

Informe Técnico de Análisis Exploratorio de Datos (EDA)



Proyecto: ComercioYA - Comportamiento de clientes

Autora: Cintia Gisele García Ortega

Fecha: 23-02-2026

Resumen:

Este informe presenta un análisis exploratorio de datos aplicado a un dataset simulado de clientes de una tienda online, con el objetivo de identificar patrones y relaciones entre variables que faciliten la toma de decisiones comerciales.

Introducción:

El objetivo de este análisis es aplicar técnicas de Análisis Exploratorio de Datos (EDA) para comprender el comportamiento de los clientes de la tienda online ficticia ComercioYA. Se busca analizar variables como visitas al sitio, número de compras, monto gastado, devoluciones y calificación del servicio, identificando patrones y relaciones entre ellas mediante estadísticas descriptivas y visualizaciones gráficas.

Metodología:

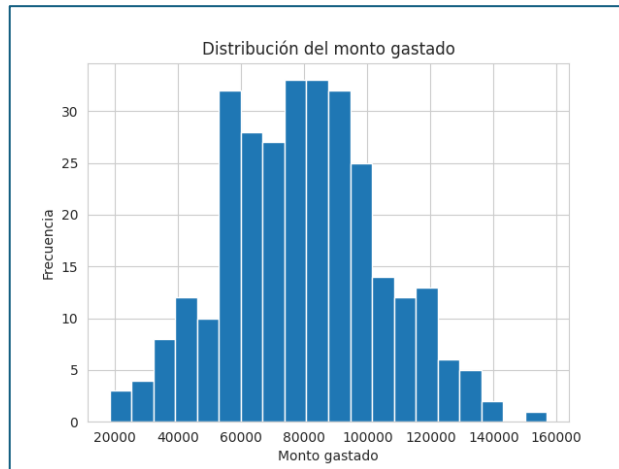
Se siguieron las siguientes etapas:

1. Generación de un dataset simulado de 300 registros de clientes.
2. Exploración inicial del dataset: revisión de tipos de variables, valores faltantes y consistencia.
3. Estadística descriptiva: cálculo de media, mediana, desviación estándar, varianza y cuartiles.
4. Análisis de correlación entre variables mediante matriz de correlación y coeficiente de Pearson.
5. Modelado mediante regresión lineal para estudiar relación entre número de compras y monto gastado.
6. Visualización de datos con Seaborn y Matplotlib: histogramas, boxplots, scatter plots, pairplots, violin plots, jointplots y FacetGrid.

Resultados principales

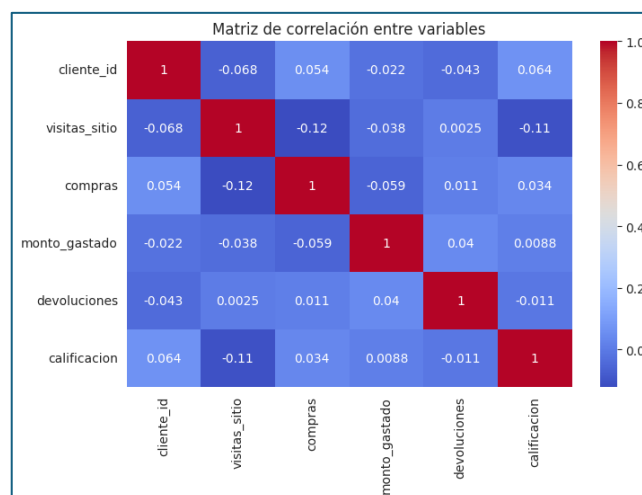
Estadística descriptiva

- Media de visitas al sitio: 15 visitas/clientes.
- Media de compras: ~4 compras.
- Promedio monto gastado: ~\$79.000 CLP.
- Valores atípicos detectados: un cliente con monto cercano a \$160.000.



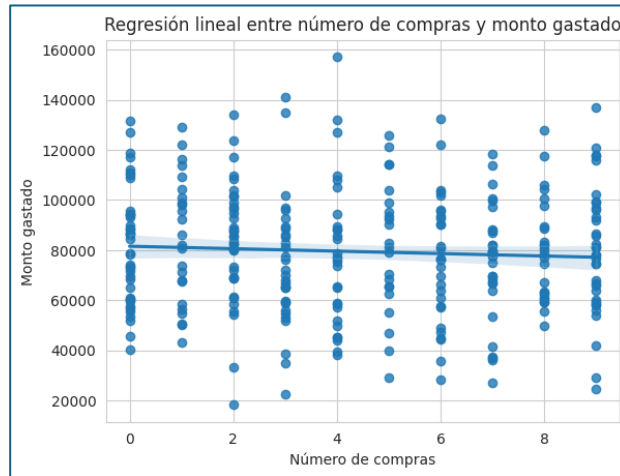
Correlación entre variables

- Las variables presentan correlaciones débiles.
- Correlación número de compras – monto gastado: -0.05 \rightarrow prácticamente nula.



Relación compras vs. Gasto

- Regresión lineal: $R^2 = 0.003 \rightarrow$ modelo explica muy poco la variabilidad.



Visualizaciones adicionales

- Pairplot: muestra distribución y dispersión de todas las variables numéricas.
- Violinplot: distribución del monto gastado.
- Jointplot y FacetGrid: dispersión compras vs. monto según calificación de servicio.

Conclusiones

- El análisis permitió comprender la distribución de las variables y detectar valores atípicos.
- No se identificaron relaciones fuertes entre número de compras y monto gastado.
- Las visualizaciones facilitaron la interpretación de los patrones del dataset.
- La regresión lineal mostró que el número de compras no es un buen predictor del monto gastado.
- En general, los resultados muestran un comportamiento disperso de los clientes y sugieren que otras variables podrían ser más relevantes para explicar el gasto total.