Project Title	:	Bike Share Prediction Analysis
Date	:	14/12/2018
Group	:	13
Members	:	LIN, Hongzhou (Hong)
		AJWAD, Habeeb Ahamed
		LIN, Yiqi (Regina)
		LIU, Xiaoning (Eric)
		CHEN, Yihua (Alisa)

Executive Summary:

We have totally built 3 models to predict the amount of bike users in the future and all 3 models use time-related predictors such as instant, year or season to predict the amount of users. Our predictions are all based on one core assumption that the amount of users in the future will follow similar time trend and seasonal pattern as followed in 2011 - 2012. Model 1 and Model 2 are considered valid models for the prediction use since their model diagnostics satisfy all OLS assumptions. Nevertheless, Model 1 is not the optimal option to predict the demand on bikes due to its relatively low accuracy in contrast with Model 2's and Model 3's Out-Sample RMSEs and MADs. Although Model 3's combined prediction provides the best prediction performance in the test data set, the Model developed to predict the amount of casual users does not satisfy the Constant Variance Assumption even though the model developed to predict the amount of registered users satisfies all OLS assumptions. Nevertheless, we consider the prediction interval generated from this model to be unreliable. As a result, we could utilize Model 2 to predict the total number of users as this model provides the best accuracy under the condition that all OLS assumptions are satisfied, meaning that both estimated values and prediction intervals can be trusted. Additionally, we can use Model 3 to analyze and predict the bike demands from casual and registered users. Based on the different behaviors of these 2 types of users, different operation strategies can be implemented for optimization.

Introduction:

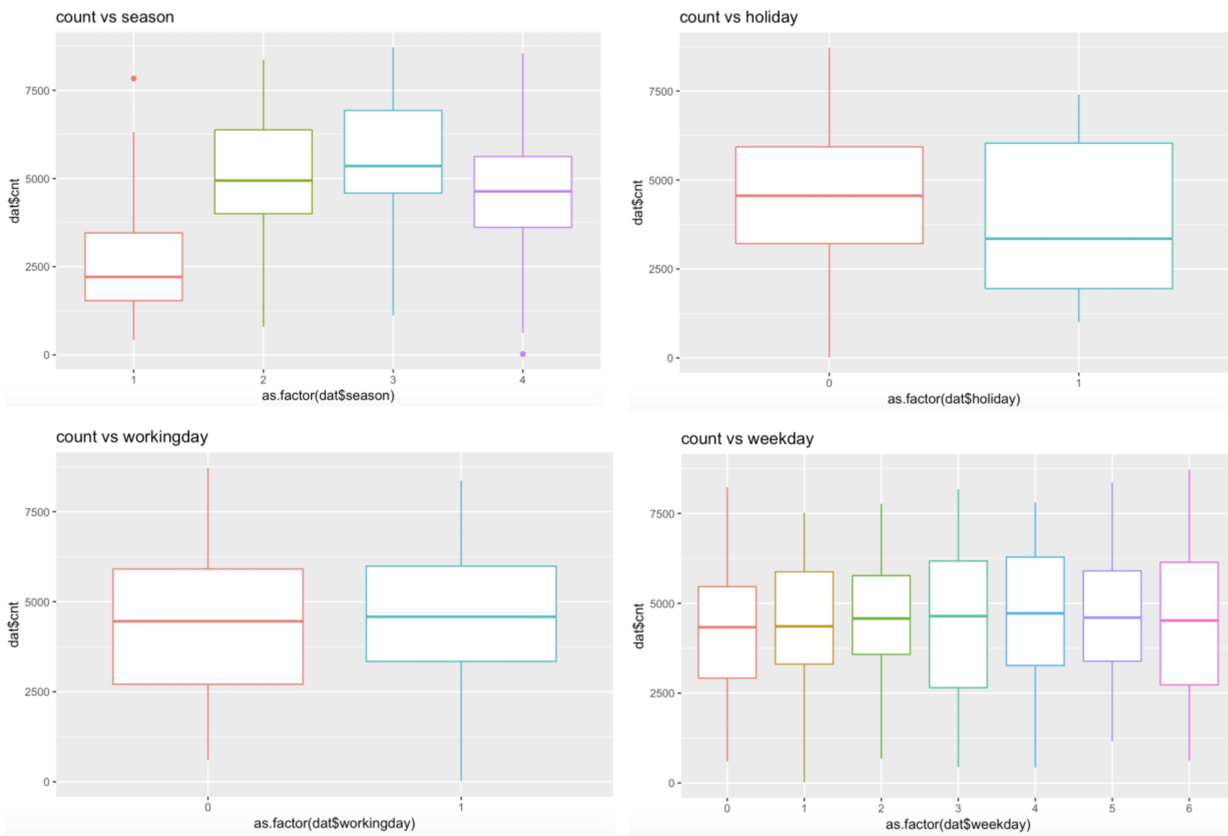
The objective of this project is to build a predictive model to estimate the amount of bike users on a given day based on a variety of conditions. Further, the purpose of this report is to elaborate on the different models that were developed to predict the usage of the bikes. Our dataset contains the daily amounts of bike users including both casual and registered users from 2011 – 2012. This dataset also records the daily conditions of the region in regard to temperature, humidity, windspeed, weather, season, etc. Furthermore, this report consists of two sections namely Analysis & Results and Conclusion. Where the Analysis & Results section focuses on the models explanation, methodologies, detailed studies on metrics of different predictors and regression outcomes. The Conclusion section summarizes how different models work, provides business insights, and gives a verdict on the model with the best prediction performance.

Analysis & Results:

This section elaborates on the different models generated along with comparing and interpreting the outputs of each model.

The initial process is to explore the dataset and study possible correlations or patterns among dependent variables and predictors involved; a correlation matrix is used to examine the distributions of the correlation among them. Based on the correlation plot (Refer to the correlation plot in the appendix), we observe that most numerical predictors have significant correlations with the dependent variable. Distributions of the dependent variable versus different numerical predictors are roughly bell-shaped. Subsequently we realize that "temp" and "atemp" are almost perfectly correlated. Since both of these two predictors are expressing similar information, we decided to include only one of them namely "atemp", which refers to the feeling temperature compared to the real temperature expressed by "temp". By doing this, the multi-collinearity problem is avoided.

Subsequently, we conducted side by side box plots which describe the relationship between the dependent variable and categorical predictors respectively. All of these categorical predictors (such as "season", "working day", and "holiday") are correlated with the dependent variable due to the clear patterns shown on box plots. Thus, the demand on sharing bikes are affected by these categorical predictors as you can see in the plots below.

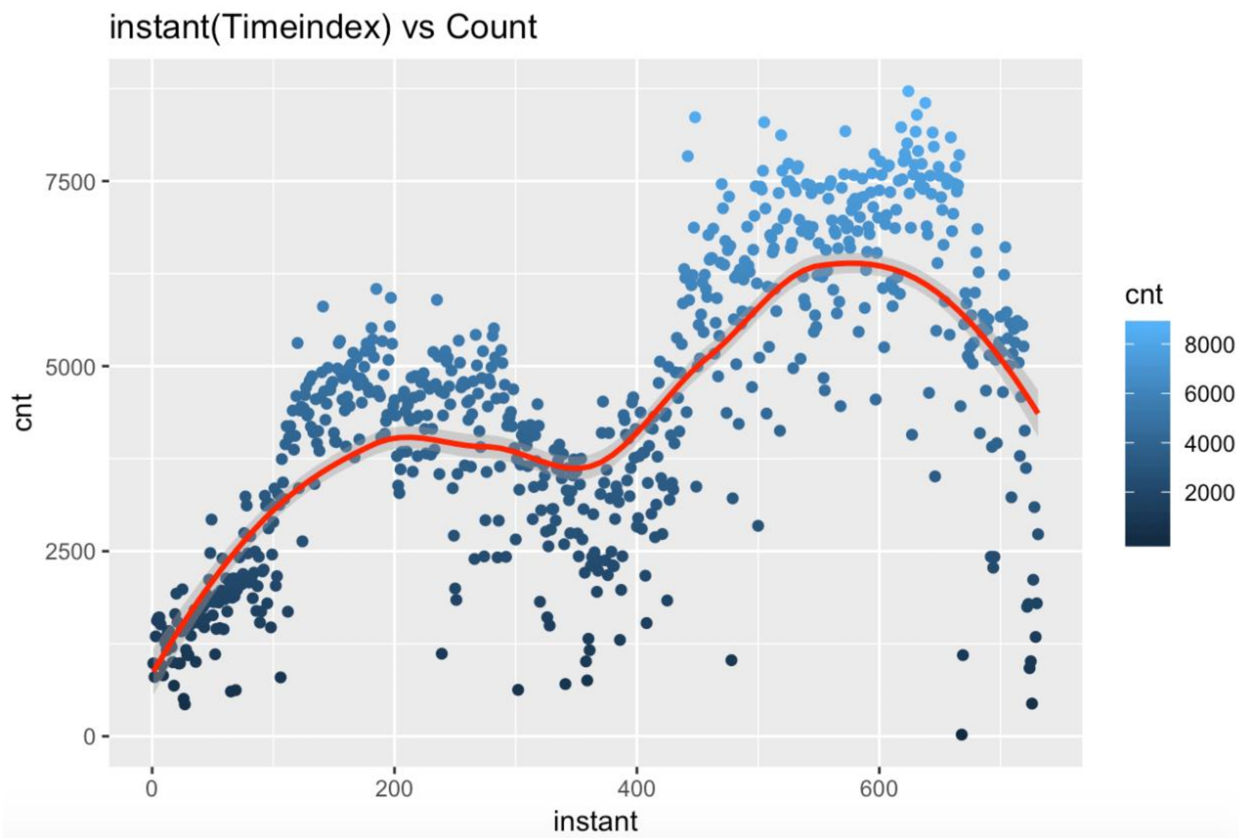


After exploring the dataset, we conclude that count (amount of bike users) is correlated with most of our predictors such as temperature, humidity, windspeed, weather, season, etc.

Since the data under the response variable ("cnt") are much larger than 100, we decided to build a multi-linear regression model for prediction instead of a Poisson model which specifies on the target with smaller count number (less than 100).

Model 1 – Additive Model to Predict Total User Count

Before building the models, we decided to split the dataset into training set and test set by the ratio of 8:2, in order to use the training set to train our models and save the test set for validation purposes. From the scatter plot below, there is an increasing trend and a seasonal patterns. Each season the demand increases or decreases by a fixed amount. Because the dependent variable is normally distributed, and this dataset has time series problem. we decided to choose the Additive Model to predict the demand.



Firstly, we built a full model (model 1) by including all the predictors that we found useful (Check appendix for model 1 summary output). Since the seasonal trend is observed, we decide to put time index ("instant") as a predictor variable in our model. Also, if we plug "workingday" into the regression modeling, its coefficient would be "NA", which may be caused by the multicollinearity problem, so this variable is omitted. Based on this model's summary output, the P-Value for the overall F-test is less than 0.05, which means that at least one of the predictors is useful in predicting our dependent variable. While some predictors used in this model are endowed with similar information associated to the change of target variable, multicollinearity problem becomes a concern. We conducted a VIF Test to examine the VIFs for each predictor before we checked their individual P-values. Based on the test output above, we realized that the GVIFs for "instant", "season", "year" and "month" exceed the threshold value of 10. Therefore, "month" and "year" are omitted in our corresponding model (model 1.1) in order to fix the multi-collinearity problem (Check model summary output for model 1.1). Next, the new GVIFs examined show that the multicollinearity problem in this model is finally fixed.

Next, we analyzed the residuals of model 1.1 to check whether this model satisfies the OLS assumptions. The process of analysis includes a normal QQ plot, a residual vs fitted value plot, a histogram of residuals and a residual vs time index plot (Check Appendix for The Model Diagnostics).

Based on the normal QQ plot and the histogram, the distribution of the residuals is roughly normal. However, we also realize that the variance of the residuals doesn't seem to be constant. Therefore, we conducted a Breusch–Pagan test to verify our observations. According to the test result above, the constant variance assumption is not satisfied. In order to identify the predictor that we need to transform for fixing this problem, we had made plots for the residuals against each predictor and later found out that "atemp" might be the cause the issue (Check appendix for all residual plots). After the quadratic transformation, violation of constant variance was fixed by adding a higher order term into the model (Check appendix for the summary output of model 1.2).

Next, we also conducted a Cochrane-Orcutt test to verify whether there is autocorrelation in the residuals (Check appendix for the CO test output). The test result indicates autocorrelation problem existing in model's residuals. Hence, we performed a Cochrane-Orcutt procedure to verify the P-values of the residuals (Check appendix for the Cochrane-Orcutt model's summary output). According to the summary output, the P-values for "season" and "holiday" are not significant at the 95% confidence level. After eliminating these two predictors in our model and applying an ANOVA test to compare our original model 1.2 with this newest model (without "season" & "holiday"). (Check appendix for the ANOVA output). The ANOVA test result shows that the original model with these two predictors are significantly useful in prediction.

model = lm(cnt instant + season + holiday + weekday + weathersit + atemp + sq_atemp + hum + windspeed , data =train)

The model shown above is used for prediction due to the ANOVA test result. Its RMSE is 1299 and MAD is 963, which falls in acceptable levels when the mean of "total user count" is around 4500.

Model 2 – MLR model to Predict Total User Count (Without Using “Instant”)

To improve the prediction performance, we build a MLR model by using different predictor sets. On the timeseries plot, the overall upward trend is represented by “yr” and the seasonal pattern is caught by “season”. The VIF test output from the full model indicates “instant”, “mnth” and “yr” have multicollinearity problem. This is because all three predictors represent the time trend. While Model 1 shown on last section keeps “instant” and obsolete “yr” in modeling, Model 2 in this section would use “yr” to replace “instant” to build a a MLR model. (Check Appendix for Model 2.1 summary output).

Subsequently, we also conducted a VIF test to verify that the multicollinearity problem in the original full model is fixed (Check appendix for the VIF test output for model 2.1). As the result indicates the problem is fixed by eliminating predictors “instant” and “mnth” from the full model. Additionally, we observed that the adjusted R- Square for model 2.1 is high than that of model 1.1. Following our established procedure, we conduct the model diagnostics for model 2.1 (Check Appendix For The Model Diagnostics of Model 2.1).

model = lm(cnt ~ yr+season+holiday+weekday+weathersit+atemp+hum+windspeed, data =train)

Based on the model diagnostics output, the distribution of the residuals for model 2.1 is also roughly normal. Nevertheless, the residual vs fitted value plot indicates that the constant variance assumption might also be violated in this model. Hence, we conduct a Breusch–Pagan test to verify our observation.

As the test result indicates, the constant variation assumption is violated. By following the procedures, we had conducted residual plots against each predictor, and then found that patterns are be similar to those in model 1.1. Therefore, we added a higher order term for “atemp” in model 2.2 to fix the problem (Check appendix for model 2.2 summary output). In regard to the BP test output for model 2.2, the violation of constant variance assumption is fixed, as expected.

Eventually, a Cochrane-Orcutt test was used to check the autocorrelation in the residuals of model 2.2 (Check appendix for Cochrane-Orcutt output for model 2.2). Based on the test results, autocorrelation exists in residuals. Thus, we need to conduct the Cochrane-Orcutt procedure again to verify the P-value for each predictor in model 2.2. (Check the appendix for Cochrane-Orcutt model summary output for model 2.2). The Cochrane-Orcutt model summary output indicates that the P-value for "holiday" is not significant at the 95% significance level. ANOVA test is conducted to compare the model with "holiday" and the one without it. The result indicated that the smaller model is statistically better. RMSE and MAD values of the two models are close, so the model, which was trained with "season", "yr", "weekday", "weathersit", "atemp", "sq_atemp", "hum", "windspeed", is chosen to predict the total demand on bikes.

model = lm(cnt~ season+yr+weekday+weathersit+atemp+sq_atemp+hum+windspeed, data =train)

Thanks for the efforts made to improve the prediction performance, accuracy of model 2 on the validation set improves significantly in terms of the latest RMSE and MAD scores (RMSE:1096; MAD: 826). The visual representation of the prediction has been attached for your reference on next page.

Model 3 – MLR Models to Predict Casual & Registered Users Respectively

In further exploration of our dataset, we observed that the number of total users are composed by registered and casual users. Relationships between some predictors and the number of Casual Users are significantly different from the relationships between them and the number of Registered users. For instance, the customer base on registered users are growing significantly faster than the casual users (Check the appendix for the time series plots of casual and registered users). Moreover, the number of casual users are significantly higher in holidays whereas the number of registered users decreases on such days (Check appendix for the correlation charts and side and side box plots for both casual and registered users). Correspondingly, we built the 3rd model(s) to predict the number of casual and registered users respectively in order to escalate the model's prediction performance.

Model 3 (Registered Users)

Based on the correlation chart of the registered users and other numerical predictors, the distribution of registered users is roughly normal and the patterns of registered users and other numerical predictors are roughly linear. From the side by side box plots, changes in categorical predictors are associated with the number change on registered users (Check appendix for the side by side box plots and the correlation matrix). Subsequently, we built a full model with all predictors except for "temp" due to the multicollinearity problem (Check appendix for model 3 registered summary output).

After conducting the VIF test, multicollinearity problems is found. To maximize the prediction performance, Stepwise Selection Method (both ways) is chosen to select the optimal combination of predictors with the lowest AIC (Check appendix for model 3.1 registered summary output). Based on the step selection model 3.1, the multicollinearity problems are fixed. Next, we conducted model diagnostics for model 3.1 (registered).

As the diagnostics output indicates, the distribution of the residuals is roughly normal. According to the Breusch–Pagan test output, the constant variance assumption is again violated. Therefore, we added a higher order term of "atemp" and conducted Model 3.2 (Registered) to fix the problem. This procedure has been effective as we have already seen in the previous models. Upon verification by the BP test for Model 3.2, we conclude that Model 3.2 is our final model for registered user prediction
In terms of the prediction performance, the Out-Sample RMSE and MAD are 927 and 684. The visualization of our model prediction performance is attached below:

Model 3 (Casual Users)

Based on the correlation matrix between casual users and other numerical predictors, the distribution of casual users is strongly right skewed and the patterns between number of casual users and some numeric predictors such as "atemp" is not linear. Hence, data transformation is needed to make the distribution normal. Both log transformation and square root transformation were taken; based on the correlation charts, the outcome of Log Transformation is better because the curved patterns turns to be more linear. Side by side boxplots also show how categorical predictors affect the number of casual users (check Appendix for correlation charts of log transformation and square root transformation, and side by side boxplots between Log_Casual and categorical predictors).
Subsequently, "Log_Casual" is chosen to be the response variable for modeling. Appendix for Model 3 (Casual) Summary Output shows the result of the VIF test conducted. While the GVIFs of few predictors exceed the threshold of 10. Step selection method (both ways) is used to form the optimal combination of predictors with lowest AIC (Check Model 3.1 Casual summary output). The multicollinearity problem does not exist here, and the selected combination of predictors also complies with common and business sense. The adjusted R Square also indicates very good performance of

Model 3.1. Hence, we perform model diagnostics on Model 3.1. According to the diagnostics output of Model 3.1, the distribution of residuals is roughly normal. However, the constant variation assumption is violated according to the Residual vs Fitted Value Plot and the Breusch Pagan Test.
Unfortunately, the constant variation violation cannot be fixed after the data transformation on numerical predictors. We did not manipulate the categorical predictors because the side by side boxplots between residuals and the categorical predictors do not show any obvious pattern. Thus, Model 3.1 (Casual) is used as the optimal model to predict the number of casual users. Due to the constant variation violation, we can't trust the prediction interval calculated from our final model.
In terms of prediction performance, our final model yields out-sample RMSE and MAD of 317 and 217. The visualization of prediction performance is attached below:

Combined for Prediction of Total Number of Users

To produce the final predictions of total number of users for model 3, we sum up the predictions of casual users and registered users. The out-sample RMSE and MAD are 1007 and 757. This prediction performance is the best out of the 3 models that we have built above. The visualization of the prediction performance is attached below:

Conclusion

Due to the fact that all 3 models use time-related predictors (instant / year / season) to predict the amount of users. Our predictions are all based on one core assumption that the amount of users in the future will follow similar time trend and seasonal pattern as 2011 - 2012. Note that only when this assumption is valid then our models and predictions will be reliable.

Model 1 and Model 2 are considered valid models for the prediction use since their model diagnostics satisfy all OLS assumptions. Nevertheless, Model 1 is not the optimal option to predict the demand on bikes due to its relatively low accuracy in contrast with Model 2's and Model 3's Out-Sample RMSEs and MADs. In terms of the Out-Sample accuracy, Model 2 is a better choice for predicting demand when the target variable is cnt (total user counts). Although Model 3's combined prediction provides the best prediction performance in the test data set, the Model developed to predict the amount of casual users does not satisfy the Constant Variance Assumption. Thus, the prediction interval generated from this model is not trustful. On the other hand, the model developed to predict the amount of registered users satisfies all OLS assumptions.

With that said, Model 3 is still very useful in a business operation standpoint because it is able to separately analyze the relationships between casual users and predictors and the relationships between registered users and predictors. The Scatter Plot Matrices and Side-by-Side Box Plots (Check Appendix) show different patterns between casual and registered user demands under different conditions (refer to appendix). This phenomenon complies with common sense. For example, casual users' demand on bike hikes up on holidays while registered users' demand would go down. In working days, regular users' demand are higher than their demand on weekend; however, demand of casual users shows an opposite pattern in this case. Thus, building two independent models to predict demand of casual and registered users respectively is a sophisticated approach based on the observations mentioned above.

From a business operation standpoint, the behaviors of casual bike users and registered bike users are likely to be very different. For example, the bike usage time of registered users will be much more consistent and predictable compared with the bike usage time of casual users. Thus, the turn-over rate of bikes can differ greatly with the same amount of total bike users, depending on ratio between casual and registered users. Moreover, the times when bikes are needed by registered users tend to cluster since many registered users use bikes to commute between home to work. On the other hand, the times when bikes are used by casual users are more scattered. Therefore, it would be very important for the company to study the behaviors of casual and registered users and predict the number of casual and registered users separately in order to optimize the business operation and resource allocation.

In conclusion, we could utilize Model 2 to predict the total number of users because this model provides the best accuracy under the condition that all OLS assumptions are satisfied, meaning that both estimated values and prediction intervals can be trusted. Additionally, we can use Model 3 to analyze and predict the bike demands from casual and registered users. Based on the different behaviors of these 2 types of users, different operation strategies can be implemented for optimization.

Appendix

Model 1

Summary Output

```
Call:
lm(formula = cnt ~ instant + season + yr + mnth + holiday + weekday +
    weathersit + atemp + hum + windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2952.40  -352.40    60.79   424.42  2911.84

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1476.5775   234.7495   6.290 6.43e-10 ***
instant        -0.7819     3.3664  -0.232  0.81642
season2       704.5961   172.0553   4.095 4.85e-05 ***
season3       564.8749   223.7072   2.525  0.01184 *
season4      1049.4399   219.2245   4.787 2.17e-06 ***
yr1          2331.2847  1230.0039   1.895  0.05856 .
mnth2         209.2672   162.9109   1.285  0.19948
mnth3         775.9748   240.7787   3.223  0.00134 **
mnth4         868.2082   358.1356   2.424  0.01566 *
mnth5        1285.7127   446.1378   2.882  0.00411 **
mnth6        1257.0639   536.2667   2.344  0.01942 *
mnth7         899.1208   634.0192   1.418  0.15671
mnth8        1144.5075   720.1486   1.589  0.11257
mnth9        1566.8886   819.0146   1.913  0.05624 .
mnth10       1296.2071   938.1915   1.382  0.16765
mnth11        929.6281  1037.9188   0.896  0.37082
mnth12        896.1370  1131.4113   0.792  0.42867
holiday1     -358.8070   187.1203  -1.918  0.05568 .
weekday1      105.9400   110.9639   0.955  0.34013
weekday2      246.2097   108.2730   2.274  0.02335 *
weekday3      203.3237   108.8821   1.867  0.06238 .
weekday4      261.8080   108.7788   2.407  0.01642 *
weekday5      278.9916   108.8031   2.564  0.01060 *
weekday6      290.5028   107.9156   2.692  0.00732 **
weathersit2  -384.9532    78.9407  -4.876 1.41e-06 ***
weathersit3 -1709.2328   193.9168  -8.814  < 2e-16 ***
atemp        4063.6006   442.0958   9.192  < 2e-16 ***
hum         -1322.6337   286.5801  -4.615 4.88e-06 ***
windspeed   -2459.2076   420.5688  -5.847 8.52e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 696.6 on 556 degrees of freedom
Multiple R-squared:  0.8562,	Adjusted R-squared:  0.8489
F-statistic: 118.2 on 28 and 556 DF,  p-value: < 2.2e-16
```

VIF (Model 1)

```
                  GVIF Df GVIF^(1/(2*Df))
instant     389.672719  1       19.740130
season      201.808543  3        2.421902
yr          428.010099  1       20.688405
mnth      14123.817110 11        1.543953
holiday       1.123051  1        1.059741
weekday       1.173688  6        1.013435
weathersit    1.929533  2        1.178591
atemp         6.633379  1        2.575535
hum           2.160245  1        1.469777
windspeed     1.235671  1        1.111607
```

Model 1.1

Summary Output

```
Call:
lm(formula = cnt ~ instant + season + holiday + weekday + weathersit +
    atemp + hum + windspeed, data = train)

Residuals:
    Min      1Q  Median      3Q     Max
-2815.4  -345.5    57.8   488.7  3168.2

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1295.8214   236.1899   5.486 6.19e-08 ***
instant         5.3417     0.1941  27.524  < 2e-16 ***
season2       929.2686   107.1799   8.670  < 2e-16 ***
season3       195.8333   147.9095   1.324 0.186033
season4       207.2874   108.3503   1.913 0.056235 .
holiday1     -457.5914   197.3687  -2.318 0.020778 *
weekday1      104.1721   118.1298   0.882 0.378234
weekday2      230.2300   115.1389   2.000 0.046021 *
weekday3      177.3823   115.7941   1.532 0.126110
weekday4      245.4508   115.5016   2.125 0.034010 *
weekday5      264.1208   115.8509   2.280 0.022987 *
weekday6      275.4644   114.9423   2.397 0.016873 *
weathersit2  -309.6739    82.9343  -3.734 0.000207 ***
weathersit3 -1637.3816   205.1871  -7.980 8.13e-15 ***
atemp        4770.3093   346.6217  13.762  < 2e-16 ***
hum         -1457.3303   289.0718  -5.041 6.22e-07 ***
windspeed   -2371.9396   440.0012  -5.391 1.03e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 742.4 on 568 degrees of freedom
Multiple R-squared:  0.8331,	Adjusted R-squared:  0.8284
F-statistic: 177.2 on 16 and 568 DF,  p-value: < 2.2e-16
```

VIF (Model 1.1)

```
               GVIF Df GVIF^(1/(2*Df))
instant    1.140108  1        1.067758
season     3.635701  3        1.240028
holiday    1.099952  1        1.048786
weekday    1.149200  6        1.011656
weathersit 1.856945  2        1.167346
atemp      3.589822  1        1.894683
hum        1.935003  1        1.391044
windspeed  1.190682  1        1.091184
```

Breusch – Pagan Test (Model 1.1)

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 28.96285, Df = 1, p = 7.378e-08
```

Model 1.2

Summary Output

```
Call:
lm(formula = cnt ~ instant + season + holiday + weekday + weathersit +
    atemp + sq_atemp + hum + windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max 
-2782.73  -356.65    17.31   462.75  2969.76 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -43.2210   285.1508  -0.152  0.87958    
instant         5.1935     0.1859  27.930  < 2e-16 ***
season2       813.3162   103.2543   7.877 1.72e-14 ***
season3       428.6477   144.1975   2.973  0.00308 ** 
season4        30.8222   105.7972   0.291  0.77090    
holiday1     -384.3398   188.3229  -2.041  0.04173 *  
weekday1       83.4978   112.6023   0.742  0.45868    
weekday2      205.0397   109.7691   1.868  0.06229 .  
weekday3      153.5576   110.3880   1.391  0.16475    
weekday4      218.8926   110.1201   1.988  0.04732 *  
weekday5      233.0285   110.4730   2.109  0.03535 *  
weekday6      258.3643   109.5551   2.358  0.01870 *  
weathersit2  -334.6017    79.0980  -4.230 2.72e-05 ***
weathersit3 -1675.5458   195.5932  -8.566  < 2e-16 ***
atemp       13377.3911  1172.8504  11.406  < 2e-16 ***
sq_atemp    -9641.8276  1260.6710  -7.648 8.79e-14 ***
hum         -1744.9696   278.0216  -6.276 6.89e-10 ***
windspeed   -2599.2203   420.3435  -6.184 1.20e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 707.4 on 567 degrees of freedom
Multiple R-squared:  0.8487,	Adjusted R-squared:  0.8441 
F-statistic: 187.1 on 17 and 567 DF,  p-value: < 2.2e-16
```

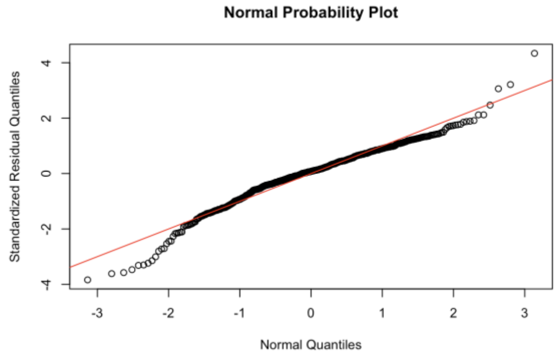
Breusch – Pagan Test (Model 1.2)

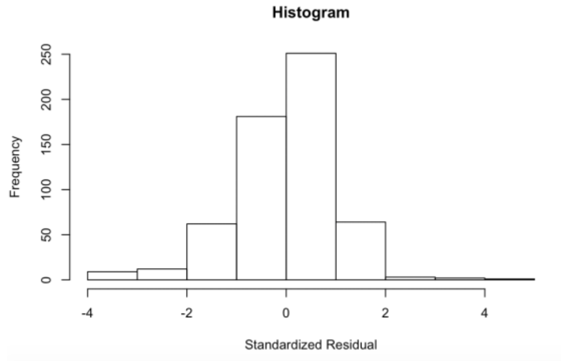
```
Non-constant Variance Score Test 
Variance formula: ~ fitted.values 
Chisquare = 0.7450636, Df = 1, p = 0.38804
```

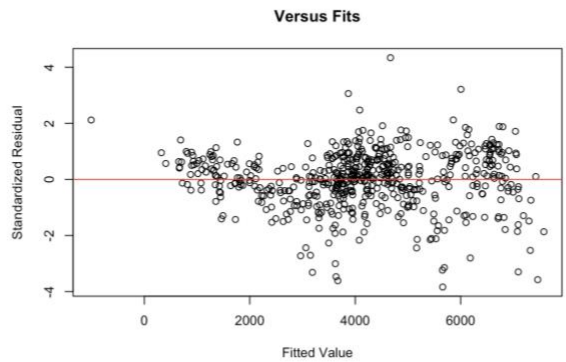
Cochrane-Orcutt Test (Model 1.2)

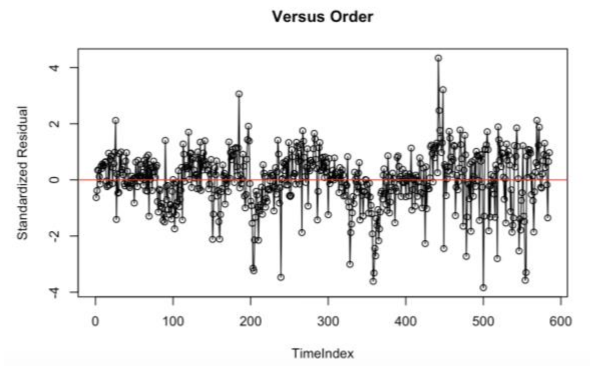
```
Call:
lm(formula = model2$residuals ~ Lag(model2$residuals, 1))

Residuals:
    Min      1Q  Median      3Q     Max 
-2833.0  -308.3     7.7   362.7  2970.0 

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)               1.76464   26.23976   0.067    0.946    
Lag(model2$residuals, 1)  0.41739    0.03769  11.073   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 634.1 on 582 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.174,	Adjusted R-squared:  0.1726 
F-statistic: 122.6 on 1 and 582 DF,  p-value: < 2.2e-16
```

Model Diagnostics (Model 1.1)





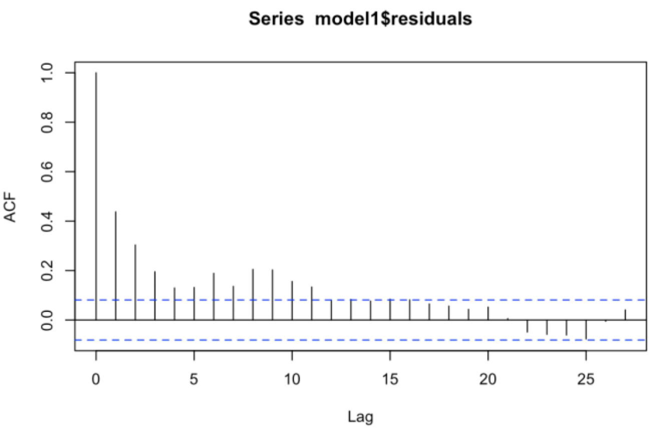




Cochrane-Orcutt Test (Model 1.2)

```
Call:
lm(formula = star_cnt ~ star_instant + star_season + star_holiday +
    star_weekday + star_weathersit + star_atemp + star_sq_atemp +
    star_hum + star_windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2993.89  -364.62     3.45   392.36  2785.43

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        631.767    281.618   2.243   0.0253 *
star_instant         5.117      0.305  16.777  < 2e-16 ***
star_season        -63.587     52.931  -1.201   0.2301
star_holiday      -547.808    301.635  -1.816   0.0699 .
star_weekday        56.006     24.637   2.273   0.0234 *
star_weathersit   -654.150    115.212  -5.678 2.17e-08 ***
star_atemp       15992.381   1393.477  11.477  < 2e-16 ***
star_sq_atemp   -11282.593   1470.636  -7.672 7.31e-14 ***
star_hum         -2201.888    293.630  -7.499 2.46e-13 ***
star_windspeed   -2067.897    408.477  -5.062 5.58e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 687.2 on 574 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.6911,    Adjusted R-squared:  0.6862
F-statistic: 142.7 on 9 and 574 DF,  p-value: < 2.2e-16
```

ACF (Model 1.2)



Cochrane-Orcutt Method (Model 1.2)

```
Call:
lm(formula = star_cnt ~ star_instant + star_season + star_holiday +
    star_weekday + star_weathersit + star_atemp + star_sq_atemp +
    star_hum + star_windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2993.89  -364.62     3.45   392.36  2785.43

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        631.767    281.618   2.243   0.0253 *
star_instant         5.117      0.305  16.777  < 2e-16 ***
star_season        -63.587     52.931  -1.201   0.2301
star_holiday      -547.808    301.635  -1.816   0.0699 .
star_weekday        56.006     24.637   2.273   0.0234 *
star_weathersit   -654.150    115.212  -5.678 2.17e-08 ***
star_atemp       15992.381   1393.477  11.477  < 2e-16 ***
star_sq_atemp   -11282.593   1470.636  -7.672 7.31e-14 ***
star_hum         -2201.888    293.630  -7.499 2.46e-13 ***
star_windspeed   -2067.897    408.477  -5.062 5.58e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 687.2 on 574 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.6911,    Adjusted R-squared:  0.6862
F-statistic: 142.7 on 9 and 574 DF,  p-value: < 2.2e-16
```

Model 2

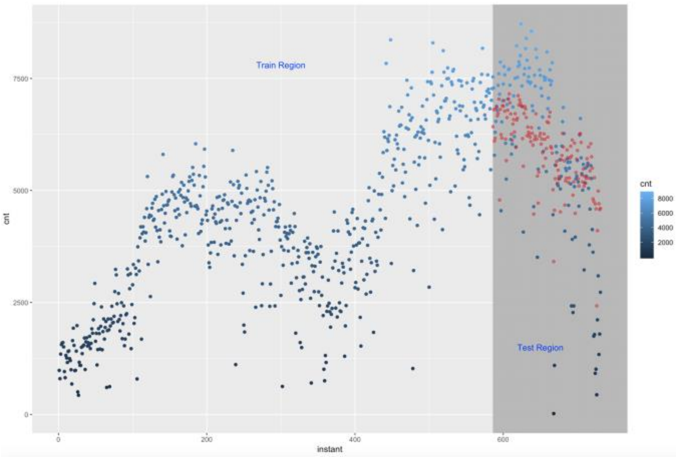
Summary Output

```
Call:
lm(formula = cnt ~ instant + season + yr + mnth + holiday + weekday +
    weathersit + atemp + hum + windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2952.40  -352.40    60.79   424.42  2911.84

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1476.5775   234.7495   6.290 6.43e-10 ***
instant       -0.7819     3.3664  -0.232  0.81642
season2      704.5961   172.0553   4.095 4.85e-05 ***
season3      564.8749   223.7072   2.525  0.01184 *
season4     1049.4399   219.2245   4.787 2.17e-06 ***
yr1         2331.2847  1230.0039   1.895  0.05856 .
mnth2        209.2672   162.9109   1.285  0.19948
mnth3        775.9748   240.7787   3.223  0.00134 **
mnth4        868.2082   358.1356   2.424  0.01566 *
mnth5       1285.7127   446.1378   2.882  0.00411 **
mnth6       1257.0639   536.2667   2.344  0.01942 *
mnth7        899.1208   634.0192   1.418  0.15671
mnth8       1144.5075   720.1486   1.589  0.11257
mnth9       1566.8886   819.0146   1.913  0.05624 .
mnth10      1296.2071   938.1915   1.382  0.16765
mnth11       929.6281  1037.9188   0.896  0.37082
mnth12       896.1370  1131.4113   0.792  0.42867
holiday1    -358.8070   187.1203  -1.918  0.05568 .
weekday1     105.9400   110.9639   0.955  0.34013
weekday2     246.2097   108.2730   2.274  0.02335 *
weekday3     203.3237   108.8821   1.867  0.06238 .
weekday4     261.8080   108.7788   2.407  0.01642 *
weekday5     278.9916   108.8031   2.564  0.01060 *
weekday6     290.5028   107.9156   2.692  0.00732 **
weathersit2 -384.9532    78.9407  -4.876 1.41e-06 ***
weathersit3 -1709.2328   193.9168  -8.814  < 2e-16 ***
atemp       4063.6006   442.0958   9.192  < 2e-16 ***
hum        -1322.6337   286.5801  -4.615 4.88e-06 ***
windspeed  -2459.2076   420.5688  -5.847 8.52e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 696.6 on 556 degrees of freedom
Multiple R-squared:  0.8562,    Adjusted R-squared:  0.8489
F-statistic: 118.2 on 28 and 556 DF,  p-value: < 2.2e-16
```

VIF (Model 2)

```
   instant     season2     season3     season4
389.672719    7.695273   11.091176    7.474382
     mnth4       mnth5       mnth6       mnth7
 14.234135   22.738247   31.915215   45.922309
    mnth11      mnth12    holiday1    weekday1
 63.192642   77.452957    1.123051    1.825600
  weekday5    weekday6 weathersit2 weathersit3
  1.737759    1.726678    1.647699    1.352109

       yr1       mnth2       mnth3
428.010099    2.814067    6.623003
     mnth8       mnth9      mnth10
 37.978751   39.348028   53.257395
  weekday2    weekday3    weekday4
  1.738133    1.740286    1.736983
     atemp         hum   windspeed
  6.633379    2.160245    1.235671
```

Prediction (Model 1.2)



Model 2.2
Summary Output

```
Call:
lm(formula = cnt ~ season + holiday + yr + weekday + weathersit +
    atemp + sq_atemp + hum + windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2748.81  -381.42    17.95   436.71  2897.74

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -220.01     271.72  -0.810  0.41845
season2          1038.78      98.77  10.517  < 2e-16 ***
season3          1097.33     139.89   7.844 2.18e-14 ***
season4          1229.02     106.59  11.530  < 2e-16 ***
holiday1         -316.44     179.13  -1.767  0.07783 .
yr1              1921.52      63.29  30.359  < 2e-16 ***
weekday1           75.09     107.11   0.701  0.48358
weekday2          201.74     104.42   1.932  0.05385 .
weekday3          161.54     105.01   1.538  0.12451
weekday4          234.72     104.76   2.241  0.02544 *
weekday5          253.73     105.10   2.414  0.01608 *
weekday6          281.00     104.22   2.696  0.00722 **
weathersit2      -399.13      75.31  -5.300 1.66e-07 ***
weathersit3     -1719.62     186.11  -9.240  < 2e-16 ***
atemp           15203.31    1106.46  13.741  < 2e-16 ***
sq_atemp       -11289.39    1194.53  -9.451  < 2e-16 ***
hum             -1514.03     265.38  -5.705 1.88e-08 ***
windspeed       -2705.67     399.61  -6.771 3.21e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 673 on 567 degrees of freedom
Multiple R-squared:  0.8631,    Adjusted R-squared:  0.859
F-statistic: 210.2 on 17 and 567 DF,  p-value: < 2.2e-16
```

Breusch – Pagan Test (Model 2.2)

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 2.786994, Df = 1, p = 0.095032
```

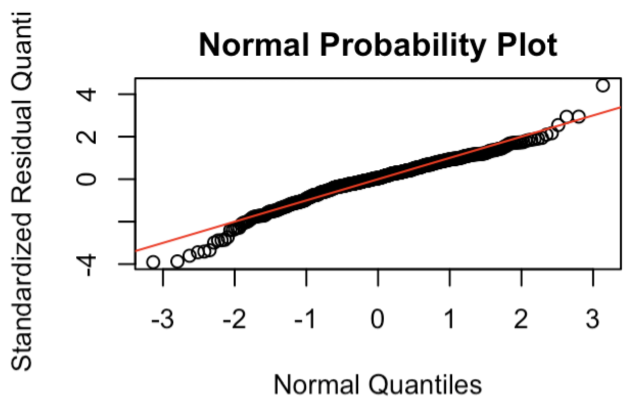
Model 2.1
Summary Output

```
Call:
lm(formula = cnt ~ season + yr + holiday + weekday + weathersit +
    atemp + hum + windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2790.15  -362.63    34.59   477.45  3140.45

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1367.95     229.54   5.959 4.45e-09 ***
season2      1177.73     104.99  11.218  < 2e-16 ***
season3       831.14     147.30   5.643 2.65e-08 ***
season4      1460.08     111.53  13.092  < 2e-16 ***
yr1          1954.87      67.93  28.778  < 2e-16 ***
holiday1     -400.56     192.31  -2.083  0.03771 *
weekday1       98.98     115.11   0.860  0.39018
weekday2      230.74     112.19   2.057  0.04018 *
weekday3      189.23     112.83   1.677  0.09407 .
weekday4      265.45     112.56   2.358  0.01869 *
weekday5      290.22     112.90   2.571  0.01040 *
weekday6      301.36     112.01   2.691  0.00734 **
weathersit2  -370.19      80.88  -4.577 5.80e-06 ***
weathersit3 -1671.46     199.98  -8.358 4.93e-16 ***
atemp        5168.01     334.35  15.457  < 2e-16 ***
hum         -1187.04     282.84  -4.197 3.14e-05 ***
windspeed   -2450.80     428.58  -5.718 1.74e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 723.4 on 568 degrees of freedom
Multiple R-squared:  0.8415,    Adjusted R-squared:  0.837
F-statistic: 188.5 on 16 and 568 DF,  p-value: < 2.2e-16
```

VIF (Model 2.1)

```
  season2    season3    season4        yr1
 2.656648   4.458501   1.793625   1.210451
 weekday3   weekday4   weekday5   weekday6
 1.732887   1.724376   1.734774   1.724661
      hum  windspeed
 1.951052   1.189817

   holiday1    weekday1    weekday2
   1.099887    1.821444    1.730399
weathersit2 weathersit3       atemp
   1.603923    1.333387    3.518054
```

Model diagnostics (Model 2.1)



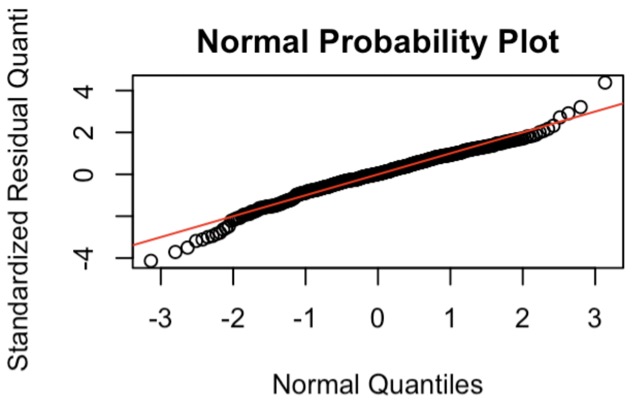
Cochrane-Orcutt test (Model 2.2)

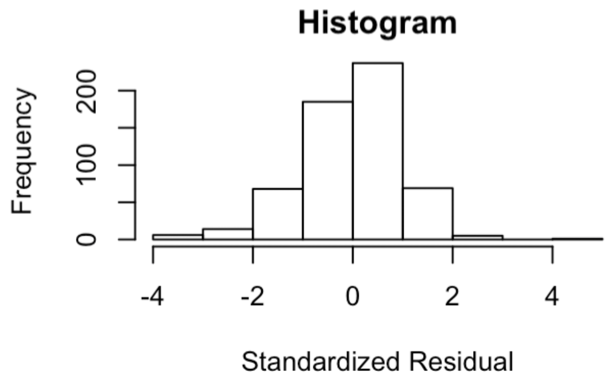
```
Call:
lm(formula = model2$residuals ~ Lag(model2$residuals, 1))

Residuals:
     Min       1Q   Median       3Q      Max 
-2837.97  -324.53     9.61   401.67  2922.69 

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)               2.35754   25.80679   0.091    0.927    
Lag(model2$residuals, 1)  0.33833    0.03899   8.678   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 623.6 on 582 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1146,	Adjusted R-squared:  0.1131 
F-statistic: 75.31 on 1 and 582 DF,  p-value: < 2.2e-16
```

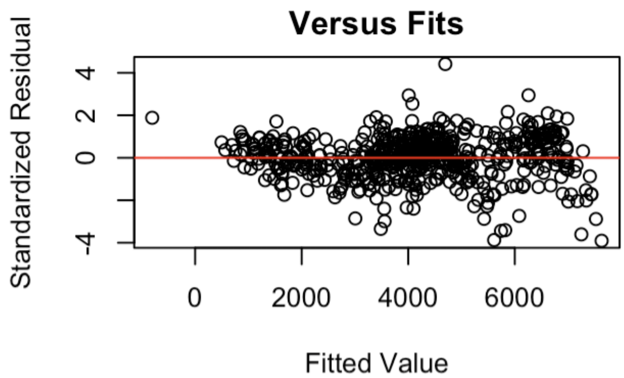
Cochrane-Orcutt Method (Model 2.2)

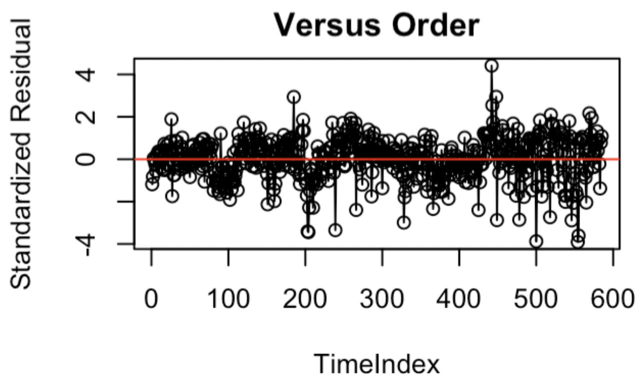
```
Call:
lm(formula = star_cnt ~ star_season + star_holiday + star_yr + 
    star_weekday + star_weathersit + star_atemp + star_sq_atemp + 
    star_hum + star_windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max 
-2857.65  -363.31    23.33   402.78  2772.56 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -1023.85     295.82  -3.461 0.000578 ***
star_season        347.44      48.85   7.112 3.42e-12 ***
star_holiday      -430.25     257.36  -1.672 0.095112 .  
star_yr           1934.76      93.59  20.673  < 2e-16 ***
star_weekday        57.17      21.03   2.719 0.006752 ** 
star_weathersit   -662.09      99.47  -6.656 6.57e-11 ***
star_atemp       17271.00    1271.60  13.582  < 2e-16 ***
star_sq_atemp   -12649.12    1336.92  -9.461  < 2e-16 ***
star_hum         -1891.61     286.62  -6.600 9.38e-11 ***
star_windspeed   -2194.22     400.55  -5.478 6.44e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 666.1 on 574 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.749,	Adjusted R-squared:  0.7451 
F-statistic: 190.3 on 9 and 574 DF,  p-value: < 2.2e-16
```

Model Diagnostics (Model 2.2)



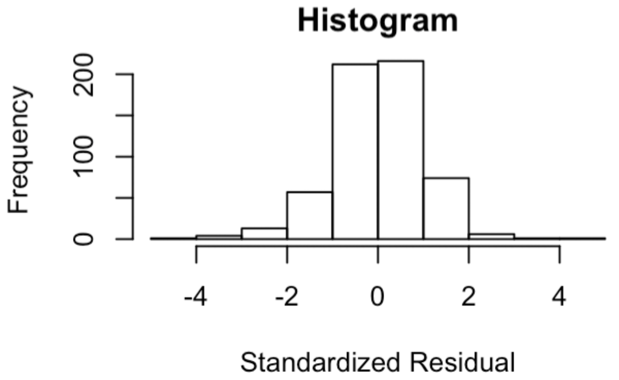


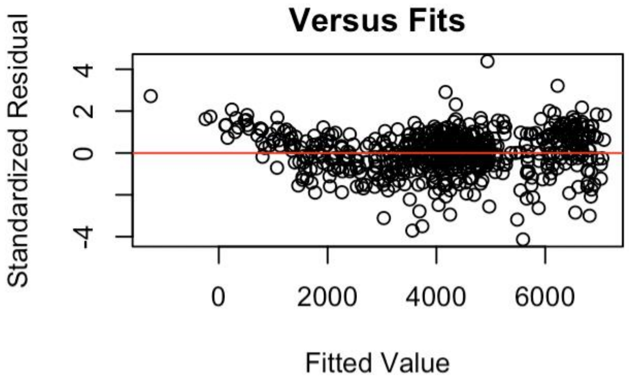


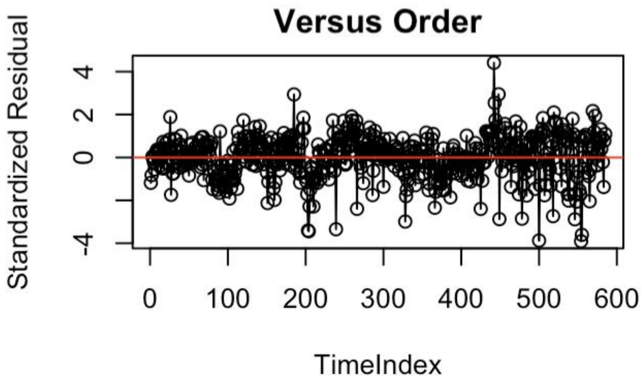


Breusch – Pagan Test (Model 2.1)

```
Non-constant Variance Score Test 
Variance formula: ~ fitted.values 
Chisquare = 46.11916, Df = 1, p = 1.1127e-11
```



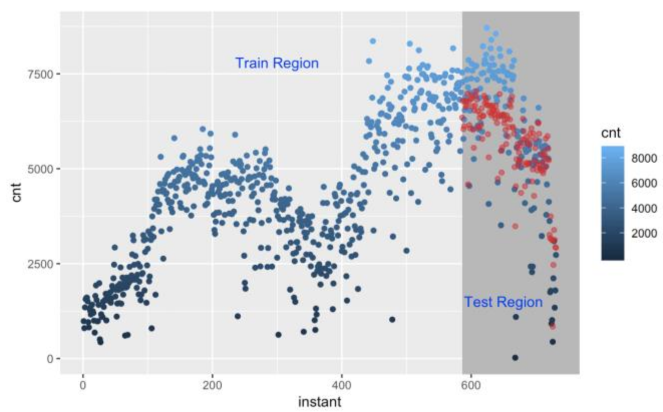




Model 3 (Registered Users) Summary Output

```
Call:
lm(formula = registered ~ instant + season + yr + mnth + holiday +
    weekday + weathersit + atemp + hum + windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2323.50  -237.46    56.98   343.96  1360.25

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.311e+06  1.839e+06  -1.800 0.072329 .
instant      3.224e-01  2.503e+00   0.129 0.897547
season2      4.866e+02  1.279e+02   3.804 0.000158 ***
season3      4.696e+02  1.663e+02   2.823 0.004927 **
season4      1.018e+03  1.630e+02   6.248 8.27e-10 ***
yr           1.647e+03  9.146e+02   1.801 0.072265 .
mnth2        1.883e+02  1.211e+02   1.554 0.120695
mnth3        4.312e+02  1.790e+02   2.408 0.016344 *
mnth4        4.928e+02  2.663e+02   1.851 0.064738 .
mnth5        8.461e+02  3.317e+02   2.551 0.011020 *
mnth6        9.114e+02  3.987e+02   2.286 0.022645 *
mnth7        5.686e+02  4.714e+02   1.206 0.228243
mnth8        7.647e+02  5.355e+02   1.428 0.153795
mnth9        9.901e+02  6.090e+02   1.626 0.104561
mnth10       5.986e+02  6.976e+02   0.858 0.391205
mnth11       3.795e+02  7.717e+02   0.492 0.623072
mnth12       4.696e+02  8.413e+02   0.558 0.576933
holiday1    -9.074e+02  1.391e+02  -6.522 1.57e-10 ***
weekday1     8.429e+02  8.251e+01  10.217  < 2e-16 ***
weekday2     1.032e+03  8.051e+01  12.819  < 2e-16 ***
weekday3     9.949e+02  8.096e+01  12.289  < 2e-16 ***
weekday4     1.050e+03  8.088e+01  12.984  < 2e-16 ***
weekday5     9.148e+02  8.090e+01  11.308  < 2e-16 ***
weekday6     1.983e+02  8.024e+01   2.472 0.013740 *
weathersit2 -2.922e+02  5.870e+01  -4.978 8.58e-07 ***
weathersit3 -1.403e+03  1.442e+02  -9.732  < 2e-16 ***
atemp        2.364e+03  3.287e+02   7.192 2.08e-12 ***
hum         -8.310e+02  2.131e+02  -3.900 0.000108 ***
windspeed   -1.364e+03  3.127e+02  -4.360 1.55e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 517.9 on 556 degrees of freedom
Multiple R-squared:  0.8687,    Adjusted R-squared:  0.8621
F-statistic: 131.4 on 28 and 556 DF,  p-value: < 2.2e-16
```

Prediction (Model 2.2)



VIF (Model 3 Registered Users)

```
   instant     season2     season3     season4
389.672719    7.695273   11.091176    7.474382
     mnth4       mnth5       mnth6       mnth7
 14.234135   22.738247   31.915215   45.922309
    mnth11      mnth12    holiday1    weekday1
 63.192642   77.452957    1.123051    1.825600
  weekday5    weekday6 weathersit2 weathersit3
  1.737759    1.726678    1.647699    1.352109
        yr       mnth2       mnth3
428.010099    2.814067    6.623003
     mnth8       mnth9      mnth10
 37.978751   39.348028   53.257395
  weekday2    weekday3    weekday4
  1.738133    1.740286    1.736983
     atemp         hum   windspeed
  6.633379    2.160245    1.235671
```

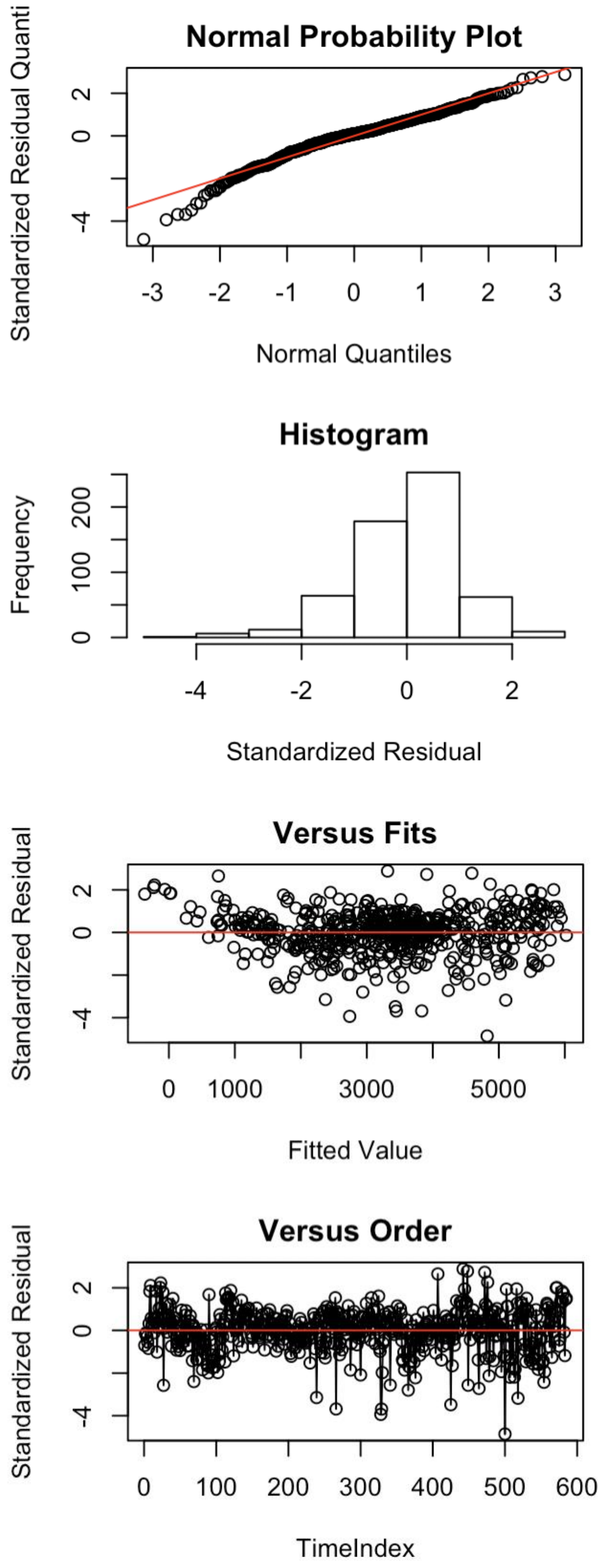
Model 3.1 (Registered Users) Summary Output

```
Call:
lm(formula = registered ~ mnth + weekday + weathersit + season +
    atemp + holiday + windspeed + hum + yr, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2323.96  -237.38    57.09   341.06  1360.91

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.548e+06  1.012e+05 -35.047  < 2e-16 ***
mnth2        1.976e+02  9.712e+01   2.034 0.042399 *
mnth3        4.489e+02  1.141e+02   3.934 9.39e-05 ***
mnth4        5.187e+02  1.747e+02   2.970 0.003106 **
mnth5        8.811e+02  1.895e+02   4.650 4.16e-06 ***
mnth6        9.556e+02  2.026e+02   4.716 3.04e-06 ***
mnth7        6.214e+02  2.337e+02   2.659 0.008068 **
mnth8        8.268e+02  2.345e+02   3.526 0.000457 ***
mnth9        1.064e+03  2.129e+02   4.996 7.86e-07 ***
mnth10       6.842e+02  2.127e+02   3.216 0.001374 **
mnth11       4.752e+02  2.082e+02   2.283 0.022794 *
mnth12       5.760e+02  1.602e+02   3.596 0.000352 ***
weekday1     8.431e+02  8.242e+01  10.230  < 2e-16 ***
weekday2     1.032e+03  8.043e+01  12.830  < 2e-16 ***
weekday3     9.949e+02  8.089e+01  12.300  < 2e-16 ***
weekday4     1.050e+03  8.081e+01  12.995  < 2e-16 ***
weekday5     9.149e+02  8.082e+01  11.320  < 2e-16 ***
weekday6     1.986e+02  8.015e+01   2.477 0.013541 *
weathersit2 -2.928e+02  5.845e+01  -5.009 7.34e-07 ***
weathersit3 -1.404e+03  1.439e+02  -9.755  < 2e-16 ***
season2      4.887e+02  1.268e+02   3.853 0.000130 ***
season3      4.731e+02  1.639e+02   2.886 0.004058 **
season4      1.020e+03  1.626e+02   6.270 7.23e-10 ***
atemp        2.369e+03  3.264e+02   7.258 1.33e-12 ***
holiday1    -9.075e+02  1.390e+02  -6.529 1.49e-10 ***
windspeed   -1.362e+03  3.121e+02  -4.363 1.53e-05 ***
hum         -8.285e+02  2.120e+02  -3.908 0.000104 ***
yr           1.765e+03  5.033e+01  35.063  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 517.5 on 557 degrees of freedom
Multiple R-squared:  0.8687,    Adjusted R-squared:  0.8623
F-statistic: 136.5 on 27 and 557 DF,  p-value: < 2.2e-16
```

VIF (Model 3.1 Registered Users)

```
    mnth2       mnth3       mnth4       mnth5
 1.812388    2.695255    6.133744    7.434709
    mnth9      mnth10      mnth11      mnth12
 4.817345    4.960811    4.605128    2.813161
 weekday4    weekday5    weekday6 weathersit2
 1.736956    1.737563    1.725985    1.636887
  season4       atemp    holiday1   windspeed
 7.451570    6.550625    1.122942    1.232934

    mnth6       mnth7       mnth8
 8.255656   11.306058    7.295660
 weekday1    weekday2    weekday3
 1.824972    1.738124    1.740286
weathersit3     season2     season3
 1.349713    7.576992   10.793773
      hum          yr
 2.142219    1.298393
```

Model Diagnostics(Model 3.1 Registered)



Breusch – Pagan Test (Model 3.1 Reg)

```
Residual standard error: 517.5 on 557 degrees of freedom
Multiple R-squared:  0.8687,	Adjusted R-squared:  0.8623
F-statistic: 136.5 on 27 and 557 DF,  p-value: < 2.2e-16
```

VIF (Model 3.2 Registered)

```
     mnth2      mnth3      mnth4      mnth5
  1.901220   2.921642   6.446150   7.527668
     mnth9     mnth10     mnth11     mnth12
  4.981405   5.247591   4.925022   2.998566
  weekday4   weekday5   weekday6 weathersit2 w
  1.737084   1.738780   1.726452   1.644053
   season4      atemp   sq_atemp   holiday1
  7.480424  62.205563  66.819774   1.129181

     mnth6      mnth7      mnth8
  8.256785  11.388910   7.297744
  weekday1   weekday2   weekday3
  1.826532   1.738913   1.740963
weathersit3   season2    season3
  1.350522   7.576994  10.934669
 windspeed        hum         yr
  1.232935   2.153346   1.321158
```

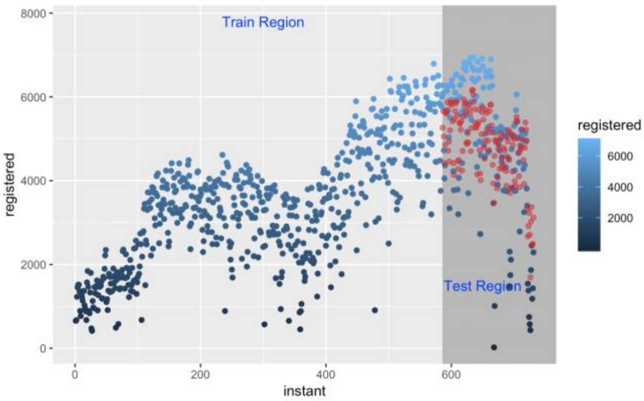
Breusch – Pagan Test (Model 3.2 Reg)

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 1.125665, Df = 1, p = 0.2887
```

Model 3.2 (Registered)
Summary output

```
Call:
lm(formula = registered ~ mnth + weekday + weathersit + season +
    atemp + sq_atemp + holiday + windspeed + hum + yr, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-2322.33  -220.65    42.28   286.10  1366.61

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.440e+06  9.592e+04 -35.864  < 2e-16 ***
mnth2        2.251e+01  9.345e+01   0.241 0.809755
mnth3        1.797e+02  1.116e+02   1.610 0.107954
mnth4        1.978e+02  1.682e+02   1.176 0.240111
mnth5        7.086e+02  1.792e+02   3.955 8.63e-05 ***
mnth6        9.364e+02  1.904e+02   4.919 1.15e-06 ***
mnth7        7.842e+02  2.204e+02   3.559 0.000404 ***
mnth8        7.945e+02  2.203e+02   3.606 0.000339 ***
mnth9        7.437e+02  2.034e+02   3.657 0.000280 ***
mnth10       2.678e+02  2.055e+02   1.303 0.193178
mnth11       2.859e+01  2.022e+02   0.141 0.887624
mnth12       2.412e+02  1.554e+02   1.552 0.121187
weekday1     8.235e+02  7.746e+01  10.631  < 2e-16 ***
weekday2     1.018e+03  7.558e+01  13.470  < 2e-16 ***
weekday3     9.819e+02  7.600e+01  12.920  < 2e-16 ***
weekday4     1.045e+03  7.592e+01  13.758  < 2e-16 ***
weekday5     8.975e+02  7.596e+01  11.816  < 2e-16 ***
weekday6     1.878e+02  7.531e+01   2.494 0.012929 *
weathersit2 -3.243e+02  5.503e+01  -5.893 6.58e-09 ***
weathersit3 -1.433e+03  1.353e+02 -10.592  < 2e-16 ***
season2      4.881e+02  1.192e+02   4.097 4.81e-05 ***
season3      6.256e+02  1.550e+02   4.035 6.21e-05 ***
season4      1.102e+03  1.531e+02   7.200 1.97e-12 ***
atemp        1.011e+04  9.449e+02  10.704  < 2e-16 ***
sq_atemp    -9.010e+03  1.040e+03  -8.666  < 2e-16 ***
holiday1    -8.232e+02  1.309e+02  -6.286 6.58e-10 ***
windspeed   -1.363e+03  2.932e+02  -4.647 4.20e-06 ***
hum         -9.529e+02  1.997e+02  -4.772 2.33e-06 ***
yr           1.710e+03  4.769e+01  35.862  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 486.1 on 556 degrees of freedom
Multiple R-squared:  0.8843,	Adjusted R-squared:  0.8785
F-statistic: 151.8 on 28 and 556 DF,  p-value: < 2.2e-16
```

Prediction (Model 3.2 Registered)



Model 3 (Casual Users)
Summary Output

```
Call:
lm(formula = Log_Casual ~ instant + season + yr + mnth + holiday +
    weekday + weathersit + atemp + hum + windspeed, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-1.83391 -0.17868  0.03908  0.22242  1.20155

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.411e+03  1.384e+03  -1.019 0.308706
instant     -8.567e-04  1.884e-03  -0.455 0.649537
season2      3.426e-01  9.630e-02   3.558 0.000406 ***
season3      1.632e-01  1.252e-01   1.303 0.192992
season4      8.955e-02  1.227e-01   0.730 0.465828
yr           7.041e-01  6.885e-01   1.023 0.306923
mnth2        2.342e-01  9.119e-02   2.568 0.010488 *
mnth3        9.015e-01  1.348e-01   6.689 5.48e-11 ***
mnth4        8.190e-01  2.005e-01   4.085 5.05e-05 ***
mnth5        7.922e-01  2.497e-01   3.172 0.001595 **
mnth6        5.680e-01  3.002e-01   1.892 0.058989 .
mnth7        4.886e-01  3.549e-01   1.377 0.169133
mnth8        6.978e-01  4.031e-01   1.731 0.083981 .
mnth9        1.051e+00  4.584e-01   2.292 0.022286 *
mnth10       1.267e+00  5.251e-01   2.413 0.016133 *
mnth11       1.042e+00  5.809e-01   1.794 0.073408 .
mnth12       7.418e-01  6.333e-01   1.171 0.241974
holiday1     5.906e-01  1.047e-01   5.639 2.72e-08 ***
weekday1    -8.005e-01  6.211e-02 -12.888  < 2e-16 ***
weekday2    -9.231e-01  6.060e-02 -15.233  < 2e-16 ***
weekday3    -9.571e-01  6.094e-02 -15.705  < 2e-16 ***
weekday4    -9.432e-01  6.089e-02 -15.492  < 2e-16 ***
weekday5    -6.645e-01  6.090e-02 -10.911  < 2e-16 ***
weekday6     5.440e-02  6.040e-02   0.901 0.368160
weathersit2 -1.908e-01  4.419e-02  -4.319 1.85e-05 ***
weathersit3 -1.022e+00  1.085e-01  -9.419  < 2e-16 ***
atemp        3.452e+00  2.475e-01  13.949  < 2e-16 ***
hum         -7.961e-01  1.604e-01  -4.963 9.24e-07 ***
windspeed   -1.243e+00  2.354e-01  -5.281 1.85e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3899 on 556 degrees of freedom
Multiple R-squared:  0.862,	Adjusted R-squared:  0.855
F-statistic:   124 on 28 and 556 DF,  p-value: < 2.2e-16
```

VIF (Model 3 Casual)

```
    instant     season2     season3     season4
 389.672719    7.695273   11.091176    7.474382
      mnth4       mnth5       mnth6       mnth7
  14.234135   22.738247   31.915215   45.922309
     mnth11      mnth12    holiday1    weekday1
  63.192642   77.452957    1.123051    1.825600
   weekday5    weekday6 weathersit2 weathersit3
   1.737759    1.726678    1.647699    1.352109

         yr       mnth2       mnth3
 428.010099    2.814067    6.623003
      mnth8       mnth9      mnth10
  37.978751   39.348028   53.257395
   weekday2    weekday3    weekday4
   1.738133    1.740286    1.736983
      atemp         hum   windspeed
   6.633379    2.160245    1.235671
```

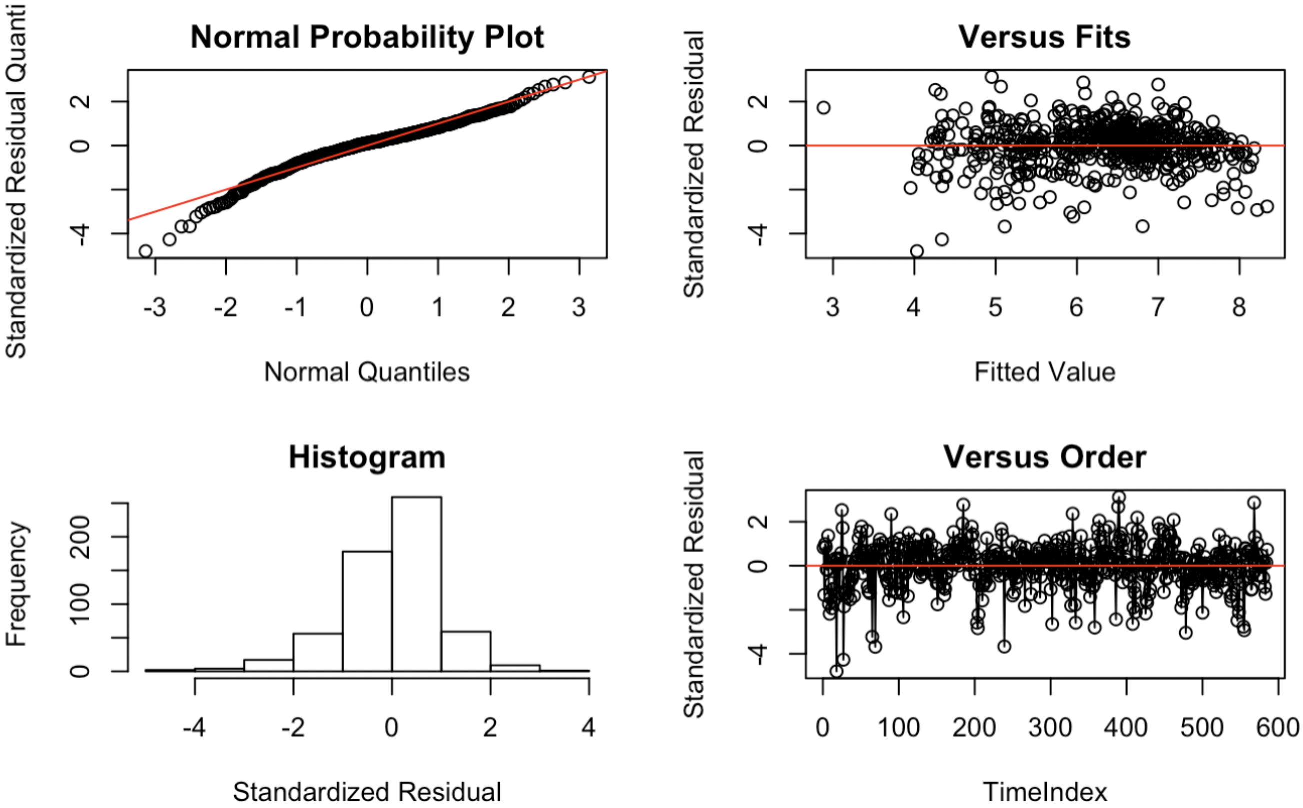
Model 3.1 (Casual Users)
Summary Output

```
Call:
lm(formula = Log_Casual ~ mnth + weekday + atemp + weathersit +
    yr + holiday + windspeed + hum + season, data = train)

Residuals:
     Min       1Q   Median       3Q      Max
-1.83518 -0.18628  0.04014  0.22005  1.19510

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -782.11591   76.21997 -10.261  < 2e-16 ***
mnth2          0.20943    0.07313   2.864 0.004341 **
mnth3          0.85432    0.08591   9.944  < 2e-16 ***
mnth4          0.75020    0.13150   5.705 1.89e-08 ***
mnth5          0.69906    0.14269   4.899 1.26e-06 ***
mnth6          0.45045    0.15255   2.953 0.003282 **
mnth7          0.34851    0.17596   1.981 0.048127 *
mnth8          0.53307    0.17654   3.020 0.002648 **
mnth9          0.85540    0.16029   5.337 1.38e-07 ***
mnth10         1.03991    0.16016   6.493 1.86e-10 ***
mnth11         0.78771    0.15672   5.026 6.75e-07 ***
mnth12         0.45912    0.12060   3.807 0.000156 ***
weekday1      -0.80101    0.06205 -12.908  < 2e-16 ***
weekday2      -0.92321    0.06056 -15.245  < 2e-16 ***
weekday3      -0.95712    0.06090 -15.716  < 2e-16 ***
weekday4      -0.94311    0.06084 -15.501  < 2e-16 ***
weekday5      -0.66477    0.06085 -10.924  < 2e-16 ***
weekday6       0.05385    0.06035   0.892 0.372583
atemp          3.43912    0.24573  13.996  < 2e-16 ***
weathersit2   -0.18922    0.04401  -4.300 2.02e-05 ***
weathersit3   -1.02029    0.10837  -9.415  < 2e-16 ***
yr             0.39151    0.03789  10.332  < 2e-16 ***
holiday1       0.59111    0.10466   5.648 2.59e-08 ***
windspeed     -1.24818    0.23497  -5.312 1.57e-07 ***
hum           -0.80278    0.15962  -5.029 6.65e-07 ***
season2        0.33720    0.09549   3.531 0.000448 ***
season3        0.15388    0.12344   1.247 0.213064
season4        0.08647    0.12243   0.706 0.480328
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3896 on 557 degrees of freedom
Multiple R-squared:  0.8619,	Adjusted R-squared:  0.8552
F-statistic: 128.8 on 27 and 557 DF,  p-value: < 2.2e-16
```

VIF (Model 3.1 Casual)

```
      mnth2       mnth3       mnth4       mnth5
   1.812388    2.695255    6.133744    7.434709
      mnth9      mnth10      mnth11      mnth12
   4.817345    4.960811    4.605128    2.813161
   weekday4    weekday5    weekday6       atemp
   1.736956    1.737563    1.725985    6.550625
   holiday1   windspeed         hum     season2
   1.122942    1.232934    2.142219    7.576992

      mnth6       mnth7       mnth8
   8.255656   11.306058    7.295660
   weekday1    weekday2    weekday3
   1.824972    1.738124    1.740286
weathersit2 weathersit3          yr
   1.636887    1.349713    1.298393
    season3     season4
  10.793773    7.451570
```

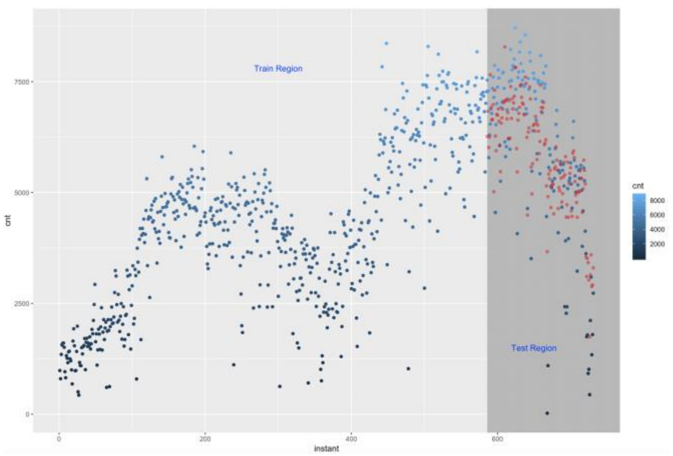
Model Diagnostics (Model 3.1 Casual Users)

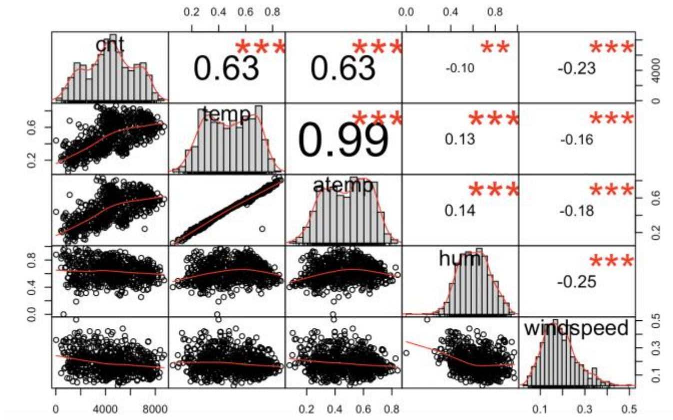


Breusch – Pagan Test (Model 3.1 Casual)

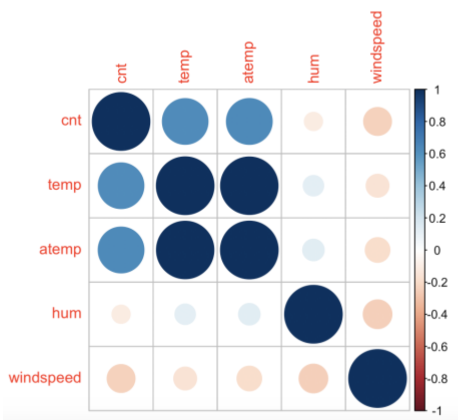
```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 28.88828, Df = 1, p = 7.6676e-08
```

Prediction Combined (Reg + Cas)





Prediction (Model 3.1 Casual)



Side by Side Box Plot (Dependent Variable vs Categorical Predictors)