

SALES FORECASTING

USING REGRESSION MODEL, RANDOM FOREST REGRESSOR

PRIYADARSHINI G.
BADM, BATCH DM7

INTRODUCTION

In this sales forecasting project, historical sales data for 45 retail stores located in different regions is provided. The goal is to predict the department-wide sales for each store. Additionally, the company runs promotional markdown events before significant holidays such as Super Bowl, Labor Day, Thanksgiving, and Christmas. The dataset includes information about stores, features (including markdowns and macroeconomic factors), and training/testing data.

Problem Statement:

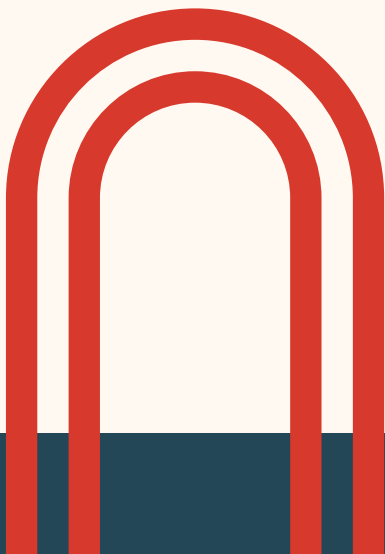
<https://drive.google.com/drive/folders/1buvhyRpnXuaOz1eoyDWq5OgHpOqcASo9?usp=sharing>

PERFORMANCE

Random Forest Regressor model was used to predict weekly sales.
The model's performance on the validation set is as follows:

- Mean Absolute Error (MAE): 1615.07
- Mean Squared Error (MSE): 18170416.78
- Root Mean Squared Error (RMSE): 4262.68
- R-squared (R^2): 0.97

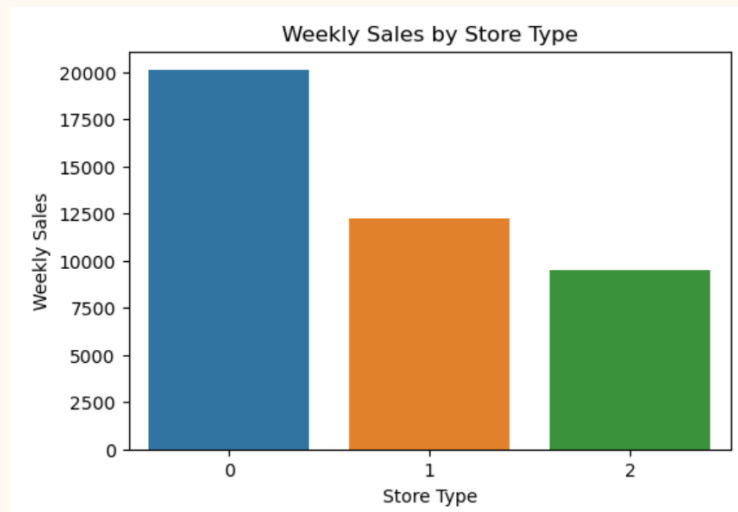
This analysis of Sales Forecasting using Random Forest Regressor,
can be accessed on Kaggle via the following link:
[Kaggle Sales Forecasting](#)



KEY INSIGHTS

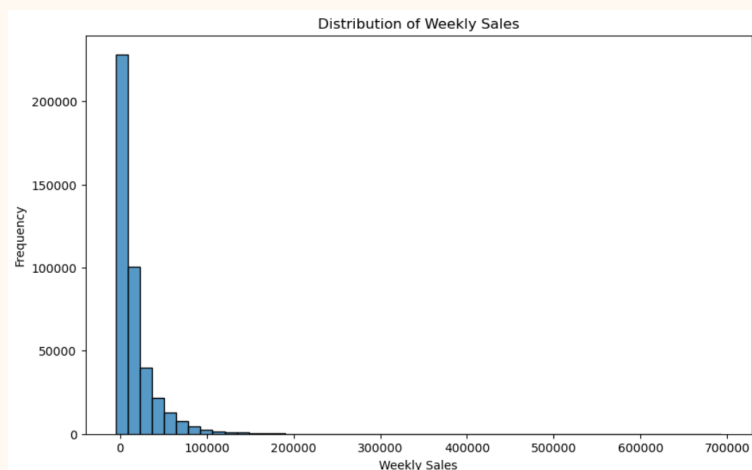
1. STORE TYPE ANALYSIS :

Store type A has the highest average weekly sales, followed by store type B and then C.



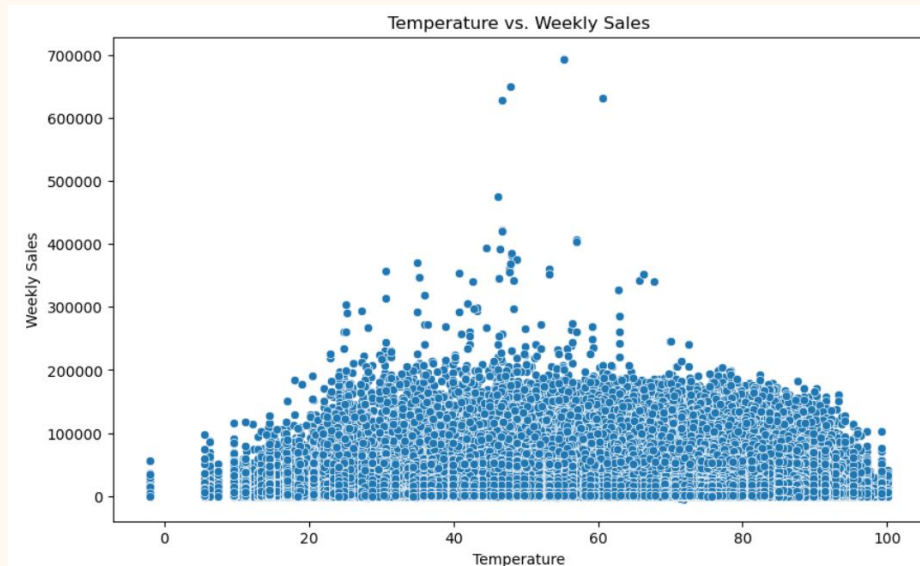
2. DISTRIBUTION OF WEEKLY SALES :

The distribution of weekly sales is right-skewed, indicating that most weeks have lower sales, while a few weeks have significantly higher sales.



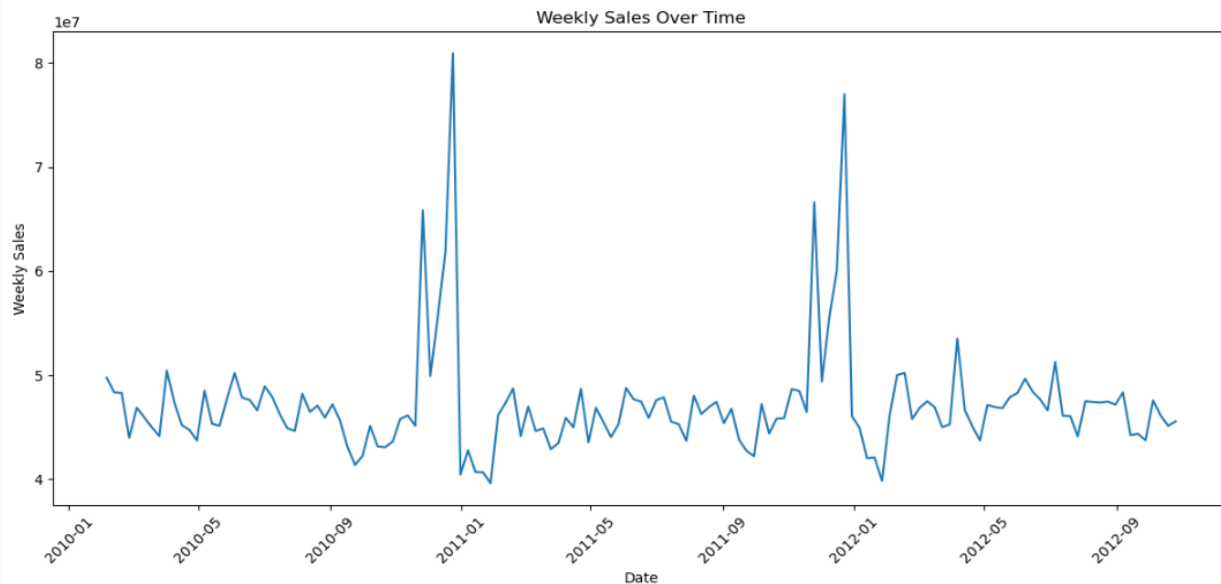
3. TEMPERATURE VS. WEEKLY SALES:

The scatter plot of temperature against weekly sales suggests that there is a higher concentration of higher sales when the temperature is between 25 to 75 degrees.



4. SEASONALITY:

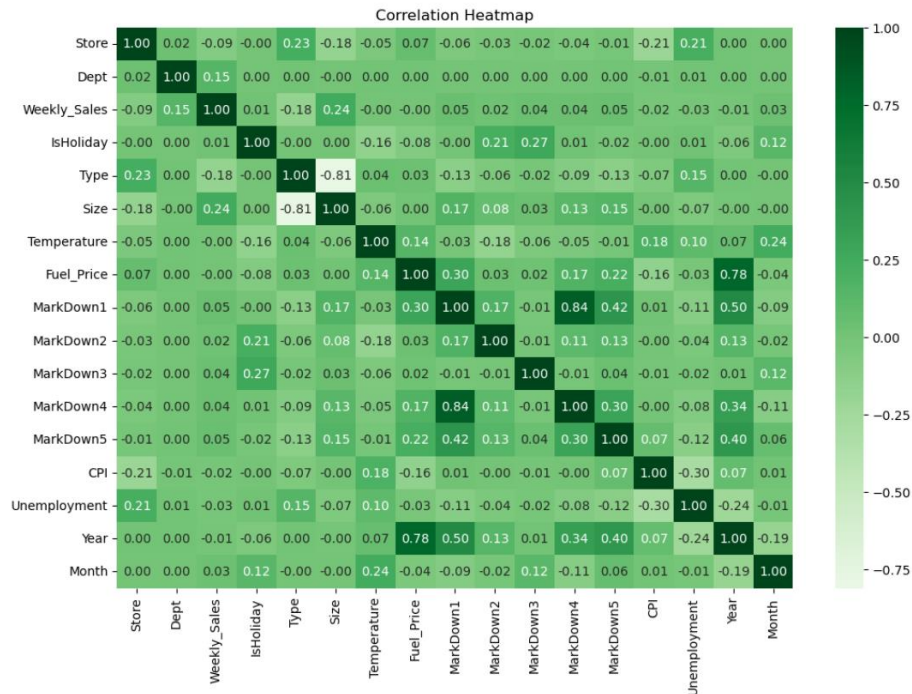
The line plot of weekly sales over time shows a clear seasonal trend, with a sudden increase in sales during November and December, which coincides with the holiday season.



5. CORRELATION INSIGHTS:

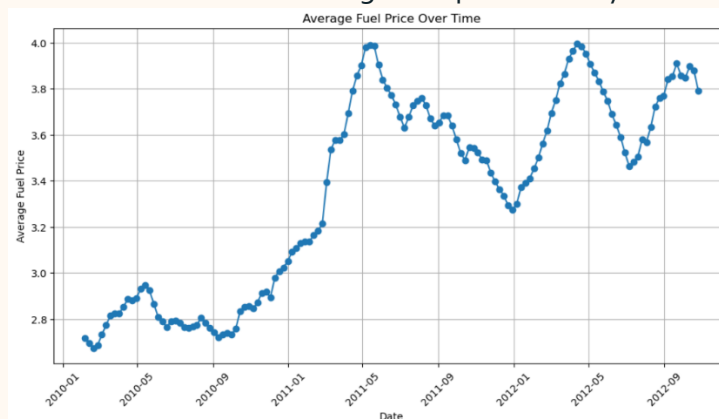
There is a strong positive correlation (0.84) between Markdown1 and Markdown4, suggesting that these two types of markdowns are often applied together.

Fuel price and year have a correlation of 0.78, indicating a general trend of increasing fuel prices over the years.



6. AVERAGE FUEL PRICE OVER TIME:

The average fuel price exhibits fluctuations over time. Notably, there was an increase in fuel price from 2.7 in September 2010 to 4.0 in May 2011. Subsequently, there were variations with an eventual decrease and increase in the average fuel price in the year 2012.



CONCLUSION

The Random Forest Regressor model demonstrates strong performance with a high R-squared value, indicating a good fit to the data. The insights obtained from visualizations and correlations provide valuable information about store performance, seasonal trends, and relationships between variables. These insights can guide decision-making for the retail company in terms of inventory management, pricing strategies, and promotional events.