Capstone Final Report: Sales Forecasting for Retail Optimization

Introduction

In this project, I focused on developing a sales forecasting model to support strategic decision-making within a retail organization. My objective was to accurately predict future sales trends to optimize inventory management, streamline supply chain operations, and improve resource allocation. Leveraging various modeling techniques, feature engineering, and data preprocessing methods, I aimed to create a reliable and robust forecasting model that would address key business challenges and provide actionable insights.

Problem Statement

Efficient inventory and supply chain management are essential to profitability in the retail industry. Overstocking leads to excessive storage costs and potential product waste, while understocking risks missed sales opportunities and dissatisfied customers. This project addresses these challenges by building a predictive model that uses historical sales data and seasonal patterns to enable precise stock management, reducing costs and enhancing decision-making.

Project Goals and Objectives

1. Develop a Predictive Model: Building a model that accurately forecasts future sales trends.

2. Enhance Decision-Making: Provide actionable insights to support decisions related to pricing, promotions, and inventory planning.

3. Optimize Business Performance: Minimize costs associated with inventory mismanagement, maximize revenue, and improve customer satisfaction by aligning stock levels with predicted demand.

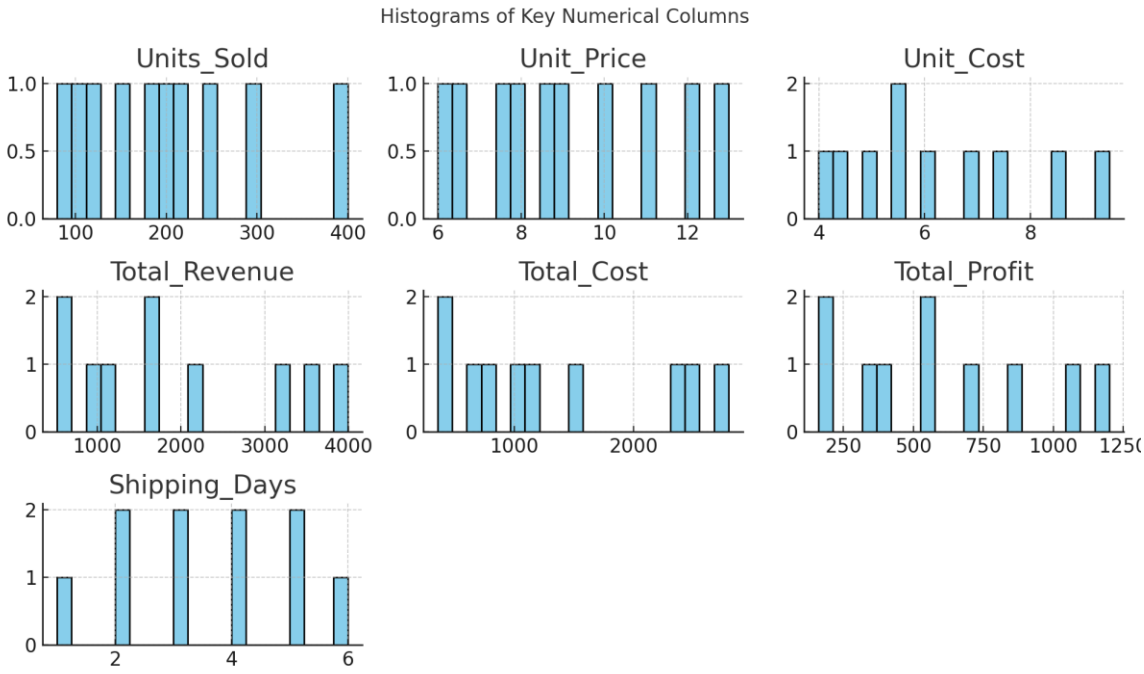
Methodology

I divided the project into several steps to ensure structured data processing, exploration, and model evaluation. Each step aimed to prepare the data, identify meaningful patterns, and implement a model that would meet the project's objectives.

1. Data Acquisition and Preprocessing: Prepared historical sales data, addressed missing values, removed irrelevant columns, and created time-based features.

2. Exploratory Data Analysis (EDA): Generated insights into data distributions, sales trends, and relationships between key variables.

Example: figure no.1



3. Feature Engineering: Applied one-hot encoding and normalization to ensure consistency across the dataset.

4. Model Selection and Training: Tested Linear Regression, Random Forest, and Decision Tree Regressors with hyperparameter tuning.

5. Evaluation Metrics: Used R^2 and RMSE scores to evaluate and compare model accuracy and error.

Model Metrics and Performance

I evaluated the performance of three regression models based on R^2 and RMSE, which assess accuracy and error in predictions. Below is a summary of the model metrics:

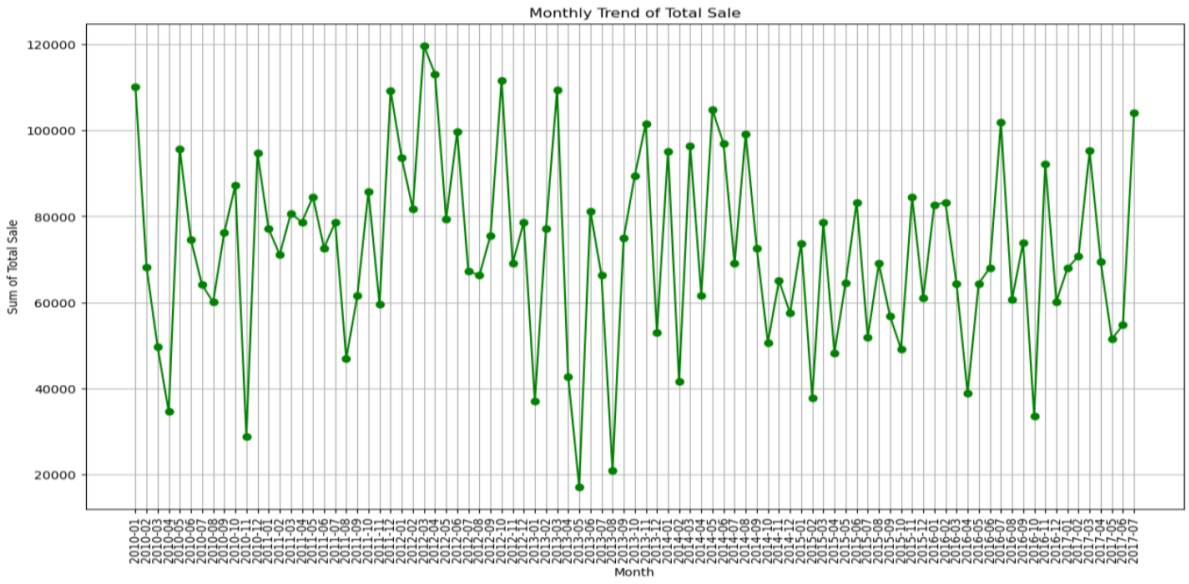
Model	R^2 (Train)	R^2 (Test)	RMSE (Train)	RMSE (Test)
Linear Regression	0.85	0.87	135,484.49	120,776.40
Random Forest Regressor	0.91	0.90	104,655.93	106,013.93
Decision Tree Regressor	0.89	0.88	113,483.33	115,176.42

The Random Forest Regressor provided the best performance, with an R^2 of 0.90 on the test set and the lowest RMSE of 106,013.93. This model was selected for its balance between high accuracy and low error in predicting sales trends.

Key Findings and Figures

1. Sales Patterns: Seasonal fluctuations were observed, with peak sales periods around major holidays.

figure no.2



2. Feature Importance: The Random Forest model highlighted key variables like Units_Sold and Item_Type, essential for inventory predictions.

3. Prediction Accuracy: The RMSE scores indicate that the model's predictions are reliable within a reasonable error range.

Recommendations

1. Implement Predictive Stock Management: Use model forecasts to optimize inventory levels.

2. Leverage Seasonal Promotions: Capitalize on peak sales periods with targeted promotions to maximize sales.

3. Regularly Update the Model: Retrain the model periodically with updated data to adapt to new trends and ensure sustained accuracy.

Conclusion and Future Research

The Random Forest Regressor emerged as the most effective model, achieving accurate predictions while avoiding overfitting. In future research, I could explore advanced models like LSTM for improved time-series forecasting, especially if the dataset grows or seasonal patterns become more pronounced. This would enable even more accurate sales predictions and offer deeper insights into seasonal demand.