Background

Many factors contribute to the needs of an innovative system to increase efficiency in health care, such as investment in the latest advancement.

- *Vanderbilt University medical center rooms a variety of surgeries*
- *Elective surgeries are scheduled based on the urgency of the patient's needs and their own scheduling preferences*
- *Although surgical staff schedules are made weeks in advance, final number of surgeries are known only the day before*
- *AJ Bose supervised the task of improving prediction of the daily surgical case volume*

Problem Statement

If elective surgery scheduled made a week prior could be used to predict the final number of surgery performed?

- *A model should be developed based on the developing elective surgery schedule to predict daily demand.*
- *Current system receive its schedule only 1 day in advance*
  - ❖ *As an example, The elective surgery schedule was not finalized until 5 p.m. the day before. At the end of each day, the charge nurse reported the schedule for the next day*
  - ❖ *6% of elective surgeries receive their schedule only 15 hours in advance*
- *It is preferred to doing the surgery early in the week / early in the day*

Therefore, the need is to make a forecasting on schedules, so it enables the medical center to get prepared as much as possible and increase efficiency

Vanderbilt Elective Surgery Scheduling Dataset

- ✓ *A large dataset providing information of 241 consecutive surgeries*
- ✓ *Gives us information of around 11 months in 2011 and 2012*
- ✓ *All surgeries are done during weekdays*
- ✓ *Last columns represents the total actual done surgeries in its corresponding day*
- ✓ *The rest of the columns provide number of scheduled surgeries which accumulate to its next day*

SurgDate	DOW	T - 28	T - 21	T - 14	T - 13	T - 12	T - 11	T - 10	T - 9	T - 8	T - 7	T - 6	T - 5	T - 4	T - 3	T - 2	T - 1	Actual
2011-10-10	Mon	38	45	60	63	65	70	73	73	73	80	84	89	94	98	100	104	106
2011-10-11	Tue	35	47	65	68	78	82	82	82	86	89	92	95	99	99	99	114	121
2011-10-12	Wed	26	43	54	62	72	72	72	74	87	94	96	101	102	102	106	114	126
2011-10-13	Thu	28	48	65	70	72	72	72	82	87	91	94	94	94	97	98	103	114
2011-10-14	Fri	31	40	50	50	50	54	62	68	71	73	73	73	78	83	87	94	106
2011-10-17	Mon	41	56	65	69	72	73	77	78	78	80	86	85	86	92	96	102	111
2011-10-18	Tue	44	55	69	74	79	83	83	83	93	92	96	103	105	105	107	114	122
2011-10-19	Wed	32	40	62	66	71	73	73	84	86	87	89	96	96	96	102	119	127
2011-10-20	Thu	33	44	62	66	67	67	79	77	88	90	98	98	98	105	111	118	116
2011-10-21	Fri	20	32	48	48	48	47	52	55	59	61	61	61	69	72	70	88	99

Analysis Part Agenda

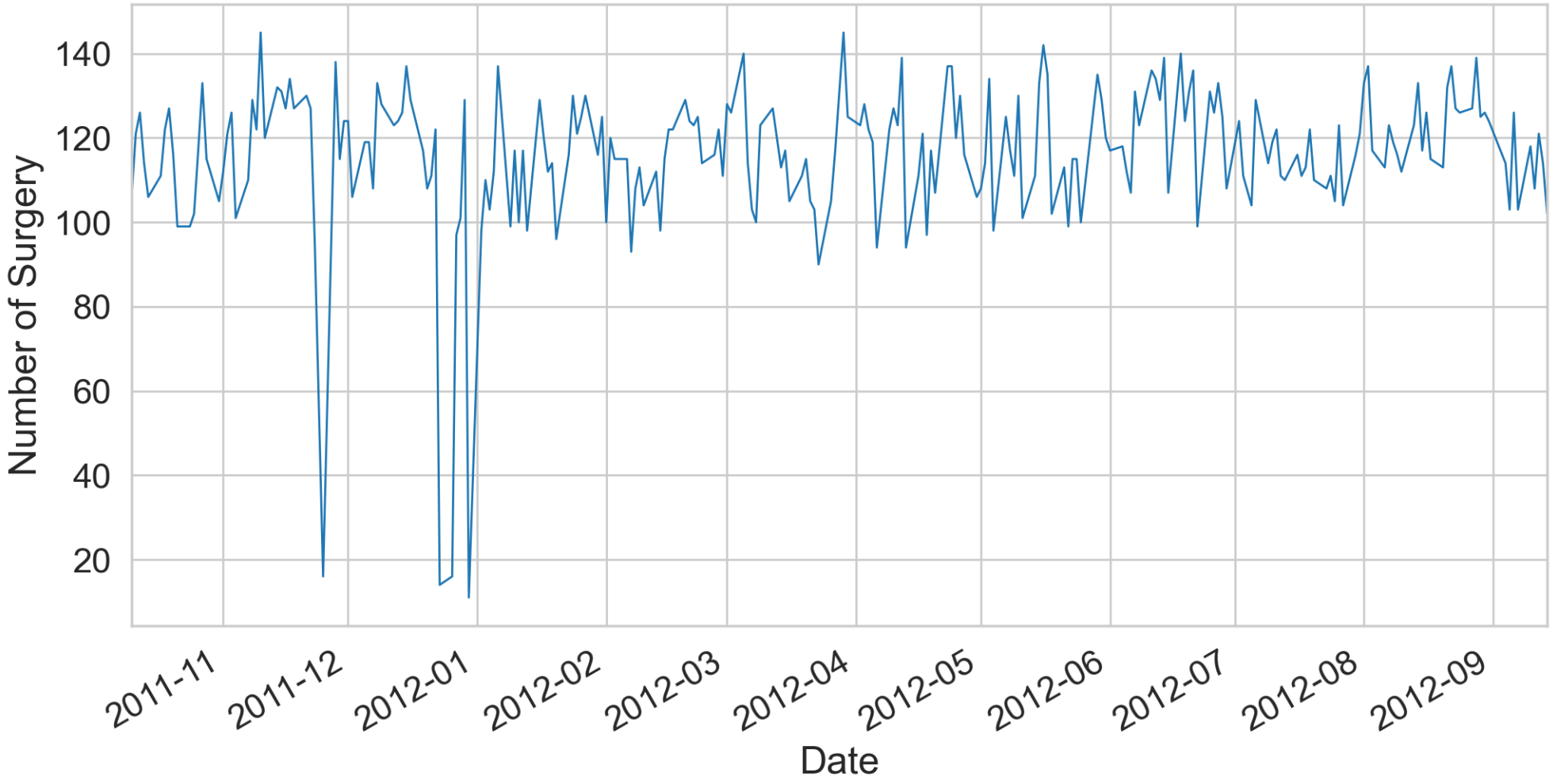
- ✓ ***Initial Analysis***
- ✓ ***Regression Analysis***
- ✓ ***Data Cleaning***
- ✓ ***Time Series Analysis and Forecasting***
- ✓ ***Forecasting with Prophet***
- ✓ ***Forecasting with ARIMA Model***

Initial Analysis

The following pages provides some graphs which were generated using Google Colab along with a brief explanation.

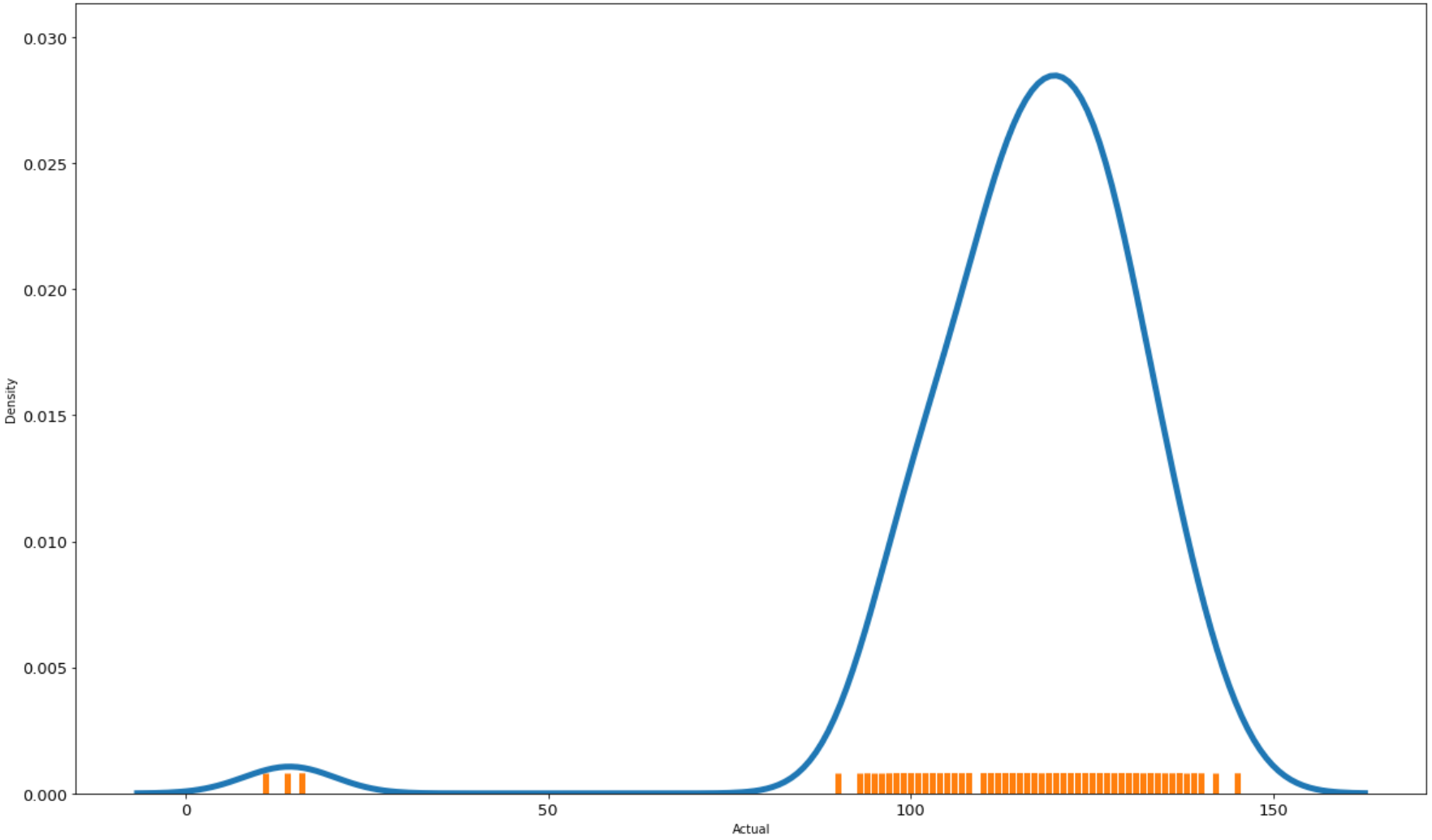
Initial Analysis

This plot shows a broad overview of how actual number of surgeries took place during the whole time interval. Generally, the number of surgeries ware around 100 and 140.

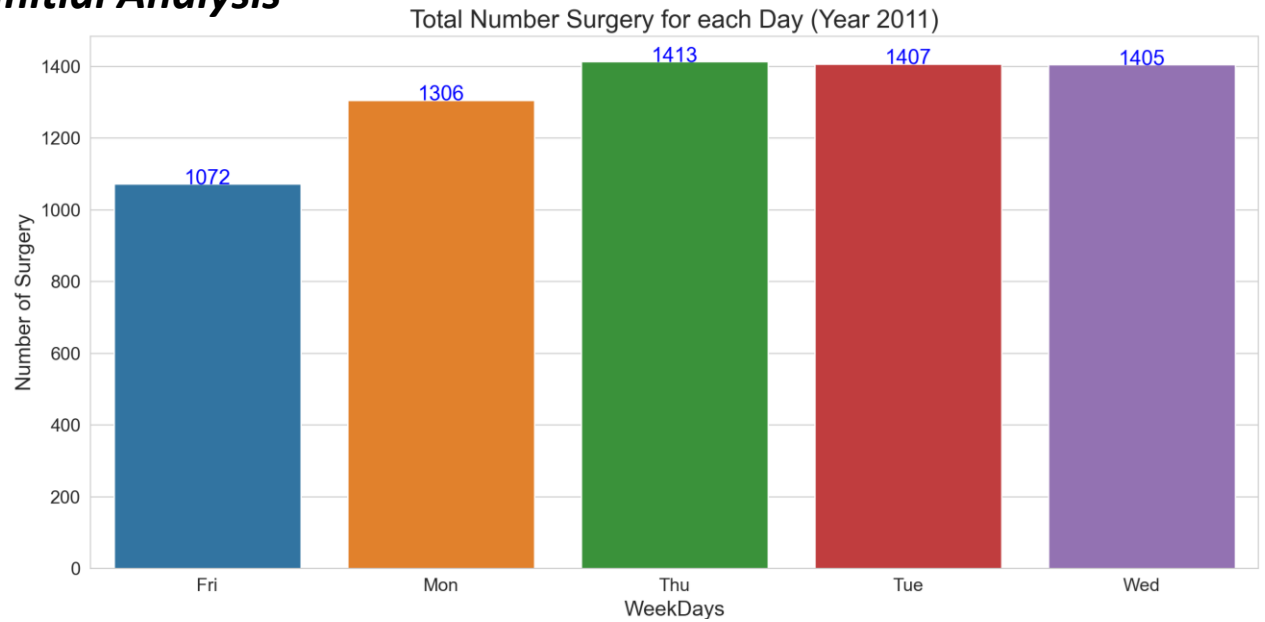


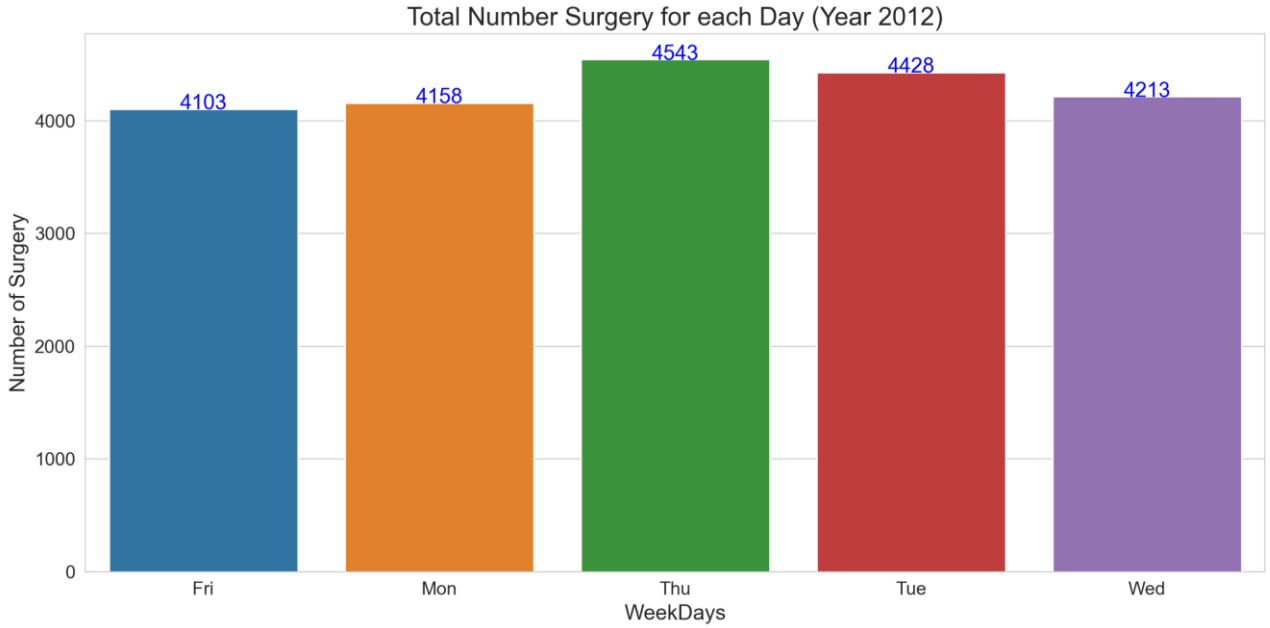
Initial Analysis

The number of surgeries are normally distributed. The average is around 120



Initial Analysis



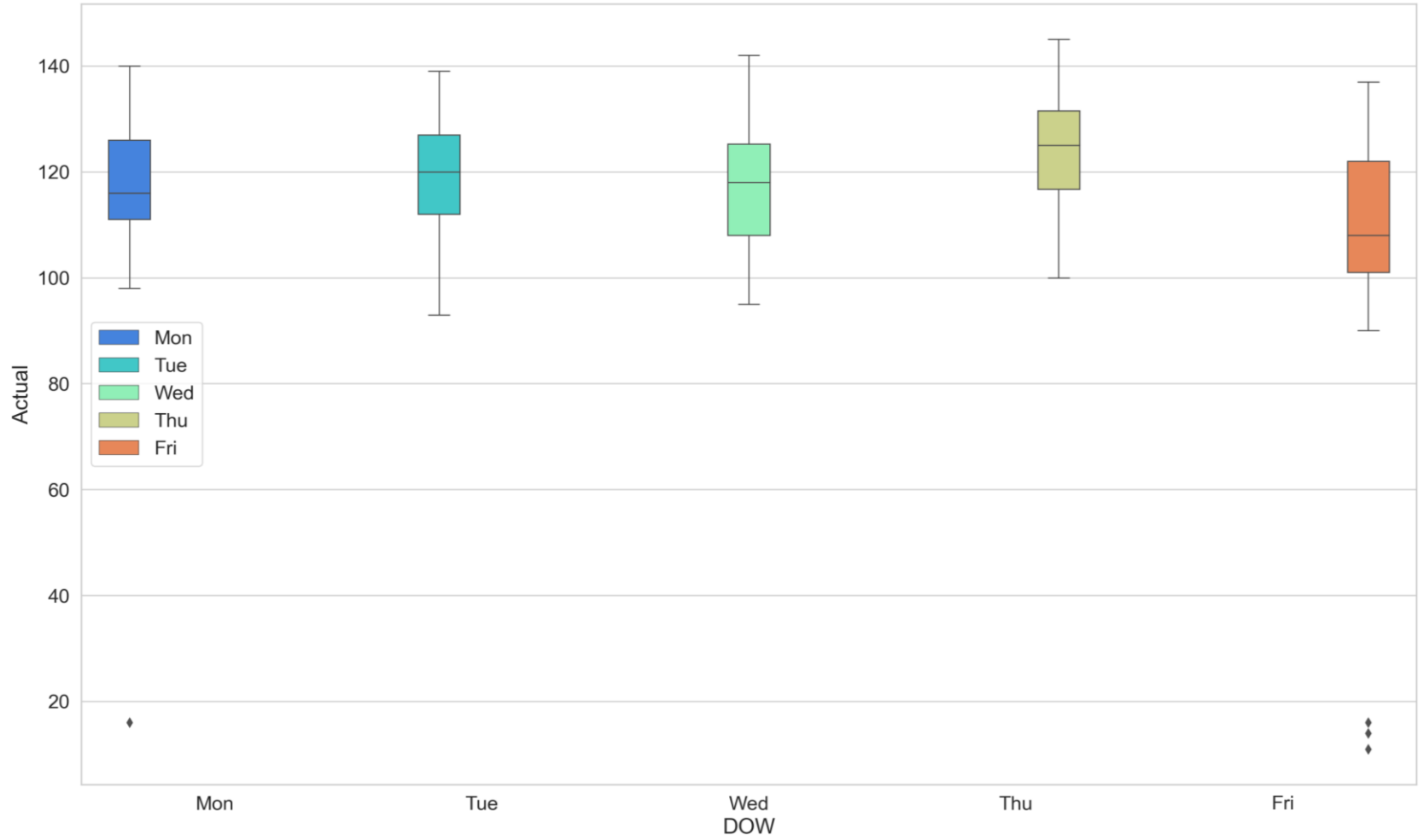


Total number of surgeries to better visualize is divided into weekdays of years 2011 and 2012.

More number of surgeries were done on Thursdays for both of the years and Fridays has the least.

Initial Analysis

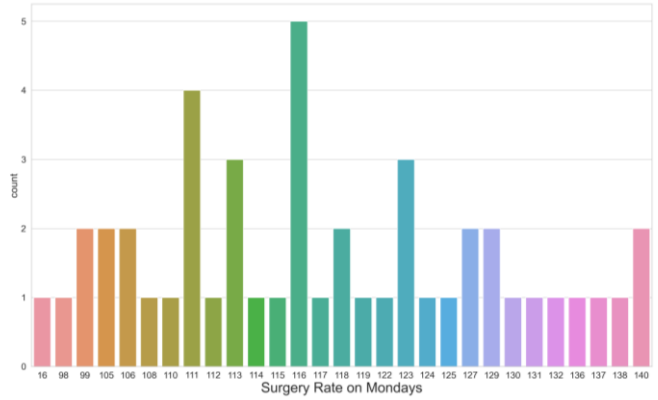
Days of the weeks and what concluded in the previous slide is evident by employing Boxplot



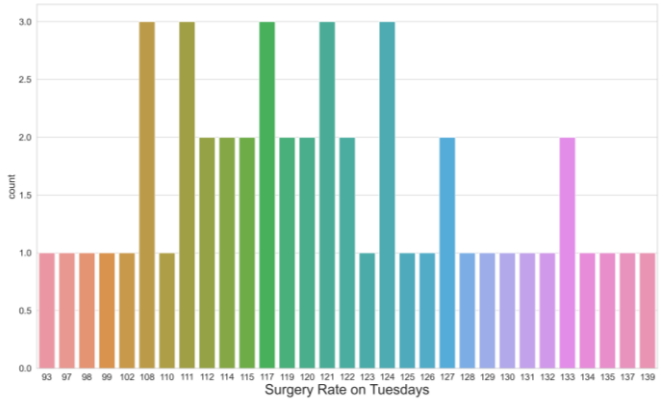
Initial Analysis

How the number of surgeries is distributed for different days of week during the whole period

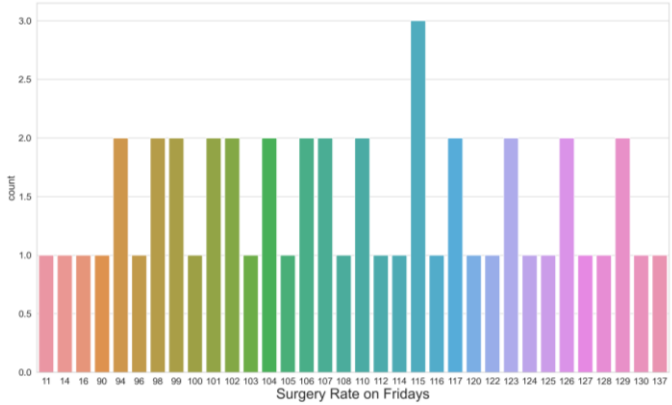
Monday



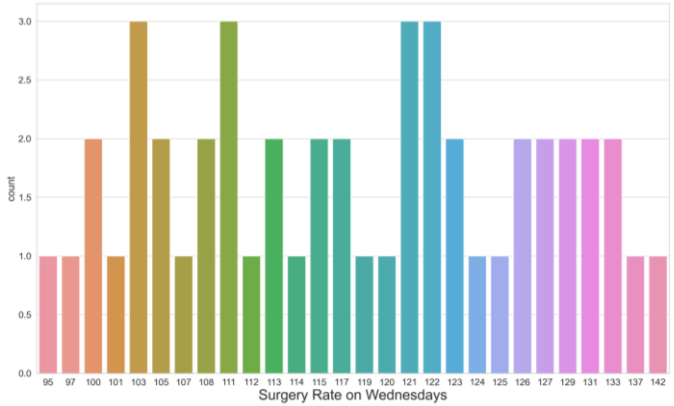
Tuesday



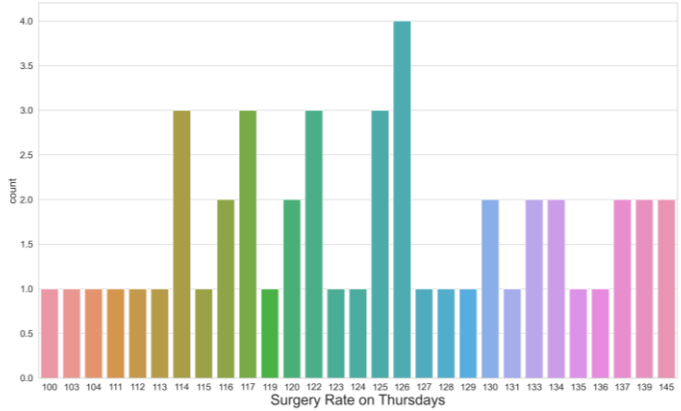
Friday



Wednesday



Thursday



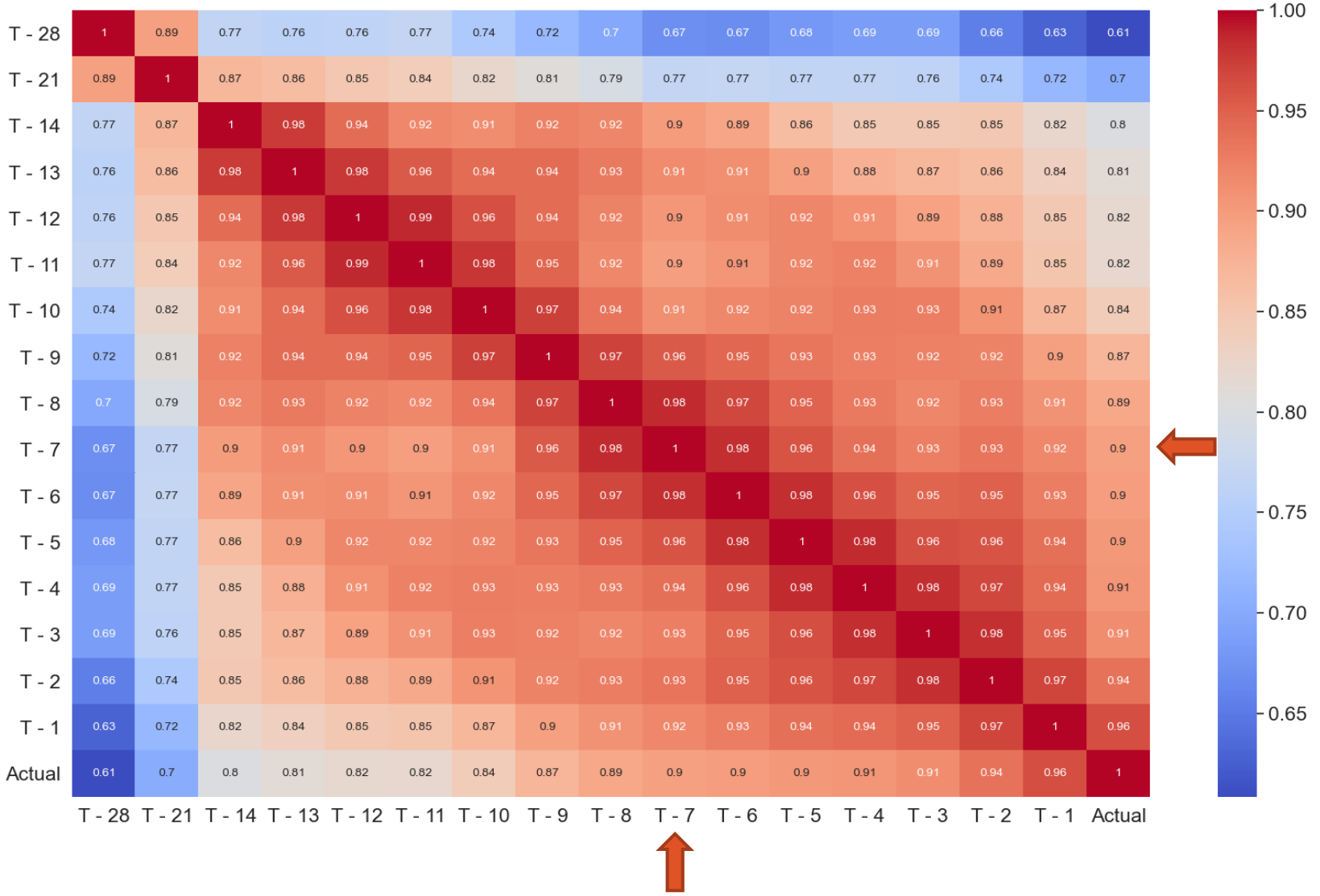
Regression Analysis

One important question needs to be answered is ***" How many days before a surgery should be scheduled"*** *.*

I tried different options among (T-28 to T-1). Based on the two different approach in the following pages, 7 days in advance would be preferable to schedule surgeries.

Regression Analysis

Correlation Heatmap, clearly shows the correlation of each day with others days and the Actual column



First Approach

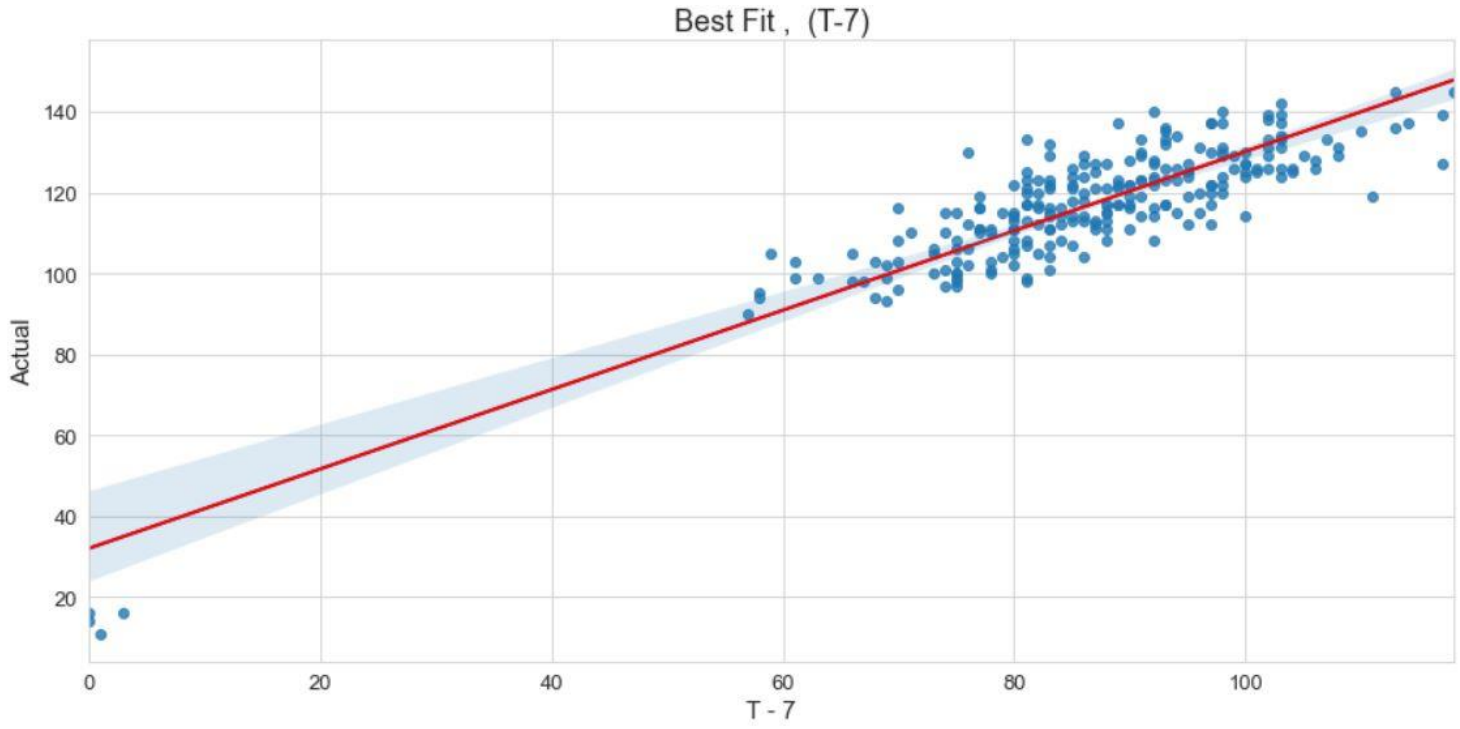
Day 7 is among the highly correlated days with the Actual column by 0.895779

Regression Analysis

By employing regression analysis among all days, T-7 day's

R-square explains good accuracy by 0.80241

Second Approach

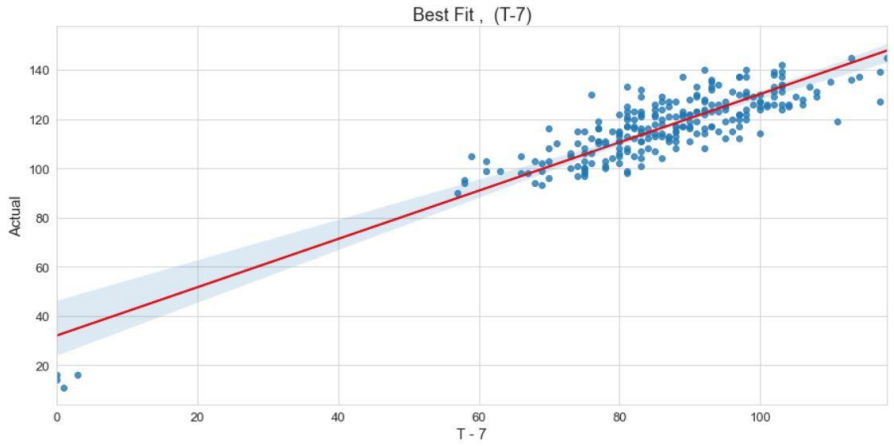


R square (T - 7) vs Actual for Surgery_data_Y is equal to 0.8024195012892146

Regression Analysis(Additional Details)

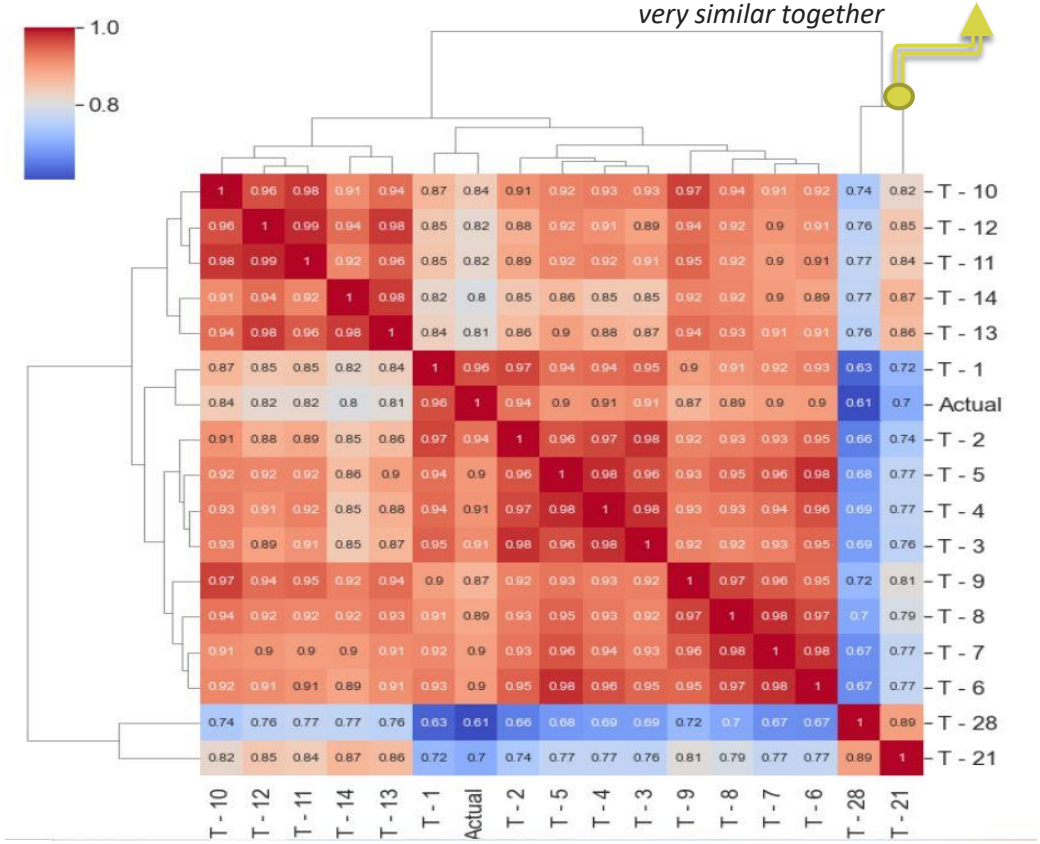
If we want to schedule a surgery on day T-7, the most accurate Actual surgery day will belong to Fridays by R-Square equal to 0.915

```
R square (T - 7) vs Actual for    Monday is equal to  0.8249855171010343
R square (T -7) vs Actual for   Tuesday is equal to  0.5374665148734139
R square (T - 7) vs Actual for Wednesday is equal to  0.6684289424818737
R square (T - 7) vs Actual for  Thursday is equal to  0.6154385108447944
R square (T - 7) vs Actual for    Friday is equal to  0.9158867220957229
Best fit for (T - 7) is equal to 0.9158867220957229
```

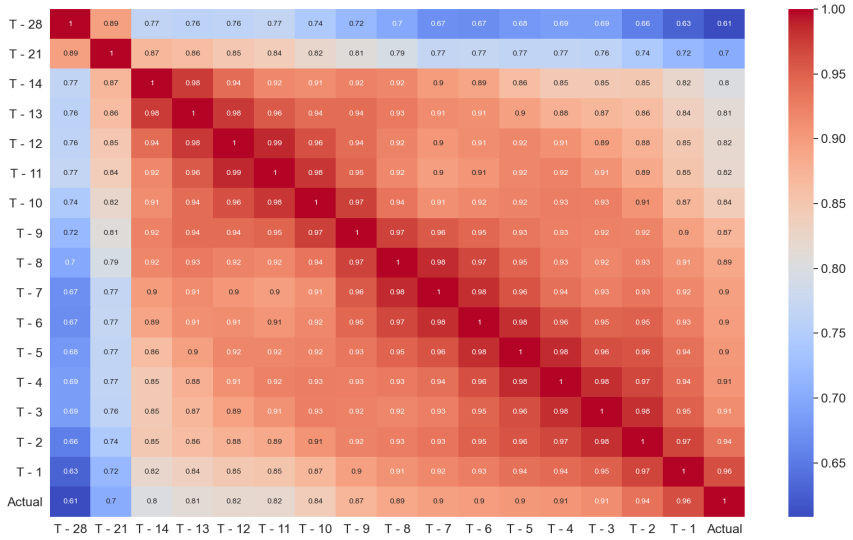


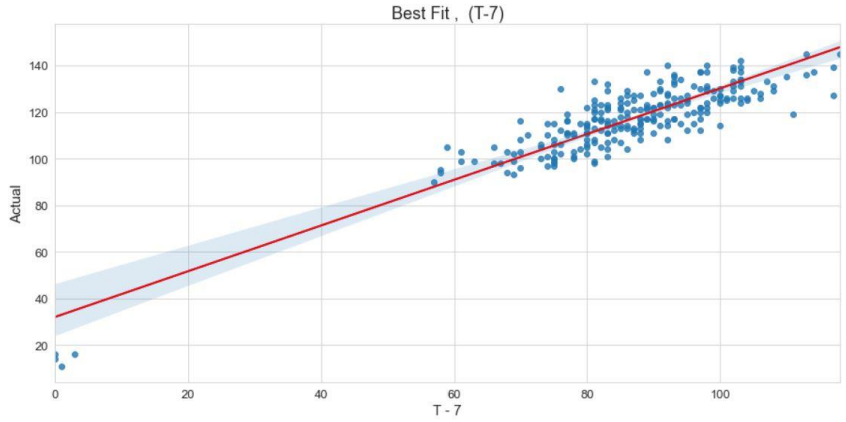
Clustering Correlation Heatmap

For example : T-21 and T-28 are very similar together



Regression Analysis (Conclusion)





Therefore, by a balance of the "highest possible accuracy" and "the soonest possible date to schedule" I choose 7 days prior to surgery date to make the schedule

Time Series Analysis and Forecasting

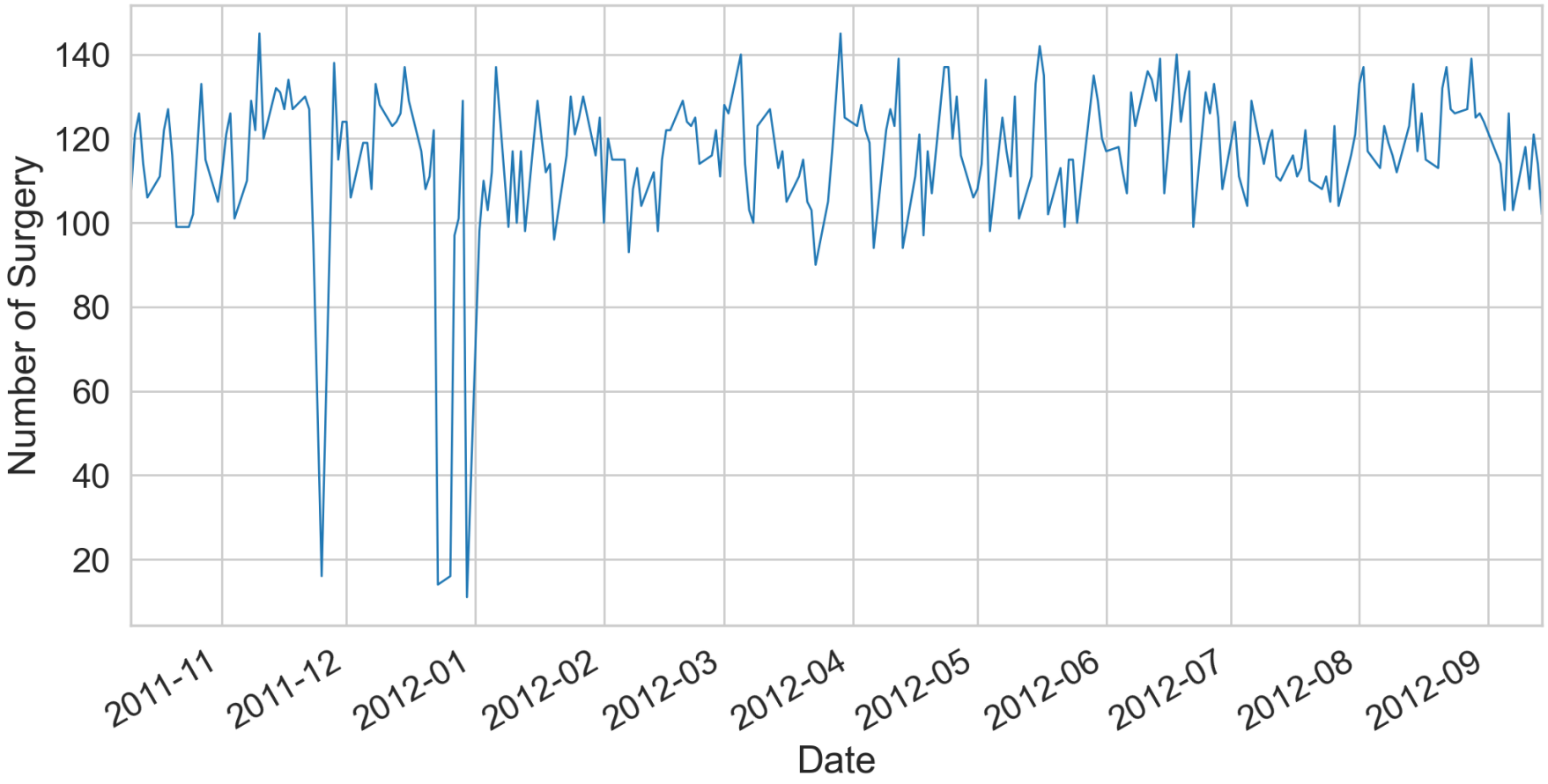
- *Time series Components*
- *Predictive Model*
- *Autocorrelation*
- *Baseline Method*
- *Histogram of Residuals*
- *Forecasting by Prophet*
- *Forecasting by ARIMA Model*

Time Series Analysis and Forecasting

- *Cleaning the Dataset*
- *Baseline Method*
- *Forecasting by Prophet*
- *Forecasting by ARIMA Model*

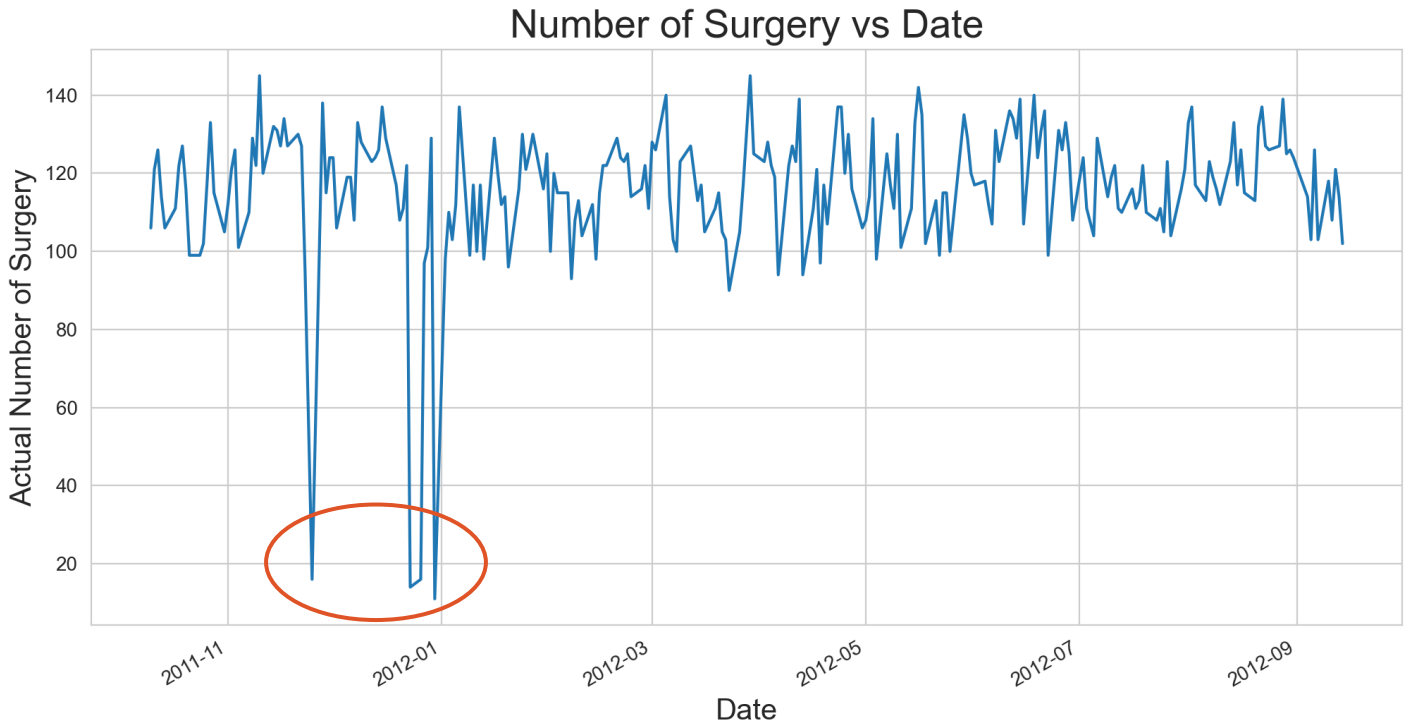
Time series Components

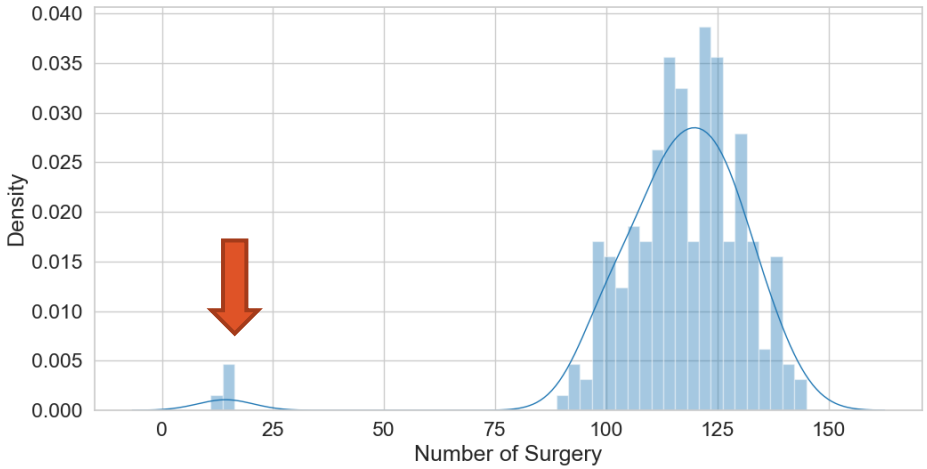
- *Random process with random fluctuation and general constant variance*
- *Number of surgeries fluctuated during the whole period of study*



Initial Dataset

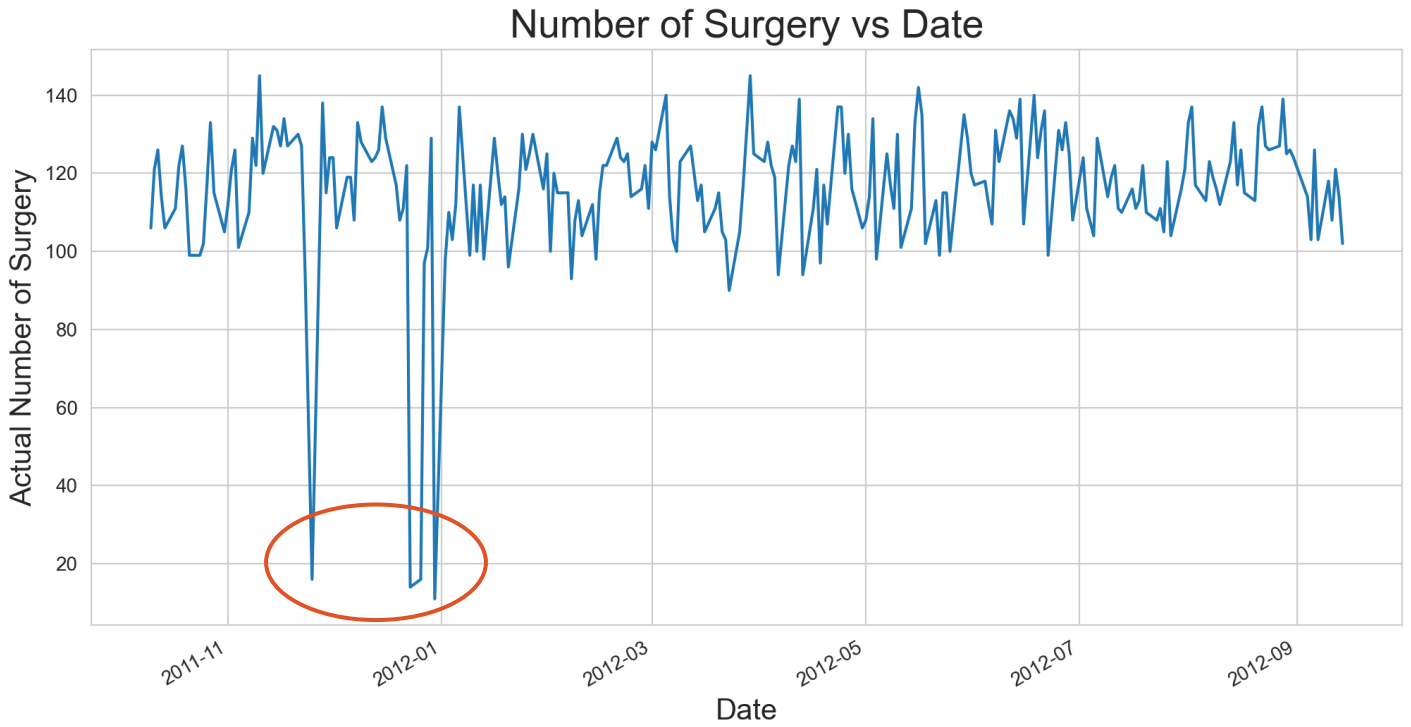
*As it is pointed, there are 4 points called **Outliers** which are far away from the average.*

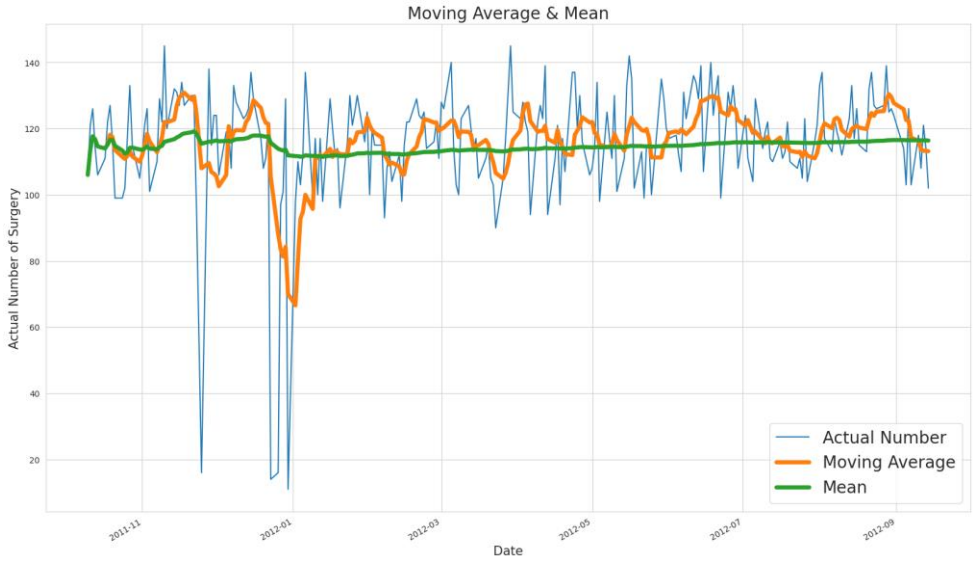




Initial Dataset

*As it is pointed, there are 4 points called **Outliers** which are far away from the average.*



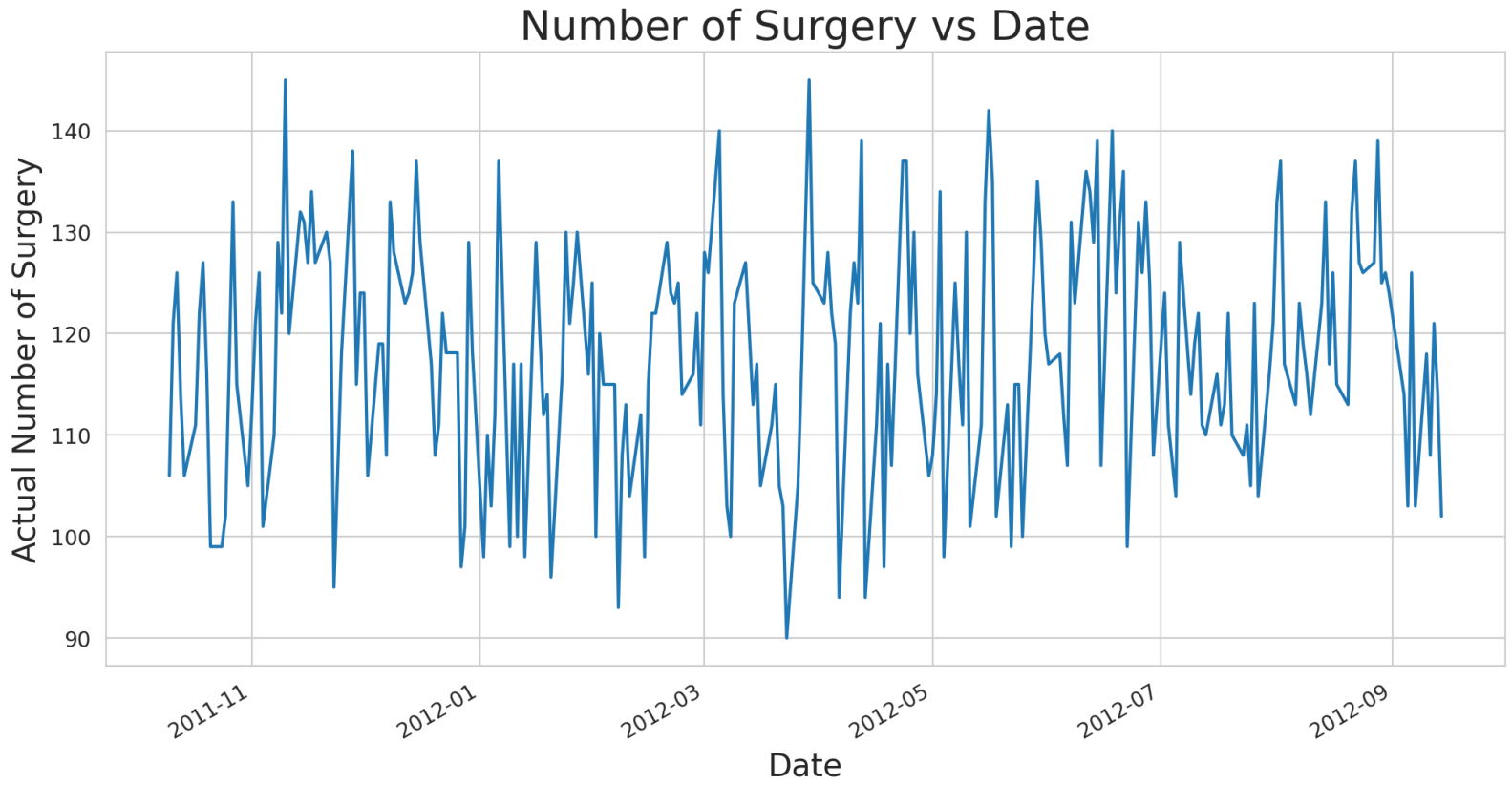


*Therefore, to better analyze the time series, **Outliers** should be ignored or replace with a value to level off our data set.*

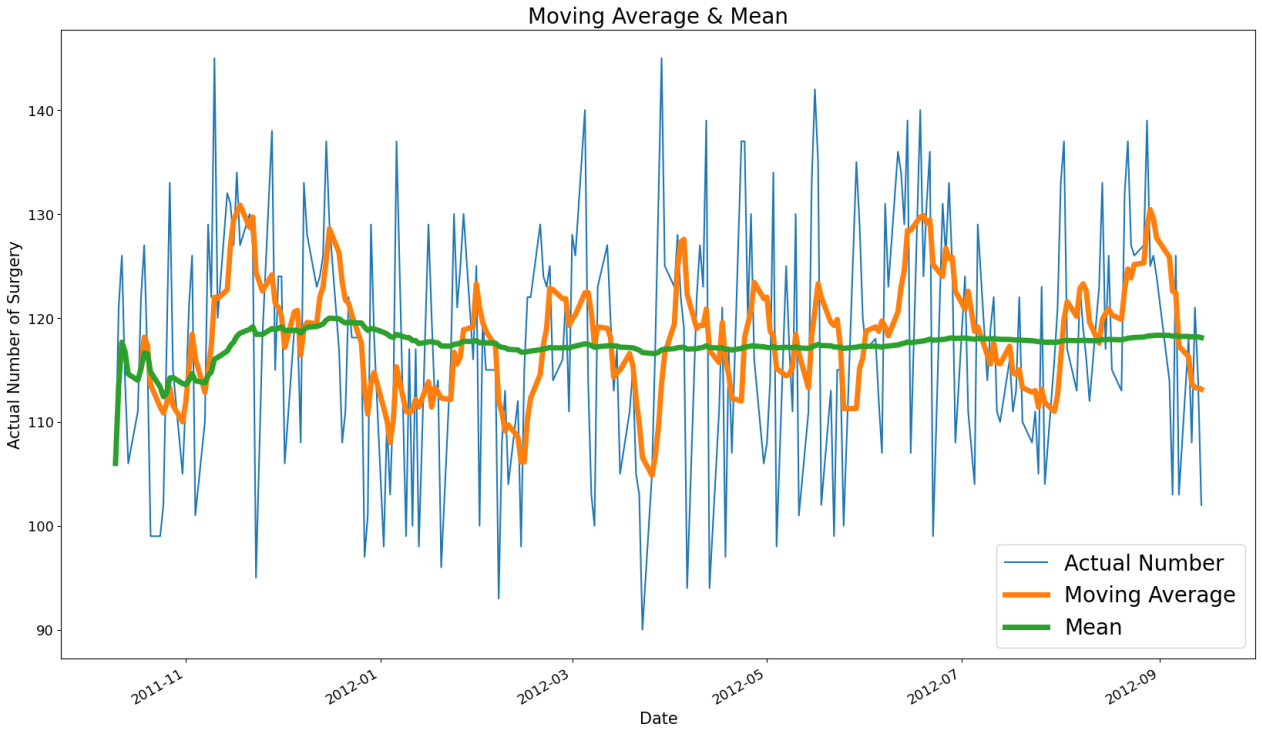
Cleaned Dataset

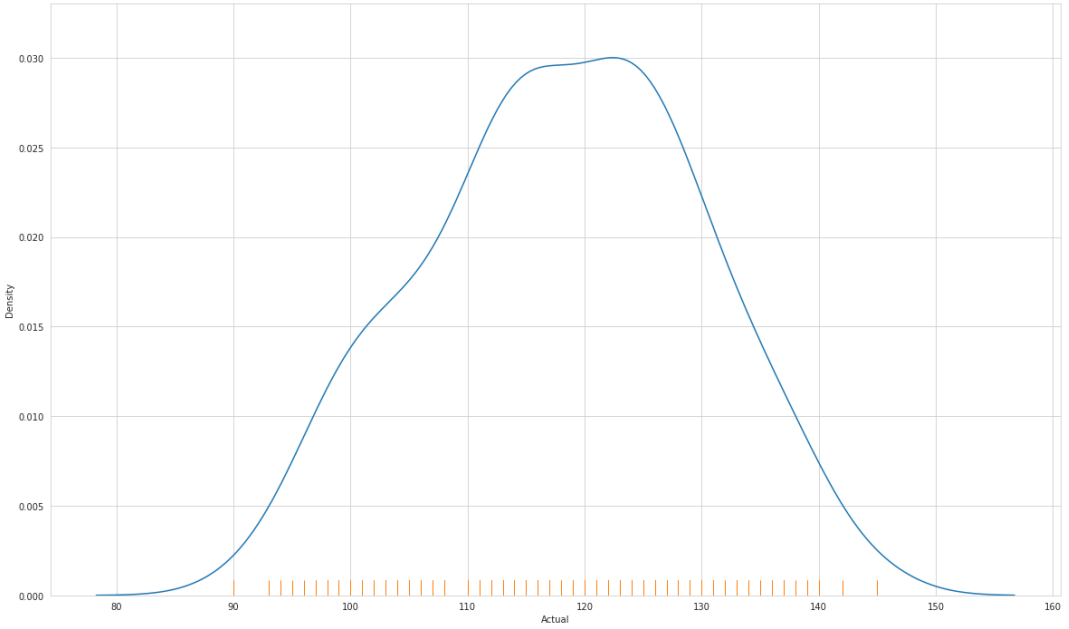
Outlier points were substituted by the whole mean of other points in 'Actual' Column.

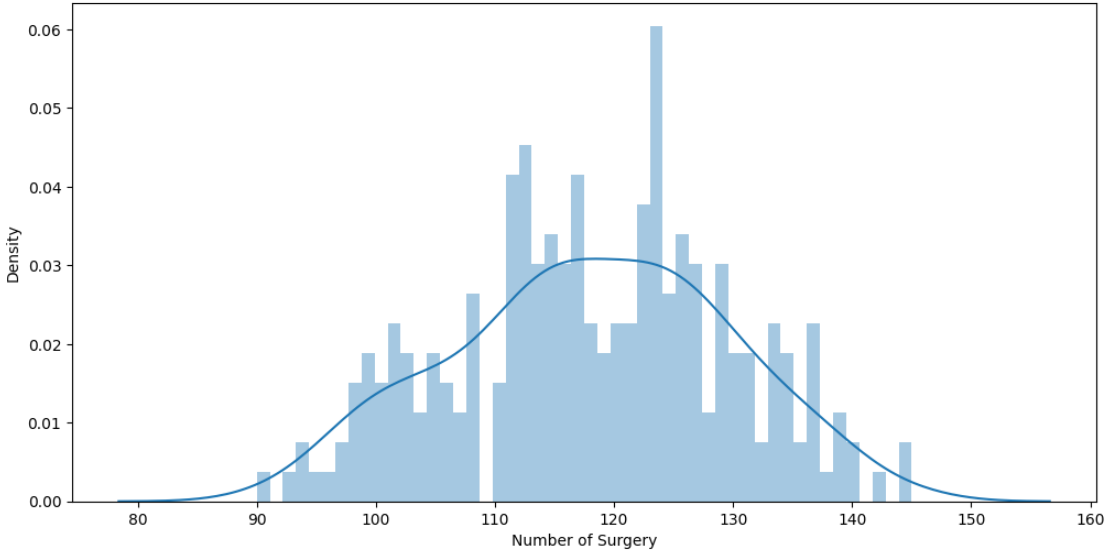
- *Generally no trend is seen (Follows a Stationary Pattern)*
- *Irregular Cyclic pattern*
- *No significant seasonality*
- *Number of surgeries experienced a significant fluctuation as of late December 2011 to late January 2012*



Cleaned Dataset

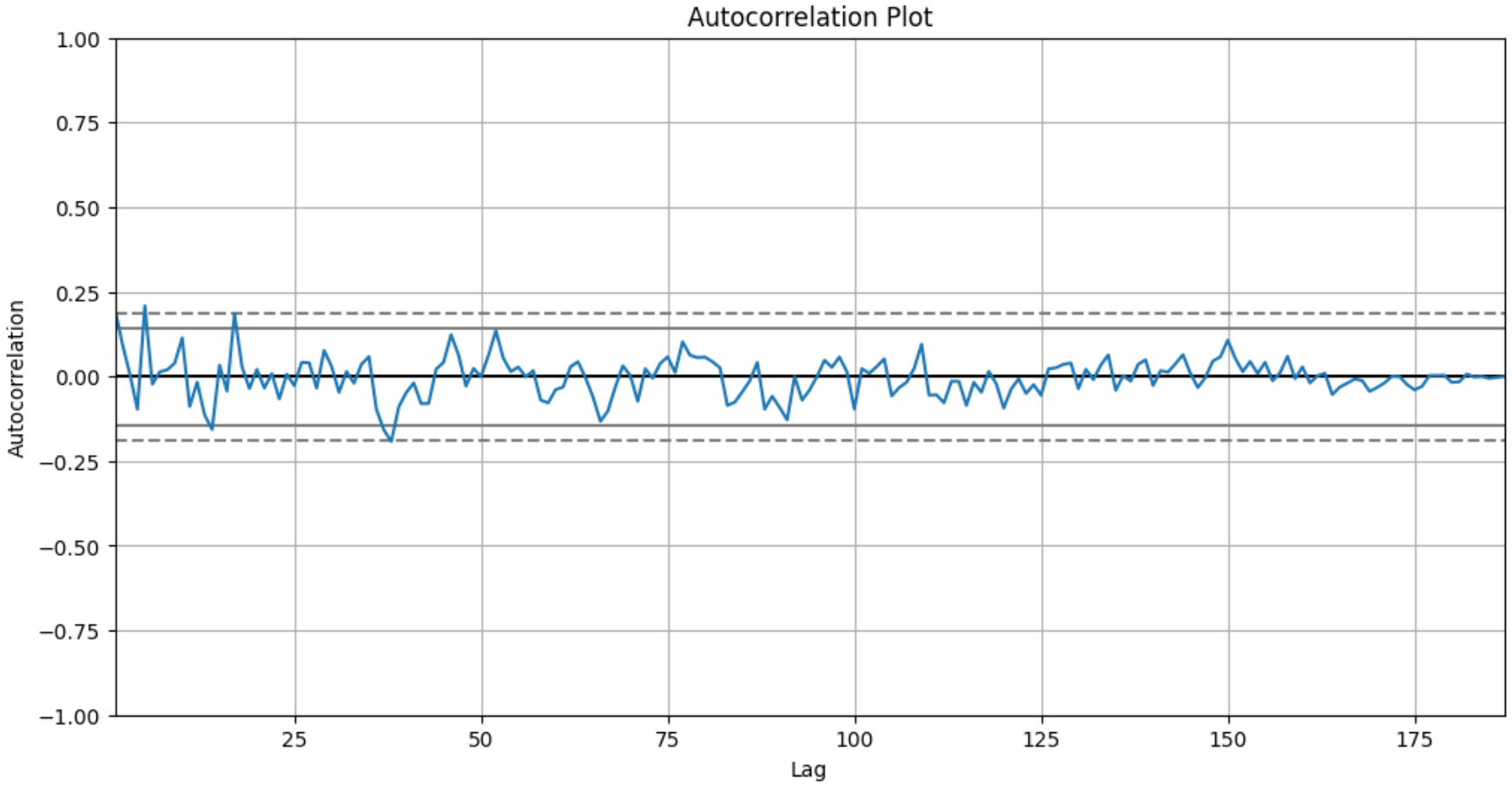






Autocorrelation

- *There is no Autocorrelation for any lag* ⟹ *White Noise (more than 95% spike inside)*
- *There is no significant relationship between an observation with its past*
- *No Trend*
- *No Seasonality*



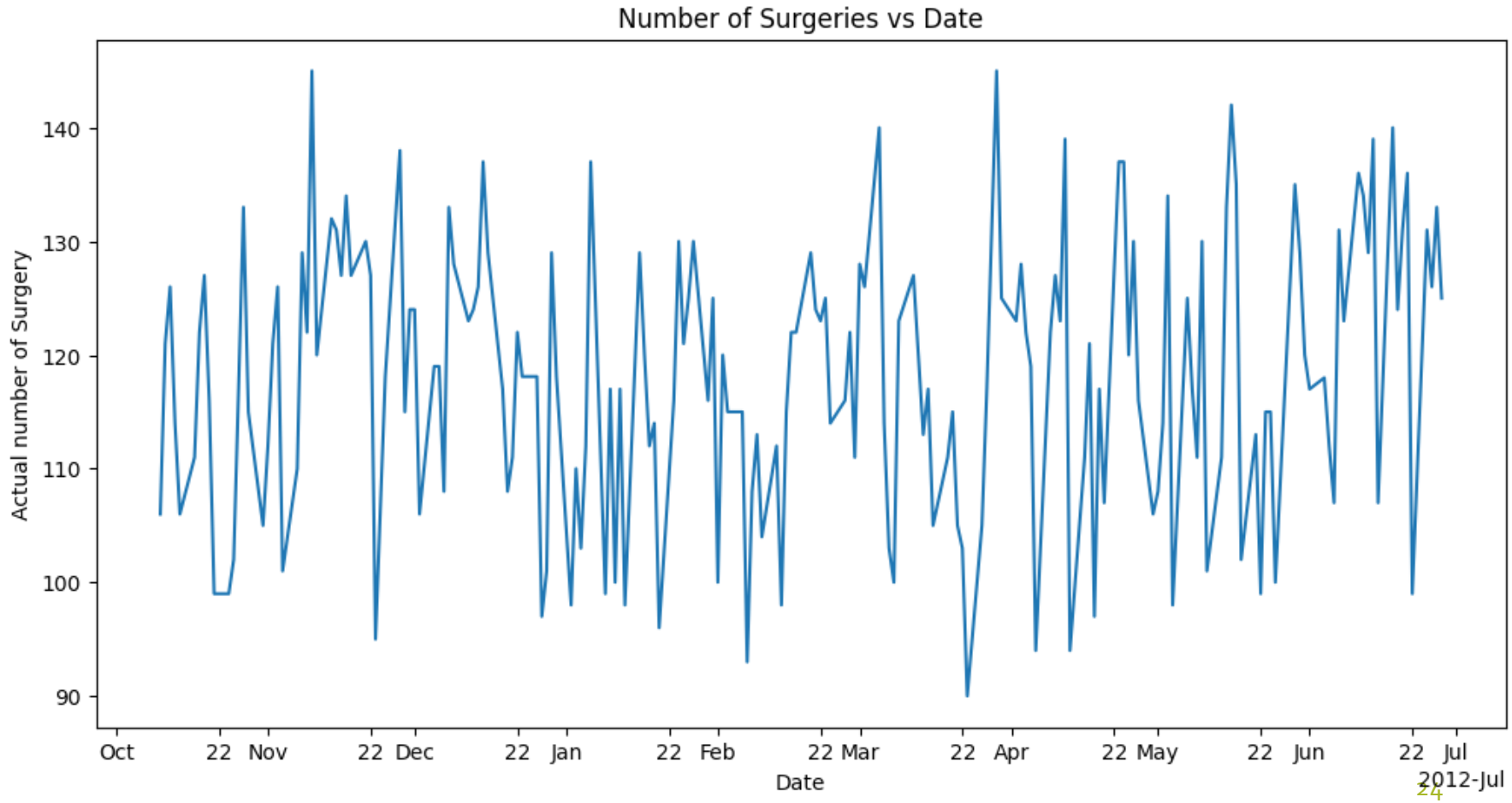
Baseline Method *(Predictive Model)*

```
Surgery_data_test = Surgery_cleaned_data[(Surgery_cleaned_data['ds'] >= dt.datetime(2012,6,30)) & (Surgery_cleaned_data['ds'] <= dt.datetime(2012,9,14))]
## Fill in the missing components in the following line:
Surgery_data_train = Surgery_cleaned_data[(Surgery_cleaned_data['ds'] < dt.datetime(2012,6,29)) & (Surgery_cleaned_data['ds'] >= dt.datetime(2011,10,10))]
Surgery_data_test = Surgery_data_test.reset_index(drop = True)
Surgery_data_train = Surgery_data_train.reset_index(drop = True)
```

Splitting the Data into Train and Test Sets

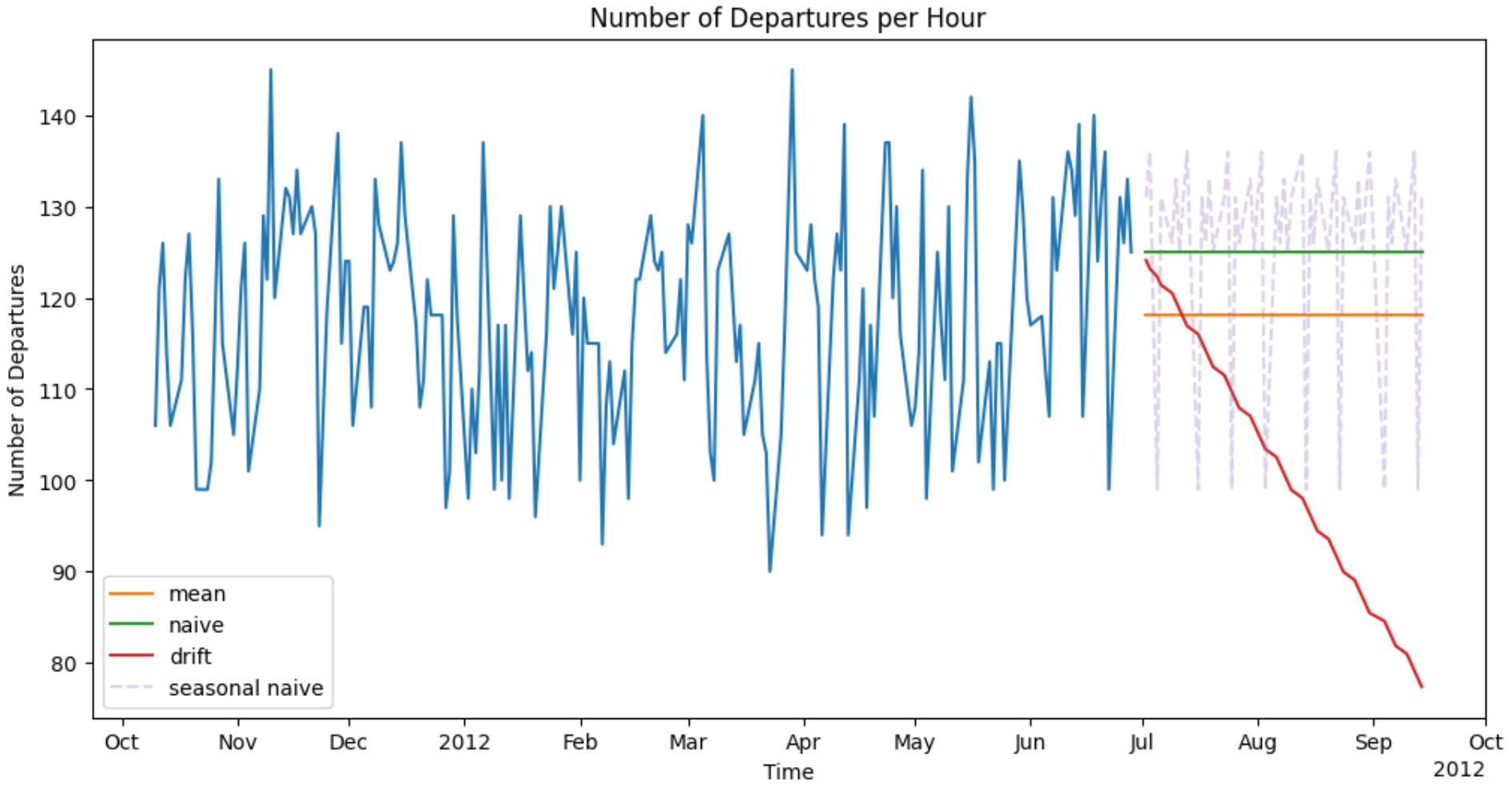
Training data from 2011/10/10 to 2012/06/29

Test set from 2012/06/30 to 2012/09/14 ,



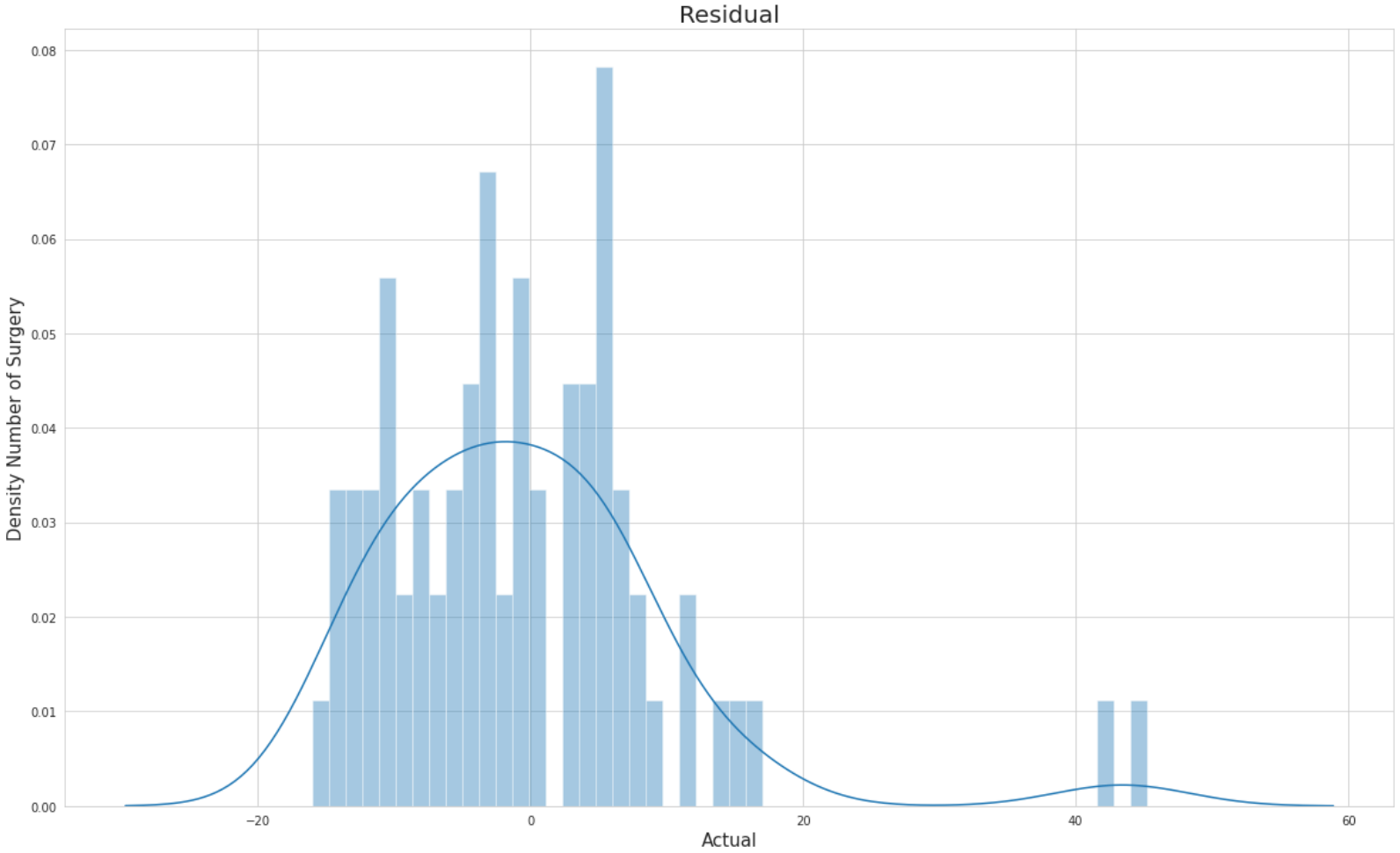
Baseline Method *(Plotting)*

yT = 125 *h_max=53* *Y1=105* *T =187*



Baseline Method *(Histogram of Residuals)*

- *Looks normal with a mean equal to zero which is good but correlation between residuals should be considered to check if the model needs improvement.*



Forecasting by Prophet

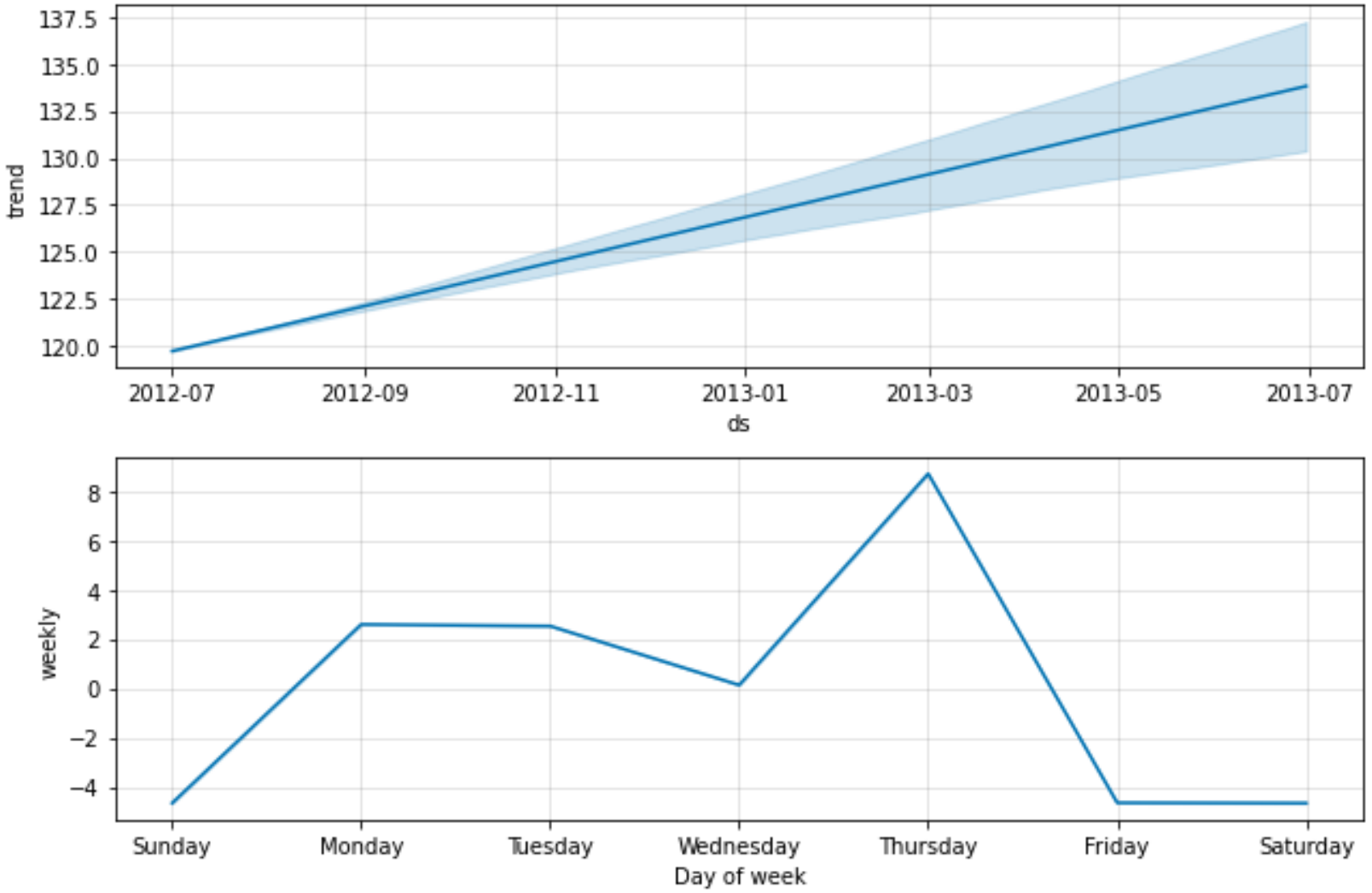
Forecast

Uncertainty Interval

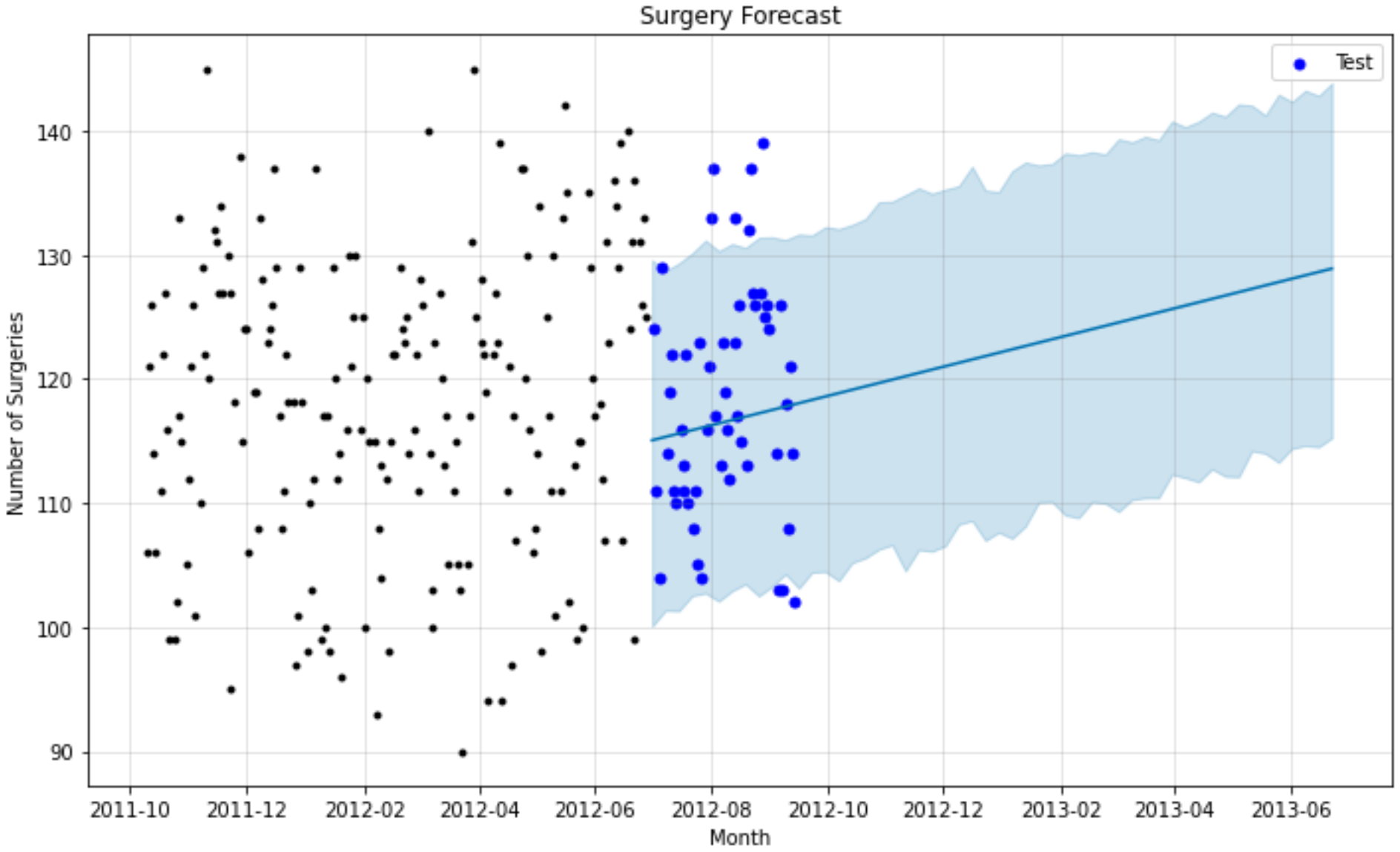
Frequency= Weekly

ds	yhat	yhat_lower	yhat_upper
2012-09-16	118.042856	103.689561	132.928726
2012-09-23	118.314128	102.875840	132.540608
2012-09-30	118.585400	103.883650	132.948909
2012-10-07	118.856673	104.480982	133.128134
2012-10-14	119.127945	105.974970	133.065786
2012-10-21	119.399217	105.167350	133.721711
2012-10-28	119.670489	105.355872	132.549623
2012-11-04	119.941761	106.397262	134.288450
2012-11-11	120.213033	105.960359	134.029288
2012-11-18	120.484305	105.641389	134.457398
2012-11-25	120.755577	107.557735	135.343080
2012-12-02	121.026849	107.186875	135.255589
2012-12-09	121.298121	107.550520	136.198361
2012-12-16	121.569393	108.601446	136.611378
2012-12-23	121.840665	107.131638	136.088603
2012-12-30	122.111937	108.149870	136.625044

Visualize the underlying forecast components



Plotting the Forecast



Computing the Error from each of the Predictions

Loss function metrics should be minimized to create the best model

Mean Squared Error

```
Mean baseline MSE: 86.15229539445873
Naive baseline MSE: 130.18867924528303
Naive seasonal baseline MSE : 293.5660377358491
Drift baseline MSE: 619.0044227852486
Prophet MSE: 107.19137271246215
```

Mean Absolute Error

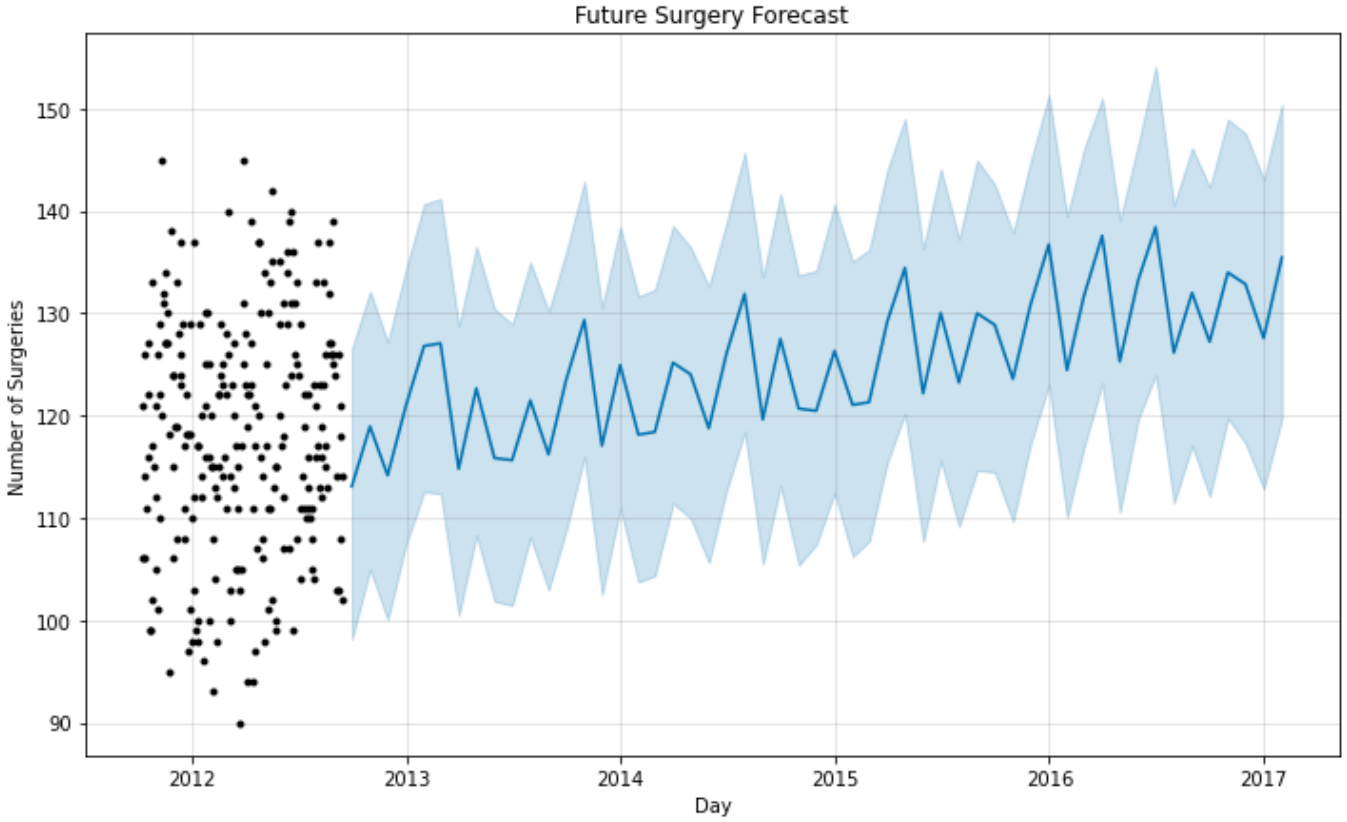
```
Mean baseline MAE: 7.72195195467509
Naive baseline MAE: 9.39622641509434
Naive seasonal baseline MAE : 14.39622641509434
Drift baseline MAE: 20.30995863182323
Prophet MAE: 8.11657790321257
```

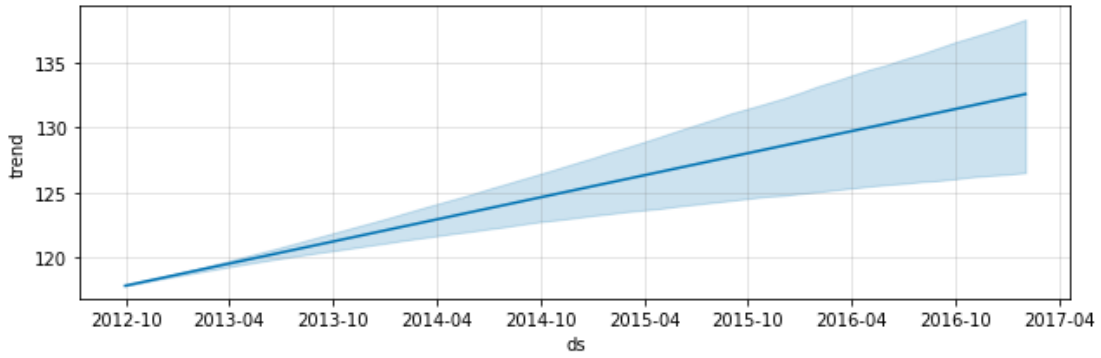
Mean Absolute Percentage Error

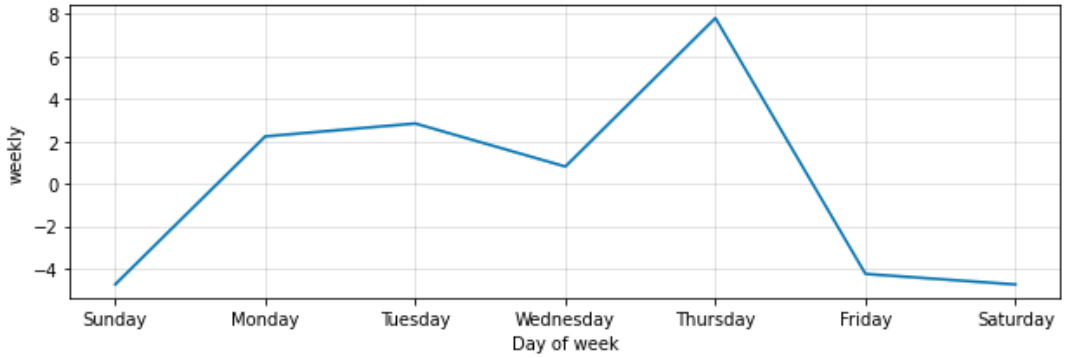
```
Mean baseline MAPE: 0.0653916051141056
Naive baseline MAPE: 0.07516981132075473
Naive seasonal baseline MAPE : 0.12243418409019506 Drift
baseline MAPE: 0.2214703246937 1752
Prophet MAPE: 0.0660460788965 2046
```

Out of Sample Prediction

- *The entire dataset is used to make a prediction*
- *Future monthly prediction pattern is shown, number of surgeries will see an increasing trend*
- *Still Thursdays will have the most Surgeries*







Forecasting by ARIMA Model

SARIMAX Results

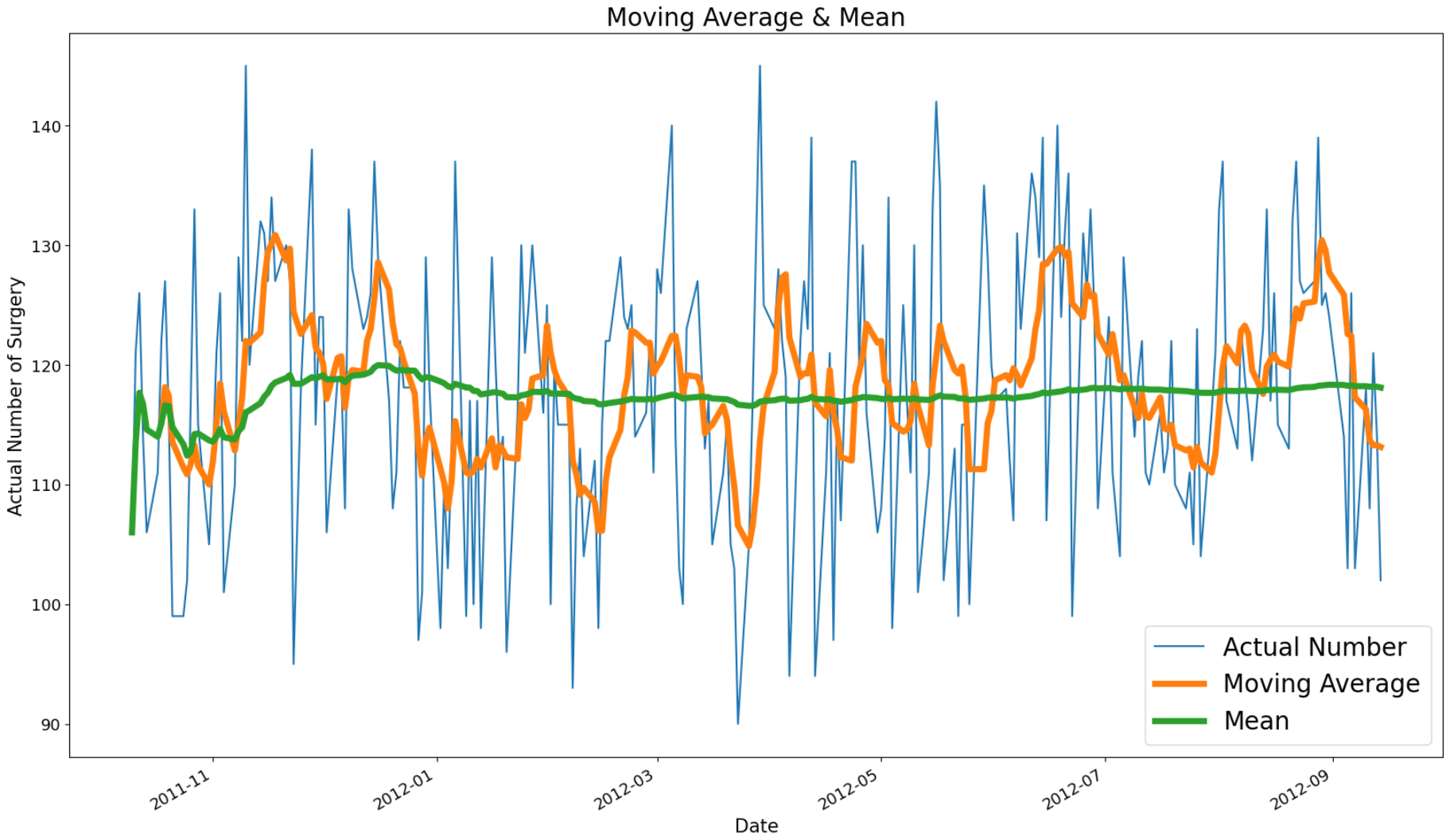
Dep. Variable:	y	No. Observations:	187
Model:	SARIMAX(2, 0, 2)x(1, 0, [1], 53)	Log Likelihood	-721.359
Date:	Sat, 30 Oct 2021	AIC	1458.718
Time:	20:32:41	BIC	1484.567
Sample:	0	HQIC	1469.192
	- 187		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	94.4621	381.434	0.248	0.804	-653.134	842.058
ar.L1	0.0449	0.082	0.546	0.585	-0.116	0.206
ar.L2	-0.7656	0.075	-10.171	0.000	-0.913	-0.618
ma.L1	0.1218	0.054	2.243	0.025	0.015	0.228
ma.L2	0.9199	0.050	18.317	0.000	0.821	1.018
ar.S.L53	0.5350	1.877	0.285	0.776	-3.145	4.215
ma.S.L53	-0.4774	1.917	-0.249	0.803	-4.236	3.281
sigma2	128.6461	16.148	7.967	0.000	96.996	160.296

Ljung-Box (L1) (Q):	0.12	Jarque-Bera (JB):	5.30
Prob(Q):	0.73	Prob(JB):	0.07
Heteroskedasticity (H):	1.42	Skew:	-0.32
Prob(H) (two-sided):	0.17	Kurtosis:	2.48

Check if our data is stationary

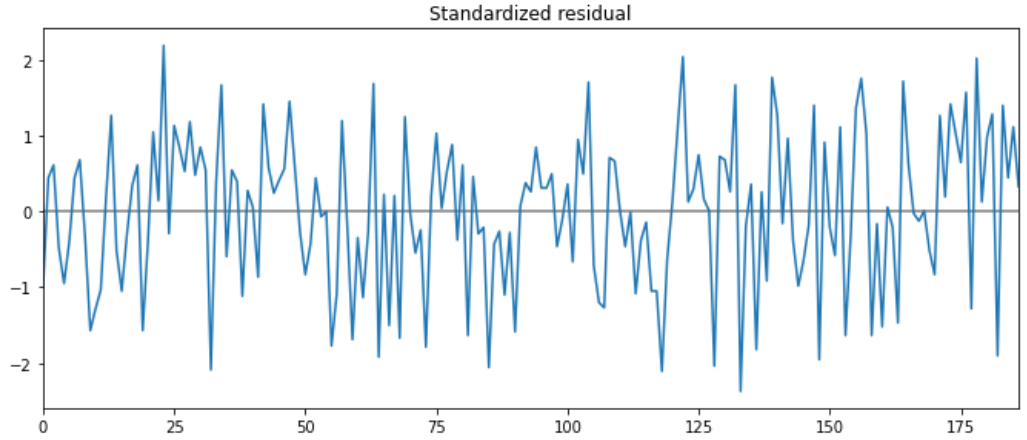
Moving average and mean are plotted. Mean is constant. Also there is no trend ⟹ *Data is stationary*

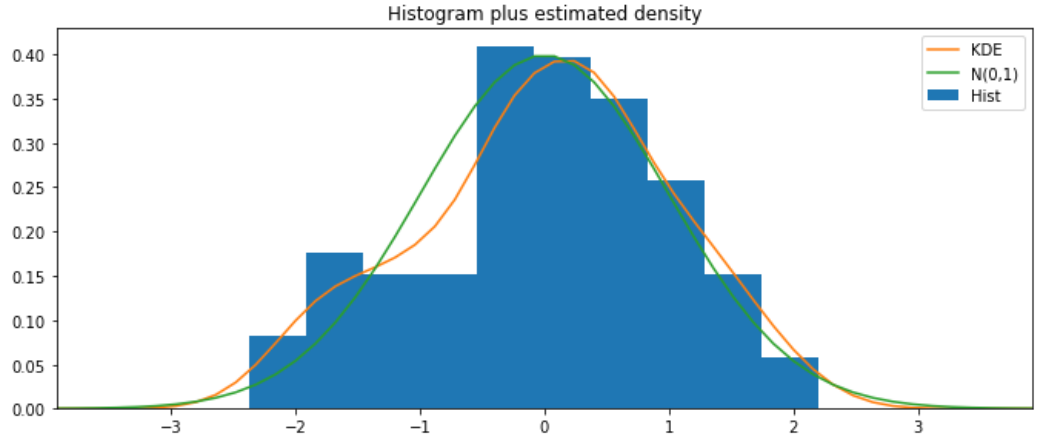


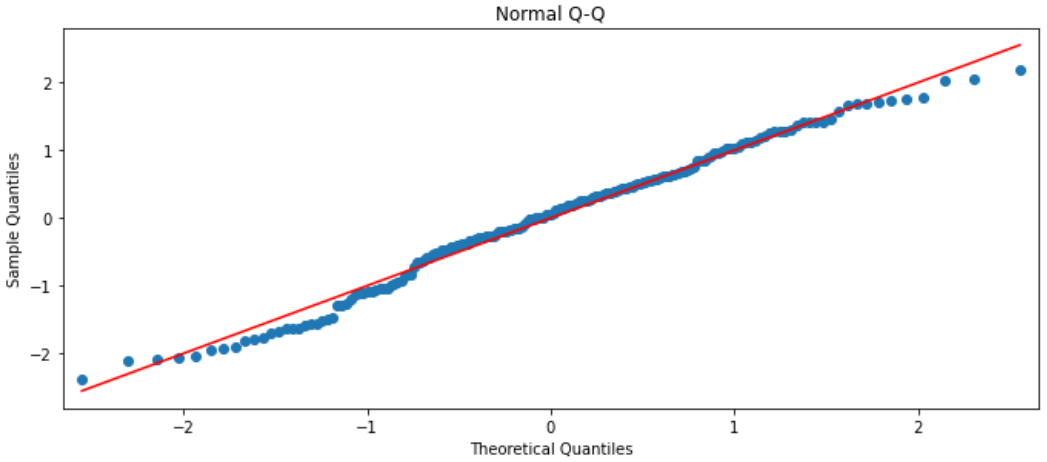
Residual Plots

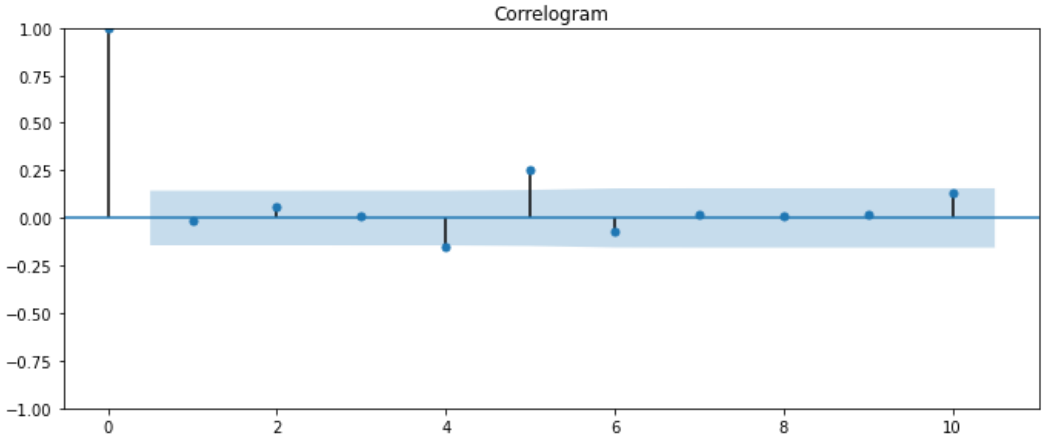
Mean is equal to zero

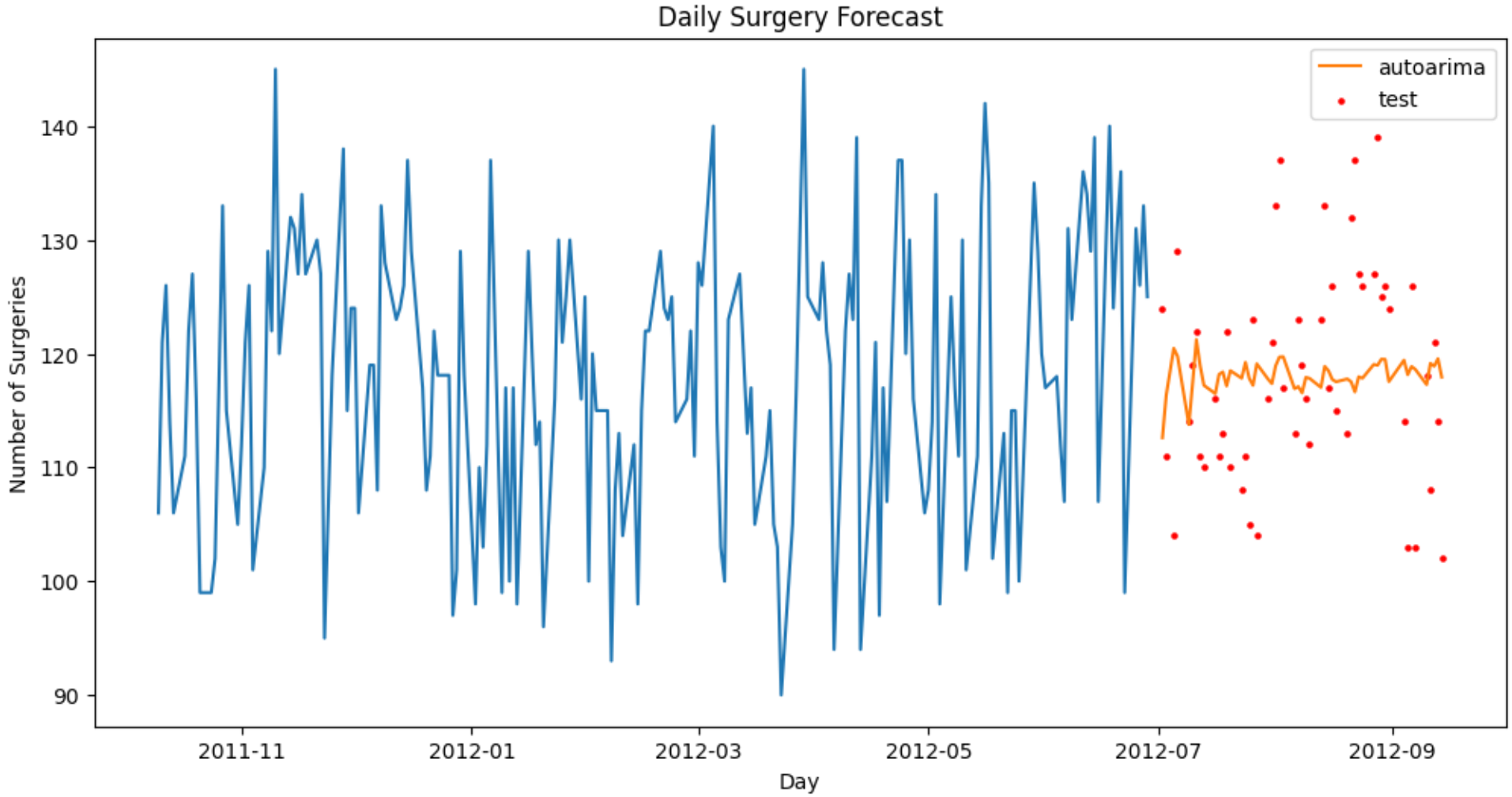
Residuals are normally distributed







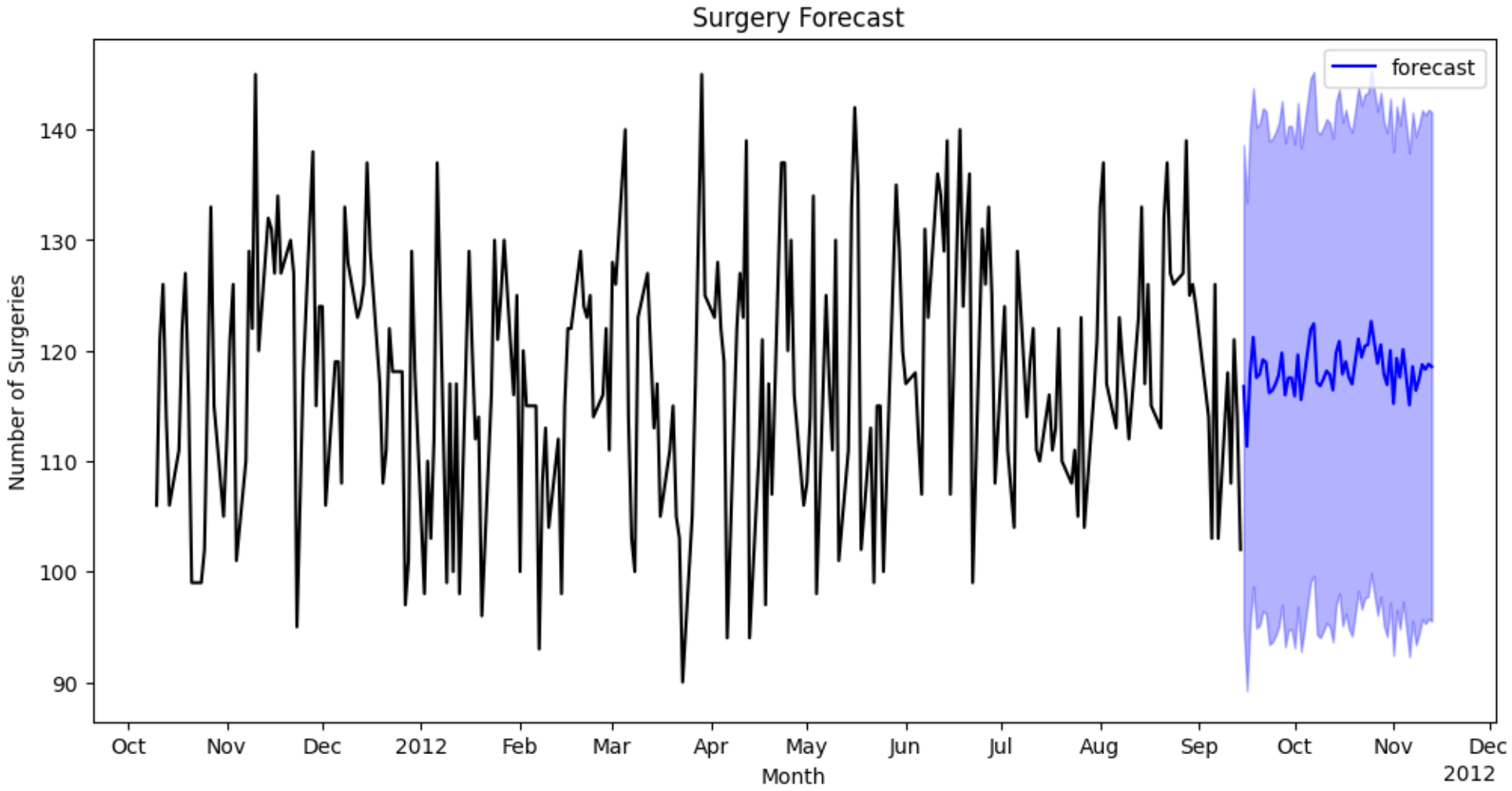




Loss function Metrics:

```
Mean baseline MSE:  86.15229539445873
Naive baseline MSE:  130.18867924528303
Naive seasonal baseline MSE :  155.9056603773585
Drift baseline MSE:  619.0044227852486
Auto Arima Model : 88.06641793210217
```

Forecast using the ARIMA model



I used different ways for ARIMA model but I could not get a good prediction.
The AIC when generating different values for (p,d,q) were high. Below some of them are provided:

```
(0, 0, 0) 1489.5667474309964
(0, 0, 1) 1485.3595422292196
(0, 0, 2) 1485.1310791220521
(0, 0, 3) 1487.1221380542775
(0, 0, 4) 1479.2617880570504
(0, 0, 5) 1472.4513069732993
(0, 0, 6) 1474.1703451490507
(0, 0, 7) 1476.1321001314122
(0, 0, 8) 1477.9501000801122
(0, 0, 9) 1479.6527396918286
(0, 0, 10) 1477.3399196416985
(0, 0, 11) 1479.3025715335489
(0, 1, 0) 1572.7559401553024
(0, 1, 1) 1488.1355788894148
(0, 1, 2) 1484.4300525983201
(0, 1, 3) 1484.1612588420608
(0, 1, 4) 1486.1504415633983
(0, 1, 5) 1478.2104166965864
(0, 1, 6) 1471.432647640444
(0, 1, 7) 1473.1300402653153
(0, 1, 8) 1475.0799529314772
(0, 1, 9) 1476.9096719734255
(0, 1, 10) 1478.5940914937573
(0, 1, 11) 1476.2474006609486
(0, 2, 0) 1763.0464728449053
(0, 2, 1) 1572.670161228788
(0, 2, 2) 1493.8287098428084
(0, 2, 3) 1490.109892699391
(0, 2, 4) 1489.9784735979547
(0, 2, 6) 1483.8267407492426
(0, 2, 7) 1477.3412851518333
(0, 2, 8) 1479.4411829150472
(0, 2, 9) 1481.3568243132768
(1, 0, 0) 1484.4556612823676
(1, 0, 1) 1486.0151608349383
```

```
(9, 2, 6) 1490.4369433078514
(10, 0, 0) 1479.9064359757806
(10, 0, 1) 1480.8117777000584
(10, 0, 2) 1478.6445915051313
(10, 0, 3) 1480.6439878813107
(10, 0, 4) 1482.5122674828663
(10, 0, 5) 1482.7236054749924
(10, 0, 6) 1484.7205275475362
(10, 0, 8) 1507.849415657502
(10, 0, 9) 1506.856770201381
(10, 0, 10) 1481.3215407985076
(10, 0, 11) 1473.096015376097
(10, 1, 0) 1485.9916041371612
(10, 1, 1) 1487.037057918516
(10, 1, 2) 1474.4696671426516
(10, 1, 3) 1478.5653835850676
(10, 1, 4) 1471.8491310397876
(10, 1, 5) 1481.489169573128
(10, 1, 6) 1473.0305967870786
(10, 2, 0) 1518.6718356512672
(10, 2, 1) 1489.4258835063315
(11, 0, 0) 1478.704275484553
(11, 0, 1) 1480.512837859147
(11, 0, 2) 1481.5917208642363
(11, 0, 3) 1481.6958013658123
(11, 0, 5) 1510.6111850752845
(11, 0, 6) 1470.3612501584253
(11, 0, 7) 1507.1078627680636
(11, 0, 8) 1502.8994196129415
(11, 0, 9) 1519.2506993167872
(11, 0, 10) 1476.1189492327471
(11, 0, 11) 1476.9658348967228
(11, 1, 0) 1485.603809033703
(11, 1, 1) 1477.7757786151715
(11, 1, 2) 1479.2808083528637
(11, 1, 3) 1474.4279230596196
```

```
(5, 2, 2) 1492.2651579212913
(5, 2, 6) 1475.9963822414466
(6, 0, 0) 1473.8973986245362
(6, 0, 1) 1475.881493026699
(6, 0, 2) 1477.8810934654298
(6, 0, 3) 1472.953180169633
(6, 0, 4) 1478.931711526146
(6, 0, 5) 1471.350554702432
(6, 0, 6) 1473.094359170187
(6, 0, 7) 1470.336105425788
(6, 1, 0) 1489.1063029680809
(6, 1, 1) 1472.8815657281484
(6, 1, 2) 1474.8573661409005
(6, 1, 3) 1473.8733463047756
(6, 1, 9) 1497.9361115648685
(6, 2, 0) 1542.0685518971727
(6, 2, 1) 1491.869602601409
(6, 2, 3) 1491.8447852487495
(6, 2, 6) 1475.6606333656628
(7, 0, 0) 1475.8805095999508
(7, 0, 1) 1477.7423439025895
(7, 0, 2) 1474.5711818921081
(7, 0, 3) 1480.8014197468276
(7, 0, 4) 1475.8623375803415
(7, 0, 5) 1473.0956652504194
(7, 0, 7) 1467.6868708584282
(7, 1, 0) 1488.6388854722586
(7, 1, 1) 1474.8562534184741
(7, 1, 2) 1474.8177699756
(7, 1, 3) 1473.2944377635663
(7, 2, 0) 1540.384560292341
(7, 2, 1) 1491.628137691521
(8, 0, 0) 1477.859376422385
(8, 0, 1) 1479.8368457955132
(8, 0, 2) 1476.9882787569932
(8, 0, 3) 1474.0392174629112
(8, 0, 4) 1475.0003517543073
(8, 0, 5) 1483.8310763298186
```