JAKARTA EMERGENCY SERVICES

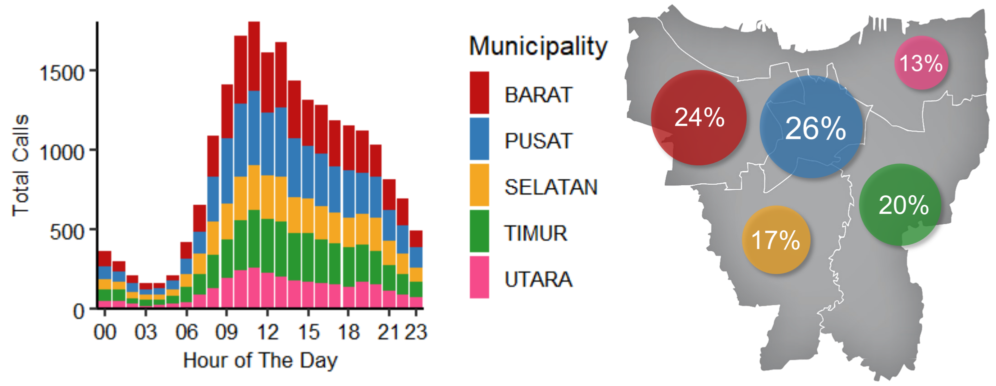
Overview

Three quantitative forecasting models were created to understand and predict the short-term demand for ambulance services in the city of Jakarta, Indonesia. The analysis used a dataset comprised of 22 540 emergency calls that were made over a period of five months (from 2019/01/01 to 2019/05/31). For evaluation purposes, a 70/30 data split was used to train and test the models, respectively.

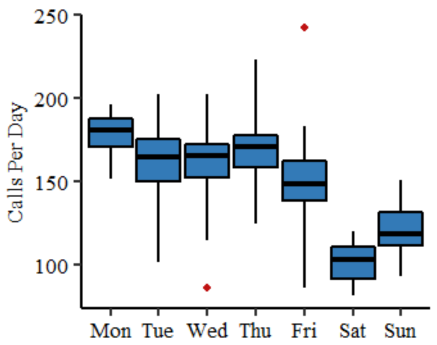
Call Data Analysis

After all the data was cleaned and segmented into administrative cities, 0.95% of the entries were deemed as either indistinguishable or as belonging to the Thousand Islands municipality. These data points have been excluded from the call data analysis. However, they were included in the time series analysis and in the forecasting models.

From the preliminary analysis (illustrated below), it can be observed that majority of the emergency calls were made between 08:00 - 20:00 with Pusat, Barat, and Timur having the highest demand requirements. It was further observed that the proportionate demand, per hour of the day, appeared to remain fairly stable across all administrative cities.



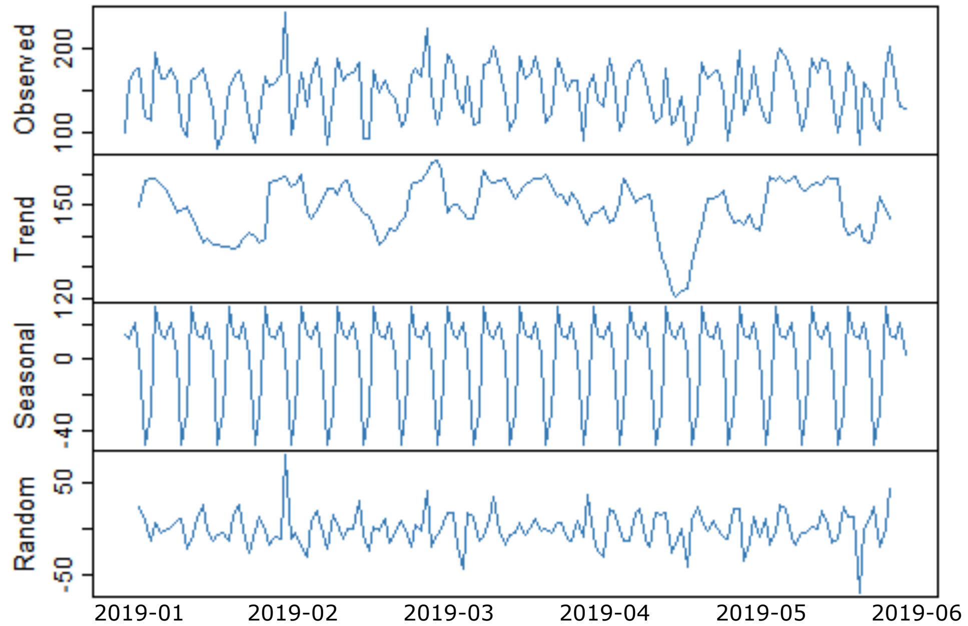
When considering the data per day, it was noted that demand levels were highest on Mondays and then appeared to slowly decline as the week progressed. Saturdays and Sundays consistently showed the lowest demand, with weekends accounting for only 22% of all calls.



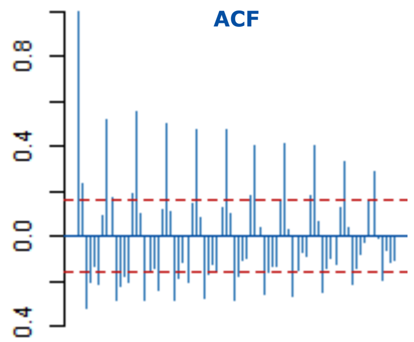
	Count	Calls	Average	Std Dev
Mon	21	17%	179	11
Tue	21	15%	160	26
Wed	21	15%	161	26
Thu	21	16%	169	21
Fri	21	15%	152	29
Sat	21	10%	102	11
Sun	21	12%	120	15

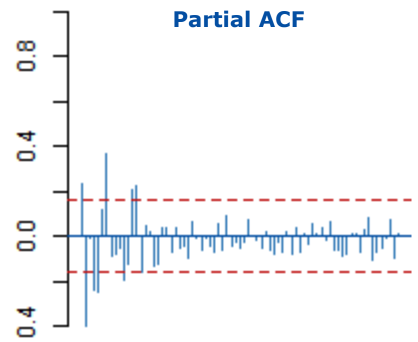
Time Series Analysis

After visually inspecting the time series, emergency calls showed strong signs of consistent (additive) seasonality, while also displaying short to medium term trend fluctuations that were mean reverting in the long term. This suggested that to make short-term forecasts the models would need to address both seasonality and trend.

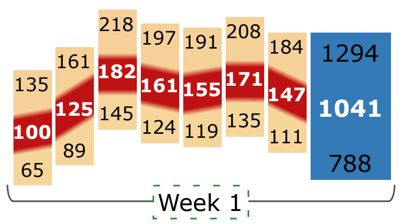


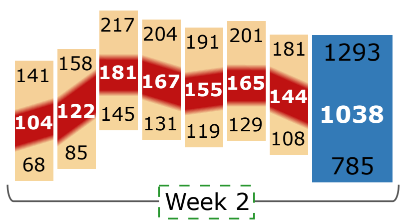
The autocorrelation function (ACF) and partial ACF below confirmed that the data was non-stationary and was correlated at lags six and seven. This verified that the data followed a weekly seasonal pattern.



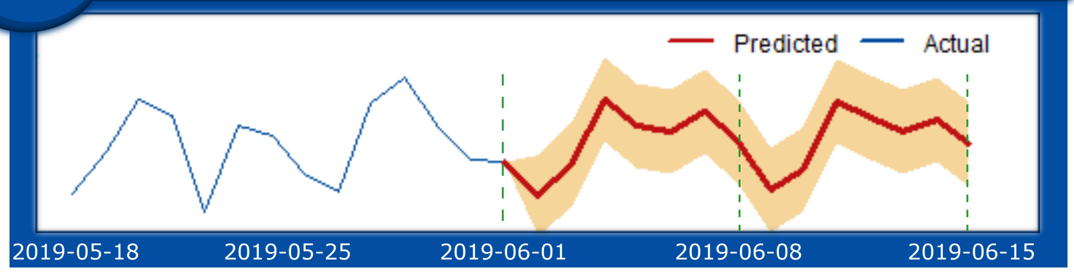


The ARIMA approach was selected due to its superior results and its dynamic ability to handle seasonal and non-stationary data. An ARIMA(0,0,1)(1,1,1)$_7$ model was developed based off all the data from 2019/01/01 to 2019/05/31. The results were forecasted as well as the range of future demand that was estimated with 90% confidence (see the lighter shaded region).





Summary and Conclusions

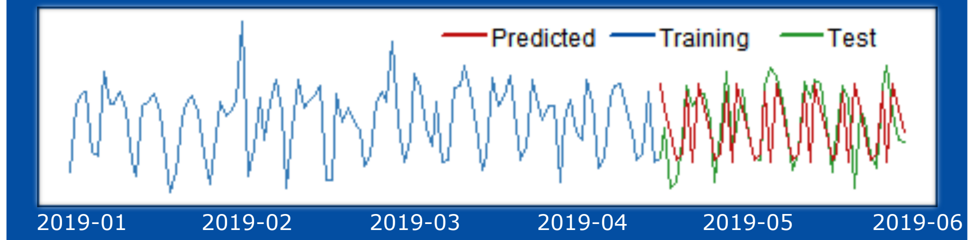


Forecasting Models

Naive Forecast 2 (NF2)

A basic NF2 baseline model which incorporates seasonality. The RMSE and MAPE were calculated and compared to the models presented below.

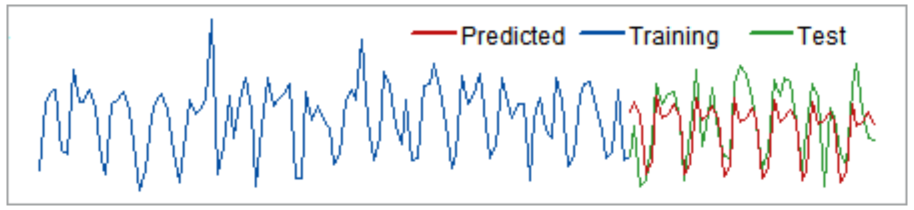
RMSE: 33.9
MAPE: 16.8



Holt Winters

The model was created with trend and additive seasonal components to address the observed properties of the data. The low RMSE and MAPE showed improvements compared to the baseline NF2 model. Furthermore, smaller estimated smoothing parameters indicated that estimates from the model placed less reliance on more recent observations.

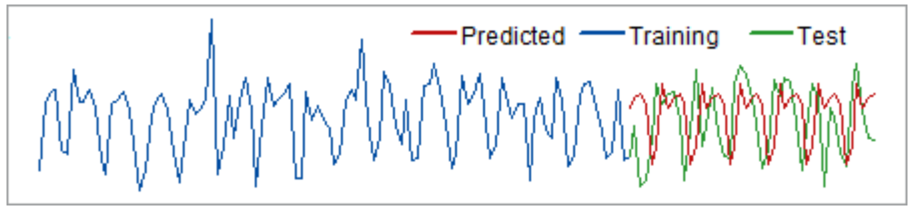
RMSE: 26.4
MAPE: 15.3



Multiple Linear Regression

The data was grouped by days of the week and seven regression models were created for each day, with dummy variables used to assign coefficients to each hour of the day. The model produced the best results. However, upon inspection the residuals of the model violated the required assumptions around normality, independence, and variance. These findings negatively impacted the reliability of the model and suggested that this approach was not appropriate for forecasting.

RMSE: 26.3
MAPE: 11.9



ARIMA

An ARIMA(2,0,0)(2,1,0)$_7$ model was estimated and tested for forecasting ability. Second order differencing was used to remove autocorrelation from the non-seasonal portion of the data. While from the seasonal portion of the data, both second and first order differencing were used to remove autocorrelation and non-stationarity. The ACF and partial ACF generated from the model confirmed that the data was stationary and the residuals were tested for normality. Ths inspired confidence in the reliability and appropriateness of this approach.

RMSE: 26.0
MAPE: 14.0