



SALES PREDICTION ACROSS MULTIPLE STORES

BY MANILA DUGAR

DATASETS

Store Data : Store , StoreType, Assortment, CompetitionDistance,
CompetitionOpenSinceMonth, CompetitionOpenSinceYear, Promo2,
Promo2SinceWeek, Promo2SinceYear, Promointerval

Train Data : Store, DayOfWeek, Date, Sales, Customers, Open, Promo,
StateHoliday, SchoolHoliday

Test Data : Id, Store, DayOfWeek, Date, Open, Promo, StateHoliday,
SchoolHoliday

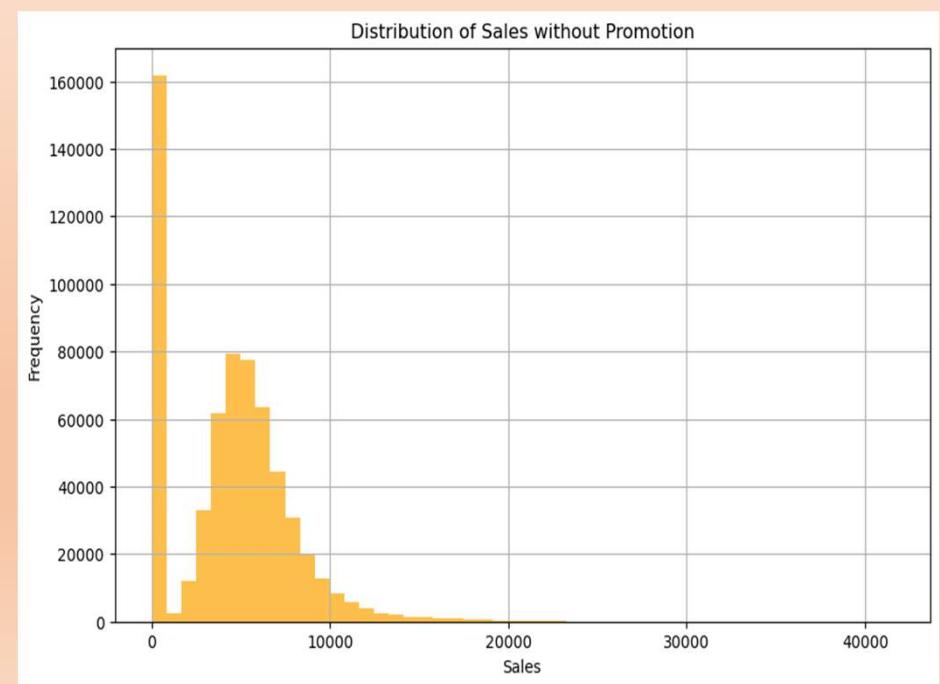
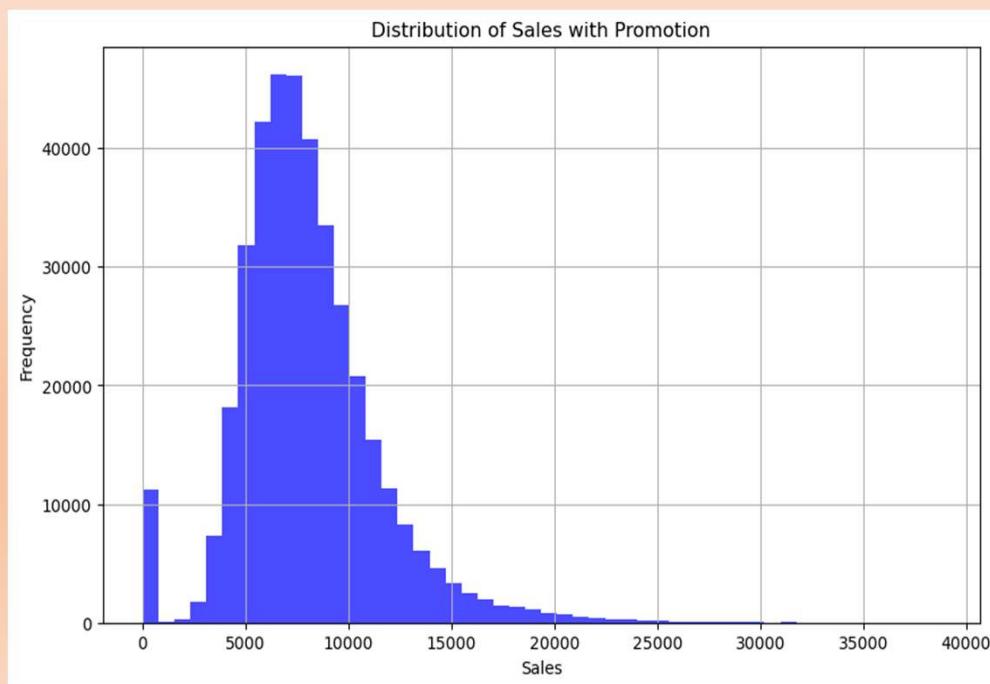
Exploration of customer purchasing behaviour



The sales is directly related to customer

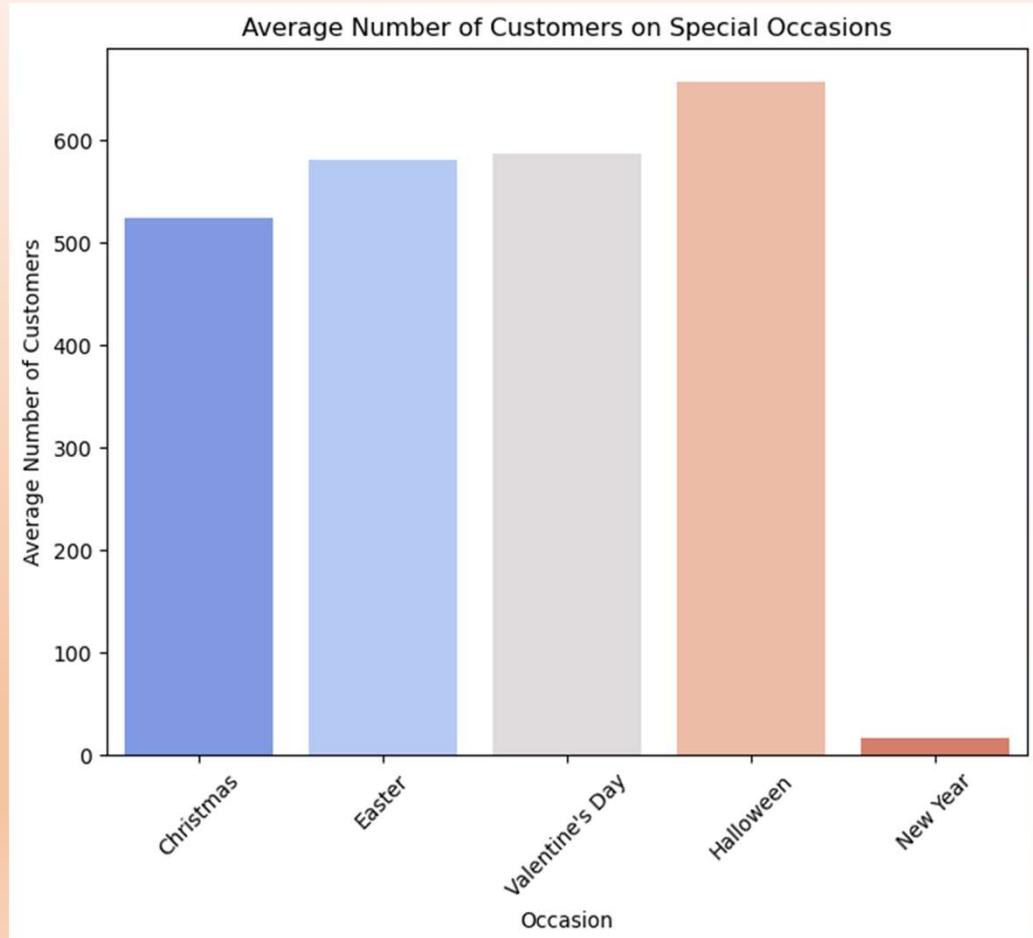


Distribution of sales is higher with promotion
compare to sales without promotion

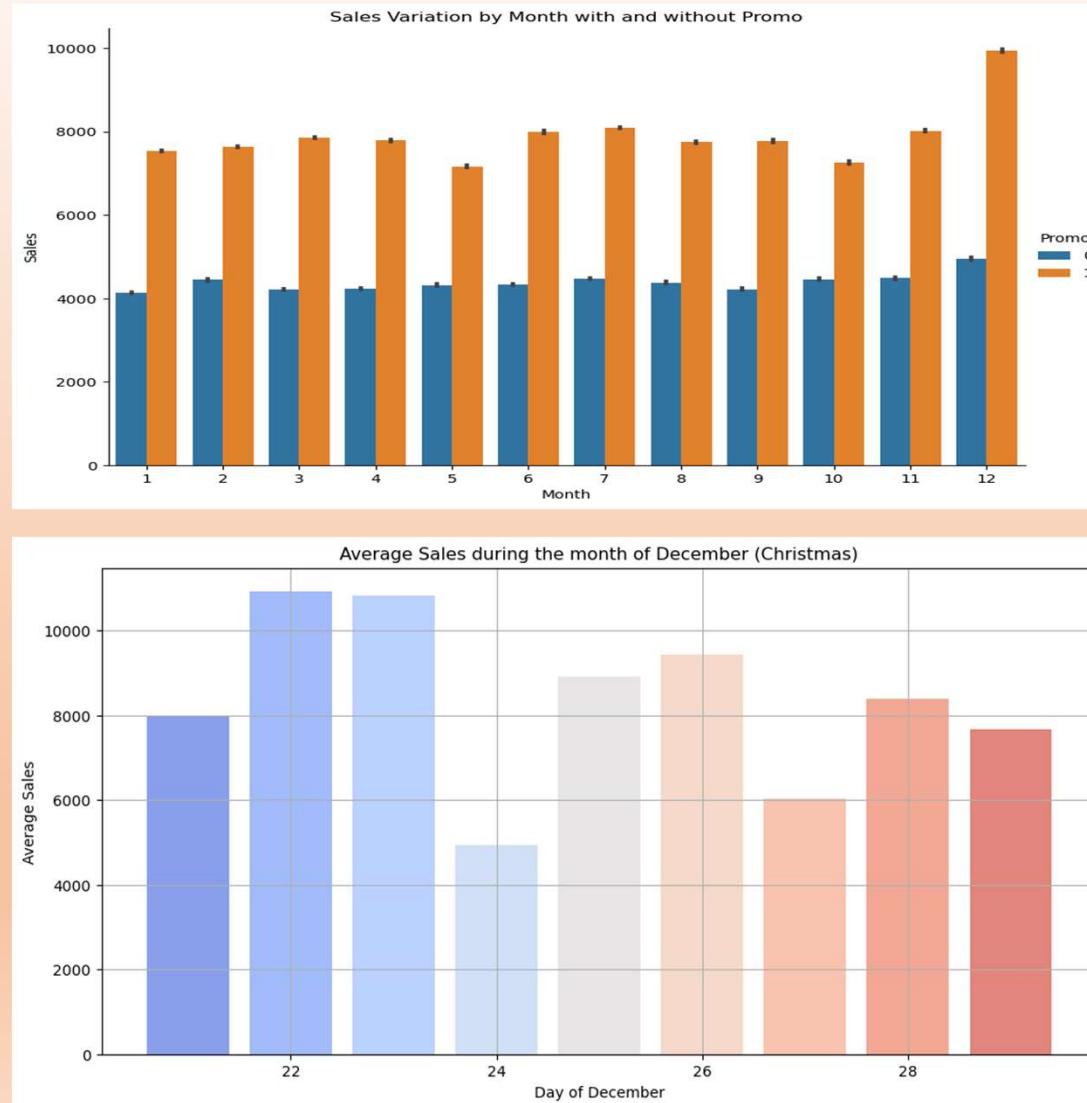


Unveiling Seasonal Sales Trends in Pharmaceutical Stores

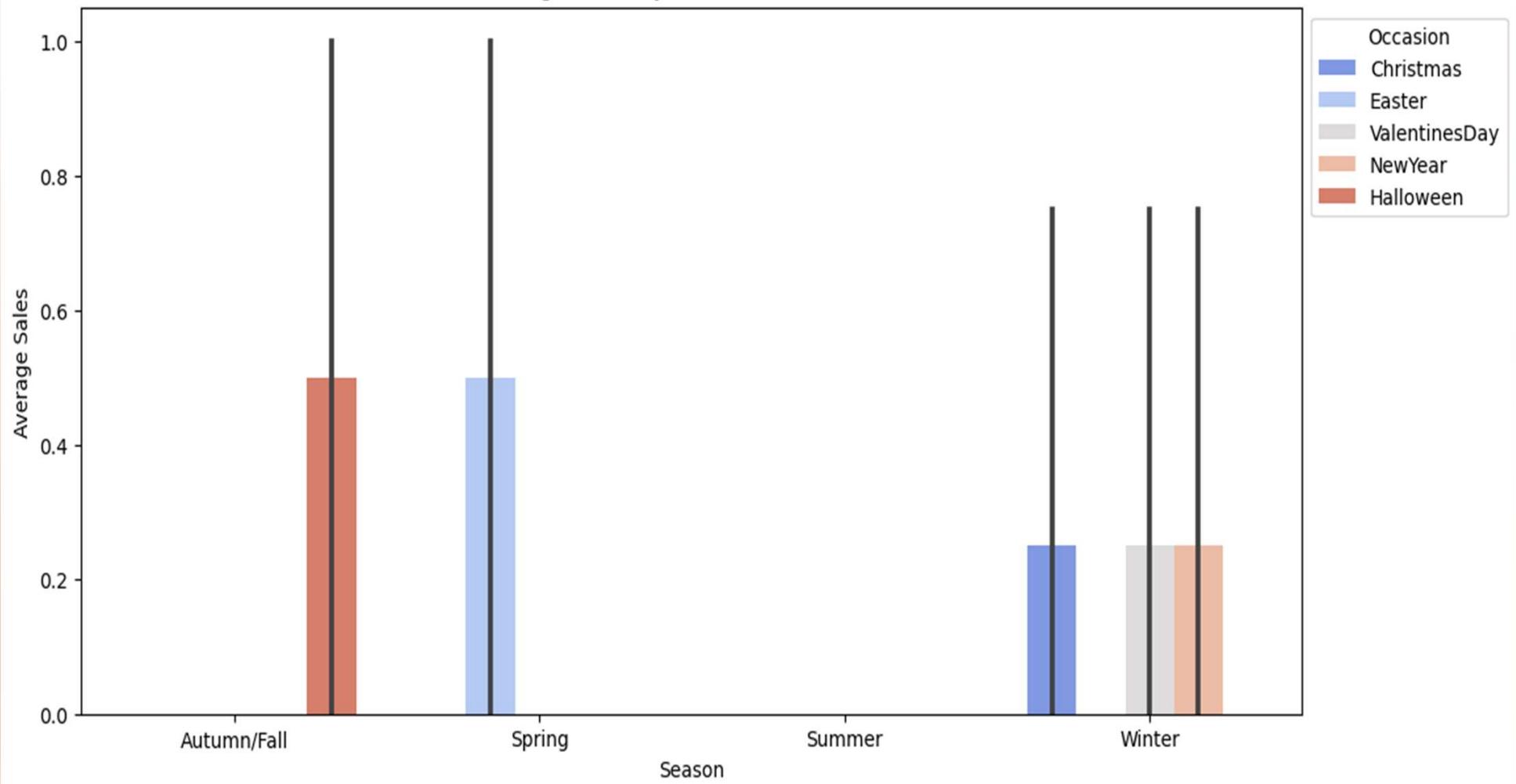
- In this presentation, we delve into the intriguing correlation between Halloween, the beginning of winter, and its impact on pharmaceutical store sales. Halloween drives consumer behavior, leading to increased spending and social gatherings .
- As Halloween arrives, it signals the onset of the winter season with colder weather and climatic changes .
- This transition has notable health implications, making individuals more susceptible to illnesses like colds, flu, and seasonal affective disorder.
- Consequently, pharmaceutical store sales surge, as consumers seek health-related products.
- To capitalize on this surge, pharmaceutical stores can employ seasonal marketing strategies while also addressing specific health needs during winter.
- In conclusion, recognizing and catering to this seasonal sales phenomenon is key for pharmaceutical stores.



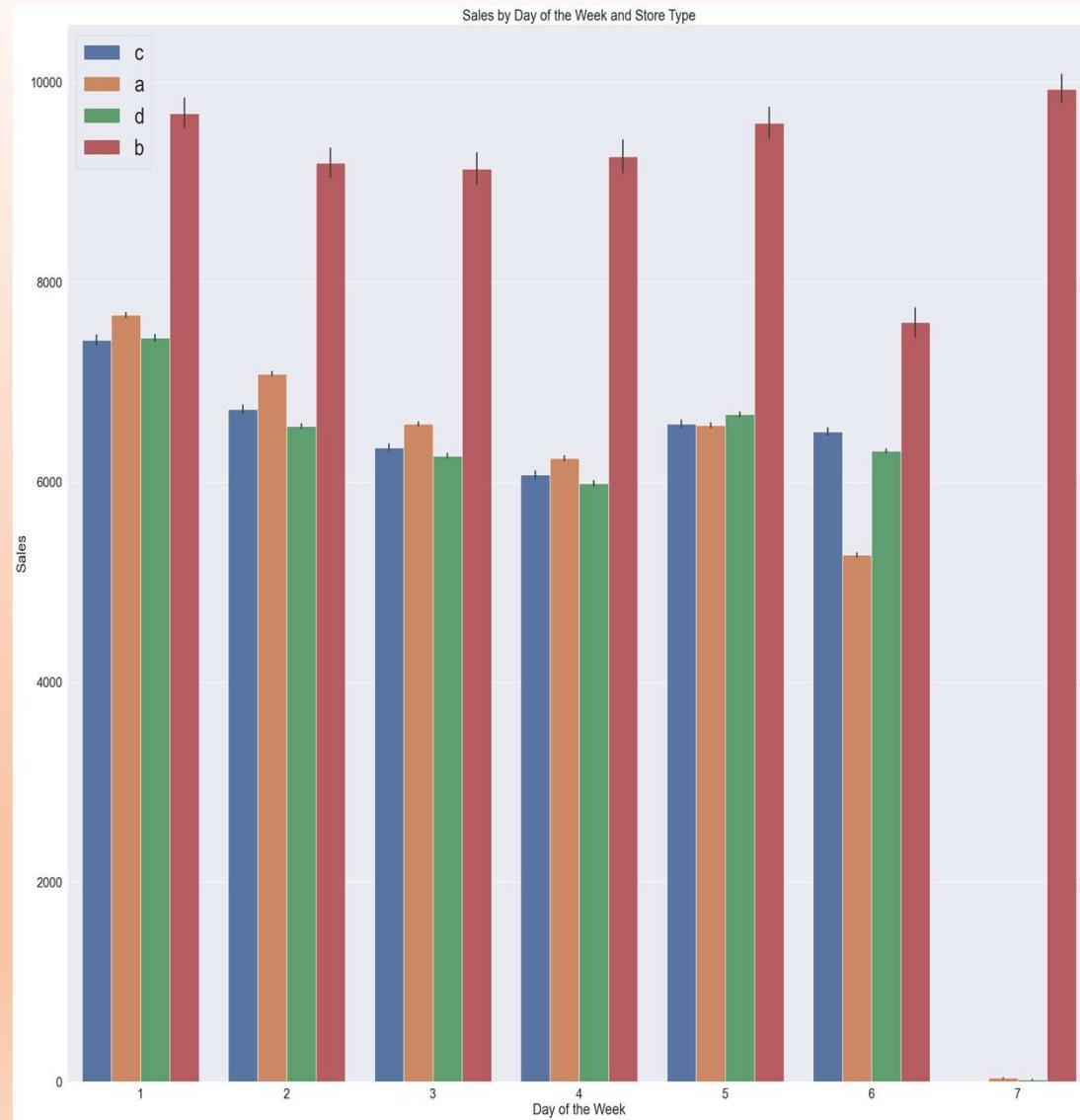
- Promotions play a pivotal role in driving sales in pharmaceutical stores, with December emerging as a strategic hotspot for these marketing endeavors.
- Notably, the period from December 22nd to 25th, encompassing the Christmas holiday, witnesses a surge in promotional activities.
- This timing aligns with the transition from winter to spring, a critical juncture when consumers seek health-related products.
- Pharmacies strategically capitalize on this consumer behavior by offering discounts, special deals, and targeted promotions during this festive season.
- Recognizing the correlation between seasonal transitions, heightened health concerns, and promotions is crucial for pharmaceutical stores to maximize their sales potential.



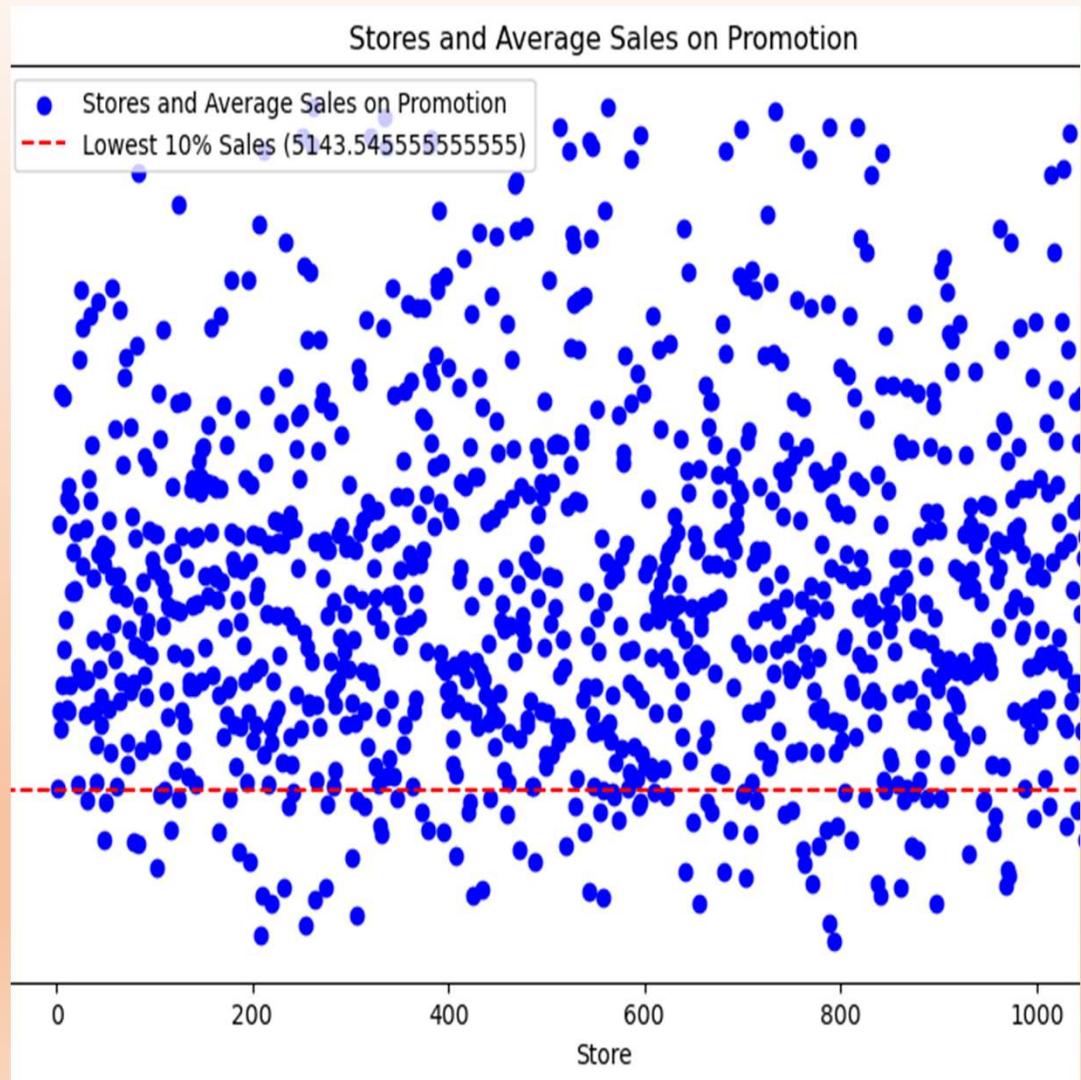
Average Sales by Season and Occasion



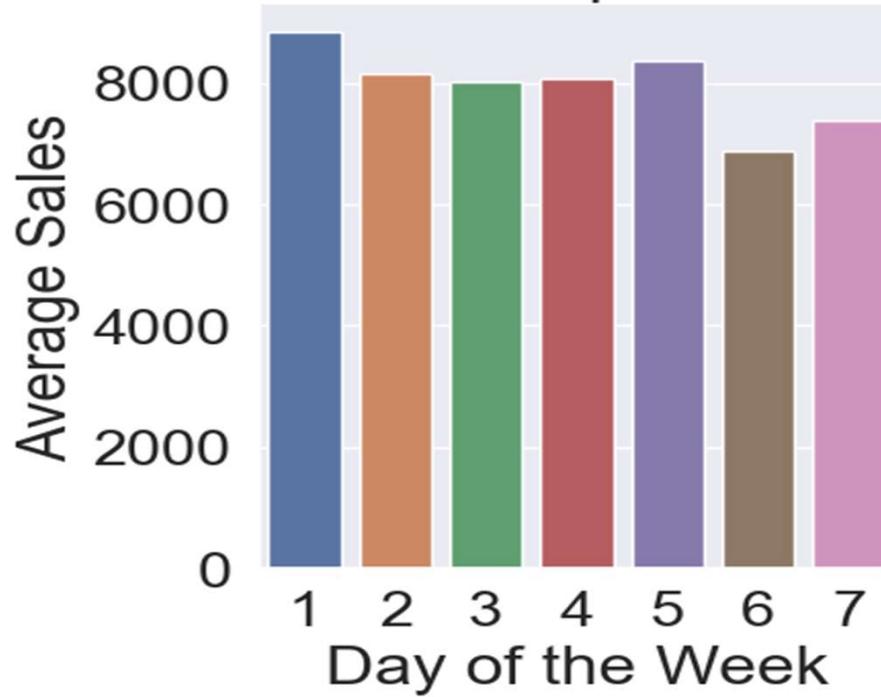
- There are four store types: A, B, C, and D.
- Store A has more locations, but store B has higher weekly sales.



- "Lowest 10% sales (5143.545555555)" means that 10% of the sales data have values lower than or equal to 5143.545555555.
- This figure serves as a reference point to identify the least performing sales in a dataset, which can be important for business analysis and improvement

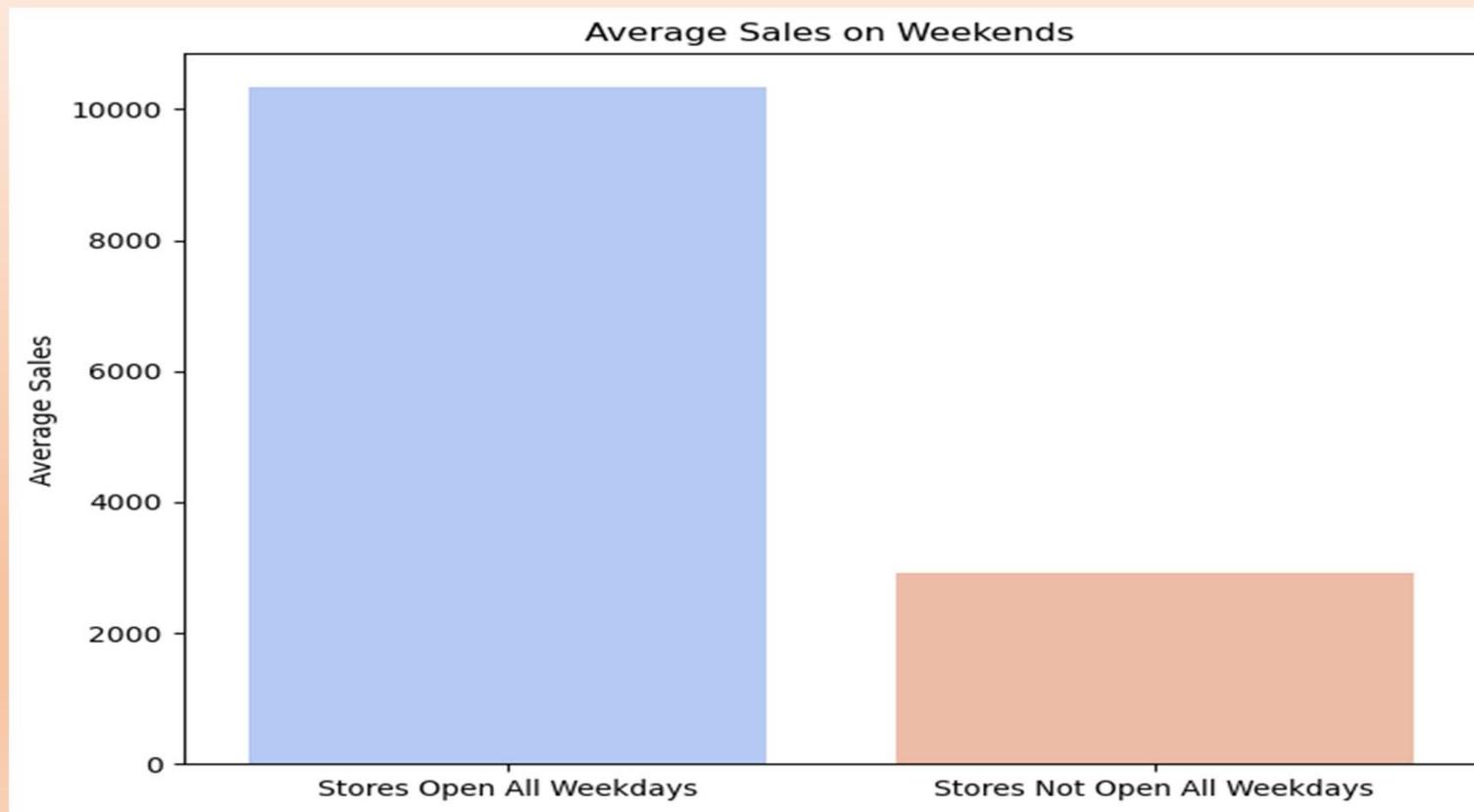


Average Sales for Stores Open All Week on Each Day

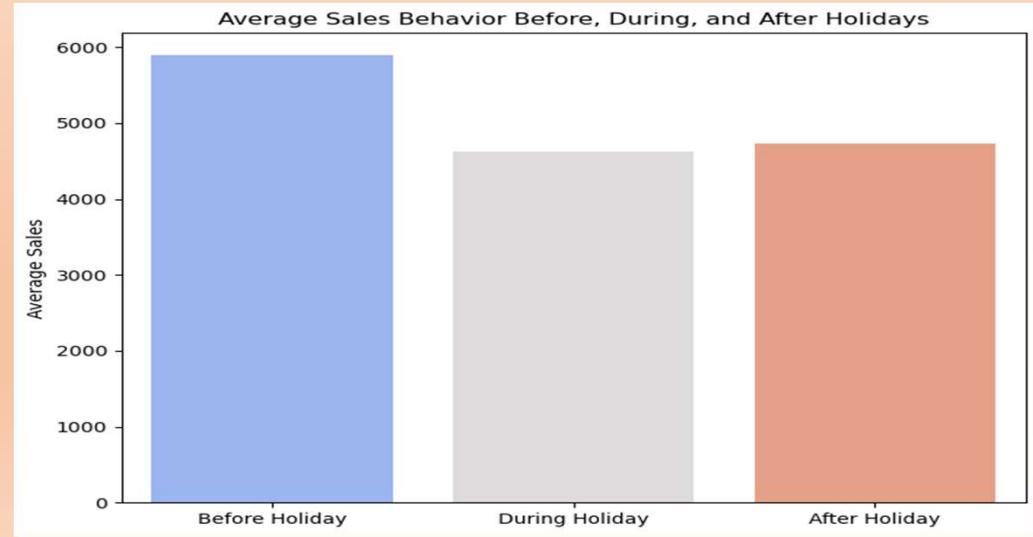
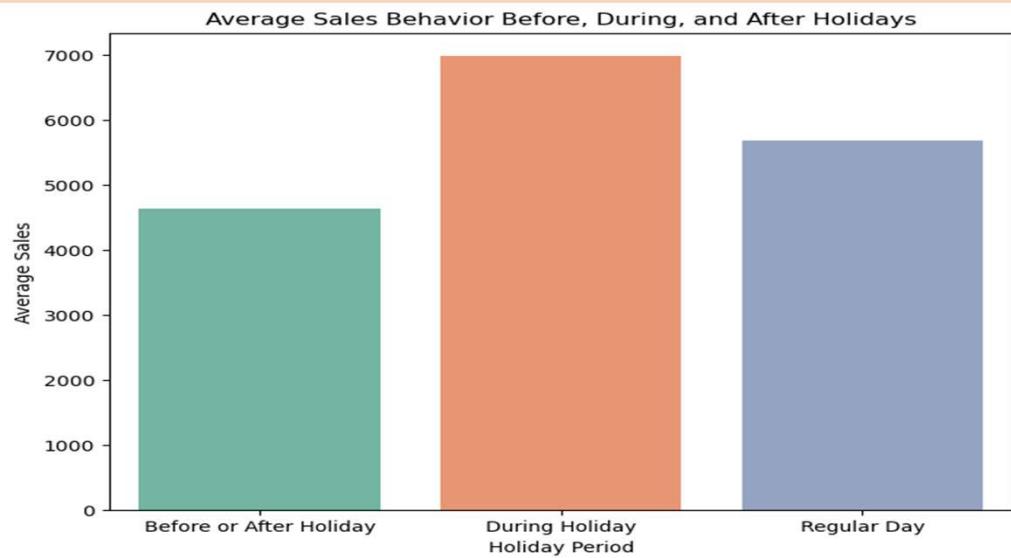


Sales are higher on Mondays and lower on weekends in comparison

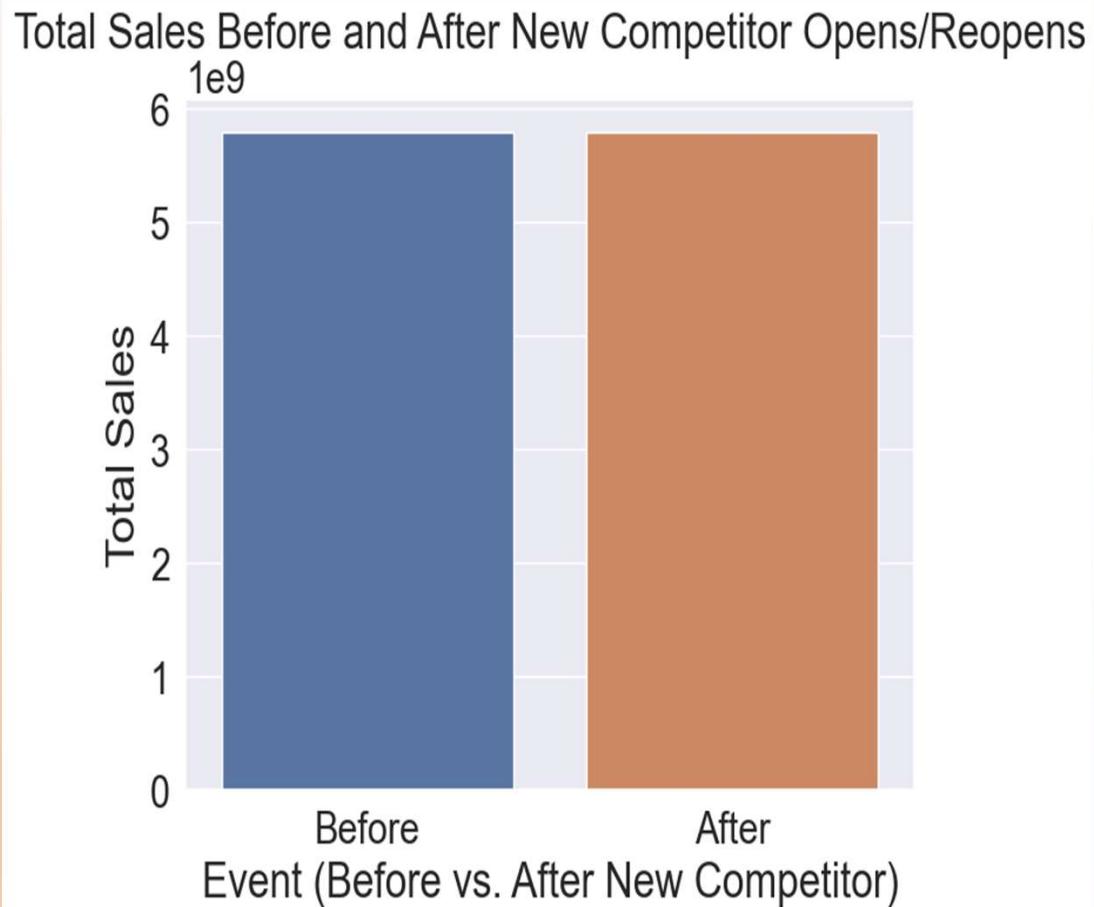
Majority of stores are open all weekdays



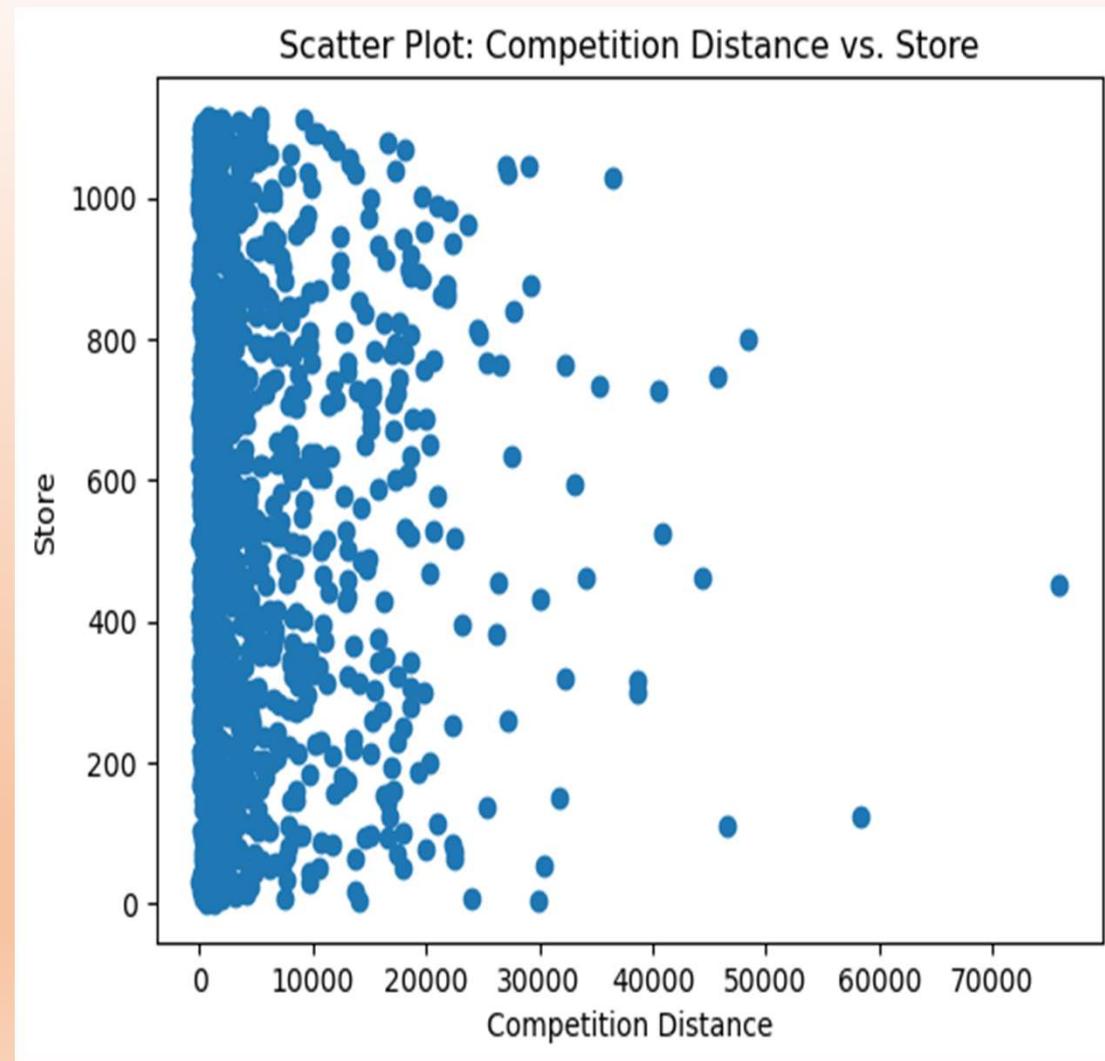
- Sales tend to exhibit two distinct patterns around holidays: either surging on the holiday itself or peaking in the days leading up to it. This phenomenon can be attributed to various factors.
- Firstly, consumers engage in holiday shopping, purchasing gifts, decorations, and supplies, boosting sales on the actual holiday.
- Secondly, retailers strategically implement promotions and discounts before holidays, luring shoppers and increasing sales.
- Additionally, people tend to stock up on essentials and seasonal items in preparation for the festivities, driving sales before the holiday.
- Last-minute shoppers and extended store hours also contribute to pre-holiday sales spikes.
- The timing of these sales fluctuations depends on consumer behavior, holiday characteristics, and retailer strategies.

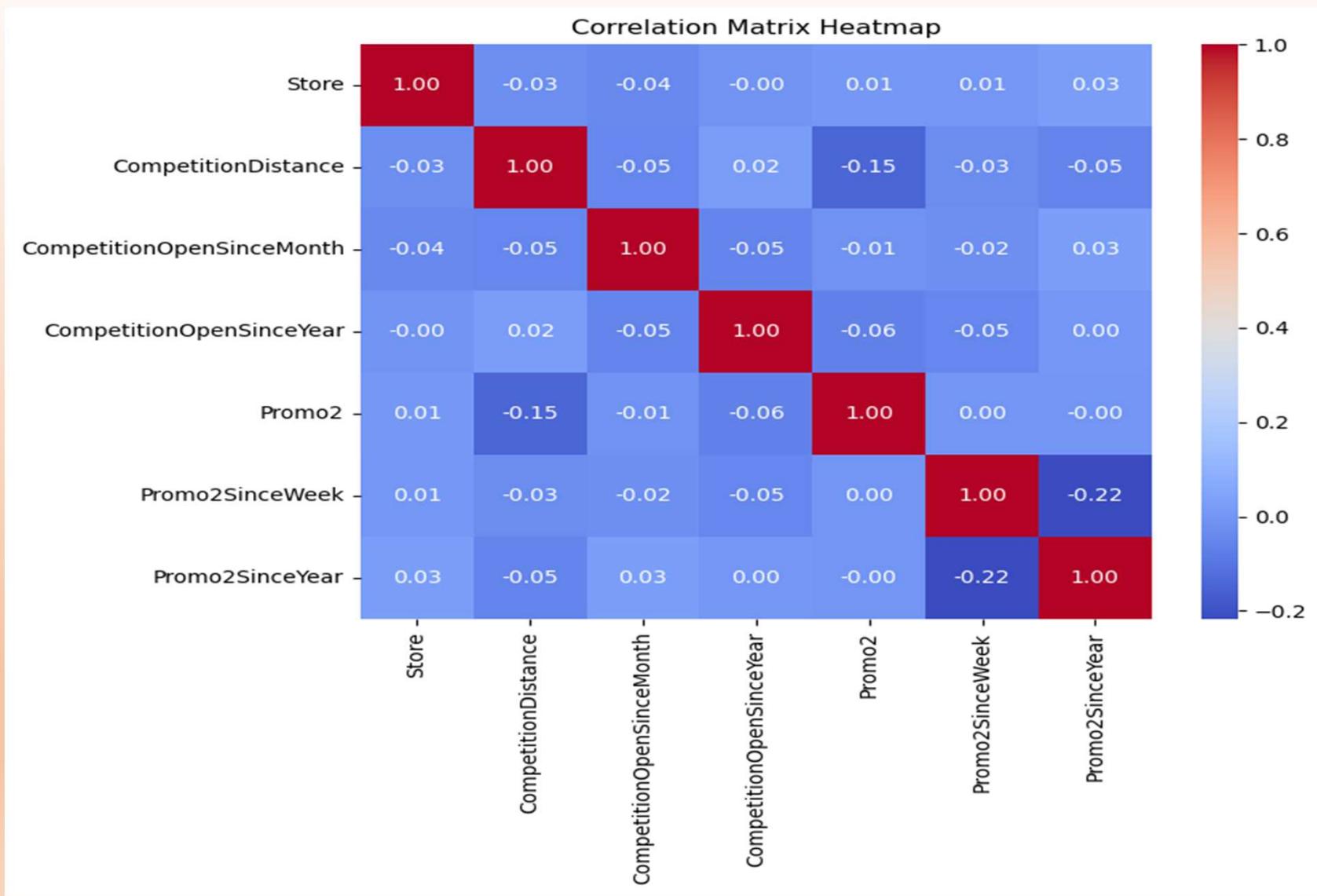


- Total sales before and after a new competitor opens and reopens are the same indicates that there is no significant change in total sales despite the introduction of a new competitor or the reopening of an existing one.
- This could imply that the impact of the competitor's entry or reopening on the market and overall sales has been minimal or that other factors have balanced out any potential sales fluctuations.
- It might also suggest a stable or competitive market environment where businesses are able to maintain consistent sales levels even in the face of increased competition.

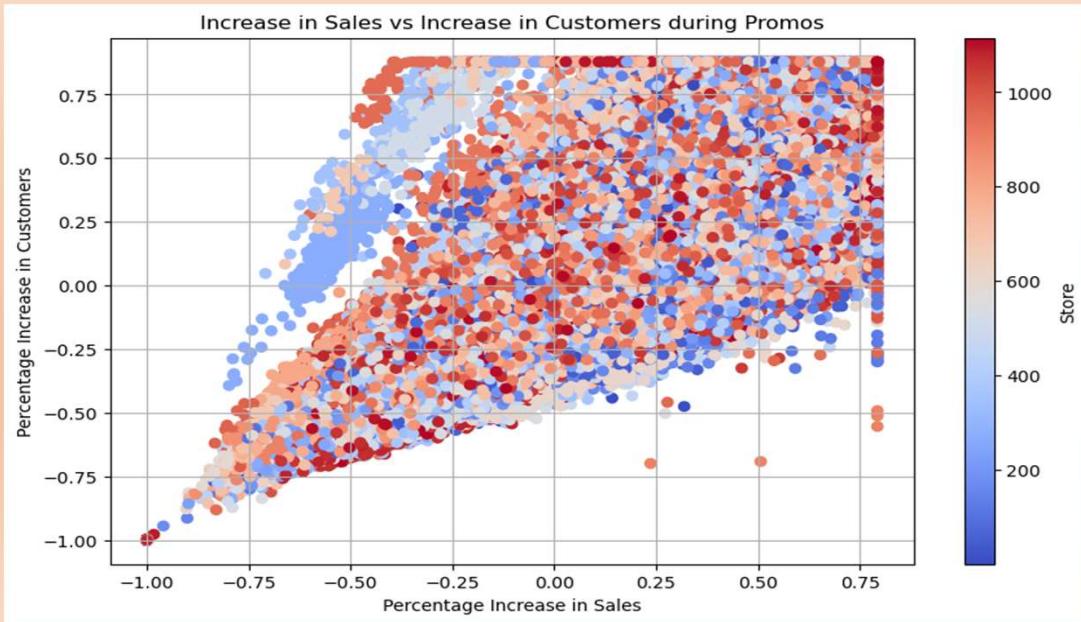
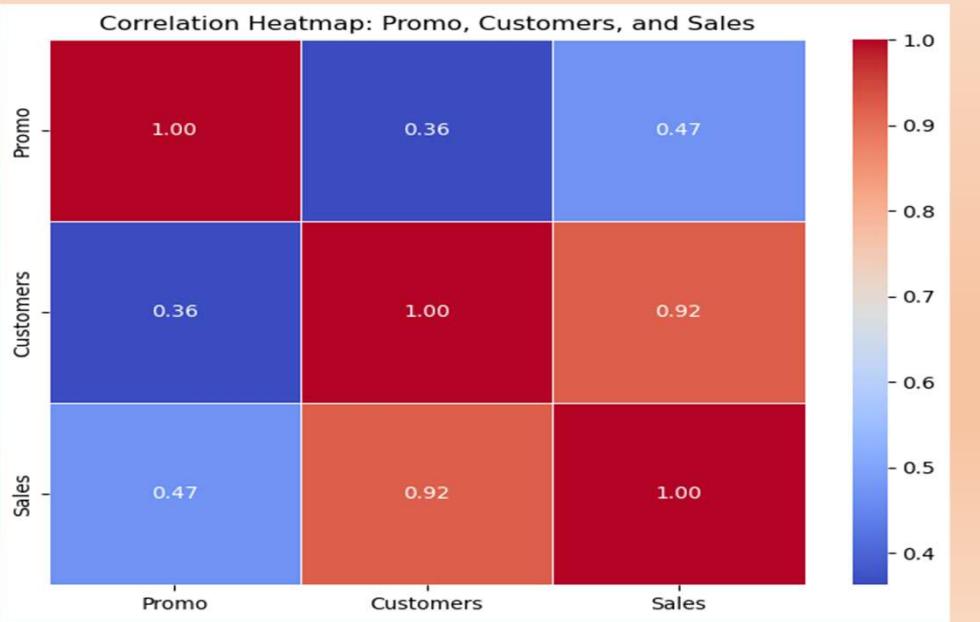


- As the competition distance is increasing, the number of stores is decreasing.
- This suggests that when the distance between competing stores or businesses increases, there is a trend of fewer stores or businesses being present in the area.
- This could be due to a variety of factors, such as reduced customer demand in areas with greater distances between stores or a consolidation of businesses as competition becomes less intense.
- It highlights the relationship between competition and the density of stores in a particular geographic area.





- There is a very strong correlation between sales and customers; an increase in customers leads to an increase in sales accurately describes a fundamental principle of business.
- It highlights the positive relationship between the number of customers a business serves and its overall sales.
- When more customers visit or engage with a business, it often results in higher sales revenue.
- This concept underscores the importance of attracting and retaining customers as a key driver of business growth and success.

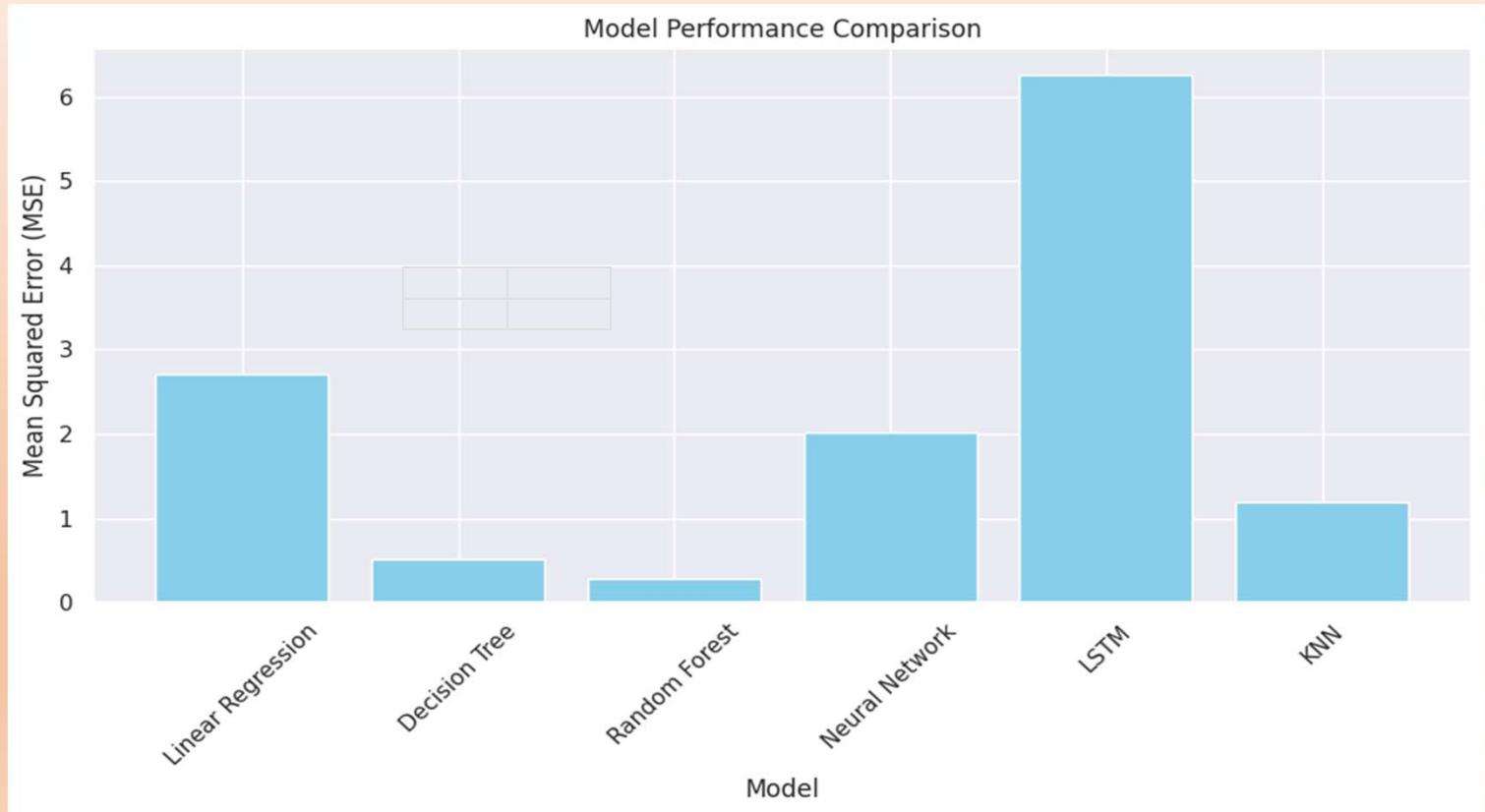


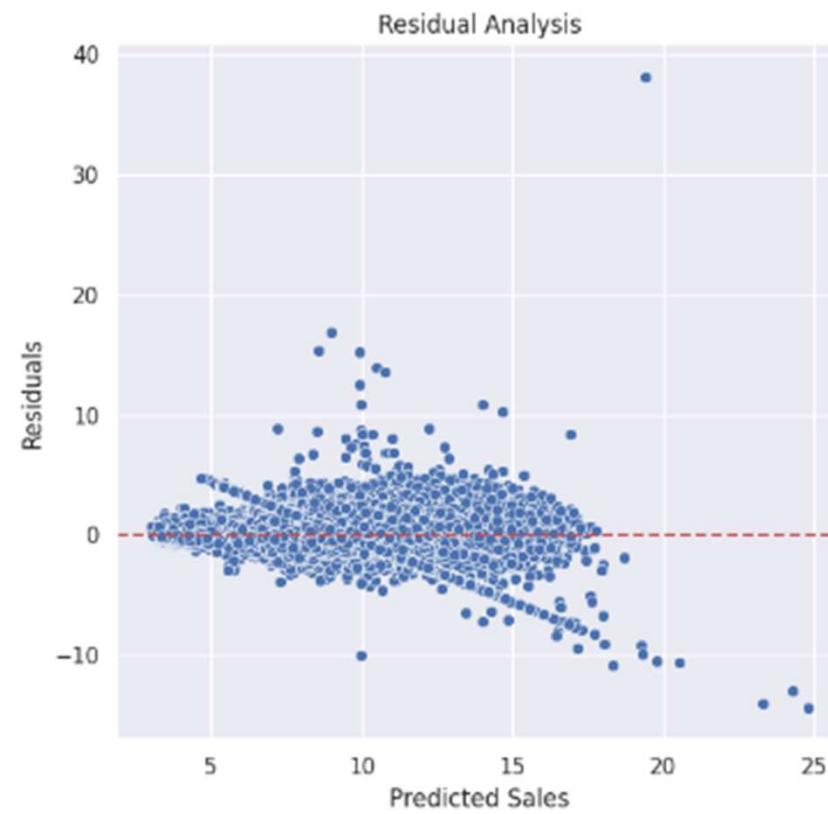
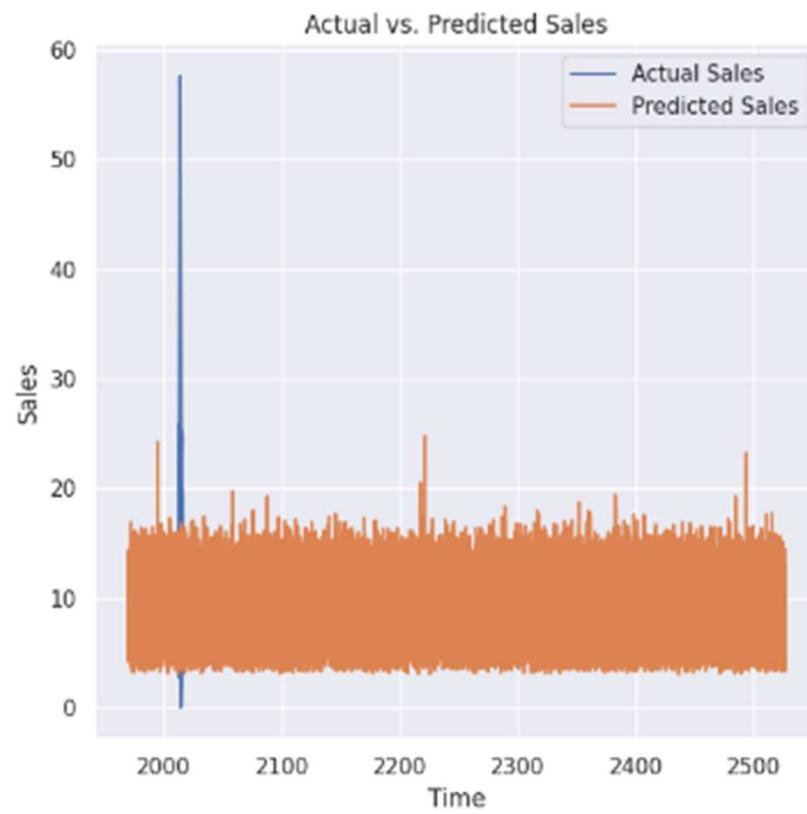
Prediction Of Store Sales



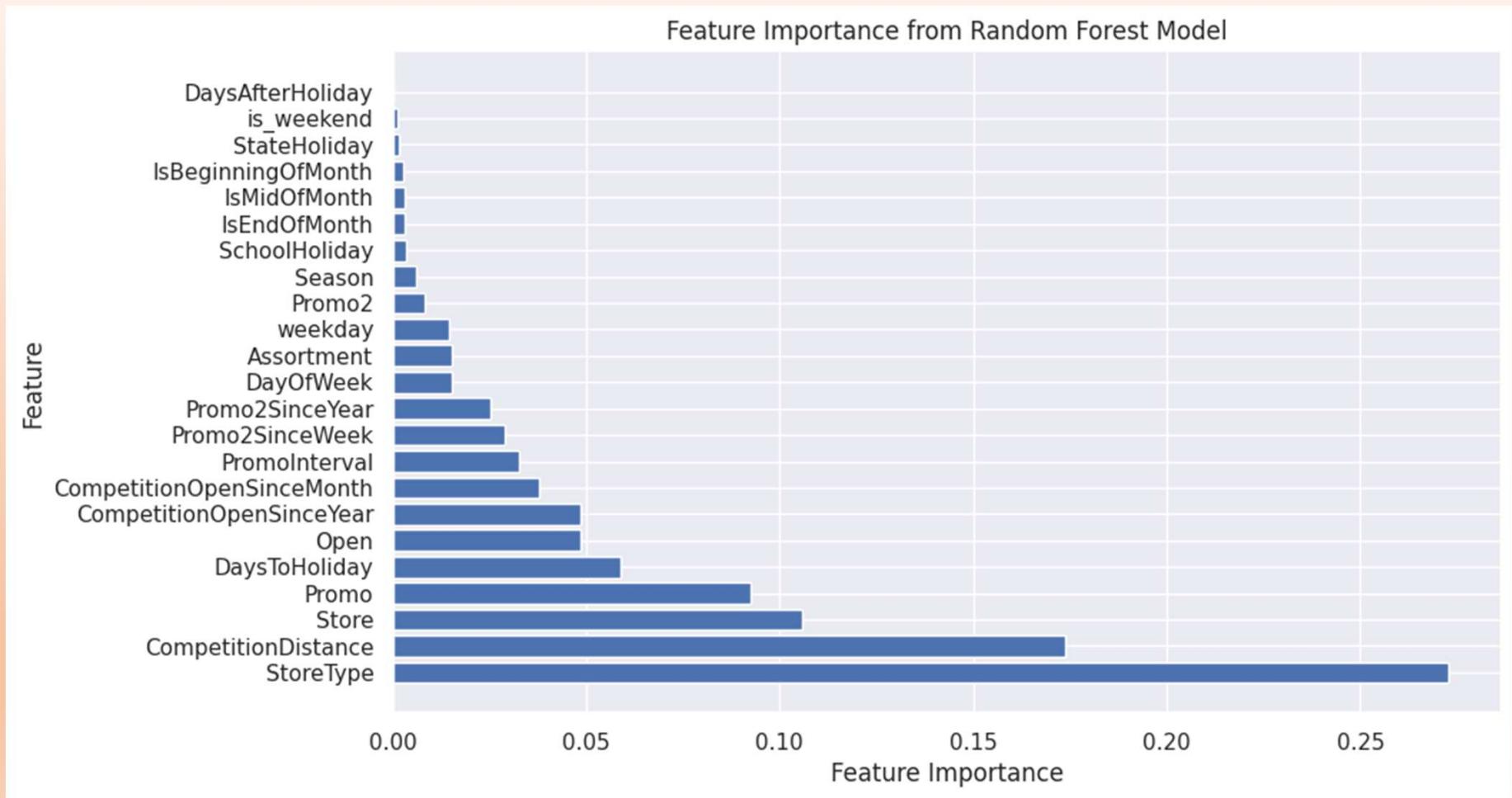
Models

	Model	MSE
0	Linear Regression	2.701617
1	Decision Tree	0.510121
2	Random Forest	0.282171
3	Neural Network	2.023631
4	LSTM	6.254384
5	KNN	1.189246





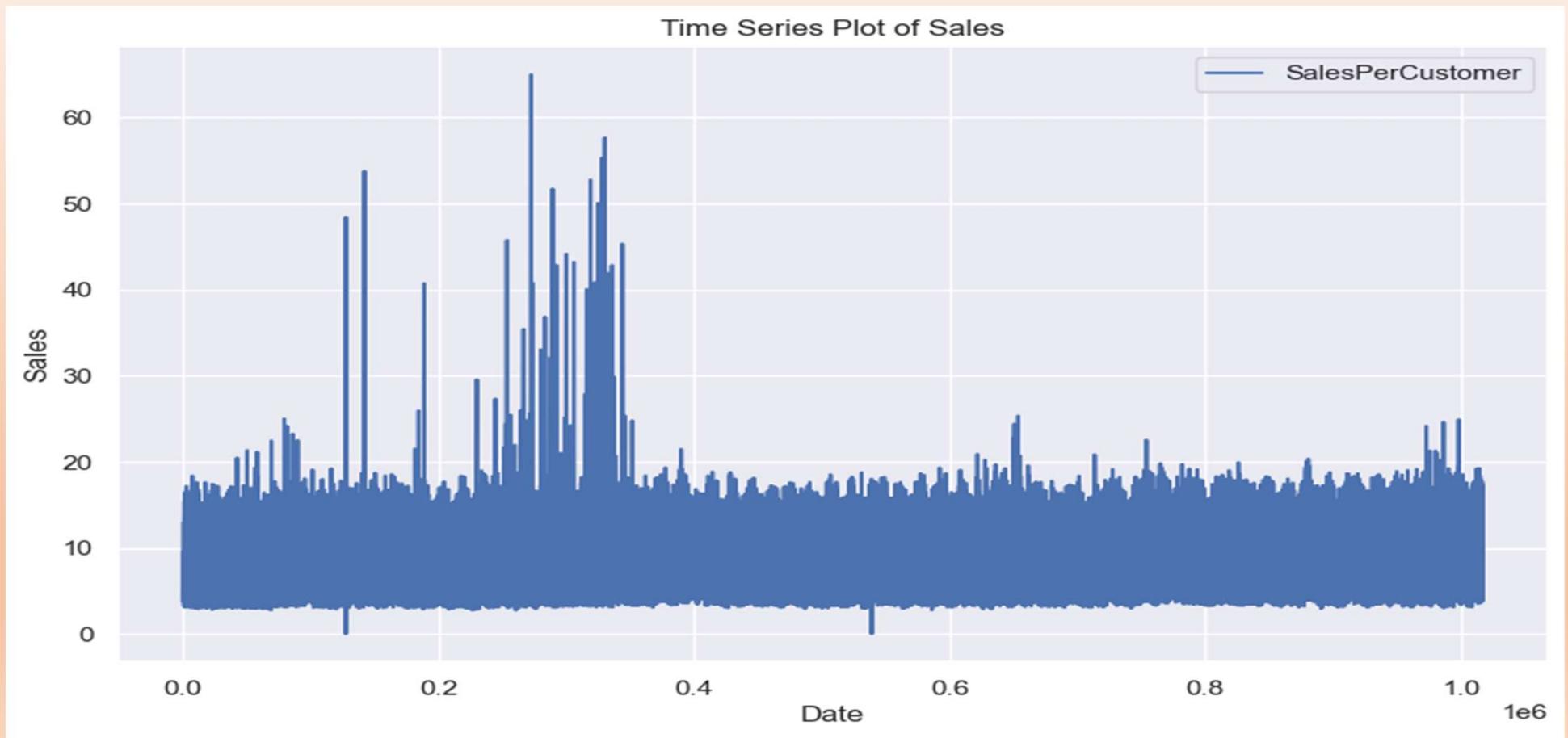
```
MAE for Store Type '0': 0.7969396724017819
MAE for Store Type '1': 0.7969396724017819
MAE for Store Type '2': 0.7969396724017819
MAE for Store Type '3': 0.7969396724017819
```



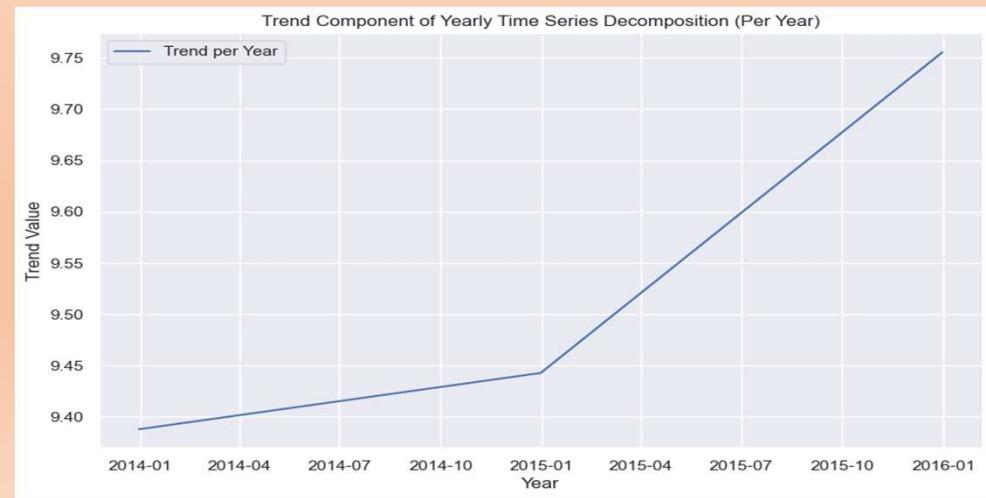
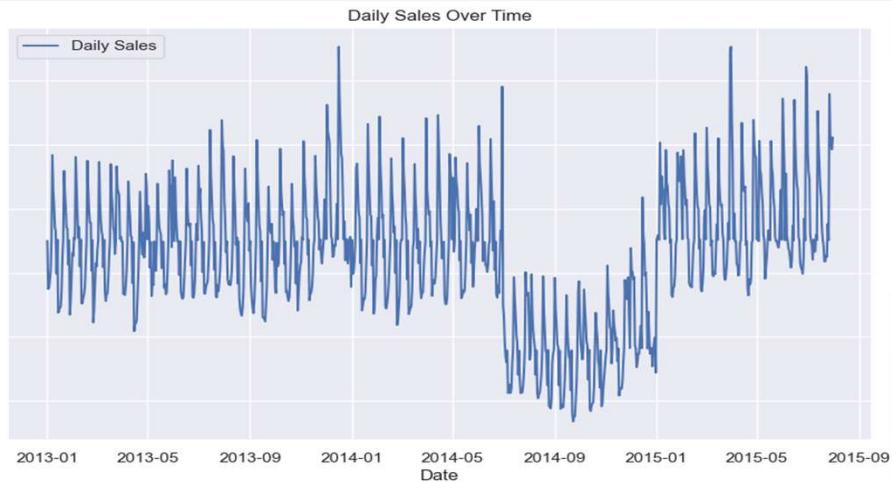
Class Interval

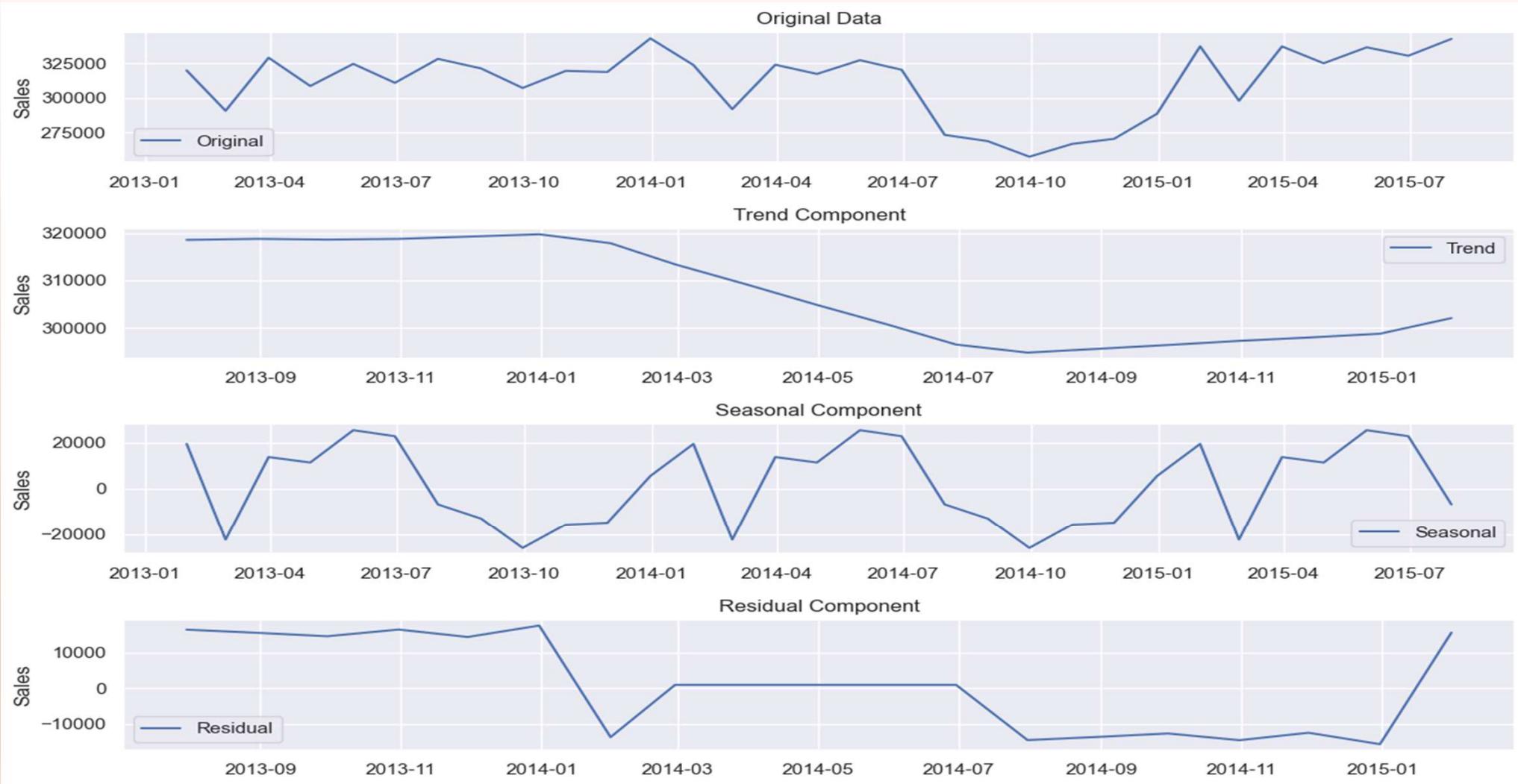


TIME SERIES ANALYSIS

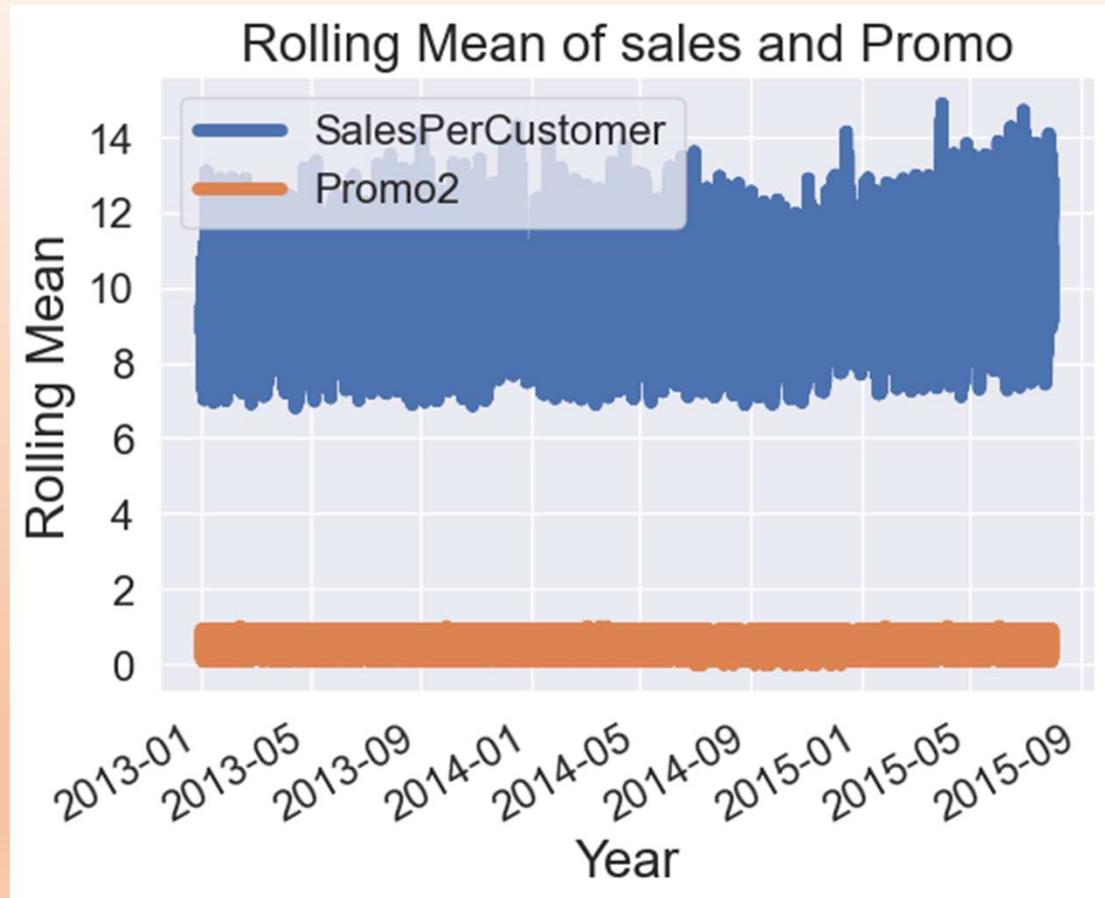


Trend Analysis

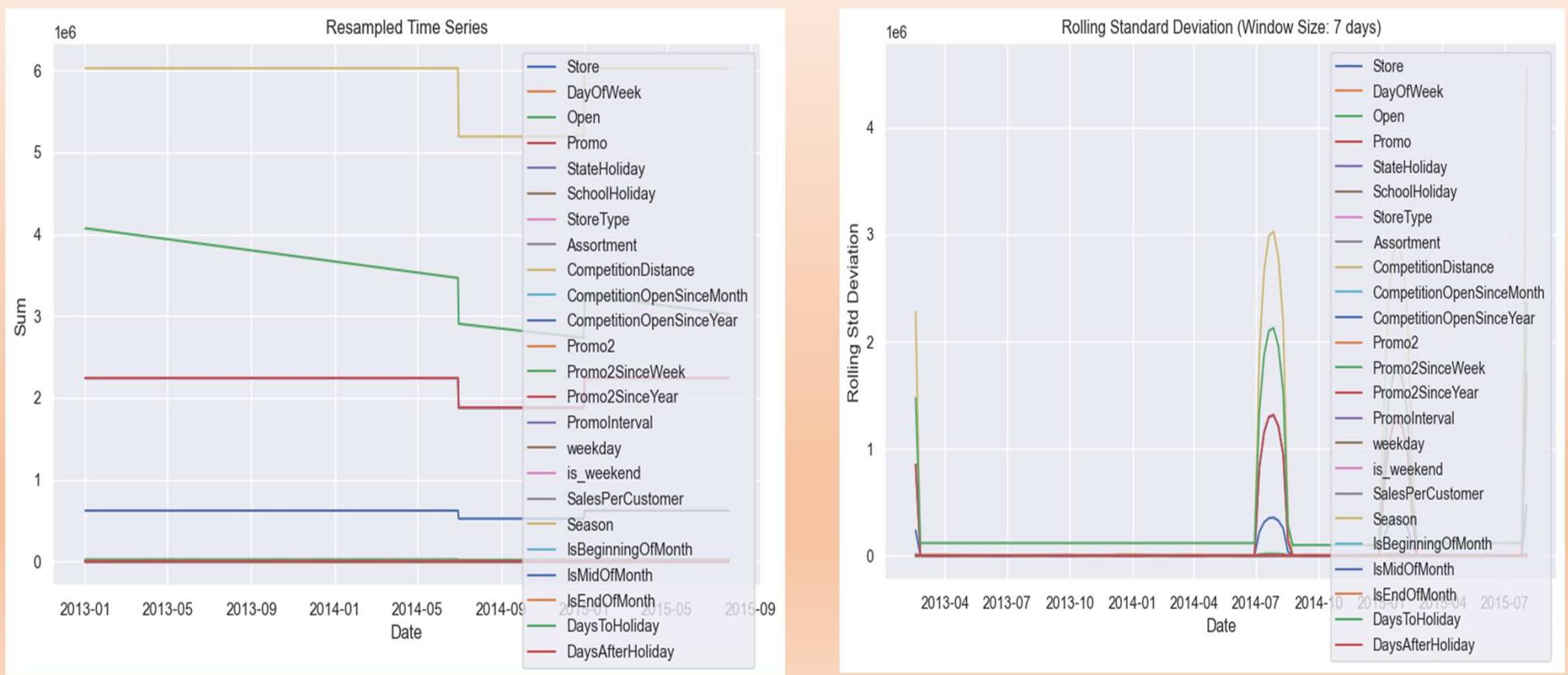




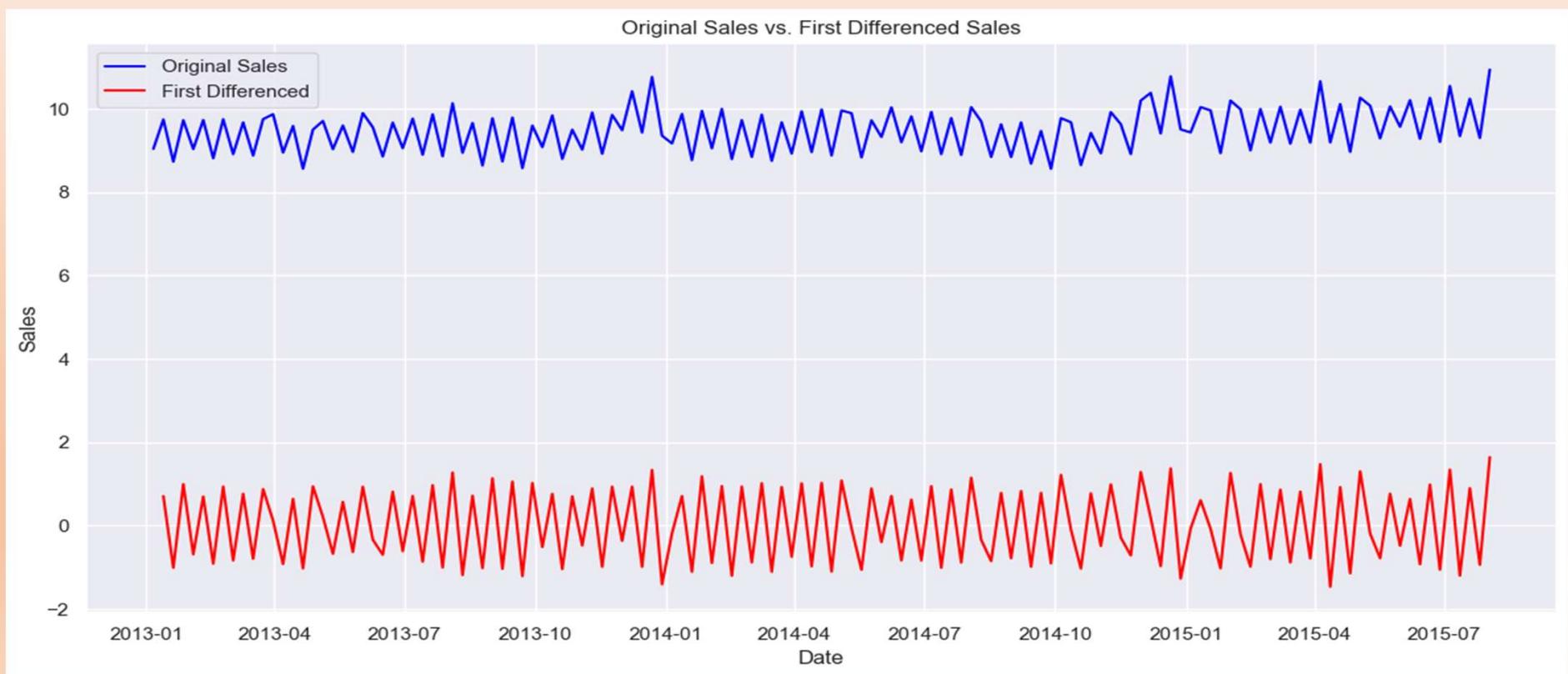
- Analyzing the rolling mean of sales within the range of 7-14 while considering promotional activity levels at a constant range of 0 to 1 can provide insights into how promotions may impact sales trends over a medium-term period.
- This analysis helps assess whether the presence or absence of promotions within this range is associated with specific patterns or fluctuations in sales behavior.



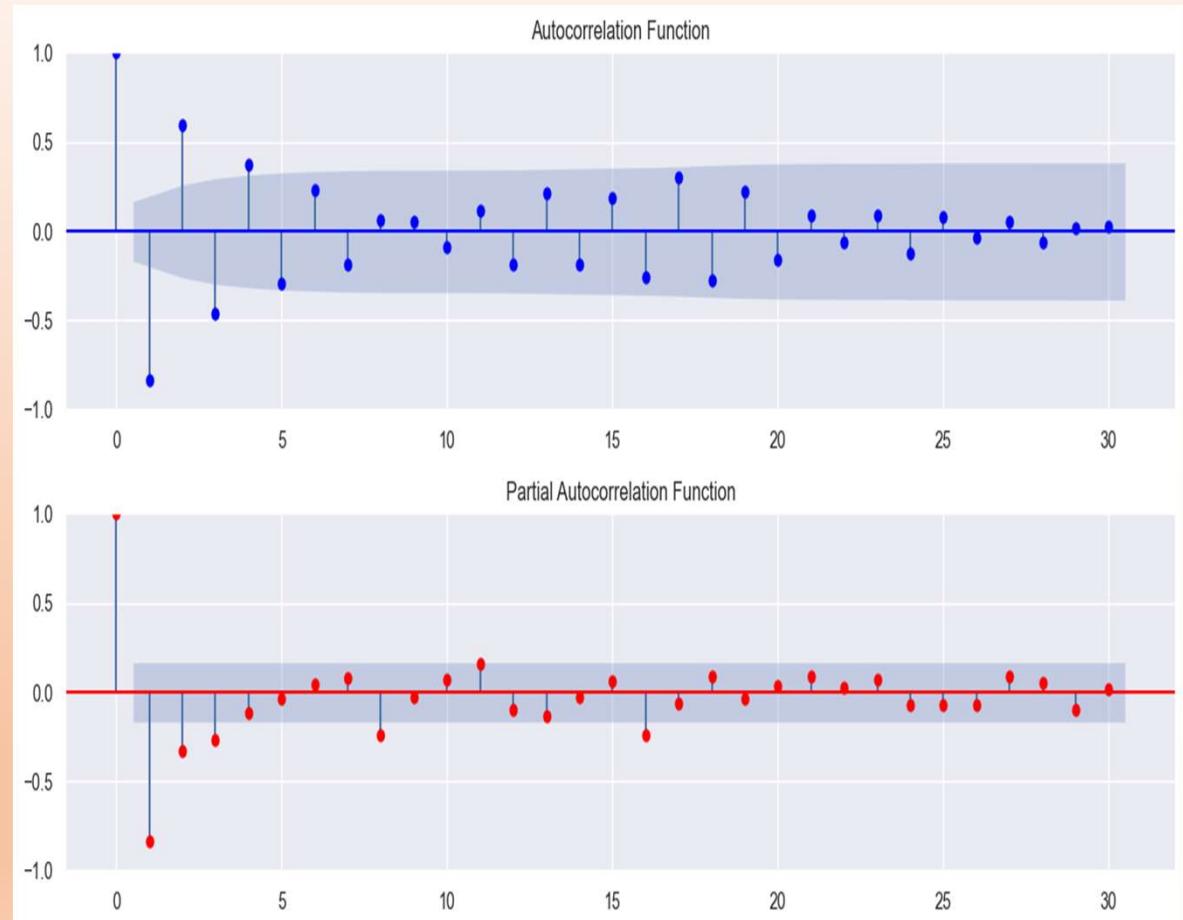
Both in Resampled Time series and Rolling Standard Deviation Plot we can see there is high fluctuation in the mid-year 2014



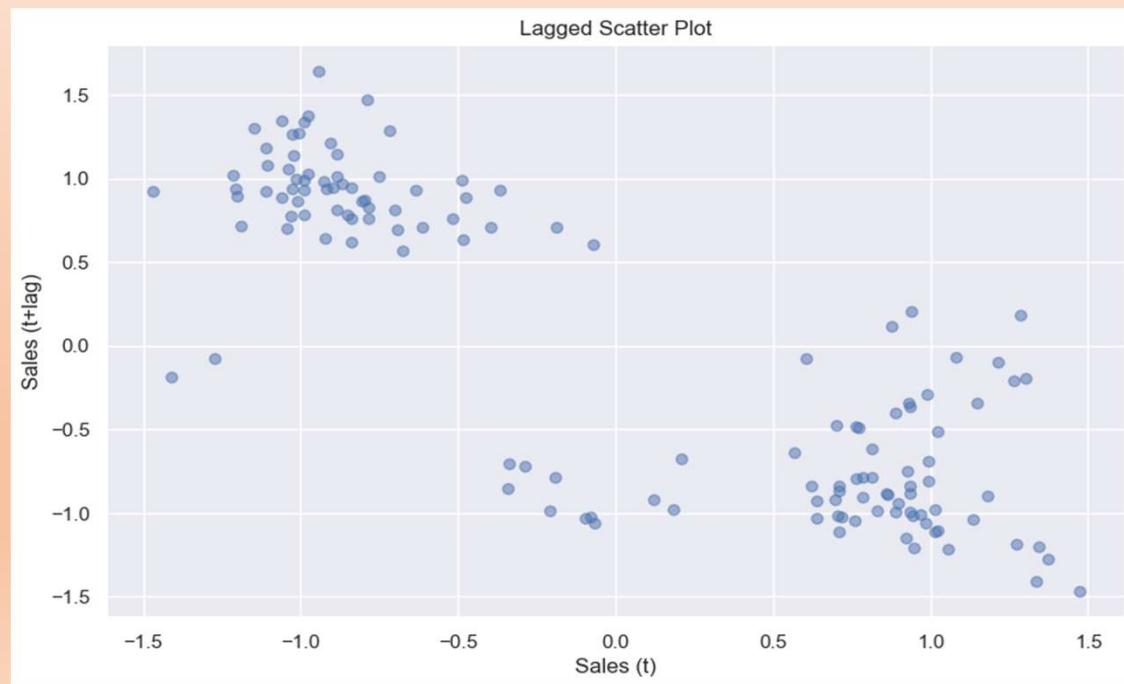
- First Differenced Sales represents the sales data after applying a differencing operation, specifically taking the difference between each data point and the previous one.
- The differencing process have reduced the overall variability or patterns in the data, making it appear more stable or stationary.



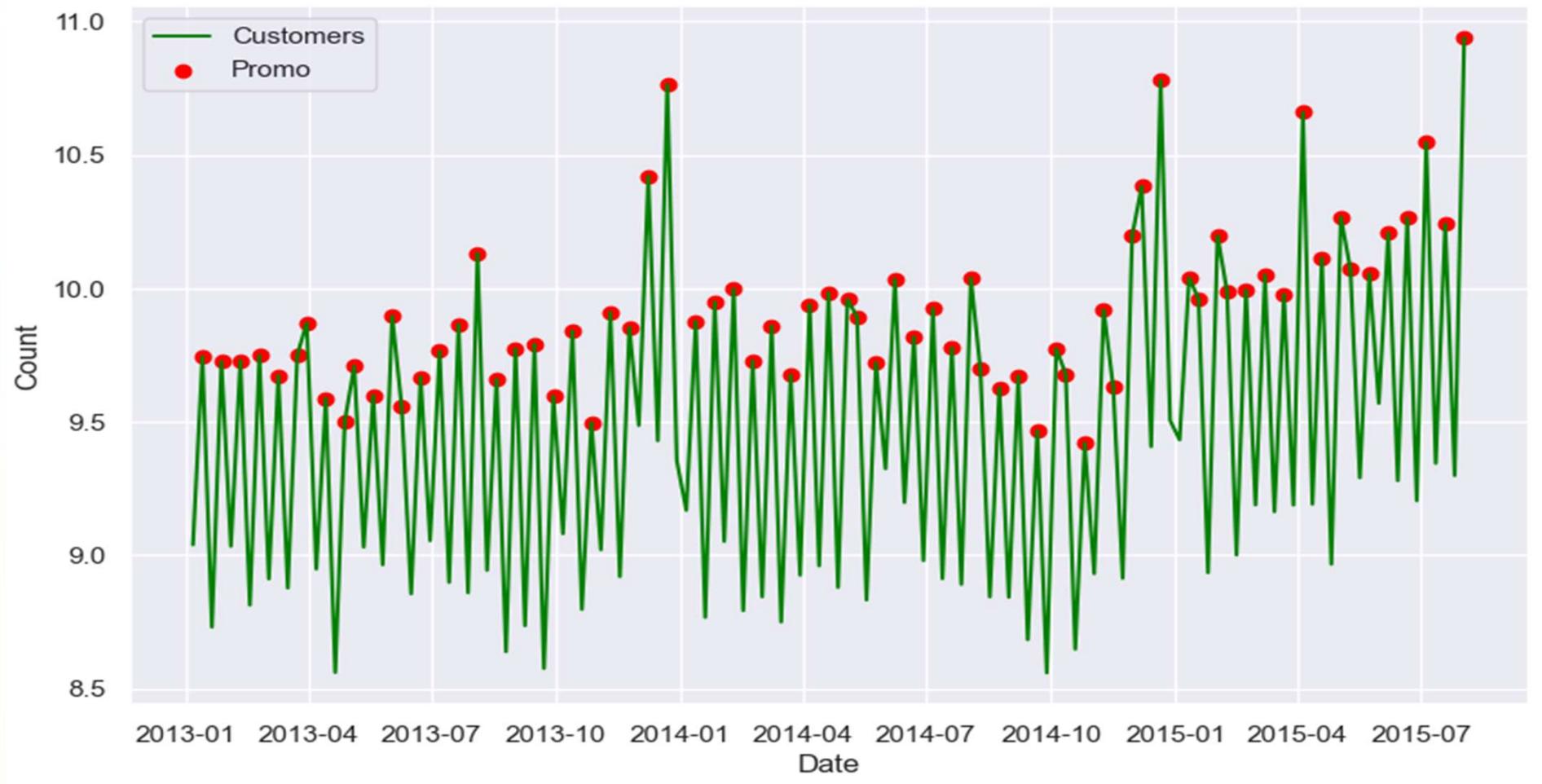
- A fluctuating ACF pattern with both falling and rising values typically signifies seasonality within the data. It suggests regular, repeating cycles at specific time intervals.
- On the other hand, when PACF values are notably lower than ACF values, it indicates that the direct correlation between the current observation and its past lags is relatively weak compared to the influence of other lags or variables.

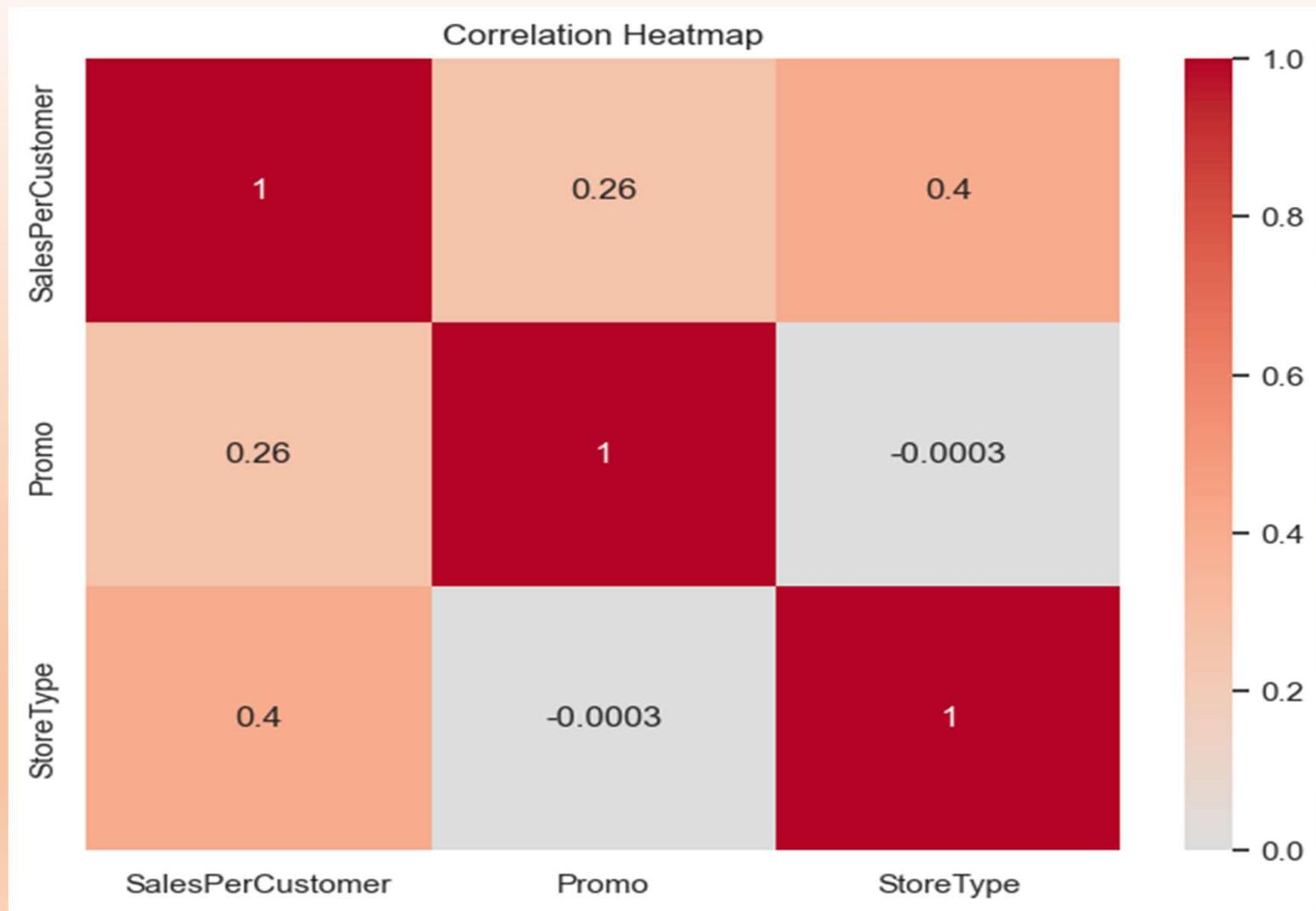


- An increase in the current time period, denoted as ' t ', leading to an increase in the future time period ' $t+lag$ ' indicates a positive correlation or dependency between the two time points.
- In other words, as time progresses from ' t ' to ' $t+lag$ ', any changes, improvements, or variations occurring at ' t ' are expected to have a corresponding impact on the future time period ' $t+lag$ '.

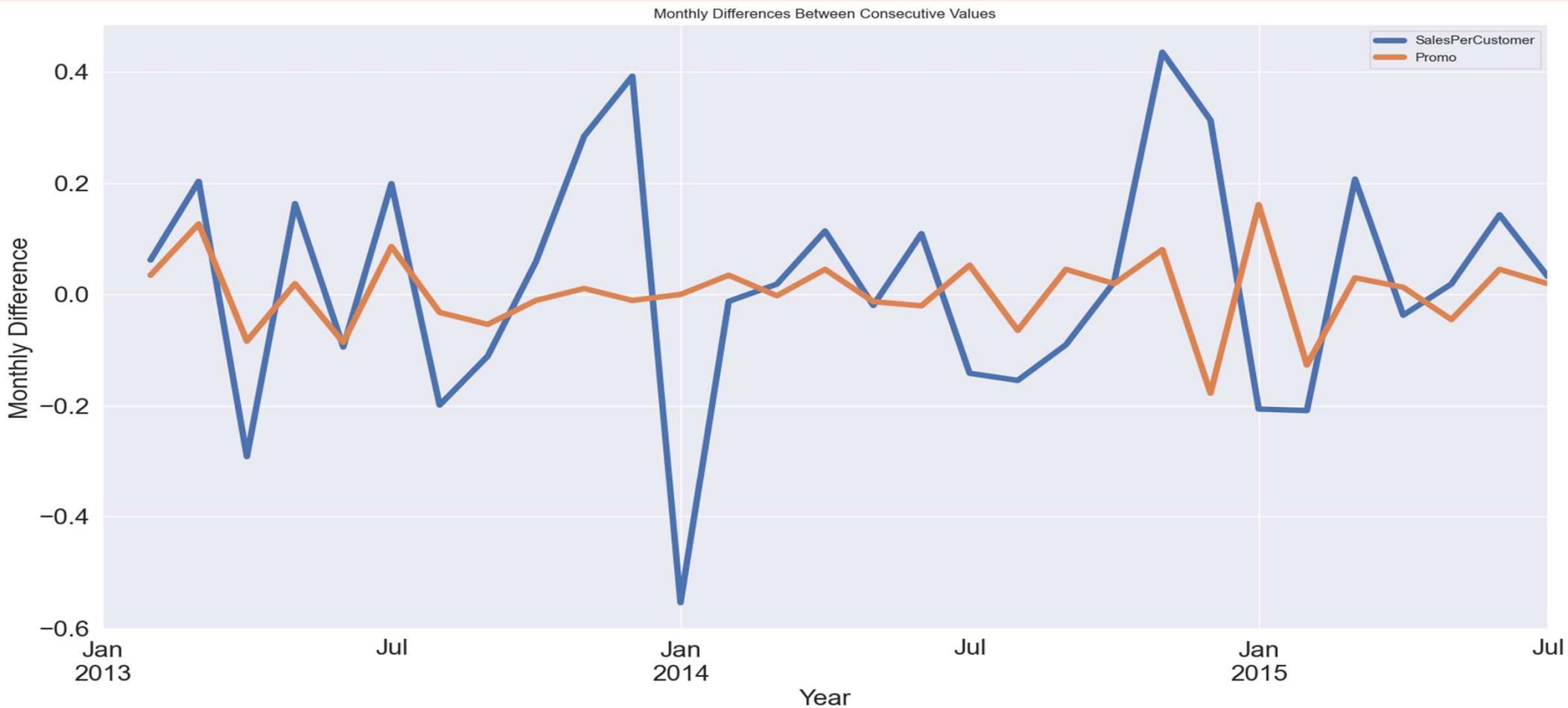


Customers and Sales Over Time with Promo Highlights

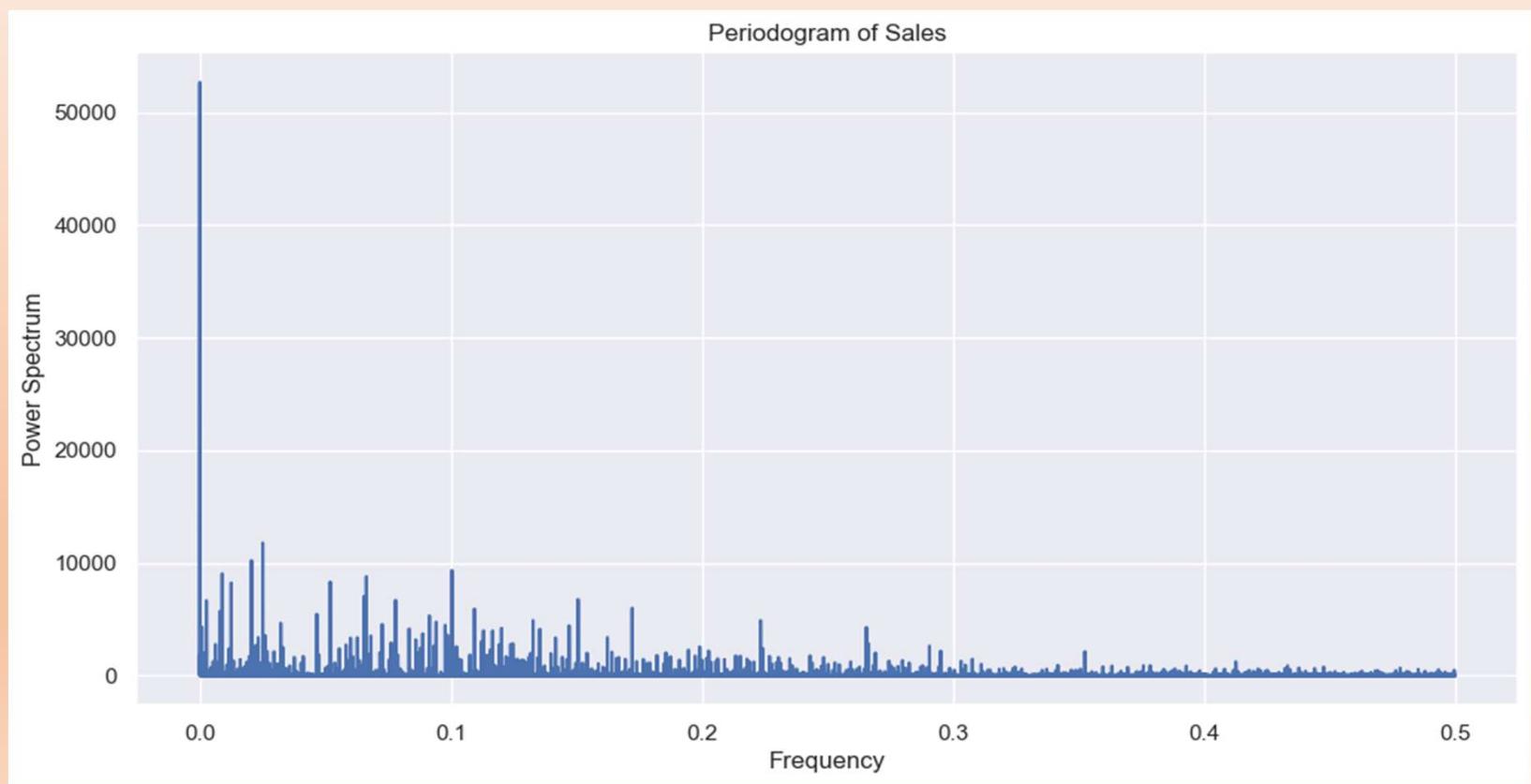


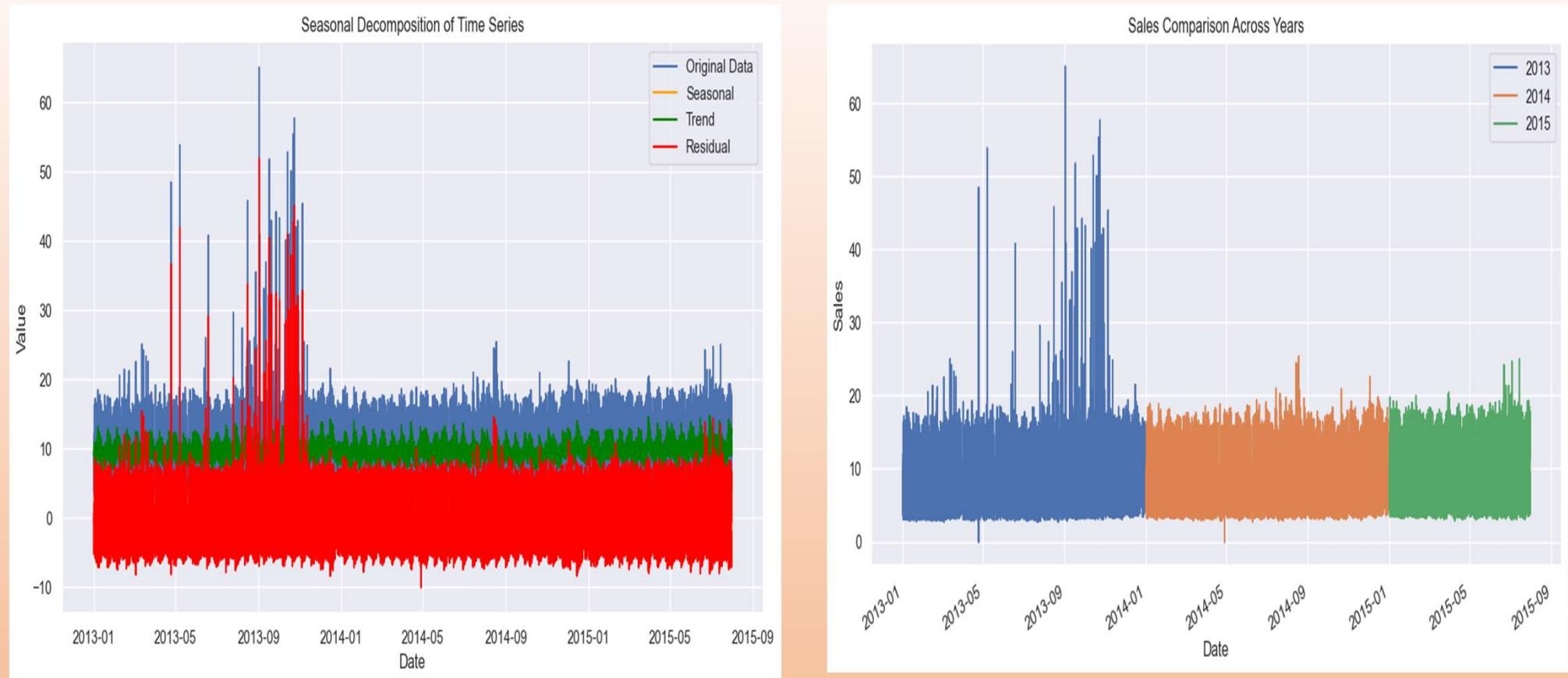


Regular promotions are conducted, but these promotional activities do not significantly increase the number of customers; however, they do lead to an increase in sales.

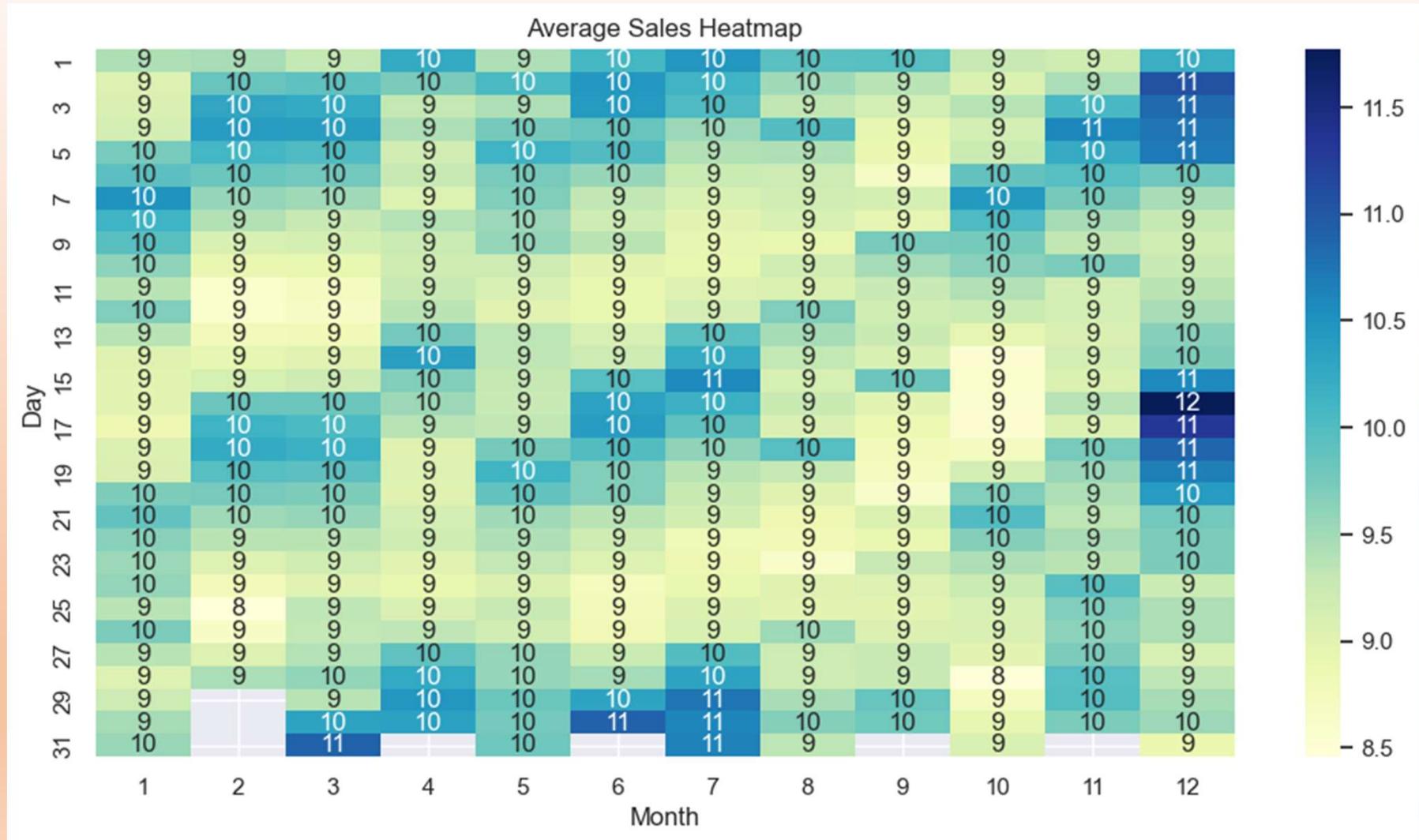


A periodogram showing high spectral power near 0 and decreasing power as it approaches 1 suggests that the sales data is characterized by a dominant low-frequency component, possibly indicating long-term trends or seasonality

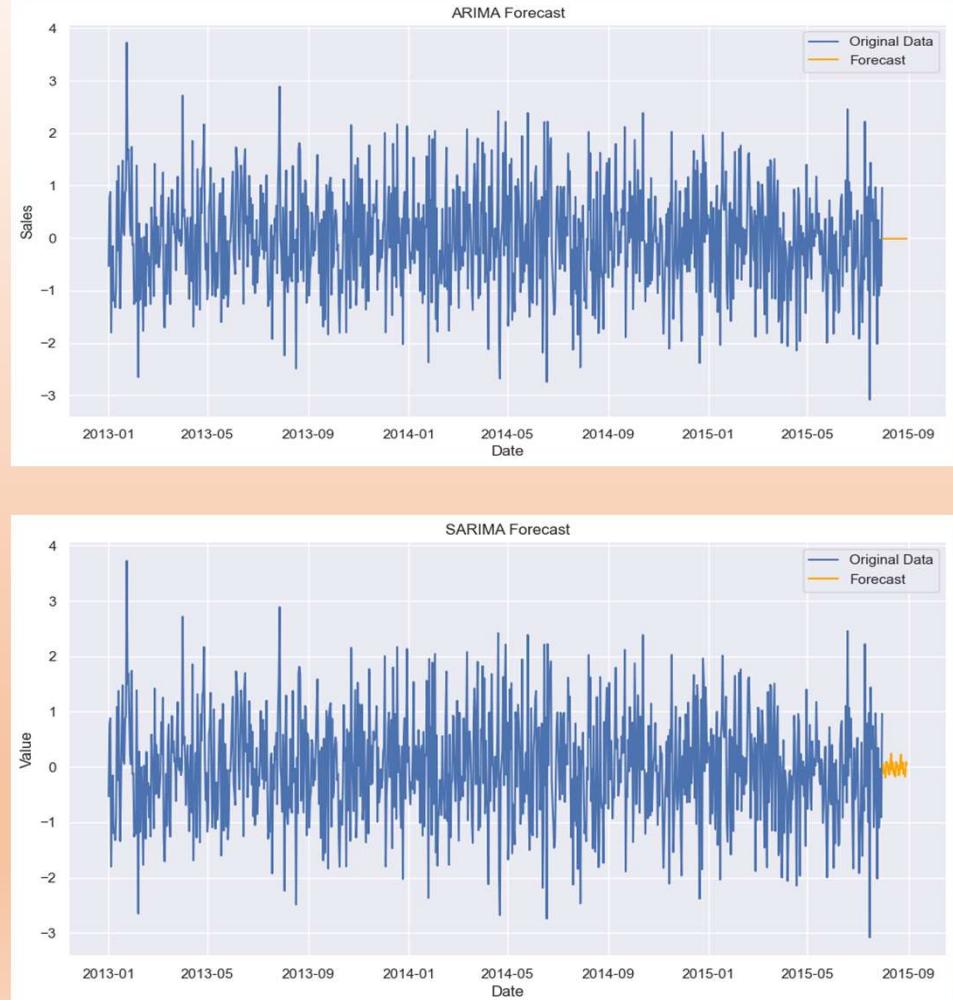




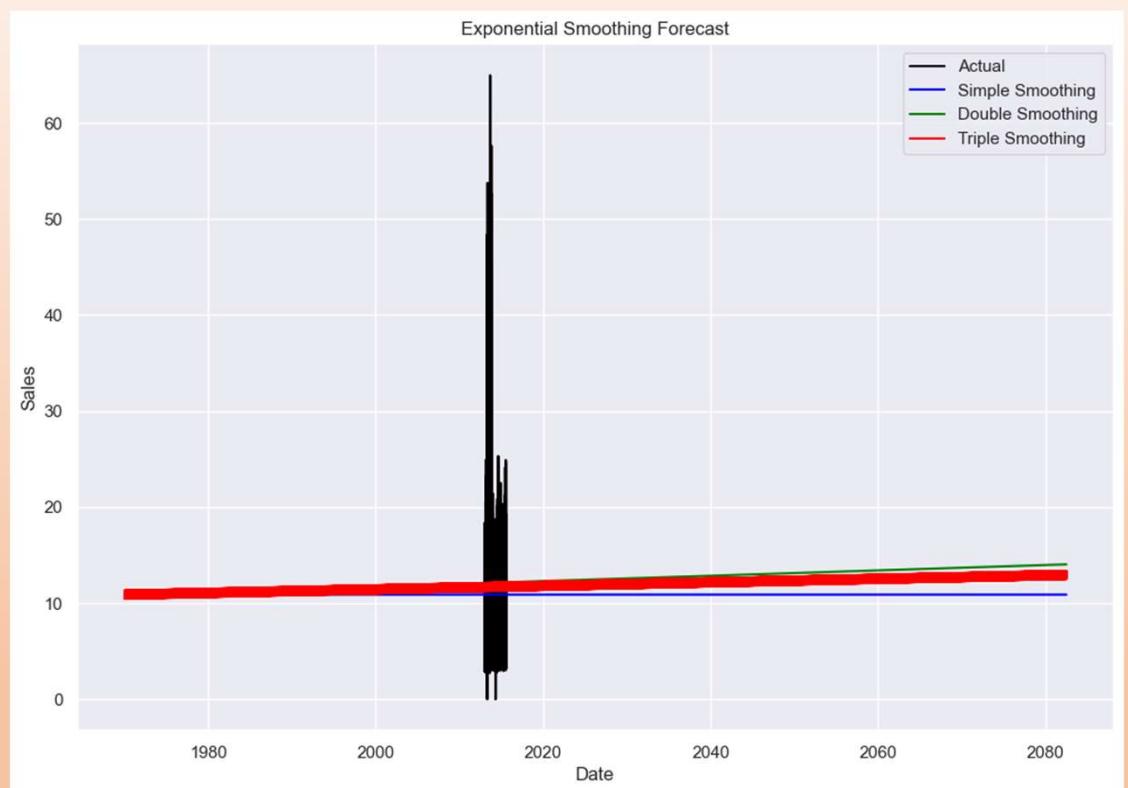
The sales data exhibits low variation, with a substantial increase in sales in the year 2013, followed by consistent sales levels in the subsequent years of 2014 and 2015.

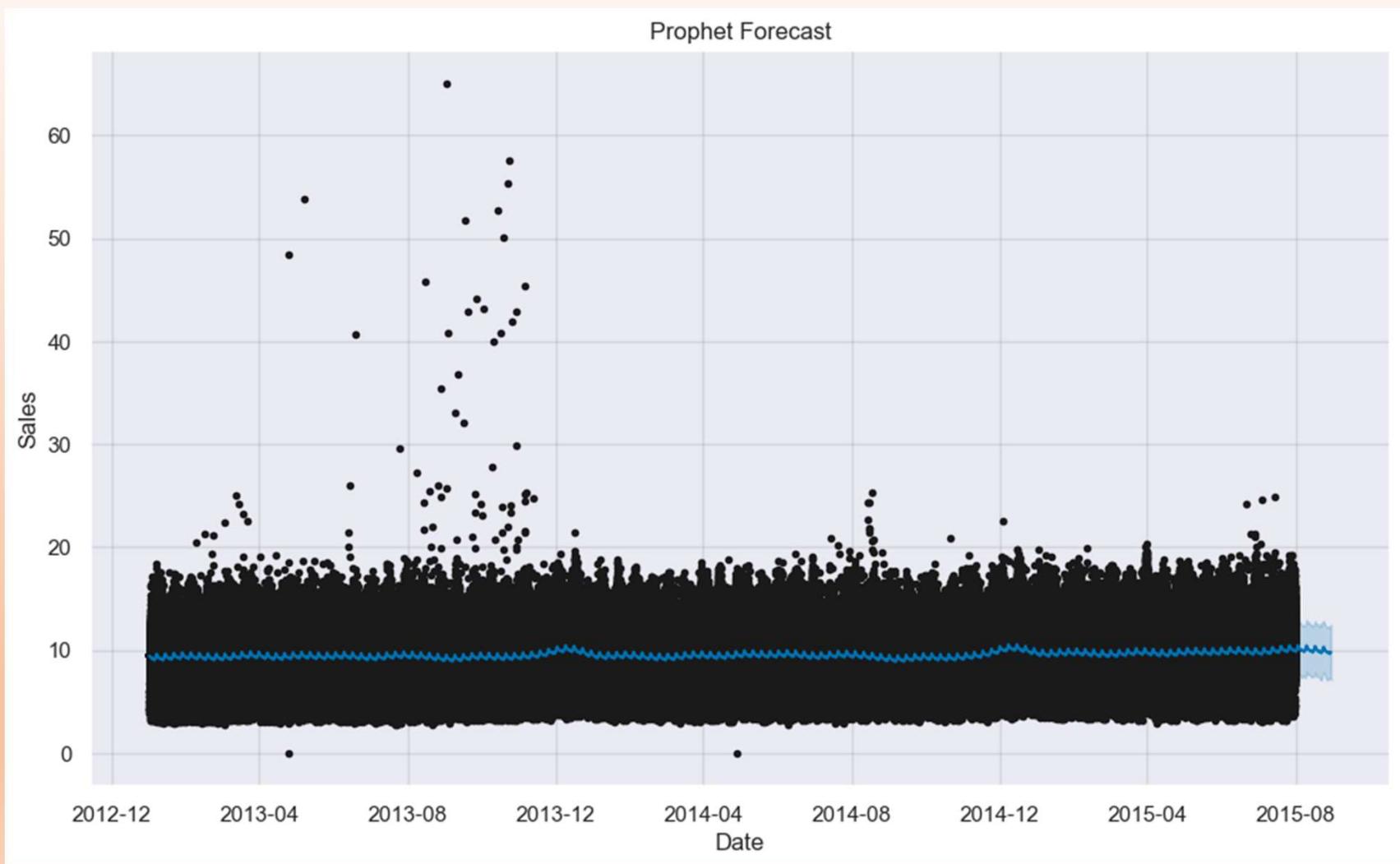


- In this stepwise search for the best ARIMA model, various combinations were tested to minimize the AIC (Akaike Information Criterion), which is a measure of a model's goodness of fit. The presented models represent different combinations of autoregressive (AR), differencing (I), and moving average (MA) components, as well as seasonal components (S).
- The best-fitting model identified is ARIMA(0,1,0)(1,1,0)[12], which suggests a non-seasonal differencing order of 1, no autoregressive or moving average components, and a seasonal differencing order of 1 with a seasonal component of 12 (indicating seasonality at a 12-month interval).
- The results of this best-fitting model show that it has the lowest AIC value (408.305), indicating a good fit to the data. The coefficients and diagnostic statistics are also provided, offering insights into the model's performance and its ability to capture the underlying patterns in the time series data.

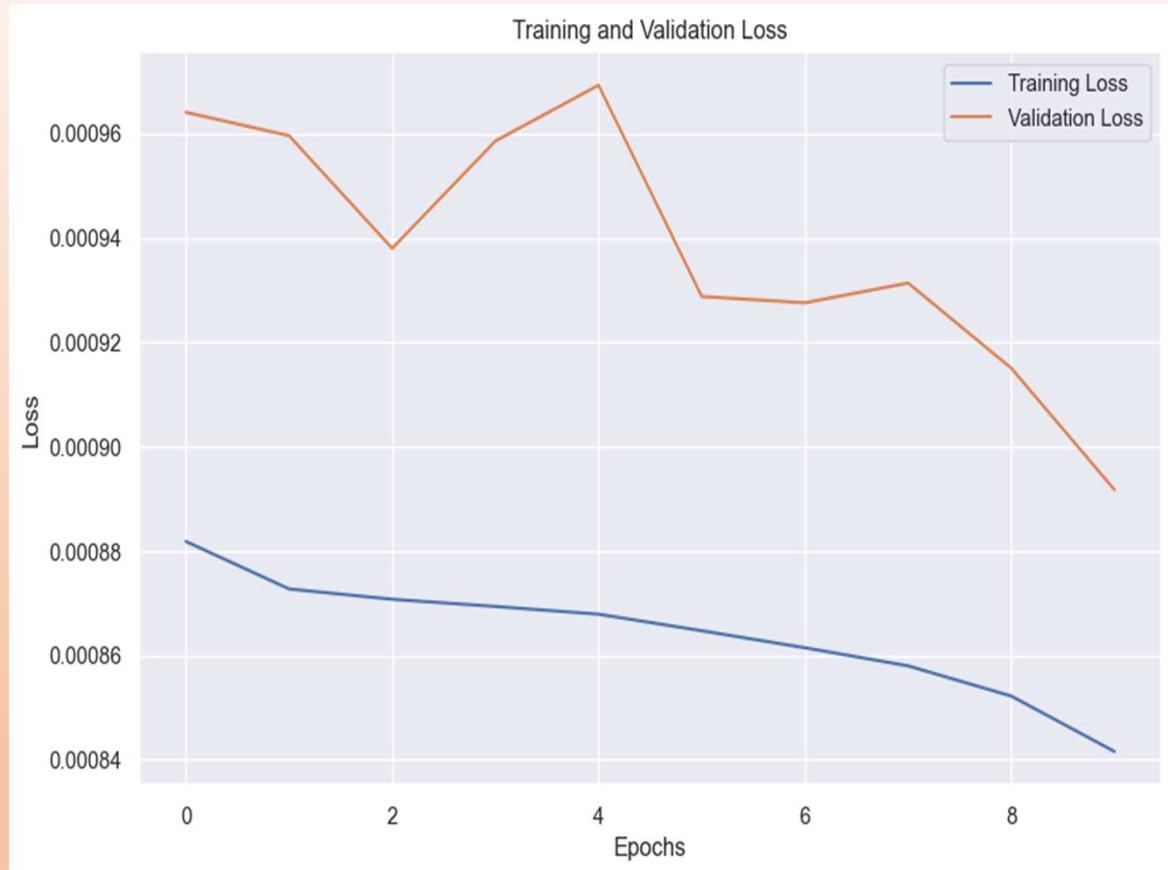


- In exponential smoothing, the actual values tend to deviate from the predicted values.
- Simple smoothing results in a relatively constant sales trend, hovering around 10.
- Double smoothing introduces a slightly more pronounced increase, with sales reaching around 15.
- Triple smoothing falls between the levels of simplicity observed in simple and double smoothing methods.

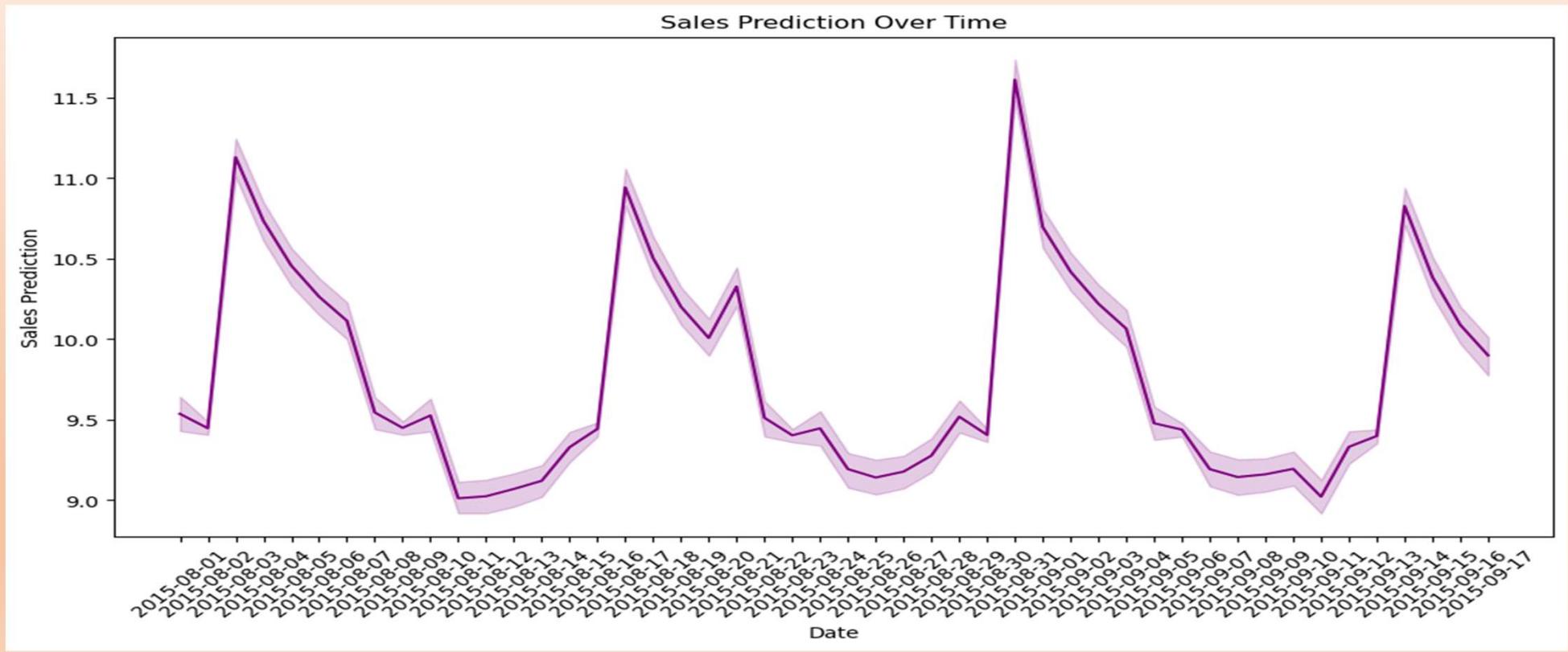




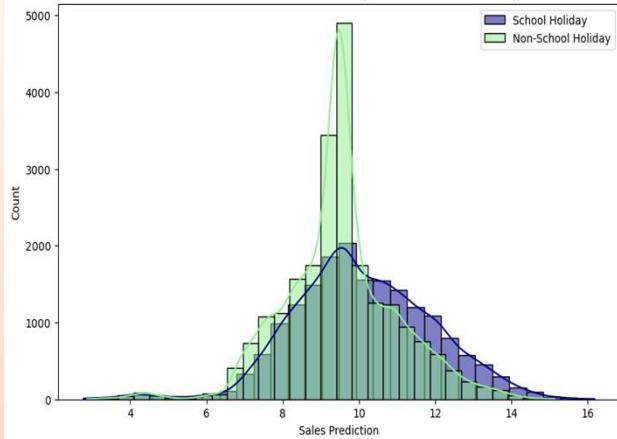
- The training loss of 0.00088 and the validation loss of 0.00096, both sustained over 8 epochs, indicate a model that has achieved a high level of accuracy and generalization.
- The training loss reflects how well the model is fitting the training data, and the validation loss gauges the model's performance on unseen data.



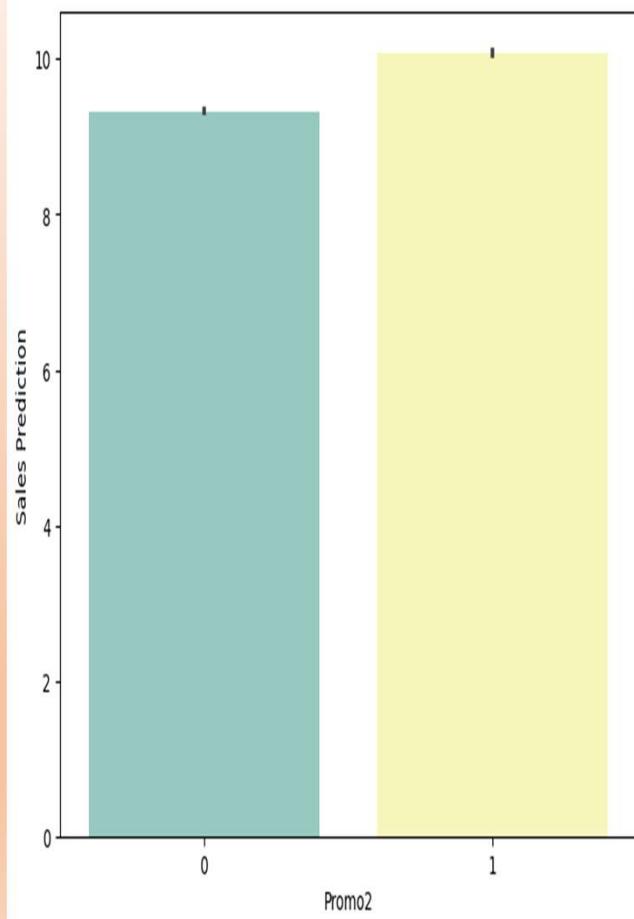
Sales Predicted for Test Data



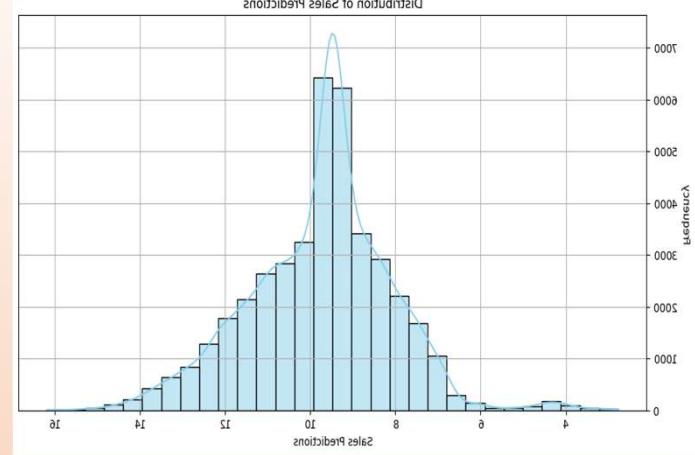
Sales Distribution on School Holidays vs. Non-School Holidays



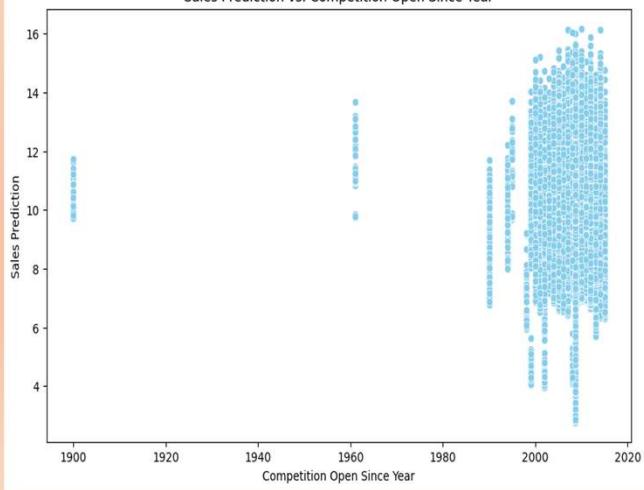
Sales Prediction for Promo2 vs. Non-Promo2 Stores



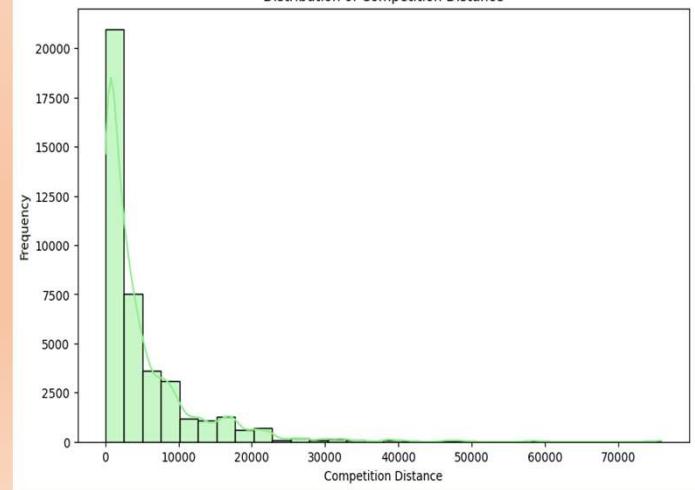
Distribution of Sales Predictions



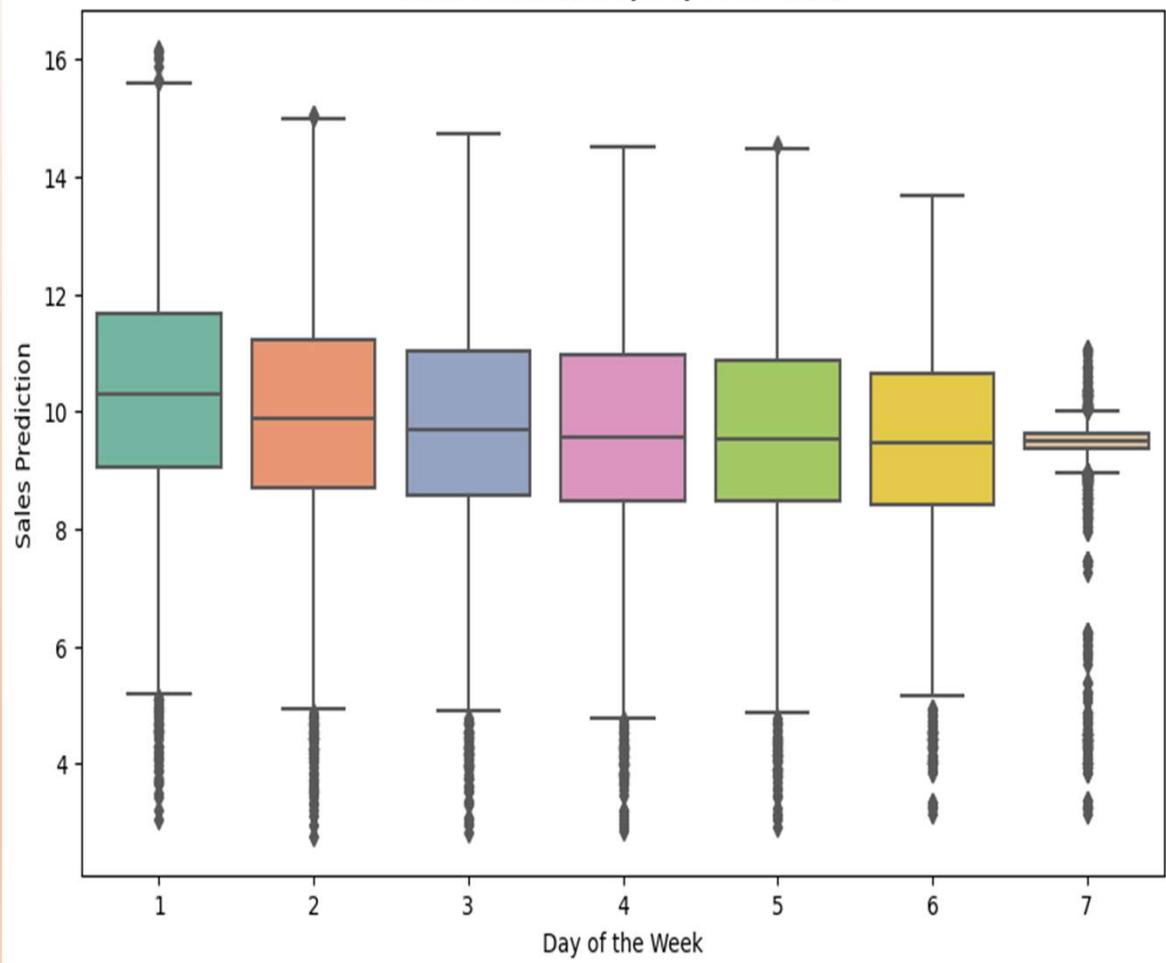
Sales Prediction vs. Competition Open Since Year



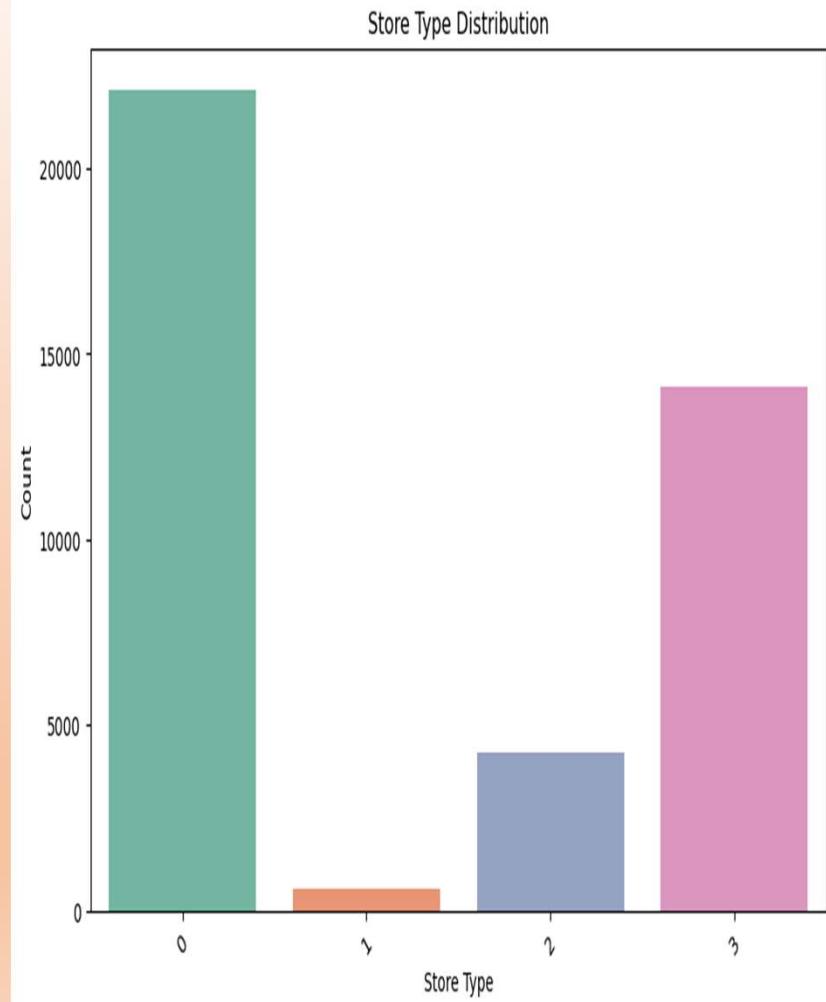
Distribution of Competition Distance

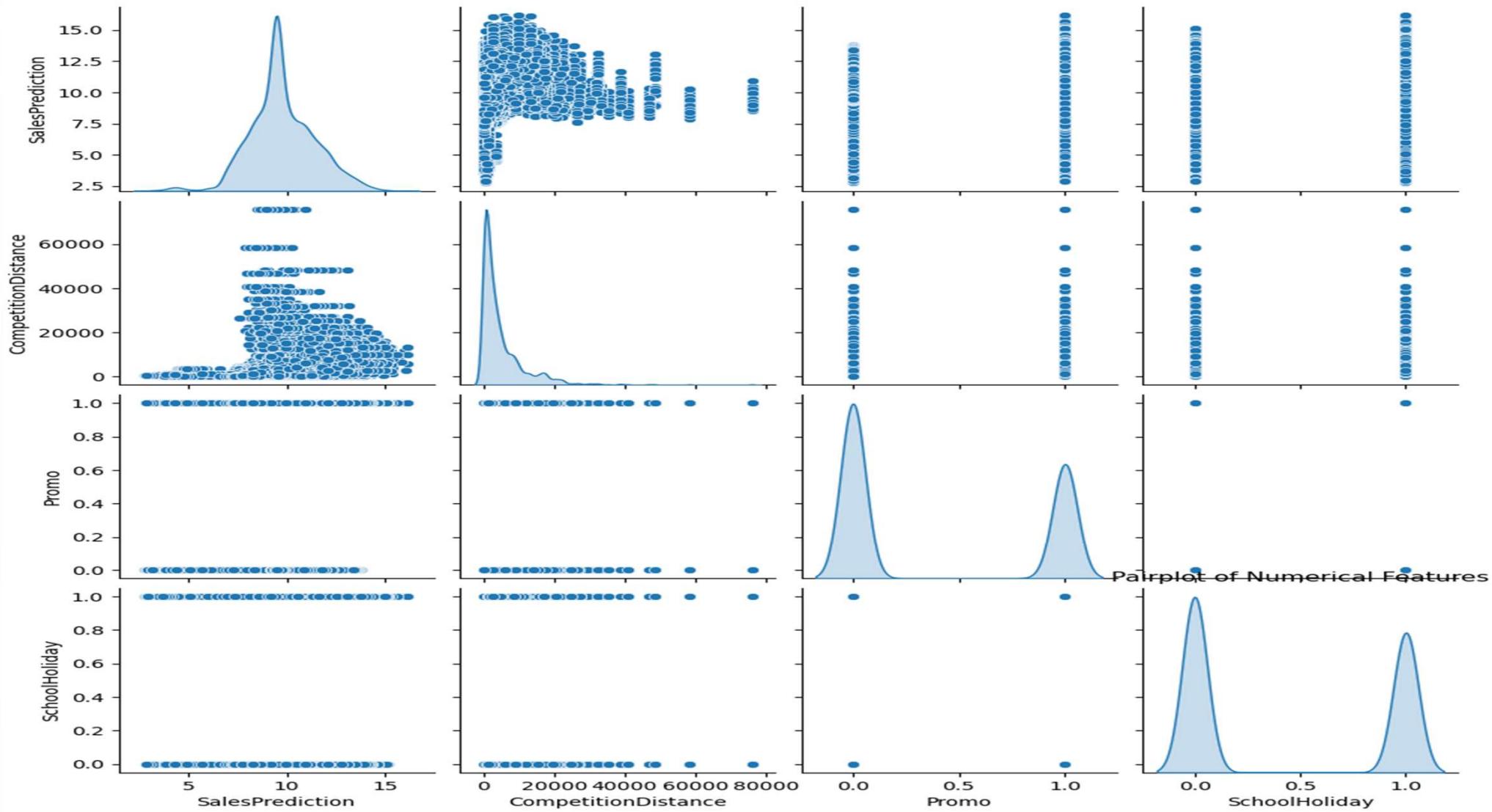


Sales Distribution by Day of the Week

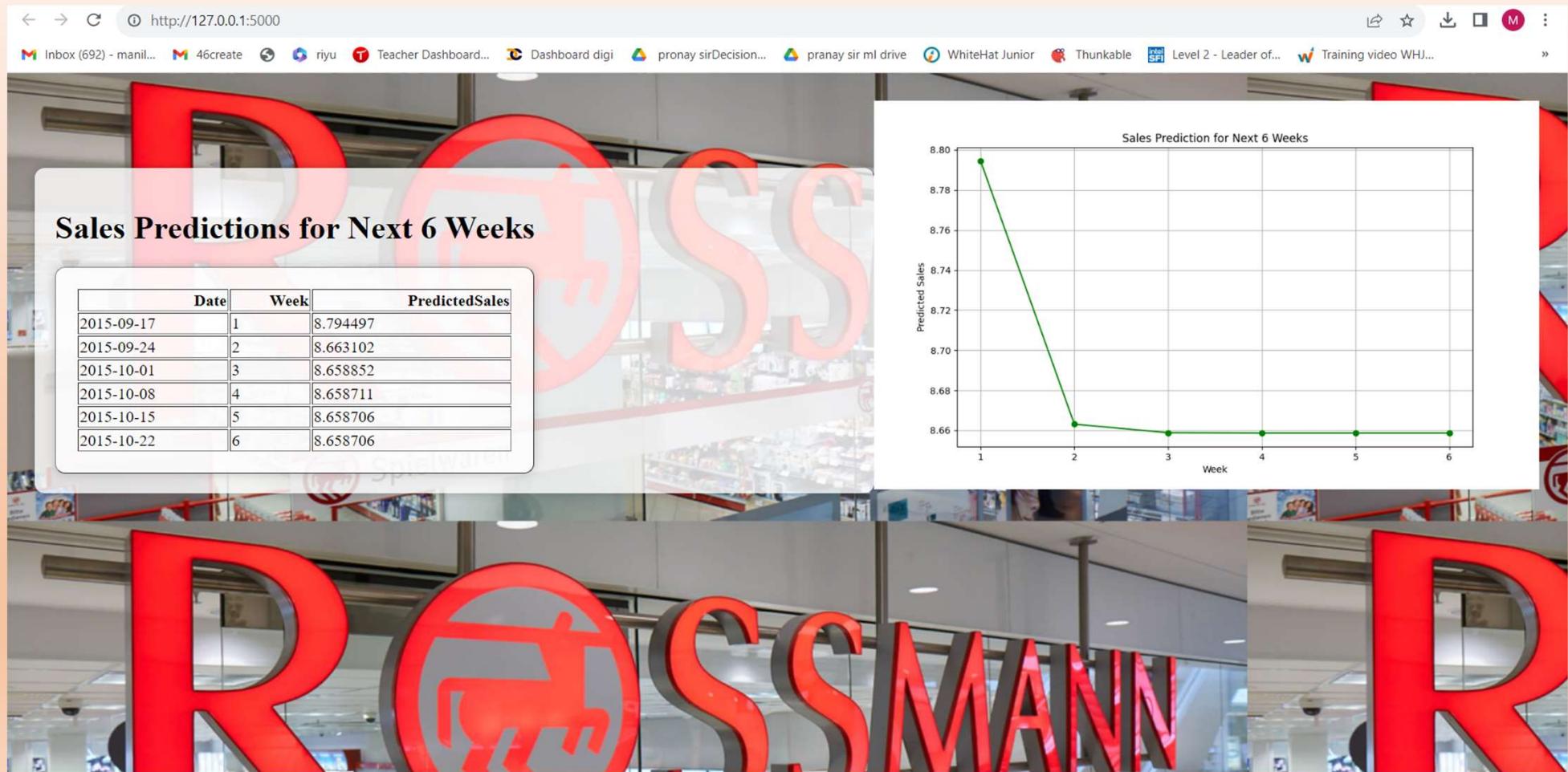


Store Type Distribution





Sales Prediction for next 6 Weeks



THANK YOU