Group No-241: Bike-Sharing System

A statistical approach to understand and predict bike-sharing demand

First Name	Last Name	Email (hawk.iit.edu)	Student ID
FNU	Pragya Pramudita	ppramudita@hawk.iit.edu	A20430719

Results

Data load and analysis

```
> #get working path
> getwd()
[1] "C:/Users/Pragya/Documents/DA"
>
> #setup working path
> setwd('C:/Users/Pragya/Documents/DA/')
>
> #load data
> myProjectdata=read.table("hour.csv",head =T,sep=',')
>
> #Find the total number of data : 7538
> dim(myProjectdata)
[1] 7538   11
>
> #Check first few data
> head(myProjectdata)
  season month hour IsHoliday IsWeekday IsWorkingday weathersituation temp humidity windspeed bikeCount
1      1     1    5         0         6            0                2 0.24     0.75    0.0896         1
2      1     1   10         0         6            0                1 0.38     0.76    0.2537        36
3      1     1   11         0         6            0                1 0.36     0.81    0.2836        56
4      1     1   12         0         6            0                1 0.42     0.77    0.2836        84
5      1     1   13         0         6            0                2 0.46     0.72    0.2985        94
6      1     1   14         0         6            0                2 0.46     0.72    0.2836       106
>
> #Check last few data
> tail(myProjectdata)
     season month hour IsHoliday IsWeekday IsWorkingday weathersituation temp humidity windspeed bikeCount
7533      1    12   17         0         6            0                1 0.42     0.54    0.1940       129
7534      1    12   18         0         6            0                1 0.42     0.54    0.1343        93
7535      1    12   19         0         6            0                1 0.42     0.54    0.2239        92
7536      1    12   20         0         6            0                1 0.42     0.54    0.2239        71
7537      1    12   21         0         6            0                1 0.40     0.58    0.1940        52
7538      1    12   22         0         6            0                1 0.38     0.62    0.1343        38
>
```

Create local variable :

```
> season = myProjectdata$season
> month =myProjectdata$month
> hour=myProjectdata$hour
> IsHoliday = myProjectdata$IsHoliday
> IsWeekday=myProjectdata$IsWeekday
> IsWorkingday = myProjectdata$IsWorkingday
> weathersituation = myProjectdata$weathersituation
> temp=myProjectdata$temp
> humidity=myProjectdata$humidity
> windspeed=myProjectdata$windspeed
> bikeCount=myProjectdata$bikeCount
>
```

Dataset before Pre-processing

```
> #load data
> myProjectdata=read.table("hour.csv",head =T,sep=',')
> #Find the total number of data : 7538
> dim(myProjectdata)
[1] 7538   11
> #Check first few data
> head(myProjectdata)
  season month hour IsHoliday IsWeekday IsWorkingday weathersituation temp humidity windspeed bikeCount
1      1     1    5         0         6            0                2 0.24     0.75    0.0896         1
2      1     1   10         0         6            0                1 0.38     0.76    0.2537        36
3      1     1   11         0         6            0                1 0.36     0.81    0.2836        56
4      1     1   12         0         6            0                1 0.42     0.77    0.2836        84
5      1     1   13         0         6            0                2 0.46     0.72    0.2985        94
6      1     1   14         0         6            0                2 0.46     0.72    0.2836       106
> #Check last few data
> tail(myProjectdata)
     season month hour IsHoliday IsWeekday IsWorkingday weathersituation temp humidity windspeed bikeCount
7533      1    12   17         0         6            0                1 0.42     0.54    0.1940       129
7534      1    12   18         0         6            0                1 0.42     0.54    0.1343        93
7535      1    12   19         0         6            0                1 0.42     0.54    0.2239        92
7536      1    12   20         0         6            0                1 0.42     0.54    0.2239        71
7537      1    12   21         0         6            0                1 0.40     0.58    0.1940        52
7538      1    12   22         0         6            0                1 0.38     0.62    0.1343        38
> #Check the datatype of data
> str(myProjectdata)
'data.frame':   7538 obs. of  11 variables:
 $ season          : int  1 1 1 1 1 1 1 1 1 1 ...
 $ month           : int  1 1 1 1 1 1 1 1 1 1 ...
 $ hour            : int  5 10 11 12 13 14 15 16 17 18 ...
 $ IsHoliday       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ IsWeekday       : int  6 6 6 6 6 6 6 6 6 6 ...
 $ IsWorkingday    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ weathersituation: int  2 1 1 1 2 2 2 2 2 3 ...
 $ temp            : num  0.24 0.38 0.36 0.42 0.46 0.46 0.44 0.42 0.44 0.42 ...
 $ humidity        : num  0.75 0.76 0.81 0.77 0.72 0.72 0.77 0.82 0.82 0.88 ...
 $ windspeed       : num  0.0896 0.2537 0.2836 0.2836 0.2985 ...
 $ bikeCount       : int  1 36 56 84 94 106 110 93 67 35 ...
> 
```

Conversion of discrete variable to nominal using cut function for variables “month” and “hour”

```
> #create a copy of data in new data frame and work on this data frame to create dummy variable
>
> NewProjectdata <- myProjectdata
> #data Pre-preocessing
> #month converted to categories by using median
>
> NewProjectdata$month <- cut(NewProjectdata$month, breaks=c(-Inf,6,Inf), labels=c("m1","m2"),right=FALSE)
> month=NewProjectdata$month
> #hour converted to catergories by using median
>
> NewProjectdata$hour <- cut(NewProjectdata$hour, breaks=c(-Inf,12,Inf), labels=c("h1","h2"),right=FALSE)
> hour=NewProjectdata$hour
> head(NewProjectdata)
  season month hour IsHoliday IsWeekday IsWorkingday weathersituation temp humidity windspeed bikeCount
1      1    m1   h1         0         6            0                2 0.24     0.75    0.0896         1
2      1    m1   h1         0         6            0                1 0.38     0.76    0.2537        36
3      1    m1   h1         0         6            0                1 0.36     0.81    0.2836        56
4      1    m1   h2         0         6            0                1 0.42     0.77    0.2836        84
5      1    m1   h2         0         6            0                2 0.46     0.72    0.2985        94
6      1    m1   h2         0         6            0                2 0.46     0.72    0.2836       106
> 
```

For the purpose of linear regression, converted **nominal variable to dummy variable :**

```
> #Convert Nominal variable to dummy variable :
>
> library(dummies)
> data=dummy.data.frame(NewProjectdata,names=c("month","hour","season","weathersituation"))
> head(data)
  season1 season2 season3 season4 monthm1 monthm2 hourh1 hourh2 IsHoliday IsWeekday IsWorkingday
1       1       0       0       0       1       0      1      0         0         6            0
2       1       0       0       0       1       0      1      0         0         6            0
3       1       0       0       0       1       0      1      0         0         6            0
4       1       0       0       0       1       0      0      1         0         6            0
5       1       0       0       0       1       0      0      1         0         6            0
6       1       0       0       0       1       0      0      1         0         6            0
  weathersituation1 weathersituation2 weathersituation3 weathersituation4 temp humidity windspeed bikeCount
1                 0                 1                 0                 0 0.24     0.75    0.0896         1
2                 1                 0                 0                 0 0.38     0.76    0.2537        36
3                 1                 0                 0                 0 0.36     0.81    0.2836        56
4                 1                 0                 0                 0 0.42     0.77    0.2836        84
5                 0                 1                 0                 0 0.46     0.72    0.2985        94
6                 0                 1                 0                 0 0.46     0.72    0.2836       106
> 
```

```
> season1 = data$season1
> season2 = data$season2
> season3 = data$season3
> season4 = data$season4
> monthm1=data$monthm1
> monthm2=data$monthm2
> hourh1=data$hourh1
> hourh2=data$hourh2
> wsit1=data$weathersituation1
> wsit2=data$weathersituation2
> wsit3=data$weathersituation3
> wsit4=data$weathersituation4
> 
```

Check Correlation

```
> cor(cbind(bikecount,season1,season2,season3,season4,monthm1,monthm2,hourh1,hourh2,IsHoliday,Isweekday,
+           Isworkingday,wsit1,wsit2,wsit3,wsit4,temp,humidity,windspeed))
               bikecount     season1      season2      season3      season4       monthm1       monthm2        hourh1        hourh2     IsHoliday
bikecount    1.000000000 -0.312711332  0.057852736  0.201772945  0.048924477 -0.2361658338  0.2361658338 -0.4045386625  0.4045386625 -0.0347833904
season1     -0.312711332  1.000000000 -0.344220093 -0.336300666 -0.310570275  0.5203218881 -0.5203218881 -0.0069242351  0.0069242351  0.0336972114
season2      0.057852736 -0.344220093  1.000000000 -0.356685648 -0.329395600  0.4414350426 -0.4414350426  0.0111047497 -0.0111047497 -0.0273196515
season3      0.201772945 -0.336300666 -0.356685648  1.000000000 -0.321817238 -0.5075634872  0.5075634872  0.0068009680 -0.0068009680 -0.0223142872
season4      0.048924477 -0.310570275 -0.329395600 -0.321817238  1.000000000 -0.4687297635  0.4687297635 -0.0116937235  0.0116937235  0.0175312583
monthm1     -0.236165834  0.520321888  0.441435043 -0.507563487 -0.468729763  1.0000000000 -1.0000000000  0.0006481131 -0.0006481131 -0.0002219369
monthm2      0.236165834 -0.520321888 -0.441435043  0.507563487  0.468729763 -1.0000000000  1.0000000000 -0.0006481131  0.0006481131  0.0002219369
hourh1      -0.404538662 -0.006924235  0.011104750  0.006800968 -0.011693723  0.0006481131 -0.0006481131  1.0000000000 -1.0000000000 -0.0009530697
hourh2       0.404538662  0.006924235 -0.011104750 -0.006800968  0.011693723 -0.0006481131  0.0006481131 -1.0000000000  1.0000000000  0.0009530697
IsHoliday   -0.034783390  0.033697211 -0.027319652 -0.022314287  0.017531258 -0.0002219369  0.0002219369 -0.0009530697  0.0009530697  1.0000000000
Isweekday   -0.006607487  0.013872068 -0.010606783 -0.012178595  0.009660359 -0.0121454377  0.0121454377  0.0100182861 -0.0100182861 -0.0713559396
Isworkingday 0.019496321 -0.033570484  0.025007141  0.015194043 -0.007808435 -0.0057522845  0.0057522845 -0.0068453822  0.0068453822 -0.2460678961
wsit1        0.125146638 -0.016468049 -0.034873515  0.088223603 -0.038316053 -0.0750122867  0.0750122867 -0.0413694433  0.0413694433 -0.0330135272
wsit2       -0.040377772  0.026721852  0.027387009 -0.058216166  0.004449292  0.0662880082 -0.0662880082  0.0471757185 -0.0471757185  0.0441194036
wsit3       -0.143842286 -0.013943683  0.016249882 -0.056755304  0.056275617  0.0227628698 -0.0227628698 -0.0026504877  0.0026504877 -0.0120998772
wsit4       -0.009462538  0.020219089 -0.006959817 -0.006799693 -0.006279448  0.0133967342 -0.0133967342 -0.0110866369  0.0110866369 -0.0019259137
temp         0.460527820 -0.611299928  0.127130920  0.630994426 -0.165563758 -0.4829831048  0.4829831048 -0.1579455300  0.1579455300 -0.0353871437
humidity    -0.277636852 -0.195294456  0.064786768  0.016180666  0.114779265 -0.0872359871  0.0872359871  0.3025263547 -0.3025263547 -0.0238946071
windspeed    0.068797761  0.118899665  0.031399828 -0.091718101 -0.059216368  0.1574386488 -0.1574386488 -0.1386749104  0.1386749104  0.0063997074
                Isweekday Isworkingday        wsit1        wsit2        wsit3         wsit4        temp    humidity    windspeed
bikecount   -0.0066074869  0.019496321  0.125146638 -0.040377772 -0.143842286 -0.0094625377  0.46052782 -0.27763685  0.068797761
season1      0.0138720684 -0.033570484 -0.016468049  0.026721852 -0.013943683  0.0202190892 -0.61129993 -0.19529446  0.118899665
season2     -0.0106067834  0.025007141 -0.034873515  0.027387009  0.016249882 -0.0069598168  0.12713092  0.06478677  0.031399828
season3     -0.0121785953  0.015194043  0.088223603 -0.058216166 -0.056755304 -0.0067996931  0.63099443  0.01618067 -0.091718101
season4      0.0096603591 -0.007808435 -0.038316053  0.004449292  0.056275617 -0.0062794481 -0.16556376  0.11477927 -0.059216368
monthm1     -0.0121454377 -0.005752284 -0.075012287  0.066288008  0.022762870  0.0133967342 -0.48298310 -0.08723599  0.157438649
monthm2      0.0121454377  0.005752284  0.075012287 -0.066288008 -0.022762870 -0.0133967342  0.48298310  0.08723599 -0.157438649
hourh1       0.0100182861 -0.006845382 -0.041369443  0.047175719 -0.002650488 -0.0110866369 -0.15794553  0.30252635 -0.138674910
hourh2      -0.0100182861  0.006845382  0.041369443 -0.047175719  0.002650488  0.0110866369  0.15794553 -0.30252635  0.138674910
IsHoliday   -0.0713559396 -0.246067896 -0.033013527  0.044119404 -0.012099877 -0.0019259137 -0.03538714 -0.02389461  0.006399707
Isweekday    1.0000000000  0.011928004 -0.030482470  0.017540801  0.023572187 -0.0002022029 -0.04983499 -0.04544987  0.029671506
Isworkingday 0.0119280038  1.000000000 -0.049329255  0.016776436  0.055245658  0.0078267574  0.05443902  0.02000164  0.012287660
wsit1       -0.0304824704 -0.049329255  1.000000000 -0.799894603 -0.435625785 -0.0155487139  0.09610342 -0.39282855  0.009516063
wsit2        0.0175408012  0.016776436 -0.799894603  1.000000000 -0.191231260 -0.0068255834 -0.05665324  0.20558363 -0.063176026
wsit3        0.0235721868  0.055245658 -0.435625785 -0.191231260  1.000000000 -0.0037172399 -0.07164178  0.33369513  0.078757477
wsit4       -0.0002022029  0.007826757 -0.015548714 -0.006825583 -0.003717240  1.0000000000 -0.01578650  0.01742033  0.011855207
temp        -0.0498349852  0.054439025  0.096103416 -0.056653237 -0.071641777 -0.0157865047  1.00000000 -0.01978602 -0.033196397
humidity    -0.0454498672  0.020001637 -0.392828551  0.205583632  0.333695133  0.0174203305 -0.01978602  1.00000000 -0.260367736
windspeed    0.0296715058  0.012287660  0.009516063 -0.063176026  0.078757477  0.0118552074 -0.03319640 -0.26036774  1.000000000
> 
```

5.1.3 Transformation

I have tried various transformation on x and y variables .From the transformation given in screenshot ,I got the best correlation value. Even the correlation is still weak we will try to build the model based on new value after correlation and try to visualize the model.

```
> #Transformation
> 
> bikeCount1=log(bikeCount)
> humidity1 = humidity*humidity
> windspeed1 = 1/windspeed
> 
```

Correlation has improved to some extent after transformation as highlighted below

```
> cor(cbind(bikeCount1,season1,season2,season3,season4,monthm1,monthm2,hourh1,hourh2,IsHoliday,Isweekday,
+           Isworkingday,wsit1,wsit2,wsit3,wsit4,temp,humidity1,windspeed1))
              bikeCount1      season1      season2      season3      season4       monthm1       monthm2        hourh1        hourh2     IsHoliday     Isweekday
bikeCount1   1.000000000 -0.272332175  0.049019004  0.160808779  0.059574357 -0.2010472739  0.2010472739 -0.5055476858  0.5055476858 -0.0281480845  0.0095809642
season1     -0.272332175  1.000000000 -0.344220093 -0.336300666 -0.310570275  0.5203218881 -0.5203218881 -0.0069242351  0.0069242351  0.0336972114  0.0138720684
season2      0.049019004 -0.344220093  1.000000000 -0.356685648 -0.329395600  0.4414350426 -0.4414350426  0.0111047497 -0.0111047497 -0.0273196515 -0.0106067834
season3      0.160808779 -0.336300666 -0.356685648  1.000000000 -0.321817238 -0.5075634872  0.5075634872  0.0068009680 -0.0068009680 -0.0223142872 -0.0121785953
season4      0.059574357 -0.310570275 -0.329395600 -0.321817238  1.000000000 -0.4687297635  0.4687297635 -0.0116937235  0.0116937235  0.0175312583  0.0096603591
monthm1     -0.201047274  0.520321888  0.441435043 -0.507563487 -0.468729763  1.0000000000 -1.0000000000  0.0006481131 -0.0006481131 -0.0002219369 -0.0121454377
monthm2      0.201047274 -0.520321888 -0.441435043  0.507563487  0.468729763 -1.0000000000  1.0000000000 -0.0006481131  0.0006481131  0.0002219369  0.0121454377
hourh1      -0.505547686 -0.006924235  0.011104750  0.006800968 -0.011693723  0.0006481131 -0.0006481131  1.0000000000 -1.0000000000 -0.0009530697  0.0100182861
hourh2       0.505547686  0.006924235 -0.011104750 -0.006800968  0.011693723 -0.0006481131  0.0006481131 -1.0000000000  1.0000000000  0.0009530697 -0.0100182861
IsHoliday   -0.028148085  0.033697211 -0.027319652 -0.022314287  0.017531258 -0.0002219369  0.0002219369 -0.0009530697  0.0009530697  1.0000000000 -0.0713559396
Isweekday    0.009580964  0.013872068 -0.010606783 -0.012178595  0.009660359 -0.0121454377  0.0121454377  0.0100182861 -0.0100182861 -0.0713559396  1.0000000000
Isworkingday 0.009155695 -0.033570484  0.025007141  0.015194043 -0.007808435 -0.0057522845  0.0057522845 -0.0068453822  0.0068453822 -0.2460678961  0.0119280038
wsit1        0.091426891 -0.016468049 -0.034873515  0.088223603 -0.038316053 -0.0750122867  0.0750122867 -0.0413694433  0.0413694433 -0.0330135272 -0.0304824704
wsit2       -0.002120364  0.026721852  0.027387009 -0.058216166  0.004449292  0.0662880082 -0.0662880082  0.0471757185 -0.0471757185  0.0441194036  0.0175408012
wsit3       -0.146205345 -0.013943683  0.016249882 -0.056755304  0.056275617  0.0227628698 -0.0227628698 -0.0026504877  0.0026504877 -0.0120998772  0.0235721868
wsit4       -0.005931378  0.020219089 -0.006959817 -0.006799693 -0.006279448  0.0133967342 -0.0133967342 -0.0110866369  0.0110866369 -0.0019259137 -0.0002022029
temp         0.416820396 -0.611299928  0.127130920  0.630994426 -0.165563758 -0.4829831048  0.4829831048 -0.1579455300  0.1579455300 -0.0353871437 -0.0498349852
humidity1   -0.302950553 -0.177974207  0.071607490 -0.001618298  0.108410107 -0.0617418711  0.0617418711  0.2895647402 -0.2895647402 -0.0276348160 -0.0373802898
windspeed1  -0.117861708 -0.076786515 -0.049340813  0.053311935  0.075019054 -0.1351612243  0.1351612243  0.1314076761 -0.1314076761  0.0079157856 -0.0362060944
             Isworkingday        wsit1        wsit2        wsit3         wsit4         temp    humidity1   windspeed1
bikeCount1    0.009155695  0.091426891 -0.002120364 -0.146205345 -0.0059313780  0.416820396 -0.302950553 -0.117861708
season1      -0.033570484 -0.016468049  0.026721852 -0.013943683  0.0202190892 -0.611299928 -0.177974207 -0.076786515
season2       0.025007141 -0.034873515  0.027387009  0.016249882 -0.0069598168  0.127130920  0.071607490 -0.049340813
season3       0.015194043  0.088223603 -0.058216166 -0.056755304 -0.0067996931  0.630994426 -0.001618298  0.053311935
season4      -0.007808435 -0.038316053  0.004449292  0.056275617 -0.0062794481 -0.165563758  0.108410107  0.075019054
monthm1      -0.005752284 -0.075012287  0.066288008  0.022762870  0.0133967342 -0.482983105 -0.061741871 -0.135161224
monthm2       0.005752284  0.075012287 -0.066288008 -0.022762870 -0.0133967342  0.482983105  0.061741871  0.135161224
hourh1       -0.006845382 -0.041369443  0.047175719 -0.002650488 -0.0110866369 -0.157945530  0.289564740  0.131407676
hourh2        0.006845382  0.041369443 -0.047175719  0.002650488  0.0110866369  0.157945530 -0.289564740 -0.131407676
IsHoliday    -0.246067896 -0.033013527  0.044119404 -0.012099877 -0.0019259137 -0.035387144 -0.027634816  0.007915786
Isweekday     0.011928004 -0.030482470  0.017540801  0.023572187 -0.0002022029 -0.049834985 -0.037380290 -0.036206094
Isworkingday  1.000000000 -0.049329255  0.016776436  0.055245658  0.0078267574  0.054439025  0.027155653 -0.019047767
wsit1        -0.049329255  1.000000000 -0.799894603 -0.435625785 -0.0155487139  0.096103416 -0.417325560 -0.003360387
wsit2         0.016776436 -0.799894603  1.000000000 -0.191231260 -0.0068255834 -0.056653237  0.198771723  0.045427046
wsit3         0.055245658 -0.435625785 -0.191231260  1.000000000 -0.0037172399 -0.071641777  0.383913526 -0.062220263
wsit4         0.007826757 -0.015548714 -0.006825583 -0.003717240  1.0000000000 -0.015786505  0.019771070 -0.011299045
```

Final data set for building model

I will build models on data with and without pre-processing ,so have used two data-frames "myProjectdata" contains variable without pre-processing and "data" contains variables after pre-processing

Head(myProjectdata)

```
> myProjectdata[,"bikeCount1"]=bikeCount1
> myProjectdata[,"humidity1"]=humidity1
> myProjectdata[,"windspeed1"]=windspeed1
> head(myProjectdata)
  season month hour IsHoliday IsWeekday IsWorkingday weathersituation temp humidity windspeed bikeCount bikeCount1 humidity1 windspeed1
1      1     1    5         0         6            0                2 0.24     0.75    0.0896         1   0.000000    0.5625  11.160714
2      1     1   10         0         6            0                1 0.38     0.76    0.2537        36   3.583519    0.5776   3.941663
3      1     1   11         0         6            0                1 0.36     0.81    0.2836        56   4.025352    0.6561   3.526093
4      1     1   12         0         6            0                1 0.42     0.77    0.2836        84   4.430817    0.5929   3.526093
5      1     1   13         0         6            0                2 0.46     0.72    0.2985        94   4.543295    0.5184   3.350084
6      1     1   14         0         6            0                2 0.46     0.72    0.2836       106   4.663439    0.5184   3.526093
> |
```

head(data)

```
> data[,"bikeCount1"]=bikeCount1
> data[,"humidity1"]=humidity1
> data[,"windspeed1"]=windspeed1
> head(data)
  season1 season2 season3 season4 monthm1 monthm2 hourh1 hourh2 IsHoliday IsWeekday IsWorkingday weathersituation1 weathersituation2 weathersituation3
1       1       0       0       0       1       0      1      0         0         6            0                 0                 1                 0
2       1       0       0       0       1       0      1      0         0         6            0                 1                 0                 0
3       1       0       0       0       1       0      1      0         0         6            0                 1                 0                 0
4       1       0       0       0       1       0      0      1         0         6            0                 1                 0                 0
5       1       0       0       0       1       0      0      1         0         6            0                 0                 1                 0
6       1       0       0       0       1       0      0      1         0         6            0                 0                 1                 0
  weathersituation4 temp humidity windspeed bikeCount bikeCount1 humidity1 windspeed1
1                 0 0.24     0.75    0.0896         1   0.000000    0.5625  11.160714
2                 0 0.38     0.76    0.2537        36   3.583519    0.5776   3.941663
3                 0 0.36     0.81    0.2836        56   4.025352    0.6561   3.526093
4                 0 0.42     0.77    0.2836        84   4.430817    0.5929   3.526093
5                 0 0.46     0.72    0.2985        94   4.543295    0.5184   3.350084
6                 0 0.46     0.72    0.2836       106   4.663439    0.5184   3.526093
> |
```

Hypothesis Testing Results

Hypothesis testing is a statistical method that is used in making statistical decisions using experimental data. Hypothesis Testing is basically an assumption that we make about the population parameter. In my project I performed both **one sample and two sample hypothesis testing on the basis of two assumptions .**

a)The average count of bike rental per hour is 120

b)Average count of bike rented on working day and non-working day are same

I have used 95% confidence level to validate my assumption.

One sample hypothesis testing :

Null hypothesis , H0 : $\mu = 120$

Alternative hypothesis , Ha : $\mu \neq 120$

```
> z.test(bikeCount,NULL,alternative="two.sided",mu=145,sigma.x=sd(bikeCount),conf.level=0.95)

        One-sample z-Test

data:  bikeCount
z = 1.076, p-value = 0.2819
alternative hypothesis: true mean is not equal to 145
95 percent confidence interval:
 143.6283 149.7111
sample estimates:
mean of x
 146.6697
```

From the z-test results ,p-value is greater than 0.05. Hence, we can conclude that we have enough evidence to accept null hypothesis and reject alternative hypothesis. With 95% confidence level we can say that average count on bike rented per hour would be 120**.**

Two sample Hypothesis Testing :

Null Hypothesis (H0) : The total rented bike count on working and non-working day will be same ; $\mu_d = 0$

Alternative Hypothesis(Ha) : The total rented bike count on working and non-working same will not be same : $\mu_d \neq 0$

```
> z.test(diff,NULL,alternative="two.sided",mu=0,sigma.x=sd(diff),sigma.y=NULL,conf.level=0.95)

        One-sample z-Test

data:  diff
z = -6.7132, p-value = 1.904e-11
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -31.12035 -17.05520
sample estimates:
mean of x
-24.08778
```

From the Z-test result p-value is less than 0.05 hence, we can conclude that we have enough evidence to reject our null hypothesis and accept alternative hypothesis. Total rented bike will be different on working and non-working day based on 95% confidence level

Model creation using feature selection techniques:

My data size was small, so I have used 10-fold cross validation method to evaluate them model. In N-fold cross validation model is built considering the whole data set ,so splitting of data is not required in this case. I have used whole piece of data to build the models and then validated the model using F-test and residual analysis.

Evaluated the model based on N-fold cross validation using glm() to build the model cv.glm for evaluation. To evaluate the performance of the models, I have used AIC and MSE to compare the models. The model with smaller AIC and MSE value would be qualified as better model .

Models to be built

The models to be built using feature selection techniques are given below

1)Full model with only numerical variable

2)full model considering all the variables without data pre-processing

3)full model using as.factor for discrete variable directly when building the model

4)full model considering all the dummy variables after pre-processing

After building all the above model I will build other model using stepwise forward ,backward and both selections respectively based on the step() function on each of the full model one by one. Then will compare the best models found by stepwise selection, by using the AIC value as evaluation criteria and will decide the best model for further analysis.

In first 3 models I will use the data frame " myProjectdata" and in last model I will use data frame "data" which contains all the dummy variables after data pre-processing

Implementation : MODEL 1 (Full model with only numerical variable)

a) Full model with numerical variables only

```
> full = glm(bikeCount1~temp+humidity+humidity1+windspeed1,data=myProjectdata)
> summary(full)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1,
    data = myProjectdata)

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-4.7801  -0.5778   0.2721   0.8330   3.4365

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.299517   0.127308  25.918  < 2e-16 ***
temp         2.964369   0.071876  41.243  < 2e-16 ***
humidity     1.552993   0.409926   3.788 0.000153 ***
humidity1   -2.831805   0.323108  -8.764  < 2e-16 ***
windspeed1  -0.029602   0.005393  -5.488 4.19e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.525153)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance: 11489  on 7533  degrees of freedom
AIC: 24581

Number of Fisher Scoring iterations: 2

> |
```

AIC value for this model is 24581

b) Base model considering only one independent variable

```
> base=glm(bikeCount1~temp,data=myProjectdata)
> summary(base)

Call:
glm(formula = bikeCount1 ~ temp, data = myProjectdata)

Deviance Residuals:
    Min       1Q    Median       3Q      Max
-4.7130  -0.5991   0.3489   0.9099   2.6323

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.83678    0.04028   70.42   <2e-16 ***
temp         3.02608    0.07602   39.81   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.709305)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance: 12881  on 7536  degrees of freedom
AIC: 25437

Number of Fisher Scoring iterations: 2
```

AIC value for this model is 25437

Model creation using stepwise forward ,backward and both selections respectively based on the step() function:

c) Backward stepwise Model

```
> backward=step(full,direction="backward",trace=T,data=myProjectdata)
Start:  AIC=24580.67
bikeCount1 ~ temp + humidity + humidity1 + windspeed1

             Df Deviance   AIC
<none>             11489 24581
- humidity    1    11511 24593
- windspeed1  1    11535 24609
- humidity1   1    11606 24655
- temp        1    14083 26113
> summary(backward)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1, 
    data = myProjectdata)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.7801  -0.5778   0.2721   0.8330   3.4365  

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.299517   0.127308  25.918  < 2e-16 ***
temp         2.964369   0.071876  41.243  < 2e-16 ***
humidity     1.552993   0.409926   3.788 0.000153 ***
humidity1   -2.831805   0.323108  -8.764  < 2e-16 ***
windspeed1  -0.029602   0.005393  -5.488 4.19e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.525153)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance: 11489  on 7533  degrees of freedom
AIC: 24581

Number of Fisher Scoring iterations: 2
```

AIC value for this model is 24581

d) Forward stepwise model

```
> forward = step(base,scope=list(upper=full,lower=~1),direction= "forward",trace=F)
> summary(forward)

Call:
glm(formula = bikeCount1 ~ temp + humidity1 + windspeed1 + humidity, 
    data = myProjectdata)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.7801  -0.5778   0.2721   0.8330   3.4365  

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.299517   0.127308  25.918  < 2e-16 ***
temp         2.964369   0.071876  41.243  < 2e-16 ***
humidity1   -2.831805   0.323108  -8.764  < 2e-16 ***
windspeed1  -0.029602   0.005393  -5.488 4.19e-08 ***
humidity     1.552993   0.409926   3.788 0.000153 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.525153)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance: 11489  on 7533  degrees of freedom
AIC: 24581

Number of Fisher Scoring iterations: 2
```

AIC value for this model is 24581

e) Both Stepwise model

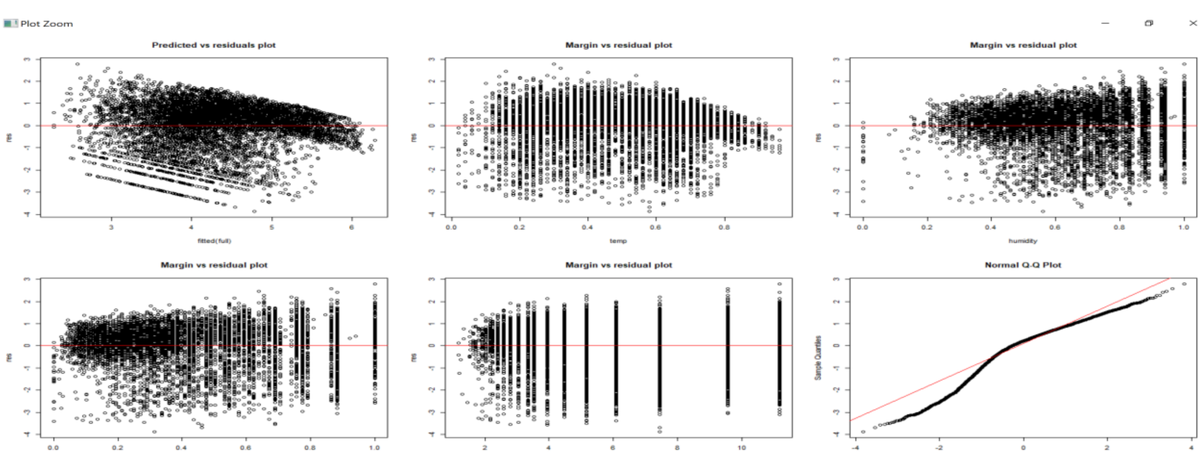
```
> both = step(base,scope=list(upper=full,lower=~1),direction= "both",trace=F)
> summary(both)

Call:
glm(formula = bikeCount1 ~ temp + humidity1 + windspeed1 + humidity,
    data = myProjectdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.7801  -0.5778   0.2721   0.8330   3.4365

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.299517   0.127308  25.918  < 2e-16 ***
temp         2.964369   0.071876  41.243  < 2e-16 ***
humidity1   -2.831805   0.323108  -8.764  < 2e-16 ***
windspeed1  -0.029602   0.005393  -5.488 4.19e-08 ***
humidity     1.552993   0.409926   3.788 0.000153 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.525153)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance: 11489  on 7533  degrees of freedom
AIC: 24581

Number of Fisher Scoring iterations: 2
```

AIC value for this model is 24581

Comparing all the models obtained from stepwise backward ,forward and both we could see that AIC values of all the models are same including the full model .Hence I have performed residual analysis on any one of the model and also check for multicollinearity and interaction term on any one of the model . Here I will perform all the operation on full model .

f) Residual analysis :

```
> par(mfrow=c(2,3))
> plot(fitted(full),res,main="Predicted vs residuals plot")
> abline(a=0,b=0,col='red')
> plot(temp, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(humidity, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(humidity1, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(windspeed1, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> qqnorm(res)
> qqline(res,col=2
+
```



If we consider the residual analysis, we see that the residual analysis fails because the model does not follow constant variance even though it follows normal distribution with few outliers.

g) Multicollinearity by VIF

```
> library(car)
> vif(full)
     temp   humidity  humidity1 windspeed1
 1.001929  31.657674  31.470550   1.068774
```

There is no multicollinearity problem in the variables of the model as the values are less than 4 and for humidity and humidity1 it is very large because humidity is factor if humidity1.By removing humidity I tried building the model ,AIC value for the model was higher as compared to previous model .So I will keep this model for further analysis .

h) Calculation of MSE value

```
> library(boot)
> mse_full = cv.glm(myProjectdata,full,K=10)$delta
> mse_full
[1] 1.525680 1.525599
>
```

I) Finding the influential points and then rebuild the model for better accuracy

```
> #remove influencial points
> cooksd <- cooks.distance(full)
> influential <- as.numeric(names(cooksd)[(cooksd > (4/7538))])
> data1<-myProjectdata
> data1<-data1[-influential,]
> dim(data1)
[1] 7128   14
>
```

As we can see after removing the influential points approx. 450 rows are removed which contained the influential points .I have created another data frame "data1" which contains the data after removing influential points .Created full model again and checked MSE

```
> mse_m1 = cv.glm(data1,m1,K=10)$delta
> mse_m1
[1] 1.091742 1.091640
>
```

After removing the influential points from the data set ,MSE has decreased as compared with the previous model .

Model m1 is the improvement of model "full" after removing influential points.

```
> m1 = glm(bikeCount1~temp+humidity+humidity1+windspeed1,data=data1)
> summary(m1)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1,
    data = data1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0137  -0.5302   0.1903   0.7234   2.6103

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.860472   0.122291  31.568  < 2e-16 ***
temp         2.641961   0.063045  41.906  < 2e-16 ***
humidity     0.625205   0.396201   1.578    0.115
humidity1   -2.147025   0.311677  -6.889 6.12e-12 ***
windspeed1  -0.026893   0.004749  -5.663 1.55e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.090569)

    Null deviance: 11082.7  on 7127  degrees of freedom
Residual deviance:  7768.1  on 7123  degrees of freedom
AIC: 20853

Number of Fisher Scoring iterations: 2
```

Thus, the equation of the model obtained as follows :

bikeCount1 = 3.860472 + 2.641961(temp) + .625205(humidity)-2.147(humidity1)-0.026893(windspeed1) +e

Implementation : MODEL 2 (full model considering all the variables without data pre-processing**)**

a)Full model with all variables without data pre-processing

```
> full1 = glm(bikeCount1~temp+humidity+humidity1+windspeed1+season+month+hour+IsWeekday+IsWorkingday+weathersituation+
+               IsHoliday,data=myProjectdata)
> summary(full1)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1 +
    season + month + hour + IsWeekday + IsWorkingday + weathersituation +
    IsHoliday, data = myProjectdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0462  -0.6142   0.1260   0.6716   3.1558

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.617662   0.121511  13.313  < 2e-16 ***
temp              2.225560   0.065523  33.966  < 2e-16 ***
humidity          2.017979   0.351471   5.742 9.75e-09 ***
humidity1        -2.776236   0.282573  -9.825  < 2e-16 ***
windspeed1       -0.020282   0.004562  -4.445 8.90e-06 ***
season            0.140455   0.019900   7.058 1.84e-12 ***
month             0.014011   0.006230   2.249  0.02455 *
hour              0.099297   0.001807  54.957  < 2e-16 ***
IsWeekday         0.014969   0.005975   2.505  0.01226 *
IsWorkingday     -0.041851   0.026425  -1.584  0.11328
weathersituation  0.038120   0.021318   1.788  0.07379 .
IsHoliday        -0.243914   0.075731  -3.221  0.00128 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.059664)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance:  7975  on 7526  degrees of freedom
AIC: 21843

Number of Fisher Scoring iterations: 2
```

AIC value for this model is 21843

Model creation using stepwise forward ,backward and both selections respectively based on the step() function using the model created above.

b) Stepwise Backward model :

```
> backward1=step(full1,direction="backward",trace=T)
Start:  AIC=21842.75
bikeCount1 ~ temp + humidity + humidity1 + windspeed1 + season + 
    month + hour + IsWeekday + IsWorkingday + weathersituation + 
    IsHoliday

                   Df Deviance   AIC
<none>                  7975.0 21843
- IsWorkingday      1   7977.7 21843
- weathersituation  1   7978.4 21844
- month             1   7980.4 21846
- IsWeekday         1   7981.7 21847
- IsHoliday         1   7986.0 21851
- windspeed1        1   7996.0 21861
- humidity          1   8010.0 21874
- season            1   8027.8 21891
- humidity1         1   8077.3 21937
- temp              1   9197.6 22916
- hour              1  11175.5 24384
> summary(backward1)

Call:
```

```
> summary(backward1)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1 + 
    season + month + hour + IsWeekday + IsWorkingday + weathersituation + 
    IsHoliday, data = myProjectdata)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.0462  -0.6142   0.1260   0.6716   3.1558  

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)       1.617662   0.121511  13.313  < 2e-16 ***
temp              2.225560   0.065523  33.966  < 2e-16 ***
humidity          2.017979   0.351471   5.742 9.75e-09 ***
humidity1        -2.776236   0.282573  -9.825  < 2e-16 ***
windspeed1       -0.020282   0.004562  -4.445 8.90e-06 ***
season            0.140455   0.019900   7.058 1.84e-12 ***
month             0.014011   0.006230   2.249  0.02455 *  
hour              0.099297   0.001807  54.957  < 2e-16 ***
IsWeekday         0.014969   0.005975   2.505  0.01226 *  
IsWorkingday     -0.041851   0.026425  -1.584  0.11328    
weathersituation  0.038120   0.021318   1.788  0.07379 .  
IsHoliday        -0.243914   0.075731  -3.221  0.00128 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.059664)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance:  7975  on 7526  degrees of freedom
AIC: 21843

Number of Fisher Scoring iterations: 2

> |
```

AIC value for this model is 21843

c) Stepwise Forward model :

```
> full1 = glm(bikeCount1~temp+humidity+humidity1+windspeed1+season+month+hour+Isweekday+Isworkingday+weathersituation+
+             IsHoliday,data=myProjectdata)
> summary(full1)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1 +
    season + month + hour + Isweekday + Isworkingday + weathersituation +
    IsHoliday, data = myProjectdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0462  -0.6142   0.1260   0.6716   3.1558

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.617662   0.121511  13.313  < 2e-16 ***
temp              2.225560   0.065523  33.966  < 2e-16 ***
humidity          2.017979   0.351471   5.742 9.75e-09 ***
humidity1        -2.776236   0.282573  -9.825  < 2e-16 ***
windspeed1       -0.020282   0.004562  -4.445 8.90e-06 ***
season            0.140455   0.019900   7.058 1.84e-12 ***
month             0.014011   0.006230   2.249  0.02455 *
hour              0.099297   0.001807  54.957  < 2e-16 ***
Isweekday         0.014969   0.005975   2.505  0.01226 *
Isworkingday     -0.041851   0.026425  -1.584  0.11328
weathersituation  0.038120   0.021318   1.788  0.07379 .
IsHoliday        -0.243914   0.075731  -3.221  0.00128 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.059664)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance:  7975  on 7526  degrees of freedom
AIC: 21843

Number of Fisher Scoring iterations: 2
```

AIC value for this model is 21843

d) Stepwise both model :

```
> both1 = step(base,scope=list(upper=full1,lower=~1),direction= "both",trace=F)
> summary(both1)

Call:
glm(formula = bikeCount1 ~ temp + hour + humidity1 + season +
    humidity + windspeed1 + Isweekday + IsHoliday + month + weathersituation +
    Isworkingday, data = myProjectdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0462  -0.6142   0.1260   0.6716   3.1558

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.617662   0.121511  13.313  < 2e-16 ***
temp              2.225560   0.065523  33.966  < 2e-16 ***
hour              0.099297   0.001807  54.957  < 2e-16 ***
humidity1        -2.776236   0.282573  -9.825  < 2e-16 ***
season            0.140455   0.019900   7.058 1.84e-12 ***
humidity          2.017979   0.351471   5.742 9.75e-09 ***
windspeed1       -0.020282   0.004562  -4.445 8.90e-06 ***
Isweekday         0.014969   0.005975   2.505  0.01226 *
IsHoliday        -0.243914   0.075731  -3.221  0.00128 **
month             0.014011   0.006230   2.249  0.02455 *
weathersituation  0.038120   0.021318   1.788  0.07379 .
Isworkingday     -0.041851   0.026425  -1.584  0.11328
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.059664)

    Null deviance: 15590  on 7537  degrees of freedom
Residual deviance:  7975  on 7526  degrees of freedom
AIC: 21843

Number of Fisher Scoring iterations: 2
```

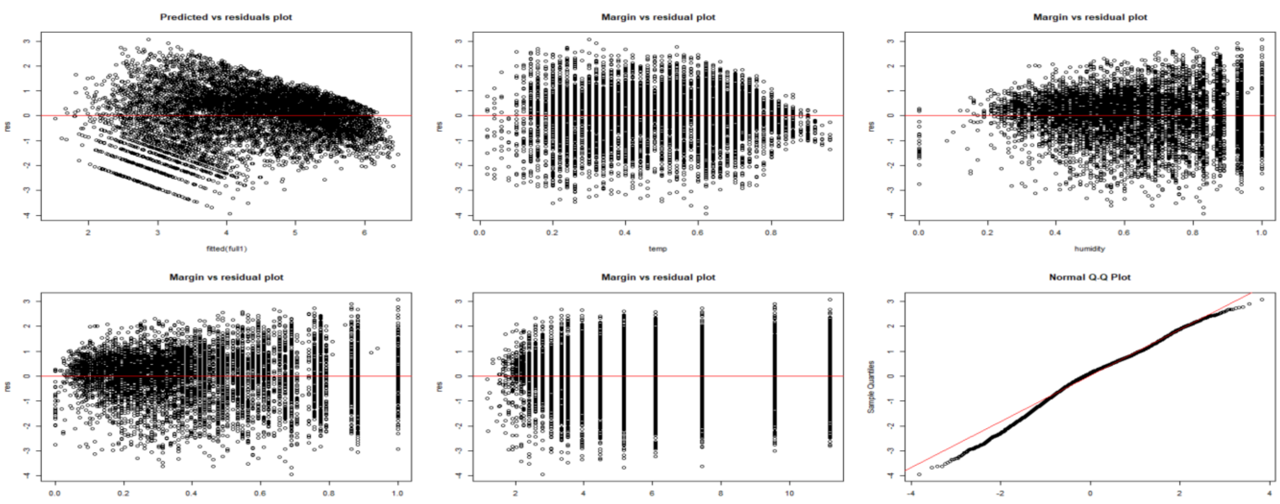
AIC value for this model is 21843

As we could see the AIC value of all the models are same using stepwise regression .So we will consider any one of the model for further analysis . Now performed residual analysis on any one of the model

and have also checked for multicollinearity and influential points . Here I have performed all the operation on full1 model .

e) Residual Analysis :

```
> par(mfrow=c(2,3))
> plot(fitted(full1),res,main="Predicted vs residuals plot")
> abline(a=0,b=0,col='red')
> plot(temp, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(humidity, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(humidity1, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(windspeed1, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> qqnorm(res)
> qqline(res,col=2)
>
```



If we consider the residual analysis, we see that the residual analysis has not passed because the model does not follow constant variance even though it follows normal distribution with few outliers.

f) Multicollinearity using VIF

```
> library(car)
> vif(full1)
        temp      humidity     humidity1    windspeed1        season         month          hour
    1.198402     33.495954     34.643054      1.100723      3.373327      3.200938      1.086753
   IsWeekday  IsWorkingday weathersituation  IsHoliday
    1.018544      1.073370      1.412161      1.079346
>
```

There is no multicollinearity problem in the variables of the model as the values are less than 4 and for humidity and humidity1 it is very large because humidity is factor of humidity1.I have tried removing humidity1 from the model, but this has resulted into higher MSE and AIC value. So, I have decided to keep humidity1

g) Calculation of MSE value

```
> library(boot)
> mse_full1 = cv.glm(myProjectdata,full,K=10)$delta
> mse_full1
[1] 1.525674 1.525593
>
```

h) Finding the influential points and then rebuild the model for better accuracy

```
> #remove influencial points
> cooksd <- cooks.distance(full1)
> influential <- as.numeric(names(cooksd)[(cooksd > (4/7538))])
> data2<-myProjectdata
> data2<-data2[-influential,]
> dim(data2)
[1] 7154   14
```

As we can see after removing the influential points 384 rows were removed which contained the influential points .I have created another data frame "data2" which contains the data after removing influential points .Created full model again and checked MSE

```
> mse_m2 = cv.glm(data2,m2,K=10)$delta
> mse_m2
[1] 0.8178342 0.8177027
>
```

After removing the influential points from the data set ,MSE has decreased significantly as compared to previous model. So, I will use this model for final analysis .**Model m2 is the final model after removing influential points.**

```
> m2 = glm(bikeCount1~temp+humidity+humidity1+windspeed1+season+month+hour+IsWeekday+IsWorkingday+weathersituation+
+                    IsHoliday,data=data2)
> summary(m2)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1 + 
    season + month + hour + IsWeekday + IsWorkingday + weathersituation + 
    IsHoliday, data = data2)

Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-3.11323  -0.54621   0.08924   0.58814   2.68427  

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)       2.044080   0.115454  17.705  < 2e-16 ***
temp              2.102067   0.059119  35.557  < 2e-16 ***
humidity          1.758940   0.340377   5.168 2.44e-07 ***
humidity1        -2.793598   0.271441 -10.292  < 2e-16 ***
windspeed1       -0.023451   0.004151  -5.649 1.67e-08 ***
season            0.087111   0.019319   4.509 6.61e-06 ***
month             0.026745   0.006107   4.379 1.21e-05 ***
hour              0.090611   0.001643  55.165  < 2e-16 ***
IsWeekday         0.002899   0.005383   0.539   0.5901    
IsWorkingday     -0.106692   0.023767  -4.489 7.26e-06 ***
weathersituation  0.124594   0.019509   6.386 1.80e-10 ***
IsHoliday        -0.156554   0.077470  -2.021   0.0433 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.8167056)

    Null deviance: 12193.9  on 7153  degrees of freedom
Residual deviance:  5832.9  on 7142  degrees of freedom
AIC: 18868

Number of Fisher Scoring iterations: 2
```

AIC value of the model is 18868

Implementation : MODEL 3(model using as.factor to create dummy variable for discrete variable directly when building the model)

a)Full model using as.factor for discrete variable when building the model

```
> full2 = glm(bikeCount1~as.factor(season)+as.factor(month)+as.factor(hour)+as.factor(IsHoliday)+as.factor(IsWeekday)
+           +temp+humidity+humidity1+windspeed1,data=myProjectdata)
> summary(full2)

Call:
glm(formula = bikeCount1 ~ as.factor(season) + as.factor(month) +
    as.factor(hour) + as.factor(IsHoliday) + as.factor(IsWeekday) +
    temp + humidity + humidity1 + windspeed1, data = myProjectdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.7020  -0.2992   0.0387   0.3838   2.0672

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.659218   0.086365  19.212  < 2e-16 ***
as.factor(season)2     0.261809   0.044012   5.949 2.83e-09 ***
as.factor(season)3     0.378309   0.053513   7.069 1.70e-12 ***
as.factor(season)4     0.543918   0.045941  11.840  < 2e-16 ***
as.factor(month)2      0.246208   0.037445   6.575 5.19e-11 ***
as.factor(month)3      0.231076   0.040533   5.701 1.24e-08 ***
as.factor(month)4      0.351008   0.062839   5.586 2.41e-08 ***
as.factor(month)5      0.635345   0.067814   9.369  < 2e-16 ***
as.factor(month)6      0.421857   0.072307   5.834 5.63e-09 ***
as.factor(month)7      0.223692   0.082370   2.716 0.006629 **
as.factor(month)8      0.252603   0.079453   3.179 0.001482 **
as.factor(month)9      0.406800   0.071449   5.694 1.29e-08 ***
as.factor(month)10     0.318558   0.063607   5.008 5.62e-07 ***
as.factor(month)11     0.246734   0.061671   4.001 6.37e-05 ***
as.factor(month)12     0.273722   0.047611   5.749 9.32e-09 ***
as.factor(hour)1      -0.620278   0.050762 -12.219  < 2e-16 ***
as.factor(hour)2      -1.057839   0.051465 -20.555  < 2e-16 ***
as.factor(hour)3      -1.675421   0.051779 -32.357  < 2e-16 ***
as.factor(hour)4      -1.991078   0.051436 -38.710  < 2e-16 ***
as.factor(hour)5      -1.039364   0.051015 -20.374  < 2e-16 ***
as.factor(hour)6       0.193088   0.050779   3.802 0.000144 ***
as.factor(hour)7       1.157202   0.050473  22.927  < 2e-16 ***
as.factor(hour)8       1.776972   0.050347  35.295  < 2e-16 ***
as.factor(hour)9       1.478148   0.049900  29.622  < 2e-16 ***
as.factor(hour)10      1.136071   0.049847  22.791  < 2e-16 ***
```

```
as.factor(hour)11       1.239111   0.050125  24.721  < 2e-16 ***
as.factor(hour)12       1.405574   0.050586  27.786  < 2e-16 ***
as.factor(hour)13       1.375349   0.050831  27.057  < 2e-16 ***
as.factor(hour)14       1.323836   0.050927  25.995  < 2e-16 ***
as.factor(hour)15       1.363078   0.051192  26.627  < 2e-16 ***
as.factor(hour)16       1.630651   0.050796  32.102  < 2e-16 ***
as.factor(hour)17       2.041355   0.050442  40.470  < 2e-16 ***
as.factor(hour)18       1.962548   0.049714  39.476  < 2e-16 ***
as.factor(hour)19       1.679720   0.049770  33.749  < 2e-16 ***
as.factor(hour)20       1.395405   0.049838  27.999  < 2e-16 ***
as.factor(hour)21       1.179889   0.050170  23.518  < 2e-16 ***
as.factor(hour)22       0.944192   0.050556  18.676  < 2e-16 ***
as.factor(hour)23       0.543948   0.049881  10.905  < 2e-16 ***
as.factor(IsHoliday)1  -0.187152   0.046527  -4.022 5.82e-05 ***
as.factor(IsWeekday)1  -0.032058   0.027910  -1.149 0.250750
as.factor(IsWeekday)2  -0.041627   0.027114  -1.535 0.124751
as.factor(IsWeekday)3  -0.046723   0.027223  -1.716 0.086148 .
as.factor(IsWeekday)4  -0.001670   0.026935  -0.062 0.950552
as.factor(IsWeekday)5   0.091131   0.027155   3.356 0.000795 ***
as.factor(IsWeekday)6   0.076755   0.026804   2.864 0.004200 **
temp                    1.436774   0.088691  16.200  < 2e-16 ***
humidity                2.636280   0.217513  12.120  < 2e-16 ***
humidity1              -2.720080   0.169719 -16.027  < 2e-16 ***
windspeed1              0.017762   0.002812   6.315 2.85e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.3875573)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  2902.4  on 7489  degrees of freedom
AIC: 14298

Number of Fisher Scoring iterations: 2

> |
```

Here AIC value of model is 14298

Model creation using stepwise forward ,backward and both selections respectively based on the step() function:

b) Stepwise Backward model :

```
> backward2=step(full2,direction="backward",trace=T)
Start:  AIC=14297.55
bikeCount1 ~ as.factor(season) + as.factor(month) + as.factor(hour) +
    as.factor(IsHoliday) + as.factor(Isweekday) + temp + humidity +
    humidity1 + windspeed1

                       Df Deviance   AIC
<none>                      2902.4 14298
- as.factor(IsHoliday)  1   2908.7 14312
- windspeed1            1   2917.9 14336
- as.factor(Isweekday)  6   2921.9 14336
- as.factor(season)     3   2959.3 14438
- humidity              1   2959.3 14442
- as.factor(month)     11   2974.2 14460
- humidity1             1   3002.0 14550
- temp                  1   3004.1 14555
- as.factor(hour)      23  10570.4 23995
> summary(backward2)
```

```
> summary(backward2)

Call:
glm(formula = bikeCount1 ~ as.factor(season) + as.factor(month) +
    as.factor(hour) + as.factor(IsHoliday) + as.factor(Isweekday) +
    temp + humidity + humidity1 + windspeed1, data = myProjectdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.7020  -0.2992   0.0387   0.3838   2.0672

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.659218   0.086365  19.212  < 2e-16 ***
as.factor(season)2    0.261809   0.044012   5.949 2.83e-09 ***
as.factor(season)3    0.378309   0.053513   7.069 1.70e-12 ***
as.factor(season)4    0.543918   0.045941  11.840  < 2e-16 ***
as.factor(month)2     0.246208   0.037445   6.575 5.19e-11 ***
as.factor(month)3     0.231076   0.040533   5.701 1.24e-08 ***
as.factor(month)4     0.351008   0.062839   5.586 2.41e-08 ***
as.factor(month)5     0.635345   0.067814   9.369  < 2e-16 ***
as.factor(month)6     0.421857   0.072307   5.834 5.63e-09 ***
as.factor(month)7     0.223692   0.082370   2.716 0.006629 **
as.factor(month)8     0.252603   0.079453   3.179 0.001482 **
as.factor(month)9     0.406800   0.071449   5.694 1.29e-08 ***
as.factor(month)10    0.318558   0.063607   5.008 5.62e-07 ***
as.factor(month)11    0.246734   0.061671   4.001 6.37e-05 ***
as.factor(month)12    0.273722   0.047611   5.749 9.32e-09 ***
as.factor(hour)1     -0.620278   0.050762 -12.219  < 2e-16 ***
as.factor(hour)2     -1.057839   0.051465 -20.555  < 2e-16 ***
as.factor(hour)3     -1.675421   0.051779 -32.357  < 2e-16 ***
as.factor(hour)4     -1.991078   0.051436 -38.710  < 2e-16 ***
as.factor(hour)5     -1.039364   0.051015 -20.374  < 2e-16 ***
as.factor(hour)6      0.193088   0.050779   3.802 0.000144 ***
as.factor(hour)7      1.157202   0.050473  22.927  < 2e-16 ***
as.factor(hour)8      1.776972   0.050347  35.295  < 2e-16 ***
as.factor(hour)9      1.478148   0.049900  29.622  < 2e-16 ***
```

```
as.factor(hour)10         1.136071   0.049847   22.791   < 2e-16 ***
as.factor(hour)11         1.239111   0.050125   24.721   < 2e-16 ***
as.factor(hour)12         1.405574   0.050586   27.786   < 2e-16 ***
as.factor(hour)13         1.375349   0.050831   27.057   < 2e-16 ***
as.factor(hour)14         1.323836   0.050927   25.995   < 2e-16 ***
as.factor(hour)15         1.363078   0.051192   26.627   < 2e-16 ***
as.factor(hour)16         1.630651   0.050796   32.102   < 2e-16 ***
as.factor(hour)17         2.041355   0.050442   40.470   < 2e-16 ***
as.factor(hour)18         1.962548   0.049714   39.476   < 2e-16 ***
as.factor(hour)19         1.679720   0.049770   33.749   < 2e-16 ***
as.factor(hour)20         1.395405   0.049838   27.999   < 2e-16 ***
as.factor(hour)21         1.179889   0.050170   23.518   < 2e-16 ***
as.factor(hour)22         0.944192   0.050556   18.676   < 2e-16 ***
as.factor(hour)23         0.543948   0.049881   10.905   < 2e-16 ***
as.factor(IsHoliday)1    -0.187152   0.046527   -4.022  5.82e-05 ***
as.factor(Isweekday)1    -0.032058   0.027910   -1.149  0.250750
as.factor(Isweekday)2    -0.041627   0.027114   -1.535  0.124751
as.factor(Isweekday)3    -0.046723   0.027223   -1.716  0.086148 .
as.factor(Isweekday)4    -0.001670   0.026935   -0.062  0.950552
as.factor(Isweekday)5     0.091131   0.027155    3.356  0.000795 ***
as.factor(Isweekday)6     0.076755   0.026804    2.864  0.004200 **
temp                      1.436774   0.088691   16.200   < 2e-16 ***
humidity                  2.636280   0.217513   12.120   < 2e-16 ***
humidity1                -2.720080   0.169719  -16.027   < 2e-16 ***
windspeed1                0.017762   0.002812    6.315  2.85e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.3875573)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  2902.4  on 7489  degrees of freedom
AIC: 14298

Number of Fisher Scoring iterations: 2
```

AIC value for this model is 14298

c) Stepwise forward model

```
> forward2 = step(base,scope=list(upper=full2,lower=~1),direction= "forward",trace=F)
> summary(forward2)

Call:
glm(formula = bikeCount1 ~ temp + as.factor(hour) + as.factor(month) + 
    humidity1 + humidity + as.factor(season) + as.factor(Isweekday) + 
    windspeed1 + as.factor(IsHoliday), data = myProjectdata)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.7020  -0.2992   0.0387   0.3838   2.0672  

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)            1.659218   0.086365  19.212  < 2e-16 ***
temp                   1.436774   0.088691  16.200  < 2e-16 ***
as.factor(hour)1      -0.620278   0.050762 -12.219  < 2e-16 ***
as.factor(hour)2      -1.057839   0.051465 -20.555  < 2e-16 ***
as.factor(hour)3      -1.675421   0.051779 -32.357  < 2e-16 ***
as.factor(hour)4      -1.991078   0.051436 -38.710  < 2e-16 ***
as.factor(hour)5      -1.039364   0.051015 -20.374  < 2e-16 ***
as.factor(hour)6       0.193088   0.050779   3.802 0.000144 ***
as.factor(hour)7       1.157202   0.050473  22.927  < 2e-16 ***
as.factor(hour)8       1.776972   0.050347  35.295  < 2e-16 ***
as.factor(hour)9       1.478148   0.049900  29.622  < 2e-16 ***
as.factor(hour)10      1.136071   0.049847  22.791  < 2e-16 ***
as.factor(hour)11      1.239111   0.050125  24.721  < 2e-16 ***
as.factor(hour)12      1.405574   0.050586  27.786  < 2e-16 ***
as.factor(hour)13      1.375349   0.050831  27.057  < 2e-16 ***
as.factor(hour)14      1.323836   0.050927  25.995  < 2e-16 ***
as.factor(hour)15      1.363078   0.051192  26.627  < 2e-16 ***
as.factor(hour)16      1.630651   0.050796  32.102  < 2e-16 ***
as.factor(hour)17      2.041355   0.050442  40.470  < 2e-16 ***
as.factor(hour)18      1.962548   0.049714  39.476  < 2e-16 ***
as.factor(hour)19      1.679720   0.049770  33.749  < 2e-16 ***
as.factor(hour)20      1.395405   0.049838  27.999  < 2e-16 ***
as.factor(hour)21      1.179889   0.050170  23.518  < 2e-16 ***
as.factor(hour)22      0.944192   0.050556  18.676  < 2e-16 ***
as.factor(hour)23      0.543948   0.049881  10.905  < 2e-16 ***
as.factor(month)2      0.246208   0.037445   6.575 5.19e-11 ***
as.factor(month)3      0.231076   0.040533   5.701 1.24e-08 ***
```

```
as.factor(month)4      0.351008   0.062839   5.586 2.41e-08 ***
as.factor(month)5      0.635345   0.067814   9.369  < 2e-16 ***
as.factor(month)6      0.421857   0.072307   5.834 5.63e-09 ***
as.factor(month)7      0.223692   0.082370   2.716 0.006629 ** 
as.factor(month)8      0.252603   0.079453   3.179 0.001482 ** 
as.factor(month)9      0.406800   0.071449   5.694 1.29e-08 ***
as.factor(month)10     0.318558   0.063607   5.008 5.62e-07 ***
as.factor(month)11     0.246734   0.061671   4.001 6.37e-05 ***
as.factor(month)12     0.273722   0.047611   5.749 9.32e-09 ***
humidity1             -2.720080   0.169719 -16.027  < 2e-16 ***
humidity               2.636280   0.217513  12.120  < 2e-16 ***
as.factor(season)2     0.261809   0.044012   5.949 2.83e-09 ***
as.factor(season)3     0.378309   0.053513   7.069 1.70e-12 ***
as.factor(season)4     0.543918   0.045941  11.840  < 2e-16 ***
as.factor(Isweekday)1 -0.032058   0.027910  -1.149 0.250750    
as.factor(Isweekday)2 -0.041627   0.027114  -1.535 0.124751    
as.factor(Isweekday)3 -0.046723   0.027223  -1.716 0.086148 .  
as.factor(Isweekday)4 -0.001670   0.026935  -0.062 0.950552    
as.factor(Isweekday)5  0.091131   0.027155   3.356 0.000795 ***
as.factor(Isweekday)6  0.076755   0.026804   2.864 0.004200 ** 
windspeed1             0.017762   0.002812   6.315 2.85e-10 ***
as.factor(IsHoliday)1 -0.187152   0.046527  -4.022 5.82e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.3875573)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  2902.4  on 7489  degrees of freedom
AIC: 14298

Number of Fisher Scoring iterations: 2

> |
```

The AIC value for this model is 14298

d) Both Stepwise:

```
> both2 = step(base,scope=list(upper=full2,lower=~1),direction= "forward",trace=F)
> summary(both2)

Call:
glm(formula = bikeCount1 ~ temp + as.factor(hour) + as.factor(month) + 
    humidity1 + humidity + as.factor(season) + as.factor(Isweekday) + 
    windspeed1 + as.factor(IsHoliday), data = myProjectdata)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.7020  -0.2992   0.0387   0.3838   2.0672  

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)            1.659218   0.086365  19.212  < 2e-16 ***
temp                   1.436774   0.088691  16.200  < 2e-16 ***
as.factor(hour)1      -0.620278   0.050762 -12.219  < 2e-16 ***
as.factor(hour)2      -1.057839   0.051465 -20.555  < 2e-16 ***
as.factor(hour)3      -1.675421   0.051779 -32.357  < 2e-16 ***
as.factor(hour)4      -1.991078   0.051436 -38.710  < 2e-16 ***
as.factor(hour)5      -1.039364   0.051015 -20.374  < 2e-16 ***
as.factor(hour)6       0.193088   0.050779   3.802 0.000144 ***
as.factor(hour)7       1.157202   0.050473  22.927  < 2e-16 ***
as.factor(hour)8       1.776972   0.050347  35.295  < 2e-16 ***
as.factor(hour)9       1.478148   0.049900  29.622  < 2e-16 ***
as.factor(hour)10      1.136071   0.049847  22.791  < 2e-16 ***
as.factor(hour)11      1.239111   0.050125  24.721  < 2e-16 ***
as.factor(hour)12      1.405574   0.050586  27.786  < 2e-16 ***
as.factor(hour)13      1.375349   0.050831  27.057  < 2e-16 ***
as.factor(hour)14      1.323836   0.050927  25.995  < 2e-16 ***
as.factor(hour)15      1.363078   0.051192  26.627  < 2e-16 ***
as.factor(hour)16      1.630651   0.050796  32.102  < 2e-16 ***
as.factor(hour)17      2.041355   0.050442  40.470  < 2e-16 ***
as.factor(hour)18      1.962548   0.049714  39.476  < 2e-16 ***
as.factor(hour)19      1.679720   0.049770  33.749  < 2e-16 ***
as.factor(hour)20      1.395405   0.049838  27.999  < 2e-16 ***
as.factor(hour)21      1.179889   0.050170  23.518  < 2e-16 ***
as.factor(hour)22      0.944192   0.050556  18.676  < 2e-16 ***
as.factor(hour)23      0.543948   0.049881  10.905  < 2e-16 ***
```

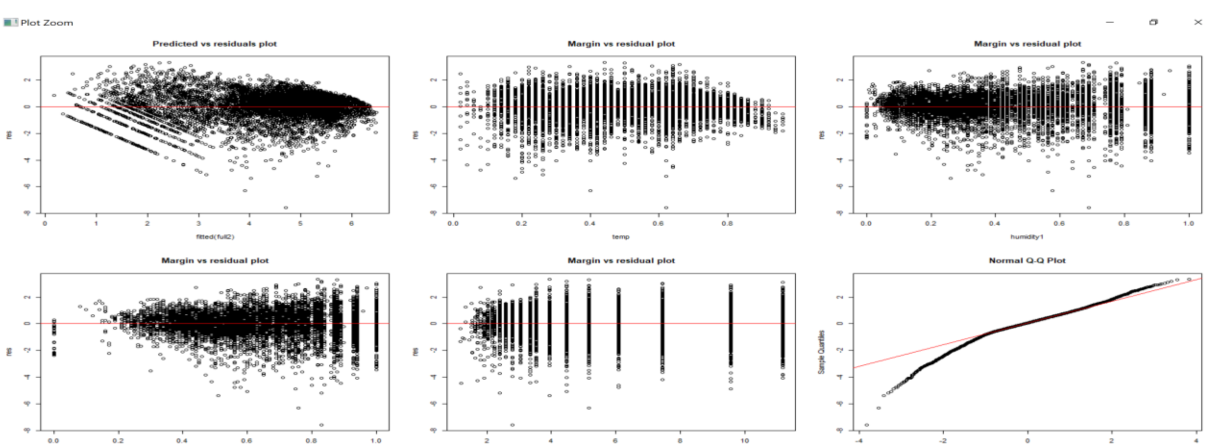
```
as.factor(month)2          0.246208   0.037445   6.575 5.19e-11 ***
as.factor(month)3          0.231076   0.040533   5.701 1.24e-08 ***
as.factor(month)4          0.351008   0.062839   5.586 2.41e-08 ***
as.factor(month)5          0.635345   0.067814   9.369  < 2e-16 ***
as.factor(month)6          0.421857   0.072307   5.834 5.63e-09 ***
as.factor(month)7          0.223692   0.082370   2.716 0.006629 **
as.factor(month)8          0.252603   0.079453   3.179 0.001482 **
as.factor(month)9          0.406800   0.071449   5.694 1.29e-08 ***
as.factor(month)10         0.318558   0.063607   5.008 5.62e-07 ***
as.factor(month)11         0.246734   0.061671   4.001 6.37e-05 ***
as.factor(month)12         0.273722   0.047611   5.749 9.32e-09 ***
humidity1                 -2.720080   0.169719 -16.027  < 2e-16 ***
humidity                   2.636280   0.217513  12.120  < 2e-16 ***
as.factor(season)2         0.261809   0.044012   5.949 2.83e-09 ***
as.factor(season)3         0.378309   0.053513   7.069 1.70e-12 ***
as.factor(season)4         0.543918   0.045941  11.840  < 2e-16 ***
as.factor(Isweekday)1 -0.032058   0.027910  -1.149 0.250750
as.factor(Isweekday)2 -0.041627   0.027114  -1.535 0.124751
as.factor(Isweekday)3 -0.046723   0.027223  -1.716 0.086148 .
as.factor(Isweekday)4 -0.001670   0.026935  -0.062 0.950552
as.factor(Isweekday)5  0.091131   0.027155   3.356 0.000795 ***
as.factor(Isweekday)6  0.076755   0.026804   2.864 0.004200 **
windspeed1             0.017762   0.002812   6.315 2.85e-10 ***
as.factor(IsHoliday)1 -0.187152   0.046527  -4.022 5.82e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.3875573)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  2902.4  on 7489  degrees of freedom
AIC: 14298

Number of Fisher Scoring iterations: 2

>
```

AIC value for this model is 14298

As we could see the AIC value of all the models are also same using stepwise regression .So we will consider any one of the model for further analysis . Now we will perform residual analysis on any one of the model and also check for multicollinearity and influential points **. Here I have performed all the operation on full2 model .**

e) Residual Analysis:

```
> par(mfrow=c(2,3))
> plot(fitted(full2),res,main="Predicted vs residuals plot")
> abline(a=0,b=0,col='red')
> plot(temp, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(humidity1, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(humidity, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(windspeed1, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> qqnorm(res)
> qqline(res,col=2)
>
```



If we consider the residual analysis, we see that the residual analysis passed because the model follow constant variance with few deviations and it also follows normal distribution with few outliers.

f) Multicollinearity by VIF

```
> vif(full2)
                              GVIF Df GVIF^(1/(2*Df))
as.factor(season)       166.788276  3        2.346178
as.factor(month)        398.819631 11        1.312856
as.factor(hour)           1.591118 23        1.010148
as.factor(IsHoliday)      1.113918  1        1.055423
as.factor(Isweekday)      1.175410  6        1.013559
temp                      6.003511  1        2.450206
humidity                 35.076467  1        5.922539
humidity1                34.170248  1        5.845532
windspeed1                1.143657  1        1.069419
>
```

Multicollinearity between season and month are very high. So, have removed month from the given model "full2" and have also removed "humidity1"

```
> vif(full2)
                          GVIF Df GVIF^(1/(2*Df))
as.factor(season)     2.996312  3        1.200691
as.factor(hour)       1.404333 23        1.007409
as.factor(IsHoliday)  1.085342  1        1.041798
as.factor(Isweekday)  1.117054  6        1.009267
temp                  2.964981  1        1.721912
humidity              1.332881  1        1.154504
windspeed1            1.112251  1        1.054633
>
```

There is no multicollinearity issue now.

g) Check MSE :

```
> mse_full2 = cv.glm(myProjectdata,full2,K=10)$delta
> mse_full2
[1] 0.4123208 0.4120928
>
```

New model after removing month and humidity1 from the model.

```
> summary(full2)

Call:
glm(formula = bikeCount1 ~ as.factor(season) + as.factor(hour) +
    as.factor(IsHoliday) + as.factor(Isweekday) + temp + humidity +
    windspeed1, data = myProjectdata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.7828  -0.3047   0.0455   0.3968   2.2009

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.655332   0.055182  48.120  < 2e-16 ***
as.factor(season)2   0.475696   0.026694  17.821  < 2e-16 ***
as.factor(season)3   0.429165   0.034520  12.432  < 2e-16 ***
as.factor(season)4   0.643848   0.024122  26.691  < 2e-16 ***
as.factor(hour)1    -0.626999   0.052207 -12.010  < 2e-16 ***
as.factor(hour)2    -1.070706   0.052917 -20.234  < 2e-16 ***
as.factor(hour)3    -1.679071   0.053232 -31.543  < 2e-16 ***
as.factor(hour)4    -2.007988   0.052863 -37.984  < 2e-16 ***
as.factor(hour)5    -1.053490   0.052421 -20.097  < 2e-16 ***
as.factor(hour)6     0.172191   0.052176   3.300 0.000971 ***
as.factor(hour)7     1.148051   0.051889  22.125  < 2e-16 ***
as.factor(hour)8     1.782041   0.051777  34.418  < 2e-16 ***
as.factor(hour)9     1.484821   0.051305  28.941  < 2e-16 ***
as.factor(hour)10    1.144853   0.051191  22.364  < 2e-16 ***
as.factor(hour)11    1.234268   0.051403  24.011  < 2e-16 ***
as.factor(hour)12    1.387809   0.051779  26.802  < 2e-16 ***
as.factor(hour)13    1.339081   0.051908  25.797  < 2e-16 ***
as.factor(hour)14    1.278666   0.051943  24.617  < 2e-16 ***
as.factor(hour)15    1.306765   0.052159  25.053  < 2e-16 ***
as.factor(hour)16    1.582795   0.051812  30.549  < 2e-16 ***
as.factor(hour)17    2.002213   0.051535  38.852  < 2e-16 ***
as.factor(hour)18    1.938355   0.050914  38.071  < 2e-16 ***
as.factor(hour)19    1.670789   0.051073  32.713  < 2e-16 ***
as.factor(hour)20    1.390683   0.051193  27.166  < 2e-16 ***
as.factor(hour)21    1.172029   0.051559  22.732  < 2e-16 ***
as.factor(hour)22    0.934519   0.051976  17.980  < 2e-16 ***
as.factor(hour)23    0.536137   0.051294  10.452  < 2e-16 ***
```

```
as.factor(hour)20        1.390683   0.051193  27.166  < 2e-16 ***
as.factor(hour)21        1.172029   0.051559  22.732  < 2e-16 ***
as.factor(hour)22        0.934519   0.051976  17.980  < 2e-16 ***
as.factor(hour)23        0.536137   0.051294  10.452  < 2e-16 ***
as.factor(IsHoliday)1   -0.158383   0.047238  -3.353 0.000804 ***
as.factor(Isweekday)1   -0.046818   0.028608  -1.637 0.101773
as.factor(Isweekday)2   -0.055508   0.027817  -1.995 0.046028 *
as.factor(Isweekday)3   -0.082619   0.027873  -2.964 0.003045 **
as.factor(Isweekday)4   -0.021118   0.027562  -0.766 0.443574
as.factor(Isweekday)5    0.095712   0.027840   3.438 0.000589 ***
as.factor(Isweekday)6    0.067220   0.027504   2.444 0.014548 *
temp                     1.702077   0.064109  26.550  < 2e-16 ***
humidity                -0.676413   0.043612 -15.510  < 2e-16 ***
windspeed1               0.020206   0.002853   7.083 1.54e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.410016)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  3075.5  on 7501  degrees of freedom
AIC: 14710

Number of Fisher Scoring iterations: 2
```

AIC value of the model is 14710

h) Finding the influential points and then rebuild the model for better accuracy

```
> #remove influencial points
> cooksd <- cooks.distance(full2)
> influential <- as.numeric(names(cooksd)[(cooksd > (4/7538))])
> data3<-myProjectdata
> data3<-data3[-influential,]
> dim(data3)
[1] 7084   14
> 
```

As we can see after removing the influential points 455 rows were removed which contained the influential points .I have created another data frame "data3" which contains the data after removing influential points .Created full model again and checked MSE. **Model m3 is the final model after removing influential points.**

```
> mse_m3 = cv.glm(data3,m3,K=10)$delta
> mse_m3
[1] 0.2385545 0.2384116
> 
```

```
> m3 = glm(bikeCount1~as.factor(season)+as.factor(hour)+as.factor(IsHoliday)+as.factor(Isweekday)
+          +temp+humidity+windspeed1,data=data3)
> summary(m3)

Call:
glm(formula = bikeCount1 ~ as.factor(season) + as.factor(hour) + 
    as.factor(IsHoliday) + as.factor(Isweekday) + temp + humidity + 
    windspeed1, data = data3)

Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-1.72566  -0.26916   0.03064   0.33313   1.41555  

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)            2.746904   0.043515  63.126  < 2e-16 ***
as.factor(season)2     0.411956   0.021078  19.544  < 2e-16 ***
as.factor(season)3     0.364573   0.027244  13.382  < 2e-16 ***
as.factor(season)4     0.600185   0.019089  31.442  < 2e-16 ***
as.factor(hour)1      -0.663105   0.041467 -15.991  < 2e-16 ***
as.factor(hour)2      -1.258632   0.044713 -28.149  < 2e-16 ***
as.factor(hour)3      -1.816518   0.043833 -41.442  < 2e-16 ***
as.factor(hour)4      -2.012676   0.041104 -48.965  < 2e-16 ***
as.factor(hour)5      -0.940734   0.041106 -22.885  < 2e-16 ***
as.factor(hour)6       0.494989   0.041820  11.836  < 2e-16 ***
as.factor(hour)7       1.515390   0.042022  36.062  < 2e-16 ***
as.factor(hour)8       1.945415   0.040725  47.770  < 2e-16 ***
as.factor(hour)9       1.487746   0.039471  37.692  < 2e-16 ***
as.factor(hour)10      1.157580   0.039415  29.369  < 2e-16 ***
as.factor(hour)11      1.256106   0.039587  31.730  < 2e-16 ***
as.factor(hour)12      1.427894   0.039990  35.706  < 2e-16 ***
as.factor(hour)13      1.374026   0.040073  34.288  < 2e-16 ***
as.factor(hour)14      1.320929   0.040148  32.901  < 2e-16 ***
as.factor(hour)15      1.338183   0.040234  33.260  < 2e-16 ***
as.factor(hour)16      1.614603   0.039974  40.391  < 2e-16 ***
as.factor(hour)17      2.039539   0.039776  51.276  < 2e-16 ***
as.factor(hour)18      1.979531   0.039383  50.264  < 2e-16 ***
```

```
as.factor(hour)19       1.699709   0.039394  43.147  < 2e-16 ***
as.factor(hour)20       1.394326   0.039360  35.425  < 2e-16 ***
as.factor(hour)21       1.180332   0.039694  29.736  < 2e-16 ***
as.factor(hour)22       0.929966   0.039987  23.257  < 2e-16 ***
as.factor(hour)23       0.543221   0.039461  13.766  < 2e-16 ***
as.factor(IsHoliday)1 -0.056430   0.038944  -1.449 0.147380
as.factor(Isweekday)1 -0.087121   0.022892  -3.806 0.000143 ***
as.factor(Isweekday)2 -0.085774   0.022296  -3.847 0.000121 ***
as.factor(Isweekday)3 -0.116295   0.022365  -5.200 2.05e-07 ***
as.factor(Isweekday)4 -0.031506   0.022150  -1.422 0.154963
as.factor(Isweekday)5  0.044940   0.022271   2.018 0.043642 *
as.factor(Isweekday)6  0.072907   0.022452   3.247 0.001171 **
temp                    1.616931   0.050451  32.050  < 2e-16 ***
humidity               -0.583082   0.034932 -16.692  < 2e-16 ***
windspeed1              0.014376   0.002253   6.381 1.87e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.2370907)

    Null deviance: 13026.6  on 7083  degrees of freedom
Residual deviance:  1670.8  on 7047  degrees of freedom
AIC: 9946.3

Number of Fisher Scoring iterations: 2

>
```

Implementation : MODEL 4(Considering all the numerical and dummy variable created after data pre-processing.

a)Full model considering all the dummy variables

```
> full3 = glm(bikeCount1~temp+humidity+humidity1+windspeed1+as.factor(Isweekday)+as.factor(season1)+as.factor(season2)
+           +as.factor(season3)+as.factor(monthm1)+as.factor(hourh1)+as.factor(IsHoliday),data=data)
> summary(full3)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1 + 
    as.factor(Isweekday) + as.factor(season1) + as.factor(season2) + 
    as.factor(season3) + as.factor(monthm1) + as.factor(hourh1) + 
    as.factor(IsHoliday), data = data)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.2157  -0.5549   0.0500   0.6709   3.4163  

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)            3.632521   0.137749  26.371  < 2e-16 ***
temp                   2.683986   0.112626  23.831  < 2e-16 ***
humidity               2.989332   0.369239   8.096 6.58e-16 ***
humidity1             -3.458191   0.289109 -11.962  < 2e-16 ***
windspeed1            -0.017657   0.004792  -3.684 0.000231 ***
as.factor(Isweekday)1 -0.041477   0.048448  -0.856 0.391966    
as.factor(Isweekday)2 -0.026597   0.047130  -0.564 0.572546    
as.factor(Isweekday)3 -0.023248   0.047330  -0.491 0.623315    
as.factor(Isweekday)4 -0.018479   0.046786  -0.395 0.692881    
as.factor(Isweekday)5  0.097671   0.047166   2.071 0.038413 *  
as.factor(Isweekday)6  0.076741   0.046636   1.646 0.099901 .  
as.factor(season1)1   -0.706687   0.055637 -12.702  < 2e-16 ***
as.factor(season2)1   -0.444330   0.056768  -7.827 5.68e-15 ***
as.factor(season3)1   -0.543465   0.047178 -11.519  < 2e-16 ***
as.factor(monthm1)1    0.169797   0.050059   3.392 0.000698 ***
as.factor(hourh1)1    -1.130431   0.027157 -41.625  < 2e-16 ***
as.factor(IsHoliday)1 -0.176735   0.080007  -2.209 0.027206 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.175879)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  8843.8  on 7521  degrees of freedom
AIC: 22632

Number of Fisher Scoring iterations: 2
```

AIC value of this model is 22632

Model creation using stepwise forward ,backward and both selections respectively based on the step() function:

b) Stepwise Backward model

```
> backward3=step(full3,direction="backward",trace=T)
Start:  AIC=22632.18
bikeCount1 ~ temp + humidity + humidity1 + windspeed1 + as.factor(IsWeekday) +
    as.factor(season1) + as.factor(season2) + as.factor(season3) +
    as.factor(monthm1) + as.factor(hourh1) + as.factor(IsHoliday)

                       Df Deviance   AIC
<none>                      8843.8 22632
- as.factor(IsHoliday)  1   8849.5 22635
- as.factor(IsWeekday)  6   8863.1 22637
- as.factor(monthm1)    1   8857.3 22642
- windspeed1            1   8859.8 22644
- as.factor(season2)    1   8915.8 22691
- humidity              1   8920.9 22696
- as.factor(season3)    1   8999.8 22762
- humidity1             1   9012.0 22772
- as.factor(season1)    1   9033.5 22790
- temp                  1   9511.6 23179
- as.factor(hourh1)     1  10881.2 24193
> 
```

```
> summary(backward3)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1 +
    as.factor(IsWeekday) + as.factor(season1) + as.factor(season2) +
    as.factor(season3) + as.factor(monthm1) + as.factor(hourh1) +
    as.factor(IsHoliday), data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.2157  -0.5549   0.0500   0.6709   3.4163

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            3.632521   0.137749  26.371  < 2e-16 ***
temp                   2.683986   0.112626  23.831  < 2e-16 ***
humidity               2.989332   0.369239   8.096 6.58e-16 ***
humidity1             -3.458191   0.289109 -11.962  < 2e-16 ***
windspeed1            -0.017657   0.004792  -3.684 0.000231 ***
as.factor(IsWeekday)1 -0.041477   0.048448  -0.856 0.391966
as.factor(IsWeekday)2 -0.026597   0.047130  -0.564 0.572546
as.factor(IsWeekday)3 -0.023248   0.047330  -0.491 0.623315
as.factor(IsWeekday)4 -0.018479   0.046786  -0.395 0.692881
as.factor(IsWeekday)5  0.097671   0.047166   2.071 0.038413 *
as.factor(IsWeekday)6  0.076741   0.046636   1.646 0.099901 .
as.factor(season1)1   -0.706687   0.055637 -12.702  < 2e-16 ***
as.factor(season2)1   -0.444330   0.056768  -7.827 5.68e-15 ***
as.factor(season3)1   -0.543465   0.047178 -11.519  < 2e-16 ***
as.factor(monthm1)1    0.169797   0.050059   3.392 0.000698 ***
as.factor(hourh1)1    -1.130431   0.027157 -41.625  < 2e-16 ***
as.factor(IsHoliday)1 -0.176735   0.080007  -2.209 0.027206 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 1.175879)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  8843.8  on 7521  degrees of freedom
AIC: 22632

Number of Fisher Scoring iterations: 2
```

AIC value of this model is 22632

c) Stepwise Forward model

```
> forward3 = step(base1,scope=list(upper=full3,lower=~1),direction= "forward",trace=F,data=data)
> summary(forward3)

Call:
glm(formula = bikeCount1 ~ temp + as.factor(hourh1) + humidity1 + 
    as.factor(season1) + as.factor(season3) + as.factor(season2) + 
    humidity + windspeed1 + as.factor(monthm1) + as.factor(Isweekday) + 
    as.factor(IsHoliday), data = data)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.2157  -0.5549   0.0500   0.6709   3.4163  

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)            3.632521   0.137749  26.371  < 2e-16 ***
temp                   2.683986   0.112626  23.831  < 2e-16 ***
as.factor(hourh1)1    -1.130431   0.027157 -41.625  < 2e-16 ***
humidity1             -3.458191   0.289109 -11.962  < 2e-16 ***
as.factor(season1)1   -0.706687   0.055637 -12.702  < 2e-16 ***
as.factor(season3)1   -0.543465   0.047178 -11.519  < 2e-16 ***
as.factor(season2)1   -0.444330   0.056768  -7.827 5.68e-15 ***
humidity               2.989332   0.369239   8.096 6.58e-16 ***
windspeed1            -0.017657   0.004792  -3.684 0.000231 ***
as.factor(monthm1)1    0.169797   0.050059   3.392 0.000698 ***
as.factor(Isweekday)1 -0.041477   0.048448  -0.856 0.391966    
as.factor(Isweekday)2 -0.026597   0.047130  -0.564 0.572546    
as.factor(Isweekday)3 -0.023248   0.047330  -0.491 0.623315    
as.factor(Isweekday)4 -0.018479   0.046786  -0.395 0.692881    
as.factor(Isweekday)5  0.097671   0.047166   2.071 0.038413 *  
as.factor(Isweekday)6  0.076741   0.046636   1.646 0.099901 .  
as.factor(IsHoliday)1 -0.176735   0.080007  -2.209 0.027206 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.175879)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  8843.8  on 7521  degrees of freedom
AIC: 22632

Number of Fisher Scoring iterations: 2

> |
```

AIC value of this model is 22632

d) Stepwise both model

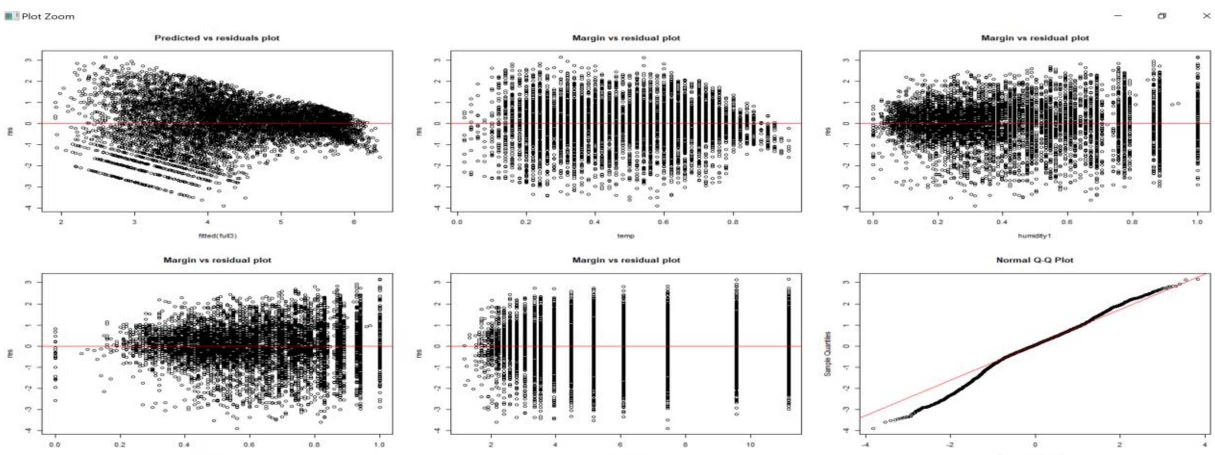
```
> both3 = step(base1,scope=list(upper=full3,lower=~1),direction= "both",trace=F,data=data)
> summary(both3)

Call:
glm(formula = bikeCount1 ~ temp + as.factor(hourh1) + humidity1 + 
    as.factor(season1) + as.factor(season3) + as.factor(season2) + 
    humidity + windspeed1 + as.factor(monthm1) + as.factor(Isweekday) + 
    as.factor(IsHoliday), data = data)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-4.2157  -0.5549   0.0500   0.6709   3.4163  

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)            3.632521   0.137749  26.371  < 2e-16 ***
temp                   2.683986   0.112626  23.831  < 2e-16 ***
as.factor(hourh1)1    -1.130431   0.027157 -41.625  < 2e-16 ***
humidity1             -3.458191   0.289109 -11.962  < 2e-16 ***
as.factor(season1)1   -0.706687   0.055637 -12.702  < 2e-16 ***
as.factor(season3)1   -0.543465   0.047178 -11.519  < 2e-16 ***
as.factor(season2)1   -0.444330   0.056768  -7.827 5.68e-15 ***
humidity               2.989332   0.369239   8.096 6.58e-16 ***
windspeed1            -0.017657   0.004792  -3.684 0.000231 ***
as.factor(monthm1)1    0.169797   0.050059   3.392 0.000698 ***
as.factor(Isweekday)1 -0.041477   0.048448  -0.856 0.391966    
as.factor(Isweekday)2 -0.026597   0.047130  -0.564 0.572546    
as.factor(Isweekday)3 -0.023248   0.047330  -0.491 0.623315    
as.factor(Isweekday)4 -0.018479   0.046786  -0.395 0.692881    
as.factor(Isweekday)5  0.097671   0.047166   2.071 0.038413 *  
as.factor(Isweekday)6  0.076741   0.046636   1.646 0.099901 .  
as.factor(IsHoliday)1 -0.176735   0.080007  -2.209 0.027206 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1.175879)

    Null deviance: 15589.9  on 7537  degrees of freedom
Residual deviance:  8843.8  on 7521  degrees of freedom
AIC: 22632

Number of Fisher Scoring iterations: 2
```

AIC value for this model is 22632

As we could see the AIC value of all the models are same using stepwise regression .So we will consider any one of the model for further analysis . Now we will perform residual analysis on any one of the model and also check for multicollinearity and influential points . **Here I will perform all the operation on full3 model .**

e) Residual Analysis :

```
> par(mfrow=c(2,3))
> plot(fitted(full3),res,main="Predicted vs residuals plot")
> abline(a=0,b=0,col='red')
> plot(temp, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(humidity1, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(humidity, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> plot(windspeed1, res, main="Margin vs residual plot")
> abline(a=0,b=0,col='red')
> qqnorm(res)
> qqline(res,col=2)
> 
```



If we consider the residual analysis, we see that the residual analysis has passed because the model does follow constant variance and also follows normal distribution with few outliers.

f) Multicollinearity by VIF

```
> library(car)
> vif(full3)
                              GVIF Df GVIF^(1/(2*Df))
temp                      3.190763  1        1.786271
humidity                 33.314457  1        5.771868
humidity1                32.680087  1        5.716650
windspeed1                1.094385  1        1.046129
as.factor(Isweekday)      1.130149  6        1.010248
as.factor(season1)        3.670873  1        1.915952
as.factor(season2)        4.047455  1        2.011829
as.factor(season3)        2.734427  1        1.653610
as.factor(monthm1)        3.925775  1        1.981357
as.factor(hourh1)         1.180253  1        1.086395
as.factor(IsHoliday)      1.085619  1        1.041931
> 
```

There is no multicollinearity problem in the variables of the model as the VIF values are less than 4 and for humidity and humidity1 it is very large because humidity is factor if humidity1.

g) Check MSE :

```
> mse_full3 = cv.glm(data,full3,K=10)$delta
> mse_full3
[1] 1.178232 1.177968
>
```

h) Finding the influential points and then rebuild the model for better accuracy

```
> #remove influencial points
> cooksd <- cooks.distance(full3)
> influential <- as.numeric(names(cooksd)[(cooksd > (4/7538))])
> data4<-data
> data4<-data4[-influential,]
> dim(data4)
[1] 7101   22
>
```

As we can see after removing the influential points 437 rows were removed which contained the influential points .I have created another data frame "data4" which contains the data after removing influential points .Created full model again and checked MSE:

```
> mse_m4 = cv.glm(data4,m4,K=10)$delta
> mse_m4
[1] 0.8347881 0.8345950
>
```

```
> m4 = glm(bikeCount1~temp+humidity+humidity1+windspeed1+as.factor(IsWeekday)+as.factor(season1)+as.factor(season2)
+              +as.factor(season3)+as.factor(monthm1)+as.factor(hourh1)+as.factor(IsHoliday),data=data4)
> summary(m4)

Call:
glm(formula = bikeCount1 ~ temp + humidity + humidity1 + windspeed1 + 
    as.factor(IsWeekday) + as.factor(season1) + as.factor(season2) + 
    as.factor(season3) + as.factor(monthm1) + as.factor(hourh1) + 
    as.factor(IsHoliday), data = data4)

Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-2.63933  -0.52615   0.04311   0.59286   2.41361  

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)            3.719916   0.121697  30.567  < 2e-16 ***
temp                   2.736696   0.097793  27.984  < 2e-16 ***
humidity               3.026759   0.331742   9.124  < 2e-16 ***
humidity1             -3.618725   0.259956 -13.921  < 2e-16 ***
windspeed1            -0.027038   0.004182  -6.466 1.08e-10 ***
as.factor(IsWeekday)1 -0.072568   0.041963  -1.729   0.0838 .  
as.factor(IsWeekday)2 -0.062106   0.040653  -1.528   0.1266    
as.factor(IsWeekday)3 -0.048975   0.040866  -1.198   0.2308    
as.factor(IsWeekday)4 -0.033898   0.040361  -0.840   0.4010    
as.factor(IsWeekday)5  0.055459   0.040573   1.367   0.1717    
as.factor(IsWeekday)6  0.041974   0.039769   1.055   0.2913    
as.factor(season1)1   -0.682999   0.049273 -13.861  < 2e-16 ***
as.factor(season2)1   -0.426870   0.050095  -8.521  < 2e-16 ***
as.factor(season3)1   -0.584685   0.040797 -14.332  < 2e-16 ***
as.factor(monthm1)1    0.176444   0.044386   3.975 7.10e-05 ***
as.factor(hourh1)1    -0.988914   0.023584 -41.932  < 2e-16 ***
as.factor(IsHoliday)1 -0.038340   0.076138  -0.504   0.6146    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.8331168)

    Null deviance: 11630.3  on 7100  degrees of freedom
Residual deviance:  5901.8  on 7084  degrees of freedom
AIC: 18874
```

Note: I have compared all the models created above based on AIC and then AIC and selected the best model out of them in order to further improve the model by introducing interaction term in the model

Comparison of models built before introducing interaction terms in the model :

```
> errs
     mse_full mse_full1 mse_full2 mse_full3    mse_m1    mse_m2    mse_m3    mse_m4
[1,] 1.526515  1.061772 0.4117262  1.285773 1.090921 0.8172783 0.2382123 0.8353048
[2,] 1.526390  1.061572 0.4115300  1.285689 1.090862 0.8171760 0.2380879 0.8350843
>
```

From the models in above table ,model number 3 and model number 7 are the best models with least MSE value .This model is the improved version model number 3 after resolving multicollinearity issue and influential points. **I have introduced interaction term in these model in order to further improve the model .**

Then I have also removed influential points from the model after introducing interaction term and then calculated MSE of model in order to find out the best model .

I added dummy variables to the numerical variables as an interaction term and below dummy variable gave the best result :

a) hour as an interaction term with numerical variable temp

b)hour as an interaction term with dummy variable "IsWeekday"

1) New Model 3(full2) after introducing interaction term :

I have renamed the new model from m3 to m5 after adding interaction terms.

```
#renamed the new model to m5 after introducing interaction term

m5= glm(bikeCount1~as.factor(season)+as.factor(hour)+as.factor(IsHoliday)+as.factor(IsWeekday)+
      as.factor(IsWeekday)*as.factor(hour)+temp+temp*as.factor(hour)+humidity+windspeed1,data=myProjectdata)
summary(m5)    #AIC :9257.9
```

Now check for influential points and remove if any from the model m5:

```
> cooksd <- cooks.distance(m5)
> influential <- as.numeric(names(cooksd)[(cooksd > (4/7538))])
> data5<-myProjectdata
> data5<-data5[-influential,]
> dim(data5)
[1] 7103   14
>
```

Rebuild the model : As you can see below the AIC value of new model is much less as compared to previous model.

```
#Rebuild the model with new data set

m5= glm(bikeCount1~as.factor(season)+as.factor(hour)+as.factor(IsHoliday)+as.factor(IsWeekday)+
          as.factor(IsWeekday)*as.factor(hour)+temp+temp*as.factor(hour)+humidity+windspeed1,data=data5)
summary(m5)
#AIC : 3486.9
```

MSE value

```
> mse_m5 = cv.glm(data5,m5,K=10)$delta
> mse_m5
[1] 0.09603777 0.09573909
>
```

2) New Model 7(m3) after introducing interaction term :

```
#renamed the new model to m6 after introducing interaction term

m6 = glm(bikeCount1~as.factor(season)+as.factor(hour)+as.factor(IsHoliday)+as.factor(IsWeekday)+
            as.factor(IsHoliday)+as.factor(IsWeekday)*as.factor(hour)
          +temp+as.factor(hour)*temp+humidity+windspeed1,data=data3)

summary(m6)   #AIC :5270.1
```

Now check for influential points and remove if any from the model:

```
> cooksd <- cooks.distance(m6)
> influential <- as.numeric(names(cooksd)[(cooksd > (4/7538))])
> data6<-data3
> data6<-data6[-influential,]
> dim(data6)
[1] 6648   14
```

Removed influential point from new model “m6,Rebuilt the model with new data set

```
m6= glm(bikeCount1~as.factor(season)+as.factor(hour)+as.factor(IsHoliday)+as.factor(IsWeekday)+
          as.factor(IsWeekday)*as.factor(hour)+temp+temp*as.factor(hour)+humidity+windspeed1,data=data6)
summary(m6)
#AIC : 4878.9
```

MSE

```
> mse_m6 = cv.glm(data6,m6,K=10)$delta
> mse_m6
[1] 0.1230360 0.1225995
>
```

After adding interaction term and removing influential point from the model, both models have improved significantly .Compared the MSE value of all models after adding interaction terms ,**model m5 turned out to be the best model with least MSE value of 0.09585098.**I will discuss the findings of best model in next section

```
> errs
     mse_full mse_full1 mse_full2 mse_full3   mse_m1    mse_m2    mse_m3    mse_m4     mse_m5    mse_m6
[1,] 1.526515  1.061772 0.4117262  1.285773 1.090921 0.8172783 0.2382123 0.8353048 0.09615609 0.1227470
[2,] 1.526390  1.061572 0.4115300  1.285689 1.090862 0.8171760 0.2380879 0.8350843 0.09585098 0.1223259
>
```

Final model :

```
> summary(m5)

Call:
glm(formula = bikeCount1 ~ as.factor(season) + as.factor(hour) +
    as.factor(IsHoliday) + as.factor(Isweekday) + as.factor(Isweekday) *
    as.factor(hour) + temp + temp * as.factor(hour) + humidity +
    windspeed1, data = data5)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0566  -0.1606   0.0307   0.1900   0.8897

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                            3.157024   0.065883  47.919  < 2e-16 ***
as.factor(season)2                     0.450750   0.013226  34.082  < 2e-16 ***
as.factor(season)3                     0.421022   0.017076  24.655  < 2e-16 ***
as.factor(season)4                     0.648827   0.011946  54.313  < 2e-16 ***
as.factor(hour)1                      -0.272555   0.091232  -2.988 0.002823 **
as.factor(hour)2                      -0.360981   0.092795  -3.890 0.000101 ***
as.factor(hour)3                      -0.985774   0.095009 -10.376  < 2e-16 ***
as.factor(hour)4                      -2.450294   0.097346 -25.171  < 2e-16 ***
as.factor(hour)5                      -2.536438   0.096281 -26.344  < 2e-16 ***
as.factor(hour)6                      -2.064427   0.096597 -21.372  < 2e-16 ***
as.factor(hour)7                      -0.950750   0.092376 -10.292  < 2e-16 ***
as.factor(hour)8                       0.320171   0.092319   3.468 0.000527 ***
as.factor(hour)9                       1.034654   0.089876  11.512  < 2e-16 ***
as.factor(hour)10                      1.326301   0.089569  14.808  < 2e-16 ***
as.factor(hour)11                      1.441437   0.089662  16.076  < 2e-16 ***
as.factor(hour)12                      1.563869   0.090836  17.216  < 2e-16 ***
as.factor(hour)13                      1.597538   0.092634  17.246  < 2e-16 ***
as.factor(hour)14                      1.565558   0.091942  17.028  < 2e-16 ***
```

```
as.factor(hour)15                      1.563947   0.093030  16.811  < 2e-16 ***
as.factor(hour)16                      1.361711   0.092620  14.702  < 2e-16 ***
as.factor(hour)17                      1.296226   0.091445  14.175  < 2e-16 ***
as.factor(hour)18                      1.037546   0.090114  11.514  < 2e-16 ***
as.factor(hour)19                      0.767158   0.090657   8.462  < 2e-16 ***
as.factor(hour)20                      0.333955   0.089891   3.715 0.000205 ***
as.factor(hour)21                      0.096293   0.091744   1.050 0.293948
as.factor(hour)22                     -0.225219   0.091725  -2.455 0.014099 *
as.factor(hour)23                     -0.487865   0.091064  -5.357 8.71e-08 ***
as.factor(IsHoliday)1                 -0.044759   0.026841  -1.668 0.095446 .
as.factor(Isweekday)1                 -1.150365   0.069187 -16.627  < 2e-16 ***
as.factor(Isweekday)2                 -1.289525   0.067594 -19.078  < 2e-16 ***
as.factor(Isweekday)3                 -1.122821   0.068439 -16.406  < 2e-16 ***
as.factor(Isweekday)4                 -0.977758   0.066538 -14.695  < 2e-16 ***
as.factor(Isweekday)5                 -0.609166   0.067019  -9.089  < 2e-16 ***
as.factor(Isweekday)6                 -0.059334   0.066219  -0.896 0.370276
temp                                   2.048549   0.101631  20.157  < 2e-16 ***
humidity                              -0.505409   0.022426 -22.536  < 2e-16 ***
windspeed1                             0.015348   0.001425  10.769  < 2e-16 ***
as.factor(hour)1:as.factor(Isweekday)1 -0.509433  0.098555  -5.169 2.42e-07 ***
as.factor(hour)2:as.factor(Isweekday)1 -0.763737  0.102283  -7.467 9.22e-14 ***
as.factor(hour)3:as.factor(Isweekday)1 -0.642349  0.103457  -6.209 5.65e-10 ***
as.factor(hour)4:as.factor(Isweekday)1  0.944381  0.101055   9.345  < 2e-16 ***
as.factor(hour)5:as.factor(Isweekday)1  2.079707  0.105395  19.732  < 2e-16 ***
as.factor(hour)6:as.factor(Isweekday)1  3.231628  0.103152  31.329  < 2e-16 ***
as.factor(hour)7:as.factor(Isweekday)1  3.329601  0.102365  32.527  < 2e-16 ***
as.factor(hour)8:as.factor(Isweekday)1  2.796721  0.099440  28.125  < 2e-16 ***
as.factor(hour)9:as.factor(Isweekday)1  1.569254  0.095544  16.424  < 2e-16 ***
as.factor(hour)10:as.factor(Isweekday)1 0.573047  0.093588   6.123 9.68e-10 ***
as.factor(hour)11:as.factor(Isweekday)1 0.405145  0.093525   4.332 1.50e-05 ***
as.factor(hour)12:as.factor(Isweekday)1 0.539146  0.094477   5.707 1.20e-08 ***
as.factor(hour)13:as.factor(Isweekday)1 0.429901  0.095758   4.489 7.26e-06 ***
as.factor(hour)14:as.factor(Isweekday)1 0.343560  0.095348   3.603 0.000317 ***
as.factor(hour)15:as.factor(Isweekday)1 0.491912  0.094574   5.201 2.04e-07 ***
as.factor(hour)16:as.factor(Isweekday)1 0.854226  0.093789   9.108  < 2e-16 ***
as.factor(hour)17:as.factor(Isweekday)1 1.576505  0.094227  16.731  < 2e-16 ***
as.factor(hour)18:as.factor(Isweekday)1 1.702789  0.093972  18.120  < 2e-16 ***
```

```
as.factor(hour)19:as.factor(Isweekday)1  1.499999   0.094257  15.914  < 2e-16 ***
as.factor(hour)20:as.factor(Isweekday)1  1.474262   0.094227  15.646  < 2e-16 ***
as.factor(hour)21:as.factor(Isweekday)1  1.445681   0.095475  15.142  < 2e-16 ***
as.factor(hour)22:as.factor(Isweekday)1  1.297615   0.095749  13.552  < 2e-16 ***
as.factor(hour)23:as.factor(Isweekday)1  1.092621   0.095274  11.468  < 2e-16 ***
as.factor(hour)1:as.factor(Isweekday)2  -0.435209   0.095223  -4.570 4.95e-06 ***
as.factor(hour)2:as.factor(Isweekday)2  -0.849003   0.097620  -8.697  < 2e-16 ***
as.factor(hour)3:as.factor(Isweekday)2  -0.750764   0.102497  -7.325 2.66e-13 ***
as.factor(hour)4:as.factor(Isweekday)2   0.996895   0.098743  10.096  < 2e-16 ***
as.factor(hour)5:as.factor(Isweekday)2   2.450727   0.100914  24.285  < 2e-16 ***
as.factor(hour)6:as.factor(Isweekday)2   3.437272   0.099196  34.651  < 2e-16 ***
as.factor(hour)7:as.factor(Isweekday)2   3.525753   0.095823  36.794  < 2e-16 ***
as.factor(hour)8:as.factor(Isweekday)2   2.992878   0.095569  31.317  < 2e-16 ***
as.factor(hour)9:as.factor(Isweekday)2   1.725440   0.095214  18.122  < 2e-16 ***
as.factor(hour)10:as.factor(Isweekday)2  0.607976   0.093592   6.496 8.82e-11 ***
as.factor(hour)11:as.factor(Isweekday)2  0.480707   0.093710   5.130 2.98e-07 ***
as.factor(hour)12:as.factor(Isweekday)2  0.509772   0.093579   5.447 5.28e-08 ***
as.factor(hour)13:as.factor(Isweekday)2  0.549124   0.093549   5.870 4.56e-09 ***
as.factor(hour)14:as.factor(Isweekday)2  0.426615   0.093080   4.583 4.66e-06 ***
as.factor(hour)15:as.factor(Isweekday)2  0.505263   0.094892   5.325 1.04e-07 ***
as.factor(hour)16:as.factor(Isweekday)2  0.957289   0.092391  10.361  < 2e-16 ***
as.factor(hour)17:as.factor(Isweekday)2  1.770080   0.092602  19.115  < 2e-16 ***
as.factor(hour)18:as.factor(Isweekday)2  1.923854   0.091741  20.971  < 2e-16 ***
as.factor(hour)19:as.factor(Isweekday)2  1.712156   0.092363  18.537  < 2e-16 ***
as.factor(hour)20:as.factor(Isweekday)2  1.694510   0.093584  18.107  < 2e-16 ***
as.factor(hour)21:as.factor(Isweekday)2  1.667348   0.094884  17.573  < 2e-16 ***
as.factor(hour)22:as.factor(Isweekday)2  1.596739   0.094871  16.831  < 2e-16 ***
as.factor(hour)23:as.factor(Isweekday)2  1.348175   0.093557  14.410  < 2e-16 ***
as.factor(hour)1:as.factor(Isweekday)3  -0.525626   0.096413  -5.452 5.16e-08 ***
as.factor(hour)2:as.factor(Isweekday)3  -1.200991   0.100005 -12.009  < 2e-16 ***
as.factor(hour)3:as.factor(Isweekday)3  -0.898649   0.101814  -8.826  < 2e-16 ***
as.factor(hour)4:as.factor(Isweekday)3   0.712301   0.102955   6.919 4.97e-12 ***
as.factor(hour)5:as.factor(Isweekday)3   2.111721   0.100817  20.946  < 2e-16 ***
as.factor(hour)6:as.factor(Isweekday)3   3.228438   0.101169  31.911  < 2e-16 ***
as.factor(hour)7:as.factor(Isweekday)3   3.263418   0.097675  33.411  < 2e-16 ***
```


```
as.factor(hour)20:as.factor(Isweekday)2  1.694510   0.093584  18.107  < 2e-16 ***
as.factor(hour)21:as.factor(Isweekday)2  1.667348   0.094884  17.573  < 2e-16 ***
as.factor(hour)22:as.factor(Isweekday)2  1.596739   0.094871  16.831  < 2e-16 ***
as.factor(hour)23:as.factor(Isweekday)2  1.348175   0.093557  14.410  < 2e-16 ***
as.factor(hour)1:as.factor(Isweekday)3  -0.525626   0.096413  -5.452 5.16e-08 ***
as.factor(hour)2:as.factor(Isweekday)3  -1.200991   0.100005 -12.009  < 2e-16 ***
as.factor(hour)3:as.factor(Isweekday)3  -0.898649   0.101814  -8.826  < 2e-16 ***
as.factor(hour)4:as.factor(Isweekday)3   0.712301   0.102955   6.919 4.97e-12 ***
as.factor(hour)5:as.factor(Isweekday)3   2.111721   0.100817  20.946  < 2e-16 ***
as.factor(hour)6:as.factor(Isweekday)3   3.228438   0.101169  31.911  < 2e-16 ***
as.factor(hour)7:as.factor(Isweekday)3   3.263418   0.097675  33.411  < 2e-16 ***
```


```
as.factor(hour)8:as.factor(Isweekday)3   2.813333   0.097972  28.716  < 2e-16 ***
as.factor(hour)9:as.factor(Isweekday)3   1.627000   0.095256  17.080  < 2e-16 ***
as.factor(hour)10:as.factor(Isweekday)3  0.422303   0.094979   4.446 8.87e-06 ***
as.factor(hour)11:as.factor(Isweekday)3  0.352197   0.092974   3.788 0.000153 ***
as.factor(hour)12:as.factor(Isweekday)3  0.450793   0.093673   4.812 1.52e-06 ***
as.factor(hour)13:as.factor(Isweekday)3  0.381943   0.093688   4.077 4.62e-05 ***
as.factor(hour)14:as.factor(Isweekday)3  0.315039   0.092743   3.397 0.000685 ***
as.factor(hour)15:as.factor(Isweekday)3  0.315970   0.094009   3.361 0.000781 ***
as.factor(hour)16:as.factor(Isweekday)3  0.826027   0.094417   8.749  < 2e-16 ***
as.factor(hour)17:as.factor(Isweekday)3  1.561697   0.093437  16.714  < 2e-16 ***
as.factor(hour)18:as.factor(Isweekday)3  1.650137   0.092985  17.746  < 2e-16 ***
as.factor(hour)19:as.factor(Isweekday)3  1.495644   0.093670  15.967  < 2e-16 ***
as.factor(hour)20:as.factor(Isweekday)3  1.488801   0.093450  15.931  < 2e-16 ***
as.factor(hour)21:as.factor(Isweekday)3  1.425632   0.095222  14.972  < 2e-16 ***
as.factor(hour)22:as.factor(Isweekday)3  1.440224   0.096055  14.994  < 2e-16 ***
as.factor(hour)23:as.factor(Isweekday)3  1.254267   0.093932  13.353  < 2e-16 ***
as.factor(hour)1:as.factor(Isweekday)4  -0.445464   0.095513  -4.664 3.16e-06 ***
as.factor(hour)2:as.factor(Isweekday)4  -1.090756   0.097281 -11.212  < 2e-16 ***
as.factor(hour)3:as.factor(Isweekday)4  -0.890059   0.096564  -9.217  < 2e-16 ***
as.factor(hour)4:as.factor(Isweekday)4   0.641074   0.098908   6.481 9.71e-11 ***
as.factor(hour)5:as.factor(Isweekday)4   2.077930   0.098749  21.042  < 2e-16 ***
as.factor(hour)6:as.factor(Isweekday)4   3.148284   0.098466  31.973  < 2e-16 ***
as.factor(hour)7:as.factor(Isweekday)4   3.207698   0.095792  33.486  < 2e-16 ***
as.factor(hour)8:as.factor(Isweekday)4   2.702471   0.095534  28.288  < 2e-16 ***
as.factor(hour)9:as.factor(Isweekday)4   1.523718   0.093154  16.357  < 2e-16 ***
as.factor(hour)10:as.factor(Isweekday)4  0.287513   0.093072   3.089 0.002015 **
as.factor(hour)11:as.factor(Isweekday)4  0.258377   0.091840   2.813 0.004917 **
as.factor(hour)12:as.factor(Isweekday)4  0.371053   0.093094   3.986 6.79e-05 ***
as.factor(hour)13:as.factor(Isweekday)4  0.294263   0.092559   3.179 0.001483 **
as.factor(hour)14:as.factor(Isweekday)4  0.208603   0.092054   2.266 0.023477 *
as.factor(hour)15:as.factor(Isweekday)4  0.232288   0.093596   2.482 0.013095 *
```

```
as.factor(hour)16:as.factor(Isweekday)4  0.643372   0.092288   6.971 3.43e-12 ***
as.factor(hour)17:as.factor(Isweekday)4  1.410631   0.092061  15.323  < 2e-16 ***
as.factor(hour)18:as.factor(Isweekday)4  1.539626   0.091589  16.810  < 2e-16 ***
as.factor(hour)19:as.factor(Isweekday)4  1.385021   0.092053  15.046  < 2e-16 ***
as.factor(hour)20:as.factor(Isweekday)4  1.461091   0.091378  15.990  < 2e-16 ***
as.factor(hour)21:as.factor(Isweekday)4  1.405616   0.093081  15.101  < 2e-16 ***
as.factor(hour)22:as.factor(Isweekday)4  1.381877   0.094131  14.680  < 2e-16 ***
as.factor(hour)23:as.factor(Isweekday)4  1.340513   0.092895  14.430  < 2e-16 ***
as.factor(hour)1:as.factor(Isweekday)5  -0.528499   0.096918  -5.453 5.12e-08 ***
as.factor(hour)2:as.factor(Isweekday)5  -0.920517   0.096392  -9.550  < 2e-16 ***
as.factor(hour)3:as.factor(Isweekday)5  -1.104433   0.097266 -11.355  < 2e-16 ***
as.factor(hour)4:as.factor(Isweekday)5   0.427039   0.100678   4.242 2.25e-05 ***
as.factor(hour)5:as.factor(Isweekday)5   1.746660   0.099600  17.537  < 2e-16 ***
as.factor(hour)6:as.factor(Isweekday)5   2.582197   0.099910  25.845  < 2e-16 ***
as.factor(hour)7:as.factor(Isweekday)5   2.656031   0.095874  27.703  < 2e-16 ***
as.factor(hour)8:as.factor(Isweekday)5   2.307517   0.096153  23.998  < 2e-16 ***
as.factor(hour)9:as.factor(Isweekday)5   1.205830   0.093719  12.866  < 2e-16 ***
as.factor(hour)10:as.factor(Isweekday)5  0.153392   0.092430   1.660 0.097052 .
as.factor(hour)11:as.factor(Isweekday)5  0.101589   0.092791   1.095 0.273636
as.factor(hour)12:as.factor(Isweekday)5  0.197398   0.093263   2.117 0.034331 *
as.factor(hour)13:as.factor(Isweekday)5  0.188222   0.093460   2.014 0.044055 *
as.factor(hour)14:as.factor(Isweekday)5  0.113041   0.093220   1.213 0.225314
as.factor(hour)15:as.factor(Isweekday)5  0.160996   0.093442   1.723 0.084942 .
as.factor(hour)16:as.factor(Isweekday)5  0.507862   0.093315   5.442 5.44e-08 ***
as.factor(hour)17:as.factor(Isweekday)5  1.087209   0.093416  11.638  < 2e-16 ***
as.factor(hour)18:as.factor(Isweekday)5  1.083871   0.093162  11.634  < 2e-16 ***
as.factor(hour)19:as.factor(Isweekday)5  0.883915   0.092234   9.583  < 2e-16 ***
as.factor(hour)20:as.factor(Isweekday)5  0.859212   0.092949   9.244  < 2e-16 ***
as.factor(hour)21:as.factor(Isweekday)5  0.861621   0.093209   9.244  < 2e-16 ***
as.factor(hour)22:as.factor(Isweekday)5  1.059945   0.094859  11.174  < 2e-16 ***
as.factor(hour)23:as.factor(Isweekday)5  1.133866   0.092921  12.202  < 2e-16 ***
as.factor(hour)1:as.factor(Isweekday)6  -0.059268   0.092843  -0.638 0.523255
as.factor(hour)2:as.factor(Isweekday)6  -0.129013   0.093419  -1.381 0.167318
as.factor(hour)3:as.factor(Isweekday)6  -0.267389   0.094891  -2.818 0.004848 **
as.factor(hour)4:as.factor(Isweekday)6   0.167023   0.097566   1.712 0.086960 .
as.factor(hour)5:as.factor(Isweekday)6   0.202557   0.100931   2.007 0.044802 *
as.factor(hour)6:as.factor(Isweekday)6   0.489021   0.098060   4.987 6.28e-07 ***
as.factor(hour)7:as.factor(Isweekday)6   0.387349   0.095891   4.039 5.42e-05 ***
as.factor(hour)8:as.factor(Isweekday)6   0.264957   0.095359   2.779 0.005476 **
as.factor(hour)9:as.factor(Isweekday)6   0.198468   0.093390   2.125 0.033610 *
as.factor(hour)10:as.factor(Isweekday)6  0.011618   0.091642   0.127 0.899124
as.factor(hour)11:as.factor(Isweekday)6 -0.008648   0.091880  -0.094 0.925016
```

```
as.factor(hour)16:as.factor(Isweekday)6  0.118338   0.092125   1.285 0.198997
as.factor(hour)17:as.factor(Isweekday)6  0.127258   0.092314   1.379 0.168080
as.factor(hour)18:as.factor(Isweekday)6  0.163785   0.091384   1.792 0.073133 .
as.factor(hour)19:as.factor(Isweekday)6  0.117214   0.091616   1.279 0.200795
as.factor(hour)20:as.factor(Isweekday)6  0.158540   0.091629   1.730 0.083634 .
as.factor(hour)21:as.factor(Isweekday)6  0.292391   0.093104   3.140 0.001694 **
as.factor(hour)22:as.factor(Isweekday)6  0.465000   0.094180   4.937 8.11e-07 ***
as.factor(hour)23:as.factor(Isweekday)6  0.604136   0.093153   6.485 9.46e-11 ***
as.factor(hour)1:temp                    0.008555   0.141125   0.061 0.951666
as.factor(hour)2:temp                   -0.128650   0.145490  -0.884 0.376588
as.factor(hour)3:temp                   -0.130846   0.147792  -0.885 0.376006
as.factor(hour)4:temp                   -0.216912   0.145642  -1.489 0.136442
as.factor(hour)5:temp                    0.053305   0.143541   0.371 0.710382
as.factor(hour)6:temp                    0.095926   0.140934   0.681 0.496119
as.factor(hour)7:temp                   -0.309383   0.136164  -2.272 0.023109 *
as.factor(hour)8:temp                   -0.971644   0.133894  -7.257 4.40e-13 ***
as.factor(hour)9:temp                   -1.356809   0.131381 -10.327  < 2e-16 ***
as.factor(hour)10:temp                  -0.916482   0.130066  -7.046 2.02e-12 ***
as.factor(hour)11:temp                  -0.799093   0.130875  -6.106 1.08e-09 ***
as.factor(hour)12:temp                  -0.822228   0.130259  -6.312 2.92e-10 ***
as.factor(hour)13:temp                  -0.882419   0.131013  -6.735 1.77e-11 ***
as.factor(hour)14:temp                  -0.861154   0.130670  -6.590 4.71e-11 ***
as.factor(hour)15:temp                  -0.865089   0.130588  -6.625 3.75e-11 ***
as.factor(hour)16:temp                  -0.541349   0.130509  -4.148 3.39e-05 ***
as.factor(hour)17:temp                  -0.599496   0.130169  -4.606 4.19e-06 ***
as.factor(hour)18:temp                  -0.388492   0.128743  -3.018 0.002557 **
as.factor(hour)19:temp                  -0.148534   0.130929  -1.134 0.256640
as.factor(hour)20:temp                   0.107336   0.131668   0.815 0.414987
as.factor(hour)21:temp                   0.165230   0.134431   1.229 0.219072
as.factor(hour)22:temp                   0.302042   0.136921   2.206 0.027420 *
as.factor(hour)23:temp                   0.163878   0.136513   1.200 0.230002
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.09303882)

    Null deviance: 13183.07  on 7102  degrees of freedom
Residual deviance:   642.43  on 6905  degrees of freedom
AIC: 3486.9

Number of Fisher Scoring iterations: 2
```

The final model has also passed residual analysis test with constant variance and is normal.