



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

**EVALUACIÓN Y REDISEÑO DE UN MOTOR DE RECOMENDACIONES
PERSONALIZADAS A NIVEL DE CONSULTORA, PARA UNA EMPRESA DE
VENTAS POR CATÁLOGO DE LA INDUSTRIA DE BELLEZA**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

JAVIERA PAZ CARVAJAL GUAJARDO

PROFESORA GUÍA:
ALEJANDRA PUENTE CHANDIA

MIEMBROS DE LA COMISIÓN:
PABLO MARÍN VICUÑA
LUIS ABURTO LAFOURCADE

SANTIAGO DE CHILE
2019

EVALUACIÓN Y REDISEÑO DE UN MOTOR DE RECOMENDACIONES
PERSONALIZADAS A NIVEL DE CONSULTORA, PARA UNA EMPRESA DE VENTAS
POR CATÁLOGO DE LA INDUSTRIA DE BELLEZA

La siguiente memoria tiene como objetivo evaluar y proponer mejoras al actual motor de recomendaciones de la empresa Penta Analytics, el cual genera ofertas personalizadas a nivel de consultora-campaña para una empresa peruana dedicada a la venta por catálogo, con marco en la industria de belleza.

Una de las principales motivaciones para realizar el trabajo de título es que la empresa no posee una evaluación del motor de recomendaciones, por lo que no se tiene claridad sobre los resultados que este tiene en la empresa peruana. Por otro lado, también es parte de la memoria evaluar cómo y cuánto afecta el sesgo de popularidad al motor. Este sesgo está presente en gran parte de los sistemas de recomendaciones, generando que a los clientes se les muestren principalmente los productos populares del catálogo. Este tipo de recomendaciones resultan ser muy poco atractiva tanto para los clientes como para la empresa.

Para poder realizar la evaluación del motor de diseño de ofertas (MDO), es necesario en primer lugar realizar un levantamiento de este proceso, detallando los modelos ocupados e identificando oportunidades de mejora. De este levantamiento se proponen nuevos modelos de propensión de compra y de bundle.

La realización del diagnóstico de las recomendaciones resulta ser una de las etapas claves del trabajo de esta memoria, dado que en base a los resultados de este diagnóstico, se podrá definir si las propuestas realizadas logran mejorar el sesgo de popularidad que tiene el modelo actual del MDO. De la evaluación de las recomendaciones se obtiene que el motor cuenta con una tasa de conversión igual a 0,22%, y un efecto directo igual a USD\$145.000 equivalente al 1,1% de la venta total de la empresa.

Como propuestas para el nuevo modelo de propensión de compra se usa un algoritmo de filtros colaborativos a nivel de MCT (subconjunto de productos por su marca, categoría y tipo) y una versión de filtros colaborativos incluyendo un indicador de novelty score, el cual indica que tan novedosa es cada una de las MCT de la empresa. Con ambos modelos se busca recomendar solo productos nuevos a las consultoras.

Dentro de los principales resultados de este trabajo, se destaca el hecho de que es posible recomendar productos más novedosos a las consultoras y aún así, mejorar la tasa de conversión que tienen las ofertas alcanzando un 3,98% con el modelo de novelty score en las ofertas de productos individuales, y un 1,64% en las ofertas de packs de productos.

TABLA DE CONTENIDO

1. INTRODUCCIÓN.....	5
1.1. ANTECEDENTES GENERALES.....	5
1.2. JUSTIFICACIÓN	7
1.3. OBJETIVOS.....	10
1.3.1. Objetivo General	10
1.3.2. Objetivos Específicos.....	10
1.4. ALCANCES	11
2. MARCO CONCEPTUAL.....	12
2.1. SISTEMAS DE RECOMENDACIÓN	12
2.1.1. Filtros colaborativos	12
2.1.2. Novelty Score en filtros colaborativos	14
2.1.3. Reglas de asociación	16
3. METODOLOGÍA.....	17
3.1. METODOLOGÍA CRISP-DM.....	17
3.1.1. Comprensión del negocio	17
3.1.2. Entendimiento de los datos.....	18
3.1.3. Preparación de los datos	18
3.1.4. Modelado	18
3.1.5. Evaluación.....	18
4. DESARROLLO METODOLÓGICO	19
4.1. COMPRESIÓN DEL NEGOCIO.....	19
4.1.1. Motor de diseño de ofertas (MDO).....	19
4.1.2. Propuestas de mejora para el MDO.....	21
4.2. ENTENDIMIENTO DE LOS DATOS	22
4.3. PREPARACIÓN DE LOS DATOS	29
4.3.1. Indicadores típicos	30
4.3.2. Resultados indicadores típicos	31
4.3.3. Indicadores de sesgo de popularidad	32
4.3.4. Resultados indicadores de sesgo de popularidad	33
4.4. MODELO DE PROPENSIÓN DE COMPRA.....	34
4.4.1. Modelo de filtros colaborativos.....	34
4.4.2. Modelo de filtros colaborativos con novelty score.....	38
4.5. EVALUACIÓN DE LOS MODELOS DE PROPENSIÓN DE COMPRA.....	44
4.6. MODELO DE BUNDLE	46
4.6.1. Modelo basado en reglas de asociación.....	46
4.7. EVALUACIÓN DEL MODELO DE BUNDLE	48
5. CONCLUSIONES.....	50
5.1. CONCLUSIONES DEL PROYECTO.....	50
5.2. TRABAJOS FUTUROS	52

6. BIBLIOGRAFÍA.....	53
7. ANEXOS.....	54

1. INTRODUCCIÓN

1.1. Antecedentes generales

La presente memoria se realiza en la empresa Penta Analytics la cual brinda soluciones analíticas a través del análisis de Big Data, todo esto con el propósito de mejorar los resultados económicos y hacer mas eficientes los procesos de negocios de sus clientes. Las principales industrias a las cuales la empresa les brinda sus servicios son consumo masivo, educación e innovación [1].

La misión de Penta Analytics es:

“Mejorar los resultados económicos de nuestros clientes a través de la implementación de soluciones analíticas – integradas a sus sistemas y con seguimiento de resultados en tiempo real – para el apoyo efectivo a la toma de decisiones.” [2]

Mientras que la visión es:

“Ser una empresa líder en soluciones analíticas innovadoras a nivel Latinoamérica, buscando ser la primera alternativa para nuestros clientes y exigiéndonos continuamente en mejorar su rentabilidad.” [2]

El trabajo de la memoria se realiza para una empresa peruana que cuenta con 50 años de trayectoria en el negocio de la venta por catálogos en la industria de la belleza, con productos de cuidado personal y cosméticos. La empresa está presente en 15 países de América y cuenta con más de 800 mil consultoras, que son las encargadas de ofrecer y vender los productos a los clientes.

La venta por catálogo es considerada un tipo de venta directa y representa a un mercado que ha ido creciendo en todo el mundo, donde la principal fuerza de trabajo son las mujeres emprendedoras, que buscan poder administrar de forma independiente sus horarios de trabajo[3]. En Latino América el negocio de la venta directa alcanzó los US\$23,4 billones en ventas durante en año 2016 y logró un crecimiento de un 5,1% con respecto al año anterior, situándose así en la segunda región del mundo con mayor crecimiento en ventas, superada sólo por África-Medio Oriente. Durante el año 2016 la empresa alcanzó ventas por US\$1.090 millones posicionándose cómo la segunda empresa de venta directa con mayores ventas en Latinoamérica, siendo superada sólo por la empresa brasileña Natura.

El siguiente gráfico muestra la cantidad de consultoras y pedidos que tuvo la empresa durante el año 2017, en los 12 países de los cuales se tienen registros:



Gráfico 1: Cifras por país, año 2017.
Fuente: Elaboración propia

En el gráfico se puede observar cómo Colombia lidera en número de consultoras que realizaron pedidos durante ese año, y en el total de pedidos realizados. Dada estas cifras es que se decidió acotar el trabajo a este país.

La compañía cuenta con tres marcas, que se denominarán Marca A, Marca B y Marca C, donde cada una busca un público objetivo distinto y cuenta con su propio catálogo de ventas. La primera es la marca Premium de la empresa que busca llegar a un público más exclusivo. La segunda es una marca clásica y masiva, con productos enfocados principalmente en las mujeres, pero también algunos destinados a los hombres y los niños. Y la tercera se enfoca en el público más juvenil.

El modelo de ventas de la empresa está basado en campañas, las cuales ocurren en periodos de tiempo de 21 días donde se presenta un set acotado de productos para los clientes a través de un catálogo. Existe un catálogo distinto en cada campaña y para cada marca, donde los productos y las temáticas van cambiando y adecuándose a las festividades de cada país, sumando un total de 18 campañas al año.

Como se menciona anteriormente una campaña tiene una duración de 21 días. En los primeros 17 días las consultoras reciben los productos solicitados en la campaña anterior, realizan las entregas y hacen las cobranzas a sus clientes. De forma paralela se pueden mostrar y ofrecer los productos del nuevo catálogo, correspondiente a la campaña actual. Del día 17 al 21 es el periodo de facturación, donde las consultoras realizan los pedidos a través de dos posibles canales: por medio de una hoja de pedidos con los gerentes de zona y socias empresarias o a través del portal personal de la plataforma digital.

ETAPAS DE UNA CAMPAÑA	
VENTA Del día 1 al 17	FACTURACIÓN Del día 18 al 21
<ul style="list-style-type: none"> • Mostrar catálogo campaña actual • Entregar productos pedidos en la campaña anterior • Realizar cobranza de pedidos campaña anterior 	<ul style="list-style-type: none"> • Ingresar pedidos de la campaña actual eligiendo una de las dos alternativas: <ul style="list-style-type: none"> • Hoja de pedido física • Perfil personal consultora en sitio web

*Ilustración 1: Etapas de una campaña que deben seguir las consultoras.
Fuente: Elaboración propia*

En cuanto a su fuerza de venta, esta tiene una estructura piramidal, donde la base de esta pirámide son las consultoras que, a pesar de ser externas a la empresa, son el pilar dentro del modelo de venta directa dado que son las que tienen el contacto directo con los clientes. Las consultoras son las encargadas de mostrar los catálogos y ofrecer los productos a sus clientes, a través de un catálogo físico o de su versión online, luego deben realizar los pedidos con los productos elegidos por sus clientes para finalmente entregarlos en la siguiente campaña. Cabe destacar que las consultoras toman un papel activo al momento de ofrecer los productos, haciendo recomendaciones a sus clientas sobre que comprar y ofreciendo descuentos o promociones adicionales a las mostradas en el catálogo.

El ingreso del pedido realizado por la consultora puede ser notificado a través de una hoja de pedidos o a través de la plataforma digital, siendo esta última la de interés de estudio en esta memoria.

1.2. Justificación

Actualmente la empresa peruana brinda promociones personalizadas para cada consultora y en cada campaña, a través de portal web. Estas ofertas son creadas por el motor de diseño de ofertas (denomino MDO) de Penta Analytics. Se estima que en cada campaña se muestran entre 50 y 100 promociones distintas por cada consultora lo que suma un total de 12 millones de ofertas por ciclo, llegando a un total de 216 millones de promociones al año generadas por el MDO.

La promoción es personalizada a nivel de consultora-campaña, es decir, van variando campaña a campaña y para cada consultora. Estas ofertas se definen por la selección de uno o varios productos, un descuento y un tipo de promoción, los cuales son explicados en la siguiente tabla:

TIPO DE PROMOCIÓN	CARACTERÍSTICAS
Oferta individual	Un producto con un descuento en su precio normal
Oferta Volumen	2, 3, 4 o 5 productos iguales con un descuento en el precio total del conjunto
Oferta bundle	Conjunto de productos distintos con un descuento en el precio total del conjunto
Oferta niveles	Combinación de una oferta volumen con una oferta bundle, con un descuento extra o algún producto de regalo

Tabla 1: Tipos de promociones.

Fuente: Elaboración propia

Uno de los objetivos de esta memoria es poder evaluar que tan efectivas son estas recomendaciones, es decir, si son aceptadas o no por la consultora y los beneficios en cuanto a ventas que tienen para la compañía. Actualmente no existen cifras al respecto, por lo que se desconocen los efectos del MDO en la empresa. Esto genera una posición de desventaja de Penta frente a su cliente al momento de realizar nuevas negociaciones y renovaciones de contrato. Por otra parte, dificulta la mejora del modelo dado que no existe una situación base con la cual se poder comparar posibles mejoras o cambios en los modelos.

Por otro lado, también es parte del proyecto poder evaluar la calidad del MDO en cuanto a los productos que ofrece, como se menciona en el paper *What recommenders recommend*, unos de los sesgos de los recomendadores son el de popularidad y el de concentración[5] donde, a pesar de mostrar buenos indicadores como la tasa de conversión o fidelización de sus clientes, tienden a ofrecer los productos más vendidos o una pequeña muestra del catálogo lo que puede contrastar con algunos objetivos de las promociones. Para poder evaluar estas características los autores recomiendan construir indicadores que muestren los niveles de diversidad, novedad o familiaridad que hay en el set de artículos recomendados, como también la cobertura del catálogo que se tiene. Para evaluar el MDO se definen dos métricas claves: novedad y cobertura de catálogo. La novedad indica si los productos recomendados son nuevos para la consultora, es decir, si los ha comprado o no en las últimas 6 campañas. La cobertura de catálogo evaluará que porcentaje del catálogo se recomienda a las consultoras dentro de todas las ofertas que genera el MDO en cada campaña.

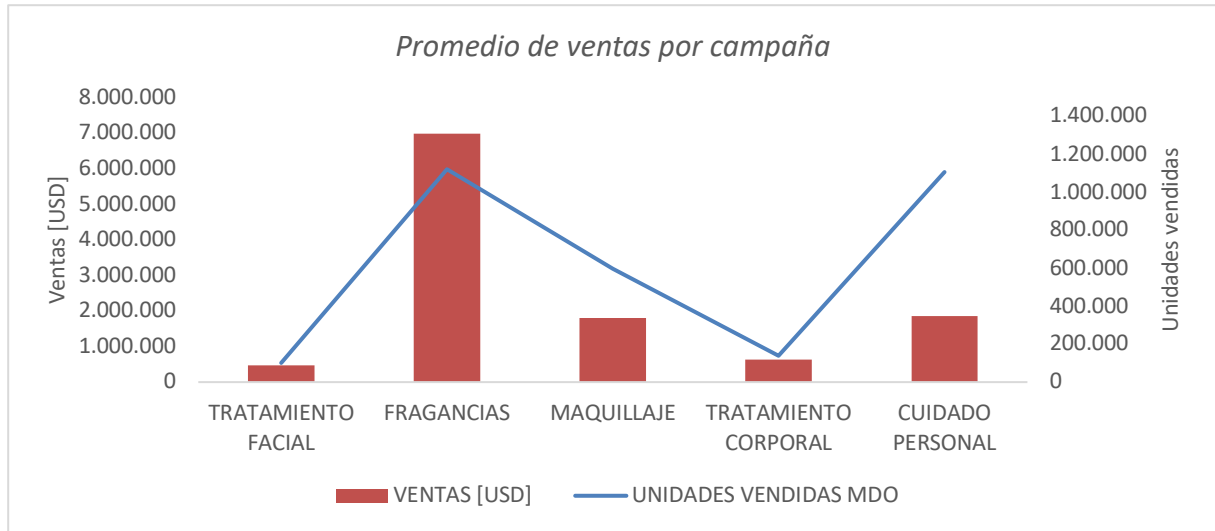


Gráfico 2: Promedio de ventas por campaña y promedio de unidades vendidas por campaña, agrupados por categoría. Entre las campañas 09 y 14 del 2018.
Fuente: Elaboración propia

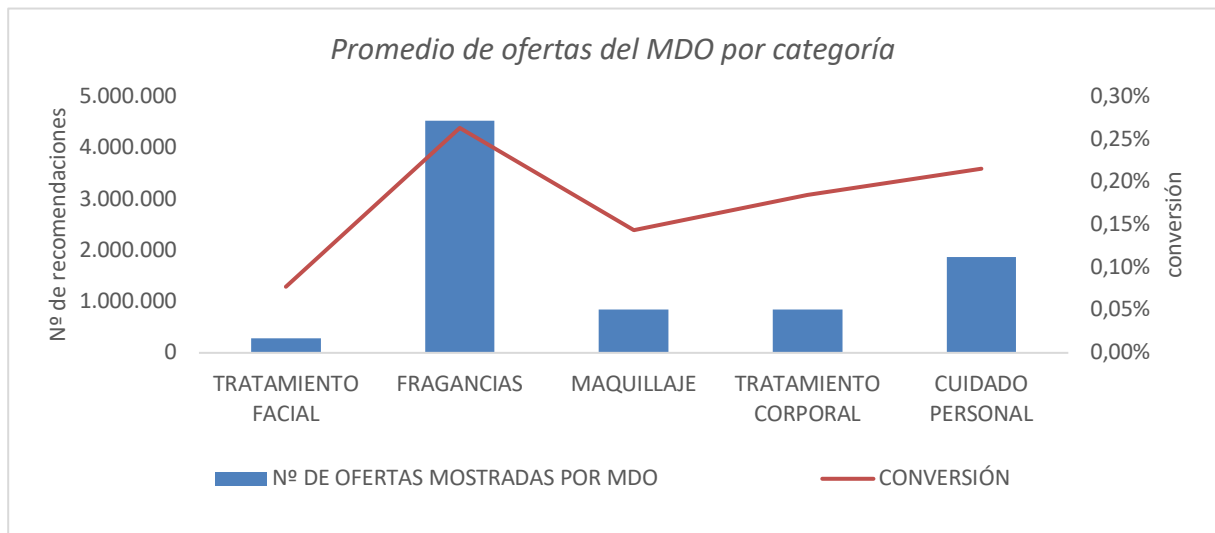


Gráfico 3: Promedio del número de ofertas del MDO por campaña y tasa de conversión promedio por campaña, agrupados por categoría. Entre las campañas 09 y 14 del 2018.
Fuente: Elaboración propia

Como primer análisis se tiene el gráfico 2, de este se puede concluir que la categoría más vendida en la empresa es la de fragancias, seguido por la categoría de cuidado personal. También se observa que son las categorías que generan mayores ventas en dólares. En el gráfico 3 se observa que las categorías más mostradas por el MDO son fragancias y cuidado personal, es decir, coincide con las categorías que más se venden. Con esto se puede concluir que las recomendaciones del MDO tienen presente

el sesgo de popularidad y se debe ahondar más en el tema para hacer conclusiones más específicas al respecto.

Mejorando las recomendaciones de la plataforma digital se lograría ampliar la canasta de compra de las consultoras, y a su vez lograr que estas ofrezcan mayor variedad de productos a sus clientes. Adicionalmente podría evitar posibles sobre stock de las consultoras, que perjudiquen las ventas de las campañas futuras o la disminución del margen en las compras dado los descuentos ofrecidos.

Actualmente se estima que la tasa de conversión del MDO es de un 0,22%, lo que equivale a USD\$ 145.000 por cada campaña, representando un 1,11% de la venta total de la empresa. Si se logra subir la tasa de conversión en un 1%, se obtendrían ventas por USD\$ 146.450 por medio de las promociones, aumentando las ventas totales en USD\$ 26.100 al año.

1.3. Objetivos

1.3.1. Objetivo General

Evaluar y rediseñar un motor de promociones personalizadas a nivel de consultora– campaña para una empresa de ventas por catálogo, con el fin de disminuir el sesgo de popularidad de los productos recomendados manteniendo la tasa de conversión actual.

1.3.2. Objetivos Específicos

- Hacer un levantamiento del proceso actual del MDO, identificando subprocesos y oportunidades de mejora en base a brechas proponiendo un rediseño en el proceso.
- Realizar un diagnóstico a las recomendaciones entregadas por el MDO, mediante indicadores de conversión, porcentajes de participación y beneficios monetarios, además de nuevos indicadores de variedad, novedad y concentración.
- Diseñar y aplicar las mejoras propuestas al MDO, realizando cambios en las etapas de propensión y de Bundle del proceso mediante desarrollo de nuevos modelos de recomendación de productos y de paquetes de productos.
- Evaluar las mejoras realizadas al MDO usando la información histórica y comparar los resultados con los del modelo actual.

1.4. Alcances

- En esta memoria se trabajará sólo con los datos de la empresa existentes en las bases de Penta Analytics correspondientes al País de Colombia durante las campañas del 2017 en adelante.
- Se plantearán mejoras sólo para las etapas de propensión de compra y modelo de bundle del proceso del MDO.
- Se crearán recomendaciones para usuarios que hayan comprado en las últimas 6 campañas, por lo que los usuarios nuevos (que realizan su primer pedido) no serán considerados
- Se recomendarán productos que hayan aparecido en las últimas 6 campañas, por lo que no se incluirán productos nuevos (que no se hayan expuesto en campañas anteriores)
- No es parte de esta memoria implementar las mejoras propuestas en el modelo del MDO, ni llevar a cabo experimentos para evaluarlas.

2. MARCO CONCEPTUAL

2.1. Sistemas de recomendación

Los sistemas de recomendación ayudan a las empresas seleccionar productos o contenido que podría gustarle a sus usuarios, gestionando con esto una compra o manteniendo por más tiempo interesado al usuario en el sitio web de dicha empresa.

Existen dos enfoques para los sistemas de recomendación, el primero se basa en estrategias de asociación donde se identifican comportamientos similares al momento de comprar productos. Un ejemplo de esto es cuando un usuario compra un computador, se le recomendará comprar accesorios como un mouse o un bolso para transportarlo. La segunda estrategia se basa en encontrar usuarios similares, es decir, que han tenido comportamientos de compra similares en el pasado. Esta estrategia se basa en el supuesto de que si dos usuarios han tenido gustos similares en el pasado tendrán gustos similares en el futuro.

2.1.1. Filtros colaborativos

Los sistemas de filtros colaborativos ocupan la estrategia de similitud de usuario, usando distintas técnicas para calcular la semejanza entre estos y así poder recomendarles productos. A pesar de que existen distintos tipos de filtros colaborativos todos tienen una visión general parecida del proceso:

Típicamente se tienen dos vectores principales: un vector con m usuarios denominado U y otro vector con n productos identificado como I [11].

Cada usuario tendrá distintas calificaciones para distintos productos, estas pueden ser explícitas o implícitas. Las primeras corresponden a los casos donde el usuario califica con una nota un productos o servicio. Mientras que las calificaciones implícitas son aquellas en que se identifica como calificación las veces que el usuario ha comprado el producto. Con estas evaluaciones se crea lo que denomina la matriz de rating o de interacciones (A).

La matriz de rating contiene las calificaciones de cada consumidor a cada producto, donde se puede denomina como NA o cero cuando el usuario no ha comprado ese producto antes. Los vectores antes mencionados tienen los siguientes formatos:

$$U = \{ u_1, u_2, \dots, u_m \}$$

$$I = \{ i_1, i_2, \dots, i_n \}$$

$$A = (a_{iu})$$

- **Filtros colaborativos basados en el usuario**

Este algoritmo predice transacciones futuras a través de transacciones observadas de consumidores similares, utilizando la información usuario-producto para generar las predicciones. El algoritmo genera una matriz de similitud de consumidores denominada WC , donde un valor alto de wc_{uv} indica que los usuarios u y v podrían tener preferencias similares[¿?].

$$WC = (wc_{uv}) , \quad u, v = 1, 2, \dots, m$$

Luego se calcula $WC \cdot A$, matriz que indica los puntajes potenciales de los productos para cada consumidor.

- **Filtros colaborativos basados en el producto**

Este algoritmo tiene un enfoque probabilístico, generando recomendaciones de productos desarrollando primero un modelo clasificación de los usuarios. Utiliza algoritmos de *machine learning* como Redes Bayesianas, Clustering, y Rule-based. En primer lugar, el modelo genera una matriz de similitud de productos identificada como WP donde un valor elevado en wp_{ij} indica que los productos i y j son similares, es decir, muchos usuarios han comprado ambos productos.

$$WP = (wp_{ij}) , \quad i, j = 1, 2, \dots, n$$

En segundo lugar, se calcula $A \cdot WC$, esta matriz indica los puntajes potenciales de los productos para cada consumidor.

Como se mencionó anteriormente, los filtros colaborativos basan sus predicciones comparando el comportamiento de usuarios similares, esta similitud se puede calcular de distintas formas, donde las más usadas son:

- **Medida de similitud de Pearson**

Una forma de calcular la similitud es usando el coeficiente de correlación de Pearson. Para calcular la semejanza entre el usuario u y el usuario v se utiliza la siguiente fórmula, basándose en la información de la matriz de rating:

$$Sim(u, v) = \frac{\sum_{i \in I} (R_{u,i} - \bar{R}_i)(R_{v,i} - \bar{R}_i)}{\sqrt{\sum_{i \in I} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{i \in I} (R_{v,i} - \bar{R}_i)^2}}$$

, con $R_{u,i}$ la calificación del usuario u al producto i , \bar{R}_i la calificación promedio del producto i .

- **Medida de similitud Coseno**

La métrica coseno mide la similitud entre dos vectores basándose en el ángulo entre ellos. La similitud entre el usuario u y el usuario v se calcula como:

$$Sim(\vec{u}, \vec{v}) = \cos(u, v) = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}| * |\vec{v}|}$$

- **Medida de similitud de Coseno ajustada**

La similitud de coseno sin ajustar no toma en cuenta las diferencias entre las escalas de calificación promedio de los usuarios, este problema se corrige restando el promedio de las calificaciones del usuario. La métrica se calcula de la siguiente forma:

$$Sim(u, v) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}}$$

, con \bar{R}_u el promedio de las calificaciones del usuario u.

2.1.2. Novelty Score en filtros colaborativos

Los modelos de filtros colaborativos tienen presente lo que se conoce como sesgo de popularidad, donde se tiende a recomendar los productos más populares del set y por ende resultan ser poco novedosos para los clientes. El problema con esto es que los productos populares pueden ya ser conocidos por los usuarios y no representan un real aporte. La métrica de puntuación de novedad en los filtros colaborativos está basada en la frecuencia con que son comprados los productos y la frecuencia inversa de los usuarios con los ítems que no conocen, la cual ayuda a identificar los productos más novedosos para los usuarios.

El artículo *“Using novelty score of unseen items to handle popularity bias in recommender systems”*[12] propone una nueva técnica de filtros colaborativos denominada *“Enfoque de filtros colaborativo modificado para recomendaciones novedosas”* (siendo MCFNR sus siglas en inglés). Para aplicar esta técnica se deben segmentar a los usuarios en dos tipos: los usuarios ocasionales, a quienes se les recomendarán productos usando el modelo tradicional de filtros colaborativos y los usuarios frecuentes a los cuales se les recomendarán productos más novedosos usando el Novelty Score.

Para poder realizar la segmentación de los usuarios es necesario definir la métrica θ_{uc} , la cual nos indicará el número de interacciones que ha tenido el usuario u con el sistema. Luego, se debe definir un θ_{uc}^* que será el valor límite para segmentar a los usuarios. Los usuarios con su $\theta_{uc} \geq \theta_{uc}^*$ serán considerados usuarios frecuentes, mientras que los que tengan $\theta_{uc} < \theta_{uc}^*$ se considerarán usuarios ocasionales.

Para los usuarios frecuentes se utilizará la siguiente metodología:

1. Categorización del usuario objetivo u como frecuentes

Se deben separar todos los usuarios clasificados como usuarios frecuentes, que tengan un $\theta_{uc} \geq \theta_{uc}^*$.

2. Predecir la relevancia de los elementos para las consultoras

Al igual que en los filtros colaborativos conocidos, se debe saber que tan propenso es el usuario para comprar un producto, para esto se realiza una comparación con las preferencias de los usuarios similares. La relevancia se calcula como:

$$\hat{r}(u, i) = \bar{r}(u) + C \sum_{v \in N_k(u, i)} sim(u, v)(r(v, i) - \bar{r}(v))$$

, con $\hat{r}(u, i)$ es la relevancia del producto i para el usuario u , $\bar{r}(u)$ el promedio de los rating del usuario u , C es la constante normalizada, N_k es el barrio del usuario u , $sim(u, v)$ es el puntaje de similitud entre el usuario y el vecino v calculada con la medida de similitud de Pearson, $r(v, i)$ es el rating del usuario u al productos i y $\bar{r}(v)$ es el rating promedio del usuario v .

3. identificación de los elementos desconocidos para el usuario objetivo u

Se deben identificar los elementos “desconocidos”, los cuales son elementos que los usuarios no han comprado en un periodo definido de tiempo.

I = set de productos a recomendar

r_i = rating de un usuario para el producto i

I_{su} = set de productos que el usuario a comprado en un periodo definido de tiempo

$I_{ns,u}$ = set de productos desconocidos por el usuario

4. Calcular la novedad de los elementos desconocidos usando Novelty Score (NS)

Una vez que se sabe que elementos son desconocidos para el usuario y cual es la relevancia que tienen, se debe calcular el Novelty Score $NS(i)$ que indicará que tan novedoso es un producto. Para calcular el NS se usa la frecuencia de un producto denominada $IF(i)$, la cual indica que tanto se compra ese elemento comparado al elemento más comprado del catálogo. Y la frecuencia inversa de un usuario llamada $IUF(i)$, que evalúa de forma negativa si muchos usuarios han comprado un producto comparado con el total de usuarios del sistema.

$$IF(i) = \frac{\text{n}^\circ \text{ de rating recibidos por el producto } i}{\text{máximo n}^\circ \text{ de rating recibidos por un producto}}$$

$$IUF(i) = \log \left\{ \frac{\text{n}^\circ \text{ total de usuarios}}{\text{n}^\circ \text{ de usuarios que evalúan el producto } i} \right\}$$
$$NS(i) = IF(i) * IUF(i)$$

2.1.3. Reglas de asociación

Las reglas de asociación son algoritmos que tienen como objetivo encontrar relaciones significativas entre elementos dentro de un conjunto de transacciones. Estas relaciones ocurren entre dos o más elementos, como productos, clientes o eventos, dependiendo del problema donde se desean aplicar las reglas.

Las transacciones (T) pueden ser las compras que se realizan en un supermercado, o como en el caso de esta memoria los pedidos que realizan las consultoras. Los elementos que conforman las transacciones son denominados items, y los conjuntos de items se denomina itemsets. Por ejemplo, se puede tener una transacción conformada por los items A, B y C:

$$T = \{A, B, C\}$$

Los posibles itemsets de esta transacción son:

$$Itemsets = \{A, B, C\}, \{A, B\}, \{A, C\}, \{B, C\}, \{A\}, \{B\} \text{ y } \{C\}$$

Una regla de asociación se define como una relación entre X e Y, siendo X e Y items o itemsets. La regla se escribe de la forma $X \Rightarrow Y$, la parte de la izquierda se conoce como antecedente y la parte de la derecha se denomina consecuente.

Para evaluar una regla de asociación se calculan tres principales variables[8]:

- El soporte indica la probabilidad que tiene un producto de aparecer en una transacción:

$$Soporte(X \Rightarrow Y) = \mathbb{P}(X \cup Y)$$

- La confianza es la probabilidad de que una transacción que contenga un antecedente, también contenga el consecuente:

$$Confianza(X \Rightarrow Y) = \frac{Soporte(X \cup Y)}{Soporte(X)}$$

- El lift muestra la probabilidad de que aparezca el conjunto A, B en un pedido, dividido por la probabilidad de que aparezca cada producto por separado:

$$Lift(X \Rightarrow Y) = \frac{\mathbb{P}(X, Y)}{\mathbb{P}(X)\mathbb{P}(Y)}$$

3. METODOLOGÍA

3.1. Metodología CRISP-DM

Para cumplir con los objetivos descritos en los puntos anteriores se desarrolla la presente memoria utilizando la metodología CRISP-DM, la que es una de las más usadas en proyectos de minería de datos. Esta metodología tiene 6 etapas: Comprensión del negocio, Entendimiento de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue. [4]

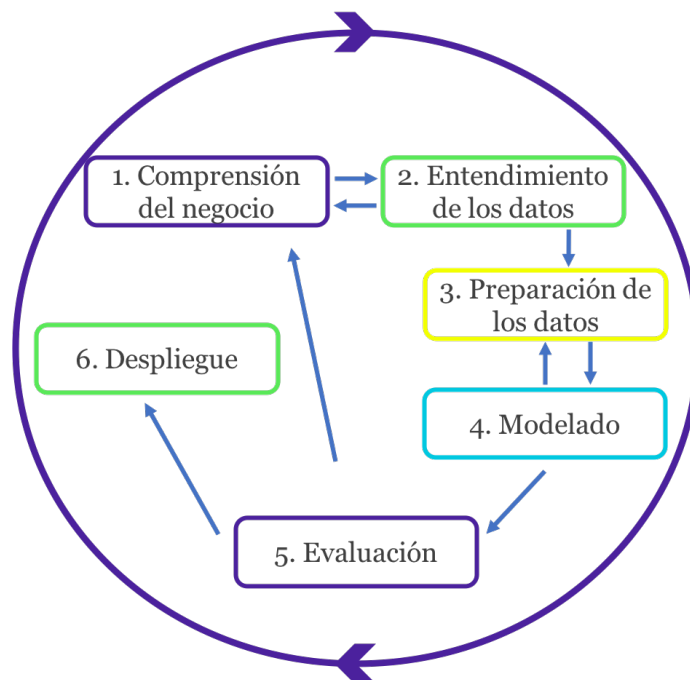


Ilustración 2: Etapas del proceso de KDD.

Fuente: Elaboración propia en base a papper "From data mining to knowledge discovery in databases".

Las etapas de la metodología a seguir en la presente memoria, son las siguientes:

3.1.1. Comprensión del negocio

En esta esta etapa se debe entender el mercado en el que se mueve la empresa y las principales características del negocio, comprender como funciona la venta por catálogo y sus principales características. También es parte de la comprensión del negocio el reclutar información para realizar el levantamiento del proceso del motor de recomendaciones de la empresa , describir sus etapas, evaluarlas y proponer mejoras.

3.1.2. Entendimiento de los datos

Luego de finalizar la primera etapa y entender el funcionamiento del motor de recomendaciones, se deben explorar los datos disponibles en las bases, entenderlos y seleccionar las principales tablas con las que se debe trabajar. Se verifican los datos disponibles y se identifican los faltantes, para ver la posibilidad de calcularlos por medio de los existentes o realizar las mejores aproximaciones posibles. También es parte de esta etapa realizar una descripción de los datos a ocupar.

3.1.3. Preparación de los datos

En esta etapa, una vez que ya se conocen las bases de datos se deben seleccionar las variables con las que se va a trabajar. Se crean nuevas tablas seleccionando, por ejemplo, el país y las fechas con las que se trabaja en la memoria. Por otro lado, se debe verificar la presencia de outliers y campos vacíos en los datos seleccionados, eliminándolos o rellenando según sea el caso. Finalmente, en esta etapa se deben crear las tablas necesarias para poder aplicar los modelos de minería de datos.

3.1.4. Modelado

En base a las mejoras propuestas para el motor de recomendaciones, se deben seleccionar las técnicas de minería de datos a usar en la memoria, se deben construir los modelos y diseñar la forma en que se evaluarán los resultados y se compararán con el modelo anterior. Se crea un modelo para la etapa de propensión de compra del MDO u otro modelo para la etapa de ofertas tipo bundle.

3.1.5. Evaluación

Evaluar los resultados obtenidos en la etapa de modelado y compararlos con los resultados del motor de recomendaciones inicial. Verificar si las mejoras realizadas son útiles a las necesidades de la empresa, revisar las etapas del proceso y establecer los siguientes pasos.

4. DESARROLLO METODOLÓGICO

4.1. Comprensión del negocio

4.1.1. Motor de diseño de ofertas (MDO)

El MDO es un motor que recomienda ofertas personalizadas a las consultoras que realizan pedidos a través del sitio web de la empresa. Este motor no sólo recomienda productos de forma personalizada a las consultoras en cada campaña, también asigna el tipo de oferta en que será mostrado el producto y su descuento. El MDO contiene tanto algoritmos en SQL como modelos de Machine Learning en el software R.

Cabe mencionar que al momento de validar el proceso con los trabajadores de Penta Analytics, se comentó que existía una nueva versión del MDO, la cual había empezado a funcionar en Colombia desde la campaña 08 del año 2018. En la siguiente ilustración se observan las etapas del MDO, las que son explicadas más adelante:



*Ilustración 3: Etapas del motor de diseño de ofertas.
Fuente: Elaboración propia*

- **Perfil de consultoras**

En esta etapa, el modelo busca crear perfiles de las consultoras para poder segmentarlas según sus preferencias y aplicar ciertas reglas del modelo según los perfiles. Para esto usa sus datos transaccionales como por ejemplo, productos que compran, marcas que frecuentan o las categorías por las que se inclinan. También ocupa los datos sociodemográficos, como país y región a las que pertenecen.

Usando el algoritmo k-means en dos etapas se genera un total de seis clústers para agrupar a las consultoras, según sus preferencias de marca y de categoría en sus transacciones. En una primera instancia, se generan dos grupos basados en las marcas que compran y luego, basándose en las categorías de los productos transados, se obtienen tres clústers. La unión de estos perfiles genera seis combinaciones. Los perfiles de las consultoras son del tipo marca - categoría, y se recalibran cada 6 campañas correspondientes a 4,2 meses.

- **Modelo de propensión de compra**

En esta etapa se calcula la probabilidad de que una consultora compre una MCT (marca-categoría-tipo) determinada. La MCT es un subconjunto de productos de una

misma marca, categoría y tipo. Por ejemplo, los productos de la marca A, categoría maquillaje y tipo maquillaje de ojos.

La propensión de compra es calculada a través de árboles de decisión, generando un árbol para cada MCT calificada como relevante, es decir, las MCT que representan el 80% de la venta. En total se crean 27 árboles. Luego, para determinar a qué productos es más propensa la consultora, se realiza una bajada a nivel de producto priorizando según tres objetivos comerciales: propensión (MCT probable para la consultora), presencia (productos que la consultora compra) y recompra (productos de compra frecuente).

- **Ofertas tipo volumen**

En esta etapa se predice el número de productos que comprará una consultora en el futuro, para esto se utilizan 2 algoritmos: árboles de regresión y Tóbit. La predicción se realiza a nivel de MCT, generando un total de 27 modelos en cada algoritmo. El número de productos que comprará de una MCT se calcula como el promedio de los resultados obtenidos en las predicciones de cada modelo, de la siguiente forma:

$$N^{\circ} \text{ de productos que comprará} = \frac{\text{Árbol de regresión} + \text{TÓBIT}}{2}$$

Finalmente, para saber cuantos productos recomendar en las ofertas de este tipo, al valor obtenido con los promedios de los modelos se le agrega una unidad más, esto para forzar la venta incremental, obteniendo así la cantidad de productos a ofrecer en las promociones de volumen. Por reglas del negocio se puede ofrecer un máximo de 5 unidades.

Por ejemplo, para una MCT dada, si el árbol de regresión predice que en la próxima campaña se comprarán 3 unidades de la MCT y el modelo Tobit predice que se comprarán 5 unidades, el promedio de ambos resultados es 4 unidades. Luego, se agrega un producto extra, por lo que finalmente se crea una oferta de volumen recomendando un pack de 5 unidades.

- **Selección del descuento**

Utilizando la información histórica de los descuentos y productos que han comprado las consultoras, se calcula el mínimo descuento moda por cada relación producto-consultora-descuento, y así saber que descuento aplicar a las promociones. Se le llama descuento moda al descuento que más se repite entre los productos comprados por la consultora. En caso de que no exista la relación producto-consultora, es decir, la consultora no haya comprado ese producto antes, se calcula para las interacciones consultora-MCT-descuento o perfil-MCT-descuento. Por reglas de negocio se puede aplicar un descuento máximo de 60%.

Uno de los cambios relacionados con la nueva versión del MDO, es que el descuento de las ofertas ya no sería personalizado, si no que sería único por país creando la relación país-producto-descuento.

- **Ofertas tipo bundle**

Utilizando como input la información transaccional de las consultoras, en esta etapa se crean los packs de productos distintos a ofrecer en las promociones. Se crean reglas de asociación con los datos transaccionales de la empresa, para luego filtrar las 300 mejores reglas por perfil según 3 parámetros de ajuste: lift, confianza y soporte (en orden de importancia). Este algoritmo se realiza a nivel de perfil de las consultoras, por lo que las consultoras pertenecientes al mismo perfil tienen las mismas ofertas.

4.1.2. Propuestas de mejora para el MDO

Uno de los puntos débiles que se puede observar en el MDO, es el modelo de propensión de compra utilizado, el cuál calcula la probabilidad que tiene una consultora a comprar una MCT, ocupando como input las transacciones de las consultoras pertenecientes a su perfil. Estos perfiles se crean agrupando consultoras que tienen un comportamiento similar en cuanto a la marca y categoría de los productos comprados, para luego bajar a nivel de producto nuevamente filtrando productos que la consultora ya ha comprado antes. Es de esperarse que el modelo termine recomendando productos que la consultora ya compra de manera frecuente, lo cual no agrega valor a la recomendación en sí.

Otra desventaja en este proceso es que el recomendador solo trabaja con 27 de las 233 MCT existentes, por lo que gran parte del catálogo de productos queda fuera de las recomendaciones. Además, esto aumenta la probabilidad de recomendar solo productos populares entre las consultoras, dado que son las MCT relevantes en cuanto a ventas.

Para poder mejorar el punto anterior se propone un modelo de recomendación basado en filtros colaborativos, agregando todas las MCT al modelo y calculando de forma adicional los puntajes de novedad de cada MCT, esto para poder recomendar productos menos populares, y por ende mas novedosos para cada consultora.

Por otro lado, en el modelo de las promociones de tipo Bundle, nuevamente se recomienda lo más comprado por las consultoras, esto dado que las reglas de asociación funcionan identificando conjunto de elementos que se compran frecuentemente. El modelo del MDO al momento de crear los packs, no presenta restricciones que generen la recomendación de un producto nuevo para la consultora. La recomendación de productos nuevos dentro de las ofertas de packs de productos, tiene como objetivo aumentar el mix de productos de la canasta de compra (up-selling).

Se propone incorporar nuevas restricciones en los packs de ofertas con el fin de recomendar productos que las consultoras no hayan comprado antes, y con esto aumentar la novedad en las recomendaciones. Además, al agregar las restricciones se genera una recomendación personalizada por consultora, y no se ofrece lo mismo por cada perfil. Al igual que en el modelo de propensión de compra se propone eliminar la regla de negocio que solo permite recomendar MCT pertenecientes al 80% de la venta, ampliando la canasta de recomendaciones a todas las MCT pertenecientes a las 3 principales marcas de la empresa.

4.2. Entendimiento de los datos

Penta Analytics mantiene una amplia base de datos de la empresa, donde las principales tablas con las que se va trabajar son:

- **Maestro de consultoras (debelista):**

Información sociodemográfica de las consultoras como país, estado civil, fecha de nacimiento, campaña de ingreso, etc. Contiene un total de 10.773.085 de registros¹.

El maestro de consultoras cuenta con un total de 1.552.510 consultoras activas en el año 2017, es decir, que realizaron alguna transacción en ese año.



Gráfico 4: Consultoras por país en el año 2017.
Fuente: Elaboración propia

Se puede observar en el gráfico anterior que Colombia representa el país con mayor número de consultoras, seguido por Perú y México. Esto demuestra que Colombia es un país atractivo para trabajar dentro de los países que se tienen registros.

¹ Detalles de los registros en anexo J: tabla 1.

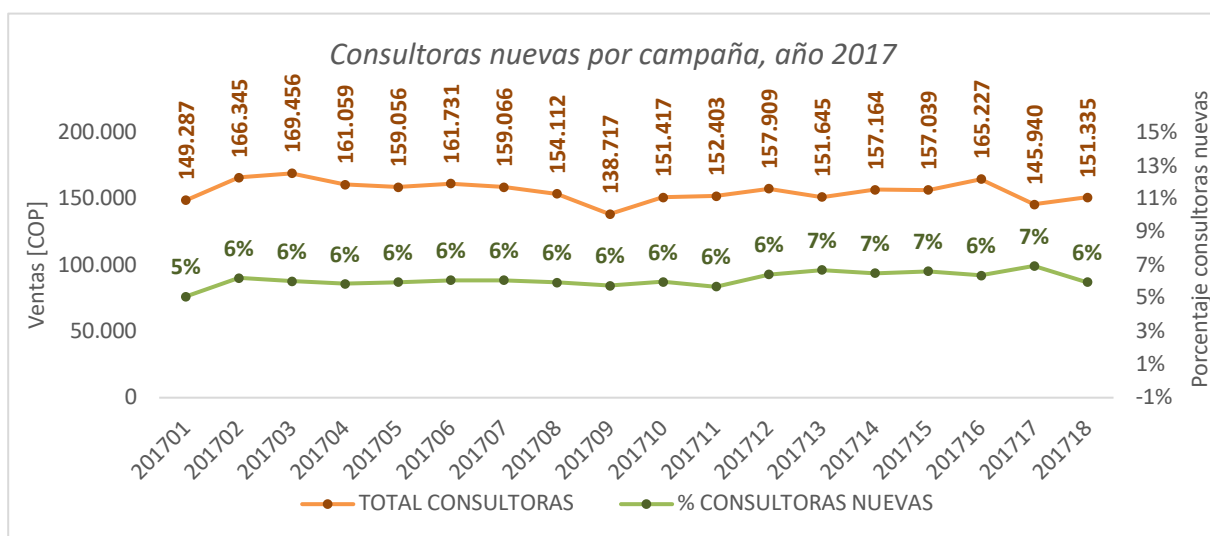


Gráfico 5: Total consultoras y porcentaje de consultoras nuevas en Colombia.
Fuente: Elaboración propia

Se define como consultoras nuevas aquellas consultoras que realizan su primer pedido en la campaña seleccionada.

Del gráfico anterior se concluye que en general la cantidad de consultoras que realizan pedidos por campaña es bastante constante y que en promedio un 6% son consultoras nuevas, por lo que se trabaja en general con consultoras que son frecuentes y tienen información histórica.

- **Transacciones (fvtaprobecam)**

Información de todas las transacciones de la empresa, contiene datos sobre el país donde se realizó la transacción, la campaña a la que pertenece, el producto que se compró, los valores de este, entre otros. Contiene un total de 550.178.153² de registros, donde los principales son los montos de venta, unidades vendidas y el de productos distintos por pedido, los que se detallan en la siguiente tabla.

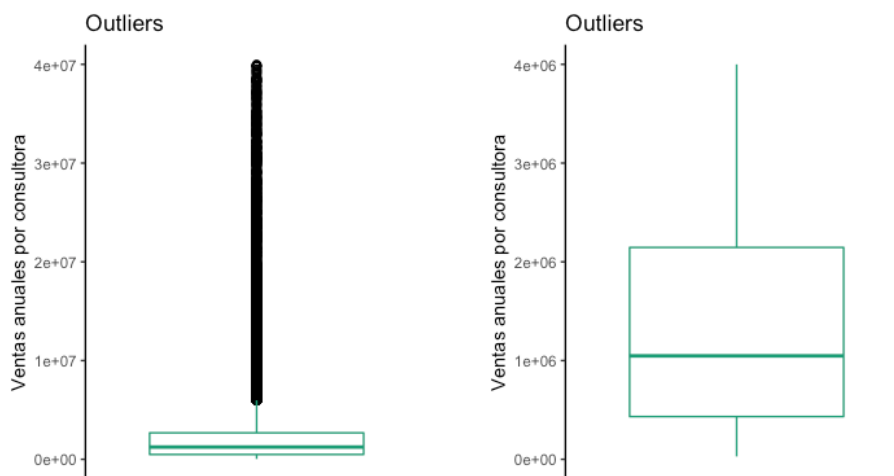
VARIABLE POR PEDIDO	MÍN.	1ST QU.	MEDIANA	PROMEDIO	3RD QU.	MÁX.
Ventas [COP ³]	13.200	168.114	207.124	254.110	272.444	113.519.729
Total unidades	1	11	15	19,57	22	9.413
Productos distintos	1	8	11	12,11	15	308

Tabla 2: Información de las principales variables de las transacciones
Fuente: Elaboración propia

² Detalle de los registros en el anexo K: tabla 2.

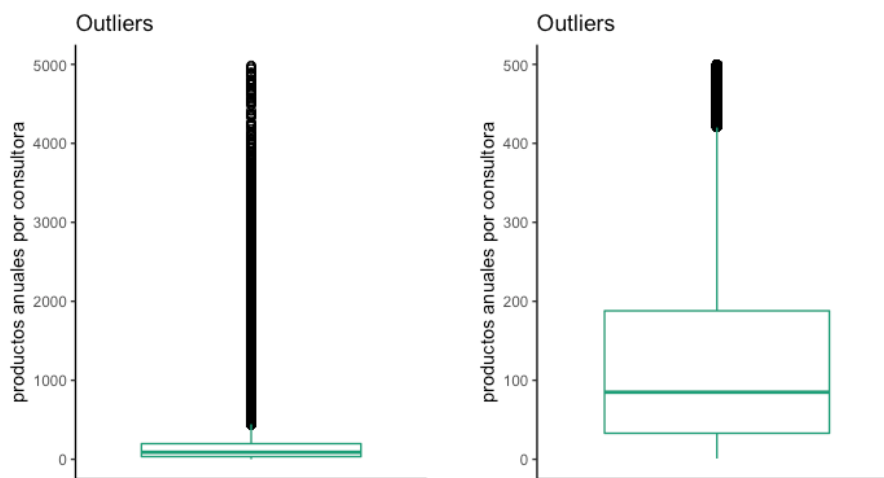
³ COP= peso colombiano

De la *tabla 2* se puede concluir que tanto las ventas, el total de unidades y los productos distintos promedio por pedido varían bastante en el año 2017, además se observa que los valores máximos se alejan bastante del promedio, lo que se refleja también en los siguientes gráficos de estas variables. En el *gráfico 7* se puede observar que en promedio una consultora vende COP\$ 1 millón al año, lo que es equivalente a USD\$ 320. De la base de transacciones adicionalmente se puede extraer la cantidad de consultoras que realizaron pedidos por campaña y las ventas que se realizan en cada una de estas⁴.



*Gráfico 6 y 7: Gráficos Box-plot para identificar outliers en las ventas [COP] anuales por consultora.
Fuente: Elaboración propia*

El *gráfico 9* muestra que en promedio se venden 100 productos en un año por cada consultora.



*Gráfico 8 y 9: Gráfico Box-plot para identificar outliers en el número de productos por consultora.
Fuente: Elaboración propia*

⁴ Ver detalle de las consultoras que realizan pedidos por campaña en anexo A: gráfico 1.

Para la limpieza de datos se desea identificar posibles outliers, como consultoras con ventas fuera de lo normal, o que venden grandes volúmenes de productos. En un inicio se hace el análisis utilizando gráficos Box-plot, donde todos aquellos casos que queden bajo el primer cuartil y sobre el tercer cuartil son considerados outliers, sin embargo, este método puede no ser muy preciso dejando una cantidad considerable de datos fuera.

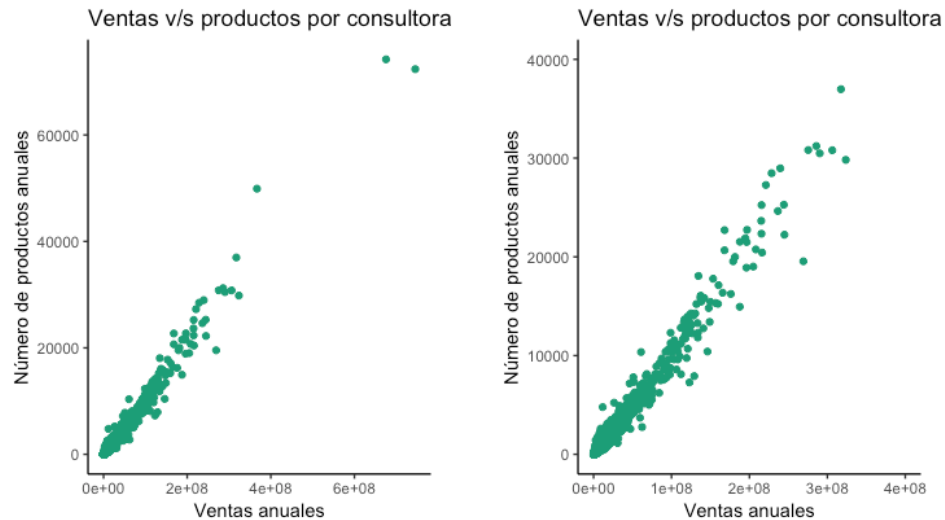


Gráfico 10: Número de productos por monto de venta [COP], por pedido.

Gráfico 11: Número de productos por monto de venta [COP], por pedido. Ejes reducidos.

Fuente: Elaboración propia

El gráfico 10 muestra la relación que existe entre el número de productos vendidos y el monto del pedido, como es de esperarse a mayor número de productos mayor es el valor del pedido. También se observan casos de consultoras que realizan pedidos fuera de lo normal, especialmente los casos ubicados en la esquina superior derecha del gráfico, de pedidos que se alejan del resto dada su gran dimensión. Estos casos serán considerados outliers y eliminados de las bases de datos.

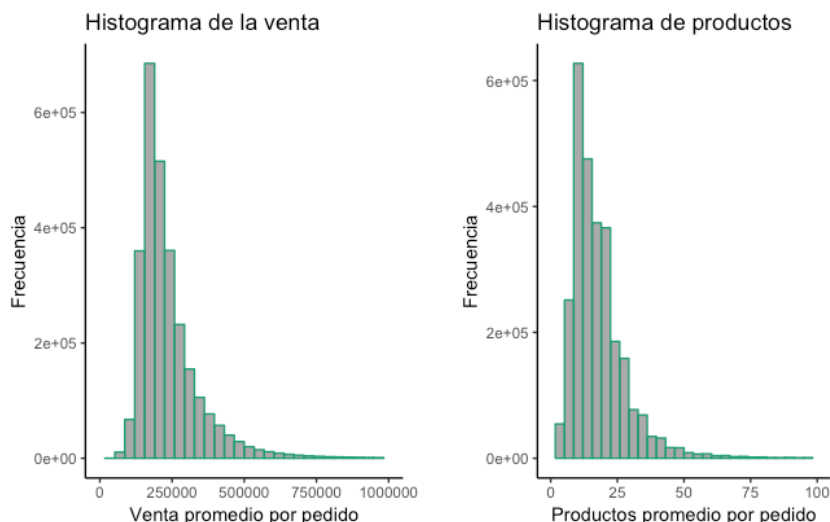


Gráfico 12: Histograma de las ventas promedio por pedido.

Gráfico 13: Histograma de los productos promedio por pedido.

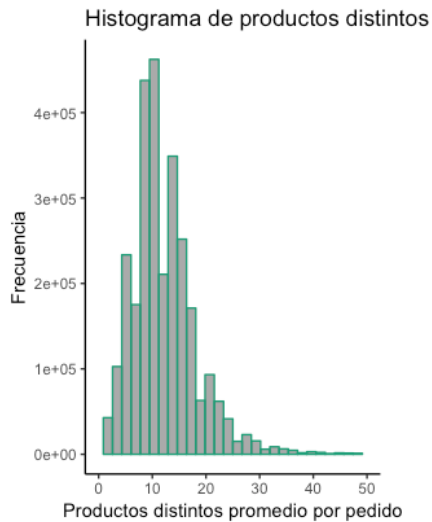
Fuente: Elaboración propia

Cabe mencionar que para obtener los gráficos 12, 13 y 14 se debió limitar el eje X cortando la cola del gráfico, esto era de esperarse dados los datos de la tabla 10 donde se observa que los valores máximos están muy alejados del valor promedio. En el gráfico 12, de ventas promedio por pedido⁵, se observa que las ventas se concentran alrededor de los COP\$ 250.000 (equivalente a USD\$ 80), donde una menor cantidad de consultoras logra superar los COP\$ 500.000. En cuanto a los productos promedio por pedido, se identifica que la mayoría de las consultoras compra alrededor de 10 productos por pedido⁶, y no muchas superan los 50 productos.

Finalmente, el gráfico 14 muestra la cantidad de productos distintos en promedio que compran las consultoras, donde se ilustra que en su mayoría tienden a comprar aproximadamente 10 productos distintos.

⁵ Ver detalle del valor promedio por pedido, agrupado por campaña en anexo C: gráfico 3.

⁶ Ver detalle del promedio de productos por pedido, agrupado por campaña en anexo B: gráfico 2.



*Gráfico 14: Histograma de los productos distintos en promedio por pedido.
Fuente: Elaboración propia*

- **Productos por campaña (dmatrizcamapaña)**

Información de los productos asociadas a cada campaña, donde se especifican los productos disponibles⁷ para los clientes tanto en el catálogo físico como en el sitio web. Contiene un total de 2.089.078 de registros⁸.

Adicional a esta tabla, existe otra tabla que muestra información específica de cada producto denominada como maestro de productos (dproductos en la base de datos). En ella se muestra la información detallada de cada producto que la empresa vende como su tamaño, marca, categoría, tipo de producto, entre otros⁹.

⁷ Detalle de la distribución de los productos disponibles por categoría en anexo N: tabla 5.

⁸ Detalle de los registros en el anexo M: tabla 4.

⁹ Detalle de los registros en el anexo L: tabla 3.

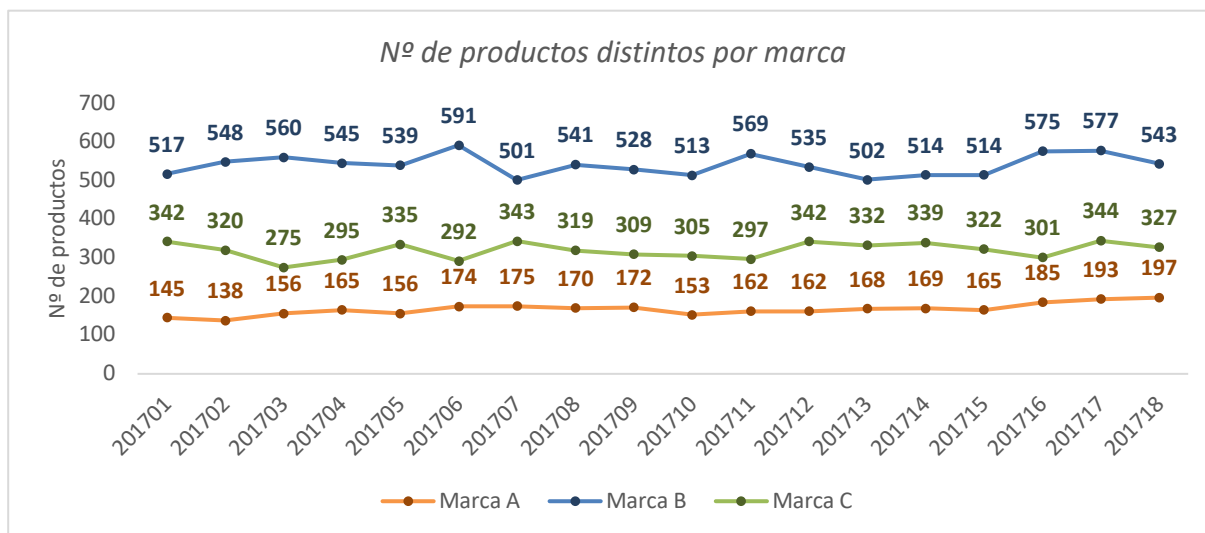


Gráfico 15: Número de productos distintos por marca, por campaña.
Fuente: Elaboración propia

El gráfico 15 muestra los distintos productos disponibles en cada campaña, por cada marca. Se concluye que la marca B es la que tiene el catálogo más amplio en cuanto a variedad de productos. Por otro lado se observa que en promedio se cuenta con 1.000 productos distintos en cada campaña.

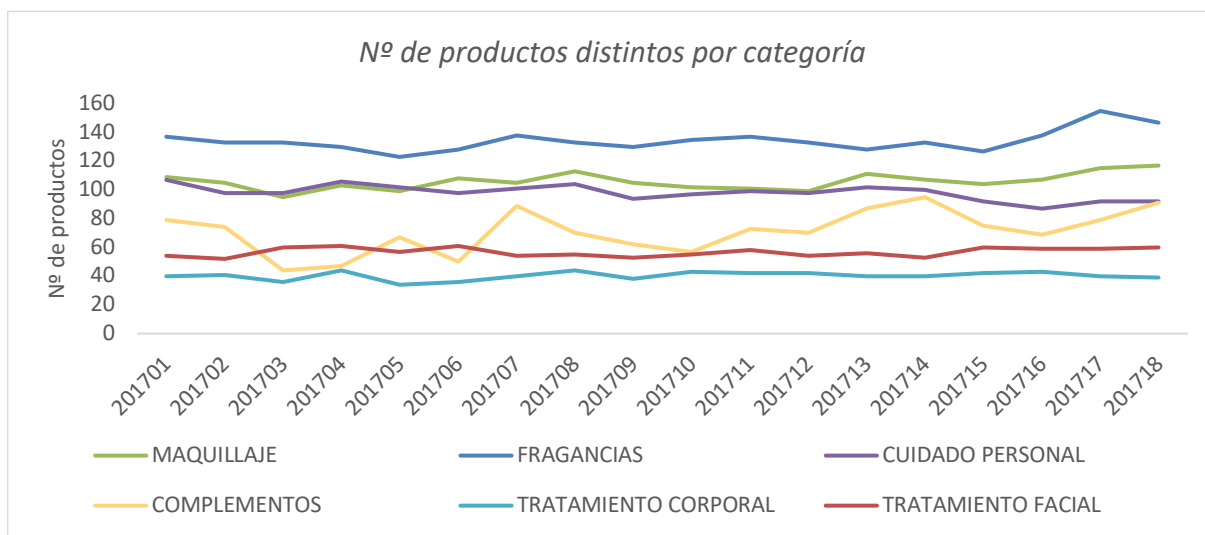


Gráfico 16: Número de productos distintos por categoría, por campaña.
Fuente: Elaboración propia

La empresa cuenta con un total de 17 categorías distintas, donde las 6 principales dada la variedad de productos con los que cuentan son: fragancias, maquillaje, cuidado personal, complementos, tratamiento facial y tratamiento corporal, tal como se muestra en el gráfico 16. Estas categorías representan más del 90% de las ventas anuales de la

empresa, alcanzando montos por USD\$ 224.122.539 en el año 2017. Se observa que en general la cantidad de productos distintos por categoría es constante, a excepción de la categoría complementos que varía entre los 40 y 100 productos distintos por campaña. En el *gráfico 17* que se muestra más abajo, se puede concluir que la marca B, la cual tenía la mayor variedad de productos, es la que lidera con un mayor número de productos vendidos en las principales categorías durante ese año.

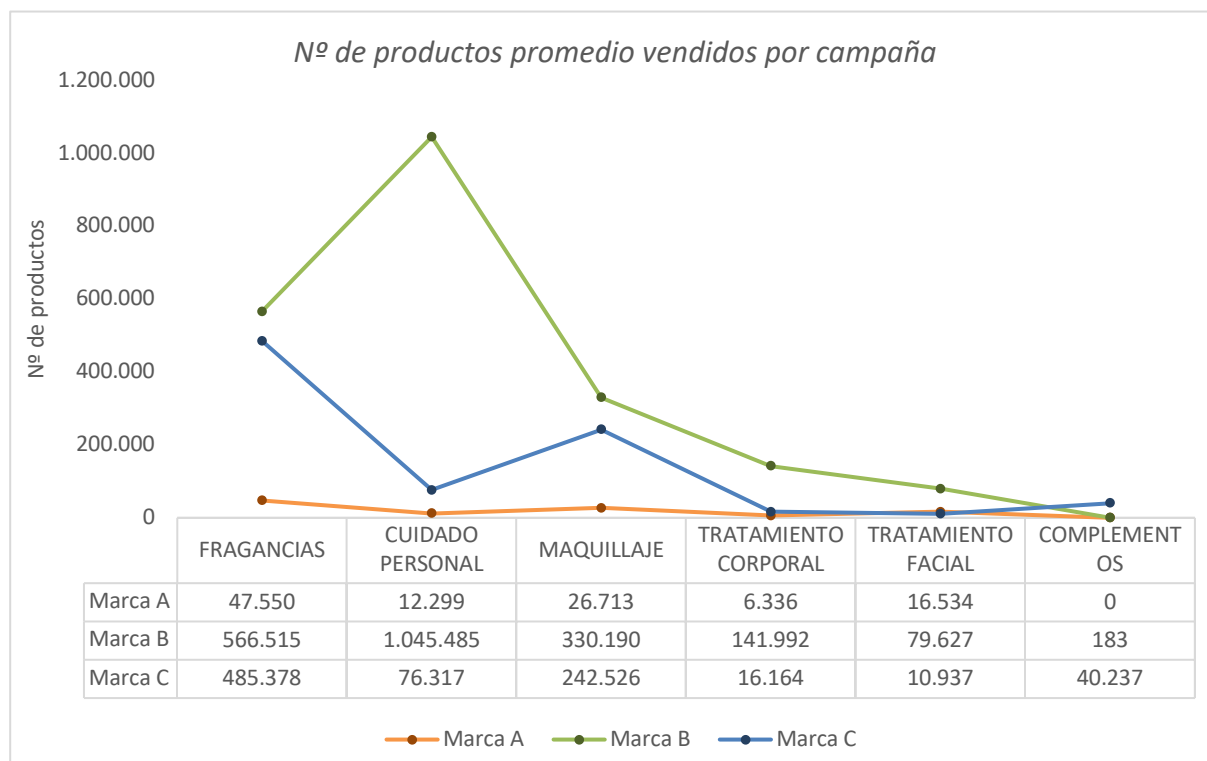


Gráfico 17: Número de productos promedio por campaña, vendidos en las campañas del año 2017. Agrupados por categoría. Fuente: Elaboración propia

4.3. Preparación de los datos

Para poder comenzar con el diagnóstico del MDO se seleccionaron los datos con los que se decidió trabajar en las secciones anteriores, filtrando las bases de datos principales por país y por fecha de las campañas, de manera de facilitar la ejecución de las consultas en SQL. Algo importante a considerar es que, dado que cambió la versión del MDO, los resultados de los indicadores se calculan sólo para las campañas donde se ejecutó el modelo mejorado, esto para no ensuciar las conclusiones mezclando datos de ambos modelos. Las campañas por evaluar son desde la 08 a la 15 del año 2018.

Penta Analytics no contaba con una evaluación anterior del MDO de la empresa, por lo que fue necesario partir desde cero, creando indicadores acordes con lo que se

deseaba saber de las recomendaciones y verificar la existencia de los datos necesarios para poder realizar los cálculos. Se decide dividir los indicadores en dos grupos, el primero correspondiente a indicadores típicos para evaluar un motor de recomendaciones, como la tasa de conversión o los montos en dólares que se obtienen por el canje de ofertas. Mientras que el segundo grupo corresponde a una propuesta para poder evaluar el sesgo de popularidad de dicho recomendador.

4.3.1. Indicadores típicos

$Efecto\ directo_t = Monto\ de\ venta\ de\ las\ ofertas\ MDO\ compradas_t$

$$\% \text{ consultoras}_t = \frac{N^{\circ} \text{ de consultoras que compran ofertas MDO}_t}{\text{Consultoras totales que realizan un pedido}_t}$$

$$\% \text{ venta}_t = \frac{\text{Monto de venta de las ofertas MDO compradas}_t}{\text{Monto de venta del total de pedidos realizados}_t}$$

$$\text{Conversión}_t = \frac{N^{\circ} \text{ de ofertas MDO compradas}_t}{N^{\circ} \text{ de ofertas MDO mostradas}_t}$$

Donde existen las siguientes definiciones:

- $N^{\circ} \text{ de consultoras que compran ofertas MDO}_t$: consultoras que han comprado al menos un producto de las recomendaciones del MDO en la campaña t .
- $\text{Consultoras totales que realizan un pedido}_t$: consultoras totales que han realizado un pedido en la campaña t .
- $\text{Monto de venta de las ofertas MDO compradas}_t$: monto de las ventas en dólares de los productos recomendados por el MDO en la campaña t .
- $\text{Monto de venta del total de pedidos realizados}_t$: monto de las ventas en dólares del total de productos vendidos en los pedidos de la campaña t .
- $N^{\circ} \text{ de ofertas MDO compradas}_t$: número de ofertas recomendadas por MDO que fueron compradas en la campaña t .
- $N^{\circ} \text{ de ofertas MDO mostradas}_t$: número del total de ofertas del MDO que fueron mostradas en los perfiles de las consultoras en la campaña t .

4.3.2. Resultados indicadores típicos

Para las campañas 2018-08 a 2018-15 se obtuvieron los siguientes resultados de los indicadores:

EFEECTO DIRECTO	% CONSULTORAS	% VENTA	CONVERSIÓN
USD \$145.000	4,88%	1,11 %	0,22 %

Tabla 3: Promedio de indicadores típicos por campaña.
Fuente: Elaboración propia

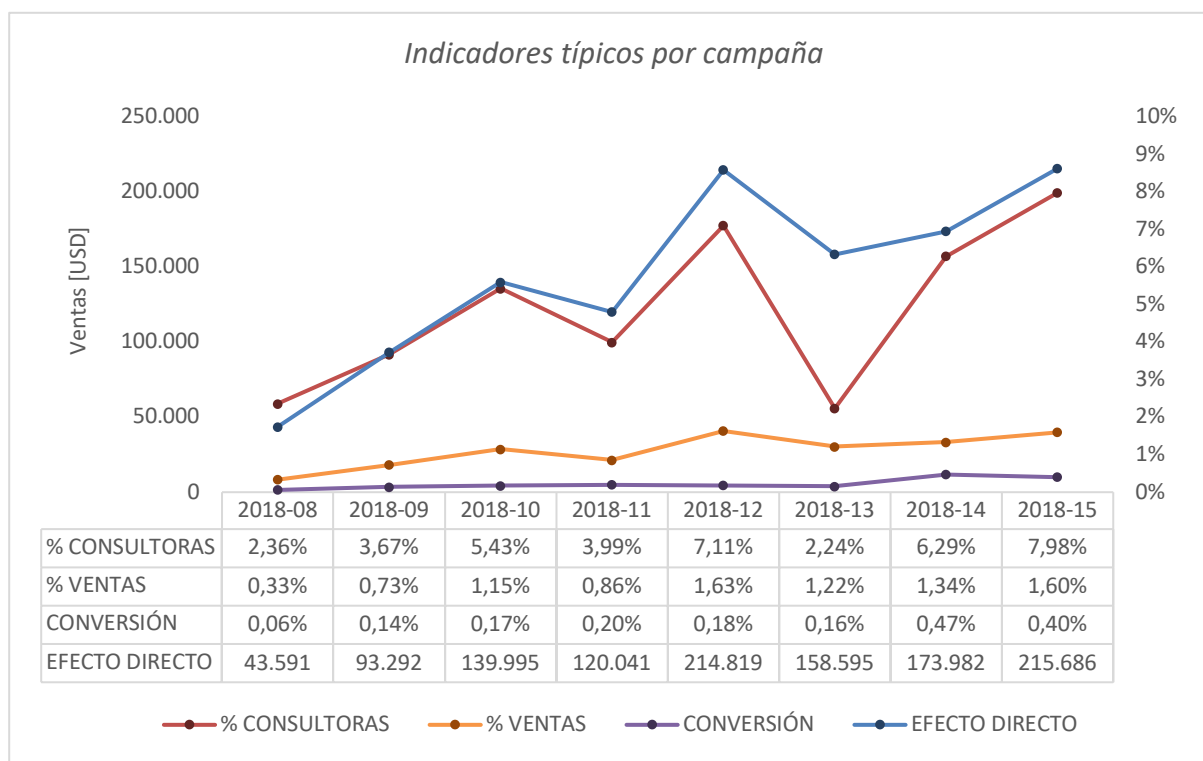


Gráfico 18: Resultados de indicadores típicos por campaña.
Fuente: Elaboración propia

Se puede concluir que en las últimas campañas los indicadores han tenido una tendencia al alza, especialmente el de efecto directo¹⁰ y el porcentaje de consultoras que canjean ofertas en el sitio web. Esta alza puede explicarse debido a que el modelo del MDO con sus mejores lleva poco tiempo en funcionamiento, por lo que esta en un proceso de adaptación. Es de esperar que en las siguientes campañas las indicadores se establezcan mostrando un crecimiento menos pronunciado.

¹⁰ Detalle de los resultados del indicador efecto directo por tipo de oferta en anexo E: gráfico 5.

Se observa que la tasa de conversión¹¹ ha presentado mejoras a lo largo de las campañas, presentando su mejor resultado durante la campaña 14 del 2018, con una conversión de 0,47%.

4.3.3. Indicadores de sesgo de popularidad

$$\text{Cobertura de catálogo}_t = \frac{N^{\circ} \text{ de productos distintos comprados}_t}{N^{\circ} \text{ de productos distintos del catálogo}_t}$$

$$\text{Cobertura MDO}_t = \frac{N^{\circ} \text{ de productos distintos en las ofertas MDO}_t}{N^{\circ} \text{ total de productos distintos del catálogo}_t}$$

$$\text{Novedad}_t = \frac{N^{\circ} \text{ de MCT novedosas compradas}_t}{N^{\circ} \text{ total de MCT compradas}_t}$$

$$\text{Novedad MDO}_t = \frac{N^{\circ} \text{ de MCT novedosas en las ofertas MDO}_t}{N^{\circ} \text{ total de MCT en las ofertas del MDO}_t}$$

Donde existen las siguientes definiciones:

- *Nº de productos distintos comprados_t*: número de productos distintos comprados por el total de consultoras en la campaña *t*
- *Nº de productos distintos del catálogo_t*: número de productos distintos disponibles en los catálogos de la campaña *t*.
- *Nº de productos distintos en las ofertas MDO_t*: número de productos distintos dentro del total de productos ofertados por el MDO la campaña *t*.
- *Nº de MCT novedosas compradas_t*: número de MCT compradas por las consultoras en la campaña *t*, y que no han sido compradas en las últimas 6 campañas por la consultora.
- *Nº total de MCT compradas_t*: número total de MCT compradas por las consultoras en la campaña *t*.
- *Nº de MCT novedosas en las ofertas MDO_t*: número de MCT de las ofertas del MDO mostradas a las consultoras en la campaña *t*, y que no han sido compradas en las últimas 6 campañas por la consultora.
- *Nº total de MCT en las ofertas del MDO_t*: número total de MCT que se muestran en las ofertas del MDO en la campaña *t*.

¹¹ Detalle de la tasa de conversión por tipo de oferta en anexo D: gráfico 4.

4.3.4. Resultados indicadores de sesgo de popularidad

Para las campañas 2018-08 a 2018-15 se obtuvieron los siguientes resultados de los indicadores:

COBERTURA	COBERTURA MDO	NOVEDAD	NOVEDAD MDO
84,3 %	6,2 %	42 %	49 %

Tabla 4: Promedio de indicadores de sesgo de popularidad por campaña.
Fuente: Elaboración propia

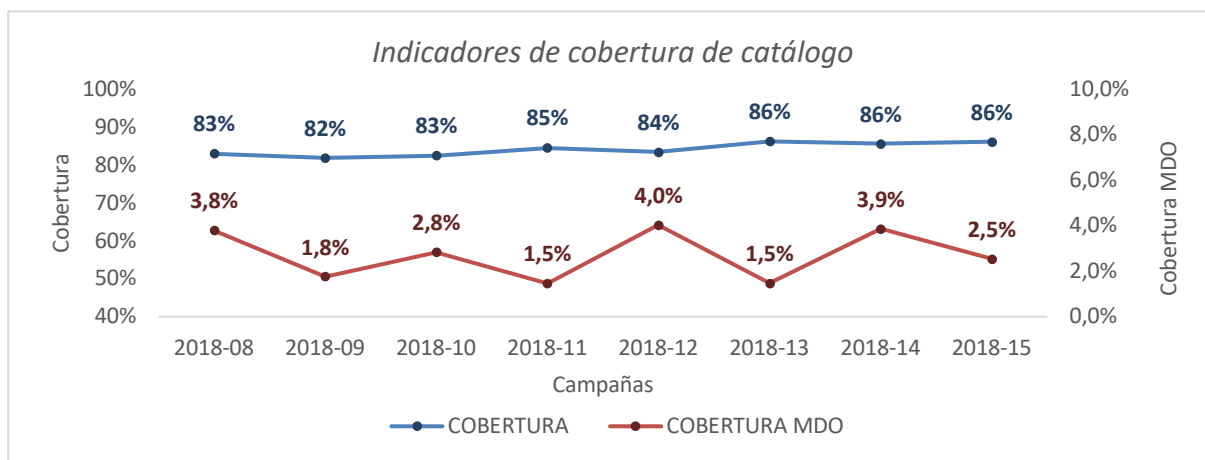


Gráfico 19: Resultados de indicadores de cobertura de catálogo por campaña.
Fuente: Elaboración propia

Con el indicador de cobertura, se puede concluir que las consultoras compran un porcentaje alto del total de productos disponibles en los catálogos. Por otro lado, con el indicador de novedad¹² se observa que tienden a comprar un porcentaje no menor de MCT nuevas, lo que muestra que no son consultoras que compren siempre lo mismo.

En cuanto a los indicadores del MDO, se cubre sólo un 6,2% del catálogo, esto podría no ser tan bueno dada la gran cantidad de ofertas que se muestra por consultora, lo que podría evidenciar que en general a las consultoras se les están mostrando ofertas similares. En cuanto a que tan novedosas son las MCT para las consultoras, se tiene que casi la mitad de las MCT mostradas no han sido compradas en los últimos 6 meses por las consultoras.

¹² Detalle del indicador de novedad por campaña en anexo O: tabla 6.

4.4. Modelo de propensión de compra


El objetivo del modelo de propensión de compra es encontrar los productos que tienen mayor probabilidad de ser comprados por cada consultora, de forma personalizada en el futuro.

4.4.1. Modelo de filtros colaborativos

En un inicio para el modelo de propensión de compra, se usa la técnica de filtros colaborativos a nivel de MCT, por lo que es necesario crear la matriz de rating o interacciones de las consultoras con las MCT. Los datos se calculan usando los periodos de tiempo entre la campaña 2018-09 y la campaña 2018-14.

La matriz de rating de este modelo tiene en la parte superior todas las MCT, que en total son 233, y en sus columnas la información de cada consultora que haya comprado dicha MCT en las campañas establecidas anteriormente. Estos datos van de 1 a 6, donde 1 significa que la consultora compró 1 vez la MCT en las campañas, y 6 si compró la MCT en las 6 campañas (Ver *ilustración 4*). También cabe la posibilidad de que la consultora no haya comprado nunca la MCT, en dicho caso el valor es NA.

Es importante mencionar que se cuenta con un total de 260.665 consultoras que realizan pedidos entre las campañas 9 y 14 del 2018. Para poder cumplir con la capacidad de procesamiento del software se selecciona una muestra aleatoria de 10.000 consultoras, conjunto con el que se trabajará en todos los modelos futuros.



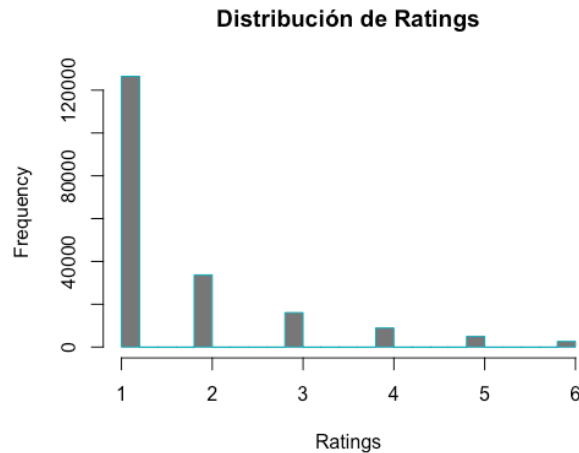
	MCT_1	MCT_2	...	MCT_m
u_1	1	6		NA
u_2	NA	NA		1
...				
u_n	4	5		1

	MCT_1	MCT_2	...	MCT_m
u_1	NA	NA		3,5
u_2	2	1		NA
...				
u_n	NA	NA		NA

*Ilustración 4: Matriz de rating filtros colaborativos.
Fuente: Elaboración propia*

A la izquierda de la *ilustración 4* se puede observar la matriz de rating de input para el modelo de filtros colaborativos, mientras que al lado derecho se ve la matriz de rating de output del modelo, donde se predice un valor para cada NA existente en la primera matriz y se asigna un NA en las MCT que ya tenían un rating dado. Los modelos de filtros colaborativos se programan usando el software R, específicamente el paquete *recommenderlab* [7].

Para poder entender como se comportan los rating de la matriz, se genera el siguiente gráfico que muestra la distribución que tienen los rating de las MCT:



*Gráfico 20: Distribución de ratings de las MCT.
Fuente: Elaboración propia*

Mín.	1st Qu.	Mediana	Promedio	3rd Qu.	Máx.	NA's
1	1	1	1,6	2	6	2.281.237

*Tabla 5: Distribución de ratings de las MCT.
Fuente: Elaboración propia*

Se observa que la mayor parte de los rating está entre 1 y 2, y muy pocas consultoras compran una MCT en todas las campañas seleccionadas. Además, un 91% de los datos son NA's, lo cual podría afectar la precisión del modelo.

De estos datos se puede inferir que las consultoras no son constantes con sus compras, es decir, van variando los productos que compran campaña a campaña. Analizando el rating promedio por consultora se observa que dentro de todas las MCT que compran en un periodo de 6 campañas, la mayoría de las MCT son comprada una sola vez, lo que genera un promedio de rating igual a 1,5.

Mín.	1st Qu.	Mediana	Promedio	3rd Qu.	Máx.
1	1,19	1,522	1,515	1,792	3,218

*Tabla 6: Distribución de rating promedio por consultora.
Fuente: Elaboración propia*

El modelo de filtros colaborativos que se aplica calcula las predicciones de la matriz de rating usando la correlación de Pearson, esto para calcular la similitud que existe entre las consultoras. Se calculan 3 modelos distintos, diferenciados en la normalización de los datos:

El primer modelo denominad UBCF_N_P en el software, deja los datos tal cual fueron ingresados sin normalizarlos, el segundo (UBCF_C_P) realiza una normalización restando la media a los ratings, y el último modelo (UBCF_Z_P) normaliza los datos restando la media y dividiendo los rating por su desviación estándar. Los resultados de los distintos tipos de errores se observan en la siguiente ilustración:

	RMSE	MSE	MAE
UBCF_N_P	0,5899152	0,3480000	0,3030769
UBCF_C_P	0,9280948	0,8613600	0,8980156
UBFC_Z_P	0,9549455	0,9119209	0,9132601

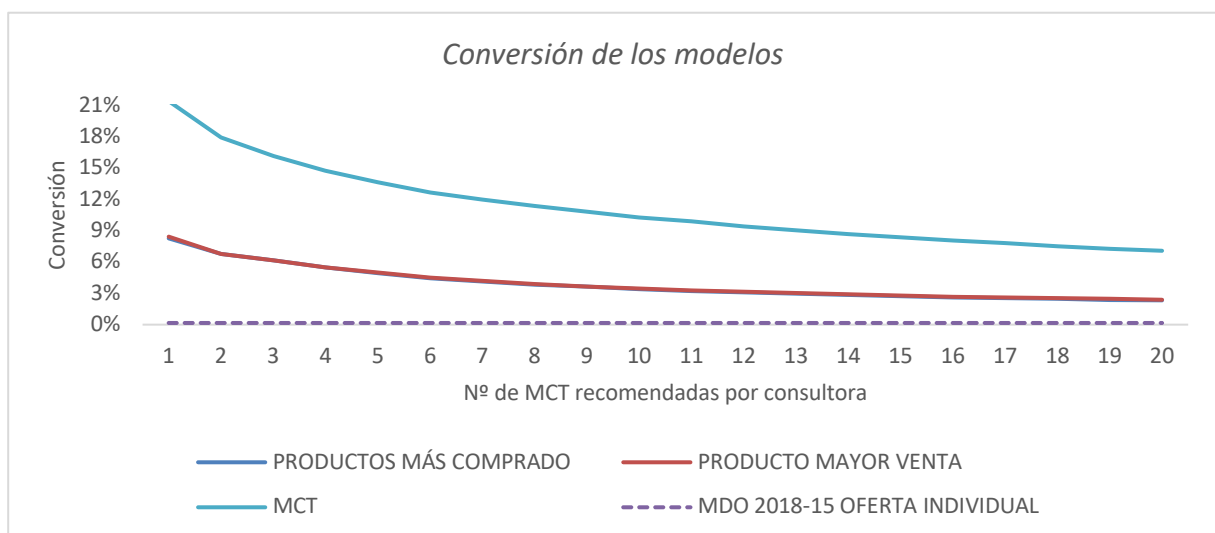
*Ilustración 7: Errores de predicciones matriz de rating.
Fuente: Resultados de errores del modelo en el software R.*

Dado que el modelo que tiene errores mas bajos es el UBCF_N_P, cuando no hay normalización de los datos, se continúa trabajando solo con este en las etapas futuras.

La matriz de rating de output del modelo muestra que tan propensa será una consultora para comprar una MCT que no había adquirido en los periodos anteriores. Entre más alto es el valor, más probabilidades hay de que la consultora compre dicha MCT en el futuro.

Una vez que se tiene la matriz de output que arroja el modelo, se ordenan las MCT de mayor a menor rating y se seleccionan las 20 primeras. Se decide seleccionar 20 MCT dado que el MDO recomienda en promedio 19 ofertas individuales a las consultoras. Esto ayudará a que al momento de comparar resultados de ambos modelos, las conclusiones sean más certeras.

Al momento de seleccionar el productos se decide probar dos posibles modelos, el primero selecciona el producto más comprado en cuando a unidades de cada MCT, y el segundo el producto con mayores ventas en dinero para cada MCT. Para verificar que modelo es mejor se calcula la conversión que hubieran tenido estas recomendaciones en la campaña 2018-15.



*Gráfico 21: Conversión a nivel de MCT y de producto.
Fuente: Elaboración propia*

En el *gráfico 21* se observan los resultados de la conversión de los modelos de filtros colaborativos descritos anteriormente, además se integra la conversión a nivel de MCT y la conversión promedio que tuvo el MDO en las ofertas individuales. Se puede concluir que ambos modelos tienen una conversión mayor al MDO a nivel de producto, y que el modelo con el producto de mayor venta está levemente por encima que el más comprado¹³. También se ve que a medida que se recomienda un mayor número de MCT por consultora la conversión va bajando, esto se explica porque cada MCT nueva que se va agregando tiene un menor rating que la anterior y por ende una menor probabilidad de ser comprada.

En cuanto a temas de sesgo de popularidad, el nuevo modelo de filtros colaborativos presenta mejoras con respecto al MDO. El siguiente gráfico muestra variedad de categorías, número de MCT presentes en cada modelo y el porcentaje del total de MCT que los modelos logran mostrar en las recomendaciones entregadas a las consultoras, respecto al total de MCT con las que trabaja el modelo.

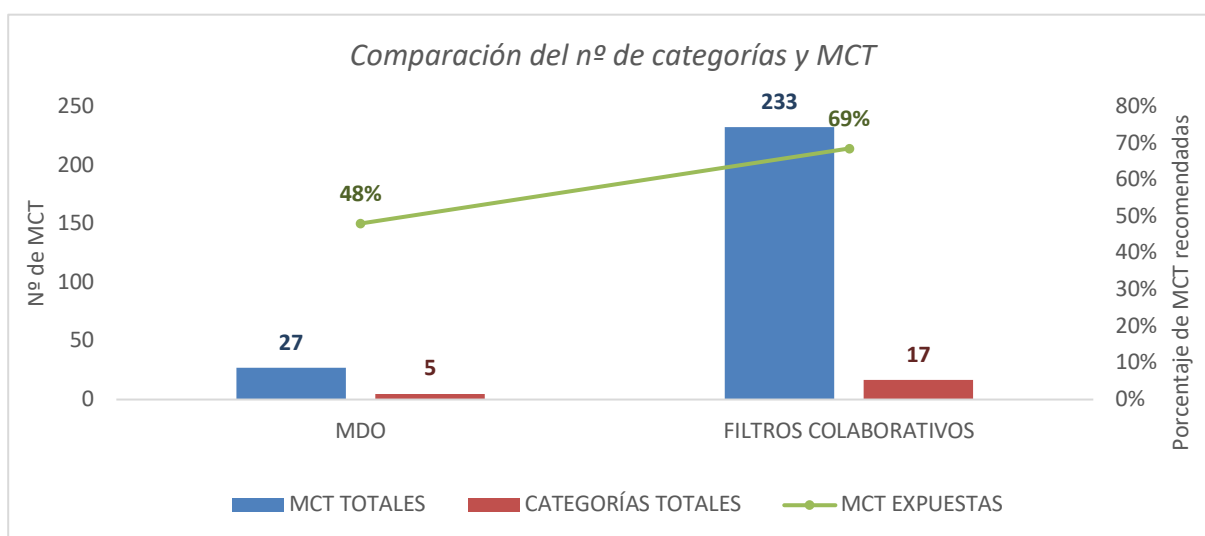


Gráfico 22: Comparación nº de categorías y de MCT de ambos modelos.
Fuente: Elaboración propia

¹³ Detalle de los resultados de la tasa de conversión para cada modelo en anexo 3, tabla 6.

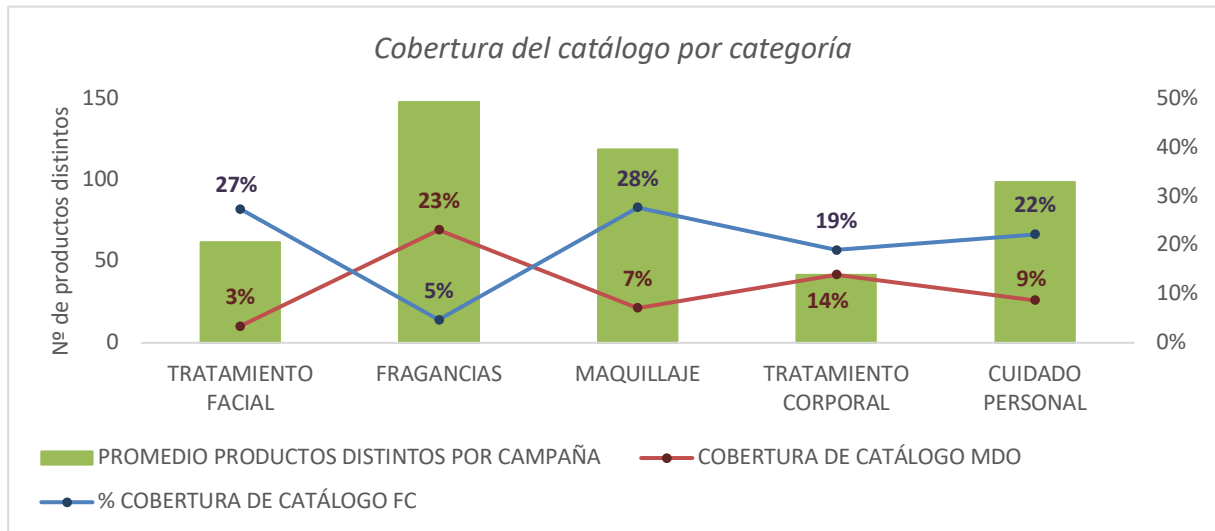


Gráfico 23: Cobertura del catálogo por categoría.
Fuente: Elaboración propia

En el gráfico 23 se observa que en casi todas las categorías el modelo de filtros colaborativos logra una mayor cobertura de catálogo, alcanzando su máximo con un 28% en la categoría de maquillaje. En el caso de la categoría fragancias se observa una cobertura de catálogo bastante menor en comparación con el resto, esto se debe a que esta es la categoría más comprada y también la más recomendada por el MDO, por lo que es de esperar que sea menos recomendada en el nuevo modelo por su popularidad.

4.4.2. Modelo de filtros colaborativos con novelty score

Con el modelo de filtros colaborativos se obtienen mejores resultados en temas de conversión y cobertura que con el MDO, sin embargo, se desea resolver de mejor forma los temas ligados al sesgo de popularidad. Para esto se ocupa una metodología de filtros colaborativos con novelty score, en la cual se segmentan las consultoras y se agrega la variable de novedad, la cual nos indica cuáles son las categorías menos populares. Esta nueva metodología tiene 4 principales etapas:

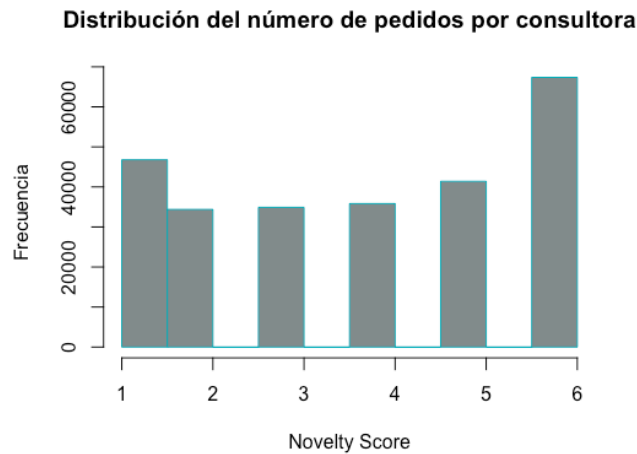
1. Categorización de las consultoras en frecuentes y ocasionales, según la variable θ que indica la cantidad de pedidos que realiza una consultora en un periodo dado.
2. Calcular la relevancia de cada MCT para todas las consultoras, la cual indica que tan propensa es la consultora a comprar una MCT específica en el futuro.
3. Seleccionar elementos no comprados para cada consultora.

4. Calcular el novelty score de cada MCT, para identificar que tan novedosa o popular es dicha MCT

A continuación, se presenta el desarrollo de cada una de las etapas del filtro colaborativo con novelty score.

1. Categorización de las consultoras

Se define la variable θ como el número de pedidos que realiza una consultora entre la campaña 09 y la campaña 14 del año 2018. Esta variable toma valores entre 1 y 6 para cada consultora.



*Gráfico 24: Distribución de la variable θ .
Fuente: Elaboración propia*

En el *gráfico 24* se observa que gran parte de las consultoras han realizado pedidos en las 6 campañas seleccionadas, y más de la mitad de estas realizan 4 o más pedidos en el periodo seleccionado.

Mín.	1st Qu.	Mediana	Promedio	3rd Qu.	Máy.
1	2	4	3,74	6	6

*Tabla 8: Distribución de la variable θ .
Fuente: Elaboración propia*

Una vez que se entiende la distribución de la variable θ se define la variable θ^* , la cual será el valor límite para diferenciar entre consultoras frecuentes y ocasionales. Las consultoras frecuentes serán aquellas que posean un θ igual o mayor a θ^* , mientras que las consultoras ocasionales deben tener un θ menor a θ^* . Por ejemplo, cuando θ^* es igual a 4, las consultoras frecuentes serán aquella que hayan realizado 4 o más pedidos durante el periodo estudiado. Luego, las consultoras ocasionales serán todas aquellas que hayan realizado 3 o menos pedidos.

El objetivo de segmentar las consultoras es aplicar el nuevo modelo de FC con novelty score sólo a las consultoras frecuentes, mientras que a las consultoras ocasionales se les aplica el modelo de filtros colaborativos regular.

θ^* toma los valores 1, 3, 4 y 5 dejando de lado los casos de θ^* igual a 2 e igual a 6, esto dado que con el primer valor se crea un conjunto de consultoras ocasionales muy poco atractivo (por el número de consultoras) para inspeccionar. De igual forma al tomar θ^* igual a 6, el conjunto de consultoras frecuentes es muy pequeño, por lo que no es de interés analizarlo. Se realizan las simulaciones necesarias para comprobar como se comporta el modelo en cada uno de los posibles valores de θ^* , para luego seleccionar el mejor valor.

2. Relevancia para cada conjunto consultora – MCT

La relevancia nos indica que tan propensa es una consultora a comprar una MCT específica. Para esto, en primer lugar es necesario calcular la similitud, según sus ratings, entre las consultoras ocupando la correlación de Pearson. Se obtiene el siguiente gráfico de la distribución de las similitudes:

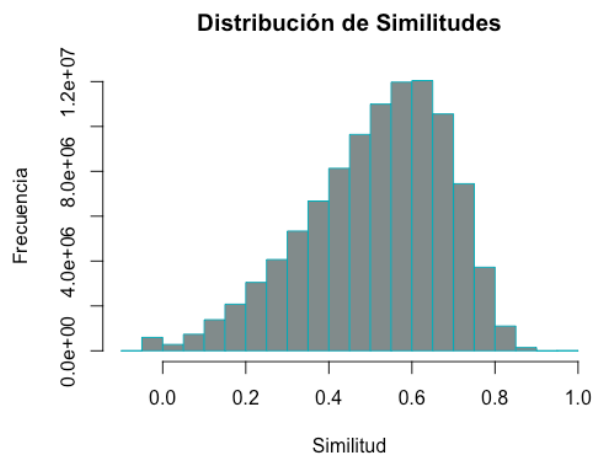


Gráfico 25: Distribución de las similitudes de rating entre consultoras
Fuente: Elaboración propia

Del gráfico 25 se puede concluir que las consultoras son bastante similares entre sí, donde gran parte de las similitudes entre consultoras está por sobre el 0.5. Esto quiere decir, que el comportamiento de compra entre las consultoras es bastante similar.

A continuación, se presenta la fórmula que ocupa el modelo para calcular la relevancia que tiene una MCT (i) para una consultora (u):

$$\hat{r}(u, i) = \bar{r}(u) + \sum_{v \in N_k(u, i)} sim(u, v)(r(v, i) - \bar{r}(v))$$

Dada esta ecuación, se pueden identificar otros datos importantes para calcular la relevancia, como el rating que cada consultora tiene respecto a las MCT ($r(v, i)$) y el rating promedio ($\bar{r}(v)$) que tiene cada consultora. El primero es definido como el número de campañas en que una consultora compra dicha MCT para un periodo de 6

campañas. Este rating va de 0 a 6, tomando valor cero cuando nunca ha sido comprada la MCT por la consultora, y seis cuando la MCT es comprada en todas las campañas del periodo seleccionado. Por otro lado, el rating promedio¹⁴ es un promedio simple de todos los rating que tiene una consultora.

La relevancia esta conformada por dos términos principales, el primero es el rating promedio de la consultora ($\bar{r}(u)$), y el segundo es una sumatoria que toma un vecindario de las consultoras mas similares, ponderando según que tan similares son y por la diferencia entre el rating de la MCT y el rating promedio que tiene esa consultora. Dada esta ecuación se obtiene el siguiente gráfico con la distribución de las relevancias entre cada consultora y MCT:

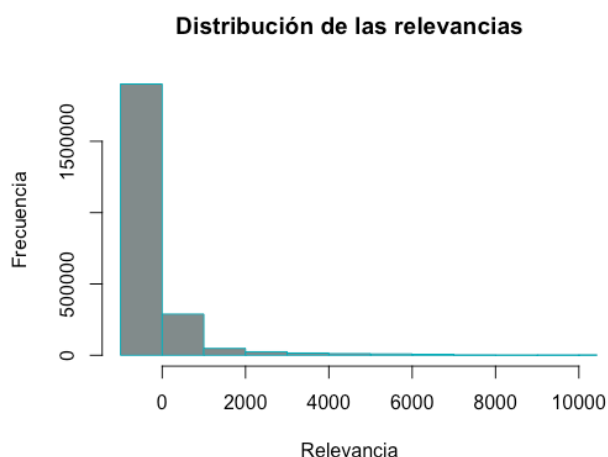


Gráfico 26: Distribución de las relevancias entre cada consultora y MCT.
Fuente: Elaboración propia

La relevancia posee algunos valores mayores que cero y otros menores. Los valores negativos nos muestran los casos en que la mayor parte de las consultoras, similares a la consultora objetivo, no han comprado la MCT antes, o la han comprado menos veces que su promedio de compras por MCT.

3. Elementos no comprados por las consultoras

Se identifican los elementos que las consultoras no han comprado en las últimas seis campañas, con el objetivo de recomendar de forma personalizada, sólo productos nuevos para cada consultora.

4. Novelty score para cada MCT

El novelty score se calcula para cada MCT y es independiente de las consultoras. El objetivo de este paso es poder crear un corte entre MCT novedosas y populares, siendo estas últimas las que siempre son compradas por las consultoras. Una vez que se identifican las MCT populares, estas son dejadas fuera del conjunto de MCT a recomendar a las consultoras, es decir, sólo se recomiendan MCT novedosas. El siguiente gráfico indica como distribuyen los valores de novelty score de las MCT:

¹⁴ Detalle de la distribución del rating promedio de las consultoras en anexo O: gráfico 6.

Distribución de Novelty Score

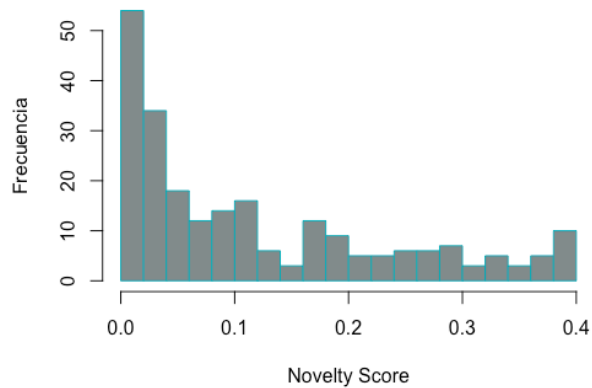


Gráfico 27: Distribución del novelty score para las MCT.
Fuente: Elaboración propia

Se prueban distintos cortes de novelty score (NS*), para poder identificar cual entrega mejores resultados. Del gráfico 27 se concluye que de las 233 MCT existentes más de 50 MCT son muy populares, dado que tienen un novelty score muy cercano a cero.

Luego de calcular todos los datos, y tener los valores finales de relevancia de los productos a recomendar a las consultoras, dejando de lado las MCT populares, se seleccionan las 20 MCT con mayor relevancia para cada consultora. Estas MCT a recomendar son testeadas con los datos existentes de las transacciones de las consultoras en la campaña 15. Esto, con el objetivo de obtener resultados sobre como se hubieran comportado las recomendaciones en caso de ser expuestas a las consultoras.

Es importante mencionar que dado el costo en tiempo que significa probar cada caso, se testearon solo 7 modelos (ver tabla 6). Para esto, se fueron iterando las variables a testear en cada modelo: θ^* , Sim^* y NS^* . Se comenzó con datos bases (Modelo 1), luego para el siguiente modelo se mantienen dos variables fijas y se modifica la tercera. Por ejemplo, el Modelo 2 mantiene Sim^* y NS^* del Modelo 1, pero varía θ^* . Algunos de los resultados de los 7 modelos probados se encuentran a continuación en la tabla 6.

MODELOS	Variables a testear			Resultados obtenidos			
	θ^*	Sim*	NS*	CONSULTORAS	MCT	CONVERSIÓN	VENTAS [USD]
MODELO 1 NS	5	0,6	0,02	4.237	179	2,30%	8.066
MODELO 2 NS	4	0,6	0,02	5.566	179	2,37%	10.245
MODELO 3 NS	3	0,6	0,02	6.891	179	2,47%	12.150
MODELO 4 NS	3	0,6	0,08	6.891	115	2,78%	12.186
MODELO 5 NS	3	0,8	0,08	6.891	115	2,35%	11.480
MODELO 6 NS	3	0,75	0,08	6.891	115	2,47%	11.985
MODELO 7 NS	1	0,6	0,08	10.000	115	2,82%	16.552

Tabla 9: Resultados modelos de filtros colaborativos con novelty score.

Fuente: Elaboración propia

En la tabla 9 la columna *modelos* indica el nombre del modelo testeado. θ^* indica el valor límite de esta variable para segmentar entre consultoras frecuentes y ocasionales en el modelo, y *consultoras* indica el número de consultoras pertenecientes al segmento frecuentes. La columna *Sim** muestra el valor de similitud límite con que se selecciona el vecindario de consultoras similares. *NS** indica el corte con que se segmenta entre MCT novedosas y populares, y *MCT* indica el número de MCT, valga la redundancia, que forman parte del segmento de novedosas. Por último la columna *conversión* indica la tasa de conversión de ese modelo, y *ventas* el monto en dólares de los productos que fueron canjeados por las consultoras en la simulación.

Para calcular el mejor valor de similitud, se decide fijar en un inicio como valor *Sim** igual a 0,6, esto dado que es el valor promedio de las similitudes entre consultoras. De la *tabla 6* se puede concluir que al usar una similitud más pequeña (*Sim**) para seleccionar el vecindario de las consultoras similares, se obtienen mejores resultados de conversión. Por otro lado, al fijar el novelty score límite para definir que productos se pueden recomendar y cuales no, se observa que al ser más estricto con lo que sería un producto novedoso se obtienen mejores resultados. Sin embargo, se debe llegar a un límite de *NS** igual a 0,08 para tener un conjunto suficiente de MCT a recomendar a las consultoras. Por último, se observa que es mejor no segmentar a las consultoras dejando θ^* igual a 1 y así aplicar a todas el modelo por igual. Basándose en los resultados de conversión, se decide que es mejor el Modelo 7 NS.

4.5. Evaluación de los modelos de propensión de compra

El siguiente paso es comparar el modelo de filtros colaborativos básico y con novelty score, con el modelo del MDO, tanto MDO general como el MDO de oferta individual. Es importante destacar que dado que esta memoria no posee una evaluación experimental de los modelos propuestos, no se puede llegar a resultados 100% certeros, sin embargo, se asume que con la metodología ocupada para simular la conversión de las ofertas, se puede obtener una idea mas clara del efecto que tendría el recomendador para la empresa.

Para realizar las comparaciones se agrupan los resultados según que tan novedosas son las MCT que se recomiendan. Se crean tres segmentos de MCT según su novelty score, la *tabla 7* muestra los detalles de cada segmento:

REGLA	SEGMENTO	NÚMERO DE MCT
$NS < 0,12$	Popular	146
$0,12 \leq NS < 0,25$	Novedoso	46
$NS \geq 0,25$	Muy novedoso	41

Tabla 10: Detalle segmentación MCT.

Fuente: Elaboración propia

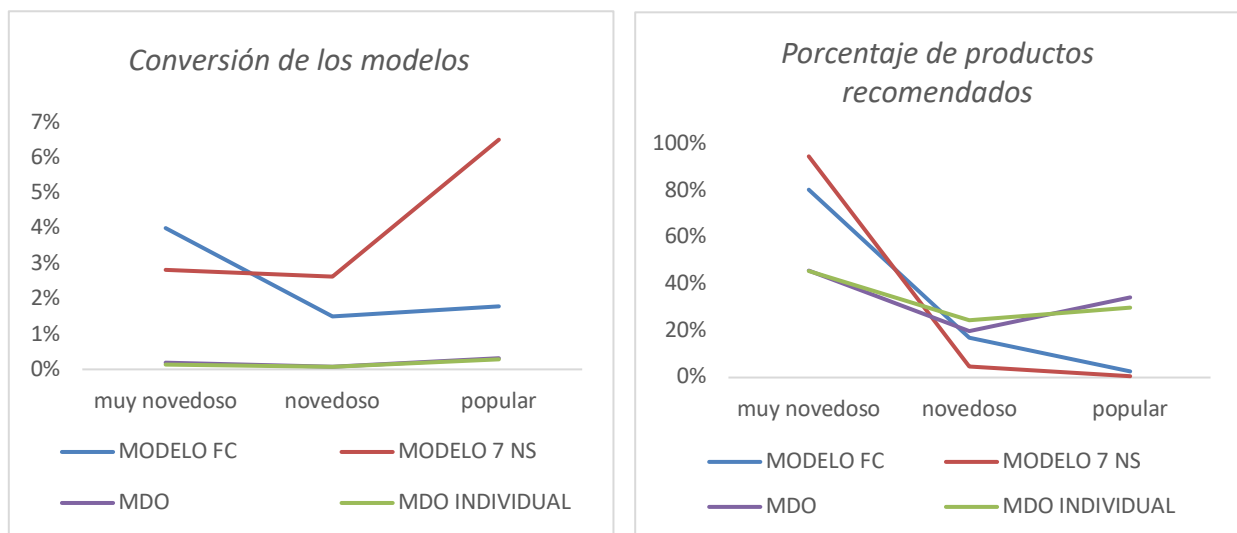


Gráfico 28: Tasa de conversión de los modelos por segmento de MCT.
Gráfico 29: Porcentaje de productos recomendados por segmento de MCT.
Fuente: Elaboración propia

Al comparar los 4 modelos, se puede notar en el *gráfico 28* que el Modelo FC (de filtros colaborativos) logra mejor conversión en la categoría de MCT muy novedoso, sin embargo, en las otras categorías es superado por el Modelo 7 NS (modelo de filtros colaborativos con novelty score). En cuanto al porcentaje de productos que recomienda cada modelo, el modelo con novelty score cumple totalmente lo esperado, que es

recomendar mayormente MCT muy novedosas y muy pocas MCT populares. Si comparamos los resultados con el MDO podemos notar que el MDO siempre es superado en la tasa de conversión, y recomienda porcentajes parecidos de MCT muy novedosos y populares. Se concluye que en este aspecto también es superado por los nuevos modelos.

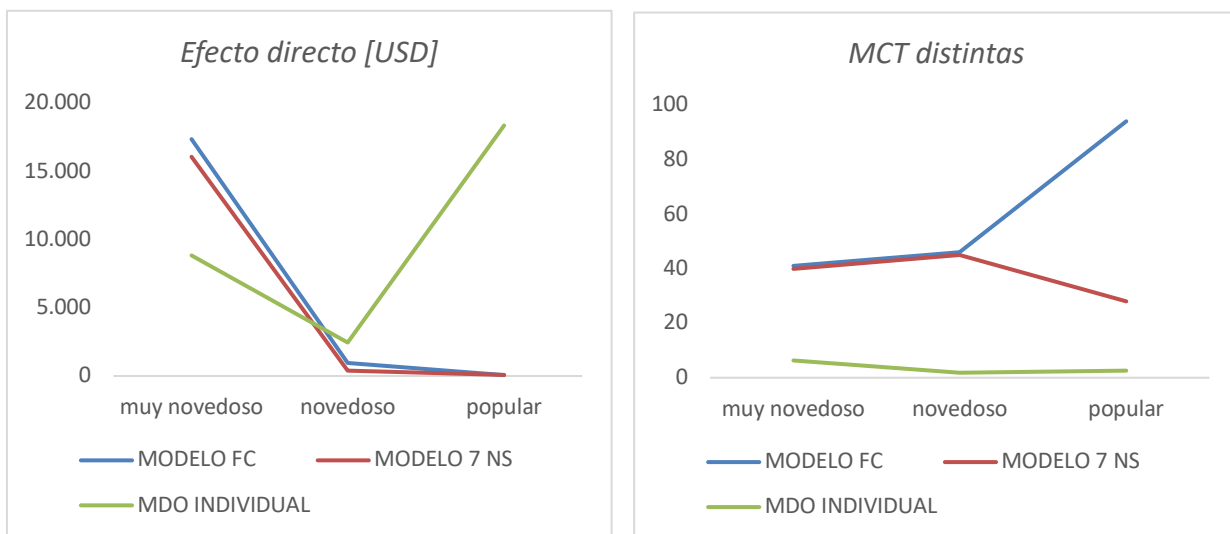


Gráfico 30: Ventas en dólares de los modelos, por segmento de MCT.

Gráfico 31: Número de MCT distintas por segmento de MCT.

Fuente: Elaboración propia

Al observar el monto en ventas de los modelos en el *gráfico 30*, se puede notar que el MDO supera ampliamente los montos en ventas de productos populares, esto se explica a las reglas de negocio presentes en el modelo, donde se ofrece lo que las consultoras más compran. Sin embargo, se logra mejorar los montos de ventas en la categoría muy novedosos, que es el enfoque de los nuevos modelos. El *gráfico 31* muestra que ambos modelos recomiendan mayor variedad de MCT que el MDO individual, y que el modelo FC recomienda mayor variedad de MCT populares que el modelo 7 NS. Esto era de esperar, dado que el modelo con novelty score recomienda un número muy reducido de MCT populares.

Respecto a los resultados de los distintos modelos se puede concluir que los nuevos modelos no son mejores que el MDO en su totalidad, sin embargo vemos que logran superar los resultados obtenidos por el MDO en la categoría de MCT muy novedosas. Esta última categoría representa el objetivo de insertar en las recomendaciones de las consultoras productos que ellas no hayan comprado antes, y así lograr un incremento en el mix de la canasta de compra.

Por otro lado, el modelo de filtros colaborativos y el modelo con novelty score muestran resultados bastante similares, esto debido a que sus metodologías son parecidas y ambos recomiendan solo productos que las consultoras no hayan comprado en los últimos 6 meses. Por esta razón no se selecciona un modelo por sobre el otro, eso

dependerá de las necesidades de la empresa y de lo que ellos privilegian en un modelo de recomendación.

4.6. Modelo de bundle

El modelo de bundle está encargado de crear las recomendaciones de packs de dos o más productos distintos para las consultoras.

4.6.1. Modelo basado en reglas de asociación

Para trabajar con este modelo, se cuenta con los datos de todos los productos comprados durante las campañas 09 y 14 del año 2018. Cada producto lleva la identificación del pedido en el que fue comprado por la consultora. Se cuenta con un total de 979.622 pedidos distintos y 2.261 productos distintos. Los modelos de reglas de asociación se programan usando el software R, específicamente el paquete arules [6], el cual permite usar de manera sencilla el algoritmo Apriori.

En promedio, las consultoras compran 11,7 productos en cada pedido¹⁵ y el 90% de los pedidos no supera los 18 productos¹⁶. El siguiente gráfico muestra como distribuye el tamaño de los pedidos de las consultoras en el periodo seleccionado:

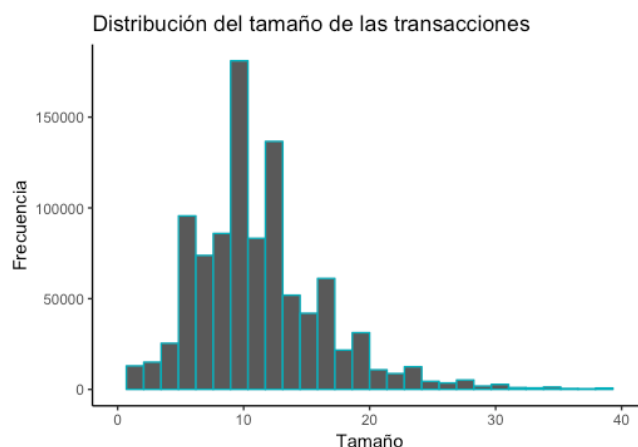


Gráfico 32: Distribución del tamaño de las transacciones.
Fuente: Elaboración propia

Las reglas de asociación muestran los set de productos que más se repiten entre los pedidos de las consultoras, denominando estos productos como antecedentes y conecuentes de la regla.

¹⁵ Detalle de la distribución de los productos por transacción en anexo Q: tabla 8.

¹⁶ Detalle de los porcentajes de consultoras por número de productos por transacción en anexo R: tabla 9.

Para este proyecto de memoria se trabaja con reglas que contienen sólo dos productos, un antecedente y un consecuente. El objetivo es recomendar sets de productos donde el antecedente haya sido comprado antes por la consultora, mientras que el consecuente es un producto que no ha sido comprado en las últimas 6 campañas por la consultora. Con esto se pretende apalancar la compra de un producto nuevo mediante la compra de un producto ya conocido.

Se crean reglas de asociación fijando valores mínimos para los parámetros de soporte y confianza. Estos valores límites son fijados mediante pruebas donde se debe encontrar un equilibrio entre la calidad de las reglas y el número total de reglas encontradas. Es importante contar con un número suficiente de reglas que permita realizar recomendaciones a la mayoría de las consultoras. La siguiente tabla muestra los distintos conjuntos de parámetros, con sus valores límites de soporte y confianza, y el número de reglas que se obtuvo.

REGLAS DE ASOCIACIÓN	SOPORTE	CONFIANZA	Nº DE REGLAS
Regla 1	0,1	0,5	2
Regla 2	0,01	0,5	93
Regla 3	0,01	0,3	300
Regla 4	0,01	0,2	742

Tabla 11: Resultados modelos de filtros colaborativos con novelty score.
Fuente: Elaboración propia

Se selecciona la regla 4, con 742 reglas distintas. Estas tienen un soporte promedio de 0,023, una confianza promedio igual a 0,325 y un lift promedio de 3,765¹⁷.

Como se mencionó anteriormente, para asignar las mejores recomendaciones personalizadas, para cada consultora se realiza un filtro donde cada una puede tener solo reglas donde el antecedente haya sido comprado antes y el consecuente no haya sido comprado en las últimas 6 campañas. Luego se ordenan las reglas por su lift y se selecciona un máximo de 30 reglas por cada consultora.

Para seleccionar el óptimo de reglas a recomendar por cada consultora, se realiza un gráfico con la tasa de conversión para cada sets desde 1 a 30 recomendaciones por consultora. Esta conversión se calcula usando los datos de las transacciones de la campaña 15, suponiendo que las reglas fueron mostradas en dicha campaña y verificando los conjuntos que se repiten en las compras de las consultoras. Se obtiene el siguiente gráfico con los resultados de las conversiones:

¹⁷ Detalle de la distribución del soporte, confianza y lift en anexo T: tabla 11.



Gráfico 33: Porcentaje de productos recomendados por segmento de MCT.
Fuente: Elaboración propia

El gráfico 33 muestra como a medida que se van agregando mas recomendaciones por consultoras la conversión va aumentando también, pero este aumento en la conversión es cada vez más pequeño hasta que llega un punto donde se estanca. Se desea recomendar elementos hasta antes de ese punto donde ya no hay aumento significativo en la conversión, por lo que se decide recomendar un máximo de 12 packs por consultora.

4.7. Evaluación del modelo de bundle

Para poder analizar el modelo, se comparan distintas métricas con los resultados obtenidos por las ofertas tipo bundle del MDO. Se calculan las tasas de conversión de cada modelo, las ventas que se obtienen en dólares, la cantidad de productos distinto que logra mostrar el recomendador en sus pack y cuantas ofertas se genera por cada consultora. Si bien en el recomendador nuevo se fija un límite de 12 recomendaciones, hay consultoras que tienen menos recomendaciones dadas las restricciones que tienen los packs.

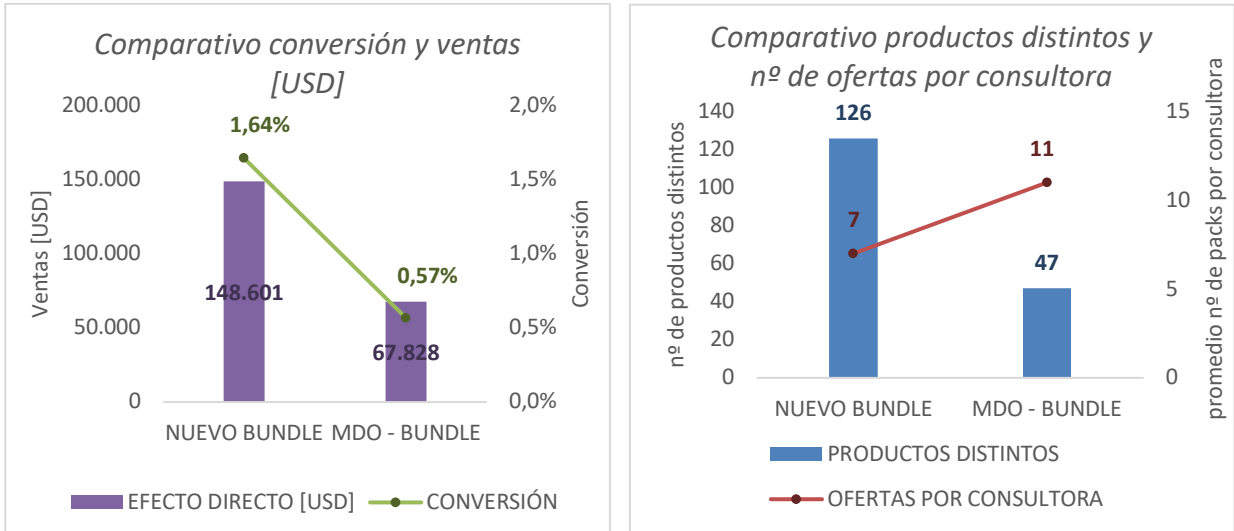


Gráfico 34: Comparativo de conversión y de ventas en dólares de los modelos de bundle.
 Gráfico 35: Comparativo de productos distintos y número de ofertas por consultora, de los modelos de bundle.
 Fuente: Elaboración propia

El gráfico 34 muestra que el nuevo modelo de bundle tiene una mayor tasa de conversión y también logra mayores ventas que el MDO. A diferencia del modelo de propensión de compra, el modelo de bundle recomienda los productos más comprados por las consultoras, dado que selecciona los packs de productos que más se repiten en las transacciones de las consultoras. Es por esto que es importante incluir la novedad de otra forma, y eso se logra al recomendar sólo consecuentes que no hayan sido comprados por las consultoras anteriormente. De esta forma se sigue la línea de incluir productos novedosos en las recomendaciones.

En el gráfico 35 se puede observar que el nuevo modelo recomienda en promedio menos pack que el modelo bundle, esto no es algo negativo dado que se logra mayor conversión y mayores ventas de todas maneras. Por otro lado, se recomienda un mix de productos mayor alcanzando un total de 126 productos distintos, esto es más del doble de los productos distintos generados por el bundle del MDO.

Se concluye que el nuevo modelo de bundle logra mejores resultados que el modelo de bundle del MDO.

5. CONCLUSIONES

5.1. Conclusiones del proyecto

El presente proyecto de memoria se llevó a cabo de forma satisfactoria, pudiendo cumplir con cada uno de los objetivos planteados en un inicio. Su principal objetivo fue evaluar y rediseñar un motor de promociones personalizadas a nivel de consultora-campaña, para una empresa peruana dedicada a la venta por catálogo, esto con el fin de disminuir el sesgo de popularidad de las recomendaciones, manteniendo o mejorando las tasas de conversión.

Cabe destacar que Penta Analytics no contaba con una evaluación previa del motor de recomendaciones, por lo que no había claridad de que se contara con toda la información necesaria en la base de datos para realizar dicho diagnóstico. Por otro lado, las personas que sabían acerca de las características del negocio y manejaban la información de las bases de datos y de los modelos, trabajaban en las oficinas de Perú de la empresa, por lo que no siempre se podían resolver las dudas de forma rápida y con total claridad. Estos fueron uno de los mayores obstáculos al momento de realizar la memoria, dado que fue necesario invertir mucho más tiempo del pronosticado para las etapas de comprensión del negocio y entendimiento de los datos.

Del levantamiento de las etapas del proceso del MDO, se concluye que el modelo desde un inicio presenta limitaciones en los productos que puede recomendar, dado que solo cuenta con 27 de las 233 MCT disponibles en los catálogos de las marcas. Estas reglas de negocio son determinadas por el área comercial de la empresa, quienes deciden en una última instancia que recomendaciones mostrar a las consultoras. También se destaca, que en ninguna etapa de los algoritmos, existen reglas o filtros que ayuden a recomendar MCT novedosas a las consultoras, por el contrario, se recomiendan las MCT que ellas más compran y las que frecuentan sus consultoras similares. Esto explicaría el porque la categoría fragancias es por lejos la más recomendada por el modelo.

Al realizar el diagnóstico de las recomendaciones del MDO, lo más complicado fue poder identificar en las bases de datos los productos pertenecientes a recomendaciones hechas por el MDO. En otras memorias relacionadas con ofertas, se identifican los productos con descuento solo viendo variaciones en el precio normal de estos. Sin embargo, en el negocio de las ventas de catálogos existen múltiples tipos de descuentos en todos los productos, por lo que no bastaba con fijarse en los precios. Fue necesario realizar múltiples cruces de tablas de datos para llegar a un resultado confiable.

De los resultados de la evaluación destaca la baja tasa de conversión encontrada, esto puede ser explicado por el gran número de ofertas que se muestra a cada consultora, siendo posible que muchas de las recomendaciones nunca sean vistas por la consultora. Esto sería muy interesante de evaluar en trabajos futuros, identificando

previamente el orden en que son mostradas las recomendaciones¹⁸, y verificar si existe relación entre la posición de la recomendación y su conversión. Con esto se podría identificar el número óptimo de recomendaciones a mostrar por cada consultora. Actualmente este número lo decide el área comercial de la empresa.

Como se mencionó en el informe, desde la campaña 08 del año 2018 comenzó a funcionar una nueva versión del MDO para Colombia. Al analizar los resultados del diagnóstico se observa una clara tendencia de mejora en los indicadores, alcanzando alzas de hasta un 100% y más. Estas mejoras no hubieran sido percatadas en su totalidad sin el diagnóstico del recomendador, por lo que se recalca la importancia de tener un sistema de evaluación dentro de los modelos de negocio de una empresa.

Por otro lado, se observa que las recomendaciones que tienen mayor tasa de conversión, son las ofertas de tipo bundle y las individuales, siendo también las que generan mayores ventas [USD]. Esto evidencia que dichos tipos de ofertas, son las que generan mayor interés a las consultoras, siendo importante enfocar en ellas las propuestas de mejoras a sus modelos en esta memoria.

En el modelo de filtros colaborativos con novelty score, se obtienen dos conclusiones que podrían no ser obvias en un inicio. En primer lugar, se obtienen mejores resultados en las tasas de conversión, cuando no se segmenta a las consultoras en frecuentes y ocasionales como sugieren las lecturas. Se obtienen mejores resultados aplicando a todas las consultoras por igual el modelo con novelty score. Esto demuestra que el modelo obtiene buenos resultados a pesar de la cantidad de información previa disponible que tenga la consultora. En segundo lugar, al momento de identificar el límite para segmentar entre productos novedosos y populares, se obtienen mejores resultados en temas de conversión cuando se es más exigente con el valor límite de novelty score.

En los modelos de propensión de compra propuestos, se observa una desventaja con respecto al MDO relacionada con los resultados del indicador de efecto directo. Esto dado que las recomendaciones del MDO están hechas en base a las categorías que representan un 80% de las ventas de la empresa, por lo que se logran mayores montos de venta con el canje de sus promociones. Esta dificultad es compensada por otras ventajas que posee el ampliar el set de productos a recomendar, como por ejemplo aumentar el mix de la canasta de compra de los clientes o disminuir las colas de productos estancados en bodega. Por supuesto, al momento de decidir que recomendador es mejor para una empresa, se debe tener en consideración las estrategias comerciales de esta.

Con el modelo de propensión de compra filtros colaborativos y filtros colaborativos con novelty score, se obtienen las mejoras en temas de sesgo de popularidad y además, se logran mejorar las tasas de conversión del MDO. Se concluye que resulta atractivo recomendar productos que las consultoras no hayan comprado antes, que estas recomendaciones tendrían aceptación según los resultados de las simulaciones hechas en el trabajo de memoria.

¹⁸ Actualmente no se conoce el orden en que son mostradas las recomendaciones a las consultoras.

Con el modelo de bundle propuesto en esta memoria, se logra mejorar la tasa de conversión y además, aumentar el efecto directo que tenía el MDO en este tipo de ofertas. Adicionalmente se cumple con el objetivo central, dado que también se observan beneficios en temas de sesgo de popularidad. El nuevo modelo de bundle basado en reglas de asociación y en reglas personalizadas, logra que la mitad de los productos recomendados a las consultoras sean nuevos para ellas, es decir, que no los hayan comprado en las últimas seis campañas.

En términos generales, se obtiene de forma satisfactoria una evaluación del MDO, logrando identificar sus etapas y señalando posibles mejoras al modelo. También se logran obtener cifras respecto a los resultados que ha logrado el recomendador para la empresa peruana, pudiendo así cuantificar su desempeño y sus ganancias. Finalmente, se pudo cumplir con el objetivo de disminuir el sesgo de popularidad, presente en las recomendaciones del motor de ofertas de la empresa, y además mejorar la tasa de conversión de estas.

5.2. Trabajos futuros

Para los futuros trabajos que estén en línea con la presente memoria, se proponen las siguientes recomendaciones con el objetivo de mejorar los resultados en el proyecto:

En relación al nivel al cual se realizan las recomendaciones, queda pendiente realizar un estudio detallado de los subconjuntos de productos que podrían ser utilizados. Las MCT no son subconjuntos homogéneos en cuanto a cantidad de productos, por lo es necesario crear un nuevo subconjunto en base a otros criterios.

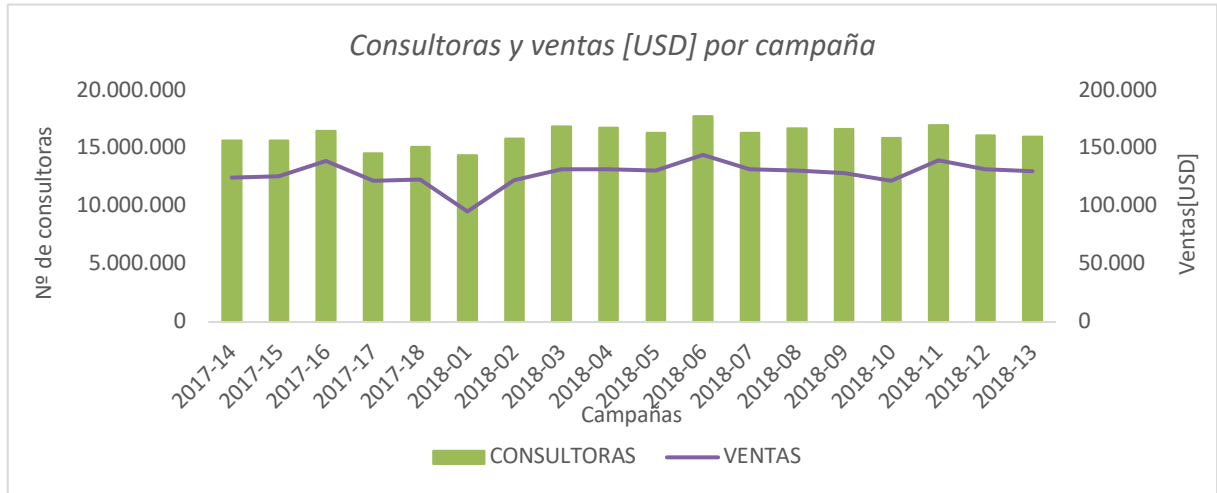
Para obtener resultados más precisos respecto a la conversión de los modelos de propensión de compra y de ofertas bundle, es necesario realizar un diseño experimental a través de una implementación piloto de las nuevas ofertas a un grupo de testeo de consultoras. Con esto se podrá medir la real efectividad de los nuevos modelos para la empresa.

6. BIBLIOGRAFÍA

- [1] Penta Analytics, Inicio. <http://www.analytics.cl/#home> [Consulta: Junio 2018]
- [2] Penta Analytics, Nosotros. <http://www.analytics.cl/nosotros/> [Consulta: Junio 2018]
- [3] World Federation of Direct Selling Associations (WFDSA), Global Statistics. <https://wfdsa.org> [Consulta: Julio 2018]
- [4] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996). FROM DATA MINING TO KNOWLEDGE DISCOVERY IN DATABASES.
- [5] Dietmar Jannach, Lukas Lerche, Iman Kamehkhosh, Michael Jugovac (2015). WHAT RECOMMENDERS RECOMMEND: AN ANALYSIS OF RECOMMENDATION BIASES AND POSSIBLE COUNTERMEASURES.
- [6] Reglas de asociación y algoritmo Apriori con R. https://rpubs.com/Joaquin_AR/397172 [Consulta: Febrero 2019]
- [7] Nicolás Ignacio Torres Rudloff (2015). SISTEMAS DE RECOMENDACIÓN BASADOS EN MÉTODOS DE FILTRADO COLABORATIVO (tesis de pregrado). Universidad técnica Federico Santa María, Santiago, Chile.
- [8] Nalda Reyes, José Antonio (2010). SISTEMA SECUENCIAL DE RECOMENDACIONES PERSONALIZADAS EN UNA EMPRESA DE HOME IMPROVEMENT (tesis de pregrado). Universidad de Chile, Santiago, Chile.
- [9] Badrul Sarwar, George Karypis, Joseph Konstan , John Riedl (2001). ITEM-BASED COLLABORATIVE FILTERING RECOMMENDATION ALGORITHMS.
- [10] Zan Huang, Daniel Zeng, Hsinchun Chen (2007). A COMPARISON IF COLLABORATIVE-FILTERING RECOMMENDATION ALGORITHMS FOR E-COMMERCE.
- [11] Data iku (2017). HOW TO: DRIVE SERENDIPITOUS DISCOVERY WITH RECOMMENDATION ENGINES (guidebook).
- [12] Punam Bedi, Anjali Gautam, Richa, Chhavi Sharma (2014). USING NOVELTY SCORE OF UNSEEN ITEMS TO HANDLE POPULARITY BIAS IN RECOMMENDER SYSTEMS.

7. ANEXOS

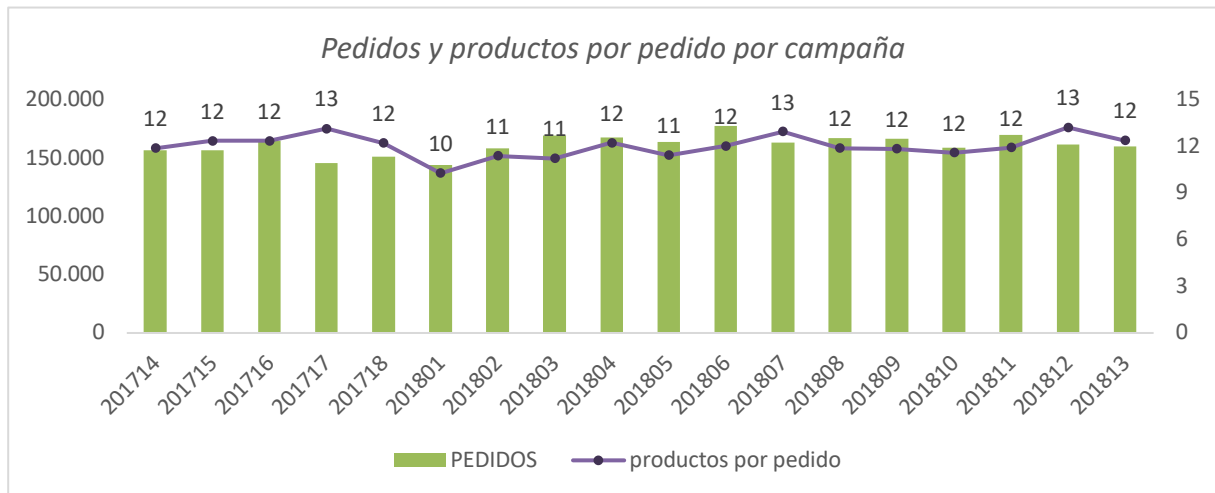
Anexo A: Gráfico 1



Número de consultoras que realizan un pedido, y monto en dólares de las ventas de los pedidos por campaña. Entre la campaña 14 del 2017 y 13 del 2018.

Fuente: Elaboración propia

Anexo B: Gráfico 2



Número total de pedidos realizados en una campaña y promedio de productos vendidos por pedido entre la campaña 14 del 2017 y 13 del 2018.

Fuente: Elaboración propia

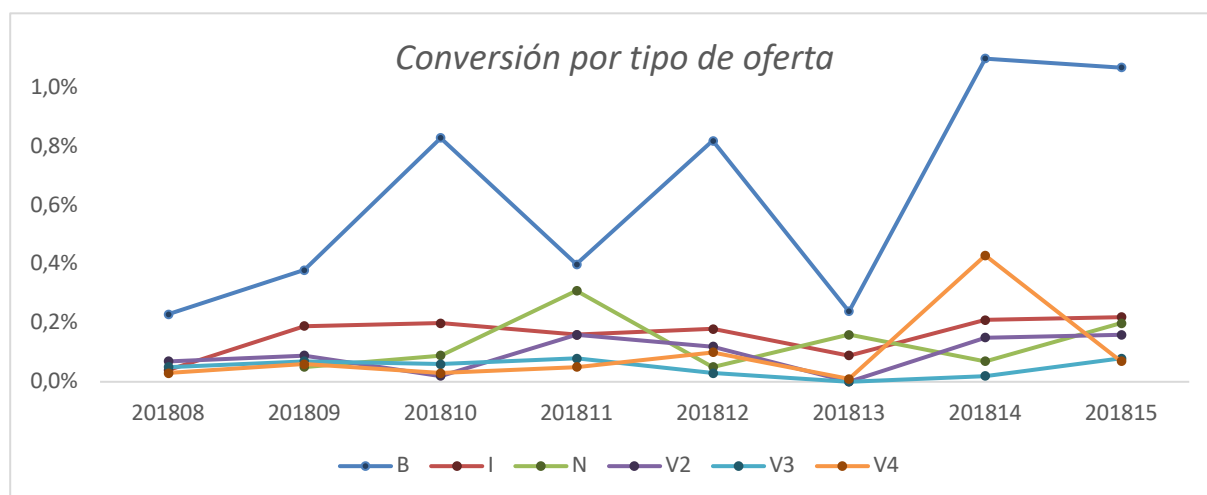
Anexo C: Gráfico 3



Valor promedio en dólares de los pedidos realizados en una campaña y valor promedio en dólares de un productos, agrupados por camapaña. Entre la campaña 14 del 2017 y 13 del 2018.

Fuente: Elaboración propia

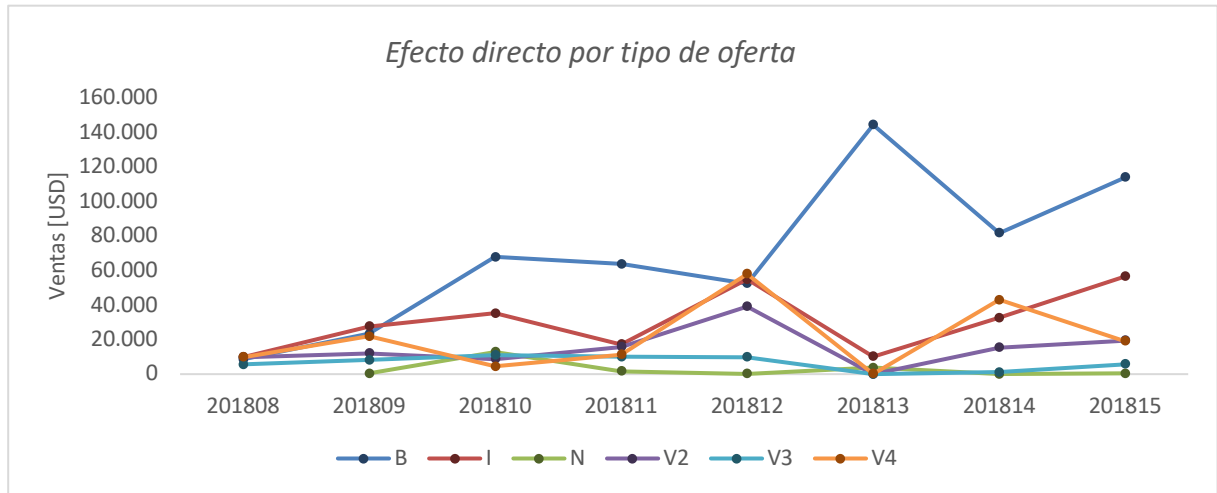
Anexo D: Gráfico 4



Resultados de la tasa de conversión por tipo de oferta, agrupados por campaña.

Fuente: Elaboración propia

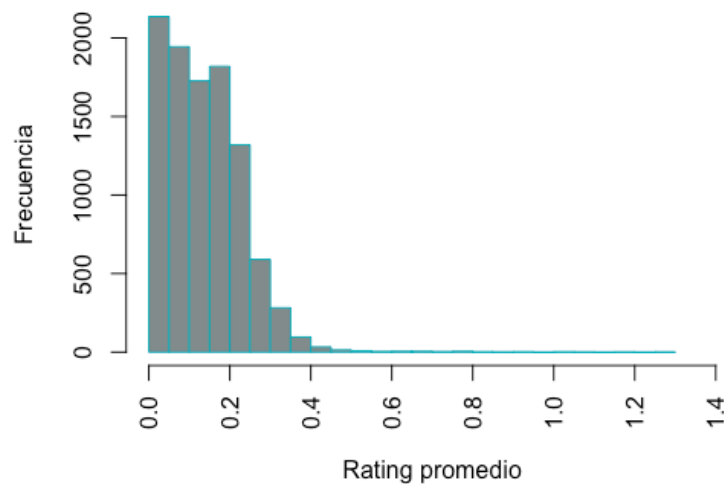
Anexo E: Gráfico 5



Resultados del indicador Efecto directo en dólares por tipo de oferta, agrupado por campaña.
Fuente: Elaboración propia

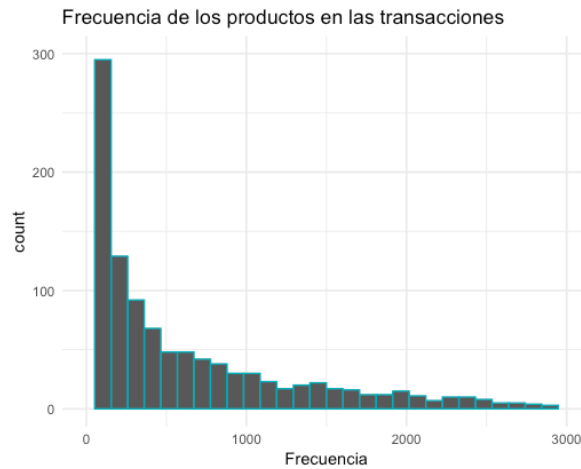
Anexo F: Gráfico 6

Distribución de los rating promedios por consultora



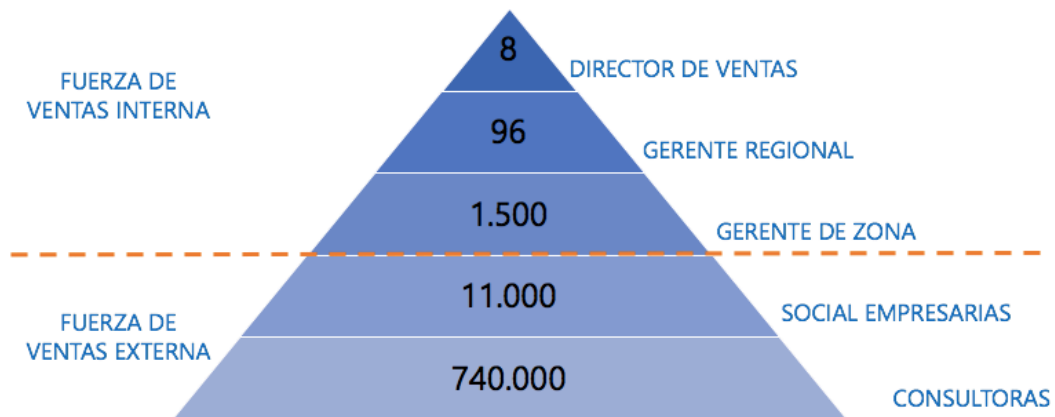
Distribución de rating promedio por consultora.
Fuente: Elaboración propia

Anexo G: Gráfico 7



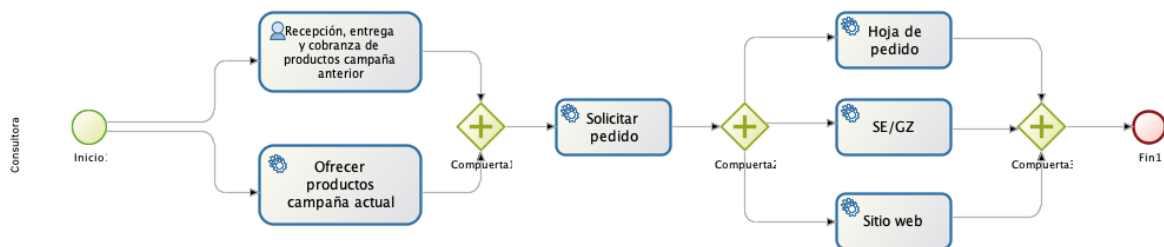
Frecuencia de los productos en las transacciones.
Fuente: Elaboración propia

Anexo H: Imagen 1



Estructura fuerza de ventas Belcorp a Julio 2016.
Fuente: Elaboración propia

Anexo I: Imagen 2



Proceso de una consultora para realizar un pedido en una campaña
Fuente: Elaboración propia

Anexo J: Tabla 1

REGISTRO	NOMBRE COLUMNA	DESCRIPCIÓN
Código del país	cod_pais	Código del país de la consultora, ej. CO
Id consultora	pk_consultora	Identificador único de la consultora (DECIMAL), ej 3074946
Código consultora	cod_ebelista	Identificador único de la consultora (VARCHAR), ej 9906849248
Campaña de ingreso	anio_capana_ingreso	Año en que ingreso la consultora a la empresa, indicando el año y luego la campaña, ej: 201808: año 2018, campaña 08
Fecha de nacimiento	fecha_nacimiento	Fecha de nacimiento de la consultora, formato aaaa-mm-dd, ej. 1990-04-10
Estado civil	des_estado_civil	Estado civil de la consultora, ej. Soltero(a)
Camapaña del último pedido	anio_campana_ultimo_pedido	Campaña en que la consultora realiza su primer pedido, ej. 201809
Campaña de primer pedido	anio_campana_primer_pedido	Campaña en que la consultora realiza su primer pedido a través del sitio web, ej. 201810

*Detalle de las principales variables de la base de datos maestro de las consultoras
Fuente: Elaboración propia*

Anexo K: Tabla 2

REGISTRO	NOMBRE COLUMNA	DESCRIPCIÓN
Código del país	cod_pais	Código del país de la consultora, ej. CO
Id producto	pk_producto	Identificador único del producto (DECIMAL), ej. 167024
Código de la marca	cod_marca	Código de la marca del producto, ej. B
Nombre de la categoría	des_categoria	Nombre de la categoría del producto, ej. MAQUILLAJE
Nombre del tipo	des_tipo_solo	Nombre del tipo del producto, ej. LAPIZ LABIAL
Nombre de la marca	des_marca	Nombre de la marca del producto
Código de la categoría	cod_categoria	Código de la categoría del producto, ej. 104530
Nombre del tipo	des_tipo	Nombre de la categoría y tipo del producto, ej. MAQUILLAJE LAPIZ LABIAL
Código del producto	cod_cuc	Identificador único del producto (VARCHAR), ej. 603979670

*Detalle de las principales variables de la base de datos maestro de los productos
Fuente: Elaboración propia*

Anexo L: Tabla 3

REGISTRO	NOMBRE COLUMNA	DESCRIPCIÓN
Código del país	cod_pais	Código del país de la consultora, ej. CO
Campaña	anio_campana	Campaña en que se realiza la transacción, ej. 201809
Id consultora	pk_consultora	Identificador único de la consultora (DECIMAL), ej. 3074946
Id pedido	pk_pedido	Identificador único del pedido(DECIMAL), ej. 39997526
Id producto	pk_producto	Identificador único del producto (DECIMAL), ej. 167024
Unidades vendidas	real_uu_vendidas	Unidades que la consultora compró del producto (DECIMAL), ej. 3
Monto del pedido	real_vta_mn_factura	Valor del producto que se compró en la moneda del país (DECIMAL), ej. 15.600
Descuento	descuento	Descuento con que fue comprado el producto (DECIMAL), ej. 20

*Detalle de las principales variables de la base de transacciones
Fuente: Elaboración propia*

Anexo M: Tabla 4

REGISTRO	NOMBRE COLUMNA	DESCRIPCIÓN
Código del país	cod_pais	Código del país de la consultora, ej. CO
Campaña	anio_campana	Campaña en que se realiza la transacción, ej. 201809
Id producto	pk_producto	Identificador único del producto (DECIMAL), ej. 167024
Tipo de oferta	des_tipo_oferta	Tipo de oferta del producto (VARCHAR), ej. OFERTA SUPER ESPECIAL
Nombre catálogo	des_tipo_catalogo	Nombre del catálogo donde aparece el producto (VARCHAR), ej. CATALOGO MARCA A
Precio de oferta	precio_oferta	Valor del producto en oferta (DECIMAL), ej. 18.900
Precio normal	precio_normal_mn	Valor del producto sin oferta (DECIMAL), ej. 43.400
Código catálogo	cod_catalogo	Código del catálogo donde aparece el producto (DECIMAL), ej. 05

*Detalle de las principales variables de la base de los productos disponibles por campaña
Fuente: Elaboración propia*

Anexo N: Tabla 5

CATEGORÍA	PROMEDIO DE PRODUCTOS DISTINTOS	DESVIACIÓN ESTANDAR	VALOR MÍNIMO	VALOR MÁXIMO
ACCESORIOS COSMETICOS	12	3	6	18
BIJOUTERIE	245	37	177	331
CABALLEROS	6	2	2	10
COMPLEMENTOS	71	15	44	95
CUIDADO PERSONAL	98	5	87	107
DAMAS	8	3	3	14
FRAGANCIAS	134	7	123	155
HOGAR	47	7	37	59
JOVENES	6	2	2	11
LENTES	10	3	5	18
MAQUILLAJE	106	6	95	117
MUESTRAS COSMETICOS	117	17	82	151
PRENDA DE VESTIR (ROPA)	32	14	13	63
PROMOCION USUARIOS	19	4	13	26
TRATAMIENTO CORPORAL	40	3	34	44
TRATAMIENTO FACIAL	57	3	52	61
VARIOS	86	11	69	114

*Información de la cantidad de productos distintos por categorías.
Fuente: Elaboración propia*

Anexo O: Tabla 6

Nº de MCT	Productos más comprado	Producto mayor venta	MCT	MDO 2018-15 oferta individual
1	8,25%	8,42%	21,4%	0,16%
2	6,74%	6,75%	17,9%	0,16%
3	6,12%	6,13%	16,1%	0,16%
4	5,48%	5,49%	14,7%	0,16%
5	4,94%	4,99%	13,6%	0,16%
6	4,46%	4,51%	12,7%	0,16%
7	4,14%	4,17%	11,9%	0,16%
8	3,84%	3,90%	11,3%	0,16%
9	3,62%	3,66%	10,8%	0,16%
10	3,41%	3,46%	10,3%	0,16%
11	3,23%	3,30%	9,9%	0,16%
12	3,09%	3,14%	9,4%	0,16%
13	2,95%	3,01%	9,0%	0,16%
14	2,83%	2,89%	8,7%	0,16%
15	2,71%	2,77%	8,4%	0,16%
16	2,62%	2,68%	8,1%	0,16%
17	2,54%	2,60%	7,8%	0,16%
18	2,46%	2,52%	7,5%	0,16%
19	2,38%	2,45%	7,3%	0,16%
20	2,33%	2,38%	7,1%	0,16%

Detalle de la conversión de los distintos modelos del modelo de propensión de compra con filtros colaborativos

Fuente: Elaboración propia

Anexo P: Tabla 7

CAMPAÑA	NOVEDAD	NOVEDAD MDO
201814	43%	54%
201815	40%	45%

Indicador de novedad por campaña.

Fuente: Elaboración propia

Anexo Q: Tabla 8

Mín.	1st Qu.	Mediana	Promedio	3rd Qu.	Máx.
1	8	10	11,68	14	298

Distribución número de productos por cada transacción.

Fuente: Elaboración propia

Anexo R: Tabla 9

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
1	6	7	8	9	10	12	13	15	18	298

Distribución en porcentaje del número de productos por cada transacción.

Fuente: Elaboración propia

Anexo S: Tabla 10

Mín.	1st Qu.	Mediana	Promedio	3rd Qu.	Máx.
1	19	196	5.061	1.594	259.699

Frecuencia de los productos en las transacciones.

Fuente: Elaboración propia

Anexo T: Tabla 11

SOPORTE		CONFIANZA		LIFT	
Mínimo	0,010	Mínimo	0,200	Mínimo	0,884
1st Qu	0,013	1st Qu	0,233	1st Qu	1,224
Mediana	0,017	Mediana	0,273	Mediana	2,213
Promedio	0,023	Promedio	0,325	Promedio	3,765
3st Qu	0,027	3st Qu	0,375	3st Qu	3,205
Máximo	0,270	Máximo	1,000	Máximo	50,111

Distribución del soporte, confianza y lift.

Fuente: Elaboración propia

Anexo U: Tabla 12

CATEGORIA MCT	PORCENTAJE DE PRODUCTOS	CONVERSION	EFEECTO DIRECTO[USD]
muy novedoso	80,36%	4,00%	17.364
novedoso	17,01%	1,50%	971
popular	2,63%	1,78%	66
PROMEDIO	33,33%	2,43%	6.134

Resultados modelo de propensión de compra de filtros colaborativos

Fuente: Elaboración propia

Anexo V: Tabla 13

CATEGORIA MCT	PORCENTAJE DE PRODUCTOS	CONVERSION	VENTA
muy novedoso	0,14%	0,14%	8.839
novedoso	0,07%	0,07%	2.453
popular	0,28%	0,28%	18.368
PROMEDIO	0,17%	0,16%	9.887

Resultados modelo 7 de propensión de compra de filtros colaborativos con novelty score

Fuente: Elaboración propia

Anexo W: Tabla 14

CATEGORIA MCT	PORCENTAJE DE PRODUCTOS	CONVERSION	VENTA
muy novedoso	0,18%	0,18%	335.924.065
novedoso	0,07%	0,07%	57.396.541
popular	0,32%	0,32%	1.117.453.085
PROMEDIO	0,19%	0,19%	503.591.230

Resultados modelo de propensión de compra del MDO

Fuente: Elaboración propia

Anexo X: Tabla 15

CATEGORIA MCT	PORCENTAJE DE PRODUCTOS	CONVERSION	VENTA
muy novedoso	0,14%	0,14%	8.839
novedoso	0,07%	0,07%	2.453
popular	0,28%	0,28%	18.368
PROMEDIO	0,17%	0,16%	9.887

Resultados modelo de propensión de compra del MDO oferta tipo individual

Fuente: Elaboración propia

Anexo Y: Tabla 16

	TIPO MCT	MCT DISTINTAS	OFERTAS COMPRADAS	OFERTAS MOSTRADAS	VENTA [USD]	CONVERSION
MODELO 1 FC	muy novedoso	41	2.994	73.869	16.884	4,05%
	novedoso	46	267	18.494	1.105	1,44%
	popular	94	48	3.677	81	1,31%
MODELO 1 NS	muy novedoso	40	1.629	68.853	7.885	2,37%
	novedoso	44	56	4.525	164	1,24%
	popular	34	6	183	18	3,28%
MODELO 2 NS	muy novedoso	40	2.054	84.692	10.059	2,43%
	novedoso	45	60	4.499	166	1,33%
	popular	36	9	250	20	3,60%
MODELO 3 NS	muy novedoso	40	2.421	96.302	11.950	2,51%
	novedoso	45	67	4.423	186	1,51%
	popular	37	9	296	14	3,04%
MODELO 4 NS	muy novedoso	40	2.421	96.348	11.971	2,51%
	novedoso	45	70	4.418	201	1,58%
	popular	24	9	255	14	3,53%
MODELO 5 NS	muy novedoso	40	2.190	82.468	10.859	2,66%
	novedoso	45	177	15.883	596	1,11%
	popular	23	12	2.670	26	0,45%
MODELO 6 NS	muy novedoso	40	2.360	90.504	11.578	2,61%
	novedoso	45	125	9.455	380	1,32%
	popular	23	12	1.062	27	1,13%
MODELO 7 NS	muy novedoso	40	3.175	112.960	16.069	2,81%
	novedoso	45	151	5.745	404	2,63%
	popular	28	44	676	79	6,51%

Resultados de todos los modelo de propensión de compra de filtros colaborativos con novelty score

Fuente: Elaboración propia