

INFORME FINAL EXAMEN MODULO 4 “COMERCIOYA”-

Proyecto: Análisis exploratorio de datos para decisiones comerciales

Evaluación del módulo ANÁLISIS EXPLORATORIO DE DATOS

Empresa: ComercioYA

Unidad solicitante: Dirección de Análisis Estratégico

Responsable: Cristóbal Hernández L.

INTRODUCCIÓN

La empresa ComercioYA, dedicada al comercio electrónico, busca comprender el comportamiento de sus clientes a partir de datos históricos de compras, visitas, devoluciones y reseñas. El objetivo de este análisis es identificar patrones relevantes que permitan optimizar estrategias de marketing, segmentación y toma de decisiones comerciales.

Para ello, se aplicaron técnicas de Análisis Exploratorio de Datos (EDA), estadística descriptiva, análisis de correlación y modelos de regresión lineal utilizando Python (Pandas, Seaborn, Matplotlib y Statsmodels).

LECCIÓN 1 – ANÁLISIS EXPLORATORIO DE DATOS (EDA)

Objetivo

Comprender la estructura del dataset “*ComercioYA_dataset_EDA_v2.csv*”, clasificar tipos de variables, detectar valores faltantes e inconsistencias y documentar hallazgos preliminares.

Análisis Realizado

El dataset contiene 2000 registros de clientes e incluye:

Variables cuantitativas:

- monto_compra

- visitas_mensuales
- numero_compras
- edad
- devoluciones

Variables categóricas:

- región
- método_pago
- categoría_producto
- reseña_score

Se realizó:

- Revisión de tipos de datos
- Identificación de valores faltantes
- Detección de inconsistencias
- Análisis inicial de máximos y mínimos

Hallazgos

- Se detectaron valores faltantes en región, método_pago y reseña_score.
- Existen registros de clientes menores de edad, lo que podría indicar errores de captura.
- Se observan valores máximos elevados en monto_compra, potenciales clientes premium o valores atípicos.
- Existe diversidad regional y de métodos de pago, lo que permitirá segmentaciones estratégicas.

Insight Comercial

La existencia de clientes con montos elevados sugiere la presencia de un segmento de alto valor. Esto representa una oportunidad para estrategias de fidelización diferenciadas.

Recomendaciones

- Depurar registros inconsistentes (edad).
 - Evaluar tratamiento de valores faltantes antes de modelar.
 - Identificar y segmentar clientes premium.
-

LECCIÓN 2 – ESTADÍSTICA DESCRIPTIVA

Objetivo

Aplicar medidas de tendencia central, dispersión y posición para comprender la distribución de las variables numéricas.

Análisis Realizado

Se calcularon:

- Media
- Mediana
- Moda
- Varianza
- Desviación estándar
- Cuartiles y percentiles
- Histogramas
- Boxplots

Hallazgos

- monto_compra presenta media superior a la mediana → distribución asimétrica positiva.
- Se detectan valores atípicos elevados en el boxplot.
- Edad muestra distribución amplia, pero sin outliers extremos.
- Variables de comportamiento (visitas y compras) presentan alta dispersión.

Insight Comercial

La asimetría positiva confirma la existencia de clientes que gastan significativamente más que el promedio. Estos clientes generan un impacto importante en los ingresos totales.

Recomendaciones

- Diseñar estrategias específicas para clientes de alto gasto.
- Evaluar ticket promedio por segmento.
- Implementar monitoreo de variabilidad en compras.

LECCIÓN 3 – ANÁLISIS DE CORRELACIÓN

Objetivo

Detectar y cuantificar relaciones entre variables numéricas.

Análisis Realizado

- Scatterplots
- Matriz de correlación
- Coeficiente de Pearson

Hallazgos

- Correlación positiva entre visitas_mensuales y monto_compra.
- Correlación positiva entre numero_compras y monto_compra.
- Correlación débil entre edad y monto_compra.
- No se detectaron correlaciones negativas relevantes.

Insight Comercial

La frecuencia de interacción con la plataforma es un factor determinante en el gasto. Clientes más activos tienden a generar mayores ingresos.

Recomendaciones

- Implementar campañas para aumentar visitas mensuales.
- Incentivar recompra.
- Priorizar segmentación basada en comportamiento y no en edad.

LECCIÓN 4 – REGRESIÓN LINEAL

Objetivo

Explicar el comportamiento del monto_compra mediante modelos predictivos.

Modelo 1 – Regresión Simple

Variable independiente: visitas_mensuales

Variable dependiente: monto_compra

- Modelo significativo ($p < 0.05$)
- $R^2 = 0.091$

Interpretación:

El modelo explica solo el 9.1% de la variabilidad del gasto, lo que indica capacidad explicativa limitada cuando se usa una sola variable.

Modelo 2 – Regresión Múltiple

Variables independientes:

- visitas_mensuales
- numero_compras
- edad

Resultados:

- $R^2 = 0.332$
- R^2 ajustado = 0.331
- MSE = 37,839
- MAE = 127.79
- Variables significativas: visitas_mensuales y numero_compras
- Edad no significativa

Interpretación:

El modelo múltiple mejora considerablemente la capacidad explicativa. Un 33.2% del comportamiento del gasto es explicado por variables conductuales.

El error promedio de predicción (MAE \approx 128 unidades monetarias) indica precisión moderada.

Insight Comercial

El gasto depende principalmente del comportamiento de compra, no de la edad.

Tabla Comparativa de Resultados

Métrica	Regresión Simple	Regresión Múltiple
R^2	0.091	0.332
R^2 Ajustado	0.091	0.331
MSE	—	37,839
MAE	—	127.79
Significancia Global	Sí ($p < 0.05$)	Sí ($p < 0.05$)

Variables Significativas visitas_mensuales visitas_mensuales, numero_compras

Recomendaciones

- Enfocar estrategias en aumentar frecuencia de visitas.
 - Desarrollar programas de fidelización.
 - Incorporar nuevas variables para mejorar el modelo (promociones, historial de descuentos, categoría de producto).
-

LECCIÓN 5 – ANÁLISIS VISUAL AVANZADO (SEABORN)

Objetivo

Visualizar relaciones complejas y segmentaciones.

Se desarrollaron:

- Pairplot
- Heatmap de correlación
- Violinplot por categoría
- Jointplot
- Segmentación por categoría mediante FacetGrid

Hallazgos

- Segmentos con mayor variabilidad de gasto.
- Confirmación visual de correlaciones.
- Diferencias claras entre categorías de producto.

Insight Comercial

La segmentación visual permite identificar nichos específicos con alto potencial de crecimiento.

Recomendaciones

- Implementar estrategias diferenciadas por categoría.
- Analizar comportamiento por región.

LECCIÓN 6 – VISUALIZACIÓN CON MATPLOTLIB

Objetivo

Crear gráficos personalizados y exportables.

Se desarrollaron:

- Subplots comparativos
- Títulos personalizados
- Etiquetas claras
- Anotaciones relevantes
- Exportación en PNG

Insight

La correcta presentación visual facilita la comprensión ejecutiva y mejora la comunicación de hallazgos.

CONCLUSIONES GENERALES

1. El comportamiento de compra está influenciado principalmente por variables conductuales.
 2. Existe un segmento de clientes premium con alto impacto en ingresos.
 3. La edad no es un predictor relevante del gasto.
 4. El modelo múltiple mejora sustancialmente la capacidad explicativa, aunque aún existen variables no observadas.
-

RECOMENDACIONES ESTRATÉGICAS FINALES

- Diseñar campañas para aumentar visitas recurrentes.
 - Implementar programa de fidelización para clientes frecuentes.
 - Segmentar clientes premium.
 - Analizar variables adicionales para mejorar modelos predictivos.
 - Tomar decisiones basadas en comportamiento, no solo demografía.
-

IMPACTO PARA LA EMPRESA

Este análisis permite a ComercioYA:

- ✓ Optimizar estrategias comerciales
- ✓ Identificar clientes de alto valor
- ✓ Mejorar segmentación
- ✓ Sustentar decisiones con evidencia estadística

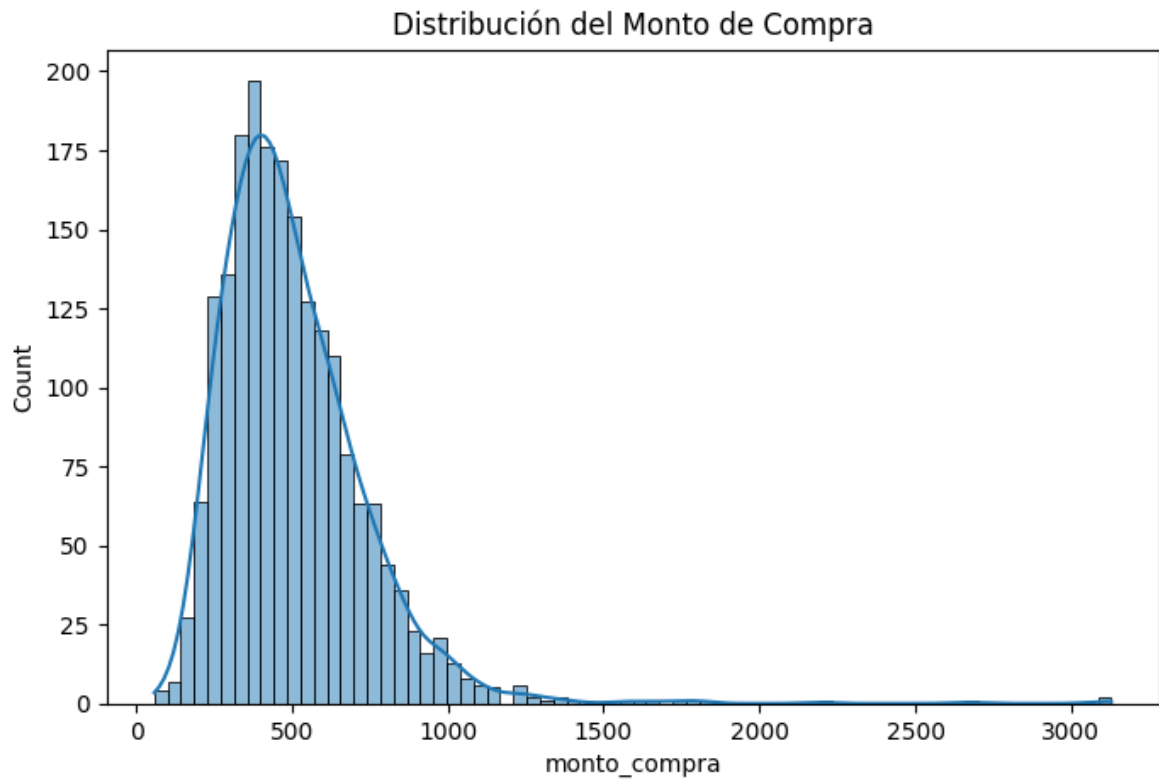
IMÁGENES RELEVANTES:

Lección 1 (df.info)

```
1 import pandas as pd
2
3 df = pd.read_csv("ComercioYA_dataset_EDA_v2.csv")
4 df.head()
5 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   cliente_id            2000 non-null   int64
1   edad                  2000 non-null   int64
2   genero                2000 non-null   object
3   region                1960 non-null   object
4   visitas_mensuales     2000 non-null   int64
5   numero_compras        2000 non-null   int64
6   devoluciones          2000 non-null   int64
7   metodo_pago           1960 non-null   object
8   reseña_score          1960 non-null   float64
9   tipo_cliente          2000 non-null   object
10  monto_compra           2000 non-null   float64
dtypes: float64(2), int64(5), object(4)
memory usage: 172.0+ KB
```

Lección 2 (distribución montocompra)



Lección 3 (Scatterplot Visitas-Monto)

Visitas vs Monto de Compra

