

# **Análisis exploratorio de datos para decisiones comerciales Evaluación del módulo ANÁLISIS EXPLORATORIO DE DATOS**

**Módulo ABP 5**

**Elaborado por:** Nicolás Pérez Cerdá

**Fecha:** 22/02/2026

# 1. Contexto y Objetivo del Análisis

La empresa ComercioYA, dedicada a las ventas online, recopila datos sobre el comportamiento de sus clientes. El objetivo de este informe es presentar los resultados del Análisis Exploratorio de Datos (EDA) aplicado a una muestra de clientes, con el fin de detectar patrones de comportamiento, identificar insights claves y para la toma de decisiones estratégicas.

El dataset analizado corresponde a una muestra **simulada** del comportamiento de clientes de la empresa ComercioYA. Se incluyen variables relevantes para el análisis comercial, tales como **visitas mensuales**, **monto de compra**, **número de devoluciones** y **categoría de producto más frecuente**.

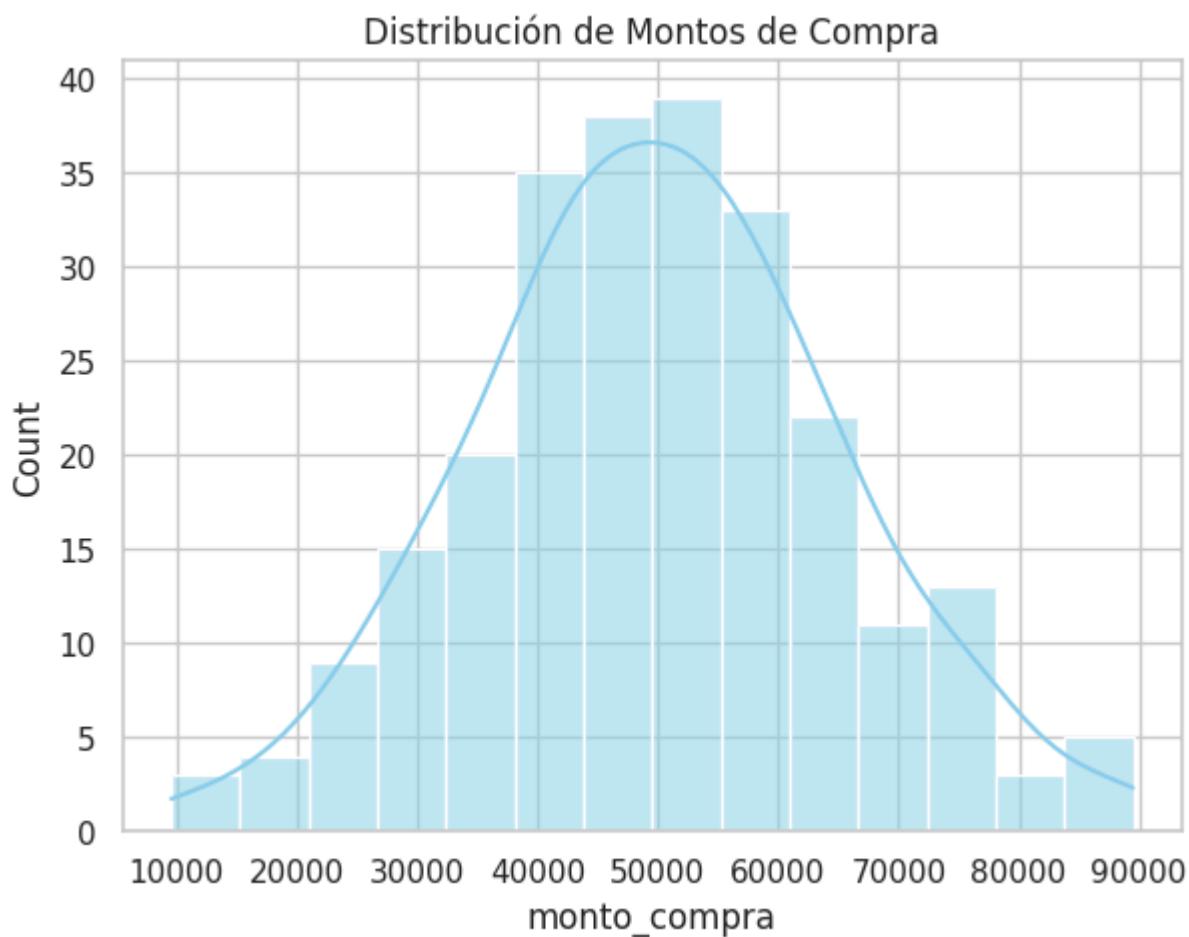
Tipos de variables analizadas:

- Variables numéricas: visitas\_mensuales, monto\_compra, devoluciones
- Variables categóricas: categoria\_frecuente

# 2. Metodología y Herramientas

El análisis se realizó utilizando el lenguaje de programación Python, procesando la información con la librería **pandas** y desarrollando el análisis visual y estadístico mediante las librerías **matplotlib** y **seaborn**. Se procesaron datos que incluyen variables como visitas mensuales, montos de compra, devoluciones y categorías frecuentes de productos.

### 3. Análisis Estadístico Descriptivo



La distribución del monto de compra muestra variabilidad moderada, con presencia de valores extremos superiores.

El análisis de la matriz de correlación muestra que no existen relaciones lineales significativas entre las variables analizadas, ya que los coeficientes de Pearson se encuentran cercanos a cero.

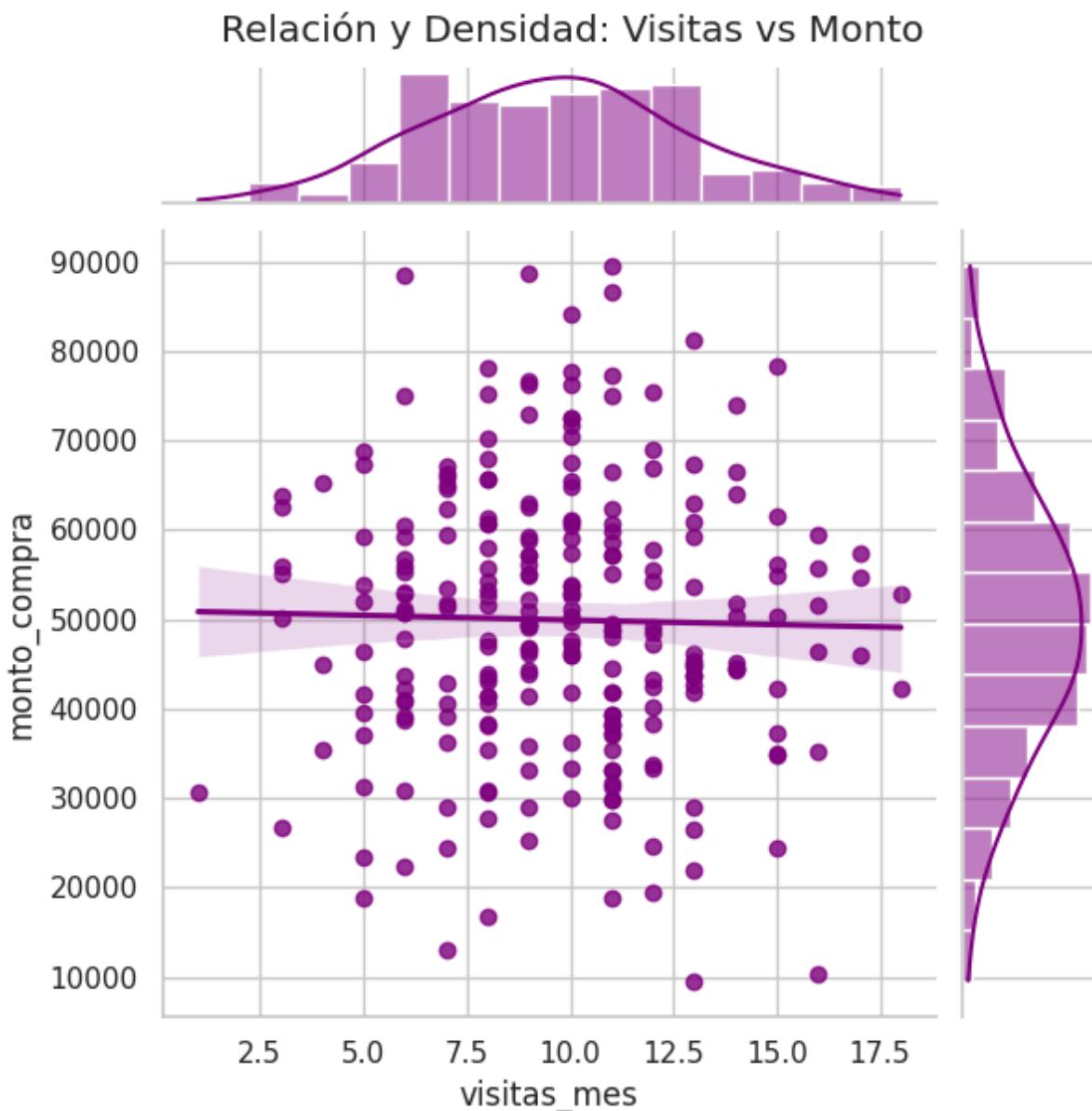
Esto indica que el comportamiento del monto de compra no está fuertemente explicado por las variables incluidas en el análisis, sugiriendo que **podrían existir otros factores** no considerados que influyen en las decisiones de compra de los clientes.

Este resultado es relevante, ya que demuestra la importancia de incorporar nuevas variables o profundizar el análisis para comprender mejor el comportamiento comercial.

## A. Relación entre variables

A través del análisis de correlación y regresión lineal, se analizó estadísticamente que el comportamiento de gasto de los clientes no es influenciado el número de visitas a la plataforma

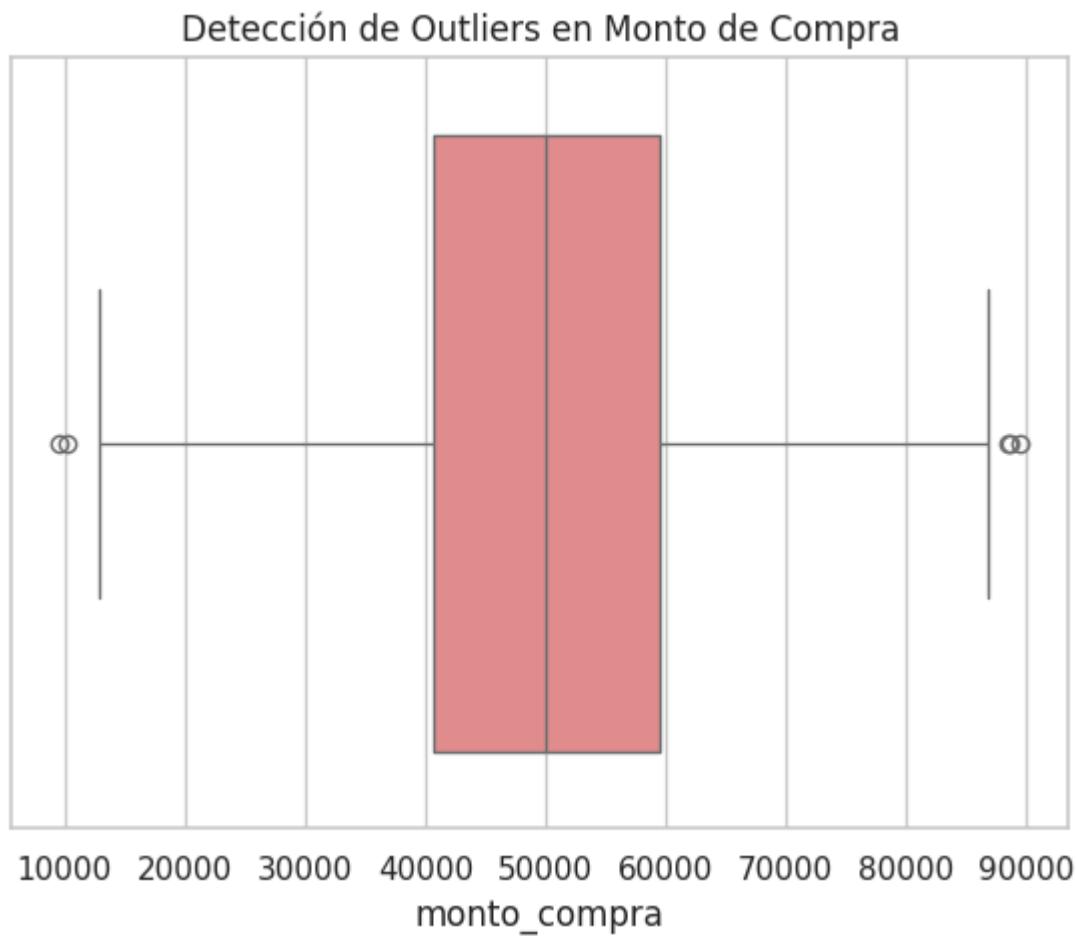
Este comportamiento puede observarse claramente en el gráfico de dispersión con línea de regresión, donde se evidencia una tendencia dispersa entre ambas variables.



Debido a que no se identificaron relaciones lineales fuertes entre las variables analizadas, se recomienda ampliar el dataset incorporando nuevas variables como campañas de marketing, antigüedad del cliente o comportamiento de navegación, con el fin de mejorar la capacidad explicativa del análisis.

## B. Segmentación de Clientes de Alto Valor (Outliers)

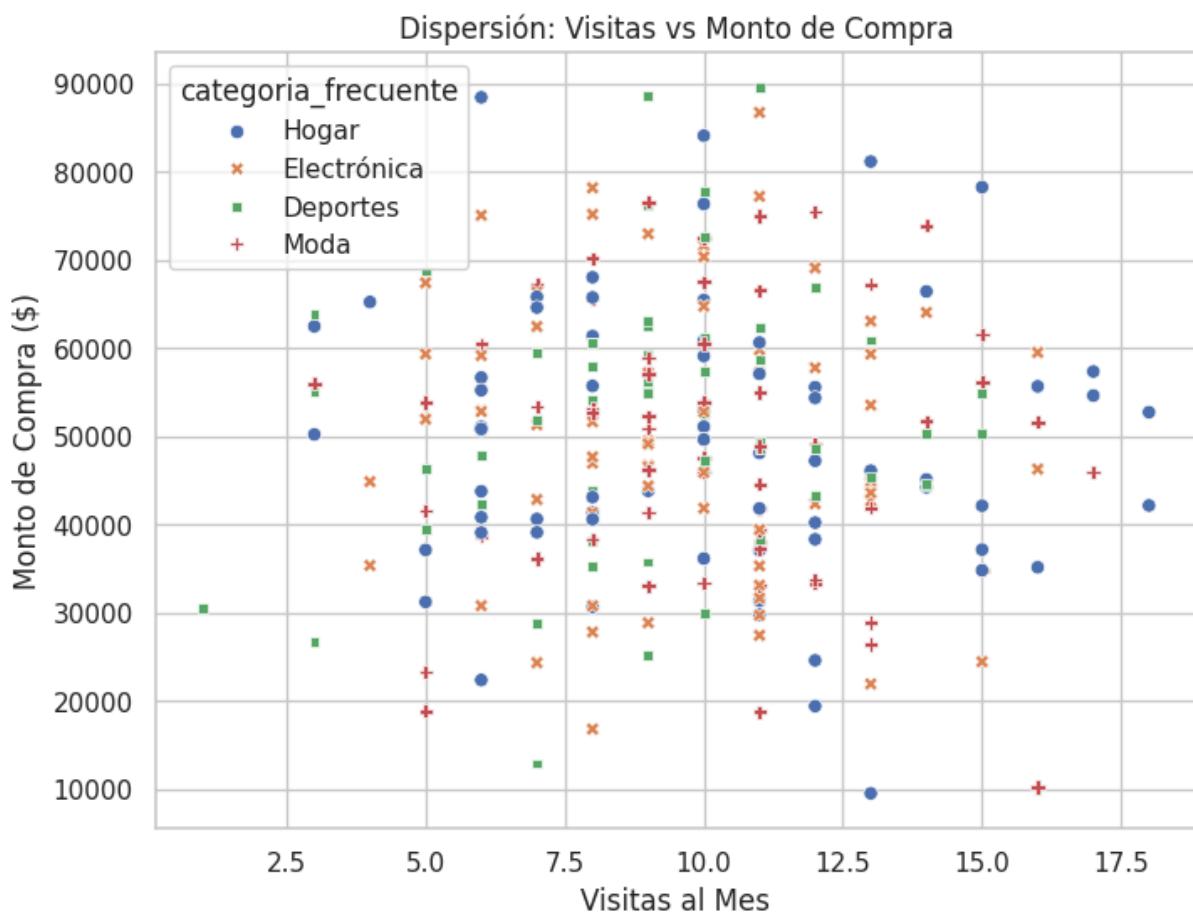
Durante el análisis de dispersión de los datos, se detectaron valores atípicos (outliers) en el monto de compra.



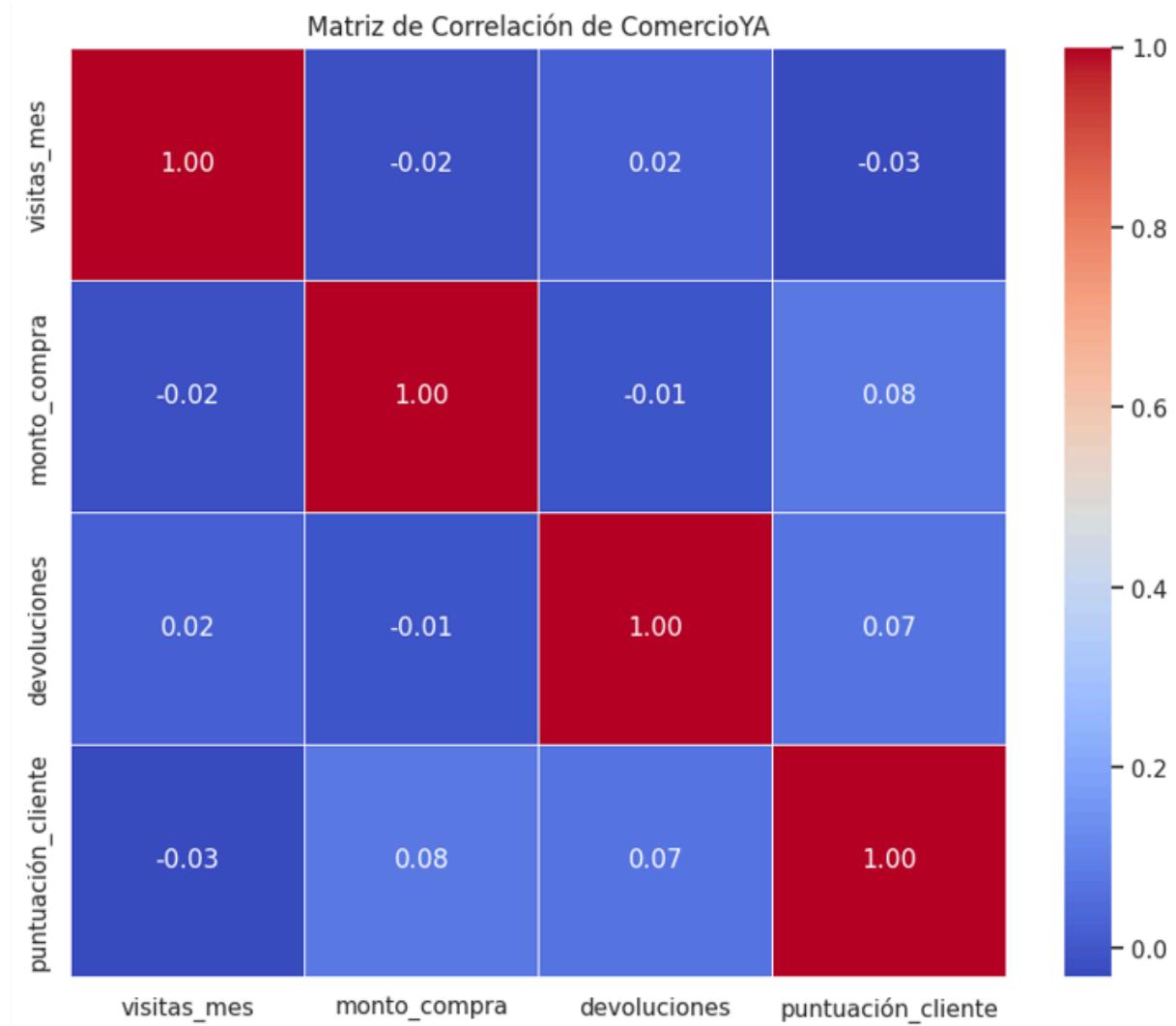
- Estos valores representan un segmento de "Clientes de Alto Valor", "VIP", etc o compras mayoristas que operan fuera del rango normal de gasto. Este grupo minoritario genera un impacto positivo en los ingresos totales.

El boxplot del monto de compra permite visualizar claramente la presencia de valores atípicos superiores, los cuales representan clientes de alto valor.

## 4. Análisis de Correlación y Modelo Predictivo



El gráfico de dispersión muestra una pendiente cercana a cero, lo que confirma la baja relación entre las variables.



El análisis de la matriz de correlación muestra que no existen relaciones lineales significativas entre las variables analizadas, ya que los coeficientes de Pearson se encuentran cercanos a cero.

Basado en los datos, se proponen las siguientes acciones a gerencia:

- Estrategia de Fidelización (Programa VIP):** Implementar un programa de recompensas o atención personalizada y exclusiva para los clientes identificados como outliers superiores. Asegurar la retención de este segmento es de real importancia para mantener el volumen de facturación.
- Reanálisis de datos:** Debido a que no se identificaron relaciones lineales fuertes entre las variables analizadas, se recomienda ampliar el dataset incorporando nuevas variables como campañas de marketing, antigüedad del cliente o comportamiento de navegación, con el fin de mejorar la capacidad explicativa del análisis.

## **5. Reflexión Técnica y Próximos Pasos**

El análisis estadístico realizado permitió evaluar la relación entre variables, evidenciando que no se identificaron correlaciones lineales fuertes dentro del dataset analizado. Esto sugiere que el comportamiento de compra de los clientes podría estar influenciado por variables adicionales no consideradas en el modelo, lo que abre la posibilidad de ampliar el análisis en futuras etapas.