo OC que obtuvo el mayor incremental en la tasa de conversión, fue el que menos gasto promedio por sesión obtuvo. Esto muestra que la opción que más aumenta conversión no necesariamente será la opción que más aumenta el

ingreso total.

Se debe mencionar que el gasto promedio por sesión está compuesto por la tasa de conversión (el número de sesiones que compran), y el gasto promedio de las sesiones que compran (ARPPS). Por lo que en cierta forma el ARPS resume la tasa de conversión y el gasto de cada grupo en una sola métrica. En este sentido, para que un incremento en la tasa de conversión resulte en una disminución del ARPS, la disminución relativa del gasto debe ser mayor que el incremento relativo de la tasa de conversión.

$$\begin{aligned} ARPS &= \frac{Ingreso\ Total}{Sesiones\ Totales} \\ &= \frac{Ingreso\ Total}{Sesiones\ con\ Compra} \times \frac{Sesiones\ con\ Compra}{Sesiones\ Totales} \quad (5.1) \\ &= ARPPS \times TC \end{aligned}$$

Ahora bien, asumiendo que las sesiones totales no son alteradas negativamente por el efecto de los carruseles de recomendaciones, un aumento en la tasa de conversión sólo puede significar una mayor cantidad de sesiones con compra. En este caso, más sesiones con compras significa un denominador mayor en ARPPS, y si el ingreso total no crece en la misma proporción que las sesiones, el ARPPS disminuirá. Por ejemplo, en el caso extremo de que las sesiones con compras adicionales hagan un gasto de \$0, el ingreso total será el mismo en ambos grupos, pero al ser mayor el número de sesiones con compra el ARPPS disminuirá en la misma proporción que el aumento en la tasa de conversión. En este caso el ARPU en ambos grupos será igual. Por lo tanto, si las sesiones con compra adicionales hacen un gasto promedio mayor a \$0 debería observarse un incremento en el ARPS.

Así mismo, la única forma de obtener una disminución en el gasto promedio por sesión al implementar las recomendaciones, siendo que hubo un aumento en la tasa de conversión, es que sesiones que hubieran realizado una compra incluso sin la implementación, terminaron gastando menos que lo que hubieran gastado en ausencia de las recomendaciones. En otras palabras terminaron encontrando productos de menor precio que siguen cumpliendo con sus necesidades.

Teniendo lo anterior en consideración, los resultados de la Tabla 5.6 también muestran que el único grupo que obtuvo un incremento estadísticamente significativo a un 95% de confianza en el ARPS es el grupo NN, que concuerda con las recomendaciones de categorías duras y blandas con mayor precio en sus productos. A pesar de que el resto de los resultados no son estadísticamente significativos, sí sugieren que para las recomendaciones en categorías blandas la mayor tasa de conversión se obtuvo al recomendar los productos más comprados, aunque en términos de ingreso total la alternativa que más aumentó el ingreso promedio por visita fue la de los productos más vistos, que terminaron mostrando productos de mayor precio. Para el caso de categorías duras, tanto en conversión como en gasto promedio por sesión el mayor incremento se obtuvo al recomendar a los visitantes los productos más vistos.

### 5.2.2.2. Resultados de uso y post-click

A pesar de que los resultados obtenidos de manera general para la conversión y el ingreso promedio en el sitio son útiles para comparar el efecto global de las recomendaciones, permitiendo realizar un contraste con el escenario consistente en una página ausente de carruseles de recomendación de cualquier tipo, lo cierto es que a través de estas métricas no se puede entender el efecto directo que se tiene sobre la navegación de los visitantes al incluir elementos de este tipo. Para esto último será conveniente analizar la cantidad de clicks realizados por los visitantes en los sistemas de recomendación implementados, así como el comportamiento posterior al uso de los carruseles.

De esta forma, en la Tabla 5.7 se muestran las tasas de uso obtenidas para los distintos sistemas de recomendación, entendidas como el número de sesiones con click en el carrusel respectivo, dividido por el número total de visitas del grupo. A pesar de que lo ideal sería usar como denominador a los visitantes que efectivamente vieron el carrusel de recomendaciones (expuestos reales), este es un número imposible de conocer con la analítica de la empresa. Como ya se ha mostrado antes, el número de expuestos reales a los sistemas de recomendación será sólo una fracción del número total de visitantes asignados dada la ubicación de los sistemas implementados (más cerca de la sección inferior de la página de inicio), pero esta proporción será difícil de determinar con seguridad. De todas formas si se asume que no habrá diferencia a priori entre los grupos experimentales con respecto a la fracción de visitantes que navegan la parte inferior del sitio, la comparación utilizando el total de visitantes asignados sigue siendo igual de válida.

En la Tabla 5.7 la tasa de uso total representa clicks en cualquiera de los recomendadores, evitando duplicar sesiones que pueden haber hecho uso de ambos sistemas. Los resultados muestran que el uso de los carruseles fue mayor en presencia de ambos carruseles (por ejemplo, el grupo NN), en comparación con los grupos a los que se les desplegó un único carrusel de recomendaciones (por ejemplo, el grupo NO). Los resultados muestran que al incluir ambos sistemas de recomendación se aumentó las sesiones con uso en poco más del doble. Más aún, se puede observar que en el caso de los grupos que presentaron ambos carruseles, la tasa de uso total (que contabiliza una sola vez sesiones con click en ambos carruseles) es prácticamente igual a la suma de las tasas de uso por separado en el recomendador de productos duros y el recomendador de productos blandos. Esto indica que los carruseles no compitieron entre sí, y que los visitantes interesados más en una macro categoría que en la otra son diferenciables, y no se vieron distraídos por la presencia de la otra macro categoría.

En particular la Tabla 5.8 permite observar la diferencia de uso entre las lógicas de recomendación, comparando aquella construida en base a los productos más comprados y en base a los más navegados, para ambas macro categorías. En esta tabla se incluye la diferencia porcentual ( $\Delta \%$ ) usando como base (denominador) el carrusel con menor uso entre ambas lógicas. Los resultados muestran tanto en duras como en blandas que el mayor uso se obtuvo en presencia de las recomendaciones basadas en los artículos más vistos, por sobre las basadas en los más comprados. Tanto en las recomendaciones duras como en blandas la diferencia es estadísticamente significativa a un 95 % de confianza, y se tiene una mayor diferencia entre lógicas dentro de las recomendaciones de productos blandos que dentro de las de productos duros. Esto indicaría que los artículos más navegados despiertan un mayor interés por parte

Tabla 5.7: Tasa de uso por carrusel para los distintos grupos no personalizados.

Grupo	TU Duras	TU Blandas	TU Total
CC	0,57 %	0,63 %	1,19 %
CN	0,58 %	0,78 %	1,34 %
CO	0,60 %		0,60 %
NC	0,64 %	0,64 %	1,26 %
NN	0,64 %	0,79 %	1,41 %
NO	0,64 %		0,64 %
OC		0,71 %	0,71 %
ON		0,88 %	0,88 %

de los visitantes que visualizan los sistemas de recomendación, pudiendo potencialmente generar una mayor cantidad de compras.

Por otro lado, la Tabla 5.9 compara de manera similar el uso de los mejores exponentes de los recomendadores de productos duros y blandos. Nuevamente la diferencia porcentual ( $\Delta$  %) se calcula utilizando como denominador al carrusel con menor tasa de uso. Esta comparativa muestra que existe una diferencia estadísticamente significativa que indica que las recomendaciones de productos blandos obtienen un 36 % más de uso que las recomendaciones de productos duros, indicando un mayor número de visitantes que muestran interés por productos de la primera macro categoría.

Tabla 5.8: Comparación tasa de uso entre lo más comprado y lo más visto.

Grupo	TU	$\Delta$ % en Uso	Z	p-valor
Duras	CO	0,60 %		
	NO	0,64 %	(+7, 12 %)	3,3307
Blandas	OC	0,71 %		
	ON	0,88 %	(+23, 33 %)	11,4510

Ahora bien, nada asegura que una mayor tasa de uso signifique un mayor incremento en la conversión o en el ingreso, por lo tanto será necesario evaluar lo que sucede de manera posterior al click. Para esto, en la Tabla 5.10 se muestra de forma separada y para cada grupo experimental la proporción de sesiones que realizaron alguna compra tras el click, la proporción de sesiones que incluyeron en su compra el mismo artículo (o artículos) en el que hicieron click, la proporción de sesiones que agregaron al carro (basket) el mismo artículo en el que hicieron click y el gasto promedio de las sesiones que incluyeron en sus compras los productos que clickearon.

Los resultados muestran que las mayores tasas de conversión post-click se obtuvieron en los grupos en presencia de recomendaciones de categorías blandas, mientras que las menores tasas fueron en los grupos que hicieron uso de las recomendaciones en categorías duras. Ade-

Tabla 5.9: Comparación tasa de uso entre recomendadores de categorías duras y blandas.

Grupo	TU	$\Delta\%$ en Uso	Z	p-valor
Duras (NO)	0,64 %			
Blandas (ON)	0,88 %	(+36, 18 %)	16,4089	0,0000

más, se observa que alrededor de un tercio de las sesiones con compras posteriores al click incluyeron en sus compras alguno de los productos clickeados en los carruseles. Al mismo tiempo, se observa que los grupos con menor tasa de conversión post-click en general también fueron los que más gasto promedio realizaron. La proporción de sesiones que agregaron el producto que clickearon al carro indica que una fracción considerable de visitantes mostraron interés directo en el producto que clickearon, existiendo una probabilidad de haber concretado la compra de dicho producto en visitas posteriores al sitio.

Tabla 5.10: Resultados de métricas posteriores al click en recomendaciones no personalizadas.

Grupo	TC post-click	TC post-click (mismo SKU)	Al Carro (Mismo SKU)	Gasto post-click (mismo SKU)
CC	2,41 %	0,95 %	9,24 %	\$ 50.988
CN	2,19 %	0,79 %	7,94 %	\$ 65.716
CO	1,46 %	0,64 %	6,54 %	\$ 93.989
NC	1,79 %	0,47 %	8,12 %	\$ 140.617
NN	1,76 %	0,49 %	7,04 %	\$ 103.330
NO	1,16 %	0,29 %	4,26 %	\$ 185.490
OC	2,68 %	1,05 %	11,19 %	\$ 20.272
ON	2,16 %	0,46 %	9,54 %	\$ 21.820

Para entender de mejor forma la implicancia de los resultados anteriores, se debe considerar que el efecto directo del carrusel sobre el ingreso adicional que se puede generar en el sitio estará determinado por (i) el porcentaje de visitantes que se interesan en el carrusel (la tasa de uso), (ii) la probabilidad de convertir de manera posterior al click y (iii) el monto gastado tras convertir. A pesar de que otros comportamientos pueden influir indirectamente en el incremento o disminución del ingreso, esta interacción describe a grandes rasgos el mayor impacto de las recomendaciones. Por ejemplo, el grupo OC (únicamente con carrusel de productos más comprados en categorías blandas) fue el que más convirtió de forma posterior al click (tanto en compra como en agregar al carro), pero al mismo tiempo fue el grupo que hizo el menor gasto promedio (dado que los productos recomendados bajo esta lógica también resultan ser más baratos). En contraste al grupo OC, se tiene el grupo NN (expuestos a lo más navegado en ambas macro categorías), que a pesar de convertir poco más de la mitad de lo que convierte el grupo OC en el post-click, su gasto promedio en productos clickeados es cinco veces superior. Sin mencionar que el grupo NN además tiene una tasa de uso dos veces mayor. Es por esto que el incremento esperado en el ingreso por parte de NN debiera ser mayor que OC, y esto concuerda con los resultados generales obtenidos de conversión y gasto mostrados en la Tabla 5.6.

Extendiendo el análisis anterior al resto de los grupos, si se asume que el ingreso adicional que pueden aportar los carruseles de recomendación corresponde (al menos de forma directa) al gasto realizado en los productos comprados posteriores al click, este ingreso adicional esperado por sesión se puede resumir en la Tabla 5.11, calculándolo al multiplicar la tasa de uso, la tasa de conversión post-click (que compran lo que clickean) y el gasto promedio en los productos clickeados y comprados. Esta métrica sugiere nuevamente que el mayor incremental en el ingreso se obtendría al colocar recomendaciones tanto para categorías duras como categorías blandas (en particular con los grupos NC y NN), siguiendo con los grupos con un único recomendador de categorías duras y en último lugar están los grupos con un único recomendador de categorías blandas.

Tabla 5.11: Estimación del ingreso adicional esperado por sesión aportado directamente por las recomendaciones no personalizadas.

<b>Grupo</b>	<b>Ingreso adicional esperado por Sesión (directo)</b>
CC	\$ 5,8
CN	\$ 7,0
CO	\$ 3,6
NC	\$ 8,3
NN	\$ 7,1
NO	\$ 3,5
OC	\$ 1,5
ON	\$ 0,9

### 5.2.2.3. Resultados por dispositivo

Como se mencionó anteriormente, más de la mitad de las sesiones que se registraron durante el experimento provinieron de navegación a través de la versión de móvil para celulares del sitio. Entendiendo que el desempeño de los carruseles de recomendación puede diferir entre dispositivos (producto de las diferencias entre la predisposición a comprar, la posición de los elementos, el número de recomendaciones desplegadas en una primera vista, entre otros factores) se procedió a analizar los resultados incluyendo esta división entre los grupos.

En la Tabla 5.12 se muestran los resultados para conversión y gasto para los grupos experimentales en el caso de las sesiones que fueron iniciadas en la versión de escritorio del sitio. Por otro lado, en la Tabla 5.13 se muestran los resultados para las mismas métricas pero en el caso de las sesiones que fueron iniciadas en la versión web móvil del sitio. Observando ambas tablas se aprecia como los visitantes que navegan a través de un dispositivo de escritorio suelen convertir casi dos veces más que los que navegan a través de un dispositivo móvil. Además, los primeros también suelen realizar compras de precios más elevados.

Al comparar ambas tablas con los resultados de conversión y compra sin desagregar, se tiene que las conclusiones que se obtienen son similares. Se debe considerar que al dividir la muestra por dispositivo, se reduce el número de unidades experimentales por grupo y por lo

Tabla 5.12: Resultados generales de conversión y gasto para dispositivos de escritorio (no personalizadas).

Grupo	TC		Z	p-valor	ARPS		t	p-valor
CC	4,64 %	(−0,41 %)	0,3527	0,3622	\$5.851	(+\$378)	1,6436	0,0501
CN	4,62 %	(−1,01 %)	0,8756	0,1906	\$5.488	(+\$15)	0,1031	0,4590
CO	4,64 %	(−0,42 %)	0,3660	0,3572	\$5.650	(+\$177)	1,2253	0,1102
NC	4,74 %	(+1,64 %)	1,4060	0,0799	\$5.653	(+\$180)	1,1778	0,1194
NN	4,70 %	(+0,91 %)	0,7828	0,2169	\$5.617	(+\$145)	1,0249	0,1527
NO	4,67 %	(+0,22 %)	0,1893	0,4249	\$5.627	(+\$155)	1,1095	0,1336
OC	4,75 %	(+1,79 %)	1,5477	0,0608	\$5.421	(−\$51)	0,3795	0,3521
ON	4,67 %	(+0,07 %)	0,0625	0,4751	\$5.637	(+\$165)	1,1858	0,1178
OO	4,66 %				\$5.472			

tanto aumenta la variabilidad.

Tabla 5.13: Resultados generales de conversión y gasto para dispositivos móviles (no personalizadas).

Grupo	TC		Z	p-valor	ARPS		t	p-valor
CC	2,67 %	(+2,01 %)	1,5525	0,0603	\$2.335	(+\$20)	0,1987	0,4212
CN	2,67 %	(+1,84 %)	1,4246	0,0771	\$2.284	(−\$31)	0,3154	0,3762
CO	2,65 %	(+0,98 %)	0,7694	0,2208	\$2.268	(−\$48)	0,4889	0,3125
NC	2,65 %	(+1,13 %)	0,8738	0,1911	\$2.408	(+\$92)	0,8081	0,2095
NN	2,62 %	(+0,00 %)	0,0008	0,4997	\$2.556	(+\$240)	1,5272	0,0634
NO	2,65 %	(+1,10 %)	0,8547	0,1964	\$2.316	(+\$0)	0,0022	0,4991
OC	2,65 %	(+1,12 %)	0,8757	0,1906	\$2.239	(−\$77)	0,8021	0,2113
ON	2,62 %	(+0,06 %)	0,0482	0,4808	\$2.235	(−\$81)	0,8825	0,1887
OO	2,62 %				\$2.316			

Además de las métricas de conversión y gasto, se analiza cómo afecta el tipo de dispositivo al uso de los carruseles de recomendaciones. En la Tabla 5.14 se resumen las tasas de uso por dispositivo y su diferencia porcentual. Los resultados muestran en todos los grupos un mayor uso al realizarse la navegación a través de un dispositivo de escritorio. La cantidad de sesiones con uso puede decrecer incluso hasta un 38%. Una de las razones que pueden explicar una menor tasa de uso en dispositivos móviles es el nivel de vistas o pantallas que el usuario debe desplazarse para encontrar los carruseles de recomendación. La Figura A.2 muestra que los usuarios de móvil deben desplazarse alrededor de 8 pantallas para llegar a las recomendaciones, mientras que en la Figura A.2 se observa que los de escritorio deben desplazarse alrededor de 6 pantallas. En este sentido se debe considerar que los distintos elementos del sitio compiten en cierto nivel por los clicks de los visitantes, y en el caso de dispositivos móviles puede que los elementos que están arriba en el sitio mermen en mayor medida la cantidad de expuestos reales a las recomendaciones. Aún así, los resultados arrojan que la disminución en el caso de recomendaciones en categorías blandas es de menos magnitud, lo que sugiere que los usuarios de dispositivos móviles muestran mayor interés en este tipo de productos.

Tabla 5.14: Comparación de tasas de uso por dispositivo para recomendaciones no personalizadas.

Grupo	TU Escritorio	TU Móvil	Lift
CC	1,35 %	1,08 %	-20 %
CN	1,55 %	1,20 %	-22 %
CO	0,78 %	0,48 %	-38 %
NC	1,42 %	1,15 %	-19 %
NN	1,61 %	1,28 %	-21 %
NO	0,78 %	0,55 %	-29 %
OC	0,73 %	0,70 %	-5 %
ON	0,93 %	0,84 %	-9 %

Considerando ambos resultados se puede determinar que en dispositivos móviles el efecto de los carruseles de recomendación será menor que en dispositivos de escritorio, a pesar de que el volumen de visitantes móviles es mayor. Esto es debido tanto a que los visitantes de móviles tendrán menos probabilidad de visualizar las recomendaciones, como por el hecho de que este grupo de visitantes en general tiene menos tendencia a concretar una compra o a gastar montos demasiado altos.

Por último, se tiene en la Figura 5.6 el recuento de clicks por posición dentro del carrusel de productos (slot), mientras que en la Figura 5.7 se tiene el mismo recuento pero para sesiones en dispositivos móviles. La primera figura muestra como los clicks de los visitantes que usan las recomendaciones suelen concentrarse en las primeras seis posiciones, que conforman la primera vista del carrusel. En específico, entre un 60-66 % de los clicks se realizan en estas primeras posiciones, variando levemente según el tipo de lógica. Esto sugiere que una gran cantidad de visitantes no visualizarán productos más allá de la sexta posición. Se estima que la mayoría de los clicks en las últimas posiciones se deben a pulsaciones involuntarias de los usuarios al intentar navegar más productos (considerando que en ocasiones los carruseles pueden haber renderizado menos de 28 productos si alguno de ellos se encuentra sin stock).

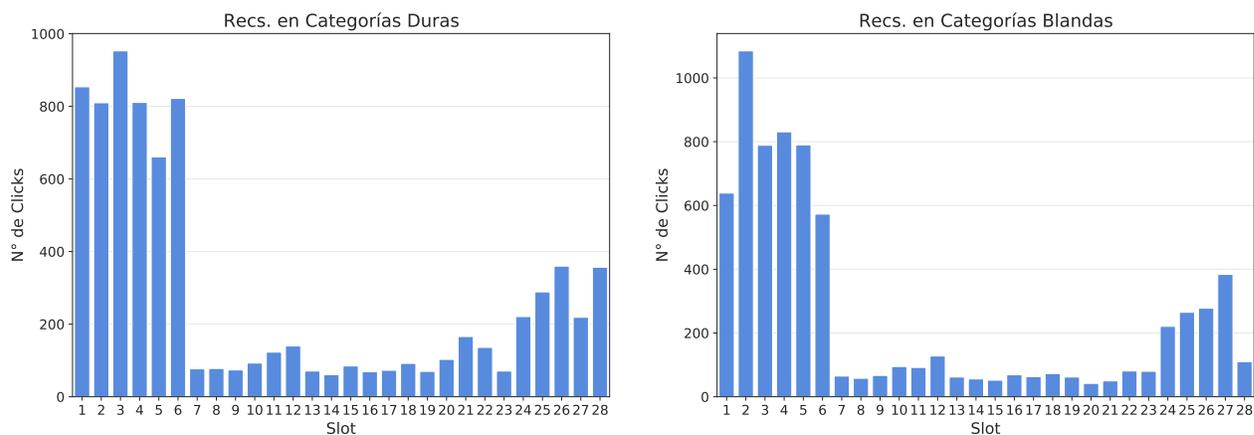


Figura 5.6: Posición de los clicks en los carruseles de recomendaciones no personalizadas para dispositivos de escritorio.

En el caso de los recuentos de clicks para dispositivos móviles la figura es similar, pero esta vez los clicks se concentran en las primeras dos posiciones, que corresponden a la primera vista en la versión de estos dispositivos (como se aprecia en la Figura A.1). En este caso, las primeras dos posiciones concentran entre un 41-48% de los clicks dependiendo de la lógica. Dado que la navegación del carrusel en móviles es a través deslizamiento táctil en comparación con el uso de botones en la versión de escritorio, el uso por posición va decreciendo de manera suave mientras se avanza entre los productos.

A pesar de que cierta forma de esta distribución de clicks puede explicarse porque las primeras posiciones debieran mostrar productos más atractivos, los quiebres pronunciados en las posiciones dos y seis sugieren que sin importar el contenido un gran porcentaje de visitantes no navegará más allá de la primera vista, por lo que puede ser importante poner especial atención al momento de decidir este contenido.

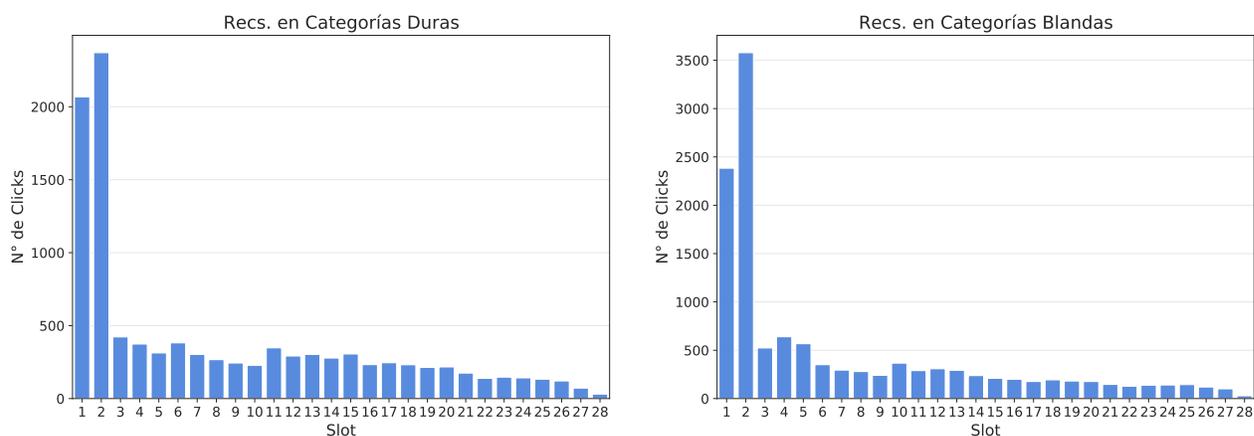


Figura 5.7: Posición de los clicks en los carruseles de recomendaciones no personalizadas para dispositivos móviles.

### 5.3. Segunda Fase: Recomendaciones Personalizadas usando Segmentos

Finalizada la experimentación con sistemas de recomendaciones que utilizan la información general del sitio y no realizan distinción por visitante, el paso siguiente será probar si es posible introducir un nivel de personalización dentro de los sistemas de recomendaciones con el objetivo de mostrar categorías de productos que puedan ser más interesantes para cada usuario o grupo de usuarios, y así aumentar las probabilidades de generar una venta adicional. Para cumplir este objetivo se detalla a continuación el proceso de generación de segmentos de clientes a partir de su información de compra y navegación, el posterior diseño y generación de los productos que de manera automatizada serían mostrados a cada segmento de clientes y finalmente los resultados del período de experimentación para estos segmentos y recomendaciones generadas.

### 5.3.1. Generación de segmentos por afinidad

Como ya se explicó en la metodología propuesta, para lograr implementar un nivel de personalización en las recomendaciones será necesario utilizar un método para cuantificar la afinidad de los clientes con respecto a las distintas categorías de productos a considerar. En este caso se utilizó el mayor nivel de agregación de la empresa, en donde se consideran 23 categorías tales como Electrodomésticos, Vestuario, Dormitorio, entre otras. Como información para cuantificar esta afinidad se hizo uso de la información histórica de compra y navegación por cliente.

#### 5.3.1.1. Afinidad según compras

En el caso de la información de compras, un primera etapa consistió en un análisis realizado en base a la construcción de variables RFM (recencia, frecuencia y gasto) de cada cliente, utilizando las bases de datos que recopilan información transaccional, considerando una ventana de tiempo de un año móvil (con respecto al cálculo de estas variables), siendo estas tanto en sitio web como en tienda física. La generación de estas variables fue realizada en primera instancia de manera general para todas las compras en la empresa, y luego se repitió el proceso para cada una de las 23 categorías de productos definidas para el análisis. Una vez procesados y organizados los datos a nivel de cliente, se analizó cada variable de manera exploratoria para entender a priori patrones que permitieran mejorar la generación de los grupos que serán la base de la personalización.

Antes de comenzar los análisis se procedió a hacer una limpieza de outliers sobre las variables de Frecuencia y Gasto, para evitar que más adelante esta información afecte las segmentaciones a realizar. Un cliente anonimizado caerá en la categoría de outlier si posee una cantidad anormal de compras o gasto, pudiendo esto representar errores, bots o compras de empresas. Para definir esta eliminación de valores anómalos se decidió realizar un modelo de segmentación basado en K-Means que agrupara los valores por cercanía, descartando aquellos grupos integrados por un porcentaje muy pequeño de clientes. En este caso el número de segmentos (K) se determinó arbitrariamente en 15, con el objetivo de inspeccionar de manera clara aquellos grupos pequeños que están muy separados del normal de datos. Este proceso fue repetido utilizando el algoritmo una vez sobre datos de Frecuencia y otra sobre datos de Gasto, eliminando finalmente cerca de 1.400 clientes anonimizados.

Tras esta limpieza, la primera variable analizada corresponde a la frecuencia total de compra, medida como el número de días en que cada cliente realizó una transacción. La Figura 5.8 muestra que una gran cantidad de los clientes para los cuales se registra alguna compra en el último año (alrededor de 5,5 millones de clientes), registra un único día con compra. En específico, alrededor del 26 % de los clientes se encontrarán en este grupo. Esto sugiere que habrá un grupo de clientes con baja frecuencia de compra (por ejemplo, igual a 1) para los cuales no se dispondrá de información suficiente para determinar una preferencia clara, por lo que será importante distinguir estos grupos con baja y alta cantidad de información, pues determinará la seguridad con la cual se podrá asignar una afinidad por cliente.

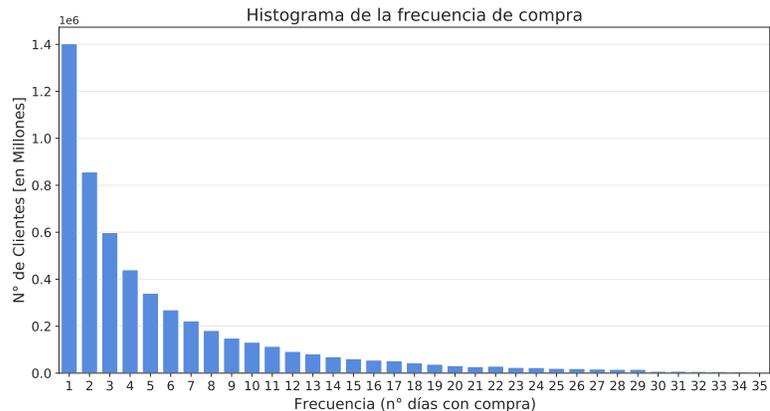


Figura 5.8: Frecuencia de compra en días por cliente.

Un fenómeno similar ocurre con la variable de gasto, una vez que se han eliminado los clientes anómalos. En general el gasto se concentra en montos bajos, relacionado con los clientes esporádicos que han comprado una sola vez en el último año. Como se puede observar en la Figura C.1 en anexos con la distribución del gasto y en los valores para sus rangos, el 25 % de los clientes con compra registran montos menores a \$30.980, indicando que además muchos de estos clientes adquirieron un único producto. Aún así, la distribución muestra una cola derecha muy larga de clientes con alto gasto y que no necesariamente son anómalos. Esto nuevamente muestra que habrá un grupo de clientes con baja y alta cantidad de información relativa a su afinidad.

Por lo tanto se consideró conveniente realizar, de manera previa a la segmentación según afinidad, un muestreo por segmentación, intentando distinguir aquellos clientes que poseen una alta cantidad de información de los clientes esporádicos con baja información. En una primera iteración de este muestreo (utilizando K-Means y un escalamiento Min-Max) se descartó la variable de Recencia del grupo de variables RFM, dado que el algoritmo asignaba demasiado peso a esta variable quedando los grupos perfectamente separados por ella (un ejemplo de esto se puede ver en la Figura C.2 en anexos). Dado que el aporte de esta variable es menor para el objetivo final de generar segmentos por categoría, se decide eliminarla para evidenciar de mejor forma los patrones de frecuencia y gasto.

Con las variables de frecuencia de compra y gasto por cliente se realizó un nuevo modelo de segmentación, teniendo las mismas consideraciones que en la iteración anterior, y seleccionando el número de segmentos a generar con apoyo de los puntajes calculados de Calinski-Harabasz (CH) y Silhouette. La Figura 5.9 muestra los resultados de estos puntajes. A pesar de que se aprecia que los valores se maximizan con 2 grupos (lo que puede sugerir la existencia de algunos valores que siguen siendo muy extremos), el gráfico también muestra que con 9 grupos se obtiene un alto puntaje para CH y Silhouette, encontrando un buen equilibrio entre cantidad de información y separación de los grupos. Esta cantidad de grupos permitirá describir de manera aproximada una combinación de tres niveles (bajo, medio, alto) para las dos variables analizadas.

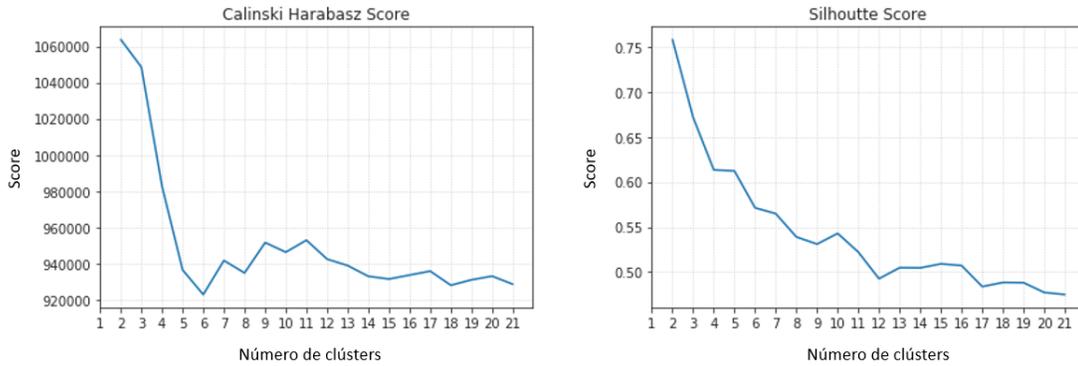


Figura 5.9: Resultados para K-Means utilizando variables FM totales

El resultado de K-Means con 9 segmentos se puede observar gráficamente en la Figura 5.10. De esta representación se desean señalar dos cosas: en primer lugar, se tiene que existirá un grupo cercano a cero tanto en gasto como en frecuencia de compra, que está muy densamente poblado y que puede ser identificado como los clientes con baja cantidad de información. En específico, cerca del 58% de los clientes fueron asignados a este grupo, lo que indica que la decisión de incluir o no incluir a estos en la segmentación por afinidad será significativa. En segundo lugar, la figura permite observar que en general todos los grupos estarán densamente poblados, y que K-Means trabajará como un particionamiento fijando límites entre los bloques de clientes. Este mismo comportamiento se evidencia más adelante (aunque en más dimensiones) al utilizar el gasto por categoría, dado que los datos serán densos y continuos.

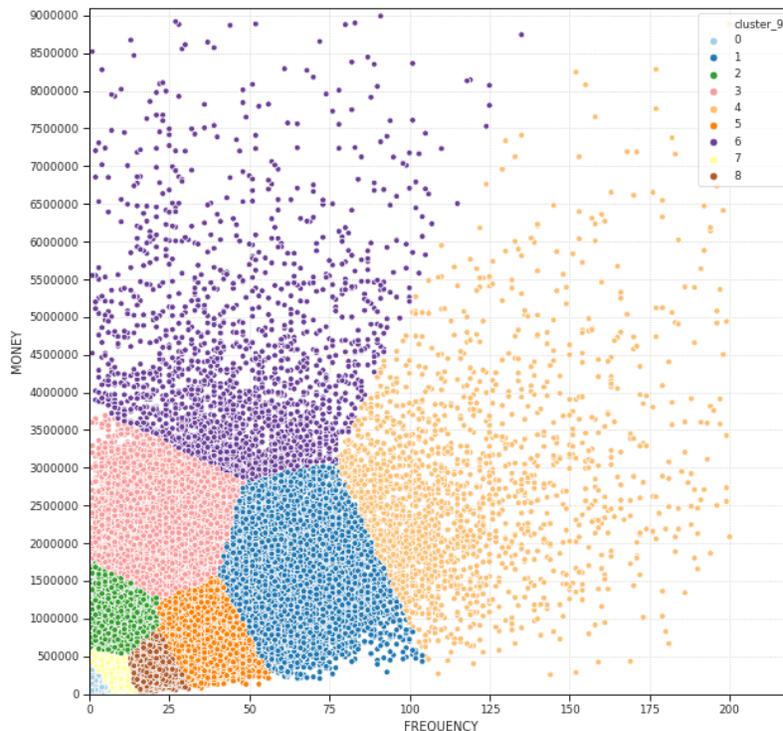


Figura 5.10: Segmentación de clientes en base a FM utilizando K-Means con 9 grupos

Una vez concluida la segmentación anterior, se procedió a etiquetar al total de clientes según el resultado del algoritmo, asignándolos como de baja cantidad de información y de alta cantidad de información. Esto para realizar dos segmentaciones en base a compras por separado: una incluyendo a todos los clientes y otra incluyendo sólo a los clientes con alta cantidad de información.

Para la primera segmentación que utiliza la totalidad de los clientes, se registraron como variables sus gastos totales en el último año para cada una de las 23 categorías de productos seleccionadas. La razón para considerar sólo el gasto y no la frecuencia de compra por categoría en esta etapa es que al añadir otras 23 variables a la base de datos, la dimensionalidad y complejidad del problema crece enormemente, contrario al objetivo final de esta etapa que es generar un número reducido y abordable de segmentos generales de gran tamaño, y no realizar una personalización exhaustiva. Esto considerando tanto los alcances del proyecto como las limitaciones técnicas con las que cuenta la empresa. Por esto se estima que los gastos por categoría proporcionarán una mayor cantidad de información que sólo considerando las frecuencias de compra.

Con el objetivo de que los gastos en cada una de las categorías sean comparables, teniendo en consideración las diferencias de precios de los distintos artículos que comercializa la empresa, se realizó un escalamiento convirtiendo el gasto absoluto por categoría en un gasto porcentual por categoría, calculado en base al gasto total del cliente (la suma de los gastos de todas las categorías), de forma que todos los valores estuvieran entre 0 y 1. Una vez realizada esta transformación, el número de segmentos a generar se decidió en base a las métricas de compacidad y separación. En particular, en la Figura 5.11 se muestra el resultado para el puntaje de Sum of Squares, que indica que un buen equilibrio entre separación y número de clusters se logra al utilizar 20 segmentos, dado que con este valor el puntaje se minimiza (dentro del rango de evaluación considerado).

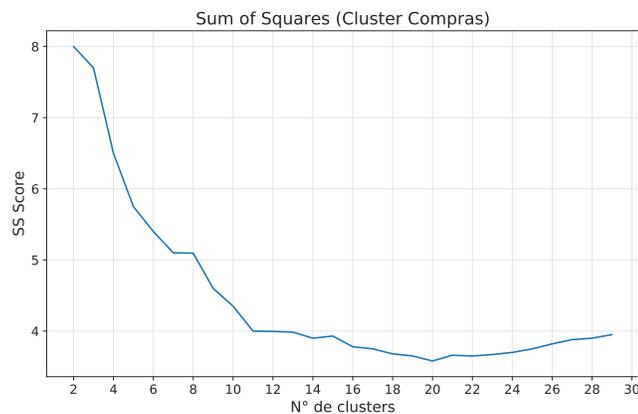


Figura 5.11: Resultados para el puntaje de Sum of Squares en segmentación por compras por categoría.

Una vez generados los 20 segmentos, se extraen y analizan los centros de los grupos. En la Figura C.3 en anexos se detallan estos valores, en donde los centros pueden interpretarse como el gasto porcentual promedio del grupo en dicha categoría. Centros con valores altos en ciertas categorías simbolizarían entonces una alta propensión a comprar y por lo tanto cierta

afinidad hacia la categoría. Con miras a generar las recomendaciones para cada grupo, se opta por caracterizar cada uno de ellos a través de su centro con valor más alto, lo que indicaría la categoría con mayor preferencia de este grupo de clientes. Al repetir este proceso con todos los segmentos generados por el algoritmo, se caracterizan 18 grupos (dado que habrán grupos distintos que compartirán categoría con valor máximo), más un grupo “indefinido” que no presentará una preferencia clara por ninguna categoría, correspondiente a cerca del 8% de los clientes.

Concluida esta primera segmentación en base a compras, se repite el mismo proceso pero esta vez seleccionando sólo a los clientes que poseen una mayor cantidad de información en cuanto a compras, determinados por el muestreo por segmentación realizado previamente. Es así como se filtra de la base de datos el grupo densamente poblado descrito anteriormente en la Figura 5.10, trabajando con el 42% aproximado de clientes restantes. Para esta segmentación también se escalan los gastos por categoría de forma que representen el gasto porcentual por cliente, y se decidió utilizar el mismo número de segmentos que al utilizar la totalidad de los clientes.

Los resultados para esta segmentación se muestran en la Figura C.4 en anexos, en donde una de las diferencias que se puede apreciar entre las dos segmentaciones es que esta vez los valores máximos para los centros de cada grupo son menores que para los centros generados al utilizar a todos los clientes. Esto muestra que al tener mayor cantidad de información de compras para los clientes, los grupos formados muestran tendencias menos extremas hacia una categoría en específico, y por lo tanto habrá un mayor número de clientes que distribuyen su gasto en distintas líneas de productos.

Tras la identificación de los centros para cada segmento, nuevamente se determina el valor máximo para cada grupo y se cataloga al grupo según la categoría de producto a la que pertenece este centro, en miras de las futuras recomendaciones. Este proceso arroja 16 grupos de acuerdo a la categoría con mayor gasto promedio (3 categorías se repiten) y 1 grupo “indefinido” que no presentará una preferencia clara por ninguna categoría, correspondiente a cerca del 6,6% de los clientes segmentados.

### **5.3.1.2. Afinidad según navegación**

Considerando la segmentación en base a compras, ya se tiene una primera forma básica para poder asignar una categoría de preferencia a cada uno de los clientes, que servirá posteriormente para probar la personalización en las recomendaciones de productos. Ahora bien, la navegación de los usuarios también podrá aportar información relevante sobre las preferencias de los visitantes del sitio, sobre todo de aquellos usuarios que sólo navegan y son candidatos a convertirse en compradores. Además de lo anterior, otra ventaja de incluir en el análisis los datos de navegación es que se tiene una gran cantidad de interacciones por usuario cada día, en comparación a lo que se tiene con las compras. Es decir que en un período más breve de tiempo se puede tener bastante información sobre los productos que interesan a cada uno de los visitantes, pudiendo utilizar intervalos de tiempo más pequeños y que respondan de mejor forma a las tendencias y cambios de intereses que suelen ocurrir.

Para utilizar la información de navegación para la generación de segmentos, de manera similar a lo realizado en la segmentación por compras, se calculó la frecuencia con la que cada visitante visualizó una página de producto (que se denominará en adelante como *hits*) y el número de sesiones por visitante. Ambas métricas por visitante se calcularon para las 23 categorías de productos definidas por la empresa, así como también se calculó un total por usuario que no hiciera distinción entre categorías. La ventana de tiempo utilizada fue de 3 meses móviles. La ventaja de usar los *hits* por sobre el número de sesiones es que reflejan mejor si la navegación de cierta categoría se hizo con mayor intensidad, por ejemplo visualizando muchos productos distintos dentro de una categoría (actividad que se reflejaría como una única sesión).

Al igual que con los datos de compras, antes de comenzar la segmentación se consideró conveniente realizar una eliminación de usuarios anómalos que estarían representados por un número excesivamente alto de hits o sesiones totales. En general estos datos pueden ser atribuidos a bots que revisan el sitio de forma automática. La eliminación de outliers se realizó de forma análoga a la sección anterior, es decir, utilizando K-Means sobre ambas variables totales para generar grupos pequeños y separados de usuarios. Realizando este proceso se eliminaron cerca de 550 usuarios con valores anormales, dejando alrededor de 1.3 millones de usuarios con valores registrados.

Un análisis exploratorio de estos valores arroja una distribución de valores similar a la distribución de la frecuencia de compras. Como se puede observar en la Figura 5.12, una gran cantidad de usuarios estarán registrados con una única sesión con visita a un producto. En específico cerca del 22 % de los usuarios alcanzarán solamente 2 sesiones. Una distribución similar se puede observar para el número de hits por usuario, graficado en la Figura C.5 en anexos, teniendo en consideración que en promedio un usuario registra alrededor de 8 hits por sesión con visita a una página de producto. Sobre todo considerando la información de los hits, una baja cantidad de datos no permitirá saber con seguridad la preferencia del usuario, y de forma similar a lo realizado con los datos de compras será conveniente separar los usuarios en aquellos con baja y alta cantidad de información.



Figura 5.12: Distribución del número de sesiones con visita a PDP por usuario.

El resultado para este muestreo por segmentación en base a la navegación se observa

gráficamente en la Figura C.6 en anexos. Este proceso determinó un particionamiento del conjunto densamente poblado de cliente, formando un grupo integrado por alrededor del 52% de los usuarios, que serán etiquetados como de baja calidad de información ya que registraron en promedio menos de 13 hits o interacciones con una página de producto. Esta etiqueta servirá nuevamente para filtrar a los clientes y realizar dos etapas de segmentación.

Para las segmentaciones por afinidad a cada categoría, se decidió utilizar sólo los hits o número de interacciones con las páginas de producto, con el objetivo de reducir la dimensionalidad del problema y dado que se consideró que esta métrica almacena mayor información acerca del interés de los usuarios que utilizando solamente el número de sesiones con visita a cada categoría. Para poder utilizar valores comparables entre categorías se calcularon las interacciones porcentuales por categoría por usuario (dividiendo cada valor por el número total de interacciones), de forma que todos los valores estuvieran entre 0 y 1. Una vez realizada esta transformación el número de segmentos a generar se decidió en base al puntaje de Sum of Squares que puede observarse en la Figura 5.13, para equilibrar el nivel de separación con el número de clústers. En la figura se observa que el mínimo es alcanzado utilizando 18 segmentos.

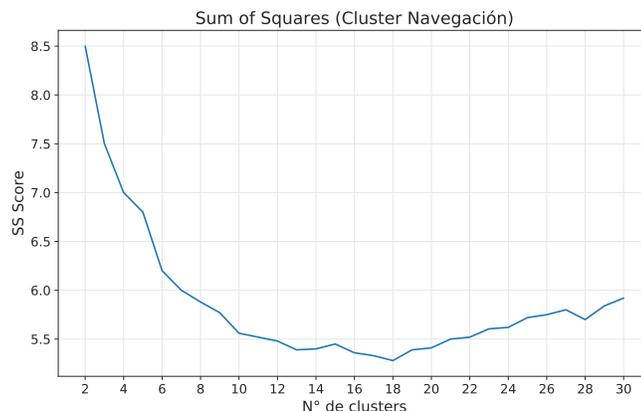


Figura 5.13: Resultados para el puntaje de Sum of Squares en segmentación por navegación por categoría.

Los centros para estos 18 segmentos generados y el tamaño relativo de ellos se pueden observar en la Figura C.7 en anexos. Con estos resultados (y de manera análoga a lo realizado en la segmentación en base a compras) se buscó el centro con el valor máximo para cada grupo con el objetivo de poder caracterizar la categoría con mayor afinidad presentada por cada uno. El resultado de este proceso arrojó 17 grupos con categoría de producto definida (ninguna se repitió), además de un grupo etiquetado como “indefinido” dado que no se observó una preferencia clara por ninguna categoría y que corresponde a cerca del 18,2% de los usuarios registrados.

El mismo procedimiento anterior se repitió, pero esta vez sólo para los clientes que poseen una mayor cantidad de información, definido por el muestreo por segmentación realizado inicialmente y que significará reducir la cantidad de clientes a considerar a cerca de un 48% del total. Los datos por cliente nuevamente fueron tratados de manera porcentual y se utilizó para la segmentación 18 grupos. Los resultados para los centros y los tamaños relativos

de estos grupos se encuentran en la Figura C.8, donde se observa que, de manera similar a lo ocurrido con la segmentación en base a compras, los valores máximos por grupo ya no resultan tan altos dado que este grupo de clientes con más información suele navegar con cierta intensidad más de una sola categoría. Tras repetir el proceso de selección de la categoría con el mayor porcentaje de hits por grupo, se logran identificar 16 grupos con categoría de producto definida (una categoría fue compartida por dos grupos), además de un grupo “indefinido” que esta vez consiste en el 11,8% de los usuarios considerados.

### 5.3.1.3. Asignación de segmentos para recomendaciones

El procesamiento anteriormente descrito de segmentación en base a compras y navegación resulta en hasta cuatro asignaciones de categorías por cliente, con valores nulos para aquellos clientes sin información. En base a estas cuatro asignaciones se decidieron reglas para definir una única categoría a asignar por cliente, intentando reflejar de buena forma su afinidad y aprovechando las distintas fuentes de información con las que se disponían. Esto definirá finalmente las recomendaciones de productos que serán exhibidas en el período experimental.

De esta forma, se creó una serie de reglas con el objetivo de utilizar la información de las cuatro segmentaciones para seleccionar una única categoría. Se debe tener en consideración que finalmente los segmentos definen un total de 18 categorías, más un grupo indefinido (tanto en compra o navegación). La idea fue por lo tanto reducir al mínimo el tamaño del grupo indefinido, valiéndose de la información analizada sobre el usuario. Además se debe considerar que este proceso permitirá llenar los vacíos tanto de personas que no tengan compra, pero sí navegación y viceversa. Por otro lado, se tuvo en consideración aquellas situaciones en las que la asignación en base a compra concordaba con la asignación en base a navegación, así como aquellas en que no había concordancia.

El resultado de este proceso de asignación se puede observar en la Figura C.9, en donde se grafican las reglas ejecutadas sobre cada cliente. En resumen, las reglas generadas dan preferencia (siempre que se tenga información) a la asignación realizada según los datos de navegación, dado que en general mostrará preferencias más recientes de los usuarios y mayor potencial a comprar. En caso de no contar con esta información, se utilizará la segmentación realizada con la información de compra. Además de esto, las reglas intentarán siempre asignar alguna categoría a los clientes etiquetados como “indefinido”, ya sea complementando la información de navegación con la de compra o utilizando los muestreos por segmentación en donde se utilizaron sólo los clientes con alta cantidad de información.

Finalmente la aplicación de estas reglas resultó en la asignación de categorías sobre aproximadamente 5.6 millones de clientes, de los cuales el 21% contaba tanto con información de compra como de navegación, el 77% sólo con información de compra y el 2% restante sólo con información de navegación. De este total de clientes, sólo el 3% no pudo ser asignado a una categoría en específico, permaneciendo como “indefinido”.

### 5.3.2. Generación de recomendaciones por segmento

Con los clientes ya asignados a las 18 categorías de productos resultantes, el próximo paso fue la generación de los productos que serían recomendados en cada una de las categorías, así como la generación del proceso de transformación de las bases de datos de la empresa y la carga de datos de manera periódica hacia la plataforma de experimentación-personalización del sitio, con el objetivo de que las recomendaciones se mantuvieran actualizadas al avanzar los días.

**Lo más popular en Vestuario**

UNIVERSITY CLUB Sweater Cuello-Mujer	BASEMENT Jeans Skinny	BASEMENT Chaleco Angora	BASEMENT Sweater de Vestir-Mujer	HANGO Sweater Manga Larga-Mujer	BASEMENT Jeans Skinny
\$ 9.990 (Oferta) \$ 16.990	\$ 12.990 (Oferta) \$ 16.990	\$ 11.990 (Oferta) \$ 24.990	\$ 11.990 (Oferta) \$ 24.990	\$ 19.990 (Oferta) \$ 24.990	\$ 12.990 (Oferta) \$ 19.990

**Lo más popular en Calzado**

SYBILLA Zapatilla Urbana Mujer Negra	NEW BALANCE 373 Zapatilla Urbana Mujer	SYBILLA Zapatilla Urbana Mujer Rosada	NEW BALANCE 009 Zapatilla Urbana Mujer	ADIDAS Crazychaos Zapatilla Urbana-Mujer	NEW BALANCE 373 Zapatilla Urbana Mujer
\$ 7.990 (Oferta) \$ 19.990	\$ 19.990 (Oferta) \$ 39.990	\$ 7.990 (Oferta) \$ 19.990	\$ 25.990 (Oferta) \$ 42.990	\$ 29.990 (Oferta) \$ 49.990	\$ 19.990 (Oferta) \$ 39.990

**Lo más popular en Perfumería y Belleza**

LANCÔME Set Rénergie Multi-Lift Rutina Resfirmante	LANCÔME Set Sérums Advanced Génifique 20 ml +	RALPH LAUREN Trilogía Polo Blue	RALPH LAUREN Polo Blue EDP-40ml	CLINIQUE Set Clinique SOS Get The Most Glow	LANCÔME Set Miniaturas Lancome
\$ 39.990 (Oferta) \$ 90.990	\$ 16.990 (Oferta) \$ 32.990	\$ 39.990 (Oferta) \$ 89.990	\$ 19.990 (Oferta) \$ 29.990	\$ 16.990 (Oferta) \$ 49.990	\$ 24.990

**Lo más popular en Tecnología y Hogar**

RECCO LCD 32" RLED-L32D1200SMTHD Smart	SAMSUNG LED SAMSUNG 43" NU7090 UHD 4K Smart	WURDEN Estufa a Gas RF 5700CL-C	ACER Notebook Aspire Intel Core i3-1005G1 12GB	APPLE Macbook Air 13.3" Intel Core i5 8GB RAM 128GB	SAMSUNG Smartphone Galaxy A30s 128GB
\$ 119.990 (Oferta) \$ 169.990	\$ 229.990 (Oferta) \$ 279.990	\$ 84.990 \$ 89.990 (Oferta) \$ 99.990	\$ 449.990 (Oferta) \$ 479.990	\$ 799.990	\$ 189.990 \$ 259.990

Figura 5.14: Ejemplo de recomendaciones personalizadas desplegadas en el sitio.

Las recomendaciones se basaron en productos “populares”, que fueron construidos aprovechando tanto la información de compra como de navegación de los productos. Para generar estos rankings de productos populares se consideraron todos los artículos de las 18 catego-

rías. Por el lado de las compras se contabilizaron la cantidad de unidades vendidas de cada producto (y todas sus variaciones) durante una ventana de 15 días móviles, y por el lado de la navegación se utilizó el número de visitas a cada producto en una ventana de 3 días móviles. Dado que cada iteración de este proceso generaría dos rankings de productos por categoría, para obtener un único ordenamiento en cada una de ellas se combinaron ambos rankings promediando las posiciones relativas de cada producto dentro de su categoría correspondiente, asignando el mismo peso a la posición en base a compras y a la basada en navegación. De esta forma si un producto es muy comprado y a la vez muy visitado se mantendrá arriba en el ranking. Esto para lograr equilibrar recomendaciones con alta probabilidad de conversión, pero que además sean atractivos por otras razones para los usuarios.

Un ejemplo de los carruseles generados para cada uno de los segmentos se observa en la Figura 5.14, donde se muestran las recomendaciones expuestas a cuatro grupos en un momento del período de experimentación. Se debe recordar que cada usuario visualizó solo uno de estos carruseles de productos recomendados durante las visitas que realizó al sitio web.

De los productos recomendados durante el período de experimentación, se registra por categoría la cantidad de artículos distintos que fueron mostrados, el precio promedio de ellos y el porcentaje de usuarios asignados a cada categoría. Estos valores pueden observarse en la Tabla 5.15. Esta tabla muestra la diferencia de precios de las recomendaciones dependiendo de cada categoría, lo que influirá en la decisión de compra de los usuarios, más allá del nivel de afinidad que cada uno pueda tener con la categoría. Como se observa en el porcentaje de usuarios asignados, gran parte de las recomendaciones personalizadas como conjunto estarán dominadas por un par de categorías. Si se suman los usuarios de las 4 categorías con mayor participación, se tiene que representan poco más del 60% del total. De todas formas habrá cierto equilibrio en el número de usuarios asignados a categorías de precio bajo y categorías de precios altos.

### 5.3.3. Resultados experimentales

Concluida la segunda fase de experimentación, en donde se implementaron los segmentos y las recomendaciones personalizadas correspondientes a cada uno de ellos, se recopilaron y analizaron los resultados capturados por la plataforma de analítica. Durante los 21 días de duración de este experimento se recopiló la navegación de alrededor de 3.300.000 sesiones, equivalentes a cerca de 1.600.000 visitantes que pudieron ser individualizados por la plataforma (los cuales no necesariamente representan la cantidad de usuarios únicos que navegaron en este período).

Tal como en el primer experimento descrito con anterioridad, se eliminaron del análisis aquellas sesiones que fueron registradas con más de un grupo experimental, así como alrededor del 10% de las sesiones que cumplieron con la condición de tener una única página vista, ya que fueron consideradas como rebotes. Además se debe recordar que dada la naturaleza del experimento en el que el sistema debe lograr reconocer al usuario que navega (con el objetivo de determinar su sexo y segmento asignado), la mayoría de las sesiones fueron realizadas por visitantes con su sesión iniciada, las cuales representan alrededor del 33% de las sesiones totales registradas en la página de inicio del sitio.

Tabla 5.15: Estadísticas de los productos recomendados por categoría.

Categoría	N° Productos Recomendados	Precio Promedio	% de Usuarios Asignados
J01	83	\$25.200	3,13 %
J02	55	\$25.372	1,48 %
J03_O	70	\$211.441	2,46 %
J03_VM	88	\$28.035	1,18 %
J03_VH	56	\$36.933	1,31 %
J04	151	\$21.121	10,33 %
J05	127	\$17.056	6,65 %
J06	89	\$11.608	3,08 %
J07	88	\$13.937	2,47 %
J08	94	\$45.601	5,65 %
J09	140	\$12.467	13,33 %
J10_M	141	\$46.220	1,86 %
J10_H	66	\$48.828	5,50 %
J11	98	\$269.924	31,87 %
J12	66	\$32.898	3,03 %
J13	69	\$301.526	3,14 %
J14	44	\$234.294	1,79 %
J15	67	\$23.457	1,72 %

Como ya se mencionó en el diseño experimental, la estrategia de recomendación personalizada utilizando segmentos fue contrastada no sólo con un grupo de control (al cual no se le expuso a ningún carrusel de recomendaciones), si no que también se expuso a un tercio de las sesiones a los productos más comprados en categorías blandas, que fue la estrategia de recomendación que mejores resultados entregó en la fase previa en términos de la tasa de conversión. Por decisiones de la empresa, y dado que el experimento aplicaría sobre usuarios que el sitio web podría identificar, se decidió diferenciar los productos más comprados mostrados en categorías de mujeres y hombres, siempre que se tuviera información del sexo del usuario.

En cuanto al dispositivo utilizado por los visitantes registrados, se tiene que el 44% de las sesiones analizadas fueron registradas desde la versión de escritorio del sitio (PC o Mac), mientras que el restante 56% lo hizo a través de la versión web para dispositivos móviles, excluyendo del análisis aquellas visitas realizadas a través de la aplicación para smartphones.

### 5.3.3.1. Resultados generales de conversión y gasto

De forma análoga a la primera fase de experimentación, se comienza analizando los resultados en base a la conversión (proporción de sesiones con compra) y el gasto general de los grupos experimentales, comparándolos a su vez con el grupo control para determinar y validar si hubo un impacto positivo de colocar los carruseles de recomendaciones en este nuevo período. Para esto se resume en la Tabla 5.16 las tasas de conversión (TC) y el gasto promedio por sesión (ARPS) de los grupos experimentales. Los incrementales (en paréntesis)

y los p-valores calculados son con respecto al grupo control a un 95 % de confianza.

Tabla 5.16: Resultados de conversión y gasto tras segunda fase de experimentación.

Grupo	TC		Z	p-valor	ARPS		t	p-valor
Más Comprados Blandas	5,05 %	(+0,80 %)	1,0271	0,1522	\$5.278	(-\$5)	0,0699	0,4722
Personalizado	5,01 %	(+0,01 %)	0,0122	0,4951	\$5.311	(\$28)	0,3857	0,3498
Control	5,01 %				\$5.283			

La Tabla 5.16 muestra un aumento en la tasa de conversión de los grupos en presencia del carrusel, aunque este incremento no resultó ser significativo en ninguno de los dos grupos. En términos generales, se observa que el mayor aumento en la conversión se registró en el grupo expuesto a las recomendaciones de productos más comprados, modificados por sexo del usuario. A pesar de no encontrarse significancia estadística a un 95 % de confianza, los incrementales positivos que se obtienen concuerdan con los resultados de la primera fase de experimentación que muestran incrementales positivos al exponer a los visitantes a carruseles de recomendaciones.

A pesar de la mayor conversión en el grupo expuesto a los productos más comprados en categorías blandas, si se observa el gasto promedio realizado por sesión se tiene que este grupo tuvo un decrecimiento con respecto al grupo control. Mientras tanto, el grupo expuesto a recomendaciones personalizadas en base a los segmentos generados obtuvo un incremento en el gasto promedio por sesión, mostrando nuevamente que la opción que más aumenta conversión no necesariamente será la que más aumente el ingreso total. De todas formas la alta variabilidad de esta métrica, así como la escala de la diferencia encontrada no permiten concluir que esta diferencia es estadísticamente significativa.

Más allá de la significancia estadística determinada, los resultados tienen sentido si se recuerda que la métrica de ARPS se puede descomponer como el producto de la tasa de conversión (TC) y el gasto promedio por sesión con compra. Con esto en consideración, un mayor ARPS, a pesar de una menor tasa de conversión, puede deberse a que los visitantes adicionales que compran terminan gastando una mayor cantidad de dinero (a pesar de ser menores en número). Mientras tanto, un decrecimiento en el ingreso, aún con un incremental positivo en la conversión, puede indicar que ciertos clientes terminaron gastando menos de lo que hubieran hecho en ausencia del carrusel de productos. Lo anterior concuerda en este caso dado que en promedio los productos mostrados en las recomendaciones personalizadas por segmentos tienen un precio más elevado que los productos más comprados en categorías blandas.

### 5.3.3.2. Resultados de uso y post-click

Los resultados generales de conversión y gasto sirven para determinar la efectividad de los carruseles de recomendación en comparación con la ausencia total de estos sistemas en la

página, pero también será importante entender el efecto directo que las variaciones de sistemas de recomendación tuvieron sobre la navegación de los visitantes, a través de los clicks realizados y su comportamiento posterior.

De esta forma, en la Tabla 5.17 se resumen las tasas de uso obtenidas por ambos grupos experimentales, entendida como la proporción de sesiones con click sobre alguna de las recomendaciones, por sobre la cantidad total de sesiones del grupo. En esta tabla se incluye la diferencia porcentual ( $\Delta\%$ ) usando como base (denominador) el carrusel con menor uso entre ambas lógicas. Tal como ya se mencionó en la primera fase de experimentación, se estima que la cantidad de visitantes expuestos reales (los que efectivamente visualizaron la sección del sitio donde estaban localizados los recomendadores) es sólo una proporción menor del total de visitantes registrados, pero conocer los verdaderos expuestos es una tarea imposible de realizar con seguridad a través de las herramientas de analítica actuales de la empresa. Aún así la tasa de uso calculada sobre el total de sesiones debiese indicar la misma tendencia que si se hiciera sobre los expuestos reales.

Tabla 5.17: Comparación de la tasa de uso entre grupo expuesto a recomendaciones de categorías blandas y personalizadas por segmento.

Grupo	TU	$\Delta\%$ de Uso	Z	p-valor
Más Comprados Blandas	0,73 %			
Personalizado	0,84 %	(+14, 48 %)	8,8323	0,0000

Los resultados para las tasas de la Tabla 5.17 muestran que el uso del carrusel de recomendaciones fue significativamente mayor en el caso del grupo personalizado por segmentos, dado que registró más de un 14 % de incremento de sesiones con click que el grupo expuesto a recomendaciones en categorías blandas. Lo anterior permite suponer que las recomendaciones mostradas al utilizar la lógica personalizada generaron un mayor interés en los visitantes expuestos que simplemente utilizar los productos más comprados en categorías blandas, diferenciados por sexo identificado. Se debe considerar que este mayor interés por parte de los usuarios puede sugerir una mayor probabilidad de generar una venta adicional o un incremento en el gasto.

Ahora bien, para completar el análisis anterior se deben considerar los resultados obtenidos para la conversión y el gasto posterior al click, que se resumen en la Tabla 5.18, para entender que sistema resultó ser más efectivo luego del uso; incluyendo la tasa de conversión post-click (considerando compras en el mismo SKU en el que se hizo click o sin esta restricción), la proporción de sesiones que agregan al carro y el gasto promedio en productos clickeados.

Los resultados post-click muestran que la mayor proporción de visitantes que compraron un producto luego de haber hecho uso del carrusel de recomendaciones (ya sea comprando el mismo producto clickeado o algún otro) se obtuvo en el grupo expuesto a los productos más comprados en categorías blandas. En este grupo, alrededor de un 43 % más de visitantes se convirtieron en compradores, en comparación a las conversiones del grupo expuesto a recomendaciones personalizadas por segmento. Esta diferencia es menor si se consideran las

proporciones de sesiones que agregan al carro, en donde el grupo expuesto a los productos más comprados en categorías blandas obtuvo alrededor de un 20 % más de estas conversiones que el grupo personalizado por segmentos. Esta métrica también debe tenerse en consideración dado que denota el interés de los clientes por adquirir el producto y la probabilidad de conversión en sesiones posteriores.

Tabla 5.18: Resultados de métricas posteriores al click en recomendaciones personalizadas.

Grupo	TC post-click	TC post-click (mismo SKU)	Al Carro (mismo SKU)	Gasto post-click (mismo SKU)
Más Comprados Blandas	3,60 %	1,29 %	13,63 %	\$24.009
Personalizado	2,92 %	0,87 %	11,39 %	\$53.131

A pesar de que los resultados anteriores muestran una mayor conversión post-click con las recomendaciones de lo más comprado en categorías blandas, también se debe considerar que este grupo tuvo un menor gasto promedio en productos del carrusel. Esto es debido a que en ciertos segmentos (del grupo personalizado) los valores de los productos recomendados superaban varias veces los precios de los productos recomendados en categorías blandas. En concreto, los expuestos a las recomendaciones personalizadas gastaron poco más del doble en promedio en productos clickeados que los expuestos a recomendaciones de categorías blandas.

En definitiva, dadas las distintas tasas y gasto entre los distintos grupos, para comparar el efecto medible directo de los carruseles se puede estimar el ingreso adicional esperado por sesión multiplicando la tasa de uso, conversión post-click (en el mismo SKU con click) y el gasto promedio post-click. El resultado de esta multiplicación por grupo se muestra en la Tabla 5.19. Aquí se puede observar que considerando las métricas antes mencionadas, los expuestos a las recomendaciones personalizadas tendrán un ingreso adicional esperado mayor que los expuestos a las recomendaciones de productos más comprados en categorías blandas. Por lo tanto, considerando una cantidad igual de sesiones expuestas a cada carrusel de recomendaciones, el aporte monetario (de forma directa) del sistema personalizado en base a segmentos será mayor que el dinero aportado por el sistema basado en lo más comprado de categorías blandas.

Tabla 5.19: Estimación del ingreso adicional esperado por sesión aportado directamente por las recomendaciones personalizadas.

Grupo	Ingreso adicional esperado por Sesión (directo)
Más Comprados Blandas	\$2,3
Personalizado	\$3,9

# Capítulo 6

## Conclusiones

El trabajo descrito nace de la oportunidad presentada a la empresa de incorporar un nuevo elemento basado en recomendaciones de productos para incrementar la conversión y la venta de su sitio web, aprovechando de incluir una nueva vía por la que los visitantes de la página puedan navegar a través de las categorías y productos que sean de su interés. Dado que el sitio no contaba con ningún sistema similar, el trabajo realizado permite entender los primeros conceptos acerca del uso de carruseles de recomendaciones dentro del sitio, así como del impacto de ellos sobre los usuarios. Se espera que los hallazgos y sistemas generados en este informe sirvan como punto de partida y base de comparación para crear y evaluar sistemas de recomendación más sofisticados en el futuro.

En general, los resultados de la experimentación entregan conclusiones positivas acerca de la implementación de recomendaciones en forma de carruseles de productos dentro del sitio. Se pudo evidenciar cómo lógicas de recomendación sencillas basadas en los artículos más vendidos o más vistos lograron generar incrementales estadísticamente significativos sobre la conversión y el ingreso. Si bien el uso de los carruseles dentro del sitio es bajo en comparación a otros elementos con mayor exposición, ya solo con lograr que alrededor del 1% de los visitantes que ingresan a la página se vean interesados en las recomendaciones se pueden generar ganancias adicionales, producto de un mejor aprovechamiento de un espacio poco considerado, redireccionando la navegación de ciertos usuarios hacia una sección más pertinente o con mayor probabilidad de conversión. Estos resultados evidencian que la inclusión de uno, dos o más carruseles de recomendación pueden superar el rendimiento de la versión actual del sitio.

En particular, el análisis realizado muestra en primera instancia que en base a la maximización del ingreso adicional los mejores resultados se obtienen al incluir tanto recomendaciones de productos de categorías duras (precio elevado, baja frecuencia de compra) como productos de categorías blandas (precio reducido, alta frecuencia de compra). Como se pudo observar en la utilización de ambos carruseles de recomendación, los grupos de clientes que se ven interesados en cada conjunto de categorías son distintos. Si bien a través del análisis exploratorio se pudo evidenciar que las distintas secciones del sitio “compiten” entre sí por los clicks de los visitantes (por esto mientras más elementos antecedan menor será el uso), en definitiva la inclusión de ambos carruseles tuvo más bien un efecto complementario. Finalmente lo que logra la inclusión de ambos carruseles es aumentar el espectro de productos a mostrar y por lo tanto la probabilidad de que el usuario encuentre algo de interés para él.

Por otro lado, la distinción del método de recomendación resultó en mayores incrementales al utilizar el interés del público medido a través de la navegación por sobre las compras. Esto estaría explicado debido a que con el primer método se obtuvieron productos de un precio superior que al utilizar el segundo, pero aún así se recomendaron productos interesantes para los visitantes. En definitiva, la alta cantidad de visualizaciones de un producto estaría implícitamente relacionada con atributos positivos del mismo, pero dando algo menos de importancia a su precio (pudiendo destacar más por su calidad, renombre, novedad, etc.). Este hallazgo sugiere que al momento de realizar recomendaciones futuras se debe buscar un equilibrio entre productos de interés para el usuario, pero que sigan generando un ingreso monetario considerable para la empresa. Así mismo, este hallazgo sugiere la importancia de definir una métrica de evaluación que considere el equilibrio antes mencionado, como es el caso del gasto promedio por sesión o visitante, dado que la tasa de conversión por sí sola no permitirá entender completamente el impacto del sistema probado.

Los resultados del trabajo realizado también sugieren que la complementariedad antes mencionada puede ser aprovechada de mejor forma si se logra caracterizar a los usuarios con el objetivo de determinar las categorías más (o menos) interesantes para ellos. Si bien los resultados muestran que una buena estrategia en caso de no tener información de los clientes puede ser aumentar el número y la amplitud de recomendaciones, puede existir un punto en que esta mayor cantidad de elementos deje de tener un aporte significativo o incluso empeore el desempeño del sitio. Es aquí donde la personalización puede jugar un rol importante al mejorar el enfoque y la eficiencia de las recomendaciones, pudiendo lograr un incremento igual o mayor con un menor espacio utilizado. Los resultados son consistentes en mostrar que recomendaciones personalizadas, incluso utilizando agrupaciones muy generales de clientes, pueden invocar un mayor interés por parte de los visitantes, resultando finalmente en una mayor probabilidad de generar un ingreso adicional.

En este caso se hizo uso de una personalización basada en segmentos de acuerdo a las limitantes técnicas de la empresa, por alcances de tiempo y bajo el supuesto que un número considerable de visitantes quedarían correctamente descritos por ella. Aunque es claro que la calidad y el nivel de detalle de la personalización tendrá un impacto en el desempeño final de las recomendaciones a probar. Si bien el método utilizado en esta ocasión para personalizar las recomendaciones puede considerarse general, incluso con este nivel de generalización se evidencian resultados que muestran el valor de utilizar la información de los usuarios, por sobre un sistema de recomendación idéntico para todos. De esta forma la personalización puede mejorarse ya sea utilizando un mayor número de información, ajustando el número de grupos o el nivel de agregación de las categorías a utilizar, aumentando el número de artículos a recomendar o utilizando otro sistema para definir afinidad hacia categorías o productos.

Los resultados empíricos también muestran que la efectividad de los carruseles de recomendaciones probados, o cualquier otro carrusel de recomendaciones que sea implementado en el sitio, estará limitada por el número de expuestos reales, que dependerá a su vez de la posición del elemento dentro de la página. Idealmente, la comparación entre variaciones al evaluar distintos sistemas de recomendación debiera realizarse solamente sobre los verdaderos expuestos, con el objetivo de reducir el ruido en el cálculo de las métricas y evitar que el efecto se diluya al considerar un denominador más grande. Como ya se mencionó antes en el desarrollo de este trabajo, las actuales herramientas de analítica de la empresa no permiten

identificar a estos verdaderos expuestos, pero será clave tenerlos en consideración para estimar con mayor precisión el potencial efecto que las recomendaciones implementadas pueden tener sobre la venta total del sitio.

Finalmente, a través del desarrollo del trabajo se pudo evidenciar como existen una gran cantidad de parámetros que se pueden definir al momento de diseñar sistemas de recomendación para un sitio de comercio electrónico. Por mencionar algunos se tiene el número de artículos a recomendar, la diversidad de ellos (en términos de categorías), la posición de los elementos, el nivel de personalización a utilizar, el precio de los productos, entre muchas otras. Este cantidad enorme de posibles configuraciones a utilizar hace imposible la tarea de encontrar una que sea óptima, en el sentido que maximiza de manera absoluta la conversión o el ingreso. Por lo tanto, la mejor forma de encontrar buenas configuraciones será a través de una heurística que vaya buscando incrementales por sobre la versión actual, teniendo en consideración que mientras más agresivos sean los cambios realizados, mayor es la probabilidad de poder evidenciar un efecto tras una semanas de experimentación.

En base a estas conclusiones, se propone a la empresa la implementación en la página de inicio del sitio de al menos dos carruseles de recomendaciones, enfocados en las categorías de productos que más ingresos reporten tanto en categorías duras como en blandas para lograr una mayor amplitud, utilizando como método de elección de los productos a recomendar el número de visualizaciones de cada artículo. Se propone utilizar estas recomendaciones como punto de inicio para probar nuevas implementaciones en el futuro. Por ejemplo, se pueden realizar nuevas pruebas aumentando el número de carruseles de recomendaciones, hasta determinar el punto en que no se registra un incremental o este es negativo. Se debe considerar que a pesar de que el incremento en las ventas puede ser reducido en comparación con intervenciones en otras partes del sitio, el costo de mantener estos elementos automatizados es bajo, y consiste principalmente en los recursos computacionales utilizados y el tiempo de monitoreo del sistema que se sugiere realizar de forma periódica. Esta ventaja permite considerar la posibilidad de utilizar más de esto elementos como reemplazo de otros que requieren un mayor esfuerzo en diseño e implementación.

La configuración anteriormente mencionada se propone para la mayoría de los usuarios, pero siempre que se tenga información sobre la persona que se encuentra navegando, se sugiere implementar un grado de personalización intentando definir las categorías que más pueden ser de su interés. Aunque el trabajo muestra resultados positivos al utilizar un único carrusel de recomendaciones personalizadas, se sugiere probar la utilización de más carruseles, definiendo no sólo una, si no dos o más categorías con alta afinidad por cliente, desplegándolas de manera ordenada en el sitio. De esta forma, es más probable realizar una recomendación que despierte interés en cada visitante, y se mitiga en cierta forma el riesgo de que la asignación de afinidad se haya hecho de manera incorrecta.

# Capítulo 7

## Trabajo Futuro

Como extensión del trabajo realizado, se propone la utilización de otros modelos que permitan definir productos recomendados, ya sean estos personalizados o no. Un ejemplo muy utilizado en gran número de sistemas de recomendación en la actualidad son los modelos de filtros colaborativos, que establecen similitudes entre clientes o usuarios en base a sus preferencias explícitas o implícitas. Una de las ventajas de estos modelos es que permiten introducir un grado de personalización sin definir a priori un número de segmentos. Ahora bien, una de las desventajas es que no funcionan bien cuando los datos son dispersos (como al existir un gran número de productos), lo que se puede solucionar si las preferencias se calculan sobre categorías y no sobre productos. Al realizar este procedimiento se pueden obtener un gran número de categorías a recomendar, dependiendo del nivel de agrupación seleccionado, y se debe tener en cuenta las capacidades técnicas para desplegar este gran número de recomendaciones diferentes. En caso de desarrollar un sistema con las condiciones antes descritas, se deben tener consideraciones similares a las analizadas en este trabajo sobre el método de elección de los artículos a recomendar y el número de elementos a mostrar.

Otros análisis adicionales que se pueden realizar consisten en medir los resultados de caruseles de recomendación a un plazo mayor y aumentando el nivel de seguimiento de las unidades experimentales, intentando determinar efectos más duraderos en el tiempo, y que difícilmente podrían evaluarse en el transcurso de dos o tres semanas. Algunos de estos análisis pueden consistir en el nivel de satisfacción del usuario o el nivel de recompra. También se puede considerar una evaluación más exhaustiva e incluso cualitativa de los sistemas de recomendación sobre ciertos clientes, para entender en mucho más detalle los factores que determinan que la implementación de un sistema de recomendación en el sitio sea exitoso.

Por último, se propone en trabajos futuros considerar la utilización de métodos bayesianos para la evaluación de los experimentos controlados en línea, que permitan responder en términos de probabilidades qué variación es mejor que la otra [12]. Para esto, se debe entender muy bien las implicancias de la elección de un prior y los supuestos que están detrás del análisis. Las ventajas de analizar los experimentos desde un enfoque bayesiano es que en general necesitará un tamaño muestral menor, que al entregar probabilidades de mejoría los resultados serán más fáciles de interpretar, y que incluso en períodos cortos de experimentación entregará información interpretable (no así necesariamente concluyente).

# Bibliografía

- [1] Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., Yin, F. (2010). Empirical Analysis of the Impact of Recommender Systems on Sales. *Journal of Management Information Systems*, 27(2), 159-188.
- [2] MacKenzie, I. (2018, 14 febrero). How retailers can keep up with consumers. Recuperado de <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>
- [3] Schafer, J. B., Konstan, J., Riedi, J. (1999). Recommender systems in e-commerce. *Proceedings of the 1st ACM conference on Electronic commerce - EC '99*, 158-166.
- [4] Aggarwal, C. C. (2016). *Recommender Systems*. Springer Publishing.
- [5] Billsus, D., Pazzani, M. J. (1998). Learning Collaborative Information Filters. *Proceedings of the Fifteenth International Conference on Machine Learning*, 46-54.
- [6] Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., Bouras, A. (2014). A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis. *IEEE Transactions on Emerging Topics in Computing*, 2(3), 267-279.
- [7] Baarsch, J., Celebi, M. E. (2012). Investigation of internal validity measures for K-means clustering. *Proceedings of the international multiconference of engineers and computer scientists*, 1, s.n.
- [8] Berger, P. D., Maurer, R. E., Celli, G. B. (2017). *Experimental Design*. Springer Publishing.
- [9] Kohavi, R., Longbotham, R., Sommerfield, D., Henne, R. M. (2008). Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1), 140-181.
- [10] Dmitriev, P., Gupta, S., Kim, D. W., Vaz, G. (2017). A Dirty Dozen: Twelve Common Metric Interpretation Pitfalls in Online Controlled Experiments . *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1427-1436.
- [11] Wirth, R. Hipp, Jochen. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*.
- [12] Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2), 573-603.

# Anexo A

## Visualización de las recomendaciones

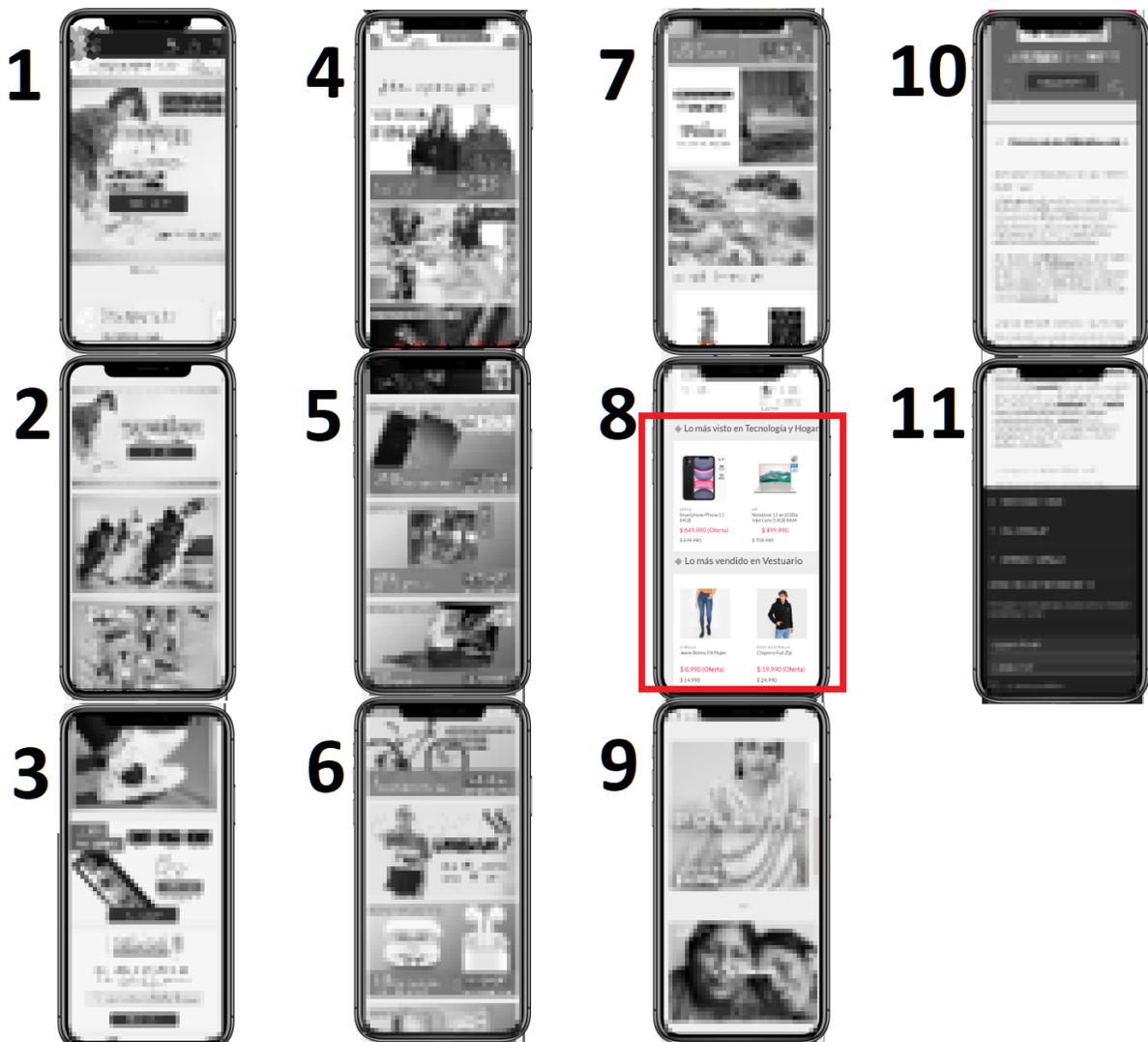


Figura A.1: Despliegue del sitio web en dispositivos móviles y distribución de las vistas.

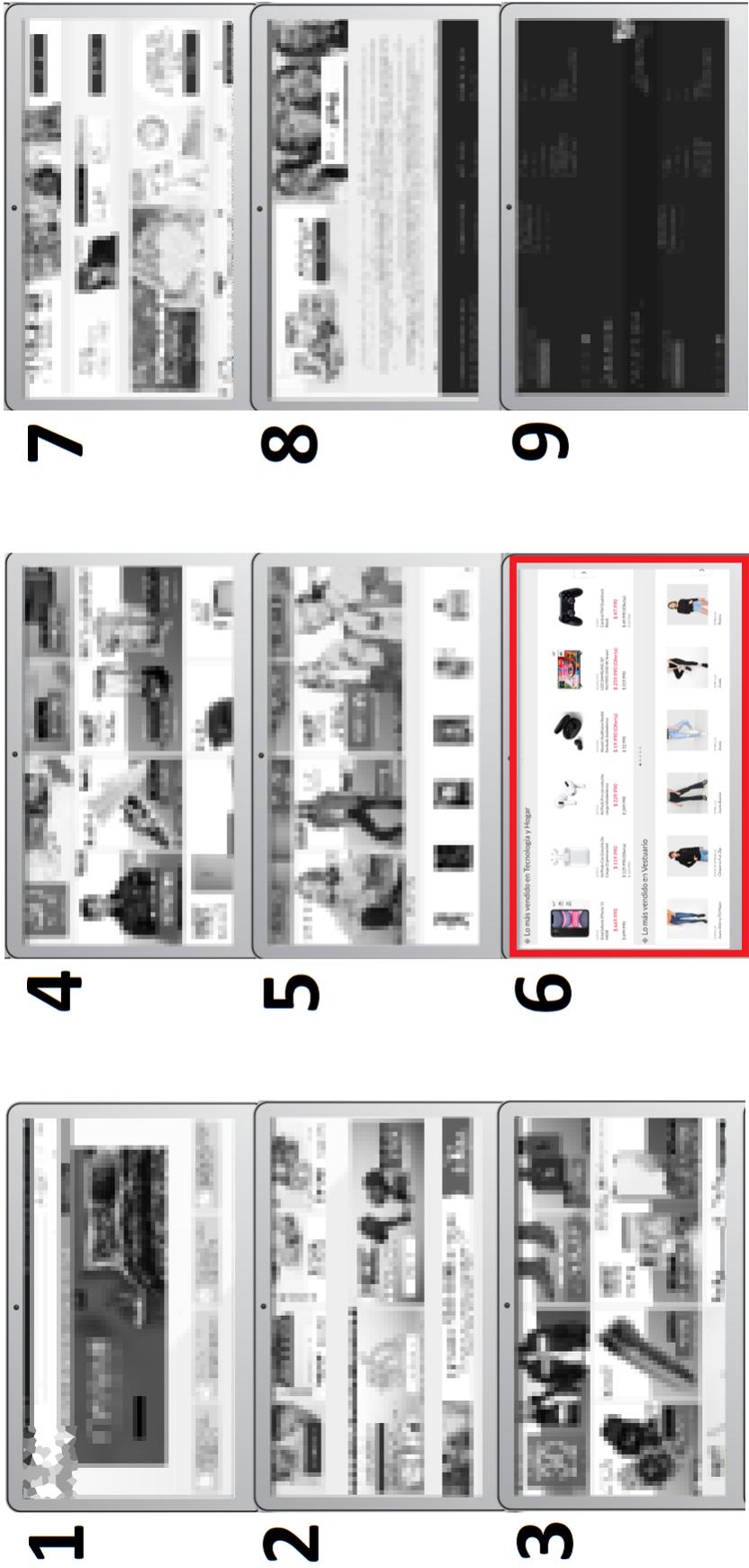


Figura A.2: Despliegue del sitio web en dispositivos de escritorio y distribución de las vistas.

# Anexo B

## Análisis Exploratorio

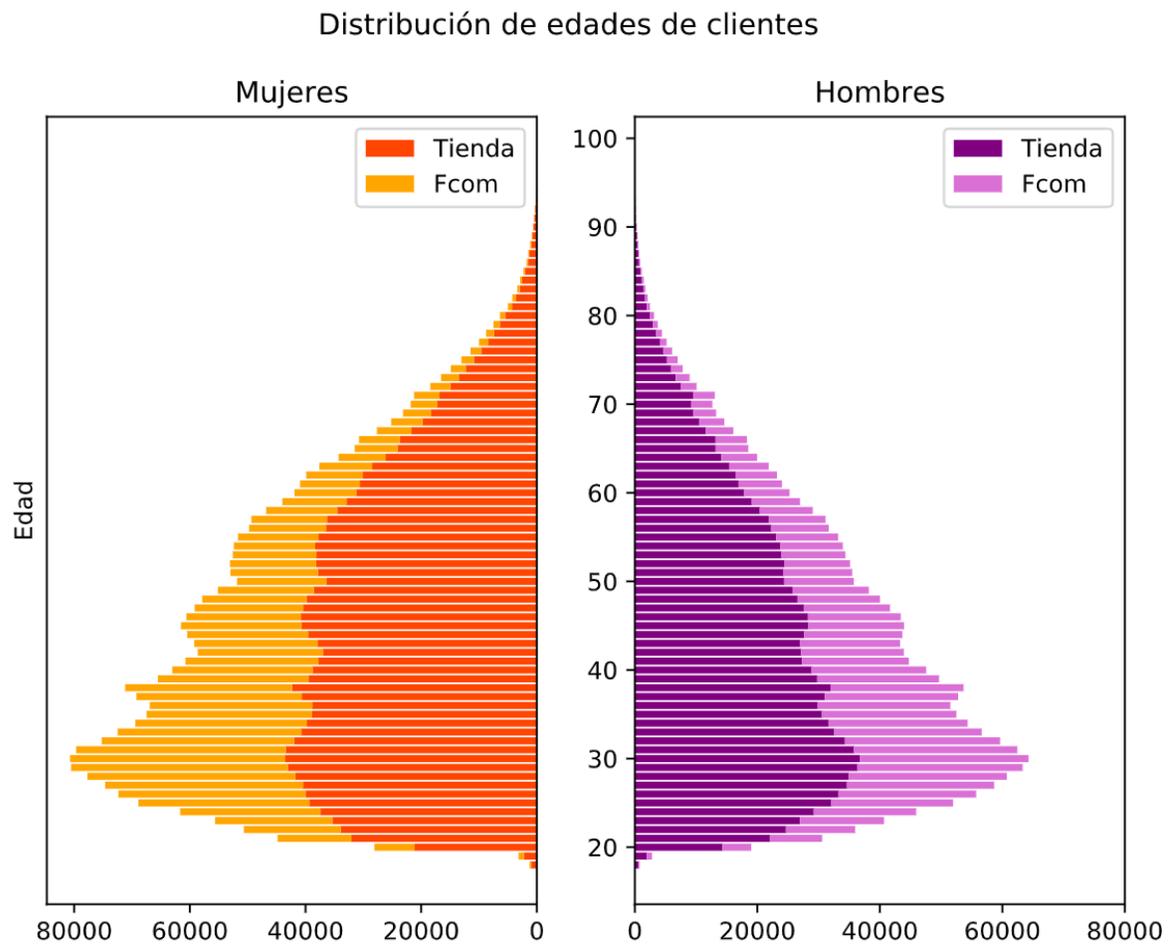


Figura B.1: Pirámide demográfica de los clientes activos.

Precio promedio y unidades vendidas por nivel de agrupación

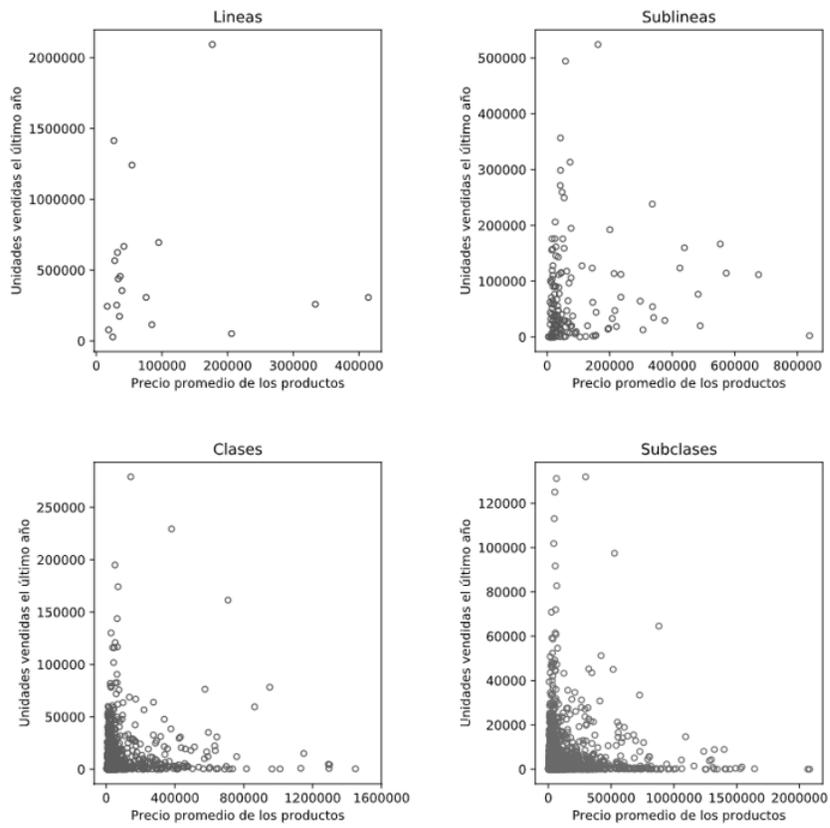


Figura B.2: Precio promedio y unidades vendidas por producto para las distintas agrupaciones jerárquicas.

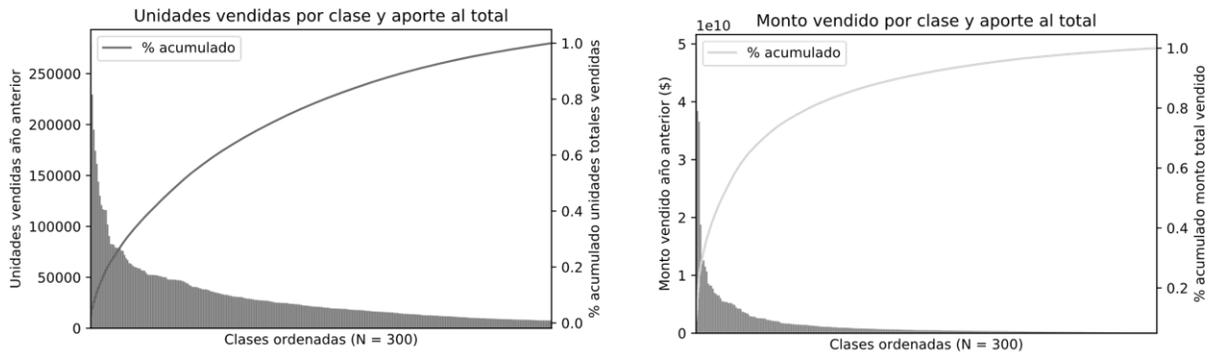


Figura B.3: Distribución de las unidades y el monto vendido dentro de cada clase.

# Anexo C

## Resultados Segmentación

Tabla C.1: Cuantiles para la distribución del gasto total por cliente tras eliminación de outliers.

Percentil	Gasto
Mínimo	\$ 490
25 %	\$ 30.980
50 %	\$ 88.840
75 %	\$ 268.880
Máximo	\$ 10.589.900

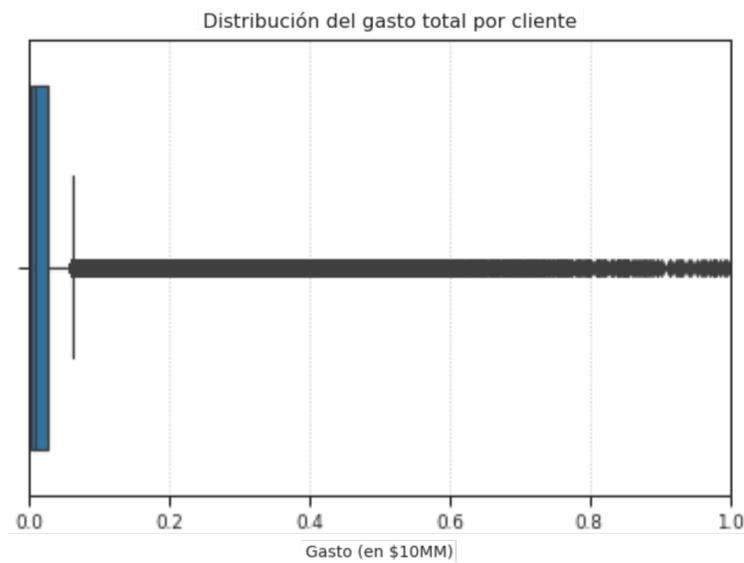


Figura C.1: Distribución del gasto total por cliente tras eliminación de outliers.

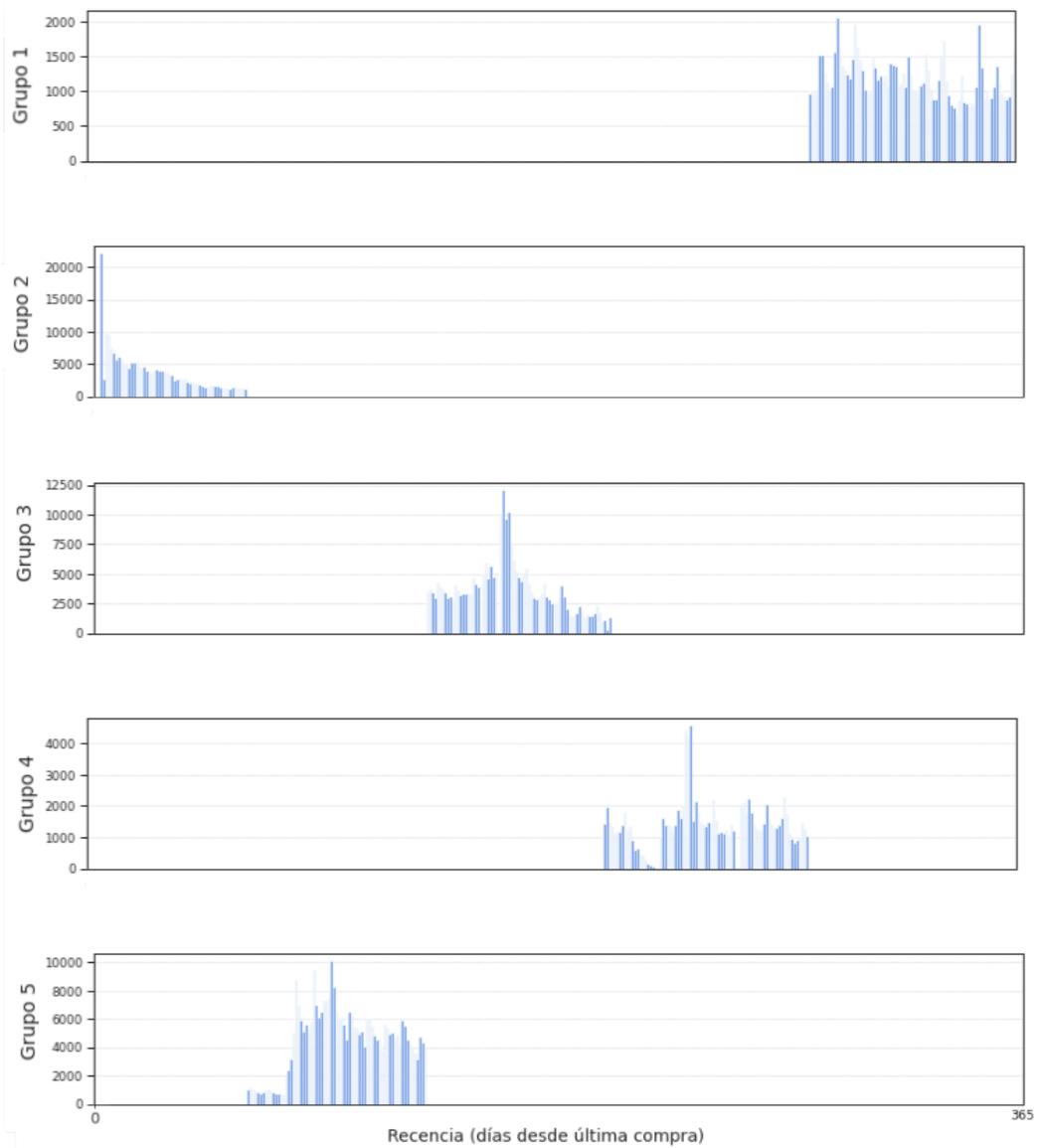


Figura C.2: Distribución de la recencia de compra al realizar segmentación en base a RFM.

		CENTROS POR CATEGORÍA																				TAMAÑO		
	J01	J02	J03_VH	J03_VM	J03_O	J04	J05	J06	J07	J08	J09	J10_H	J10_M	J11	J12	J13	J14	J15	J16	J17	J18	J21	J32	
1	1,8%	1,3%	0,5%	0,3%	0,4%	2,2%	1,8%	1,3%	1,1%	81,4%	2,4%	0,7%	1,1%	2,0%	0,5%	0,1%	0,1%	0,5%	0,1%	0,2%	0,3%	0,0%	0,0%	3,7%
2	0,7%	0,5%	0,2%	0,1%	0,3%	0,6%	0,5%	0,3%	0,3%	0,6%	1,1%	0,5%	0,5%	92,8%	0,2%	0,1%	0,2%	0,4%	0,1%	0,1%	0,1%	0,0%	0,0%	26,6%
3	3,2%	2,7%	1,5%	0,2%	0,8%	1,0%	0,9%	0,4%	0,5%	0,8%	2,1%	82,2%	1,3%	1,6%	0,3%	0,0%	0,0%	0,3%	0,0%	0,1%	0,1%	0,0%	0,0%	3,7%
4	11,2%	9,9%	4,1%	1,8%	2,7%	6,8%	5,5%	4,0%	4,3%	8,6%	8,0%	7,1%	4,5%	8,6%	2,8%	1,1%	1,3%	2,5%	2,5%	1,4%	0,7%	0,2%	0,3%	8,0%
5	5,0%	82,7%	1,6%	0,3%	0,9%	1,1%	1,4%	0,5%	0,4%	0,7%	1,9%	1,2%	0,4%	1,1%	0,2%	0,0%	0,0%	0,3%	0,0%	0,1%	0,1%	0,0%	0,0%	4,2%
6	0,4%	0,3%	0,1%	0,2%	0,2%	1,1%	0,8%	0,6%	0,4%	0,4%	0,9%	0,2%	93,6%	0,4%	0,1%	0,0%	0,0%	0,2%	0,0%	0,1%	0,1%	0,0%	0,0%	6,5%
7	0,3%	0,2%	0,1%	0,1%	0,1%	0,3%	0,4%	0,3%	0,1%	0,2%	96,8%	0,1%	0,1%	0,3%	0,1%	0,0%	0,0%	0,2%	0,0%	0,1%	0,1%	0,0%	0,0%	13,2%
8	0,4%	0,2%	0,1%	0,1%	0,1%	94,9%	0,9%	1,0%	0,4%	0,4%	0,6%	0,1%	0,3%	0,2%	0,1%	0,0%	0,0%	0,2%	0,0%	0,1%	0,1%	0,0%	0,0%	3,7%
9	0,2%	0,2%	0,1%	0,1%	0,1%	0,7%	95,7%	0,9%	0,4%	0,3%	0,6%	0,0%	0,2%	0,2%	0,1%	0,0%	0,0%	0,1%	0,0%	0,0%	0,1%	0,0%	0,0%	8,5%
10	1,7%	1,3%	0,5%	0,3%	0,4%	2,3%	2,1%	1,3%	80,2%	2,1%	2,3%	0,8%	1,3%	1,9%	0,4%	0,1%	0,1%	0,5%	0,1%	0,2%	0,2%	0,0%	0,0%	2,0%
11	2,9%	3,7%	80,2%	1,1%	2,3%	1,0%	1,0%	0,5%	0,4%	0,7%	2,2%	1,5%	0,5%	1,2%	0,2%	0,0%	0,0%	0,3%	0,0%	0,1%	0,1%	0,0%	0,0%	1,7%
12	1,7%	1,4%	0,9%	0,4%	79,0%	1,3%	1,3%	0,6%	0,7%	1,2%	2,9%	1,3%	1,1%	4,1%	0,4%	0,3%	0,3%	0,6%	0,1%	0,2%	0,2%	0,1%	0,0%	1,8%
13	84,8%	3,2%	0,8%	0,2%	0,8%	1,4%	0,8%	0,7%	0,5%	0,9%	1,7%	1,4%	0,4%	1,2%	0,3%	0,0%	0,0%	0,4%	0,0%	0,1%	0,2%	0,0%	0,0%	4,0%
14	1,7%	1,4%	0,6%	0,3%	1,0%	2,1%	1,7%	0,9%	0,9%	1,8%	3,7%	1,4%	1,8%	8,5%	1,6%	1,3%	67,8%	0,8%	0,2%	0,2%	0,1%	0,0%	0,0%	1,9%
15	1,8%	1,0%	0,4%	0,2%	0,4%	2,2%	1,4%	1,4%	0,7%	1,3%	3,1%	0,6%	0,9%	2,1%	78,7%	0,1%	0,6%	1,9%	0,2%	0,6%	0,4%	0,0%	0,1%	1,0%
16	1,7%	1,3%	0,6%	0,3%	1,0%	2,2%	1,8%	0,8%	1,0%	1,7%	3,6%	1,3%	1,7%	7,6%	1,0%	68,4%	1,6%	1,1%	0,5%	0,5%	0,1%	0,0%	0,1%	1,3%
17	1,5%	1,6%	2,9%	76,3%	1,4%	2,1%	3,3%	1,6%	0,8%	1,1%	3,0%	0,6%	1,3%	1,2%	0,4%	0,0%	0,0%	0,5%	0,0%	0,2%	0,1%	0,0%	0,0%	1,0%
18	1,3%	0,6%	0,3%	0,2%	0,3%	1,3%	0,8%	0,7%	0,5%	0,9%	2,0%	0,5%	0,5%	2,4%	1,0%	0,1%	0,1%	85,0%	0,2%	0,7%	0,4%	0,0%	0,2%	1,1%
19	1,4%	0,6%	0,2%	0,3%	0,3%	3,1%	2,1%	84,3%	0,8%	1,3%	2,3%	0,2%	0,7%	0,8%	0,5%	0,0%	0,0%	0,5%	0,0%	0,2%	0,2%	0,0%	0,0%	1,5%
20	4,8%	2,9%	1,1%	1,0%	0,7%	46,5%	7,7%	5,8%	3,3%	4,6%	6,7%	1,2%	5,0%	3,6%	1,7%	0,2%	0,2%	1,5%	0,3%	0,6%	0,4%	0,1%	0,1%	5,0%

Figura C.3: Resultados para los centros de K-Means utilizando el gasto porcentual por categoría.

		CENTROS POR CATEGORÍA																				TAMAÑO		
	J01	J02	J03_VH	J03_VM	J03_O	J04	J05	J06	J07	J08	J09	J10_H	J10_M	J11	J12	J13	J14	J15	J16	J17	J18	J21	J32	
1	4,2%	3,1%	1,3%	0,9%	1,0%	6,3%	4,6%	2,4%	48,0%	5,4%	5,6%	2,3%	4,2%	5,9%	1,3%	0,5%	0,4%	1,2%	0,4%	0,5%	0,3%	0,0%	0,1%	2,0%
2	7,2%	8,5%	37,1%	4,7%	3,3%	3,7%	4,0%	1,7%	1,7%	3,3%	7,2%	5,7%	3,4%	4,8%	1,1%	0,2%	0,3%	1,0%	0,4%	0,3%	0,3%	0,1%	0,0%	1,4%
3	0,7%	0,5%	0,2%	0,1%	0,2%	0,6%	0,5%	0,2%	0,2%	0,6%	1,1%	0,5%	0,5%	93,1%	0,2%	0,1%	0,1%	0,3%	0,1%	0,1%	0,1%	0,0%	0,0%	23,7%
4	4,5%	2,7%	1,1%	2,0%	0,9%	12,6%	7,4%	36,7%	3,0%	5,2%	7,3%	1,3%	5,2%	3,7%	2,2%	0,2%	0,3%	1,7%	0,5%	0,7%	0,5%	0,1%	0,1%	1,5%
5	1,8%	2,5%	0,8%	1,2%	0,5%	7,4%	62,3%	4,1%	2,6%	2,6%	5,6%	0,7%	3,7%	2,4%	0,6%	0,1%	0,1%	0,5%	0,1%	0,2%	0,2%	0,0%	0,0%	7,0%
6	2,2%	1,7%	0,8%	0,5%	1,3%	3,0%	2,3%	1,1%	1,3%	2,2%	4,6%	1,6%	2,4%	10,8%	1,3%	57,9%	2,2%	1,3%	0,7%	0,6%	0,2%	0,1%	0,1%	2,3%
7	2,5%	2,5%	1,2%	0,7%	0,9%	3,8%	4,3%	2,1%	1,3%	2,0%	68,5%	1,4%	2,3%	3,1%	1,1%	0,2%	0,2%	1,1%	0,1%	0,4%	0,3%	0,0%	0,0%	3,4%
8	3,2%	2,7%	1,3%	1,7%	1,0%	9,5%	6,9%	3,4%	2,9%	3,9%	6,5%	2,8%	45,8%	4,0%	1,5%	0,2%	0,3%	1,3%	0,4%	0,5%	0,3%	0,1%	0,0%	4,4%
9	2,2%	1,9%	0,8%	0,5%	1,2%	2,9%	2,3%	1,1%	1,1%	2,3%	4,7%	1,8%	2,4%	5,0%	1,9%	1,5%	64,6%	0,9%	0,3%	0,2%	0,2%	0,0%	0,0%	2,7%
10	2,7%	2,5%	1,4%	0,7%	57,2%	2,8%	2,6%	1,1%	1,4%	2,6%	5,4%	2,4%	2,6%	10,0%	0,9%	0,8%	1,0%	1,0%	0,3%	0,3%	0,2%	0,1%	0,1%	2,2%
11	2,9%	1,7%	0,6%	0,6%	0,4%	62,3%	6,4%	4,2%	2,5%	3,0%	4,5%	0,7%	4,3%	2,5%	1,2%	0,1%	0,1%	1,1%	0,1%	0,4%	0,3%	0,0%	0,0%	2,9%
12	5,2%	5,1%	2,2%	1,8%	1,7%	8,4%	7,8%	3,5%	2,7%	4,5%	37,1%	3,1%	5,9%	3,9%	2,1%	0,6%	0,8%	1,7%	0,5%	0,7%	0,4%	0,1%	0,1%	6,4
13	47,2%	7,4%	2,4%	0,8%	1,3%	5,6%	3,2%	2,0%	2,1%	3,8%	5,8%	5,1%	2,5%	5,5%	1,5%	0,3%	0,4%	1,5%	0,4%	0,5%	0,5%	0,1%	0,1%	3,0%
14	5,2%	4,3%	1,9%	1,7%	2,4%	8,8%	6,1%	3,1%	3,4%	6,5%	5,9%	3,5%	7,1%	30,3%	2,2%	1,8%	1,0%	2,1%	0,9%	0,8%	0,5%	0,1%	0,2%	6,6%
15	5,7%	3,7%	1,4%	1,7%	1,0%	33,2%	7,1%	4,9%	4,0%	6,1%	8,0%	2,0%	9,0%	4,5%	2,3%	0,6%	0,6%	1,9%	0,7%	0,9%	0,5%	0,1%	0,1%	5,5%
16	2,2%	1,9%	0,9%	0,4%	1,2%	2,2%	1,8%	0,9%	1,0%	2,2%	4,2%	1,8%	2,0%	72,9%	0,8%	0,8%	1,1%	1,0%	0,2%	0,2%	0,2%	0,1%	0,0%	13,6%
17	4,2%	3,2%	1,3%	1,0%	1,0%	6,7%	4,6%	3,1%	2,8%	48,7%	5,9%	2,0%	4,2%	5,9%	1,6%	0,4%	0,4%	1,4%	0,4%	0,5%	0,5%	0,1%	0,1%	3,9%
18	7,8%	7,4%	3,6%	1,2%	1,8%	3,8%	3,5%	1,5%	1,8%	3,2%	6,6%	42,9%	5,3%	5,6%	1,2%	0,3%	0,3%	1,2%	0,4%	0,3%	0,3%	0,1%	0,0%	2,9%
19	9,9%	44,4%	3,8%	1,3%	1,8%	4,4%	5,1%	1,7%	1,8%	3,3%	6,8%	4,9%	2,7%	4,6%	1,1%	0,2%	0,2%	1,0%	0,3%	0,3%	0,3%	0,0%	0,0%	2,9%
20	4,7%	2,8%	1,2%	1,3%	1,2%	6,1%	3,5%	3,0%	2,2%	4,2%	6,5%	1,9%	3,9%	6,7%	22,4%	1,0%	1,4%	20,1%	2,0%	2,2%	0,8%	0,1%	0,8%	1,8%

Figura C.4: Resultados para los centros de K-Means utilizando el gasto porcentual por categoría y muestreo por segmentación.

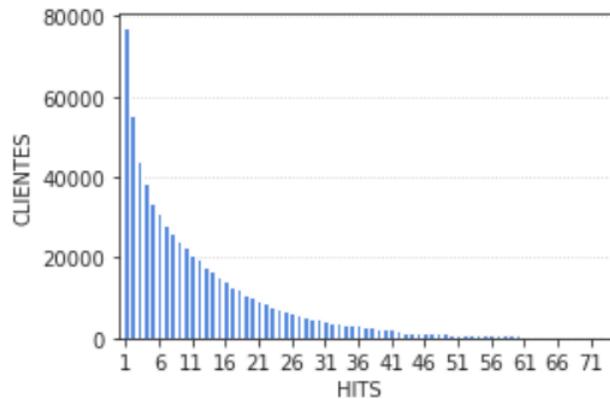


Figura C.5: Distribución del número de hits por usuario.

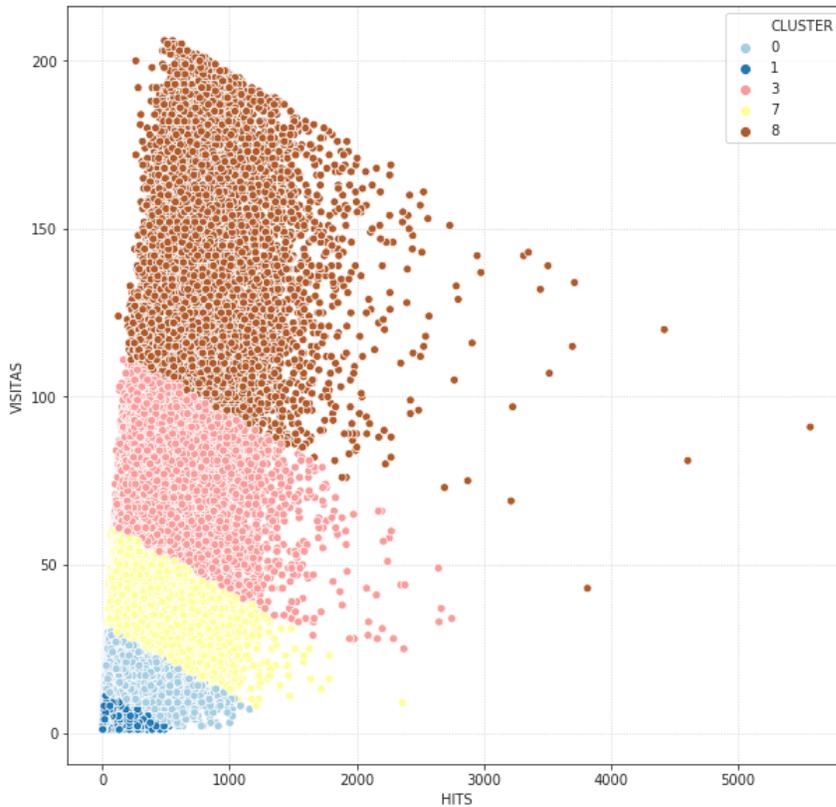


Figura C.6: Segmentación de clientes en base a al número de hits y sesiones totales utilizando KMeans.

		CENTROS POR CATEGORÍA																						
		J01	J02	J03_VH03_VA	J03_O	J04	J05	J06	J07	J08	J09	J10_H	J10_M	J11	J12	J13	J14	J15	J16	J17	J18	J21	J32	TAMAÑO
1	4,8%	4,8%	57,8%	2,4%	3,2%	0,9%	1,0%	0,7%	0,8%	1,5%	2,6%	4,0%	1,5%	9,0%	1,0%	1,2%	0,8%	0,7%	0,2%	0,3%	0,4%	0,3%	0,0%	1,4%
2	0,9%	0,6%	0,6%	0,6%	1,4%	1,6%	1,4%	1,8%	0,7%	1,5%	70,7%	0,7%	1,6%	8,8%	1,6%	1,6%	0,9%	1,0%	0,5%	0,6%	0,7%	0,3%	0,1%	8,2%
3	1,0%	0,5%	0,5%	0,6%	1,4%	1,6%	1,2%	1,4%	0,9%	2,1%	3,6%	0,8%	2,0%	9,4%	62,1%	2,4%	3,1%	1,9%	1,3%	1,1%	0,5%	0,4%	0,3%	3,1%
4	1,2%	0,7%	0,5%	0,7%	1,6%	2,2%	1,6%	1,2%	65,0%	3,3%	3,0%	1,0%	2,9%	8,4%	1,3%	1,4%	0,7%	1,1%	0,5%	0,7%	0,7%	0,4%	0,1%	1,7%
5	0,7%	0,5%	0,8%	0,5%	75,5%	0,5%	0,4%	0,4%	0,5%	1,2%	1,9%	1,0%	0,9%	10,0%	0,7%	1,2%	0,7%	0,8%	0,3%	0,4%	0,5%	0,5%	0,1%	3,2%
6	0,6%	0,6%	0,5%	1,2%	1,3%	3,1%	2,6%	1,3%	1,3%	2,5%	2,6%	2,2%	66,6%	7,4%	1,4%	1,5%	0,9%	0,8%	0,4%	0,5%	0,4%	0,3%	0,1%	4,0%
7	0,6%	0,4%	0,4%	0,4%	1,6%	0,8%	0,6%	0,6%	0,5%	1,4%	2,6%	0,6%	1,2%	10,7%	1,6%	68,7%	2,5%	1,2%	1,2%	1,4%	0,4%	0,4%	0,4%	3,8%
8	1,3%	0,8%	0,6%	2,3%	1,2%	49,6%	10,8%	3,4%	2,1%	3,6%	3,8%	0,7%	5,6%	6,3%	2,1%	1,6%	0,7%	1,2%	0,6%	0,8%	0,5%	0,3%	0,2%	4,4%
9	2,5%	1,9%	2,0%	4,8%	4,0%	4,2%	3,7%	3,2%	2,6%	5,4%	8,3%	2,6%	5,3%	21,2%	4,3%	4,8%	2,5%	3,3%	3,5%	4,0%	3,3%	2,3%	0,6%	18,2%
10	0,7%	0,4%	0,4%	0,7%	1,6%	1,8%	1,5%	1,3%	1,4%	71,1%	2,3%	0,6%	2,1%	8,1%	1,3%	1,1%	0,7%	0,9%	0,4%	0,6%	0,7%	0,4%	0,1%	3,9%
11	0,4%	0,3%	0,3%	0,3%	1,3%	0,5%	0,5%	0,4%	0,4%	1,0%	1,8%	0,7%	1,0%	9,3%	2,5%	2,9%	74,0%	0,7%	0,5%	0,4%	0,3%	0,3%	0,1%	2,0%
12	2,2%	0,7%	0,6%	1,8%	1,5%	3,8%	3,2%	57,2%	1,4%	3,2%	5,0%	0,7%	2,7%	7,7%	2,5%	1,4%	0,8%	1,4%	0,5%	0,7%	0,6%	0,4%	0,1%	2,4%
13	0,7%	0,5%	0,2%	0,1%	0,3%	0,6%	0,5%	0,3%	0,3%	0,6%	1,1%	0,5%	0,5%	90,1%	1,0%	0,9%	0,6%	0,5%	0,2%	0,3%	0,3%	0,2%	0,0%	33,8%
14	0,7%	0,4%	0,4%	0,4%	1,8%	1,0%	0,7%	0,9%	0,7%	1,9%	2,8%	0,7%	1,4%	11,6%	2,2%	2,3%	1,1%	64,2%	1,0%	1,2%	1,1%	0,5%	1,1%	1,9%
15	0,8%	1,2%	0,7%	2,9%	1,2%	10,3%	55,4%	2,6%	1,6%	2,9%	3,4%	0,8%	4,9%	6,0%	1,4%	1,1%	0,6%	0,7%	0,3%	0,5%	0,4%	0,2%	0,0%	2,9%
16	8,4%	57,8%	4,6%	0,7%	2,1%	1,2%	1,7%	0,6%	0,9%	1,4%	2,5%	4,0%	1,3%	8,1%	0,9%	0,9%	0,7%	0,7%	0,3%	0,4%	0,4%	0,3%	0,0%	1,4%
17	1,9%	1,9%	2,0%	0,4%	2,1%	0,5%	0,5%	0,3%	0,7%	1,3%	2,0%	68,7%	4,3%	8,9%	0,7%	1,1%	0,9%	0,6%	0,2%	0,3%	0,4%	0,3%	0,0%	2,3%
18	60,8%	5,0%	3,1%	0,6%	2,2%	1,1%	0,7%	1,5%	1,2%	1,9%	2,8%	3,1%	1,2%	8,6%	1,4%	1,2%	0,7%	1,0%	0,3%	0,4%	0,6%	0,3%	0,1%	1,6%

Figura C.7: Resultados para los centros de K-Means utilizando intensidad de navegación por categoría.

		CENTROS POR CATEGORÍA																						
	J01	J02	J03_VH03_VV	J03_O	J04	J05	J06	J07	J08	J09	J10_H	J10_M	J11	J12	J13	J14	J15	J16	J17	J18	J21	J32	TAMAÑO	
1	1,5%	1,3%	3,2%	38,3%	3,1%	6,4%	7,0%	4,0%	1,8%	3,8%	4,4%	1,3%	5,7%	9,3%	2,3%	1,9%	1,0%	1,4%	0,6%	0,7%	0,6%	0,4%	0,1%	1,8%
2	1,2%	0,7%	0,7%	1,3%	2,1%	3,9%	2,9%	2,7%	2,5%	52,8%	3,4%	0,9%	3,8%	10,7%	2,4%	1,9%	0,9%	1,7%	0,7%	1,1%	1,0%	0,5%	0,2%	4,3%
3	0,9%	0,6%	0,6%	0,6%	1,1%	1,7%	1,4%	2,2%	0,8%	1,5%	72,8%	0,7%	1,5%	6,8%	1,8%	1,6%	0,8%	0,9%	0,4%	0,5%	0,6%	0,2%	0,1%	4,5%
4	1,9%	1,1%	1,0%	1,2%	2,0%	4,2%	2,9%	2,4%	48,1%	4,6%	4,1%	1,3%	4,5%	10,4%	2,3%	2,1%	1,0%	1,7%	0,8%	1,0%	1,0%	0,4%	0,2%	2,0%
5	3,6%	3,7%	4,1%	0,8%	3,1%	1,1%	1,1%	0,7%	1,4%	1,8%	3,0%	50,6%	5,5%	12,2%	1,4%	1,8%	1,2%	1,0%	0,4%	0,5%	0,5%	0,4%	0,1%	2,2%
6	2,7%	0,9%	0,9%	2,2%	1,9%	5,8%	4,3%	44,6%	1,9%	4,2%	5,7%	0,8%	3,7%	9,2%	3,4%	1,9%	1,0%	1,8%	0,7%	1,0%	0,8%	0,4%	0,2%	3,1%
7	1,2%	1,7%	1,0%	3,1%	1,5%	13,7%	41,0%	3,8%	2,3%	3,8%	4,1%	1,0%	6,7%	7,7%	2,0%	1,5%	0,8%	1,0%	0,5%	0,7%	0,5%	0,3%	0,1%	4,2%
8	1,1%	0,9%	0,9%	1,7%	1,7%	5,5%	4,3%	2,2%	2,1%	3,5%	3,6%	2,7%	51,5%	9,2%	2,3%	2,1%	1,1%	1,2%	0,7%	0,7%	0,5%	0,4%	0,1%	4,7%
9	1,0%	0,7%	0,7%	0,6%	2,0%	1,6%	1,2%	1,1%	0,9%	1,9%	3,7%	1,1%	2,1%	12,9%	3,0%	54,1%	3,0%	2,1%	2,2%	2,5%	0,5%	0,4%	0,7%	3,8%
10	1,5%	1,0%	0,7%	2,2%	1,3%	43,2%	10,6%	4,2%	2,5%	4,2%	4,0%	0,8%	6,7%	7,1%	2,8%	1,9%	0,8%	1,5%	0,7%	1,0%	0,6%	0,3%	0,2%	5,7%
11	2,5%	1,7%	2,0%	2,1%	3,8%	5,4%	4,2%	3,7%	3,0%	6,0%	5,2%	2,4%	6,0%	22,8%	5,2%	5,5%	2,4%	3,6%	4,0%	4,2%	2,4%	1,3%	0,8%	11,3%
12	1,1%	0,8%	0,9%	0,6%	2,0%	1,4%	1,3%	1,0%	0,9%	2,0%	3,6%	1,4%	2,4%	13,6%	5,5%	5,6%	51,4%	1,5%	1,0%	0,9%	0,5%	0,5%	0,1%	1,6%
13	1,5%	1,2%	1,9%	1,1%	56,8%	1,3%	1,1%	1,0%	1,1%	2,4%	3,4%	2,0%	1,9%	13,9%	1,5%	2,4%	1,2%	1,5%	0,5%	0,8%	0,7%	0,7%	0,1%	2,5%
14	1,4%	0,8%	0,8%	0,9%	2,4%	2,4%	1,5%	2,0%	1,4%	3,1%	4,3%	1,0%	2,4%	14,5%	4,4%	3,8%	1,5%	43,4%	1,9%	2,3%	1,4%	0,6%	1,8%	2,1%
15	2,2%	1,9%	0,9%	0,4%	1,2%	2,2%	1,8%	0,9%	1,0%	2,2%	4,2%	1,8%	2,0%	72,9%	0,8%	0,8%	1,1%	1,0%	0,2%	0,2%	0,2%	0,1%	0,0%	30,4%
16	2,1%	1,5%	1,5%	1,6%	2,5%	4,6%	3,6%	4,0%	1,9%	3,4%	37,9%	1,6%	4,0%	14,3%	3,9%	3,3%	1,6%	2,3%	1,2%	1,4%	1,2%	0,5%	0,2%	7,8%
17	1,5%	0,8%	0,8%	1,0%	1,7%	2,8%	1,9%	2,4%	1,4%	2,9%	4,6%	1,0%	2,9%	11,0%	48,6%	3,5%	3,5%	2,5%	2,0%	1,8%	0,6%	0,4%	0,5%	3,7%
18	21,2%	18,2%	17,2%	1,5%	3,4%	1,9%	1,7%	1,4%	1,5%	2,1%	3,3%	5,4%	2,0%	11,4%	1,7%	1,7%	1,0%	1,2%	0,5%	0,6%	0,7%	0,4%	0,1%	4,4%

Figura C.8: Resultados para los centros de K-Means utilizando intensidad de navegación por categoría y muestreo por segmentación.

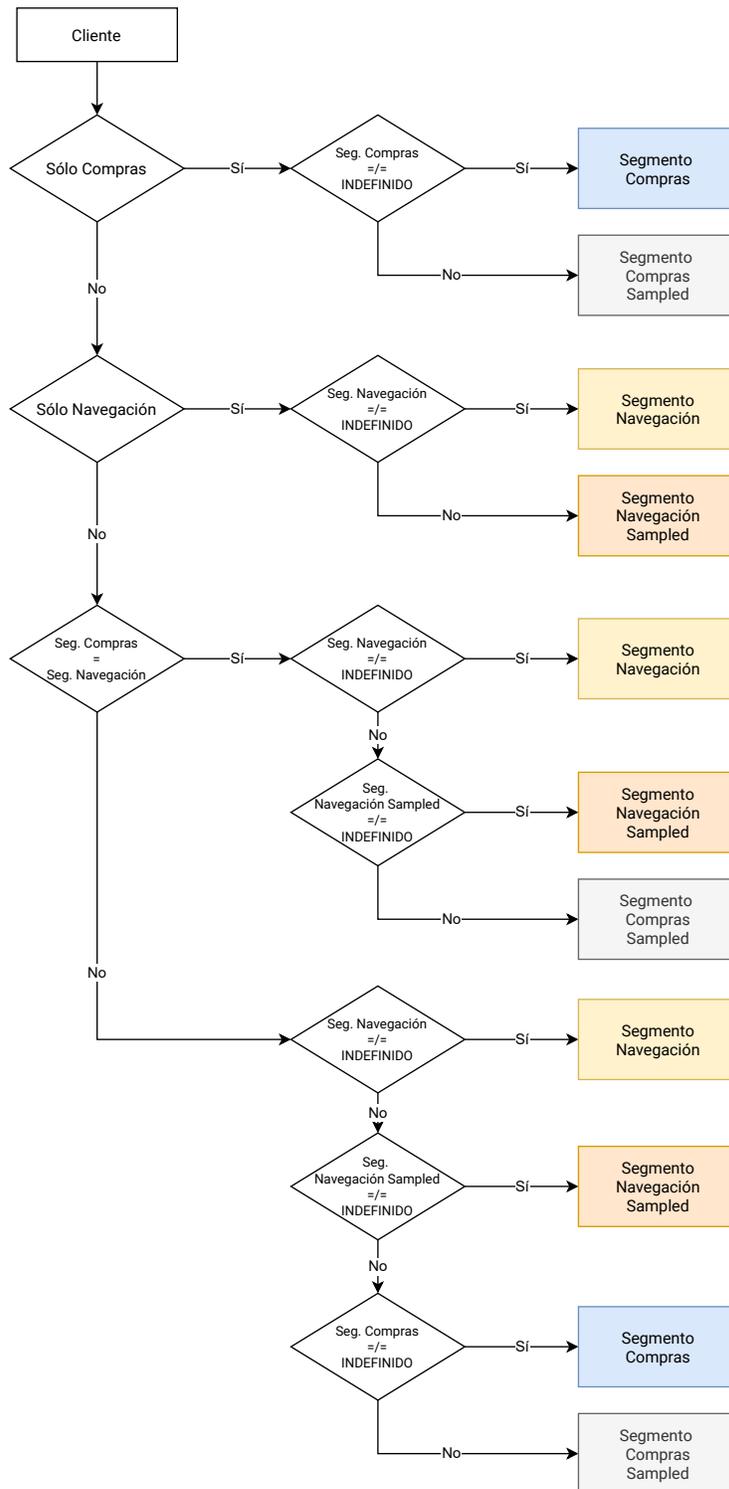


Figura C.9: Reglas de asignación de categorías utilizando los resultados de las etapas de segmentación.