Dulari Narayanar

Forecasting Public Transportation Demand

**Forecast Report**

**Final Project (PASS)**

Contents

[**Executive Summary** 2](#_Toc79744430)

[Problem Description 2](#_Toc79744431)

[Data Analysis 2](#_Toc79744432)

[Proposed Model and Forecasting Method 2](#_Toc79744433)

[Conclusion & Recommendations 2](#_Toc79744434)

[**Technical Summary** 3](#_Toc79744435)

[Data Analyzed 3](#_Toc79744436)

[Forecasting Method 3](#_Toc79744437)

[a) Forecasting Selection Process 3](#_Toc79744438)

[b) Forecast Method and Equations 4](#_Toc79744439)

[c) Model Performance Evaluation 5](#_Toc79744440)

[d) Forecast Results and Learnings 7](#_Toc79744441)

[e) Result Conclusion 8](#_Toc79744442)

[**Appendix** 8](#_Toc79744443)

[Appendix A: Data Analysis and Time Plots 8](#_Toc79744444)

[Appendix B: Saturdays and Sundays Data Training, Validation and Forecasted Plots 12](#_Toc79744445)

[Appendix C: Weekdays Data Training and Validation Residual Plots and Errors - Trend Holt-Winter Smoothing -No Trend Method 14](#_Toc79744446)

# **Executive Summary**

## Problem Description

A public transportation company, the prime stakeholder would like to plan and acquire new buses and extend its terminals based on the forecasted demand. This report elaborates on the analysis and modelling performed to reliably forecast a future 3-day local transport demand based on existing 3 weeks of data received from the company’s data warehouse.

## Data Analysis

The data obtained showed different patterns on different times in a week and thus was partitioned into 3 parts namely Weekdays, Saturdays, and Sundays and smoothened in 2 different ways for further analysis.

## Proposed Model and Forecasting Method

The Saturday and Sunday data were smoothened using Trailing Moving Average Method and the Weekdays were smoothened using Trend Holt-Winter Smoothing - No Trend Method in XLMiner prior to forecasting. Trend Holt-Winter Smoothing was again used to forecast based on trained data for the 3 days as the method ideally uses the most recent historical data and assigns more weight in forecasting than the older results. This was crucial in ensuring a more accurate predicted value.

## Conclusion & Recommendations

The Actual and forecasted values, residuals and MAPE and MAE were evaluated on the final forecasted model, and the method was found to have high forecasting accuracy and suitable for forecasting the 3-weekdays.

The Forecasted 3-day values do indicate that the public transportation company should investigate increasing their bus service during weekdays during the period of 12pm to 2pm and 6pm to 8 pm.

Although the final forecasted values and demand pattern were in line with what was observed in the last 2 weeks, the reliability of the forecasting model can be further improved if the following points are taken note during data collection:

1. The forecast model was based on 2 weeks of training and 1 week of validation which is only 9 seasons for training and 5 seasons for validation. The model was not able to capture any trend or seasonality that may have been observed if a few months of data was used. Moreover, basing on a larger training and validation period would achieve a more accurate forecasting model. A larger data set would thus be useful to revalidate the results of this report, to ensure the model encompasses any trend or seasonality which was not observed in the smaller data.
2. Due to the limited data period used for modelling, it is crucial that every data value provided is accurate and does not have any errors, outliers, or null values.
3. The reliability of the data is important and thus should be collected from the same source with the same measurement methods.
4. There is no detail in the given dataset on whether any of the days were public holidays or any other events occurred which could have affected the demand on a particular day or time. This would have a direct impact on the modelling method chosen and the forecasted values.

# **Technical Summary**

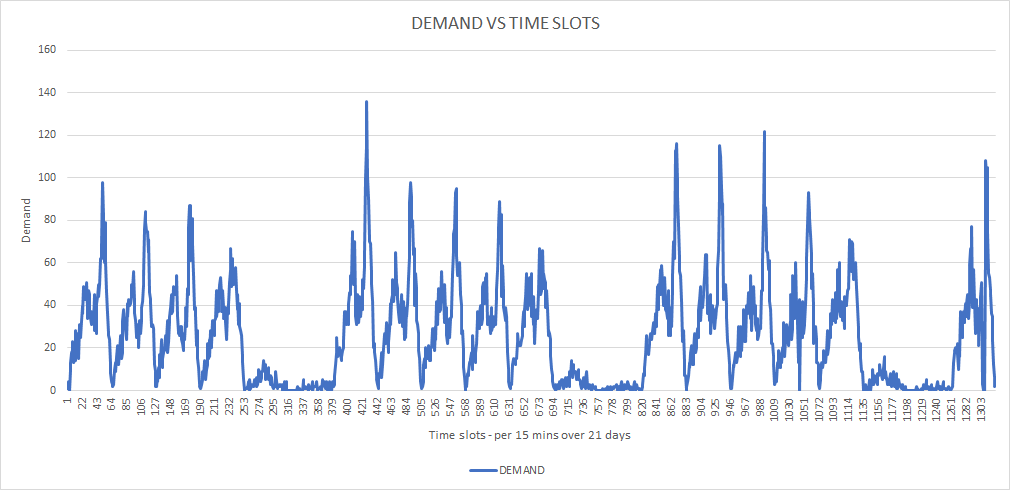
## Data Analyzed

* The data provided for forecasting was for 21 days from a period of 6:30 am to 10 pm at 15 min intervals.
* The Actual source and means of collection of the Data is not specified and the reliability of it is in question. It is crucial the data was all based on the same source to ensure the same level of accuracy.
* The Data was reasonably clean with no missing values.
* There were some null values especially during the weekends which can affect the lower margin of the forecasts.
* The following analysis assumes that all data provided can be used for the forecast modelling and the data does not include any pattern which is the effect of an unusual event which happened during the data collection period and would not happen on a regular basis.

## Forecasting Method

### **Forecasting Selection Process**

**Figure A: The time series plot of every 15 mins over the 21 days was plotted versus the daily demand**



The time plot from the given data clearly shows no trend with an additive seasonality on weekdays on certain timings of the day. But to further analyze the data before selecting the apt forecasting method, other plots of the data were created to see possible trends and seasonality. Details of the plots can be seen in Appendix A.

With reference to Appendix A, we can observe the following:

Trend

* 3 different patterns namely a pattern per hour for Weekdays and one for Saturdays and for Sundays are noted. (Refer to Appendix A: Figure 1, 2 & 5).
* The Weekdays hourly pattern doesn’t follow any trend and is rather horizontal and can be deduced as no-Trend.
* The Saturdays and Sundays seems to follow no trends as well, but the Saturday does have a slightly higher demand than Sunday.
* The demand is higher during the weekdays in comparison to Saturdays and Sundays. (Refer to Appendix A: Figure 3).

Seasonality

* There is a seasonality for weekdays peaking at 1.30 pm and 6.45 pm and dropping at 3.30pm. This could be indicating the lunch hour rush as well as the evening rush after work. The seasonality is of global nature over weekdays. (Refer to Appendix A: Figure 1, 2 & 5).
* From Figure A, it is observed that Saturdays and Sundays has a slight seasonality, but for simplicity we shall assume the seasonality is minimal.

Data Split

* To capture the seasonality during weekdays, the data was split into 3 parts namely weekdays, Saturdays, and Sundays.

### **Forecast Method and Equations**

Saturdays and Sundays Data Training, Validation and Forecasted Values

Saturday and Sunday data were extracted and recorded in separate worksheets for Forecasting separately. The Trailing Moving Average Smoothing method was used with the corresponding values of the series (Saturday or Sunday every 15 mins) as it lacked trend and seasonality.

The training period used for both Saturday and Sunday were the first 2 weeks and the Validation period used was the third week. Separate Training, Validation and Forecasts were done for Saturday and Sunday as the demand values were slightly higher on Saturday than Sunday.

The forecasted values were calculated as follows for every season:

* + For Training Period: For the first 2 weeks, the average of the values of the same day and time was taken.
    - Ft+k =
  + For Validation Period: The same values of first week as calculated above was repeated
  + For Future Forecast: For the first 3 weeks, the average of the values of the same day and time was taken.
    - Ft+k =

The Actual and Validated timelines and residuals were plotted, and it showed that the Actual and forecasted values showed pretty good representation and the residuals were well balanced (refer to Appendix B).

Weekdays Data Training and Validation

* The Weekday data after removing Saturday and Sunday data was recorded in a separate worksheet for partitioning.
* Data was partitioned 2 weeks for training period and 1 week for Validation period.
* Since the Weekday Data is of no Trend, additive nature, **a Trend Holt-Winter Smoothing - No Trend Method in XLMiner** is used on the partitioned data.
* Two versions were created, one with an alpha=0.2 and another version which was optimized. The plots and error measures are shown in detail in Appendix C.

The following observations were made from the plots:

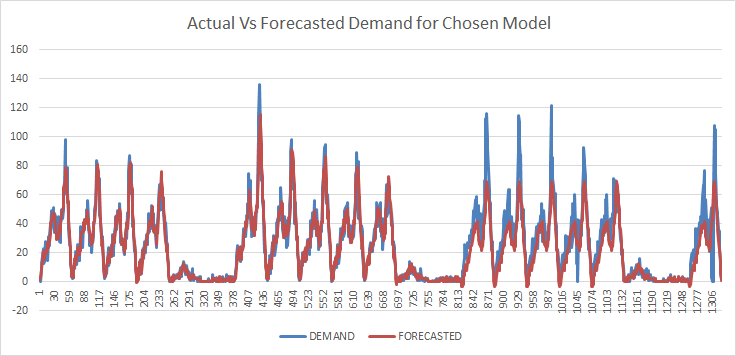
* + The Actual and Forecasted training data for the alpha0.2 and optimized versions fitted reasonably well.
  + Comparing Actual and Forecasted Validation data, the optimized version’s forecasted plot seemed to have less of a lag in comparison to alpha 0.2
  + The Residual validation data of the optimized version is slightly lower and nearer to the 0 level in comparison to the alpha 0.2 plot
  + The MAPE (42.09) and MSE (196.94) of the Optimized model was lower than The MAPE (54.25) and MSE (236.85) of the alpha 0.2 model.

Since the optimized version of Trend Holt-Winter Smoothing method showed better results, the rest of the forecasting was based on the values obtained from this.

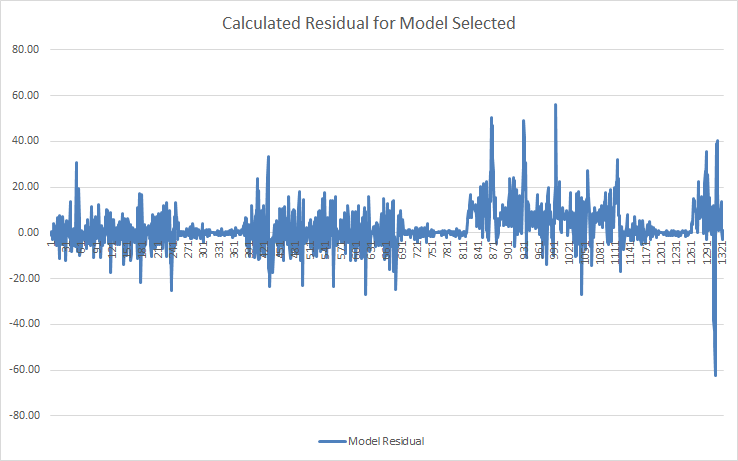
### **Model Performance Evaluation**

* The forecasts obtained in the 3 different periods as described above are combined and the Actual and forecasted value plot and residual plots were plotted to assess the chosen model.
* The MAE and MAPE for the Combined Model were calculated.

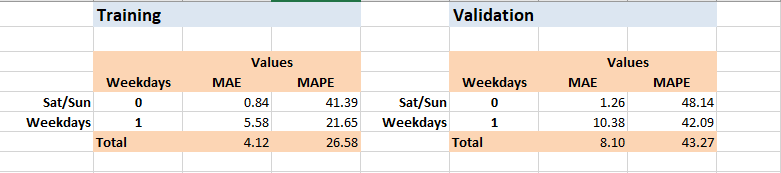
**Figure B: Actual Vs Forecasted Plot of Model**



**Figure C: Residual Plot**



**Figure D: MAE and MAPE values**

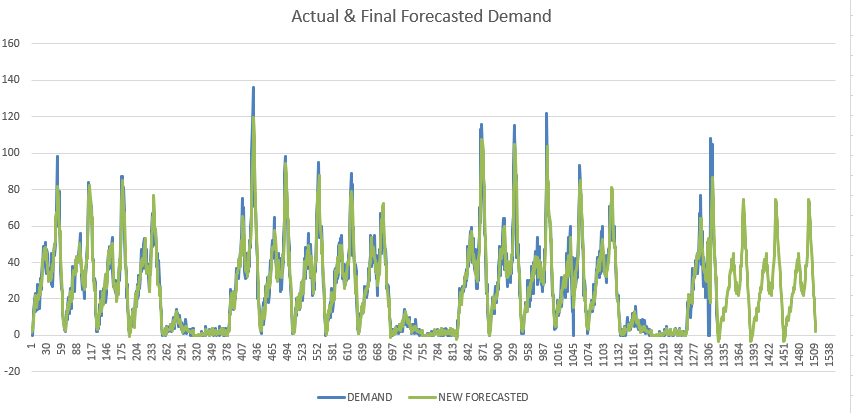


The Actual and Forecasted plots generated reasonably good forecasts for each day and time slot, and the residual plots showed a good balance along the 0 axis. The MAPE and MAE also had low values, indicating good forecasting accuracy. Based on these results, it was decided to use the Trend Holt-Winter Smoothing - No Trend Method in XLMiner to predict the next 3 days demand forecast.

### **Forecast Results and Learnings**

The XLminer Holt-Winter Smoothing - No Trend Method only predicts up to 100 values in XLminer. It was run once on the combined weekday original data for 3 weeks. The first 100 values forecasted was repeatedly copied in the table until all the 3 days forecast values were reached. Instead, we can try to run the model again to achieve 3 days’ worth of forecasts, but it may over smooth the previous values on each run.

**Figure E: Actual Demand plus forecasted 3 days Demand**



**Validated Period**

**Forecasted Section**

**Training Period**

The forecasted demand for the 3 days follows the seasonality of the previous 3 weeks well. Thus, the Holt-Winter smoothing did manage to predict the 3-day forecast quite accurately.

* The MSE and MAPE values were low and thus the Forecasting accuracy of the final forecasted model was predicted to be good.

**Figure F: Final MAE and MAPE values for Forecasted Model**

**Table

Description automatically generated**

There were 2 observations on the plot which would need further analysis on more date to make further improvements. These observations could be the result of the forecast model having over smoothened the peak values as well.

* Each weekday’s peak highest demand values were lower in the forecasted model in comparison with the Actual demand values.
* The forecasted 3 days’ first 3 timings (15mins blocks) of the day i.e., 6: 00a.m to 6: 45a.m show negative demands. The final Holt-Winter smoothing was run on 3 weeks data instead of the 2-weeks done during the training period. This could have adjusted the forecasts down further. The null values in the original data set on certain days during the 6:00 to 7:00 a.m. could also have had an impact on lowering the smoothened curve’s values at this time.

### **Result Conclusion**

The seasonality of demand hikes during the weekdays at the time from 1:00 pm to 2:00pm and 6:00pm to 8:00 pm was again observed in the new forecasted 3 days. Thus, without doubt the Transport company does need to investigate acquiring new buses and extend its terminals during this period.

The trend and seasonality assessment were based on the limited data obtained and the forecasted model was chosen and evaluated. A larger data set can open up and show further trends and seasonality which was not observed in this analysis. So, it is advisable to re assess this model on a larger Data set.

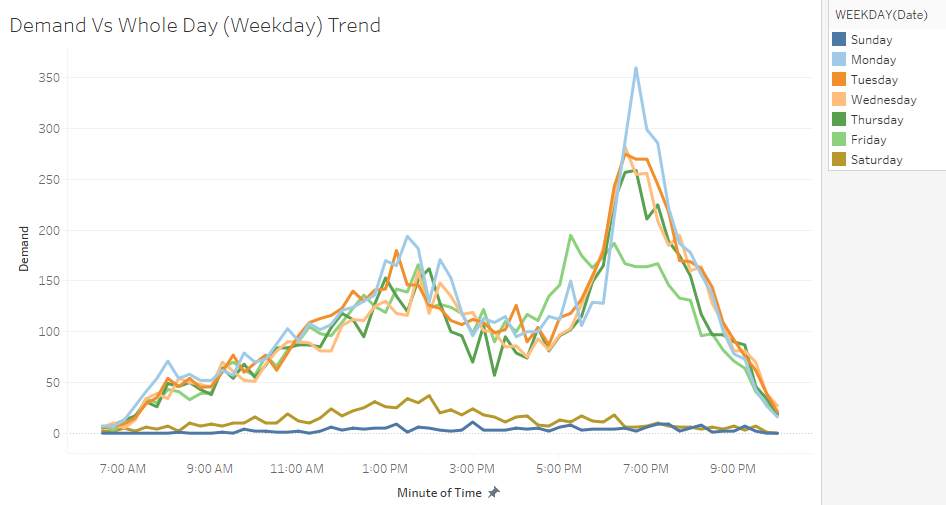
# **Appendix**

## Appendix A: Data Analysis and Time Plots

The following plots were visualized and analyzed in Tableau to locate possible Trends and Seasonality.

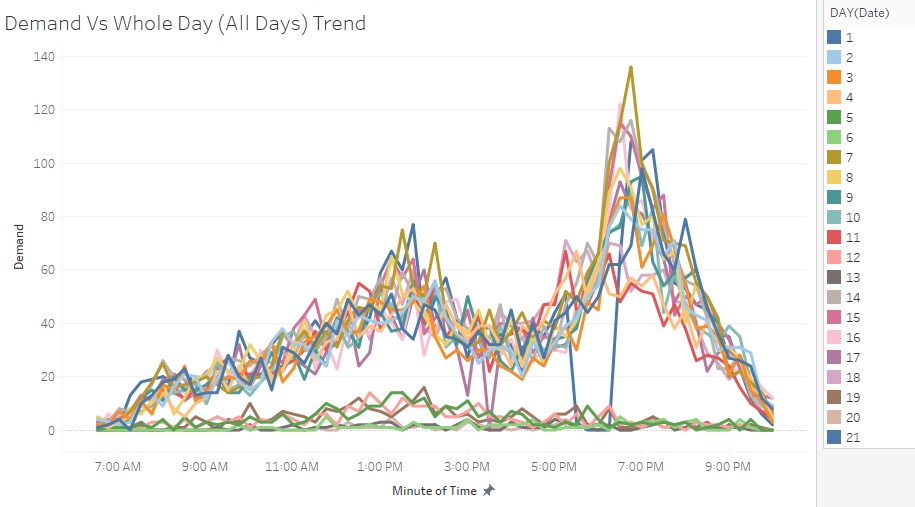
**Figure 1: Demand VS Whole Day Trend Plot**

The Plot indicated there seems to 3 different patterns namely a pattern per hour for weekdays, for Saturdays and for Sundays. There is a seasonality for weekdays peaking at 1.30 pm and 6.45 pm and dropping at 3.30pm. This could be indicating the lunch hour rush as well as the evening rush after work. The hourly seasonality for Saturday and Sunday seems to be insignificant.



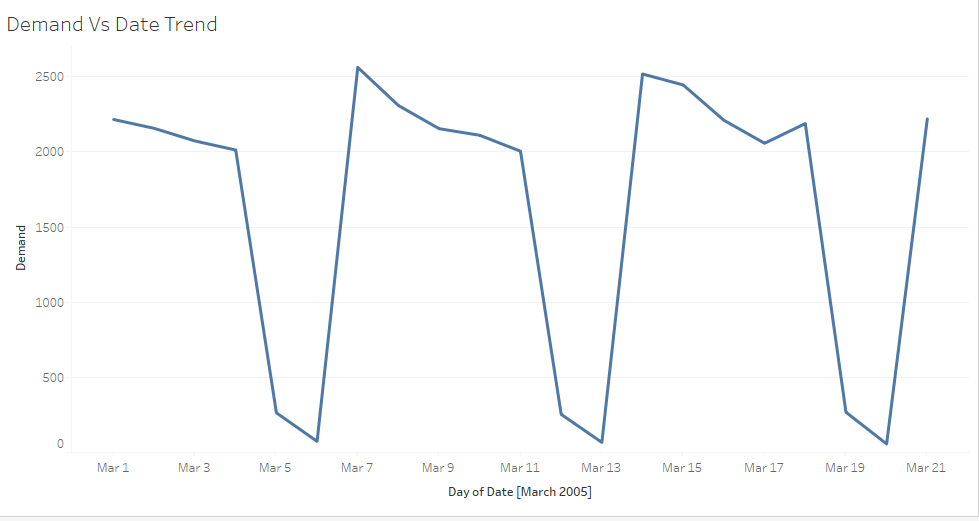
**Figure 2: Demand VS Whole Day (All days) Trend Plot**

The plot of hourly demand for all 21 days reconfirms the no trend and seasonality observed in Figure 1 is Global.



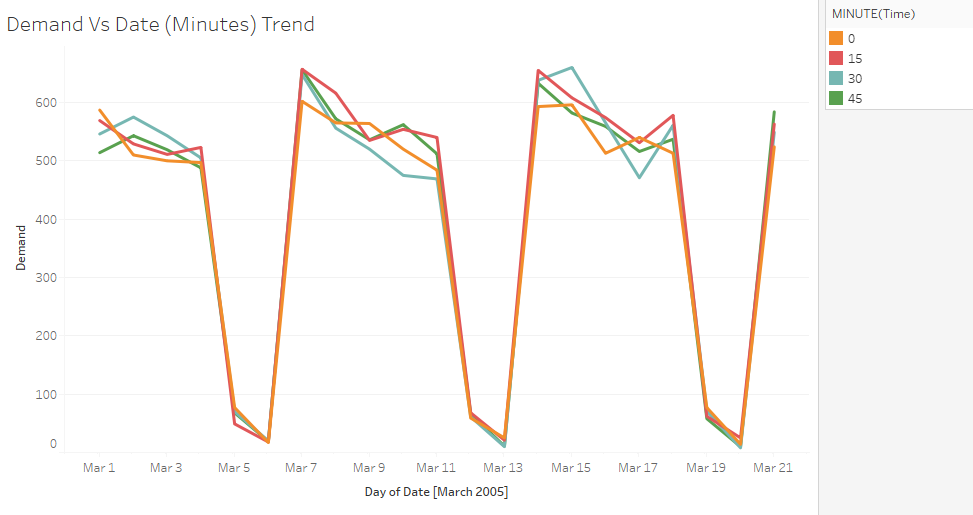
**Figure 3: Demand VS Date Trend Plot**

The plot indicates the demand is higher during the weekdays in comparison to Saturdays and Sundays. It shows no trend but a seasonality when comparing with just total demand daily.



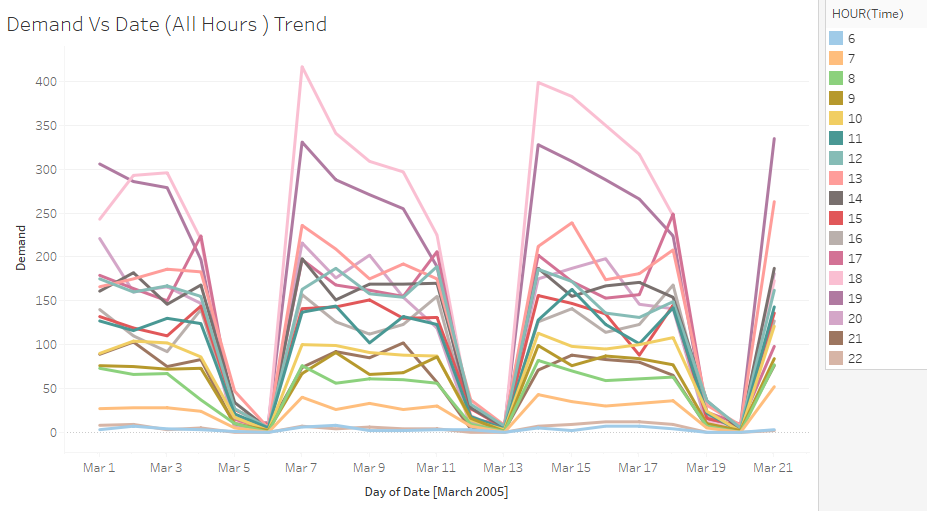
**Figure 4: Demand VS Date (Minutes) Trend Plot**

This was plotted to verify whether the different minutes of the hour showed any differences. It was observed the minutes did not affect the seasonality.



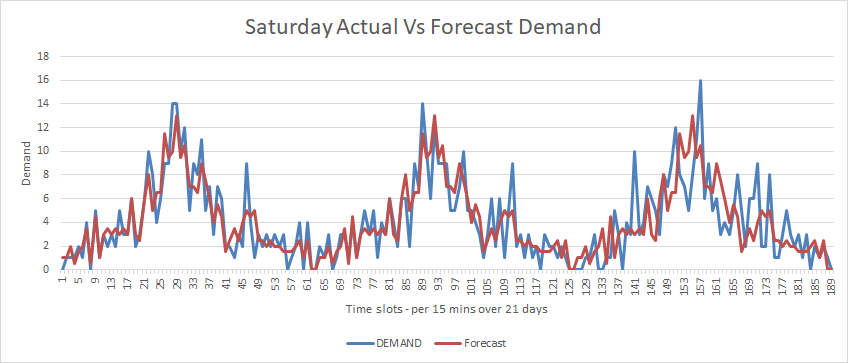
**Figure 5: Demand VS Date (Hours) Trend Plot**

This was plotted to verify whether the different hours in the day showed any differences over the days. It was observed the hours did affect the seasonality uniformly on all days. It reconfirmed the observation that we had seen earlier in Figure 1 and 2 that the peak hours were between 1 to 2 pm and 6 to 7 pm globally for all weekdays.

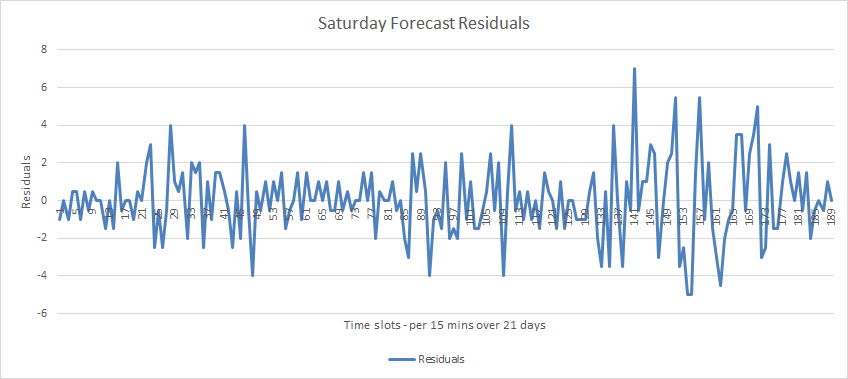


## Appendix B: Saturdays and Sundays Data Training, Validation and Forecasted Plots

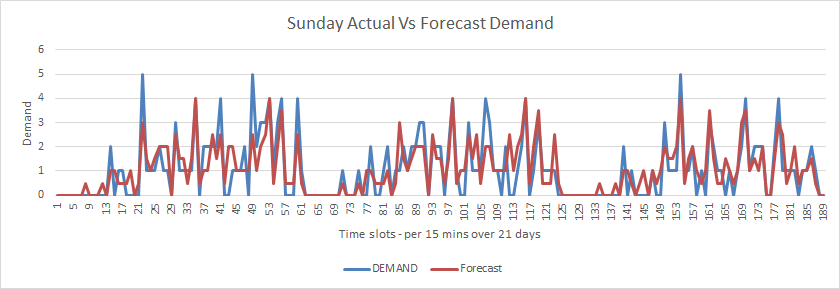
**Figure 6: Saturday Actual Vs Forecast Demand Plot**



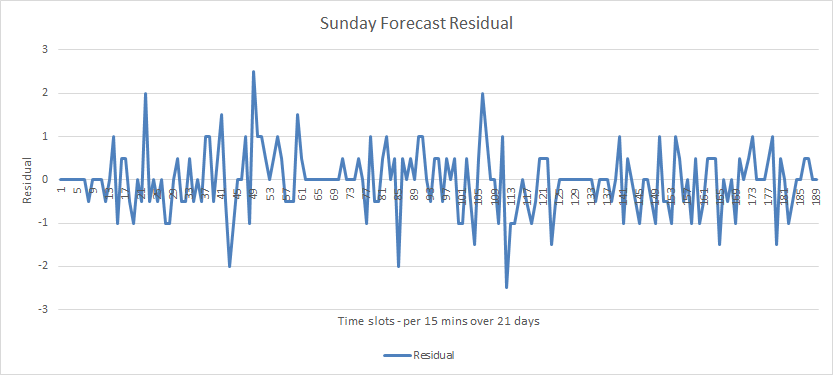
**Figure 7: Saturday Forecast Residuals Plot**



**Figure 8: Sunday Actual Vs Forecast Demand Plot**

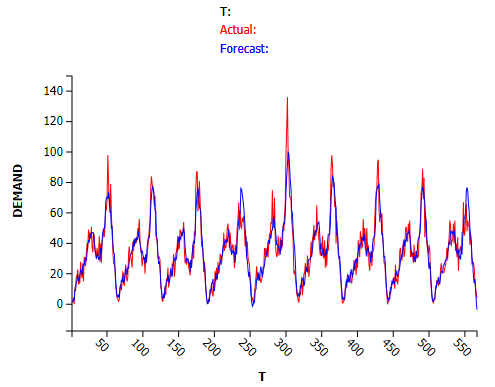


**Figure 9: Sunday Forecast Residuals Plot**

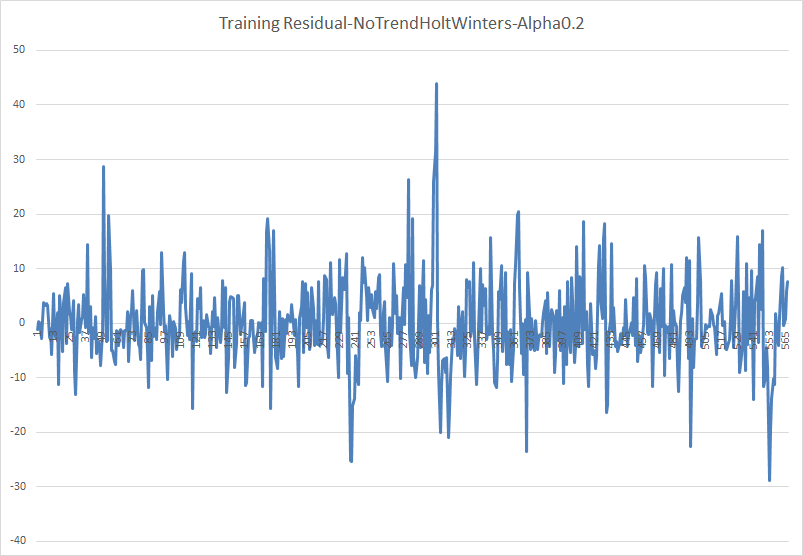


## Appendix C: Weekdays Data Training and Validation Residual Plots and Errors - Trend Holt-Winter Smoothing -No Trend Method

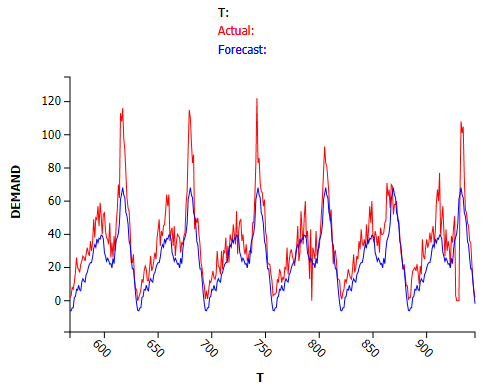
**Figure 10: Training Actual Vs Forecast Demand Plot – using Alpha 0.2**



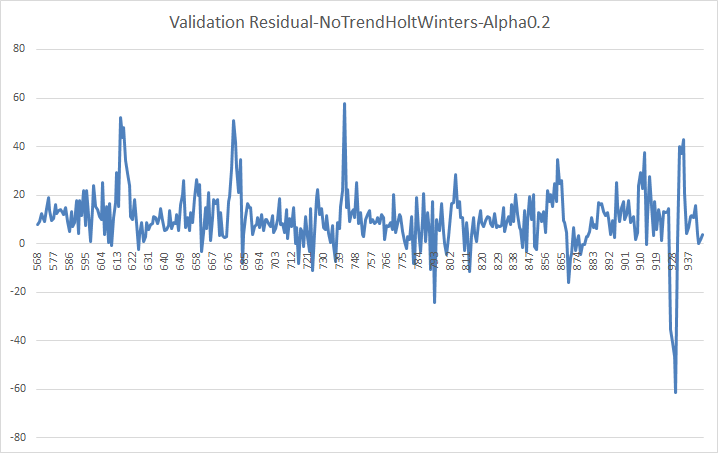
**Figure 11: Training Residual – using Alpha 0.2**



**Figure 12: Validation Actual Vs Forecast Demand Plot – using Alpha 0.2**



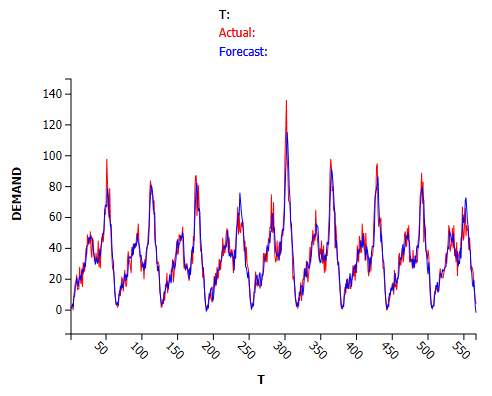
**Figure 13: Validation Residual – using Alpha 0.2**



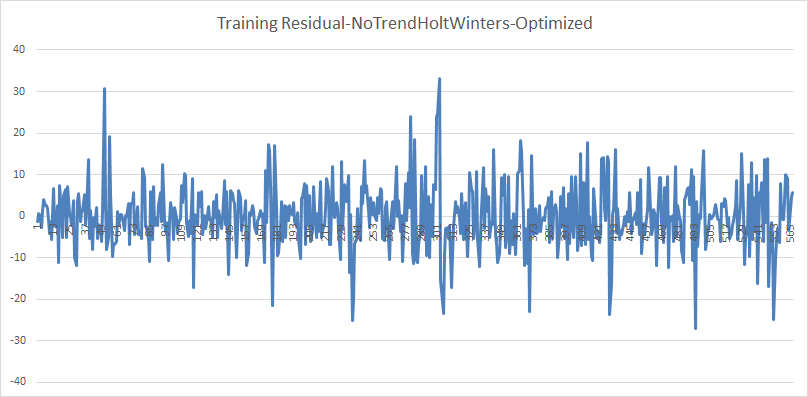
**Figure 14: Error Measures – Using Alpha 0.2**



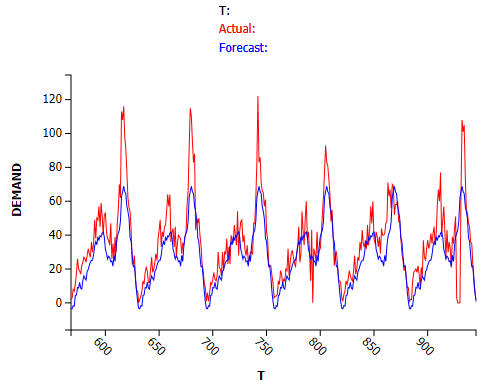
**Figure 15: Training Actual Vs Forecast Demand Plot – Optimized**



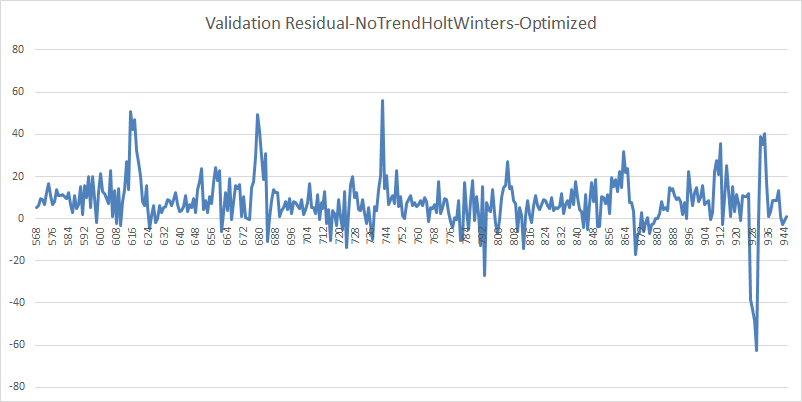
**Figure 16: Training Residual – Optimized**



**Figure 17: Validation Actual Vs Forecast Demand Plot – Optimized**



**Figure 18: Validation Residual – Optimized**



**Figure 19: Error Measures – Optimized**

