

Exploratory Data Analysis (EDA) Report

This report presents the findings from the Exploratory Data Analysis (EDA) conducted on the provided sales dataset. The EDA process involved data inspection, cleaning, visualization, and deriving insights to better understand the data and uncover trends, patterns, and relationships.

1. Data Overview

The dataset consists of 9800 records with 18 columns, including order details, customer details, shipping information, product details, and sales amounts. Most columns are of object type, with two float64 columns (Postal Code, Sales) and one int64 column (Row ID).

2. Missing Values & Data Types

- Only the 'Postal Code' column has 1 missing value.
- All other columns have complete data.
- Data types were reviewed, and date columns ('Order Date', 'Ship Date') were parsed for time-based analysis.

3. Descriptive Statistics

The `.describe()` method was used to summarize numerical data. This provided measures such as mean, standard deviation, and quartiles for columns like 'Sales'. Categorical variables were summarized using `.value_counts()` to understand the distribution of values.

4. Univariate Analysis

Histograms and countplots were created to understand the distribution of key variables:

- Sales distribution shows right-skewness, indicating a small number of very high-value sales.
- Most orders belong to a few key product categories.
- Segment distribution shows most customers belong to the 'Consumer' segment.

5. Bivariate Analysis

Scatterplots and boxplots were used to identify relationships:

- Positive correlation between Sales and Profit.
- Certain categories show more variability in sales amounts.
- Shipping mode impacts delivery time, with 'Same Day' offering shortest delays.

6. Correlation Analysis

A heatmap of numerical features revealed strong correlation between Sales and Profit, while other numeric variables showed weak or no correlation.

7. Key Findings

- Sales are concentrated in a few key product categories.
- Consumer segment contributes the largest share of orders.

- There is a positive relationship between sales and profit, but some products may yield low or negative profits despite high sales.
- Shipping mode impacts delivery time significantly.

8. Conclusion

The dataset provides valuable insights into sales performance, customer segmentation, and shipping patterns. These findings can help guide business decisions such as targeting profitable product categories, improving delivery efficiency, and addressing low-profit items.