Datasets

https://www.kaggle.com/datasets/manjeetsingh/retaildataset?resource=download

Cleaning

Date transformation DD/MM/YY
IsHoliday encoding
Get rid of missing value: your choice (markdowns)
Log sales

Summary

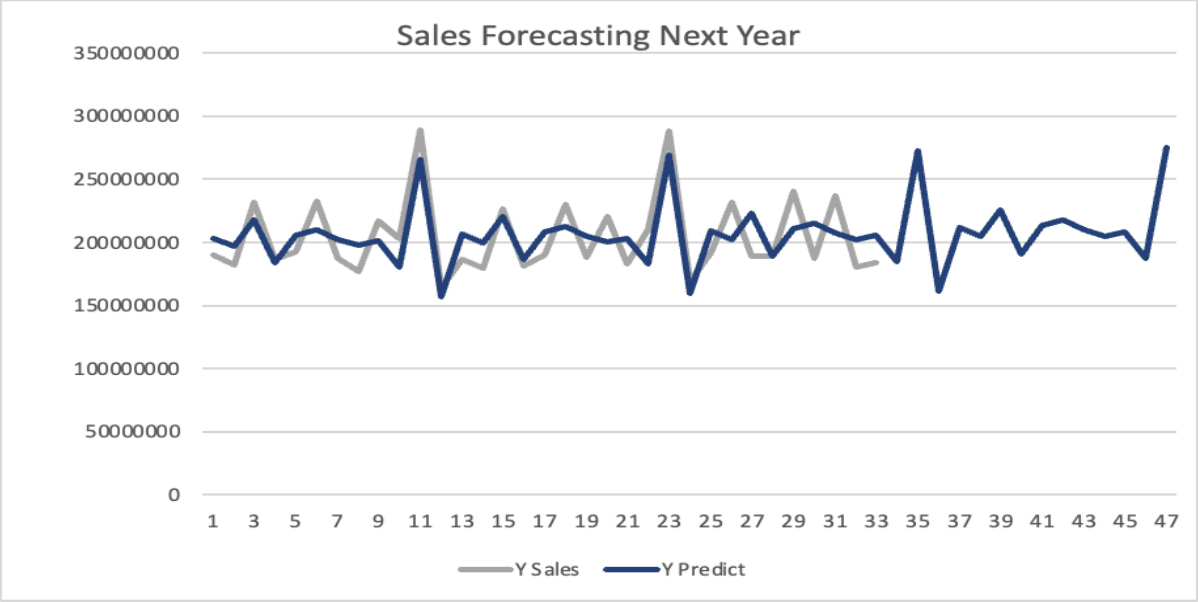
Research Question 1: What is the monthly sales trend over time? (the trend in 3 months later) (monthly sales trend of each store)

Research Question 2: The correlation of features variable and the weekly_sales

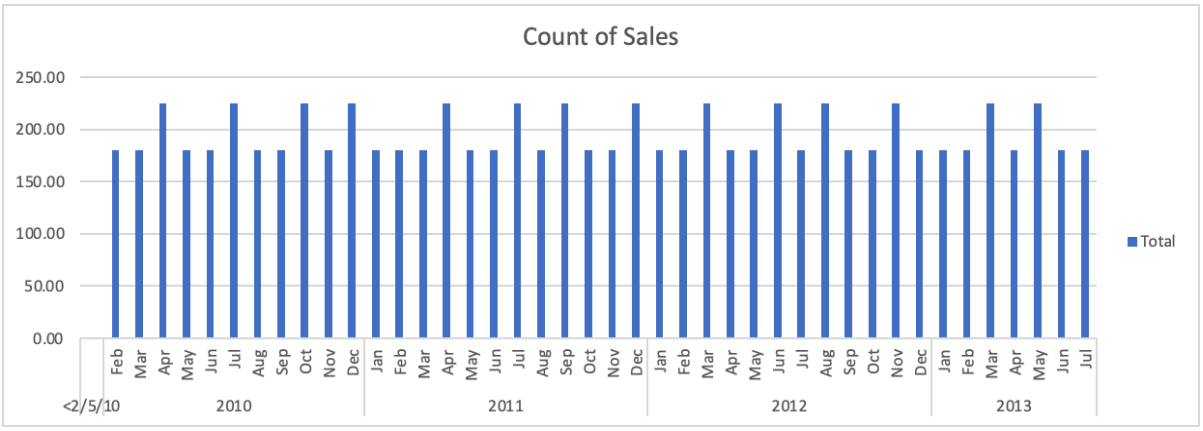
Research Question 3: Model the effects of markdowns on holiday weeks

Results

Research Question 1: What is the monthly sales trend over time?



Regarding question 1, we applied the ETS method to predict the base trendline of sales for the next 2 months. From the plot, we observed the pattern of sales has been fluctuating throughout the year and hit the peak toward the transition of the new year. This illustrates the reasonable seasonality in sales when sales reach its climax during the Holiday seasons between December and January. The predicting blue line indicates the similar trend, meaning there isn't any unexpected fluctuation for the next two months and next year. However, there are more factors that could possibly influence the sales prediction. We would need more examinations to identify the impact of other treatment factors.

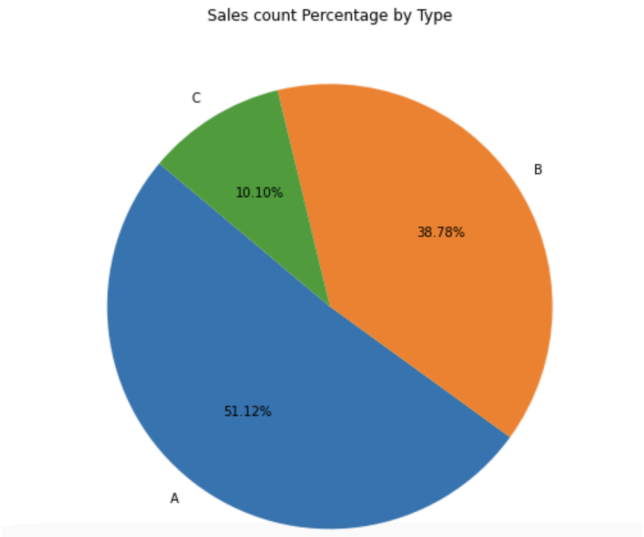


Research Question 2: The correlation of features variable and the weekly sales

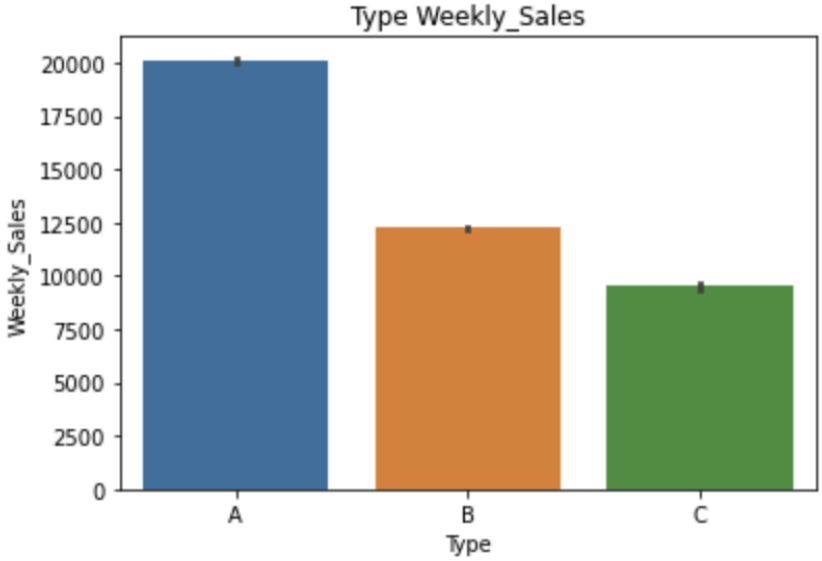
Type	Store	Sales_count	Sales_count_perstore	Sales_percentage	Weekly_Sales	Weekly_Sales_perstore	Weekly_Sales_percentage
A	22	215478	4788.40	51.11	20099.57	446.66	64.28
B	17	163495	3633.22	38.78	12237.08	271.94	29.70
C	6	42597	946.60	10.10	9519.53	211.55	6.02

Stores with type A are the most optimal store in return profit, with the highest average sales count and sales revenue per store.

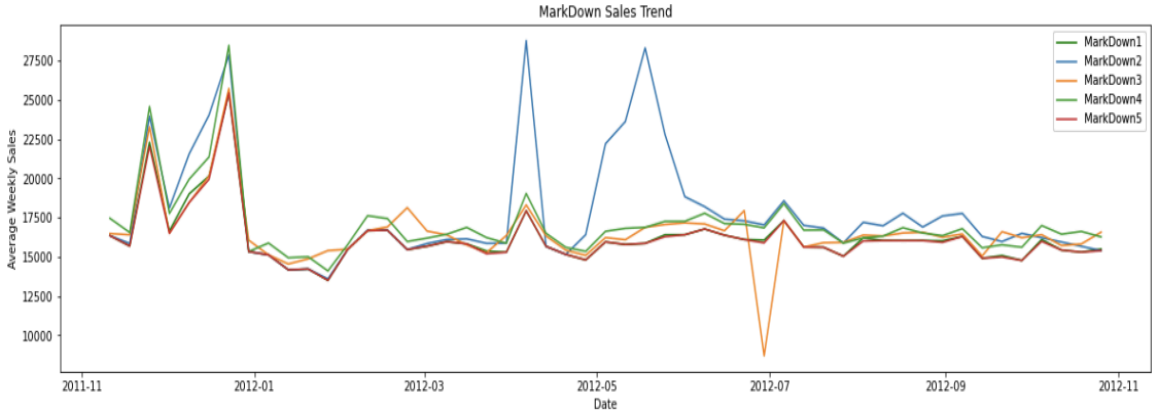
The table shows most stores sell type A products, while type c with the least stores. From the sales record count, Type A with 51.11% of all records, type B with 38.78% and type C with 10.10%.



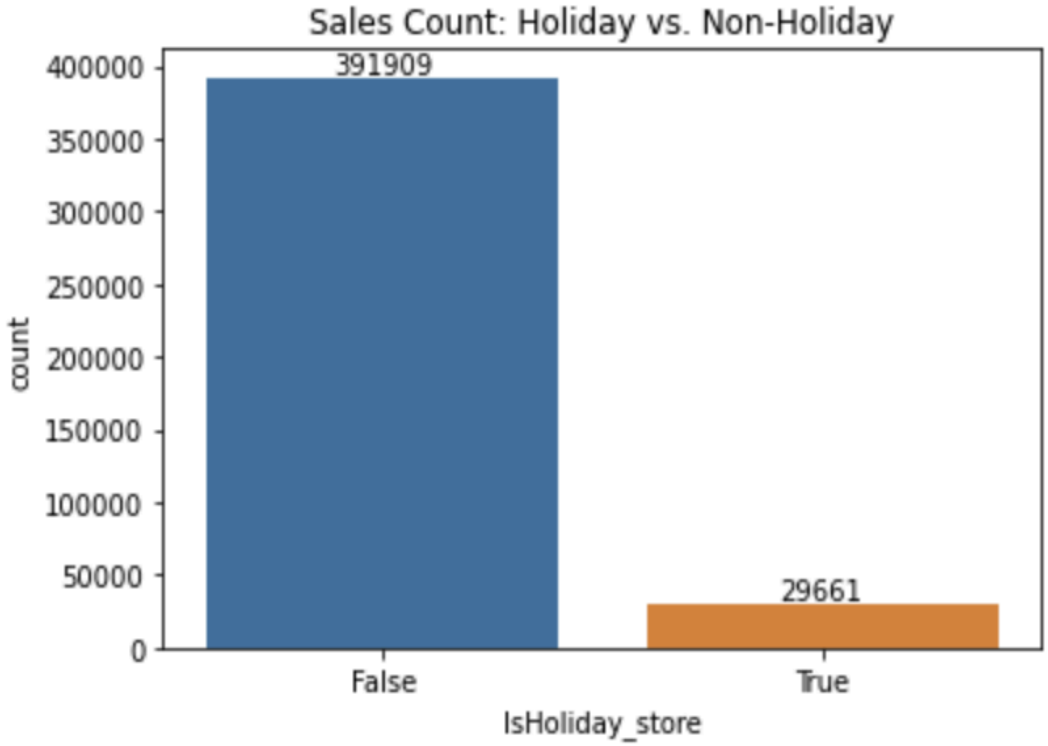
From the average sales trends across product types, Type A and Type B show similar patterns, though Type A consistently yields higher weekly sales. In contrast, Type C products display a steadier trend with less fluctuation and, rather than peaking on specific dates, show a slight dip in weekly sales during those periods.



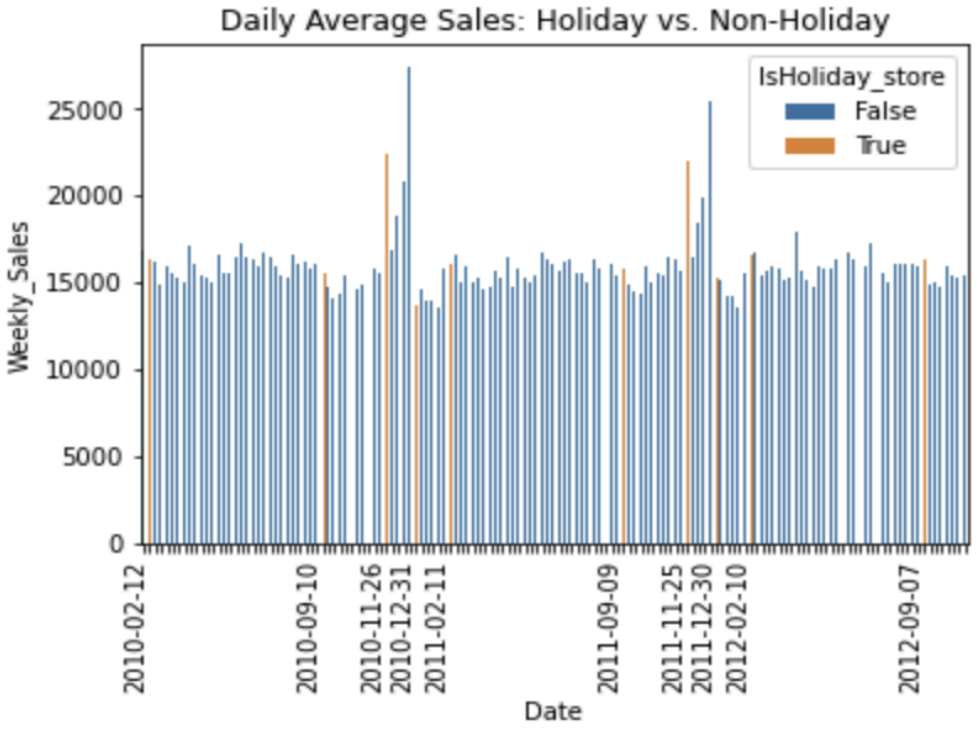
From the bar plot Type A displays the highest total weekly sales, while type c is the lowest, which is almost half of type A profits.



Among the five markdown sale trends, Markdown 2 and Markdown 3 show notable differences in timing during 2012. Markdown 2 peaks twice, in April and June, while Markdown 3 experiences a significant drop in July 2012. The other three markdown trends remain relatively consistent throughout the period. All five markdown trends reached their peak in the winter of 2011.

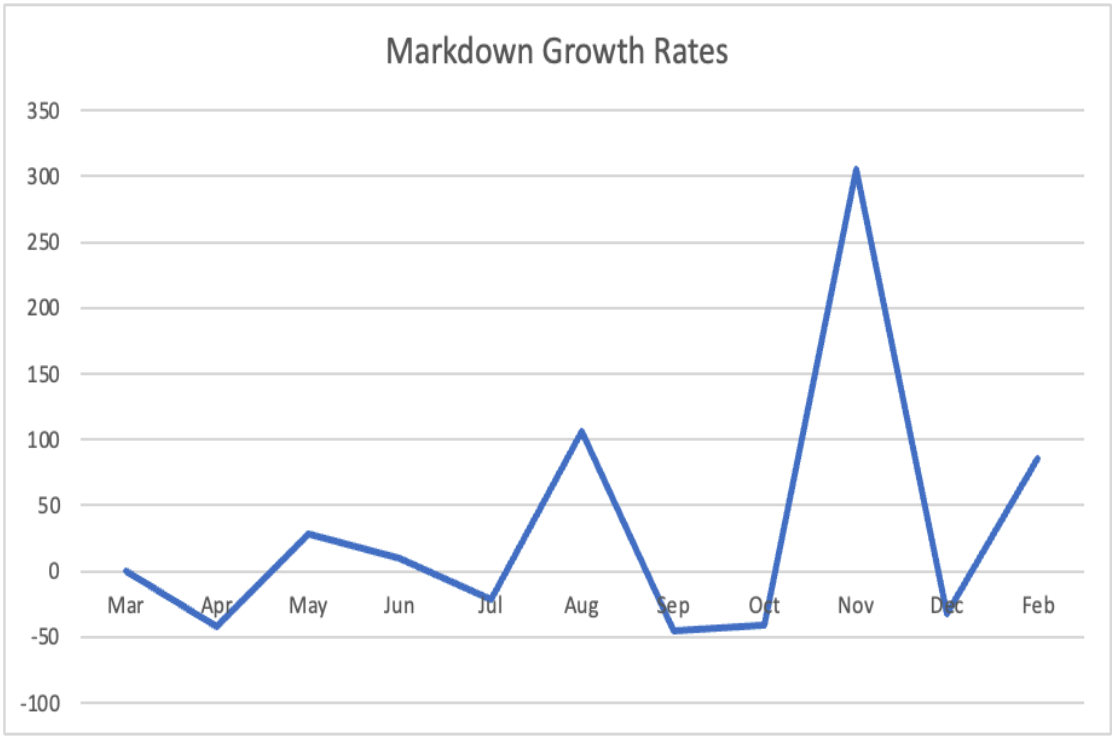


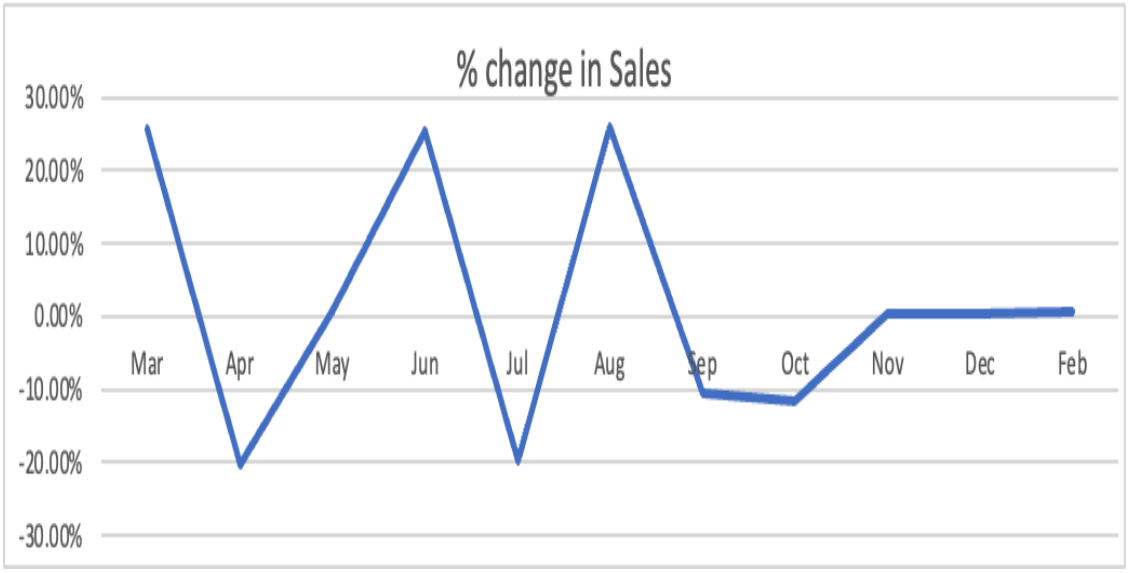
The majority of sales occurred outside of holidays, with only 10% of transactions taking place on holiday dates.



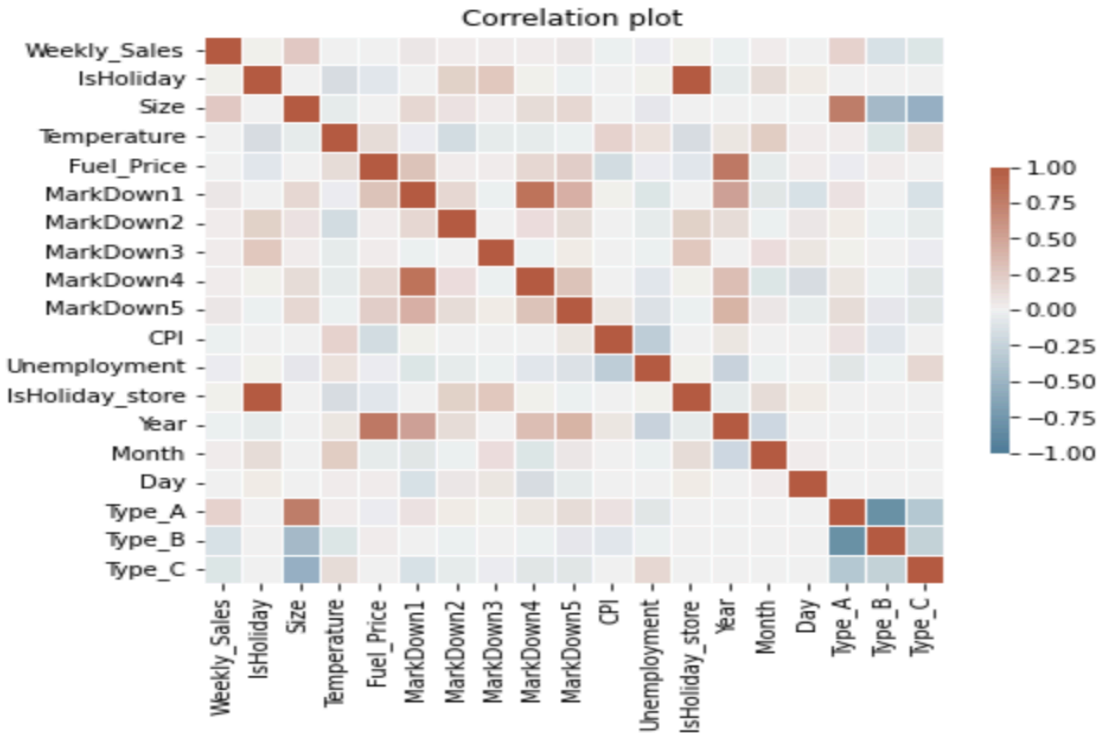
The bar plot, displaying holiday dates from the sales records, shows that holiday sales are not higher than non-holiday sales.

Markdowns

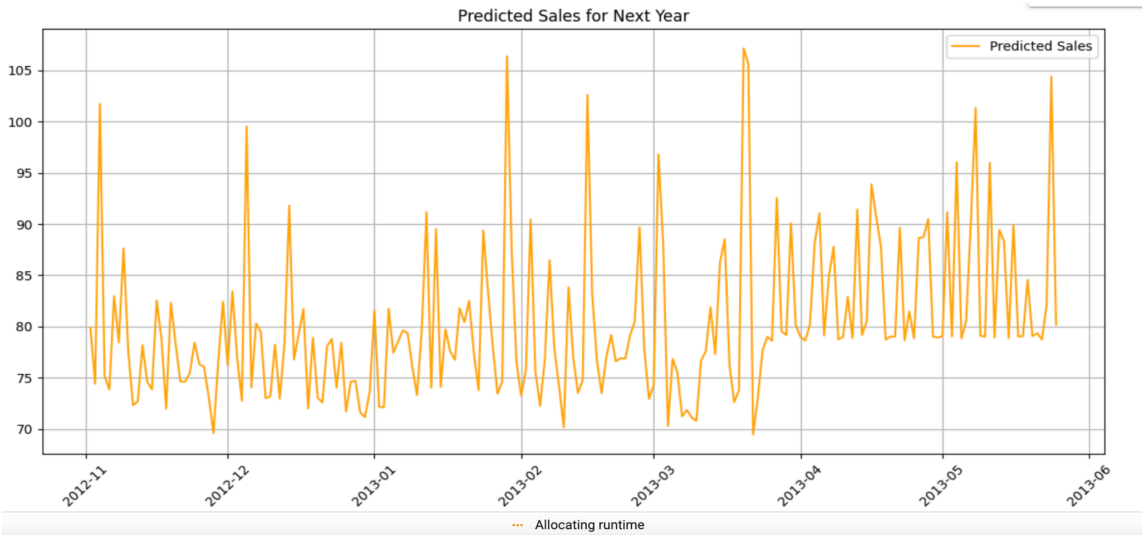




These two plots visualize the changes in Markdowns and Sales growth rates, providing the better observation in relationship between markdowns and sales. From the plots, we see the differences in growth rates between those two observations, however, the overall pattern from March to October indicates the positive correlation between those two. On the other hand, there is a spike of markdown rate in November, while the sales growth remains constant. We could imply that markdowns effectively improve sales during March to October and wouldn't affect much sales during the end of the year from November to February.



Question 3: Model the effects of markdowns, holiday, CPI, Temperature, and Fuel on Monthly Sales



In order to predict the sales affected by external factors such as markdowns, holiday, CPI, Temperature, and Fuel, we have built the Random Forest Model to predict the sales of the next 8 months. In the forecasting, there will be a pattern of sales which those beginning days of the month will achieve the most sales and relatively fluctuate for the rest of the month. Besides, starting from April next month, the base of total sales will increase, however, the peak of sales will be expected to be lower than the first quarter of 2013.

Suggestions

- Concentrating on boosting markdowns through March - October since the markdowns have a significant impact during that period.
- Allocating more type A stores to maximize sales as the sales per store in type A is nearly twice the others.
- Holiday appears to not have an obvious effect on the weekly sales, except Christmas time, suggesting on providing more sales strategies on other holidays to maximize the revenues.
- The Markdown 2 strategy produces more noticeable sales peaks than the others, so it might be worth integrating more aspects of this approach. Meanwhile, Markdown 3 sees a dip during the summer, indicating that a strategic adjustment could be beneficial.
- The prediction model shows that the overall business strategy is effective in increasing total sales. However, the sales peak is not very pronounced, suggesting that incorporating more short-term sales events during forecasted peak periods could help further boost sales.