Forecasting Daily Call Volumes for Ambulance services in Jakarta

Rutvik Randive - 23013832

Cardiff University

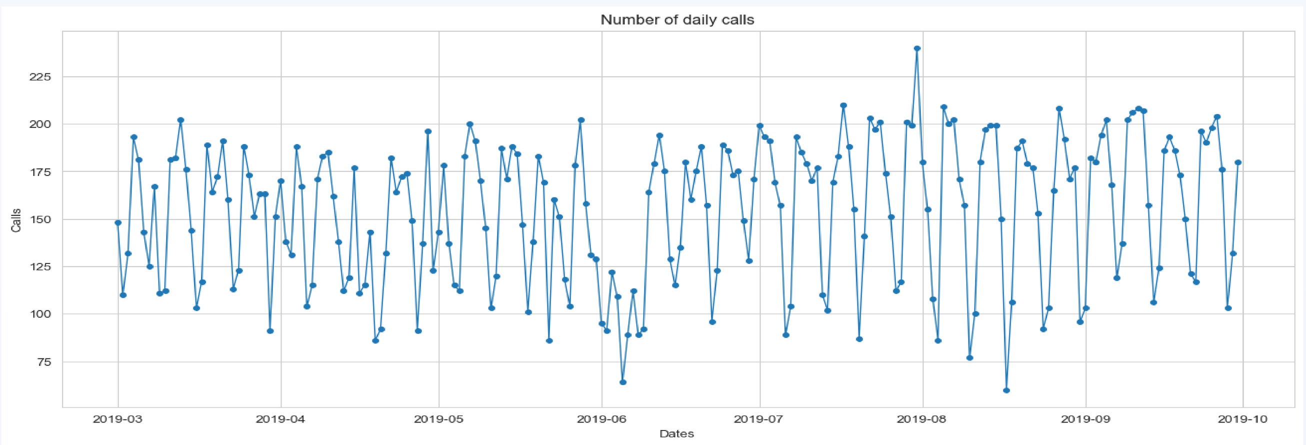
Problem Statement

In the dynamic environment of emergency medical services, efficient resource allocation plays a critical role in ensuring timely response to emergency calls and optimizing operational effectiveness. The ambulance company aims to optimize resource allocation for the next two months by analyzing six months of historical data. Additionally the company would also be interested in any other insights that can be provided to them.

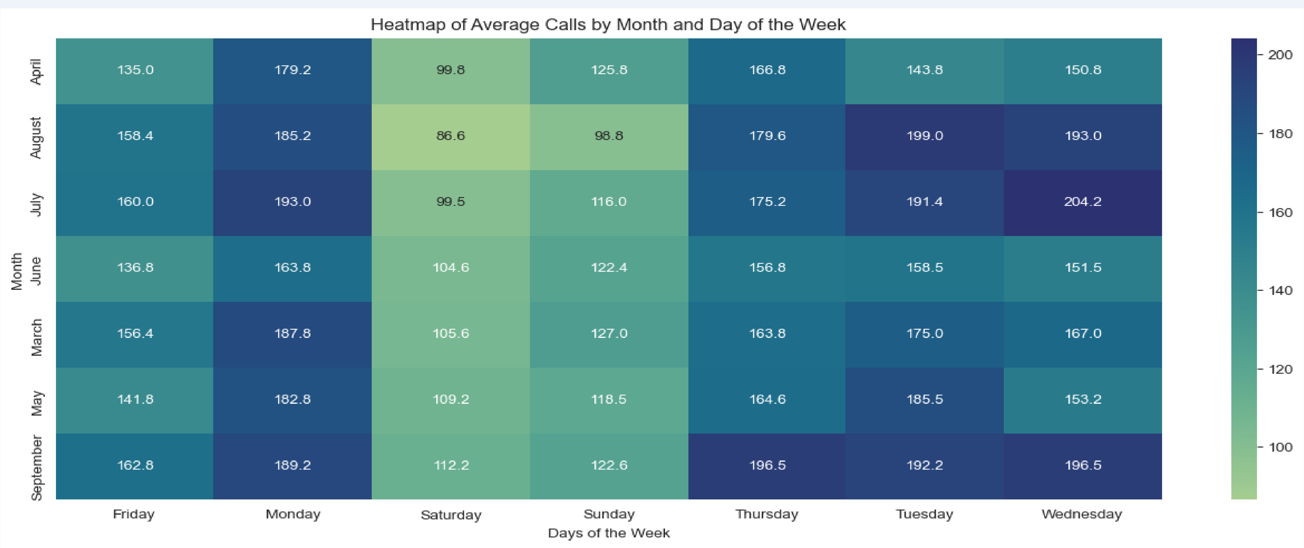
Preliminary Analysis and Summary

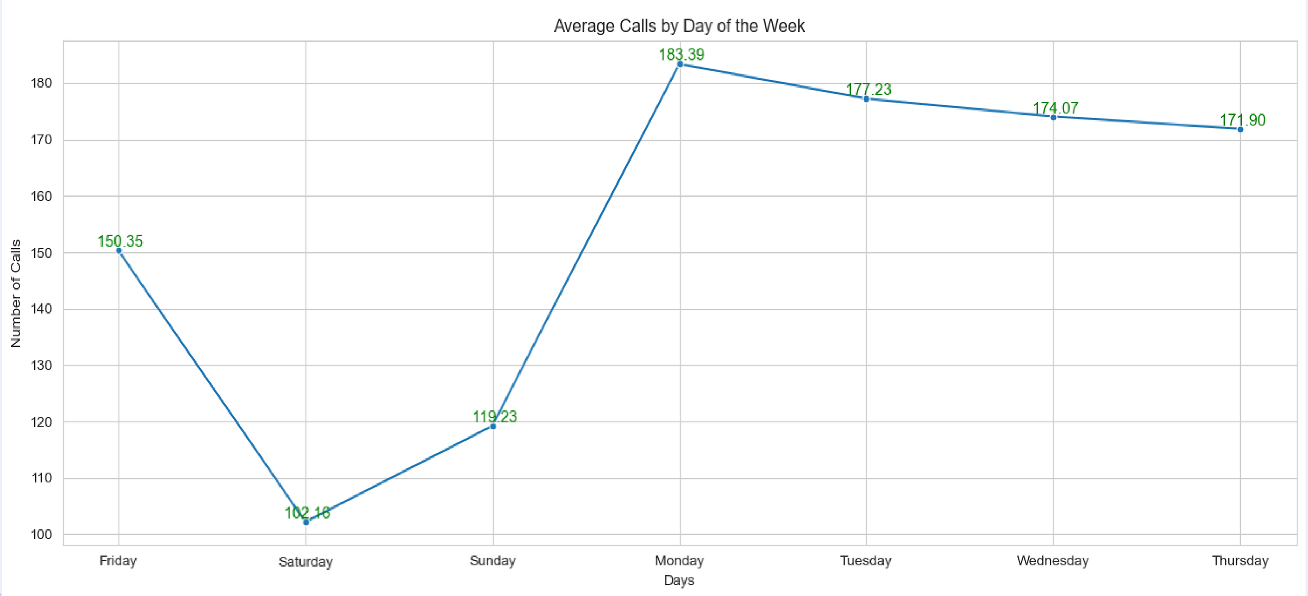
The data initially had to be preprocessed because it was just a column of calls and their timings mashed up together. I extracted the dates from the data and number of times each call was made on a particular date. This is the numerical summary of the calls made from March 2019 to September 2019 to Ambulance services in Jakarta. The dates then had to be converted to a datetime index to make use of the dataset easier in the further steps. The average calls made during these 7 months is 153 calls per day.

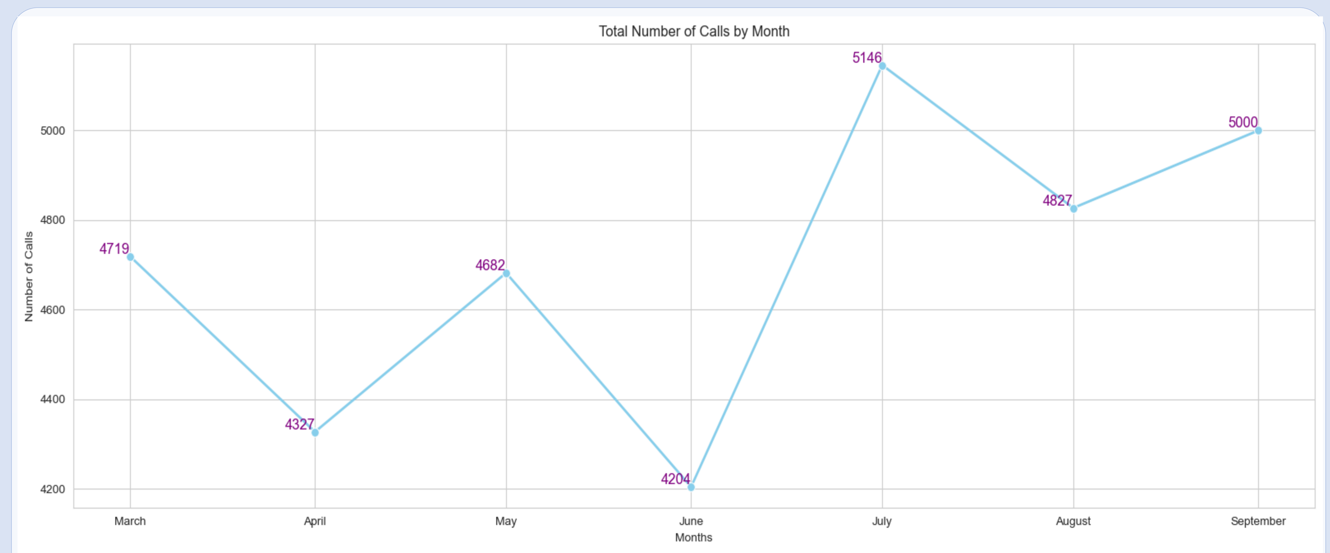
	Min	Max	Median	Mean	IQR	Variance	Standard Deviation
1	60	240	163.5	153.761682	64.5	1400.266882	37.42014



The highest number of calls made on a day is 240 while 60 being the lowest. We can also see that there is a significant difference between the calls made on weekdays and weekends. There are some outliers too. The data also shows weekly seasonality suggesting seasonal models like SARIMA and Holt Winters will work well on it. Due to its popularity as a travel destination, tourists will typically go out on the weekends to have fun. If they become ill, they will tend to call for an ambulance the following day, which is Monday with the highest average. This is one of the reasons I can think of weekends having very low average overall.



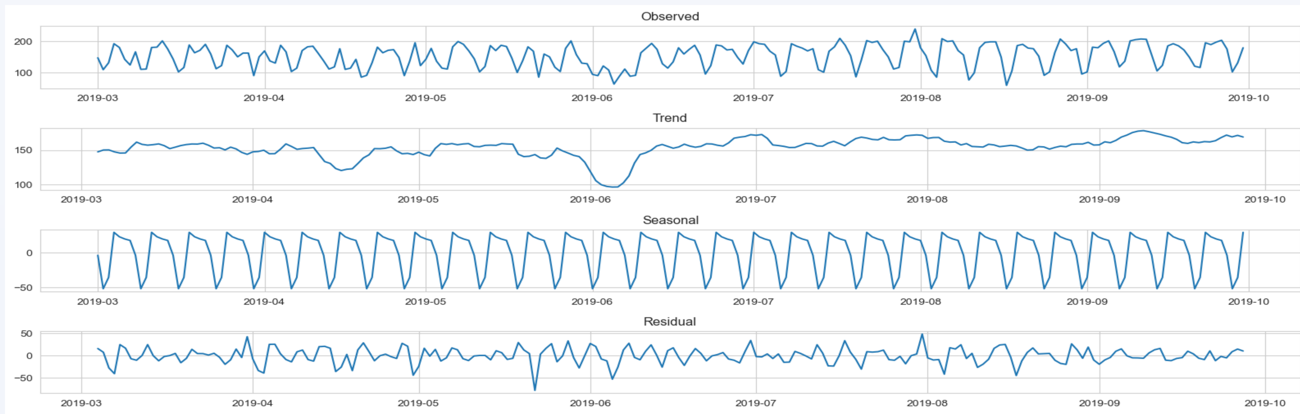


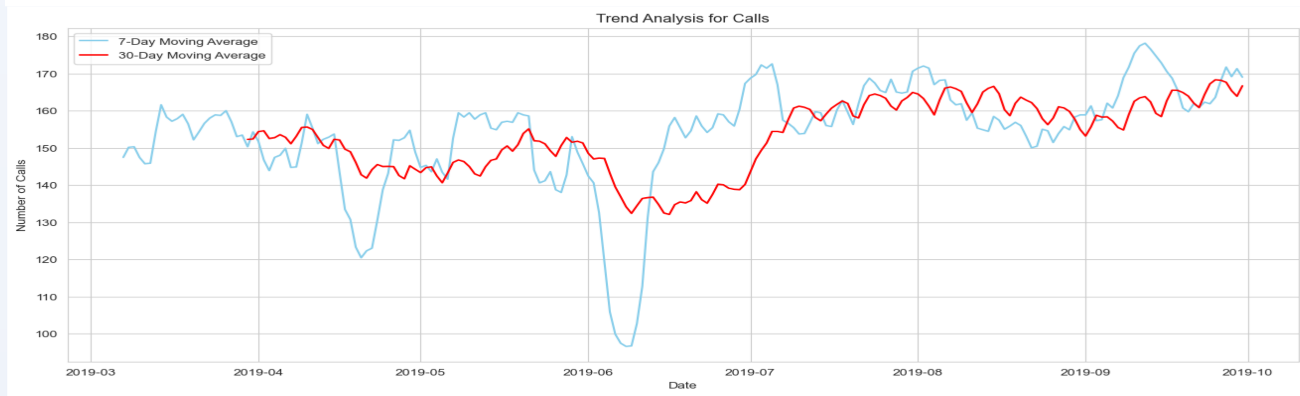


July is considered the peak season for tourism in Indonesia which further fits the bill for tourists needing emergency services along with the residents and this also explain why it has the highest number of recorded calls. June is a mid year break for school students in Indonesia and families usually go out for vacation out of town during this period which might explain why June has the lowest number of calls.

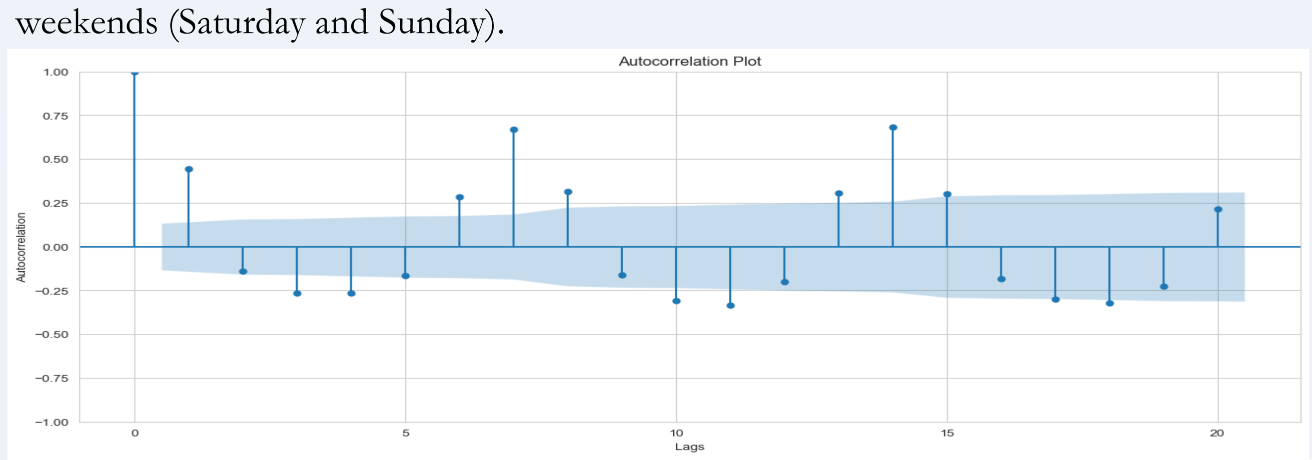
Decomposition and ACF/PACF plots

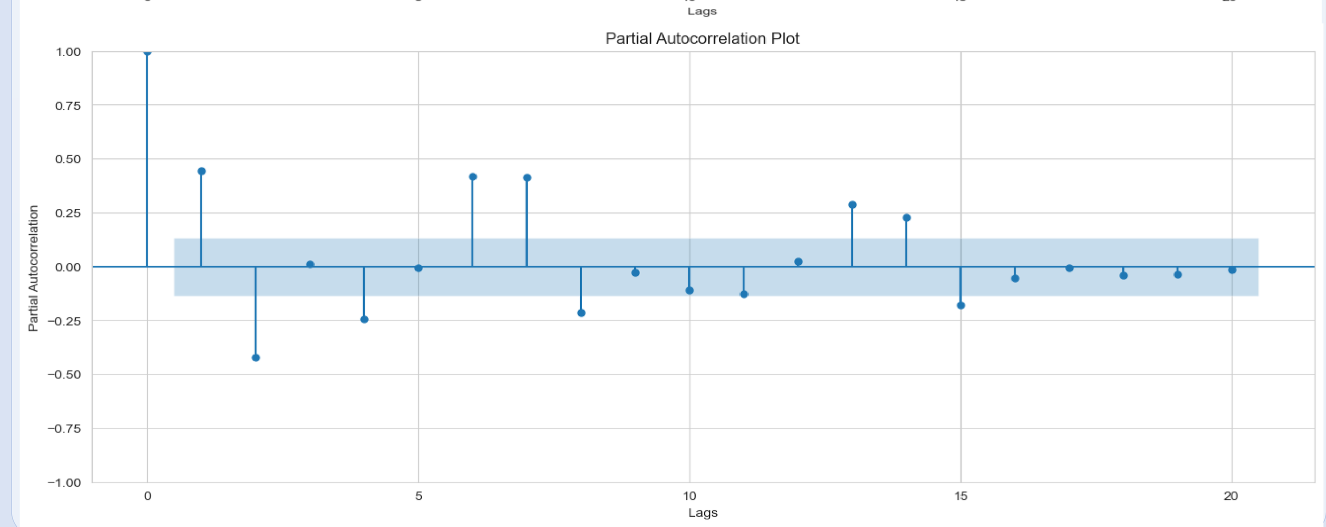
I used seasonal decomposition with an additive decomposition here to further understand the data and find any trends and insights and kept the period to 7 representing a weekly seasonal pattern. I also calculated Moving Averages for 7 and 30 days to find trends in the data.





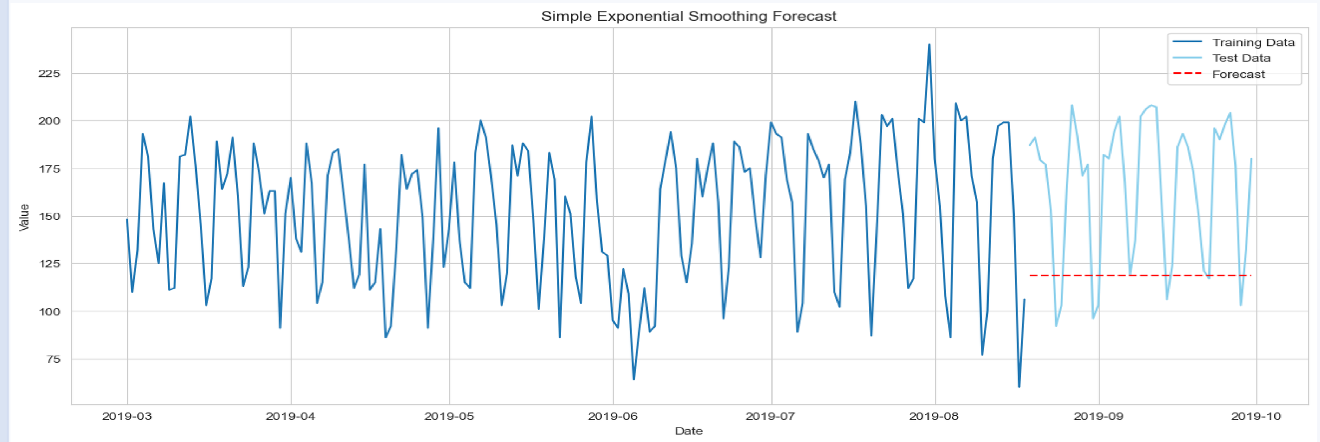
The ACF (Autocorrelation Function) plot shows significant lags 7 and 14 indicating a weekly trend which I already discussed. The PACF (Partial Autocorrelation Function) plot shows spikes at 6th and 7th lags which indicates a direct relationship between the weekends (Saturday and Sunday).

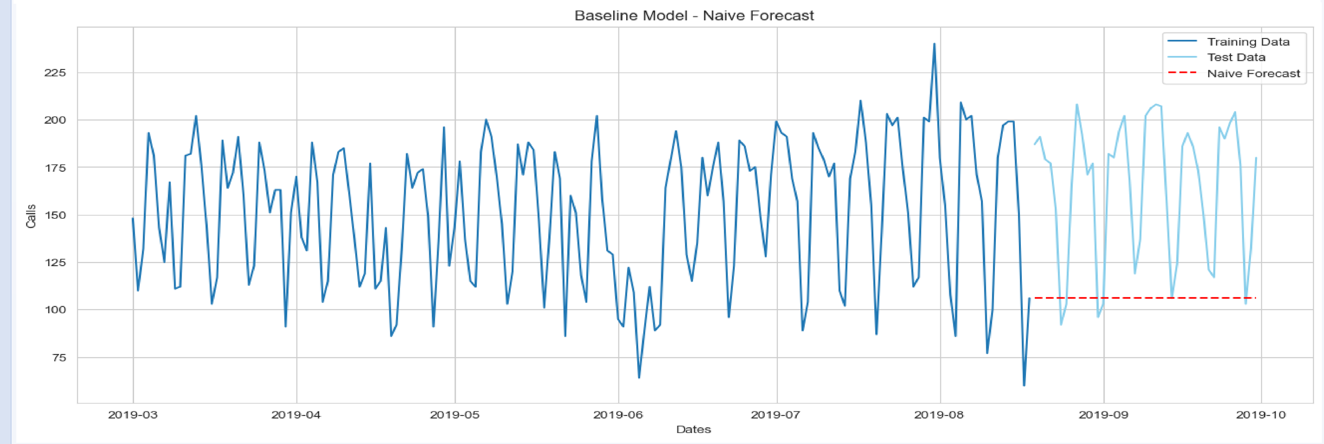




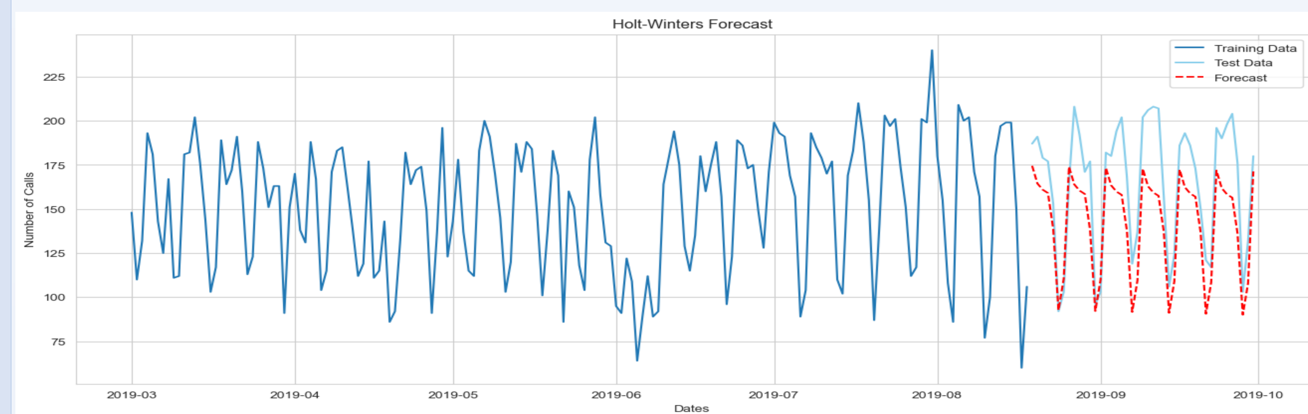
Baseline Models

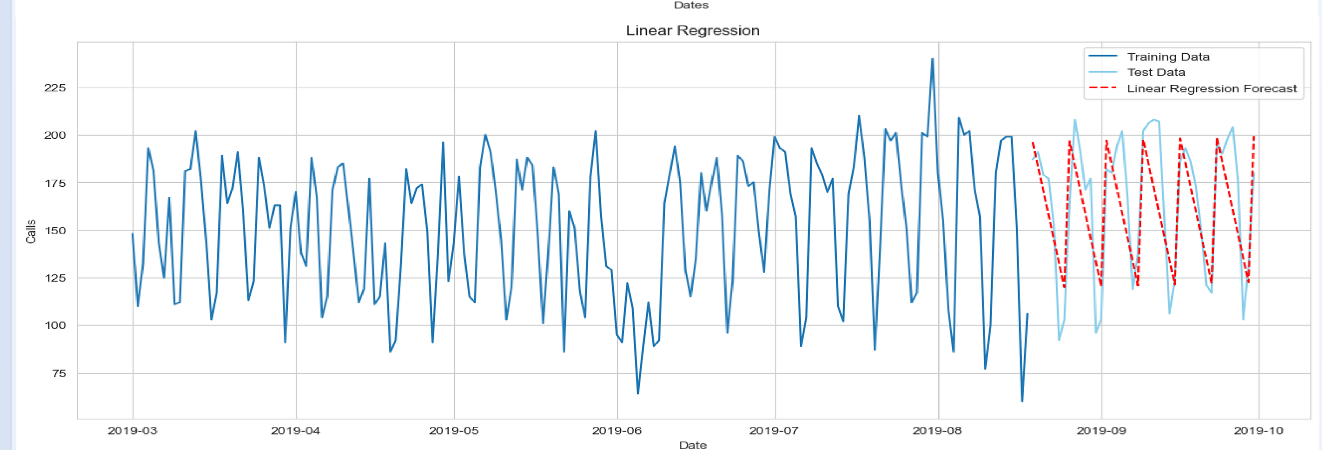
I made a training and a testing set from the data and the training set would be used to train the models and the testing one would be used to cross check out forecast accuracy and how well the model performs. The split was 80% and 20%. I implemented SES and Naïve models as the first ones and I was not confident that they will fit well.



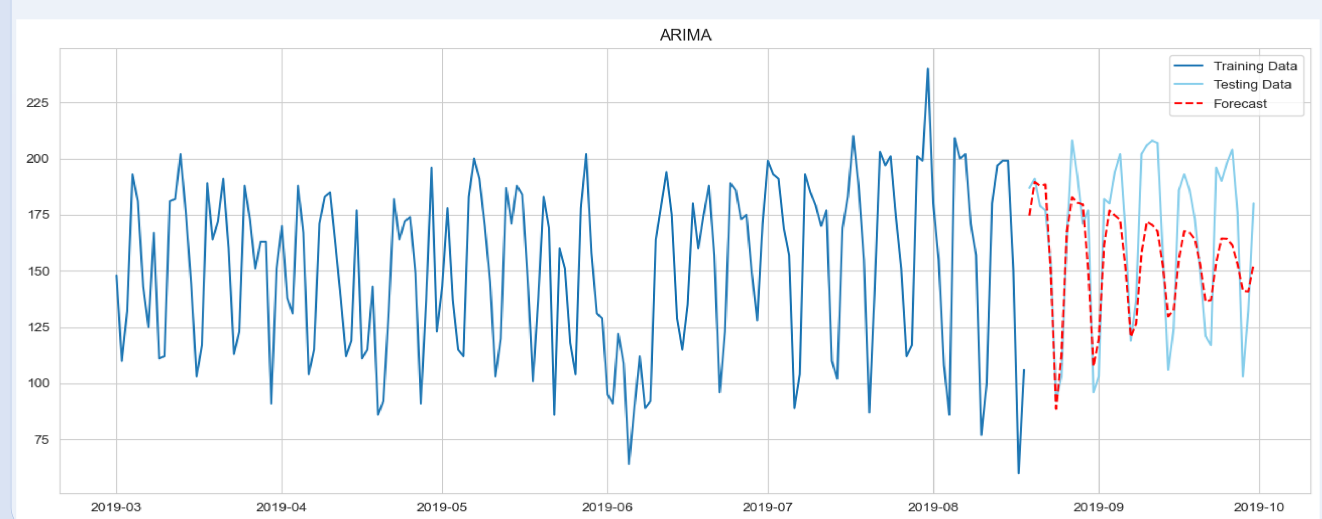


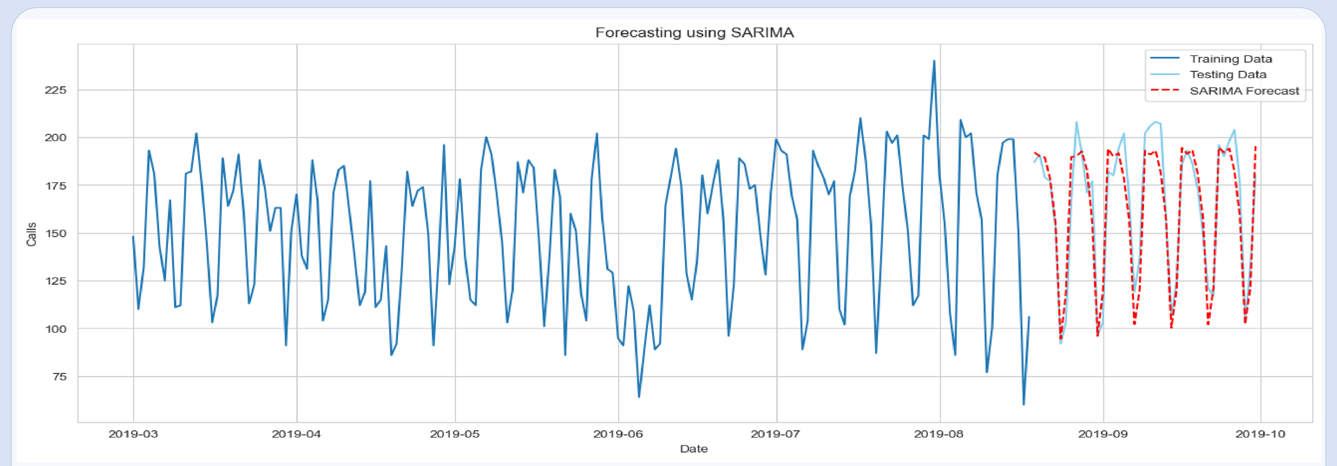
They were followed by Holt Winters model and Linear Regression. Holt Winters model accounts for seasonality so it performs a lot better since we have weekly seasonality here in our data. Linear Regression performs better the more features there are in the data so I added some more by extracting the 'day of the year' and 'day of the week' as representing numbers respectively.





ARIMA and SARIMA





I used the order for the ARIMA model based on the ACF plot. Since we had a weekly seasonality in our data I also wanted to try SARIMA which incorporates additional seasonal components and SARIMA indeed performed better than ARIMA when taking the weekly seasonality into account.

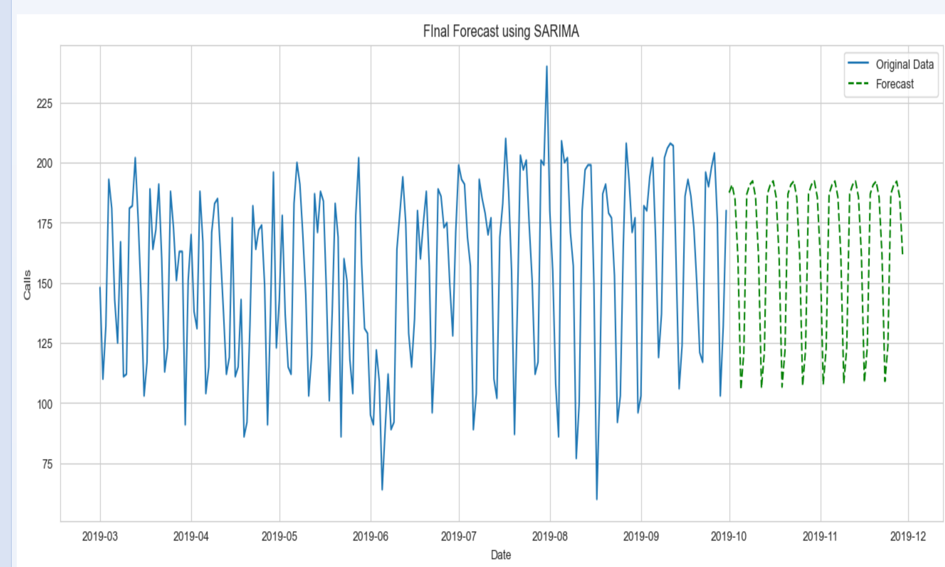
Error Statistics

As we can see from the comparison of the models Linear Regression performed very well with the extra features added. SES and Naïve didn't do well. Holt Winters performed well but SARIMA had the best fit overall because of the seasonality component.

Models	MAE	RMSE	MAPE	SMAPE	Bias
Simple Exp	51.328748	58.462094	28.777717	34.263353	46.153051
Naive	60.441860	68.976403	33.367454	41.885518	58.906977
HWES	23.667660	27.137616	13.820409	15.082491	22.541079
ARIMA	18.894300	22.802970	11.363003	11.681466	10.192748
SARIMA	9.911659	12.443632	6.153957	6.241203	2.274921
Linear Regression	18.647217	22.304755	12.200044	12.010434	-4.719845

I have also added an error statistic called as Forecast bias. It basically helps us to understand a consistent error in the forecast, where the forecasted values differ from the actual values in a particular direction either overestimating or underestimating. So basically the model is predicting correctly if the forecast bias is close to zero.

Forecast Values and Conclusion



Dates	Forecasted Values
2019-10-01	187.493158
2019-10-02	190.746665
2019-10-03	184.758969
2019-10-04	160.187747
2019-10-05	105.565354
2019-10-06	120.877088
2019-10-07	186.881403
2019-10-08	190.794449
2019-10-09	192.323251
2019-10-10	185.547787

In conclusion I used seven months of historical data to forecast daily call volumes for the upcoming two months. By predicting daily call volumes, our ambulance company can allocate resources, optimize staff scheduling, and ensure prompt emergency medical services for the residents of Jakarta. For future work I believe I would be able to find more insights and patterns if more data was available.

Finally I want to provide one more insight which could be worth taking a look is about the public holidays in Indonesia. Almost all of the holidays have very low number of calls made on that day and there is a public holiday on most of the lowest points in the data.