

Índice

[Caracterización de los datos 3](#_Toc189414167)

[ Ejercicio 1: 3](#_Toc189414168)

[Análisis exploratorio de datos 5](#_Toc189414169)

[ Ejercicio 2: 5](#_Toc189414170)

[ Ejercicio 3: 6](#_Toc189414171)

[Análisis multidimensional 7](#_Toc189414172)

[ Ejercicio 4: 7](#_Toc189414173)

[Sistemas para análisis descriptivo 8](#_Toc189414174)

[ Ejercicio 5: 8](#_Toc189414175)

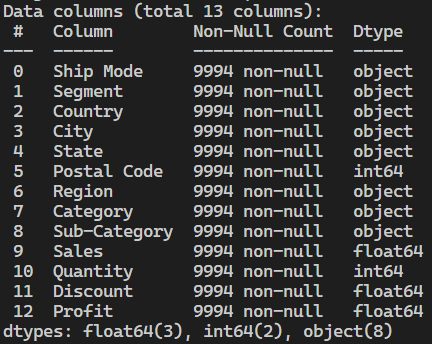
# Caracterización de los datos

# Ejercicio 1:

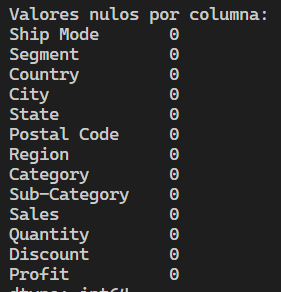
* + 1. **Determinar:**
       - **El tamaño del dataset (número de filas y columnas).**



* + - * **Los tipos de datos de cada variable.**



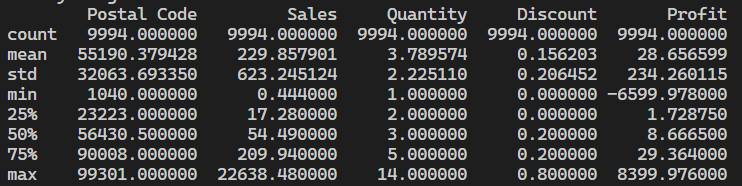
* + - * **Los valores faltantes, si existen, y cómo se podrían manejar.**



No hay.

* + - * **La descripción estadística de las variables numéricas**

**(mínimo, máximo, media, mediana, desviación estándar).**



* + 1. Responder las preguntas:
       - **¿Cuáles son las variables categóricas y cuáles las numéricas?**
* **Variables categóricas (tipo object):** Ship Mode, Segment, Country, City, State, Region, Category, Sub-Category.
* **Variables numéricas (tipo float64 e int64):** Postal Code, Sales, Quantity, Discount, Profit.
  + - * **¿Existen valores atípicos en las variables? ¿Cómo los identificaste?**
* Valores atípicos identificados:
  + - * Sales: 1167 valores atípicos
      * Profit: 1881 valores atípicos
      * Quantity: 170 valores atípicos
      * Discount: 856 valores atípicos

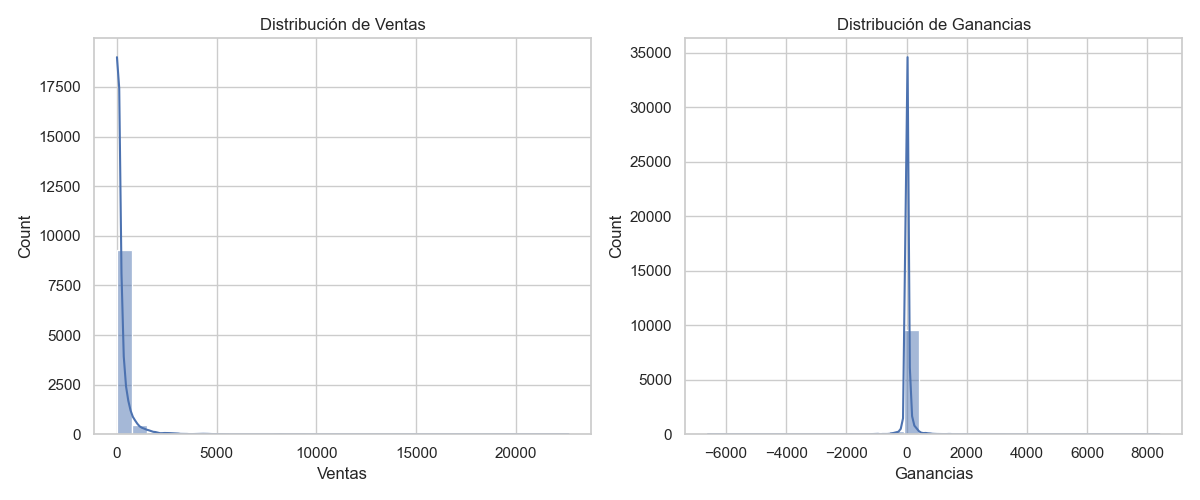
Lo calculé sacando el IQR

# Análisis exploratorio de datos

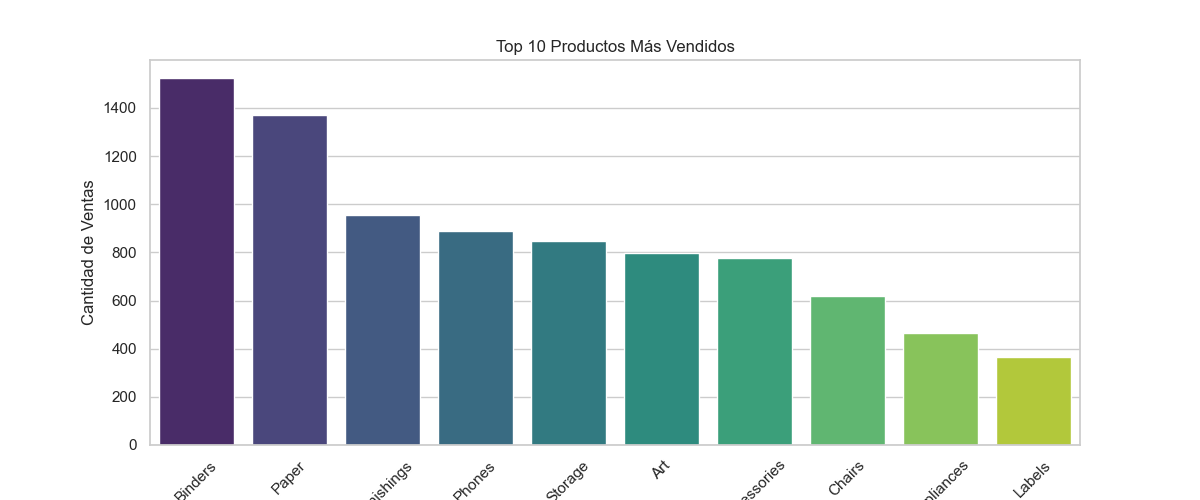
1. **Análisis univariante**

# Ejercicio 2:

* + 1. Crear gráficos para las variables principales:
       - **Histogramas para las variables numéricas (Ventas, Ganancia)**



* + - * **Diagramas de barras para variables categóricas (Región, Producto).**



* + 1. Interpretar los gráficos:
       - **¿Qué distribución tienen las ventas?**

Basándonos en él, es probable que la distribución sea **sesgada a la derecha**, lo que indica que la mayoría de las ventas son bajas, pero hay algunos valores extremadamente altos que actúan como valores atípicos.

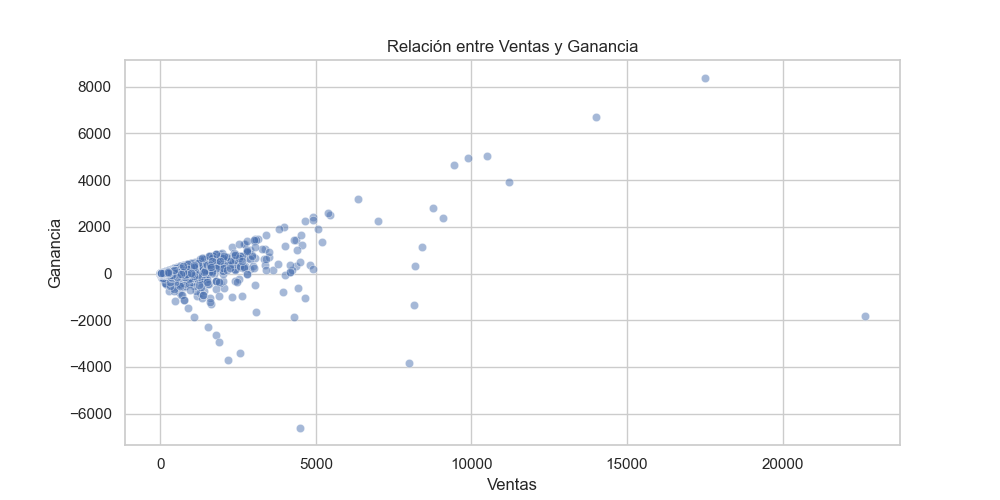
* + - * **¿Qué producto es el más vendido?**

Producto más vendido: Binders

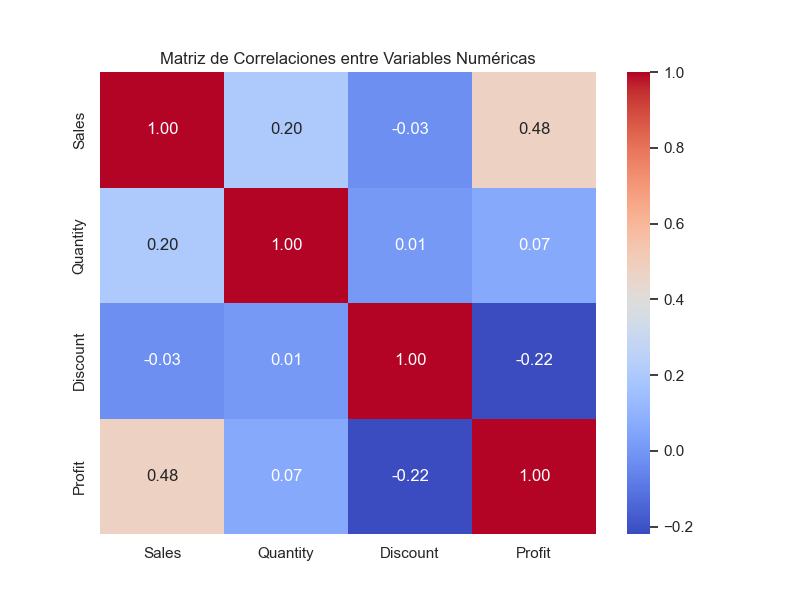
1. **Análisis multivariante**

# Ejercicio 3:

* + 1. Generar gráficos como:
       - **Diagramas de dispersión para analizar la relación entre Ventas y Ganancia.**



* + 1. **Calcular las correlaciones entre las variables numéricas**.



* + 1. Interpretar:
       - **¿Hay una relación entre las ganancias y las ventas?**

Si, a más ventas, más ganancias

* + - * **¿En qué región se venden más productos?**

En la región West

# Análisis multidimensional

# Ejercicio 4:

* + 1. Seleccionar indicadores como:
       - **Ventas totales por región.**

Ventas totales por región:

* Central 501239.8908
* East 678781.2400
* South 391720.9050
* West 725457.8245
  + - * **Producto más vendido.**

Producto más vendido: Binders

* + 1. Responder:
       - **¿Qué regiones tienen mejor desempeño?**

Promedio de ganancia por región:

* Central 17.091848
* East 32.135808
* South 28.857673
* West 33.848729
  + - * **¿Qué áreas necesitan más atención según los indicadores?**

La de profit

# Sistemas para análisis descriptivo

# Ejercicio 5:

* + 1. Realizar:
       - **Una reflexión sobre qué herramienta prefieres y por qué.**

Personalmente, prefiero utilizar **Python** junto con librerías como **Pandas** y **Matplotlib/Seaborn, primero que nada, porque es la única herramienta que hemos visto y con la que estamos trabajando.**

**Creo que** facilita la visualización, interpretación y extracción de los datos, o al menos lo considero que lo hace mejor que las otras herramientas.