

seccion-3-python-copy1

October 26, 2025

```
[ ]: import pandas as pd
import numpy as np
import datetime as dt
import matplotlib.pyplot as plt
import seaborn as sns
```

0.0.1 1. Carga y exploración básica

- Carga el dataset en un DataFrame.
- Muestra los primeros 5 registros.
- Describe brevemente las columnas y sus tipos.
- ¿Hay valores nulos? ¿Cómo los manejarías?

```
[ ]: ventas = pd.read_csv('ventas_operadores.csv')
ventas.head()
```

```
[ ]:
```

	fecha	id_proveedor	proveedor	prod_id	producto \
0	2025-05-26	3	TIGO	222	9.5GB+MIN.ILIM.+20D
1	2025-05-27	3	TIGO	109	10 MIN LDI + 2 DIAS
2	2025-05-28	3	TIGO	575	5 GB+MIN.ILIM+15D
3	2025-05-29	3	TIGO	101	5 MIN LDI + 1DIA
4	2025-05-30	3	TIGO	485	13 GB+MINILI+RDS+30D

	proveedor_unificado	valor_ventas	cantidad_tx
0	TIGO	5670000	252
1	TIGO	151200	28
2	TIGO	4284000	306
3	TIGO	58900	19
4	TIGO	15307500	471

```
[ ]: ventas.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 84 entries, 0 to 83
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   fecha                 84 non-null    object
```

```

1   id_proveedor      84 non-null    int64
2   proveedor         84 non-null    object
3   prod_id           84 non-null    int64
4   producto          84 non-null    object
5   proveedor_unificado 84 non-null    object
6   valor_ventas      84 non-null    int64
7   cantidad_tx       84 non-null    int64
dtypes: int64(4), object(4)
memory usage: 5.4+ KB

```

```
[ ]: ventas.isnull().sum()
```

```

[ ]: fecha              0
    id_proveedor        0
    proveedor           0
    prod_id             0
    producto            0
    proveedor_unificado 0
    valor_ventas        0
    cantidad_tx         0
dtype: int64

```

Si tuviera datos nulos en este caso específico, eliminaría el registro, ya que esto podría alterar los análisis y los cálculos a la hora de sacar porcentajes, promedios, etc. `ventas = ventas.dropna()`

0.0.2 2. Limpieza de datos

- Corrige los tipos de datos si es necesario (por ejemplo, fechas).
- Elimina columnas irrelevantes si aplica.
- Crea una columna nueva calculada, por ejemplo: `Ganancia = Ingreso - Costo`.

```
[ ]: ventas['id_proveedor'] = ventas['id_proveedor'].astype(str)
    ventas['prod_id'] = ventas['prod_id'].astype(str)
    ventas['valor_ventas'] = ventas['valor_ventas'].astype(float)
```

```
[ ]: ventas['fecha'] = pd.to_datetime(ventas['fecha'], errors="coerce")
```

```
[ ]: ventas['fecha'].info()
```

```

<class 'pandas.core.series.Series'>
RangeIndex: 84 entries, 0 to 83
Series name: fecha
Non-Null Count  Dtype
-----
84 non-null    datetime64[ns]
dtypes: datetime64[ns](1)
memory usage: 804.0 bytes

```

Todas las columnas están relacionadas así que no serían necesario eliminar ninguna.

Si fuera necesario eliminar una columna: `ventas = ventas.drop(columns=['Columna_a_eliminar'])`

```
[ ]: ventas['valor_unitario'] = (ventas['valor_ventas'] / ventas['cantidad_tx'])
```

```
[ ]: ventas.head(1)
```

```
[ ]:      fecha id_proveedor proveedor prod_id      producto \
0 2025-05-26          3      TIGO      222  9.5GB+MIN.ILIM.+20D

      proveedor_unificado  valor_ventas  cantidad_tx  valor_unitario
0                TIGO      5670000.0          252          22500.0
```

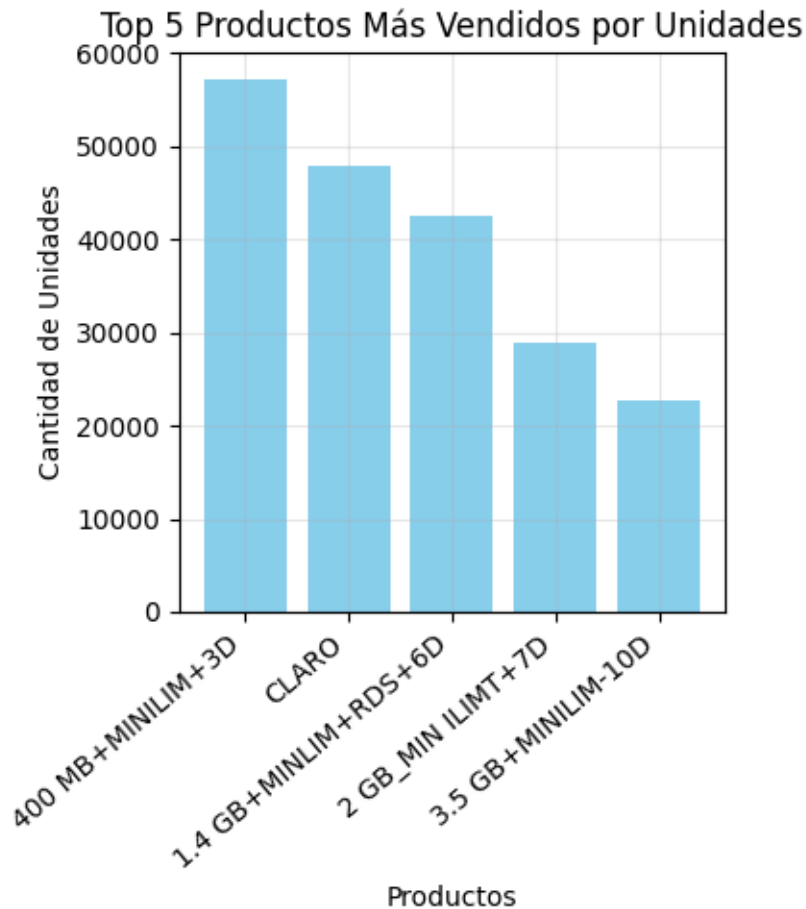
0.0.3 3. Análisis exploratorio

- ¿Cuáles son los 5 productos más vendidos por unidades?

```
[ ]: top5_productos = (ventas.groupby('producto')['cantidad_tx'].sum().
    ↪sort_values(ascending=False).head(5).reset_index())
print(top5_productos)
```

```
      producto  cantidad_tx
0    400 MB+MINILIM+3D      57218
1             CLARO      47968
2  1.4 GB+MINLIM+RDS+6D      42523
3    2 GB_MIN ILIMT+7D      28985
4    3.5 GB+MINILIM-10D      22623
```

```
[ ]: plt.figure(figsize=(4, 5))
plt.bar(top5_productos['producto'], top5_productos['cantidad_tx'],
    ↪color='skyblue')
plt.title('Top 5 Productos Más Vendidos por Unidades')
plt.xlabel('Productos')
plt.ylabel('Cantidad de Unidades')
plt.grid(True, alpha=0.3)
plt.xticks(rotation=40, ha='right')
plt.tight_layout()
```



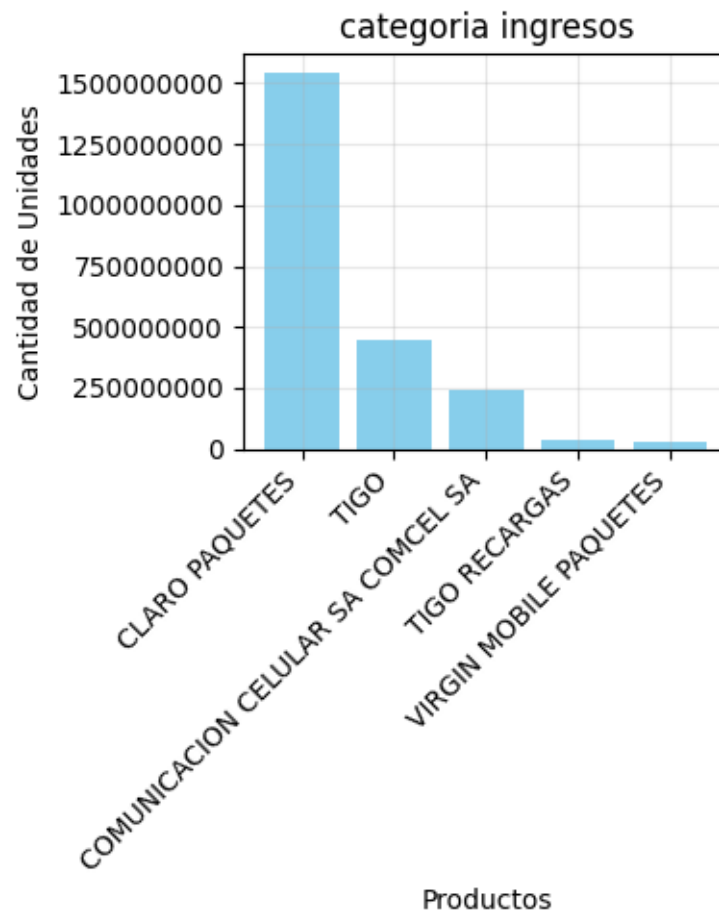
- ¿Cuál es la categoría que más ingresos genera?

```
[ ]: categoria_top = (ventas.groupby('proveedor')['valor_ventas'].sum().
    ↪sort_values(ascending=False).head(1).reset_index())
print(categoria_top)
```

```
   proveedor  valor_ventas
0  CLARO PAQUETES  1.542106e+09
```

```
[ ]: categoria_ingresos = (ventas.groupby('proveedor')['valor_ventas'].sum().
    ↪sort_values(ascending=False).head().reset_index())
plt.figure(figsize=(4, 5))
plt.bar(categoria_ingresos['proveedor'], categoria_ingresos['valor_ventas'],
    ↪color='skyblue')
plt.title('categoria ingresos')
plt.xlabel('Productos')
plt.ylabel('Cantidad de Unidades')
plt.grid(True, alpha=0.3)
```

```
plt.ticklabel_format(style='plain', axis='y')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
```



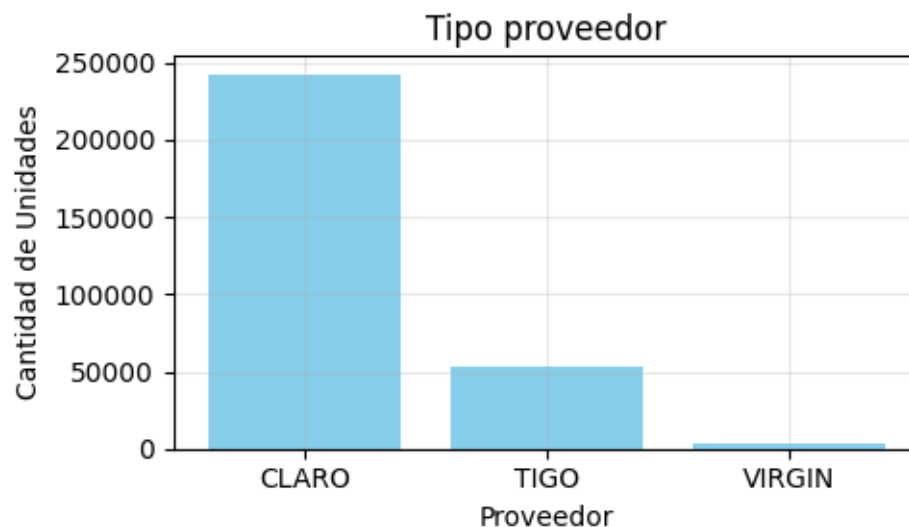
La categoría que mas ingresos genera es **CLARO PAQUETES** con \$1,542,106,250 en ventas

- ¿Hay diferencias significativas de ventas por región o canal?

```
[ ]: tipo_proveedor = (ventas.groupby('proveedor_unificado')['cantidad_tx'].sum()
    ↪sort_values(ascending=False).head().reset_index())
total_cantidad_tx = tipo_proveedor['cantidad_tx'].sum()
tipo_proveedor['porcentaje_cantidad_tx'] = (tipo_proveedor['cantidad_tx'] /
    ↪total_cantidad_tx) * 100
tipo_proveedor['porcentaje_cantidad_tx'] =
    ↪tipo_proveedor['porcentaje_cantidad_tx'].round(2)
print(tipo_proveedor)
```

	proveedor_unificado	cantidad_tx	porcentaje_cantidad_tx
0	CLARO	242152	81.01
1	TIGO	53631	17.94
2	VIRGIN	3135	1.05

```
[ ]: plt.figure(figsize=(5, 3))
plt.bar(tipo_proveedor['proveedor_unificado'], tipo_proveedor['cantidad_tx'],
        color='skyblue')
plt.title('Tipo proveedor')
plt.xlabel('Proveedor')
plt.ylabel('Cantidad de Unidades')
plt.grid(True, alpha=0.3)
plt.ticklabel_format(style='plain', axis='y')
plt.xticks(range(len(tipo_proveedor)))
plt.tight_layout()
```



CLARO domina el mercado con (81%), mientras que **TIGO** (18%) y **VIRGIN** (1%) tienen ventas considerablemente menores.

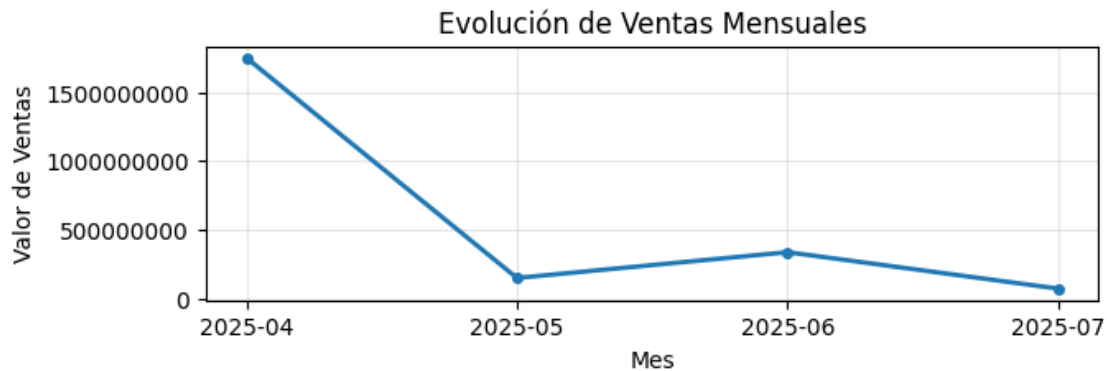
0.0.4 4. Visualización

- Crea una visualización que muestre la evolución de las ventas mensuales.

```
[ ]: ventas_mensuales = ventas.groupby(ventas['fecha'].dt.
        to_period('M'))[['valor_ventas', 'cantidad_tx']].sum().reset_index()

plt.figure(figsize=(7, 2))
plt.plot(ventas_mensuales['fecha'].astype(str),
        ventas_mensuales['valor_ventas'], marker='o', linewidth=2, markersize=4)
```

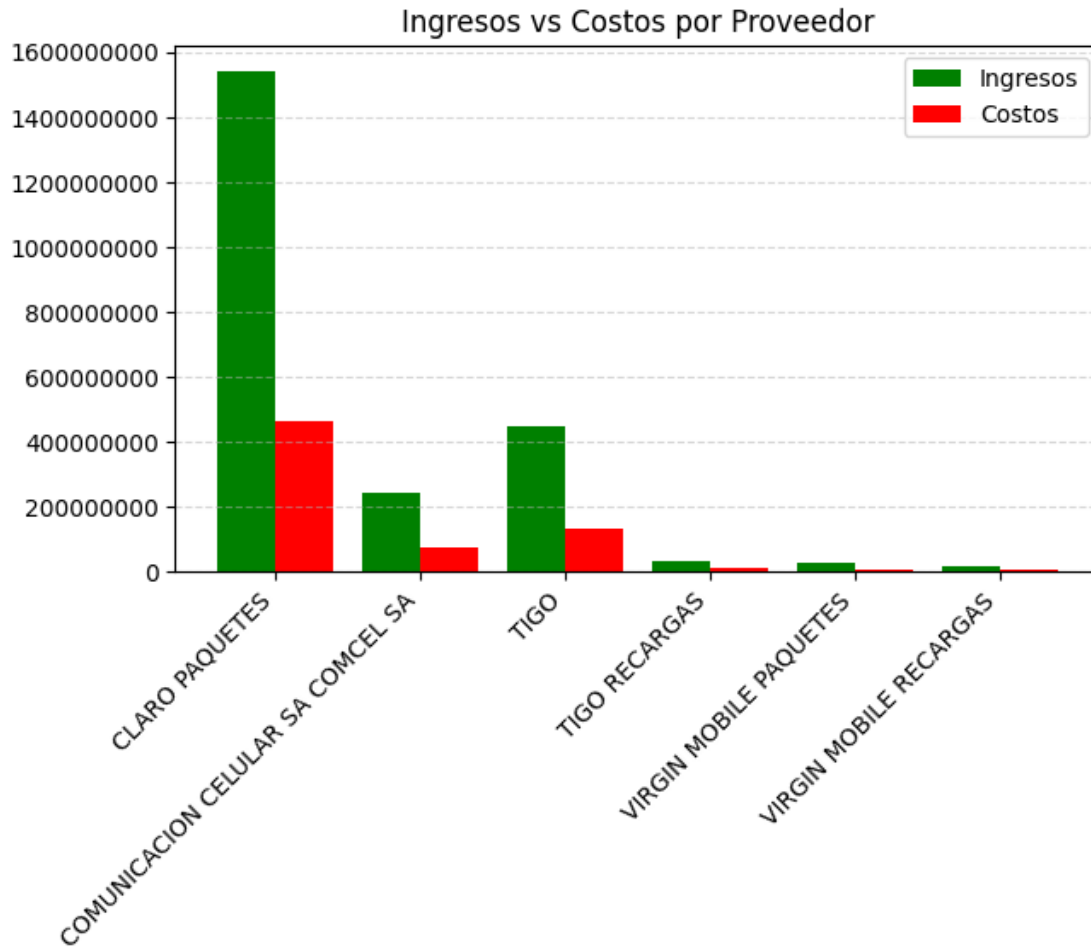
```
plt.title('Evolución de Ventas Mensuales')
plt.xlabel('Mes')
plt.ylabel('Valor de Ventas')
plt.ticklabel_format(style='plain', axis='y')
plt.xticks(rotation=0)
plt.grid(True, alpha=0.3)
```



- Muestra un gráfico comparativo entre ingresos y costos por categoría.

```
[ ]: ventas['costo'] = ventas['valor_ventas'] * 0.3 ## Costo de 30% del valor de las
    ventas
prov = ventas.groupby('proveedor')[['valor_ventas', 'costo']].sum()

plt.figure(figsize=(7, 4))
x = range(len(prov))
plt.bar([i-0.2 for i in x], prov['valor_ventas'], 0.4, label='Ingresos',
    color='green')
plt.bar([i+0.2 for i in x], prov['costo'], 0.4, label='Costos', color='red')
plt.title('Ingresos vs Costos por Proveedor')
plt.grid(axis='y', linestyle='--', alpha=0.5)
plt.ticklabel_format(style='plain', axis='y')
plt.xticks(x, prov.index, rotation=45, ha='right')
plt.legend()
plt.show()
```



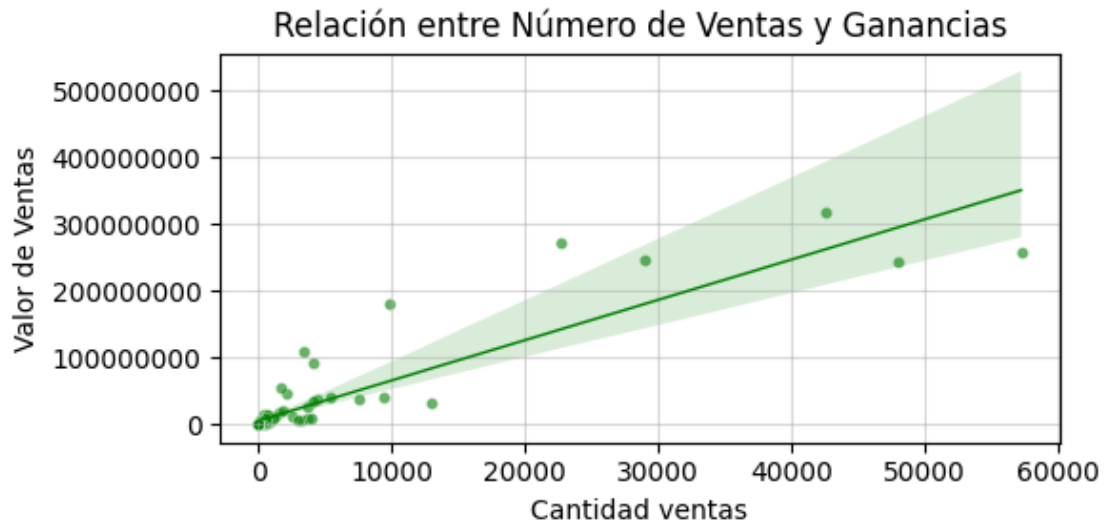
0.0.5 Debido a que no se tiene la columna de costos tomé un supuesto del 30% por costo de venta para calcularla.

- Usa un gráfico de dispersión (scatter) para mostrar la relación entre el número de ventas y las ganancias.

```
[ ]: plt.figure(figsize=(6, 3))
sns.scatterplot(data=ventas, x='cantidad_tx', y='valor_ventas', alpha=0.6,
               ↪s=19, color='green')
sns.regplot(data=ventas, x='cantidad_tx', y='valor_ventas', scatter=False,
            ↪color='green', line_kws={'linestyle': '-', 'linewidth': 1})
plt.title('Relación entre Número de Ventas y Ganancias')
plt.xlabel('Cantidad ventas')
plt.ylabel('Valor de Ventas')
plt.ticklabel_format(style='plain', axis='y')
plt.grid(True, alpha=0.5)
plt.tight_layout()
```



```
plt.show()
```



0.0.6 5. Insight y comunicación

- En máximo 5 líneas, escribe dos hallazgos interesantes del análisis.
- Propón una acción que el negocio podría tomar basada en esos hallazgos.

Hallazgo 1: CLARO domina completamente el mercado de telecomunicaciones con 242,152 unidades vendidas, representando más del 80% del total de transacciones, mientras que TIGO (53,631) y VIRGIN (3,135) tienen participaciones significativamente menores.

Hallazgo 2: Existe una correlación positiva fuerte entre la cantidad de ventas y el valor de ventas, donde el producto “400 MB+MINILIM+3D” lidera tanto en unidades vendidas (57,218) como en generación de ingresos, sugiriendo que los productos de mayor volumen también generan mayores ingresos.

Acción recomendada: El negocio debería enfocar sus esfuerzos comerciales para productos similares al “400 MB+MINILIM+3D”. Teniendo en cuenta que el volumen es el factor principal del negocio, tomar un enfoque en aumentar la frecuencia de recompra y recurrencia de clientes. Debido a la fuerte caída en ventas después del primer mes se debería diseñar estrategias para convertir ventas puntuales en ventas recurrentes.

[]: