

INFORME TÉCNICO

Proyecto: Análisis Exploratorio de Datos para Decisiones Comerciales

Empresa: Las Varas (Tienda de frutas y verduras con despacho a domicilio)

1. Introducción

El objetivo de este proyecto fue analizar los datos de clientes de la tienda Las Varas, con el propósito de comprender su comportamiento de compra e identificar información relevante que apoye la toma de decisiones comerciales.

Para el desarrollo del análisis se utilizó Python, junto con las librerías Pandas (manipulación de datos), Matplotlib y Seaborn (visualización), y Statsmodels (modelamiento estadístico).

El enfoque aplicado fue un Análisis Exploratorio de Datos (EDA), complementado con análisis de correlación y un modelo de regresión lineal.

2. Descripción del Dataset

El dataset incluye información demográfica y de comportamiento de compra de los clientes.

Variables Numéricas:

- Edad
- Número de visitas
- Número de compras
- Monto total gastado (CLP)
- Productos dañados
- Calificación del servicio

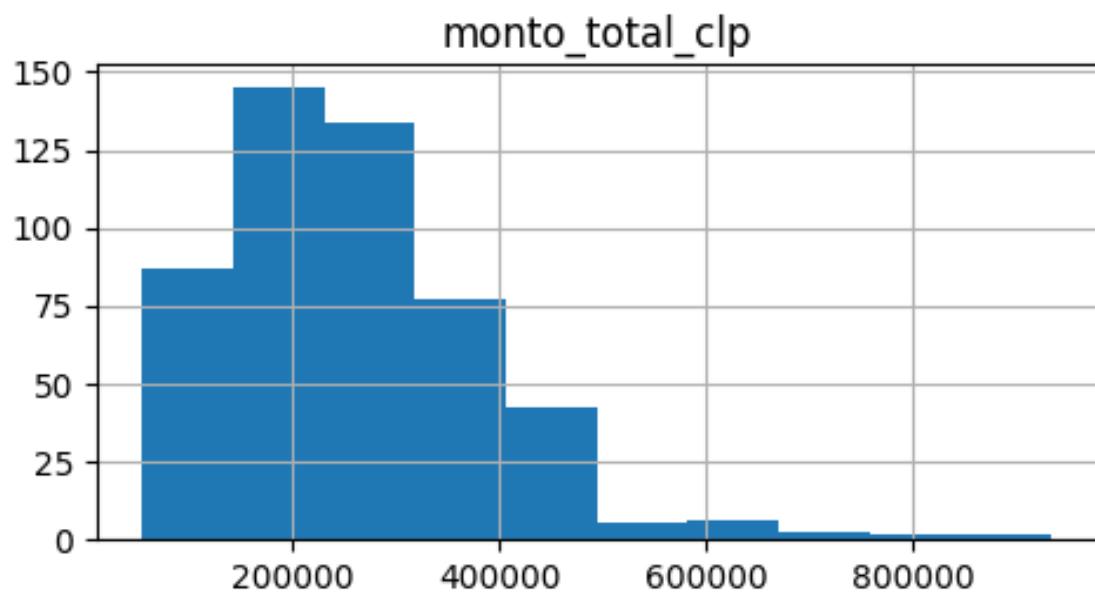
Variables Categóricas:

- Género
- Comuna
- Canal de compra

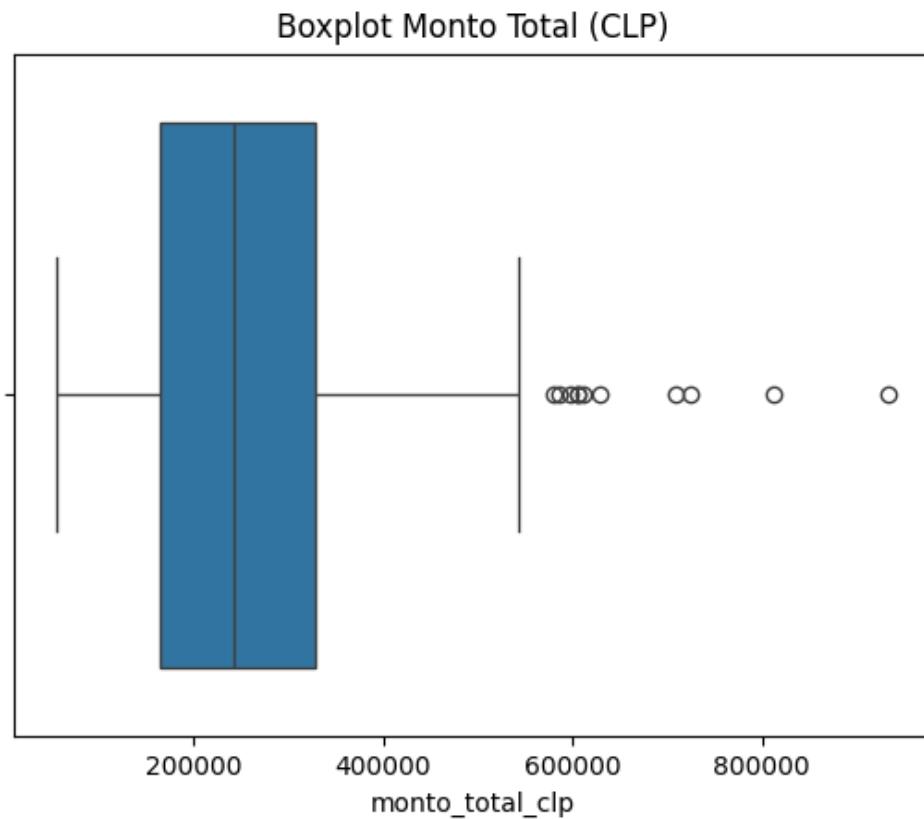
Durante la etapa de limpieza y exploración inicial no se detectaron valores faltantes ni inconsistencias estructurales relevantes, por lo que los datos se consideraron aptos para el análisis.

3. Análisis Estadístico Descriptivo

Se calcularon medidas descriptivas como media, mediana y desviación estándar para comprender el comportamiento general de las variables, especialmente del monto total gastado.



El histograma evidencia que la mayoría de los clientes se concentra en rangos medios de gasto, observándose una distribución asimétrica hacia la derecha debido a la presencia de valores elevados.



El boxplot confirma la existencia de valores atípicos, correspondientes a clientes con un nivel de gasto significativamente superior al promedio.

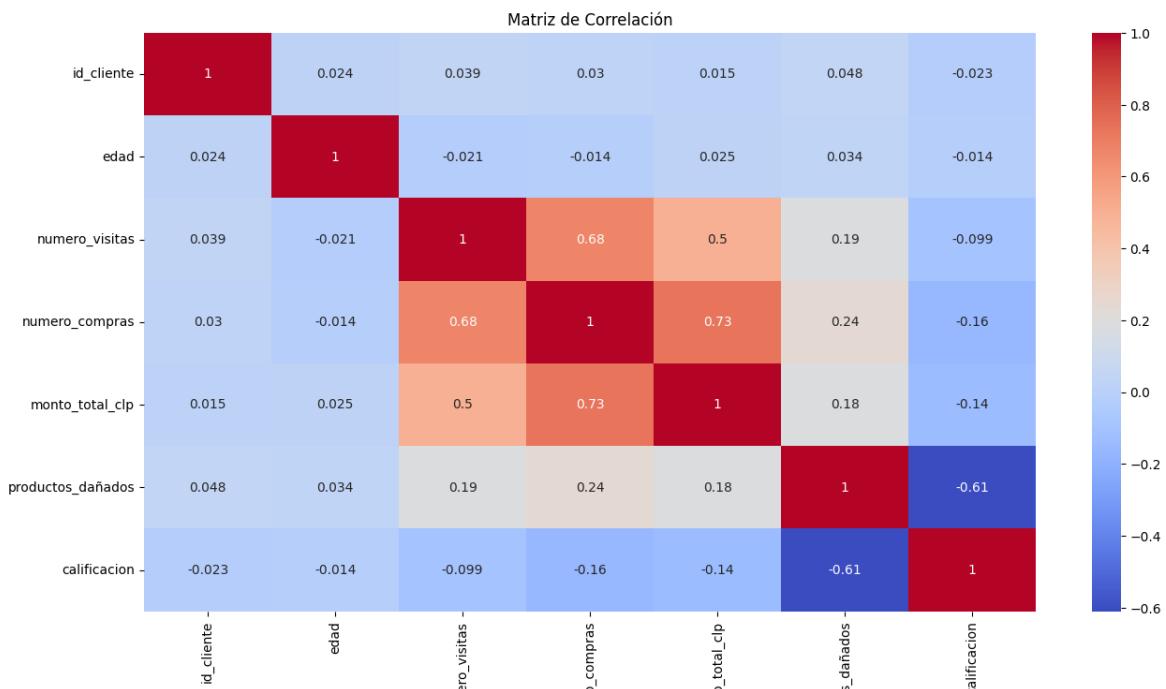
Principales observaciones:

- El monto total gastado presenta alta variabilidad.
- Existe un grupo reducido de clientes que gasta considerablemente más que el promedio (clientes de alto valor o “VIP”).
- La presencia de outliers sugiere segmentación natural dentro de la base de clientes.
- La mayoría de los clientes se concentra en rangos de gasto intermedios.

Estos resultados indican que el comportamiento de compra no es homogéneo y que existen distintos perfiles de consumo dentro de la clientela.

4. Análisis de Correlación

Se evaluó la relación entre variables numéricas mediante el coeficiente de correlación de Pearson



La matriz de correlación muestra una relación positiva moderada entre el número de visitas y el monto total gastado, lo que indica que, a mayor frecuencia de compra, mayor es el gasto acumulado.

Resultados relevantes:

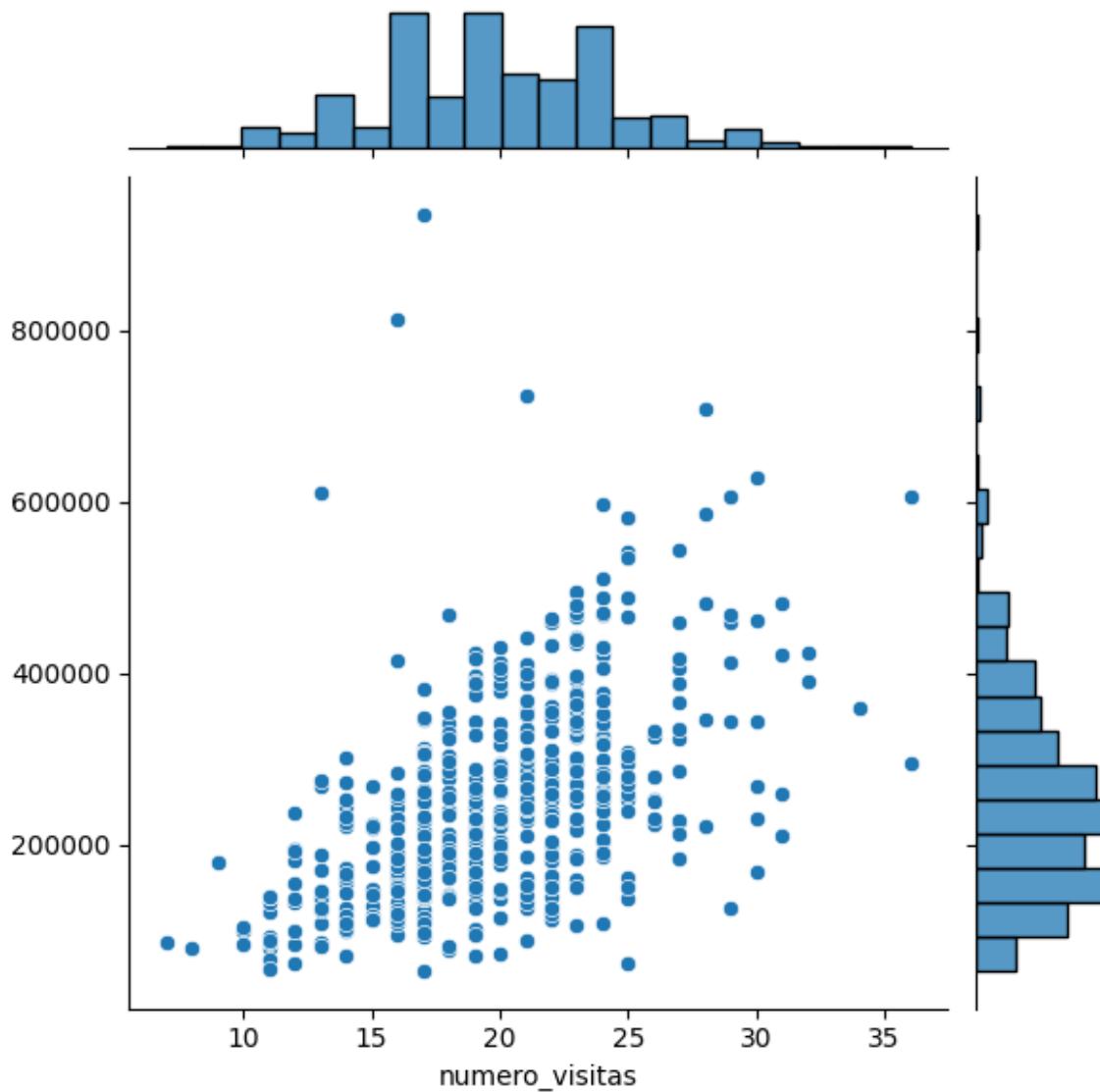
- Existe correlación positiva entre número de visitas y monto total gastado.
- Los productos dañados presentan correlación negativa con la calificación del servicio.
- No se identificaron relaciones inconsistentes o sin fundamento lógico.

En términos generales, los clientes que visitan con mayor frecuencia la tienda tienden a generar mayores ingresos para la empresa.

5. Modelo de Regresión Lineal

Se desarrolló un modelo de regresión lineal donde el monto total gastado se explica en función de:

- Número de visitas
- Productos dañados



El gráfico de dispersión evidencia una tendencia positiva entre número de visitas y gasto total, lo que respalda visualmente los resultados del modelo.

Evaluación Estadística del Modelo

El modelo presentó un R^2 de 0,249, lo que indica que aproximadamente el 24,9% de la variabilidad del monto total gastado es explicada por el número de visitas.

El modelo es estadísticamente significativo en su conjunto ($F = 165,3$; $p < 0,001$), lo que confirma que la variable explicativa utilizada aporta información relevante para predecir el gasto.

El coeficiente asociado al número de visitas fue de aproximadamente \$13.430 CLP, indicando que por cada visita adicional, el gasto total aumenta en promedio en esa cantidad.

Resultados del modelo:

- El número de visitas es un predictor estadísticamente significativo del gasto.
- El modelo logra explicar una proporción relevante del comportamiento de compra.
- Los productos dañados impactan negativamente la experiencia del cliente.

Si bien el modelo explica una parte relevante del comportamiento de compra, existen otros factores no incluidos que podrían influir en el gasto, como promociones, estacionalidad o características socioeconómicas.

Estos resultados confirman que aumentar la frecuencia de compra puede traducirse en mayores ingresos para la empresa.

6. Visualización de Datos

Para complementar el análisis numérico, se utilizaron diversas herramientas de visualización:

- Histogramas para analizar distribuciones.
- Boxplots para detectar valores atípicos.
- Scatterplots para evaluar relaciones entre variables.
- Heatmap para visualizar correlaciones.
- Pairplot para observar múltiples relaciones simultáneamente.

Las visualizaciones permitieron reforzar los hallazgos obtenidos mediante el análisis estadístico.

7. Insights Clave

A partir del análisis realizado, se identificaron los siguientes hallazgos estratégicos:

- La frecuencia de visitas es el principal factor asociado al aumento del gasto total.
- Un grupo reducido de clientes concentra una parte importante de los ingresos.
- La presencia de productos dañados afecta negativamente la percepción del servicio.
- Existen diferencias de comportamiento según el canal de compra utilizado.

Estos insights permiten comprender cómo se generan los ingresos y qué variables influyen en la experiencia del cliente.

8. Recomendaciones Estratégicas

En base a los hallazgos obtenidos, se proponen las siguientes acciones:

- Implementar estrategias de fidelización para aumentar la frecuencia de visitas.
- Identificar y segmentar clientes de alto valor para campañas personalizadas.
- Mejorar procesos logísticos para reducir productos dañados.
- Potenciar el canal de compra que muestre mayor ticket promedio.
- Realizar seguimiento continuo de métricas clave mediante análisis periódicos.

9. Conclusión

El Análisis Exploratorio de Datos permitió transformar información histórica en conocimiento estratégico para la toma de decisiones.

Se identificó que la frecuencia de compra es un factor determinante en la generación de ingresos y que la calidad del servicio impacta directamente en la experiencia del cliente.

En conclusión, el uso de herramientas de análisis de datos facilita decisiones más informadas, estratégicas y basadas en evidencia, fortaleciendo el desempeño comercial de la empresa Las Varas.