

EDA Report: Sales Dataset

Introduction

This report summarizes the analysis of a sales dataset (train (3).csv). The data included various customer and transaction details.

Missing Value Handling

Missing values were handled carefully: numerical columns were filled with median values and categorical columns with the mode.

Dataset Summary

The dataset contained multiple numerical and categorical features. Data types and distributions were inspected using `info()` and `describe()`.

Visualization Observations

- Histograms showed general distribution trends.
- Boxplots identified outliers in numeric features.
- Pairplots helped visualize relationships among important variables.

Correlation Analysis

The correlation heatmap highlighted the strength of relationships between numeric features. Strong correlations were identified and noted.

Conclusions

The dataset was cleaned without any row loss. Outliers were visualized. Relationships between major variables were successfully identified for further analysis.