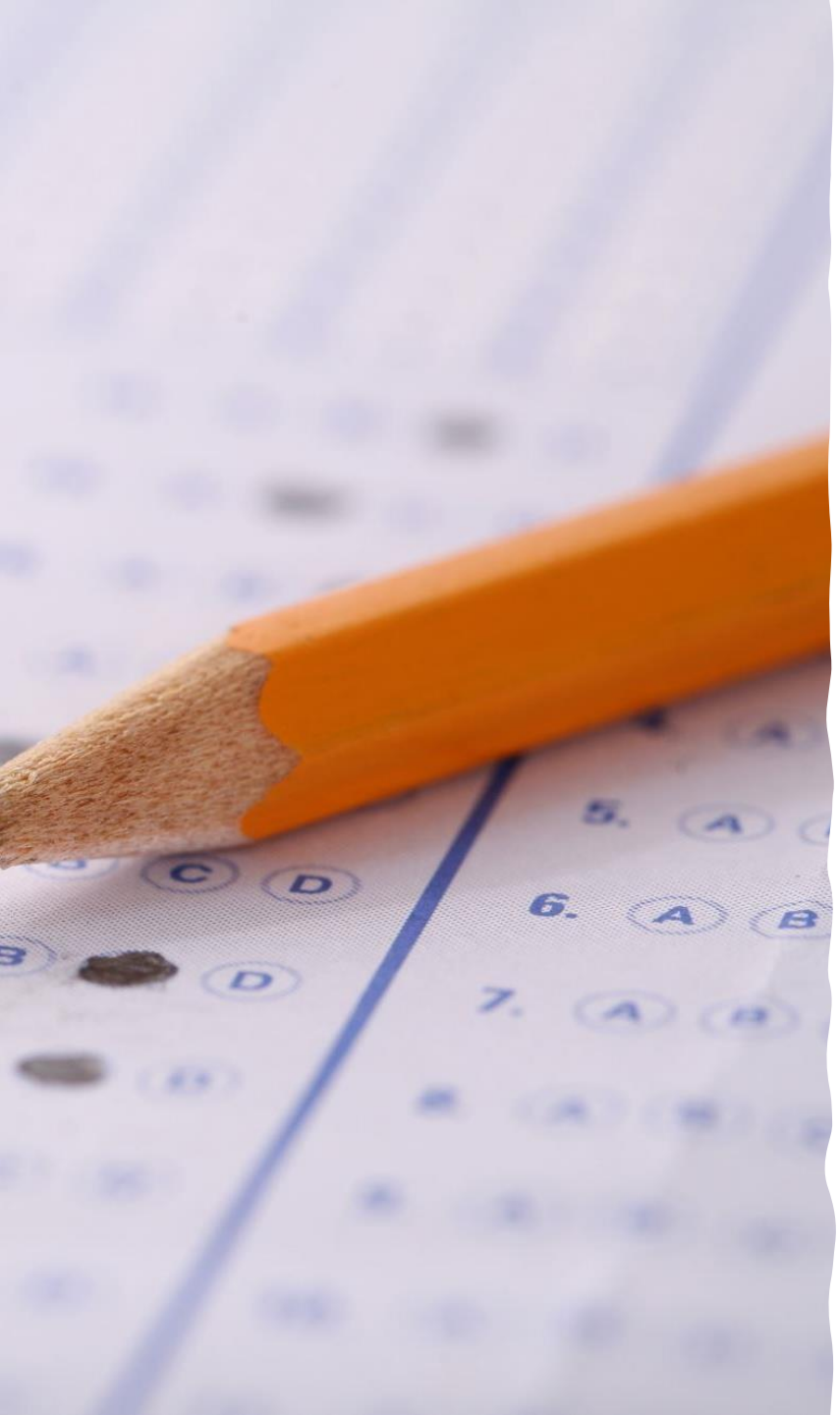


- Name : Shravan Tawri
- Student ID: 40168996
- Assignment 2 (Vanderbilt Elective Surgery Scheduling Case) Slides

# Background

Vanderbilt University Medical Center faces a problem in allocating it's resources because of uncertainty and high variation in the daily surgical volume cases

The overall monthly predictions are accurate but daily predictions have high degree of uncertainty

Late prediction in number of cases causes difficulty in making significant changes in the schedule

# Problem Statement

We need to figure out a way to make useful predictions of the daily number of surgeries with sufficient time in advance.

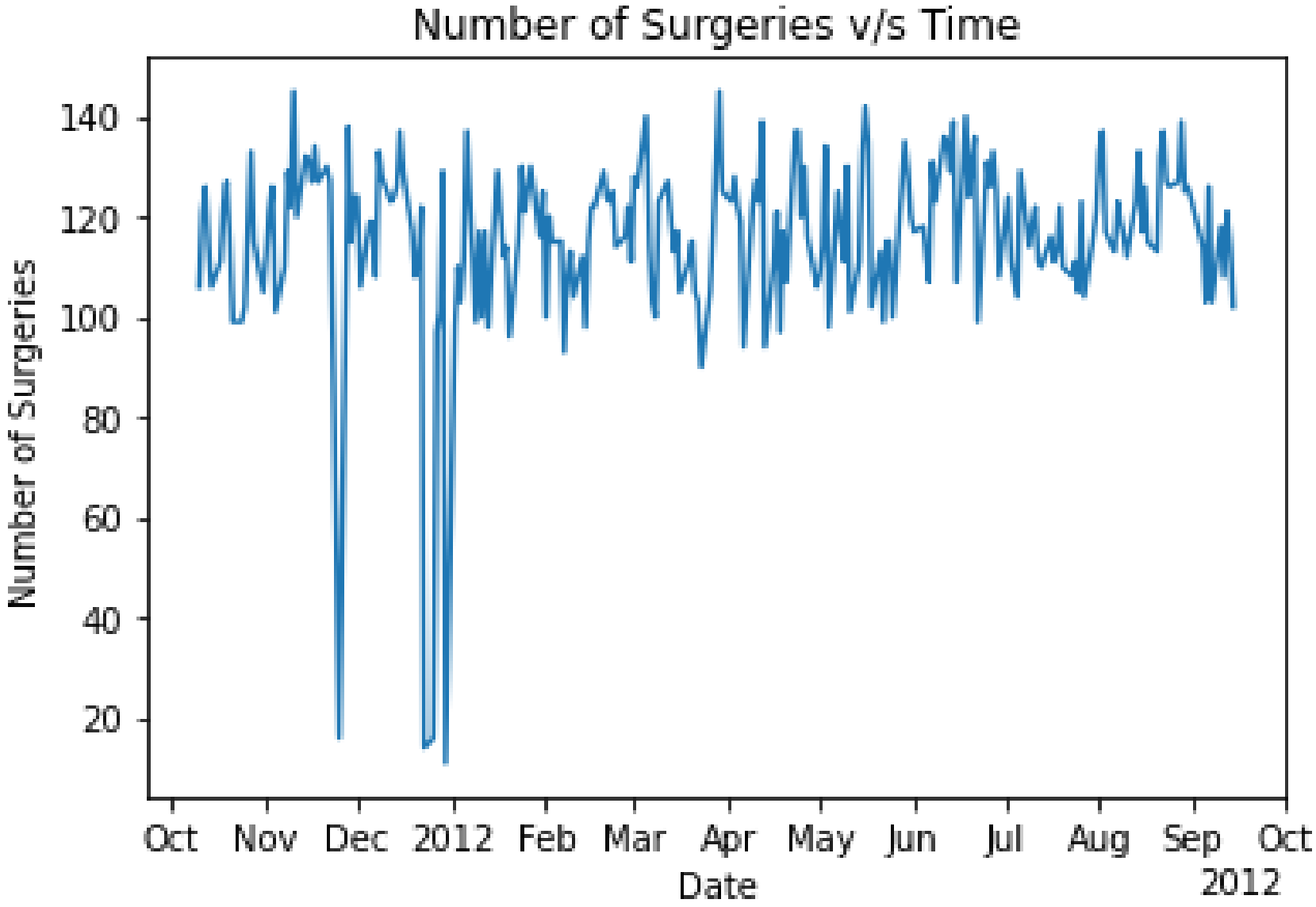
Comparison of different approaches of modeling

Figure out whether the developing surgery schedule could be more useful in predicting the daily number of surgeries than the previous days data.

Compare different timeseries model and derive useful insights

# Data Available

- The significant drops in the number of surgeries for certain days and the sudden rise in subsequent days is because on certain days, the surgeons are on vacation or attend academic meeting



# Linear Regression Approach

- For the past 48 weeks, data is available for all surgeries booked up to 30 days in the future

| Surgery Date | Day of Week | Number of surgeries scheduled in days prior to the surgery date | | | | | | | | | | | Actual Number of Surgeries* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | T – 28 | T – 21 | T – 14 | ... | T – 7 | T – 6 | T – 5 | T – 4 | T – 3 | T – 2 | T – 1 | |
| 10/10/2011 | Mon | 38 | 45 | 60 | | 80 | 84 | 89 | 94 | 98 | 100 | 104 | 106 |

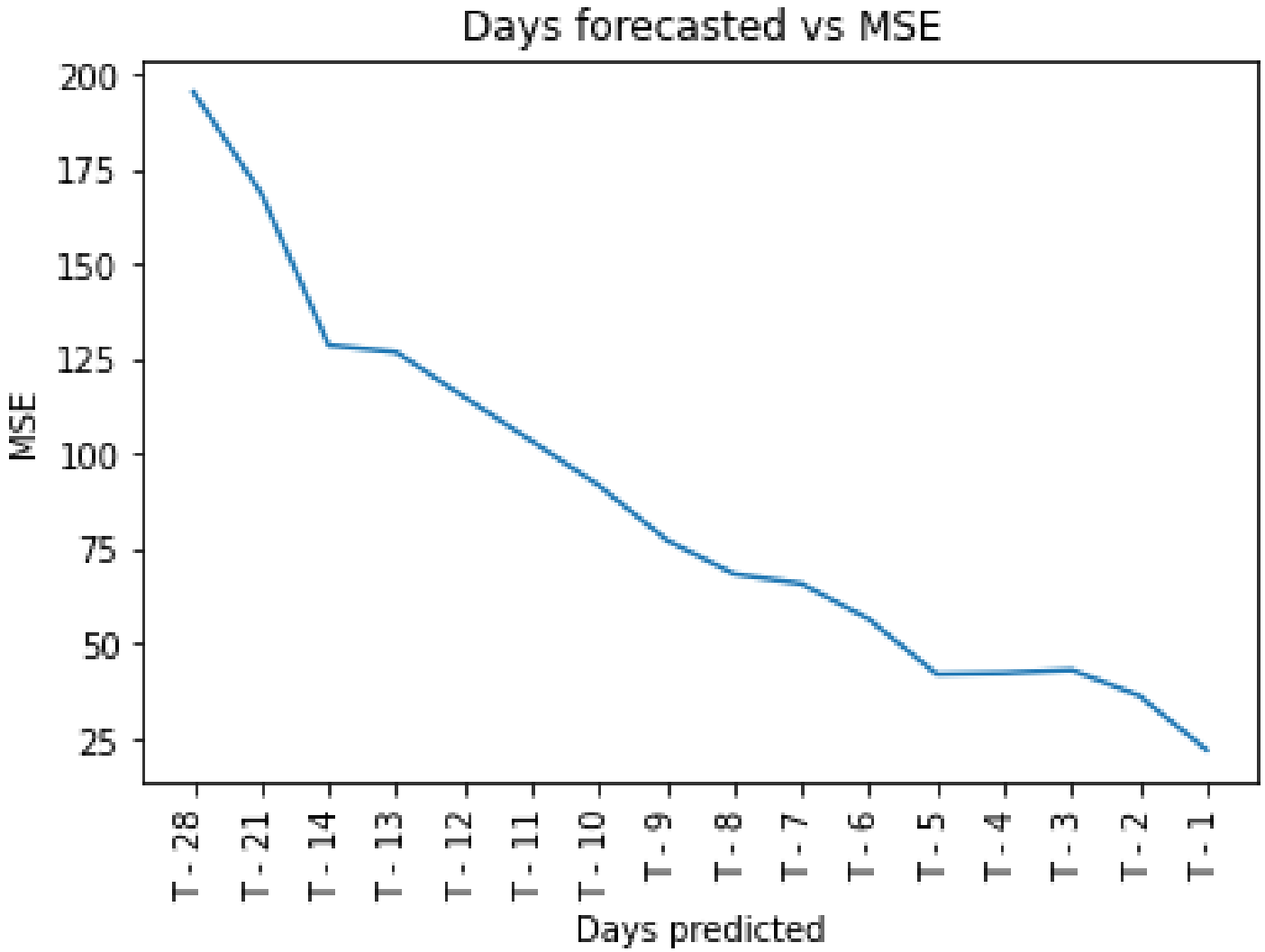
- Example: T-28 indicates number of surgeries scheduled to take place a day 28 days prior from that day.
- This can be used to build a linear regression model to see the effect of number of surgeries scheduled in prior days and the number of actual surgeries taking place on a day
- Training data : 80%
- Test Data: 20% (split randomly)

# Forecasted days vs Error in forecast

- We see that the accuracy of prediction increases as we move closer to the day of surgery but the flexibility of our schedule decreases (since its difficult to schedule certain resources in short span of time

- Therefore, we need to find a trade-off between the two.

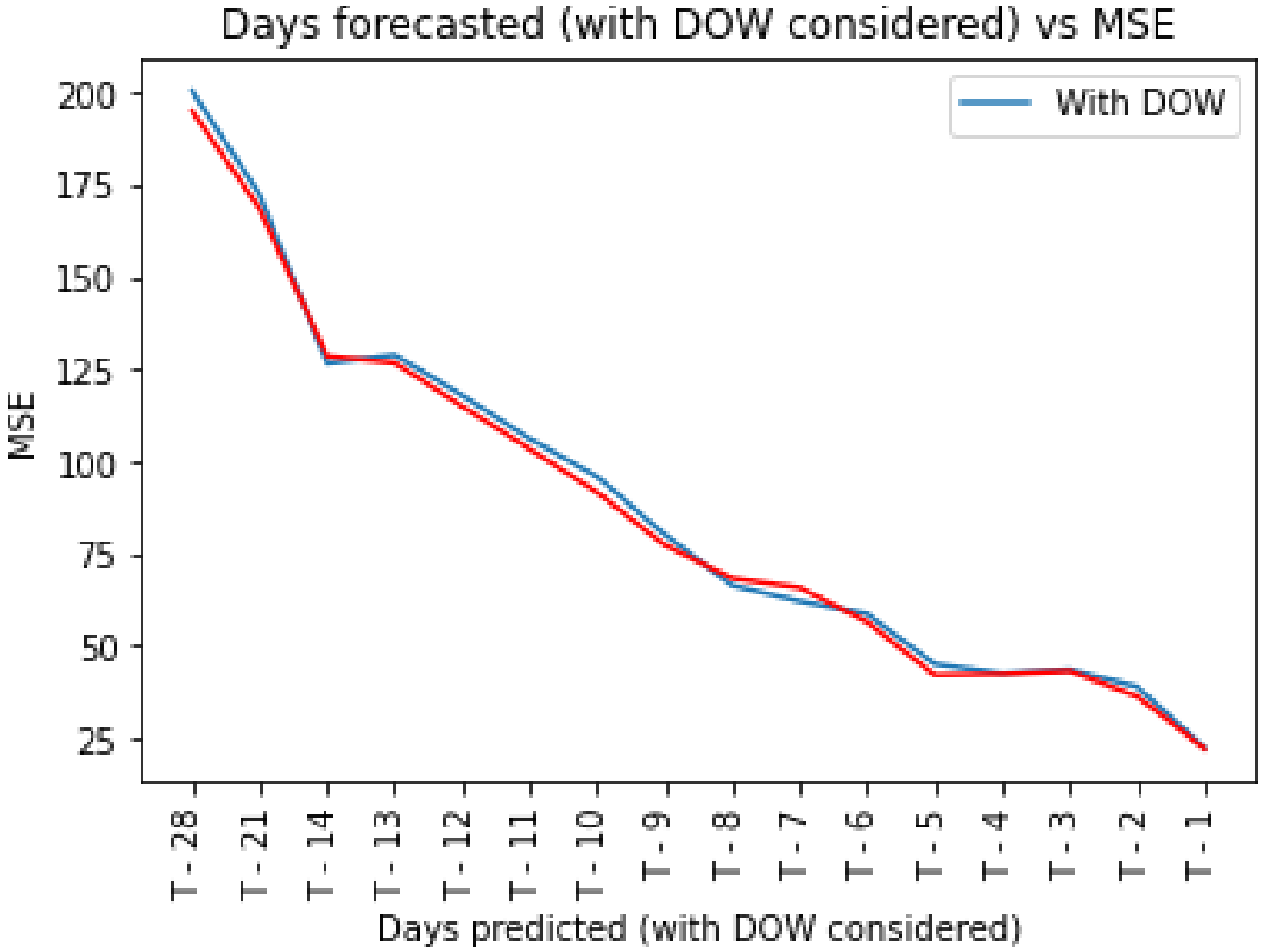
- We can make accurate enough predictions 5 days before the day of surgery with MSE = 42.1



# If day of week is considered

- There is no significant improvement in the model

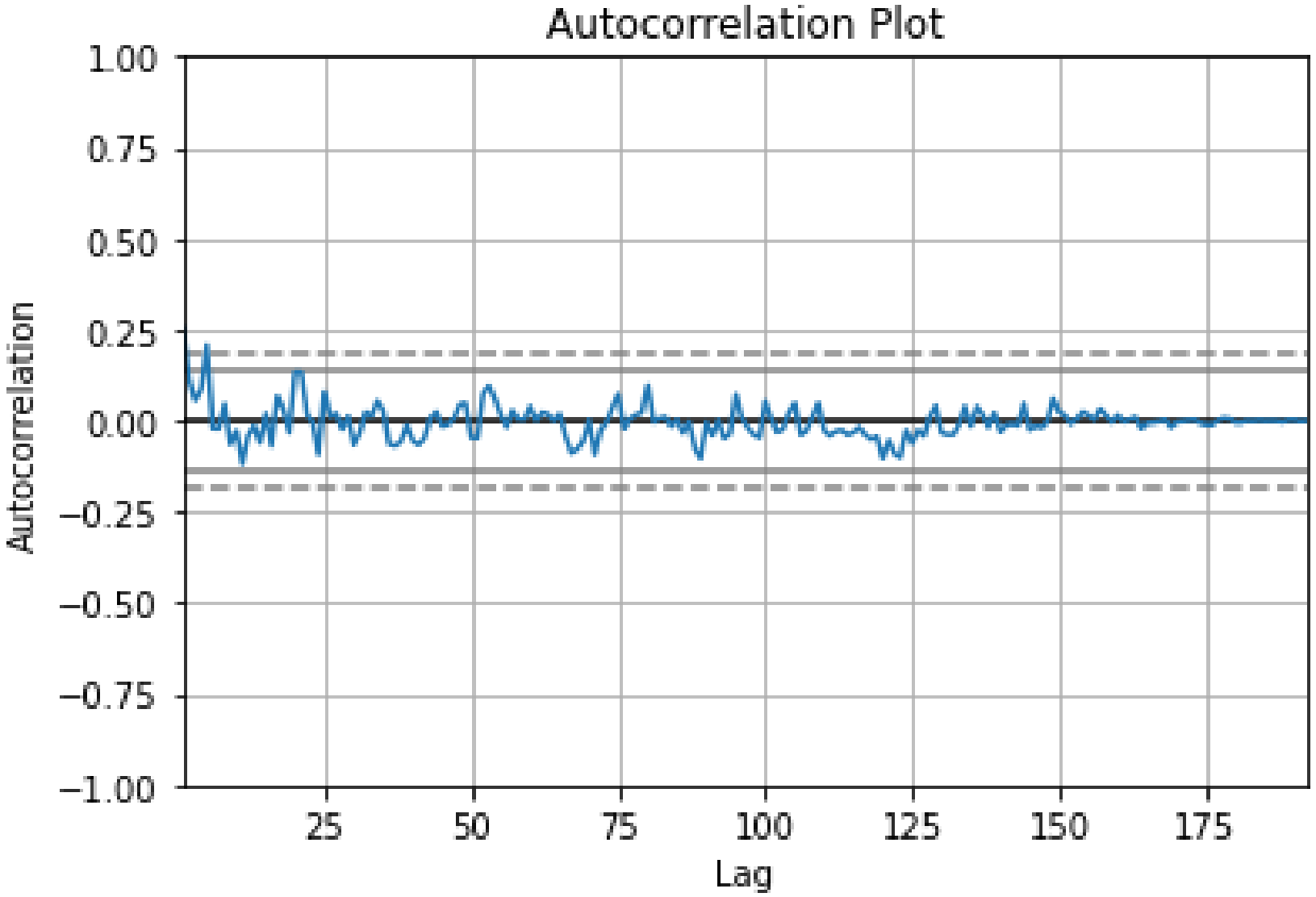
- Therefore, we can conclude that the day of the week has no impact in determining the number of surgeries taking place on the day



# Time Series

- Data of 241 consecutive weekdays is used to build a time series model

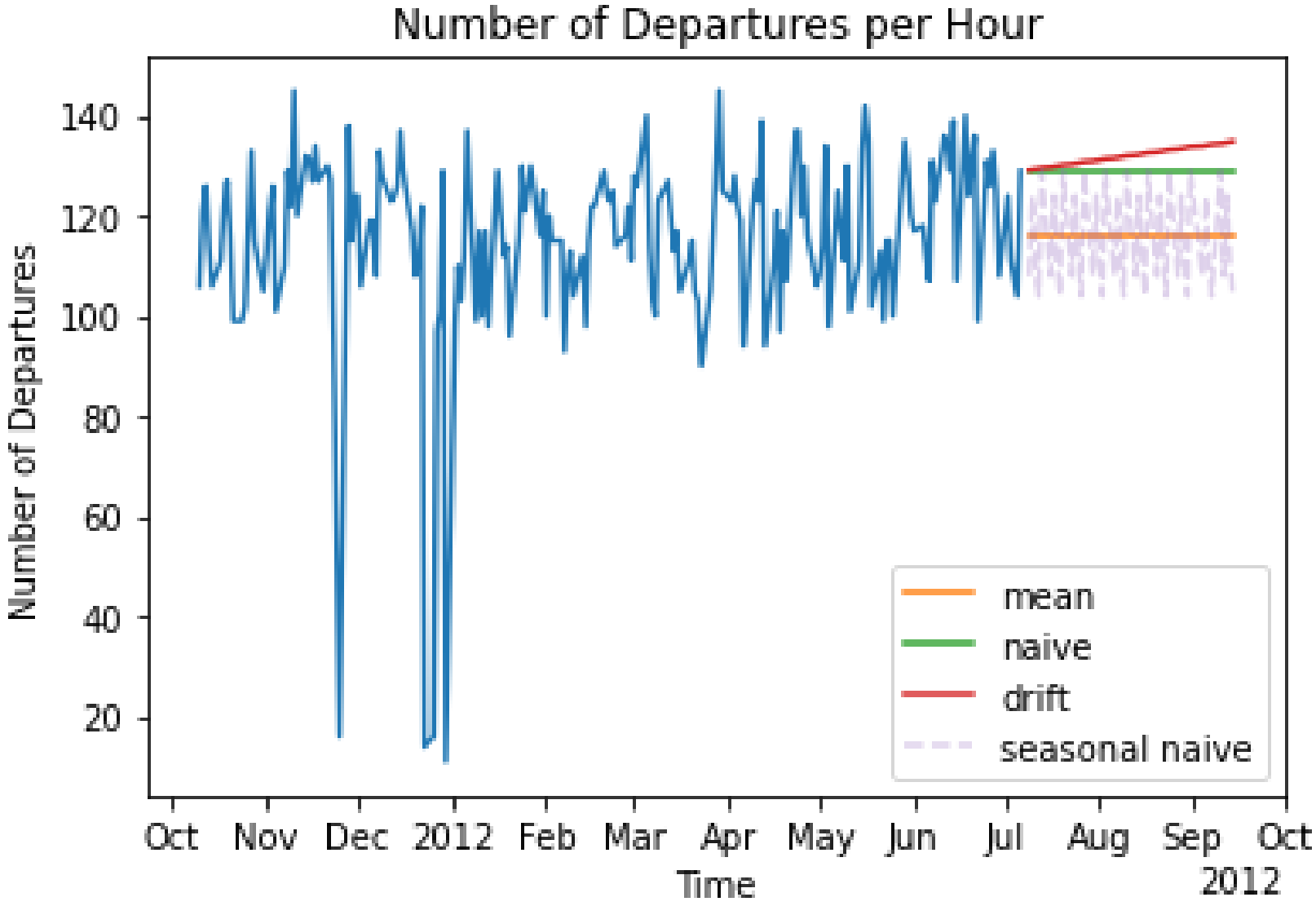
- From the autocorrelation plot we can see that there is no general trend or seasonality in the given data



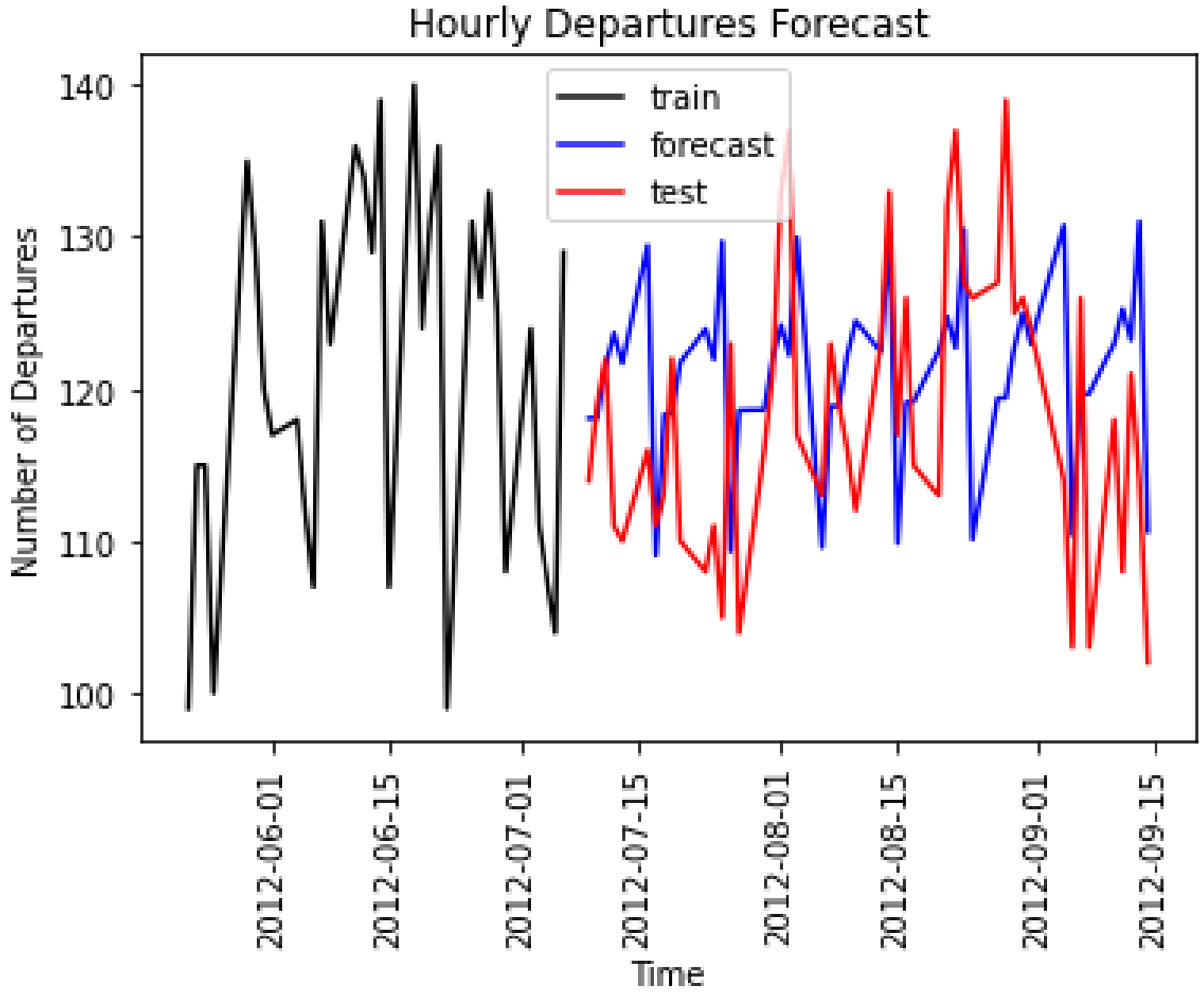
# Models used

- As in the case of Linear Regression approach, The first 80% data (192 days) is used to train the time series model.
- The remaining 20% data is used to test the accuracy of the model
- Mean baseline, Naïve Baseline, seasonal naïve baseline, Prophet and ARIMA method have been used for modeling
- The Sum of squared errors and mean squared errors is used for comparison

# Baseline Methods



# Prophet Model



# Comparison of Models

Sum of squared error

mean_baseline_error: 4492.6 (best performance)

Naive_baseline_error: 9590

seasonal_baseline_error: 9480

Prophet_error: 5290.6

autoarima_error: 4508.8

# Comparison of Models (contd.)

Mean of squared error

mean_baseline_error: 91.6 (best performance)

Naive_baseline_error: 195.7

seasonal_baseline_error: 193.4

Prophet_error: 107.9

autoarima_error: 92

# Conclusion /Recommendation

The linear regression model performs significantly better than the time series model (Linear regression MSE for (T-5) days = 42.1 vs Time Series Model best MSE = 91.6 )

Thus, we can conclude that the number of surgeries on schedule at "T-$x$ " should be used to predict the number of surgeries on the current day where $x$ is the number of days in advance we need the forecast.

*T*o find the value of $"x"$ we need further information such as the resources available, lead time to obtain resources, cost of overstocking/understocking (which is currently missing)