Exploratory Data Analysis (EDA) on Superstore Dataset

What is EDA?

Exploratory Data Analysis (EDA) is a crucial step in any data science or analysis project. It involves summarizing the main characteristics of the data, often with visual methods, to understand patterns, detect outliers, and form hypotheses.

In this project, I performed EDA on the 'Superstore' dataset using Python. The analysis includes cleaning, understanding variable distributions, studying relationships between features, and identifying trends over time.

Steps Performed

1. Data Import & Cleaning:

- Loaded the dataset using Pandas
- Checked data structure, null values, and removed duplicates

2. Date Conversion:

- Converted 'Order Date' and 'Ship Date' to datetime format

3. Univariate Analysis:

- Analyzed categorical variables using countplots (e.g., Category)
- Examined numerical features with histograms and boxplots (Sales, Profit)

4. Bivariate Analysis & Correlation:

- Explored relationships like Sales vs Profit
- Used heatmap to study correlations
- Analyzed profit distribution across segments

5. Time Series Trend:

- Visualized monthly sales trends

6. Key Insights Extracted:

- Most sold category
- Region with highest sales

Exploratory Data Analysis (EDA) on Superstore Dataset

- Segment with highest profit
- Correlation values

Key Insights

- Most sold category: Technology/Furniture/Office Supplies (based on count)
- Region with highest sales: Identified using groupby on 'Sales'
- Segment with highest profit: Corporate/Consumer/Home Office
- Strong positive correlation seen between Sales and Profit in general

These insights help in understanding customer behavior, sales dynamics, and help businesses make informed decisions.