

Práctica de Pandas - Análisis de Datos

October 21, 2025

Este colab fue desarrollado por Arnold Charry Armero.

1 Práctica de Pandas - Análisis de Datos

Primero, importamos las librerías,

```
[1]: # Importamos las librerías
import numpy as np
import pandas as pd
```

```
[2]: from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
[28]: df = pd.read_csv('/content/drive/MyDrive/Diplomados/Diplomado de Programación\u2022
    ↴en Python/Módulo 4 - Librerías y IA/Evidencia de Aprendizaje/Actividad 1/
    ↴Ventas..csv')
```

Se revisa el DataFrame,

```
[29]: df.head(10)
```

```
[29]:   ID_Venta      Fecha   Producto  Cantidad  Precio_Unitario  Vendedor \
0          1  2024-01-05  Producto A        10           100.0  Juan Pérez
1          2  2024-01-06  Producto B         5           200.0  María Gómez
2          3  2024-01-07  Producto A         2           100.0  Pedro Díaz
3          4  2024-01-08  Producto C         7           150.0  Ana Ruiz
4          5  2024-01-09  Producto B         1           200.0  Juan Pérez
5          6  2024-01-10  Producto A         3           100.0  María Gómez
6          7  2024-01-11  Producto C         8           150.0  Ana Ruiz
7          8  2024-01-12  Producto B         4           200.0  Pedro Díaz
8          9  2024-01-13  Producto A         6           100.0  Juan Pérez
9         10  2024-01-14  Producto C        10           150.0  María Gómez

          Ciudad
0    Bogotá
1     Cali
```

```
2 Medellín
3 Bogotá
4 Cali
5 Medellín
6 Cali
7 Bogotá
8 Cali
9 Medellín
```

Se verifica el tema de los faltantes,

```
[30]: df.isnull().sum()
```

```
[30]: ID_Venta      0
Fecha          0
Producto       0
Cantidad        0
Precio_Unitario 0
Vendedor        0
Ciudad          0
dtype: int64
```

```
[31]: # En caso de haber faltantes, así se llenarían con pandas
```

```
# Promedio
df['Cantidad'] = df['Cantidad'].fillna(df['Cantidad'].median())

# Mediana
df['Precio_Unitario'] = df['Precio_Unitario'].fillna(df['Precio_Unitario'].
    ↪mean())

# Moda
df['Producto'] = df['Producto'] = df['Producto'].fillna(df['Producto'].
    ↪mode()[0])
```

```
[32]: df.head(10)
```

```
[32]:   ID_Venta      Fecha     Producto  Cantidad  Precio_Unitario    Vendedor \
0           1  2024-01-05  Producto A       10        100.0  Juan Pérez
1           2  2024-01-06  Producto B        5        200.0  María Gómez
2           3  2024-01-07  Producto A       2        100.0  Pedro Díaz
3           4  2024-01-08  Producto C       7        150.0  Ana Ruiz
4           5  2024-01-09  Producto B        1        200.0  Juan Pérez
5           6  2024-01-10  Producto A       3        100.0  María Gómez
6           7  2024-01-11  Producto C       8        150.0  Ana Ruiz
7           8  2024-01-12  Producto B       4        200.0  Pedro Díaz
8           9  2024-01-13  Producto A       6        100.0  Juan Pérez
```

```
9      10 2024-01-14 Producto C      10      150.0 María Gómez  
  
      Ciudad  
0    Bogotá  
1    Cali  
2  Medellín  
3    Bogotá  
4    Cali  
5  Medellín  
6    Cali  
7    Bogotá  
8    Cali  
9  Medellín
```

```
[33]: for i in range(df.shape[0]):  
    df.loc[i, 'Total_Venta'] = df.loc[i, 'Precio_Unitario'] * df.loc[i, 'Cantidad']
```

```
[34]: df.head(10)
```

```
[34]:   ID_Venta      Fecha  Producto  Cantidad  Precio_Unitario  Vendedor \\\n0        1  2024-01-05  Producto A       10          100.0  Juan Pérez  
1        2  2024-01-06  Producto B        5          200.0  María Gómez  
2        3  2024-01-07  Producto A       2          100.0  Pedro Díaz  
3        4  2024-01-08  Producto C       7          150.0  Ana Ruiz  
4        5  2024-01-09  Producto B        1          200.0  Juan Pérez  
5        6  2024-01-10  Producto A       3          100.0  María Gómez  
6        7  2024-01-11  Producto C       8          150.0  Ana Ruiz  
7        8  2024-01-12  Producto B       4          200.0  Pedro Díaz  
8        9  2024-01-13  Producto A       6          100.0  Juan Pérez  
9       10  2024-01-14  Producto C      10          150.0  María Gómez
```

```
      Ciudad  Total_Venta  
0    Bogotá     1000.0  
1    Cali      1000.0  
2  Medellín     200.0  
3    Bogotá     1050.0  
4    Cali      200.0  
5  Medellín     300.0  
6    Cali      1200.0  
7    Bogotá     800.0  
8    Cali      600.0  
9  Medellín     1500.0
```

```
[35]: df_vendedores = df.groupby('Vendedor')['Total_Venta'].sum()  
display(df_vendedores)
```

```
Vendedor
Ana Ruiz      2250.0
Juan Pérez    1800.0
María Gómez   2800.0
Pedro Díaz    1000.0
Name: Total_Venta, dtype: float64
```

La mejor vendedora fue María Gómez con 2800 ventas.

```
[36]: df_Bogotá = df[df['Ciudad'] == 'Bogotá']
df_Bogotá.head()
```

```
[36]:   ID_Venta      Fecha  Producto  Cantidad  Precio_Unitario  Vendedor \
0           1  2024-01-05  Producto A        10          100.0  Juan Pérez
3           4  2024-01-08  Producto C         7          150.0  Ana Ruiz
7           8  2024-01-12  Producto B         4          200.0  Pedro Díaz

   Ciudad  Total_Venta
0  Bogotá      1000.0
3  Bogotá      1050.0
7  Bogotá       800.0
```

El total de ventas de Producto A fueron,

```
[37]: df_A = df[df['Producto'] == 'Producto A'].groupby('Producto')['Total_Venta'].
      ↪sum()
display(df_A)
```

```
Producto
Producto A      2100.0
Name: Total_Venta, dtype: float64
```

Si queremos ver las ventas totales de cada producto,

```
[38]: df_Producto = df.groupby('Producto')['Total_Venta'].sum()
display(df_Producto)
```

```
Producto
Producto A      2100.0
Producto B      2000.0
Producto C      3750.0
Name: Total_Venta, dtype: float64
```

Ahora se conocerán las ventas por ciudad,

```
[39]: df_Ciudad = df.groupby('Ciudad')['Total_Venta'].sum()
display(df_Ciudad)
```

```
Ciudad
Bogotá      2850.0
```

```
Cali      3000.0  
Medellín  2000.0  
Name: Total_Venta, dtype: float64
```

Ahora las ventas por vendedor y por producto,

```
[40]: df_vendedores_producto = df.groupby(['Vendedor', 'Producto'])['Total_Venta'].  
      ↪sum()  
      display(df_vendedores_producto)
```

```
Vendedor  Producto  
Ana Ruiz   Producto C    2250.0  
Juan Pérez Producto A    1600.0  
              Producto B    200.0  
María Gómez Producto A    300.0  
              Producto B    1000.0  
              Producto C    1500.0  
Pedro Díaz  Producto A    200.0  
              Producto B    800.0  
Name: Total_Venta, dtype: float64
```

```
[43]: len('DataFrame de Bogotá')
```

```
[43]: 18
```

Exportar el archivo a Excel con cada DataFrame,

```
[45]: with pd.ExcelWriter('ventas_análisis.xlsx') as writer:  
      df.to_excel(writer, sheet_name='DataFrame General', index=False)  
      df_vendedores.to_excel(writer, sheet_name='Ventas por Vendedor', index=True)  
      df_Bogotá.to_excel(writer, sheet_name='DataFrame de Bogotá', index=False)  
      df_A.to_excel(writer, sheet_name='Ventas por producto A', index=True)  
      df_Producto.to_excel(writer, sheet_name='Ventas por producto', index=True)  
      df_Ciudad.to_excel(writer, sheet_name='Ventas por Ciudad', index=True)  
      df_vendedores_producto.to_excel(writer, sheet_name='Ventas vendedor  
      ↪producto', index=True)
```