|  |
| --- |
| Edwisor |
| Bike rental count prediction |
| R-CodeFile |
|  |
| **Pavan Kumar** |
| **04-Feb-20** |

|  |
| --- |
| The aim of the project is to predict the count of bike rental based on seasonal and environmental settings. By predicting the count, it would be possible to help accommodate in managing the number of bikes required on a daily basis, and being prepared for high demand of bikes during peak periods. |

Problem statement - The objective of this Case is to Predication of bike rental count on daily based on the environmental and seasonal settings.

The details of data attributes in the dataset are as follows –

instant: Record index

dteday: Date

season: Season (1:springer, 2:summer, 3:fall, 4:winter)

yr: Year (0: 2011, 1:2012)

mnth: Month (1 to 12)

hr: Hour (0 to 23)

holiday: weather day is holiday or not (extracted fromHoliday Schedule)

weekday: Day of the week

workingday: If day is neither weekend nor holiday is 1, otherwise is 0.

weathersit: (extracted fromFreemeteo)

1: Clear, Few clouds, Partly cloudy, Partly cloudy

2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds

4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

temp: Normalized temperature in Celsius. The values are derived via (t-t\_min)/(t\_max-t\_min), t\_min=-8, t\_max=+39 (only in hourly scale)

atemp: Normalized feeling temperature in Celsius. The values are derived via (t-t\_min)/(t\_maxt\_min), t\_min=-16, t\_max=+50 (only in hourly scale)

hum: Normalized humidity. The values are divided to 100 (max)

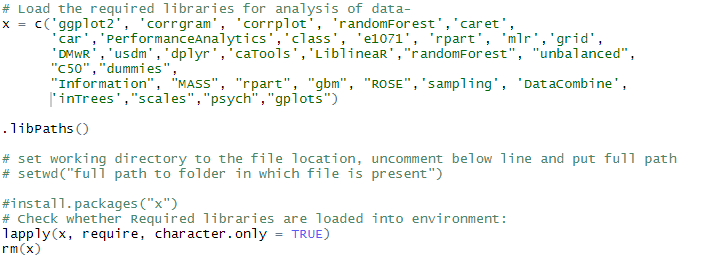
windspeed: Normalized wind speed. The values are divided to 67 (max)

casual: count of casual users

registered: count of registered users

cnt: count of total rental bikes including both casual and registered







[[1]]

[1] TRUE

[[2]]

[1] TRUE

[[3]]

[1] TRUE

[[4]]

[1] TRUE

[[5]]

[1] TRUE

[[6]]

[1] TRUE

[[7]]

[1] TRUE

[[8]]

[1] TRUE

[[9]]

[1] TRUE

[[10]]

[1] TRUE

[[11]]

[1] TRUE

[[12]]

[1] TRUE

[[13]]

[1] TRUE

[[14]]

[1] TRUE

[[15]]

[1] TRUE

[[16]]

[1] TRUE

[[17]]

[1] TRUE

[[18]]

[1] TRUE

[[19]]

[1] TRUE

[[20]]

[1] TRUE

[[21]]

[1] TRUE

[[22]]

[1] TRUE

[[23]]

[1] TRUE

[[24]]

[1] TRUE

[[25]]

[1] TRUE

[[26]]

[1] TRUE

[[27]]

[1] TRUE

[[28]]

[1] TRUE

[[29]]

[1] TRUE

[[30]]

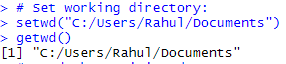
[1] TRUE

[[31]]

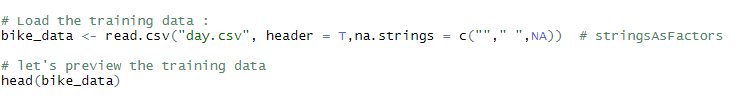
[1] TRUE

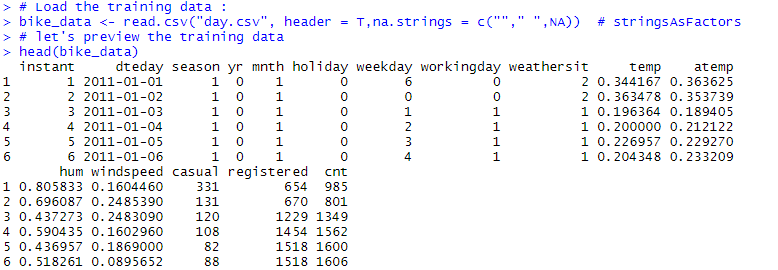
[[32]]

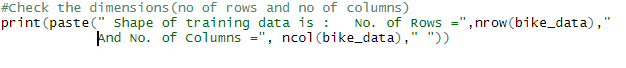
[1] TRUE



Understanding the Data:

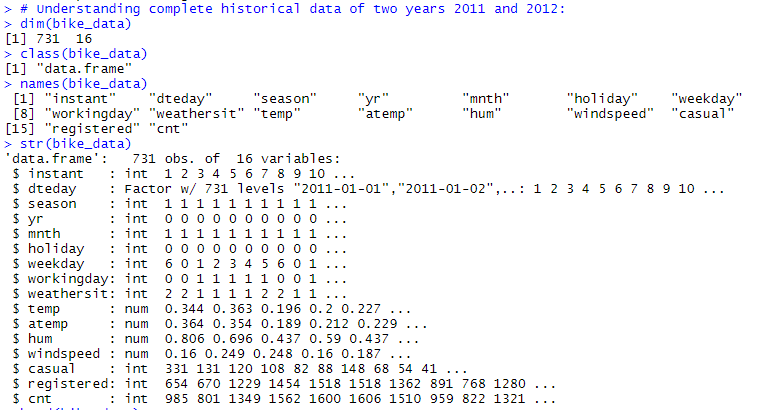


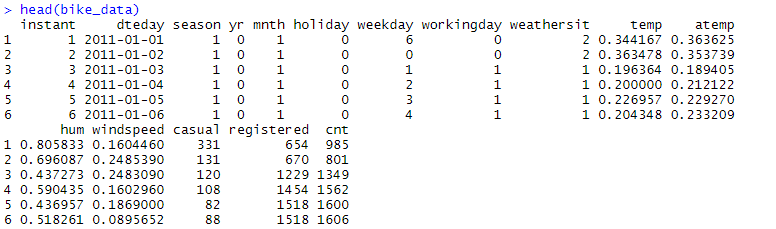


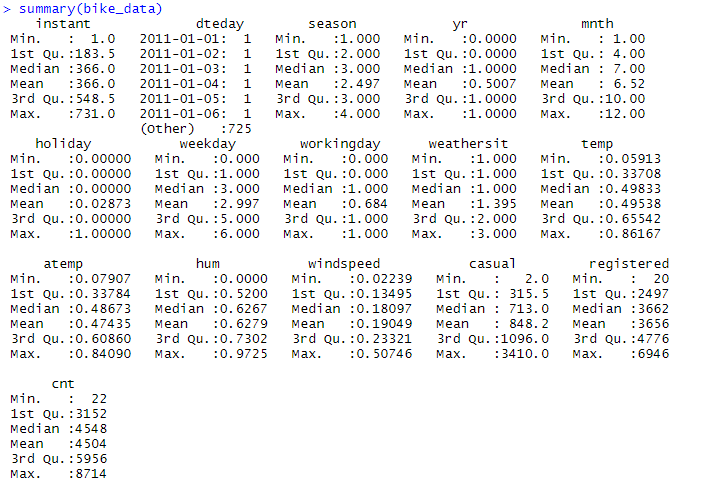


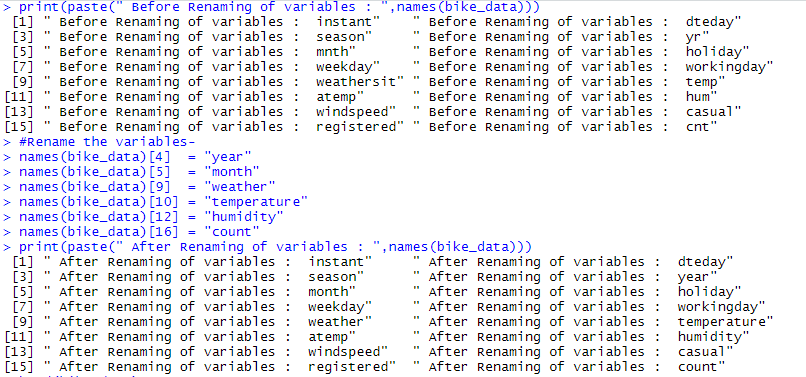


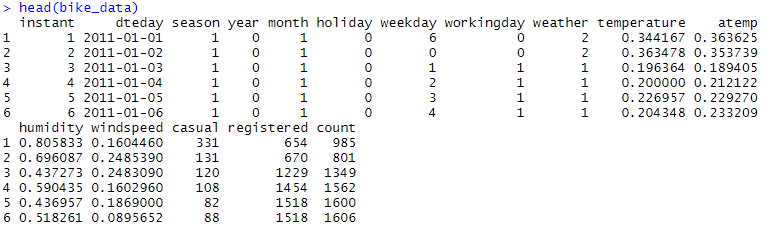
Understanding complete historical data of two years 2011 and 2012:









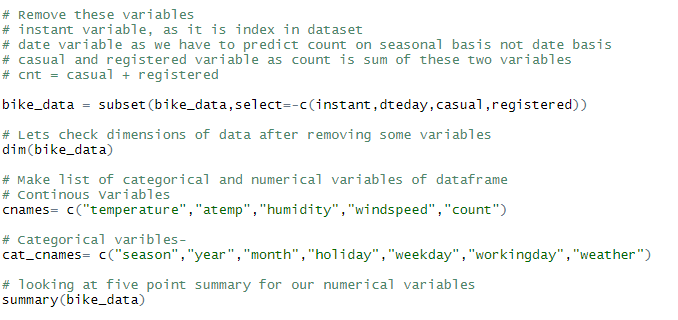


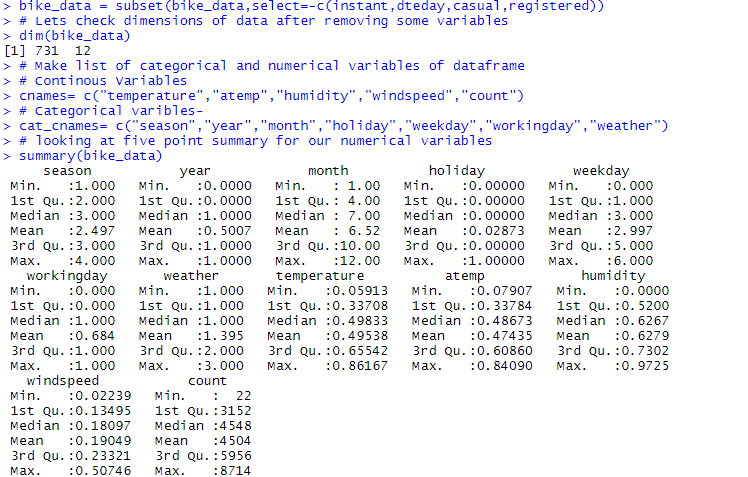
Variable identification:

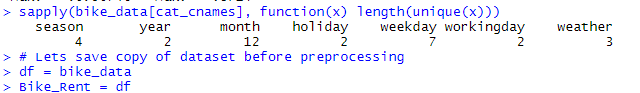


Missing value Analysis:

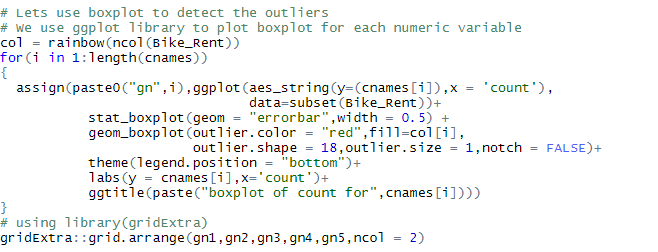


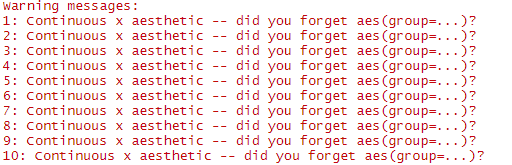


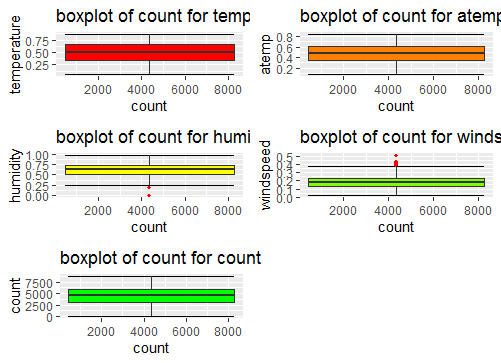


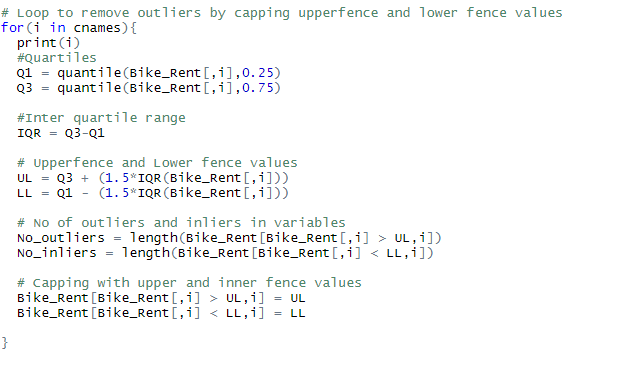


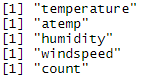
Outlier Analysis:

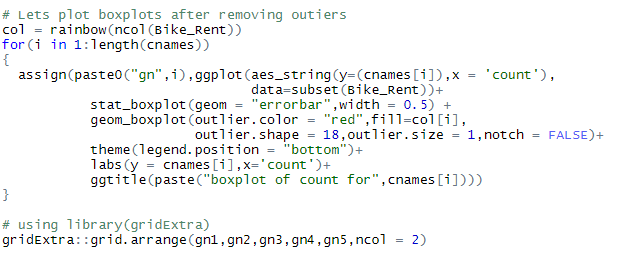


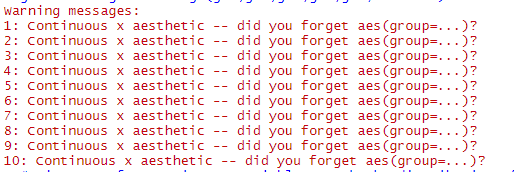


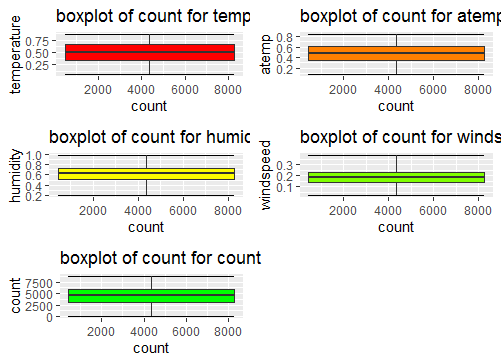








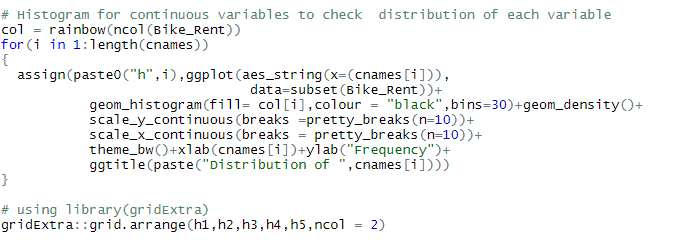


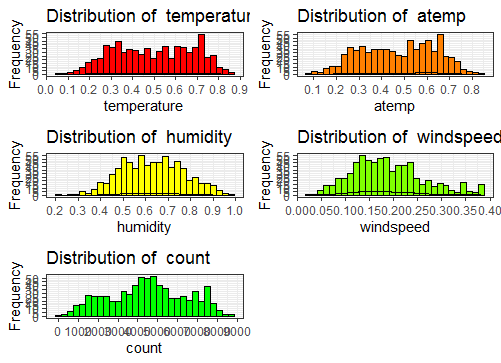


Data understanding and Visualization:

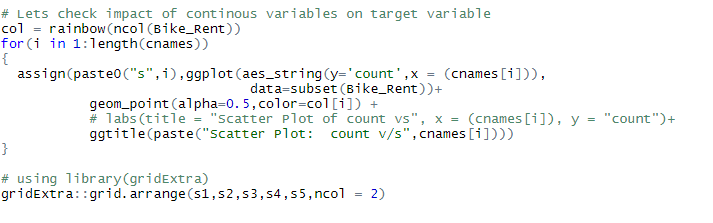
**Multi-Variate Analysis, Univariate Analysis and Bi-Variate Analysis of Train dataset**

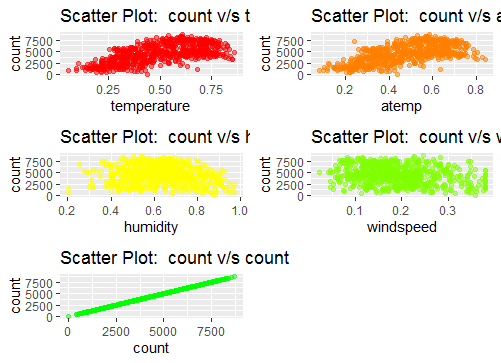
**Univariate Analysis:**

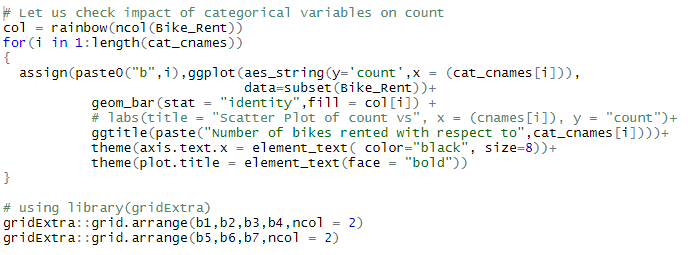


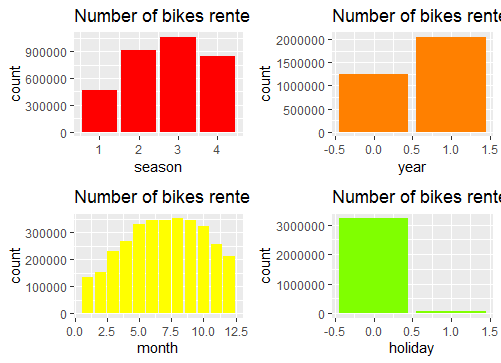


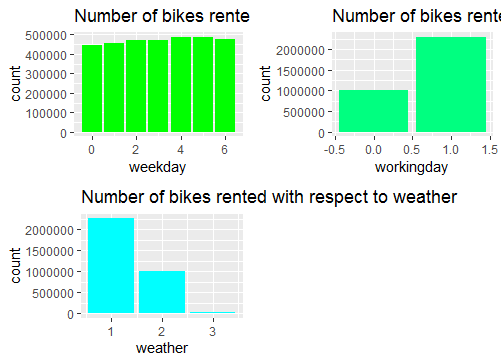
**BI-VARIATE ANALYSIS:**

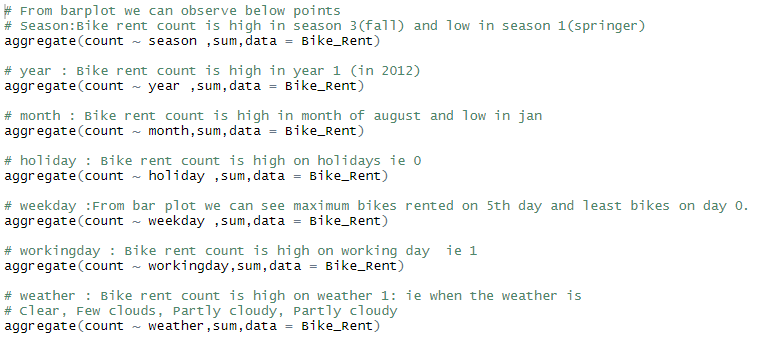


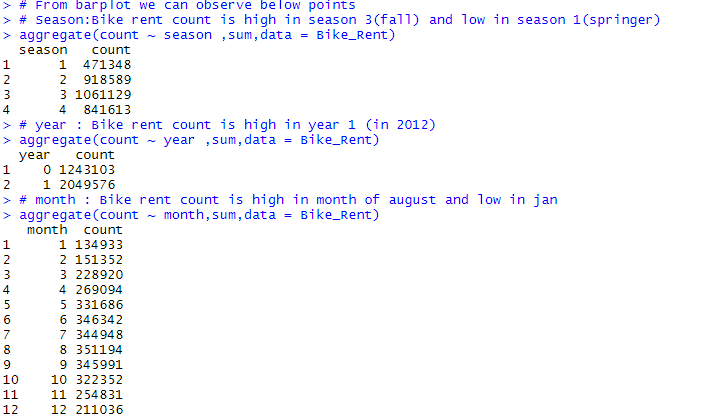


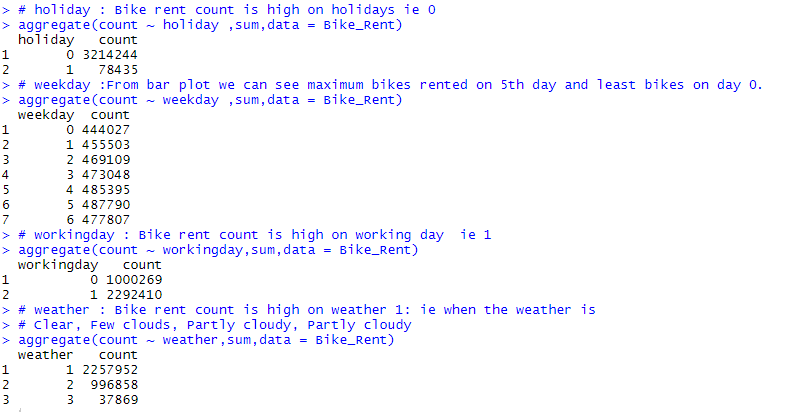


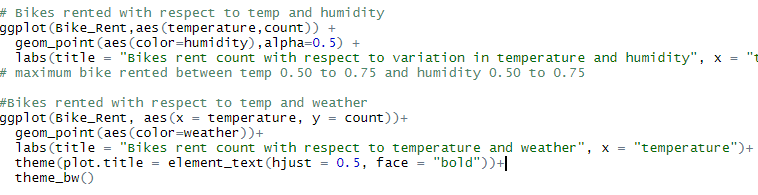


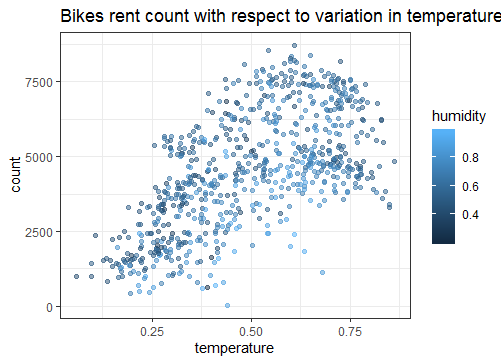


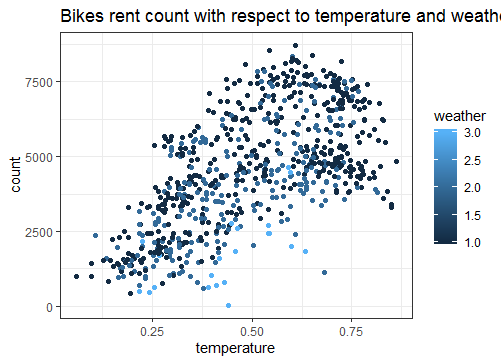


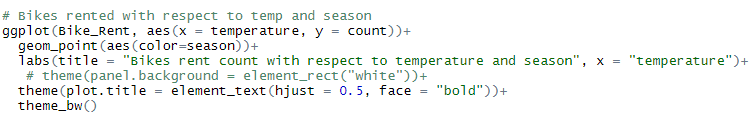


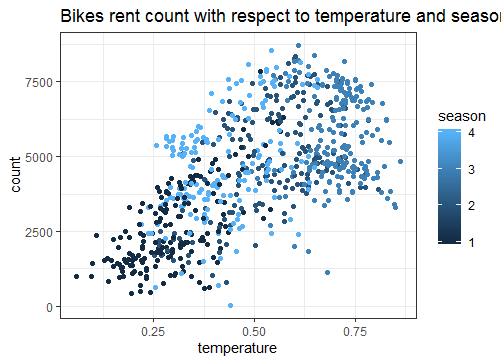




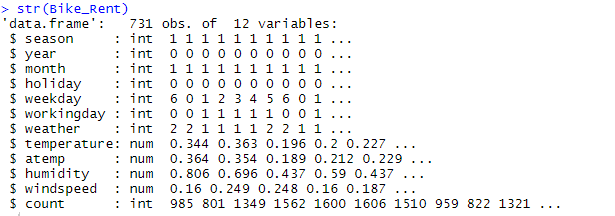


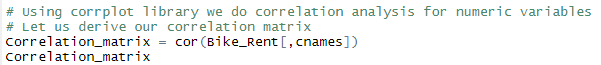


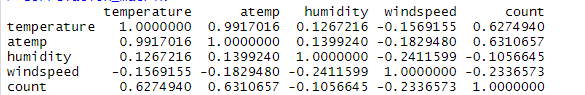


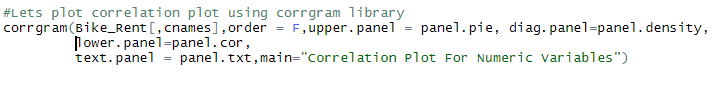


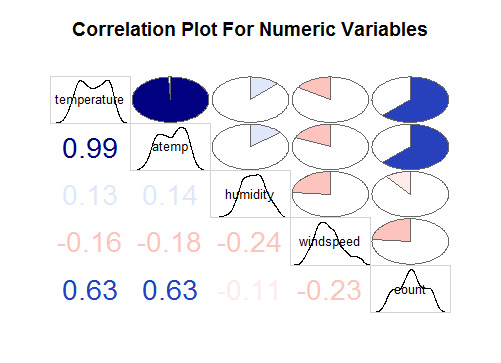
Feature Selection:

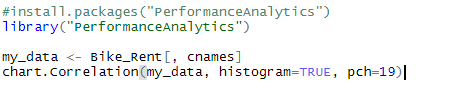


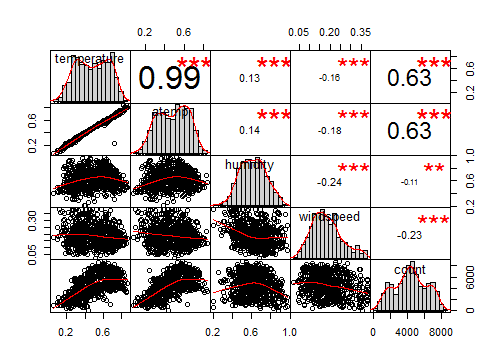


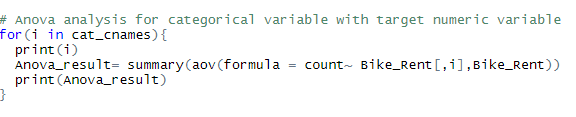


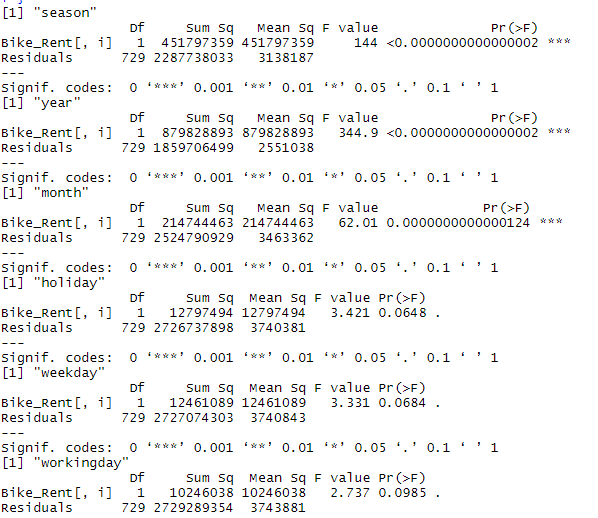


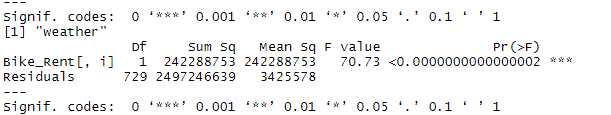




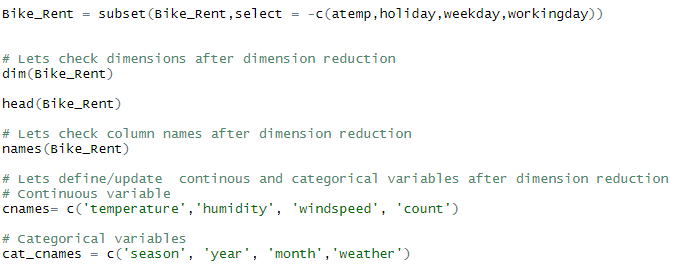


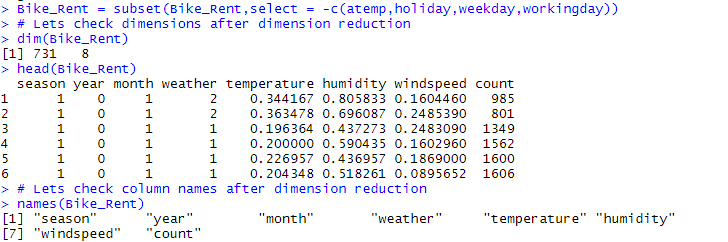






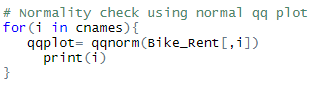
**Dimension reduction:**



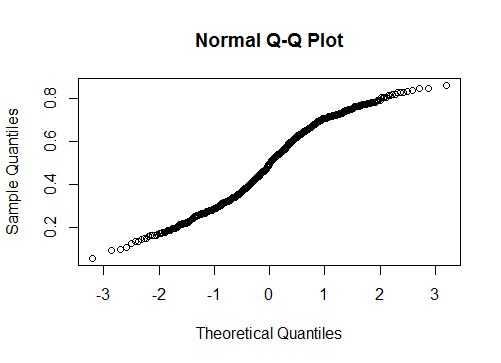


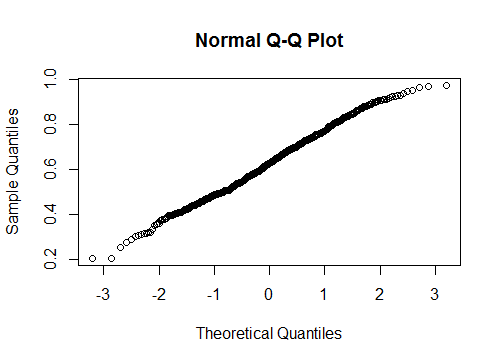
**Feature scaling:**

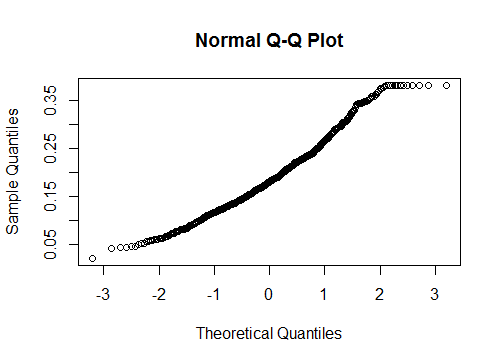
**Normality check**

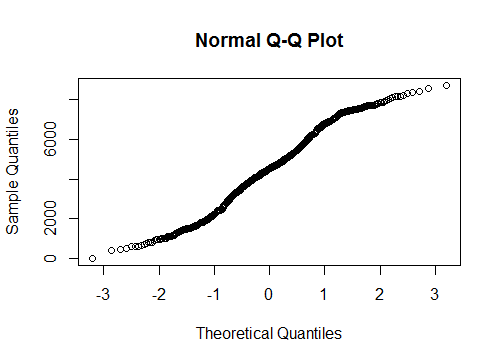




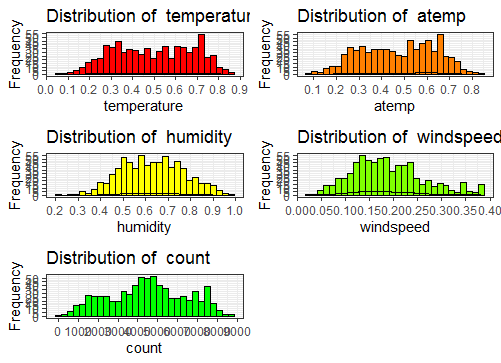


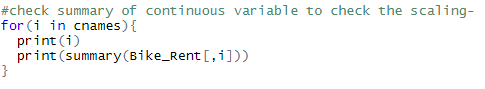


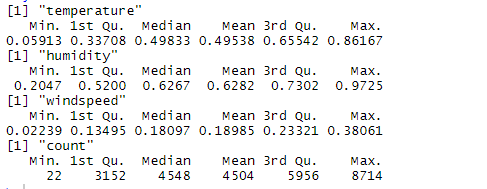


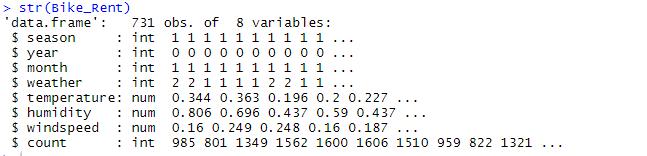




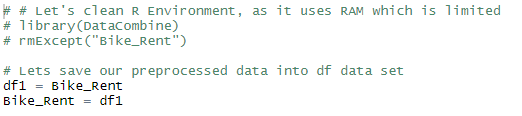


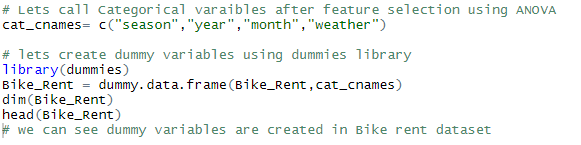


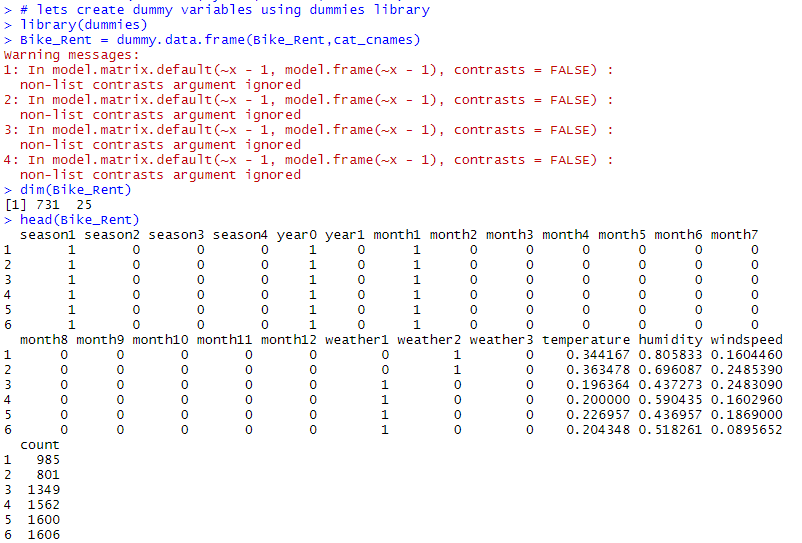


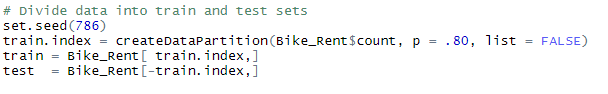


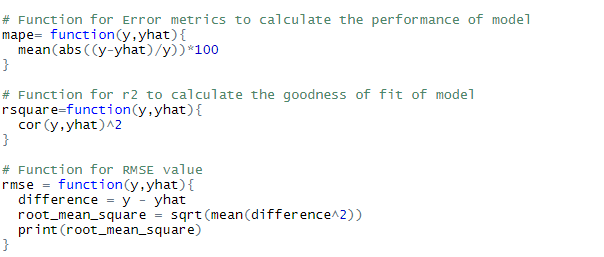
**Model development:**



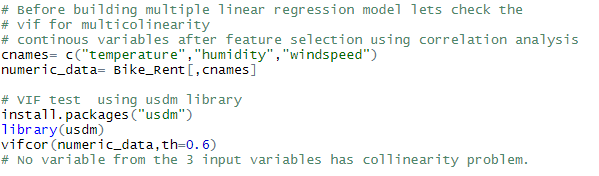


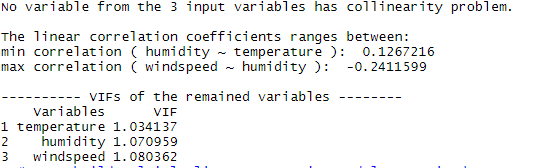


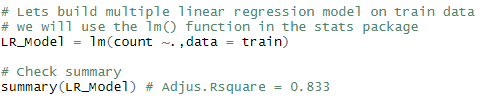


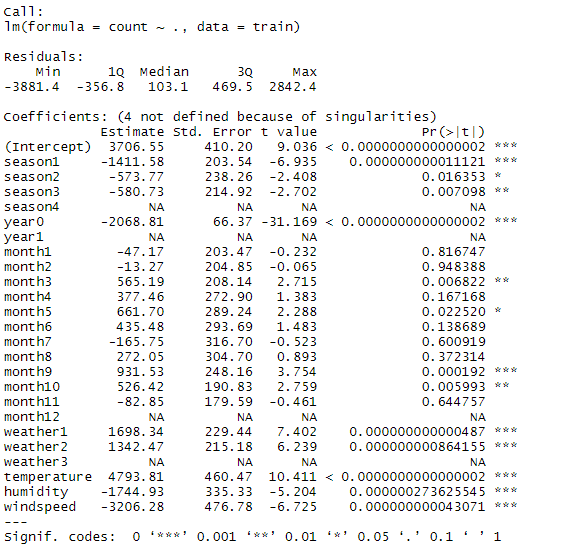


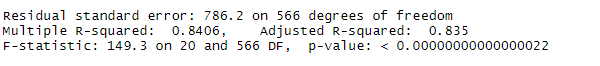
**Multiple Linear Regression Model:**

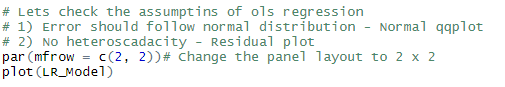


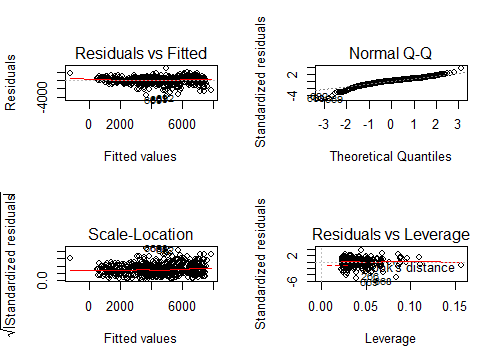


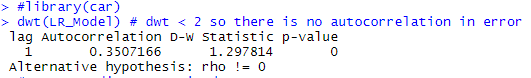


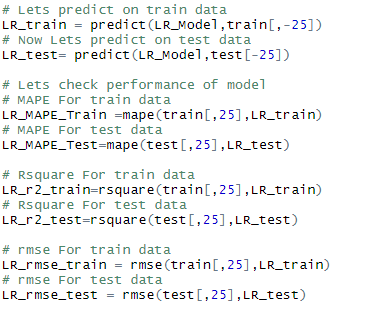


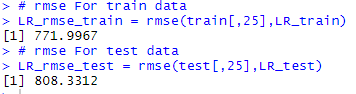


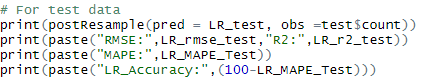


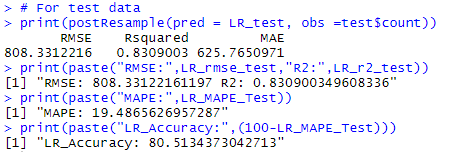




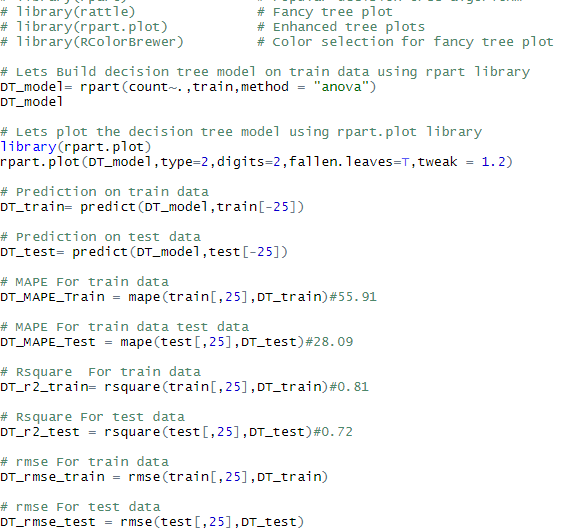


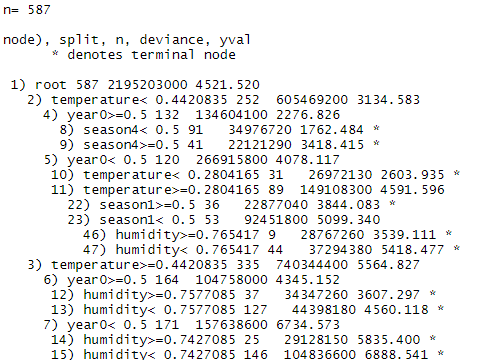


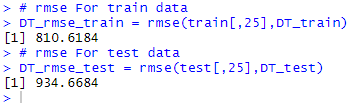


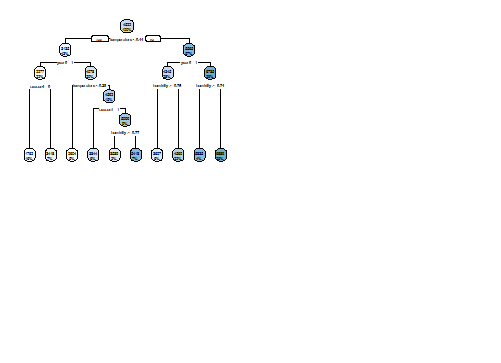


**Decision Tree:**

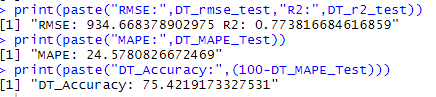




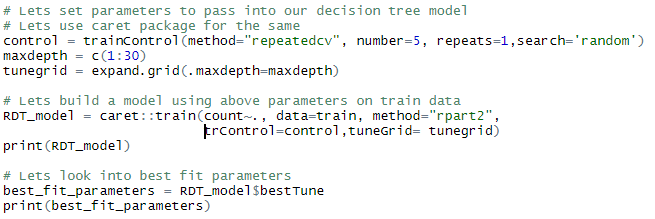


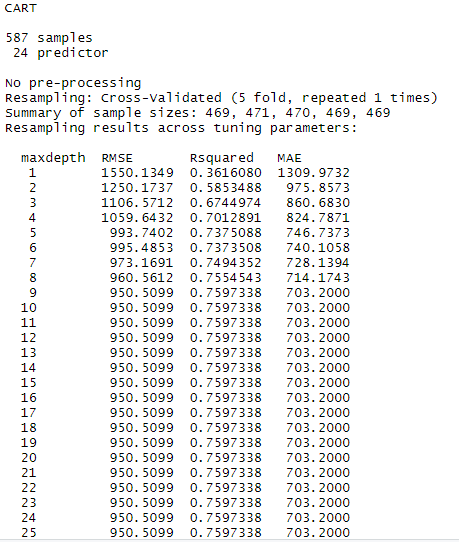


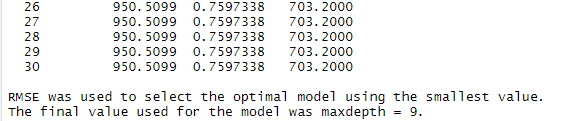


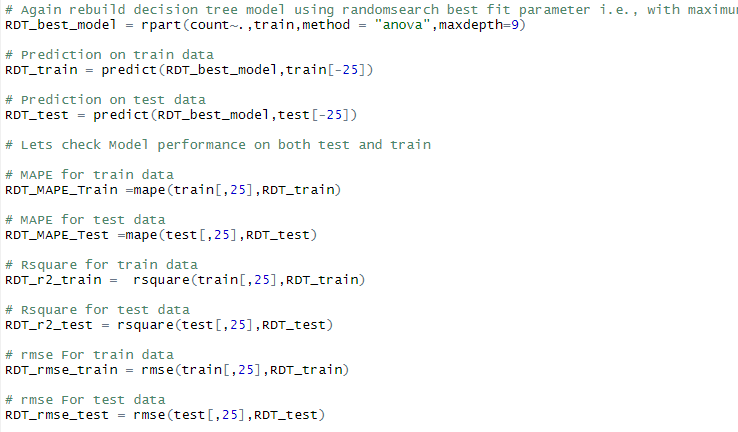


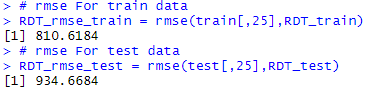
**Random Search CV in Decision Tree:**



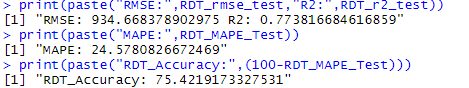




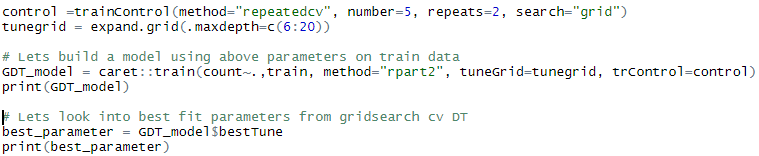


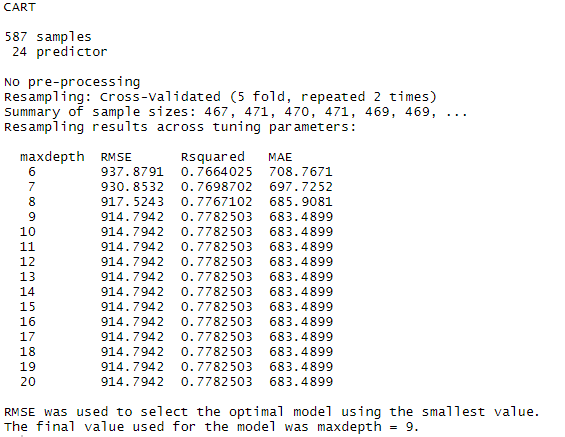


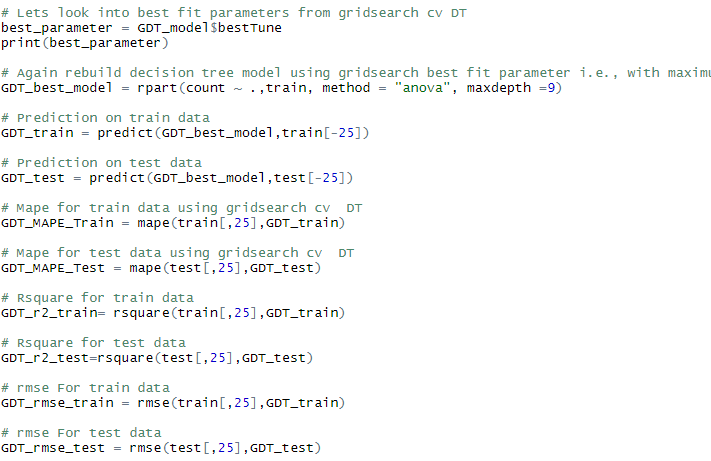


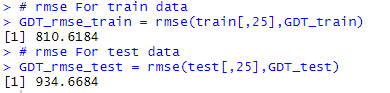


**Grid Search CV in Decision tree:**

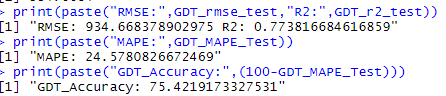




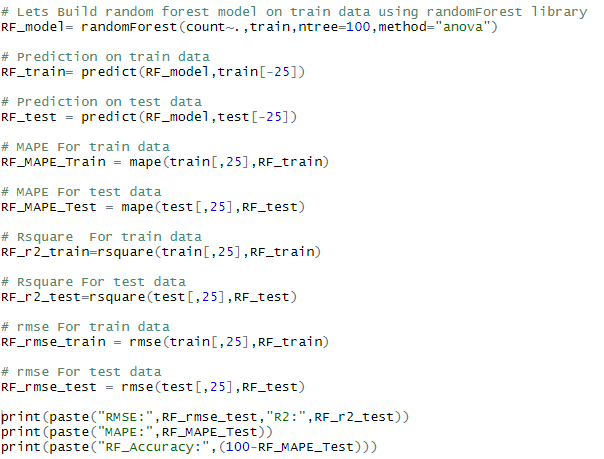


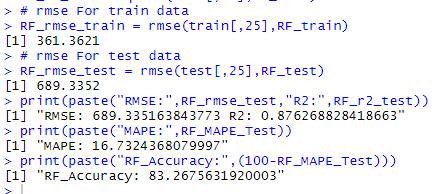




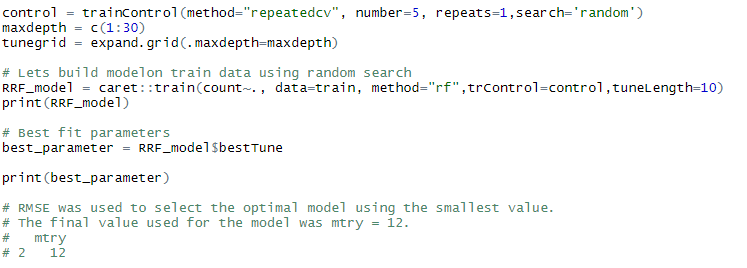


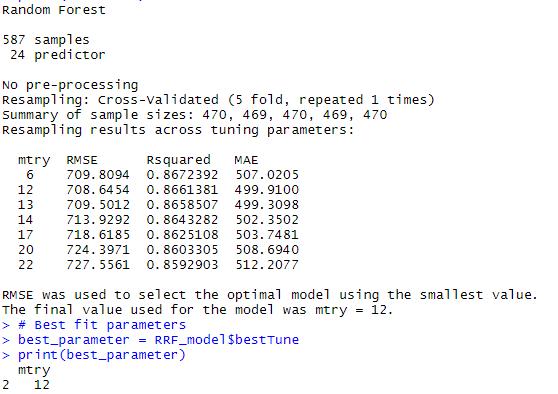
**Random Forest:**

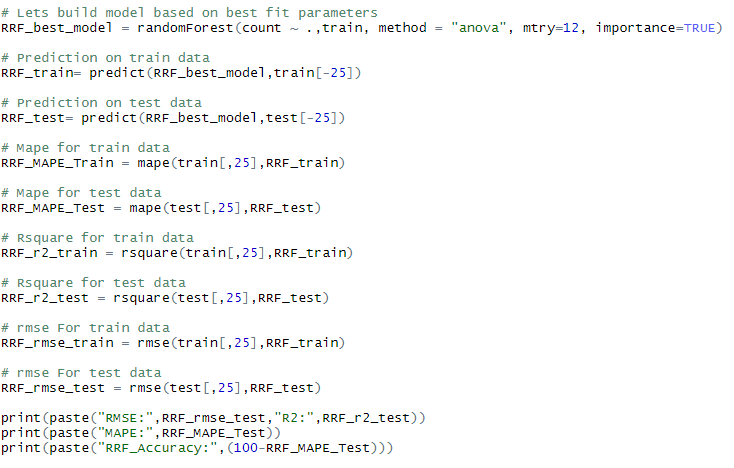


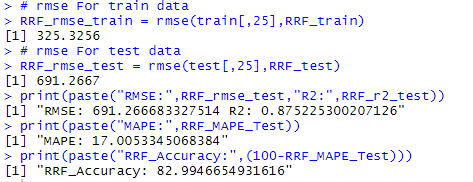


**Random Search CV in Random Forest**

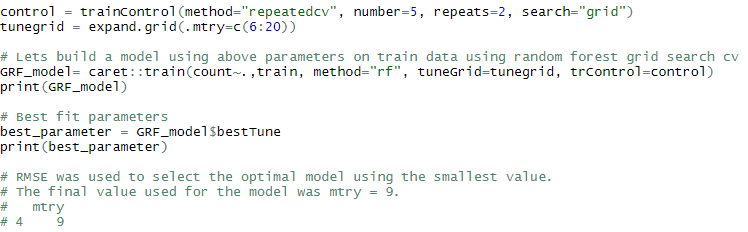


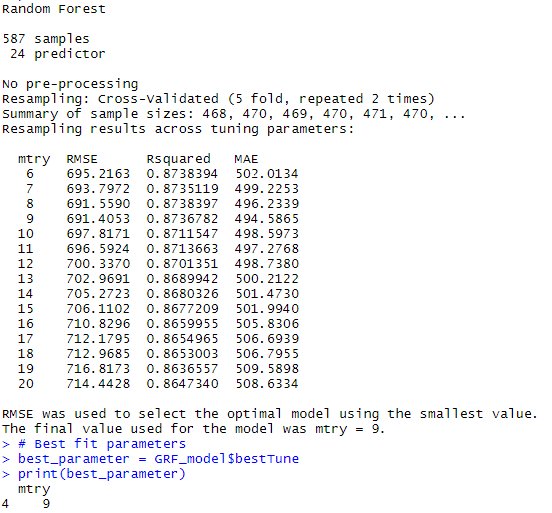


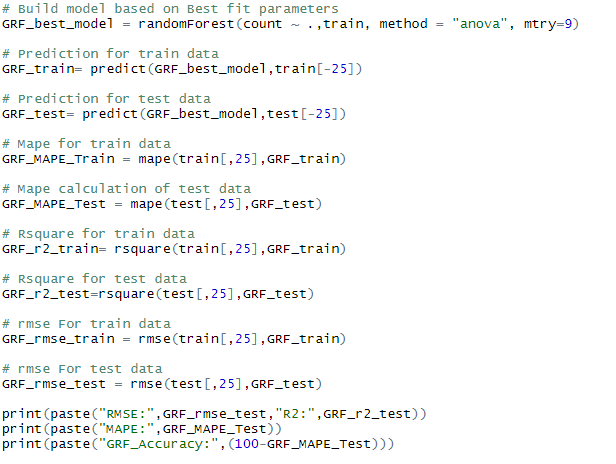


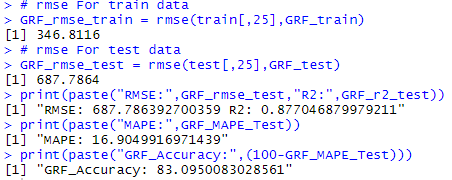


**Grid Search CV in Random Forest:**

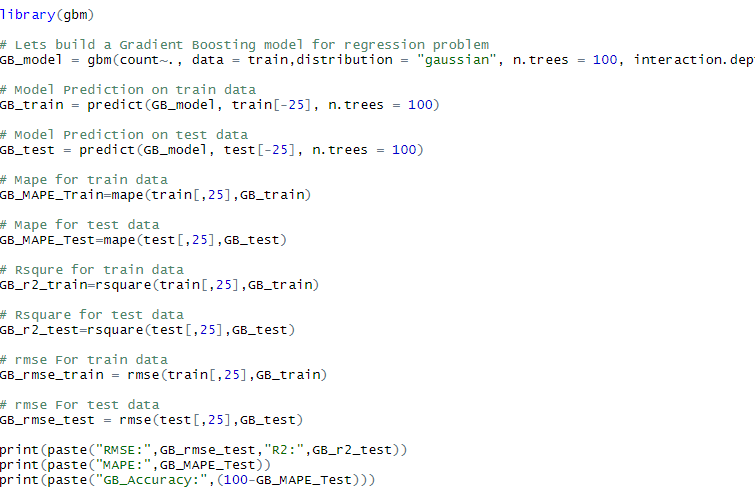


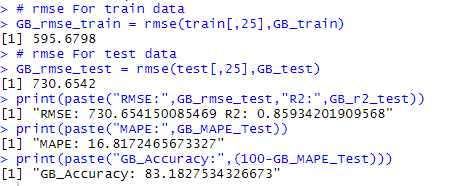




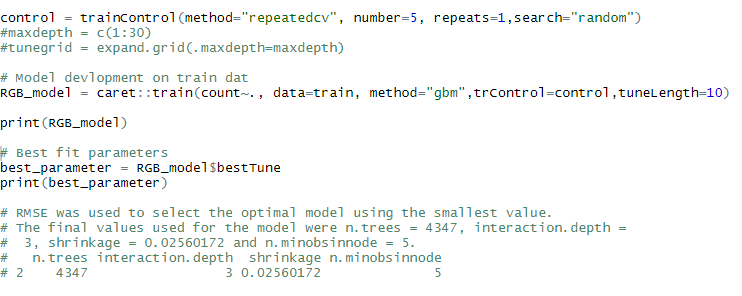


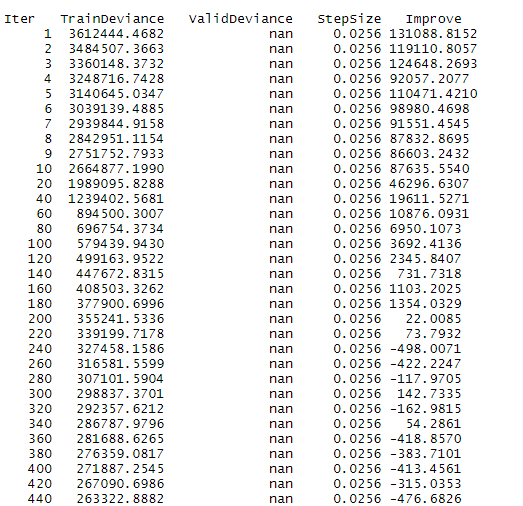
**Gradient Boosting:**

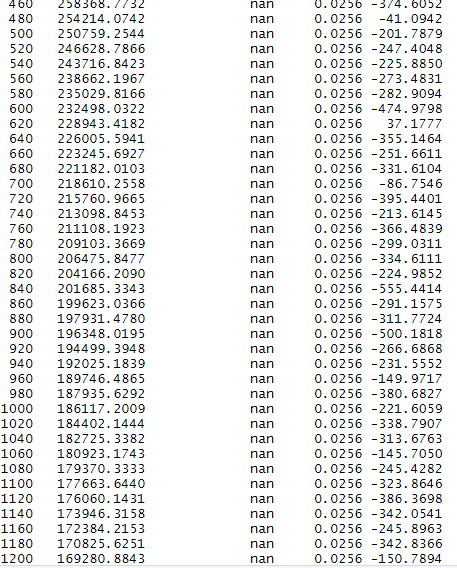


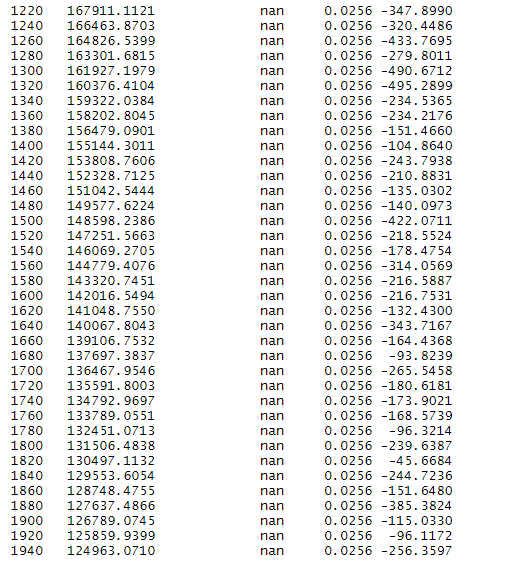


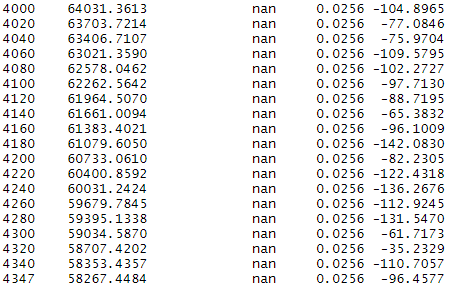
**Random Search CV in Gradient Boosting:**

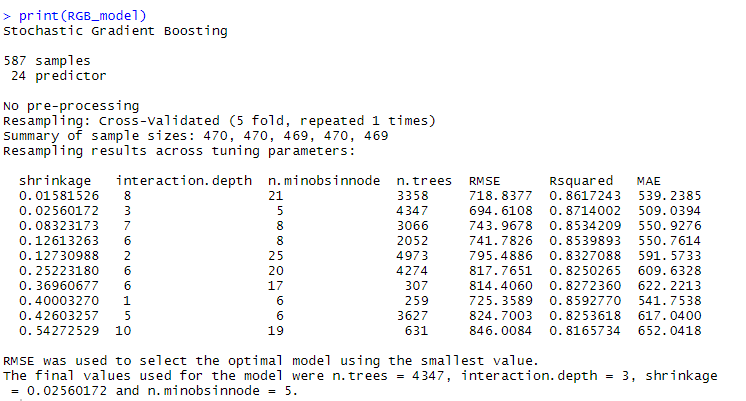




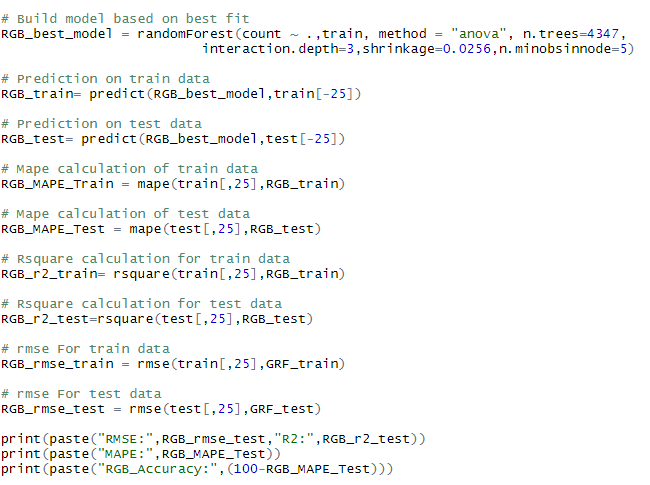


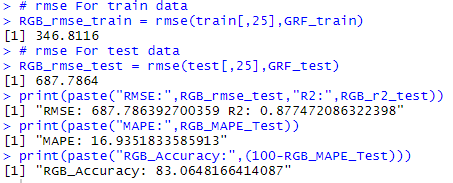




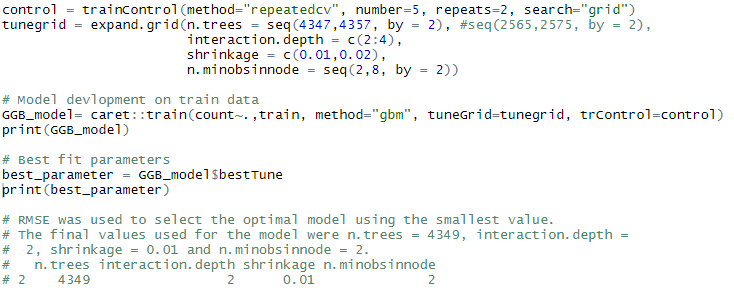


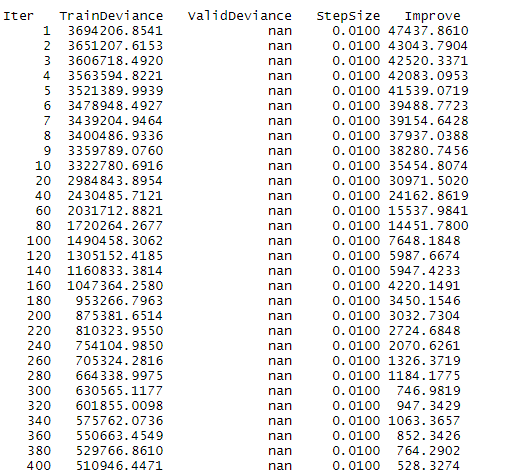


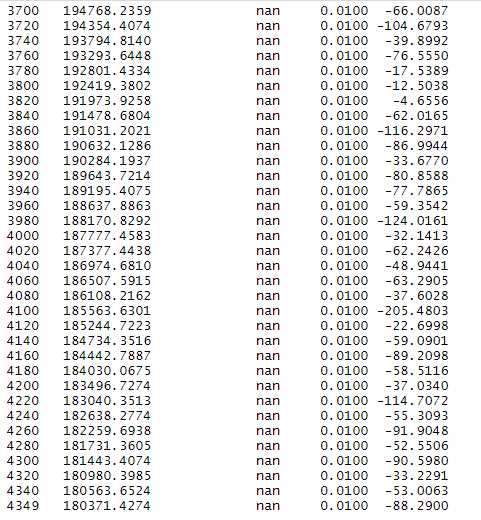


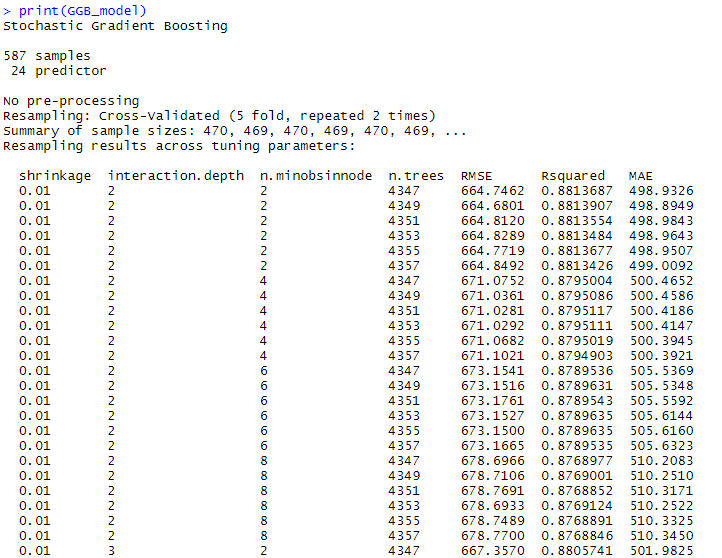


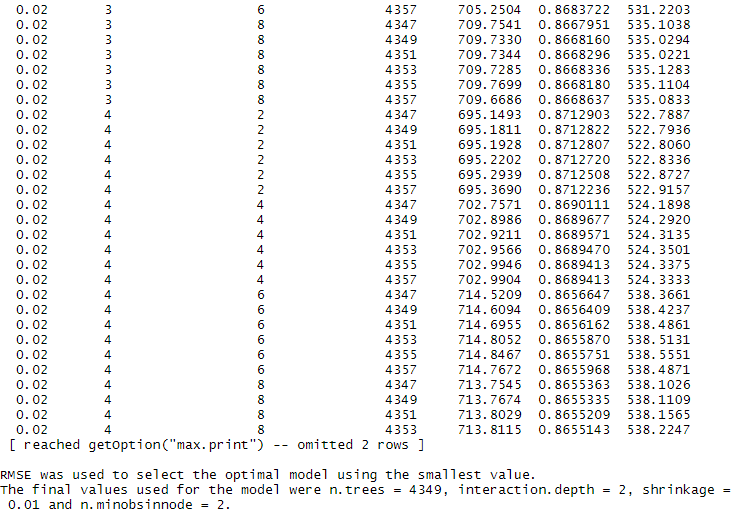
**Grid Search CV in Gradient Boosting:**



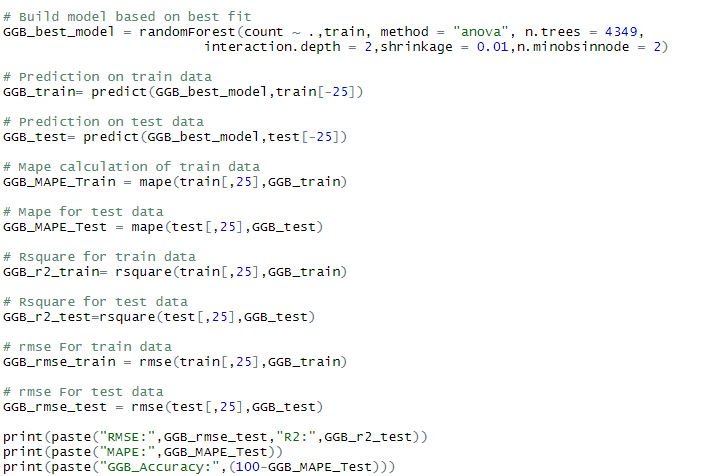


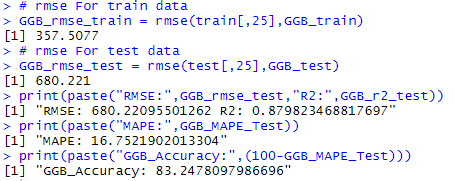












**Results:**

