

A circular wreath of various botanical illustrations surrounds the central text. The wreath includes green ferns, a red maple leaf, a green bell-shaped flower, a branch with small pink blossoms, a large green leaf, a red leaf, and purple flowers.

PHARMACEUTICAL SALES PREDICTION ACROSS MULTIPLE STORES



AGENDA

1. INTRODUCTION
2. EXPLORATION OF CUSTOMER PURCHASING BEHAVIOUR
3. PREDICTION OF STORE SALES
4. SERVING PREDICTIONS ON A WEB INTERFACE
5. SUMMARY



Introduction

We seek to explore the behaviour of customers in the various stores. Our goal is to check how some measures such as promos and opening of new stores affect purchasing behavior and also to build and serve an end-to-end product that delivers this prediction to analysts in the finance team.





Primary goals

To forecast Sales for 6 weeks ahead



Train and Test Data

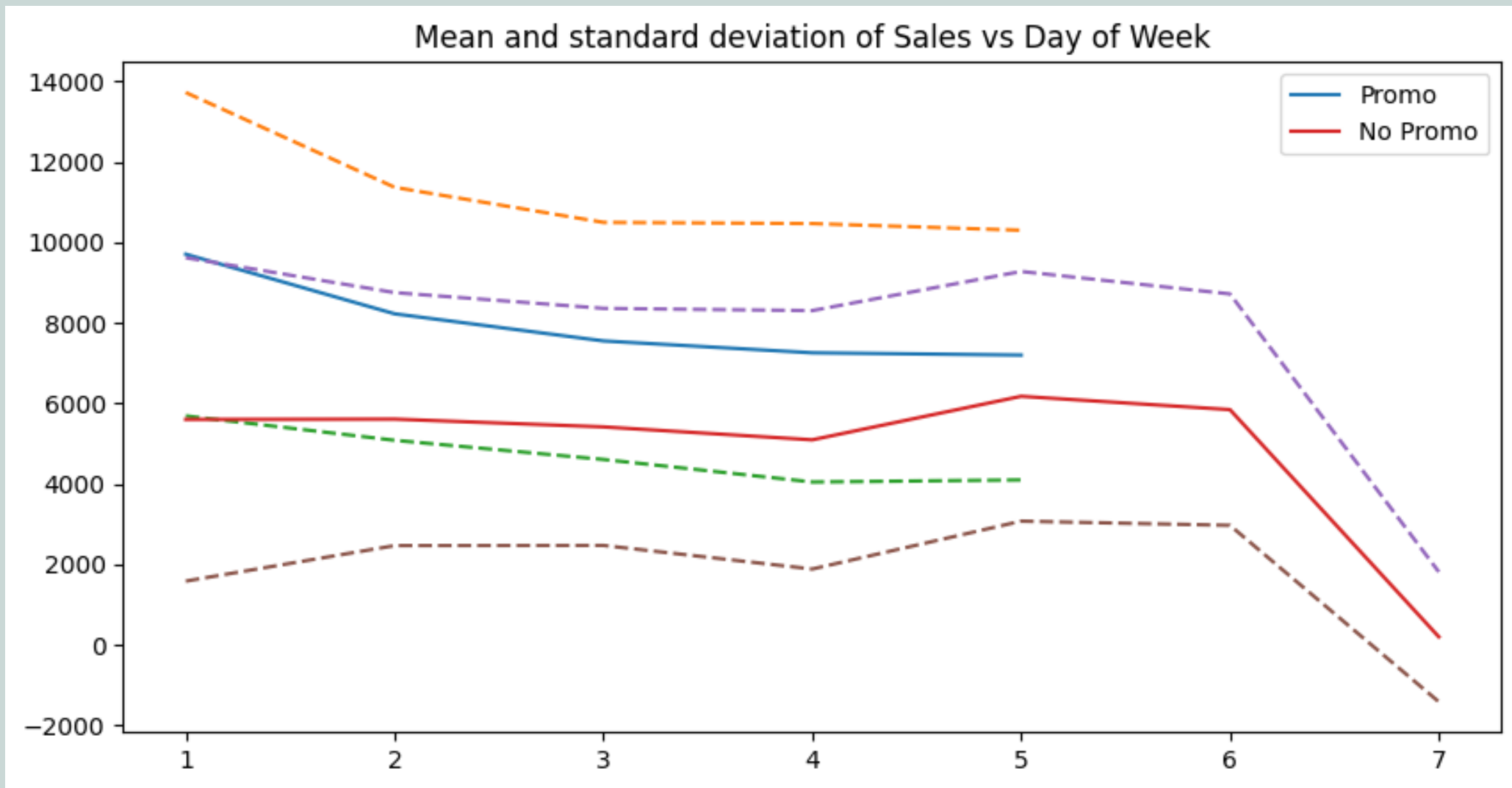
- **Train**

- min 2013-01-01 00:00:00
- max 2015-07-31 00:00:00
- delta: 941 days 00:00:00

- **Test**

- min 2015-08-01 00:00:00
- max 2015-09-17 00:00:00
- delta: 47 days 00:00:00

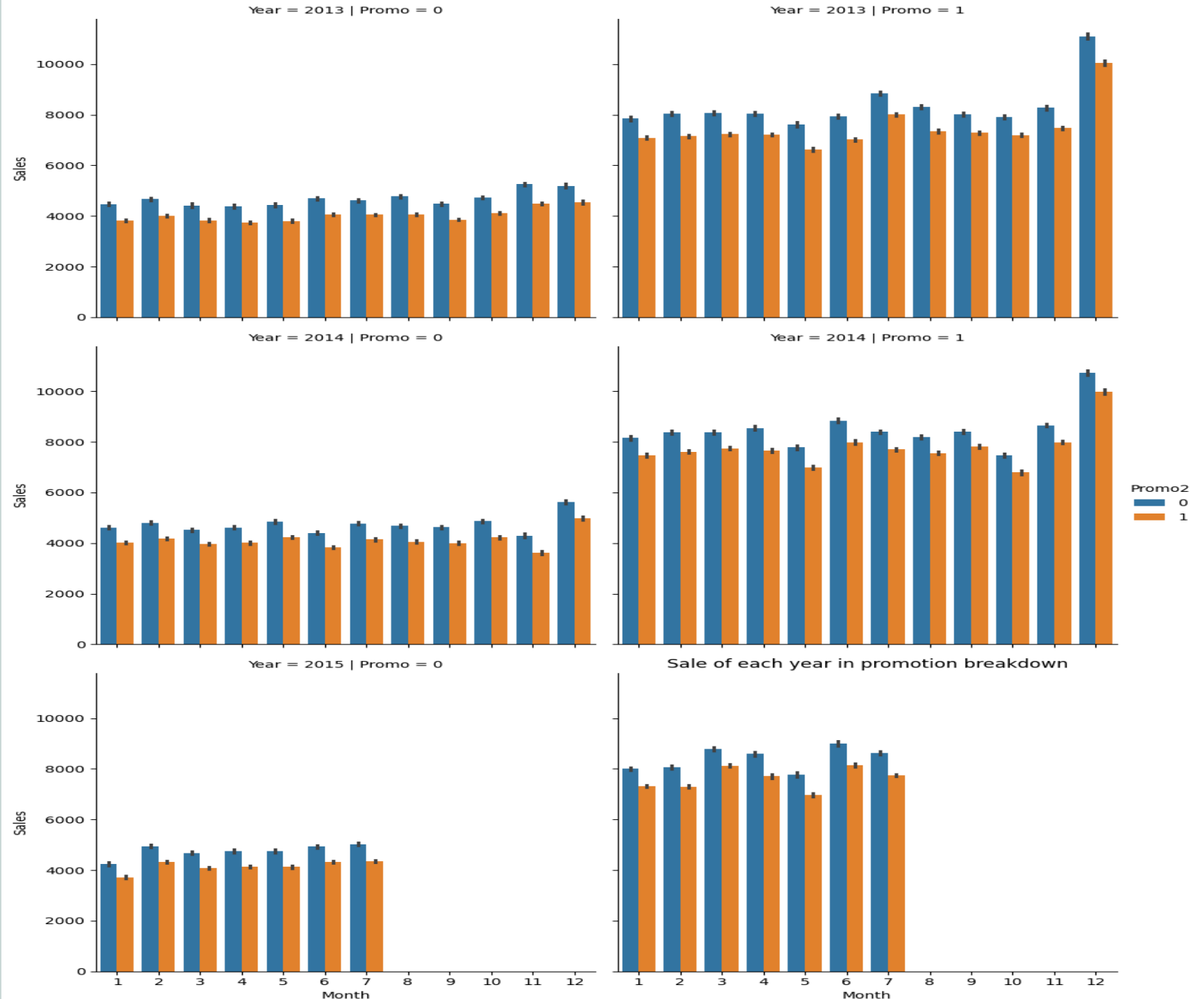
- 95% days are for training and 5% days are for testing



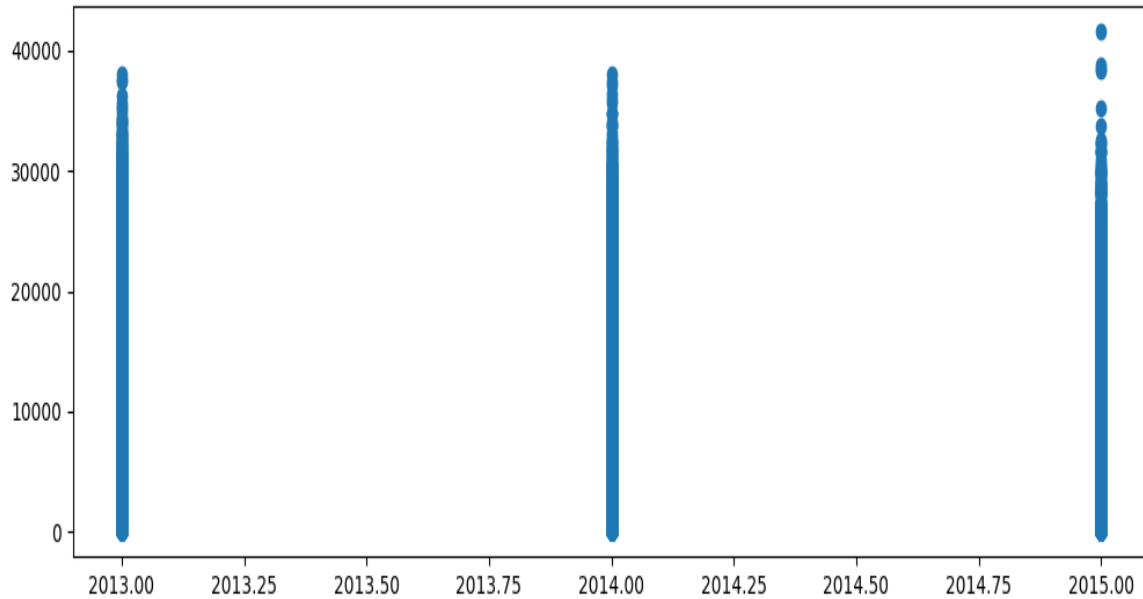
Promotions are correlated with more sales and customers

Sale of each year in promotion breakdown

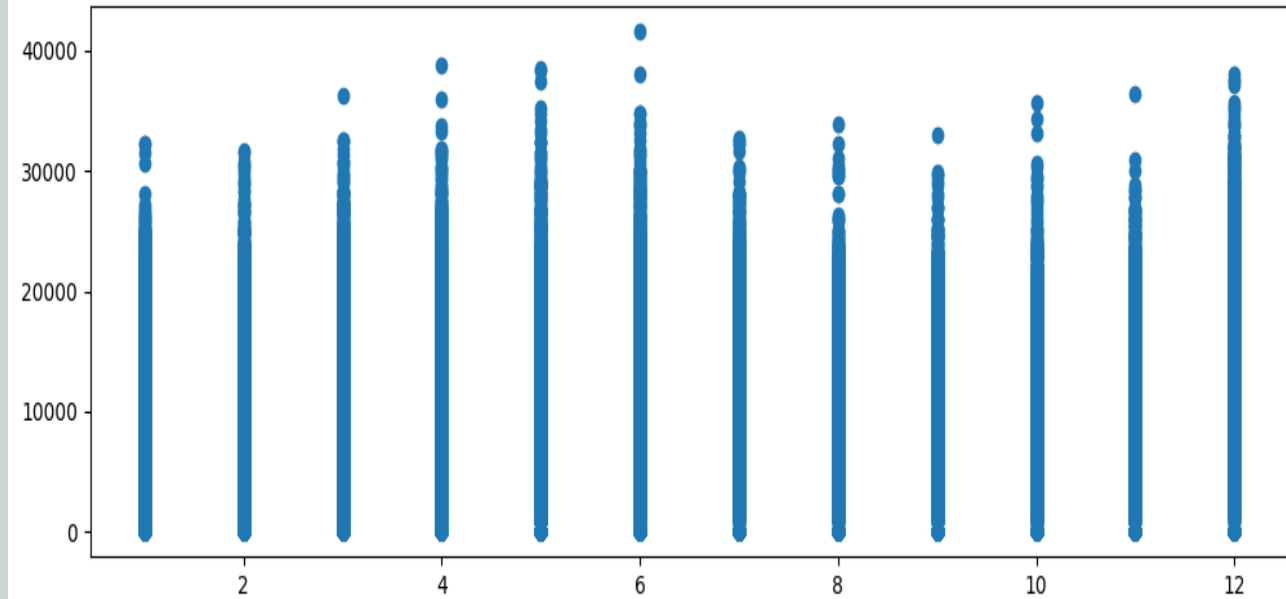
- We see that the campaigns carried out by the stores themselves and the campaigns jointly organized by the company have an increasing effect on sales.
- Strong sales are achieved when participating in the corporate campaign in the last month of the year.



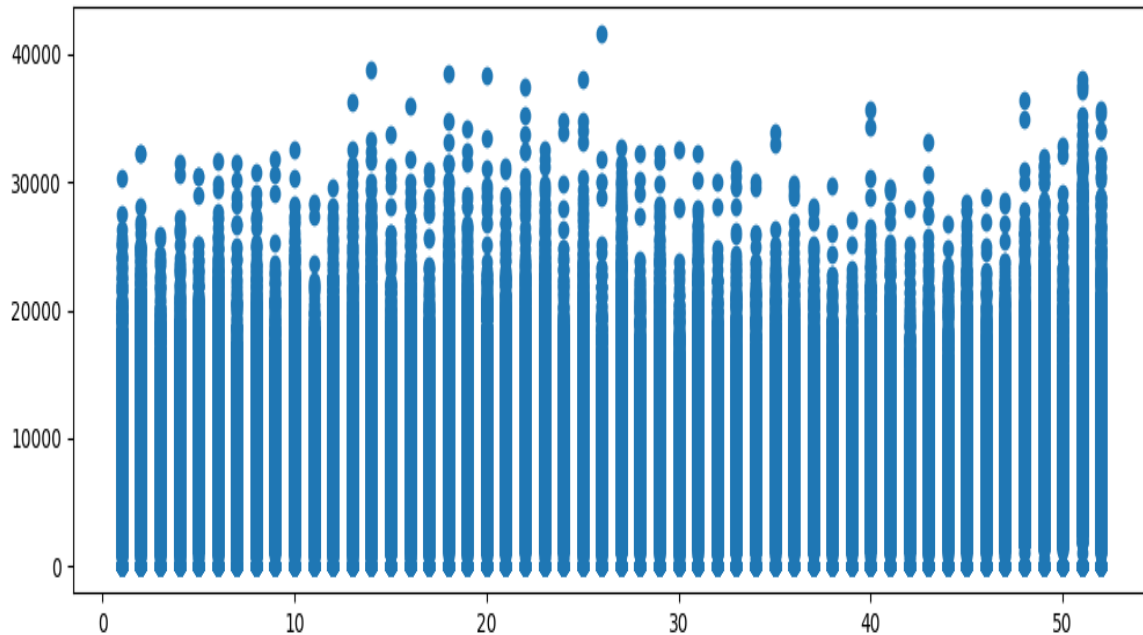
Sales per year



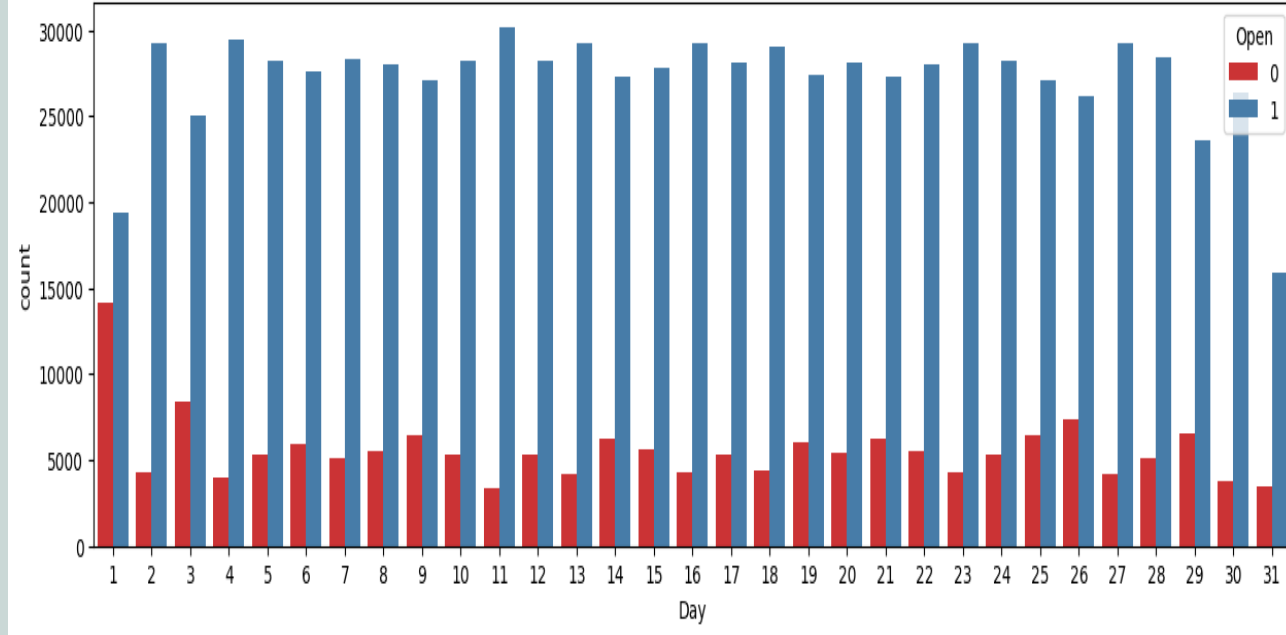
Sales per month



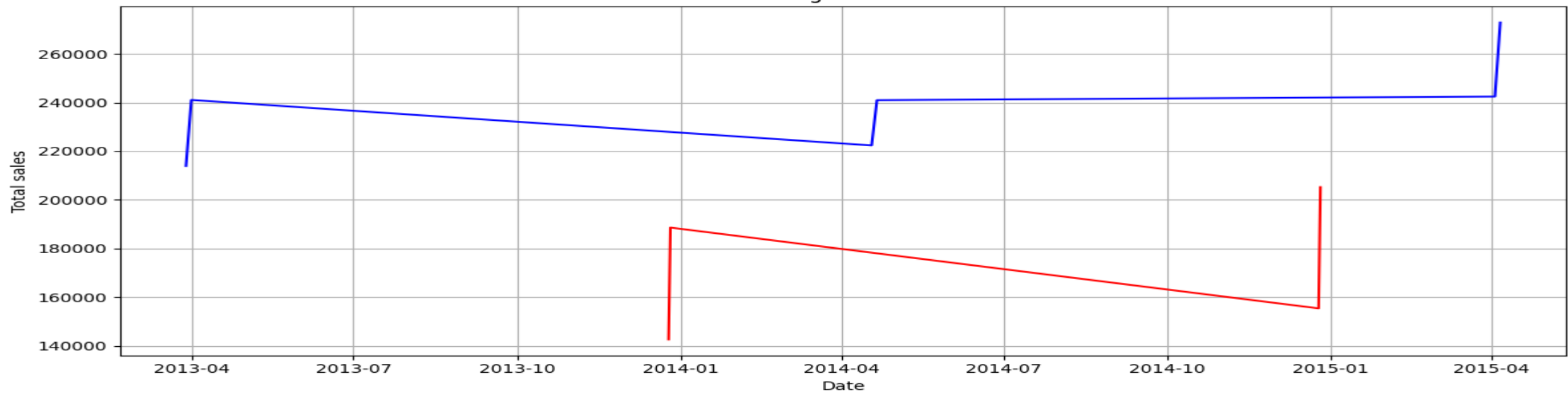
Sales per week



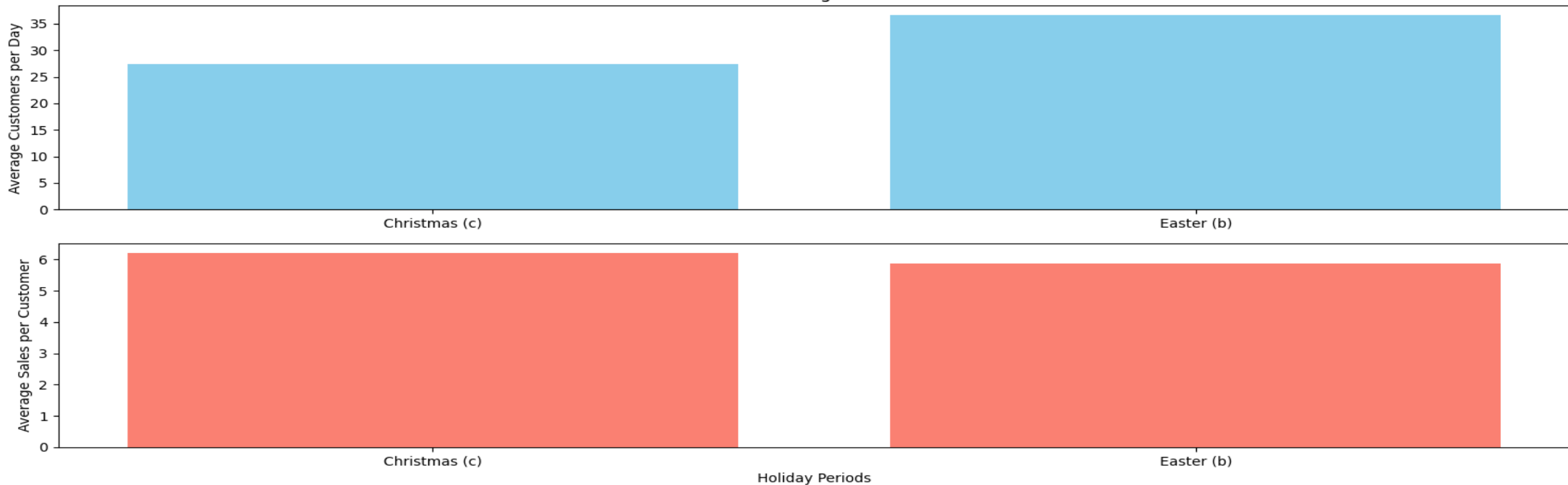
Sales on daily basis



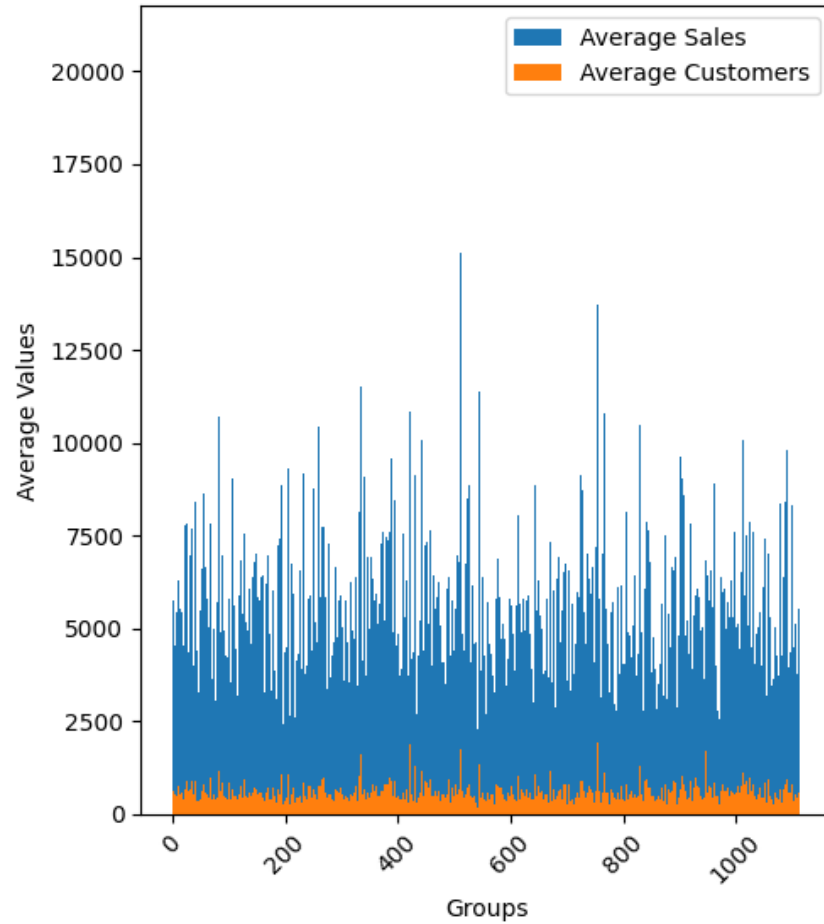
Sale Trends during Christmas and Easter



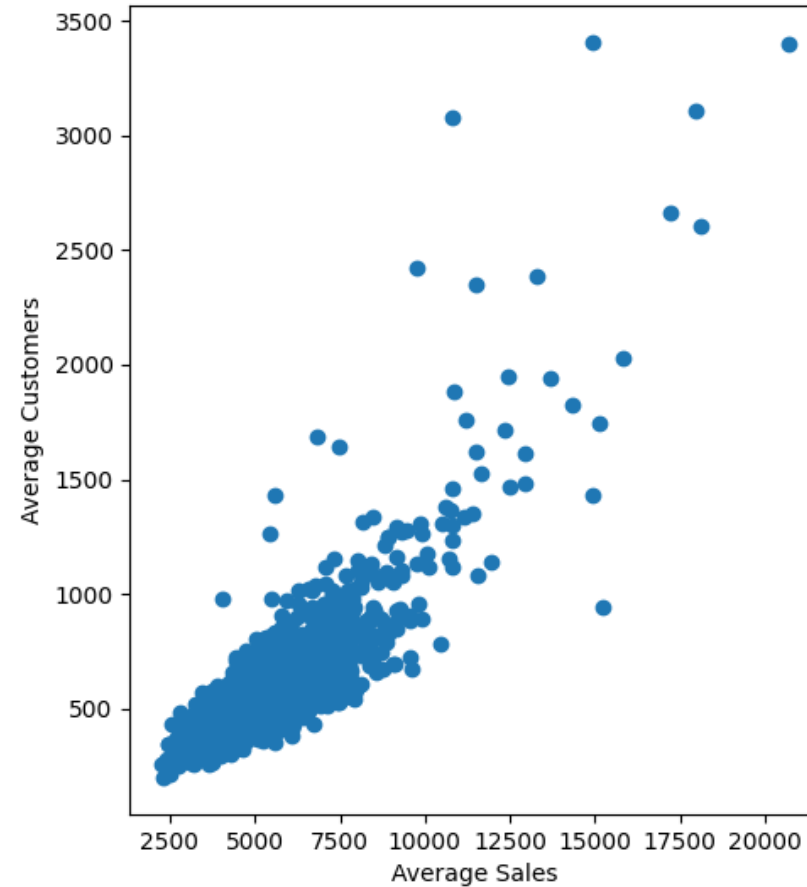
Purchase Behavior During Christmas and Easter



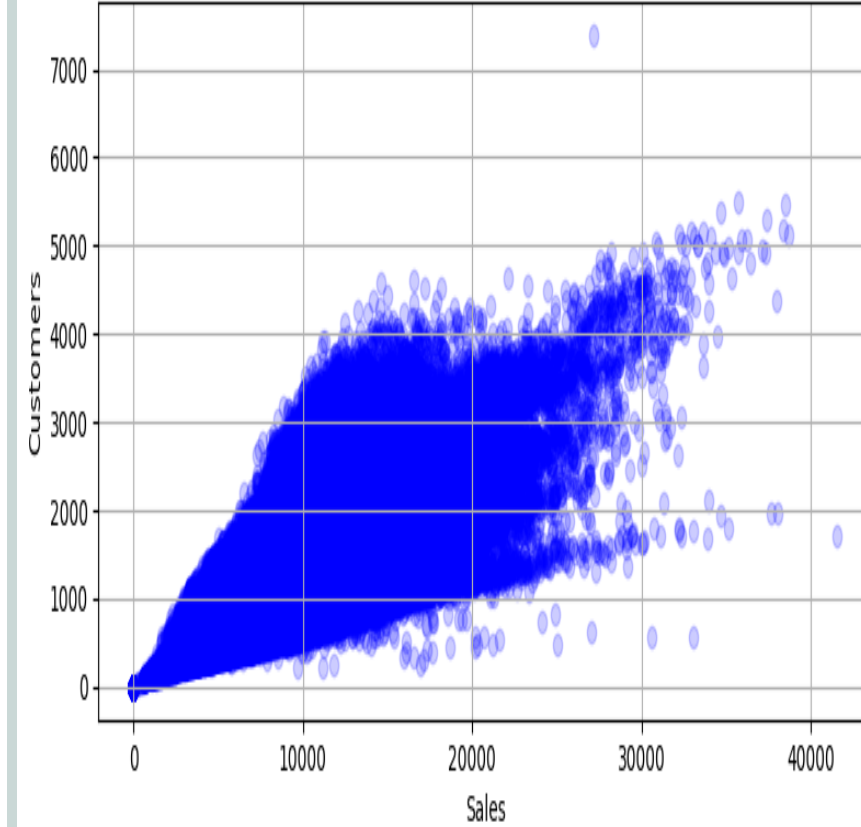
Average Sales and Customers by Group



Average Sales vs. Average Customers

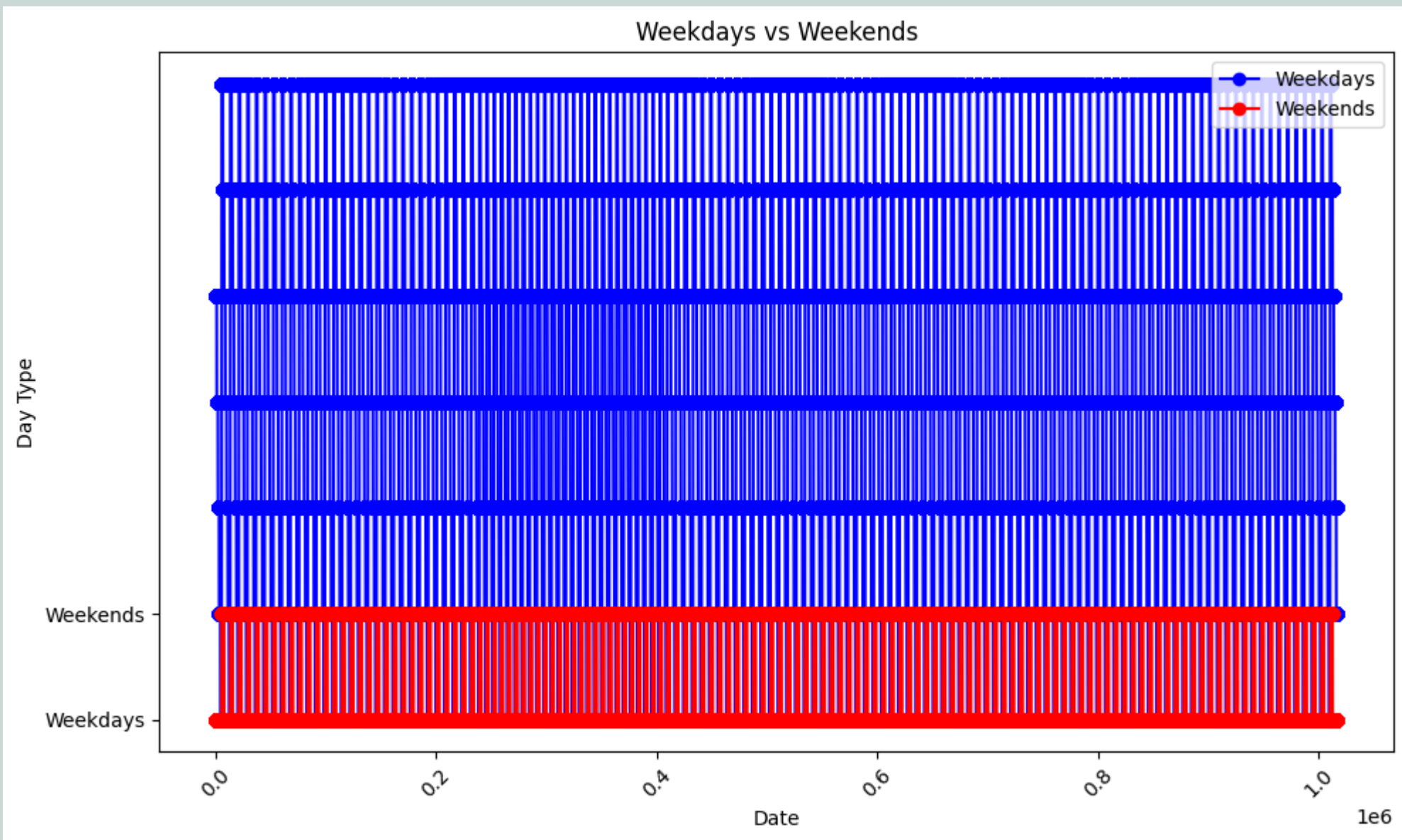


Sales vs Customers
Correlation: 0.89

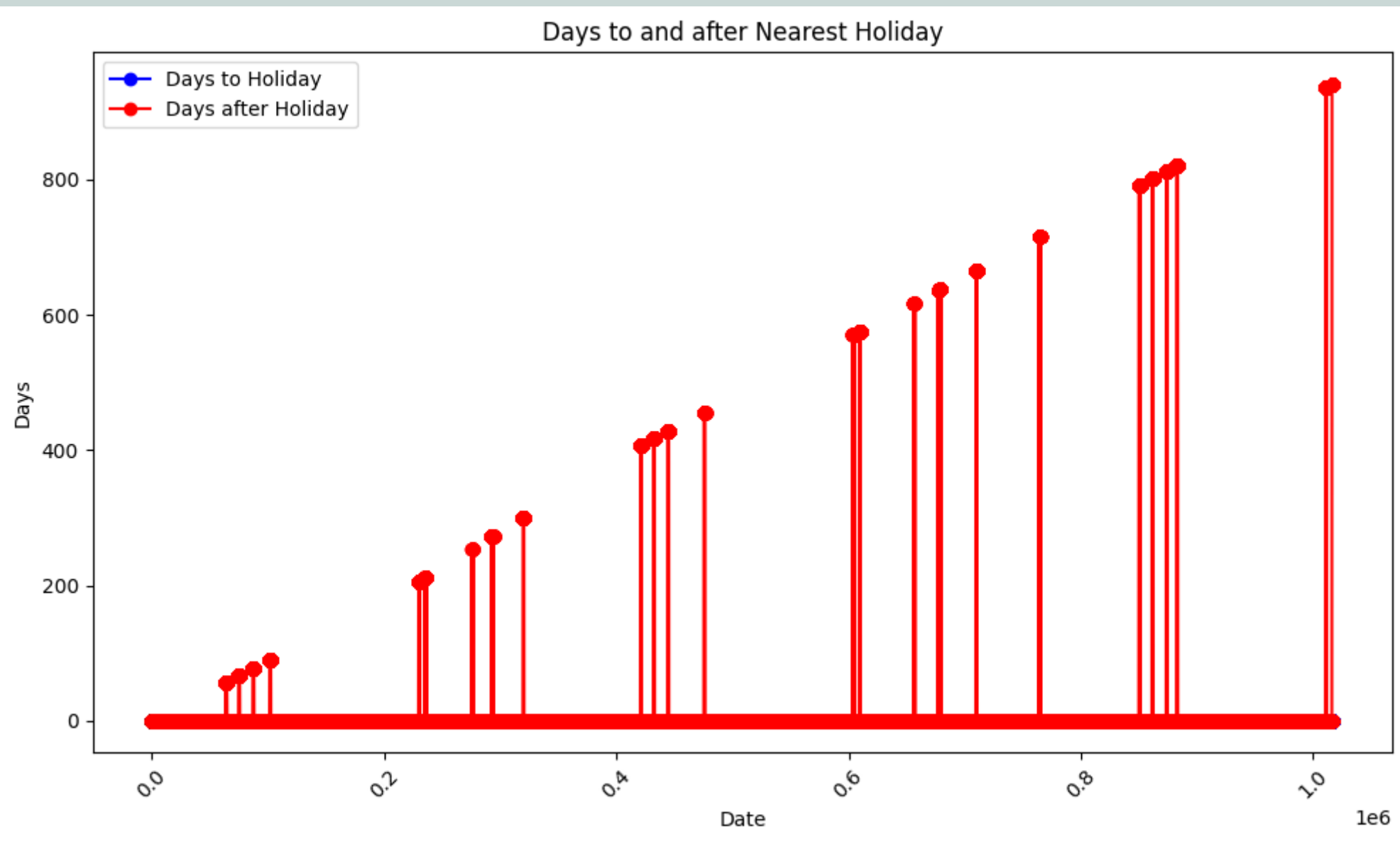


Average Sales is much more as compared to average Customers. Its Scatterplot also shows the same

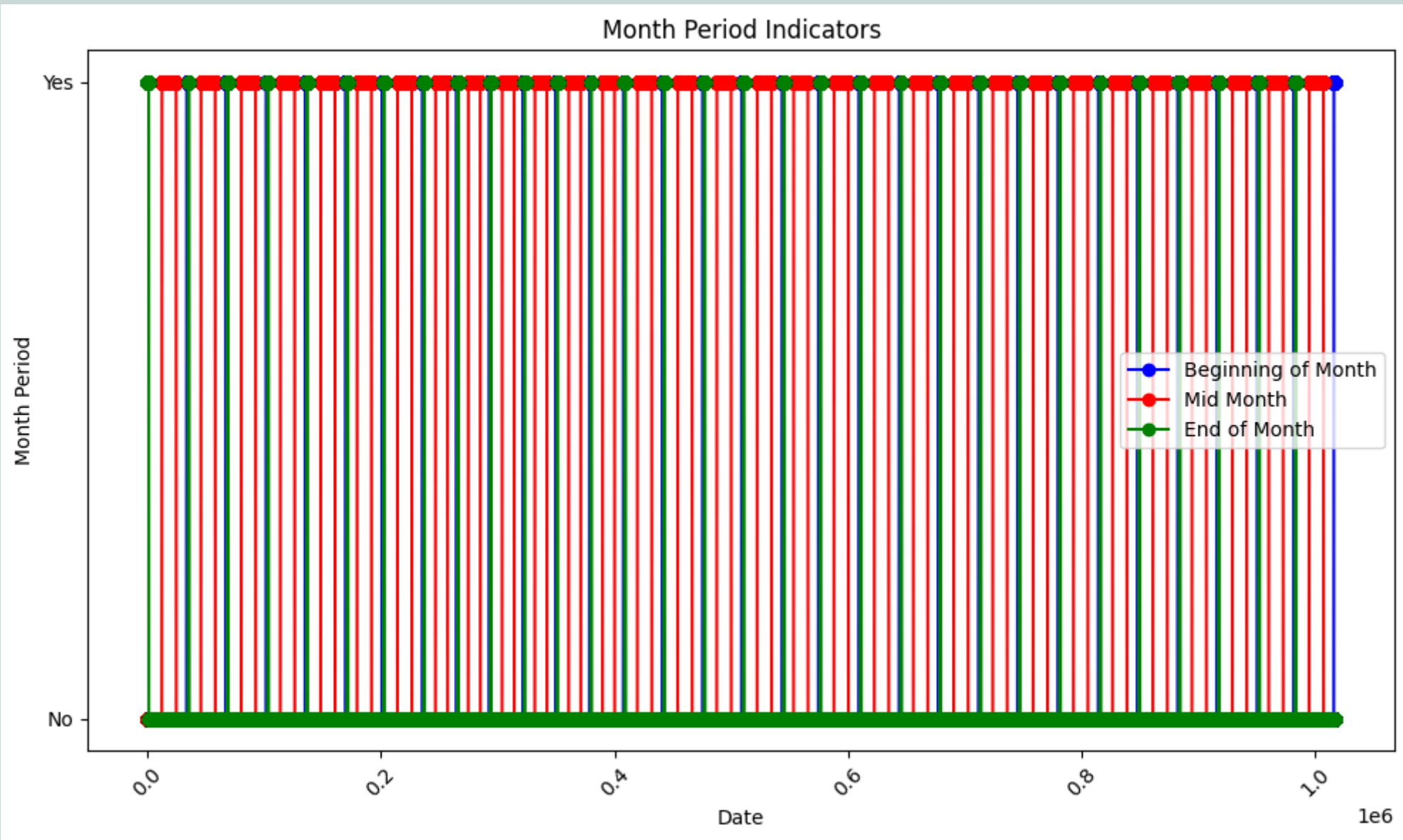
The correlation between Sales & Customers is 89% which is quite good



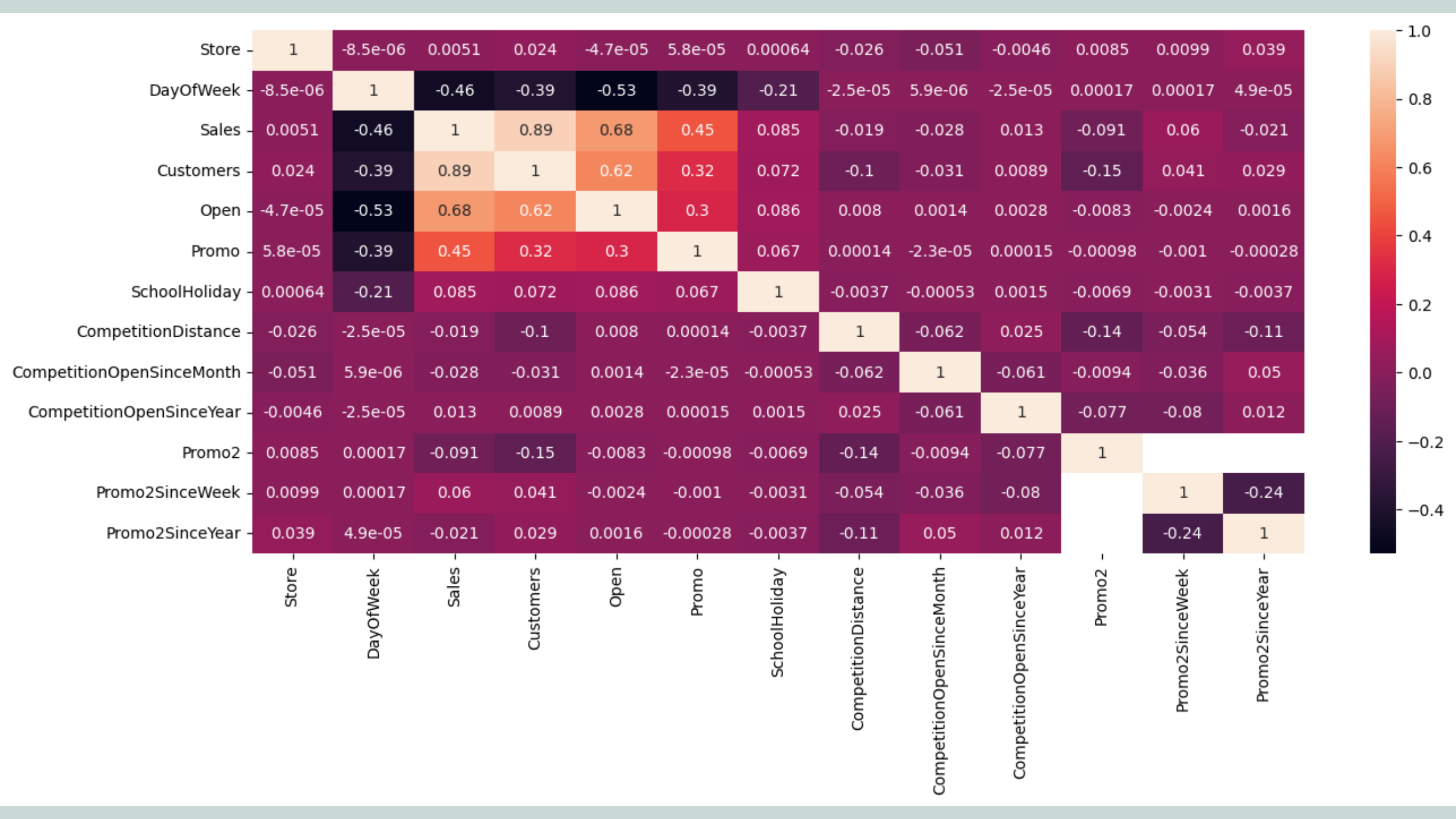
Many new columns like Weekdays and Weekends have been added in this data according to TimeSeries



Days to Holidays and Days after Holidays have been calculated

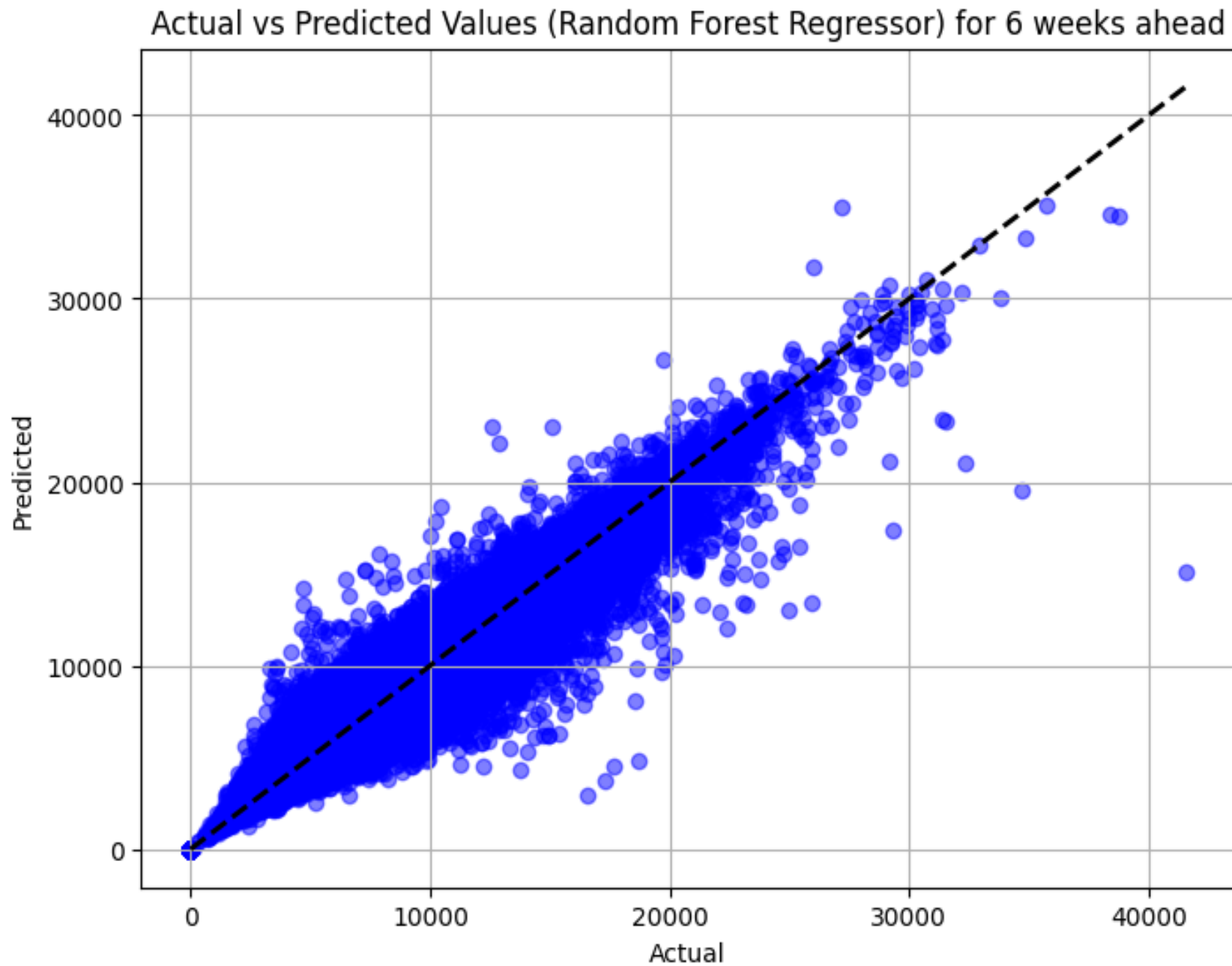


Month Period Indicators show the Beginning of month, Mid of month and End of month

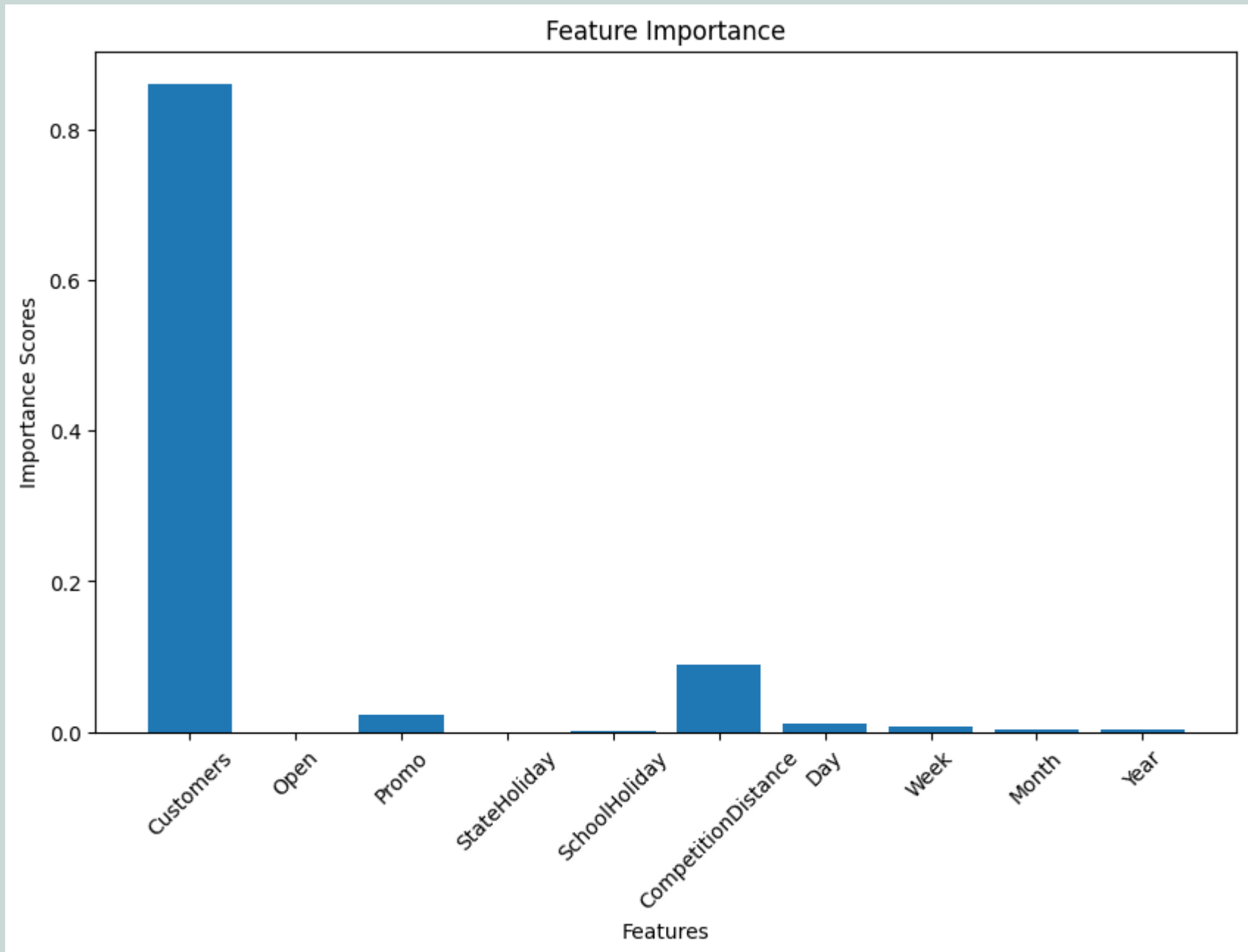


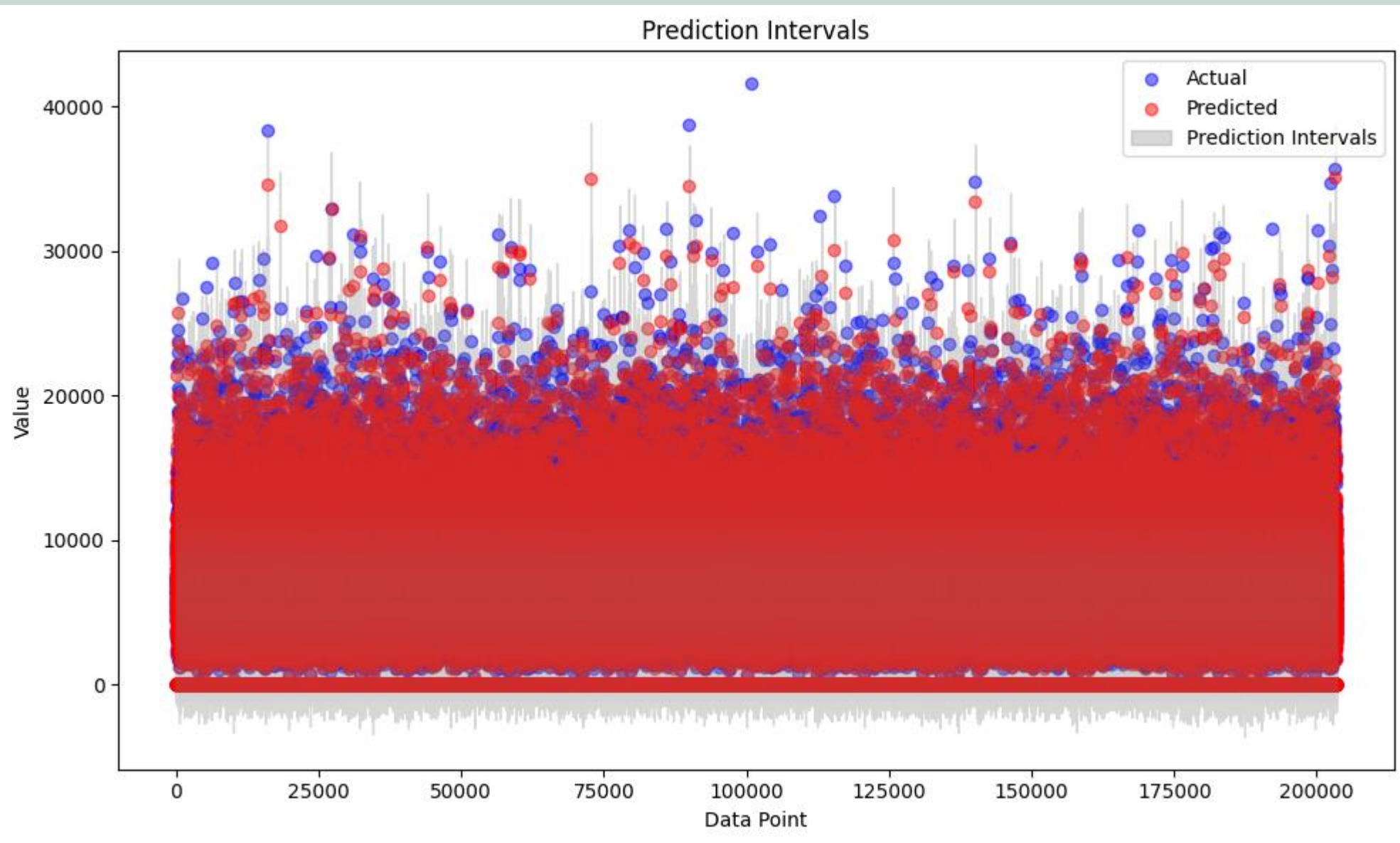
Random
Forest
Regressor with
pipeline

Score = 96.13

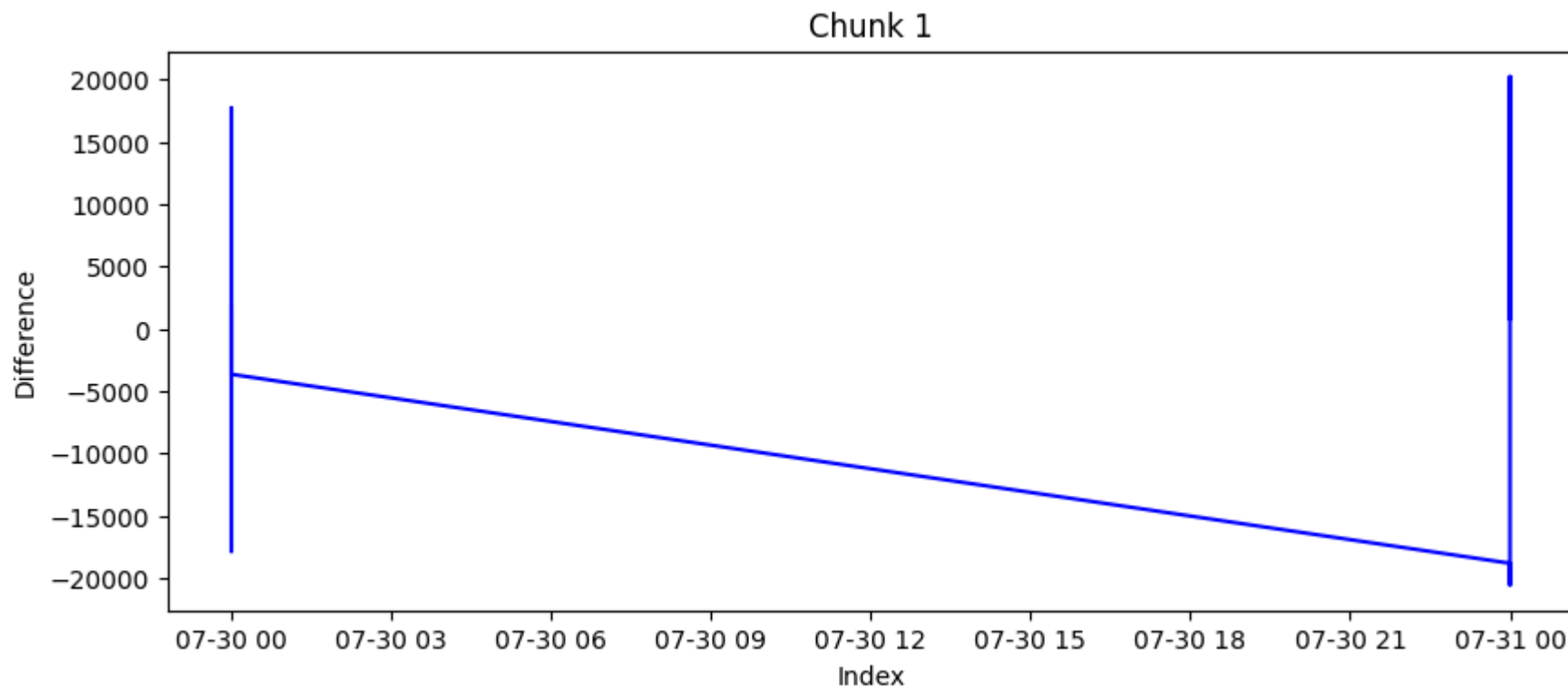


Customers is
the most
important
feature which
increases the
Sales





Prediction Intervals between Actual & Predicted according to Random Forest Regressor



Chunk 1

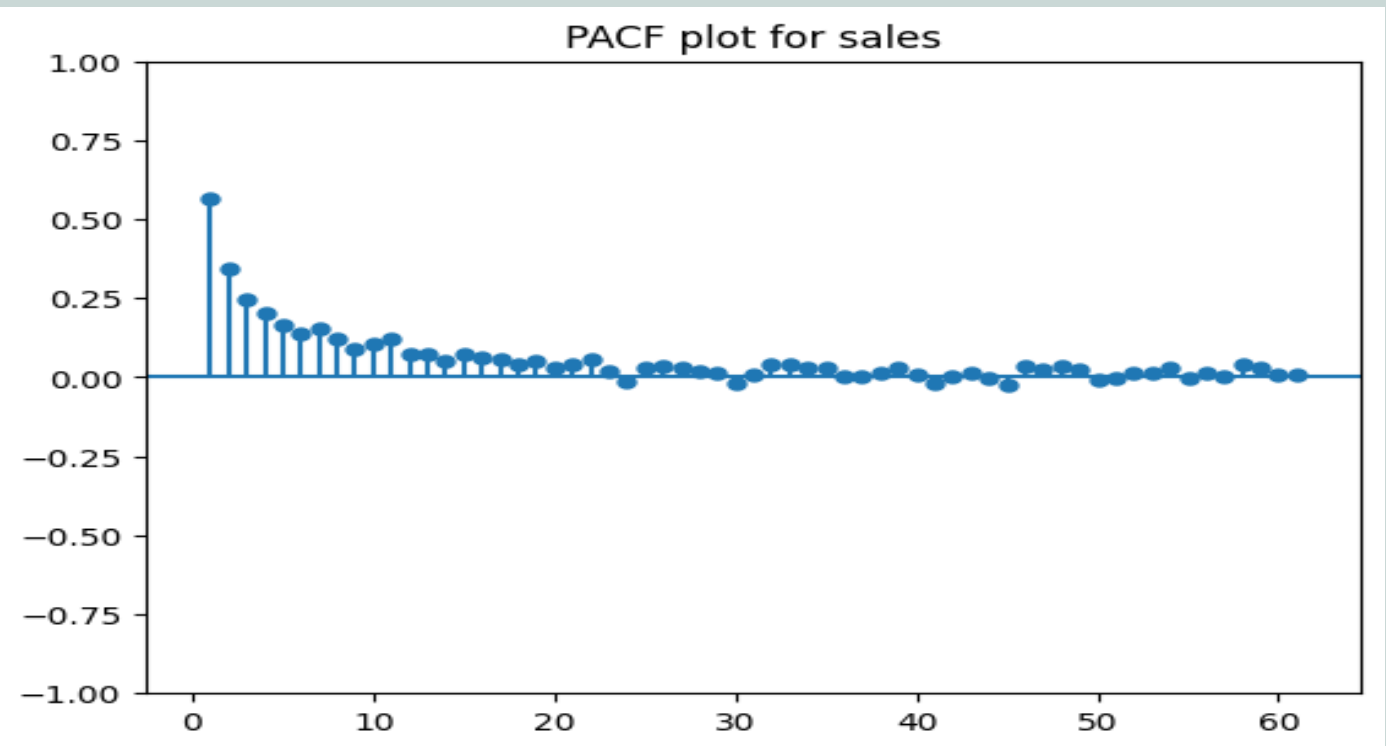
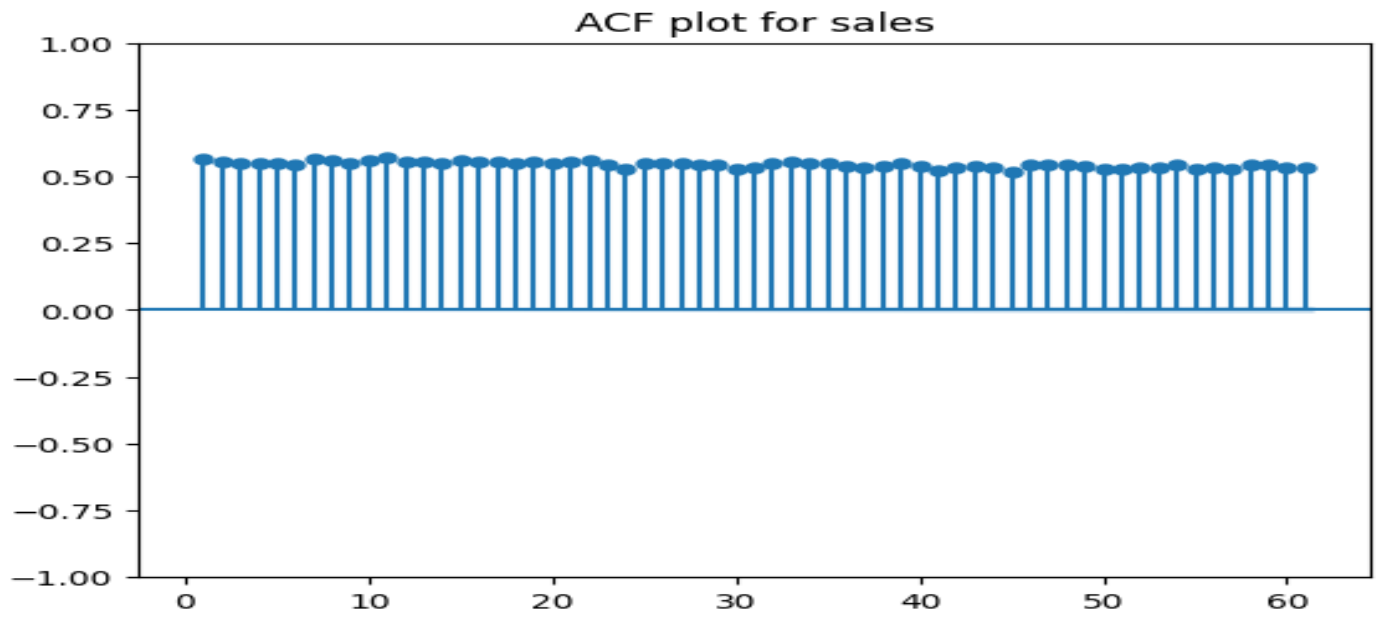
ADF statistics: -15.454115638809549

p-values: 2.7596216659777057e-28

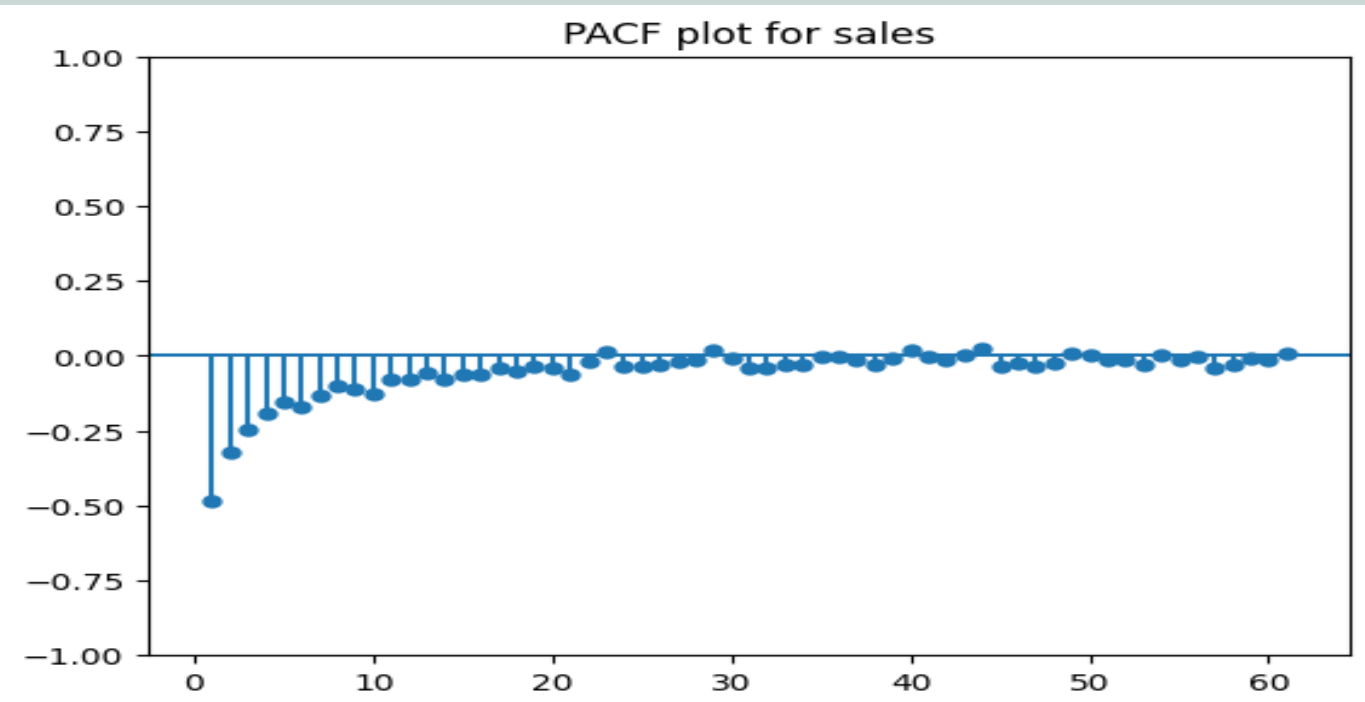
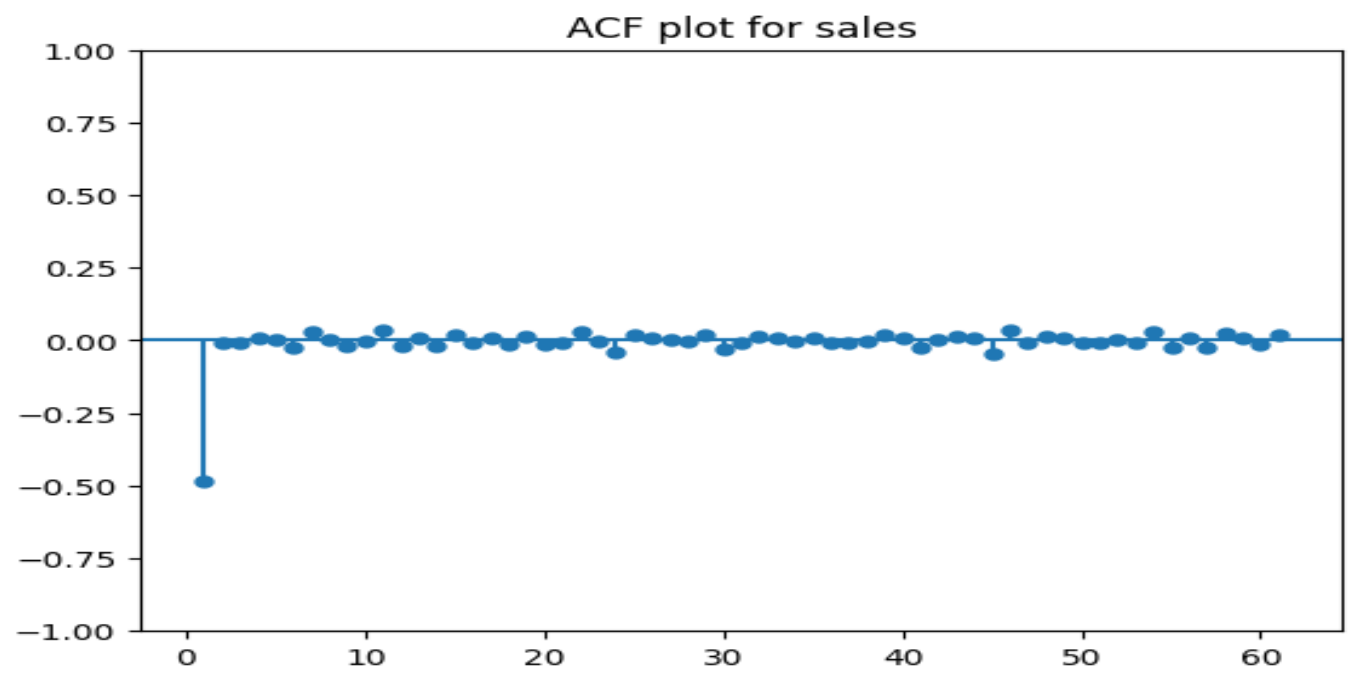
Critical Values: {'1%': -3.433665351698571, '5%': -2.8630045337025267, '10%': -2.567549656849864}

Reject the null hypothesis - Data is stationary

High autocorrelation
for lags of multiples
of 7. Maybe a
seasonal component
for day of week

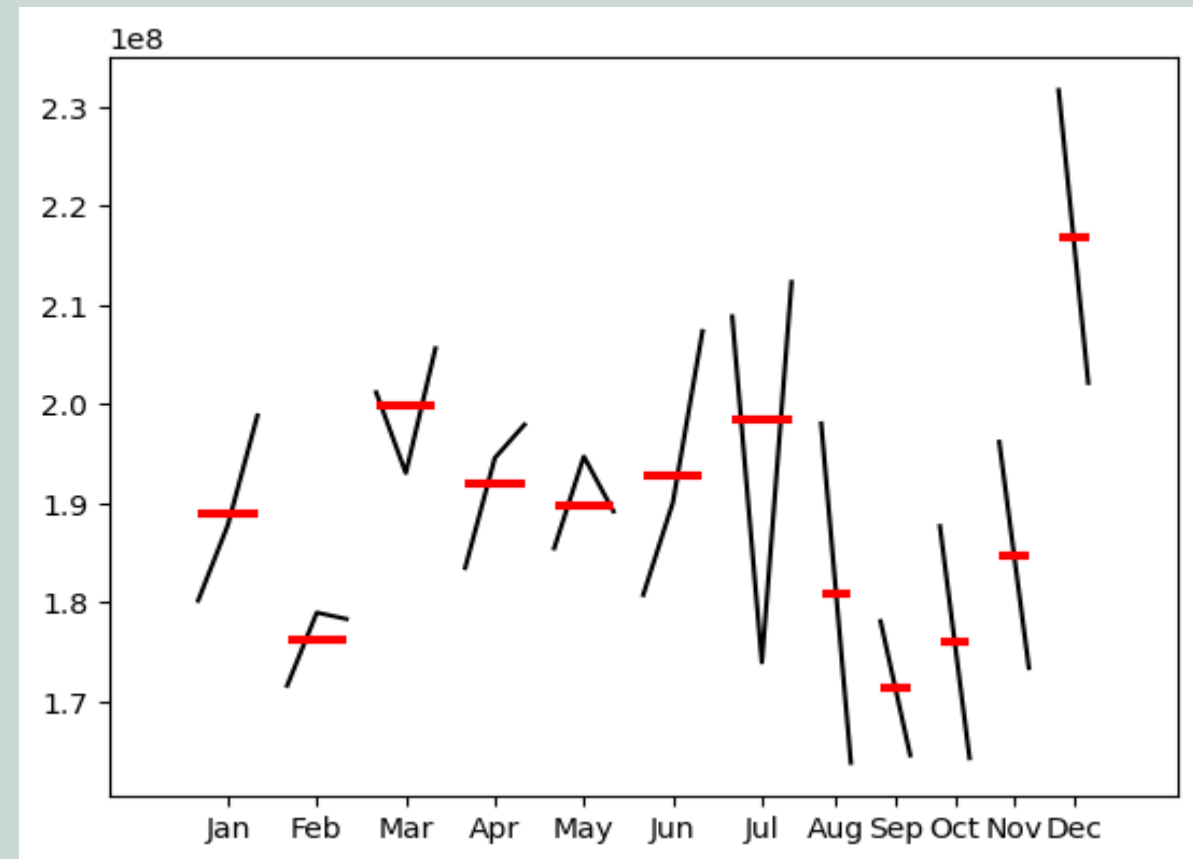
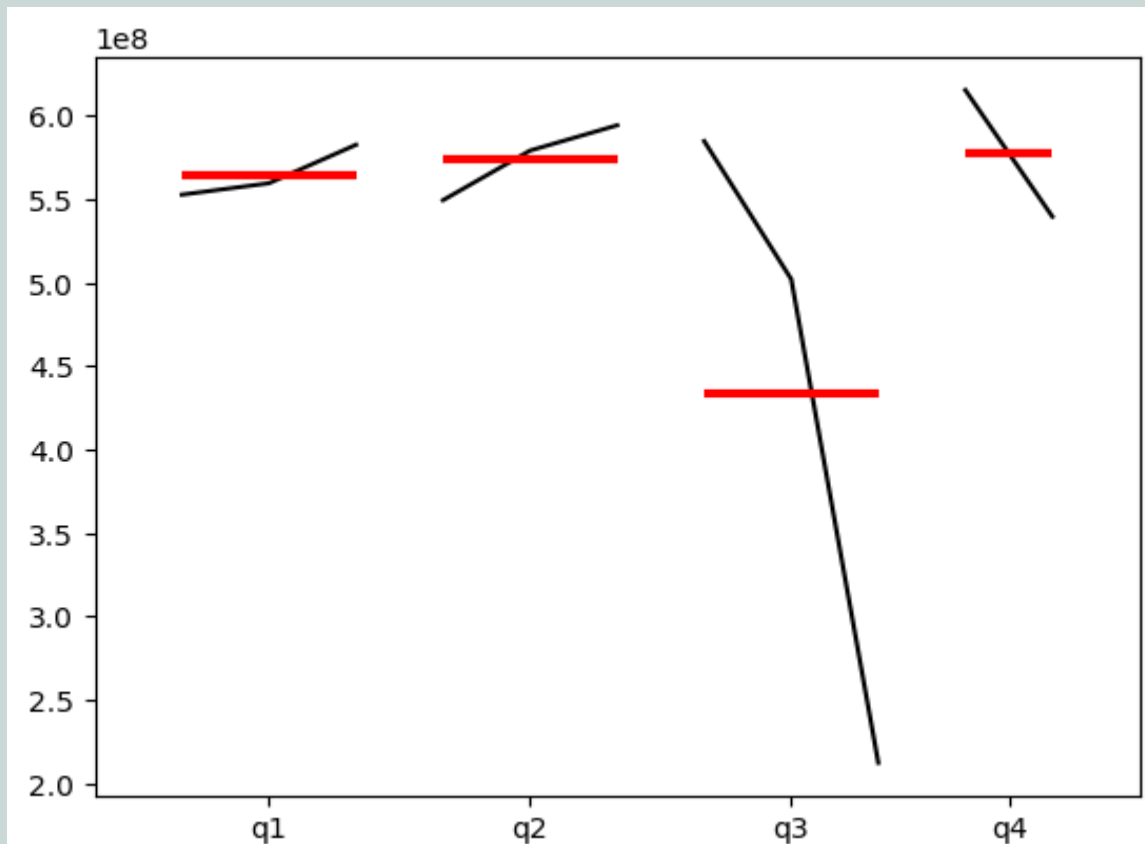


High partial
autocorrelation
for lags of
multiples of 7



Monthly Plot

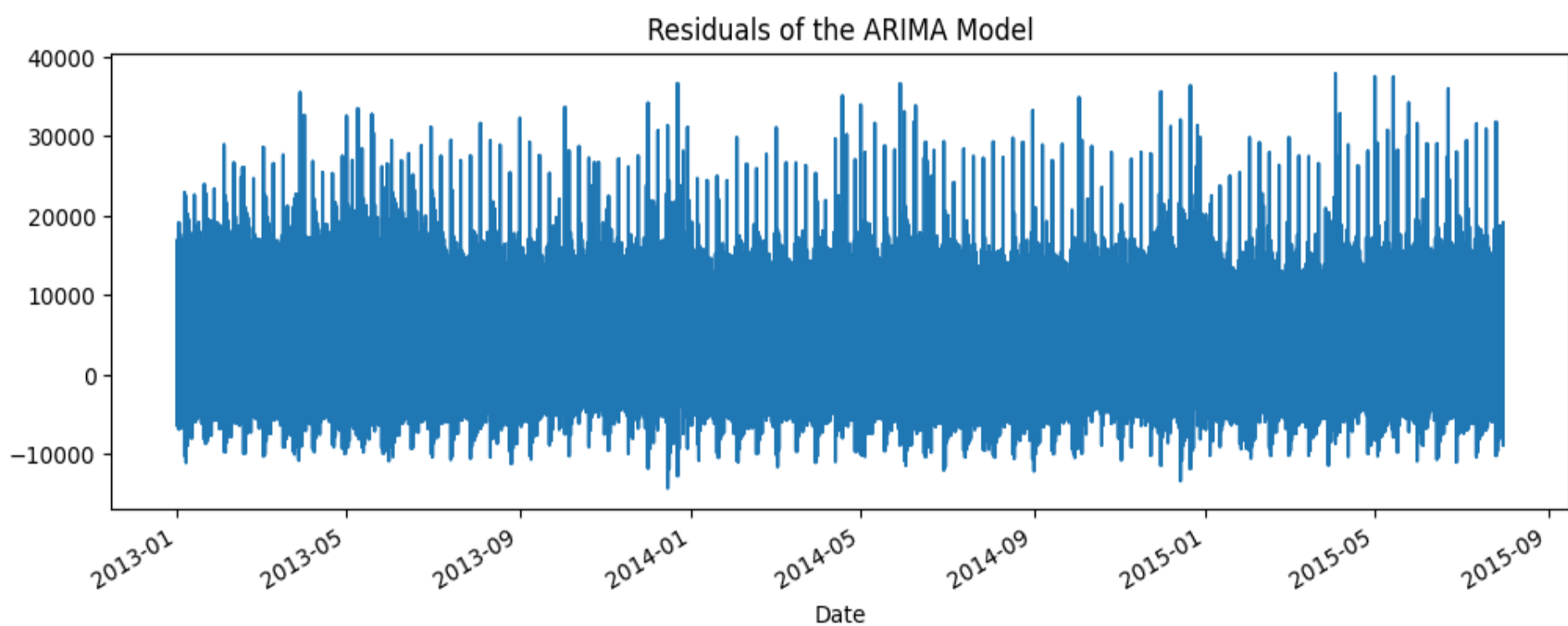
The maximum output is in the month of December i.e; Christmas



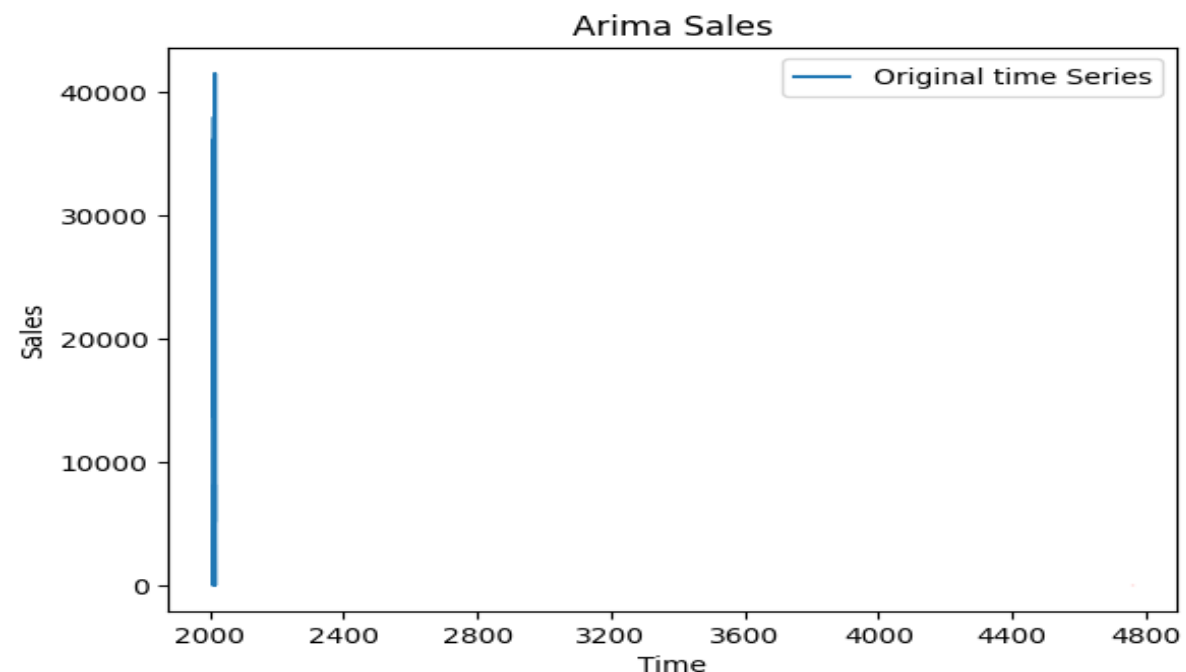
Quarter Plot

The output goes down in 3rd quarter whereas it remains same in all other quarters

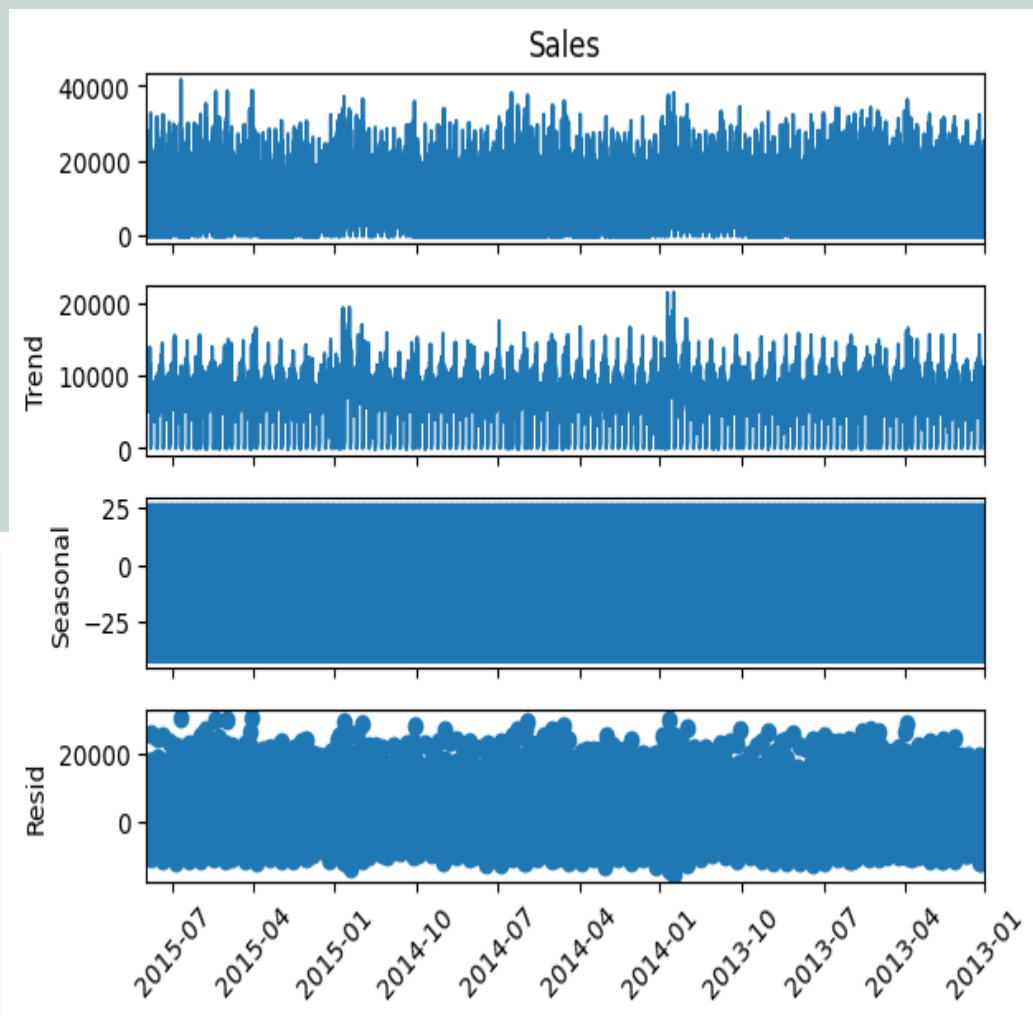
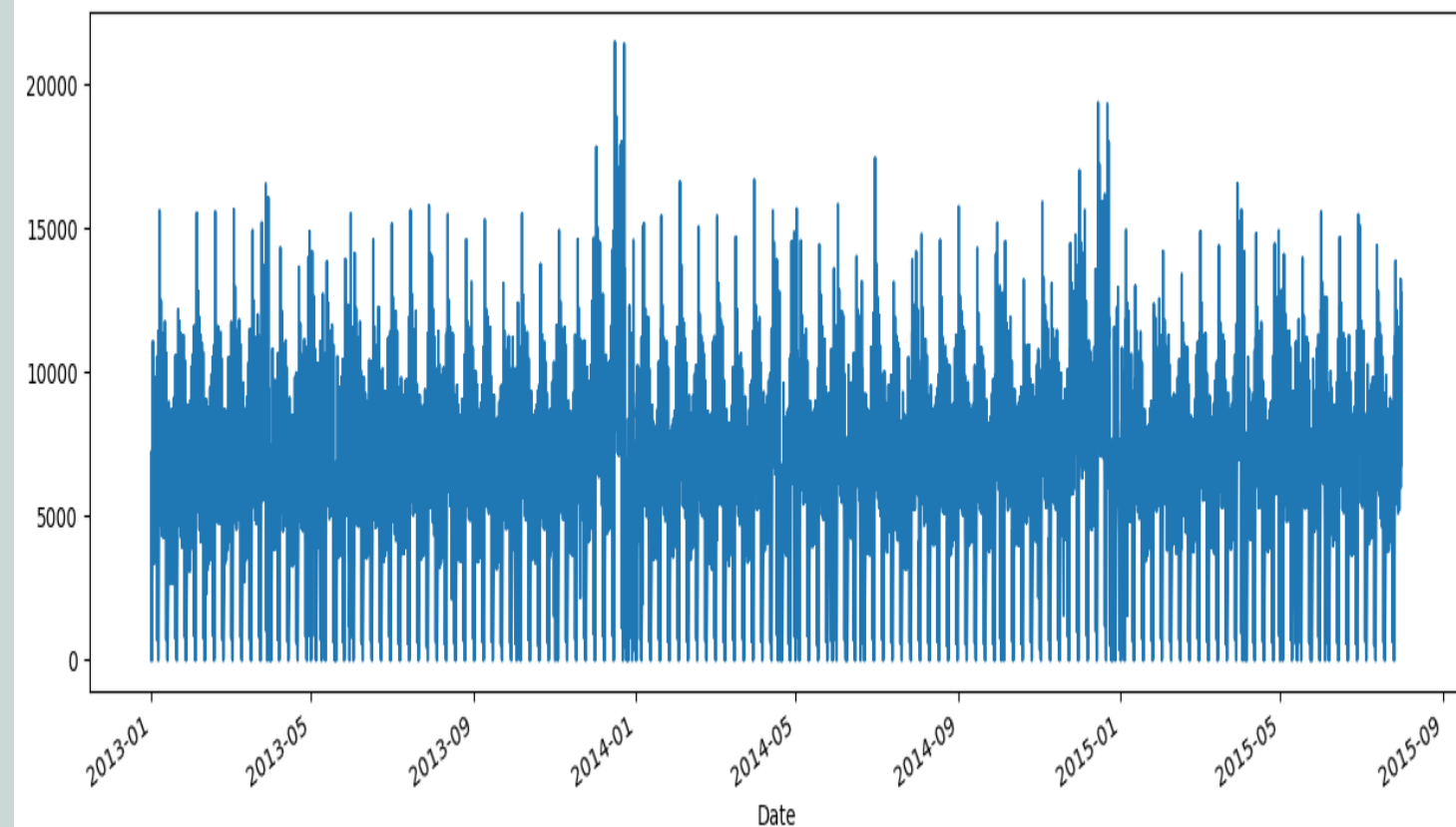
Residuals
show the
different
between
Actual Sales
and the
predicted
fitted values



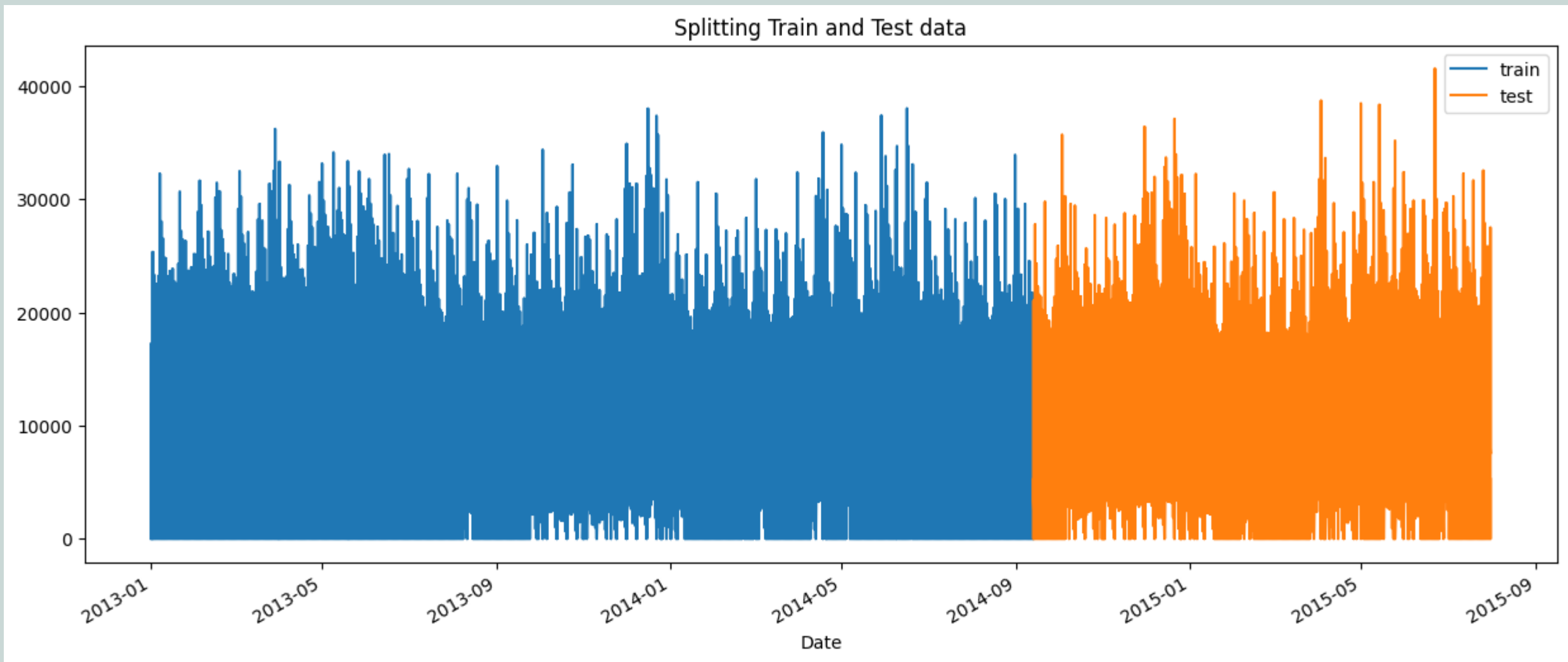
It shows the Sales
forecast by Arima
Model which is
maximum at 2000



Seasonal decompose plot

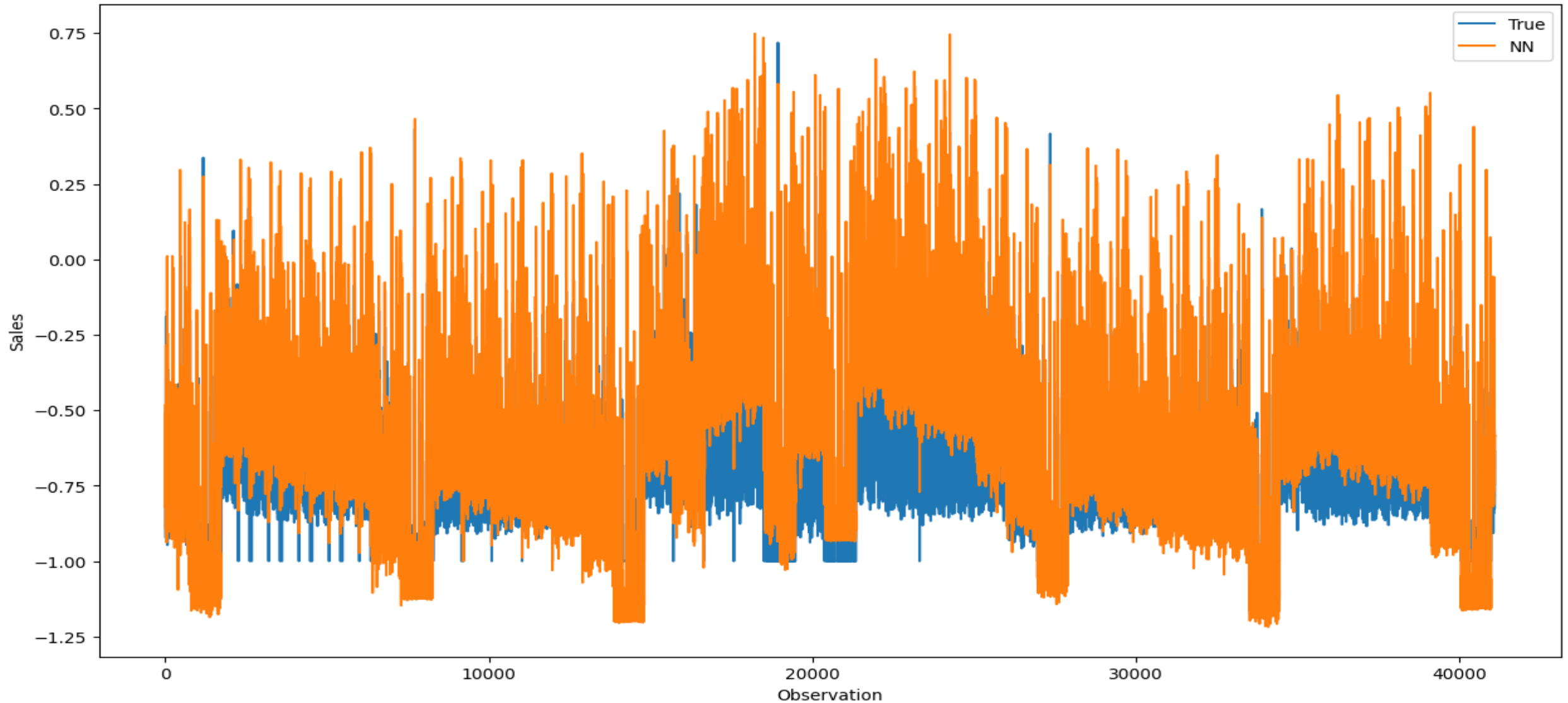


Decompose Trend Plot



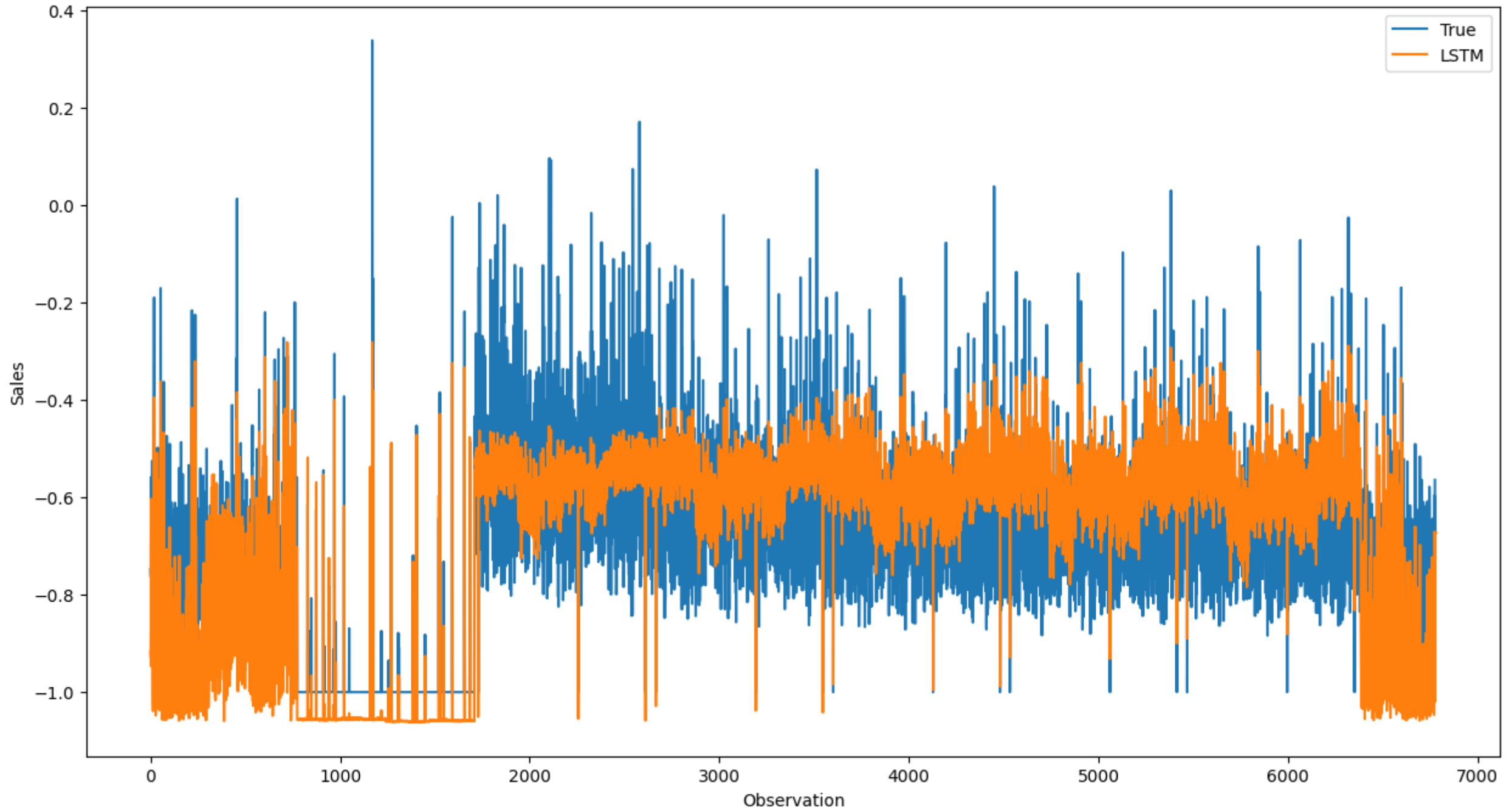
Year 2013 and 2014 have been taken in Train and year 2015 has been taken in Test data

ANN's Prediction 6 weeks ahead

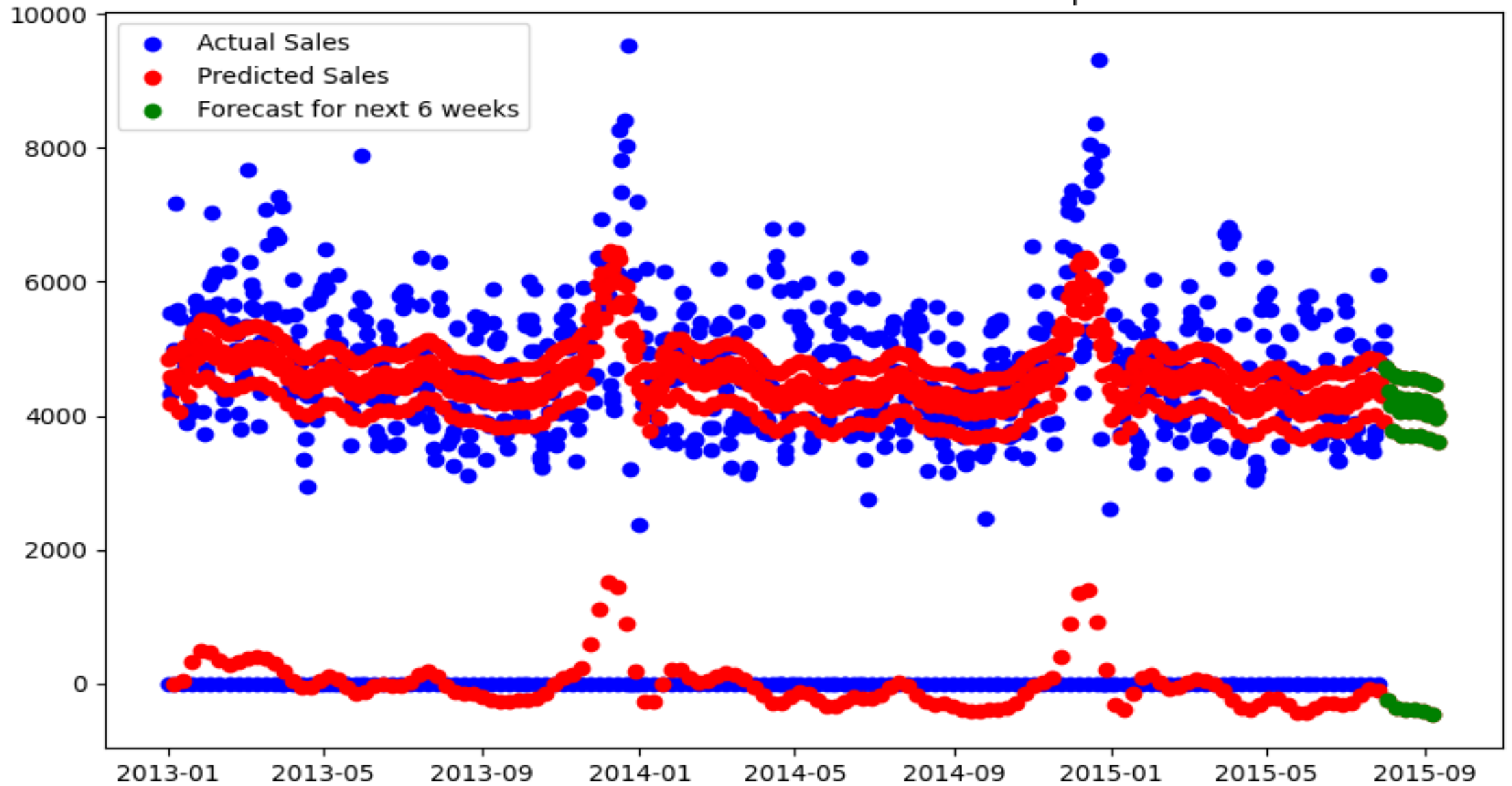


Neural Network seems to perform better than a LSTM network

LSTM's Prediction 6 weeks ahead



Actual vs 6 weeks Predicted Sales in Prophet



Actual and Predicted Sale in Prophet



“ Customer is the most important factor
which affects Sales.

”





Areas of focus

We should keep in mind that this is a simplified demonstration, using only a single store's data and making quite basic assumptions (e.g., no promotional or holiday effects other than weekly and yearly seasonality). For a more accurate and robust forecast, it would be advisable to use data from multiple stores, and potentially to consider additional features such as promotions, holidays, and other external factors that might influence sales.



Summary

The model has successfully captured some of the seasonality in the data. It predicts a rise in sales around the end of the year, which aligns with the increased shopping activity during the holiday season. This is a good sign, as it indicates that the model has learned not just from the raw sales values, but also from their timing.





Thank you



GEET GOVIND

geet13govind@gmail.com