**Retail Store Optimization — Walmart Sales Forecasting**

Prepared by: Ganesh Chandra Verma

# Executive Summary

This project focuses on optimizing retail store performance using Walmart’s historical sales data. It involves data analysis, clustering of stores based on performance metrics, and time series forecasting to predict future sales. The findings enable better resource allocation, staffing, and inventory management.

# 1. Introduction

The Walmart Store Sales Forecasting dataset provides historical weekly sales data across multiple stores in the United States. The objective is to analyze sales trends, identify store clusters based on behavior, and forecast upcoming sales using time-series modeling techniques.

# 2. Dataset Details

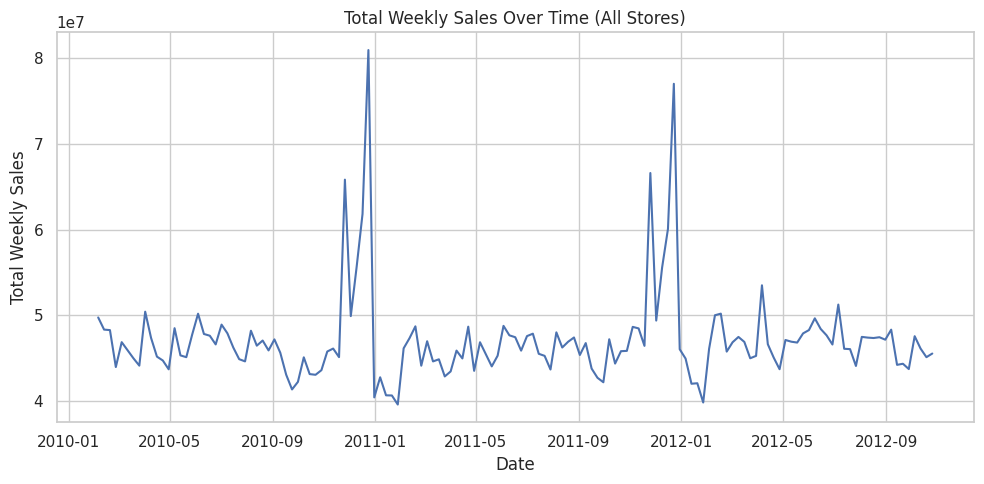
The dataset consists of three main files obtained from Kaggle’s Walmart Recruiting competition:  
• train.csv – Weekly sales data by store, department, and date  
• stores.csv – Metadata about store type and size  
• features.csv – External factors such as temperature, fuel price, CPI, and unemployment rate.  
  
Each dataset was merged on the keys (Store, Date, IsHoliday) to create a unified analytical dataset.

# 3. Exploratory Data Analysis

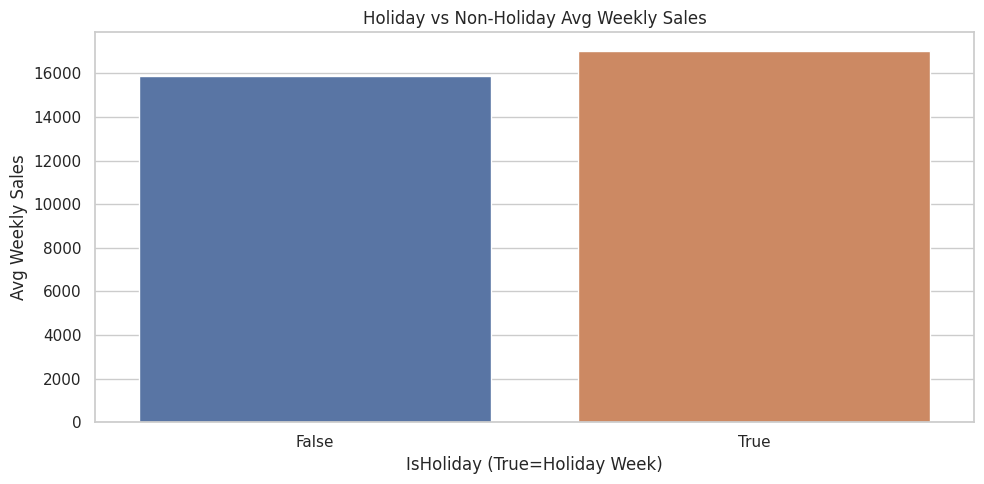
The EDA process revealed significant seasonal patterns and sales spikes around major holidays. Sales trends were analyzed at multiple granularities — across stores, months, and holiday periods.

Key Observations:  
• Strong holiday-driven sales uplift (notably Thanksgiving and Christmas weeks)  
• Store size and type are correlated with overall performance  
• Distinct seasonality with peaks in Q4 months  
• Moderate correlation between temperature and sales variation.

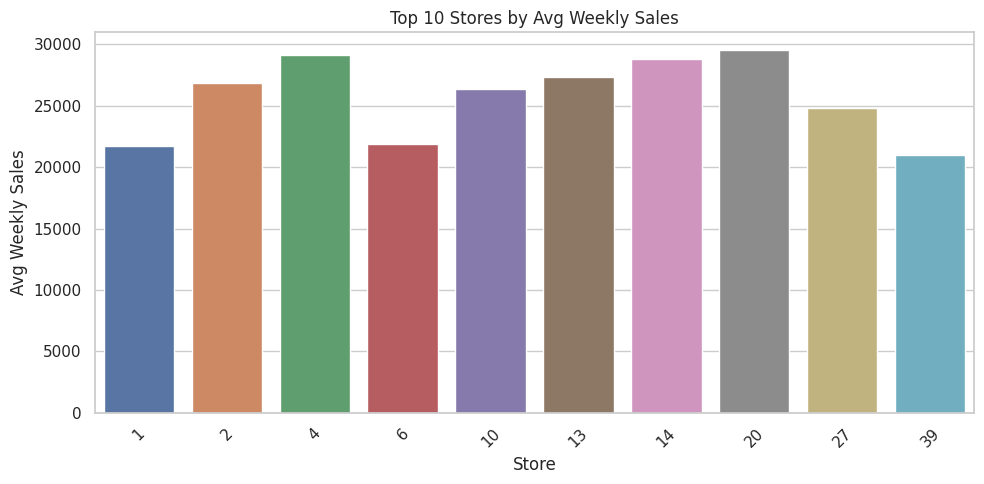
[Figure 1: Total Weekly Sales Over Time]



[Figure 2: Holiday vs Non-Holiday Average Sales]



[Figure 3: Top 10 Stores by Average Sales]



[Figure 4: Sales Seasonality Heatmap]



# 4. Store Clustering using K-Means

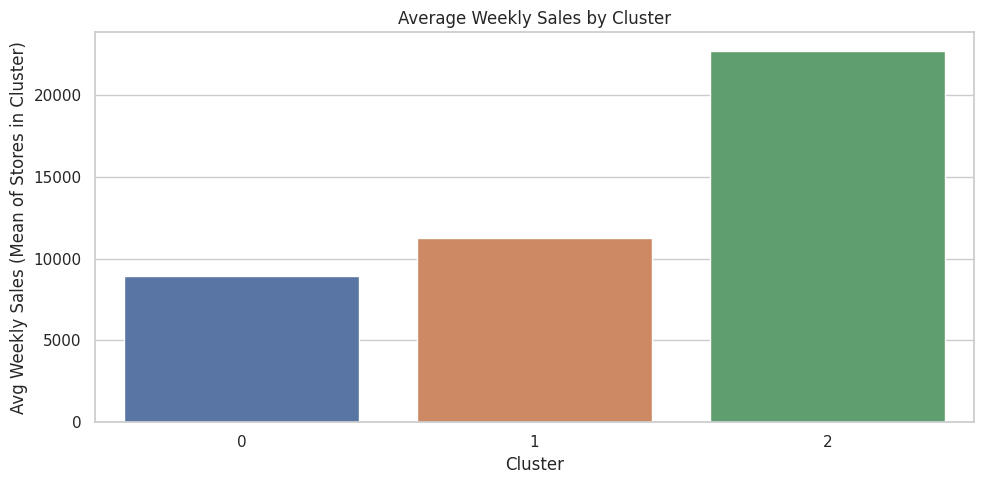
To identify performance patterns, store-level features such as average weekly sales, volatility, holiday uplift, trend slope, and store size were standardized and used for K-Means clustering. The model grouped stores into three distinct clusters.

Cluster Insights:  
• Cluster 0: High-performing stores with strong holiday response  
• Cluster 1: Stable stores with consistent weekly sales  
• Cluster 2: Low-performing or flat-trend stores

[Table 1: Cluster Profile Summary]



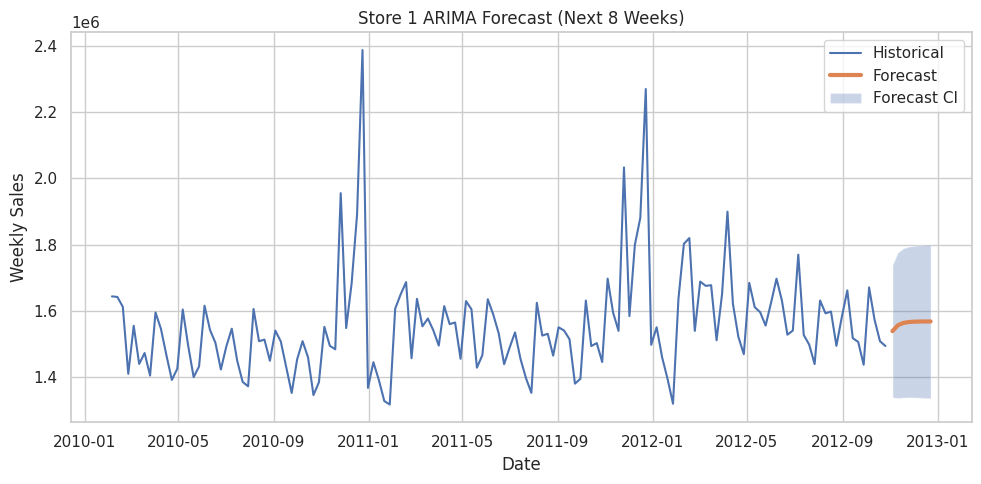
[Figure 5: Average Weekly Sales by Cluster]



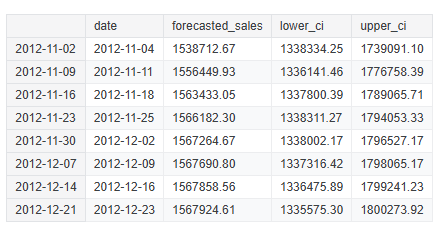
# 5. Time Series Forecasting (ARIMA)

A univariate ARIMA(1,1,1) model was fitted on weekly sales data of a selected store. The model successfully captured trend and seasonality, generating 8-week-ahead forecasts with 80% confidence intervals.

[Figure 6: Store Sales Forecast using ARIMA]



[Table 2: Forecasted Weekly Sales with Confidence Intervals]



# 6. Business Recommendations

Based on the clustering and forecasting analyses, the following business strategies are recommended:  
• Pre-stock seasonal inventory for high-performing clusters (Cluster 0)  
• Maintain consistent promotional efforts for stable-performing stores (Cluster 1)  
• Implement localized marketing and optimize product placement for low-performing stores (Cluster 2)  
• Use forecast data to anticipate demand surges and adjust staffing accordingly.

# 7. Conclusion

The project demonstrates that combining exploratory analysis, clustering, and time-series forecasting provides a holistic view of store performance. These insights support data-driven decision-making for inventory planning, staffing, and promotional strategies in large-scale retail operations.

# 8. Appendix

Tools Used: Python, Pandas, Scikit-learn, Matplotlib, Seaborn, Statsmodels.  
Dataset Source: Walmart Recruiting – Store Sales Forecasting (Kaggle).