IE5400 Spring 2019 Term Project
Natasha Jain, Neil Sanjaybhai Joshi, Veersinh Patil

Holt-Winters' Approach to ED Forecasting

1. Introduction

"Emergency" is defined as a serious, unexpected and often dangerous situation requiring immediate attention. It is worth noting the phrase 'requiring immediate attention' in the definition. Without the required attention being provided immediately, there is a high risk of adverse events. One of the major factors that plays an important role in not being able to provide immediate desired care is crowding, which stems from the mismatch between the existing capacity, various inputs, throughput, and the output factors. To avoid this mess resulting in worsening of patients situations, it is important to be able to forecast the patient demand in the emergency department which can lead to the proper allocation of resources.

2. Problem Definition

Given two years of patient admission data in a hospital emergency department, 2015 and 2016, we planned to forecast the following year, 2017, using a mathematical forecasting method. This model had to incorporate monthly seasonality, relative to the seasonality of the first year. It should had to separate out the triage levels in addition to including a forecast including all triage levels. Finally, in order for our forecast to be successful, it had to be timely, reliable, accurate, incorporate meaningful units, and be easy to use.

3. Data Description

The given data includes five measures. The "Date" is the date the patient is admitted to the Emergency Department in a "yyyy/m/d" format. The gender is given as a binary value where 0 indicates the patient is female, and 1 indicates the patient is male. The Triage column contains ordinal data that indicates the severity of the patient's condition with 1 being most severe, and 5 being least severe. Triage Time is a ratio scale of data that indicates the exact time the patient enters the system in a "yyyy/m/d hh: mm" format. Leave ED Time is also a ratio variable that indicates the exact time the patient leaves the system in the same format. There are two sheets, one with the 5 measures described for 2015 and the other for 2016.

4. Proposed Methods

4.1 Data Preprocessing

The excel data sheets had to be processed to make this data usable. First, we consolidated the 2016 and 2015 data into one sheet. Then, "Length of Stay" was calculated by subtracting the "Leave ED time" from the "Triage Time" column. This left us value in a "hh: mm" format, which we converted to minutes in the "LOS (min)" column by multiplying that column by 1440. We also used formulas in Excel to separate out the month and year from the data to make it easier for us to sort and categorize the data when using a Pivot Table.

This data was then put into a Pivot Table in order to consolidate all the patient information for each month. The Year and Month were selected as Rows. The Count of Date, which counts how many rows there are of a particular date, and the average "LOS (min)" was selected as the values for the pivot table. We also set the "Triage" column as a filter so that each triage level could be separated out. As a result, the pivot table shows the number of patients each month, as well as the average length of stay of a single patient in each month.

The data was then separated into twelve sheets: All Triage Levels Volume, Triage Level 1 Volume, Triage Level 2 Volume, Triage Level 3 Volume, Triage Level 4 Volume, Triage Level 5 Volume, All Triage Levels LOS, Triage Level 1 LOS, Triage Level 2 LOS, Triage Level 3 LOS, Triage Level 4 LOS, and Triage Level 5 LOS.

4.2 Holt-Winters' Method of Triple Exponential Smoothing

Holt-Winters' Method is a type of forecasting methodology that includes a base level, trend, and seasonal index (Chatfield 1988). It was chosen because according to our literature review, it performed well against other forecasting models, even compared to more complex ARIMA models. Unlike the Box-Jenkins method, it is an automatic approach that does not require an analyst's subjective input to output results. As this is a very limited data set, and we are unfamiliar with the system this dataset describes, it is best for us to use an approach that is objective, thus not requiring our input.

To begin, we first initialized the seasonal indices. This was done by taking each month's 2015 value and dividing it by the average of the twelve months in 2015.

$$S_t = \frac{Y_t}{average(Y_1 : Y_{12})} \qquad S_t = \text{seasonality for month } t, \quad Y_t = \text{demand in month } t$$

The next step was to initialize the Base Level (L) for month n=13, or January of 2016, for which we used the below formula. This formula essentially deseasonalized the January 2016 data using the seasonality index from January 2015.

$$L_{13} = Y_{13}/S_1 \qquad L_t = \text{base level for month } t$$

Then, we initialized the Trend (T) for January of 2016 using the below formula. This formula finds the trend by using the difference between the current period and the last period's deseasonalized values.

$$T_{13} = \frac{Y_{13}}{S_1} - \frac{Y_{12}}{S_{12}} \qquad T_{13} = \text{Trend for month } t$$

Once we have the initial values for the Level, Trend, and Seasonal Indices, we can begin using the Holt-Winters' Method formulas as follows to forecast the results of 2016.

$$L_t = \alpha(Y_t/S_{t-12}) + (1-\alpha)(L_{t-1} - T_{t-1}) \qquad \alpha = \text{base level smoothing constant}$$
$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \qquad \beta = \text{trend smoothing constant}$$
$$S_t = \gamma(Y_t/L_t) + (1-\gamma)S_{t-12} \qquad \gamma = \text{seasonality index smoothing constant}$$
$$F_t = (L_{t-1} + T_{t-1})S_{t-12} \qquad F_t = \text{Forecast for month } t$$

For each 2017 value, we calculated the mean absolute percent error (MAPE) using the following formula and taking the average of these percentages.

$$E_t : |Y_t - F_t|/Y_t \qquad E_t = \text{Absolute percent error for month } t$$

Because Triage Level 5 contains some 0 values for actual Volume and Length of Stay, we used Mean Squared Error (MSE) for the error measure for that triage level, as MAPE would result in a division by zero error.

$$E_t : (Y_t - F_t)^2$$

We then used a simple optimization function implemented in Excel Solver to find the optimal values for alpha, beta, and gamma to minimize the MAPE (or MSE for Triage Level 5).

$$Objective : \ Minimize\ \Sigma E_t/n \text{ ; n=number of months}$$
$$Subject\ To : \quad 0 \leq \alpha \leq 1$$
$$0 \leq \beta \leq 1$$
$$0 \leq \delta \leq 1$$

Finally, we forecast the 2017 values using the below formula.

$$F_{t+k} = (L_t + k * T_t) * S_{t-12} \qquad \text{; k= the number of periods ahead of the last known demand}$$

This process was repeated for each triage level for both volume and length of stay.

5. Experimental Results

After using the methodology described above, we were able to plot the following twelve graphs. The blue line indicates the actual given volume or length of stay, while the orange indicates the optimized forecast using the Holt-Winters' model.

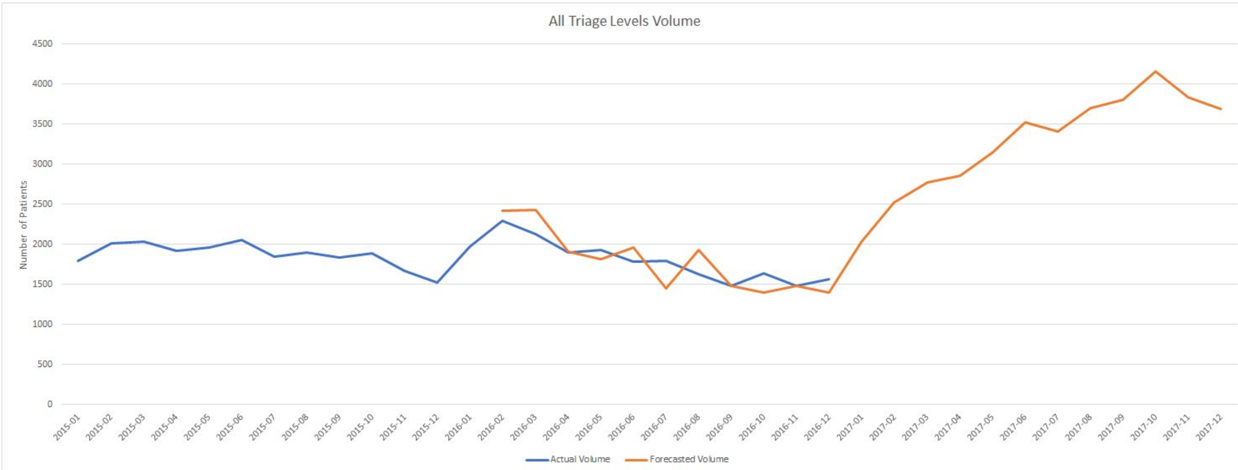

Figure 1: Volume for all Triage Levels

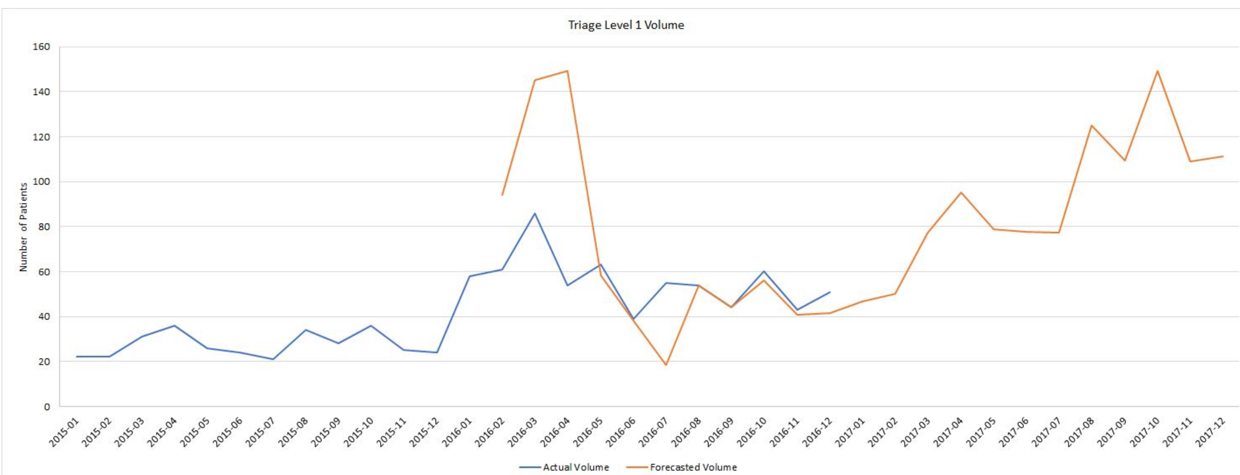

Figure 2: Volume for Triage Level 1

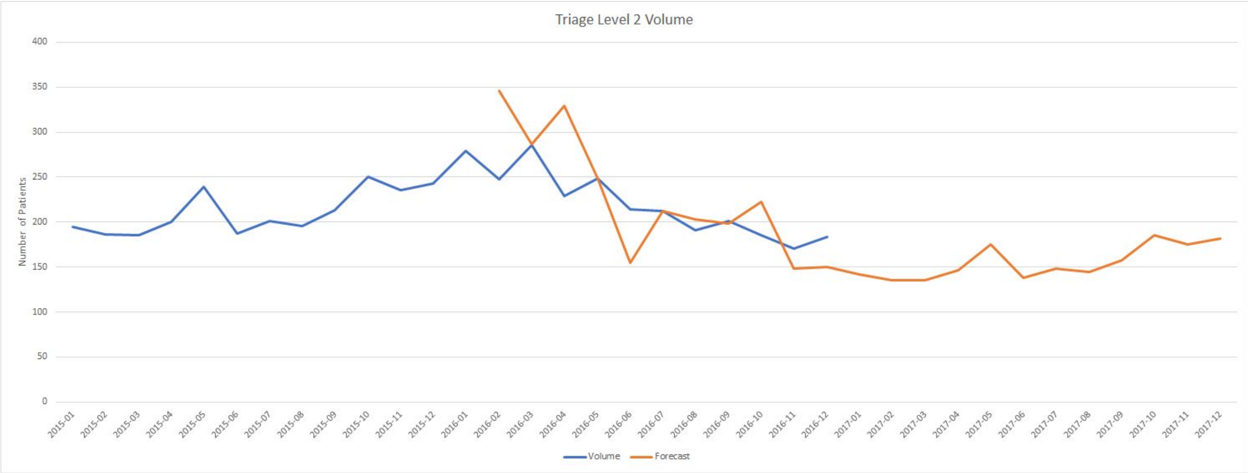

Figure 3: Volume for Triage Level 2

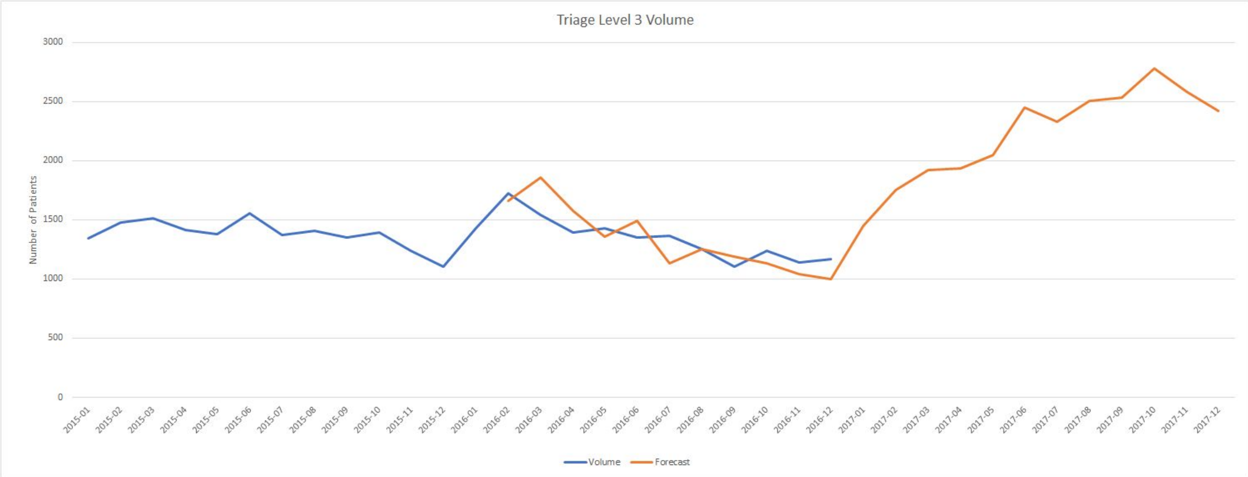

Figure 4: Volume for Triage Level 3

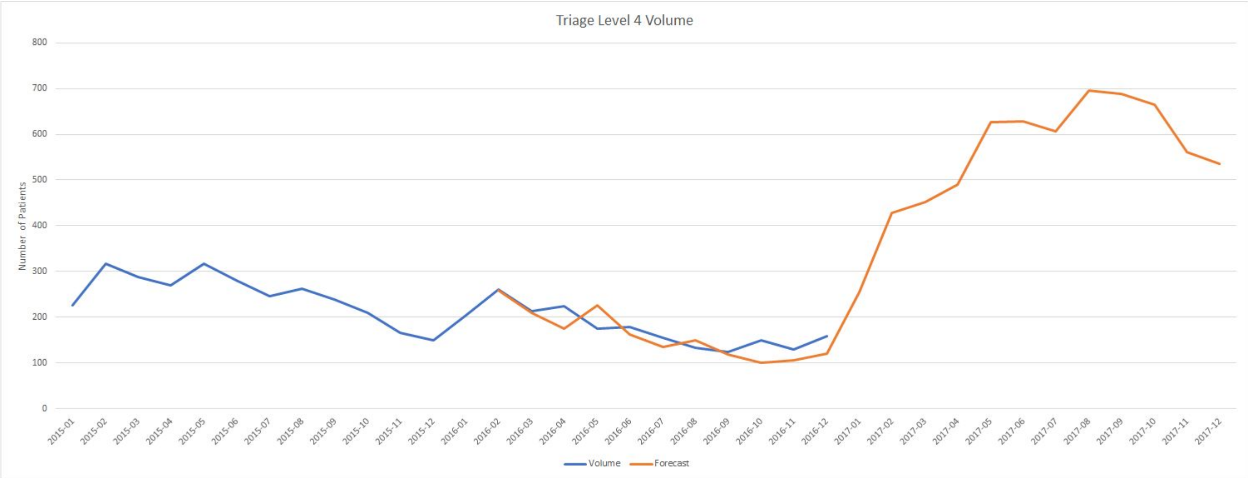

Figure 5: Volume for Triage Level 4

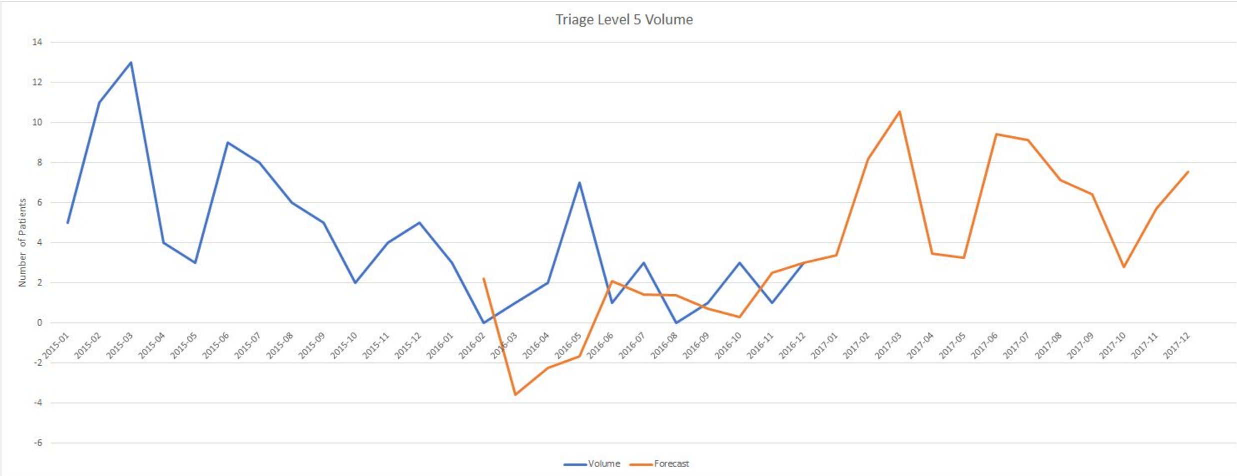

Figure 6: Volume for Triage Level 5

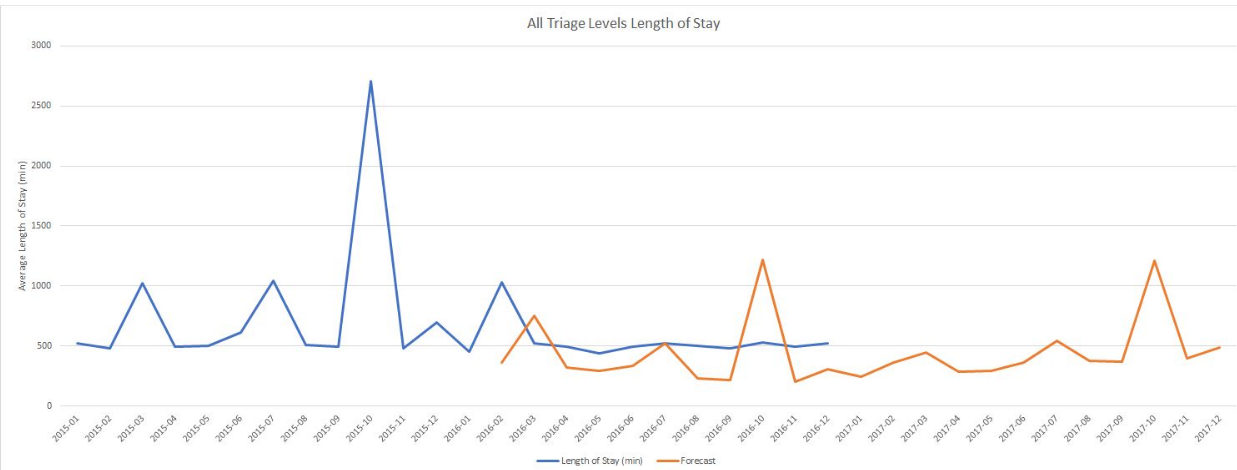

Figure 7: Average Length of Stay for All Triage Levels

Figure 8: Average Length of Stay for Triage Level 1

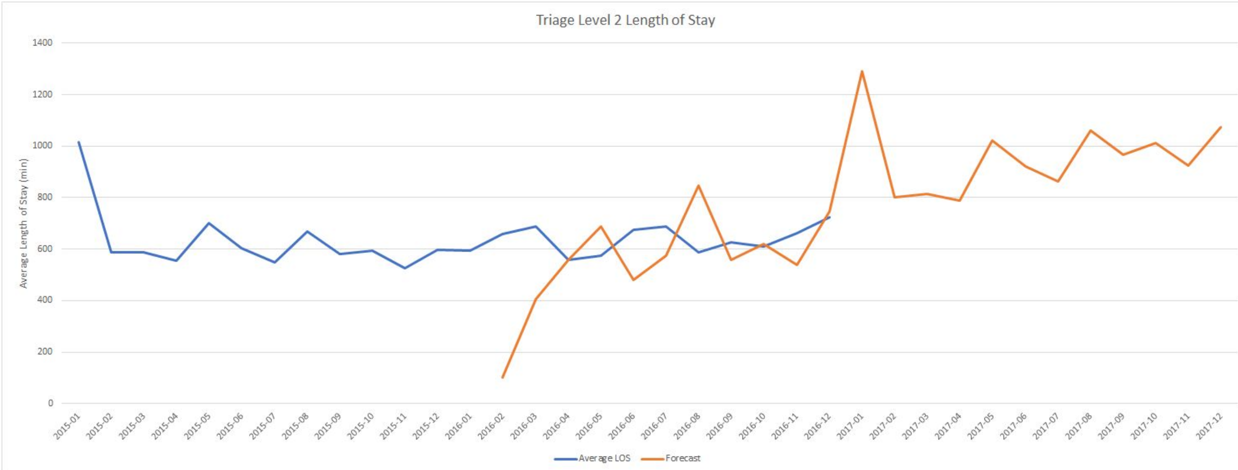

Figure 9: Average Length of Stay for Triage Level 2

Figure 10: Average Length of Stay for Triage Level 3

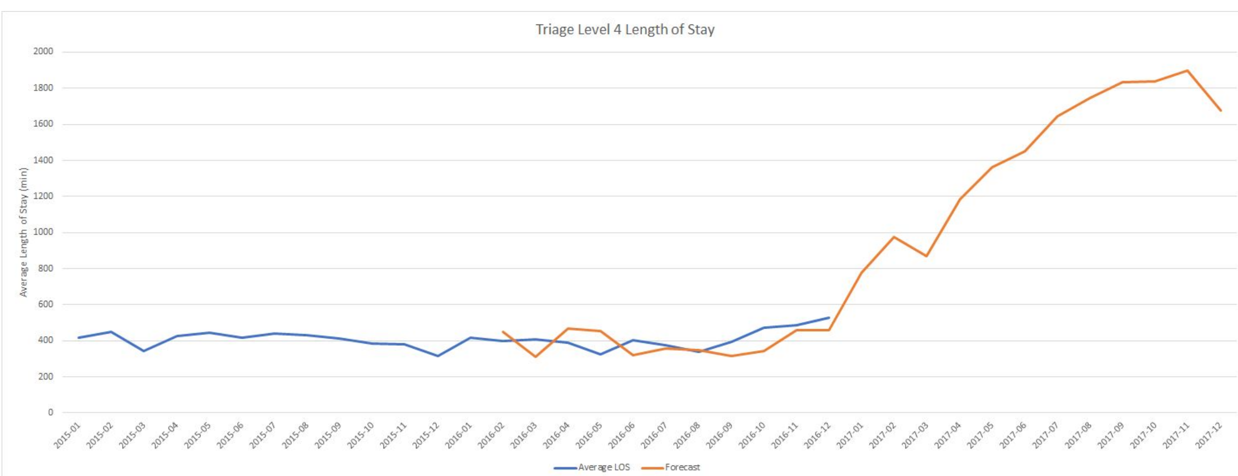

Figure 11: Average Length of Stay for Triage Level 4

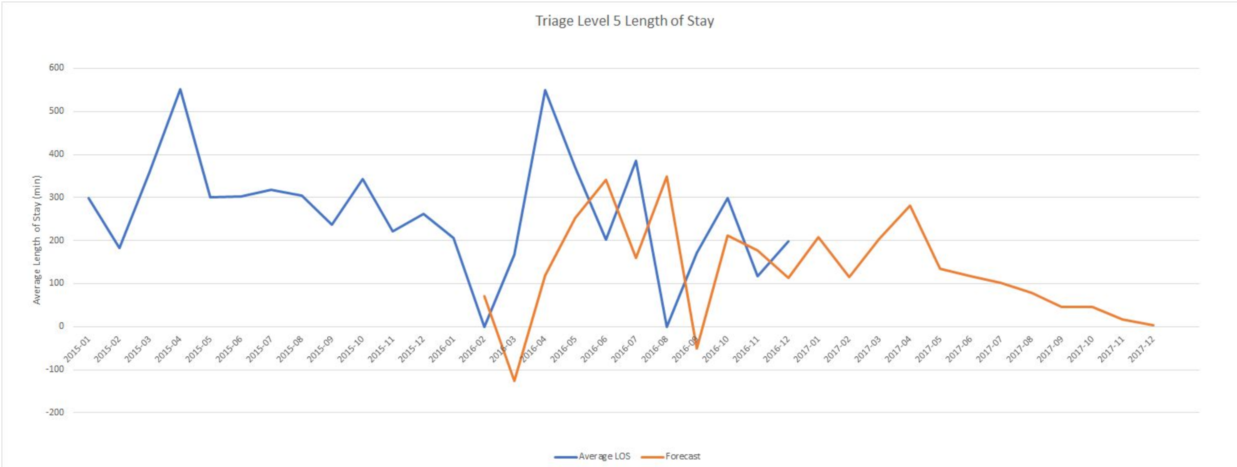

Figure 12: Average Length of Stay for Triage Level 5

The most glaring error of course is that three of these forecasts (Triage Level 5 Volume, Triage Level 1 LOS, and Triage Level 5 LOS in Figures 6, 8, and 12 respectively) have predicted negative values, which does not make sense for this system as there cannot be negative patients or a negative length of stay. This is likely due to the trend being calculated as highly negative, which is only amplified by a seasonality greater than one. Below is a summary of the errors for each forecast optimized to fit the 2016 data. Triage Level 5 also had 0 values, which may have caused errors in the calculations. Again, for Triage Level 5, we used MSE instead of MAPE due to those zero values leading to errors.

Table 1: Forecasting Error Summary

	All Triage Levels	Triage Level 1	Triage Level 2	Triage Level 3	Triage Level 4	Triage Level 5
Volume MAPE	0.09	0.37	0.16	0.16	0.15	-
Length of Stay MAPE	0.50	3.32	0.24	0.66	0.17	-
Volume MSE	-	-	-	-	-	12.19
Length of Stay MSE	-	-	-	-	-	50010.01

As we can see in Table 1, Triage Levels 2 and 4 were able to be predicted best using triple exponential smoothing for 2016 at average mean absolute percent error rates of 20% and 16% respectively. Triage Level 1 Length of Stay had an especially poor fit to the actual forecast, and would likely not be used in actual hospital implementation. Both Triage Level 5 Volume and Length of Stay forecasts would not be implemented as well, due to the high error rates and previously mentioned negative values.

6. Conclusion

A key tenant of forecasting is that the past continues into the future, which may not be a correct assumption in the given data. The given data is highly irregular, and although we calculated the seasonality using a monthly index, it does not appear as if there is a true pattern occurring on a monthly basis. This results in a seasonality index falsely predicting large increases and decreases. The data is also very limited, as it only includes two years, so perhaps a stronger pattern would be found in the data spanned several years prior. Because the forecast is fit to the 2016 data, and the data from 2016 does not appear to follow the seasonality of the 2015 data, it is difficult to be confident that the "future" 2017 forecast will be accurate.

Out of all the forecasts, we would feel most confident in the hospital adding or removing resources based on the forecasts of Triage Level 2 or 4, as their forecasts had low mean absolute percentage error rates, and did not include negative values. If we were to present the rest of the findings to a client, we would need to ensure they understand the forecasts' limitations, and possibly suggest a different measure of seasonality (day of the week or quarterly) to check if that leads to more reasonable results.

The five characteristics of a good forecast are timeliness, reliability, accuracy, meaningful units, and ease of use. We are confident that this forecast is easy to use, as the formulas for the forecast can easily be extended, and Excel is a widely accessible software, thus not requiring an organization to spend additional resources to forecast. We also feel that volume and length of stay are meaningful units, as they can be used together to estimate the bed capacity and staff (resources) needed each month. The forecast is also timely, as a forecast of one year ahead gives the system enough time to implement possible changes based on our findings. As mentioned previously, however, due to the data limitation and forecasting method, the forecast is not extremely reliable. The amount of error varied widely depending on the forecast from 9% (Volume of All Triage Levels) to 332% (Length of Stay of Triage Level 1). The forecast is also not accurate for all levels, as some forecasts had high errors and/or infeasible values. Accuracy and reliability are arguably the two most important factors in designing a good forecast, so as previously stated, further analysis is needed before implementing the results of the forecast.

7. References

1. Chatfield, C., & Yar, M. (1988). Holt-Winters Forecasting: Some Practical Issues. *Journal of the Royal Statistical Society. Series D (The Statistician), 37*(2), 129-140. doi:10.2307/2348687
2. Calegari, R., Fogliatto, F., Lucini, F., Neyeloff, J., Kuchenbecker, R., & Schaan, B. (2016). *Forecasting Daily Volume and Acuity of Patients in the Emergency Department*. Computational and Mathematical Methods in Medicine
3. Jones, S., Thomas, A., Evans, S., Welch, S., Haug, P., & Snow, G. (2008). *Forecasting Daily Patient Volumes in the Emergency Department*. Society of Academic Emergency Medicine
4. Hertzum, M. (2018). *Forecasting Hourly Patient Visits in the Emergency Department to Counteract Crowding*(Vol. 11). The Ergonomics Open Journal.