



Walmart Sales Analysis and Forecasting

Capstone Project

Subhadip Hensh

Subhadiphensh31102000@gmail.com

Table of Contents

1. Problem Statement
2. Project Objective
3. Data Description
4. Data Pre-processing Steps and Inspiration
5. Choosing the Algorithm for the Project
6. Motivation and Reasons For Choosing the Algorithm
7. Assumptions
8. Model Evaluation and Techniques
9. Inferences from the Same
10. Future Possibilities of the Project
11. Conclusion
12. References

1. Problem Statement

Managing retail inventory is all about finding the right balance between supply and demand. For a large retailer like Walmart, which runs hundreds of stores across the country, this task becomes even more complex due to differences in regional demand, seasonal changes, and broader economic conditions. When demand forecasts miss the mark, it can result in overstocking, stock shortages, and wasted resources - all of which directly affect both profit margins and customer satisfaction.

The dataset used for this project, **walmart.csv**, includes weekly sales data from 45 different Walmart stores. It consists of **6,435 records** and **eight key features**: store number, date, weekly sales, holiday flag, temperature, fuel price, consumer price index (CPI), and unemployment rate. The main goal is to analyze these patterns and develop predictive models that can accurately forecast future sales.

The problem can be broken down into two main parts:

1. Exploratory Analysis

- Understand how macroeconomic indicators like unemployment rate, CPI, and fuel price influence weekly sales.
- Study how temperature and holidays affect store performance.
- Identify seasonal sales trends and uncover their timing and possible causes.
- Rank stores based on their past performance and measure the gap between the top and bottom performers.

2. Forecasting

- Build reliable predictive models that can forecast each store's weekly sales for the next 12 weeks.
- Compare different modeling techniques, including statistical, machine learning, and deep learning approaches.

This project tackles a **real-world retail challenge** with meaningful business impact. By combining exploratory data analysis with advanced forecasting methods, the goal is to provide Walmart with actionable insights to improve inventory planning and strategic decision-making. The complexity of this problem comes from dealing with varied store behaviors, external economic factors, and changing time-based patterns - making it a rich and challenging task in the world of **time series forecasting and retail analytics**

2. Project Objective

The main objective of this project is to create a robust forecasting framework that can accurately predict weekly sales for each of Walmart's retail outlets. By analyzing past sales data and integrating key external factors, the project aims to help Walmart make smarter decisions in areas like inventory management, resource distribution, and strategic planning.

The project revolves around two core goals:

1. **Analytical Exploration :**

The first goal focuses on performing an in-depth exploratory analysis of the dataset to uncover meaningful trends and relationships. This involves studying how factors such as unemployment rate, consumer price index (CPI), fuel price, and temperature influence weekly sales. The analysis also looks into seasonal fluctuations, the impact of holidays, and performance differences among stores. These findings are expected to give Walmart a clearer picture of what drives changes in sales patterns across its network.

2. **Predictive Modeling :**

The second goal is to design and evaluate a variety of forecasting models that can predict future sales with high precision. This includes using machine learning algorithms like **XGBoost**, statistical models such as **ARIMA** and **SARIMA**, time series frameworks like **Prophet**, and deep learning models based on **Long Short-Term Memory (LSTM)** networks. Each model is tested for its ability to generalize across stores, capture time-dependent behavior, and produce reliable forecasts for the next 12 weeks.

By blending analytical insights with advanced predictive modeling, the project aims to pinpoint the most effective method for forecasting store-wise sales. The end result will be a collection of actionable insights and accurate predictions that can enhance Walmart's operational efficiency and empower data-driven decision-making across its retail ecosystem.

3. Data Description

The dataset used in this project, titled walmart.csv, contains weekly sales records from 45 Walmart stores across the United States. It comprises 6,435 rows and 8 columns, capturing a range of variables that influence retail performance. Each row represents the sales data for a specific store during a particular week, along with associated economic and environmental indicators.

Dataset Overview:

Column Name	Description
Store	Unique identifier for each Walmart outlet
Date	Week-ending date for the sales record
Weekly_Sales	Total sales for the store during that week
Holiday_Flag	Binary indicator (1 = holiday week, 0 = non-holiday week)
Temperature	Recorded temperature in the region during the week
Fuel_Price	Average fuel price in the region during the week
CPI	Consumer Price Index, reflecting inflationary trends
Unemployment	Unemployment rate in the region during the week

Key Characteristics:

- **Temporal Coverage:** The data spans multiple years, allowing for seasonal and trend analysis.
- **Granularity:** Weekly frequency enables fine-grained forecasting and short-term planning.
- **Store-Level Detail:** Each store is treated as a unique entity, allowing for localized modeling.
- **External Variables:** Inclusion of CPI, fuel price, temperature, and unemployment provides context for macroeconomic impact analysis.

This dataset offers a rich foundation for both exploratory analysis and predictive modeling. Its structure supports time series techniques, while the presence of external features enables multivariate approaches. The diversity of stores and regional conditions also introduces complexity, making it an ideal candidate for evaluating multiple forecasting strategies.

4. Data Pre-processing Steps and Inspiration

Effective forecasting begins with clean, well-structured data. Given the complexity and variability of Walmart's weekly sales data, a thorough pre-processing pipeline was essential to ensure model reliability and interpretability. This section outlines the key steps taken to prepare the dataset for analysis and modeling, along with the rationale behind each decision.

Data Cleaning and Validation

- **Missing Values:** The dataset was inspected for null entries across all columns. Missing values were handled using appropriate imputation techniques, such as forward fill for time series continuity and median imputation for numerical features.
- **Date Formatting:** The Date column was converted to a standard datetime format to enable time-based operations such as resampling, lag creation, and seasonal decomposition.
- **Duplicate Records:** The dataset was checked for duplicate rows to ensure data integrity. Any redundant entries were removed.

Feature Engineering

To enhance model performance, several new features were derived from the original dataset:

- **Time-Based Features:** Extracted week, month, year, and day of week from the Date column to capture temporal patterns.
- **Lag Features:** Created lag variables (e.g., previous week's sales) to help models learn from recent trends.
- **Rolling Statistics:** Computed moving averages and standard deviations over 4-week and 12-week windows to capture local seasonality and volatility.
- **Holiday Impact:** The Holiday_Flag was retained and cross-referenced with sales trends to assess its influence on store performance.

Store-Level Segmentation

Each store was treated as a unique entity. Data was segmented store-wise to allow for localized modeling and analysis. This approach helped capture regional differences in sales behavior and macroeconomic impact.

Inspiration Behind Pre-processing Choices

The pre-processing strategy was inspired by best practices in time series forecasting and retail analytics. Lag features and rolling statistics are commonly used in demand forecasting to capture short-term dependencies. Time-based decomposition helps isolate seasonal effects, while store-level segmentation allows models to adapt to localized patterns. These techniques were selected to balance interpretability, model readiness, and domain relevance.

5. Choosing the Algorithm for the Project

Given the complexity of retail sales forecasting and the diversity of Walmart's store-level data, a single algorithm was unlikely to capture all relevant patterns effectively. Therefore, this project adopted a multi-model strategy, implementing a range of forecasting techniques from traditional statistical models to advanced machine learning and deep learning architectures. Each algorithm was selected based on its strengths in handling time series data, interpretability, scalability, and ability to model nonlinear relationships.

The following algorithms were chosen and applied:

1. XGBoost Regression :

XGBoost was selected as the initial machine learning model due to its speed, scalability, and ability to handle structured data efficiently. Its gradient boosting framework is well-suited for capturing complex interactions between features, and GPU acceleration was leveraged to optimize training time across multiple stores.

2. ARIMA (AutoRegressive Integrated Moving Average) :

ARIMA was chosen for its simplicity and interpretability in modeling time series data. It is particularly effective for capturing linear trends and autocorrelations. As a baseline statistical model, ARIMA provided a reference point for evaluating more complex approaches.

3. SARIMA (Seasonal ARIMA) :

To account for seasonality in weekly sales, SARIMA was implemented as an extension of ARIMA. It incorporates seasonal differencing and seasonal autoregressive/moving average components, making it more suitable for retail data that exhibits periodic fluctuations.

4. Prophet :

Developed by Facebook, Prophet was selected for its ease of use and ability to automatically detect trend and seasonality. It is designed to handle missing data and holiday effects, making it a practical choice for business forecasting. Prophet also provides interpretable components and uncertainty intervals.

5. LSTM (Long Short-Term Memory) :

LSTM, a type of recurrent neural network, was chosen for its ability to model sequential dependencies and nonlinear patterns in time series data. It was used to forecast 12 weeks ahead by learning from historical sequences. While computationally intensive, LSTM offers flexibility in capturing long-term temporal relationships.

This diverse selection of algorithms allowed for a comprehensive evaluation of forecasting strategies. By comparing their performance across all 45 stores, the project aimed to identify the most reliable and scalable approach for Walmart's operational needs.

6. Motivation and Reasons for Choosing the Algorithm

The choice of forecasting algorithms in this project was guided by a combination of business needs, data characteristics, and modeling goals. Given the complexity of Walmart's store-level sales data - which includes temporal patterns, external economic indicators, and regional variability - it was important to explore a diverse set of approaches that could offer both accuracy and interpretability.

Each algorithm was selected with a specific motivation in mind:

XGBoost Regression : This model was chosen for its ability to handle structured tabular data and capture nonlinear relationships between features. Its gradient boosting framework is known for strong predictive performance, and GPU acceleration allowed for efficient training across multiple stores. XGBoost also supports feature importance analysis, which helped in understanding the relative impact of variables like CPI, temperature, and unemployment.

ARIMA : ARIMA was selected as a baseline statistical model due to its simplicity and interpretability. It is well-suited for time series data with linear trends and autocorrelation. Implementing ARIMA allowed for a foundational understanding of each store's sales behavior and served as a benchmark for more advanced models.

SARIMA : SARIMA extended ARIMA by incorporating seasonal components, making it particularly effective for retail data that exhibits weekly and annual seasonality. The motivation behind using SARIMA was to capture recurring patterns such as holiday effects and seasonal demand shifts, which are common in Walmart's sales cycles. SARIMA ultimately emerged as one of the top-performing models in terms of accuracy and consistency.

Prophet : Prophet was chosen for its ease of use and ability to automatically detect trend and seasonality. It is designed to handle missing data and holiday effects, which aligned well with the structure of the Walmart dataset. Prophet also provides interpretable components and visual diagnostics, making it a practical tool for business stakeholders.

LSTM (Long Short-Term Memory) : LSTM was introduced to explore the potential of deep learning in capturing long-term dependencies and complex temporal patterns. Unlike traditional models, LSTM can learn from sequences of past sales data without requiring manual feature engineering. Although it required more computational resources and tuning, LSTM offered a flexible framework for multistep forecasting.

The motivation behind using this combination of models was to balance predictive power with interpretability, and to compare how different modeling paradigms perform on the same dataset. This approach allowed for a comprehensive evaluation and ensured that the final recommendations were grounded in both technical rigor and practical relevance.

7. Assumptions

Throughout the course of this project, several assumptions were made to streamline the analysis, simplify modeling decisions, and ensure consistency across all 45 Walmart stores. These assumptions were necessary to manage data limitations, maintain computational feasibility, and align with the business context of retail forecasting.

Data-Related Assumptions

- **Completeness of Data:** It was assumed that the provided dataset (walmart.csv) contains all relevant historical sales records and external variables required for forecasting. No additional external datasets were merged.
- **Accuracy of Economic Indicators:** Variables such as CPI, fuel price, temperature, and unemployment rate were assumed to be correctly recorded and representative of the regional conditions affecting each store.
- **Holiday Flag Validity:** The Holiday_Flag was treated as a reliable indicator of whether a given week included a major holiday, with no further validation against external calendars.

Modeling Assumptions

- **Stationarity for ARIMA/SARIMA:** Time series models like ARIMA and SARIMA were applied under the assumption that the data could be made stationary through differencing and seasonal adjustments.
- **Independence Across Stores:** Each store was modeled independently, assuming that sales patterns and external influences vary by location and do not exhibit cross-store dependencies.
- **Feature Stability:** Engineered features such as lag variables and rolling statistics were assumed to remain stable and predictive over time, especially during the forecast horizon.

Forecasting Assumptions

- **Forecast Horizon:** All models were tasked with predicting weekly sales for a fixed 12-week horizon, assuming that this period is sufficient for short-term planning and operational decision-making.
- **No Sudden Structural Changes:** It was assumed that there would be no abrupt changes in consumer behavior, store operations, or external conditions during the forecast period (e.g., policy shifts, economic shocks, or store closures).
- **Model Generalization:** Forecasting models were expected to generalize well from historical patterns to future sales, despite potential noise and variability in the data.

These assumptions were carefully considered during model selection, evaluation, and interpretation. While they introduce certain limitations, they also provide a consistent framework for comparing forecasting approaches and deriving actionable insights.

8. Model Evaluation and Techniques

To assess the performance of each forecasting model, a consistent evaluation strategy was applied across all 45 Walmart stores. The focus was on measuring how well each model predicted weekly sales over a 12-week horizon using historical data. Rather than relying on theoretical metrics alone, the evaluation emphasized practical accuracy and visual alignment with actual sales trends.

Evaluation Metrics Used

- **Root Mean Squared Error (RMSE)** RMSE was used as the primary metric to quantify prediction accuracy. It penalizes larger errors more heavily, making it suitable for retail data where sudden spikes or drops in sales can occur.
- **Visual Comparison of Forecast vs Actual** For each store, forecasted sales were plotted alongside historical sales to visually inspect how well the model captured trends, seasonality, and fluctuations. This approach was especially useful for evaluating models like Prophet and LSTM, where interpretability of patterns mattered as much as numerical accuracy.

Evaluation Approach

- **Store-Wise Forecasting** Each store was modeled and evaluated independently to account for regional differences in sales behavior. This allowed for a granular understanding of model performance across the network.
- **Fixed Forecast Horizon** All models were tasked with predicting the next 12 weeks of sales. This uniform horizon ensured fair comparison and aligned with the business need for short-term planning.
- **Model-Specific Diagnostics**
 - For **SARIMA**, residual plots and seasonal decomposition were used to validate model fit.
 - For **Prophet**, trend and seasonal components were reviewed to assess interpretability.
 - For **LSTM**, forecast sequences were compared against actual sales to evaluate temporal alignment.

This evaluation strategy provided a balanced view of each model's strengths and limitations. SARIMA consistently delivered the most accurate and stable forecasts, followed by Prophet and ARIMA. These results informed the final recommendations for model selection and deployment.

9. Inferences from the Same

After implementing and evaluating multiple forecasting models across all 45 Walmart stores, several key insights emerged regarding sales behavior, model performance, and the influence of external factors. These inferences are based on both quantitative metrics and qualitative observations from visual comparisons.

Store-Level Sales Behavior

- **Seasonality:** Many stores exhibited clear seasonal patterns, with noticeable spikes around major holidays and year-end periods. SARIMA and Prophet were particularly effective in capturing these trends.
- **Holiday Impact:** The presence of holidays, as indicated by the `Holiday_Flag`, often correlated with increased sales. However, the magnitude of this effect varied significantly across stores.
- **Macroeconomic Influence:** Variables such as unemployment rate and CPI showed moderate correlation with sales in certain regions, but their impact was not uniformly strong across all stores.
- **Performance Disparity:** There was a significant gap between the highest and lowest performing stores. Top-performing stores maintained consistent sales volumes, while lower-performing ones showed greater volatility and weaker seasonal patterns.

Model Performance Insights

Based on RMSE scores and visual alignment with actual sales, the top three models were:

1. **SARIMA** – Delivered the most consistent and accurate forecasts across stores. Its ability to model seasonality and trend made it the most reliable choice.
2. **Prophet** – Performed well in capturing seasonal cycles and holiday effects, with minimal tuning and strong interpretability.
3. **ARIMA** – Served as a dependable baseline model, especially for stores with linear trends and minimal seasonal complexity.

Other models like XGBoost and LSTM were explored for their ability to capture nonlinear patterns and sequential dependencies. However, they did not outperform the top three models in terms of overall accuracy and consistency.

General Observations

- No single model was universally superior across all stores. Model selection should consider store-specific patterns and operational needs.
- Visual inspection of forecasts was critical in validating model behavior, especially for identifying overfitting or missed seasonal signals.
- Combining statistical and machine learning approaches provided a more holistic understanding of sales dynamics and forecasting potential.

These inferences guided the final recommendations and highlighted the importance of tailoring forecasting strategies to the unique characteristics of each store and dataset.

10. Future Possibilities of the Project

While this project successfully implemented and evaluated multiple forecasting models for Walmart's store-wise weekly sales, there are several promising directions for future development. These possibilities span technical enhancements, operational deployment, and broader research applications.

1. Model Deployment and Automation

One of the most practical next steps is to deploy the top-performing models (e.g., SARIMA and Prophet) into a production environment. This would allow Walmart to generate rolling forecasts on a weekly basis, integrate predictions into inventory systems, and automate decision-making processes. Building a dashboard or API interface could make forecasts accessible to store managers and supply chain teams in real time.

2. Incorporating External Data Sources

Future iterations of the project could benefit from integrating additional external datasets, such as:

- Regional economic indicators (e.g., consumer confidence, retail footfall)
- Weather forecasts for upcoming weeks
- Promotional calendars and marketing campaigns
- Competitor pricing and activity

These variables could improve model accuracy and provide richer context for sales behavior.

3. Ensemble Modeling

Combining the strengths of multiple models through ensemble techniques could lead to more robust forecasts. For example, blending SARIMA's seasonal precision with Prophet's trend detection or XGBoost's nonlinear capabilities may yield improved performance across diverse store profiles.

4. Store Clustering and Group Modeling

Instead of modeling each store independently, future work could explore clustering stores based on sales patterns, geography, or demographic factors. Group-level models could then be trained for each cluster, reducing computational overhead and potentially improving generalization.

5. Long-Term Forecasting and Scenario Planning

While this project focused on a 12-week horizon, extending forecasts to quarterly or annual periods could support strategic planning. Scenario-based forecasting — such as simulating the impact of economic downturns or policy changes — could further enhance Walmart's resilience and agility.

6. Deep Learning Optimization

Although LSTM did not outperform traditional models in this phase, future work could explore more advanced architectures such as attention-based models (e.g., Transformers), hybrid CNN-LSTM structures, or transfer learning from pre-trained time series models. With sufficient tuning and data augmentation, deep learning may offer competitive performance in more complex forecasting tasks.

These future possibilities reflect the evolving nature of retail analytics and the growing role of data science in operational decision-making. By continuing to refine and expand this forecasting framework, Walmart can move closer to a truly intelligent, data-driven retail ecosystem.

11. Conclusion

This project set out to address a critical operational challenge faced by Walmart: forecasting weekly sales across multiple retail outlets to improve inventory management and demand-supply alignment. Through a structured approach combining exploratory analysis and predictive modeling, the project delivered both insights and actionable forecasts that can support data-driven decision-making.

A variety of forecasting techniques were implemented and evaluated, including machine learning (XGBoost), statistical models (ARIMA, SARIMA), time series frameworks (Prophet), and deep learning (LSTM). Each model was assessed for its ability to capture sales patterns, seasonality, and external influences such as holidays and macroeconomic indicators.

Among the models tested, SARIMA emerged as the most consistent and accurate performer, followed by Prophet and ARIMA. These models demonstrated strong alignment with actual sales trends and required minimal tuning, making them suitable for operational deployment. While LSTM and XGBoost offered flexibility and nonlinear modeling capabilities, they did not outperform the top statistical models in this context.

Beyond forecasting, the project also highlighted key insights into store-level behavior, seasonal effects, and the varying impact of external factors. These findings can inform broader strategic decisions, from promotional planning to regional resource allocation.

Overall, the project showcases the value of combining statistical rigor with modern data science techniques to solve real-world business problems. It lays the groundwork for future enhancements, including model deployment, external data integration, and long-term scenario planning — all aimed at building a smarter, more responsive retail ecosystem.

12. References

1. Walmart Sales Forecasting Dataset. Provided as part of the Capstone Project assignment.
2. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts.
3. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. [Prophet Model Reference]
4. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
5. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
6. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
7. Brownlee, J. (2020). *Deep Learning for Time Series Forecasting*. Machine Learning Mastery.
8. VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media.
9. Official documentation for Prophet: <https://facebook.github.io/prophet/>
10. Official documentation for XGBoost: <https://xgboost.readthedocs.io/>
11. Official documentation for statsmodels (ARIMA/SARIMA): <https://www.statsmodels.org/>
12. Official documentation for Keras (LSTM): <https://keras.io/api/>