KUBIXSQUARE JULY INTERNSHIP

TEAM: AI ML 07

Final Report

Topic: Walmart Sales Prediction for Future Campaigns

Team Members:

Hrutik Naik

Shrumi Dedhia

Shravani Dhuri

Sneha Sable

Vikram Rao

Date of Submission: 19th September 2021

Abstract:

Many household products are sold by various branches of the retail store network, which are geographically located at multiple locations. Supply chain inefficiencies will occur at different places when the market potential is not evaluated by the retailers correctly. Many times it is not easy for retailers to understand the market condition at various locations. The organization of the retail store network has to understand the market conditions to raise its goods to be bought and sold so that they are able to attract many customers. Business forecast helps retailers to visualize the forecasting of the sales to get a general idea of the coming years and if any changes are needed then those changes are done in the retail store's goal so as to optimize the success. The forecasting of sales helps to know the retailers the demand for the products.

Problem Statement:

The model is provided with historical sales data for Walmart stores located in different regions. Each store contains several departments, and the main task of the model is predicting the department-wide sales for each store. The model tries to visualize and identify any trend useful for the company. This model is implemented for predicting the sales of the most/least sold product for future campaigns and to optimize the sales.

Requirements:

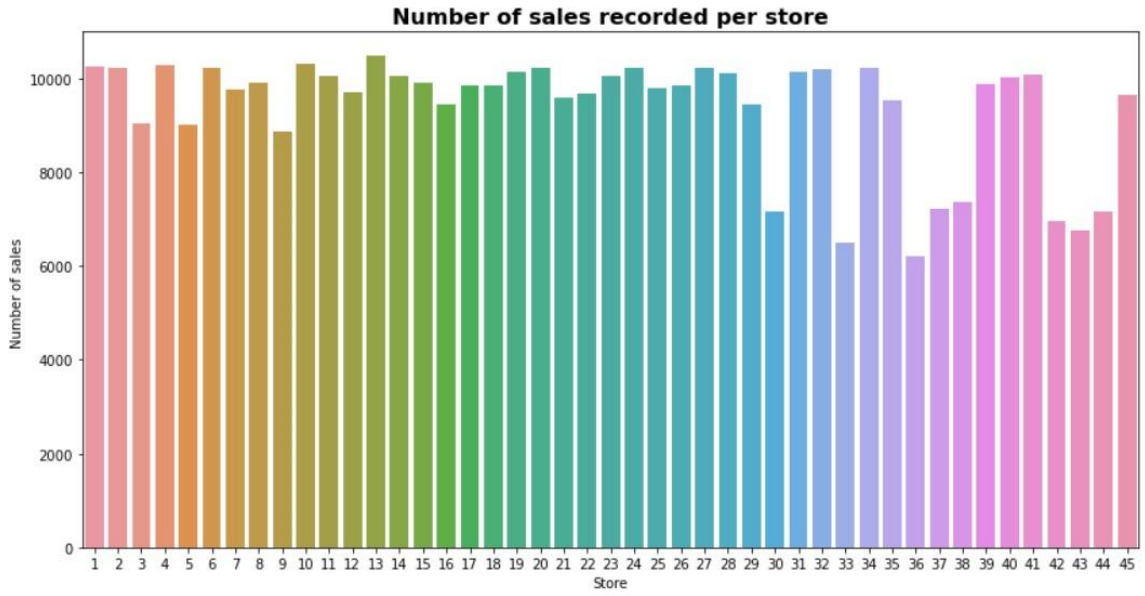
- Jupyter
- Pandas
- Matplotlib
- Seaborn
- Scikit-Learn
- Streamlit
- Numpy
- Plotly
- Pydeck
- Statsmodels
- Python 3.6 or higher

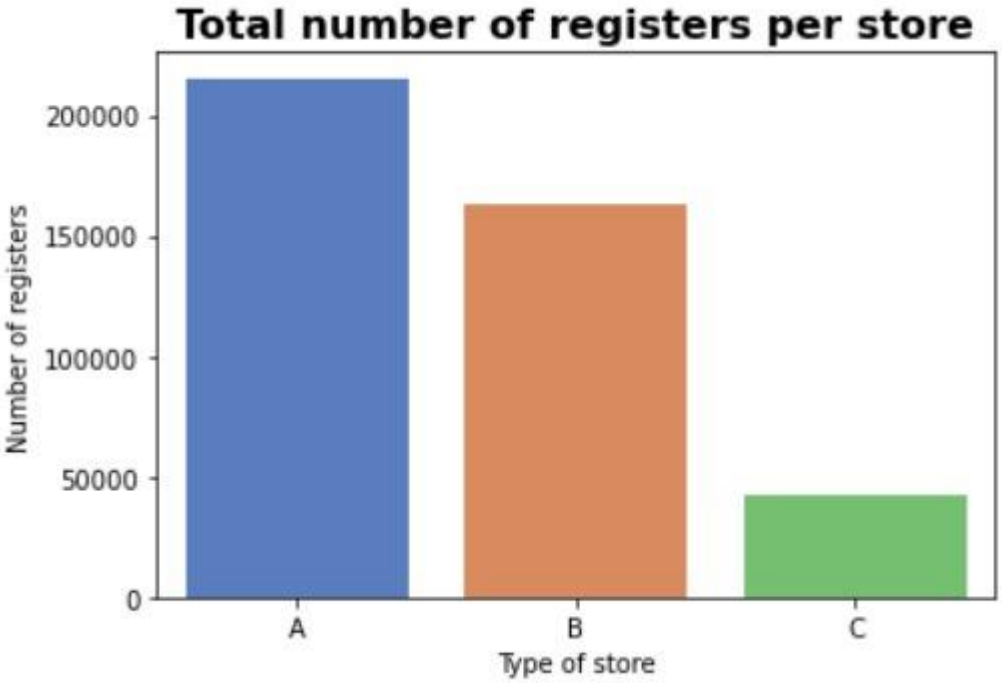
Algorithms Used:

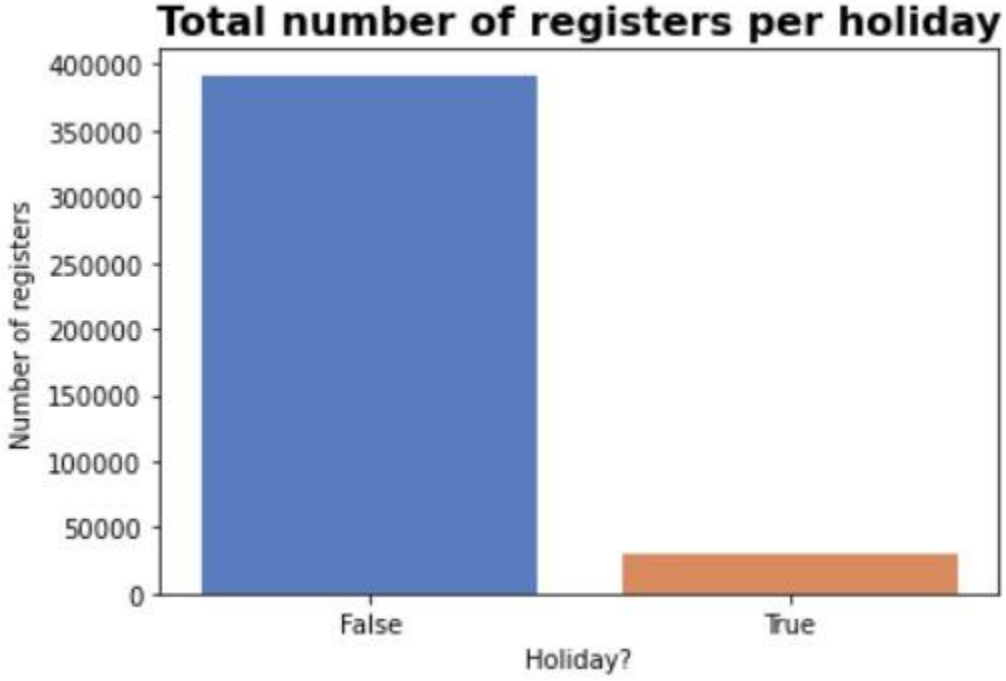
- Linear Regression
- KNN Regressor
- Decision Tree Regressor
- Sklearn
- Random Forest Regressor
- Search Grid HyperParams

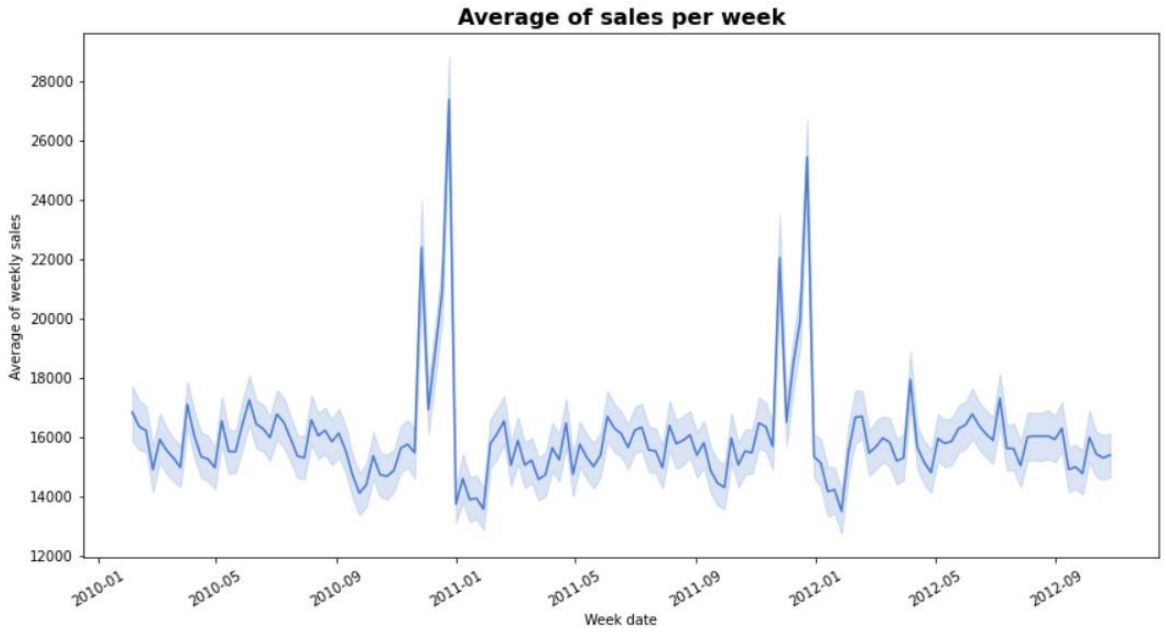
Outputs and Results:

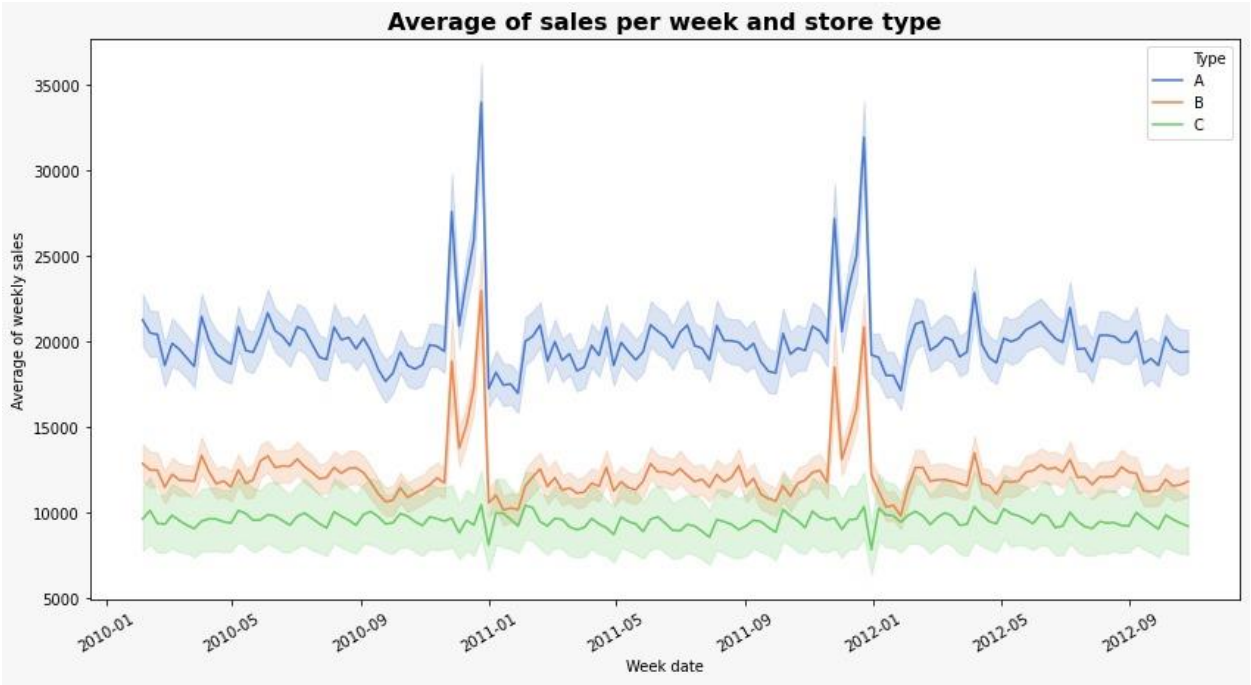
Outputs for Exploratory Data Analysis



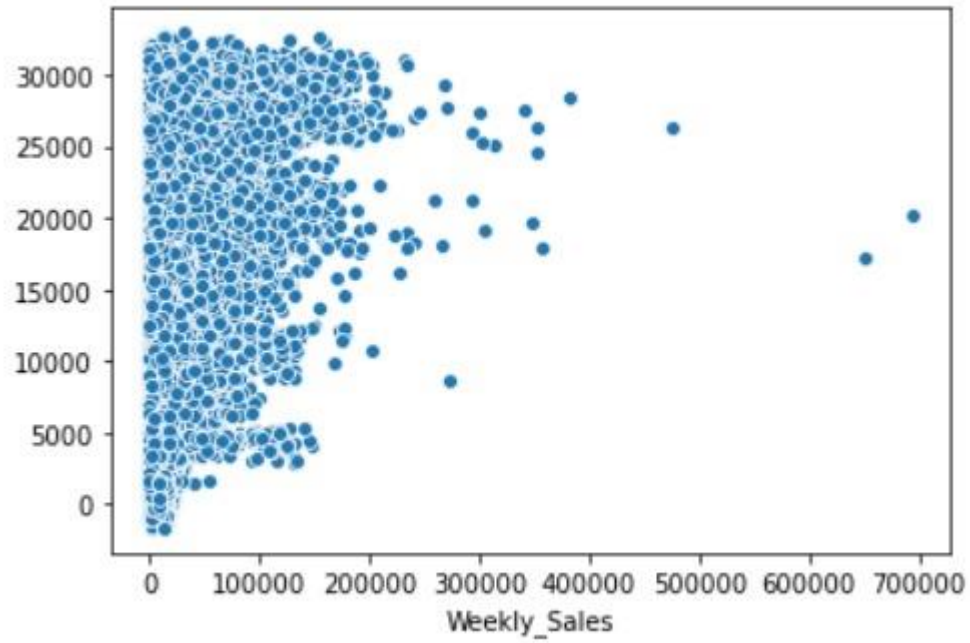

Total number of registers per store
Number of registers
200000
150000
100000
50000
0
A
B
C
Type of store


Total number of registers per holiday
Number of registers
400000
350000
300000
250000
200000
150000
100000
50000
0
False
True
Holiday?

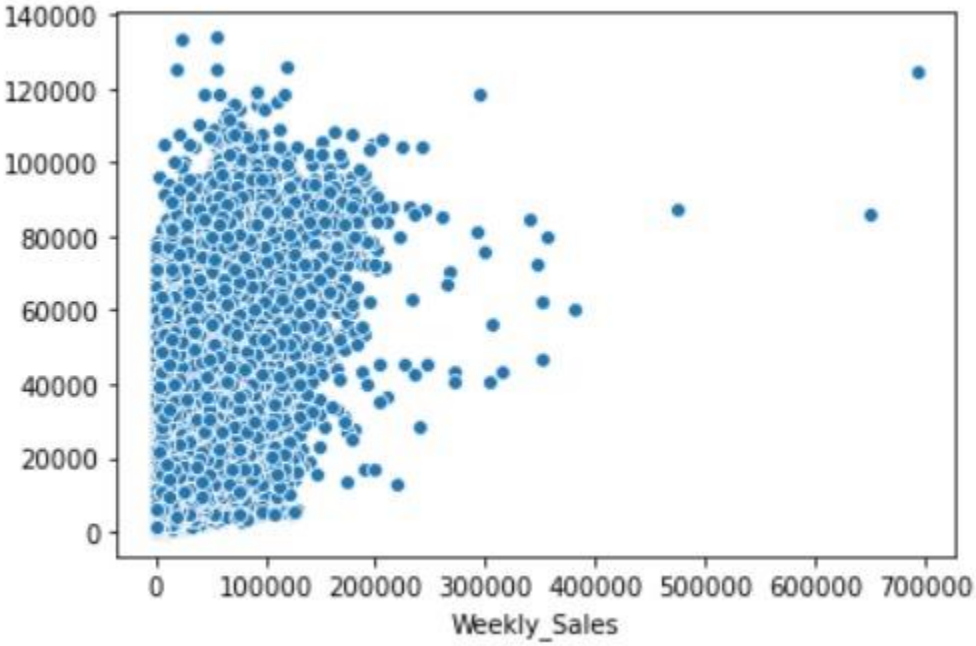

Average of sales per week
Average of weekly sales
28000
26000
24000
22000
20000
18000
16000
14000
12000
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
Week date


Average of sales per week and store type
Type
A
B
C
Average of weekly sales
35000
30000
25000
20000
15000
10000
5000
2010-01
2010-05
2010-09
2011-01
2011-05
2011-09
2012-01
2012-05
2012-09
Week date

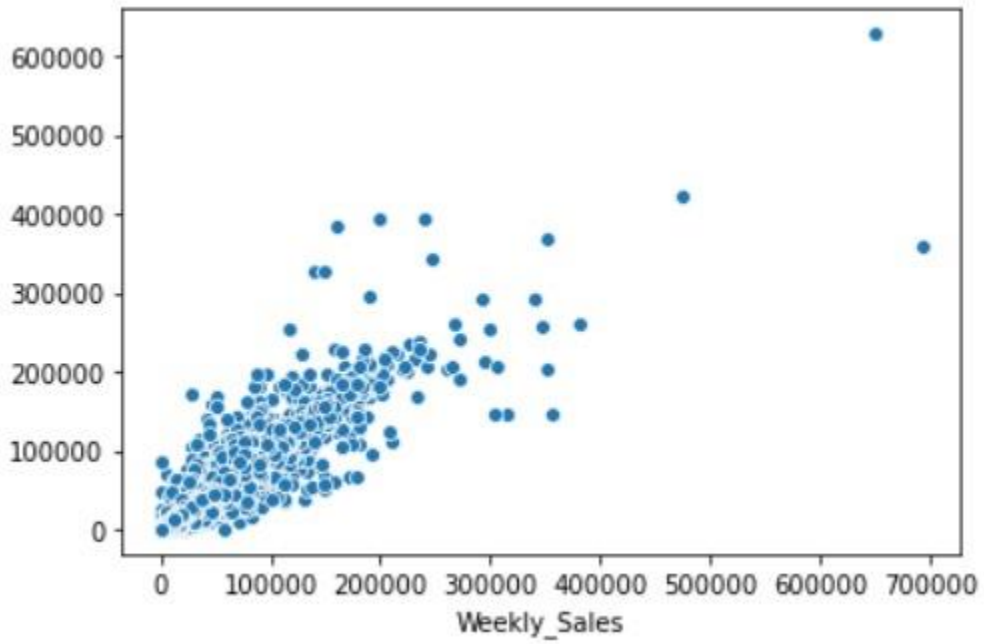
Linear Regression



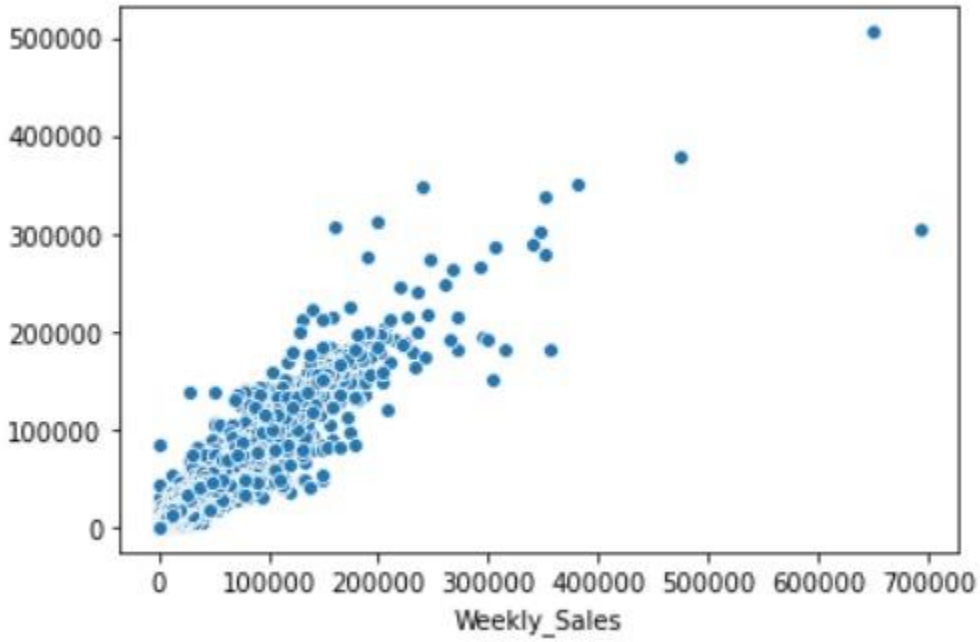
KNN Regressor



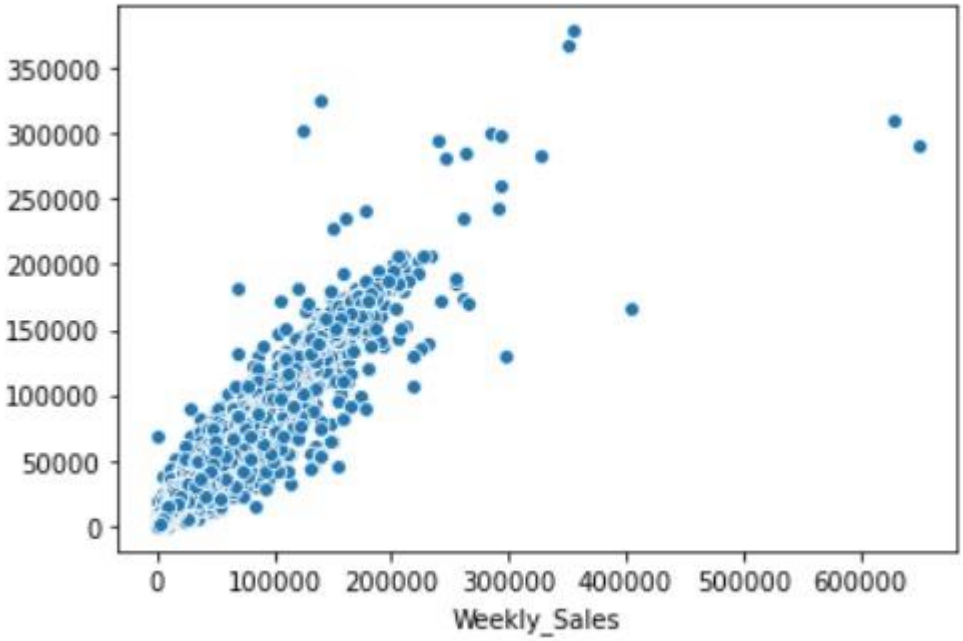
Decision Tree Regressor



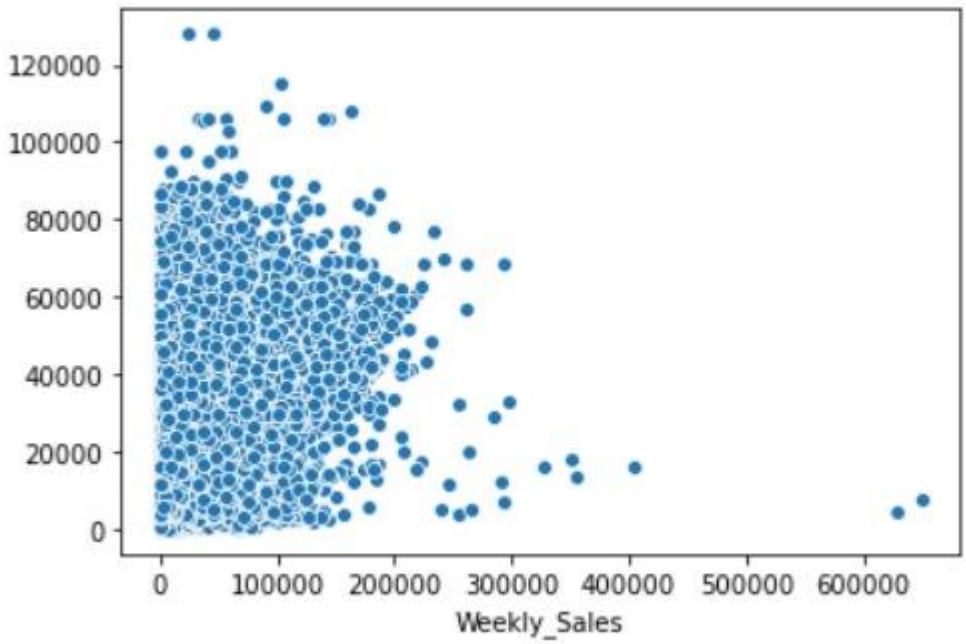
Random Forest Regressor



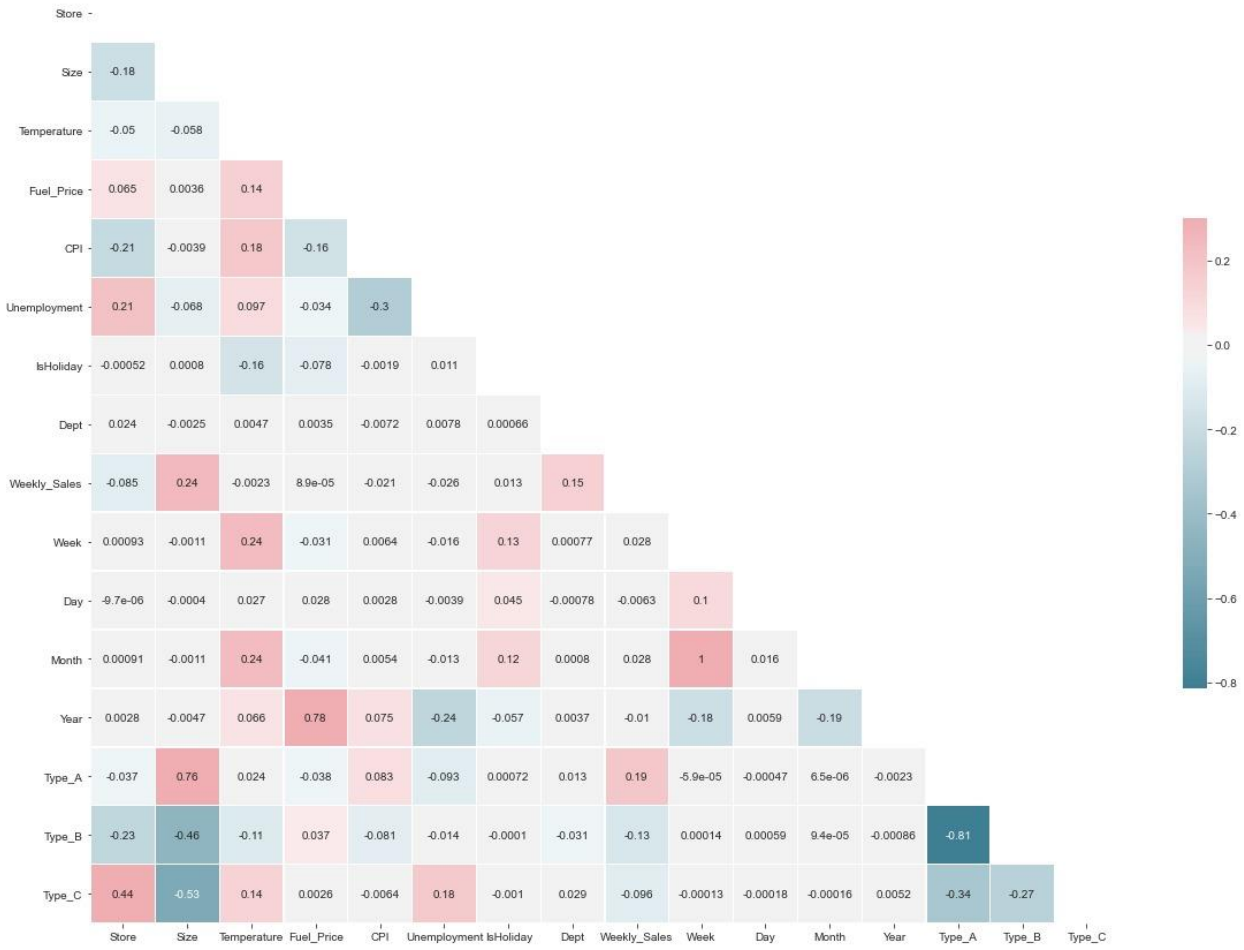
Search Grid HyperParams



KNN Data Scaled



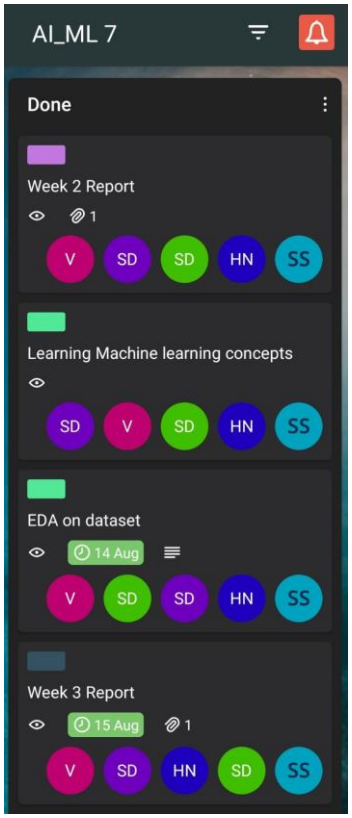
Heat Map Obtained After Cleaning the Data

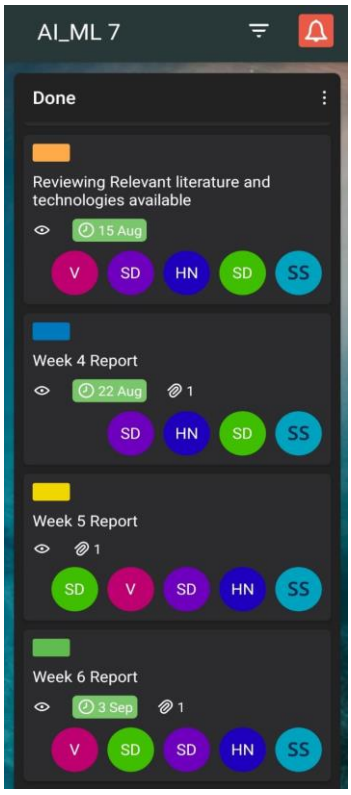


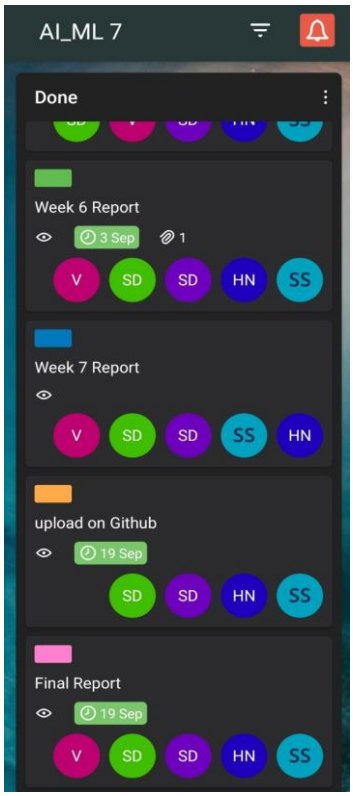
Github Link:

https://github.com/Shravani01007/Walmart-Sales-Prediction-For-Future-Campaigns

Trello Board:







Conclusion:

To encapsulate, this project dealt with the implementation of six algorithms on the Walmart dataset and a comparative analysis was carried out to determine the best algorithm. We find that our regression equation is quite accurate (78% accuracy) in predicting the weekly sales. Walmart can use it to forecast the sales better. They need to focus on the inventory planning of key departments like 38, 92 and 95. They need to overhaul the Markdowns that are given currently as they are not having the intended impact on sales. They need to focus on the year-end inventory as week 51 and 52 play a crucial part in predicting sales.