Data Analysis Report

# Step 1: Data Extraction

The dataset was loaded from a CSV file containing sales data with columns: OrderID, Product, Category, Price, Quantity, Date, and Region.

# Step 2: Data Cleaning

No missing values or duplicates were found in the dataset.

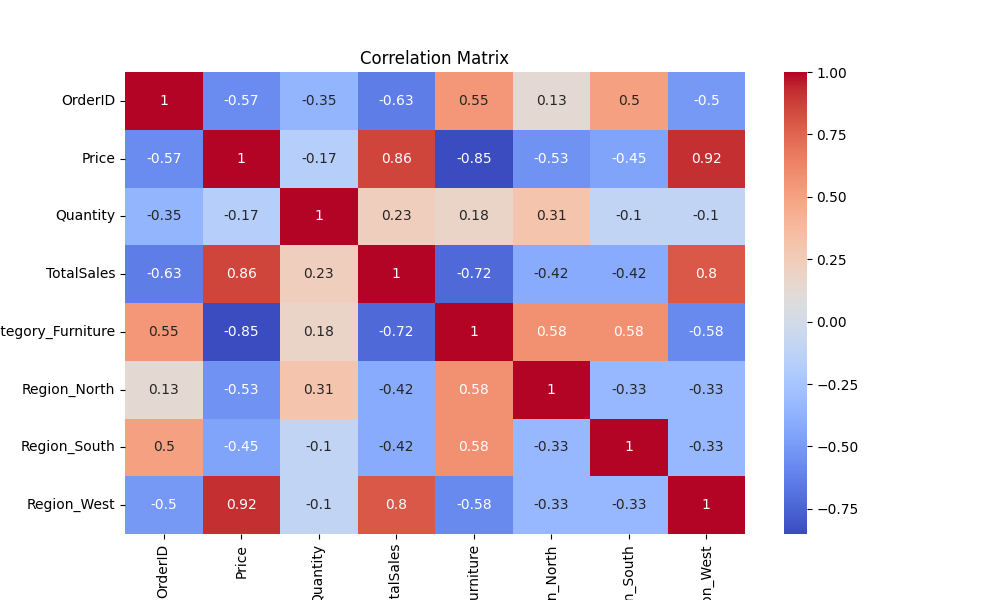
# Step 3: Data Wrangling

The data was transformed by setting OrderID as the index, converting Date to datetime, creating a TotalSales feature, and encoding categorical variables (Category and Region) using one-hot encoding.

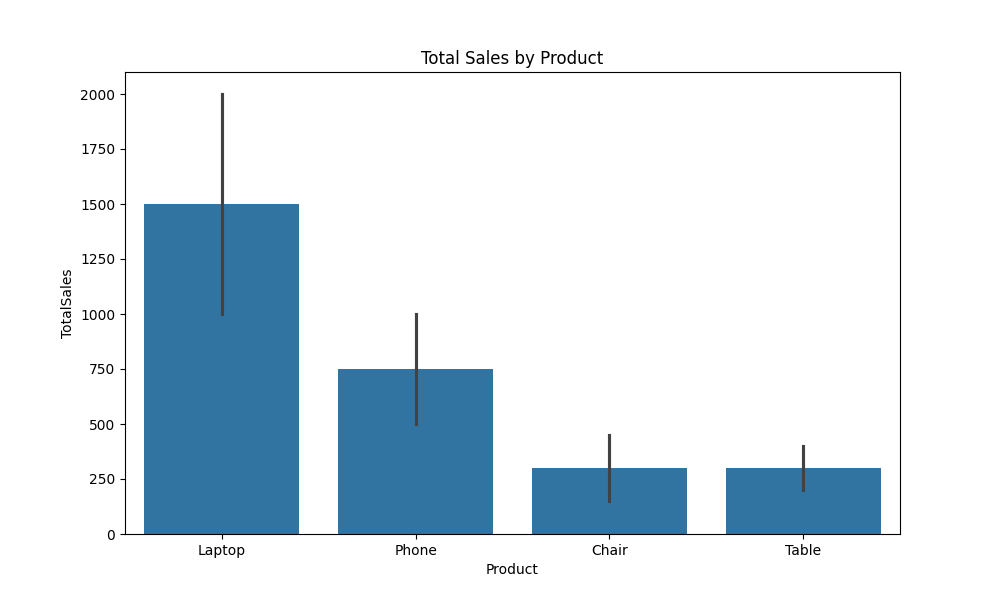
# Step 4: Data Analysis

Exploratory data analysis revealed insights about sales trends. Visualizations include:

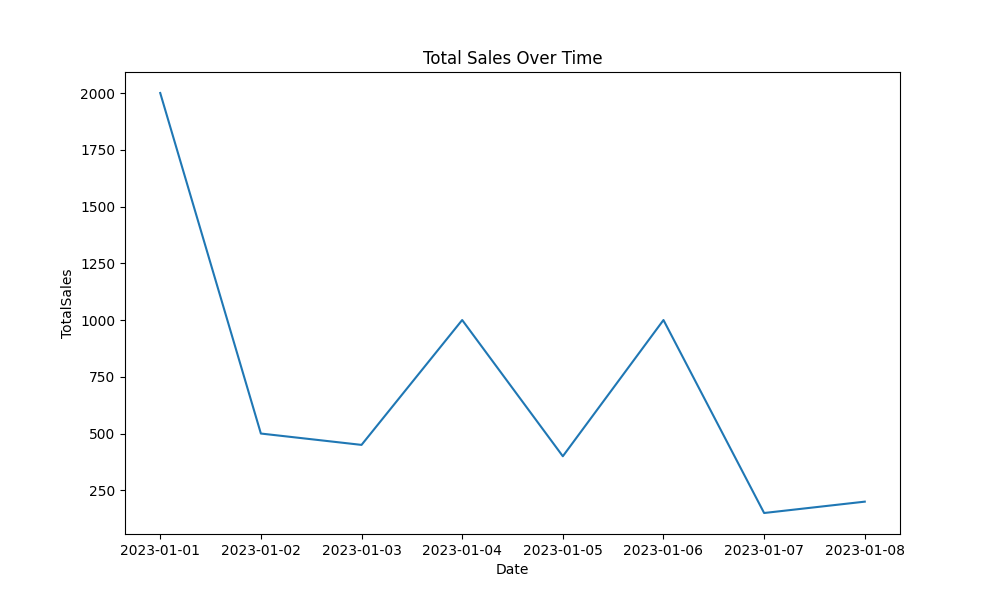
* 1. Correlation Matrix



* 1. Total Sales by Product



* 1. Total Sales Over Time



# Step 5: Action (Modeling)

A linear regression model was trained to predict TotalSales. The model achieved a Mean Squared Error of 10569.13 and an R^2 Score of 0.83.

Conclusion: The model successfully predicts TotalSales based on Price, Quantity, and categorical features. The high R^2 score indicates a good fit, suggesting that the model can be used for sales forecasting.