

Sales Analysis Report

1. Introduction

This report presents an in-depth analysis of a sales dataset, covering data cleaning, preprocessing, exploratory data analysis (EDA), and key insights derived from statistical and graphical methods. The objective is to identify sales trends, seasonality, top-performing products, and regional performance, ultimately providing actionable insights for business growth.

2. Data Preprocessing

2.1 Data Loading and Initial Exploration

- Imported essential libraries into the Jupyter Notebook.
- Loaded the dataset and inspected the first and last five rows using appropriate methods.
- Checked the total number of rows and columns.
- Examined the data types of each column.
- Identified missing values.

2.2 Handling Missing Values

- Found missing data in several columns: 89%, 52%, 2%, and 38% missing values.
- Dropped columns with excessive missing values.
- Filled missing values using mean, median, or mode, depending on the data type.

2.3 Data Cleaning and Transformation

- Checked for duplicate records and removed them if found.
- Converted OrderDate from object type to datetime format.
- Replaced incorrect Year values to reflect actual years.
- Identified and handled outliers, especially in the Sales column.

3. Exploratory Data Analysis (EDA)

3.1 Univariate Analysis

- **Histogram Analysis:** The DealSize column contains three types of data: Small, Medium, and Large, with Medium being the most preferred.
- **Pie Chart Analysis:** The Status column indicates that the 'Shipped' mode performs best.

3.2 Bivariate Analysis

- **Bar Chart Analysis:**
 - Large deal sizes result in higher sales.
 - Most sales transactions are in 'On Hold' status.
 - The year 2005 had the highest sales compared to other years.
 - The highest quantity of orders occurred in 'Shipped' mode.

3.3 Multivariate Analysis

- **Sales, Status, and Deal Size Analysis:**
 - Medium deal sizes in 'Shipped' mode perform the best.
- **Status, Deal Size, and Quantity Ordered Analysis:**
 - Confirmed that medium deal sizes in 'Shipped' mode generate the highest sales and order quantities.

4. Key Insights and Findings

4.1 Sales Trends Over Time

- Sales have decreased over the years.
- Quantity ordered has also declined over time.

4.2 Regional Sales Performance

- **State-wise Performance:** California (CA) has the highest sales.
- **Country-wise Performance:** The USA records the highest quantity of orders.

- **Territory Analysis:** 88% of sales originate from the EMEA (Europe, Middle East, and Africa) region.

5. Conclusion and Recommendations

5.1 Summary of Key Findings

- **Overall Sales Trend:** Sales and order quantities have declined over the years.
- **Seasonality:** Identified peak and low sales months.
- **Top-Performing Products:** Medium deal sizes in 'Shipped' mode generate the most revenue.
- **Regional Performance:** EMEA dominates sales, and California leads at the state level.
- **Sales Factors:** Possible reasons for declining sales include market conditions, product demand shifts, and economic factors.

5.2 Actionable Recommendations

- Investigate reasons behind declining sales and take corrective actions.
- Focus marketing efforts on top-performing regions and product categories.
- Optimize inventory and logistics for 'Shipped' mode to improve efficiency.
- Further analyze customer behavior to enhance sales strategies.

6. Next Steps

- Convert this analysis into a Jupyter Notebook for further use.
- Develop interactive visualizations using Tableau or Power BI.
- Create a Streamlit dashboard to provide real-time sales insights.

This report provides a comprehensive understanding of sales trends and patterns, helping businesses make data-driven decisions to enhance sales and profitability.