|  |
| --- |
| edwisor |
| Bike rental count prediction |
| Python project |
|  |
| **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)

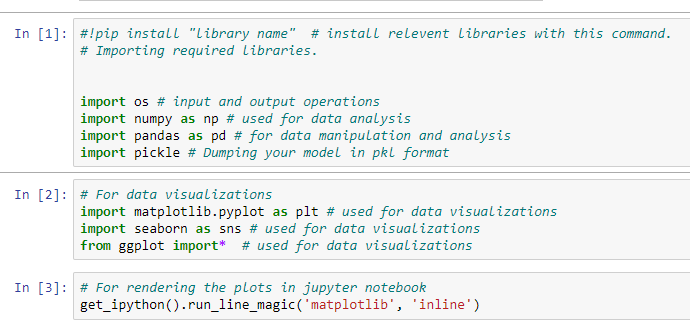
hum: Normalized humidity. The values are divided to 100 (max)

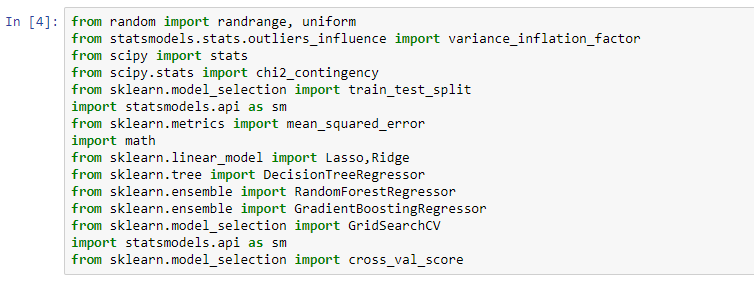
windspeed: Normalized wind speed. The values are divided to 67 (max)

casual: count of casual users

registered: count of registered users

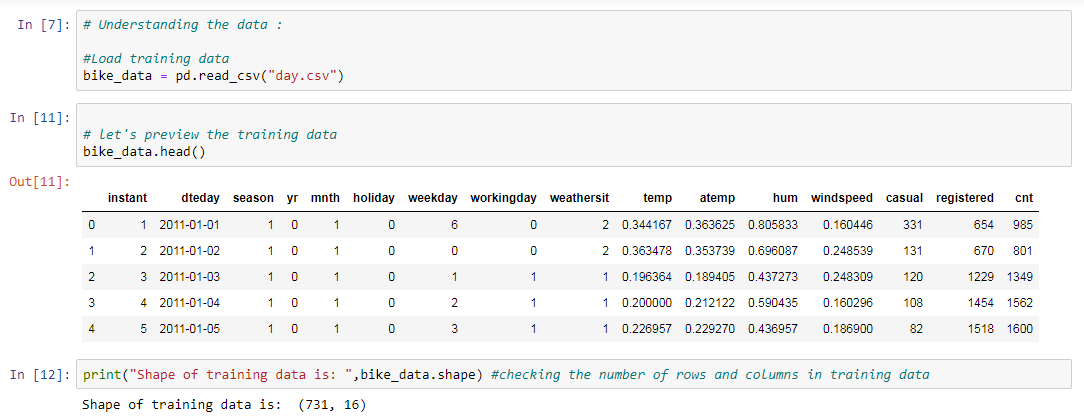
cnt: count of total rental bikes including both casual and registered





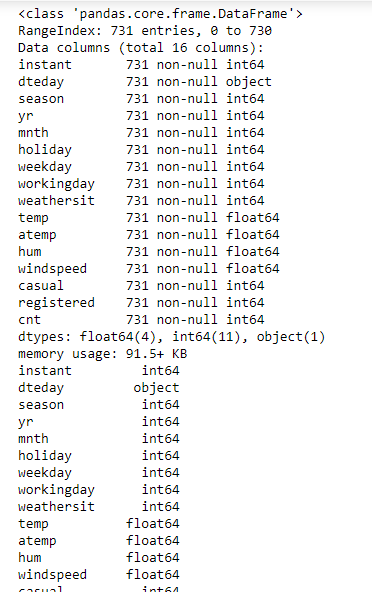


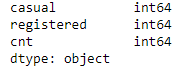
**Understanding the data**

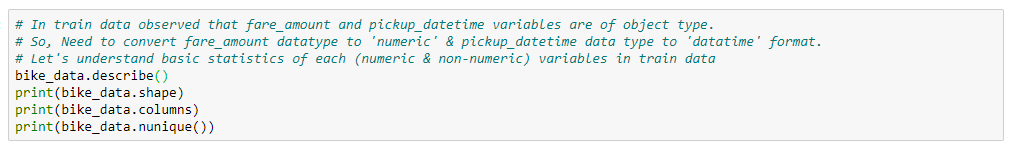


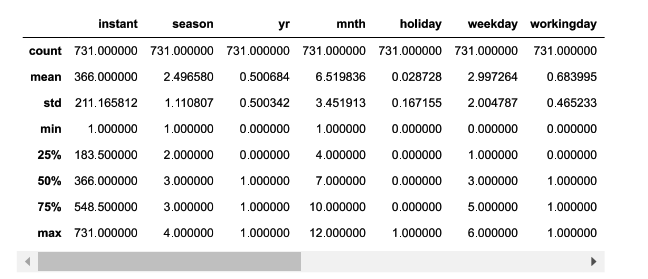
**Explorotary Data Analysis:**

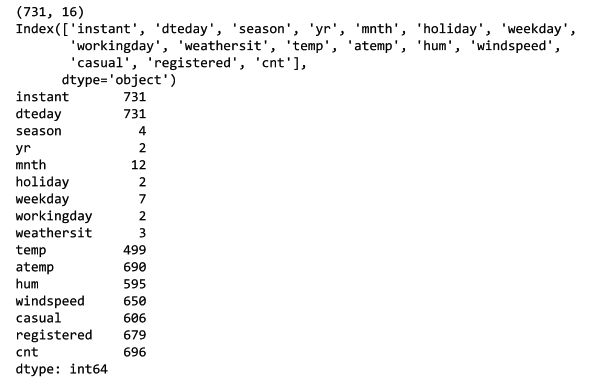


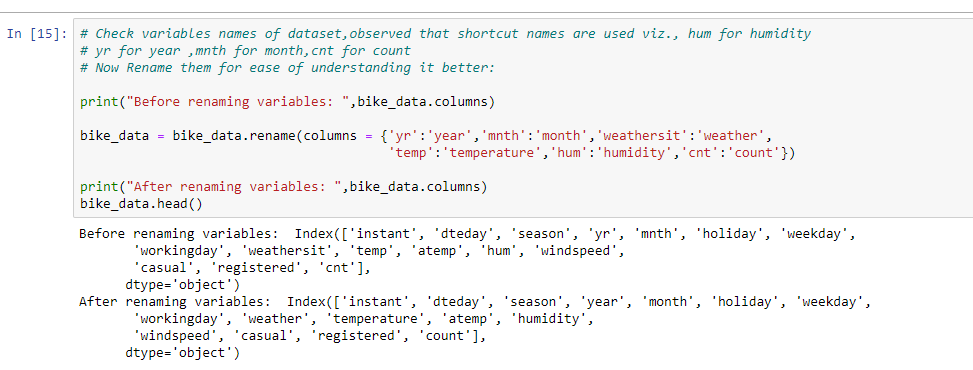


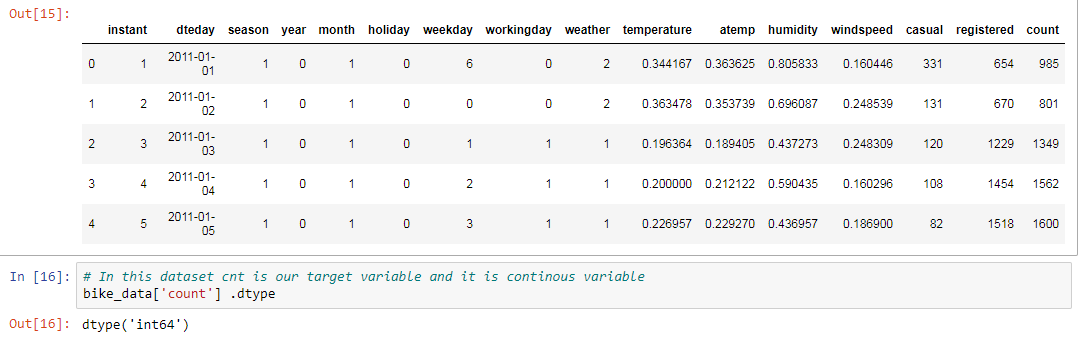




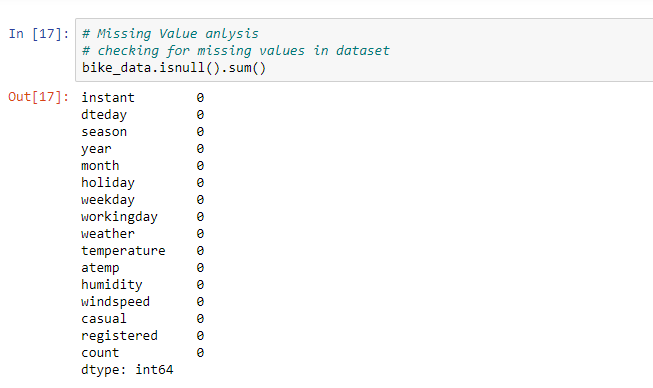




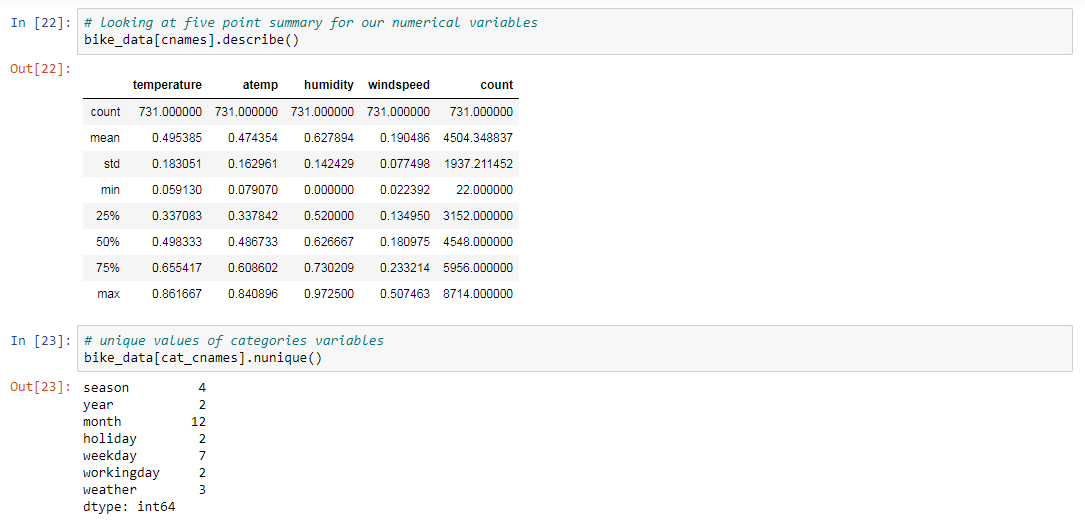




**Missing value Analysis:**

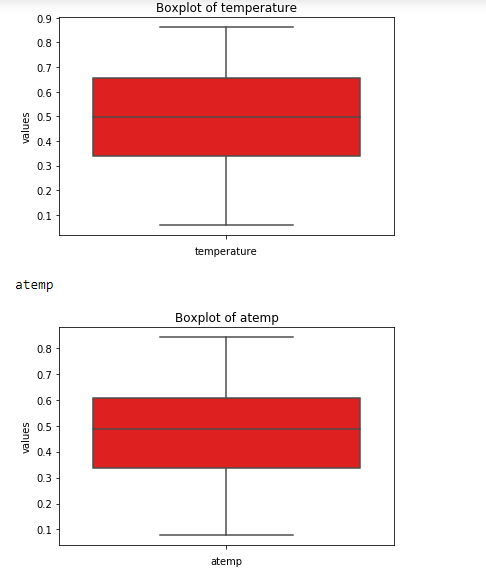


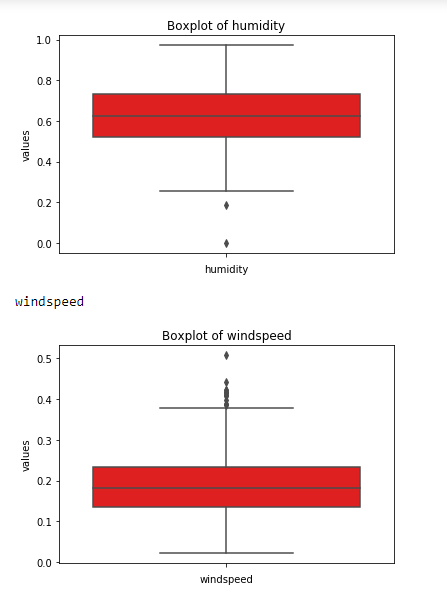


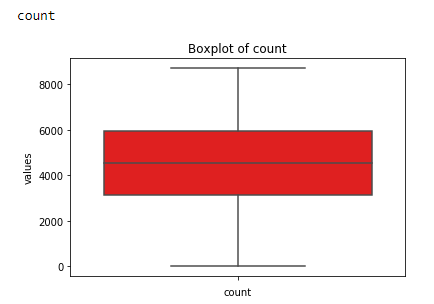


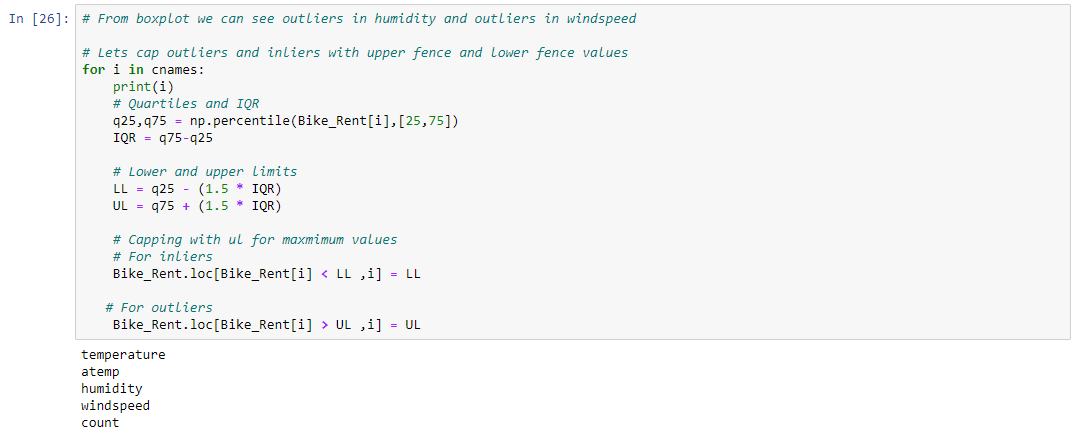
**Outlier Analysis:**

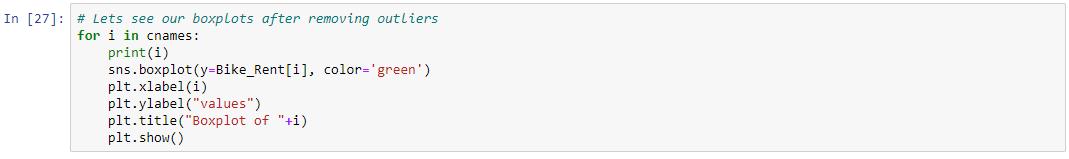


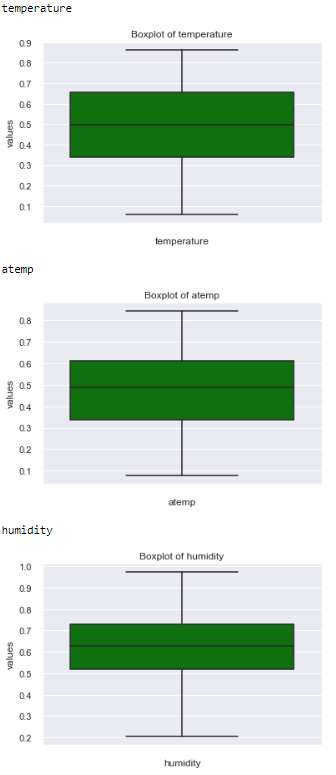


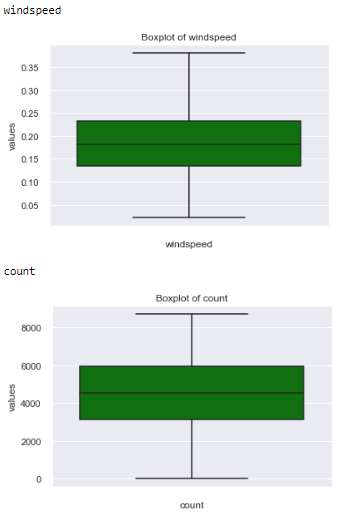












**Data understanding visualization:**

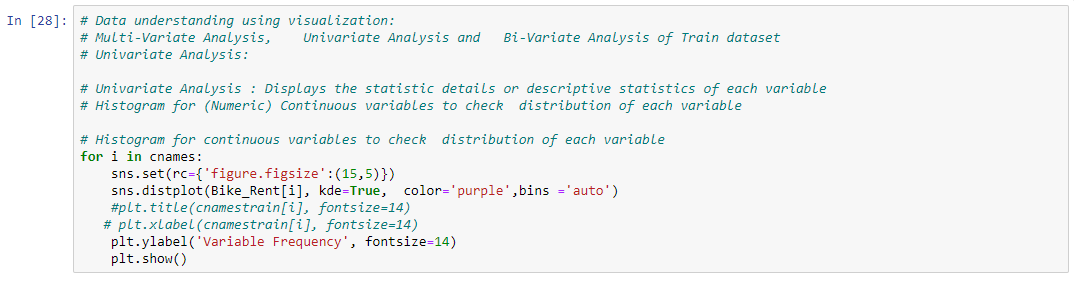
**Data understanding using visualization:**

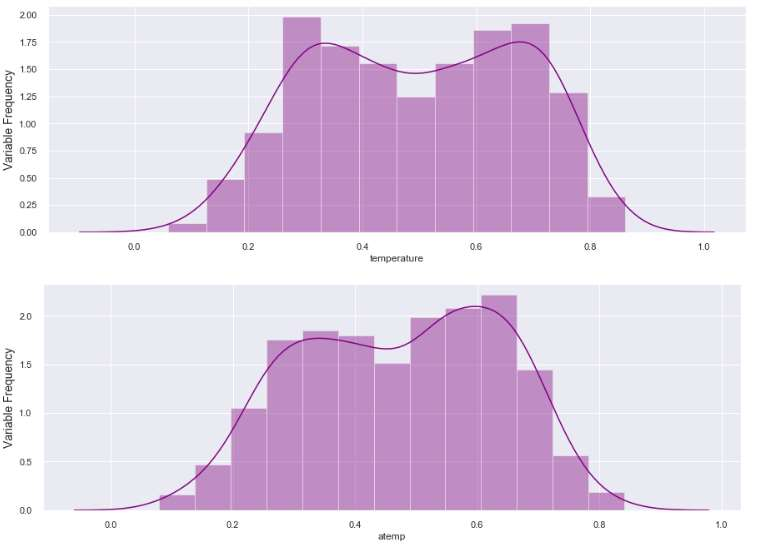
**Multi-Variate Analysis, Univariate Analysis and Bi-Variate Analysis of Train dataset**

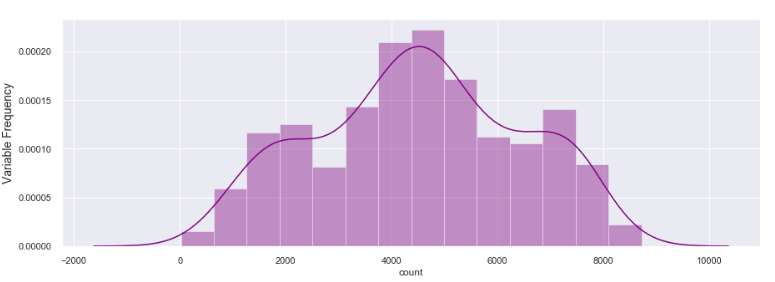
**Univariate Analysis:**

**Univariate Analysis : Displays the statistic details or descriptive statistics of each variable**

**Histogram for (Numeric) Continuous variables to check distribution of each variable**

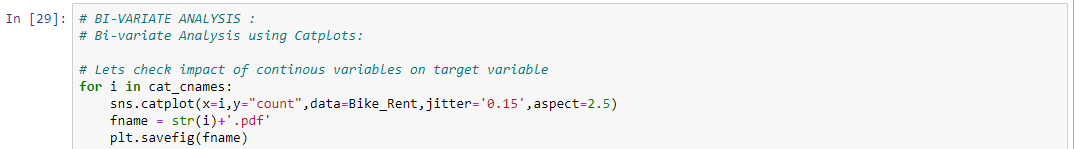


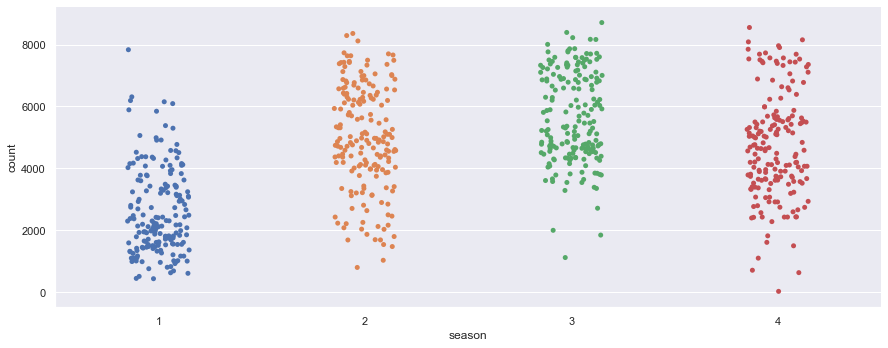


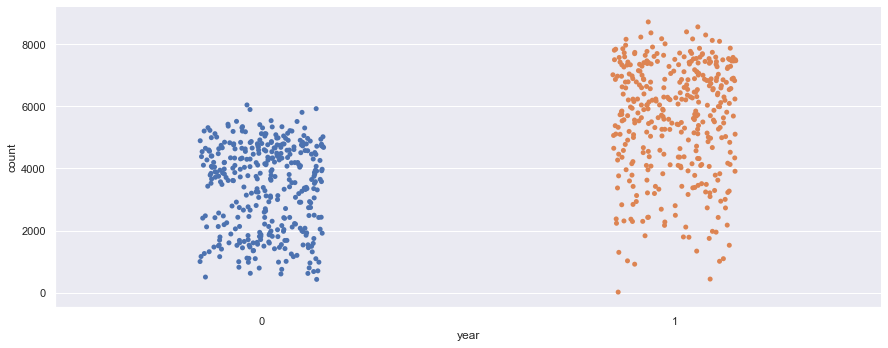


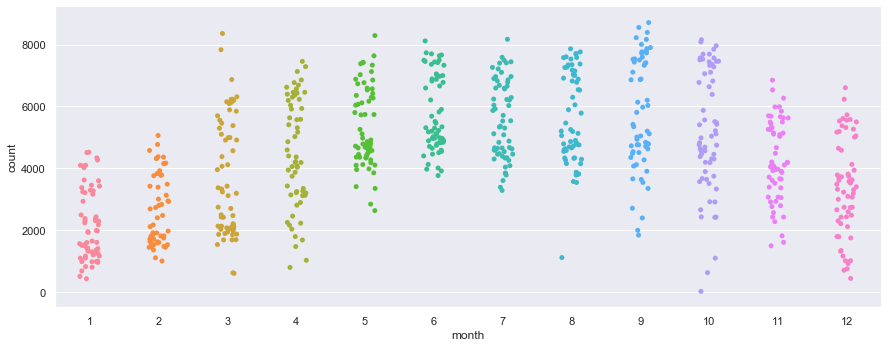
**BI-VARIATE ANALYSIS :**

**Bi-variate Analysis using Catplots:**

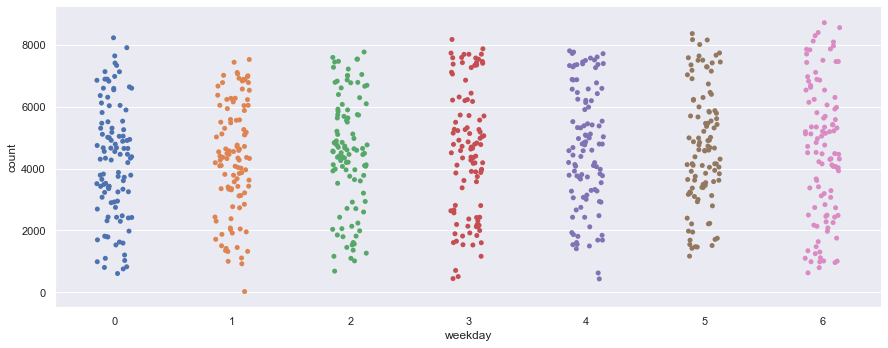


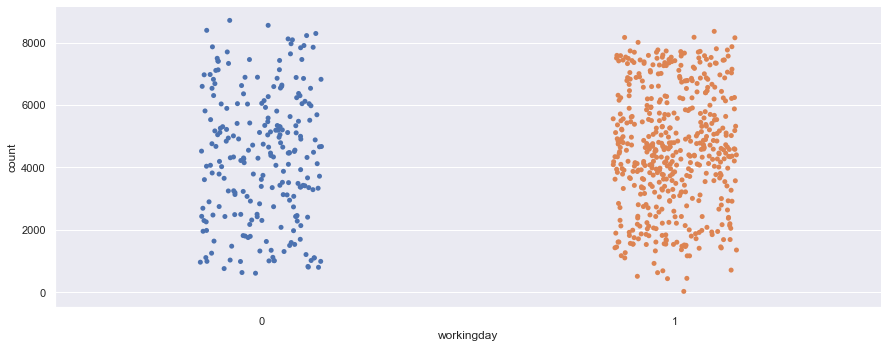
****

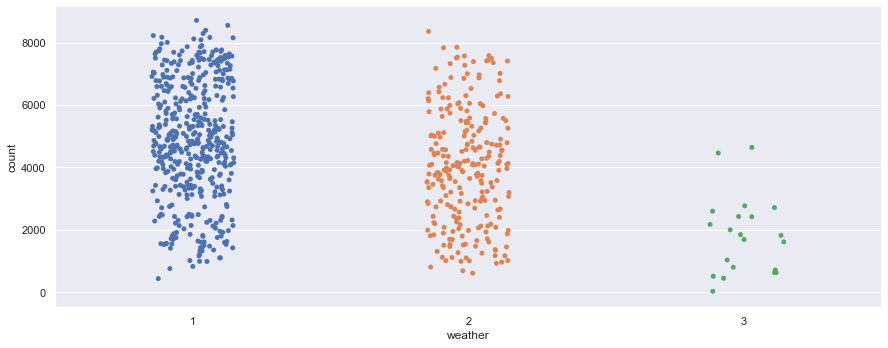
****

****

****

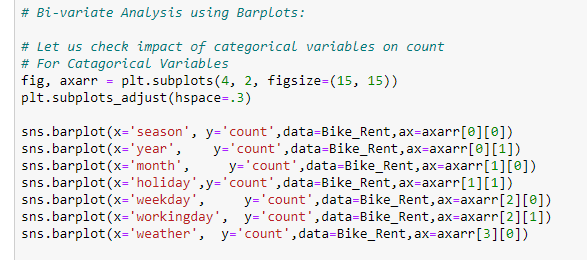
****

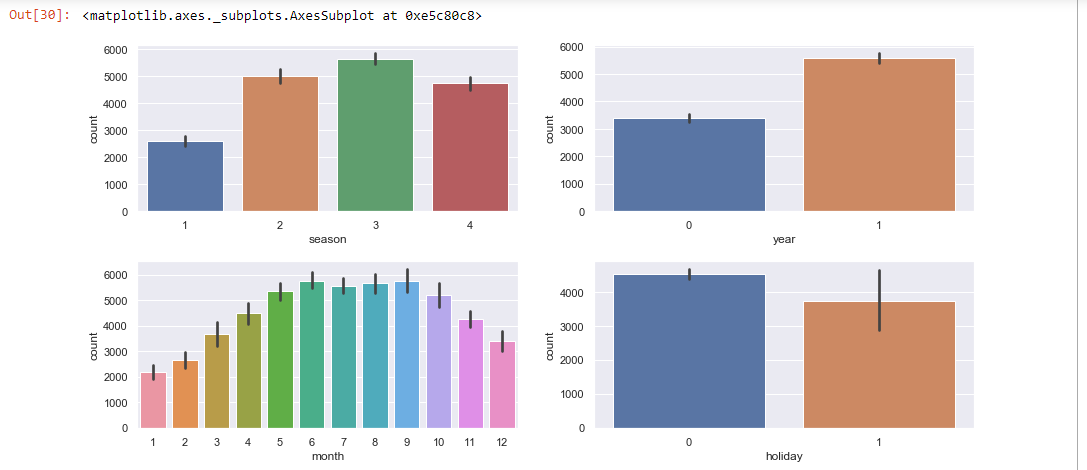
****

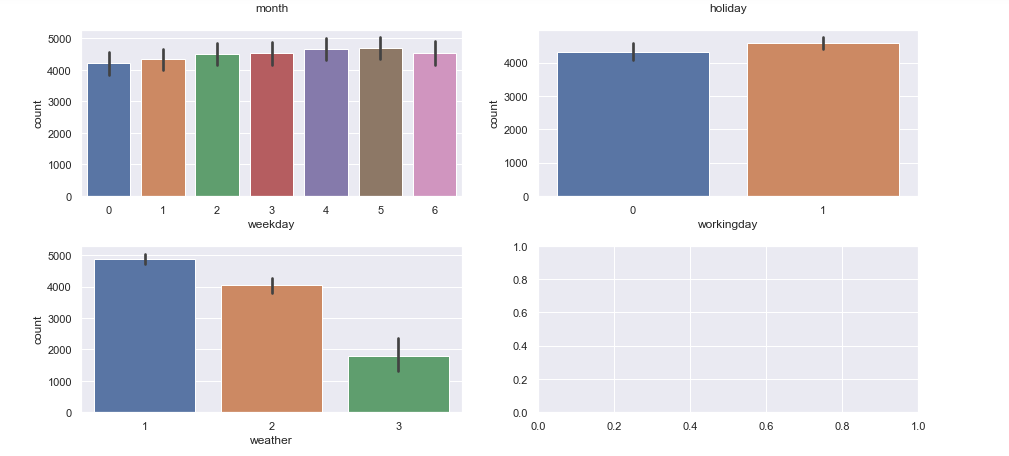
****

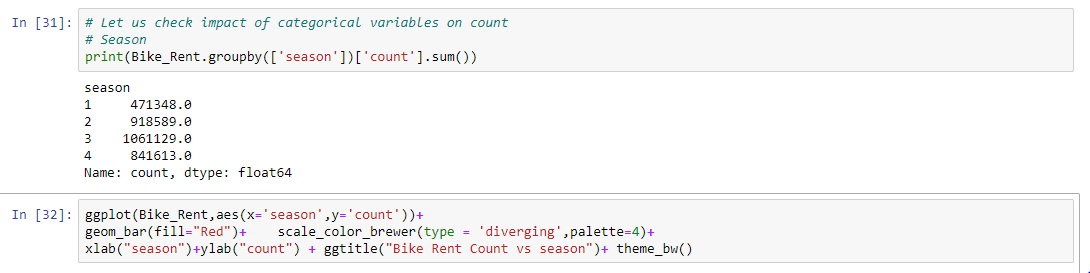
**\*From Fisrt plot we can see that season 2,3 and 4 have more bike count as comapre to season 1. the daily bike count for these season was between 4000 to 8000. \*From year plot we can see that bike count is increased in 2012 as compared to 2011. \*From month plot we can see the bike count maximum between 4 to 10 month. \*From holiday the bike count is maximum as comapre to non holiday. \*Bike count is maximum for day 0,5 and 6 as per weekday varaible. \*FOr weather 1 the count of bike is maximum, after that for weather 2.**

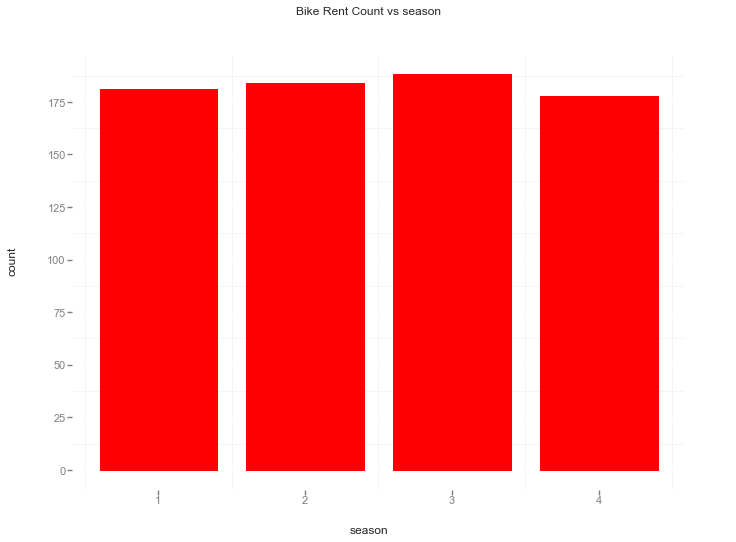
**Bi-variate Analysis using Barplots:**

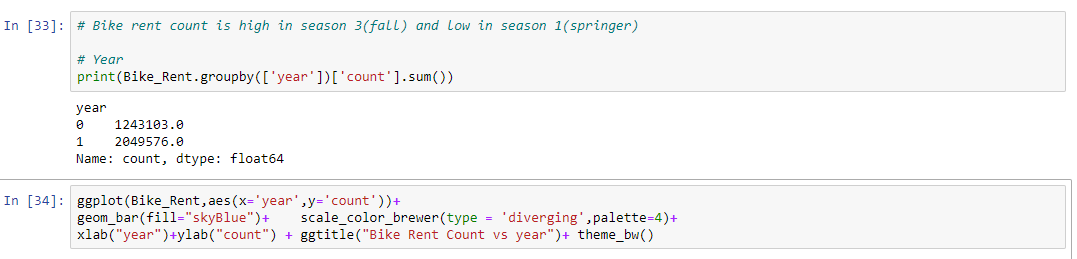


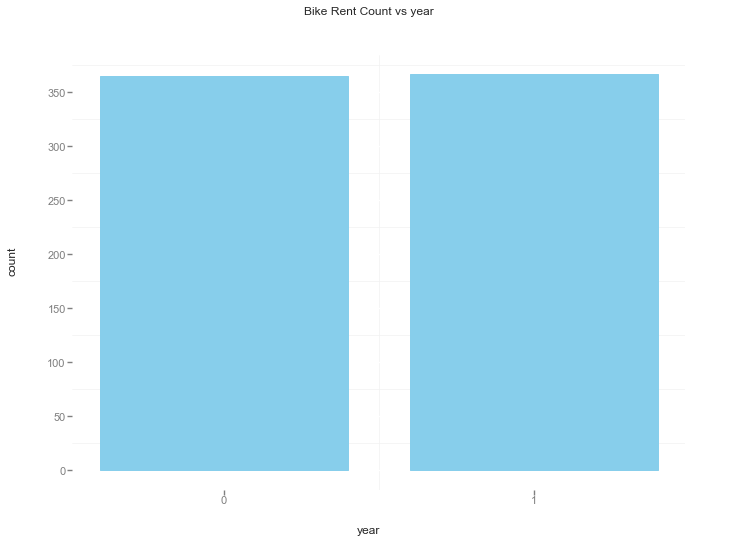




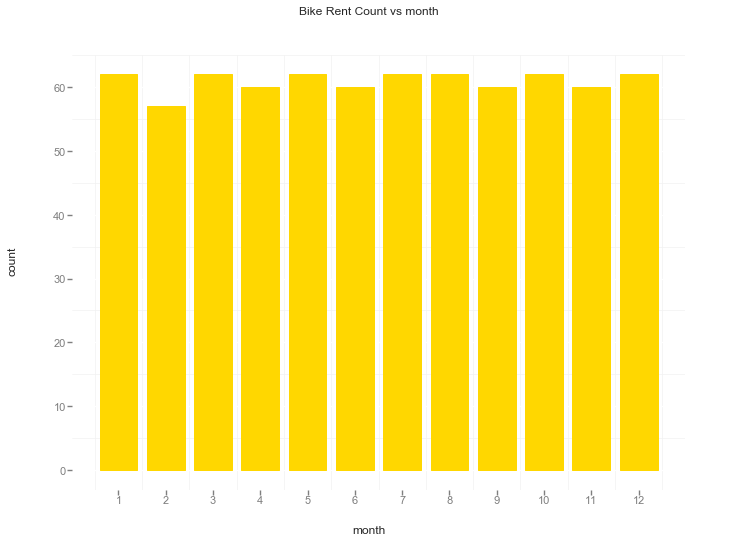


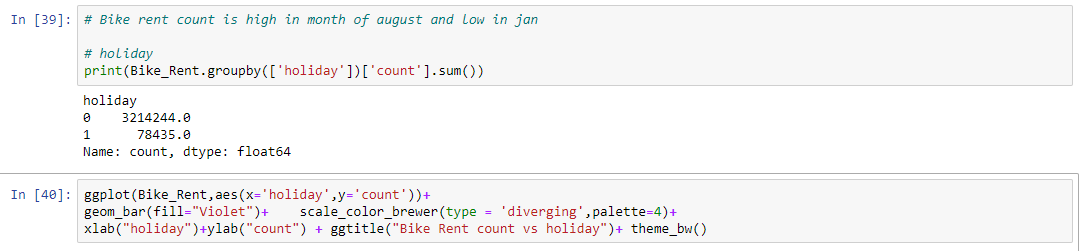
****

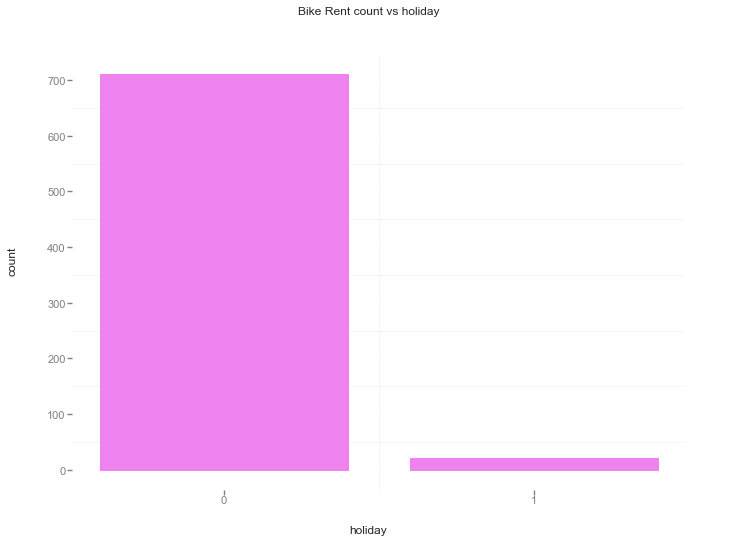


****

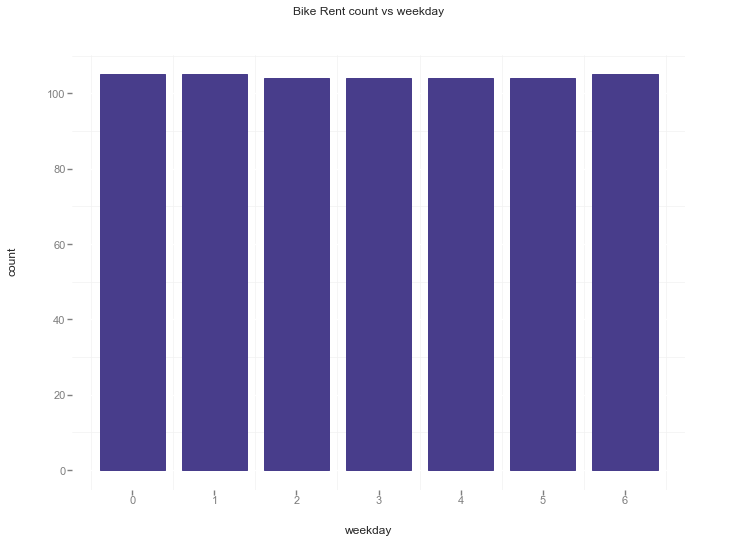


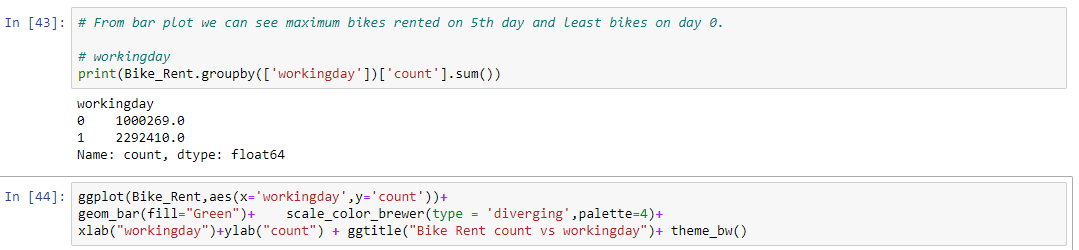
****

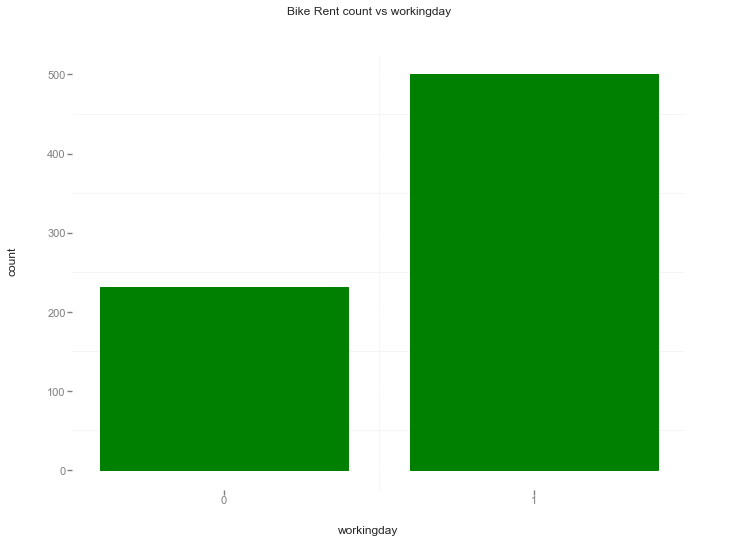


****

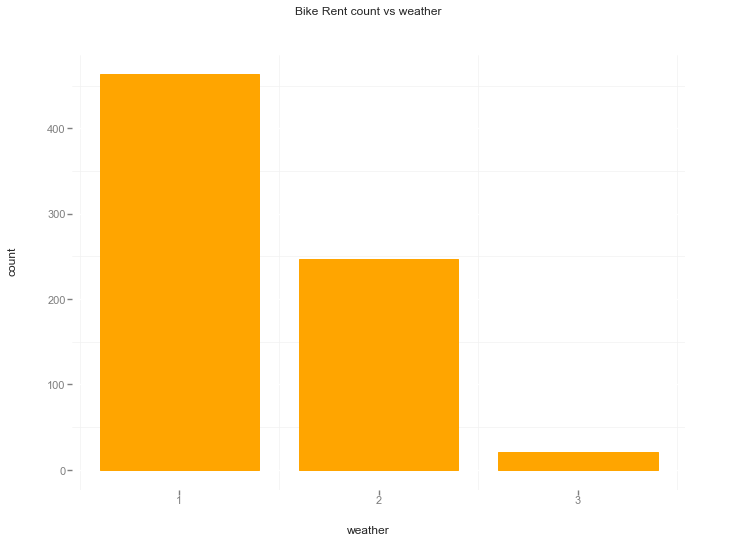


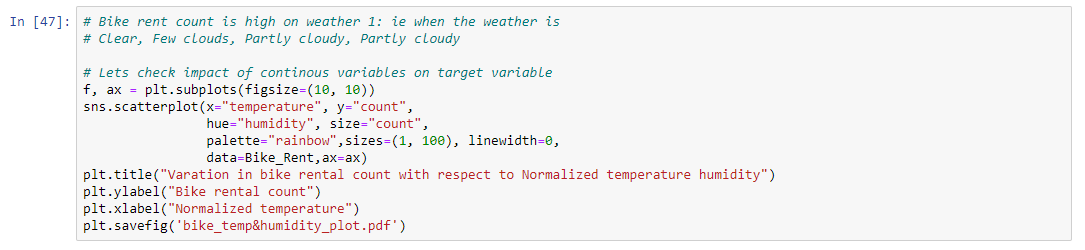
****

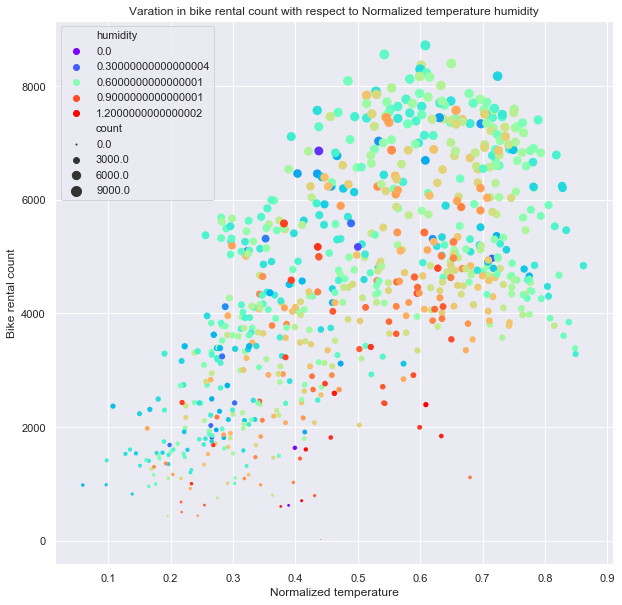


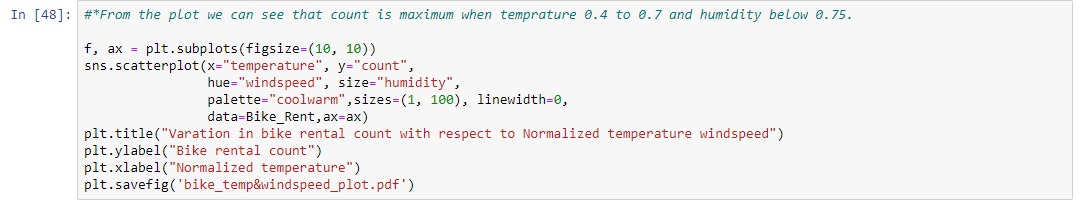
****

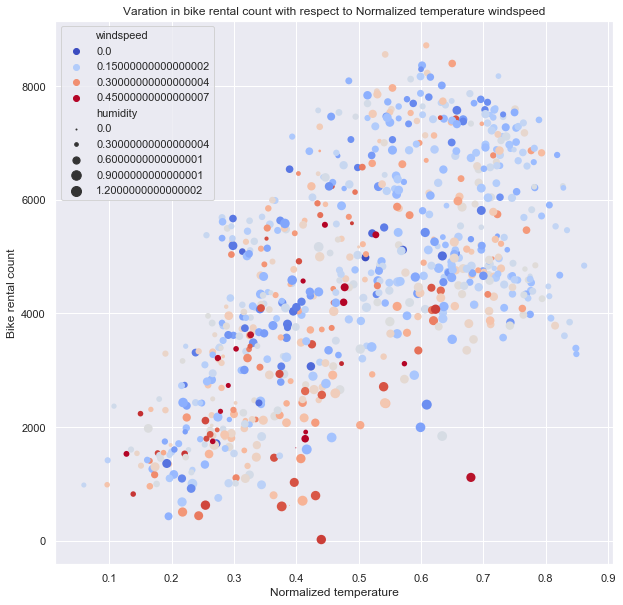


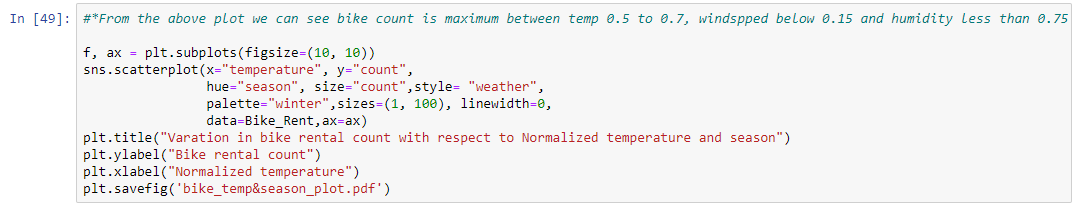
****

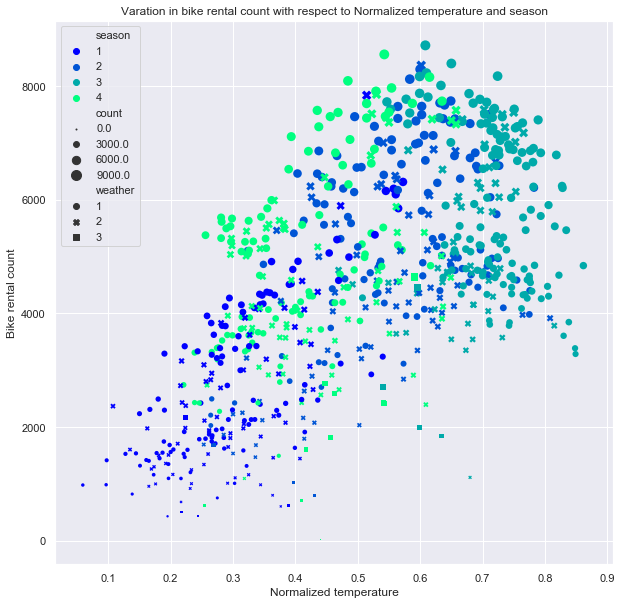


****



****



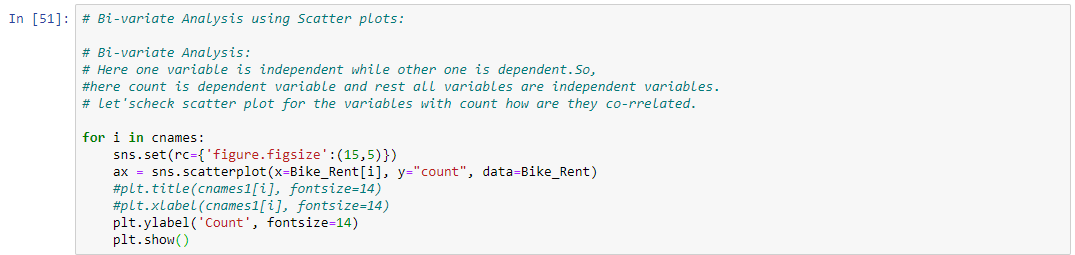
****



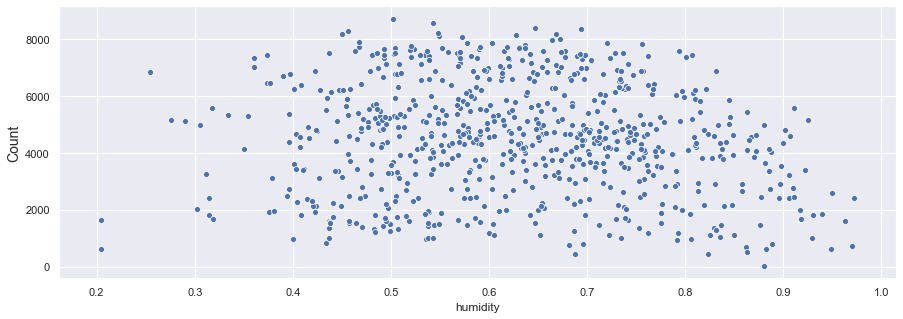
**Bi-variate Analysis using Scatter plots:**

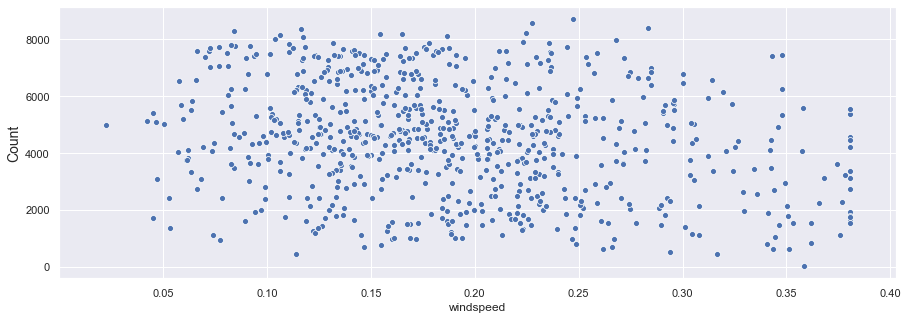
**Bi-variate Analysis:**

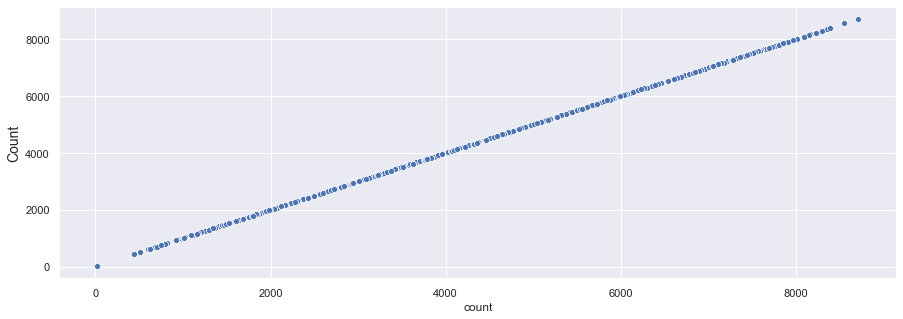
**Here one variable is independent while other one is dependent.So, here count is dependent variable and rest all variables are independent variables. let'scheck scatter plot for the variables with count how are they co-rrelated.**





****

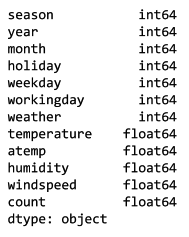
****

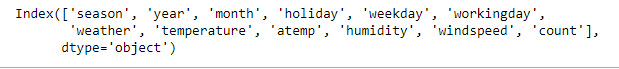
****

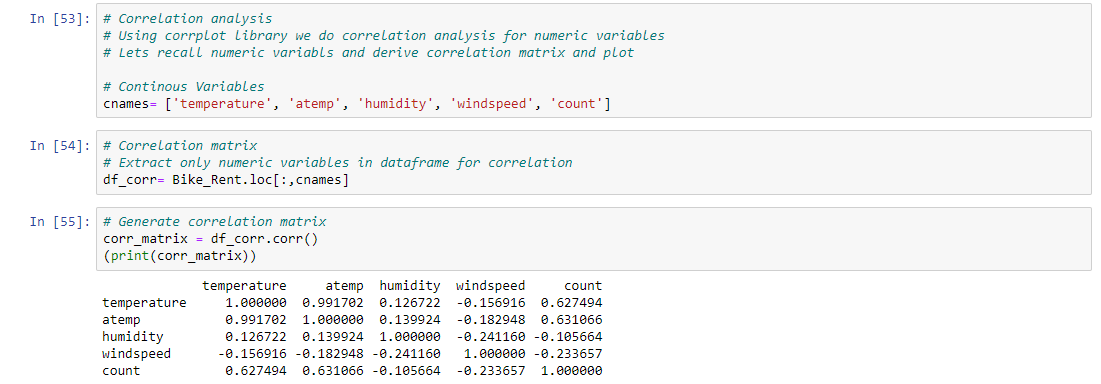
**Feature Selection :**



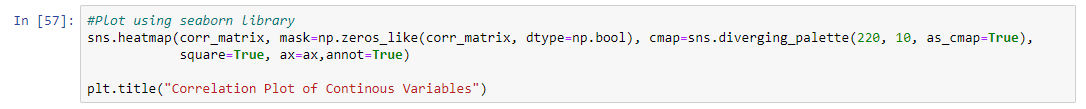


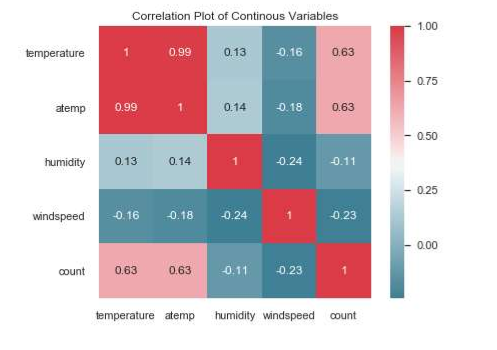








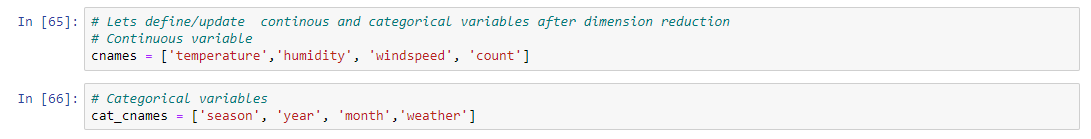




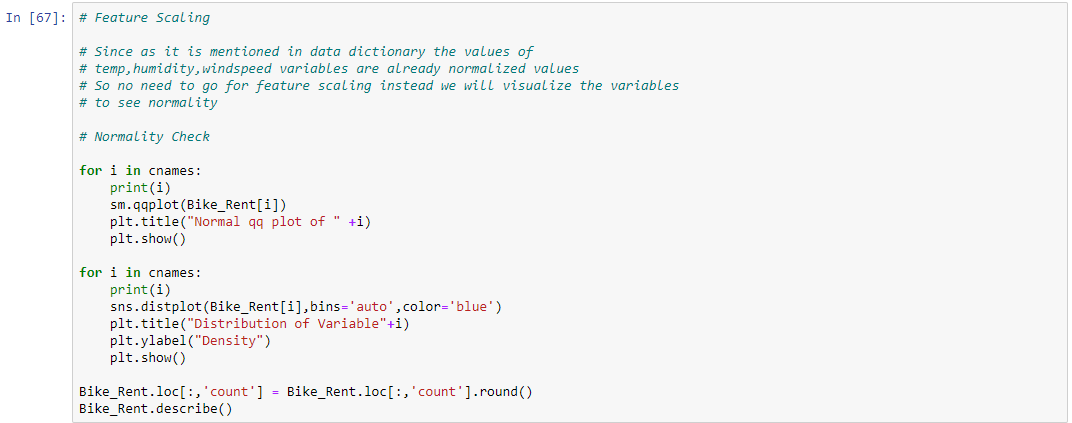


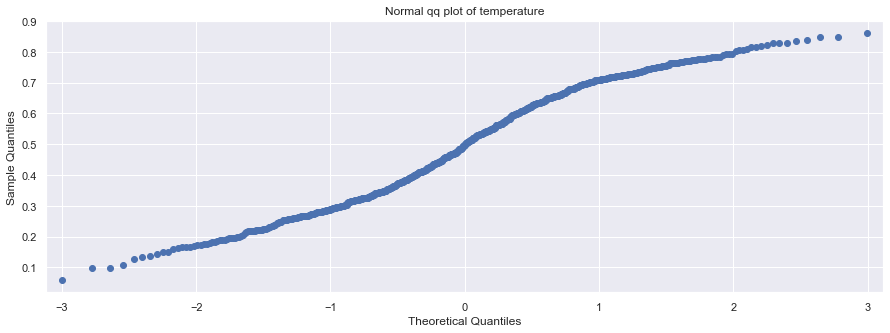


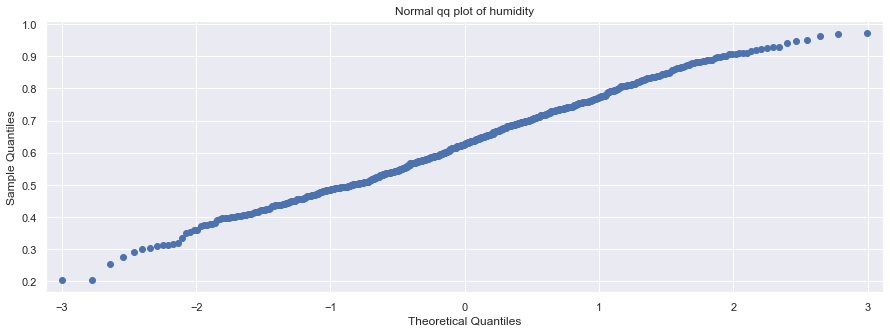


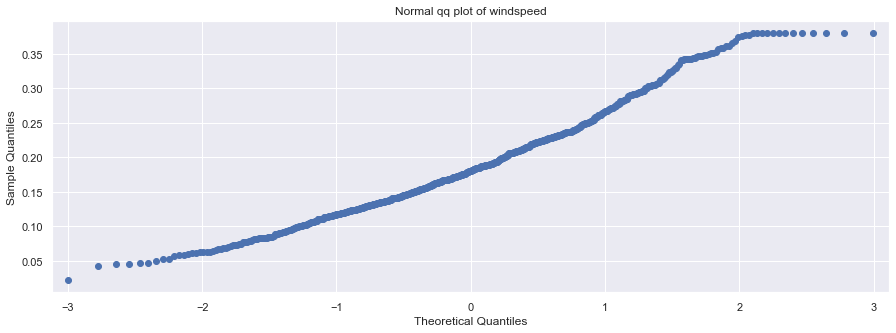


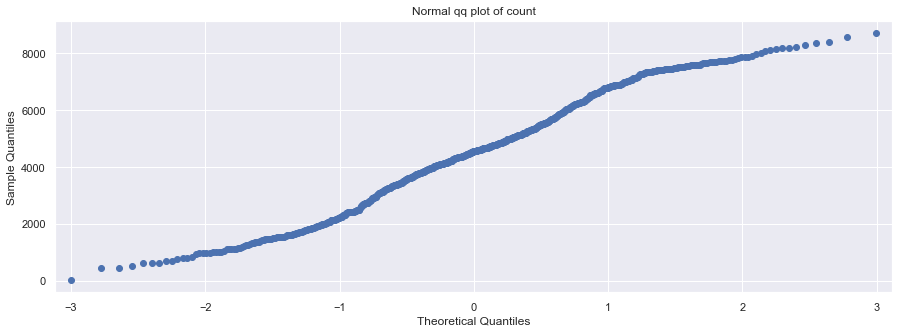
**Feature Scaling:**

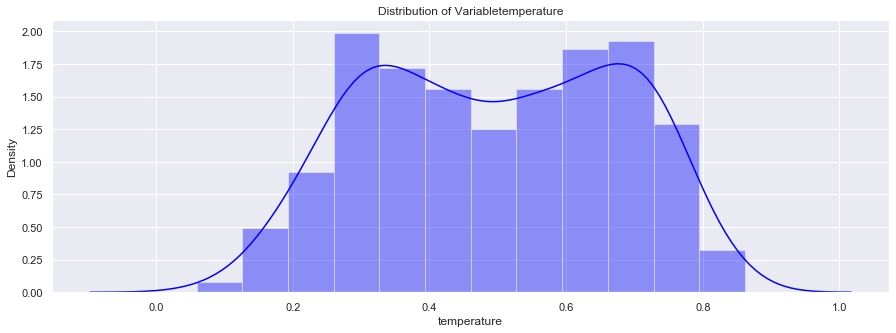


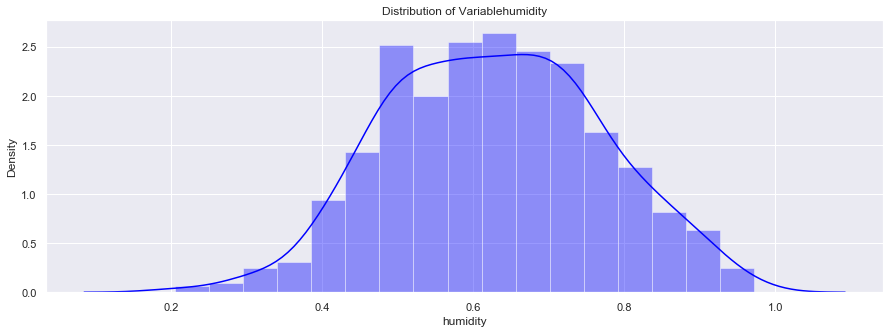
****

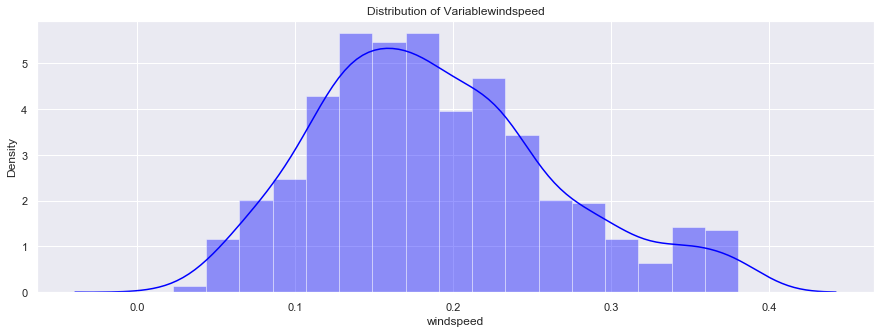
****

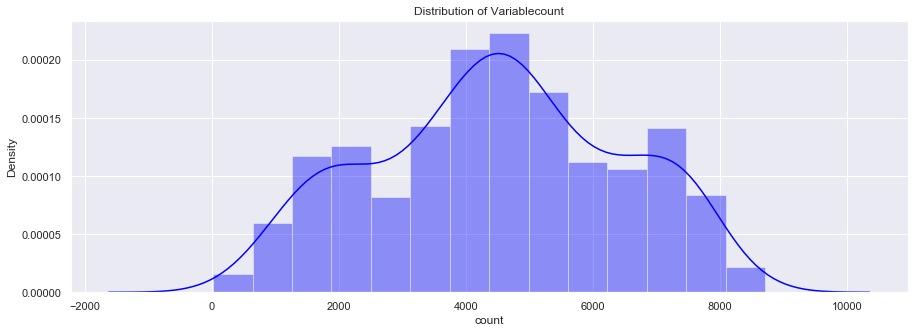
****

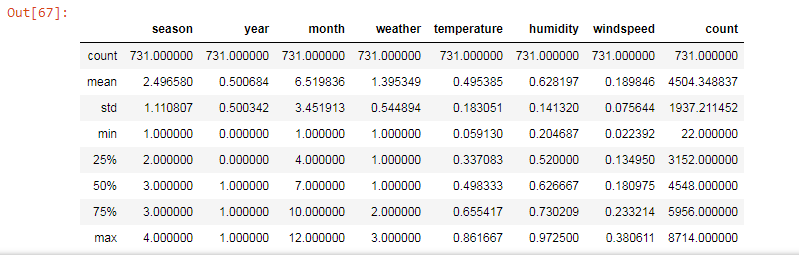
****

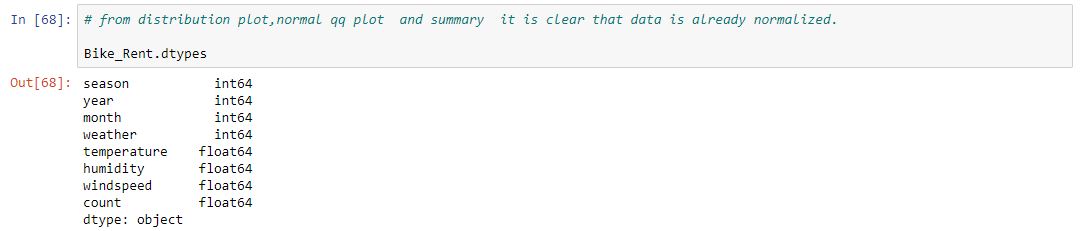
****

****

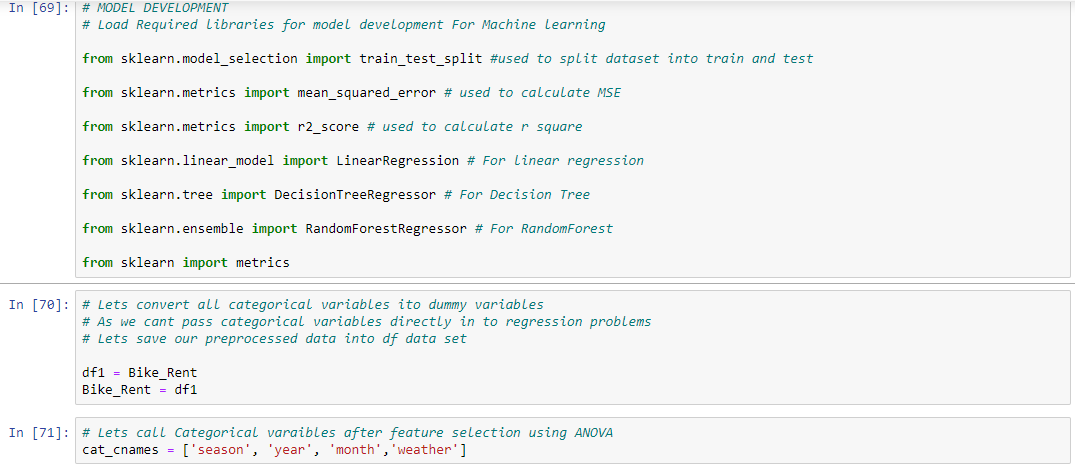
****

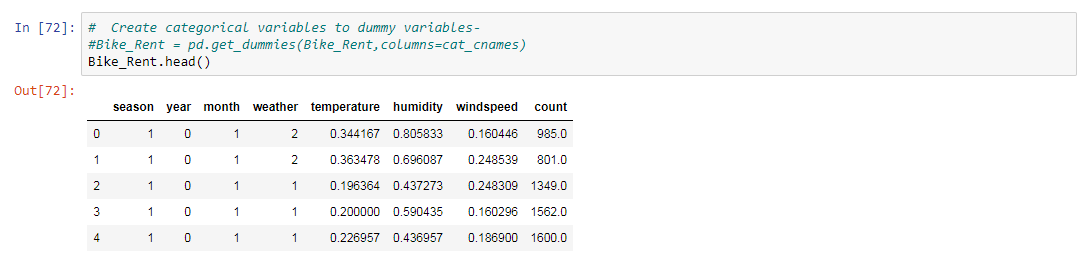
****

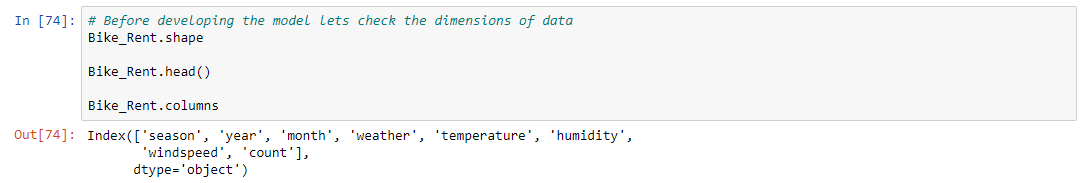


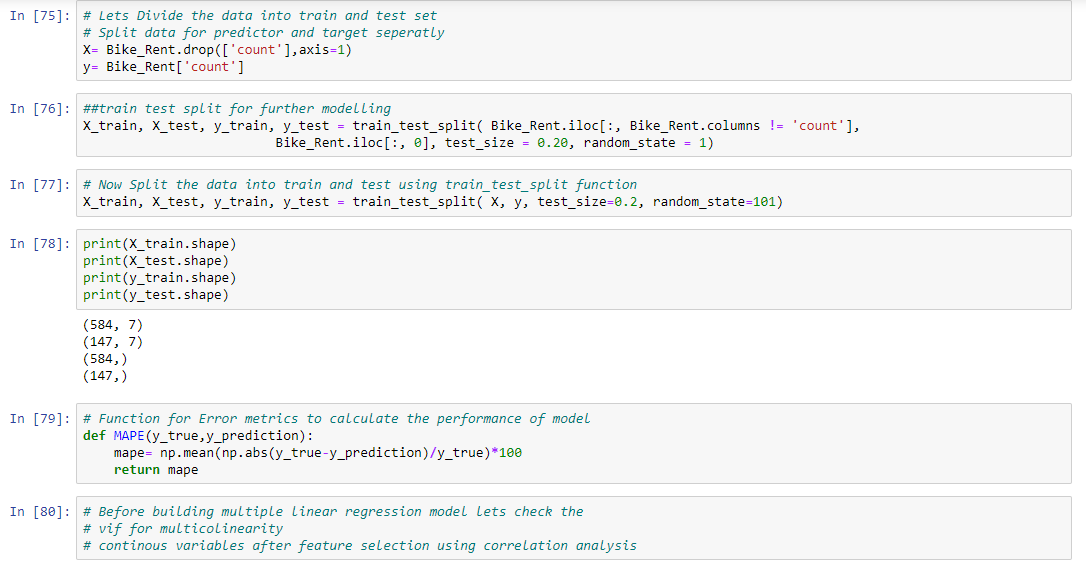


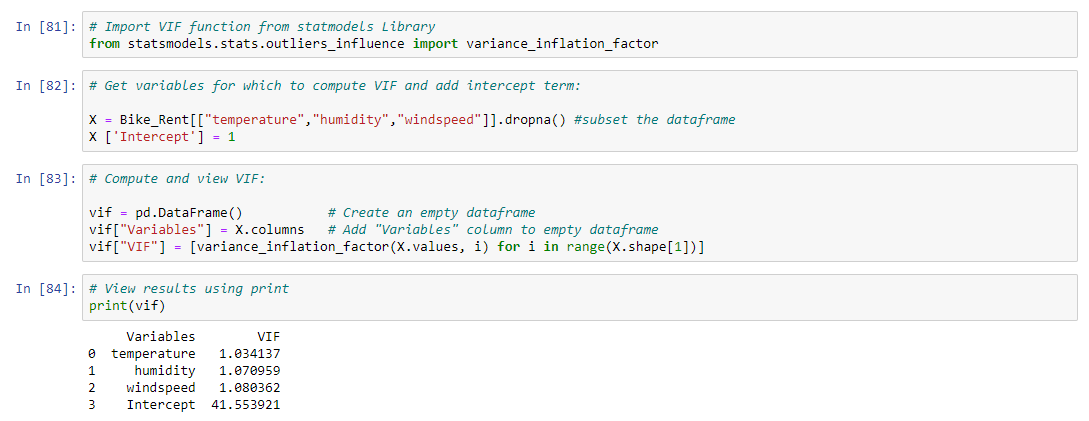
**MODEL DEVELOPMENT:**





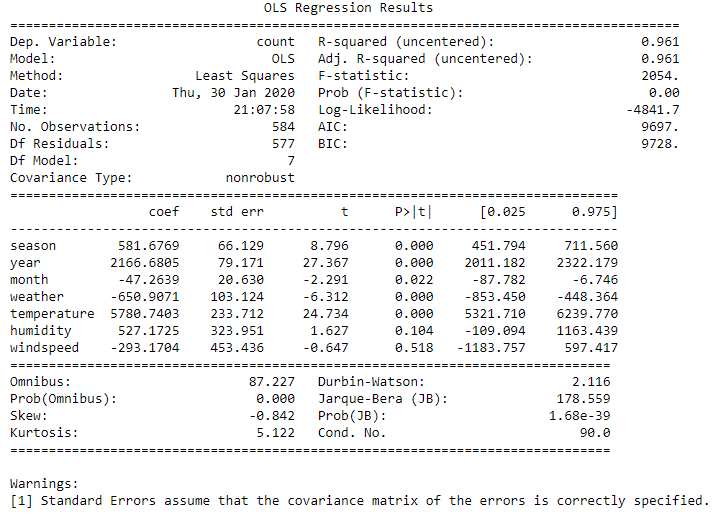


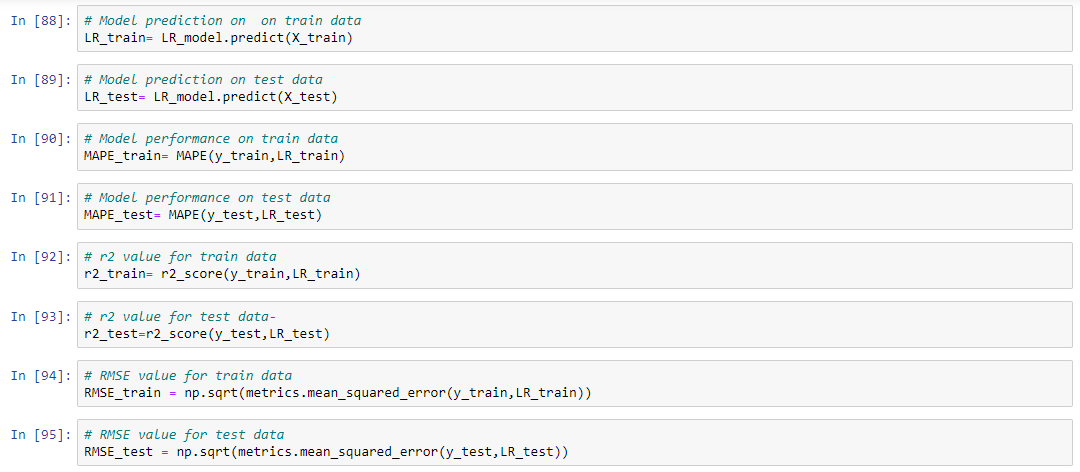


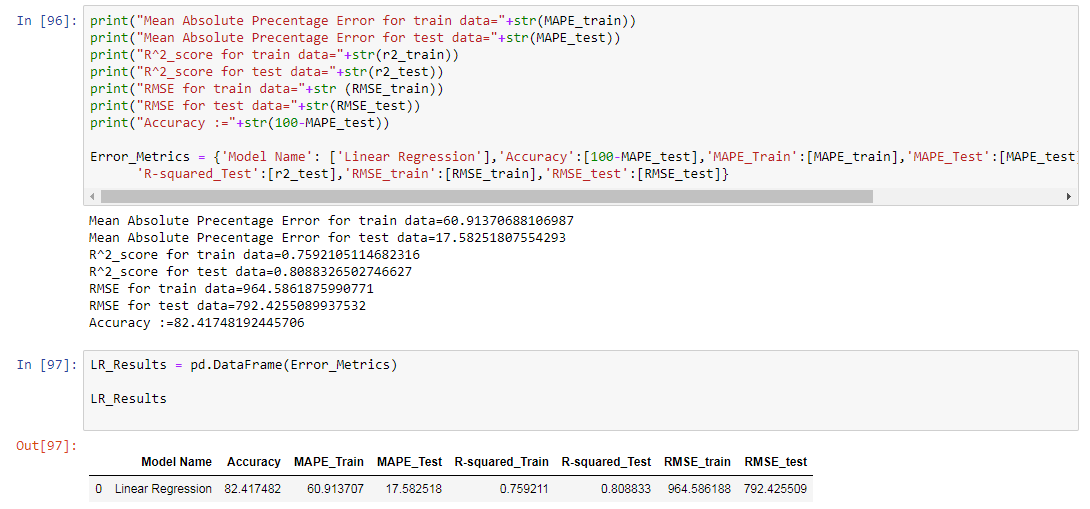


**Linear Regression Model :**



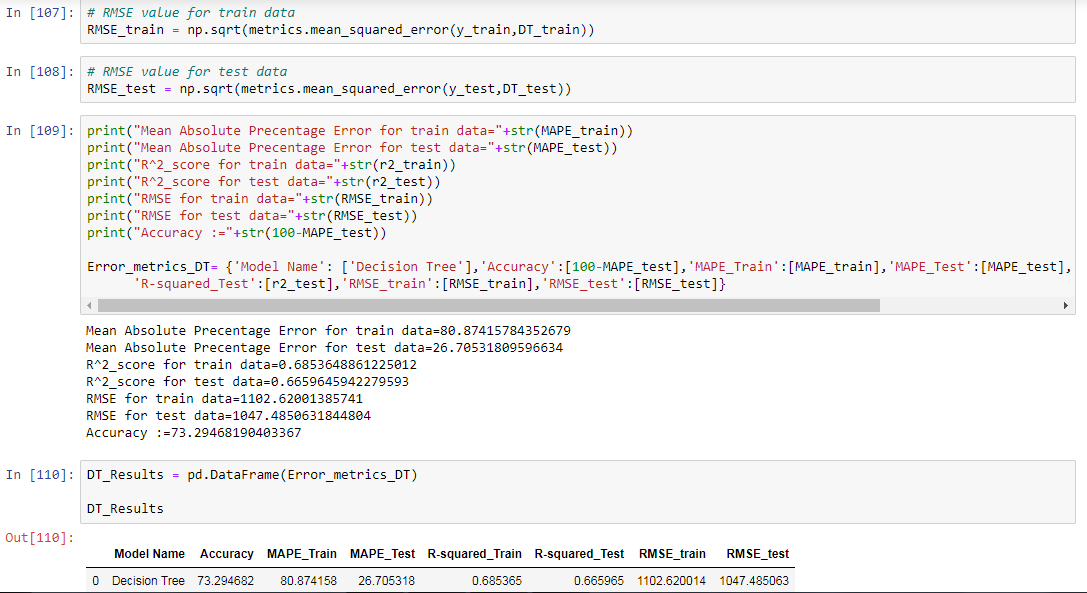






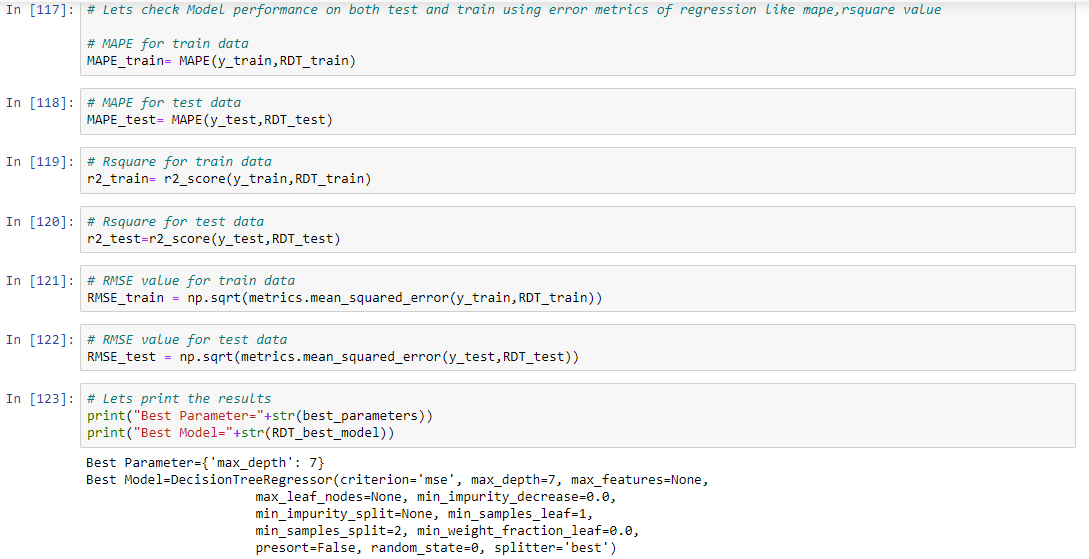
**Decision Tree:**

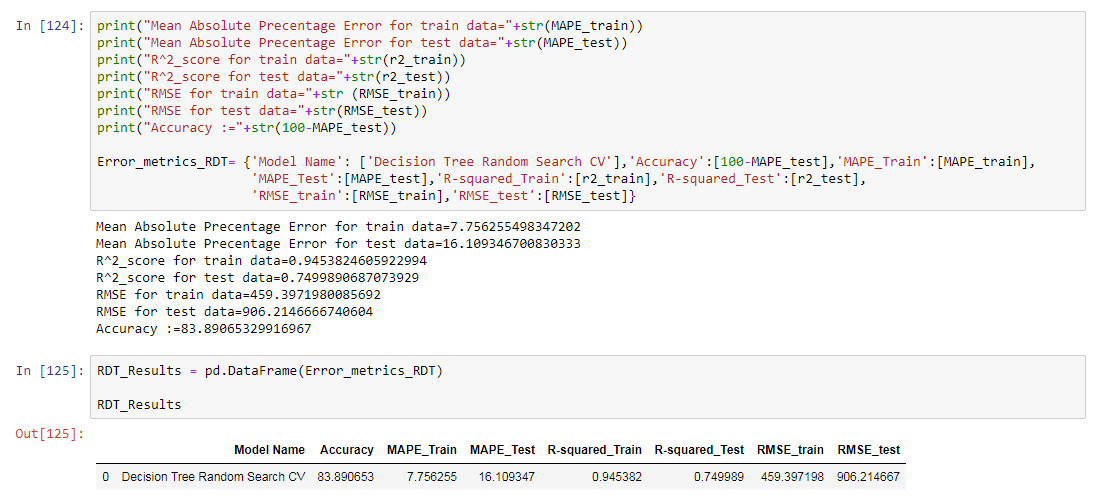




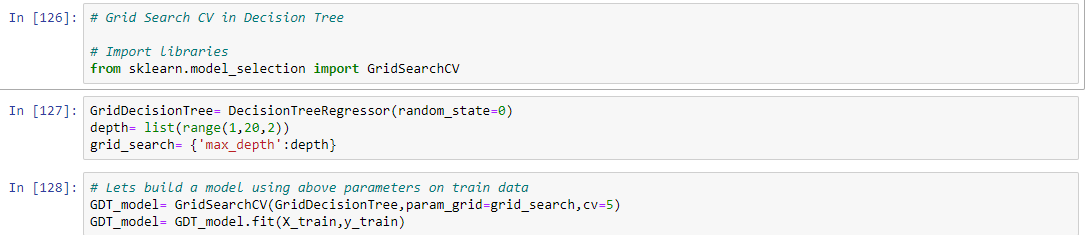
**Random Search CV In Decision Tree:**



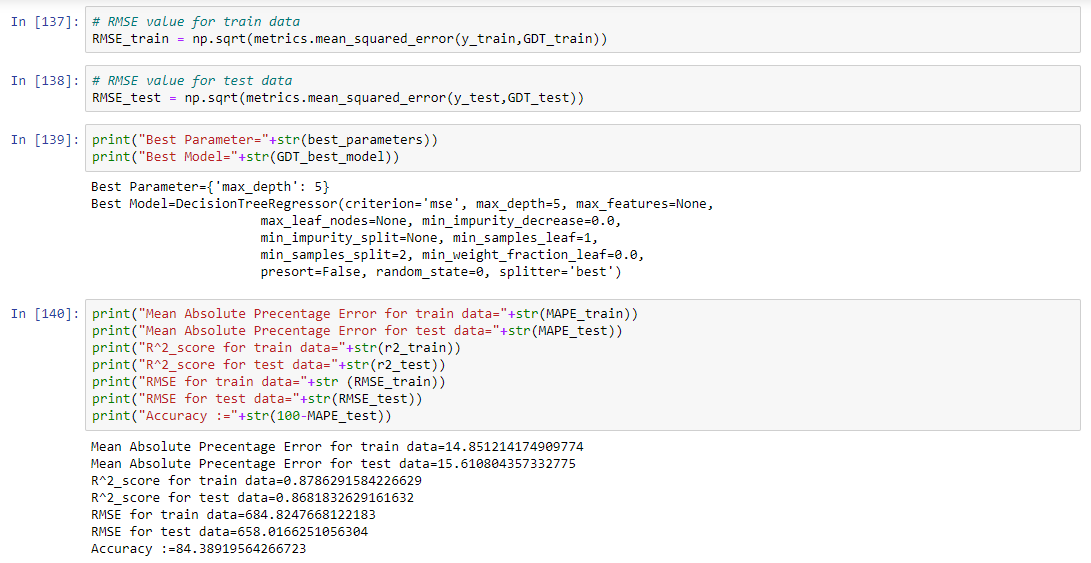


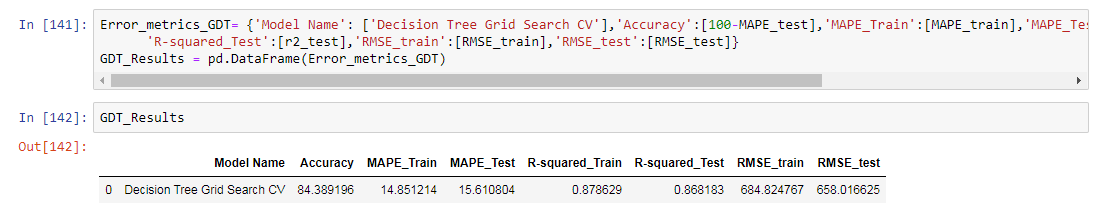


**Grid Search CV in Decision Tree:**



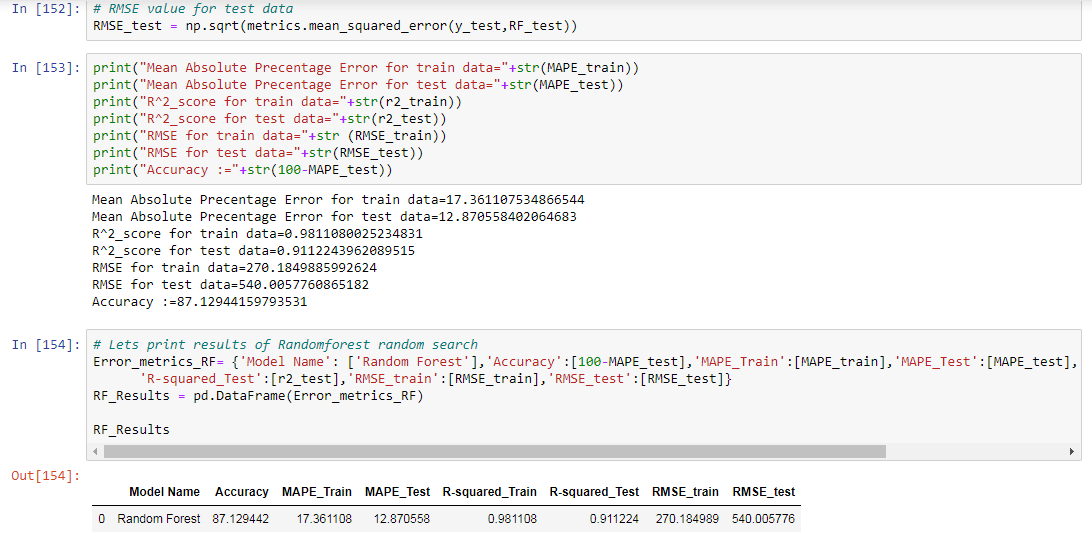






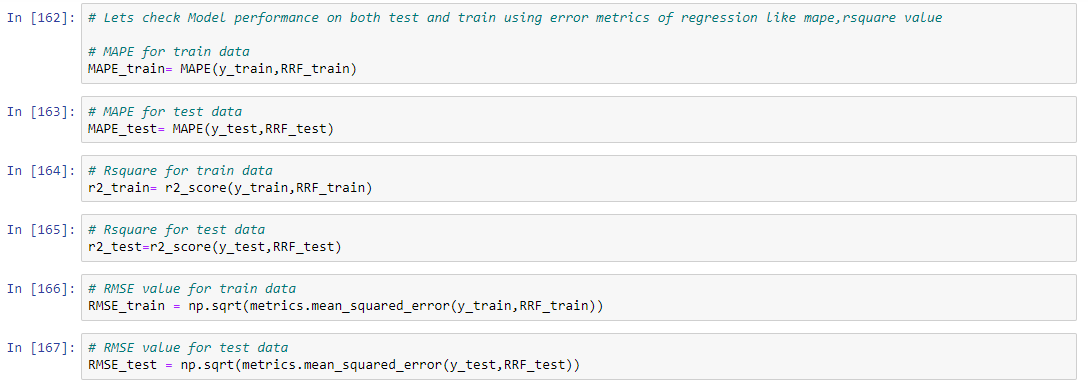
**Random Forest:**

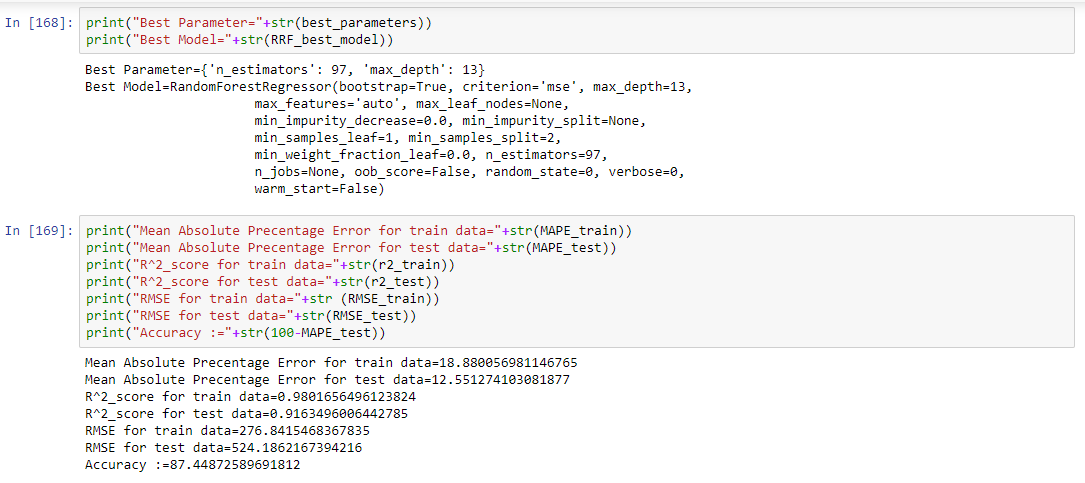


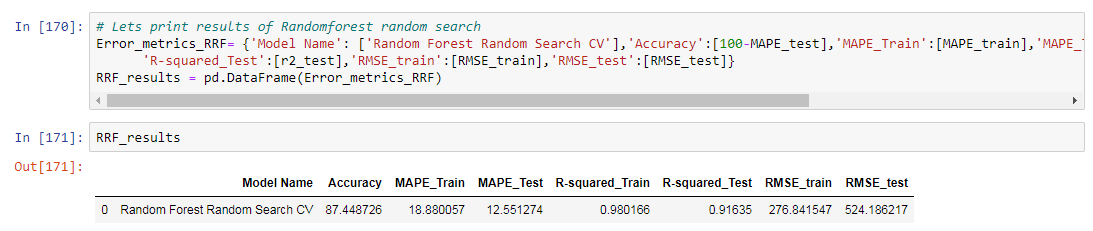


**Random Search CV in Random Forest:**

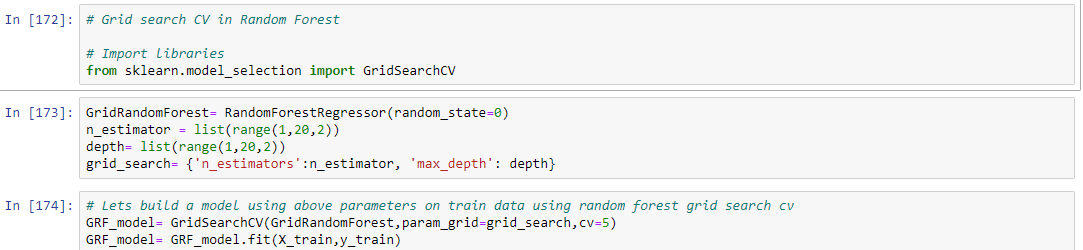


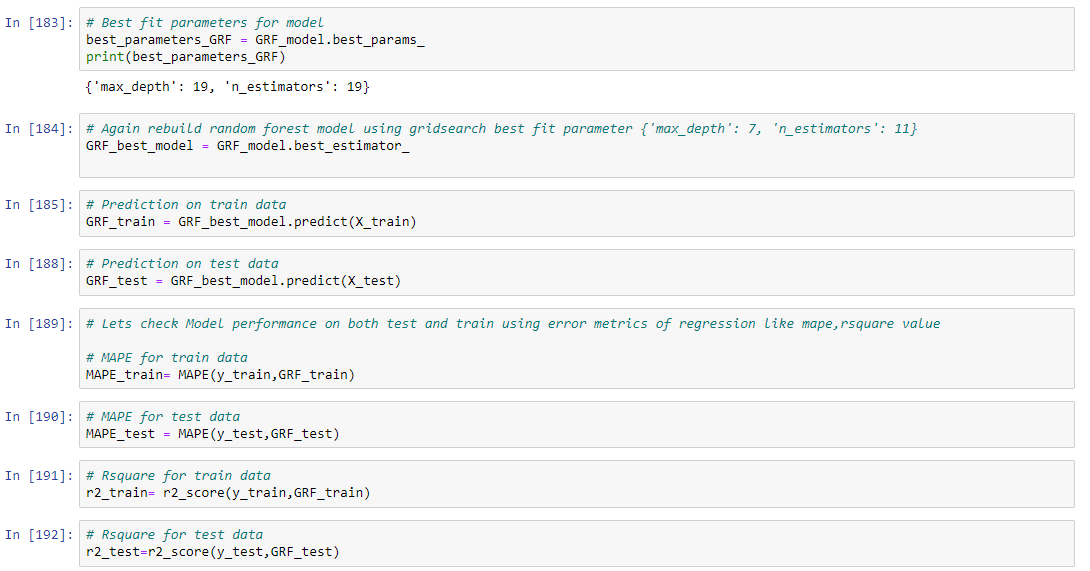


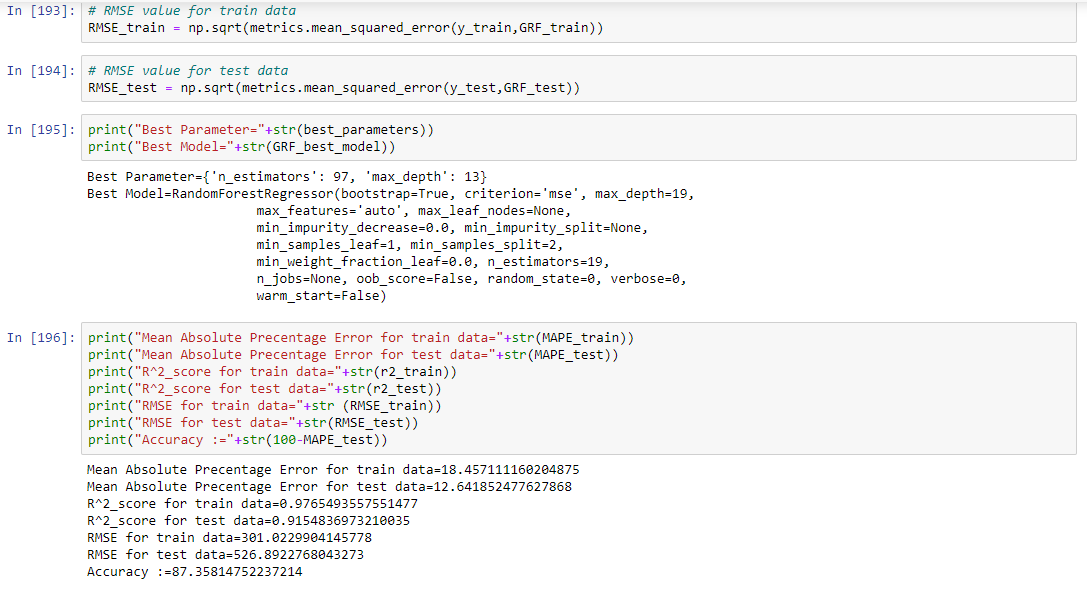


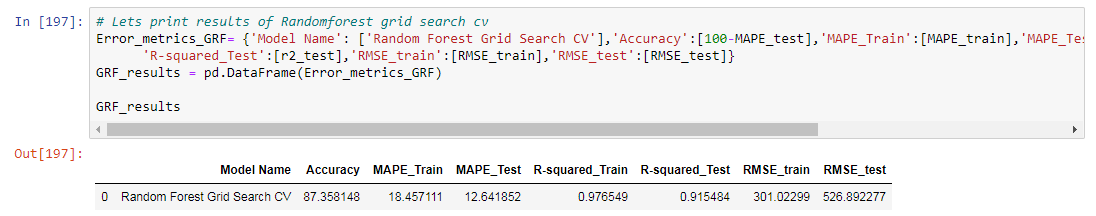


**Grid search CV in Random Forest:**



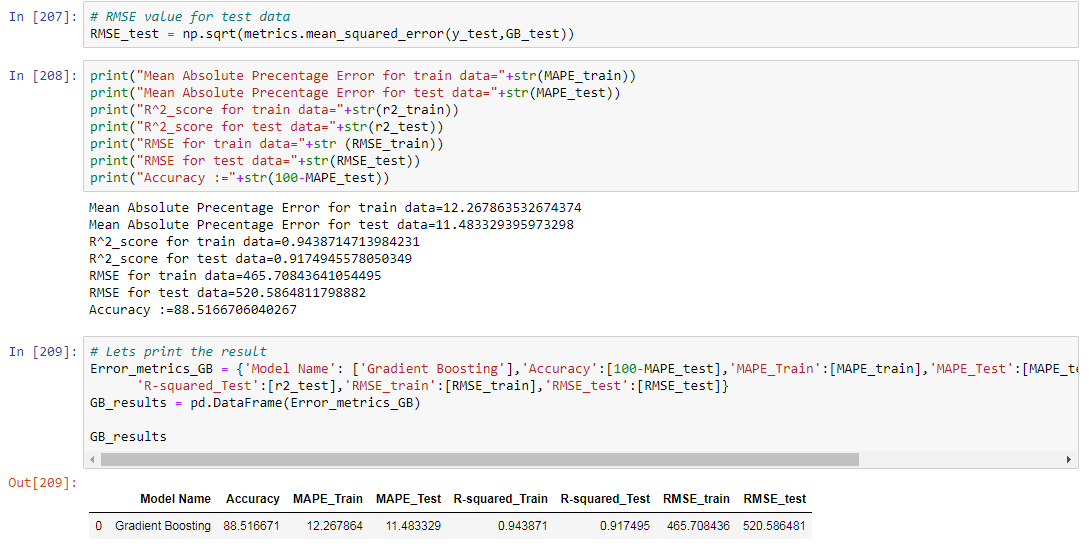






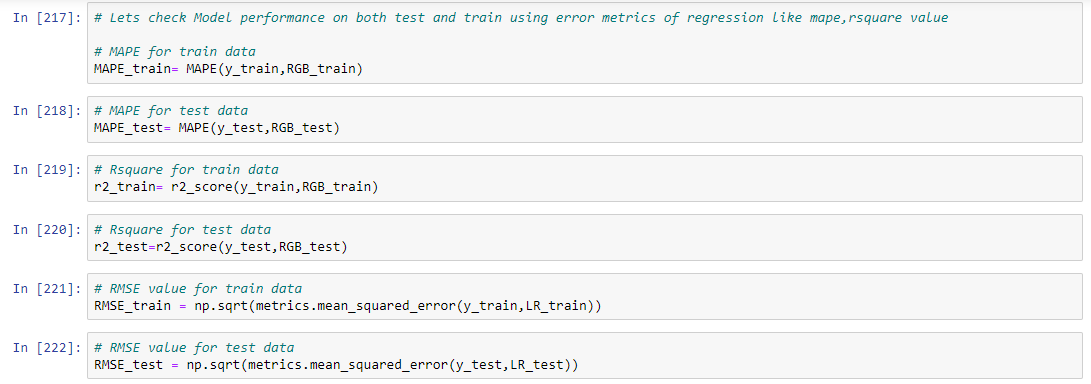
**Gradient Boosting:**



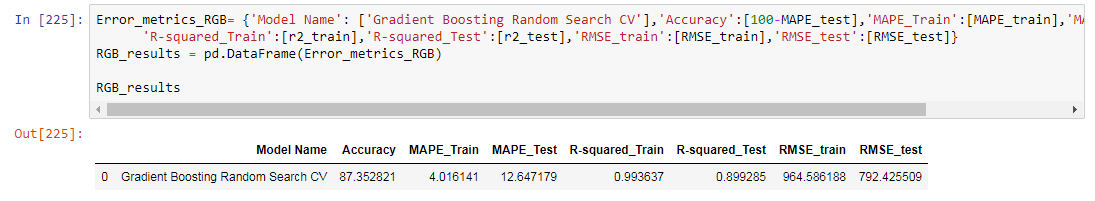


**Random Search CV in Gradient Boosting:**









**Grid Search CV in Gradient Boosting:**



