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.

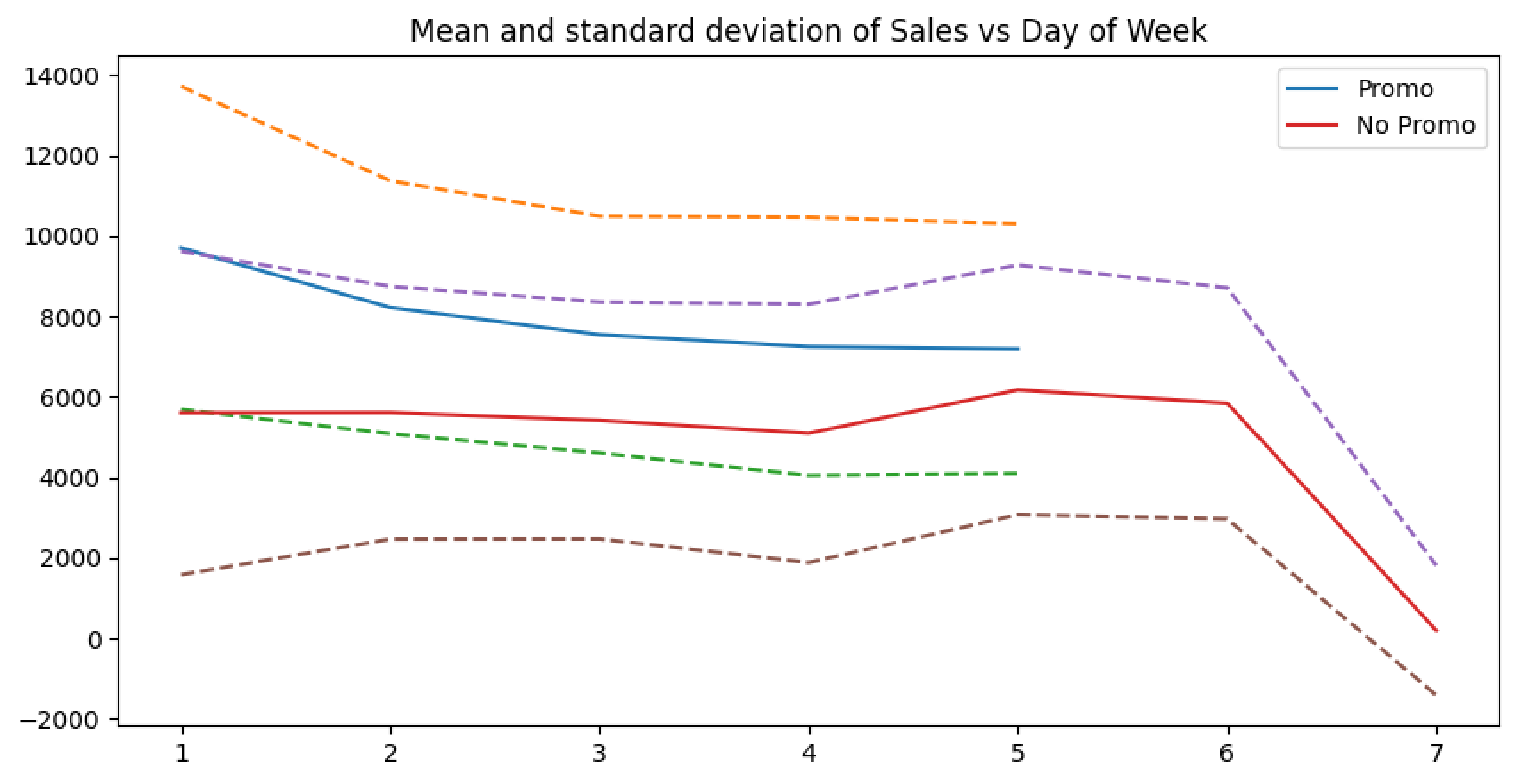
Primary goals

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

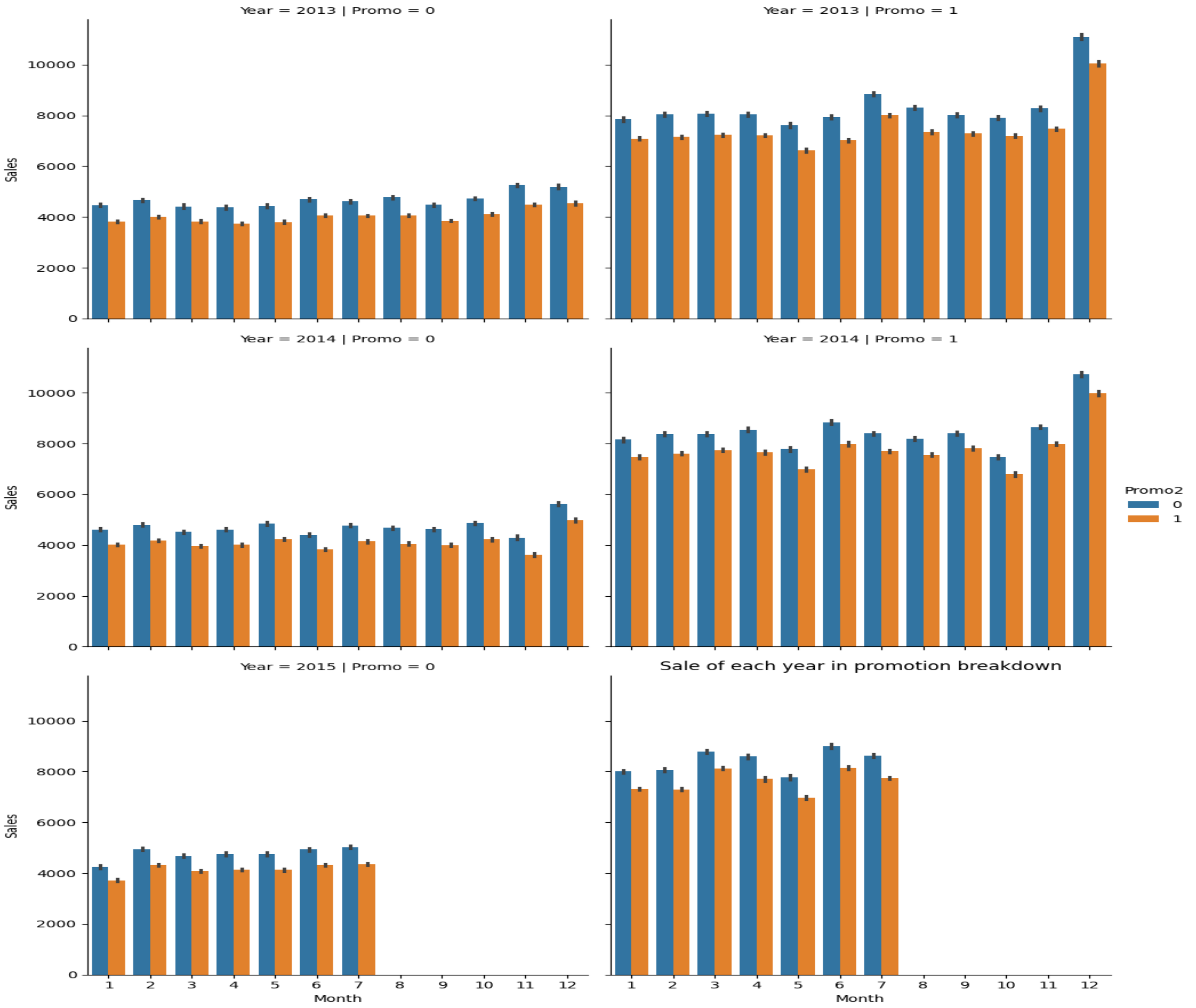


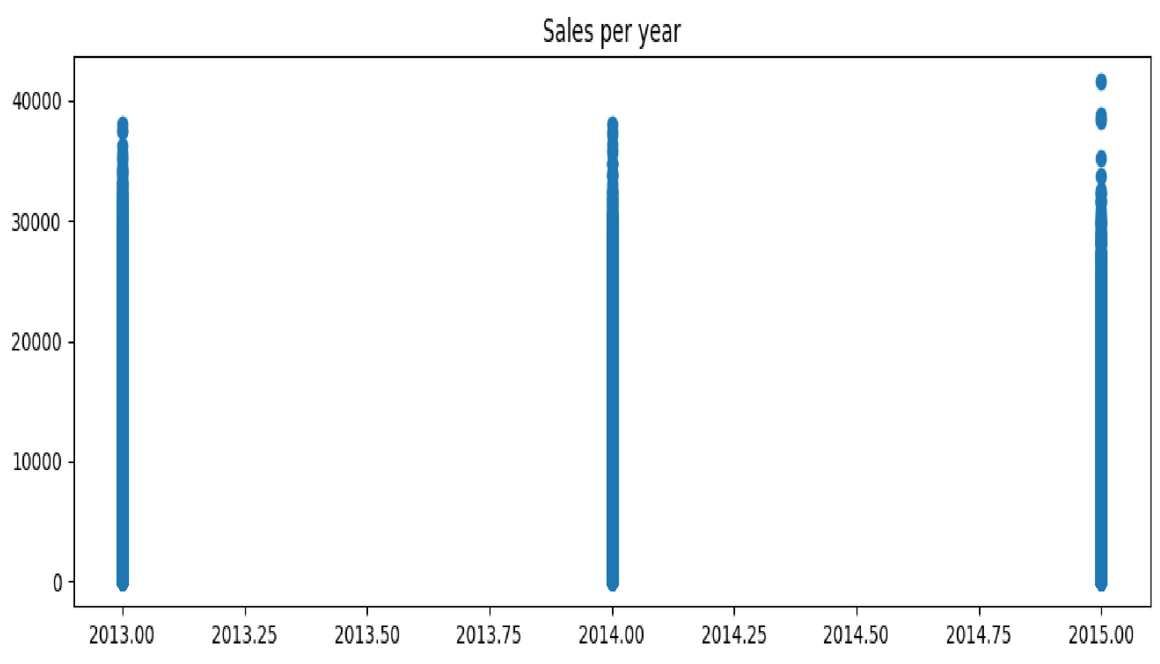
Promotions are correlated with more sales and customers

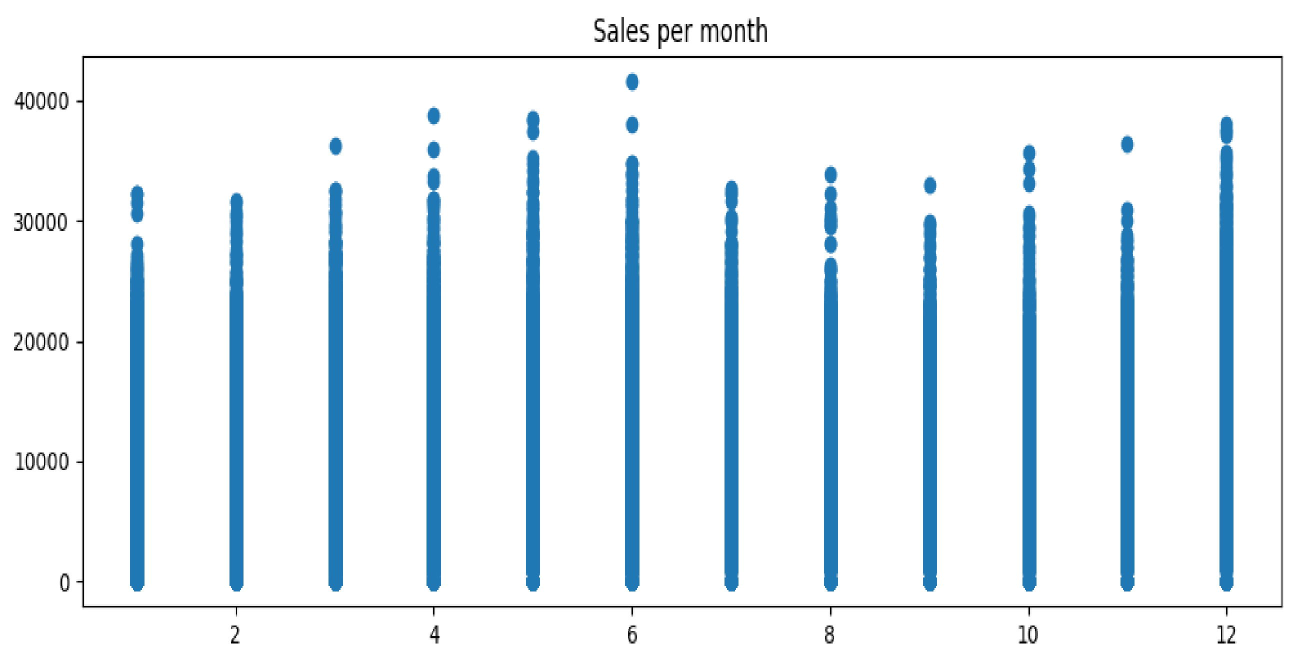
Sale of each year in promotion breakdown

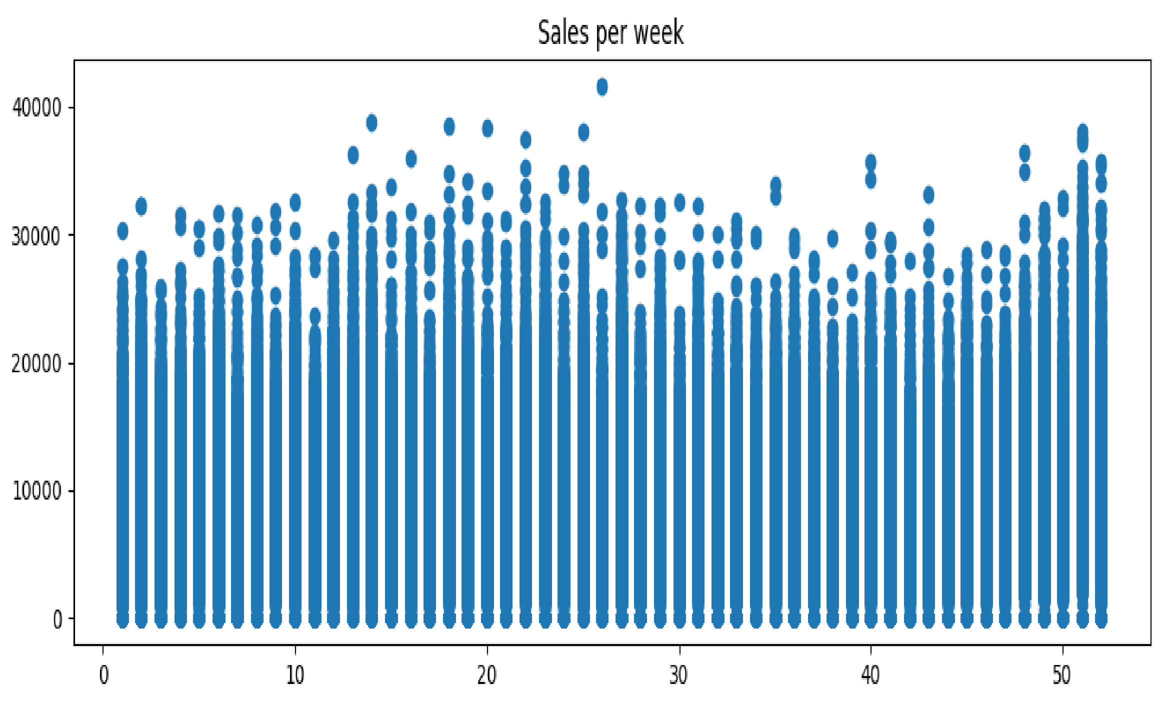
- We see that the campaigns carried out by the stores themselves and the campaigns jointly organized by the company have an increasing effect on sales.

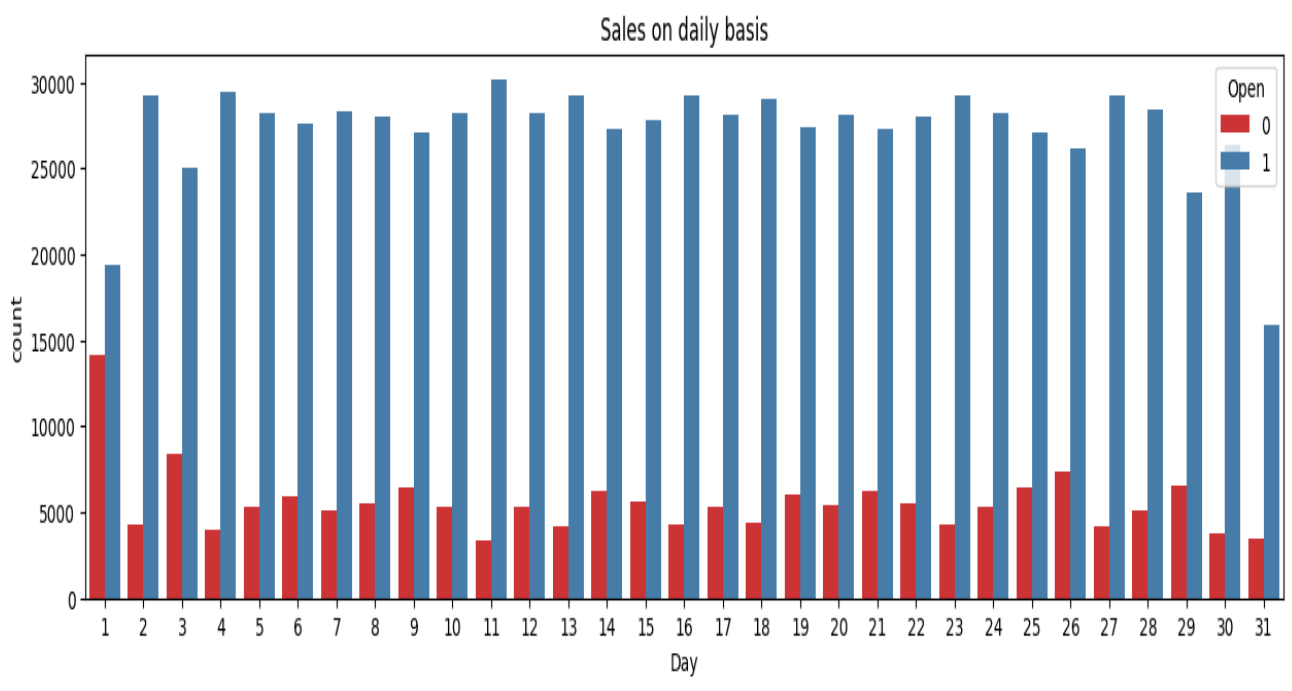
- Strong sales are achieved when participating in the corporate campaign in the last month of the year.

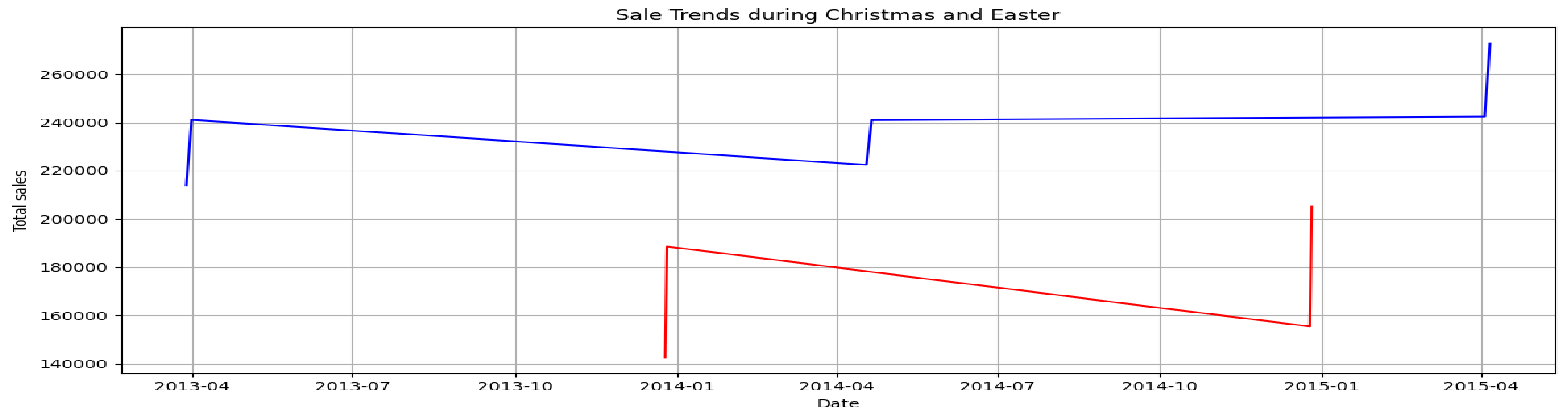


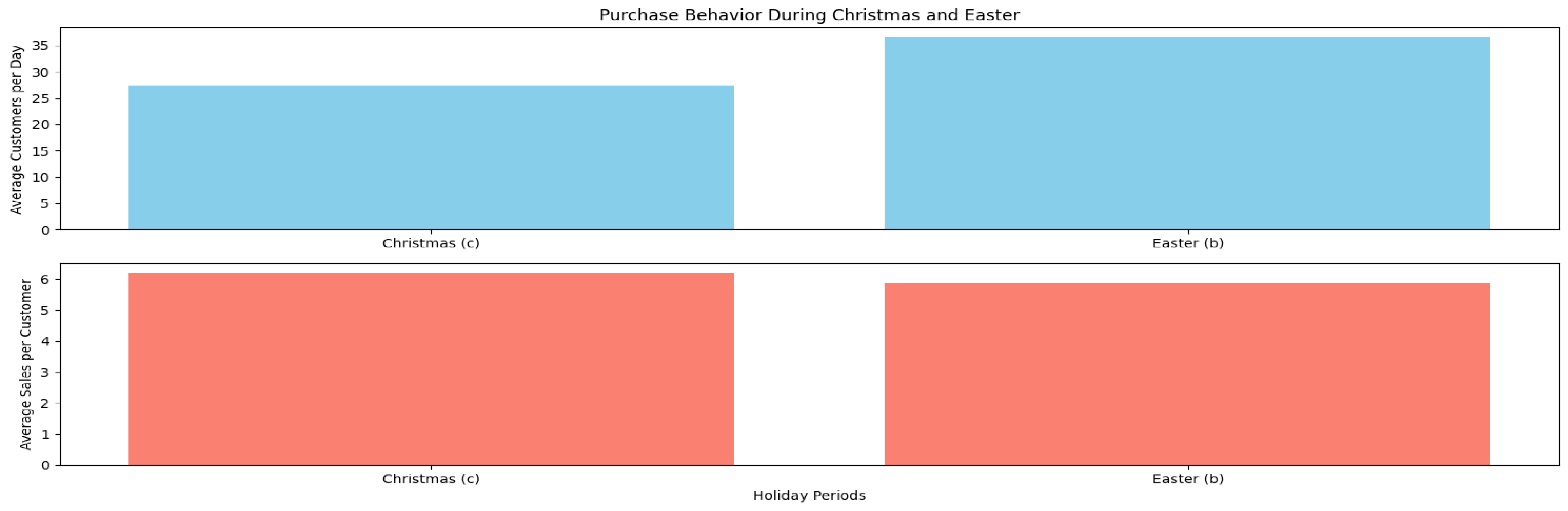

Sales per year
0
10000
20000
30000
40000
2013.00
2013.25
2013.50
2013.75
2014.00
2014.25
2014.50
2014.75
2015.00

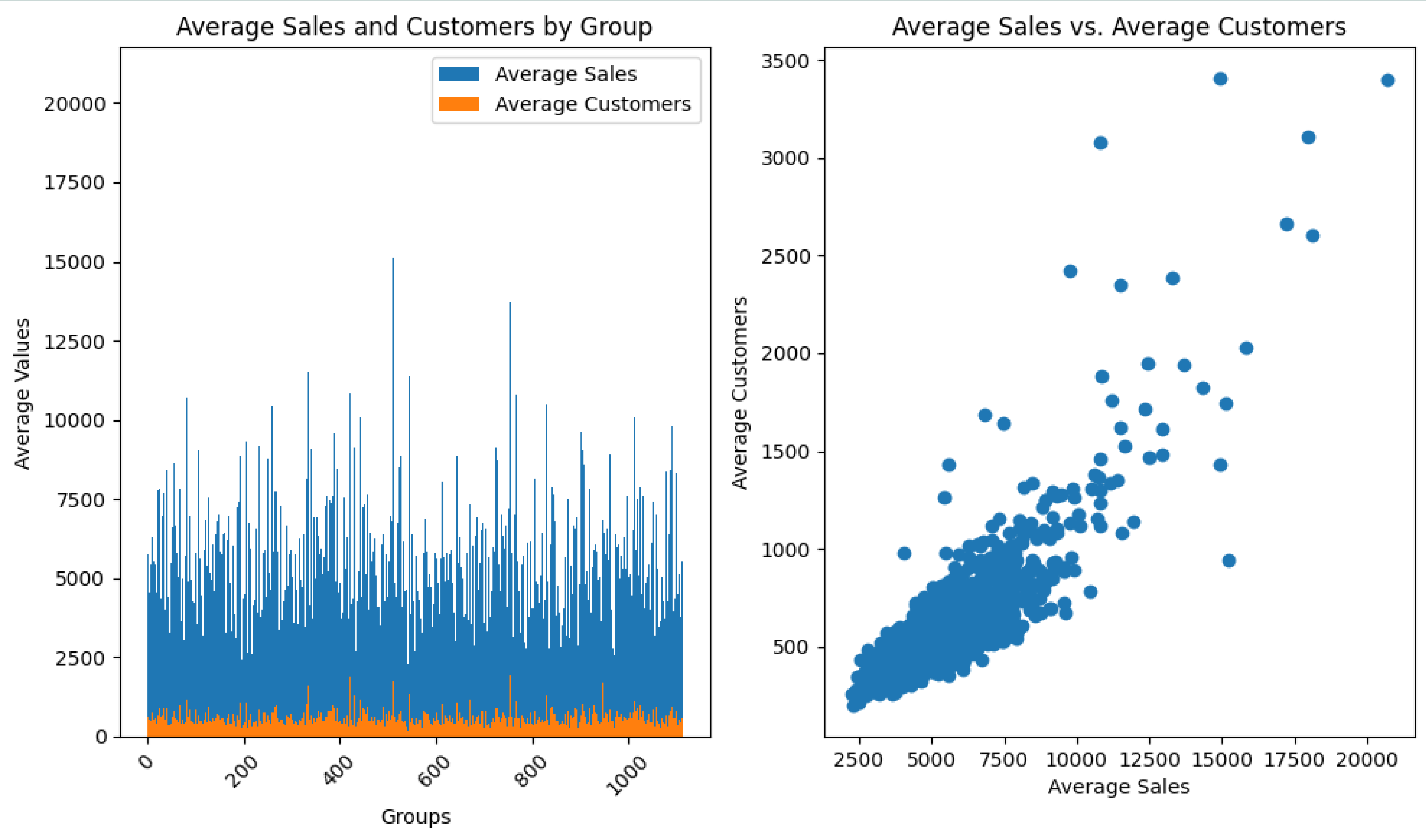

Sales per month
0
10000
20000
30000
40000
2
4
6
8
10
12

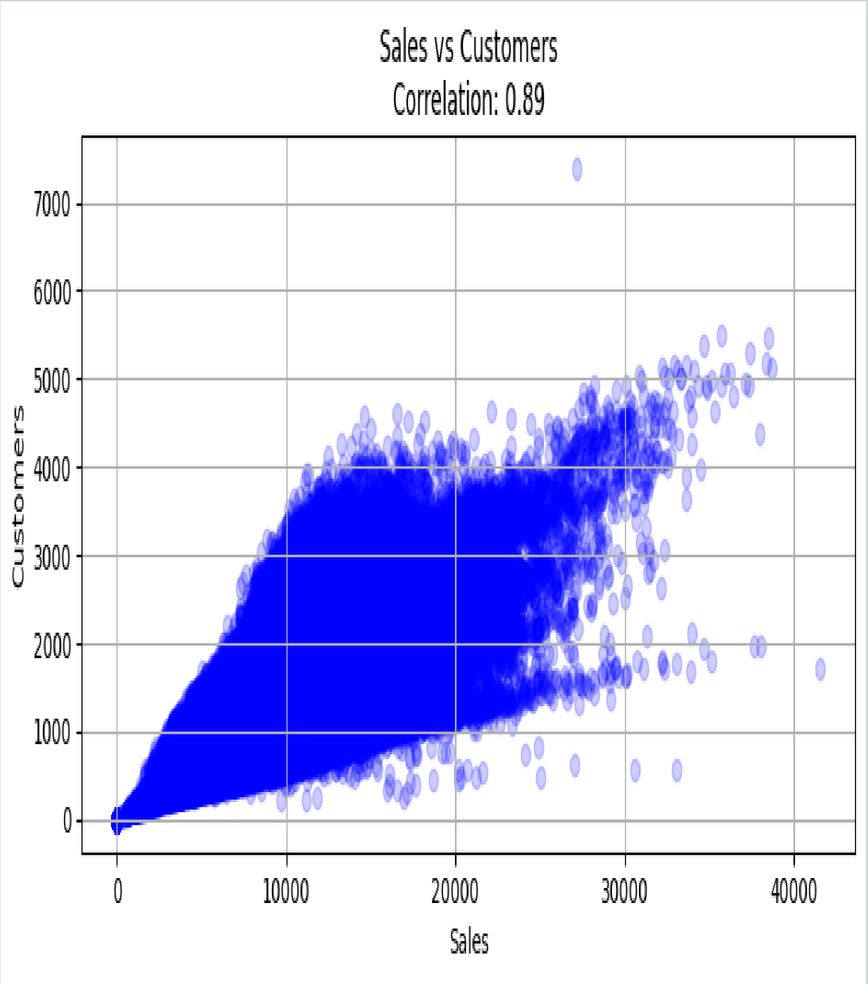

Sales per week
0
10000
20000
30000
40000
0
10
20
30
40
50


Sales on daily basis
count
0
5000
10000
15000
20000
25000
30000
Day
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Open
0
1


Sale Trends during Christmas and Easter
Total sales
260000
240000
220000
200000
180000
160000
140000
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07
2014-10
2015-01
2015-04
Date

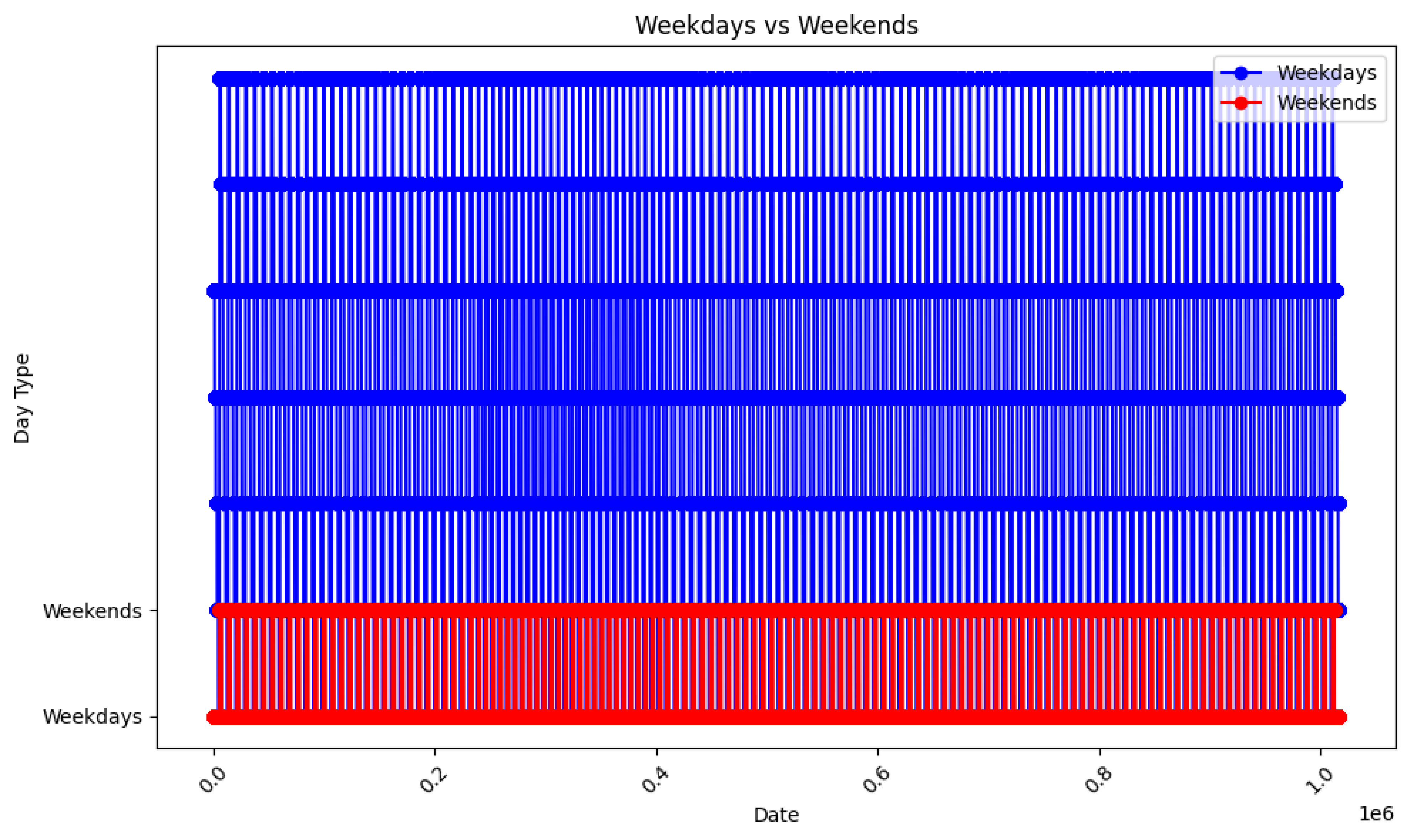

Purchase Behavior During Christmas and Easter
Average Customers per Day
35
30
25
20
15
10
5
0
Christmas (c)
Easter (b)
Average Sales per Customer
6
5
4
3
2
1
0
Christmas (c)
Easter (b)
Holiday Periods



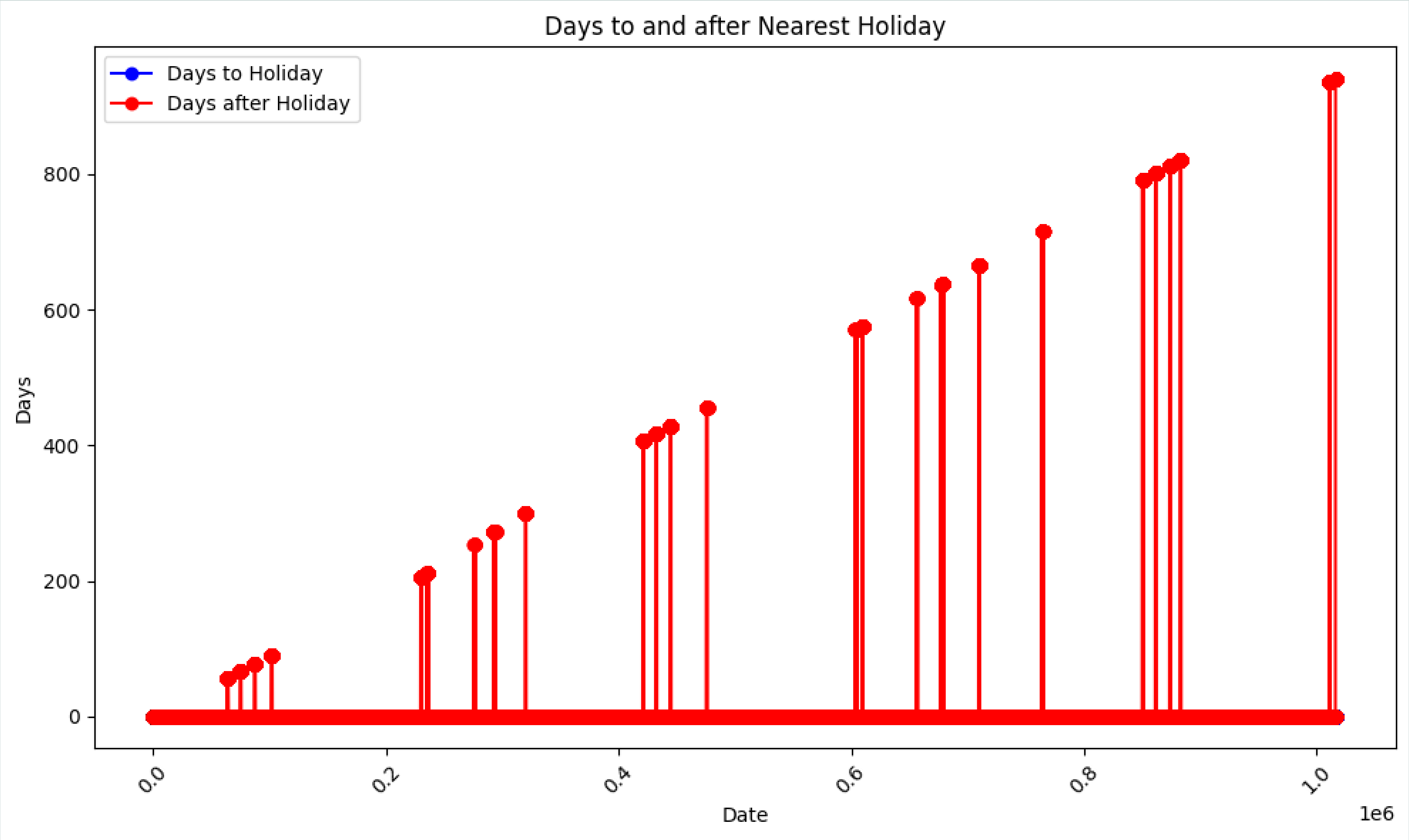


Average Sales is much more as compared to average Customers. Its Scatterplot also shows the same

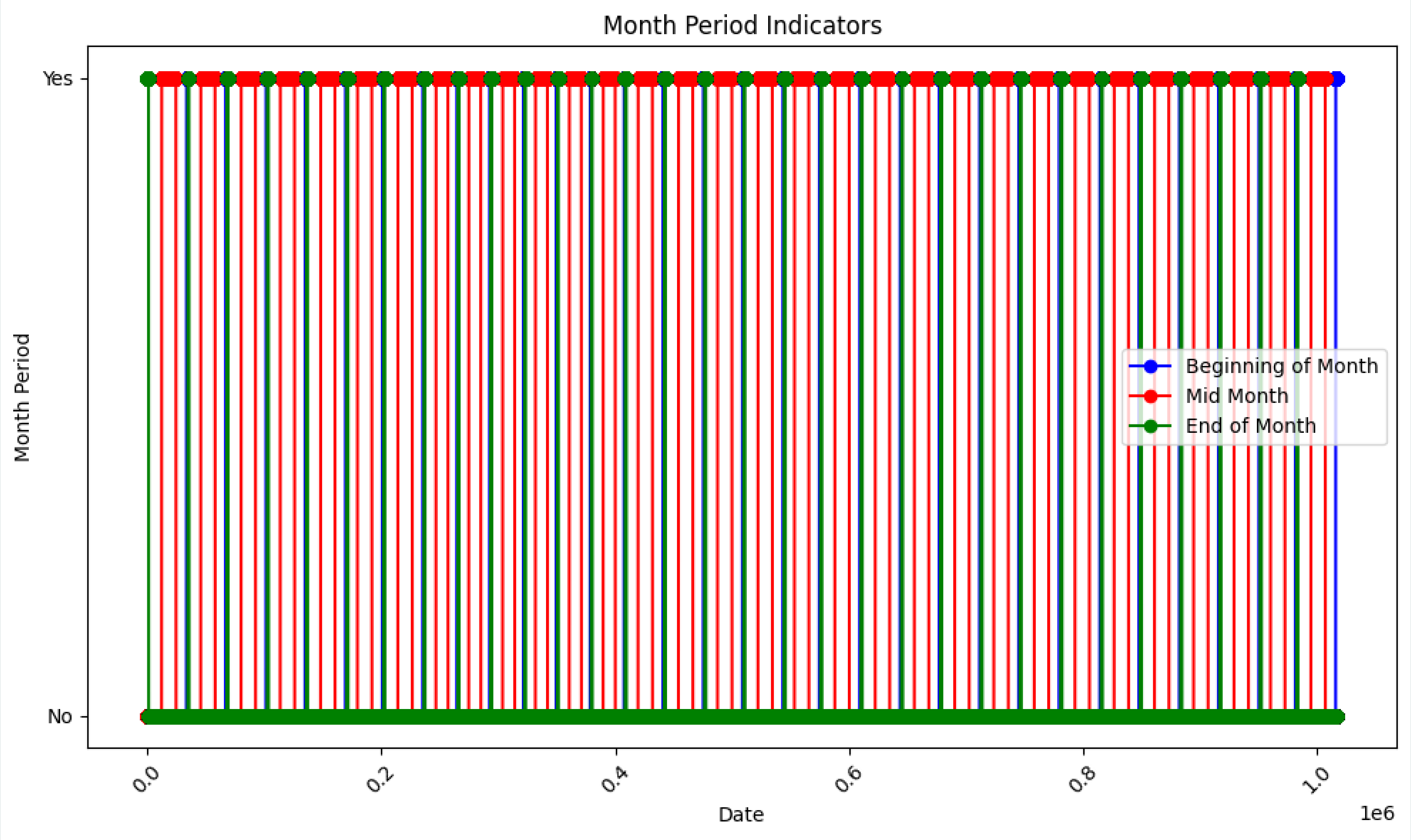
The correlation between Sales & Customers is 89% which is quite good



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



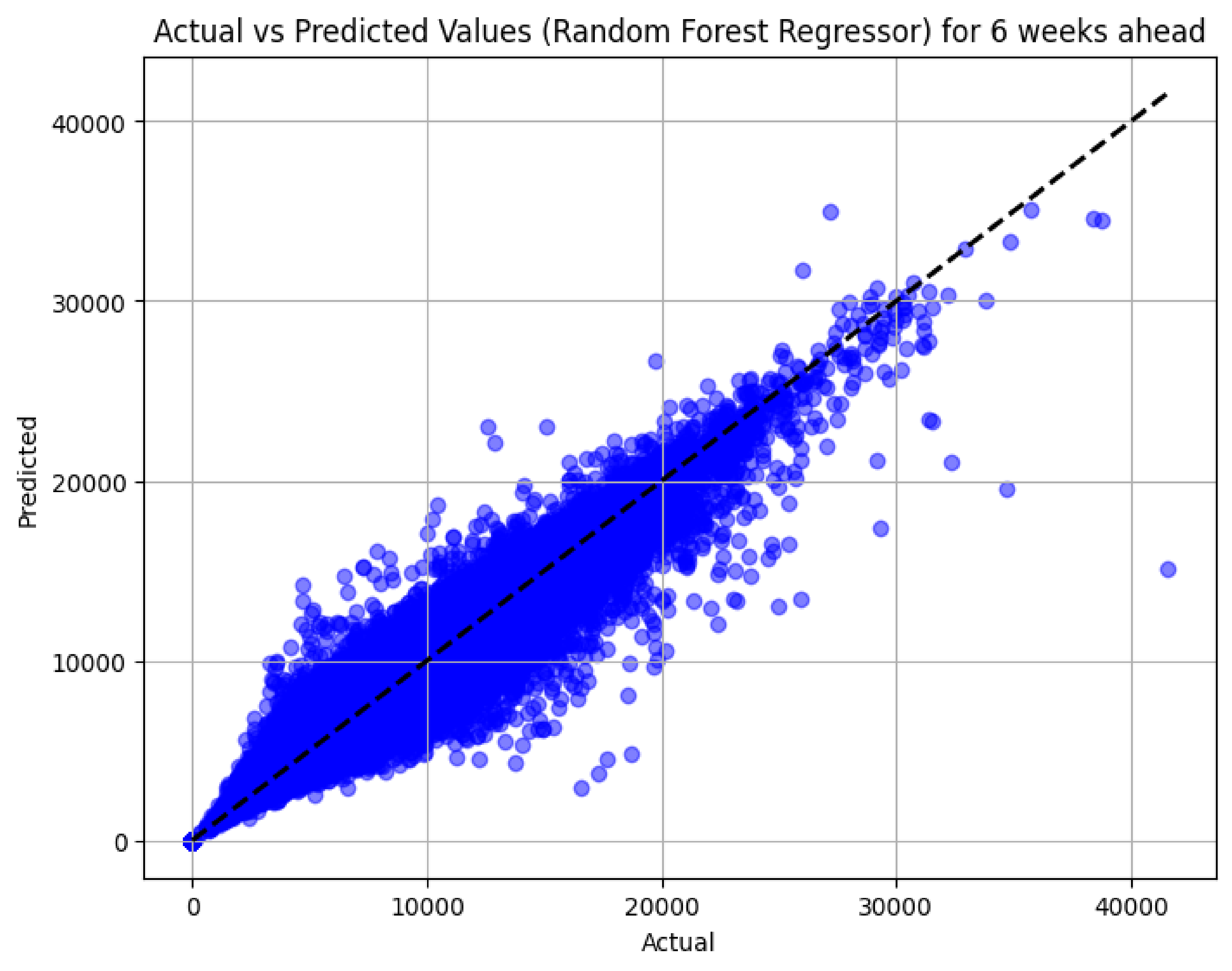
Days to Holidays and Days after Holidays have been calculated



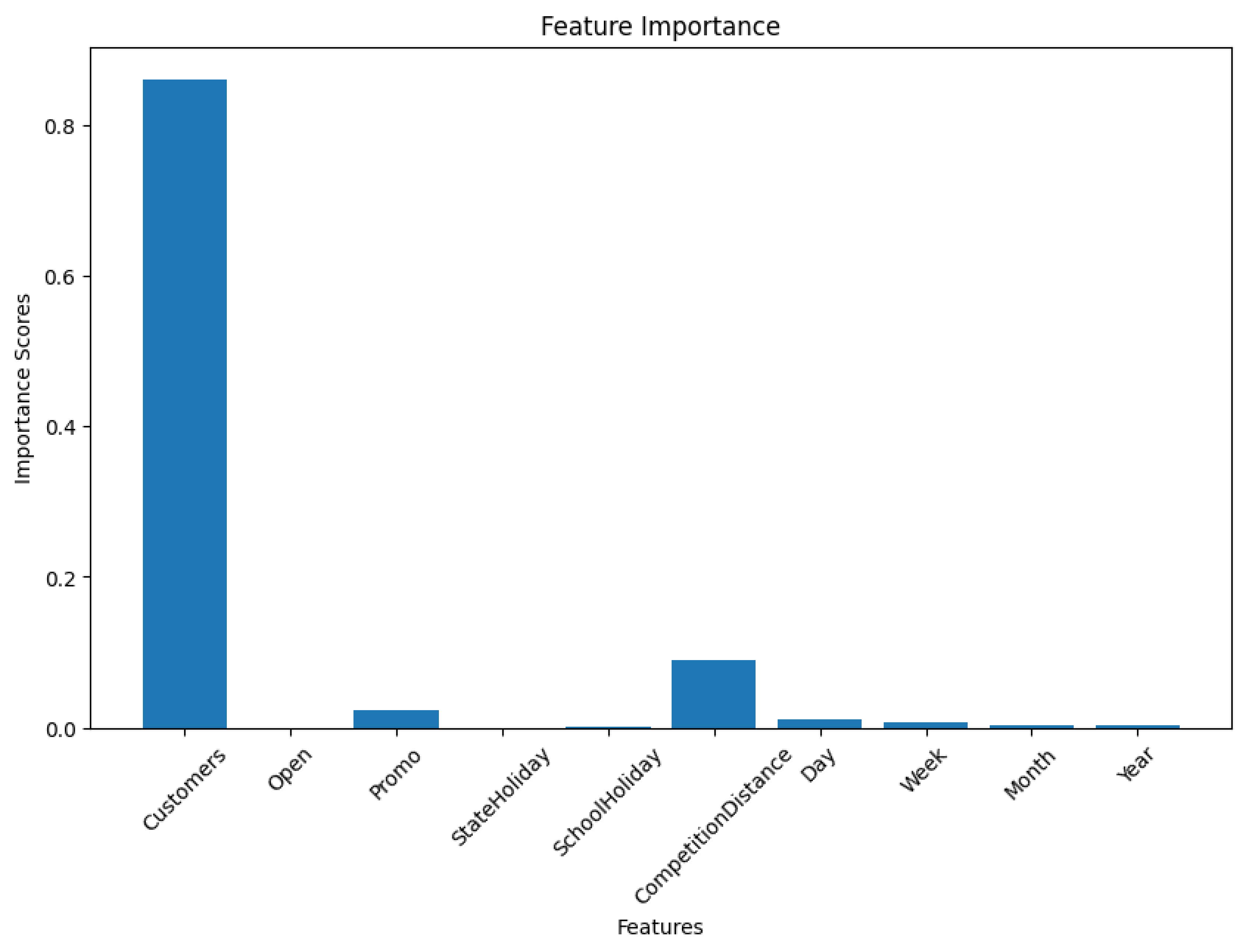
Month Period Indicators show the Beginning of month, Mid of month and End of month

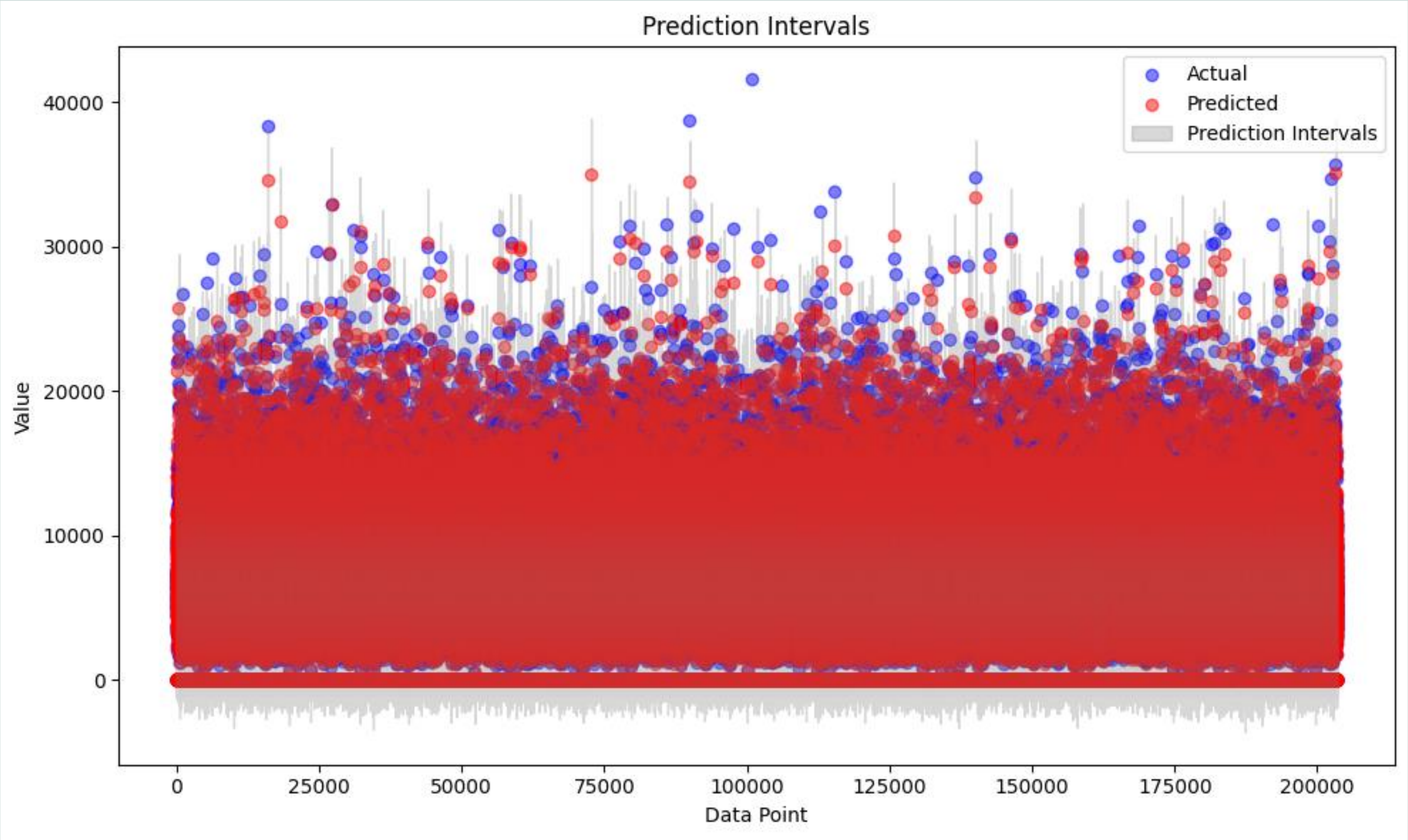
Random Forest Regressor with pipeline

Score = 96.13

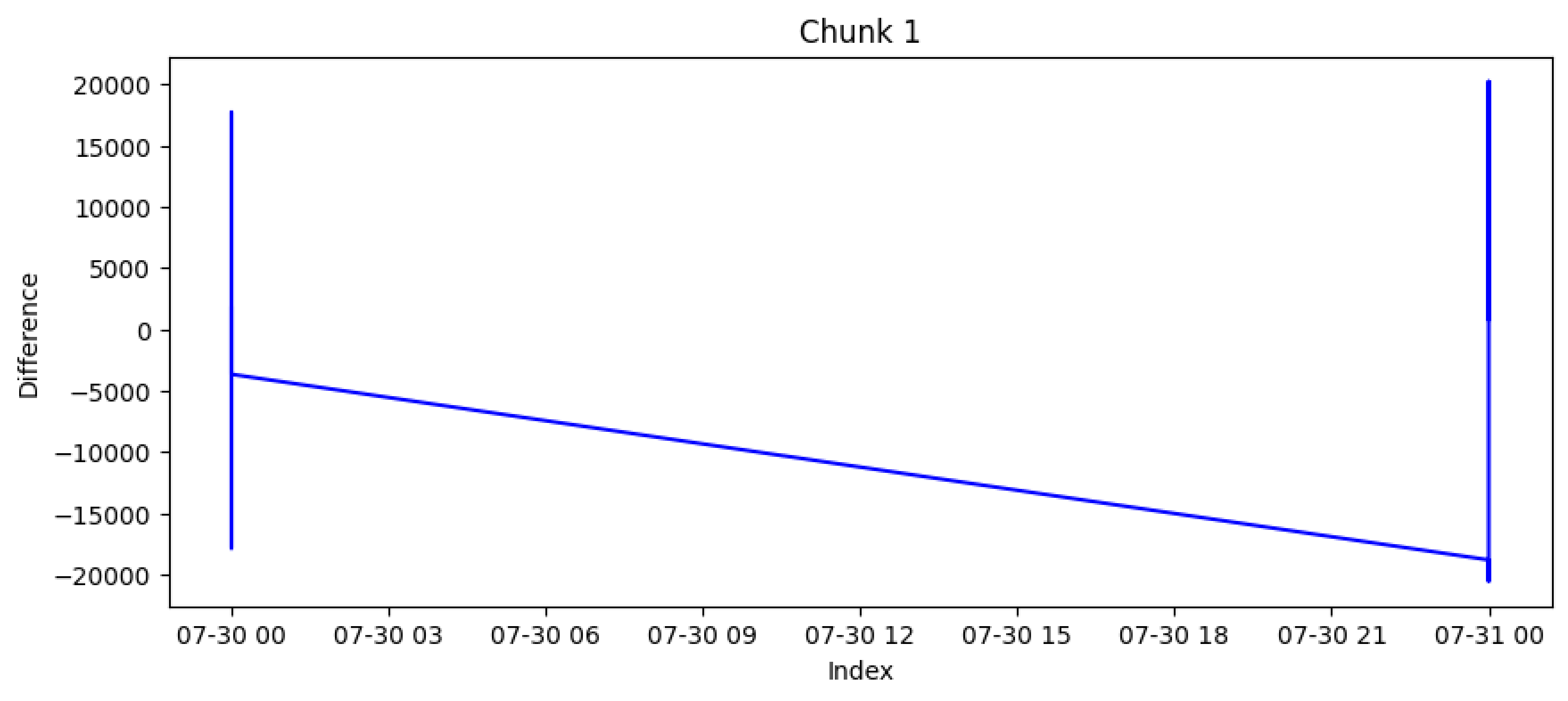


Customers is the most important feature which increases the Sales



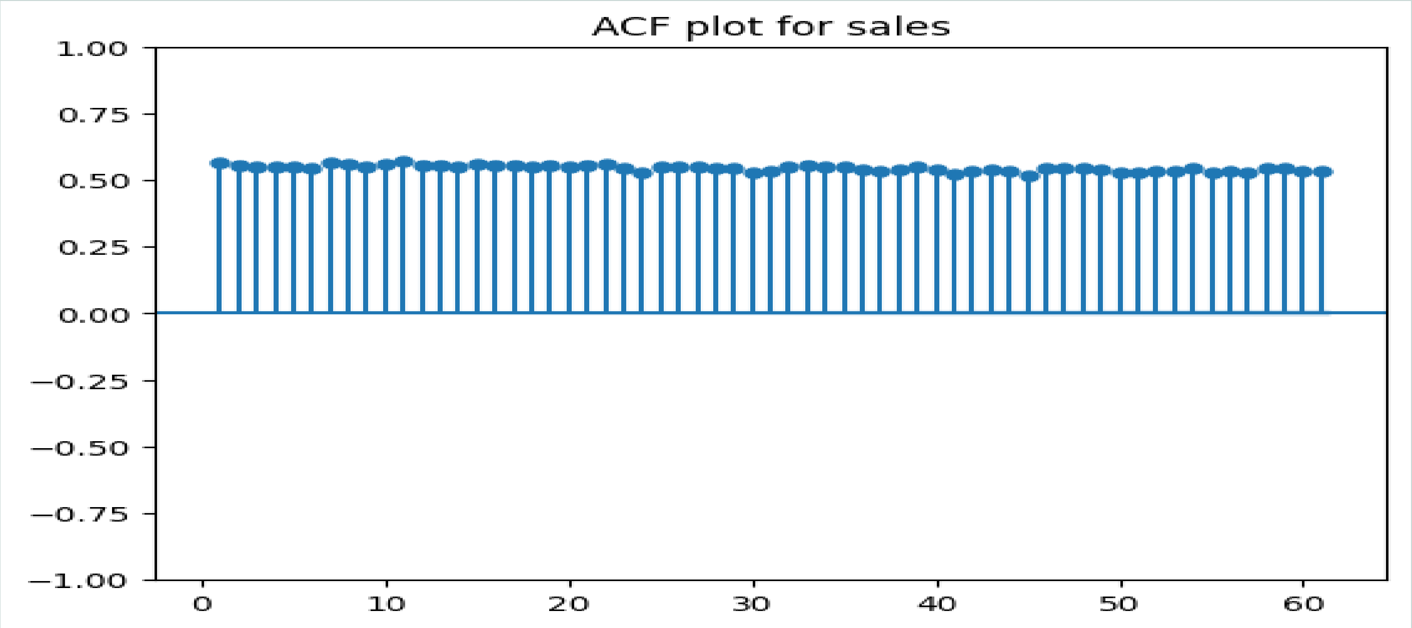


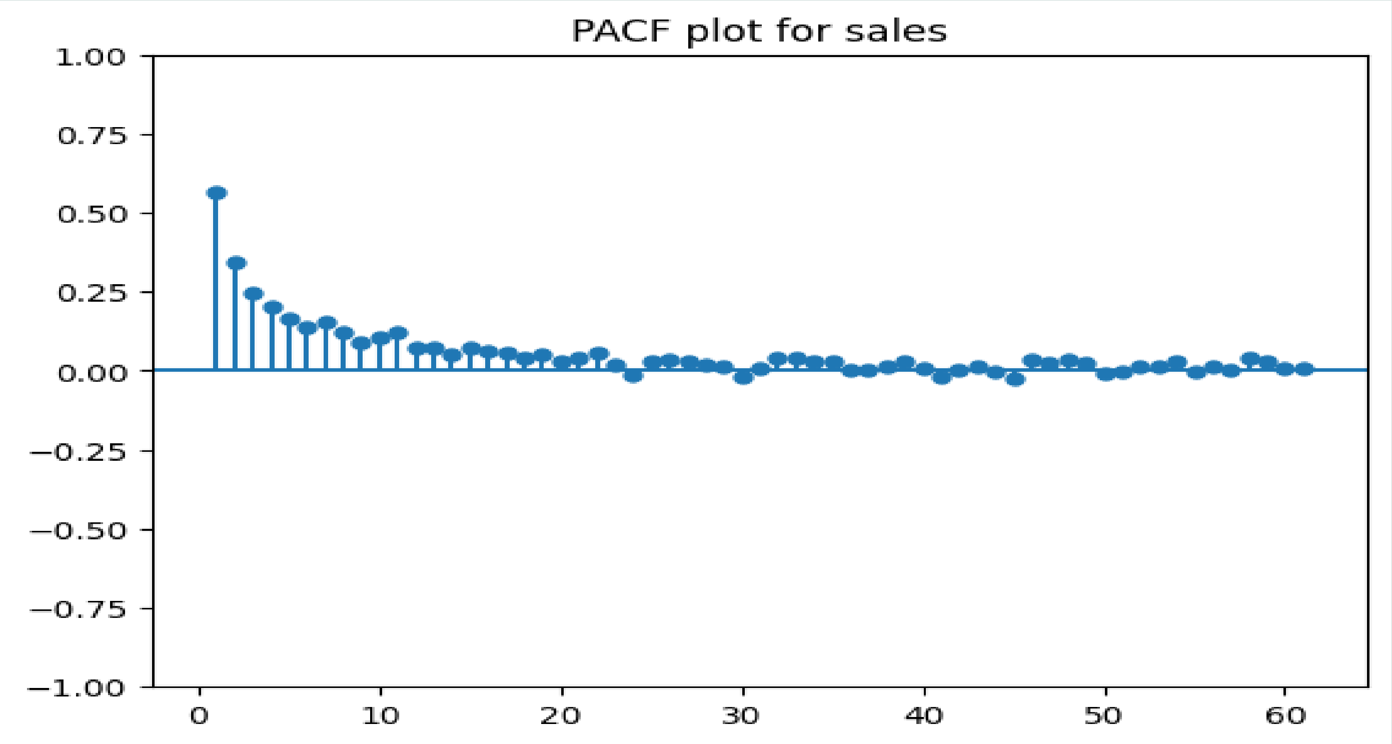
Prediction Intervals between Actual & Predicted according to Random Forest Regressor



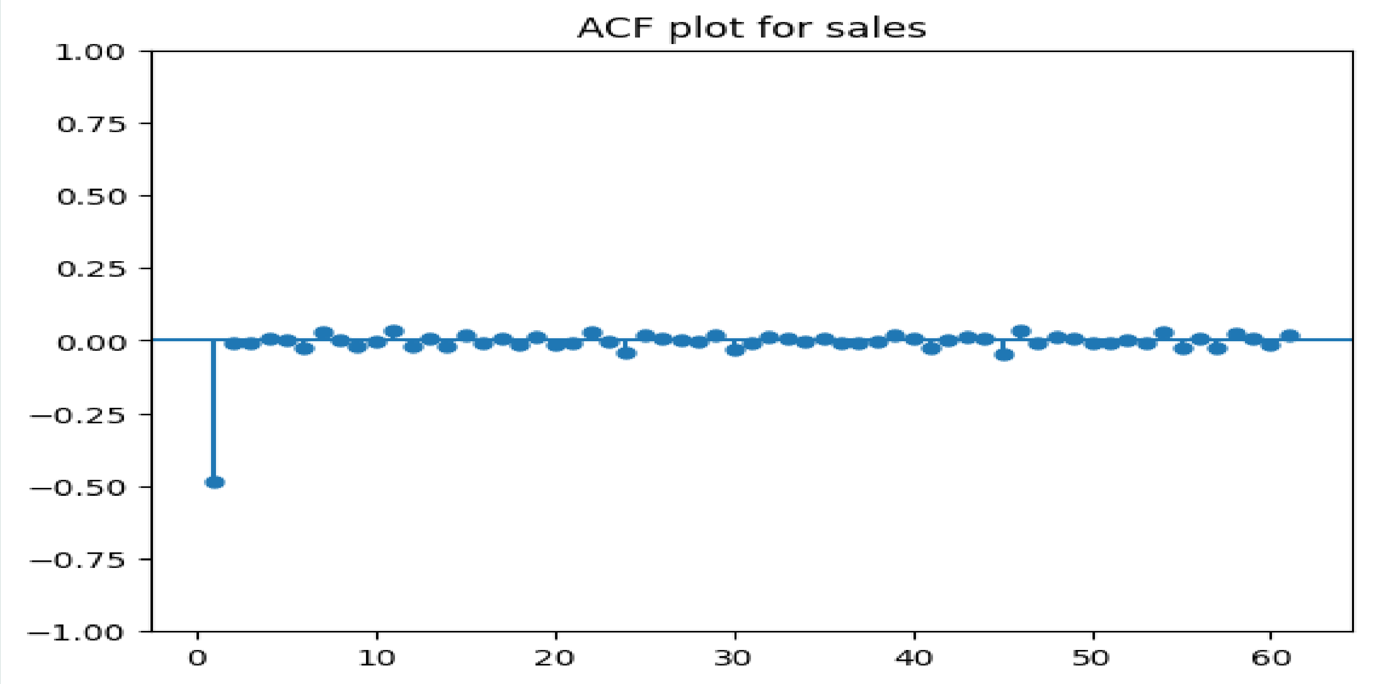
Chunk 1
ADF statistics: -15.454115638809549
p-values: 2.759621665977705e-28
Critical Values: {'1%': -3.433665351698571, '5%': -2.863004533702526, '10%': -2.567549656849864}
Reject the null hypothesis - Data is stationary

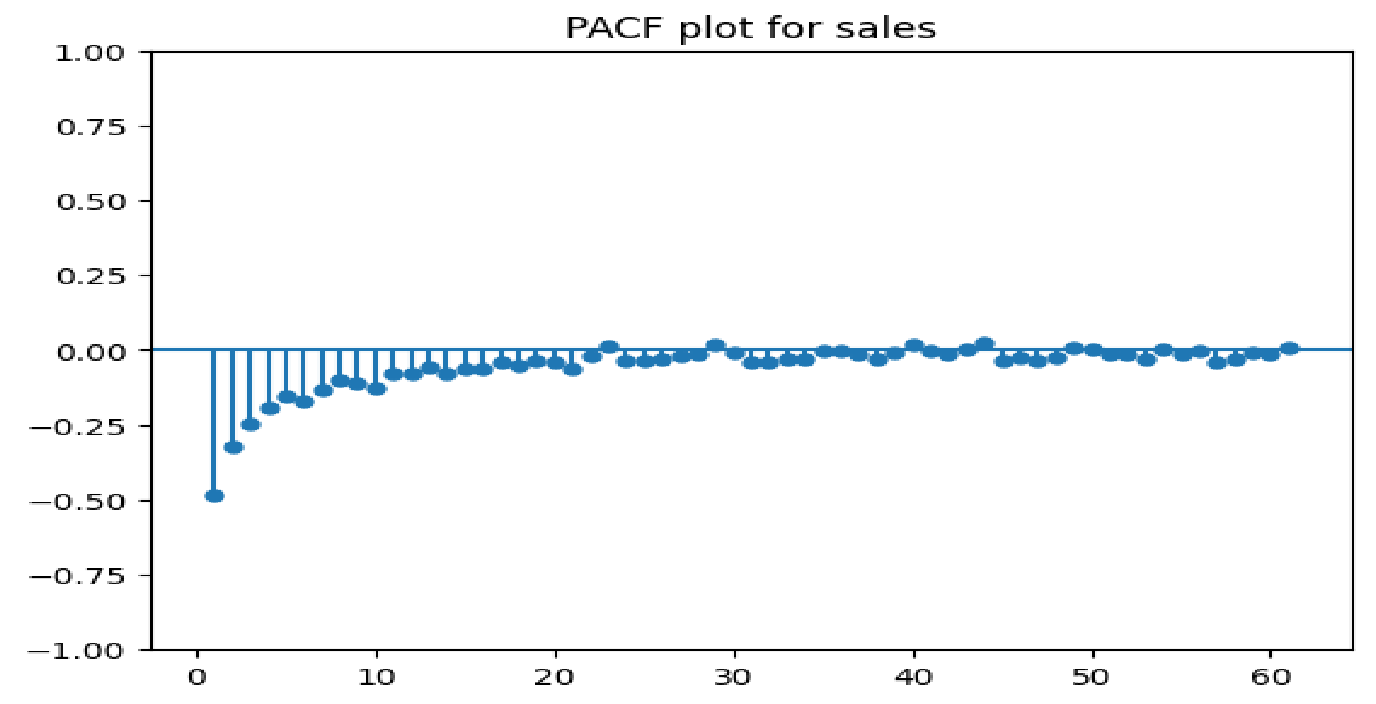
High autocorrelation for lags of multiples of 7. Maybe a seasonal component for day of week





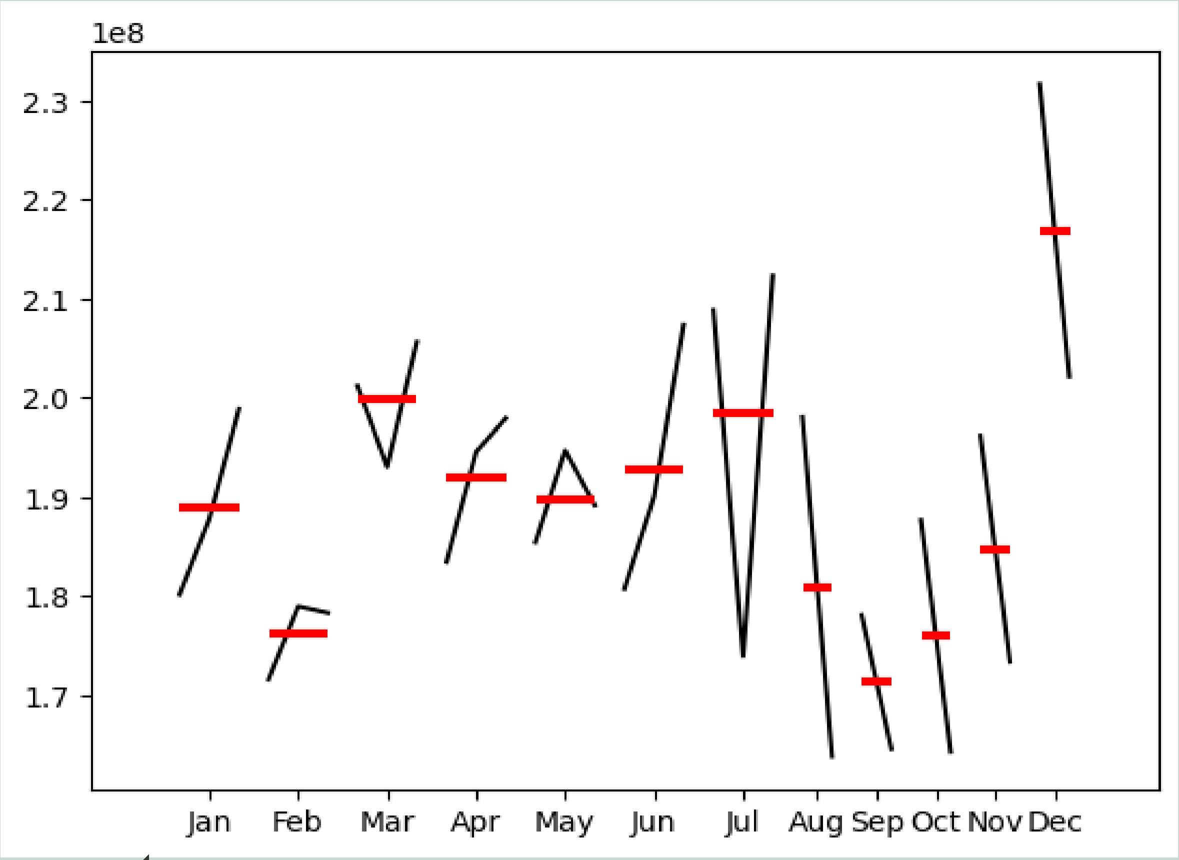
High partial autocorrelation for lags of multiples of 7

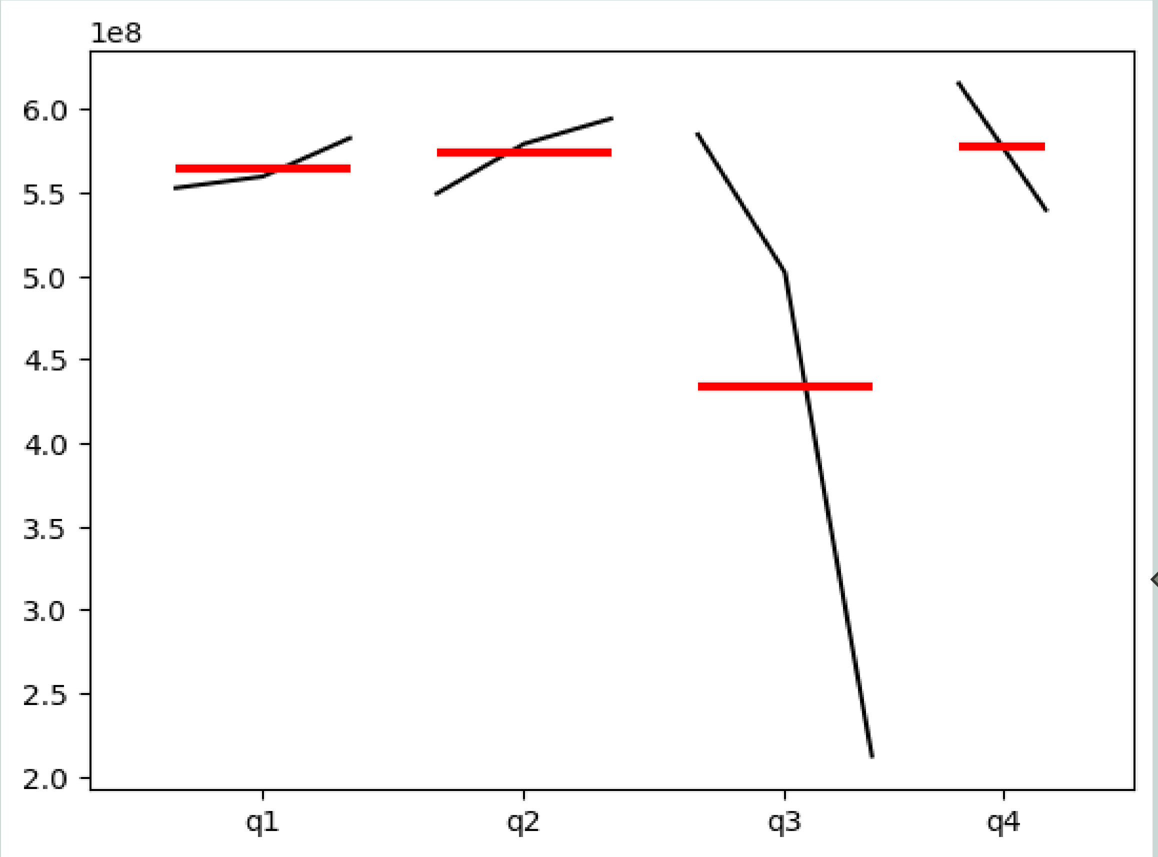




Monthly Plot

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

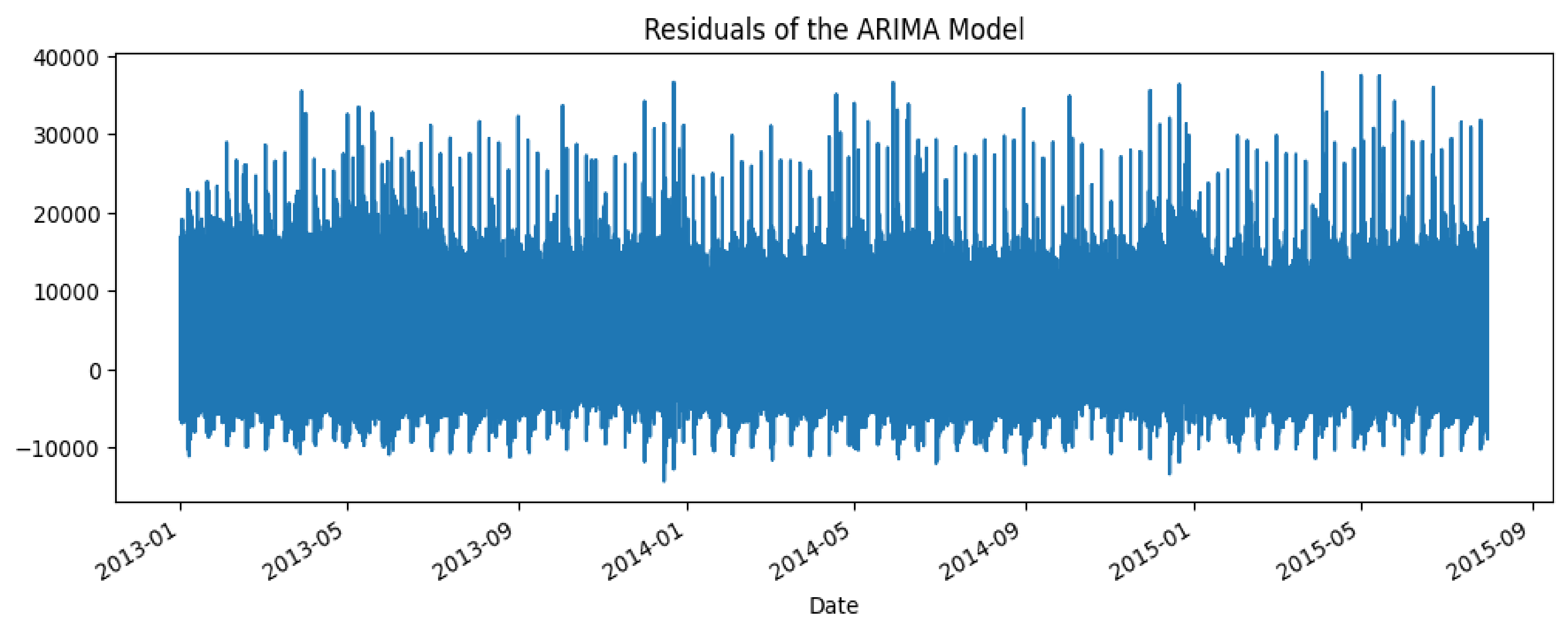




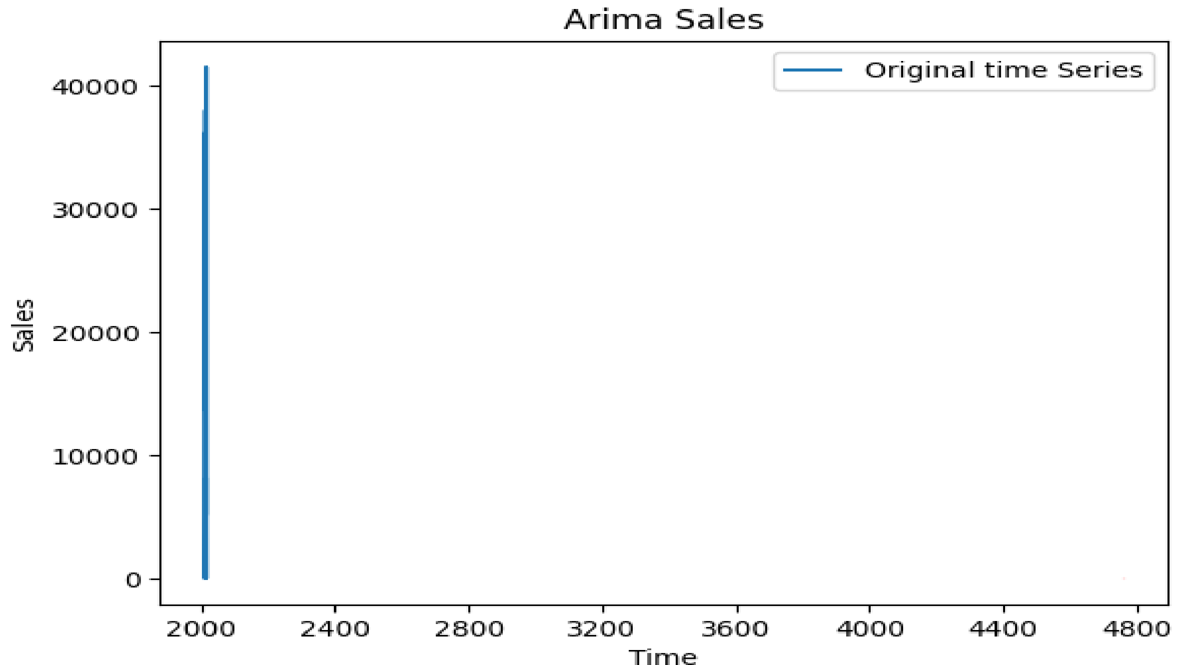
Quarter Plot

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

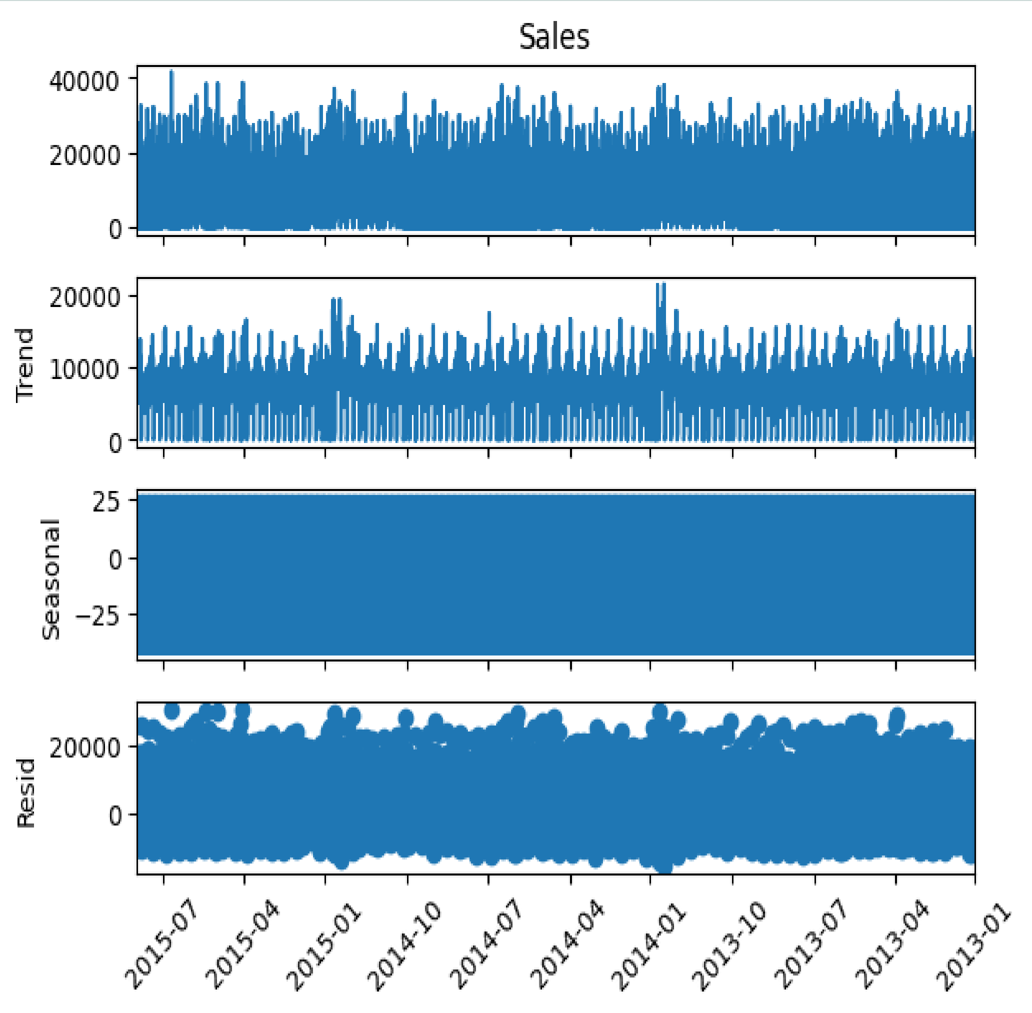
Residuals show the different between Actual Sales and the predicted fitted values

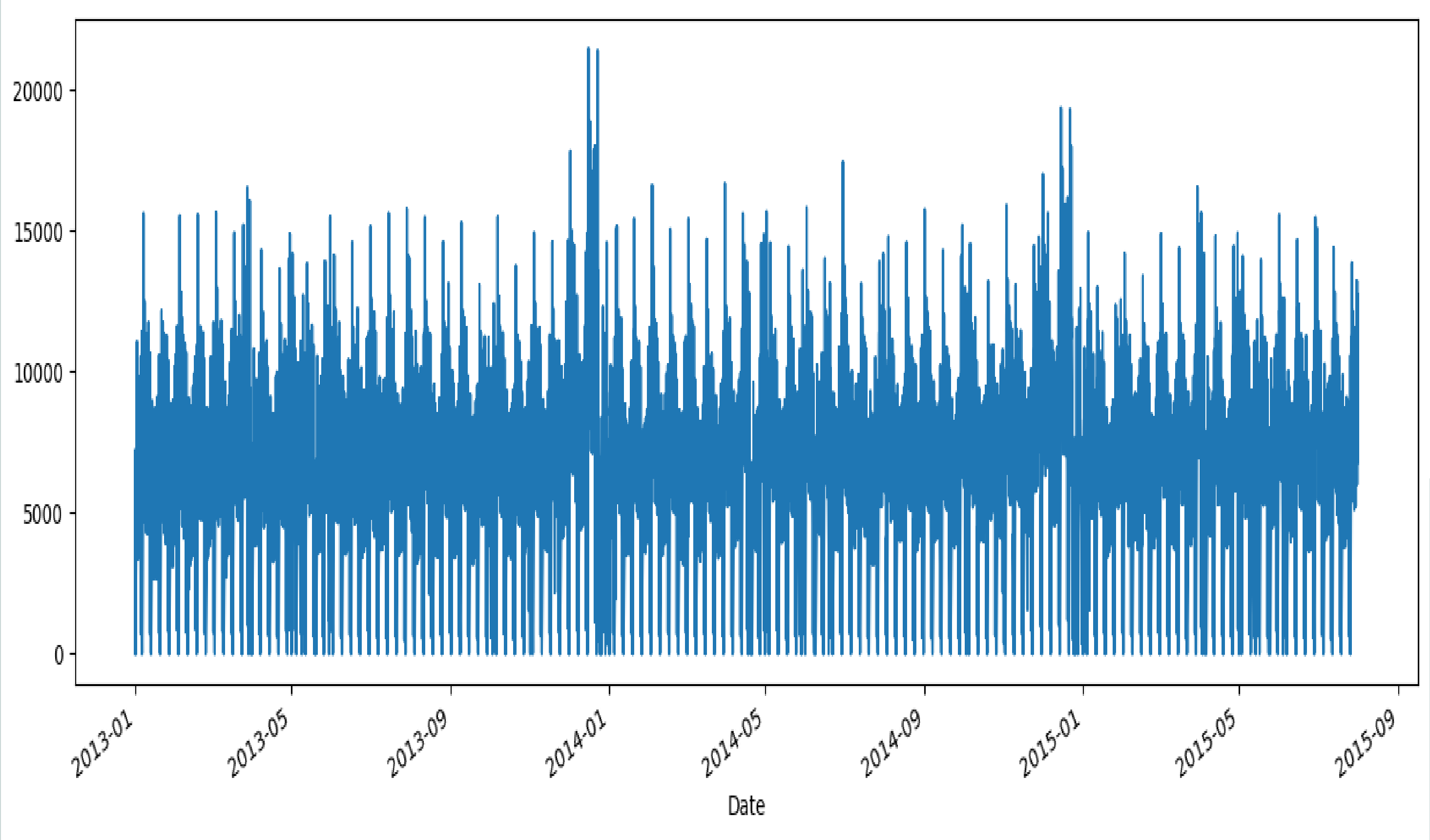


It shows the Sales forecast by Arima Model which is maximum at 2000

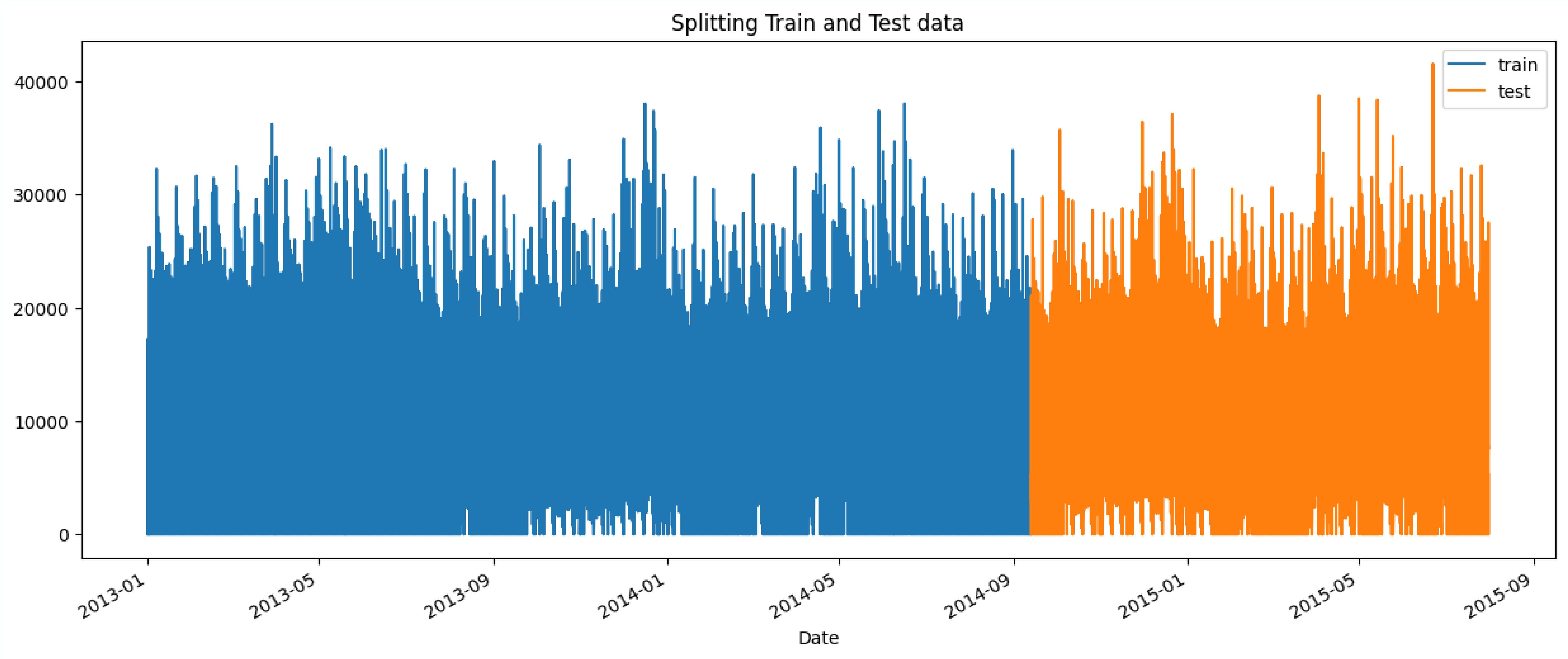


Seasonal decompose plot

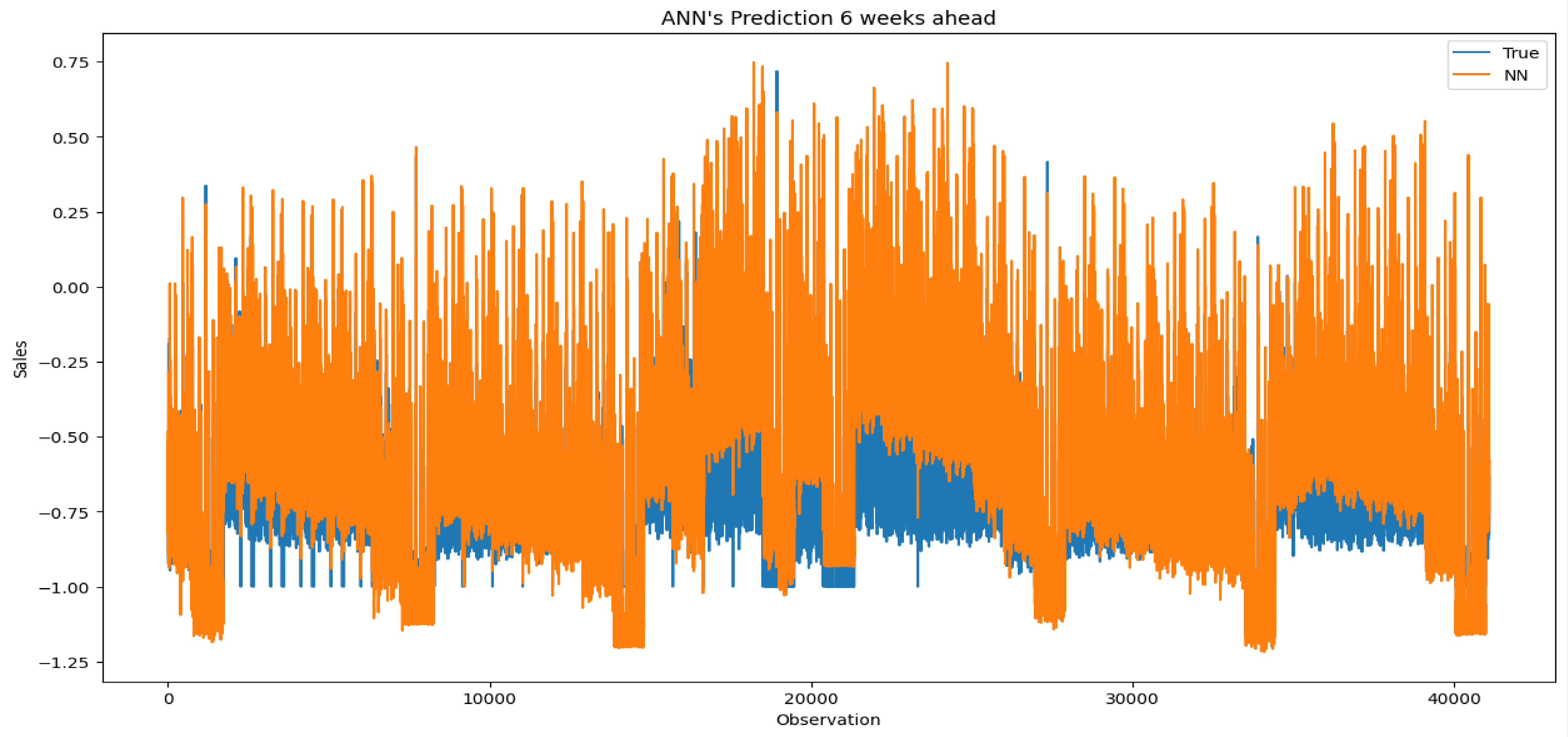




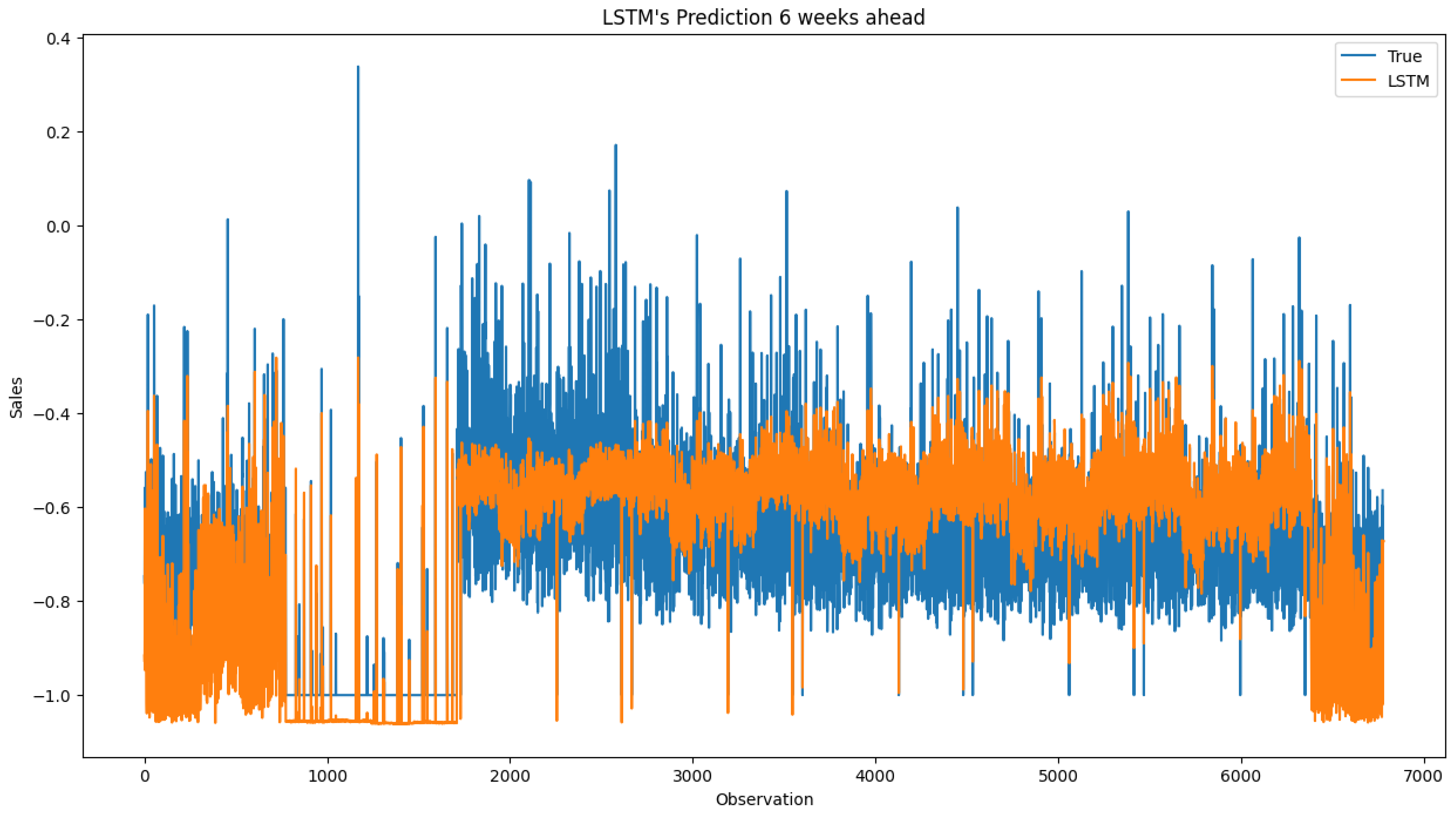
Decompose Trend Plot



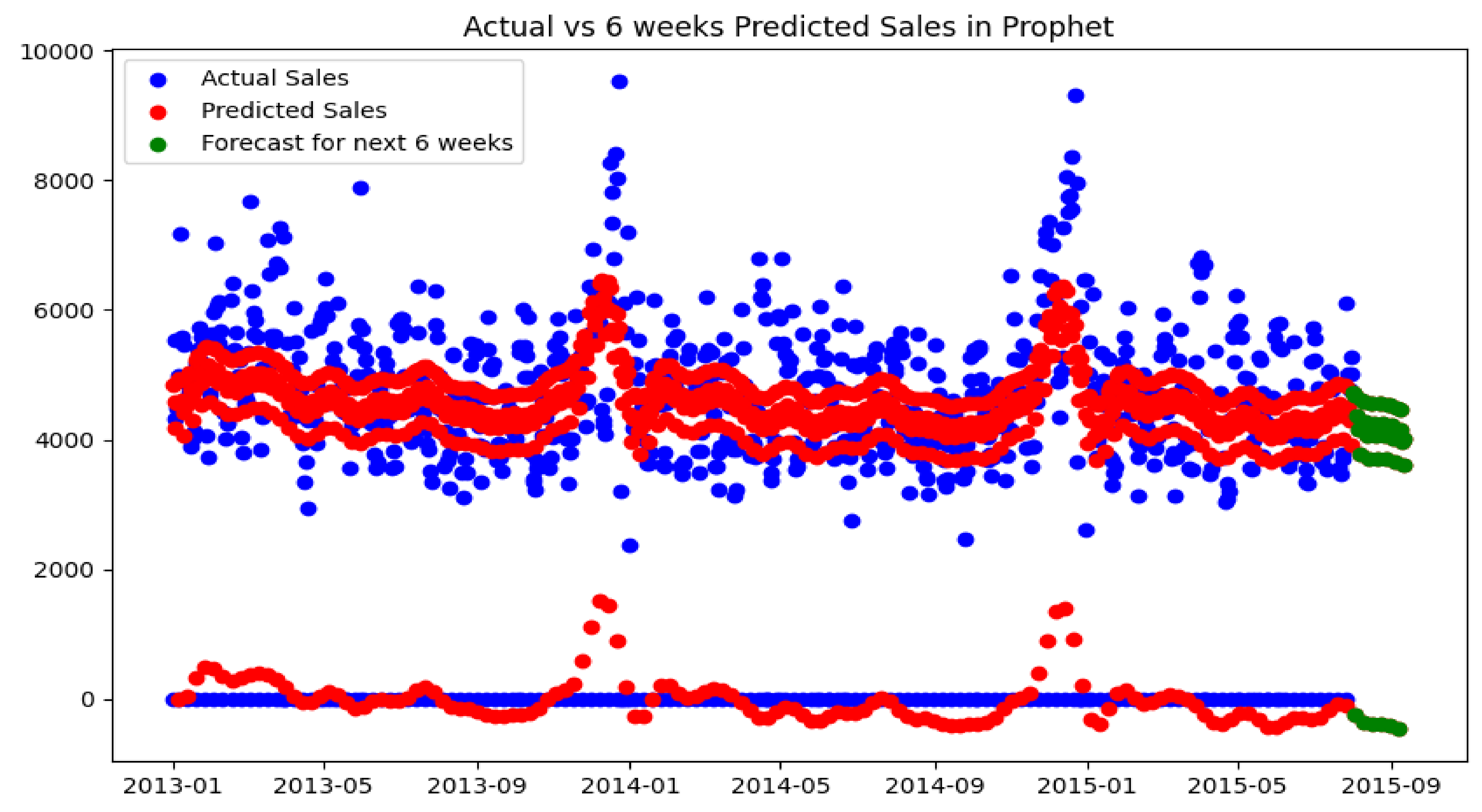
Year 2013 and 2014 have been taken in Train and year 2015 has been taken in Test data



Neural Network seems to perform better that a LSTM network



LSTM's Prediction 6 weeks ahead
True
LSTM
Sales
Observation
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
0
1000
2000
3000
4000
5000
6000
7000



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

Nitima Saigal

nitimasaigal@gmail.com