# Data Analyst Project : Sales Dataset Analysis

## 1. Data Understanding and Preparation

### Explore the Dataset:

- Check the structure:
  - Review columns and their data types (numeric, categorical, datetime).
- Sample records:
  - Inspect a random sample of records to understand data entries and potential issues.
- Missing values:
  - Identify columns with missing values and assess their impact.
- Duplicates:
  - Check for duplicate rows based on all or specific columns.

### Data Cleaning:

- Handle missing values:
  - Decide on imputation methods (mean, median, forward/backward fill) or removal based on data importance and completeness.
- Column cleanup:
  - Standardize inconsistent column names or formats (e.g., converting text to lowercase).
- Duplicate removal:
  - Eliminate duplicate rows if they skew analysis results.

## 2. Descriptive Statistics and Initial Insights

### Summary Statistics:

- Compute statistics:
  - Mean, median, mode, range, standard deviation for numerical columns (e.g., Price per Unit, Units Sold, Total Sales).

### Distribution Analysis:

- Visualize distributions:
  - Use histograms or density plots to understand the spread and skewness of numerical data.

### Explore Categorical Variables:

- Unique values:
  - Count and visualize unique values in categorical columns (e.g., Retailer, Region, Product).
- Popular items:
  - Determine top products, best-performing retailers, or preferred sales methods based on frequency.

**3. Sales Performance Analysis**

**Total Sales Trends:**

- Time analysis:
  - Plot sales trends over months or quarters using Invoice Date.
- Seasonality:
  - Detect recurring patterns or seasonal variations in sales.

**Regional Analysis:**

- Compare regions:
  - Evaluate total sales, average sales per retailer, or profitability across different regions.

**Product Analysis:**

- Best-sellers:
  - Identify and rank products by total sales or profitability.
- Contribution to revenue:
  - Determine which products generate the highest revenue contribution.

**4. Find the Correlation Matrix** : Gives the correlation among different attributes in form of a matrix.

**5. Visualization and Reporting**

**Create Visualizations:**

- Tool usage:
  - Utilize matplotlib, seaborn, or Tableau for generating clear and informative visualizations.

**Prepare Summary Report:**

- Documentation:
  - Detail methodologies, assumptions, and preprocessing steps.
- Insights:
  - Summarize key findings with supporting visualizations and statistical analyses.