Shaun Robert Commee - C22085988

Forecasting Ambulance Demand in Jakarta

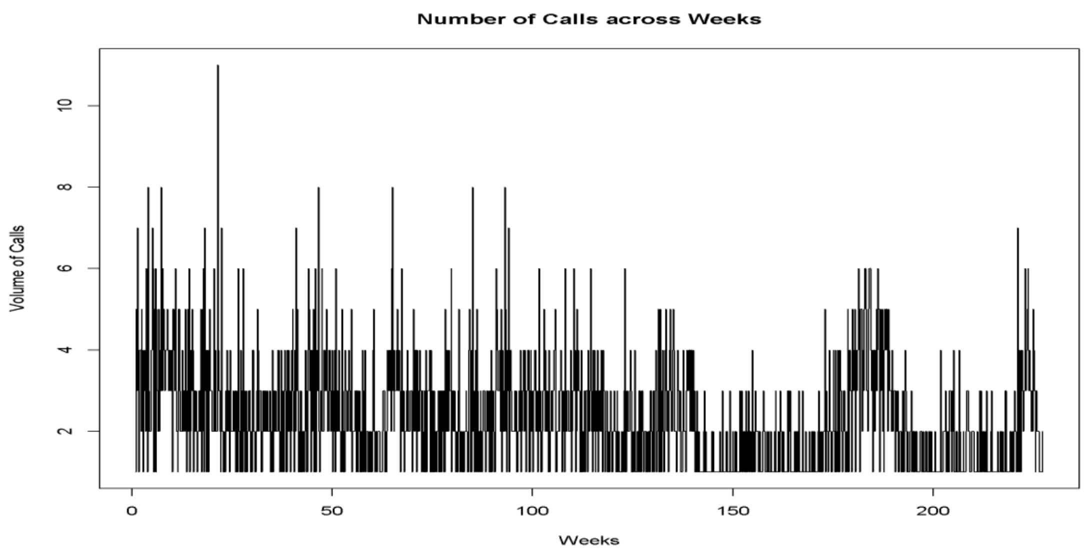
Aims & Motivations

To understand the demand for ambulances in Jakarta is challenging, however, attempting to predict it is essential to appropriate resourcing and an effective service. The aim of this project is to use data provided between January 2017 – October 2022 to forecast weekly call demand for ambulances between October 24th – 18th December 2022.

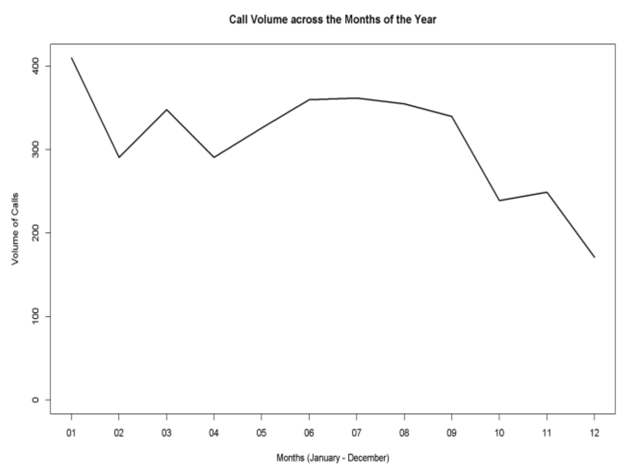
Preliminary Analysis of Data-set

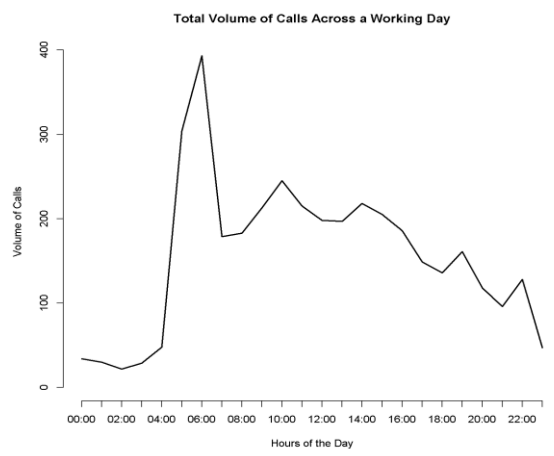
Below we can see a breakdown of the data-set into daily, weekly and annual components. In order to process the data for weekly time series for our forecasts, we have taken the daily data and implemented a frequency of 7. We have also plotted the data-set graphically across the weeks of the analysis.

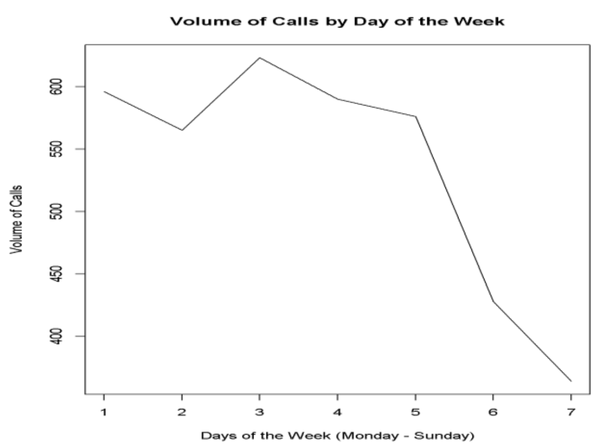
Calls	Mean	Variance	Standard Deviation	Minimum	Maximum
Daily	**2.40**	**1.9**	**1.40**	**1**	**11**
Weekly	**12.55**	**49.17**	**7.01**	**1**	**36**
Annual	**623.67**	**55,264.27**	**228.61**	**336**	**949**

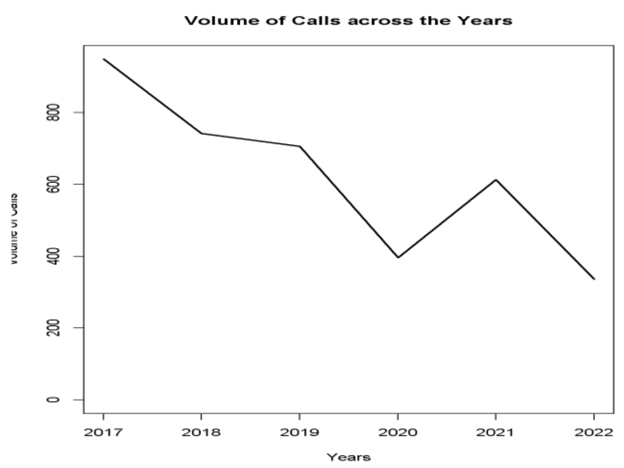


Daily, Weekly and Annual Seasonality









Patterns

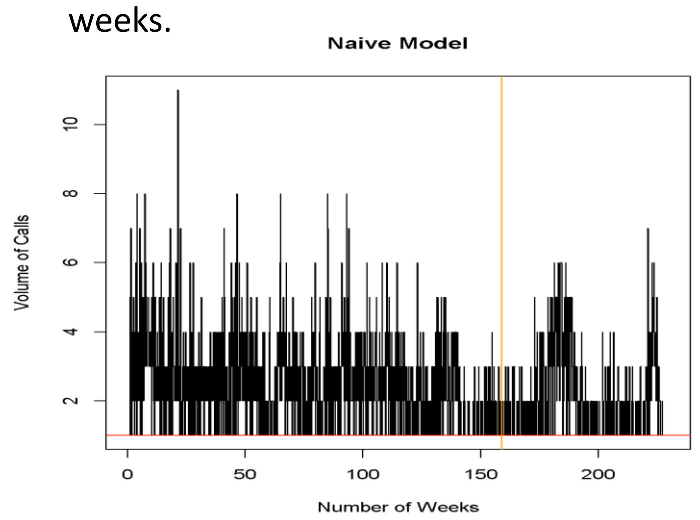
On review of the data, we observe that there is seasonality across daily call volume. There are peaks between 5:00AM - 6:00AM and lulls between midnight and 4:00AM. Across the months of the year, we see peaks during January with up to 410 calls recorded during this month. This drops off into February with a generally consistent volume up until October, November and December where we see a general decline. Across days of the week, we see a consistency amongst calls between Monday to Friday on average, but, a considerable drop going into the weekend by up to 30% in volume. Generally, we can observe that call volume has been dropping over the years.

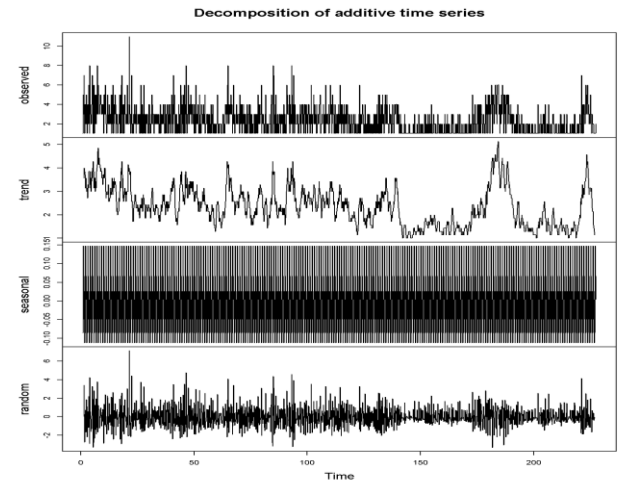
Splitting the Data-set

For us to be able to test the various forecasting methods that we will look to implement. We must split the data-set between training and test. Taking 70% of the data and training a model on this information to then be used on the remaining 30% as a measure of its performance. We will then use this accuracy to determine which model would be best placed for forecasting future events.

Naive Forecasting Model

Using a naïve forecasting model which takes the last observation and assumes that this value will continue into the future. We can observe the predictions that it has made on the right. Given that the last observation was a week on which one call took place, the naïve model has forecasted that for all future weeks.



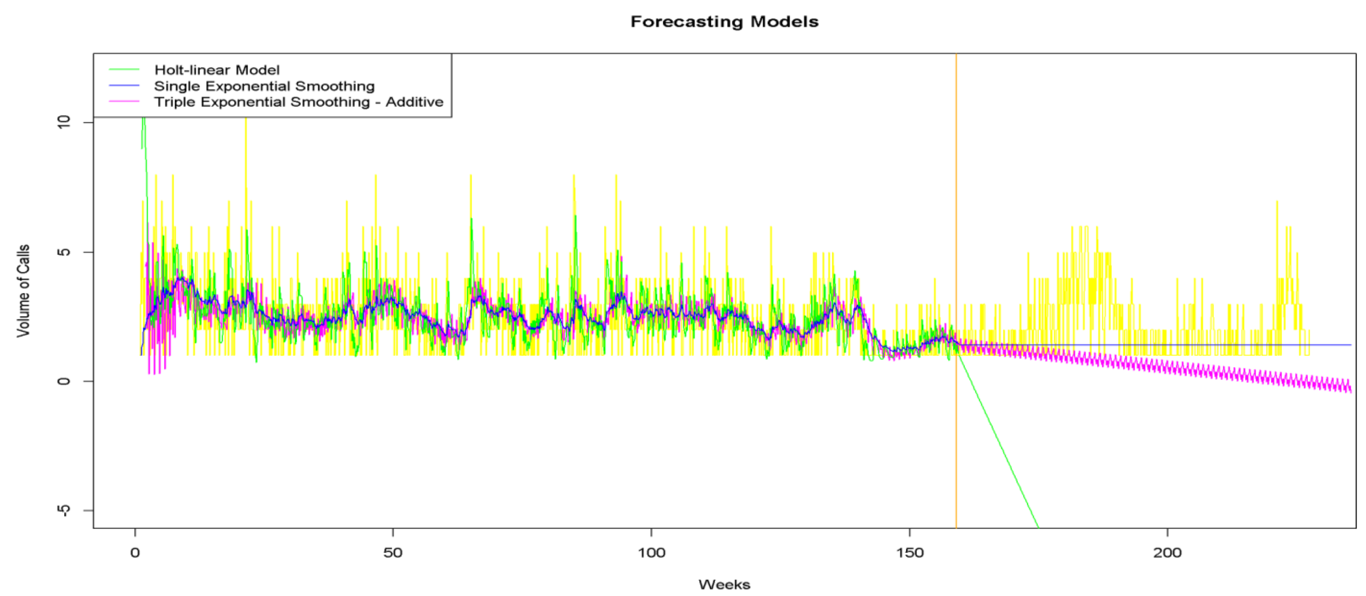


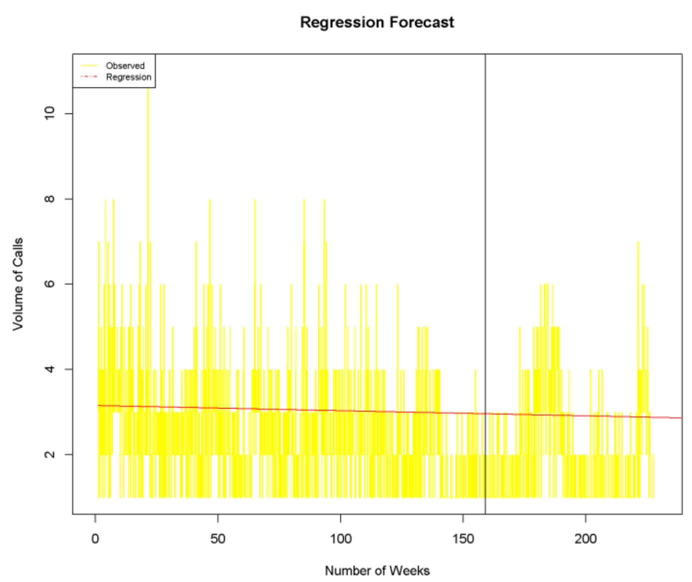
Decomposition

A time series is considered to be composed of three elements. Trend, seasonality and level. Dependent upon the data-set we can apply additive or multiplicative decomposition. In the case where we observe exponentially increasing level or seasonality with a rising trend we would opt for a multiplicative analysis which multiplies all of these elements together. In this case, the data doesn't demonstrate this pattern and so we have chosen for an additive decomposition which adds all of the elements together. We observe high levels of randomness throughout the data which will obscure any clear patterns in the data when we are forecasting.

Holt-Winters Models & Exponential Smoothing

Holt-Winters forecasting method allows us to breakdown each of the components mentioned, if selected, and smooth them appropriately. Holt-linear model takes both the level and the trend, SES takes just the level and Triple Exponential Smoothing includes all three components. Dependent upon the nature of the data-set, the more accurate each respective forecast is.



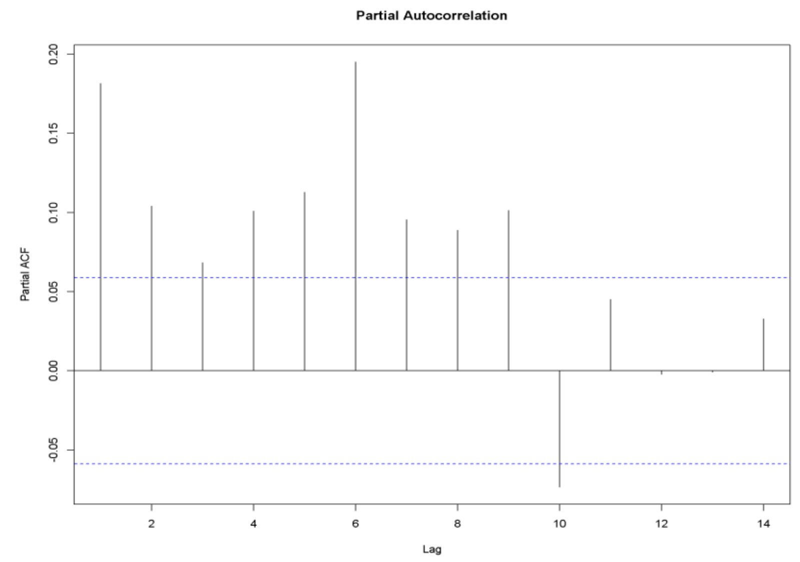


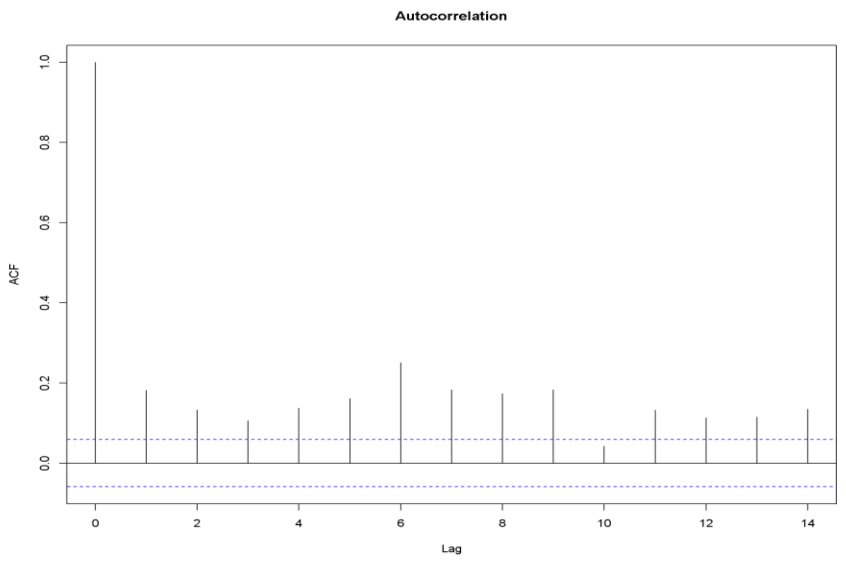
Regression

During this Regression analysis we took the volume of calls as a dependent variable and time as an explanatory variable. This then produced the forecast we see on our right. A subtle downward trend of the data. We obtained an intercept of 3.151618 and a slope of -0.001214. This shows a small decreasing relationship between time and volume of calls across this data-set.

ARIMA Models

ARIMA models review past data, perform differencing and utilise moving averages to generate forecasts. In order to establish which model is most suitable we must ascertain whether there is stationarity and seasonality in the data. By using a frequency of 1 we can examine whether there is seasonality by using an autocorrelation function. This helps us determine which model is suitable for forecasting with. Our observations provided below demonstrate that there is no clear seasonality at any particular lag in the data.

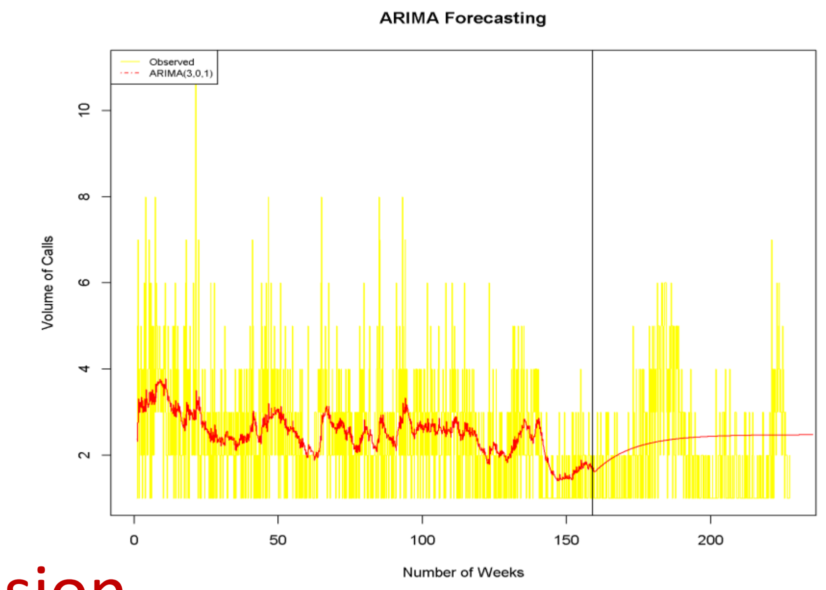




As we are aware that there is no seasonality in the data we will disregard the use of SARIMA. By undertaking auto.arima() function in R, it locates the best possible model by AIC and which therefore attains the highest predictive capabilities. Undertaking an Augmented Dickey-Fuller test, we know that the data is stationary as p-value < 0.05. Given this, we exhibit a forecast generated on an ARIMA (3,0,1).

Model Performance

Model	MAPE	MSE
ARIMA	72.91%	1.74
Naive	108.19%	2.88
Holt-Winters Triple Exponential Smoothing	62.37%	3.86
Holt-Linear Model	1003.88%	322.77
SES	44.10%	2.16
Regression	47.13%	2.35



Conclusion

In review of the models we see observe that the single exponential smoothing model was the most accurate on the test set with a MAPE of 44.10% and an MSE of 2.16. We however generally observed weak predictive accuracy across all of the models. The existence of numerous outliers particularly concentrated at the beginning and the end of the data-set led to a high degree of randomness. This analysis was backed up by the Ljung-Box test in which the p-value was significantly lower than the confidence level. This confirmed that we can't prescribe the randomness as simply white noise and that there is a high degree of it contained within the data. This factor coupled with a lack of seasonality and trend within the data made any analysis which focused on these components very poor (ARIMA, Holt-Winters Triple Exponential Smoothing). It is therefore somewhat unsurprising that given the lack of trend and seasonality that Regression and SES (given its focus on the level) performed the highest out of the forecasting models. Given the results, none of the models could be recommended for forecasting future demand.

Issues & Future Recommendations

Whilst undertaking forecasting with ARIMA, R had suggested through the auto.arima function that the model which contained the lowest AIC was a SARIMA model with differencing. Given the fact that there was no clear seasonality and the Augmented Dickey-Fuller Test stated there was stationarity of the data I opted for the ARIMA model to not consider these factors and create the ARIMA model we see above. For future forecasts, I would recommend only using the most recent data in an attempt to increase forecasting accuracy by reducing the randomness in the data and the length of the forecast.

Model Forecasts

Model	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
ARIMA	**17**	**17**	**17**	**17**	**17**	**17**	**17**	**17**
Holt Winters Triple Exponential Smoothing	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
SES	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**
Regression	**9**	**9**	**9**	**9**	**8**	**8**	**8**	**8**
Naive	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**7**
Holt-Linear	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**