

Taller 2 analítica de datos

Daniel Andrés Pinzón Jay

Id: 819793

Programa: Ingeniería de sistemas

Analítica de datos

**Universidad cooperativa de Colombia Sede:
Bucaramanga**

2024

Solución de actividades

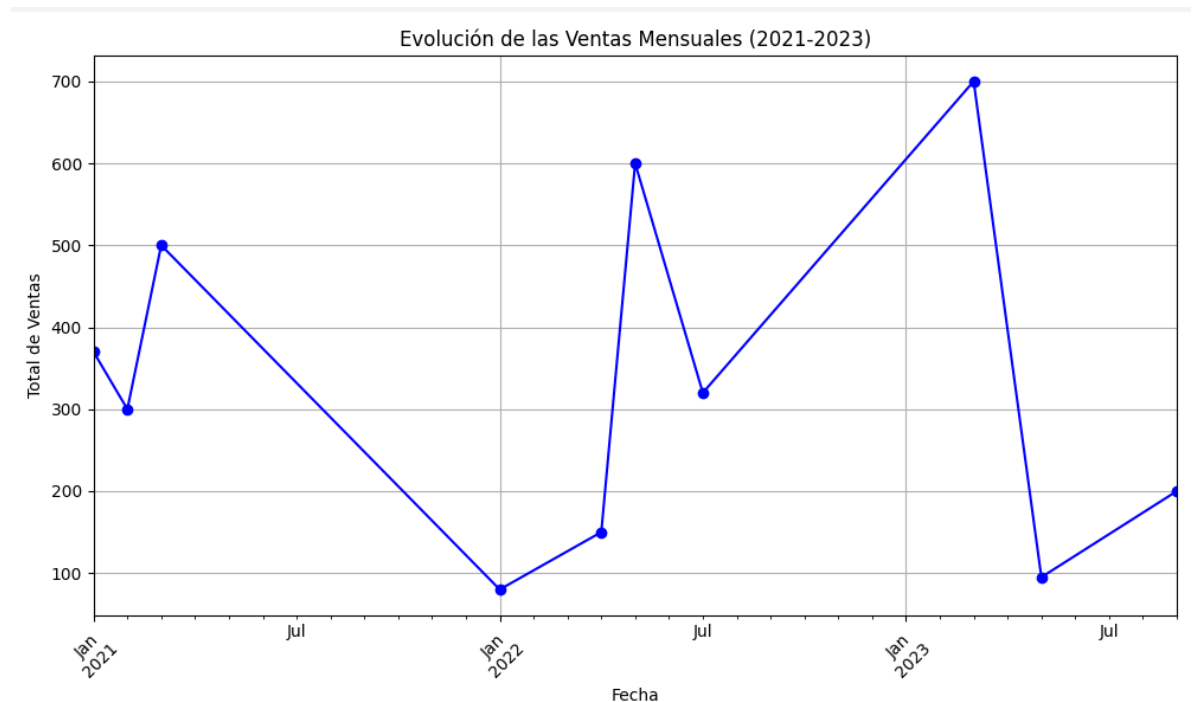
1. Limpieza de datos

```
import pandas as pd
import matplotlib.pyplot as plt

data = {
    'Date': ['2021-01-05', '2021-01-12', '2021-02-20', '2021-03-15', '2022-01-08',
            '2022-04-10', '2022-05-22', '2022-07-19', '2023-03-15', '2023-05-27', '2023-09-12'],
    'Store ID': [1, 1, 2, 1, 2, 1, 2, 2, 1, 2, 1],
    'Product ID': [1001, 1002, 1003, 1004, 1005, 1006, 1002, 1007, 1008, 1009, 1010],
    'Category': ['Electrónica', 'Ropa', 'Hogar', 'Electrónica', 'Alimentos',
                'Ropa', 'Electrónica', 'Hogar', 'Electrónica', 'Alimentos', 'Ropa'],
    'Customer ID': ['C001', 'C002', 'C001', 'C003', 'C002', 'C003', 'C001', 'C004', 'C002',
                   'C001', 'C004'],
    'Sales Amount': [250.00, 120.00, 300.00, 500.00, 80.00, 150.00, 600.00, 320.00,
                    700.00, 95.00, 200.00],
    'Units Sold': [1, 2, 3, 1, 4, 2, 1, 2, 1, 5, 3]
}

#convercion del formato fecha
df = pd.DataFrame(data)
df['Date'] = pd.to_datetime(df['Date'])
```

2. Análisis de tendencia



Meses con mayor número de ventas:

1. Marzo 2023 - 700.00
2. Mayo 2022 - 600.00
3. Marzo 2021 - 500.00

Explicaciones:

Promociones: Marzo es un mes donde muchas empresas lanzan promociones para el inicio de primavera o liquidación de productos de temporada de invierno.

Estacionalidad: En mayo puede ver un aumento de compras ya que coincide con el día de la madre lo que impulsa las compras en muchos países.

Meses con menor número de ventas:

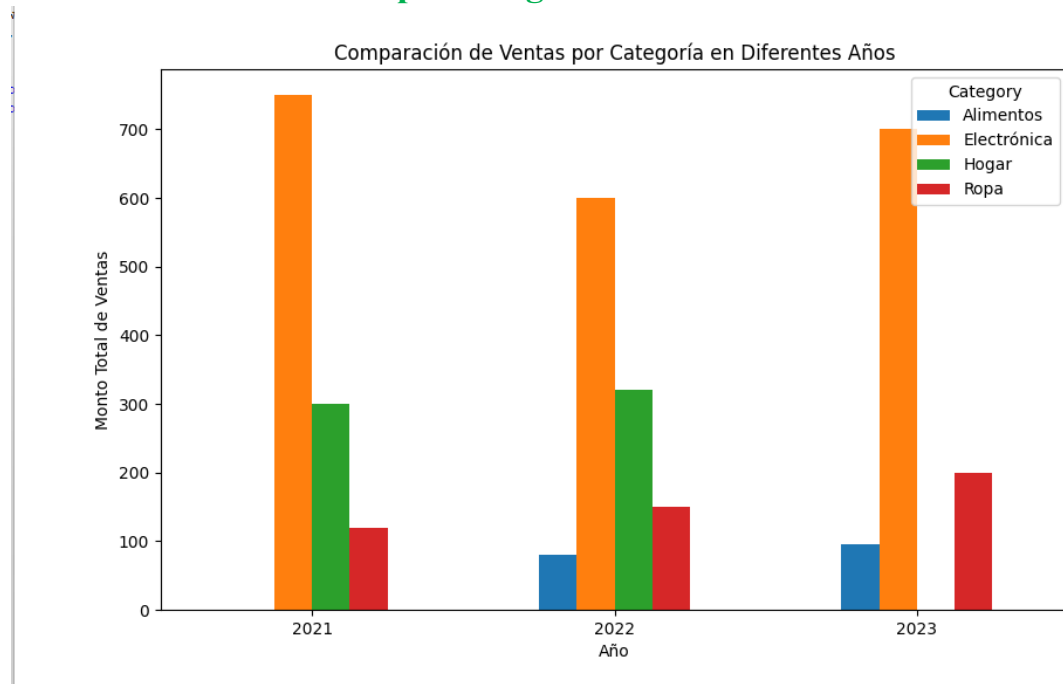
1. Enero 2022 - 80.00
2. Mayo 2023 - 95.00
3. Enero 2021 - 120.00

Explicaciones:

Época post-Diciembre: Enero es por naturaleza un mes donde el consumidor invierte menos después de los gastos de las fiestas de diciembre. Por lo que no es de extrañar que este en el top 2 veces.

Falta de promoción: Mayo suele ser un mes de alta ventas. Incluso si lo comparamos con mayo del 2022 hubo un bajón muy grande de ventas lo que supone que hubo una falta de promoción o de publicidad.

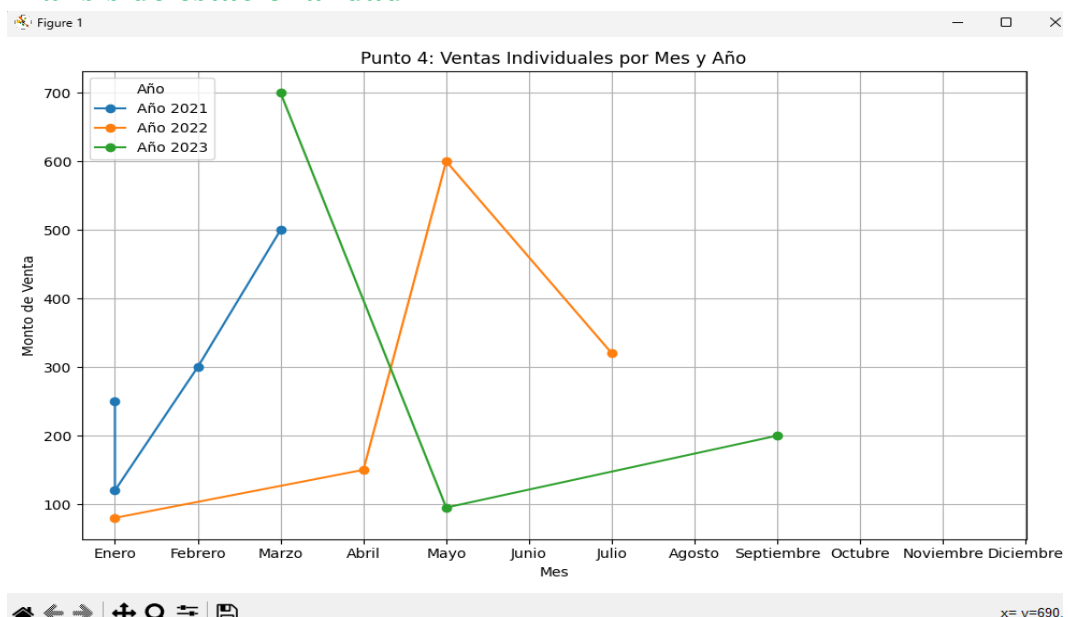
3. Análisis de distribución por categoría



Como podemos ver en los 3 años (2021, 2022 y 2023) la categoría más vendida es la **Electrónica** mientras que la categoría con menos ventas es la de **alimentos**, ya que se tuvo 0 ventas en 2021 y ventas muy mínimas en los otros dos años.

En cuanto a el **hogar** se mantuvo estable en el 2021 y 2022 pero sus ventas fueron nulas en el 2023. La única categoría que se ha mantenido estable a parte de la **electrónica** es la **ropa** ya que, aunque no representa ventas muy grandes siempre se ha mantenido estable en los 3 años.

4. Análisis de estacionalidad



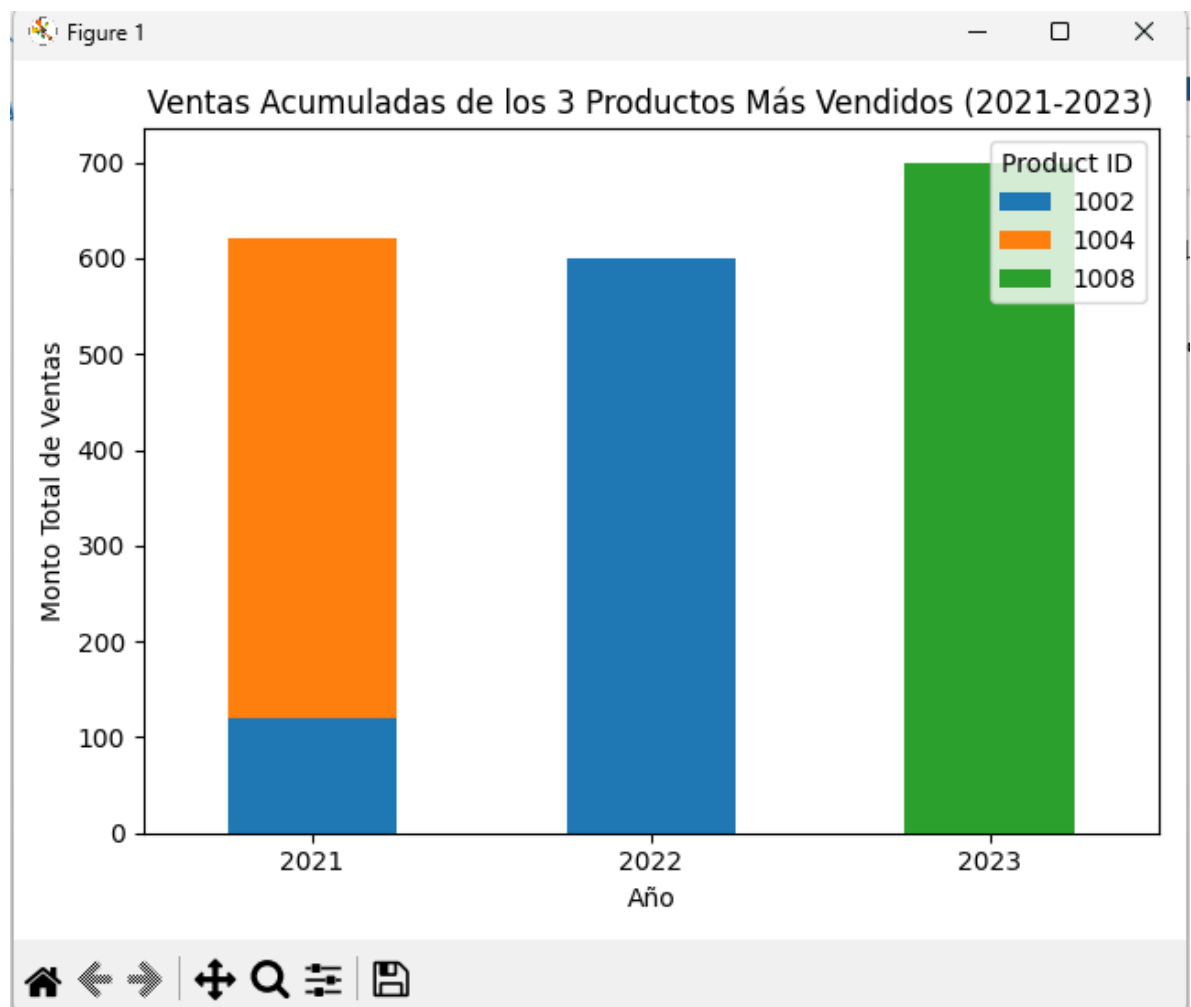
El mes que ha generado más ventas es **marzo**, seguido de **mayo** y otro mes donde se generan varias ventas es **enero**, aunque por montos menores.

Si tuviéramos que ver el año con más meses con compras sería el 2022 donde se realizaron compras en 4 meses distintos

5. Análisis de cliente y producto

Top 3 productos más vendidos:

	Product ID	Total Sales
0	1002	720.0
1	1008	700.0
2	1004	500.0

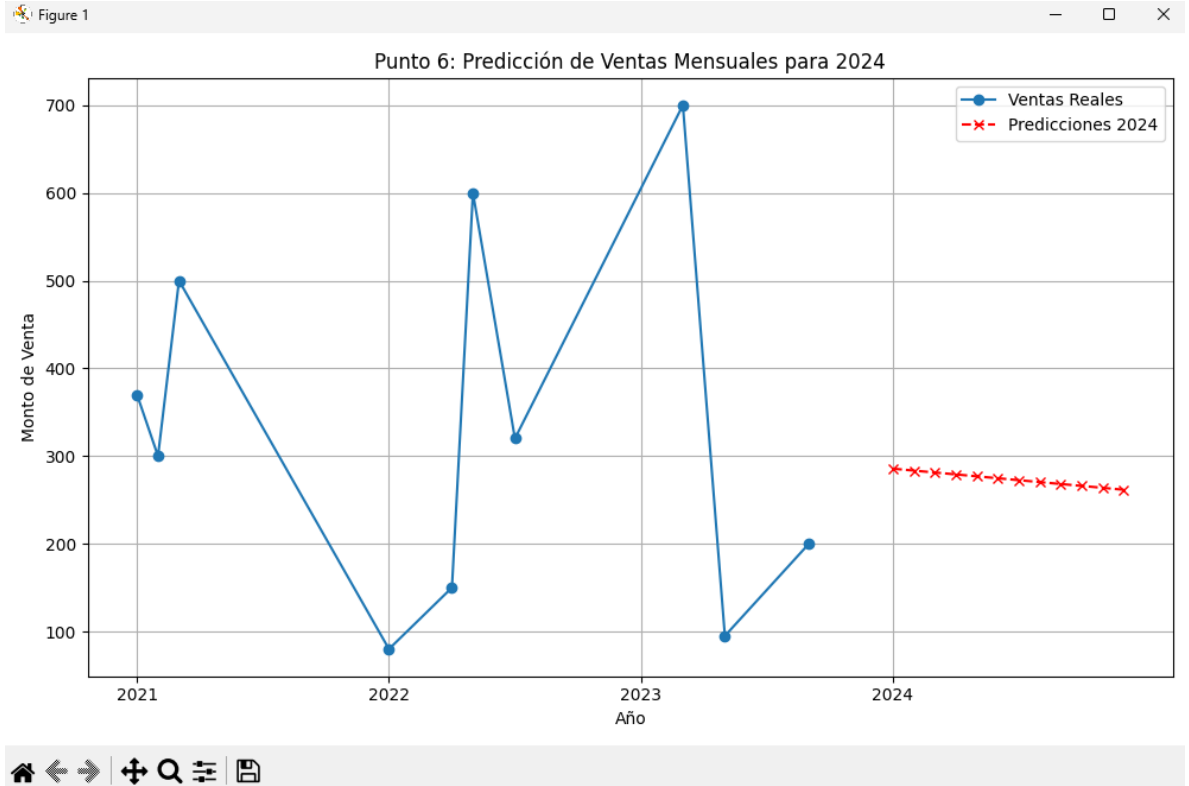


Los productos mas vendidos son el **1002** en primer lugar seguido por el **1008** y **1004**. Si realizamos un grafico de las ventas acumuladas de estos 3 productos podemos observar que el 1002 ha sido comprado en el 2021 y el 2022. Mientras que el 1008 solo en el 2023 y el 1004 solo en el 2021.

Clientes más recurrentes:		
	Customer ID	Frequency
0	C001	4
1	C002	3
2	C003	2
3	C004	2

En cuanto a los clientes mas recurrentes podemos ver que el ganador es **C001** con 4. Seguido de **C002** con 3 y en tercer lugar un empate entre **C003** y **C004** con 2 compras.

6. Predicción de ventas futuras



Como podemos ver se espera una compras en el 2024 decrecientes y bajas a los 300\$.