**Taller Final Análisis Probabilístico**

Alexander Velez

Tecnológico De Antioquia

Facultad De Ingeniería

Julián Castelblanco Benítez

Mayo de 2025

Índice

1. Introducción .......................................................................... 1

2. Objetivos ............................................................................ 2

3. Análisis Descriptivo ............................................................... 3

3.1 Inspección inicial y limpieza de datos......................................................... 3

3.2 Estadísticas descriptivas ......................................................... 3

3.3 Análisis de ventas por mes .............................................. 4

3.4 Análisis de ventas por categoría .................................................. 4

3.5 Top 5 ciudades con más ventas.................................................. 5

4. Intervalos de Confianza ...................................................... 6

5. Pruebas de Hipótesis ......................................................... 7

6. Conclusiones ...................................................................... 8

7. Referencias ......................................................................... 9

8. Anexos ................................................................................ 10

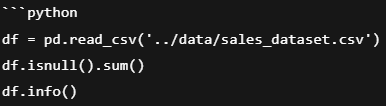
´

**3. Análisis Descriptivo**

Para iniciar el análisis probabilístico se trabajó con un conjunto de datos reales de ventas contenido en el archivo `sales\_dataset.csv`. En esta etapa se realizó una inspección preliminar del dataset, una limpieza básica, y un análisis descriptivo a través de estadísticas y visualizaciones.

**3.1 Inspección inicial y limpieza de datos**

El conjunto de datos se cargó con la librería `pandas`, y se revisaron los valores faltantes con `isnull().sum()`, identificando columnas potencialmente incompletas. Posteriormente, se verificó la estructura del dataset usando `info()`, lo cual permitió confirmar los tipos de datos de cada columna.



Esta revisión inicial permitió identificar si era necesario limpiar o transformar los datos antes del análisis. Aunque no se aplicaron técnicas avanzadas de imputación, se aseguró que las columnas clave (Amount, Profit, Quantity) estuvieran completas para el análisis.

**3.2 Estadísticas descriptivas**

Se calcularon estadísticas básicas de las columnas cuantitativas utilizando describe(), con foco en las variables Amount, Profit y Quantity:



Esto permitió obtener medidas como la media, mediana, desviación estándar, valores mínimos y máximos, fundamentales para entender la dispersión y tendencias generales de los datos. La variable Amount representa el monto total de cada venta, y Profit el beneficio correspondiente.

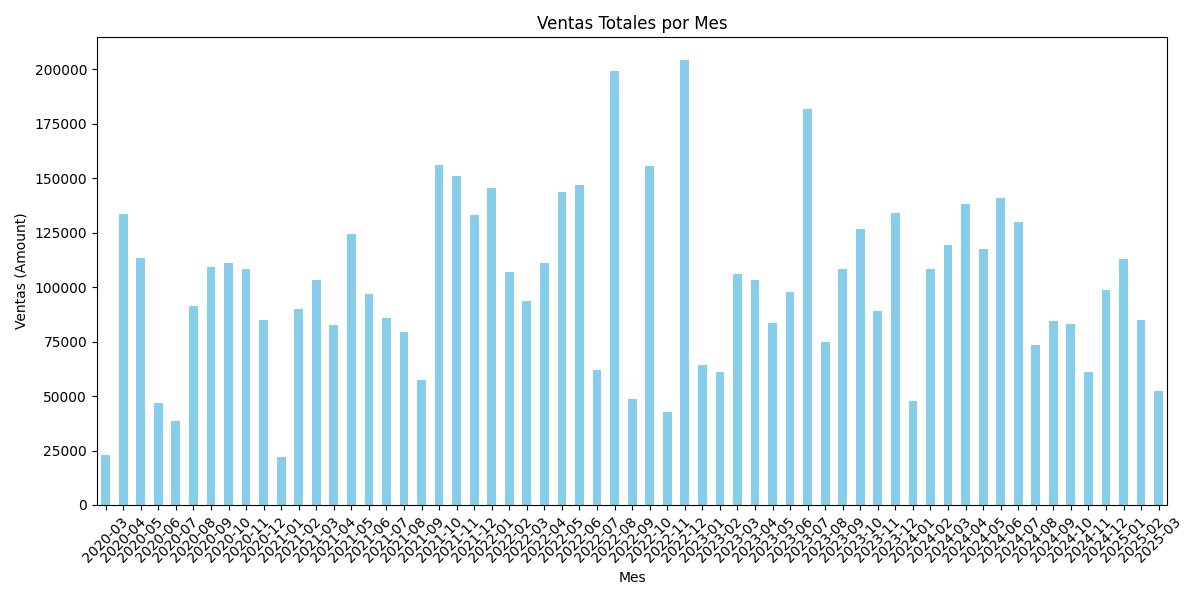
**3.3 Análisis de ventas por mes**

Se agruparon las ventas por mes utilizando la columna Year-Month, sumando los montos (Amount). Esto permite observar estacionalidades o picos de ventas:

Texto

El contenido generado por IA puede ser incorrecto.

**Figura 1**

****  
Ventas totales por mes (agrupadas por columna Year-Month).  
*Se observa una variación considerable entre meses, lo cual podría estar relacionado con campañas promocionales, temporadas altas o decisiones estratégicas internas de la empresa.*

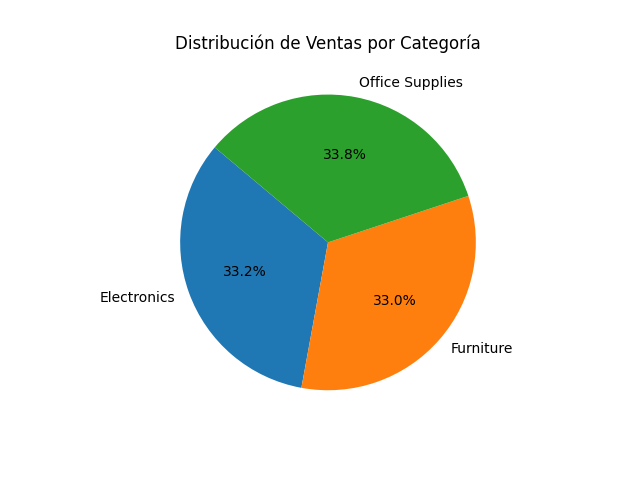
**3.4 Análisis de ventas por categoría**

Se exploró el comportamiento de las ventas según la categoría del producto (Category). Los resultados se visualizaron mediante un gráfico circular (pie chart):

Texto

El contenido generado por IA puede ser incorrecto.

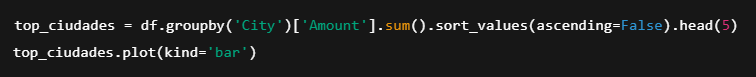
**Figura 2**

****

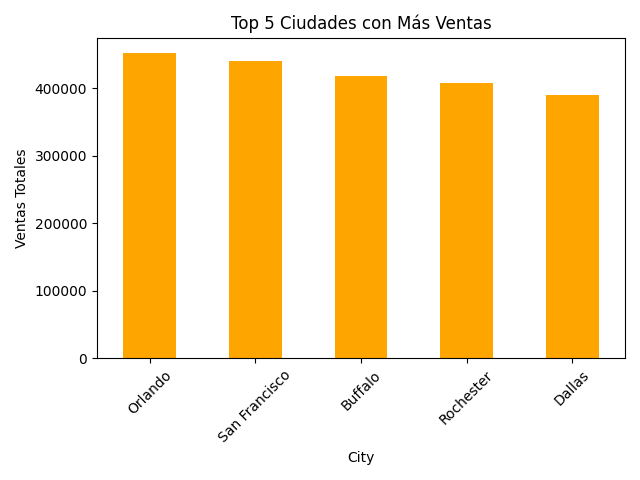
Distribución de ventas por categoría.  
*Este análisis ayuda a identificar qué tipo de productos representan la mayor parte del ingreso total. Puede usarse para tomar decisiones sobre inventario o marketing.*

**3.5 Top 5 ciudades con más ventas**

Finalmente, se identificaron las cinco ciudades con mayor volumen de ventas, lo cual permite conocer dónde se concentran los clientes más activos o rentables:



**Figura 3**



Top 5 ciudades con mayores ventas totales.  
*Este tipo de análisis geográfico puede orientar la toma de decisiones logísticas o comerciales, como la ubicación de nuevos puntos de venta o campañas regionales.*