

# Capstone Project Report: Retail Sales Forecasting

## Problem Statement

A retail store with multiple outlets across the country is facing challenges in managing inventory effectively to align demand with supply. The objective is to derive insights from historical sales data and develop predictive models to forecast sales for the upcoming 12 weeks.

## Project Objective

The primary goal is to analyze sales data from various stores and create a forecasting model that can predict weekly sales. This will assist in inventory management, ensuring that each outlet can meet customer demand without overstocking or understocking.

## Data Description

The dataset `walmart.csv` contains 6,435 rows and 8 columns, representing sales data across various stores. The features are as follows:

- Store: Store number
- Date: Week of sales
- Weekly\_Sales: Sales for the given store in that week
- Holiday\_Flag: Indicates if it is a holiday week (1 for yes, 0 for no)
- Temperature: Temperature on the day of sale
- Fuel\_Price: Cost of fuel in the region
- CPI: Consumer Price Index
- Unemployment: Unemployment Rate

## Data Pre-processing Steps and Inspiration

Data pre-processing steps include:

- Handling Missing Values: Identify and fill or remove missing entries.
- Date Formatting: Convert date strings to datetime objects for easier manipulation.
- Feature Engineering: Create additional features such as 'Year', 'Month', and 'Day of Week' from the date.
- Normalization/Standardization: Scale features like `Temperature`, `Fuel_Price`, and `CPI` to improve model performance.

Inspiration for these steps comes from standard practices in time series analysis and machine learning, focusing on enhancing data quality and model accuracy.

## Choosing the Algorithm for the Project

The forecasting model will utilize the ARIMA (AutoRegressive Integrated Moving Average) algorithm due to its effectiveness in handling time series data. ARIMA is suitable for datasets with trends and seasonality, making it ideal for retail sales forecasting.

## Motivation and Reasons For Choosing the Algorithm

ARIMA was chosen because:

- It can model complex seasonal patterns inherent in retail sales data.
- It provides a straightforward approach to forecast future points based on past values.
- The algorithm's flexibility allows for adjustments based on the characteristics of the dataset.

## Assumptions

The following assumptions are made:

- Historical sales data is representative of future trends.
- External factors (like economic changes) are consistent with historical patterns.
- Seasonal effects will repeat similarly in future periods.

## Model Evaluation and Techniques

Model evaluation will be conducted using:

- Mean Absolute Error (MAE): To measure average prediction error.
- Root Mean Squared Error (RMSE): To assess how well the model predicts actual outcomes.
- Visual Inspection: Plotting predicted vs. actual sales to visually assess model performance.

## Inferences from the Same

Initial analysis indicates that holiday weeks significantly boost sales, while temperature variations have a moderate impact. The model's predictions suggest potential spikes in sales during holiday seasons, guiding inventory decisions.

## Future Possibilities of the Project

Future enhancements could include:

- Integrating additional datasets such as local events or promotions to refine forecasts.
- Implementing machine learning models like XGBoost or LSTM for potentially improved accuracy.
- Developing a dashboard for real-time sales tracking and inventory management.