

## Capstone Project – Walmart

<b>Sr. No.</b>	<b>Contents</b>
1	Problem Statement
2	Problem Objective
3	Data Description
4	Data Pre-processing steps and Inspiration
5	Choosing the Algorithm for the Project
6	Motivation and Reasons for Choosing the Algorithm
7	Assumptions
8	Model Evaluation and Techniques
9	Inferences from the Same
10	Future Possibilities of this Project
11	Conclusion
12	References

# 1. Problem Statement :

A retail store that has multiple outlets across the country are facing issues in managing the inventory - to match the demand with respect to supply. You are a data scientist, who has to come up with useful insights using the data and make prediction models to forecast the sales for X number of months/years.

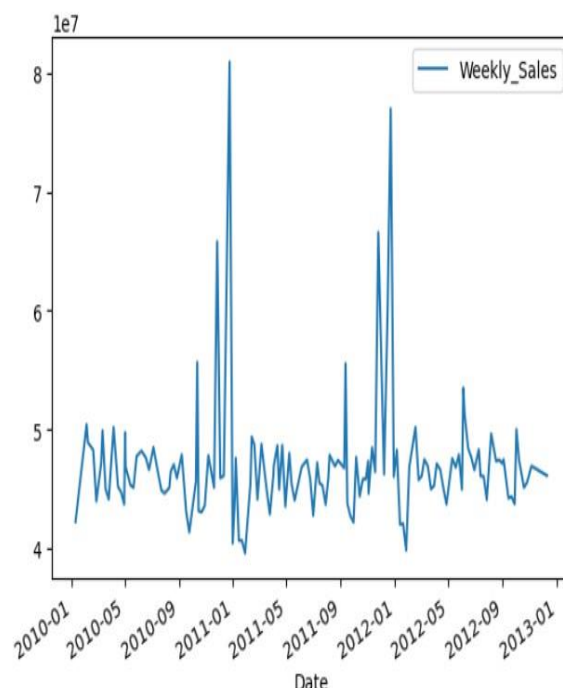
## Problem Statement

```
In [74]: #reading the Walmart.csv file
df = pd.read_csv('Walmart.csv')
df['Date'] = pd.to_datetime(df['Date'])
df.groupby('Date')[['Weekly_Sales']].sum().sort_values(by='Date', ascending=True).plot()
```

C:\Users\anilk\AppData\Local\Temp\ipykernel\_24868\1308824369.py:3: UserWarning: Parsing dates in DD/MM/YYYY format when dayfirst=False (the default) was specified. This may lead to inconsistently parsed dates! Specify a format to ensure consistent parsing.

```
df['Date'] = pd.to_datetime(df['Date'])
```

Out[74]: <Axes: xlabel='Date'>



## **2. *Project Objective***

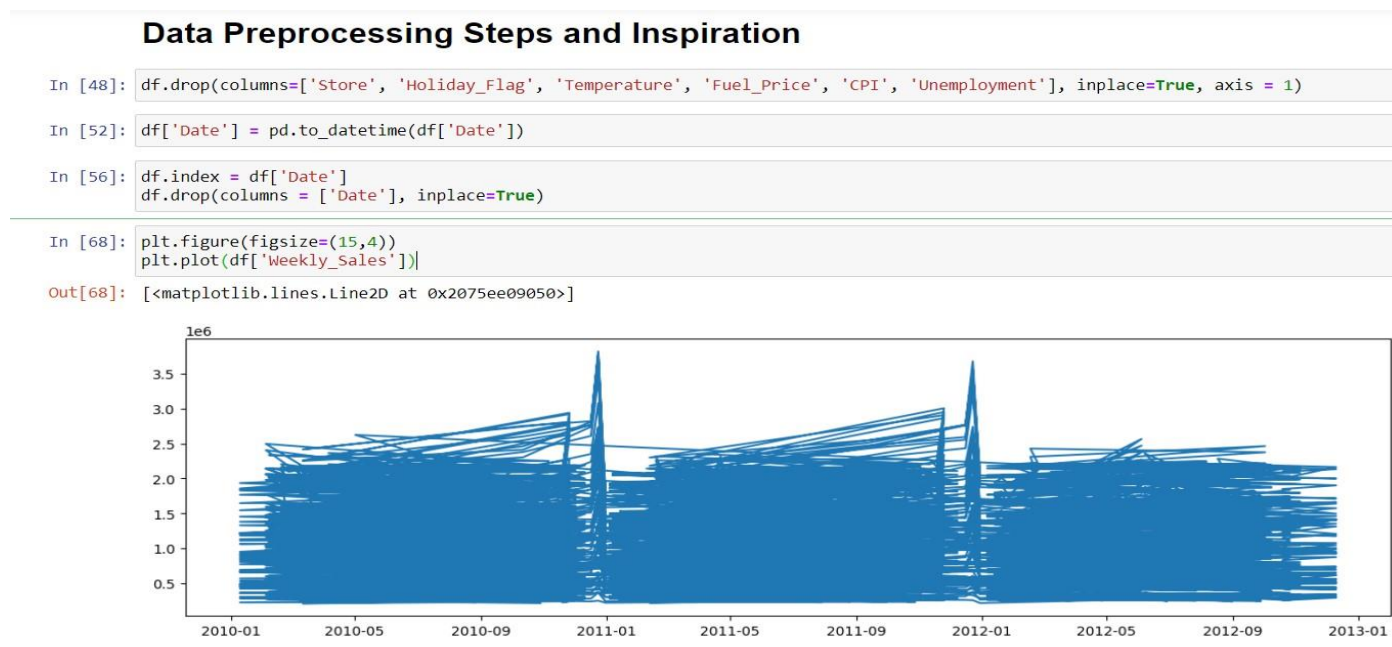
The objective of this project is to help my manager get forecasting insights of weekly sales onwards 2013-01 to at least 1 year, because the company want to sale unsold product in the start of the year. They have underutilized or undersold items that they want to push to shelves and sell it as soon as possible and they need forecast so that production team can cut down the production based on customer traffic.

### **3.    *Data Description***

1. `df.isnull().sum()`
2. `df.duplicated().sum()`
3. `df.info()`
4. `df.describe().T`
5. `df.shape`

## 4. Data Preprocessing Steps and Inspiration

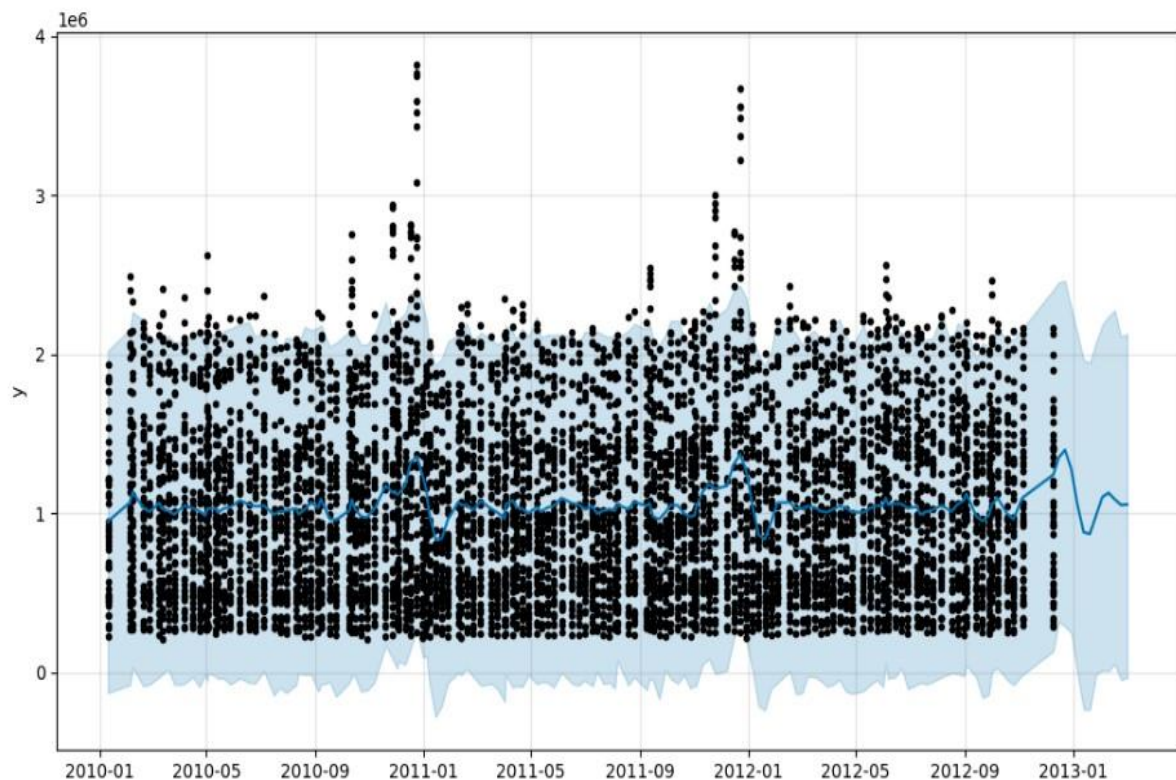
1. `df.drop(columns=['Store', 'Holiday_Flag', 'Temperature', 'Fuel_Price', 'CPI', 'Unemployment'], inplace=True, axis = 1)`
2. `df['Date'] = pd.to_datetime(df['Date'])`
3. `df.columns = ['ds', 'y']`



**Insights :** Last month and first month of the every year, i.e. December-January there is a surge on weekly sales. This seasonal pattern indicates a festival celebration, i.e. Christmas season. Once the Christmas is over, the weekly sales seems to take a dip in sales.

## 5. *Choosing The Algorithm For The Project*

```
In [94]: from prophet import Prophet  
  
In [95]: model = Prophet(interval_width=0.95)  
  
In [ ]: model.fit(df)  
  
In [143]: future_dates = model.make_future_dataframe(periods=12, freq='W')  
  
In [144]: forecast = model.predict(future_dates)  
  
In [145]: sales_forecast = model.plot(forecast)
```



Since my manager wants me to forecast weekly sales for 12 weeks and we have date and weekly sales which makes the strong candidates to tell it's a 'time series' problem, hence used time series algorithm.