Capstone Project Submission Report

# 1. Problem Statement

A retail store chain (Walmart) with multiple outlets across the country is struggling with managing inventory efficiently and matching supply with customer demand. The lack of clear insights into sales behavior across stores and time has led to a financial imbalance. The company aims to use historical data for better understanding and forecasting future sales.

# 2. Project Objective

The objective of this project is to perform data analysis and forecasting using Walmart's weekly sales data. The goals are:  
  
- To identify whether external factors like temperature, unemployment, CPI, and holidays impact sales.  
- To determine best and worst performing stores.  
- To identify seasonal patterns.  
- To forecast weekly sales for selected top-performing stores for the next 12 weeks.

# 3. Data Description

The dataset `walmart.csv` contains 6435 records with the following columns:  
  
- Store: Store number  
- Date: Week of sales  
- Weekly\_Sales: Sales amount  
- Holiday\_Flag: If the week included a holiday (1) or not (0)  
- Temperature: Temperature during the week  
- Fuel\_Price: Fuel cost in that region  
- CPI: Consumer Price Index  
- Unemployment: Unemployment rate in the region

# 4. Data Pre-processing Steps and Inspiration

1. Converted Date to datetime format.  
2. Sorted data by Store and Date.  
3. Created a Month column to study seasonal trends.  
4. Checked for missing values (none found).  
5. Grouped data weekly for time series modeling.

# 5. Choosing the Algorithm for the Project

To forecast future weekly sales, the Holt-Winters Exponential Smoothing method was chosen. This algorithm can capture trend and seasonality in time series data and works well with limited historical records.

# 6. Motivation and Reasons For Choosing the Algorithm

- Easy to implement and interpret  
- Efficient in modeling seasonal data  
- Handles both trend and seasonality  
- Performs well on weekly grouped data

# 7. Assumptions

- Historical sales trends will continue in the near future.  
- External factors like unemployment and CPI indirectly influence sales.  
- Weekly granularity is sufficient to model store-level trends.

# 8. Model Evaluation and Techniques

- Correlation analysis was used to understand the influence of variables:  
 - Sales vs Unemployment: -0.1062  
 - Sales vs CPI: -0.0726  
 - Sales vs Temperature: -0.0638  
- Monthly sales showed clear seasonality, peaking in Nov-Dec.  
- Store-wise total sales used to identify top and bottom performers.  
- Holt-Winters used to forecast next 12 weeks for Stores 20, 4, and 14.

# 9. Inferences from the Same

- Sales drop slightly with rising unemployment, CPI, and temperature.  
- November and December show the highest average sales.  
- Store 20 is the top-performing store; Store 33 is the lowest.  
- There is a performance gap of over ₹26.4 crore between best and worst stores.  
- Forecasts suggest post-holiday dip and stable recovery.

# 10. Future Possibilities of the Project

- Implement advanced models like Facebook Prophet or LSTM.  
- Include promotions, discounts, and local events as variables.  
- Deploy Power BI dashboards for real-time sales tracking.  
- Expand forecasting to all stores.

# 11. Conclusion

The analysis revealed critical insights about sales behavior, store performance, and seasonal trends. Forecasting helped predict weekly sales for selected stores with good accuracy. These insights can assist in better inventory planning and operational decisions.

# 12. References

- Walmart Sales Dataset  
- Python (pandas, matplotlib, statsmodels)  
- Holt-Winters Method (Time Series Forecasting)  
- Capstone Problem Statement from Intellipaat