Retail Demand Forecasting:

Data-Driven Inventory and Sales Optimization

Bellevue University

DSC – 680 Applied Data Science

Project 3

Trevor Zeiger

Milestone 1 – Proposal

Topic

This project, *Retail Demand Forecasting: Data-Driven Inventory and Sales Optimization*, explores the application of time series and predictive analytics to retail sales data. The goal is to examine whether retail companies can improve inventory decisions, reduce stockouts and overstock, and align marketing strategies more effectively by using historical and contextual sales data to forecast future demand.

Business Problem

Retailers often struggle to match their inventory levels with customer demand, leading to either overstock or lost sales. Inaccurate demand forecasting leads to significant consequences: excess stock results in increased holding costs and markdowns, while insufficient stock causes missed sales opportunities and customer dissatisfaction. This project aims to answer the following key questions: Can retail transaction data be used to generate reasonably accurate forecasts of future sales? How do factors, such as holidays, promotions, or seasonal patterns, have an influence on demand? By addressing these questions, the project will provide insights that help retailers optimize inventory planning, improve revenue forecasting, and reduce inefficiencies.

Retail demand forecasting helps businesses make more informed decisions around inventory, sales planning, and promotions. It can support supply chain teams in maintaining the right stock levels, especially for fast-moving or seasonal items, while reducing waste from overordering. Marketing departments can use forecast data to better time promotions and prepare for demand around holidays or key sales periods. It also gives analysts a clearer picture of sales trends over time and can help leadership spot unusual patterns or shifts in customer behavior.

Datasets

To support this analysis, I will use a combination of publicly available retail datasets, primarily sourced from Kaggle, and supplemented by external context datasets if necessary.

1. **Walmart Store Sales Forecasting Dataset** – This dataset provides weekly sales data for various Walmart stores across departments, with additional information such as store types, sizes, and whether the week contained a major holiday. It includes multiple years of data and is useful for building time series and machine learning models at both the store and department levels.
   Source: https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data

2. **Groceries Dataset** – A transactional dataset capturing individual market-basket level purchases, allowing exploration of purchasing frequency and item association. This dataset will help provide a product-level view of consumer behavior.
   Source: https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset

3. **Retail Product Sales Forecasting Dataset** – A clean, structured dataset ideal for early model testing. It includes historical sales data across products, customers,

and time, with region-level segmentation.

Source: https://www.kaggle.com/datasets/kyanyoga/sample-sales-data

Additional data, such as holiday calendars, Google Trends, or weather data, may be integrated to improve model accuracy and capture exogenous influences on consumer behavior.

Methods

Initial steps will focus on cleaning and preprocessing the data, ensuring that time stamps, categorical variables, and missing values are handled appropriately. I'll explore the data visually and statistically to identify seasonal cycles, trends, and patterns of interest. From there, time series models such as ARIMA, SARIMA, and Prophet will be used to generate initial forecasts. In parallel, I will experiment with supervised machine learning models, including Random Forest Regression, XGBoost, and potentially LSTM neural networks, to compare performance across techniques.

Model evaluation will be conducted using standard metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics will help quantify forecast accuracy and guide decisions on model selection. Visualizations will play a central role in communicating insights and results, with final outputs presented in a PowerPoint deck for stakeholders.

Ethical Considerations

Since all datasets are publicly available and anonymized, privacy concerns are minimal. However, one ethical consideration is transparency in the interpretation and communication of forecasting results. Retail demand is influenced by many factors, and models—particularly complex ones—may overfit or underperform if not validated rigorously. Presenting forecast results without accounting for uncertainty could lead to misguided business decisions. It's important to include disclaimers about model confidence and limitations. Additionally, retailers should not use demand forecasts unethically—for instance, to justify artificial scarcity or manipulate pricing beyond reasonable business practices.

Challenges/Issues

Several challenges may arise during the course of this project. First, aligning multiple datasets with different formats, time intervals, and levels of granularity (e.g., weekly vs. daily, item-level vs. store-level) will require careful preprocessing. Second, missing data and inconsistent records may impact model quality and require imputation strategies. Third, incorporating external variables like holidays or promotions can add complexity but also improve accuracy—balancing this trade-off will be important. Lastly, selecting the right modeling approach for each dataset and avoiding overfitting will be critical as some models may perform better on highly seasonal data than others.

References

- Kaggle. (n.d.). Walmart Store Sales Forecasting. https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting
- Kaggle. (n.d.). Groceries Dataset. https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset

- Kaggle. (n.d.). Sample Retail Sales Data.

  https://www.kaggle.com/datasets/kyanyoga/sample-sales-data

- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*.

  https://otexts.com/fpp3/

- Google Trends. (n.d.). https://trends.google.com/