# Walmart Sales Forecasting and Time Series Analysis

## 1. Introduction

Retail organizations rely heavily on accurate demand forecasting to manage inventory, workforce planning, and promotional strategies. Walmart, being one of the largest retail chains, experiences strong seasonal and holiday-driven fluctuations in sales. Understanding these patterns is critical for short-term and operational decision-making.

This project focuses on analyzing historical weekly sales data from Walmart stores and building a forecasting model to predict sales for the next 12 weeks. The analysis includes exploratory data analysis (EDA), outlier detection, holiday impact assessment, correlation analysis with economic indicators, and time series forecasting.

---

## 2. Dataset Description

The dataset contains weekly sales data for Walmart stores from February 2010 to October 2012.

**Key characteristics:** - Number of records: 6,435 - Number of stores: 45 - Time period: 05-02-2010 to 26-10-2012

**Variables:** - Store: Store identifier - Date: Week ending date - Weekly_Sales: Sales for the given store and week - Holiday_Flag: Indicates whether the week includes a major holiday - Temperature: Average weekly temperature - Fuel_Price: Fuel price during the week - CPI: Consumer Price Index - Unemployment: Unemployment rate

---

## 3. Data Preprocessing

The dataset was first inspected for structure, data types, and missing values. The Date column was converted from object format to datetime format, and the data was sorted by store and date to ensure correct time series ordering.

No missing values were observed in any of the columns. This allowed the analysis to proceed without the need for imputation or data removal.

---

## 4. Exploratory Data Analysis (EDA)

### 4.1 Statistical Summary

Descriptive statistics showed that weekly sales vary significantly across stores, with values ranging from approximately 210,000 to over 3.8 million. The mean weekly sales across all stores was approximately 1.05 million.

The distribution of weekly sales exhibited right skewness, indicating the presence of high-value observations.

### 4.2 Outlier Analysis

Histogram and boxplot visualizations revealed several high-value outliers in weekly sales. These outliers were primarily associated with holiday weeks and large stores.

These values were retained in the dataset, as they represent genuine business events rather than data errors.

---

## 5. Holiday Impact Analysis

To understand the effect of holidays on sales, average weekly sales were compared between holiday and non-holiday weeks.

- Average non-holiday weekly sales: ~1.04 million
- Average holiday weekly sales: ~1.12 million

Holiday weeks generated approximately 8% higher sales on average. This confirms that holidays play a significant role in driving retail demand and should be considered during forecasting and planning.

---

## 6. Correlation Analysis

Correlation analysis was conducted between weekly sales and economic variables such as temperature, fuel price, CPI, and unemployment.

Results indicated weak correlations between weekly sales and these external factors. Weekly sales showed slight negative correlation with unemployment and CPI, while temperature and fuel price had minimal impact.

This suggests that Walmart sales are primarily influenced by store-level and seasonal factors rather than macroeconomic indicators.

---

## 7. Time Series Analysis

Store-wise time series plots revealed clear seasonal patterns in weekly sales, with recurring peaks during year-end holiday periods. These patterns confirmed the suitability of seasonal time series models for forecasting.

For modeling purposes, Holt-Winters Exponential Smoothing was selected due to its ability to capture both trend and seasonality in weekly data.

---

## 8. Sales Forecasting

Holt-Winters Exponential Smoothing models were built individually for each store using weekly sales data. A seasonal period of 52 weeks was used to represent annual seasonality.

Using the trained models, sales forecasts were generated for the next 12 weeks for all 45 stores. The forecasts indicated increased sales during upcoming holiday periods, particularly toward the end of the year.

These forecasts provide actionable insights for short-term demand planning, inventory management, and workforce allocation.

---

## 9. Conclusion

This project successfully analyzed Walmart weekly sales data and generated store-level forecasts for the next 12 weeks. Exploratory analysis confirmed strong seasonality and holiday-driven demand, while correlation analysis showed limited influence from macroeconomic variables.

The Holt-Winters forecasting approach effectively captured sales trends and seasonal patterns, making it suitable for retail demand forecasting. The results of this study can support Walmart's operational planning, helping to optimize inventory levels, staffing, and promotional strategies.

---

## 10. Tools and Technologies Used

- Python
- Pandas and NumPy for data manipulation
- Matplotlib and Seaborn for visualization
- Statsmodels for time series forecasting

---

## 11. Future Scope

Future enhancements to this project could include: - Incorporating machine learning-based forecasting models - Evaluating model accuracy using backtesting techniques - Including additional external factors such as promotions or regional events - Building automated dashboards for real-time forecasting