Drug Sales Prediction - Project Documentation

Team Name : G.O.A.T.

Date:10.10.2024

This document outlines the process and results of a Machine Learning project aimed at predicting drug sales based on various factors such as marketing spend, competitor prices, seasonality, and more. The dataset contains information on sales volume and influencing factors over time.

# 1. Data Exploration

The dataset was first explored to understand the structure, data types, and basic statistics. Key variables include marketing spend, competitor price, average price, seasonality, economic index, and prescription rates.

# 2. Data Preprocessing

To prepare the data for modeling, categorical variables were converted into numerical representations using One-Hot Encoding. Continuous variables were standardized using a scaler. Missing values, if any, were handled by imputation or removed if necessary.

# 3. Feature Engineering

Feature scaling was performed on continuous variables to bring them to a common scale. New features, such as price elasticity or seasonality indexes, were explored to capture patterns in the data.

# 4. Model Selection

Multiple models were explored to predict the sales volume, including Linear Regression and Gradient Boosting. An 80-20 train-test split was used to evaluate model performance.

# 5. Model Evaluation

Model performance was evaluated using Mean Squared Error (MSE) and R-Squared. The Gradient Boosting model was selected for its superior performance.

# 6. Visualizations

The following graphs provide insights into the model's predictions and the relationships between features and sales volume.

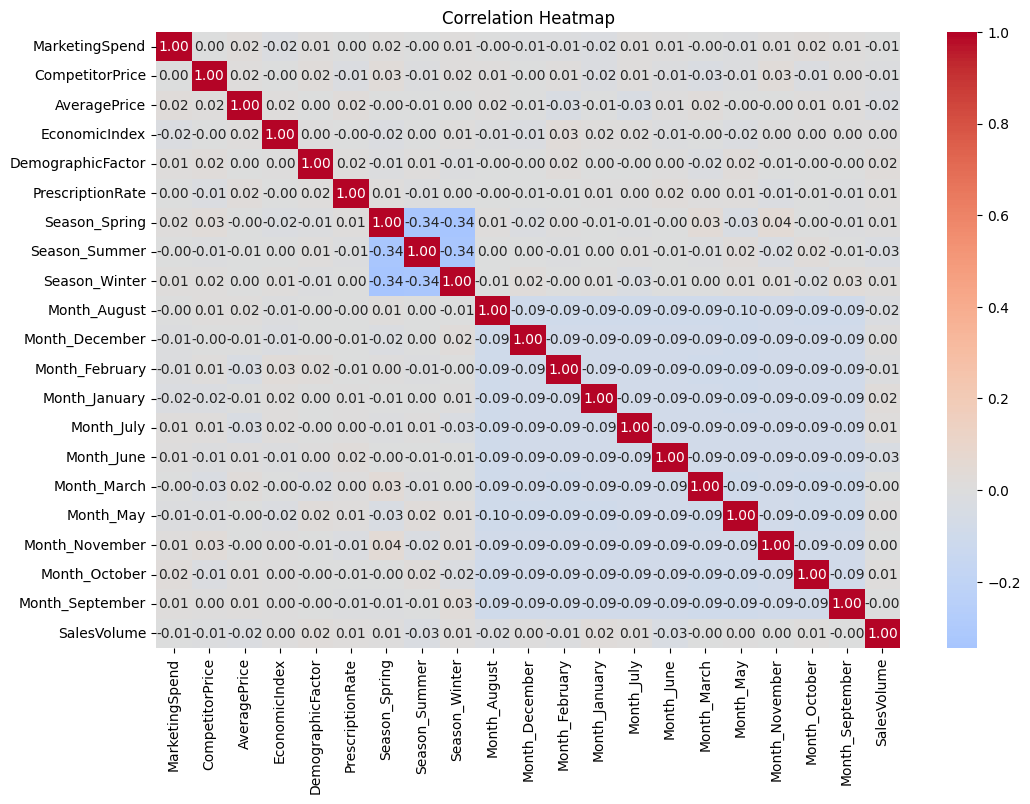


Figure 1: Correlation Heatmap

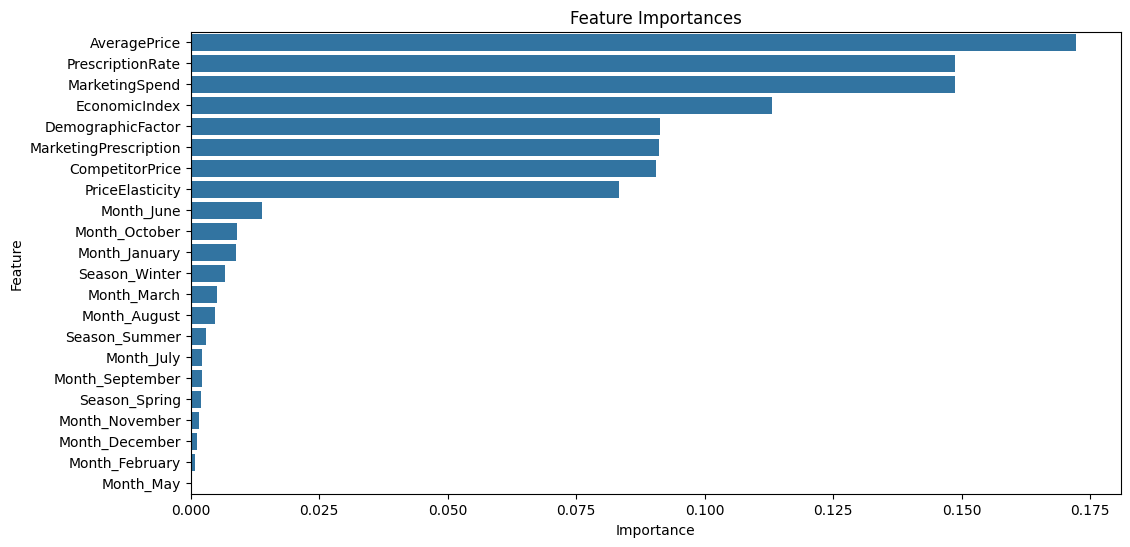
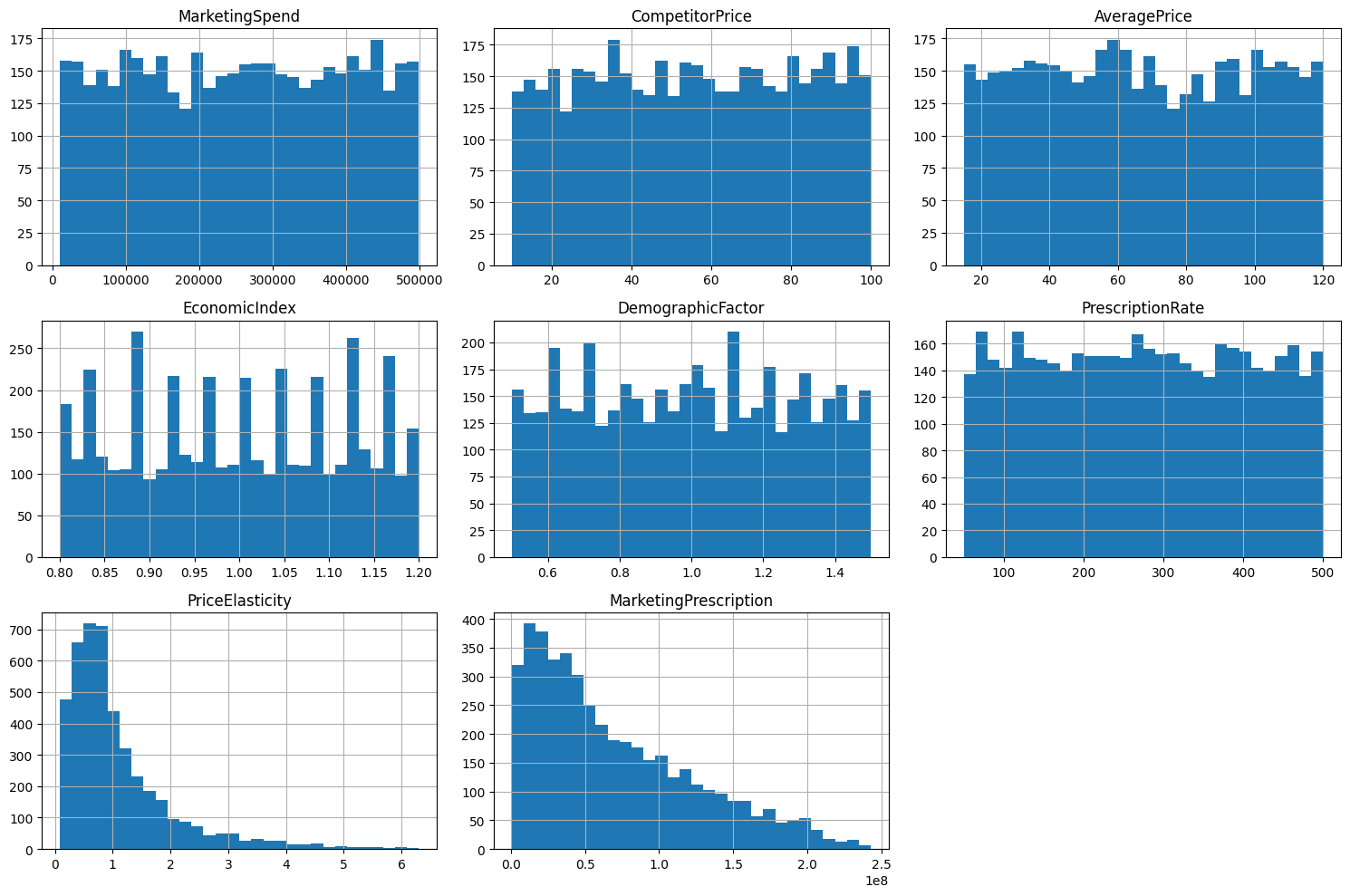
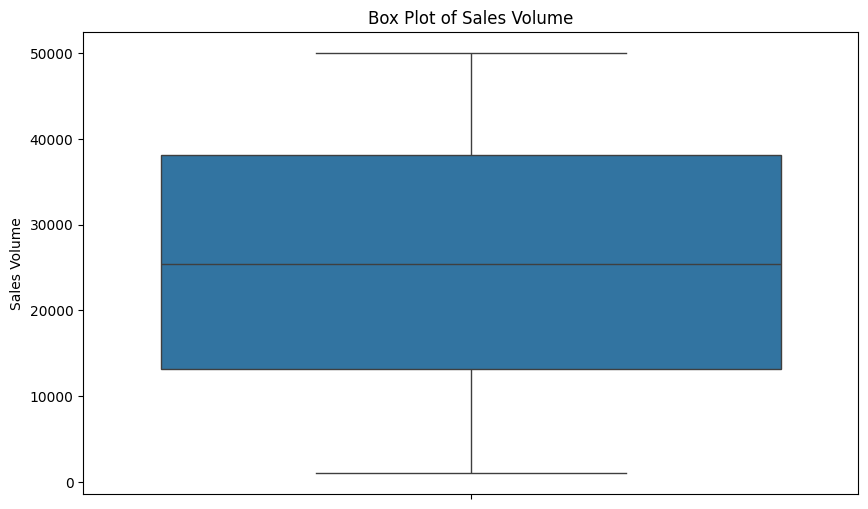


Figure 2: Feature importances





# 7. Conclusion and Recommendations

The Gradient Boosting model provided the best performance in predicting sales volume. Key factors influencing sales include marketing spend and economic conditions. Recommendations for the company include optimizing marketing spend based on these insights and adjusting pricing strategies to align with competitor prices.