

**PROBLEM**

**STATEMENT**



- **The retail store (WALMART) is facing issues in managing inventory to match demand with supply across multiple outlets in the country.**

# Project Objective:

- The objective is to analyze historical sales data and build a prediction model to forecast sales for the next 12 number of months/years.





# Data Description:

- The dataset contains the following columns:
- Store: Store number
- Date: Date of sales
- Weekly\_Sales: Sales for the week
- Holiday\_Flag: Binary indicator if the week contains a holiday (1) or not (0)
- Temperature: Temperature of the week
- Fuel\_Price: Fuel price of the week
- CPI: Consumer Price Index of the week
- Unemployment: Unemployment rate of the week

# Data Pre-processing Steps and Inspiration:

- Checked for null values: No null values found.
- Checked statistics of columns: Provides an overview of the numerical columns.
- Checked data types: Date column needs to be converted to datetime format.
- Outlier Detection: Visualized boxplots for numerical columns, decided not to remove outliers as they might contain valuable information.
- Analyzed sales by stores: Identified stores with higher and lower sales.
- Converted date to datetime format and set it as the index.



# Choosing the Algorithm for the Project:

- ARIMA (AutoRegressive Integrated Moving Average) model is chosen for time series forecasting.



# Motivation and Reasons For Choosing the Algorithm:

- ARIMA is a popular choice for time series forecasting, especially when the data exhibits a clear trend or seasonality.



# Assumptions:

- The assumption is that historical sales data contains enough information to capture patterns and trends for future sales prediction.



# Model Evaluation and Techniques:

- The model is trained on the training data and evaluated using test data. The performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) can be used for evaluation.



# Inferences from the Same:

- The ARIMA model is trained on 80% of the data and forecasted for the next 12 months.
- The forecasted sales for the next 12 months are printed.



```
# Print forecast for next 12 months  
print(forecast)
```

```
5148    279336.047953  
5149    281063.515640  
5150    281499.371111  
5151    281609.341325  
5152    281637.087788  
5153    281644.088468  
5154    281645.854802  
5155    281646.300464  
5156    281646.412908  
5157    281646.441279  
5158    281646.448437  
5159    281646.450243  
Name: predicted_mean, dtype: float64
```

# Future Possibilities of the Project:

- The model's performance can be further improved by fine-tuning the parameters of the ARIMA model.
- Other advanced time series forecasting techniques such as SARIMA, Prophet, or LSTM can be explored for potentially better performance.
- Integration with other external factors such as marketing campaigns, competitor data, etc., can enhance the accuracy of the forecast.



# Conclusion:

- The initial implementation using ARIMA shows promising results in forecasting sales for the retail store. Further optimization and integration with additional factors can improve the accuracy of the forecasts.

# References:

- Documentation of Python libraries used (pandas, numpy, matplotlib, seaborn, statsmodels, sklearn)
- Time series forecasting literature for understanding ARIMA and other techniques.



# Vote of Thanks

- Thank you for your valuable Time

