Improving Public Health Response: Data-Driven Demand Forecasting for Jakarta's Charitable Ambulance Service

Eshaq Rahmani - 22086790

1. Introduction

Efficient ambulance services are crucial for prompt medical response, impacting public health outcomes. Our project focuses on Jakarta's 118 ambulance services, aiming to enhance its efficiency through data-driven decisions. Leveraging nearly six years of historical data (01/01/17 to 23/10/22), we forecast the upcoming eight weeks' demand (24/10/22 to 18/12/22), uncover seasonality patterns, and identify prevailing trends. Our goal is to develop a practical resource allocation model, striving to significantly improve response times and service quality across the four ambulance stations in the city.

2. Preliminary Analysis

Dataset Description:

- Our dataset of 3,742 observations includes two variables: `Date_Incoming_Call` and `Incoming_Call_Time`, denoting the date and time of calls received from Jan 1, 2017, to Oct 21, 2022.

Data Preprocessing:

- Eight missing values in `Incoming_Call_Time` were identified and removed, ensuring a clean dataset for analysis.

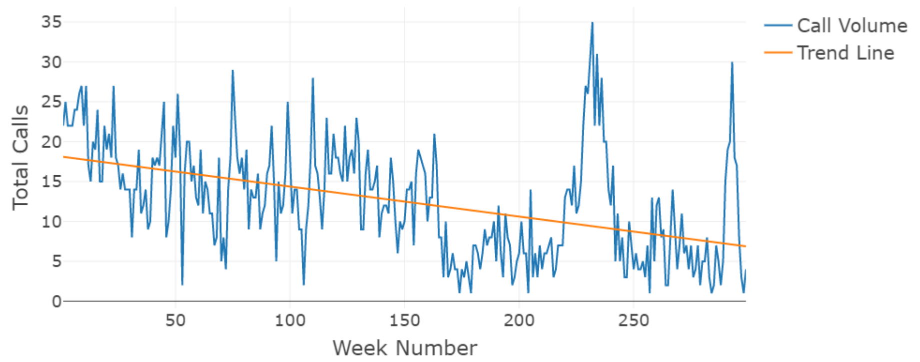

Figure 1. Total Calls by Week Number with Trend Line

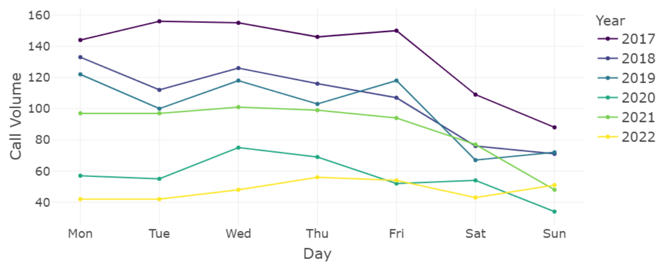

Figure 2. Daily call volume by Year

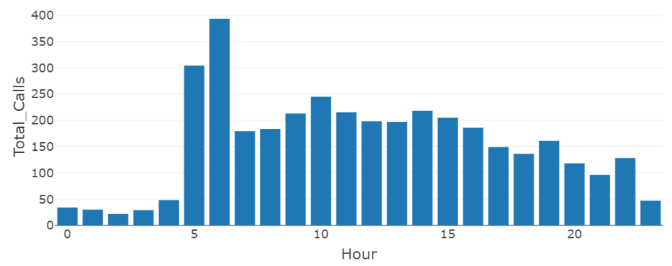

Figure 3. Aggregated Hourly call volume across all Years

Key Insights from Preliminary Analysis

- Weekday call peaks: Mondays, Wednesdays, Thursdays (Figure 2).
- Lower call volumes: weekends (Figure 2).
- Busiest hours: 5:00 AM - 6:00 AM (early morning), 10:00 AM - 11:00 AM (Figure 3).
- Lowest call volumes: 10:00 PM - 11:00 PM, 2:00 AM - 4:00 AM (Figure 3).
- Overall downward trend: 2017-2022 (Figure 1)
- Seasonality amplitude remains fairly constant. Additive decomposition will be suitable.

3. Data Decomposition

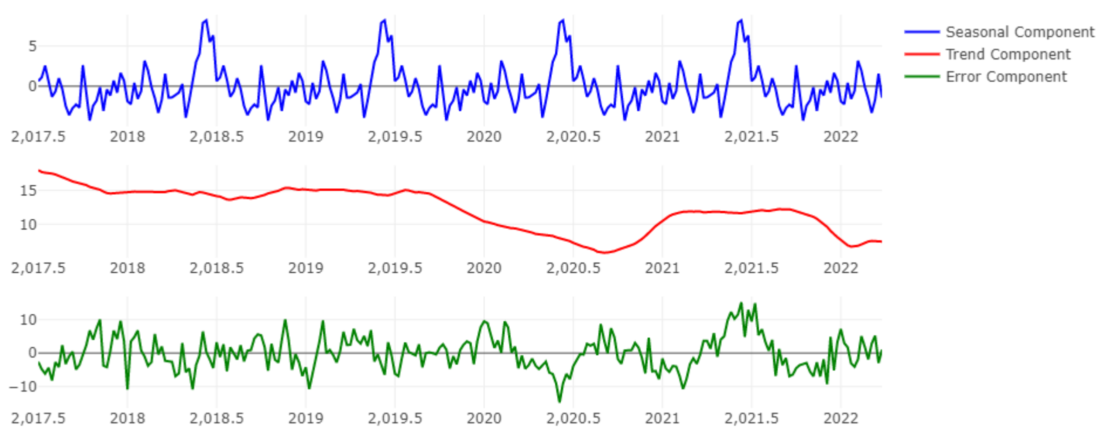

Figure 4. Decomposition plots

- Declining call volume trend (2017-2022), brief surge (mid-2020 - mid-2021).
- Annual seasonality observed.
- Volatility spike (March 2020 - August 2021), indicative of model limitations.
- Call volume spikes (specific hours, 2021 and 2022) likely due to COVID-19.

4. Baseline Model: Naïve

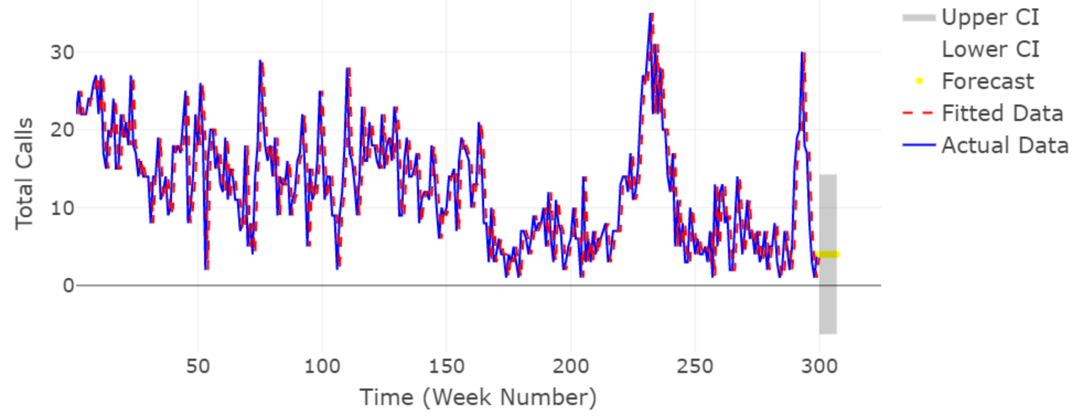

Figure 5. Holt's Linear Model

Description: Assumes future data will be the same as the last observed point. This serves as a simple baseline model. The model seems to fit well, but the forecast might not be accurate.

5. Extrapolation Models: Single Exponential Smoothing and Holt Linear

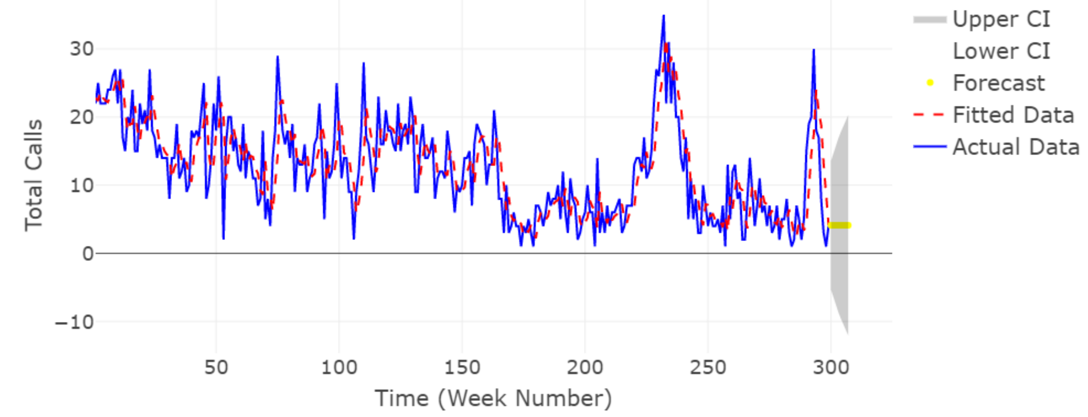

Figure 6. SES model

Description: Uses exponentially decreasing weighted averages of past data; assumes stationary data. Model seems to fit well to data, has similar forecast value as Naive above.

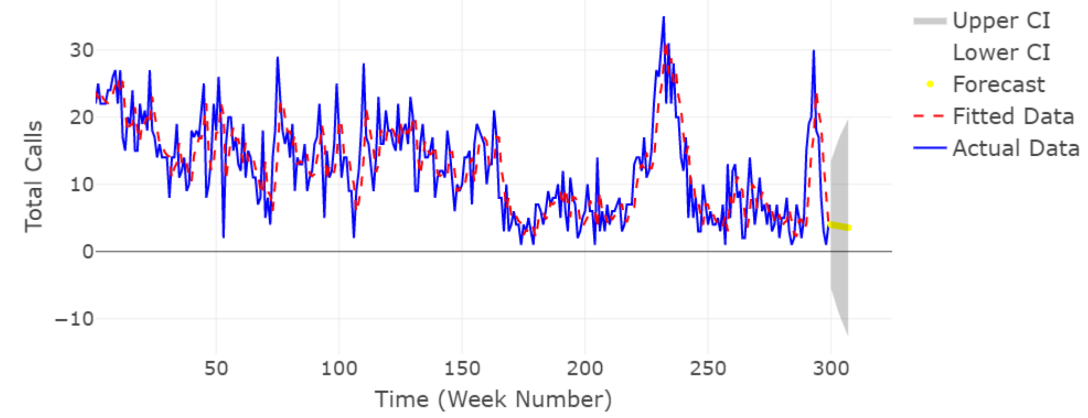

Figure 7. Holt's Linear Model

Description: An extension of SES for data with a trend; assumes constant variance over time. We observe the model has considered the trend from the forecast values.

6. Regression Model: Simple Linear Regression

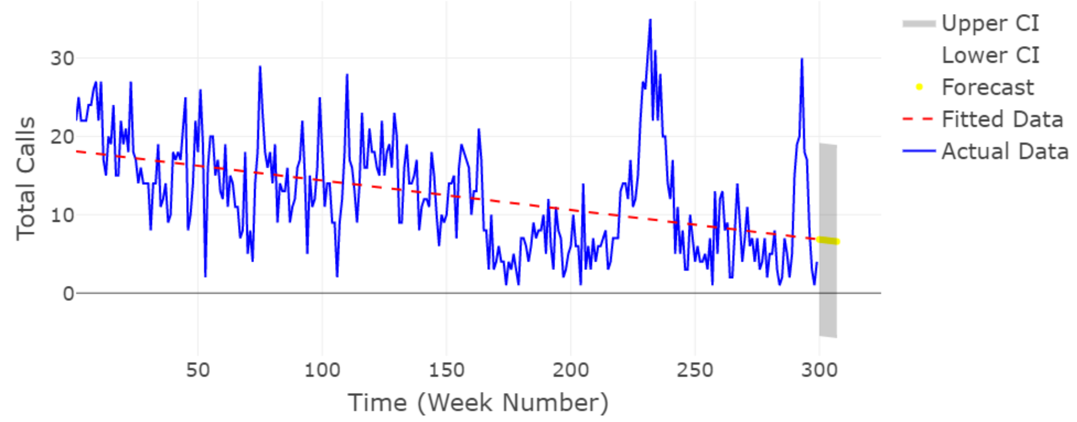

Figure 8. Simple Linear Regression

Description: Fits a linear equation to data; assumes a linear relationship, normally distributed errors and constant error variance.

7. Complex Models: ARIMA and TBATS

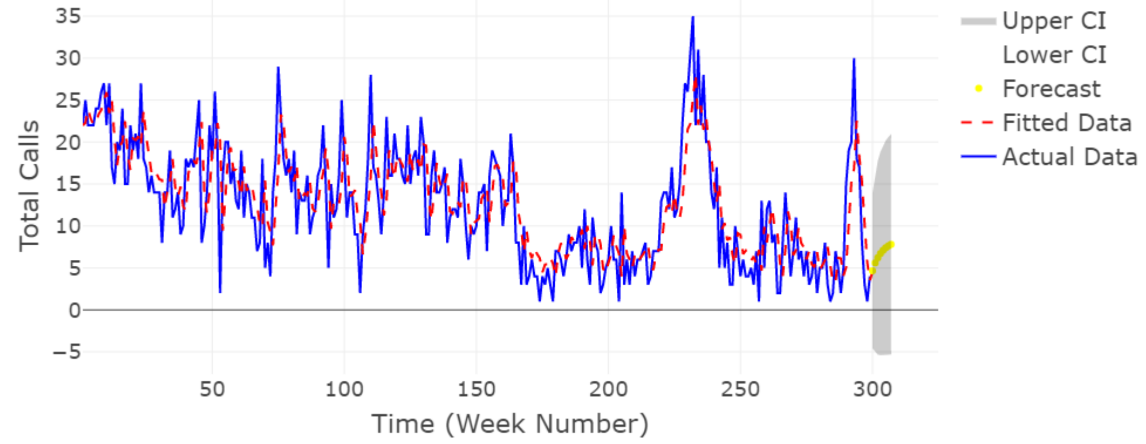

Figure 9. ARIMA Model

Description: Captures autocorrelations in time series data; assumes data is (or can be made) stationary. We observe that the ARIMA has considered some outliers and forecasted increasing values. The model looks promising.

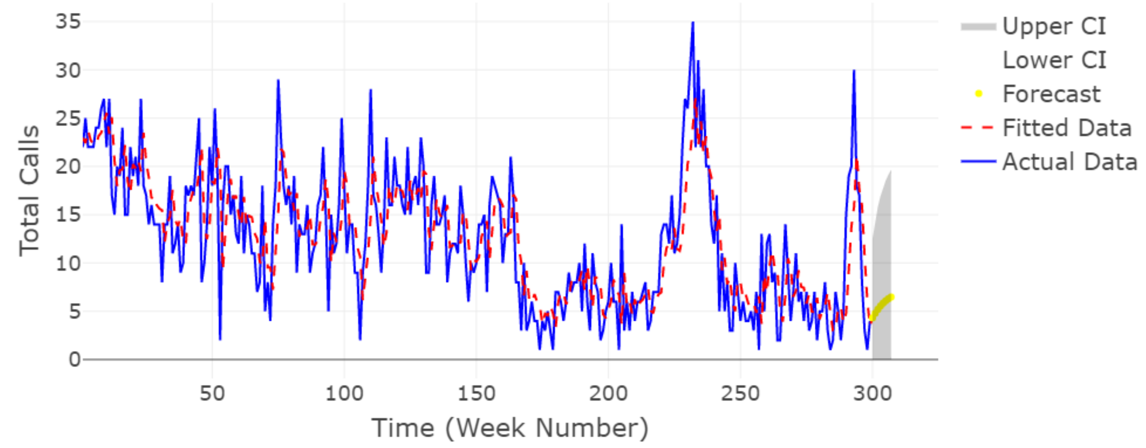

Figure 10. TBATS Model

Description: TBATS is an acronym for Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components. It's a complex model that can handle multiple seasonal patterns. Like ARIMA, TBATS seems to have considered the volatility of the data.

8. Model Evaluation and Forecast Summary

Model	MAPE_CV	RMSE_CV	MAPE_Fitted	RMSE_Fitted	Box-Ljung X-squared	p-value
Naive	1.190176	17.51009	0.5342048	5.220635	35.413	2.667e-09
SES	1.164609	16.55899	0.5172367	4.80116	0.72811	0.3935
Holt	1.124201	16.24298	0.5105186	4.800471	0.75907	0.3836
Simple Linear Regression	2.322275	21.30774	0.7604142	6.195157	125.7	< 2.2e-16
ARIMA	1.17969	16.42747	0.5498811	4.690951	0.055881	0.8131
TBATS	1.196133	16.52364	0.52854	4.710241	0.010195	0.9196

Table 1. Model Evaluation Metrics and Box-Ljung Test Results

While the Holt's model demonstrated superior forecasting accuracy in cross-validation, the TBATS model was chosen for its adeptness at capturing complex patterns, especially relevant considering the projected COVID pandemic impact on call volume. Other models, excluding Naive and Simple Linear Regression, displayed independent residuals, indicating satisfactory data capture.'

Week	300	301	302	303	304	305	306	307
Forecasted Call Volume	4.317445	4.790053	5.196109	5.542758	5.837262	6.086536	6.296913	6.474053

Table 2. 8 Week Forecasted Call Volume (TBATS)

10. Conclusion

Through our comprehensive study, we've derived actionable insights from Jakarta's 118 ambulance service call data. Despite Holt Linear model's high accuracy in cross-validation, the TBATS model was chosen due to its proficiency in handling complex patterns, a crucial factor given the unpredictable impact of the COVID-19 pandemic.

Our analysis reveals peak call volumes during early and late weekday mornings, informing optimal staff scheduling and resource allocation. Although our models fit well with historical data, ongoing monitoring and adjustments are essential due to potential pandemic influences.

By employing data-driven strategies, Jakarta's 118 ambulance service can enhance its operational efficiency, thereby improving public health outcomes within the city.