Bike Sharing Demand

TIME SERIES FORECASTING

Gokulraj Gunasekaran | Forecasting methods | December 15, 2016

Table of Contents

Dataset

Bike sharing demand, a competition held by Kaggle, has a dataset about Capital Bikeshare program in Washington D.C. This dataset has several features such as time of the day in hourly increments with a number of bike rentals per hour with a split of either casual riders or member riders. It is annotated with weather information such as temperature, 'feels like' temperature, season (four seasons), weather (cloudy, mist, rain, heavy rain), humidity and wind speed. It also has information about the day of the week and holidays.

Source: https://www.kaggle.com/c/bike-sharing-demand/data

CSV files:


train.csv


test.csv

HYPOTHESIS

Since there are many exogenous variables, several hypotheses can be made.

1. The Temperature affects the ridership.
2. Seasons: Spring, Summer, Fall, and Winter affect the ridership.
3. Weather: Cloudy, Mist, Rain, and Heavy rain affect the ridership.

Exploratory Data Analysis

Since there are already exogenous variables present in the data set, plotting bike counts against other variables may reveal patterns in the data.

Figure 1 below shows the plot of the total count over the period of two years. Since each month has only nineteen days of data, there are gaps in the plot. Figure 2, shows the boxplot total count against each month. It can be seen that there is a trend of high ridership during summer and fall months of both years and the ridership has significantly increased from 2011 to 2012.

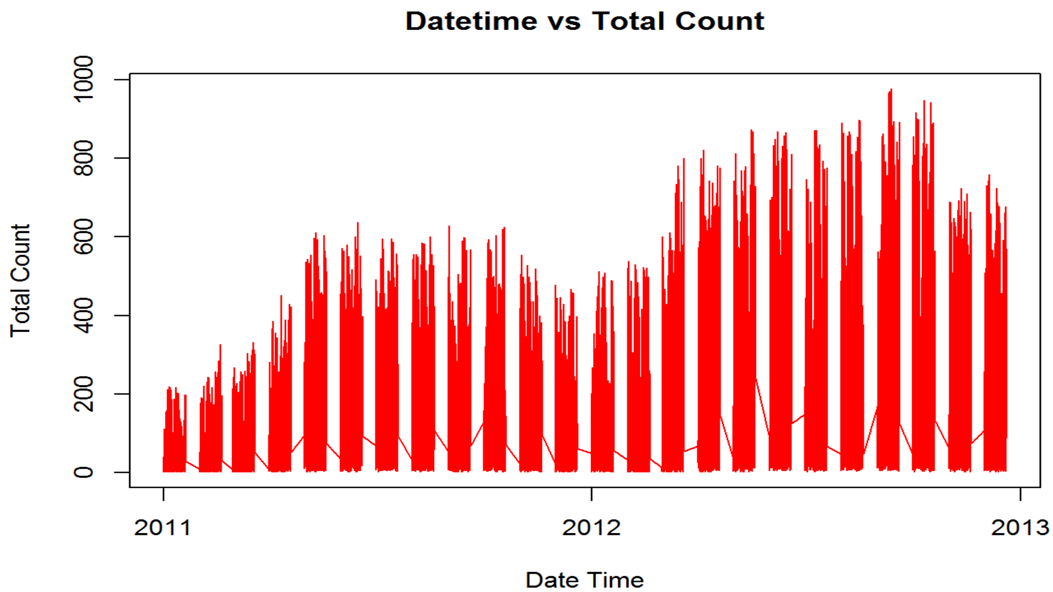

Figure 1 Date Time vs. Count of all data

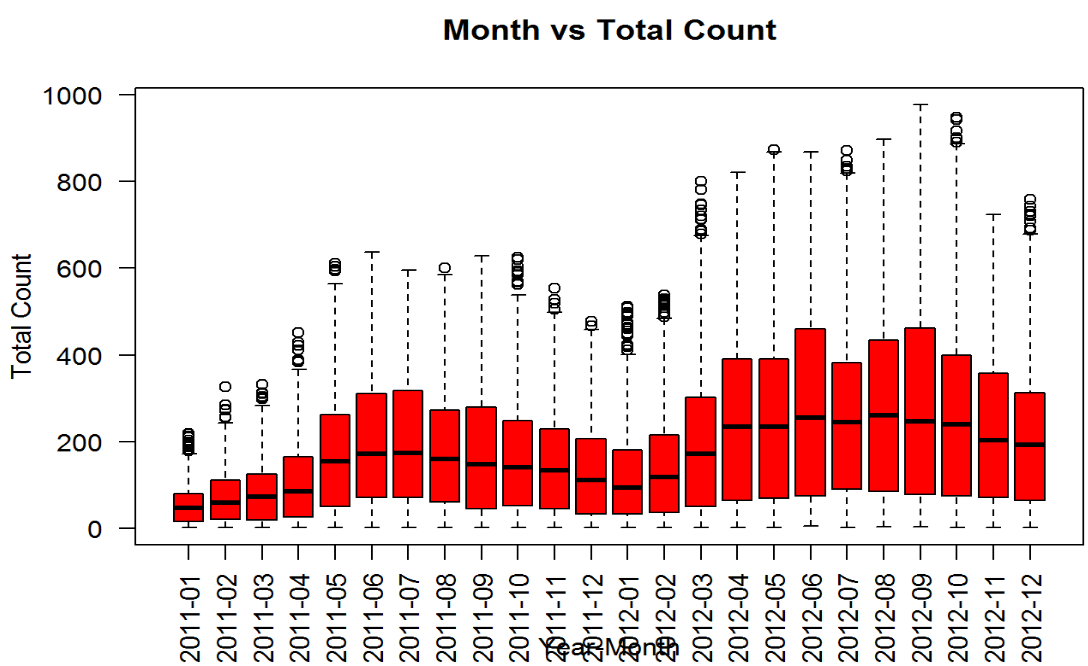

Figure 2 Month vs total count

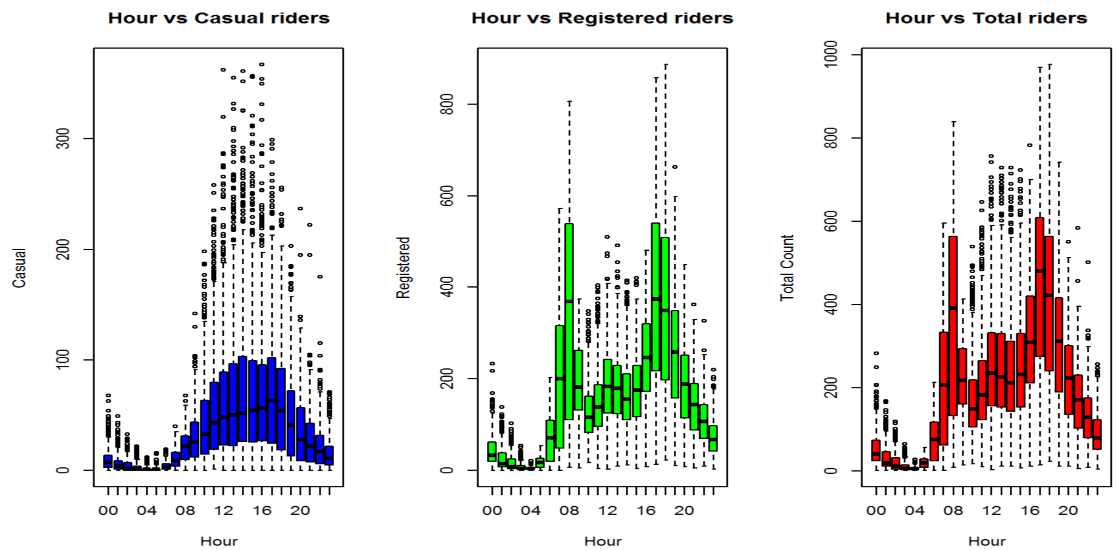

Figure 3 Hour vs Casual, Registered, Total riders

The above figure 3 shows show characteristics of daily/hourly seasonality. Casual ridership is very low during 4 AM and rises after 7 AM. It shows most ridership during early afternoon through the evening. Registered riders are very low during 4 AM. They are high during peak hours 7 Am - 8 AM and 5 PM to 6 PM, showing characteristics of office goers. The Total count resembles Registered users except that the range of riders during early afternoon through the evening is increased due to casual riders.

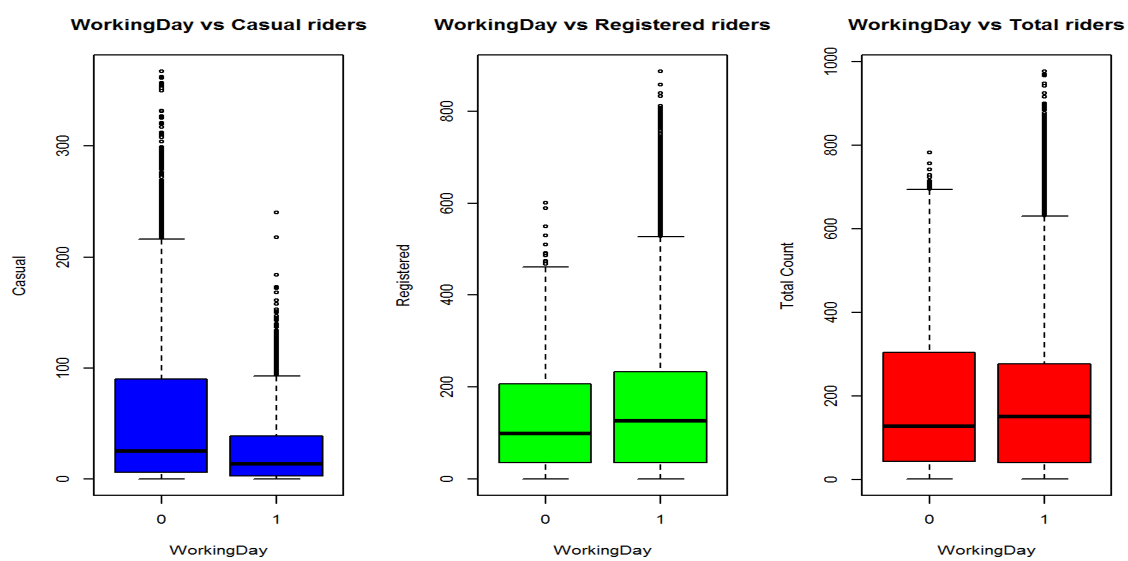

Figure 4 Working day vs Casual, Registered and Total Riders

The above figure 4 shows less number of casual riders on working days. There are relatively equal number of registered riders on both days, except the mean is higher on working days. Total ridership is similar to registered riders, but the range during the non-working days is greater due to higher casual riders.

The figure 5 below shows no significant characteristics between wind speed and ridership. In figure 6, there are some anomalies at 19 and 28 degrees in all types of riders.

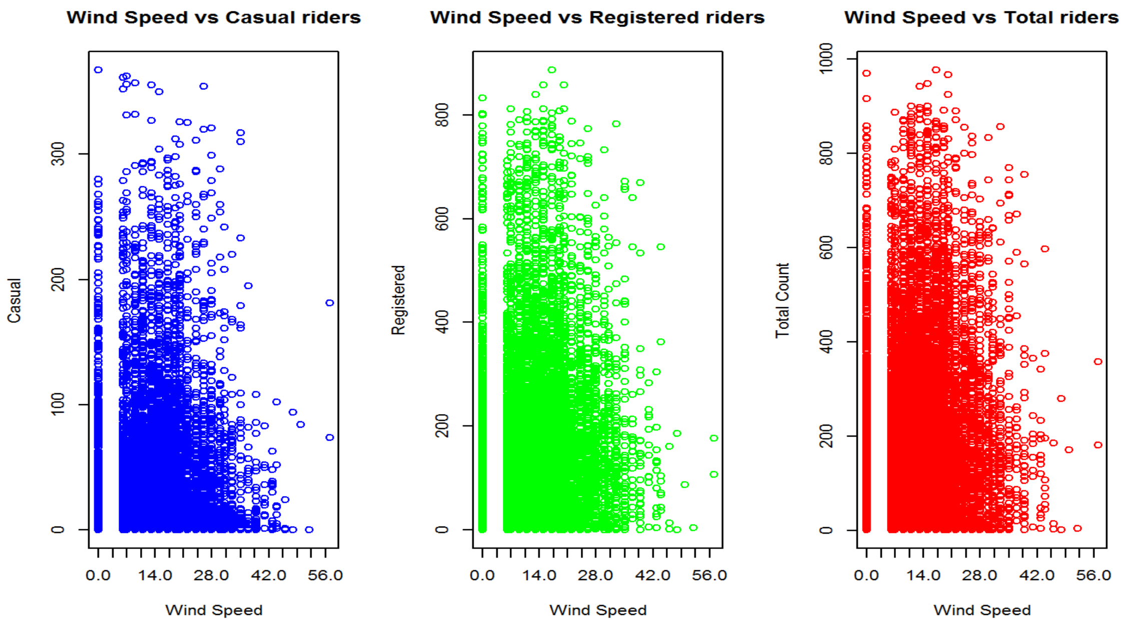

Figure 5 Wind Speed Vs Casual Registered and Total

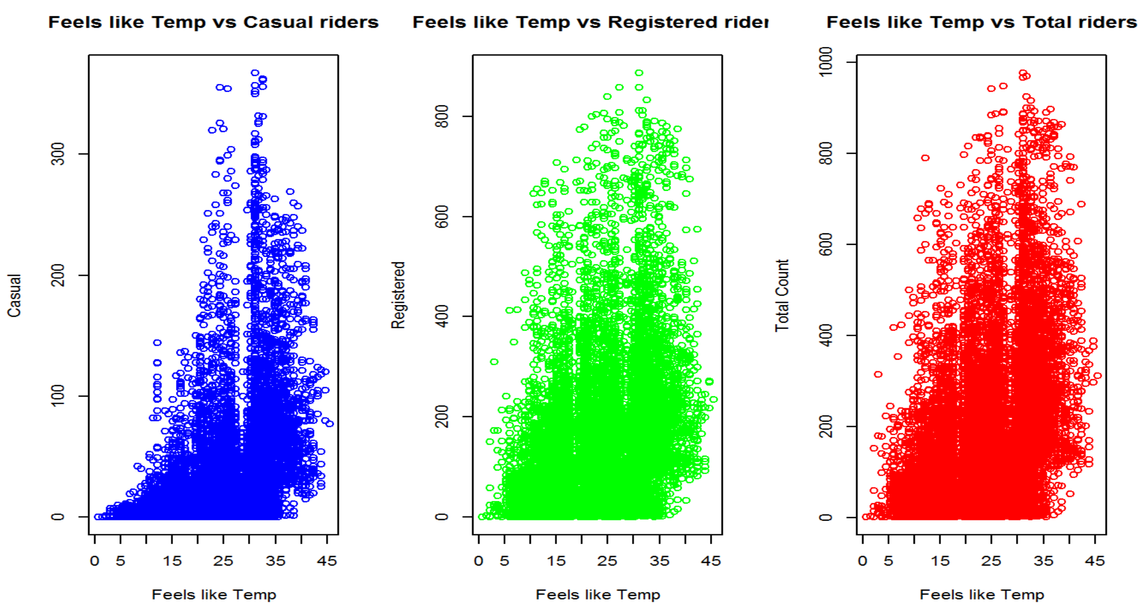

Figure 6 Feels Like Temp vs Casual Registered and Total

The below figure 7 shows effect of weather on ridership. During pleasant weather (1 – Clear, few Clouds) there are higher riders compared to weather (3 – Light snow, Light Rain + thunderstorm, 4- Heavy rain + Ice Pallets). There is only one record for weather at 4, it is an anomaly.

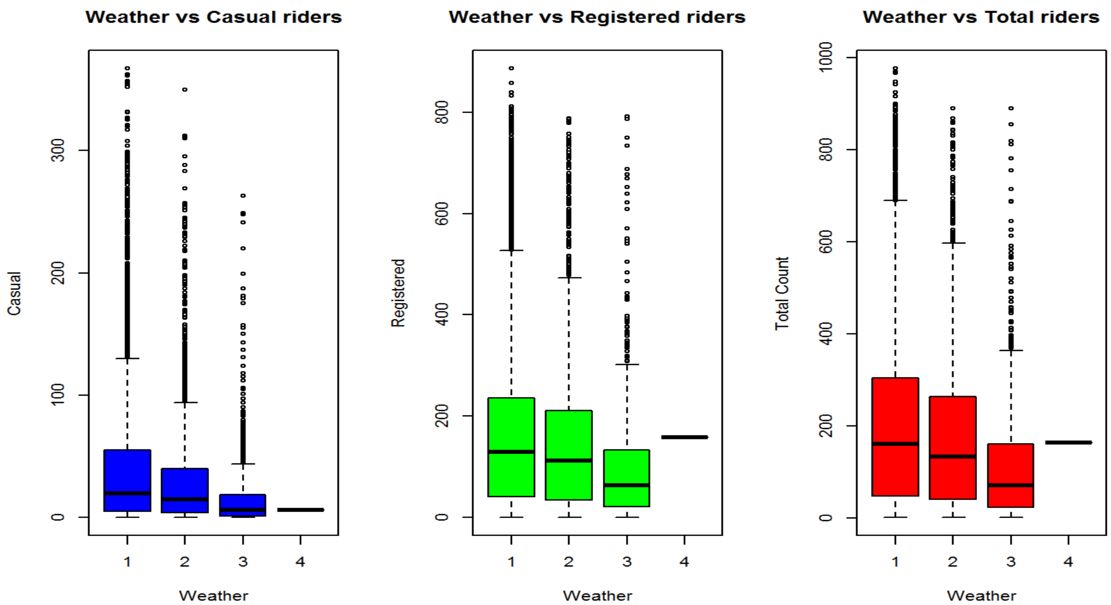

Figure 7 Weather vs Casual Registered and Total

The below plot of Temperature vs Casual, Registered and Total Riders is very similar to feels like temperature plots. Casual ridership grows more after 10 degrees and registered ridership grows after 5 degrees.

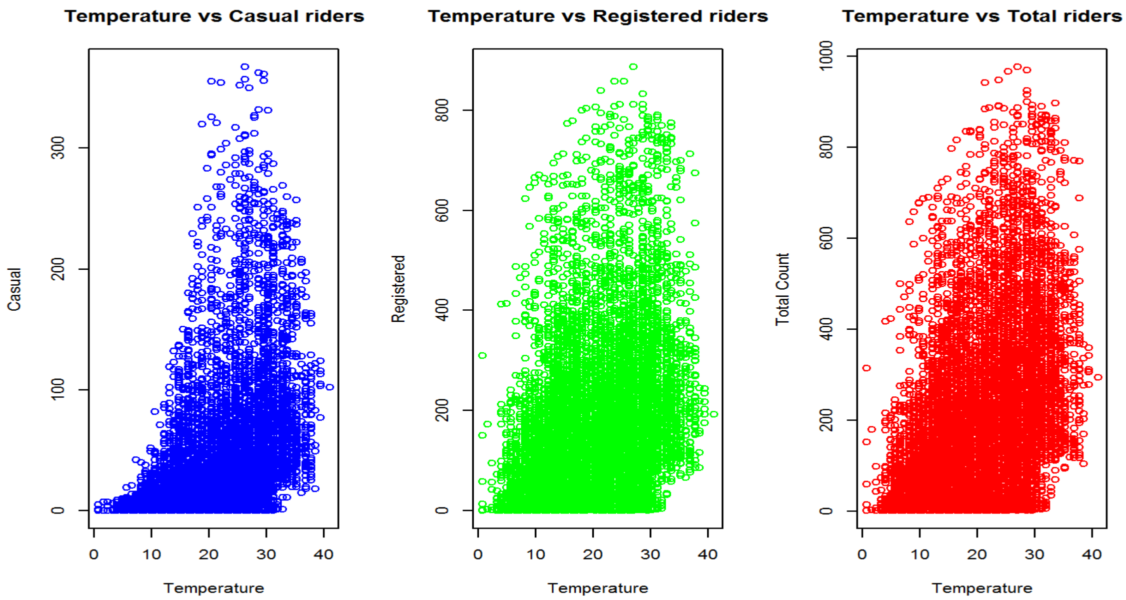

Figure 8 Temp vs Casual Registered and Total

Seasons are represented by Spring (1), Summer (2), Fall (3) and Winter (4). The below figure 9, shows that there are more riders during summer and fall and less in spring seasons.

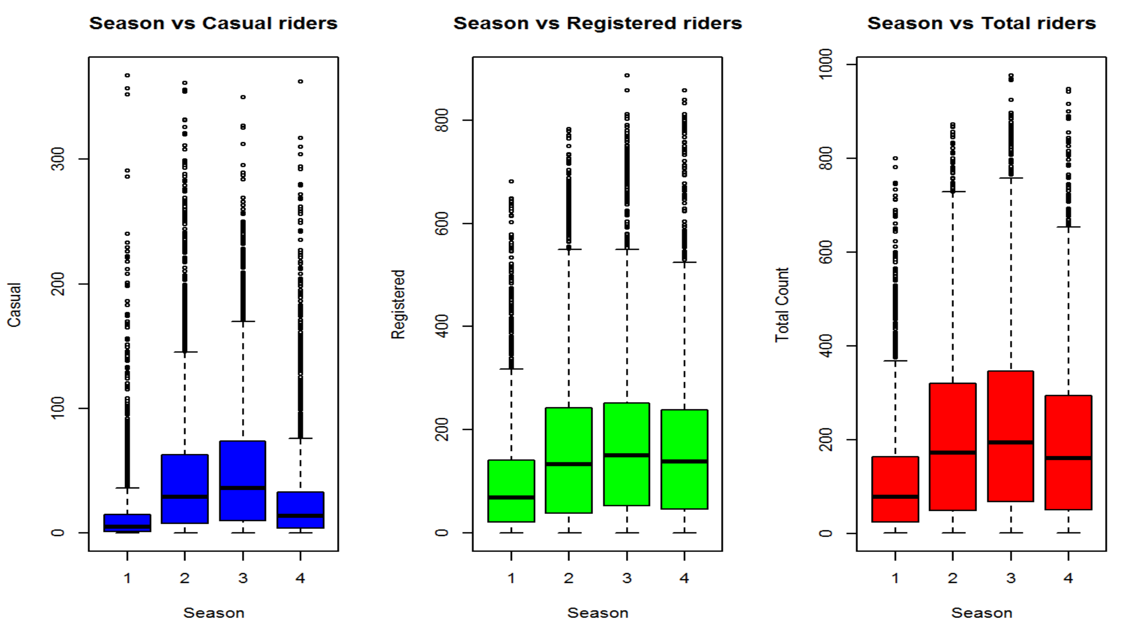

Figure 9 Season vs Casual Registered and Total

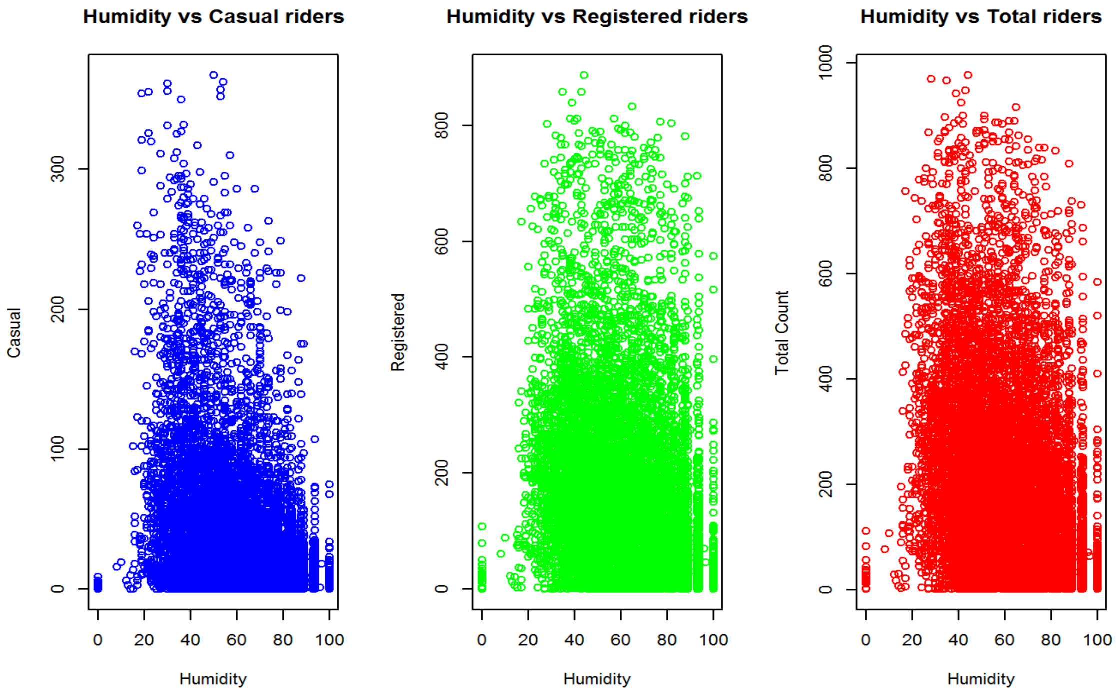

Figure 10 Humidity vs Casual Registered and Total

The above figure 10 shows the plot of humidity against riders. There are no discernible patterns.

Time Series Plot

Time series plot of the entire training data is shown in Figure 1. From figure 3, it can be seen that the total count follows the registered riders pattern. Thus, analysis will be performed only on total count of bike share per hour.

Subset Plots

Since there are 24 months with only 19 days of data, forecasting will be performed only on subsets of data. The below figures 11 and 12 show Jan 2011 and Jan 2012 total ridership respectively. It can be seen that there is high ridership during the weekdays compared to weekends indicating weekly seasonality. The Martin Luther King Day is celebrated on third Monday of January every year. This day is a non-working day/ holiday, which has a single peak pattern of ridership. Other working days have double peaks as shown in the plots below.

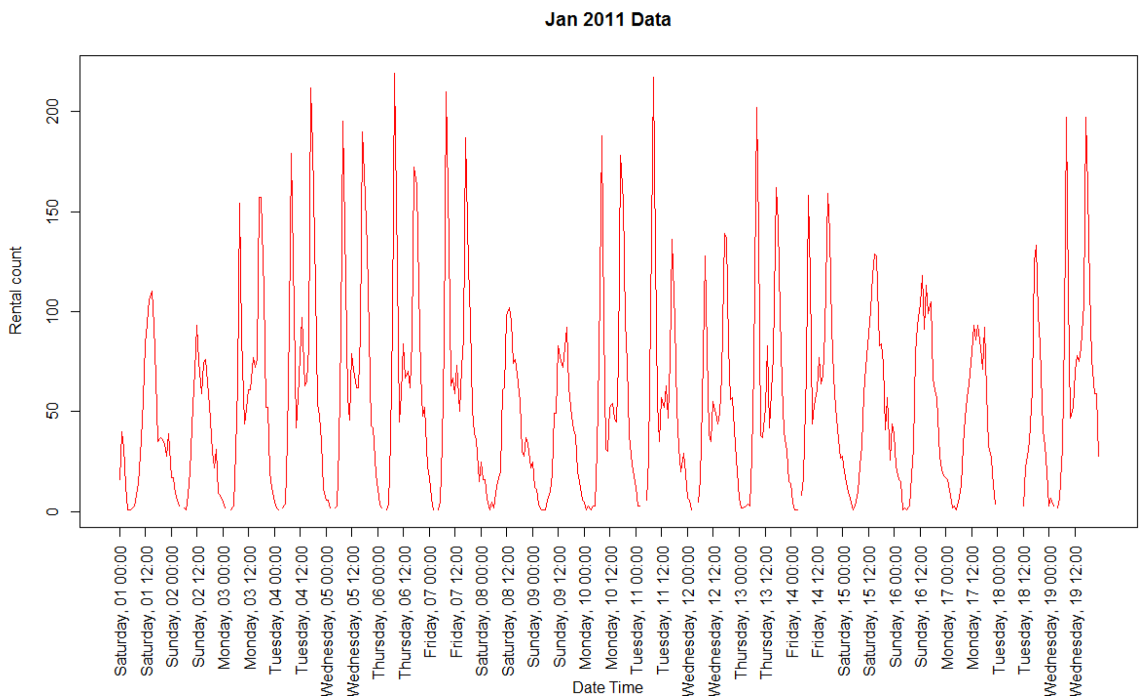

Figure 11 Jan 2011 data

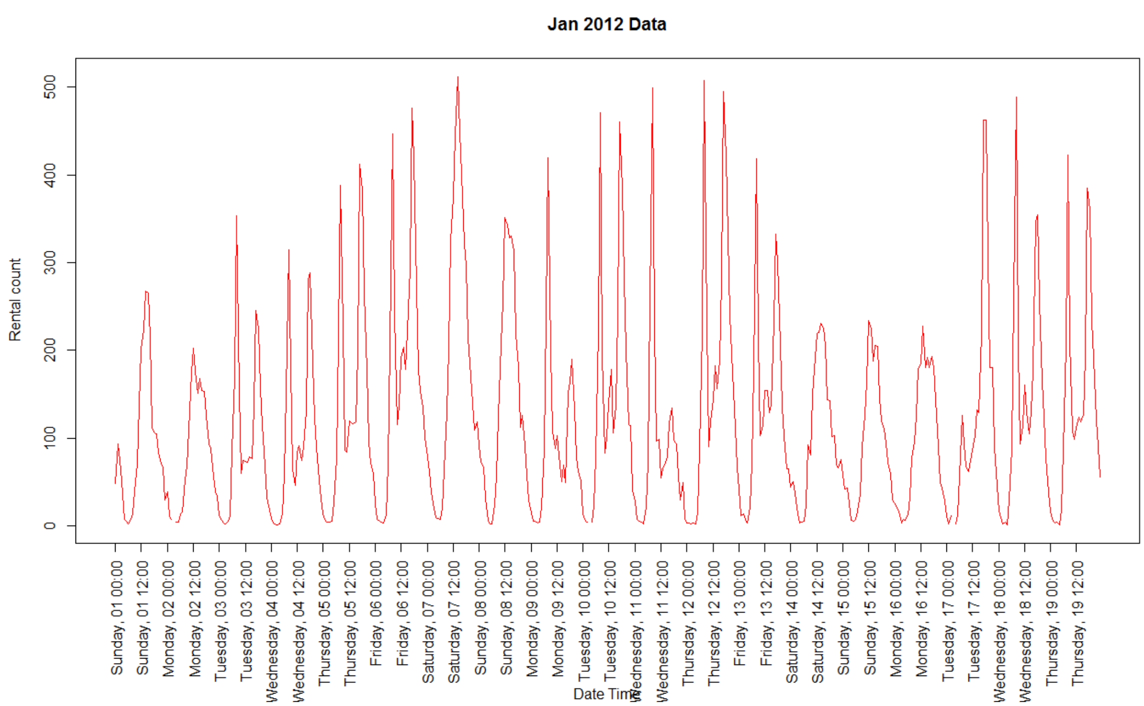

Figure 12 Jan 2012 data

Data Cleansing

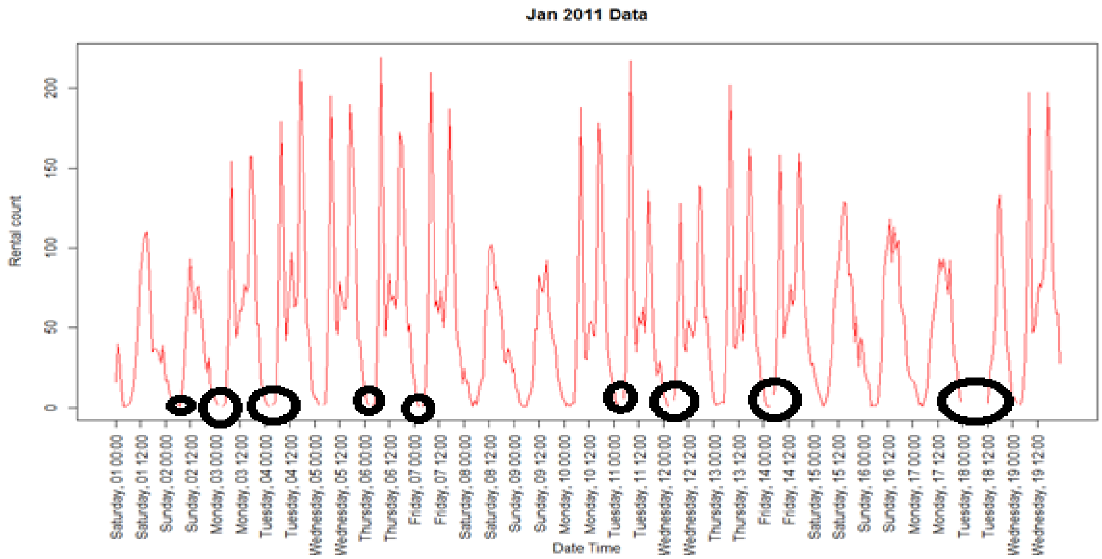

Figure 13 Missing data points

Apart from the data points after the 19th day of every month, there are some data points missing within this period. From figure 13, there are some missing values for some time, highlighted in ovals. Several imputation methods have been analyzed as shown in figures 14, 15 and 16. Spline interpolation has created negative ridership during Tuesday 18, 00:00 to 12:00 hours. The approximation technique shown in figure 15 could not impute the last five missing values in this dataset. Linear interpolation as shown in figure 16, has imputed values successfully. There are a total of 78 missing data points. Of the 78 missing points, there were no values for January 2011 during first 12 hours of Tuesday 18th. Thus going further, Jan 2012 will be used for creating models.

Spline

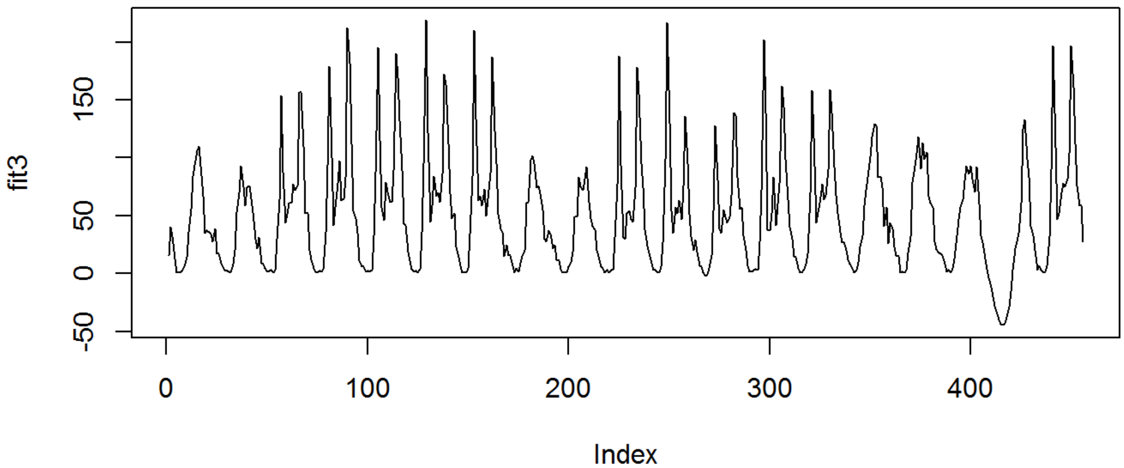

Figure 14 Spline imputation

na.approx

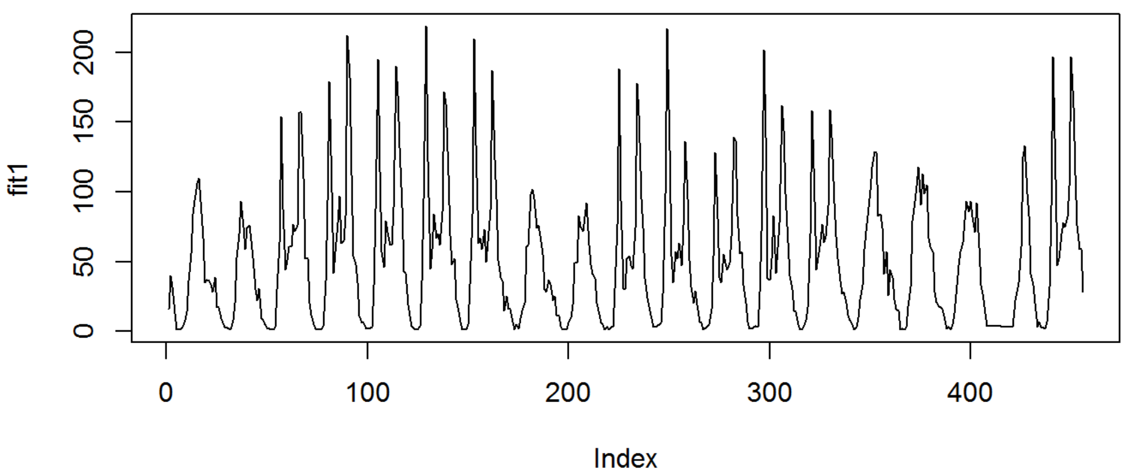

Figure 15 Approximation

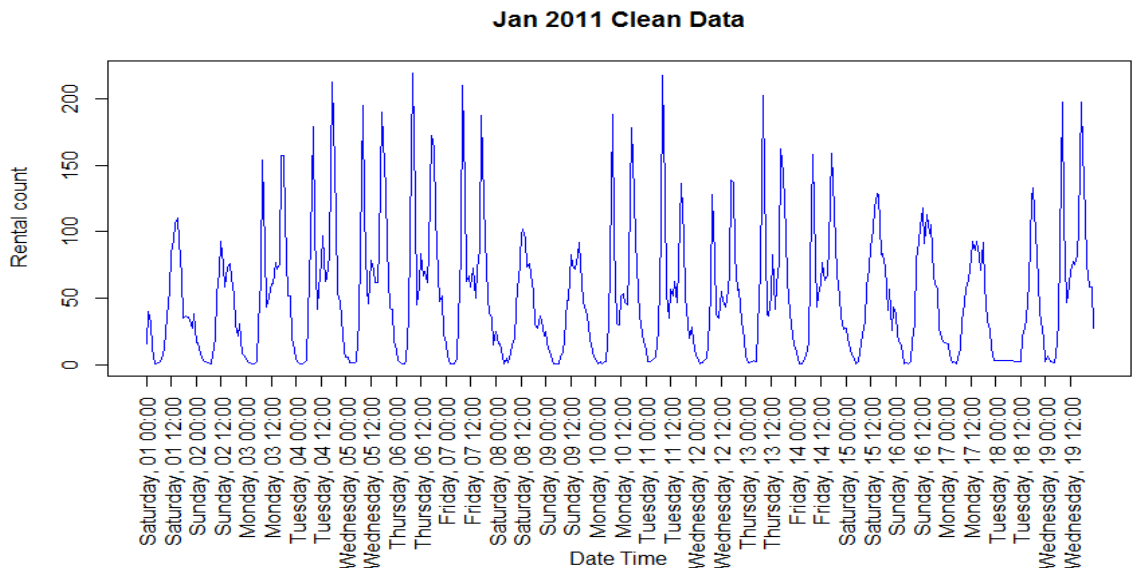

Figure 16 Jan 2011 Linear Interpolation

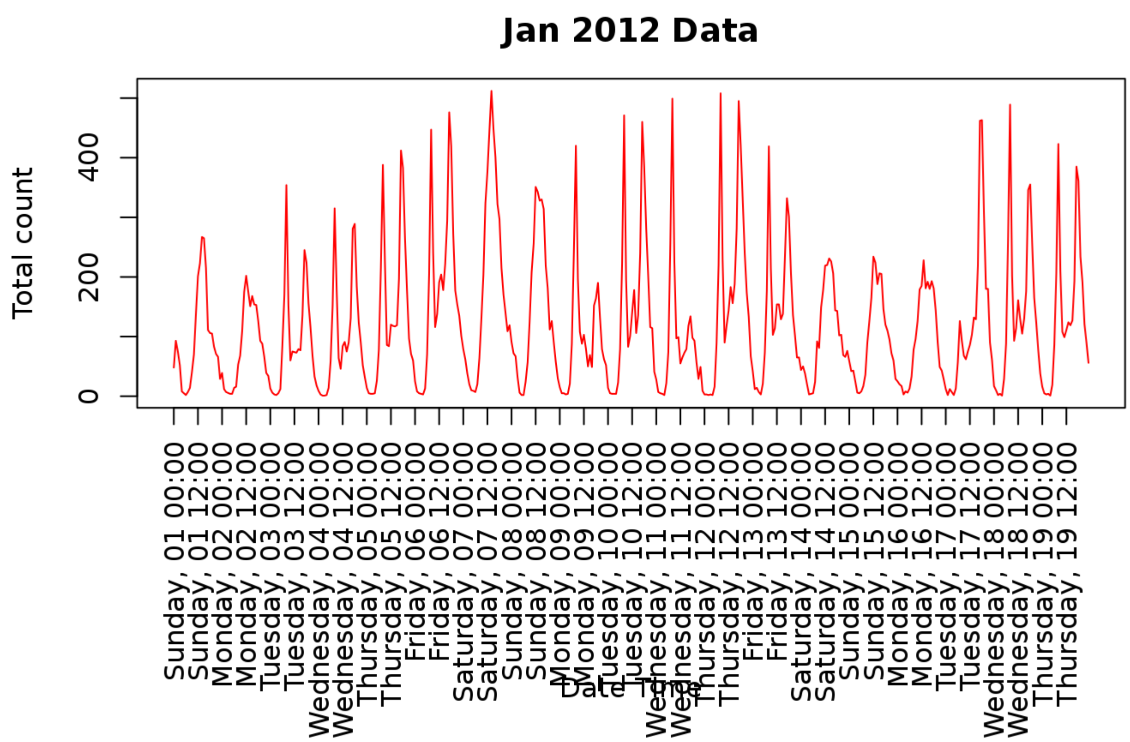

Figure 17 Jan 2012 Linear Interpolation

Smoothing

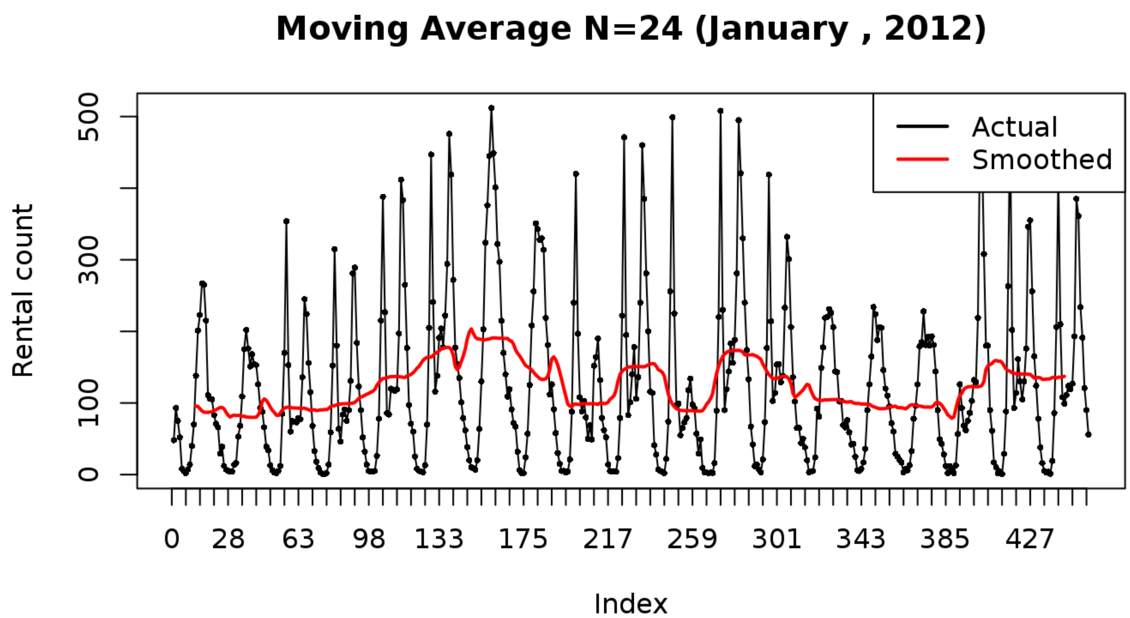

Figure 18 Smoothing with MA N= 24

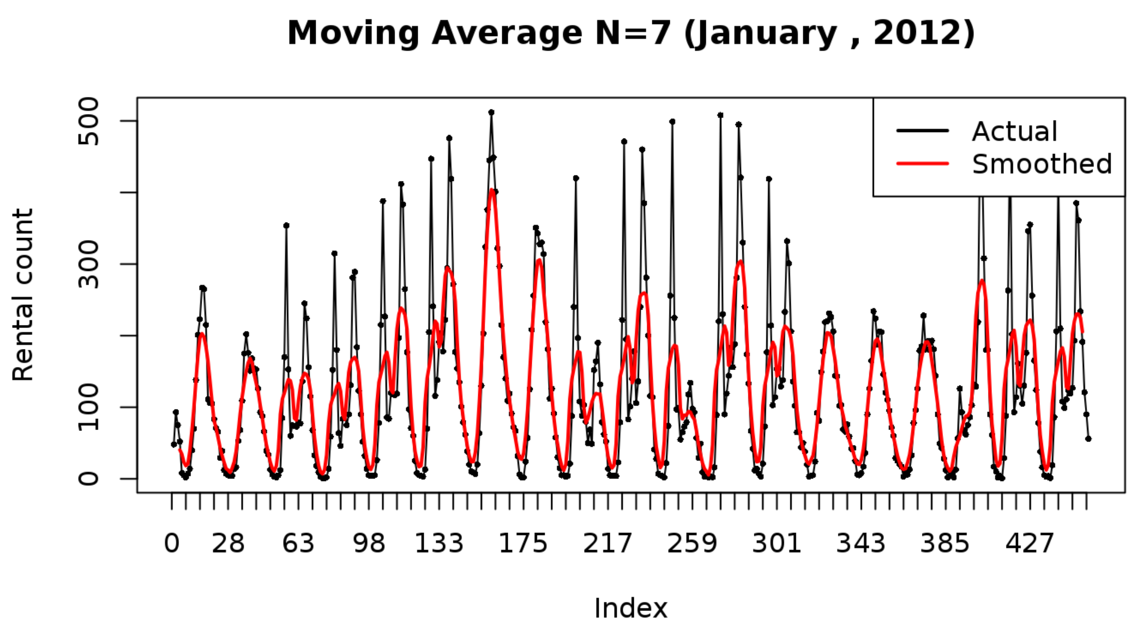

Figure 19 Smoothing with MA N=7

OBSERVATIONS

Figure 18 reveal the average ridership every 24 hours. There is a sporadic trend in the ridership over the entire time shown. In the figure 19, the ridership transitions from high to low and high again within a day, during working days. It reveals the characteristics of double peaks during working days and single peaks during non-working days.

Sample Variance, Mean, and ACF

For Jan 2012 data, sample variance is 13207.91 and the sample mean is 123.58, Below is the plot of ACF for Jan 2012 data.

ACF

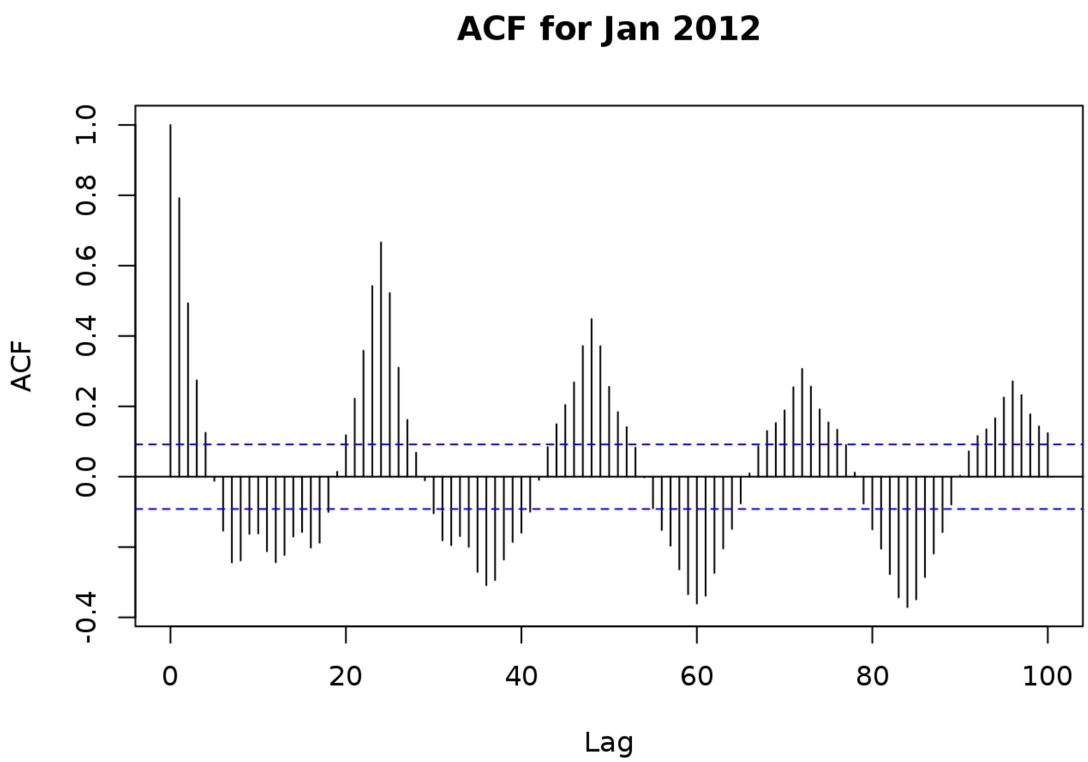

Figure 20 ACF

OBSERVATIONS

Autocorrelation function (ACF) reveal that the time series is stationary since the ACF is sinusoidal. It also reveals the high positive correlation at every 24 hours and high negative correlation at every 12 hours thus, a 24-hour cyclic pattern.

Variogram

OBSERVATIONS

From the figure 21 below, the semi-variogram is not linear. It shows very low variance at lag 24. This also reveal the weekly trend i.e. a steady increase in trend and reduces after 5 days showing a weekend/weekday pattern.

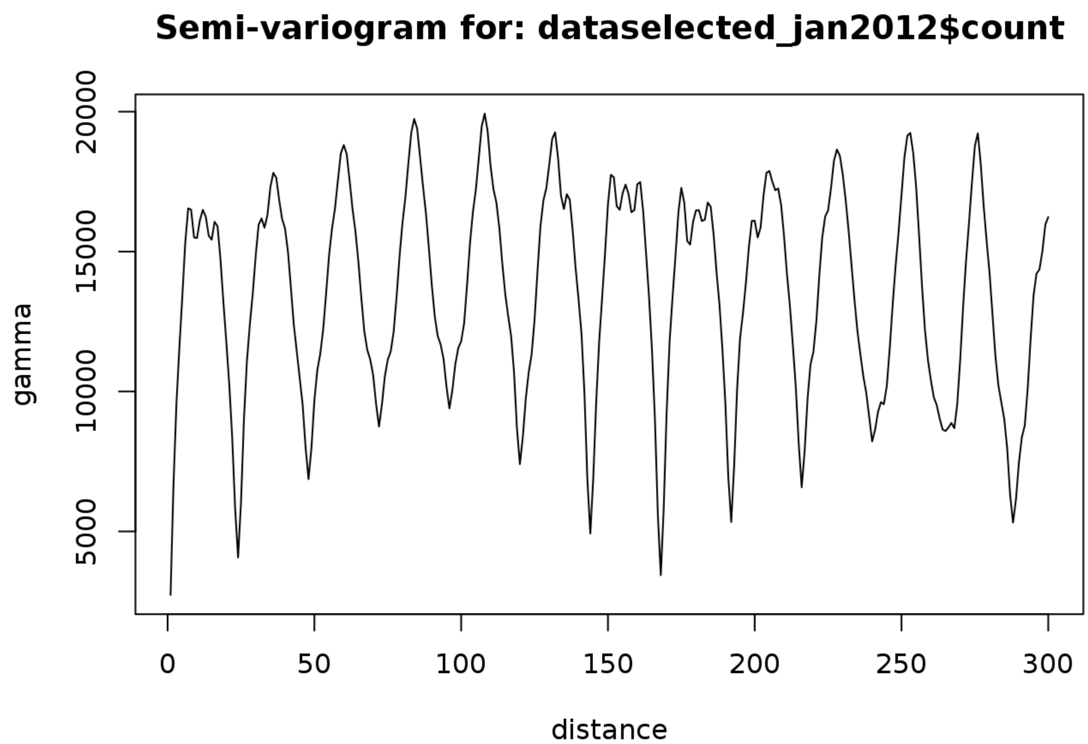

Figure 21 Variogram

Differencing

OBSERVATIONS

Figures 22 and 23 below show de-trended and de-seasoned data for Jan 2012 bike share data. The trend varies irregularly over time. The seasonal component shows a smoothed seasonality but with high variance. Thus by first order differencing, the result is a de-trended data for Jan 2012. And by differencing at lag 168, de-seasoned and de-trended data (figure 24) shows values are centered around the mean but does not have constant variance.

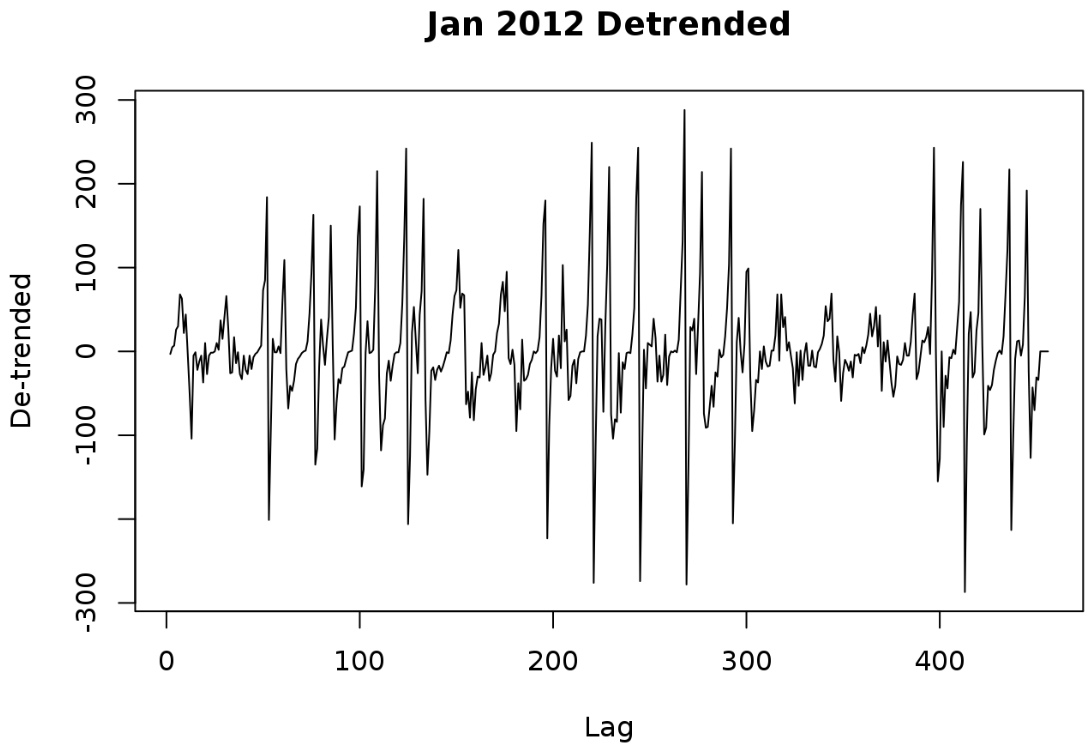

Figure 22 De Trended

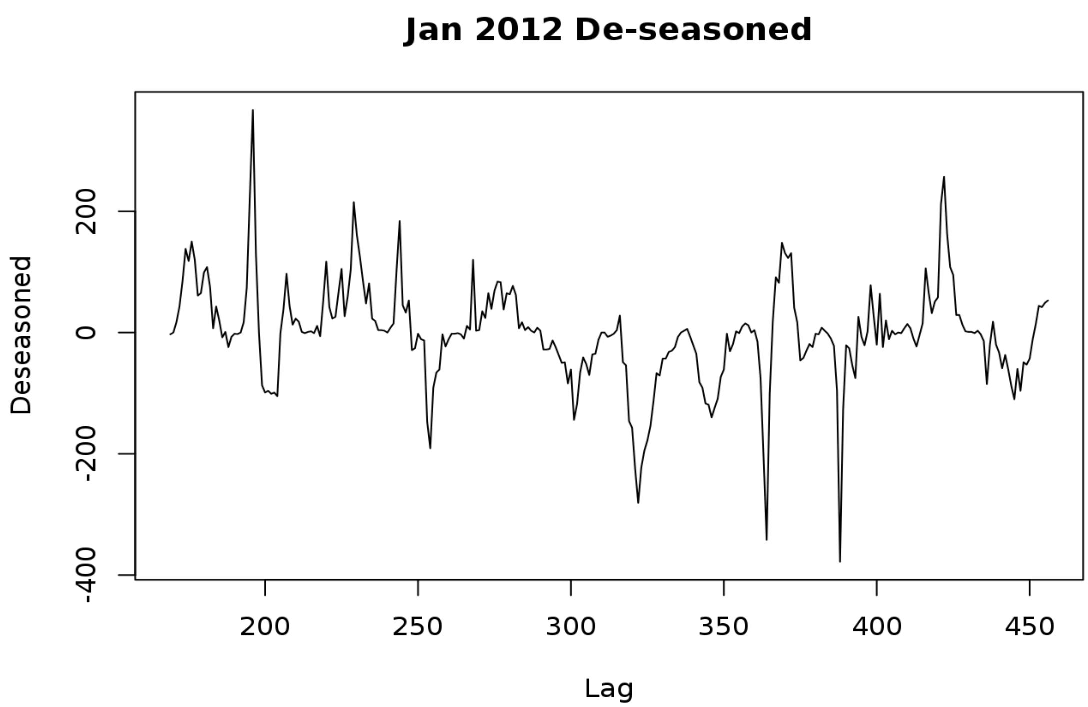

Figure 23 De Seasoned

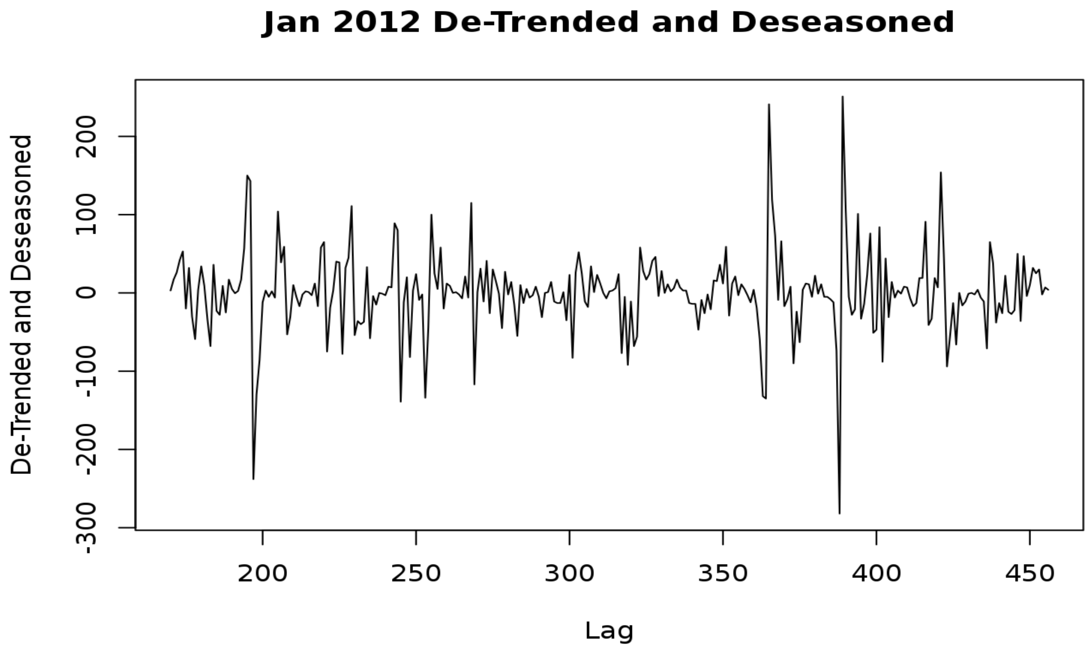

Figure 24 De-trended and De-seasoned

Power Transformation

Data transformations can be performed if the data shows high variance. In this case of Jan 2012 data, there is high variance in the data thus prompting a data transformation. Log transformation works well in reducing variance in data. The figure 25 below shows the log transformed total count for Jan 2012. It has effectively reduced the high variance with an almost equal amplitude of transformed count.

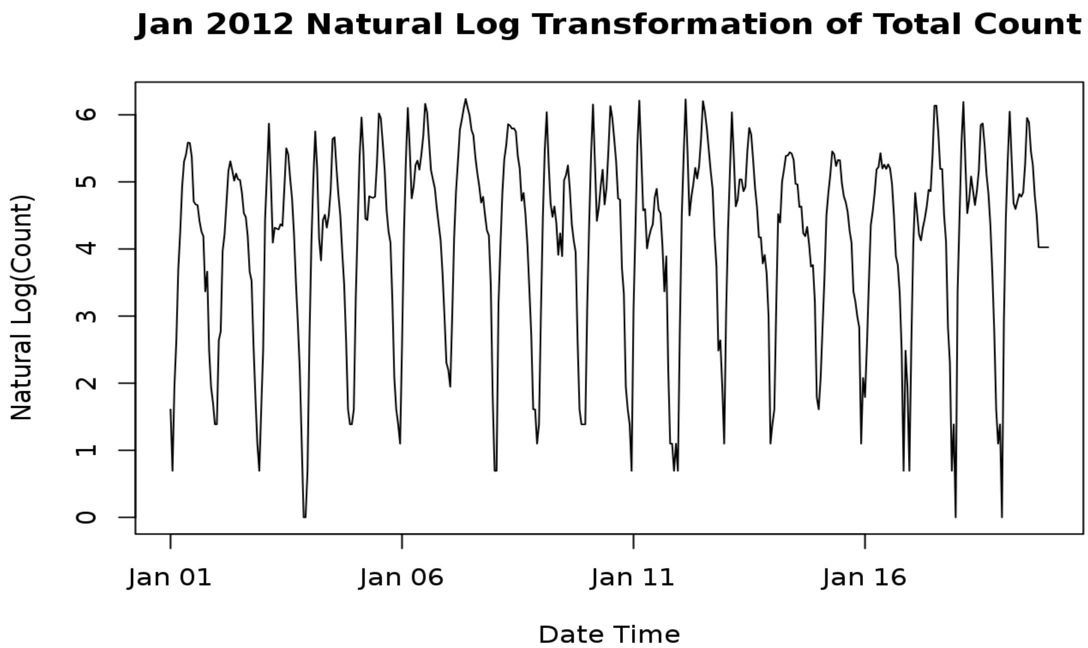

Figure 25 Log Transformed Jan 2012 data

Exponential Smoothing Decision Tree

OBSERVATIONS

The trend in data is sporadic with relatively equal amplitude of time series. By following the flow chart from figure 26, holts winter additive model is chosen to be best fit.

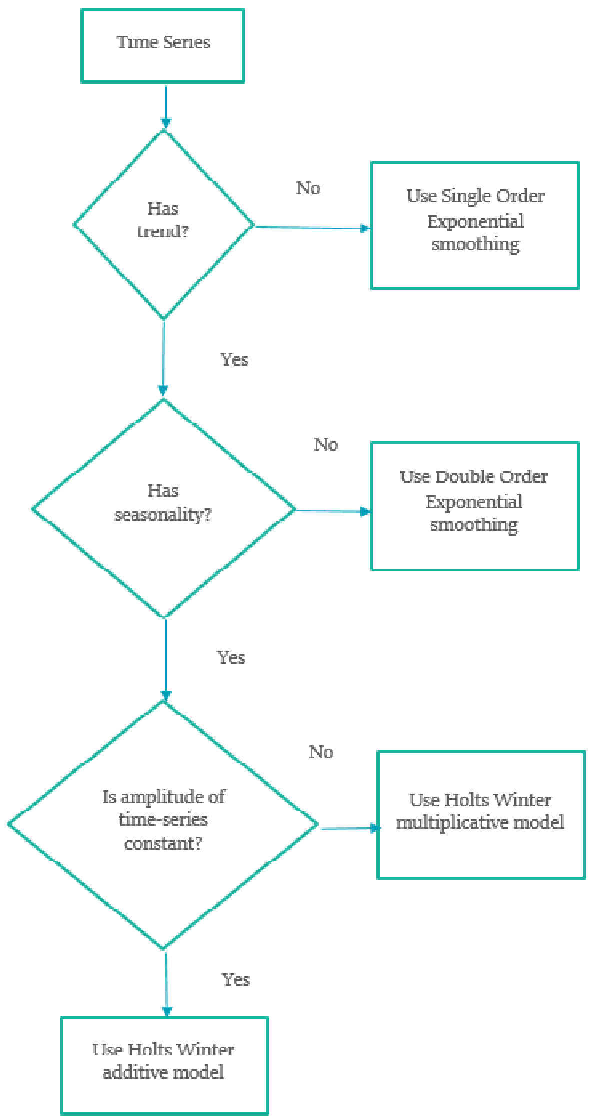

Figure 26 Exponential Smoothing Decision Tree

Fit Exponential Smoothing Model

TRAINING AND TEST SET

To create models, only a subset of data is used: January 2012. Since the data is for 19 days, training set is created with first 14 days and the remaining 5 days have been used for forecasting. The forecasting period has 120 data points i.e. 24 hours of 5 days. Smoothing models are evaluated using model fits and MAPE.

MODEL FITTING

Initially, the Holts winter model picked up high seasonality (gamma = 1), high level (alpha = 0.837), and no trend (beta = 0). The model fit showed negative values for some data points and the MAPE for the forecasted period is 231.45 %.

By varying the parameters (alpha, beta, gamma), several Holts Winter models have been tested as summarized below.

alpha	beta	gamma	MAPE	Comments
0.1	0	0	115.6087	Bad model fits and Forecasted values did not follow any required pattern
0.2	0	0	117.5885	
0.3	0	0	123.7475	
0.4	0	0	133.9161	
0.3	0	0.1	139.1429	
0.4	0	0.1	145.1948	
0.5	0	0	145.8328	
0.2	0	0.1	151.89	
0.3	0	0.2	155.7948	MAPE is high but forecast pattern is better than many models

Table 1 MAPE analysis of Holts winter model

The best model from the above analysis is the Holts winter models with alpha = 0.3, beta = 0, and gamma = 0.2, as highlighted in table 1. Even though the forecasted values have a MAPE of 155.79 %, the confidence intervals and the predicted values follow the double peaks. Thus this model picks up working day data pattern. This Holts winter model shows double peaks at days when the actuals are single peaks since it is only dependent on daily seasonality.

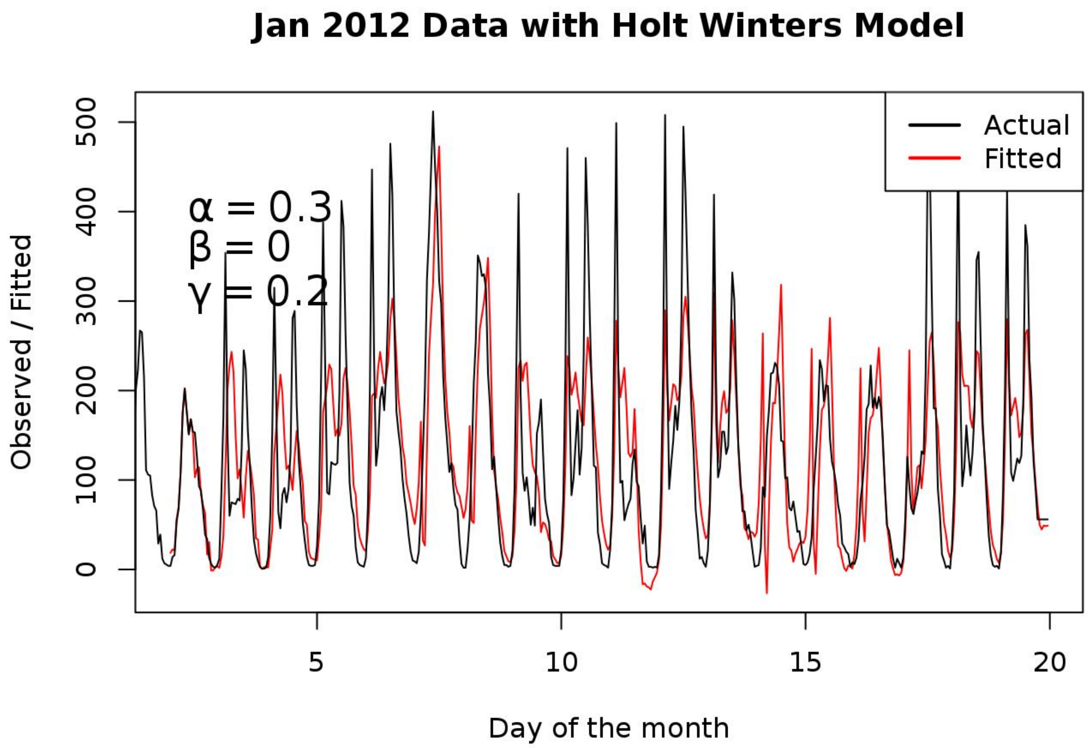

Figure 27 Actual vs Fitted Holts Winter for Jan 2012

Residuals

Figures 28 and 29 show the plots of residuals. The residuals still have correlation left as seen in figure 28. But in PACF, all values are within in the intervals. As seen in figure 29, the residuals look normally distributed except for outliers. These outliers pull the right tail of the histogram longer. The Fitted Vs Residuals plot show high variance as the value increases, showing characteristics of heteroscedasticity. The residuals are mostly centered around the mean. The box ljung test of residuals showed p-value less than 0.05 indicating non-independent residuals.

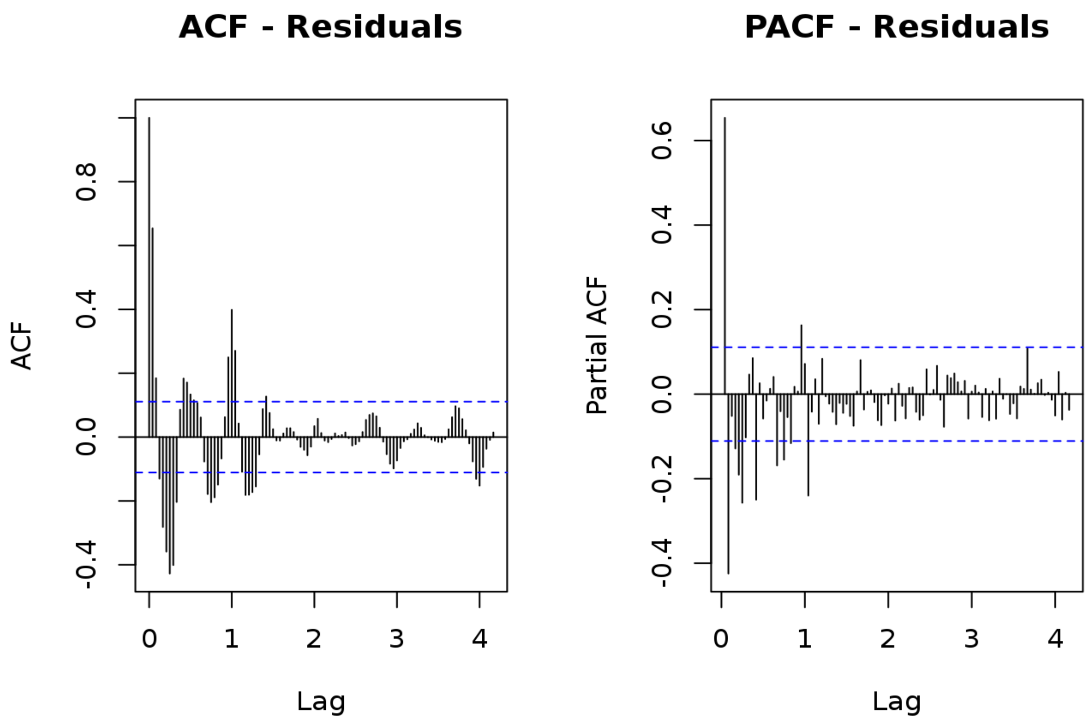

Figure 28 ACF and PACF of Residuals

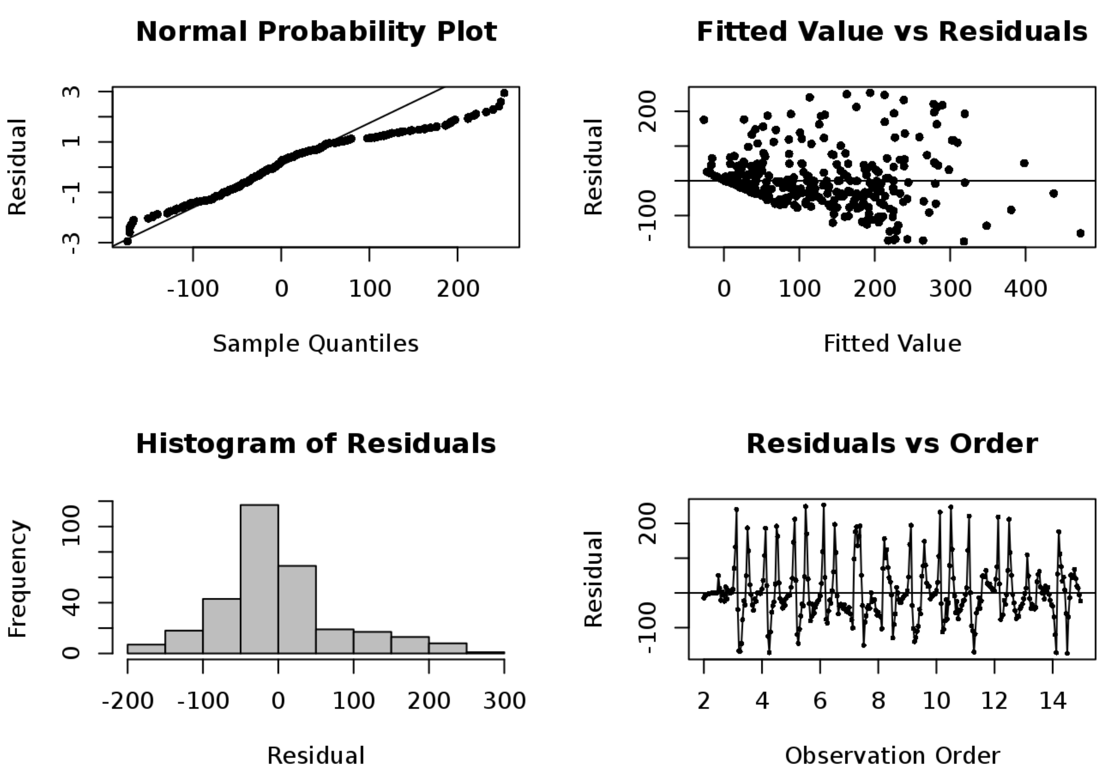

Figure 29 Residuals of HW Jan 2012

FORECAST

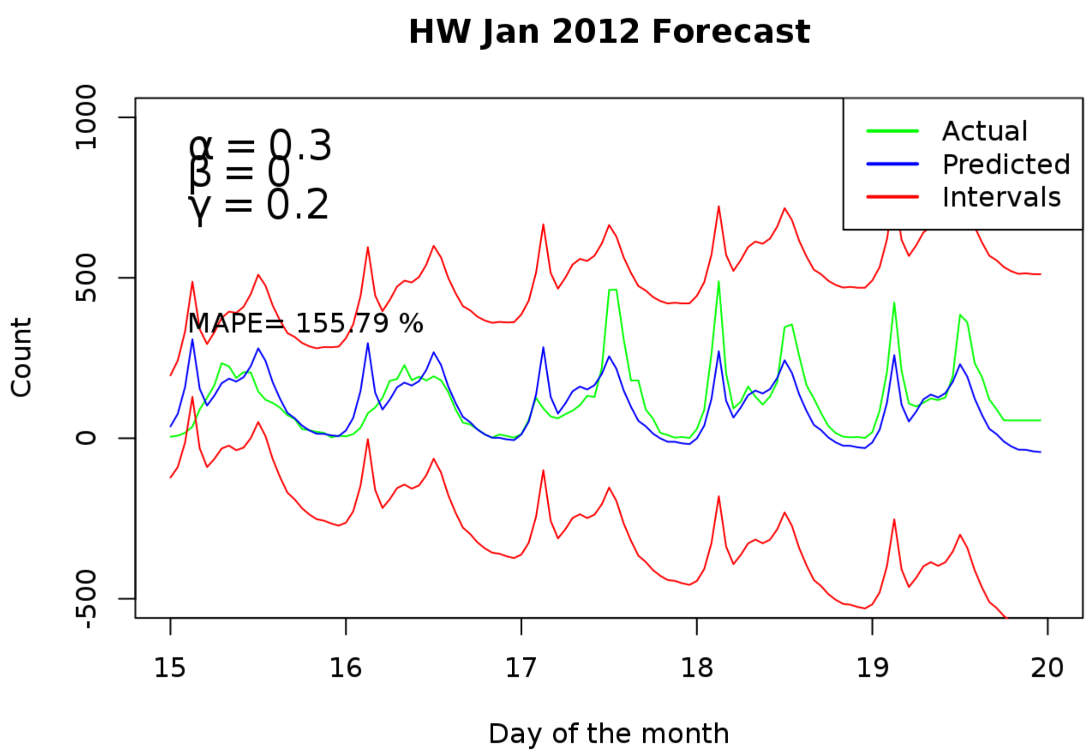

Figure 30 Holts Winter Forecast for Jan 2012

The above figure shows the forecasted values in blue against the actuals in green. The forecasted values are from Jan 15th to Jan 19th 2012, which has 120 data points. This model picks up daily seasonality and forecasts the double peaks pattern for all days.

Residuals

Figure 31 show that there is still some correlation left in the residuals with some correlation exceeding the intervals. PACF on the other hand has almost all values within the intervals. From figure 32, it can be seen that the residuals are normally distributed except for few outliers which pull the left tail of histogram longer. The mean of residuals is close to zero. There is still some high variance as the actuals increase showing characteristics of heteroscedasticity. The box ljung test of residuals showed p-value less than 0.05 indicating non-independent residuals.

Due to heteroscedasticity, a log transformation was applied to the total count variable and models were created. But all of the models provided either very bad fit or very bad forecast.

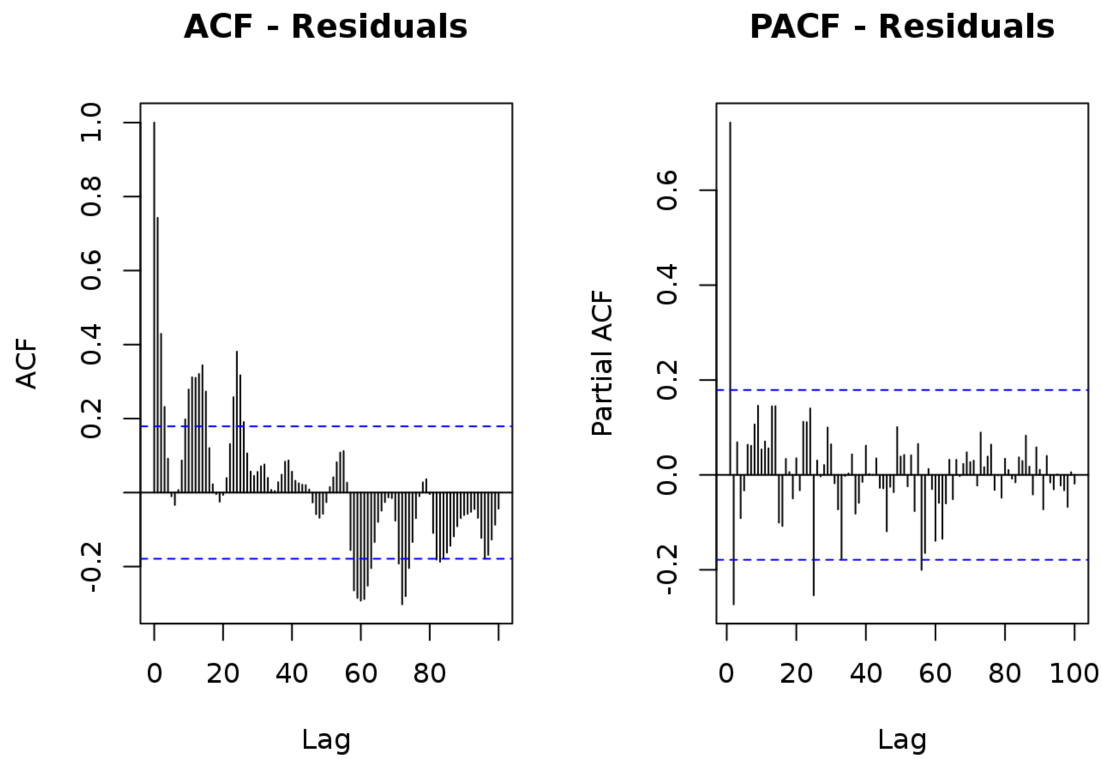

Figure 31 ACF and PACF of residuals

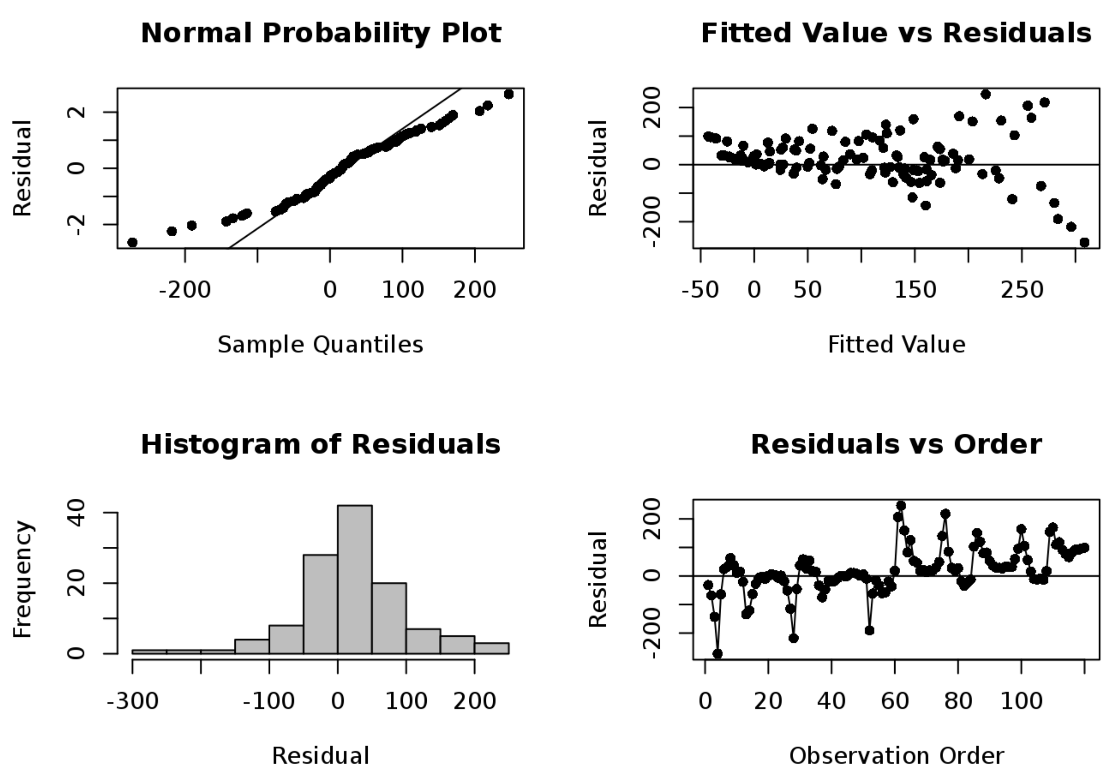

Figure 32 HW Jan 2012 Forecast residuals

Identifying MA(q) / AR(p) processes

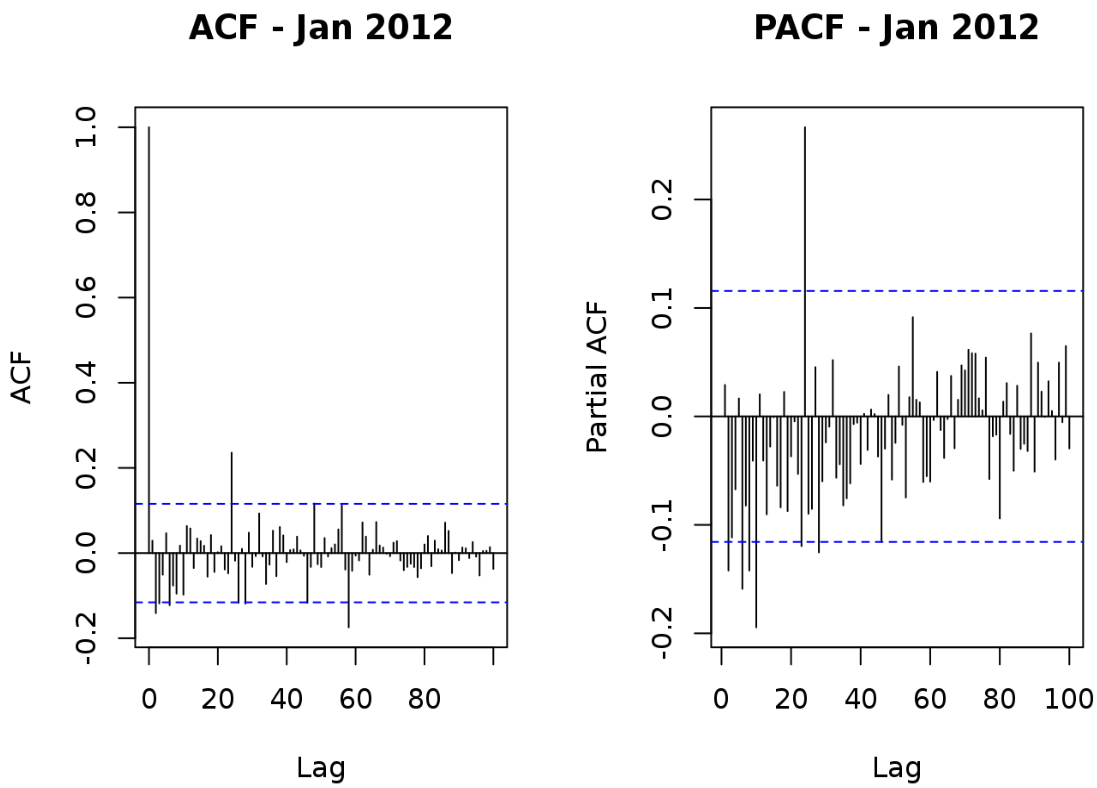

Figure 33 ACF and PACF of De-Trended and De-Seasoned Jan 2012 data

OBSERVATIONS

ACF for De-Trended and De-Seasoned count of Jan 2012 shows characteristics of damped sinusoid. PACF shows characteristics of mixture of exponential decay and damped sinusoidal pattern. ACF drops close to zero at lag 1 indicative of MA (1) process. But there is high correlation at lag 24. This requires further investigation. Thus choosing p, q, P, Q values by experimenting with several models.

ARIMA (p, d, q) x (P, D, Q) s model Fitting

The Jan 2012 data is first used to create an arima model using auto. arima function in R, which identified ARIMA (2,0,2) process This model performed poorly with a MAPE of 676% and the forecasted values are the average values (flat line) without any patterns. Thus various ARIMA models are run in R with different p, d, q, P, D, Q values for 24-hour seasonality. The parameter s is not assigned to 168 (weekly seasonality) since the arima function in R threw error. By using the training set of first 14 days of Jan 2012 and remaining 5 days as test set, MAPE and AIC analysis was performed as shown below.

p	d	q	p	d	q	MAPE	AIC	Comments
2	1	2	1	1	1	81.01582444	3359.293899	Candidate models with better fit and independent errors.
2	1	3	0	1	1	81.45388162	3353.033469	
3	0	4	1	1	1	85.05605605	3314.727441	
3	1	3	0	1	1	85.29181202	3351.13281	Models with one the following: 1) very wide confidence intervals. 2) non-independent errors. 3) forecasted values did not follow the daily pattern.
2	1	2	0	1	1	86.03077643	3394.924666	
3	0	3	1	1	1	90.19641899	3323.879063	
2	0	4	1	1	1	91.24494436	3322.037542	
2	0	3	1	1	1	91.42397498	3320.122677	
3	0	2	1	1	1	94.3497935	3328.934933	
2	0	2	1	1	1	94.98506258	3327.122676	
2	0	2	0	1	1	104.6103049	3359.84832	
2	0	4	0	1	1	105.2572841	3357.134288	
3	0	3	0	1	1	105.2769486	3357.127103	
2	0	3	0	1	1	105.2834389	3355.158176	
3	0	4	0	1	1	106.4086953	3358.22713	
3	0	2	0	1	1	106.8322083	3356.590817	
3	1	2	0	1	1	106.8775523	3350.62043	
3	1	3	1	1	1	106.9260324	3354.76729	
3	1	4	0	1	1	116.2200574	3353.050984	
2	1	4	0	1	1	120.650559	3351.216555	

Table 2 MAPE and AIC analysis of ARIMA models with different parameters

The above table is the summary of top models analyzed with different p, d, q, P, D, and Q parameters. Of all the models, the candidate models are highlighted in the first 3 rows of the table 2. Of these three models, ARIMA (2,1,3) x (0,1,1)24 is chosen as the best model after comparing other two models. These models picked up only daily seasonality and forecasted them i.e. the double peaks of count observed during working days. The forecast of ARIMA (2,1,2) x (1,1,1) 24 as shown in the figure below has very wide confidence intervals, thus removing from the candidate models.

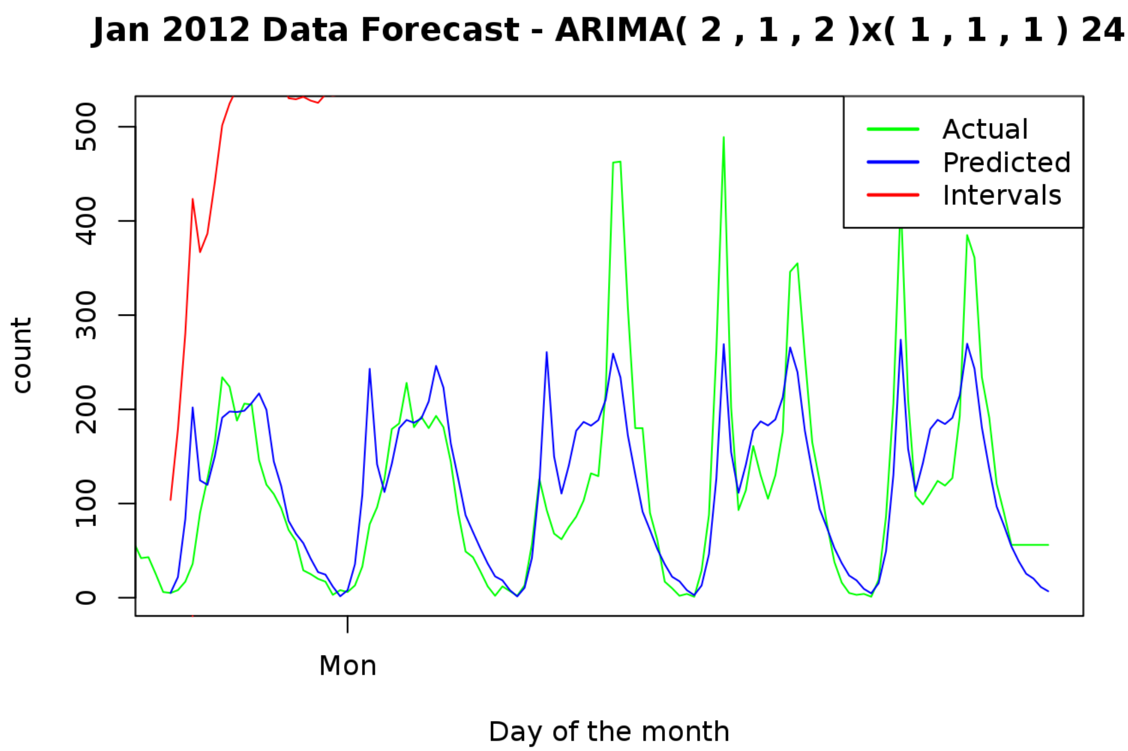

Figure 34 Forecast of ARIMA (2,1,2) x (1,1,1) 24

The next model under evaluation is ARIMA (3,0,4) x (1,1,1) 24. This model fits along the actuals and is smoothed at peak hours as shown in figure 35. Residuals however are independent with a p-value of 0.285 in box-test. The ACF and PACF plot in figure 36 show that residuals are uncorrelated. The figure 37 show that the residuals have mean close to zero. The Q-Q plot show that residuals do not follow normal distribution and are with lot of outliers, caused by underestimating the peaks in daily seasonality. Fitted vs residuals show high variance as the actuals increase, characteristics of heteroscedasticity.

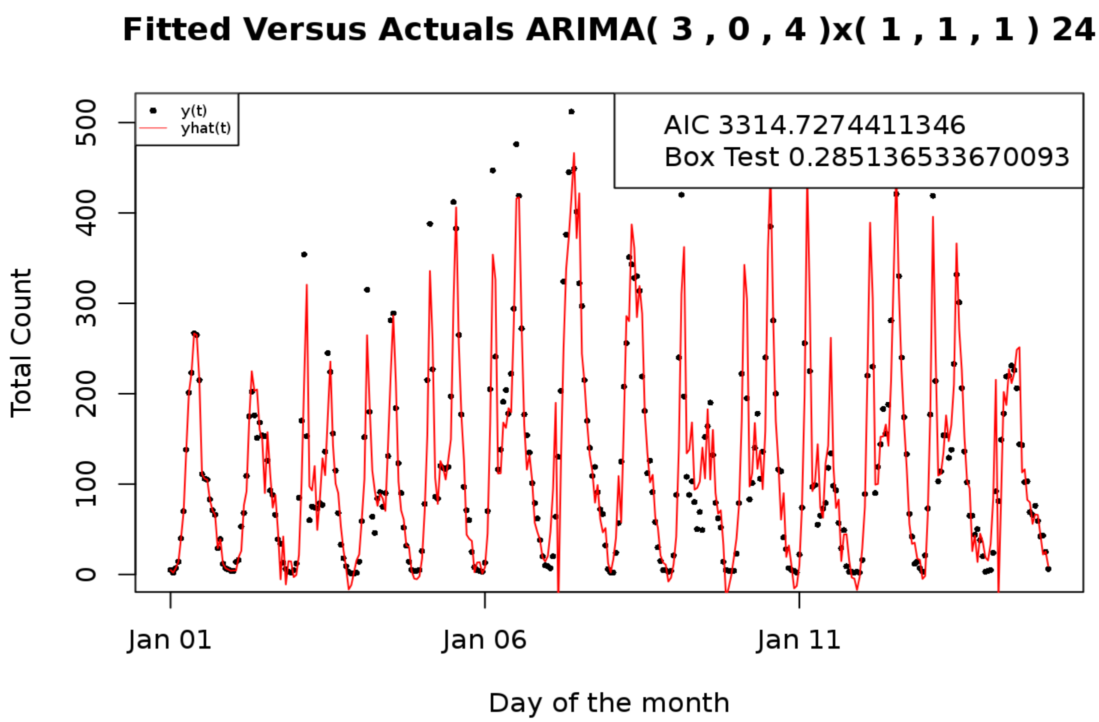

Figure 35 Model fit of ARIMA (3,0,4) x (1,1,1) 24 on Jan 2012

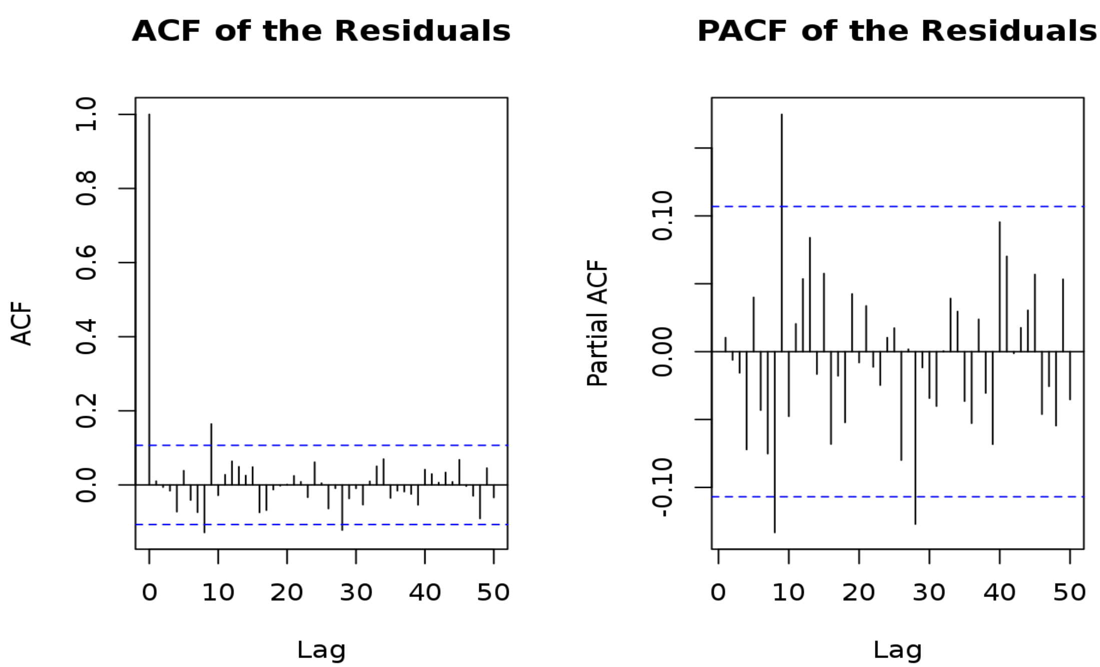

Figure 36 ACF and PACF of ARIMA (3,0,4) x (1,1,1) 24 residuals

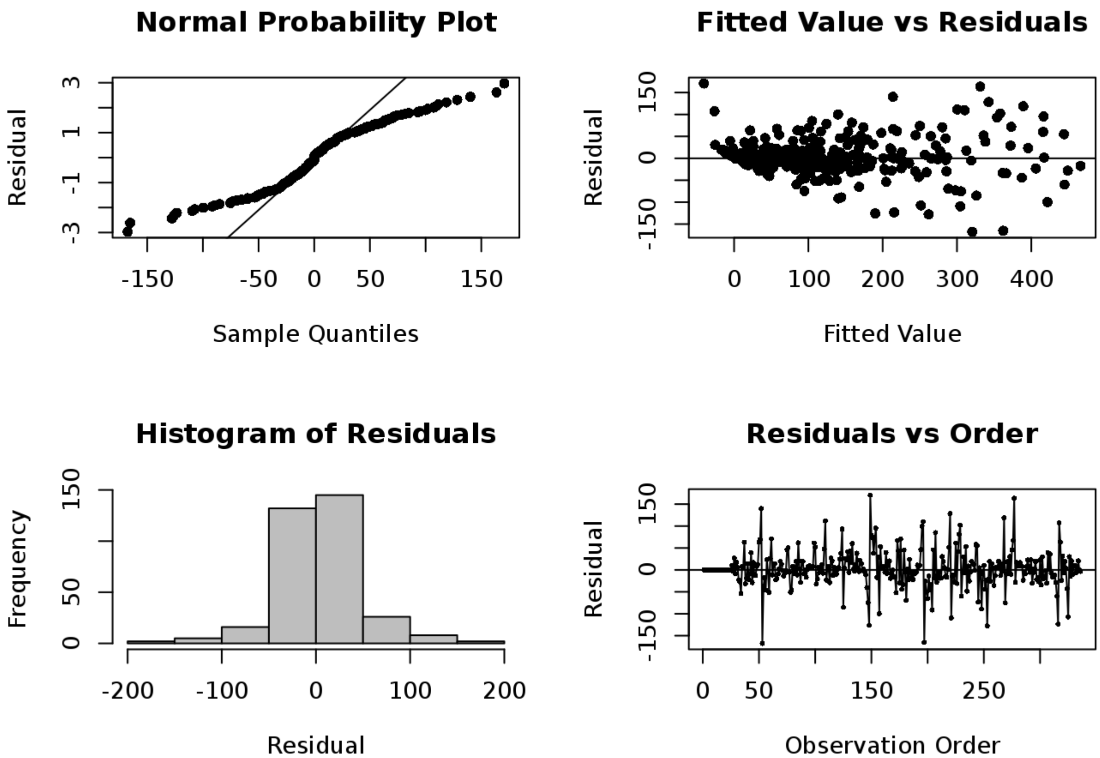

Figure 37 ARIMA (3,0,4) x (1,1,1) 24 Residuals

BEST MODEL

The best model from the above analysis is ARIMA (2,1,3) x (0,1,1) 24. This model fits along the actuals and although it is smoothed at peak hours, it fitted values are closer to actuals. Box-test revealed that the residuals are non-independent. The ACF and PACF plot in figure 39 show that residuals are uncorrelated and are within the intervals. The figure 40 show that the residuals have mean close to zero. The Q-Q plot show that residuals do not follow normal distribution and are with lot of outliers, caused by underestimating the peaks in daily seasonality. Fitted vs residuals show high variance as the actuals increase, characteristics of heteroscedasticity. This model is similar to the its competing model ARIMA (3,0,4) x (1,1,1) 24, but this model is chosen because of its forecast performance (MAPE), confidence intervals, and lesser number of parameters.

With log transformation, the same model was tested for model fits. But the residuals were auto correlated and had high variance at lower fitted values. Thus rejecting log transformation for ARIMA models.

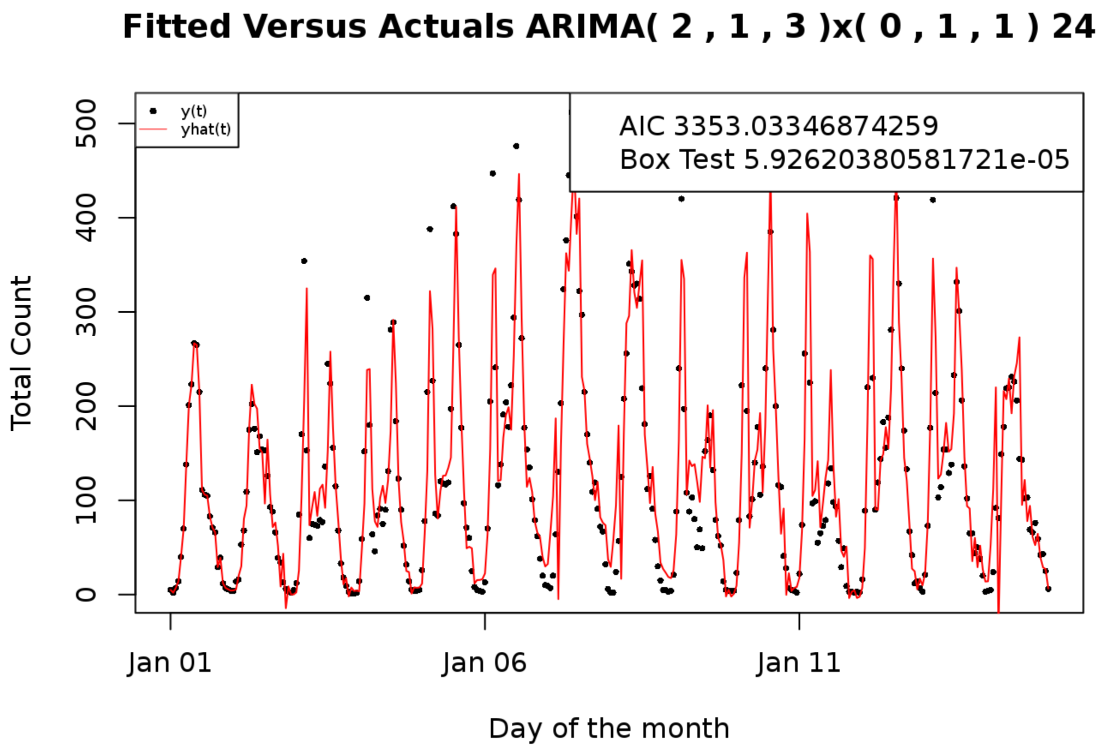

Figure 38 Jan 2012 model fit of ARIMA (2,1,3) x (0,1,1) 24

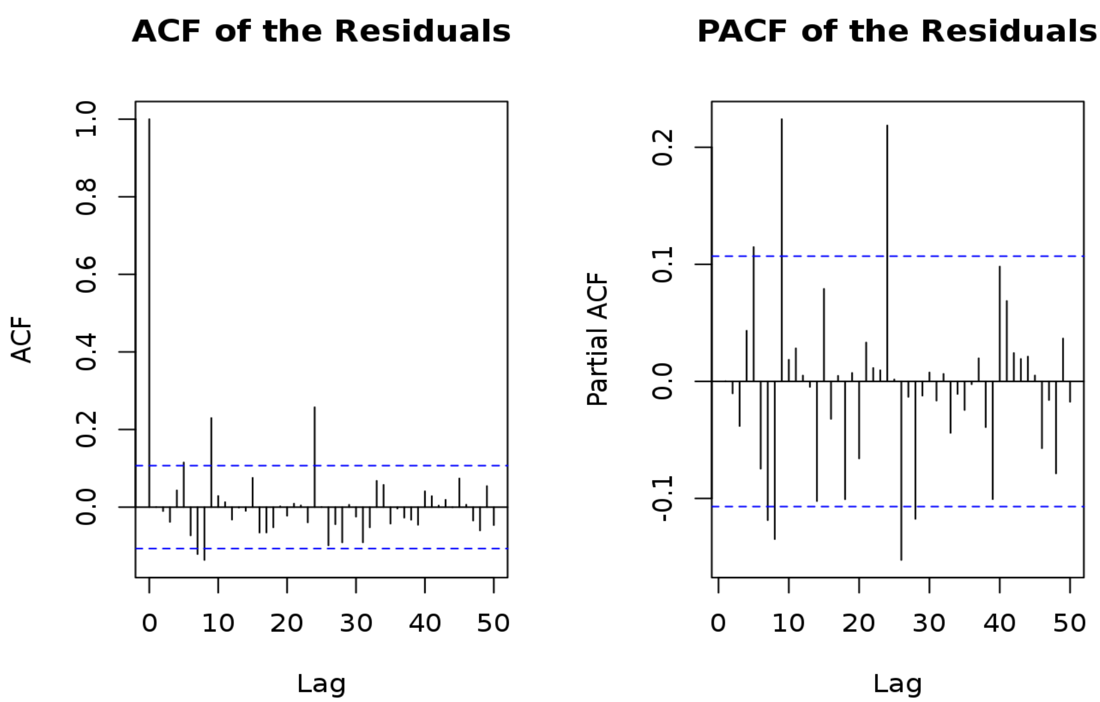

Figure 39 ACF and PACF of residuals of ARIMA (2,1,3) x (0,1,1) 24

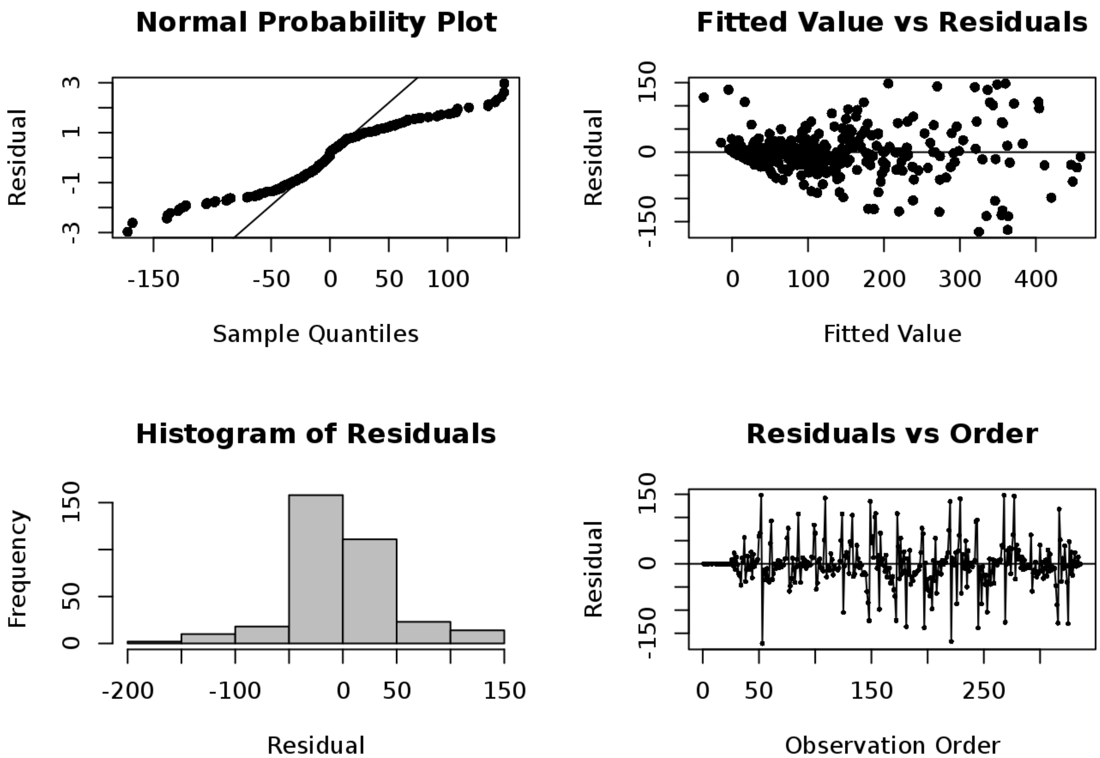

Figure 40 Residuals of ARIMA (2,1,3) x (0,1,1) 24

ARIMA (p, d, q) x (P, D, Q) s Forecasting

The below figures 41 and 42 show the forecasting using ARIMA (3,0,4) x (1,1,1)24. The confidence intervals for this forecast encompass most of the values except for peak values in working days. This model also picks up only working day pattern: double peaks since the seasonality provided was 24-hours. However, the values of these peaks are grossly underestimated with a MAPE of 85 %.

BEST MODEL FORECAST

The best model forecast chosen is from ARIMA (2,1,3) x (0,1,1) 24 as shown in figures 43 and 44. This model forecasts values that are characteristics of working days. This model forecast is very similar to ARIMA (3,0,4) x (1,1,1) 24. But its confidence intervals are better and encompass most of the values including peak counts. This model also grossly underestimates the peak counts with a MAPE of 81 %.

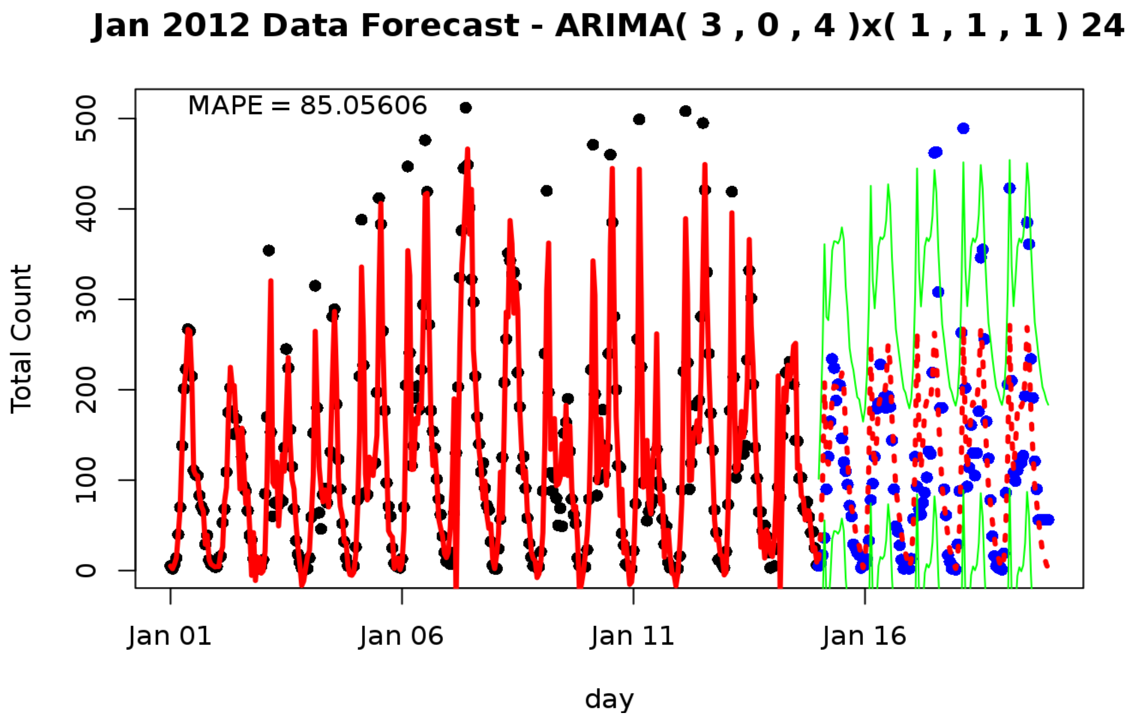

Figure 41 Forecast of Jan 2012 by ARIMA (3,0,4) x (1,1,1) 24

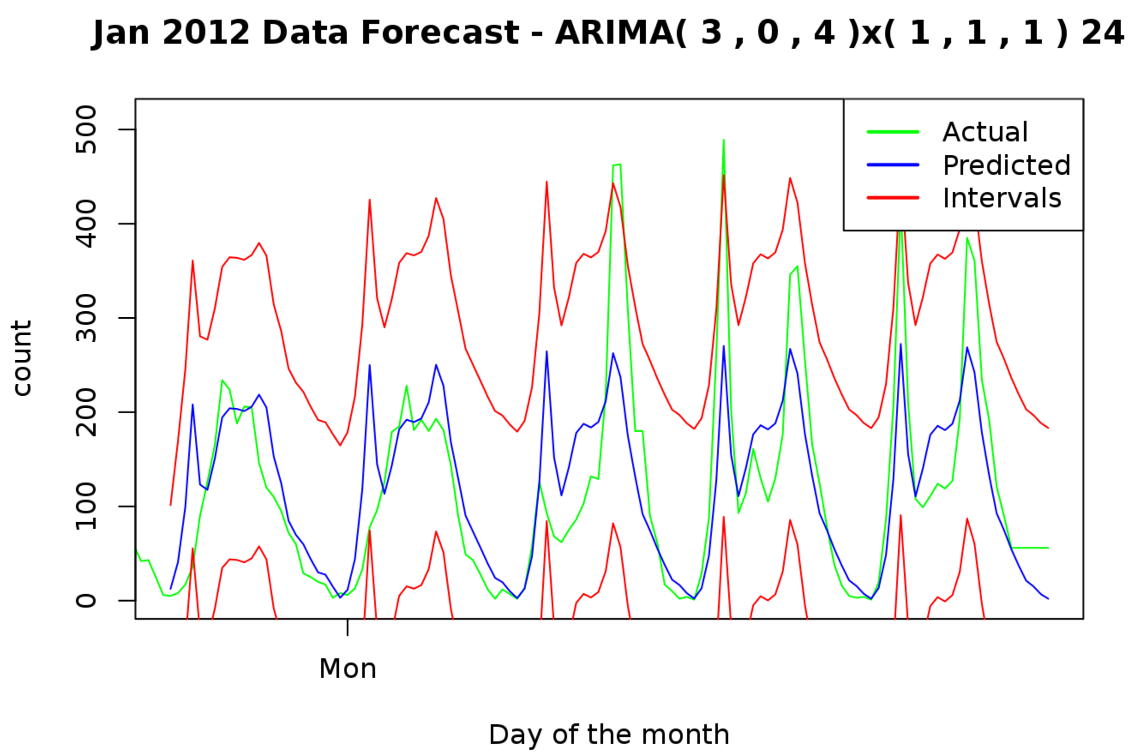

Figure 42 Forecast of ARIMA (2,1,3) x (0,1,1) 24 zoomed in

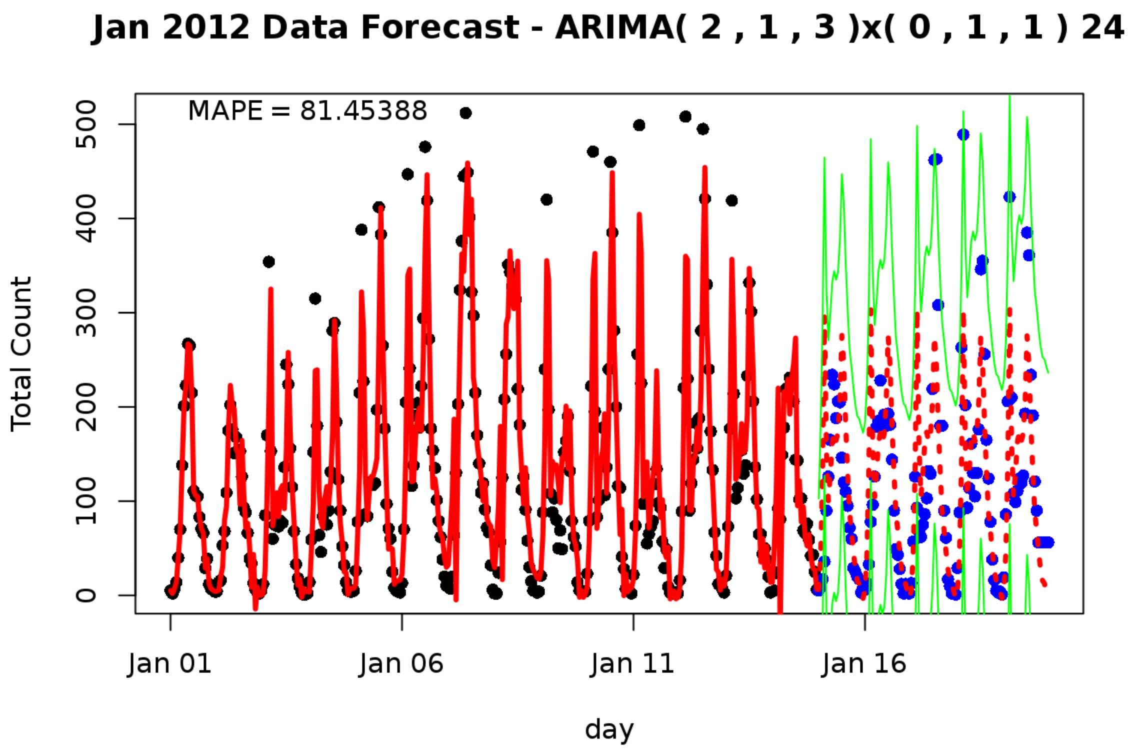

Figure 43 Forecast of Jan 2012 by ARIMA (2,1,3) x (0,1,1) 24

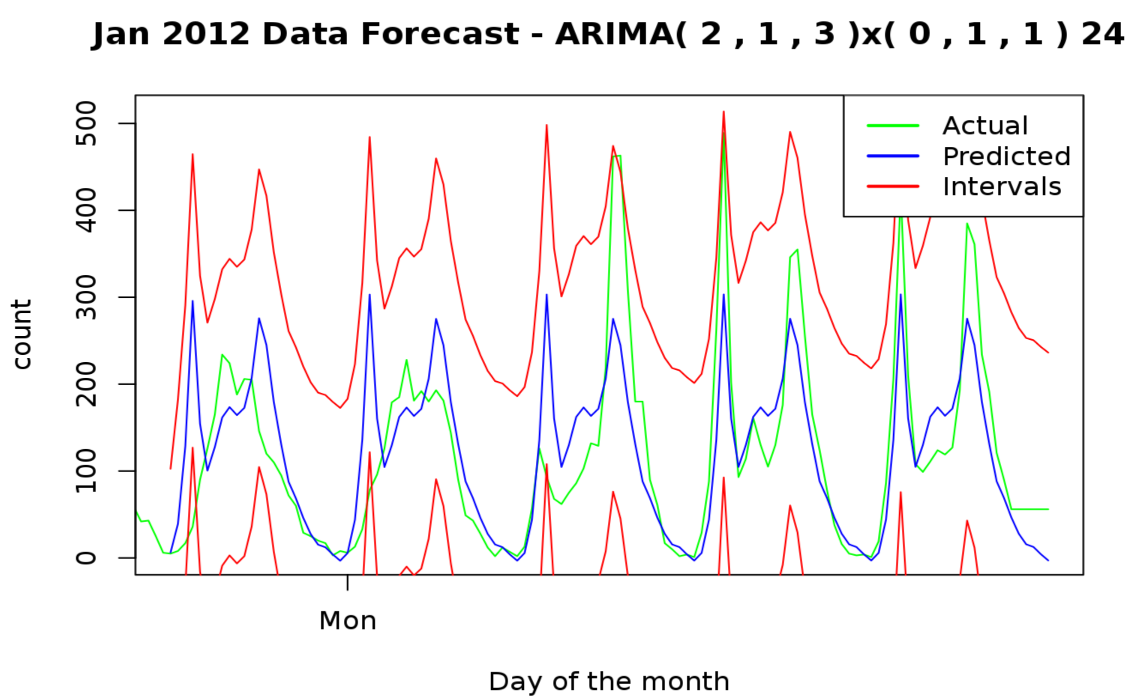

Figure 44 Forecast of ARIMA (2,1,3) x (0,1,1) 24 zoomed in

Model	Parameters	MAPE	Comments
Holts Winter	alpha = 0.3, beta = 0, gamma = 0.2	155.7948	Residuals are correlated and non-independent
ARIMA	ARIMA (2,1,3) x (0,1,1) 24	81.4538	Residuals are less correlated and independent. This is the best model so far with better forecasts.

Table 3 Summary of models

Dynamic Regression Models – Model Fitting

Since the data is already annotated with exogenous variables, it is easy to use them in ARIMA models as regressors. By viewing the total count over time, it is clear that the working days affect the pattern of ridership i.e. single peak in non-working days and double peaks in working days (commuter pattern) as shown in figure 45. Feels like temperature is used as a regressor while the actual temperature is removed because of high correlation between them. The relationship between feels like temperature and count is shown in figure 46. The table below summarizes the regressors used for dynamic regression models.

<table>
<tr><th>Type</th><th>Regressor</th><th>Comments</th></tr>
<tr><td>Discrete, Date time</td><td>Day of the week (7)</td><td>Created from Date time to represent 7 days of the week.</td></tr>
<tr><td rowspan="3">Continuous</td><td>Atemp</td><td>Feels like temperature in Celsius</td></tr>
<tr><td>Humidity</td><td>Relative humidity at the hour</td></tr>
<tr><td>Wind Speed</td><td>Wind Speed at the hour</td></tr>
<tr><td rowspan="6">Discrete</td><td>Holiday</td><td>Indicator of holiday</td></tr>
<tr><td>Working day</td><td>Indicates a day is neither holiday nor weekend</td></tr>
<tr><td>Clear</td><td rowspan="4">Weather attribute is transformed into four categorical variables representing Clear, Mist, Snow, and Heavy Rain weather.</td></tr>
<tr><td>Mist</td></tr>
<tr><td>Snow</td></tr>
<tr><td>Heavy Rain</td></tr>
</table>

Table 4 Regressors for Dynamic Regression

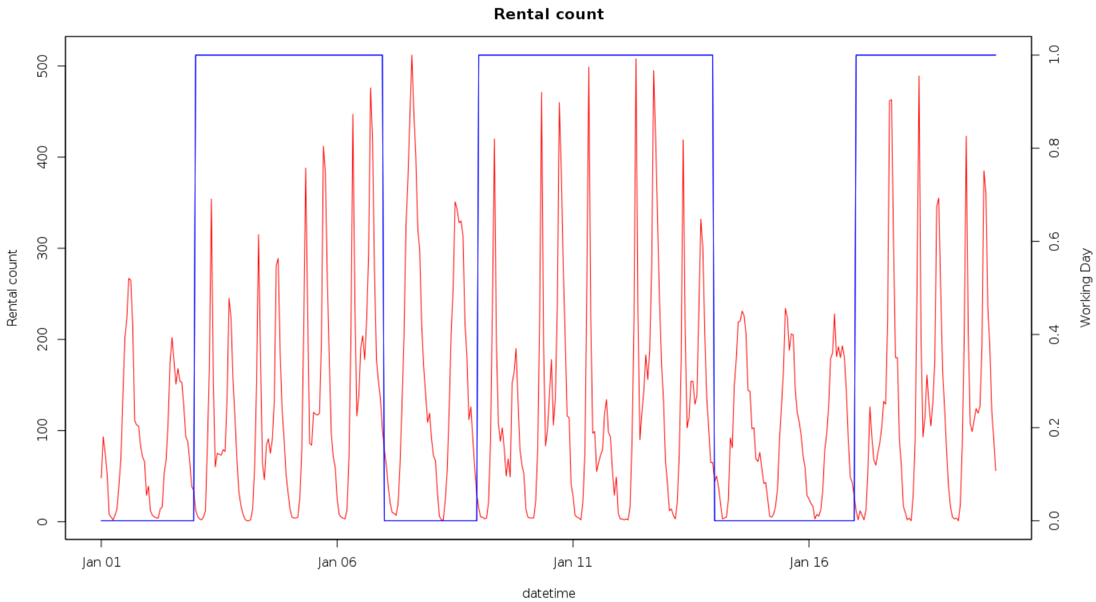

Figure 45 Relationship between Working day and Rental count Jan 2012

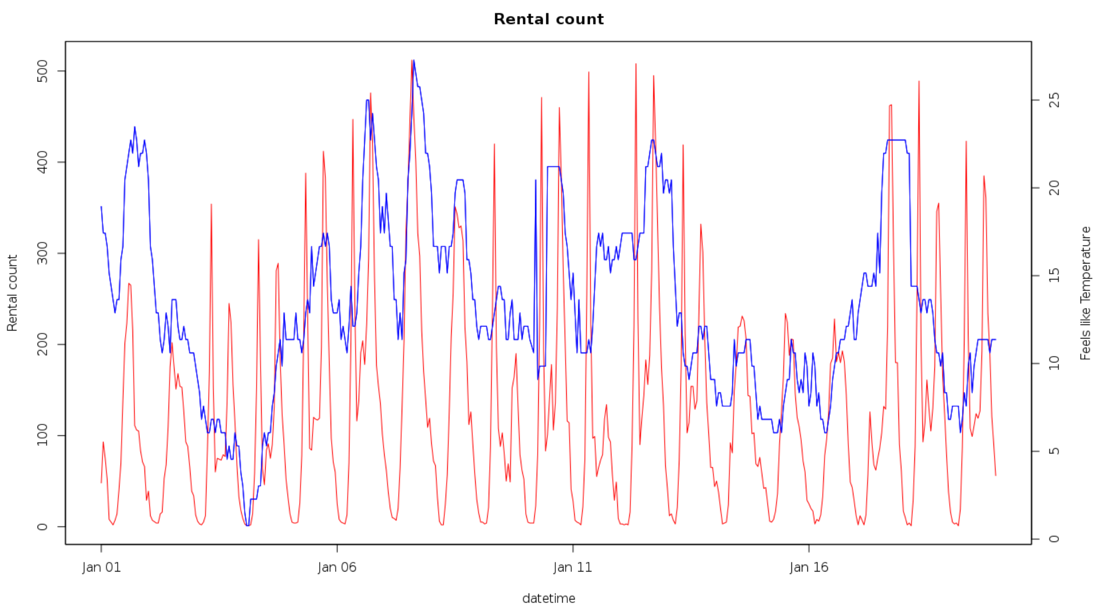

Figure 46 Relationship between feels like temperature and Rental count

The training and testing set remain the same throughout this analysis. Using different combinations of these regressors several models were created using auto.arima. The table below summarizes the models created with regressors. The ARIMA models will be analyzed with Day of the week, atemp and working day regressors as highlighted in the table below. The MAPE of this model is about 250%, thus further analysis has been done by varying the parameters of ARIMA.

Since the MAPE is too high for all these models, other models are created with lower parameters similar to ARIMA forecasts.

Regressors	Auto ARIMA Parameters	AIC
Day of the week, Atemp, Working Day	ARIMA (5,0,4) (1,0,0) 24	4847.45
Working Day, Atemp, Humidity, Wind speed	ARIMA (4,0,5) (1,0,0) 24	4843.65
Working Day, Atemp	ARIMA (4,0,5) (1,0,0) 24	4845.74
All regressors	ARIMA (0,0,0)	5336.62

Table 5 Combination of regressors

p	d	q	p	d	q	MAPE	AIC	Comments
3	1	2	1	1	1	118.143797	3334.132918	Candidate models with lower MAPE and better fits
3	1	3	1	1	1	118.890432	3327.739923	
3	1	4	0	1	1	119.3738494	3361.568977	
2	1	4	0	1	1	122.6747513	3365.953821	Rejected models due to higher MAPE and bad model fits and forecasts.
3	1	2	0	1	1	126.0365228	3359.622465	
2	1	3	1	1	1	135.5837352	3334.466326	
3	1	4	1	1	1	143.406902	3328.601714	
2	1	4	1	1	1	143.5637294	3326.602438	
2	1	3	1	1	0	149.7895138	3389.102214	
3	1	2	1	1	0	158.1929832	3380.14296	
3	1	3	0	1	0	159.2034806	3397.667049	
3	1	3	1	1	0	159.3786859	3391.081849	
2	1	3	0	1	0	160.2991237	3395.677198	
2	1	4	1	1	0	161.5326306	3377.121025	
2	1	3	0	1	1	165.3642971	3356.271813	

Table 6 MAPE analysis of Dynamic Regression models

The above table 6 summarizes the models analyzed with Day of the week, working day, and feels like temperature as regressors. Several models estimated poorly without any pattern. Even though Days of the week are used as regressors, these models could not pick up weekly seasonality or working day pattern. These regressors were not useful in forecasting for January 2012. It could also be because of lesser number of samples to fit the model.

Candidate models from the above analysis are highlighted in table 6. Of which, the ARIMA (3,1,2) x (1,1,1) 24 model underestimated most of the values, as shown in the figure below. Thus rejecting this model.

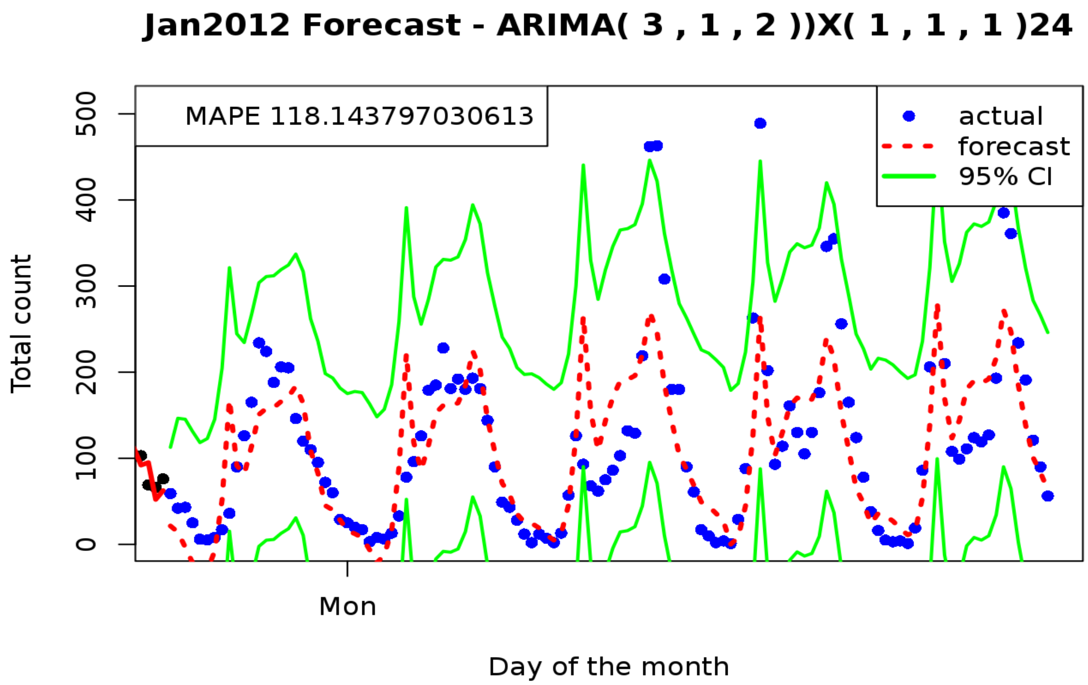

Figure 47 Forecast of ARIMA (3,1,2) x (1,1,1) 24 with regressors

The competing model is ARIMA (3,1,3) x (1,1,1) which has a MAPE of 118%. This model fits values closer to the actuals.

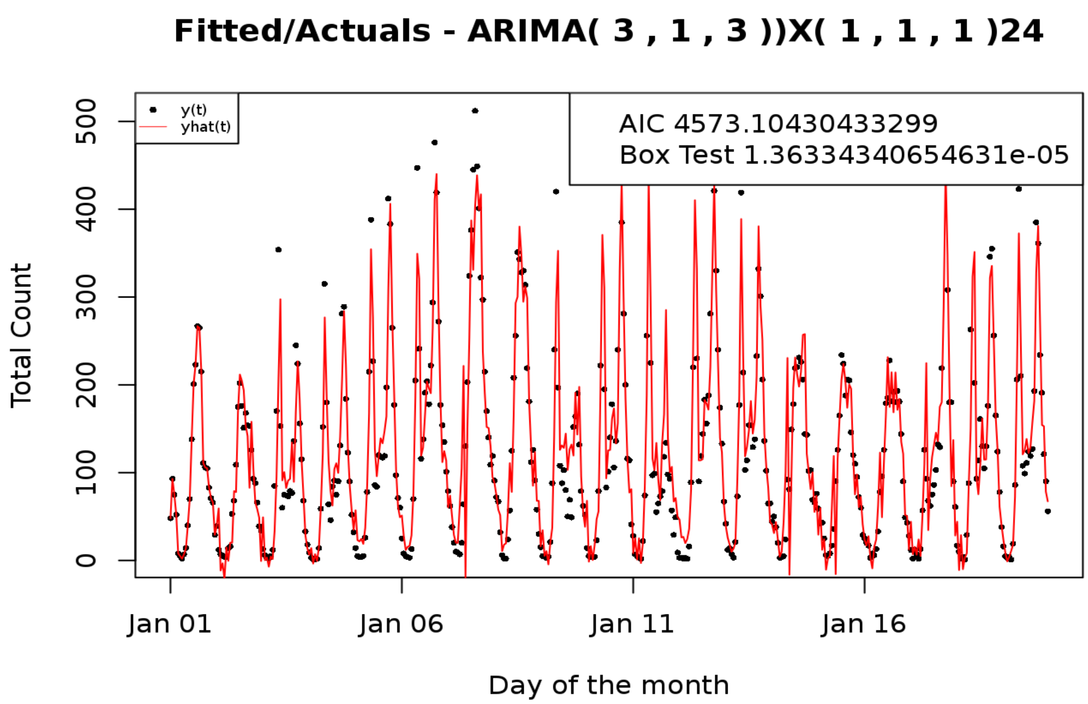

Figure 48 Model Fit of ARIMA (3,1,2) x (1,1,1) with regressors

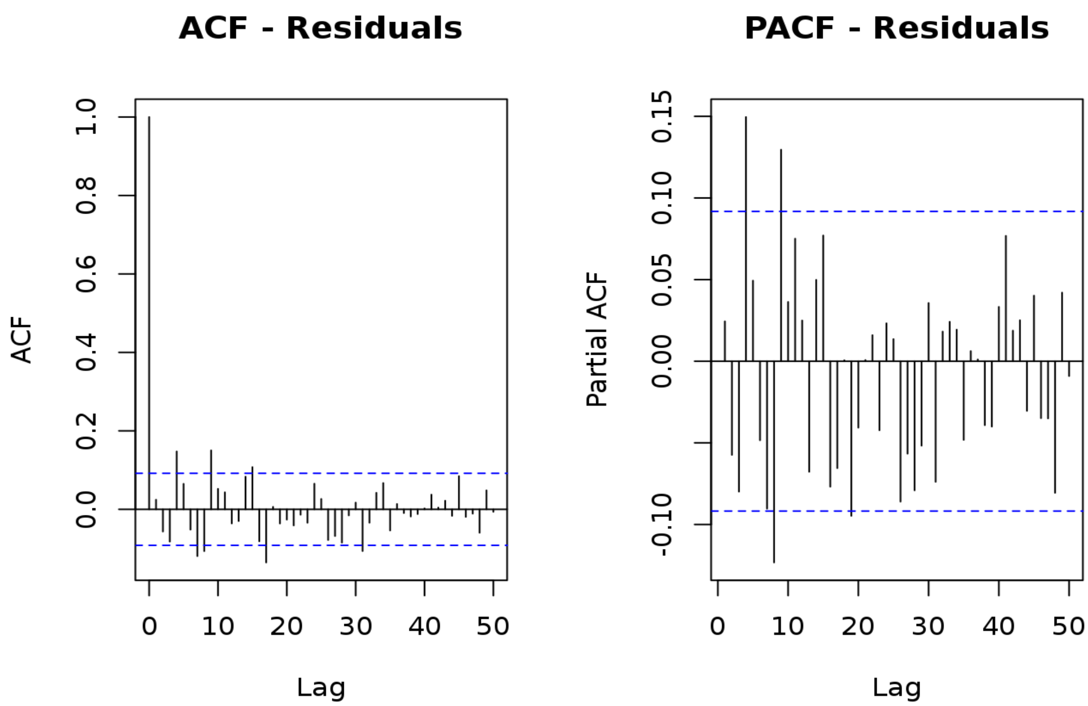

Figure 49 ACF and PACF of Residuals

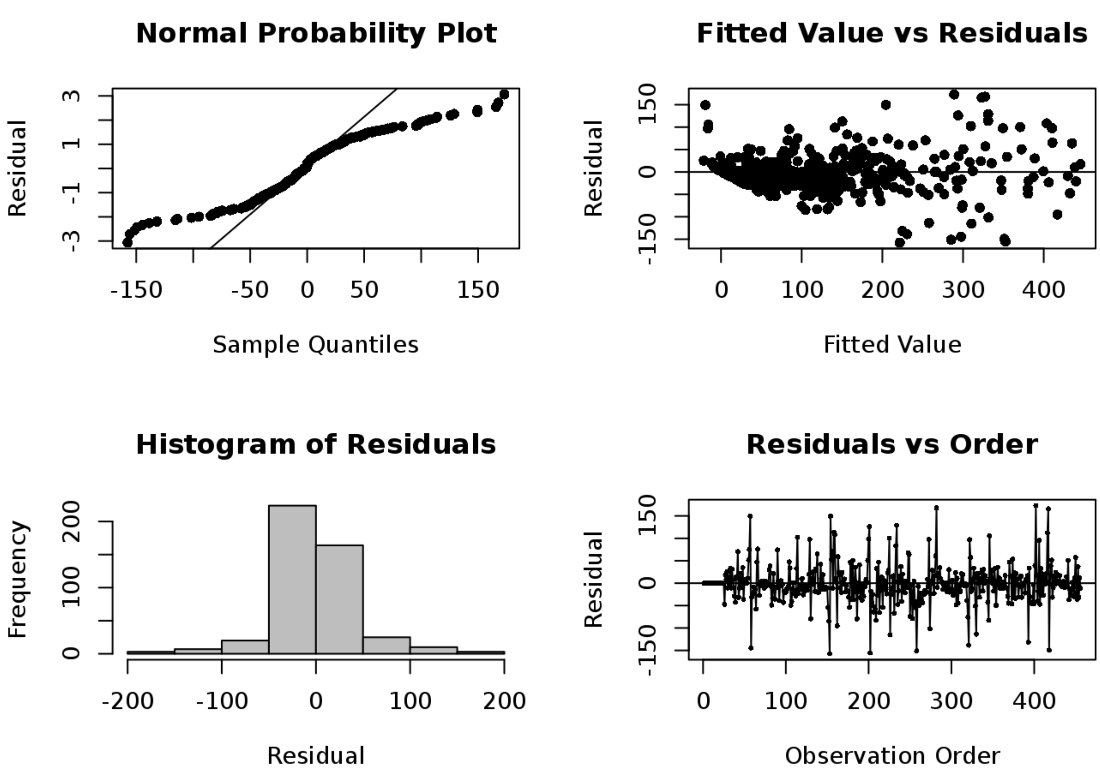

Figure 50 Residual of ARIMA (3,1,3) x (1,1,1) 24 with regressors

This model has correlated residuals as shown in figure 49 but ACF and PACF are within in the intervals. The residual plots from figure 50, show that the residuals have mean closer to zero. But there is still heteroscedasticity as there is high variance when actuals increase in value. These residuals are normally distributed with outliers on both tails.

BEST MODEL

The best model from the dynamic regression is ARIMA (3,1,4) x (0,1,1) 24. This model fits values closer to the actuals similar to the previous model.

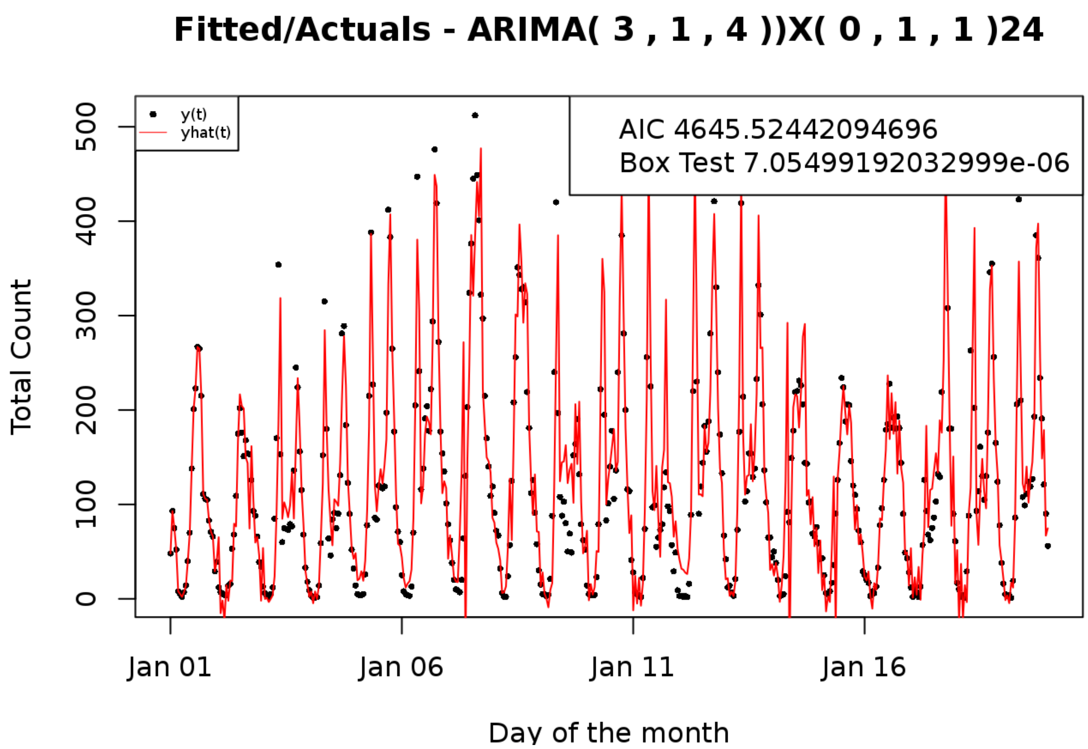

Figure 51 Fitted vs Actuals ARIMA (3,1,4) x (0,1,1) 24 with regressors

The residuals show some correlation between the residuals and are non-independent of each other. This model shows similar characteristics of residuals i.e. mean closer to zero, high variance at increasing values of count (heteroscedasticity), non-independent residuals, and normally distributed with outliers at both tails. The differentiation between these models is the forecasted values. This model estimates better than ARIMA (3,1,3) x (1,1,1) 24 i.e. the less underestimation.

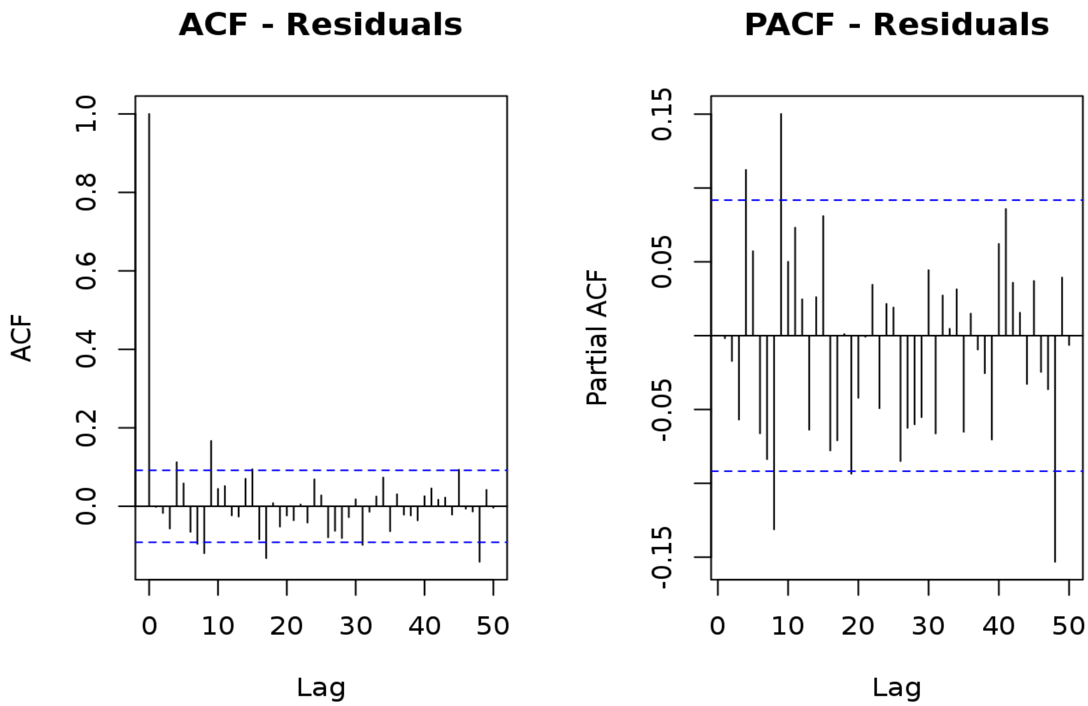

Figure 52 ACF and PACF of residuals

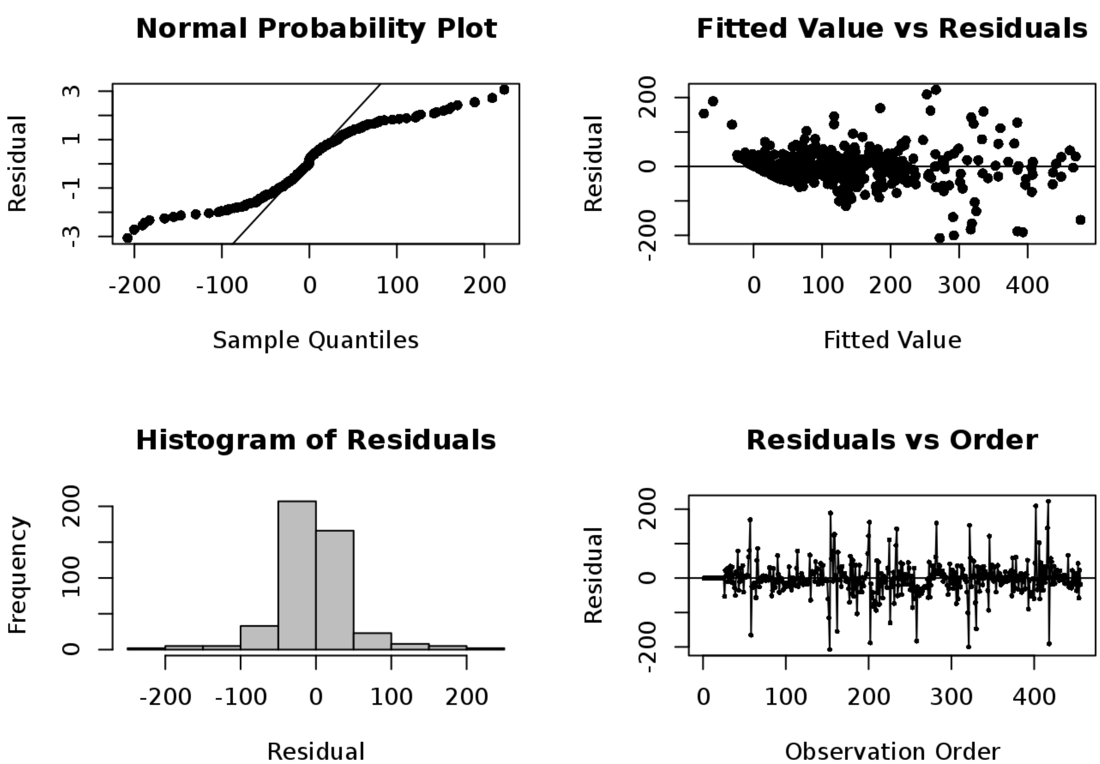

Figure 53 Residual of ARIMA (3,1,4) x (0,1,1) 24 with regressors

LAGGED PREDICTORS

Of the regressor variables, the feels like temperature is the continuous value. Thus using it as lagged predictors. The figure 46 shows the feels like temperature against rental count. It can be seen that the whenever temperature is very low, the ridership is low. Thus, using feels like temperature as the lagged predictor models were analyzed as shown in table below. Of which the best model is the one with 1 lag (AIC 3239.237), highlighted in the table below. This lagged predictor is used in ARIMA (2,1,3) x (0,1,1) 24. It's model fit is shown in the figure 54. While this model has a good fit along the actual values, the MAPE for the forecasted period is 211 %. This performance is low compared to the ARIMA (2,1,3) x (0,1,1) 24 without any predictors.

Lag 0	Lag 1	Lag 2	Lag 3	AIC
1	0	0	0	3747.688
1	1	0	0	3739.237
1	1	1	0	3740.594
1	1	1	1	3742.305

Table 7 ARIMA with Lag Predictors

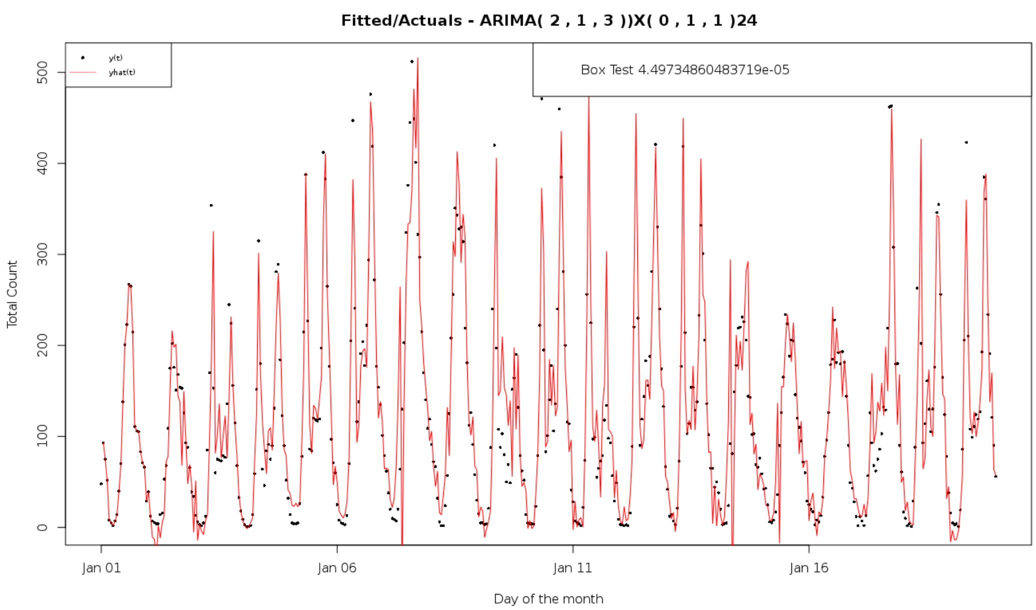

Figure 54 Model fit of ARIMA (2,1,3) x (0,1,1) 24 with lagged predictors

Dynamic Regression Models – Model Performance

From analyzing various dynamic regression models, models with regressors perform better than models with lagged predictors. The forecast period has Martin Luther King holiday on January 16th 2012. This day follows a non-working day pattern. Since all other models picked up only daily seasonality of double peaks in the forecasts, it is expected that by including working day, day of the week regressors should improve forecasts. The models built using these regressors should be able to pick up a single peak on non-working days and double peaks on working days.

The figures below show the forecasts of Jan 15th – Jan 19 2012. From figure 55, the forecast by ARIMA (2,1,3) x (0,1,1) 24 with feel like temperature at lag 0 and 1 as predictors, it is similar to the ARIMA (2,1,3) x (0,1,3) but the MAPE is very high with 211 %. This model also does not pick up any working or non-working day pattern.

The figure 56 shows the forecast of competing model ARIMA (3,1,3) x (1,1,1) 24 with Day of the week, working day, and feels like temperature. This model underestimates most of the values i.e. even at single peak pattern (non-working day), the predicted values are low. While the best model, ARIMA (3,1,4) x (0,1,1) 24 with same regressors, does not seem to have underestimation for many values, but this model does not pick up any non-working day pattern. None of the regressors seem to help, possibly because of very few samples. Given more than two weeks of data, it is possible that the dynamic regression models would work.

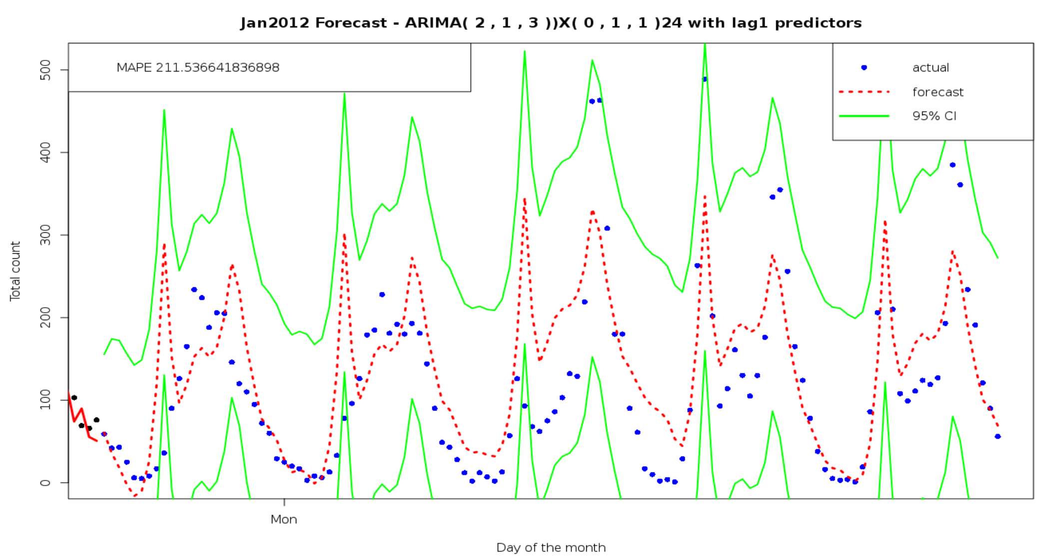

Figure 55 Forecasts with ARIMA (2,1,3) x (0,1,1) 24 with lagged predictors

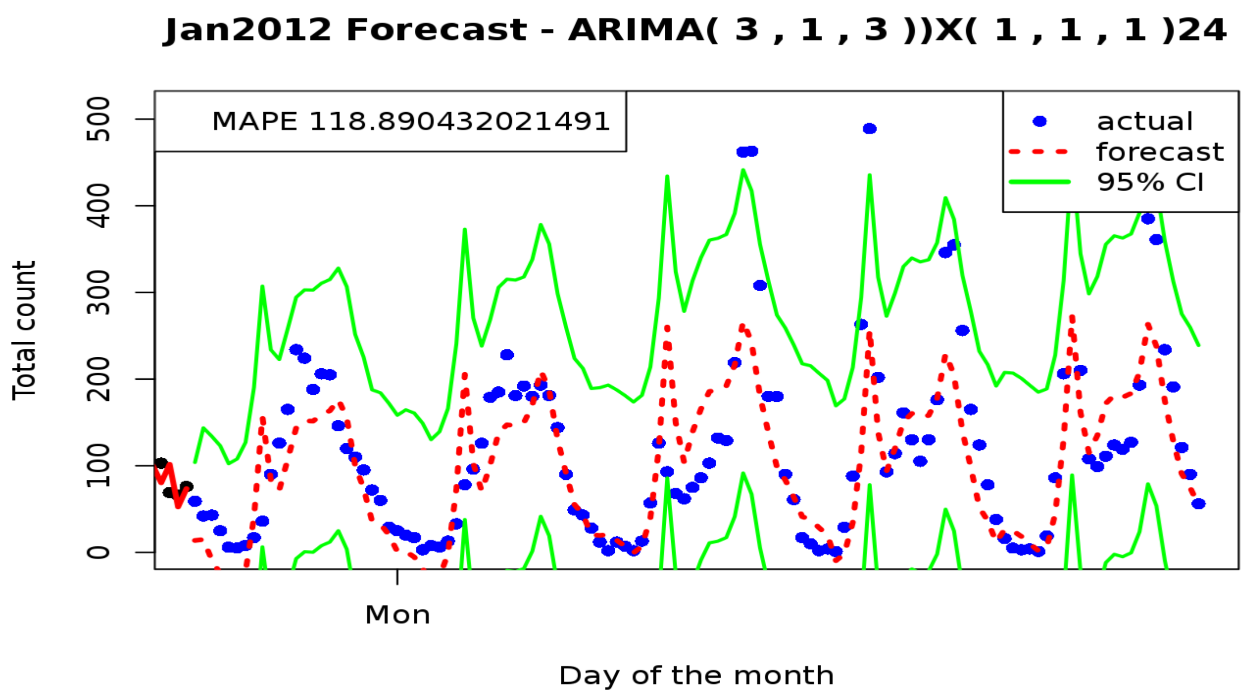

Figure 56 Forecast of ARIMA (3,1,3) x (1,1,1) 24 with regressors

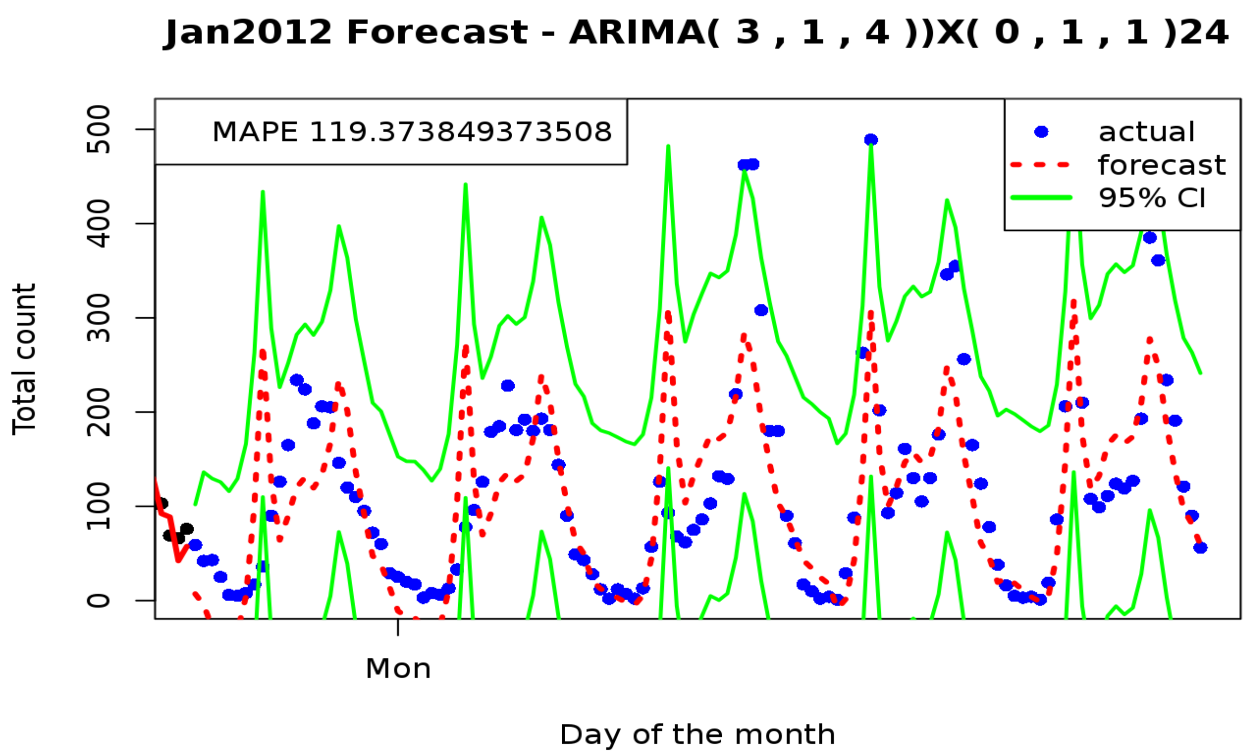

Figure 57 Forecast of ARIMA (3,1,4) x (0,1,1) 24 with regressors

Best model

The best model so far is the ARIMA (2,1,3) x (1,0,1) 24 since its MAPE is very low compared to other models. Further improvements can be done by analyzing advanced models such as neural networks. The table below summarizes the best models from other types of models.

Model	Parameters	MAPE	Comments
Holts Winter	alpha = 0.3, beta = 0, gamma = 0.2	155.7948	Residuals are correlated and non-independent
ARIMA	ARIMA (2,1,3) x (0,1,1) 24	81.4538	Residuals are less correlated and non-independent. This is the best model so far with better forecasts.
Dynamic Regression - Regressors	ARIMA (3,1,4) x (0,1,1) 24 with Day of the week, Working Day, and Feels like temperature	119.37	Residuals are correlated and non-independent
Dynamic Regression - Lagged Predictors	ARIMA (2,1,3) x (0,1,1) 24 with 1 lag of Feels like Temperature	211.53	Residuals are correlated and non-independent

Table 8 Model Summary

Neural Networks

DATASET CREATION

For the neural network model, the entire dataset is used i.e. data from all months and not just the subset of each month. Thus, 10944 records of total bike share can be used to build the model. The exogenous variables are Seasons, Weather, Holiday, working day, Temperature, Feels like temperature, Wind speed, and Humidity. The below table summarizes the repressors created for this dataset.

Type	Regressor	Comments
Discrete, Date time	Hour of the day (24)	Regressors extracted from Datetime attribute represent 24 hours of 7 days in a week over 12 months in 2 years.
	Day of the week (7)	
	Month (12)	
	Year (2011, 2012)	
Continuous	Temp	Temperature in Celsius
	Atemp	Feels like temperature in Celsius
	Humidity	Relative humidity at the hour
	Wind Speed	Wind speed at the hour
Discrete	Holiday	Indicator of holiday
	Working day	Indicates a day is neither holiday nor weekend
	Spring	Season attribute is transformed into four categorical variables representing Spring, Summer, Fall, and Winter.
	Summer	
	Fall	
	Winter	
	Clear	Weather attribute is transformed into four categorical variables representing Clear, Mist, Snow, and Heavy Rain weather.
	Mist	
	Snow	
	Heavy Rain	

Table 9 Regressors

Correlation Matix

	temp	atemp	humidity	windspeed
temp	1.00000000	0.98506910	-0.06673862	-0.01916251
atemp	0.98506910	1.00000000	-0.04564774	-0.05897954
humidity	-0.06673862	-0.04564774	1.00000000	-0.31609717
windspeed	-0.01916251	-0.05897954	-0.31609717	1.00000000

Figure 58 Correlation between continuous regressors

Figure 58 shows the correlation matrix between continuous variables in the dataset i.e. temp, atemp, humidity, and windspeed. The temp and atemp are highly correlated. The variable atemp is retained since it represents the feels like temperature, thus removing temp variable. All of these continuous variables and total count have been normalized to [0, 1] intervals.

MODEL FITTING

Training and test set

For this neural network model, Jan 15th to Jan 19th 2012 is treated as forecast period i.e. test set. It has 120 data points. This period has Jan 16th, a Monday which is a Martin Luther King holiday. This day does not follow the double peaks from other weekdays since it is not a working day. The figure below shows the forecast period in red and the training set in blue.

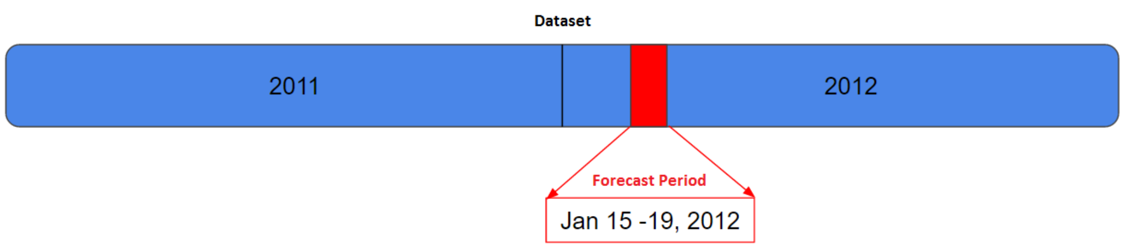

Figure 59 Training and test set

The figure below shows the single peak pattern on Monday, Jan 16th 2012. The training set consists of 10,824 records and test set has 120 records.

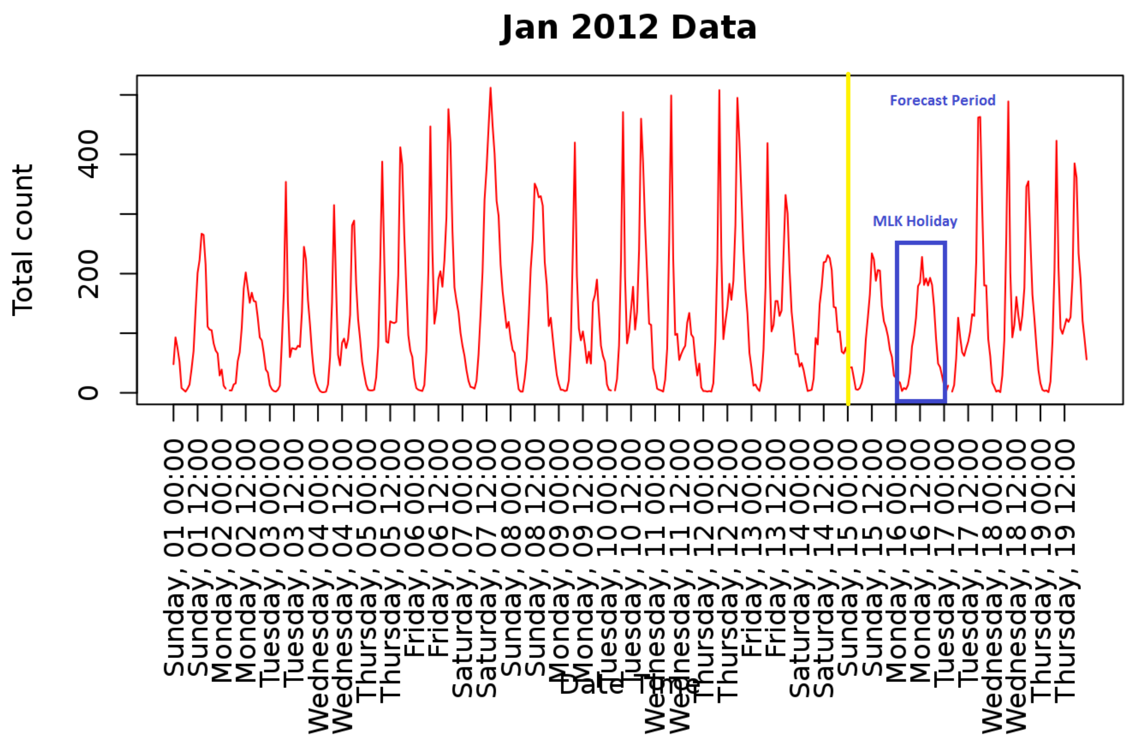

Figure 60 Jan 2012 Forecast period

MSE and MAPE vs Hidden nodes

By using MSE (Mean Squared Error) and MAPE (Mean Absolute Percentage Error), neural networks with different number of hidden nodes over different decay parameters are evaluated. These models are created with all of the regressors. A number of neural networks are created with decay parameters: 0.01, 0.05, and 0.1 with hidden nodes ranging from 1 to 20. MSE and MAPE are calculated for the forecast period. By plotting the number of hidden nodes against MSE and MAPE with different decay parameters, candidate neural networks are found. The figures 61, and 62 show MSE against hidden nodes and decay rate, and MAPE against hidden nodes and decay rate respectively.

From figure 61, MSE shows an elbow plot, where neural networks with more than 3 hidden nodes have very low MSE and does not show drastic change. From figure 62, MAPE shows oscillation with different number of hidden nodes with different decay rate. The possible candidates from MAPE versus hidden nodes and decay rate are neural networks with 3, 4, and 5 hidden nodes with 0.01 decay parameters. These neural networks are built with 500 iterations.

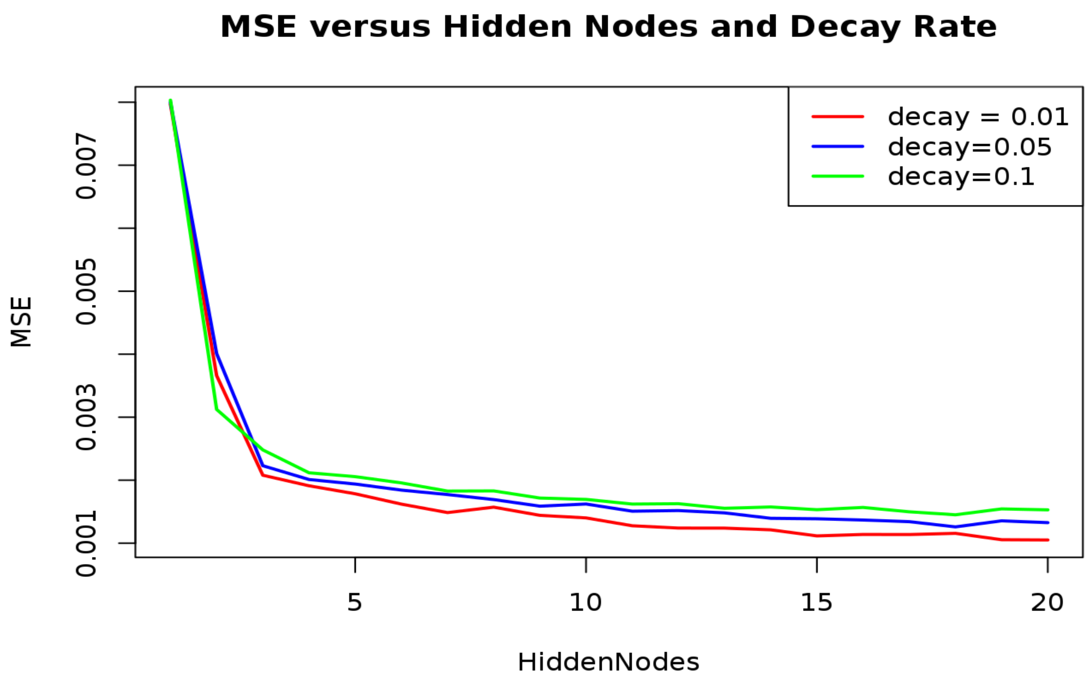

Figure 61 MSE vs Hidden nodes and Decay Rate

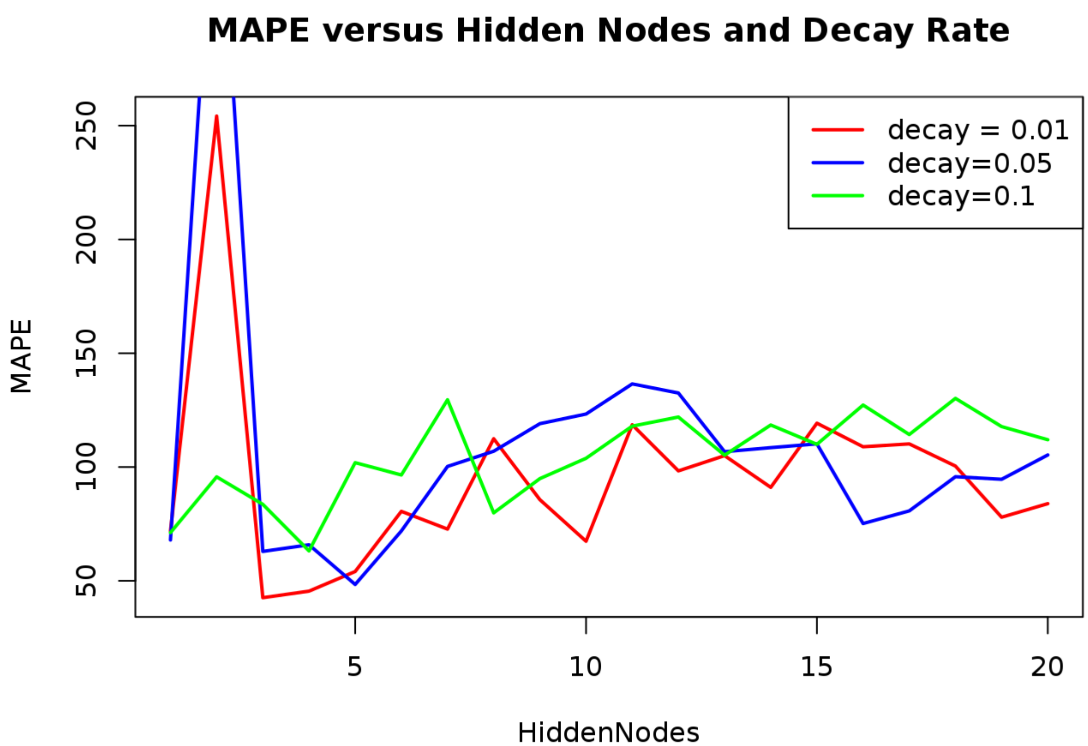

Figure 62 MAPE vs Hidden nodes and Decay Rate

MSE and MAPE vs Iterations

By using the candidate neural networks built with 500 iterations, MSE and MAPE were calculated for different number of iterations to check if these models performed better. Thus, for iterations ranging from 100 to 1000, MSE and MAPE were calculated and plotted for the candidate models with hidden nodes (3, 4, 5) with decay rate of 0.01 as shown in the figures below.

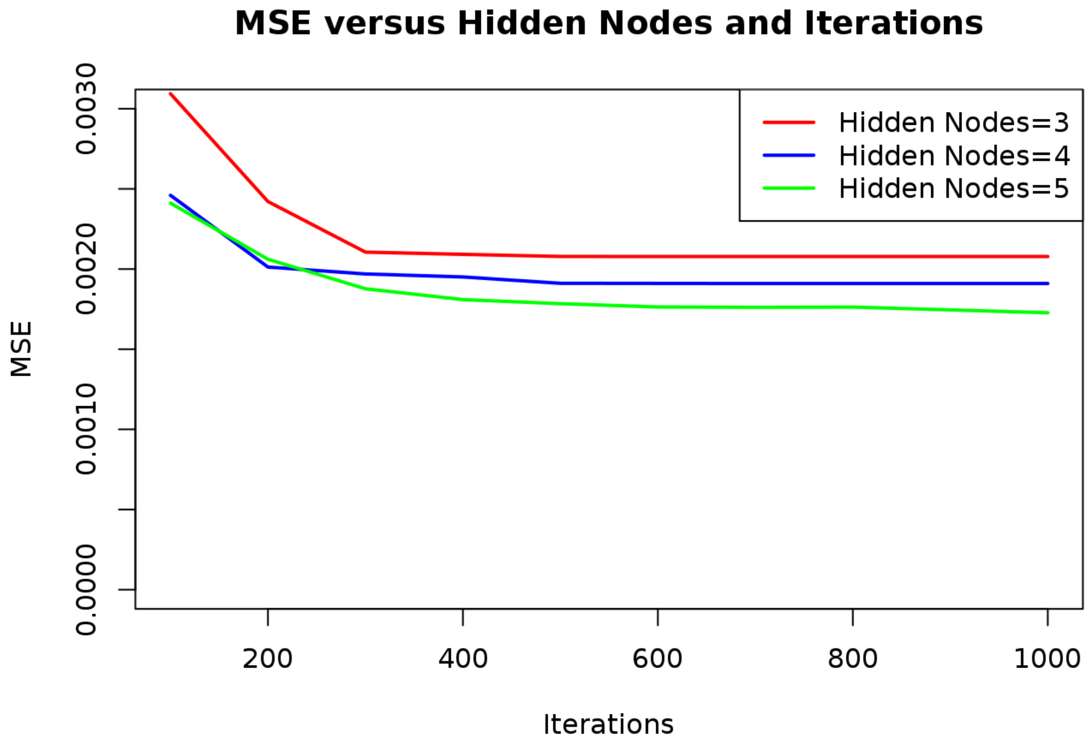

Figure 63 MSE vs Hidden nodes and Iterations

From figure 63, it can be observed that the MSE is stable for models with 3, 4, and 5 hidden nodes after 200 iterations. Figure 64, shows similar characteristics of MSE vs hidden nodes and iterations, where MAPE of all models drop after 200 iterations. The models with hidden nodes 3 and 4 stabilize after 600 iterations. Neural network with 4 hidden nodes, 0.01 decay parameter, and 400 iterations perform well with a MAPE of 41.4%. Thus choosing this model.

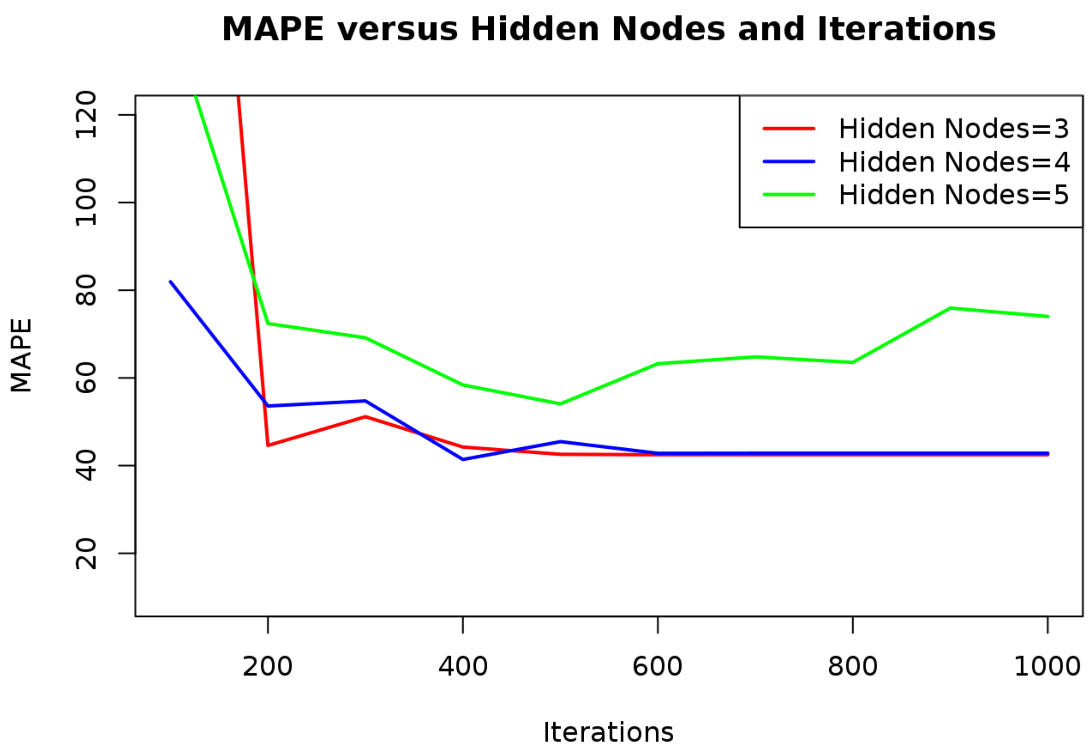

Figure 64 MAPE vs Hidden nodes and Iterations

Different regressors

The above models are created with all regressors. Since, regressors extracted from date time, and working day are important, they were not removed from any models. The table below summarizes the regressors which are removed and tested for neural network with 4 hidden nodes, 0.01 decay rate and 400 iterations.

Regressors removed	MAPE	Fit
Atemp, humidity, windspeed	54.79%	Decent fit.
Weather (Clear, Mist Snow, Heavy Rain)	48.88%	Decent fit.
Holiday	49.95%	Decent fit.
Seasons (Spring, Summer, Fall, Winter)	118.08%	Wavy fit, the forecasted points followed a wavy pattern along the actuals.

Best model

Including all of the regressors in the neural network (h = 4, d = 0.01, maxit = 400) gives a MAPE of 41.4%. The figure 65 below shows the model fit on Jan 2012 data. The fitted points follow close to the actual data except few overestimations at peak hours.

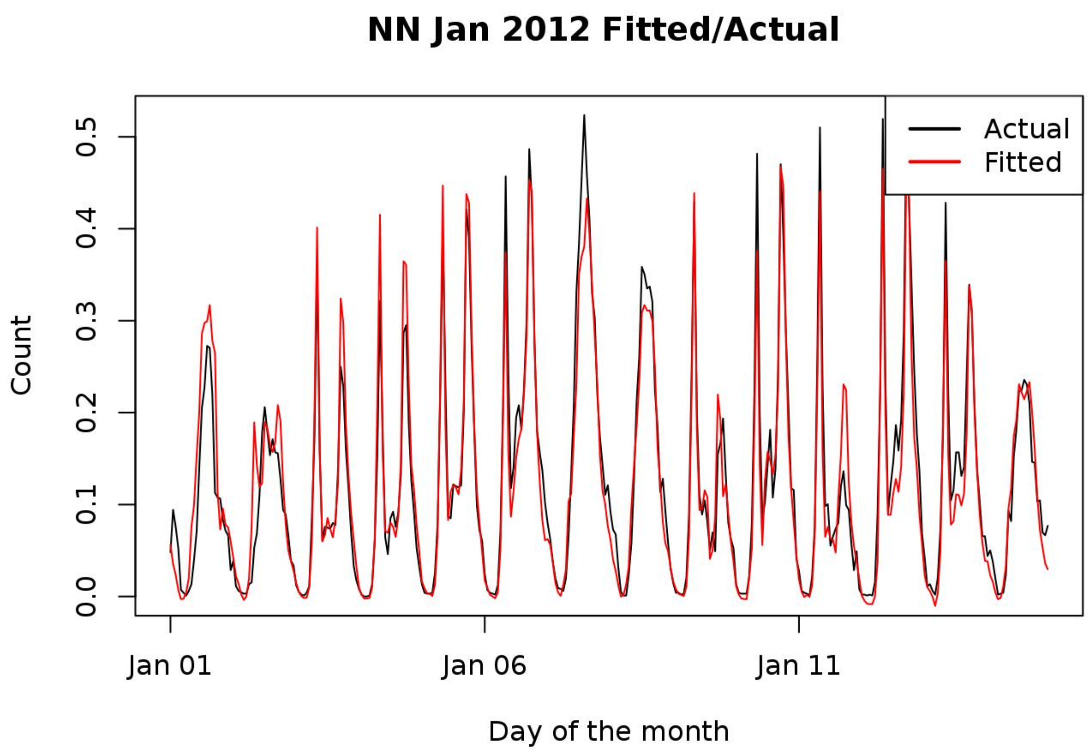

Figure 65 Jan 2012 NN Model fit

Training set Residuals

Figure 66 below shows the residuals of 14 days of Jan 2012. The residuals look normally distributed except for few outliers caused by overestimating the peak hour counts. Histogram of the residuals show the left tail a little longer. The mean of residuals is close to zero. Fitted vs Residuals show high variance with increase in fitted value, characteristics of heteroscedasticity. Figure 67 show ACF and PACF of the residuals. The ACF shows more positive correlation but the values are mostly within the intervals. The PACF also shows that the values are within the intervals.

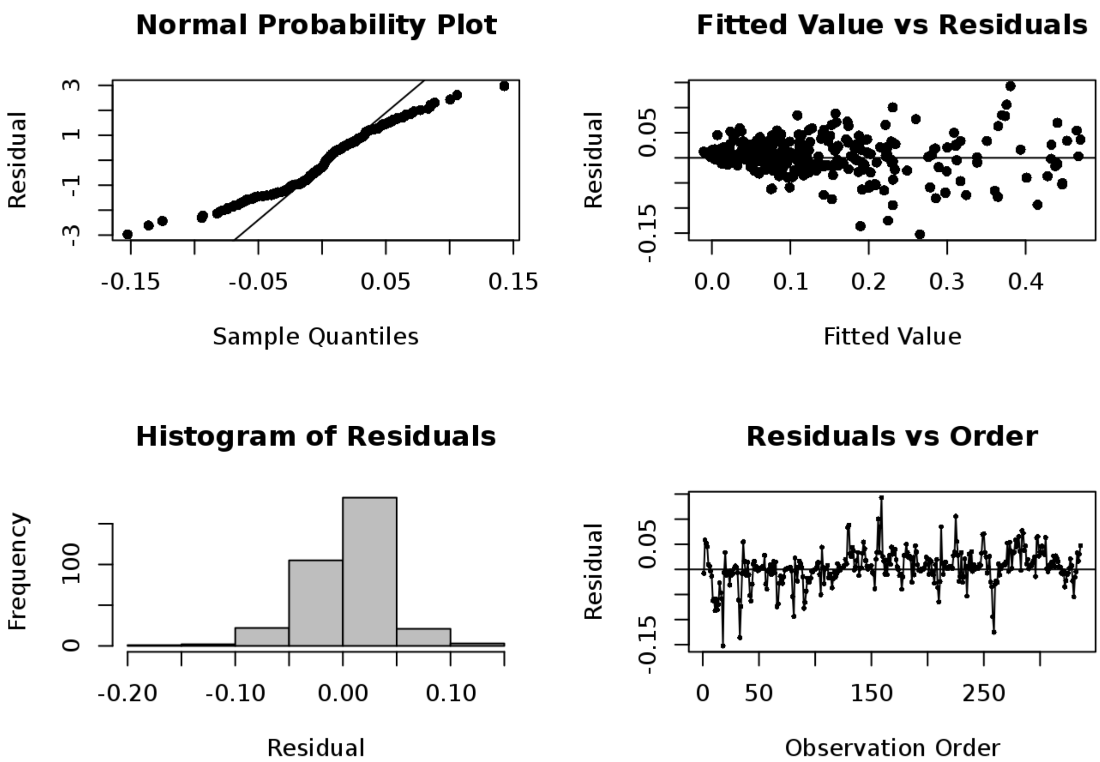

Figure 66 Residuals of first 14 days of Jan 2012

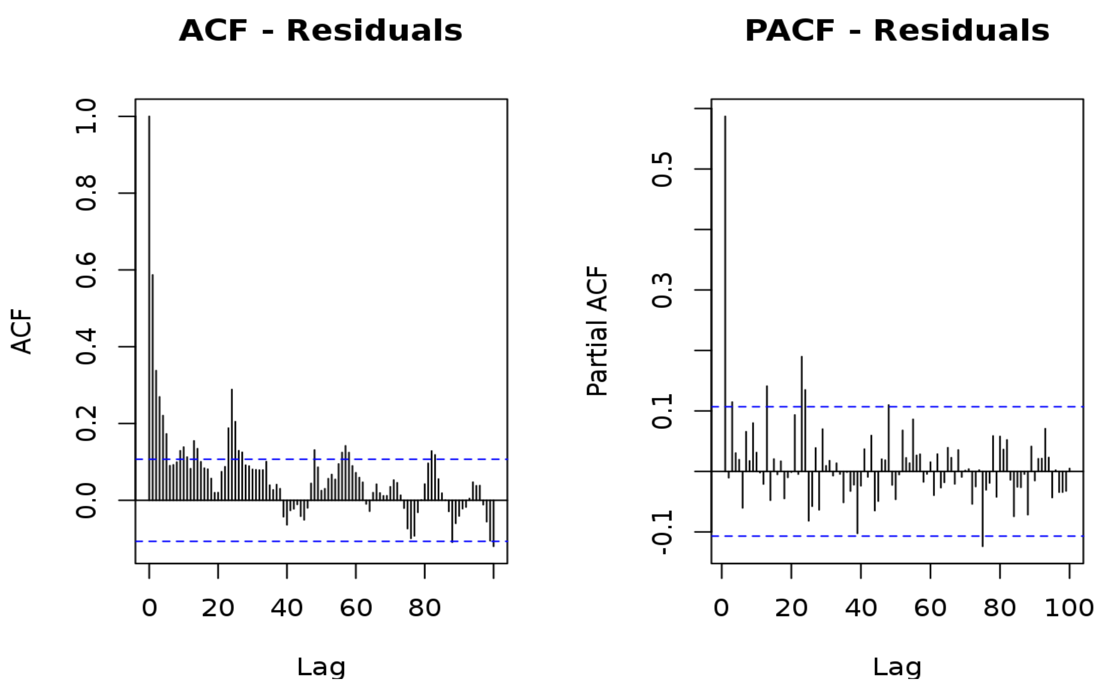

Figure 67 ACF and PACF of first 14 days of Jan 2012

MODEL FORECAST

Jan 2012 forecast

The best model chosen is a Neural Network with 4 hidden nodes, 0.01 decay rate at 400 iterations. The figure 68 shows the forecast for Jan 15- Jan 19, 2012, a total of 120 records. This forecast period has a weekday – Monday Jan 16th 2012, a Martin Luther King holiday. This model provides a MAPE of 41.4 %.

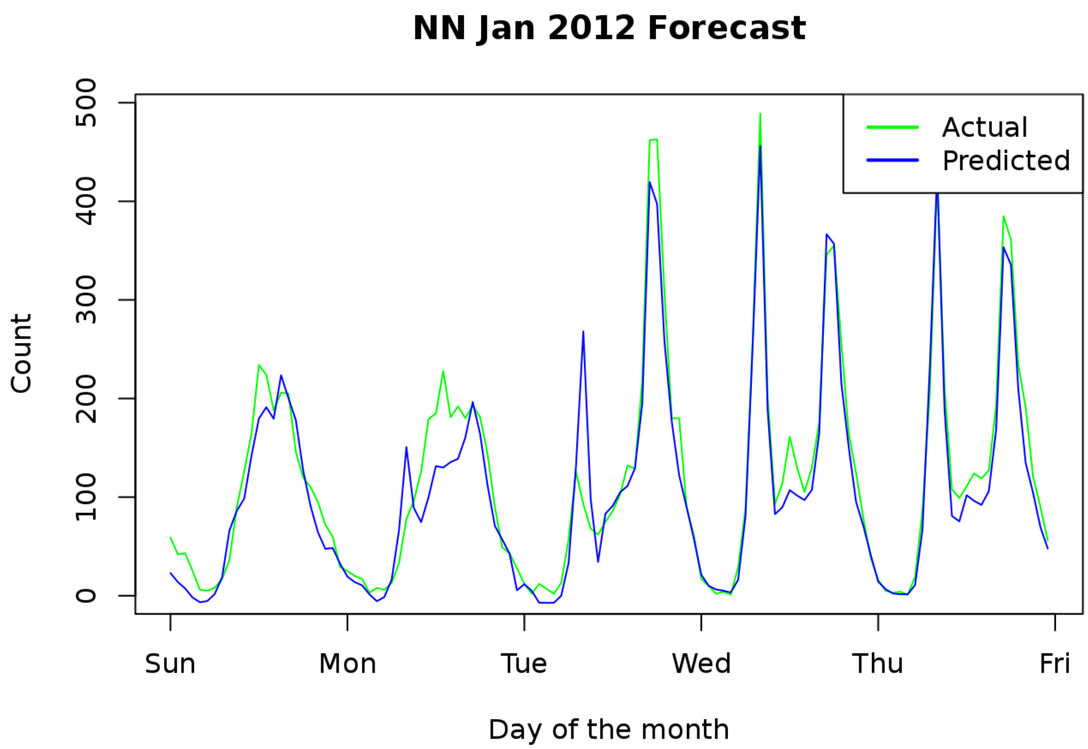

Figure 68 Jan 2012 Forecast

This model captures the Martin Luther King holiday better, i.e. a non-working day pattern. Though it inherently has a damped double peak, the forecast for Jan 16 is only scaled to a non-working day pattern. The scale of forecast varies significantly for working day and non-working day leading to better forecast than other models.

Forecast Residuals

Figure 69 shows the residuals for the forecasted values. It can be seen that the residuals are mostly centered around the mean and normally distributed, except for one forecast at Jan 17th, Tuesday, 9:00 AM. The model overestimated the count and thus causing an outlier. The left tail on the histogram is long due to this outlier. Figure 70 shows the ACF and PACF of forecasted values are mostly within the intervals and does not have discernible correlation pattern.

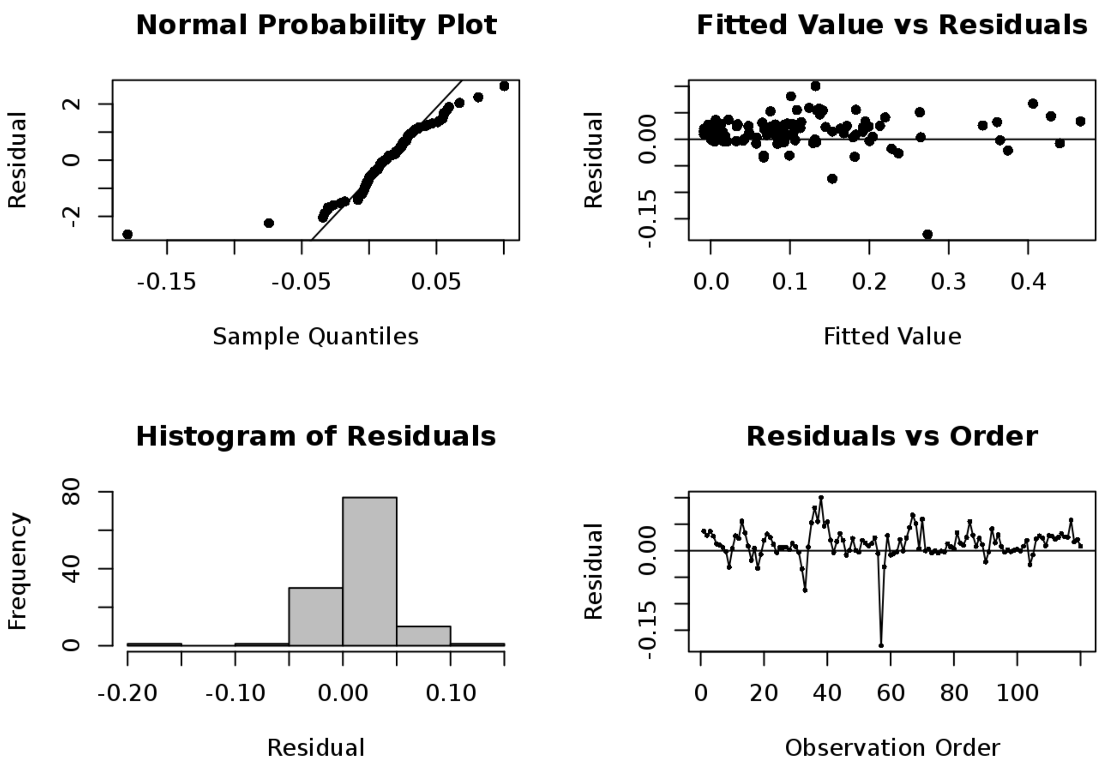

Figure 69 Residuals of forecast

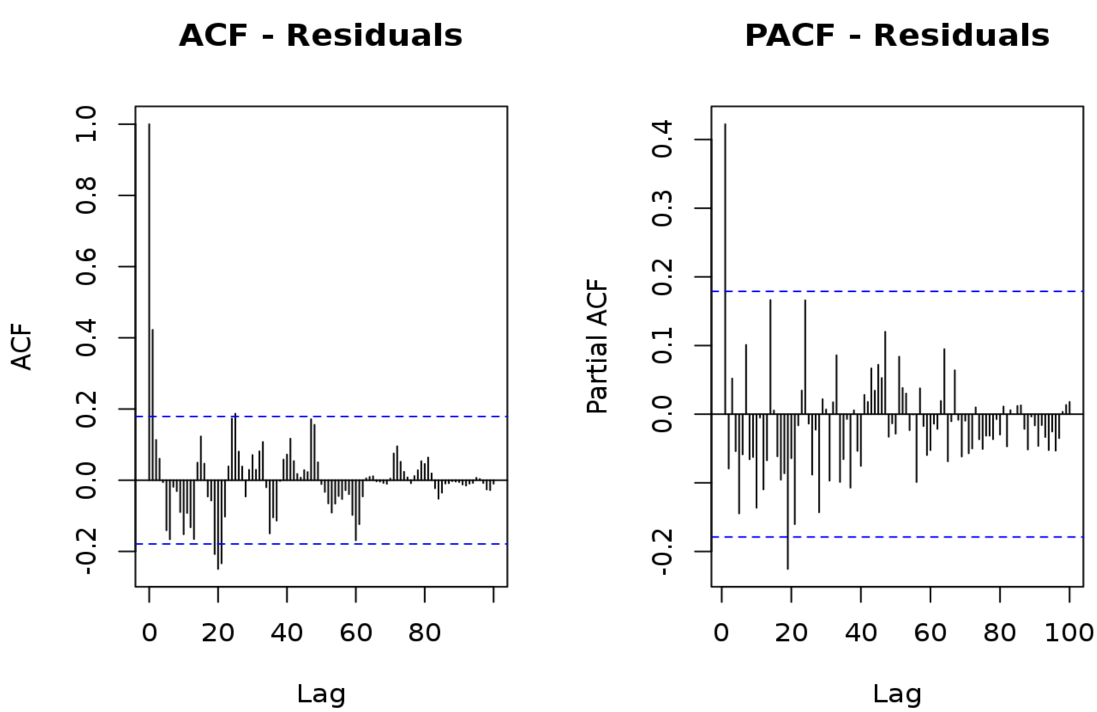

Figure 70 ACF and PACF of forecast residuals

Forecast Performance

So far, the traditional models were not able to pick up working day and non-working day patterns. Neural Networks outperform other models in picking up the scale of count and the double peaks for working days and single peaks of non-working days.

Improving neural network

The residuals analysis of the best neural network model showed characteristics of heteroscedasticity. Thus neural networks can be improved by log transformation of the count variable, leading to similar analysis as before.

MSE and MAPE vs hidden nodes

Using all regressors and log count as the response variable, several neural networks are evaluated using number of hidden nodes ranging from 1 to 20. Decay rate used are 0.01, 0.05 and 0.1. Figures below show the MSE and MAPE vs hidden nodes and decay rate. The MSE plot stabilizes after 3 nodes. MAPE drops low after 2 hidden nodes and does not vary often with the increase in hidden nodes. The candidate models are those which have 4, 6, 10, 14 hidden nodes at 0.1 decay rate. The log transformed count is transformed back to its original value to calculate MAPE.

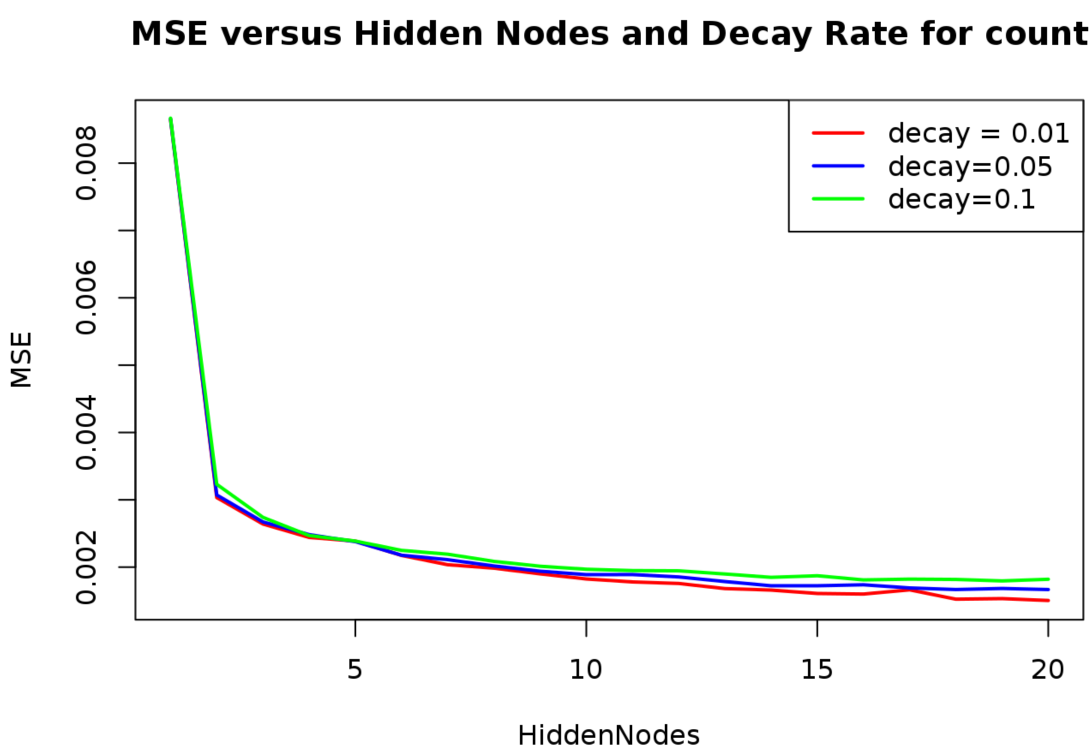

Figure 71 MSE vs Hidden nodes and Decay Rate

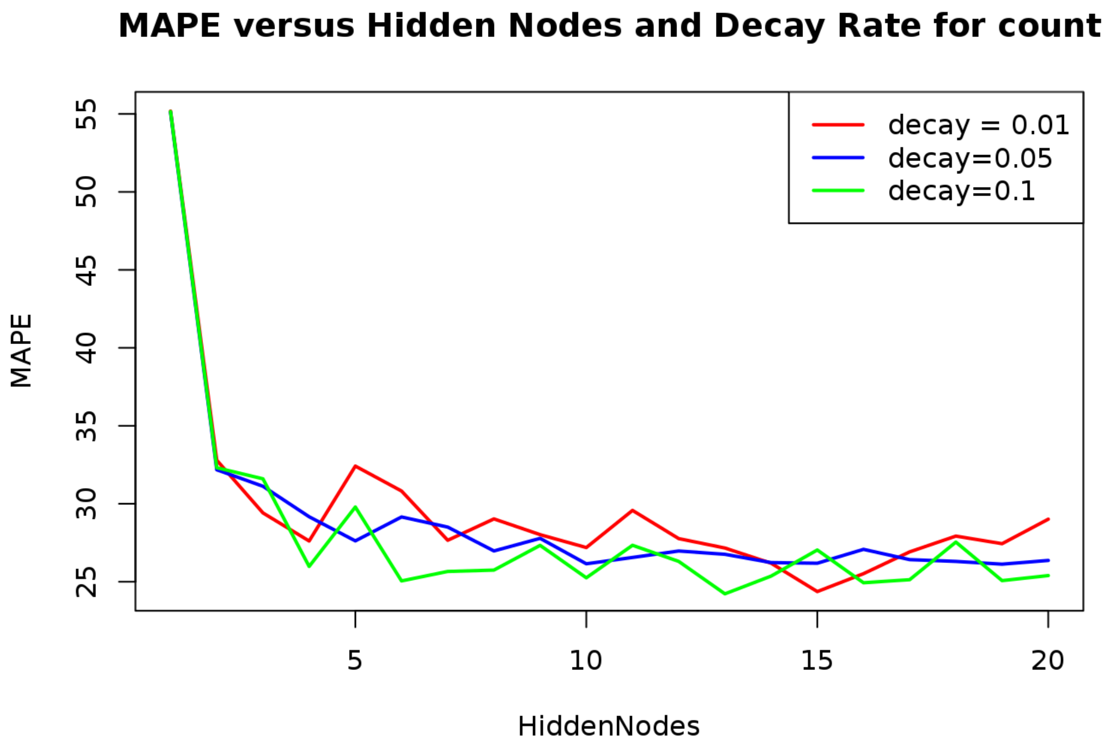

Figure 72 MAPE vs Hidden nodes and Decay Rate

MSE and MAPE vs iterations

Using the candidate hidden nodes (4,6,10,14) and decay rate of 0.1, several iterations ranging from 100 to 1000 were tested against MSE and MAPE. Figures below show the MSE and MAPE against different number of iterations for hidden nodes 4, 6, 10, 14 at a decay rate of 0.1.

The MSE stabilizes after 200 iterations for all hidden nodes. MAPE stabilizes after 500 iterations. The best model has a MAPE of 24 % with 10 hidden nodes at 0.1 decay rate at 600 iterations.

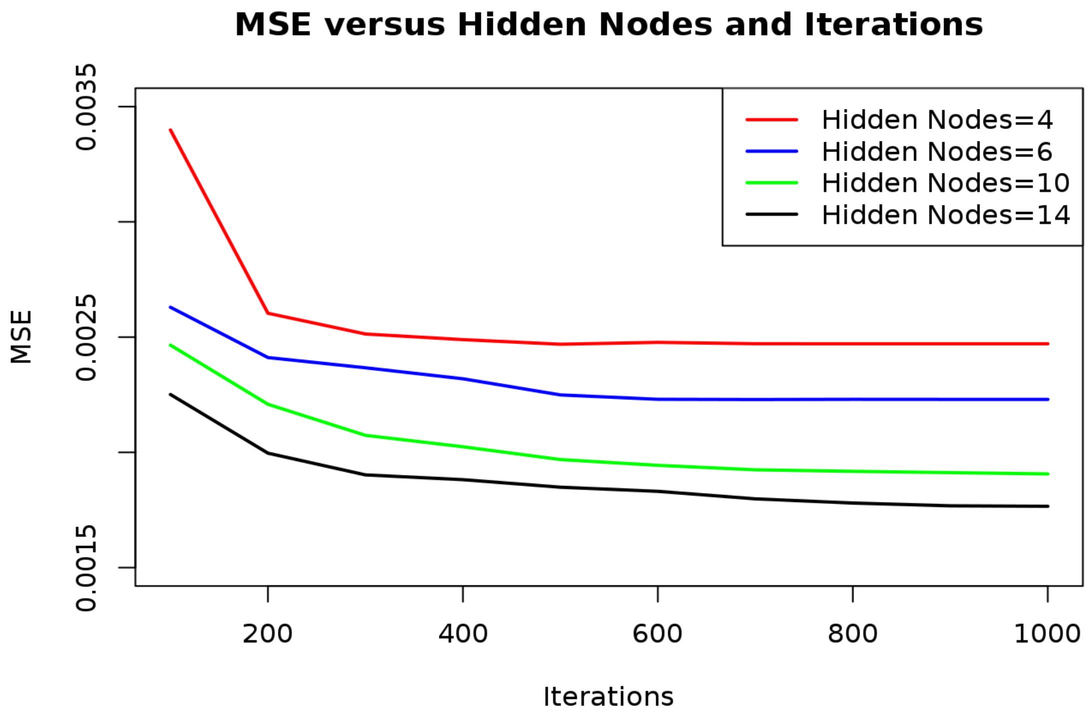

Figure 73 MSE vs Hidden nodes and Iterations

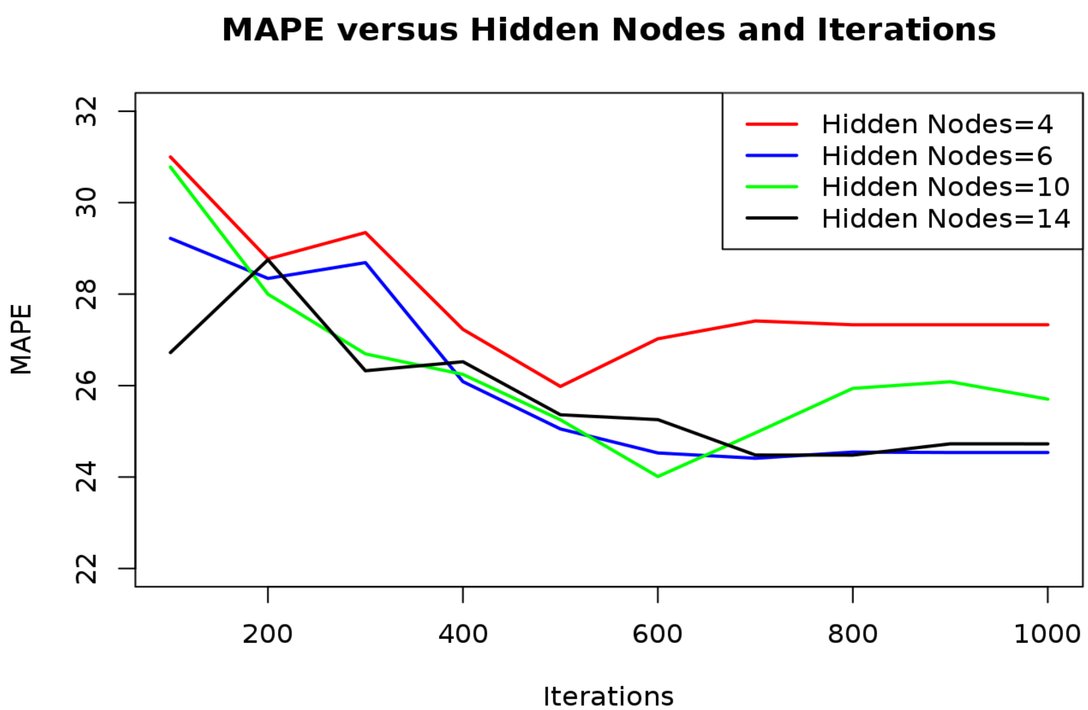

Figure 74 MAPE vs Hidden nodes and Iterations

Improved model

Including all of the regressors in the neural network (h = 10, d = 0.1, maxit = 600) gives a MAPE of 24%. The figure below shows the model fit on log transformed Jan 2012 data. The fitted points follow close to the actual data.

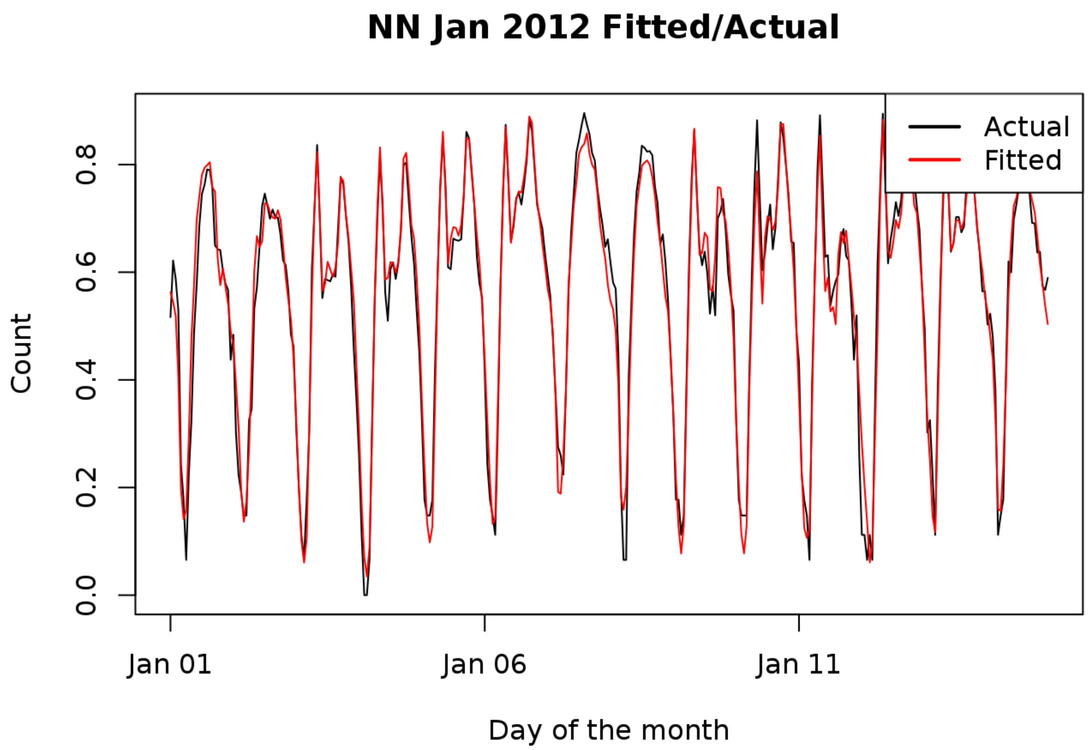

Figure 75 Model fit of Log count

Training Residuals

Figure 76 below shows the residuals of 14 days of Jan 2012. The residuals look normally distributed except for few outliers. Histogram and Q-Q plot of the residuals show normal distribution. The mean of residuals is close to zero. Fitted vs Residuals show relatively constant variance except for higher values. Figure 77 show ACF and PACF of the residuals. The ACF shows more positive correlation but the values are within the intervals. The PACF also shows that the values are within the intervals.

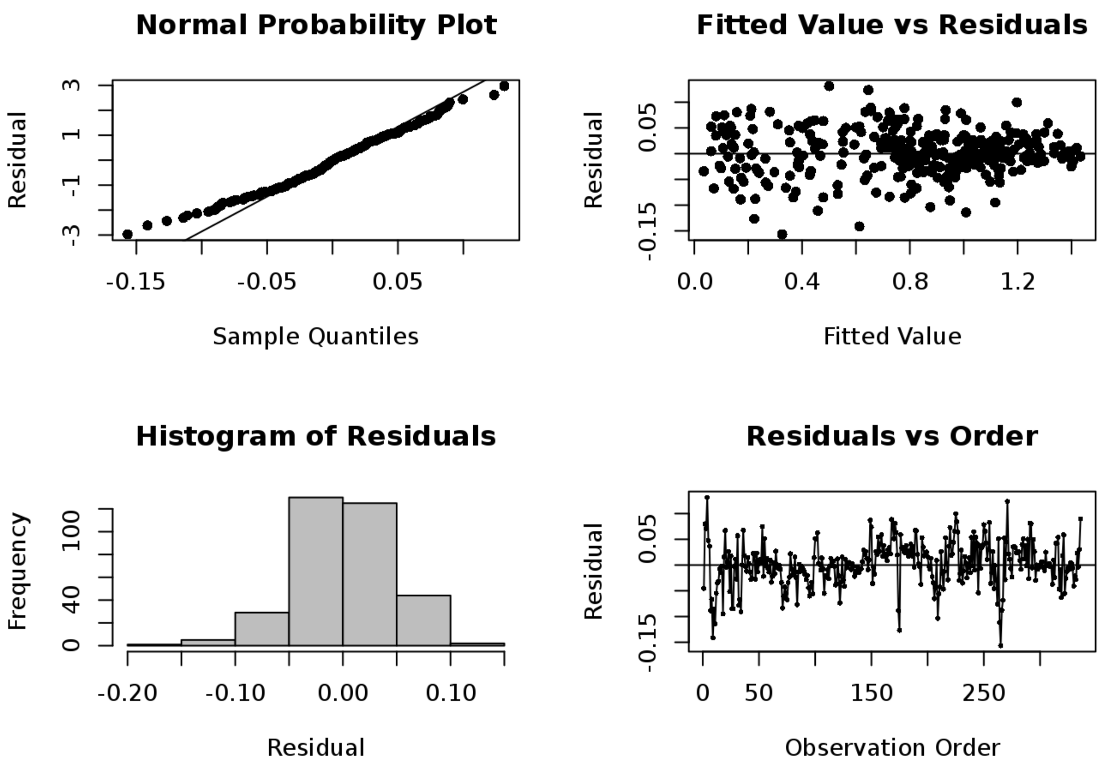

Figure 76 Training Residuals of Log Count

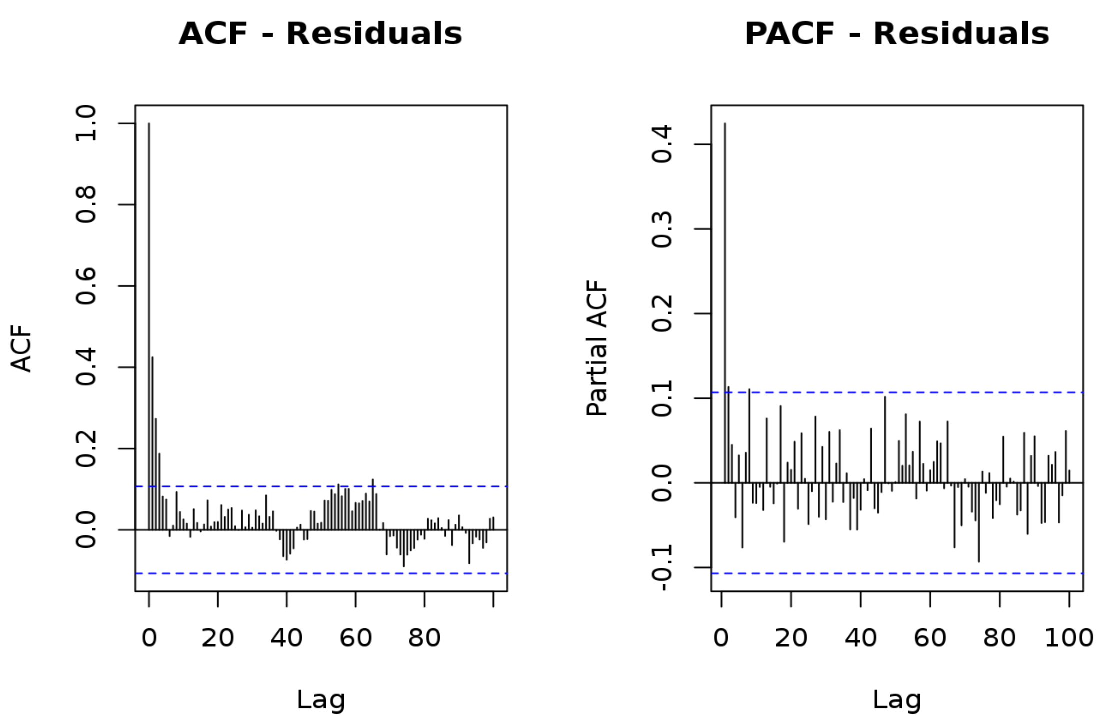

Figure 77 ACF and PACF of Residuals

MODEL FORECAST

Jan 2012 forecast

The best model chosen is a Neural Network with 10 hidden nodes, 0.1 decay rate at 600 iterations. The figure 78 shows the forecast for Jan 15- Jan 19, 2012, a total of 120 records. This forecast is better than the previous model i.e. it does not overestimate the count. The estimated count at Tuesday, Jan 17th 9:00 AM is closer to the actual count, thus improving the model.

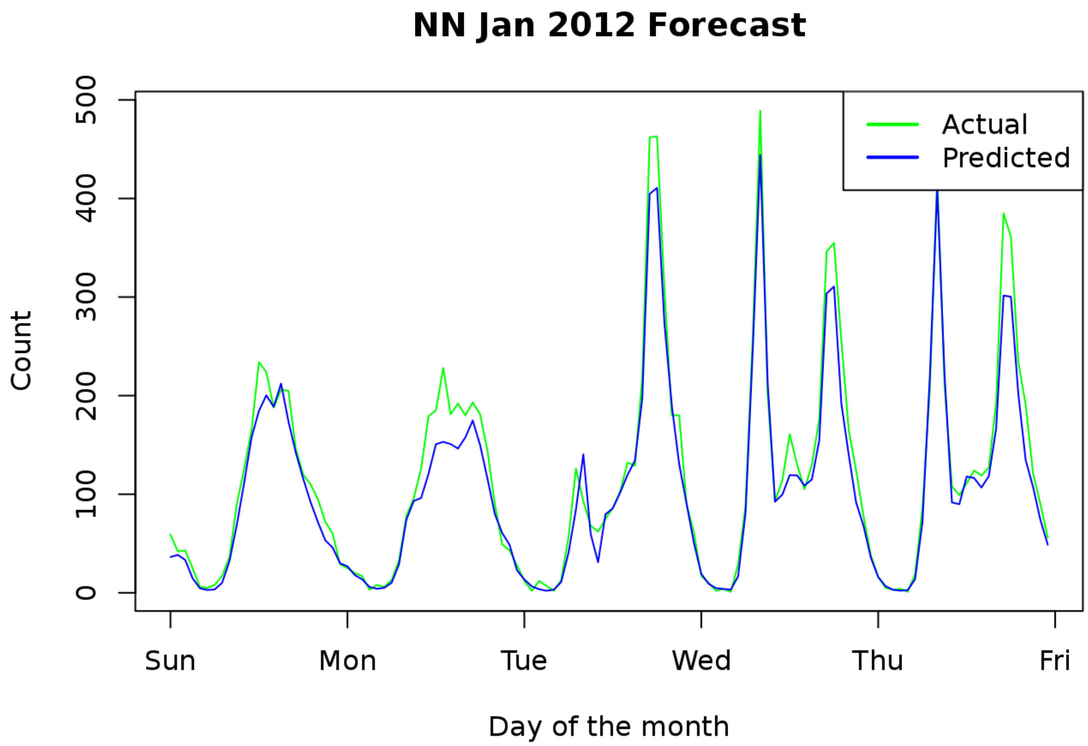

Figure 78 Jan 2012 Forecast after log transformation

Forecast Residuals

Figure 79 shows the residuals for the forecasted values. It can be seen that the residuals are mostly centered around the mean. The Q-Q plot show outliers. The model still shows some high variance at the lower values, causing outliers. Figure 80 shows the ACF and PACF of forecasted values are mostly within the intervals and does not have discernible correlation pattern.

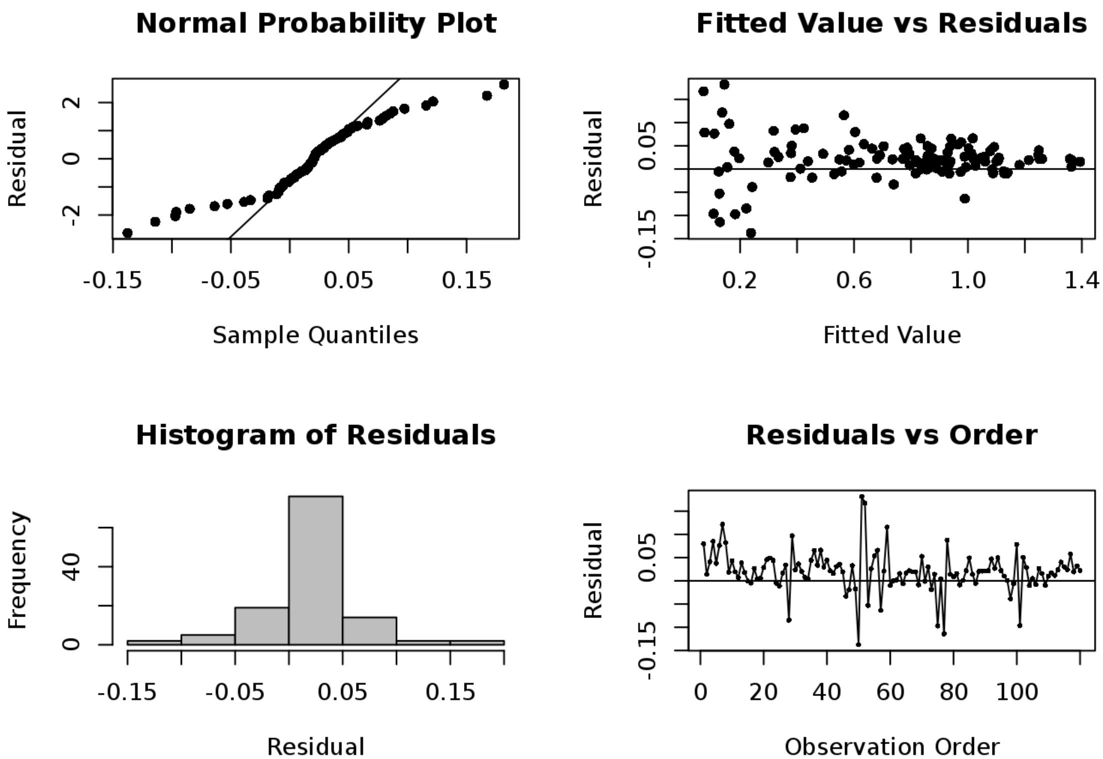

Figure 79 Forecast Residuals

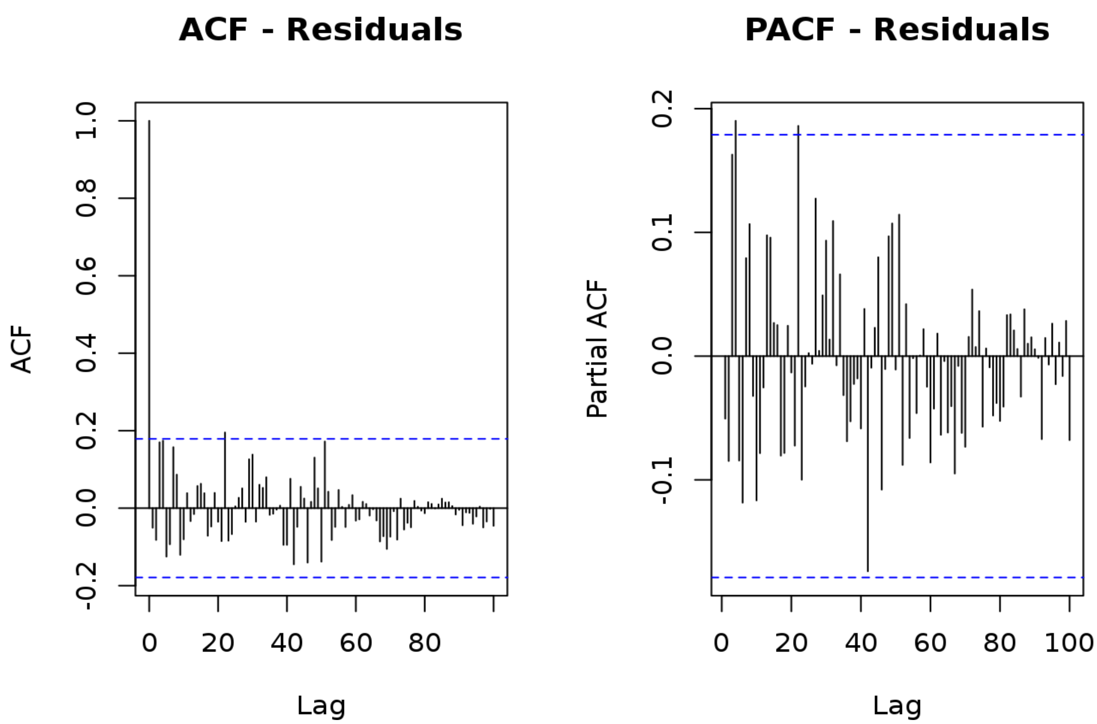

Figure 80 ACF and PACF of Forecast residuals

Model Summary

The neural network model performs better than any other model evaluated. Below is the summary of all the best models from other types of forecasting.

Model	Parameters	MAPE	Comments
Holts Winter	alpha = 0.3, beta = 0, gamma = 0.2	155.7948	Residuals are correlated and non-independent
ARIMA	ARIMA (2,1,3) x (0,1,1) 24	81.4538	Residuals are less correlated and non-independent.
Dynamic Regression - Regressors	ARIMA (3,1,4) x (0,1,1) 24 with Day of the week, Working Day, and Feels like temperature	119.37	Residuals are correlated and non-independent
Dynamic Regression - Lagged Predictors	ARIMA (2,1,3) x (0,1,1) 24 with 1 lag of Feels like Temperature	211.53	Residuals are correlated and non-independent
Neural Networks	hidden nodes = 4, decay rate = 0.01, iterations = 400	41.1	Residuals are less correlated and non-independent.
Neural Networks - Log Transformed	hidden nodes = 10, decay rate = 0.1, iterations = 600	24	Residuals are less correlated and non-independent. This is the best model so far with better forecasts.

Table 10 Model Summary

Conclusion

By evaluating several types of models over subsets of data and the whole data, model built on neural networks perform well. It is due to the number of samples, since all of the available data is used in a neural network model. The traditional models such as Holts winter, ARIMA, ARIMA with regressors can perform better if there was more continuous data.