

# Store Sales Prediction for Inventory and Revenue Optimization

Dr. S. Sreeja<sup>1</sup> and B. Sarulatha<sup>2</sup>

Assistant Professor, Department of Management Sciences<sup>1</sup>

Student II MBA, Department of Management Sciences<sup>2</sup>

Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India

**Abstract:** *The main objective of this project is to analyze store sales data and develop a sales prediction system for inventory and revenue optimization. Effective sales forecasting helps businesses manage stock efficiently, reduce inventory costs, and improve overall profitability. In this project, historical sales data is collected and analyzed using Google Colab and Python libraries.*

*The project involves several analytical techniques such as data cleaning, Exploratory Data Analysis (EDA), percentage analysis, trend analysis, correlation analysis, time series analysis, and ABC analysis to understand sales patterns and identify important factors affecting revenue. Visualizations such as heatmaps, bar charts, line graphs, and scatter plots are used to interpret the data effectively.*

*A Decision Tree machine learning model is implemented to predict future sales based on historical patterns. The analysis reveals that units sold and unit price have a strong influence on revenue, while inventory level has a limited impact. ABC analysis further identifies high-value products that contribute significantly to total sales.*

*Overall, the project helps in improving inventory management, reducing overstocking, optimizing pricing strategies, and supporting better business decision-making through accurate sales prediction and data-driven insights.*

**Keywords:** Store Sales Prediction, Inventory Optimization, Revenue Optimization

## I. INTRODUCTION

In today's competitive business environment, effective inventory management and accurate sales forecasting are essential for improving profitability and operational efficiency. Store sales prediction plays an important role in helping businesses understand customer demand, manage stock levels, and optimize revenue. By analyzing historical sales data, organizations can make better decisions related to inventory planning, pricing strategies, and resource allocation.

This project, titled "Store Sales Prediction for Inventory and Revenue Optimization," focuses on analyzing sales data to identify patterns, trends, and factors affecting business performance. The project uses Google Colab and Python-based analytical techniques to process and analyze the dataset.

Various methods such as data cleaning, Exploratory Data Analysis (EDA), percentage analysis, trend analysis, correlation analysis, time series analysis, and ABC analysis are applied to gain meaningful insights from the data. In addition, a Decision Tree machine learning model is used to predict future sales based on historical information.

The findings from this project help in identifying high-performing products, understanding seasonal demand, reducing overstocking, and improving revenue generation. Overall, the project demonstrates how data analysis and predictive modeling can support better inventory management and business decision-making



## **II. REVIEW OF LITERATURE**

### **2012 – Matthew Schneider and Sachin Gupta**

The authors studied sales forecasting using customer reviews and product information collected from Amazon and tablet company websites. They used techniques such as bag-of-words, OLS regression, and SVM. The study found that customer reviews and product ratings help improve sales prediction accuracy, especially for new products.

### **2019 – Zhi-Ping Fan, Yu-Jie Che, and Zhen-Yu Chen**

The researchers focused on predicting product sales using historical sales data and online customer reviews. They used sentiment analysis methods like Naive Bayes, SVM, and KNN to understand customer opinions. The study concluded that combining customer sentiment with sales data improves forecasting performance and helps businesses understand market demand better.

### **2020 – Devendra Swami, Alay Dilipbhai Shah, and Subhrajee K. B. Ray**

The authors worked on retail sales forecasting using machine learning techniques such as XGBoost and LSTM. They analyzed time-series sales data and found that machine learning models can effectively predict future product sales. The study showed that XGBoost performed better than LSTM for sparse retail datasets.

### **2020 – Xuan Bi, Gediminas Adomavicius, William Li, and Annie Qu**

The researchers proposed a tensor factorization approach called ATLAS for retail sales forecasting. The model used sales data from multiple stores and products to improve prediction accuracy. Their study highlighted that better forecasting supports inventory control, financial planning, and supply chain management.

### **2022 – Linda Eglite and Ilze Birzniece**

The authors conducted a systematic review on deep learning methods for retail sales forecasting. Models such as LSTM, CNN, ANN, and RNN were analyzed. The study found that deep learning techniques provide better forecasting accuracy than traditional statistical methods, especially when quality data is available.

### **2022 – Retail Forecasting: Research and Practice**

This study discussed different forecasting methods used in retail businesses. The research explained that accurate forecasting helps in inventory management, pricing decisions, and demand planning. The study also highlighted the growing importance of machine learning methods in retail forecasting.

### **2023 – Suresh B S and M. Suresh**

The authors analyzed machine learning and deep learning models used for retail sales prediction. Their study explained that sales forecasting helps businesses improve profit, demand planning, and inventory management. The research emphasized the importance of data mining and predictive analytics in retail industries.

## **III. PROBLEM STATEMENT**

Many retail stores face challenges in managing inventory and predicting future sales accurately. Inefficient inventory management can lead to overstocking or stock shortages, which directly affects revenue and customer satisfaction. Additionally, the lack of proper sales forecasting makes it difficult for businesses to understand demand patterns and make effective decisions regarding pricing, stock planning, and resource allocation.

This project aims to address these issues by analyzing historical store sales data and developing a sales prediction model to forecast future sales. The study focuses on identifying important factors influencing sales, optimizing inventory levels, and improving revenue generation through data-driven analysis and predictive techniques



#### **IV. OBJECTIVE OF THE STUDY**

- o To analyze historical store sales data
- o To predict future sales accurately
- o To optimize inventory levels
- o To maximize store revenue

#### **V. SOURCES OF DATA COLLECTION**

The data used for this project is collected from secondary sources in the form of store sales datasets. The dataset contains information related to product sales, unit price, revenue, inventory level, and sales dates. The data is stored and processed in Excel format and analyzed using Google Colab with Python libraries.

The collected data includes:

- Historical sales records
- Product details
- Revenue information
- Inventory data
- Time-based sales information

These data sources are used to perform analysis, identify sales patterns, and build predictive models for inventory and revenue optimization.

#### **VI. METHODOLOGY AND TECHNIQUES USED**

The methodology followed in this project involves collecting, cleaning, analyzing, and interpreting historical store sales data to predict future sales and optimize inventory and revenue. Initially, the dataset is collected and preprocessed by removing unnecessary columns, handling missing values, and correcting inconsistencies to improve data quality. After data cleaning, Exploratory Data Analysis (EDA) is performed to understand the structure and characteristics of the dataset.

Several analytical techniques are then applied, including percentage analysis, trend analysis, correlation analysis, time series analysis, and ABC analysis to identify sales patterns, relationships between variables, and product performance. Visualizations such as bar charts, line graphs, scatter plots, and heatmaps are used for better interpretation of the data. Finally, a Decision Tree machine learning model is implemented to predict future sales based on historical sales data and key influencing factors.

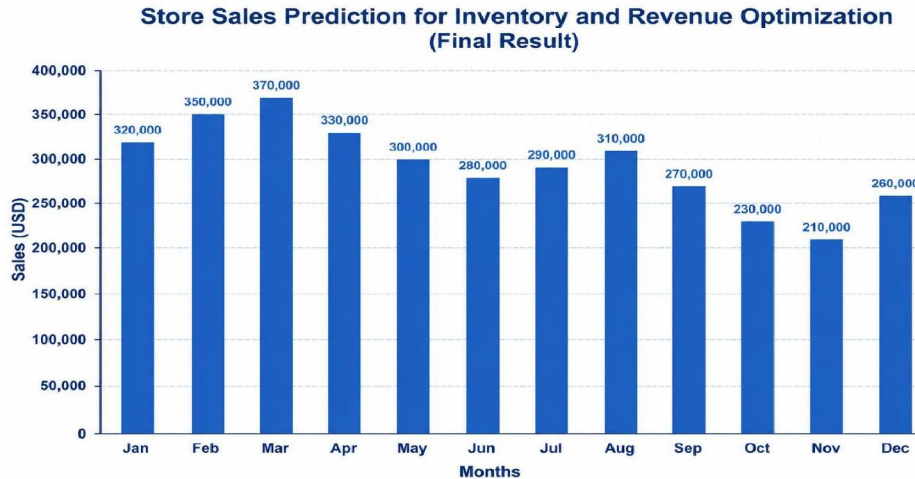
#### **VII. STATISTICAL ANALYSIS USED**

The following statistical and analytical tools are used in this project:

- Percentage Analysis – To measure the contribution of products, categories, and sales to overall revenue
- Trend Analysis – To identify sales growth and patterns over time
- Correlation Analysis – To study the relationship between variables such as units sold, price, and revenue
- Time Series Analysis – To analyze sales performance across different time periods
- ABC Analysis – To classify products based on their sales contribution and importance
- Decision Tree Algorithm – To predict future sales using machine learning techniques
- Data Visualization Tools – Bar charts, line charts, heatmaps, and scatter plots for graphical representation of data



### VIII. DATA VISUALIZATION FOR MY ANALYSIS



#### INTERPRETATION

- The highest sales are recorded in March (370,000 USD), followed by February (350,000 USD) and April (330,000 USD), indicating strong customer demand during the first half of the year.
- Sales gradually decline from May to November, with the lowest sales observed in November (210,000 USD).
- A slight recovery in sales is noticed in December (260,000 USD), which may indicate seasonal demand or year-end purchasing activity.
- The analysis shows that sales trends are not constant throughout the year, highlighting the importance of proper inventory planning and demand forecasting.
- Overall, the visualization helps identify peak and low sales periods, enabling businesses to optimize stock levels, reduce overstocking, and improve revenue management strategies.

#### IX. SUMMARY OF FINDINGS

- The analysis showed that units sold and unit price strongly influence revenue generation.
- Correlation analysis confirmed a strong positive relationship between sales, revenue, and product pricing.
- Month-wise and year-wise trend analysis identified that sales were highest during the early months of the year, especially in March and 2024 overall.
- Inventory level showed a weak impact on sales, indicating possible overstocking in some periods.
- ABC analysis revealed that only a few products contribute to the majority of total revenue, while many products generate lower sales.
- Scatter plots and predictive analysis showed that the Decision Tree model predicted sales with good accuracy.
- Time series analysis helped identify seasonal demand patterns and fluctuations in sales performance.

#### X. RECOMMENDATIONS

- Businesses should focus more on high-performing products identified in the ABC analysis to improve revenue.
- Proper inventory planning should be implemented to avoid overstocking and reduce holding costs.
- Pricing strategies should be optimized because unit price has a strong impact on revenue generation.



- Companies should use sales forecasting models regularly to predict future demand accurately.
- Seasonal sales trends should be considered while planning inventory and promotional activities.
- Machine learning techniques such as Decision Tree models can be further improved and used for better business decision-making.
- Continuous monitoring of sales data and customer demand can help improve overall operational efficiency and profitability.

### **XI. CONCLUSION**

The project “Store Sales Prediction for Inventory and Revenue Optimization” successfully analyzed historical sales data to identify patterns, trends, and factors affecting business performance. Various analytical techniques such as Exploratory Data Analysis (EDA), percentage analysis, trend analysis, correlation analysis, time series analysis, and ABC analysis were used to gain meaningful insights from the dataset. The study revealed that units sold and unit price are the primary factors influencing revenue, while inventory level has less impact on sales performance.

The Decision Tree model effectively predicted future sales using historical data, supporting better forecasting and business planning. ABC analysis further highlighted that a small number of products contribute significantly to overall revenue, emphasizing the importance of focusing on high-value products. Overall, the project demonstrates that data-driven analysis and predictive modeling can improve inventory management, optimize revenue, reduce overstocking, and support better strategic decision-making in retail businesses.

### **REFERENCES**

- [1]. Kesavan, S., Gaur, V., & Raman, A. (2010). Do Inventory and Gross Margin Data Improve Sales Forecasts for U.S. Public Retailers? *Management Science*, 56(9), 1519–1533. (Harvard Business School)
- [2]. Fildes, R., Ma, S., & Kolassa, S. (2022). Retail Forecasting: Research and Practice. *International Journal of Forecasting*, 38(4), 1283–1318. (ScienceDirect)
- [3]. Saccomanno, F. P., Trivella, A., & Guerriero, F. (2026). Integrated Sales Planning for In-Store Retail: A Multi-Stage Stochastic Optimization Approach. *European Journal of Operational Research*, 329(2), 669–686. (EconPapers)
- [4]. Beheshti-Kashi, S., Karimi, H. R., Thoben, K.-D., Lütjen, M., & Teucke, M. (2015). A Survey on Retail Sales Forecasting and Prediction in Fashion Markets. *Systems Science & Control Engineering*, 3(1), 154–161. (Taylor & Francis Online)
- [5]. Bi, X., Adomavicius, G., Li, W., & Qu, A. (2020). Improving Sales Forecasting Accuracy: A Tensor Factorization Approach with Demand Awareness. arXiv preprint arXiv:2011.03452. (arXiv)
- [6]. Shende, S., & Namdeo, A. (2024). Machine Learning-Based Sales Prediction and Inventory Management for Grocery Stores. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*. (IJARSCT)
- [7]. Fatima, A., & Salam, M. A. (2026). A Data-Driven Predictive Framework for Inventory Optimization Using Context-Augmented Machine Learning Models. arXiv preprint arXiv:2601.05033. (arXiv)
- [8]. Rekik, Y., Oliva, R., Glock, C., & Syntetos, A. (2025). Inventory Record Inaccuracy in Grocery Retailing: Impact of Promotions and Product Perishability, and Targeted Effect of Audits. arXiv preprint arXiv:2506.05357. (arXiv)
- [9]. Ping, L. Y., Wong, P., & Han, T. C. (2025). A Study of Data-Driven Methods for Inventory Optimization. arXiv preprint arXiv:2505.08673. (arXiv)
- [10]. Taylor, J. W. (2018). Considerations of a Retail Forecasting Practitioner. *International Journal of Forecasting*, 34(4), 822–829. (ScienceDirect)

