




E-SHOPPING MARKETING ANALYSIS

Paula Pareja Ramírez
Universidad Comillas
Abril - 2023





ÍNDICE

1. Presentación de los datos
2. Análisis exploratorio de datos
 - a. País
 - b. Temporalidad
 - c. Cliente
 - d. Precio
3. Modelos supervisados y no supervisados
 - a. Reglas de asociación - venta cruzada
 - b. Regresiones - predicción del número de ventas
 - c. Clusterización de productos
4. Conclusiones

1 Presentación de los datos

year	month	day	clics	country	session_ID	category	clothing_code	colour	location	model_photography	price	overpriced	page
2008	4	1	1	29	1	1	A13	1	5	1	28	2	1
2008	4	1	2	29	1	1	A16	1	6	1	33	2	1
2008	4	1	3	29	1	2	B4	10	2	1	52	1	1
2008	4	1	4	29	1	2	B17	6	6	2	38	2	1
2008	4	1	5	29	1	2	B8	4	3	2	52	1	1
2008	4	1	6	29	1	3	C56	6	1	2	57	1	4
2008	4	1	7	29	1	3	C57	5	1	2	33	2	4
2008	4	1	8	29	1	4	P67	9	5	1	38	1	4
2008	4	1	9	29	1	4	P82	6	4	2	48	1	5
2008	4	1	1	29	2	2	B31	9	5	1	57	1	2

La marca proporciona una tabla como la siguiente, en la que cada fila representa un producto vendido. Cada venta está asociada a un ID de sesión y un código de producto. Asimismo, el dataset incluye variables asociadas a los productos como localización en la web, tipo de fotografía, categoría, color o precio. También se recogen la secuencia de clics de la compra y el país donde se realizó la misma.

Sin embargo, hay información de suma importancia que no ha sido recopilada anteriormente por la empresa y deberán ser integradas en el futuro para mejorar la capacidad de análisis y optimizar las ventas.

Variables
a incluir

- ID cliente
- Usuarios de la web - no compradores
- Hora de compra
- Tiempo del usuario en la web
- Edad
- Sexo
- CP / Ciudad / Municipio

- **ID de cliente:** El ID de cliente es una variable clave porque permite identificar a los clientes individuales y realizar un seguimiento de sus compras y preferencias a lo largo del tiempo. Esto nos permite identificar patrones de compra del cliente, analizar frecuencia de compra, ofrecer experiencias de compra personalizada y realizar clusterizaciones de tipos de clientes para el lanzamiento de campañas y delimitación de público objetivo.
- **Usuarios de la web que no han comprado:** esta variable es importante pues permite a la empresa entender la tasa de conversión de la web, el tráfico e interés de los productos y páginas de la web, así como problemas de experiencia de usuario que estén afectando a las ventas.
- **Hora de compra:** muy importante para lanzar campañas de marketing.
- **Tiempo de usuario en la web:** para identificar barreras en la experiencia de usuario que puedan afectar a las ventas así como identificar mejoras del engagement de la web. Podríamos medir por ejemplo si hay más o menos tráfico en diferentes páginas de la web, y si se realizan compras dentro de las mismas. Esto nos permitiría saber si las ventas son problema de la estructura de la web o de los propios productos.
- **Edad cliente y sexo:** variables esenciales para identificar al público objetivo, ajustar la estrategia de marketing y los productos.
- **Ciudad o municipio** son importantes para entender la distribución geográfica de las ventas y ajustar estrategias de marketing.



Google analytics puede ser una herramienta muy útil para la recopilación de los datos mencionados con anterioridad.

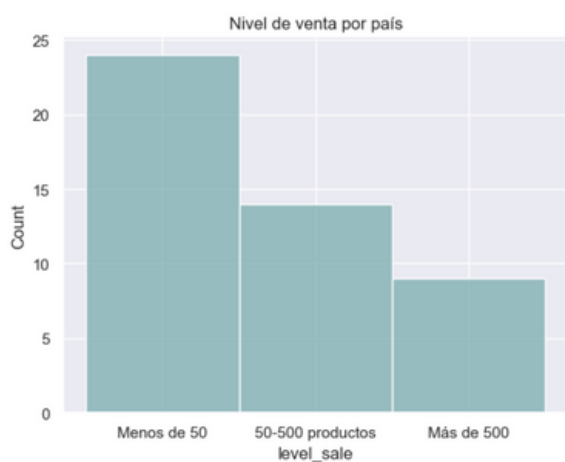
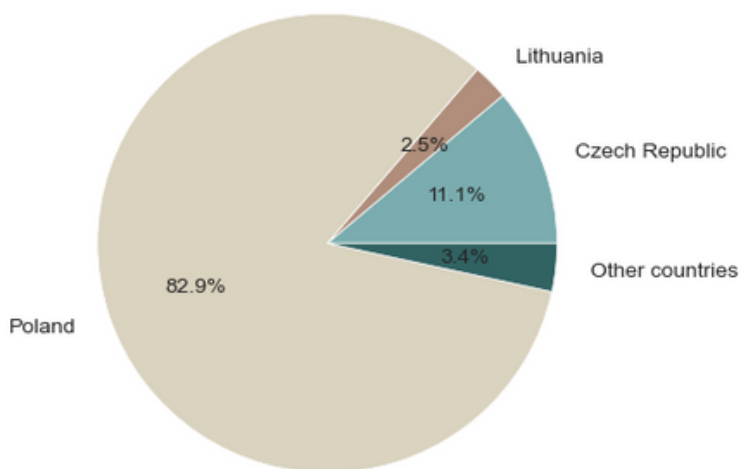
Es totalmente gratuita y ya es utilizada por la mayoría de los e-commerce.

Esta es capaz de recopilar información sobre todos los usuarios de la web y plantear las metas o KPIs que el cliente necesite. Sería interesante establecer una meta de conversión con Google Tag Manager que se active cuando el usuario completa la compra y analizar el embudo de conversión para identificar donde abandonan los usuarios cuando no compran. De esta forma podrían identificarse barreras en la experiencia de usuario, como por ejemplo dificultades a la hora de acceder a la cesta de la compra por falta de visibilidad, o problemas con la pasarela de pago.

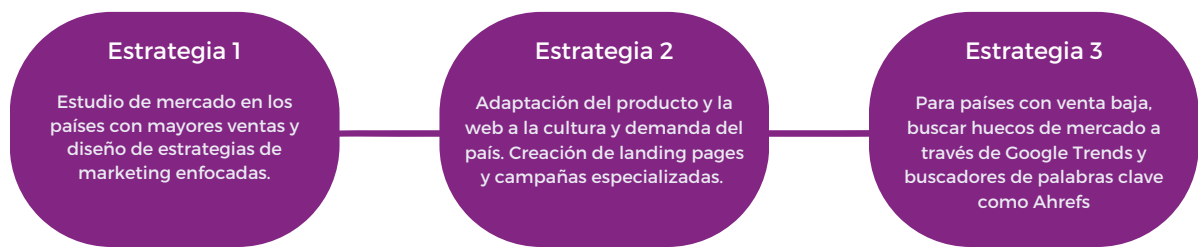
Asimismo podrían obtenerse grabaciones de pantalla con herramientas como Hotjar para analizar el comportamiento del cliente en la web y valorar la experiencia UX/UI.

2 Análisis Exploratorio de Datos

2.1. PAÍS



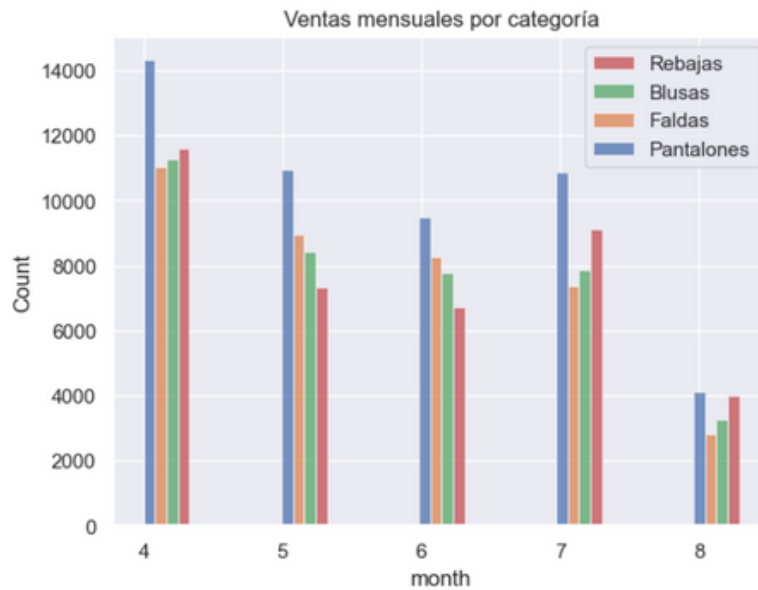
Tras el análisis de ventas por país, se ha identificado que la mayor parte de las ventas se concentran en Polonia, seguido de la República Checa y Lituania. Agrupé los países por nivel de ventas y descubrí que en la mayoría de ellos (24) se habían vendido menos de 50 productos. En 14 se habían vendido entre 50 y 500 productos y en 9 se habían vendido más de 500 productos.



Ante esta situación, se plantean tres estrategias a la marca:

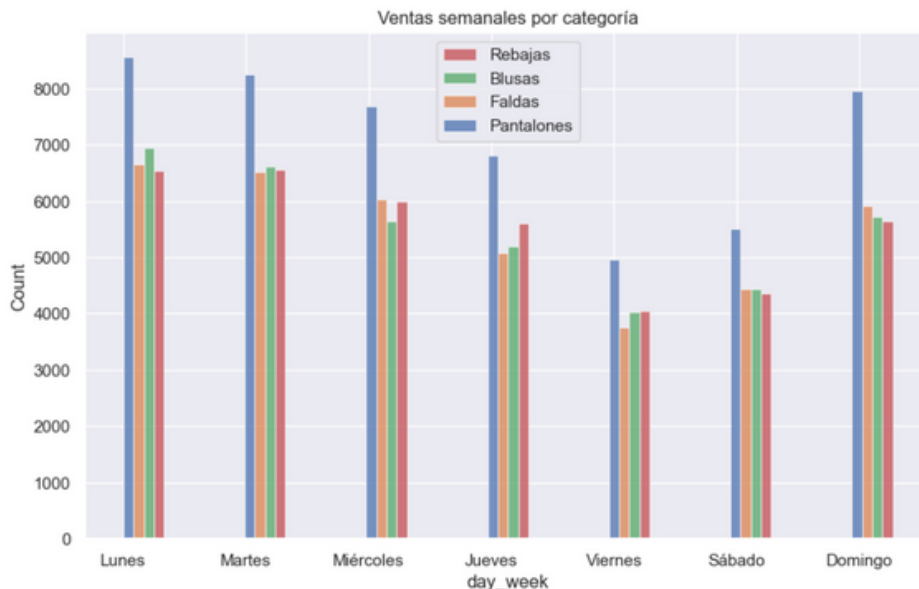
- 1 Realizar un estudio de mercado en los países con una venta alta de productos para analizar la posibilidad de aumentar las ventas en estos países. Seguir centrando esfuerzos de marketing según los resultados del estudio y aumentando la cuota de mercado en esos países. Es importante entender por qué hay mayores ventas en estos países.
- 2 Adaptar los productos y la web a la cultura y demanda de los países con venta media y alta. Esta estrategia consiste en realizar estudios de mercado en los países donde ya hay cierta presencia de la marca para identificar precios en el mercado, competidores directos y necesidades de los clientes. El producto, packaging, colores, comunicación y tipos de fotografía deben adaptarse a la cultura del país. Deben entenderse las necesidades del cliente así como a lo que el mismo da importancia para maximizar las ventas: puede que en algunos países den mayor importancia a la calidad mientras en otros prefieran precios más bajos.
 - a. Asimismo, se recomienda analizar los patrones de venta y crear landing pages, enfocadas a maximizar la tasa de conversión y persuadir al cliente a realizar la compra de los productos más vendidos en dichos países, teniendo en cuenta el público objetivo y la cultura en toda la estructuración, estilo de la web y forma de vender los productos.
- 3 La tercera estrategia se enfoca en identificar huecos de mercado en los países con pocas ventas, a través de herramientas como Google Trends y buscadores de palabras clave como Ahrefs. Esta estrategia permitiría identificar oportunidades en los mercados menos explorados, y adaptar el producto y la estrategia de marketing para satisfacer estas necesidades específicas. Esto podría contribuir a abrir nuevos mercados y a diversificar la cartera de clientes de la empresa, lo que podría reducir la dependencia de los mercados más grandes y consolidados. Deberían realizarse estudios de mercado para analizar demanda y precios de mercado en los países con mejores resultados.

2.2. TEMPORALIDAD



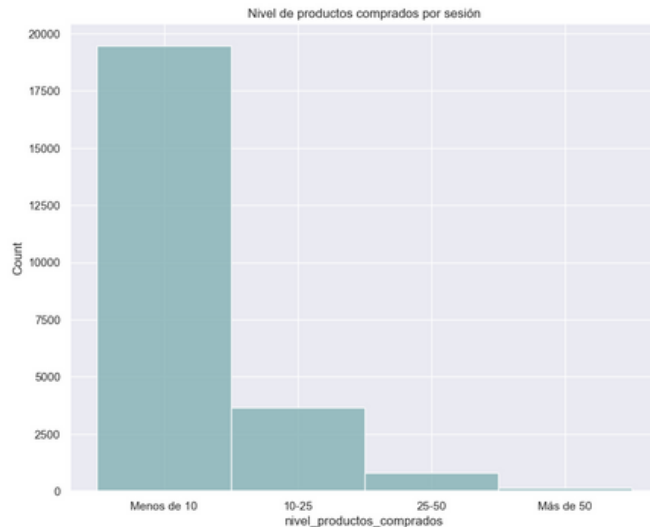
Encontramos un grave fallo en la recopilación de los datos de la variable categoría, ya que no sabemos a qué tipo de producto pertenece aquellos en rebajas (pantalones, blusas, faldas u otro). Por ende, los resultados podrían variar significativamente si todos los productos en rebajas fueran blusas, por ejemplo.

Sin embargo, en el análisis de la temporalidad que hemos llevado a cabo, se han identificado patrones significativos en la evolución de las ventas a lo largo del tiempo. Específicamente, hemos observado una disminución progresiva de las ventas desde abril hasta agosto, siendo este último mes el de menor actividad de ventas.



Respecto a las ventas semanales, podemos observar que el lunes y martes. Estos corresponden a los días en los que el cliente tiene mayor energía y entusiasmo y planifican las compras. Asimismo, el domingo es el día con mayor tiempo libre para comprar en línea, siendo un día de descanso para muchas personas.

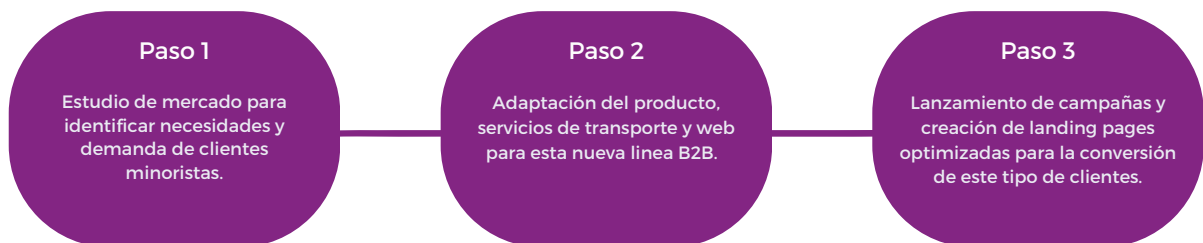
2.3. CLIENTE



ID	Cantidad de productos	Gasto	Nivel de productos comprados
22433	195	8089	Más de 50
13846	192	8538	Más de 50
22902	185	7676	Más de 50
8201	161	6456	Más de 50
19623	142	6333	Más de 50
14191	128	5232	Más de 50
8790	121	4822	Más de 50
19464	120	4447	Más de 50
20660	112	4873	Más de 50

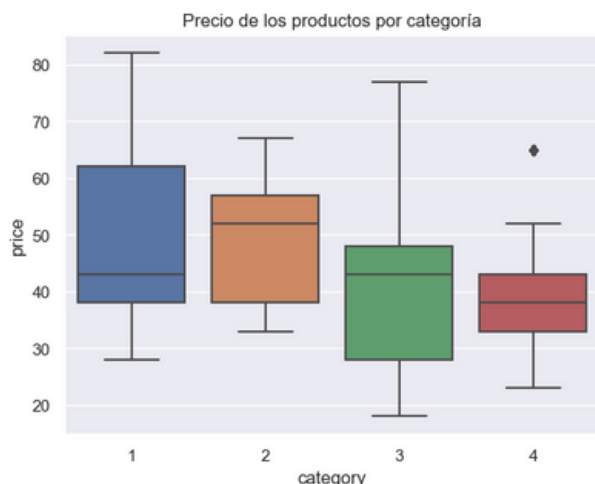
Aunque la mayoría de los usuarios han comprado menos de 10 productos en cada sesión, varias compras superaron los 150 productos. Esto puede sugerir que podría haber empresas minoristas interesadas en adquirir los productos de la marca al por mayor. Sin embargo, no se observan precios diferenciados para pedidos de gran volumen. Por ende, esto podría ser una oportunidad para una **nueva línea de negocio B2B** (business-to-business) que permita la venta directa a minoristas.

Para ello se propone la siguiente estrategia:



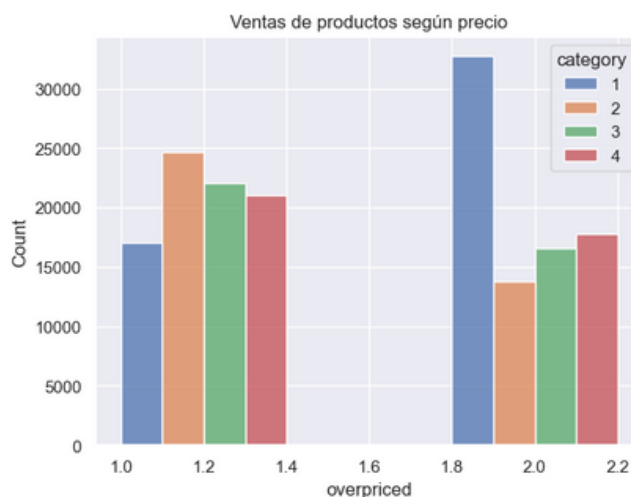
*Se considera necesario recopilar información de clientes y usuarios de la web en próximos análisis para acompañar esta estrategia.

2.4. PRECIO



Los pantalones (1) tienen los precios más altos en su totalidad. Sin embargo, la mediana más alta se encuentra en la blusas (2) lo que significa que la mayoría de las blusas tienen un precio superior a los pantalones. Como cabe esperar, los productos en rebajas (4) tienen los precios más bajos. Las faldas (3) tienen la mayor elasticidad de precios, en comparación con las blusas y los productos en rebajas.

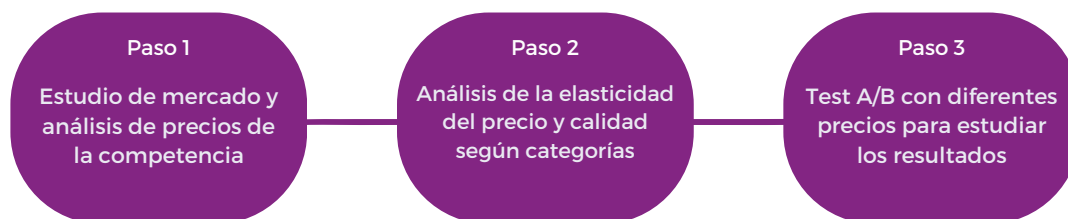
En la siguiente gráfica, donde 1 representa a los precios por encima de la media de su categoría, y 2 los que se encuentran dentro de la media, la sensibilidad de precios hacia los pantalones es mayor y los clientes prefieren precios más baratos (2) que caros (1). Sin embargo, las categorías 2, 3 y 4 tienen mayores ventas cuando los precios han sido más caros. Es decir, el precio de estas categorías parece tener una elasticidad mayor.



Esto nos lleva a los siguientes escenarios:

- 1 **Posible subida de precios:** en el caso de las blusas, faldas y productos en rebajas podríamos llevar a cabo una subida de precios, siempre y cuando se lleve a cabo un estudio de mercado, pruebas TEST A/B y los resultados confirmaran esta teoría.
- 2 **Diferenciación de productos** según calidad y precio para atender ambos públicos objetivos y aprovecharse de la elasticidad del precio.

Para ello se recomienda llevar a cabo la siguiente estrategia:



3 Modelos supervisados y no supervisados

3.1. REGLAS DE ASOCIACIÓN

El modelo de reglas de asociación es un método de aprendizaje automático que nos permite identificar patrones y relaciones entre diferentes productos en una transacción. La importancia de este modelo radica en que nos permite comprender las relaciones entre los productos y cómo se compran conjuntamente.

Su poder radica en la capacidad de recomendar productos a los clientes que podrían interesarles y que otros clientes compraron en su conjunto, incentivar la venta cruzada y llevar a cabo promociones específicas y recomendaciones basadas en las compras históricas.

Para la implementación del modelo se limpiaron los datasets de modo que cada fila representara una compra y cada columna un producto.

1	ID	clothing_codes_combined
1		A13,A16,B4,B17,B8,C56,C57,P67,P82
2		B31,B21,B24,B27,A10,A10,B27,P1,P34,P33
3		B17,CA,C7,C10,C17,P77
4		A34,A37,CA,C25
5		C21
6		C7,C15,C53,B17,B26
7		A11,C5,P60,P56,P55,P48,P50,P42,P23,P1,B4
8		C49,B23,C5,C19,C34,C40,C50,C42,B24
9		A18,A1,B26
10		A1,B1,B4,B16,B17
11		A3
12		B4,B3,B3,B30,B27,C17,P16
13		A1,A2,A5

2	A1	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A2	A20
	FALSO	FALSO	FALSO	FALSO	VERDADERO	FALSO	FALSO	VERDADERO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	VERDADERO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	VERDADERO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	VERDADERO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	VERDADERO	FALSO	FALSO	FALSO
	VERDADERO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	VERDADERO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	VERDADERO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO
	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO	FALSO

Los resultados del modelo dieron lugar a dos tipos de productos:

- 1 **Productos complementarios:** son artículos que se compran conjuntamente ya que se usan en combinación o se benefician mutuamente. Por ejemplo un conjunto de camiseta y falda.

Estrategia: crear una categoría en la página web que incentive la venta cruzada de estos artículos y realizar descuentos por lote de productos.

- Mayor probabilidad de ser comprados en conjunto
- Nueva categoría "conjuntos"
- Precio por lote

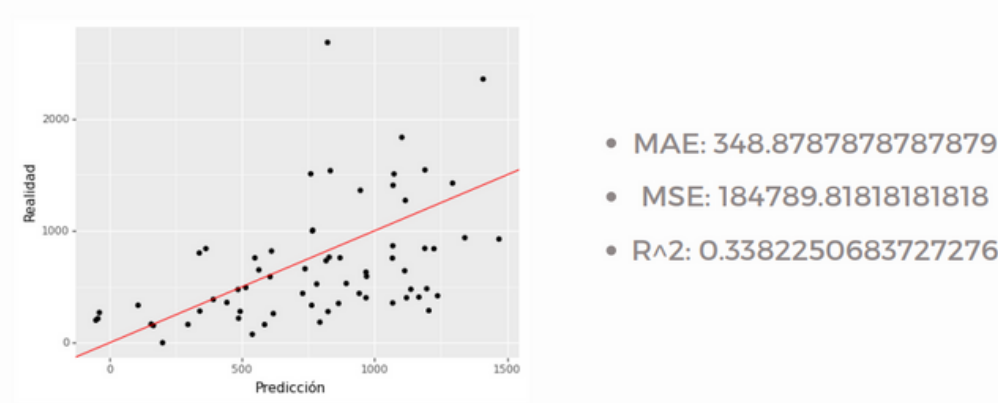
- 2 **Productos combinados:** son artículos que se venden juntos como un paquete o conjunto. Están diseñados para ser adquiridos así debido a su relación intrínseca o para ofrecer mayor valor al cliente. Por ejemplo un "conjunto de gimnasio" que incluya varios artículos para mejorar la experiencia del cliente.

Estrategia: campañas de publicidad que incentiven la compra en conjunto. Añadir en la página de compra del producto los artículos combinados con un mensaje de "combina con" u "otros también compraron"

- Interés de productos combinados
- Campañas de marketing
- Sección "combina con"

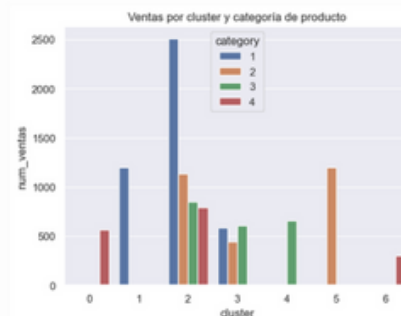
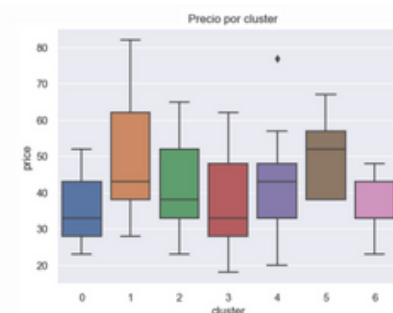
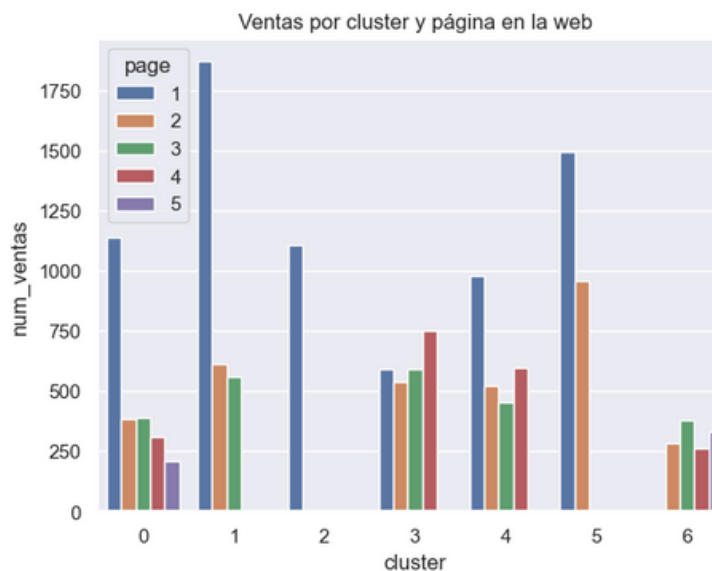
3.2. MODELOS DE REGRESIÓN

Se llevó a cabo un modelo de regresión para predecir el número de ventas según múltiples variables como precio, categoría, color, fotografía, localización y página en la web. Los resultados fueron bastante desfavorables a pesar de implementar regresión lineal, categorizar las variables e incluso árboles de regresión. Por ende, se consideró necesario mejorar la calidad de los datos e incluir las variables faltantes para desarrollar el modelo óptimo que permita a la marca predecir el número de ventas de cada producto y optimizar la producción.



3.3. CLUSTERIZACIÓN

La clusterización nos permite generar grupos de productos en base a las variables con el fin de encontrar patrones de similitud entre estas y el número de ventas. De este modo, se buscaba localizar los grupos de productos con mayores y menores ventas e implementar estrategias de mejora para la marca.



CLUSTERS

Se generaron 7 clusters (grupos) de productos con diferentes precios, categorías y localizaciones en la web. Se identificaron cuatro grupos principales (tres con las mayores ventas y uno con las peores). Las características de aquellos productos más vendidos deberán considerarse en la elaboración de la estrategia de ventas. Por otro lado, los productos con menores ventas nos ayudan a entender los desafíos de la marca y mejorar la experiencia de usuario en la web.

Los clusters identificados fueron los siguientes:



Por ende las conclusiones son las siguientes:

- Las **páginas 3,4 y 5** de la web no funcionan correctamente. Deberá llevarse a cabo un análisis de la estructura de la web para identificar el problema (falta de tráfico, mala experiencia de usuario, un problema de los productos, precios o fotografías, etc).
- Las **fotografías de frente** venden mucho más que las de perfil por lo que deberían potenciarse y colocarle en los lugares con mayor visibilidad de la web.
- Los productos con **precios más altos**, coincidentes con **faldas y pantalones**, son los que tienen mayores ventas. La venta de productos "estrella" deberá incentivarse a través de campañas y landing pages según públicos objetivo.

En primer lugar, frente a la falta de datos esenciales para los análisis y la baja calidad de los mismos, se propone la implementación de una estrategia para mejorar la recopilación de los datos. En ella se incluyen herramientas como Google Analytics, de uso gratuito.

Gracias al análisis exploratorio y a la clusterización se localizaron puntos de mejora respecto a los precios de productos y las líneas de venta. Se plantea un estudio de mercado con el fin de estudiar la venta B2B para minoristas y diferenciación de precios según la elasticidad de los mismos en algunas categorías.

De igual modo, se recomienda realizar una reestructuración de la web para mejorar el UX/UI y estudiar las páginas con menores ventas.

El modelo de reglas de asociación creado permitirá a la marca recomendar productos en la página web así como en campañas con el fin de incentivar la venta cruzada.

PROPUESTAS

Precios

Posibilidad de optimizar ventas con una subida de precios.

- Análisis de mercado
- TEST A/B

Productos

Centrarse en las categorías más vendidas.

- Productos recomendados
- Fotografías
- Otros países

Web

Mejorar UXUI y distribución de productos.

- Pantalones y faldas parte superior
- Reestructurar páginas