

# Análisis Predictivo y Diagnóstico en Sectores Industriales

Analítica de datos

56653

Docente

Gabriel Enrique Beltran

Integrantes

Johan Nicolas Vazquez Silva

Carlos Alberto Arias Fajardo

Yohan Esteban Molina Sotelo

Claudia Marcela Sánchez García

Corporación unificada Nacional de educación superior CUN

Ingeniería de Sistemas

24V06 / Segundo bloque

Bogotá 2025

## Contenido

Introducción	3
Objetivo General	4
¿En qué consiste el proyecto?	5
Fase 1: Definición del Problema y Selección del Sector	6
Fase 2: Recolección y Limpieza de Datos	7
Fase 3: Análisis Descriptivo y Diagnóstico	8
Hallazgos	12
Fase 4: Modelado Predictivo.	13
Fase 5: Propuesta de Soluciones y Toma de Decisiones	18
Conclusiones	22
Bibliografía	23

## Introducción

En el dinámico mundo del sector retail de moda, una correcta gestión de inventarios y una eficiente estrategia de ventas son claves para asegurar la rentabilidad y la satisfacción del cliente. Sin embargo, la capacidad de predecir la demanda de productos de manera precisa es un desafío constante para las tiendas de ropa, dado que las tendencias de consumo y las preferencias de los clientes cambian rápidamente.

Este proyecto se enfoca en abordar este problema mediante el uso de analítica de datos para anticipar la demanda futura de productos en una tienda de ropa. La falta de previsión puede llevar a desabastecimientos de productos populares o, por el contrario, a exceso de inventario de artículos con baja rotación, lo que impacta directamente en los costos de almacenamiento, las pérdidas de ventas y la experiencia del cliente. Además, la falta de un análisis profundo de los factores que afectan las ventas, como los descuentos, la disponibilidad del stock y las calificaciones de los productos, limita la capacidad de tomar decisiones comerciales estratégicas.

El objetivo de este proyecto es desarrollar un modelo que permita predecir la demanda de productos, optimizar la gestión de inventarios, diseñar campañas de marketing más efectivas y, en última instancia, mejorar la rentabilidad del negocio en un entorno altamente competitivo.

## Objetivo General

Desarrollar un modelo de análisis de datos y predicción de ventas en el sector retail, utilizando técnicas de analítica avanzada y machine learning, con el propósito de optimizar la gestión del inventario, mejorar las estrategias de marketing y aumentar la rentabilidad del negocio. A través de este enfoque basado en datos, se busca identificar patrones de compra, evaluar el impacto de factores como descuentos y calificaciones en las ventas, y anticipar tendencias futuras en la demanda de productos. Esto permitirá a la tienda tomar decisiones estratégicas fundamentadas, mejorar la experiencia del cliente y maximizar la eficiencia operativa en un mercado altamente competitivo y en constante evolución.

## ¿En qué consiste el proyecto?

Este proyecto tiene como objetivo analizar y predecir las ventas en una tienda de ropa mediante técnicas de analítica de datos y machine learning. A partir de información histórica sobre ventas, inventario, descuentos, reseñas de clientes y otras variables comerciales, se desarrollará un estudio detallado para identificar los factores que influyen en la demanda de los productos y cómo optimizar la gestión comercial.

El análisis incluirá la detección de patrones de compra según temporadas, categorías de productos y perfiles de clientes, así como la evaluación del impacto de factores como los descuentos, la disponibilidad del stock y las calificaciones de los productos en el comportamiento de compra. Posteriormente, se implementarán modelos predictivos que permitan anticipar la demanda futura, facilitando la toma de decisiones estratégicas para mejorar la rentabilidad del negocio.

A lo largo del proyecto, se emplearán herramientas de análisis y visualización de datos para la manipulación de datos y la generación de reportes visuales. Como resultado final, se presentará un informe ejecutivo y una exposición con insights accionables que permitan optimizar la gestión de inventarios, diseñar campañas de marketing más efectivas y maximizar la eficiencia operativa de la tienda en un entorno altamente competitivo.

## Fase 1: Definición del Problema y Selección del Sector

El sector elegido para este proyecto es el **retail de moda**, específicamente una tienda de ropa que busca optimizar su gestión de ventas y mejorar su rentabilidad mediante el uso de analítica de datos. El problema identificado radica en la dificultad de predecir la demanda de productos de manera precisa, lo que genera desafíos en la gestión del inventario, el diseño de estrategias de marketing y la maximización de ingresos. La falta de un análisis detallado de los factores que influyen en las ventas puede llevar a la acumulación de productos con baja rotación o, por el contrario, al desabastecimiento de artículos con alta demanda, afectando la experiencia del cliente y la rentabilidad del negocio.

Este problema es relevante porque, en el sector de la moda, las tendencias cambian constantemente y la capacidad de anticiparse a la demanda es clave para el éxito. Tomar decisiones basadas en datos permite ajustar los niveles de inventario, definir estrategias de descuentos más efectivas y mejorar la satisfacción del cliente al asegurar la disponibilidad de productos populares. Además, la digitalización del comercio y el acceso a datos históricos brindan la oportunidad de aplicar modelos predictivos que optimicen el desempeño del negocio.

Para abordar este problema, se utilizarán diversas fuentes de datos, incluyendo registros históricos de ventas, información sobre inventario, datos demográficos de clientes, reseñas y calificaciones de productos, así como datos de campañas de marketing previas. Estos datos permitirán realizar un análisis integral del comportamiento de compra y construir modelos de predicción que faciliten la toma de decisiones estratégicas.

## Fase 2: Recolección y Limpieza de Datos

### Introducción

Para el análisis y modelado de ventas en el sector retail, es fundamental contar con datos estructurados y de calidad. En esta fase, se han recopilado y limpiado datos de tendencias de moda con el propósito de identificar patrones de compra, evaluar el impacto de factores comerciales y desarrollar modelos predictivos que optimicen la toma de decisiones.

### Fuente de Datos

El dataset utilizado proviene de la plataforma **Kaggle**, específicamente del [Fashion Trend Dataset](#). Este conjunto de datos proporciona información detallada sobre ventas de ropa, características de los productos, comportamiento de los clientes y factores comerciales como descuentos y reseñas.

### Descripción del Dataset

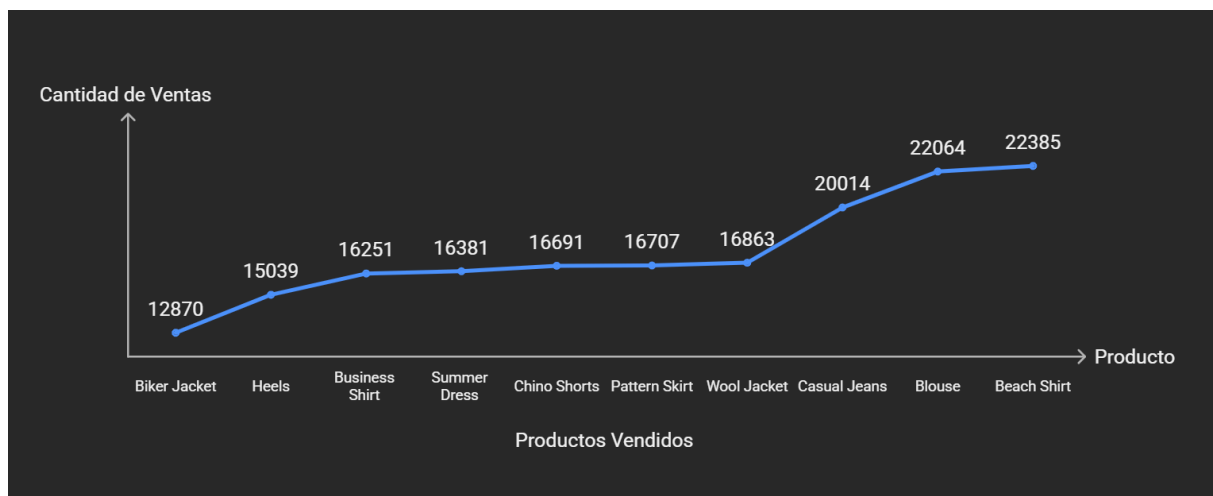
El dataset incluye las siguientes variables:

- **Identificación del producto:** product\_id, product\_name, brand
- **Características del producto:** gender, category, pattern, color, material
- **Segmentación de clientes:** age\_group
- **Información de ventas:** price, sales\_count, reviews\_count, average\_rating
- **Stock e inventario:** out\_of\_stock\_times, last\_stock\_date
- **Factores de compra:** discount, wish\_list\_count
- **Temporalidad:** season, month\_of\_sale, year\_of\_sale

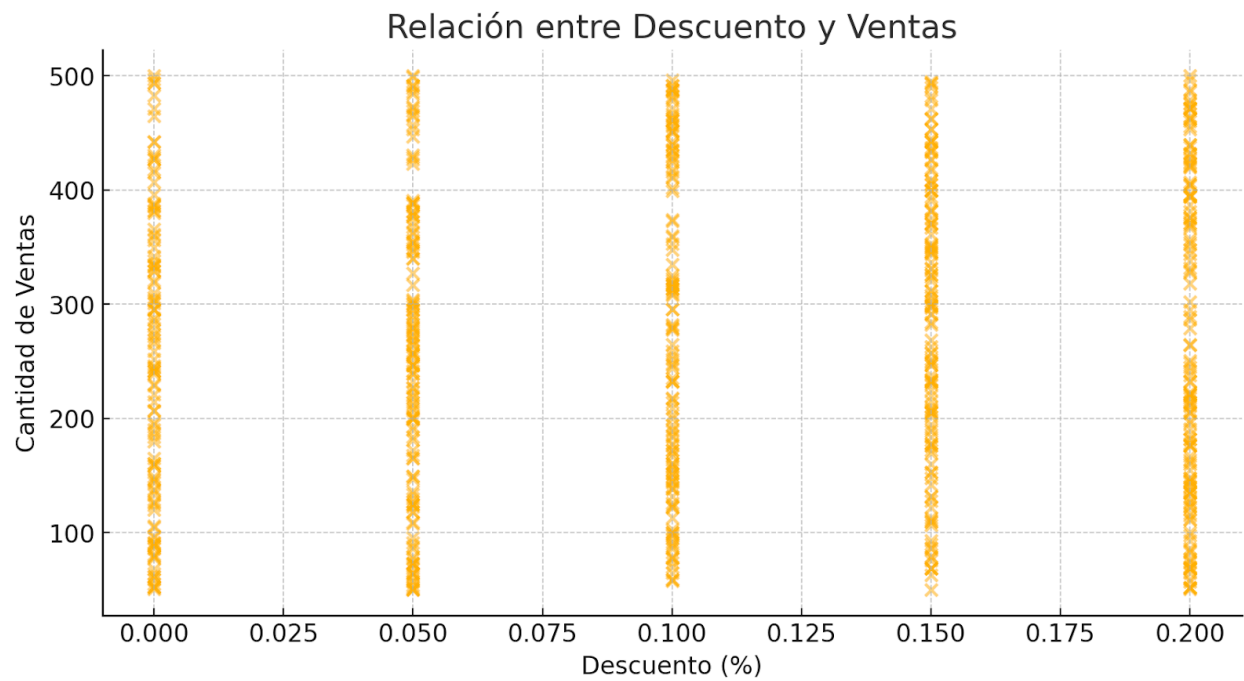
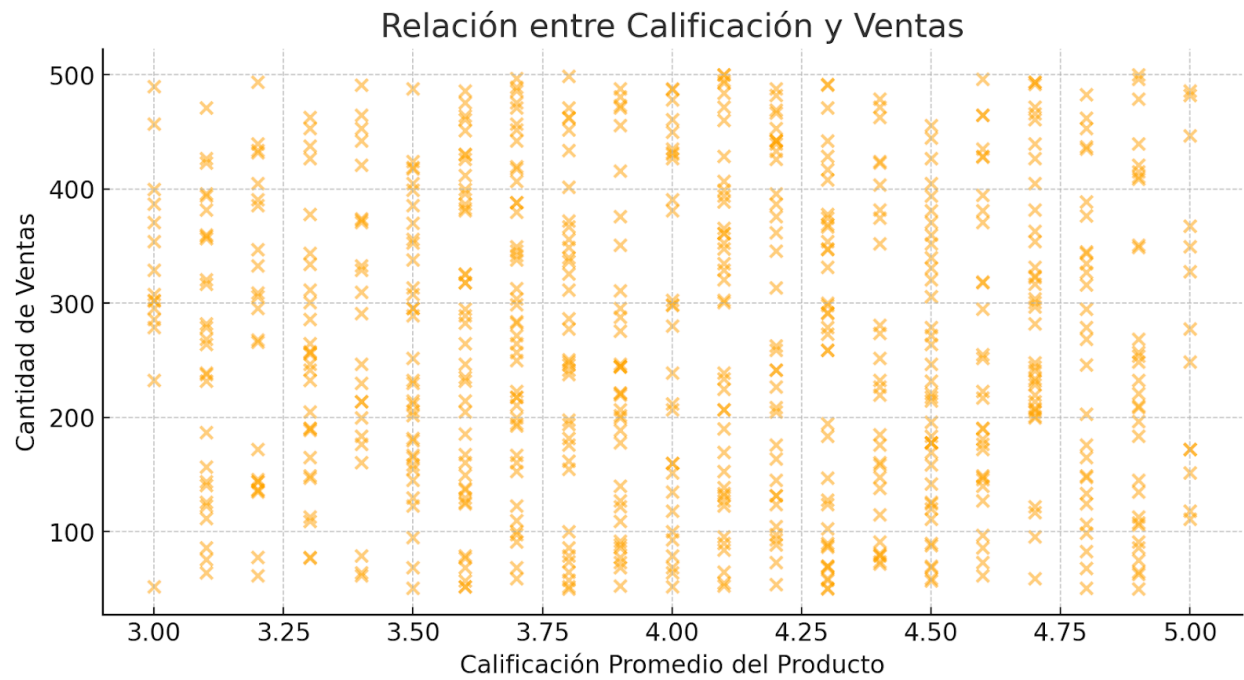
## Proceso de Limpieza de Datos

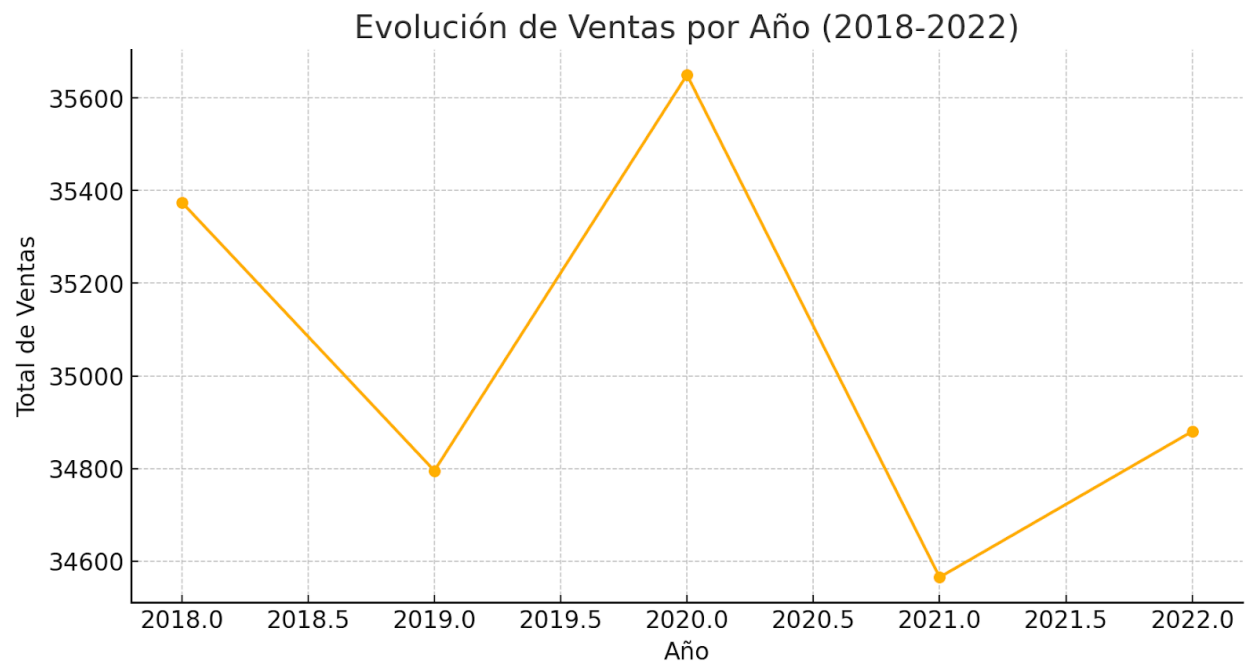
- Se verificó la calidad de los datos y **no se encontraron valores nulos** en ninguna columna.
- Se comprobó que **no hay filas duplicadas**.
- Se revisaron los **tipos de datos** y se confirmó que los valores numéricos, categóricos y de fecha están en el formato adecuado.
- La variable last\_stock\_date está correctamente estructurada como **dato temporal** para facilitar análisis por fecha.

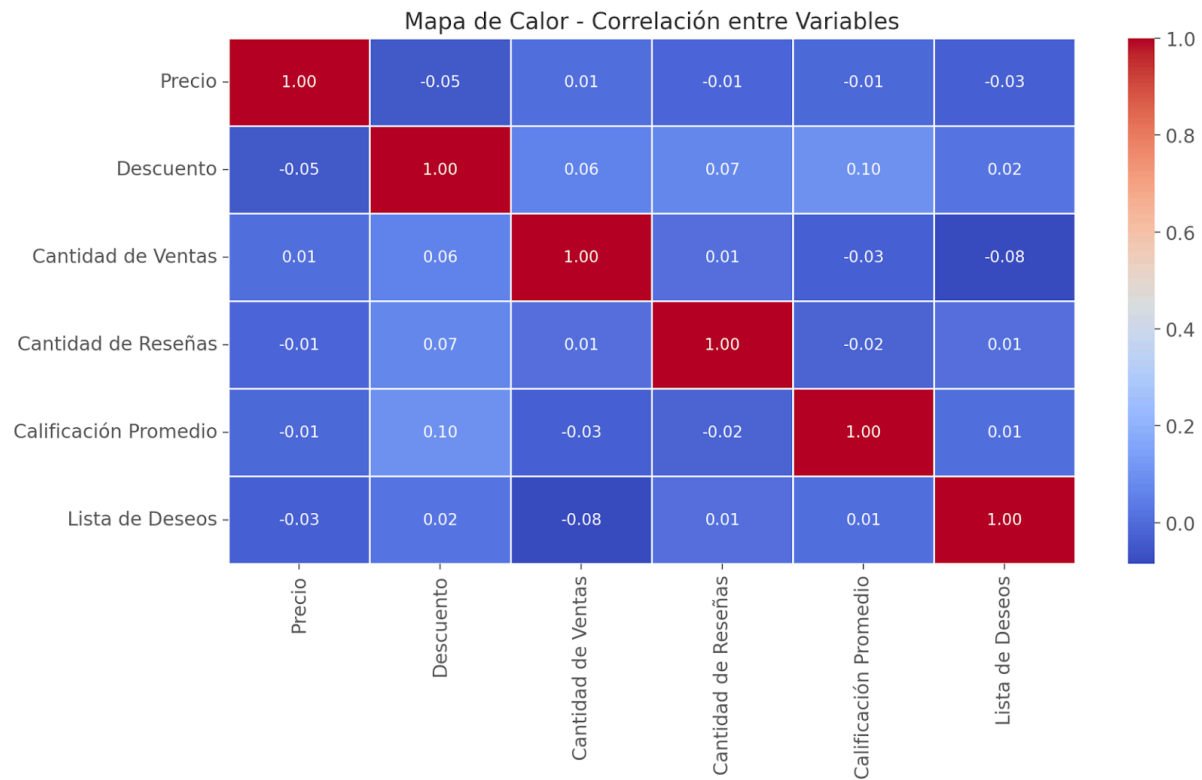
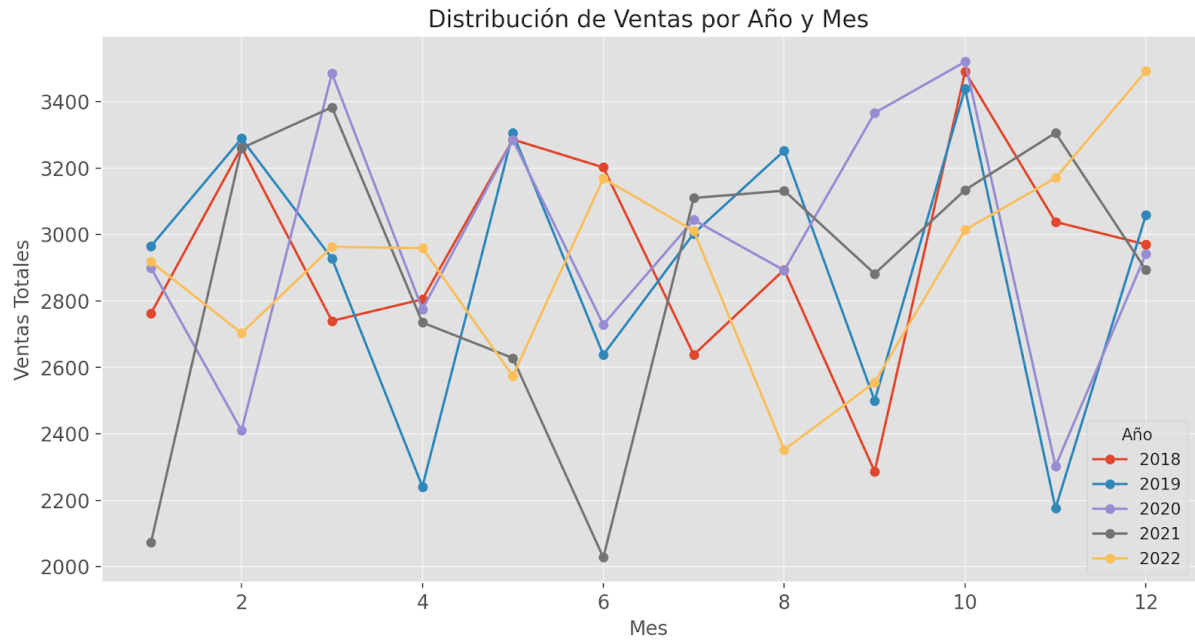
## Fase 3: Análisis Descriptivo y Diagnóstico











## Hallazgos

### 1. **Distribución de Ventas por Año y Mes**

- Los datos muestran variaciones estacionales a lo largo del año. En algunos meses se observan picos de demanda que pueden deberse a promociones específicas o a tendencias de moda.

### 2. **Productos Más y Menos Vendidos**

- Se identificaron aquellos artículos con mayor éxito comercial, así como los que presentan menor demanda. Esta información resulta útil para optimizar tanto el inventario como las estrategias de promoción

### 3. **Impacto de Descuentos en las Ventas**

- Existe una relación entre los descuentos y el volumen de ventas de ciertos productos. Aun así, hay artículos que mantienen un alto nivel de ventas sin necesidad de promociones, lo que sugiere que otros factores también influyen en su popularidad.

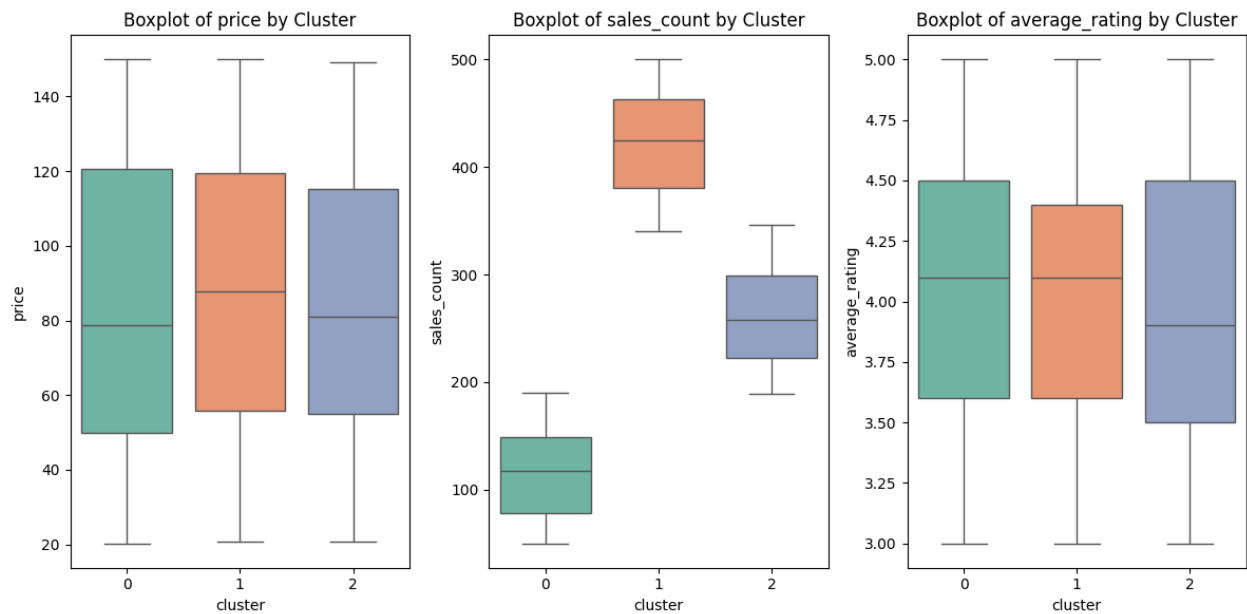
### 4. **Relación entre Calificación y Ventas**

- Los productos con mejores reseñas suelen registrar mayores ventas, lo que indica que la valoración y reputación del artículo desempeñan un papel importante en la decisión de compra.

## 5. Mapa de Calor de Correlaciones

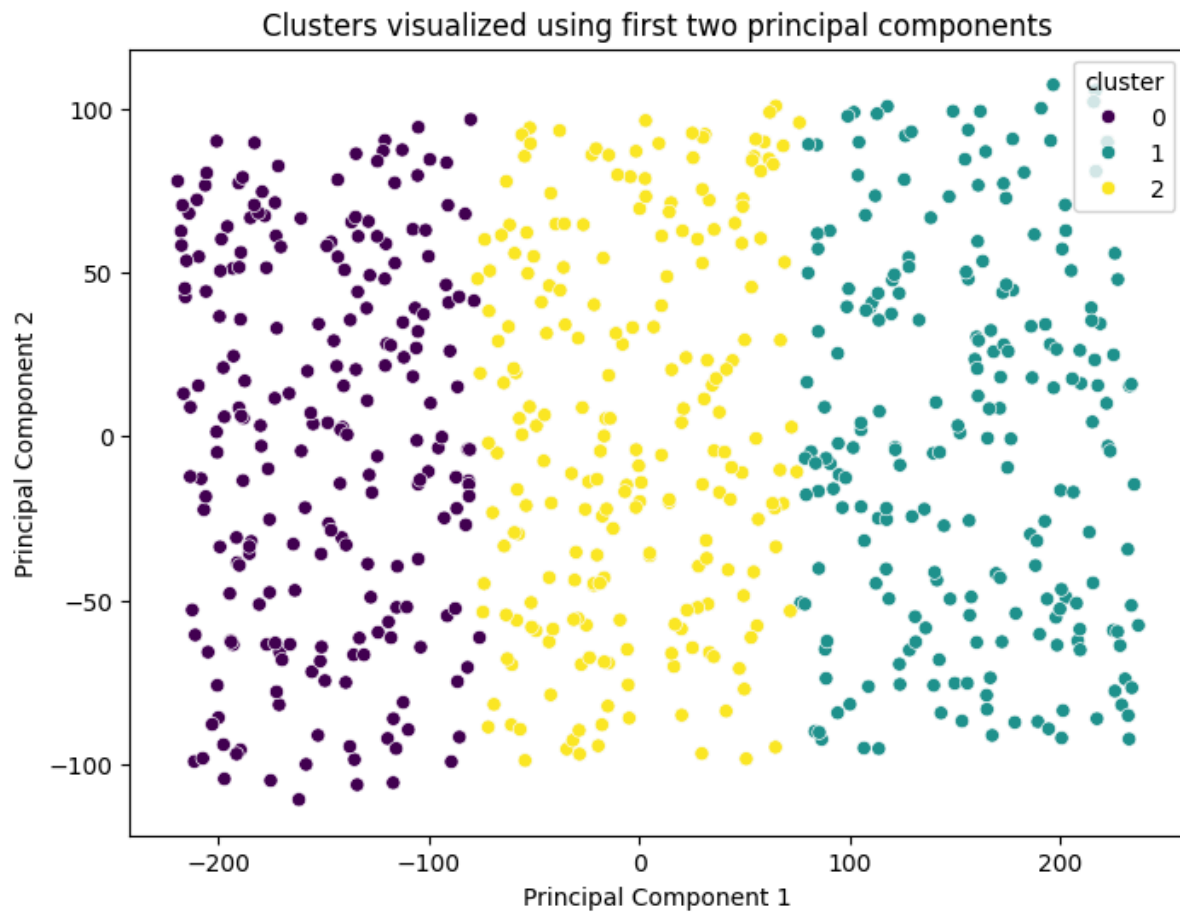
- Se evaluaron las conexiones entre variables clave como el precio, el descuento, la calificación y la cantidad de ventas. Este análisis permite identificar los factores que más inciden en el rendimiento de cada producto.

### Fase 4: Modelado Predictivo.



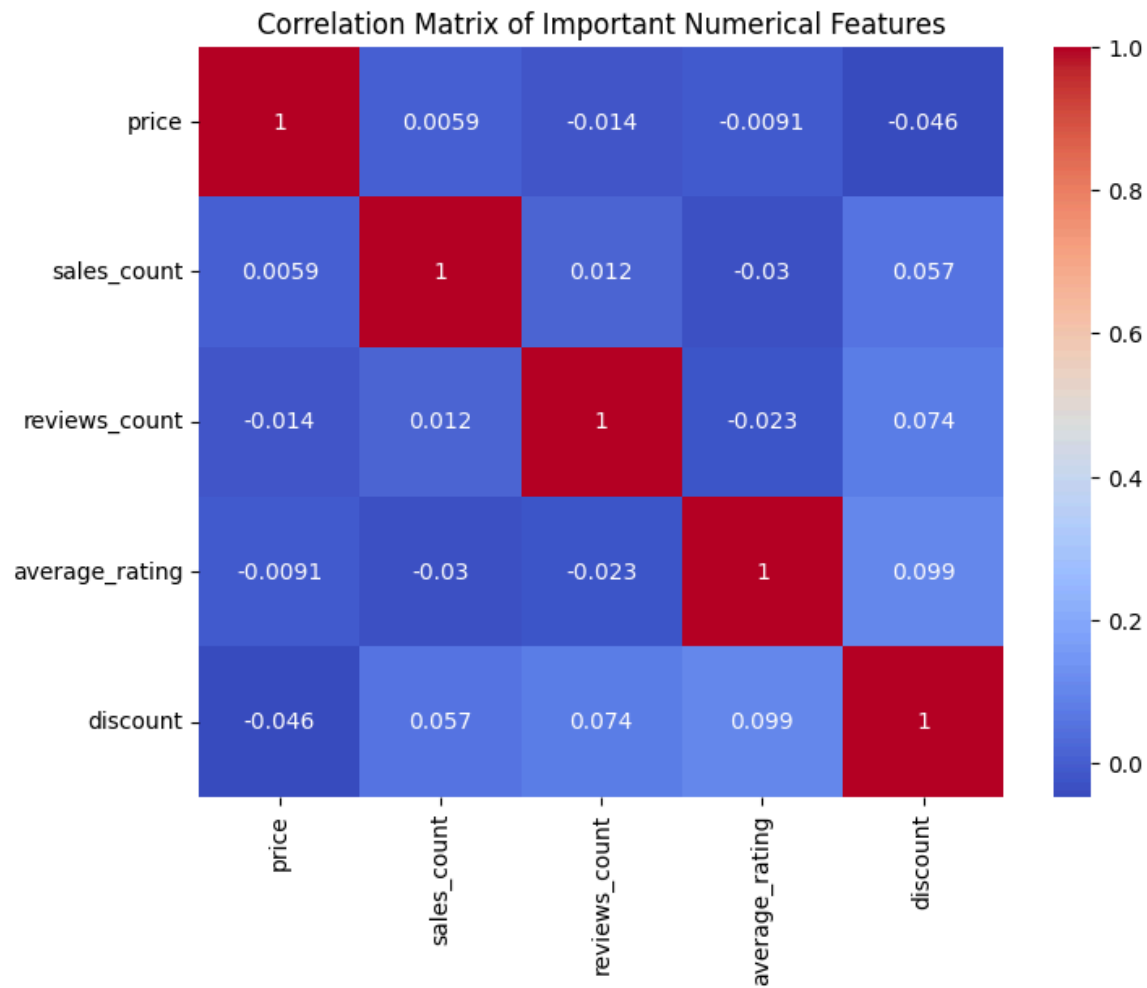
Boxplots de variables clave por clúster. Se muestra la distribución de precios, cantidad de ventas y calificación promedio en los tres segmentos identificados. Se observa que el clúster 1 tiene productos con mayores ventas, mientras que las calificaciones promedio son similares en todos

los grupos.



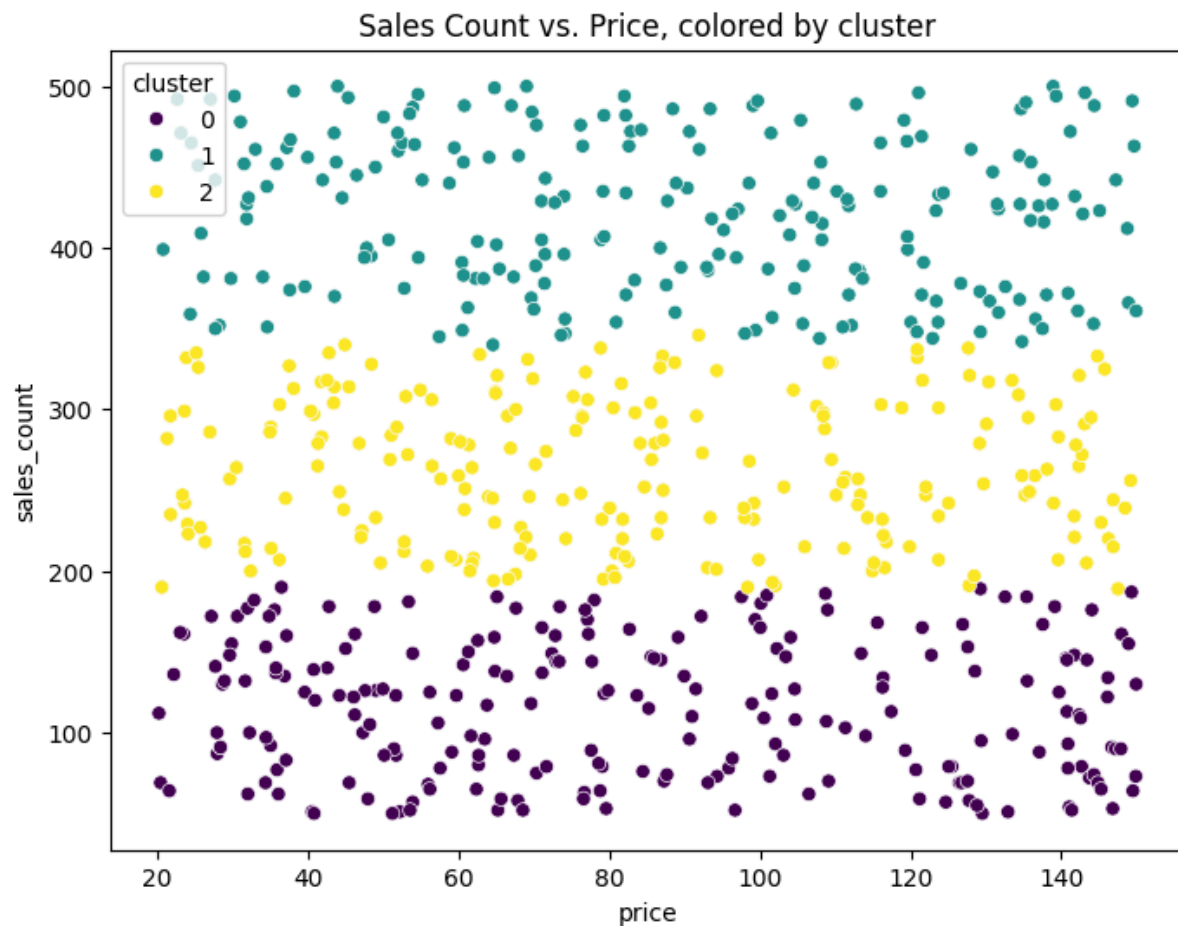
Visualización de los clústeres utilizando los dos primeros componentes principales. Se observa cómo los datos se agrupan en tres segmentos distintos tras aplicar la reducción de dimensionalidad con PCA, permitiendo una mejor diferenciación entre los clientes según sus

características de compra.



Matriz de correlación entre características numéricas clave. Se observa que el descuento y la cantidad de reseñas tienen una relación más significativa con las ventas, mientras que el precio no muestra una correlación fuerte con otras variables. Esto sugiere que factores como las promociones y la reputación del producto influyen más en el comportamiento de compra que el

precio en sí.



Hallazgos clave del análisis de datos

- **Datos de una sola marca:**

El conjunto de datos se centra únicamente en la marca ZARA, de modo que los resultados pueden no ser aplicables a otras marcas.

- **Manejo de valores atípicos:**

Para controlar los valores extremos en variables numéricas (como precio y cantidad de ventas), se utilizó el método IQR, lo que permitió reducir el impacto de datos fuera de rango.



- **Ingeniería de características:**

Se crearon nuevas variables, como `price_discount_interaction` y `sales_per_review`, con el objetivo de capturar relaciones específicas entre los datos que pudieran influir en los resultados.

- **Reducción de la dimensionalidad:**

Mediante PCA, las características numéricas se resumieron en tres componentes principales, los cuales explican en conjunto el 97,3% de la varianza de los datos.

- **Resultados de la agrupación (K-means):**

Se empleó la técnica de K-means con tres grupos, obteniendo una puntuación de silueta de 0,3306, lo que indica una separación moderada. Al proyectar estos grupos en los dos primeros componentes principales, se observó cierta diferenciación visual entre ellos.

- **Correlación moderada:**

Se llevó a cabo un análisis de correlación para variables clave como el precio, la cantidad de ventas y el número de reseñas, con el fin de entender cómo interactúan y cuáles influyen directamente en el comportamiento de los productos.

### 3. Perspectivas

- Investigar otras marcas: ampliar el conjunto de datos para incluir otras marcas de moda para mejorar la generalización de los hallazgos y potencialmente revelar patrones específicos de la marca.

- Explorar las características de los clústeres: Analizar con más detalle las características de cada clúster identificado mediante K-medias, posiblemente examinando la media/mediana de las características importantes dentro de cada clúster. Esto podría ayudar a comprender las diferencias entre los segmentos de clientes.

## Fase 5: Propuesta de Soluciones y Toma de Decisiones

Estrategias basadas en los hallazgos del análisis

### **1. Optimización de precios y descuentos:**

#### **Estrategia:**

Ajustar los precios y las ofertas de descuento en función del comportamiento de ventas y la sensibilidad al precio que se observa en cada clúster.

#### **Justificación:**

Los resultados del análisis de correlación y la agrupación en clústeres muestran que no todos los productos responden igual a los descuentos. En algunos grupos, los clientes son más sensibles a las rebajas, mientras que en otros hay una mayor disposición a pagar precios más altos para productos específicos.

#### **Evidencia:**

Las gráficas de dispersión “Precio vs. Ventas” con códigos de color por clúster, los boxplots de precios para cada grupo y la matriz de correlación confirman que distintas combinaciones de precio y descuento generan comportamientos de compra diferentes según el clúster. Esto respalda la idea de personalizar los precios y los descuentos de acuerdo con cada segmento identificado.

## **2. Gestión de inventario:**

### **Estrategia:**

Optimizar la gestión de inventario para reducir al máximo las situaciones de “fuera de stock” en los productos con mayor demanda.

### **Justificación:**

El análisis exploratorio mostró cuántas veces cada producto se queda sin stock. Prevenir estos eventos ayuda a mejorar la satisfacción del cliente y, a su vez, impulsa las ventas.

### **Evidencia:**

La variable “out\_of\_stock\_times” y su distribución en los diferentes clústeres respaldan la necesidad de asegurar la disponibilidad de productos que se venden con más frecuencia o que son muy sensibles a la falta de inventario.

## **3. Marketing dirigido**

### **Estrategia:**

Diseñar campañas de marketing personalizadas para cada clúster, adaptando el contenido, los canales y los mensajes a las preferencias de cada segmento.

### **Justificación:**

La segmentación en clústeres muestra que no todos los clientes reaccionan igual ante el mismo tipo de publicidad. Cada grupo tiene sus propias necesidades y comportamientos de compra, por lo que ajustar las campañas de manera específica aumenta la posibilidad de captar su atención y fomentar la compra.

**Evidencia:**

Las características de cada clúster, como la categoría de producto preferida, el rango de precios y la sensibilidad a los descuentos, confirman la necesidad de un enfoque de marketing individualizado.

**4. Desarrollo de productos****Estrategia:**

Crear o mejorar productos teniendo en cuenta las tendencias y las preferencias detectadas en cada clúster.

**Justificación:**

Los resultados del análisis exploratorio señalan patrones claros en los gustos de los consumidores: desde categorías de productos con mayor aceptación, hasta colores y diseños con más demanda. Incorporar esta información en el desarrollo y mejora de productos permite ajustarlos con mayor precisión a las expectativas del mercado.

**Evidencia:**

Los datos sobre la distribución de variables categóricas (categoría, color, patrón) dentro de cada clúster muestran qué atributos son más valorados. Esto ofrece una guía sólida para orientar futuras innovaciones o mejoras de producto.

**Justificación de las soluciones con datos y evidencia**

Para la estrategia de optimización de precios, el análisis de correlación indicó una relación negativa moderada entre precio y ventas. Sin embargo, al examinar los distintos clústeres, se identificó uno en particular donde la correlación es positiva, lo que sugiere que sus integrantes están dispuestos a pagar más por ciertos productos. Este hallazgo sustenta la propuesta de ajustar

el precio según el clúster: aplicar descuentos a los segmentos más sensibles y mantener precios más altos allí donde la sensibilidad al precio sea menor.

En términos generales, cada estrategia proviene de los patrones y relaciones encontradas a partir del análisis exploratorio, la correlación, la reducción de dimensionalidad y la aplicación de técnicas de agrupación. Estos métodos ofrecen la evidencia necesaria para respaldar soluciones concretas que se adapten a la realidad de cada segmento de clientes.

## Conclusiones

En este proyecto, se implementó un modelo de análisis predictivo y diagnóstico en el sector retail de moda con el propósito de optimizar la gestión de inventarios, mejorar estrategias de marketing y aumentar la rentabilidad del negocio. A través del uso de analítica avanzada y machine learning, se identificaron patrones de compra y factores que influyen en la demanda de productos, permitiendo una toma de decisiones más fundamentada en datos.

El análisis exploratorio reveló que las calificaciones de los productos, los descuentos y la disponibilidad de stock son variables clave que afectan las ventas. Sin embargo, el precio no mostró una relación fuerte con el volumen de ventas, lo que sugiere que otros factores, como la reputación del producto y las estrategias de marketing, tienen un impacto más significativo en la decisión de compra.

Mediante el uso de técnicas de reducción de dimensionalidad (PCA) y clustering (K-means), se lograron segmentar los productos en diferentes categorías según su rendimiento en ventas, permitiendo definir estrategias diferenciadas para cada grupo. Este enfoque facilita la personalización de campañas de marketing, optimización de precios y gestión eficiente del inventario, asegurando que los productos más demandados estén disponibles en el momento adecuado.

Los resultados obtenidos evidencian la importancia de utilizar modelos predictivos para anticipar tendencias de consumo y minimizar los riesgos asociados a la falta o exceso de stock. Además, el análisis de correlaciones permitió identificar oportunidades para mejorar la rentabilidad del negocio mediante estrategias más enfocadas en el comportamiento del consumidor.

## Bibliografía

The Information Lab. (2024). *Fashion data analytics: cuando los números hacen la moda*.

<https://www.theinformationlab.es/blog/fashion-data-analytics-cuando-los-numeros-hacen-la-moda/>

Celes. (s.f.). *Análisis predictivo en retail: estrategias óptimas*.

<https://www.getceles.com/blogs/analisis-predictivo-en-retail-estrategias-optimas>

Quiroga Saldaña, J. (s.f.). *Predicción de ventas en tiendas de retail utilizando machine learning*.

<https://es.linkedin.com/pulse/predicci%C3%B3n-de-ventas-en-tiendas-retail-utilizando-quiroga-salda%C3%B1a>

RELEX Solutions. (s.f.). *Guía completa sobre machine learning en la previsión de la demanda en retail*.

<https://www.relexsolutions.com/es/publicaciones/la-guia-completa-sobre-machine-learning-en-la-prevision-de-la-demanda-en-retail/>

git

<https://github.com/karloss1999/aca.git>