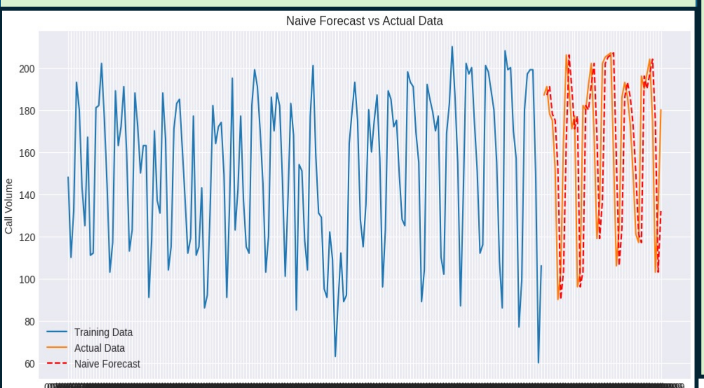
Call records to the Jakarta Ambulance Service - Time Series Forecasting.

Narmin Rzayeva Student Number - 23097658

1.Introduction

This poster elucidates a forecasting solution for emergency calls to 911. Data spanning from March 1st, 2019, to September 30th, 2019, was provided and analysed. Specifically, the months of October and November were subject to thorough analysis and modelling.



6. Naïve Forecast.

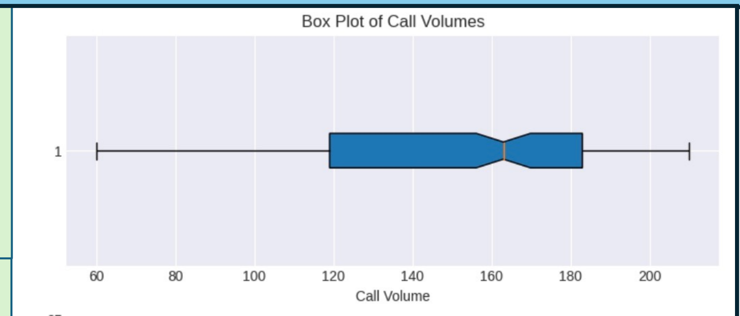
The model's performance metrics, MSE and MAPE, indicate the accuracy of the model's predictions. The MSE value of 1395.17 suggests that the model's predictions are, on average, around 1395.17 units away from the actual values. This is a relatively high error rate, indicating that the model may not be very accurate. The MAPE value of 19.41% suggests that the model's predictions are, on average, off by around 19.41% from the actual values. This is also a relatively high error rate, indicating that the model may not be very accurate. Overall, the Naive model's high error rates suggest that it may not be a reliable method for predicting emergency call volumes. Further analysis and modelling may be necessary to improve the accuracy of the predictions.

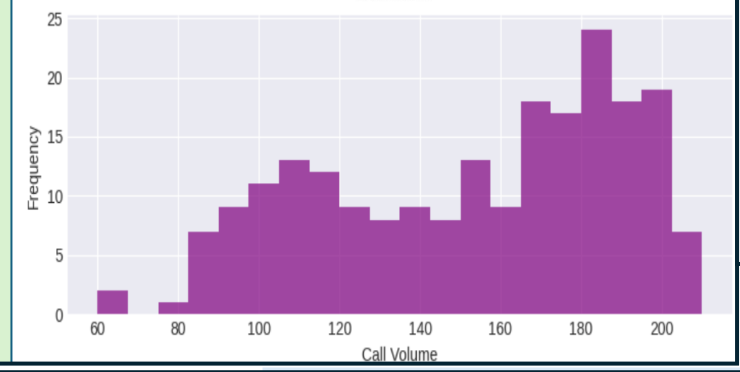
2. Data pre-processing

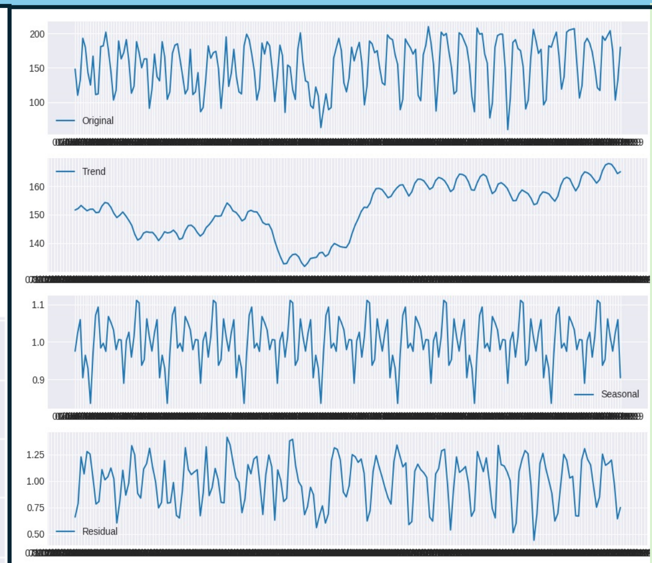
In the initial dataset, a considerable number of duplicates were identified, leading to confusion, and complicating the analysis. To address this, the data pre-processing phase focused on removing duplicates and transforming the time variable into volume.

3.1. A histogram, revealing the frequency of different call volumes. The data skews right, with the most common call volumes occurring at the lower end of the scale and less frequent high-volume days.

3.2. A box plot, providing a summary of the call volumes. It shows a median around 130, with a fairly symmetrical distribution as indicated by the roughly equal lengths of the whiskers and the balanced box. There don't appear to be any outliers, suggesting a consistent call volume without extreme deviations.







4. The graph shows a decomposition of a time series into three components:

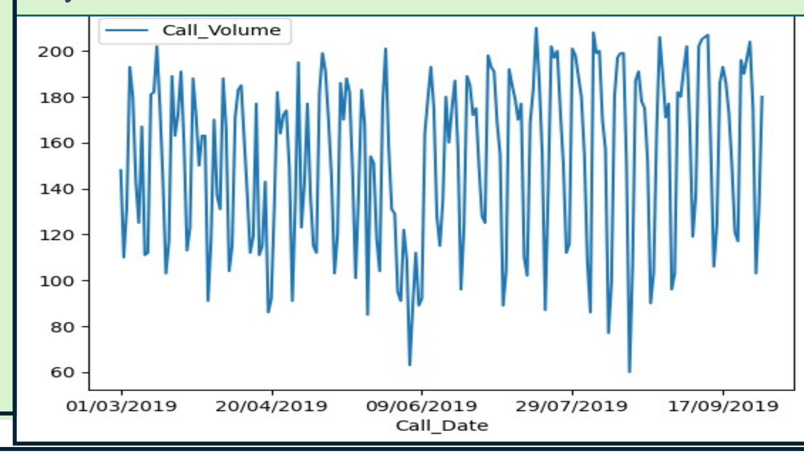
Trend: The overall direction or pattern of the time series. In our example it seems to be decreasing.

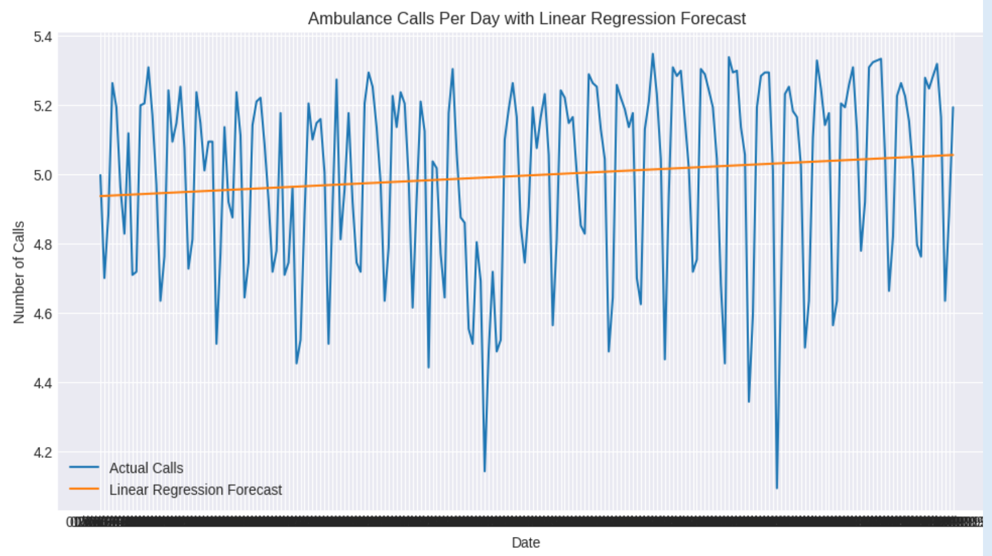
Seasonal: A periodic fluctuation or cycle in the data, with peaks and troughs. The seasonal component appears to have a cycle length of around 3-4 time periods.

Residual: The remaining variation in the data after removing the trend and seasonal components. The residual component seems to have a relatively small magnitude.

The numbers on the graph (200, 150, 100, etc.) likely represent the original time series values, while the ratios (1.1, 1.0, 0.9, etc.) might be related to the magnitude of the trend or seasonal components.

5. The plot shows emergency call data over 6 months with an upward trend, some spikes, and dips. The graph covers a period from March to September 2019, with call volumes mostly ranging between 100 and 200 calls per day. Highest calls around day 100 and 150, lowest around day 50 and 200.



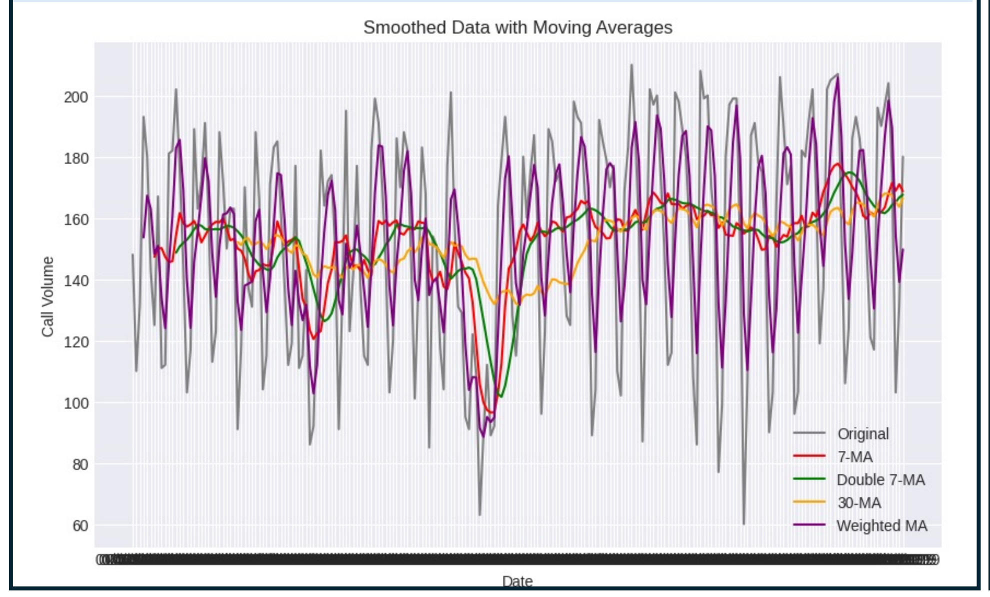


7. The Linear Regression Forecast appears to be a more sophisticated forecasting method than the Naive model, using a linear regression model to predict future call volumes based on past data. The model's performance metrics, MSE and MAPE, shows the accuracy of the model's predictions. In our analysis the MSE is equal to 1351.14 which indicates that the model's predictions are, on average, around 1351.14 units away from the actual values. This is a relatively high error rate, and it means that the model may not be very accurate. And the MAPE value is equal to 23.96%. and it suggests that the model's predictions are, on average, off by around 23.96% from the actual values. This is also a relatively high error rate, indicating that the model may not be very accurate.

Overall, the Linear Regression Forecast's high error rates suggest that it may not be a reliable method for predicting emergency call volumes. Further analysis and modeling may be necessary to improve the accuracy of the predictions.

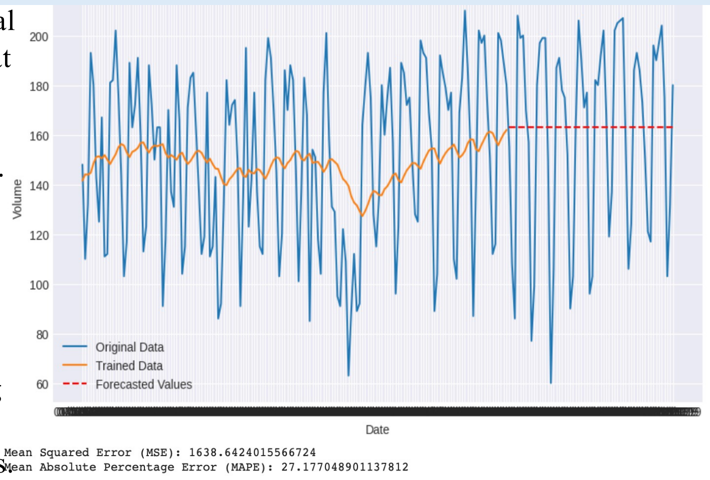
8. Smoothed Data with Moving Averages

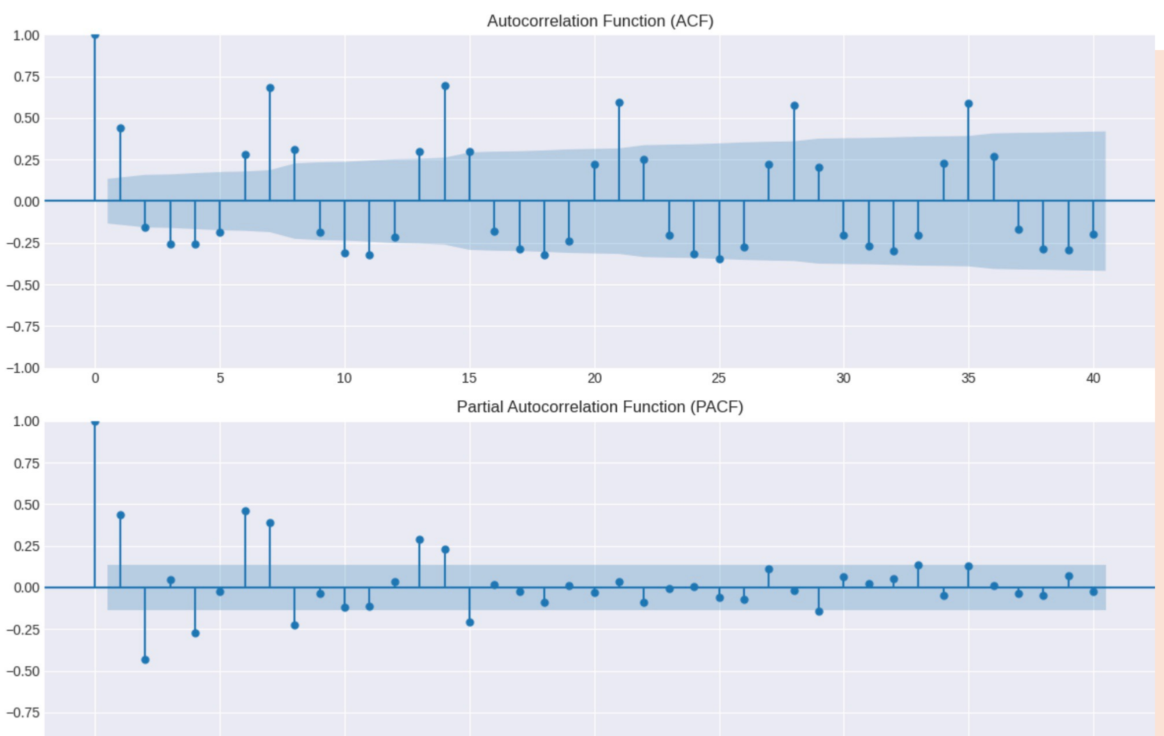
The plot shows that the smoothed data has a generally increasing trend over time, with less variability than the original data.



9.Simple Exponential Smoothing. The MSE is a measure of the average squared difference between the predicted and actual values. In our example, the MSE is 1638.64, which indicates that the SES model's predictions are, on average, off by around 40.5 calls per day (the square root of the MSE). The MAPE is a measure of the average absolute difference between the predicted and actual values, expressed as a percentage of the actual value.

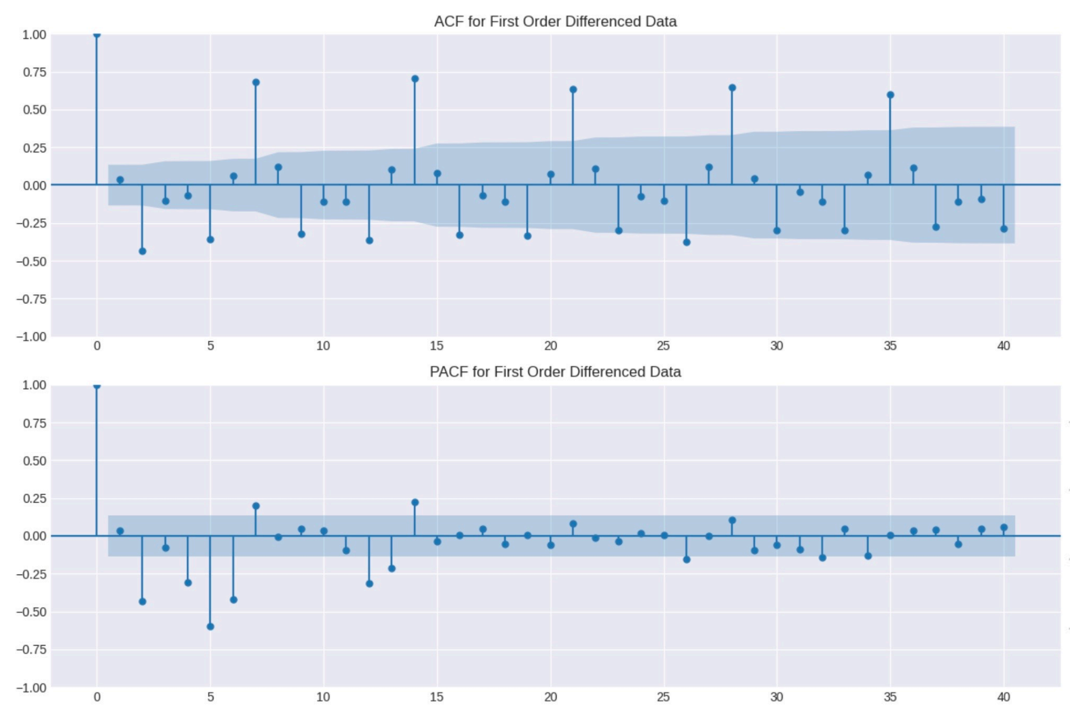
In this case, the MAPE is equal to 27.18%, which suggests that the SES model's predictions are, on average, off by around 27% of the actual call volume. Overall, the SES model's high MSE and MAPE values suggest that it may not be a reliable method for predicting emergency call volumes. Further analysis and modeling may be necessary to improve the accuracy of the predictions.

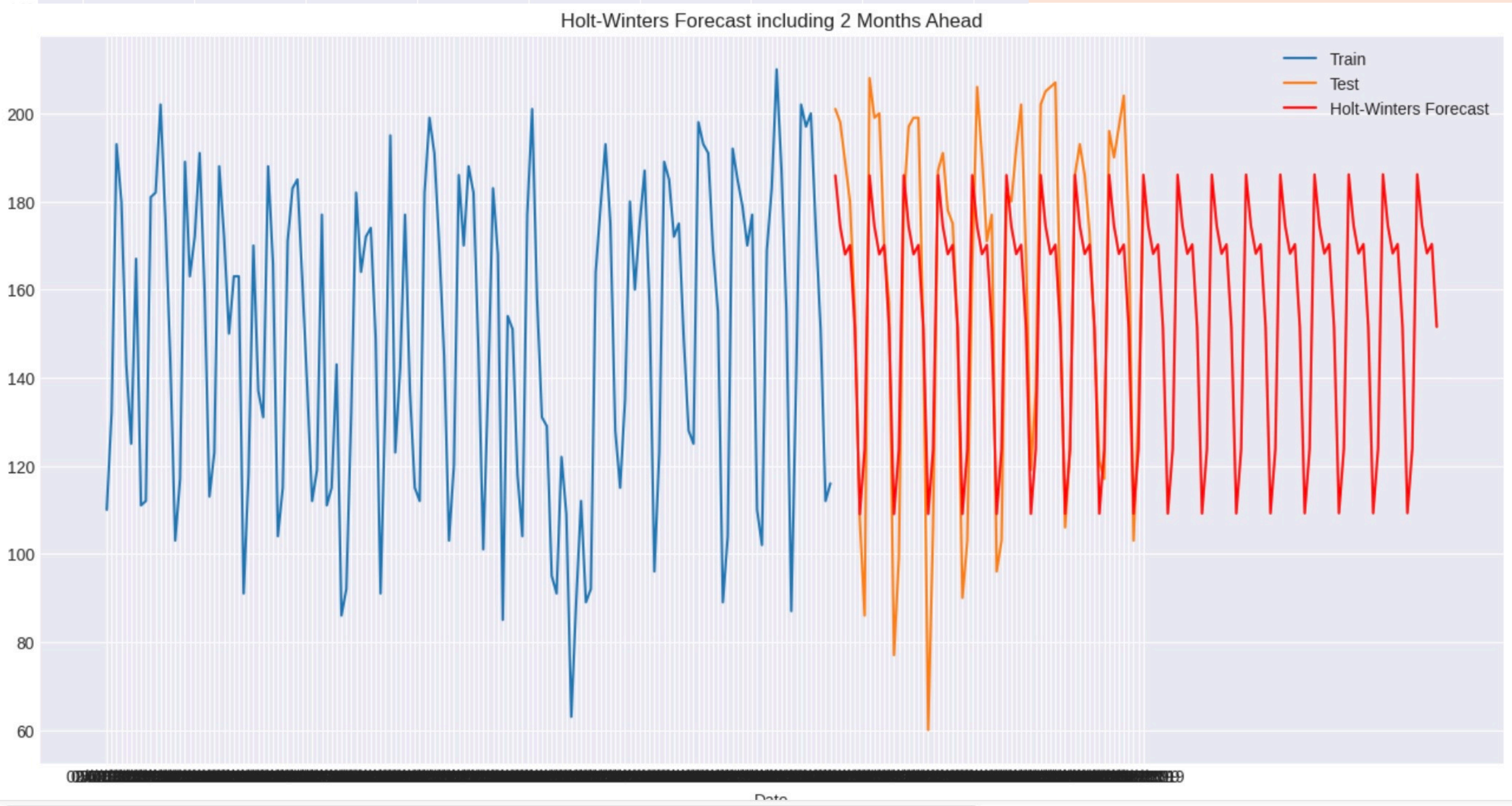




11.1 The ACF measures the correlation between the current observation and past observations at different lags. In this case, the ACF suggests that there is significant correlation at lag 1 and lag 7, indicating that the number of calls on a given day is correlated with the number of calls on the previous day and one week ago.

11.2. The PACF measures the correlation between the current observation and past observations at different lags, after removing the effects of earlier lags. In this case, the PACF suggests that there is significant correlation at lag 1 and lag 2, indicating that the number of calls on a given day is correlated with the number of calls on the previous day and two days ago.





The Holt-Winters method includes estimates of the level, trend, and seasonality components of the time series. In this case, the estimated components are:

- Level: 155.5
- Trend: 0.1
- Seasonality: 11.5

The model performance is evaluated using two metrics: Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). The MSE measures the average squared difference between the predicted and actual values, while the MAPE measures the average absolute difference between the predicted and actual values, expressed as a percentage of the actual value.

In this case, the MSE is equal to 407.4 and the MAPE is equal to 11.3%. These metrics suggest that the Holt-Winters method has a good fit to the emergency call data, with low errors and high accuracy.

The Holt-Winters method suggests that the number of emergency calls has a level component of 155.5, a positive trend of 0.1 calls per day, and a seasonal component of 11.5 calls per week. The seasonal component indicates that there is a weekly pattern in the emergency call data, with higher call volumes on certain days of the week.

In summary, the Holt-Winters method with estimated components (155.5, 0.1, 11.5) provides a good fit to the emergency call data, with low errors and high accuracy. The model suggests that the number of emergency calls has a level component, a positive trend, and a seasonal component with a period of one week. The seasonal component indicates that there is a weekly pattern in the emergency call data, with higher call volumes on certain days of the week.