**Bike Rental Prediction**

As this is a prediction task over time, we have collected time series data which consist of various Categorical and Continuous variable mentioned below:

**Categorical Data:**

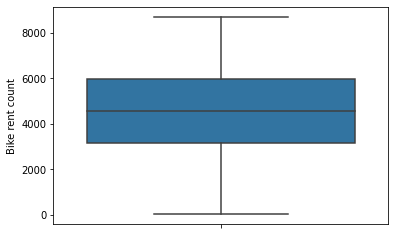
1. Season
2. Year
3. Month
4. Holiday
5. Weekday
6. Working day
7. Weather condition

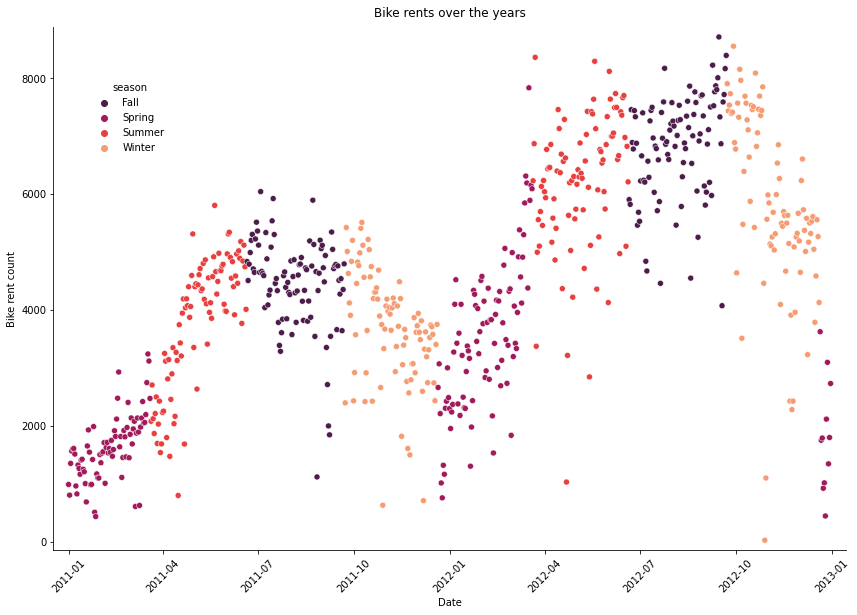
**Continuous Data:**

1. Normalized temperature in celsius
2. Normalized feeling temperature in celsius
3. Normalized humidity.
4. Normalized wind speed.
5. Count (Dependent variable)

**Exploratory Data Analysis:**

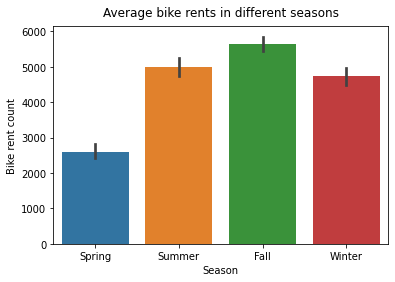
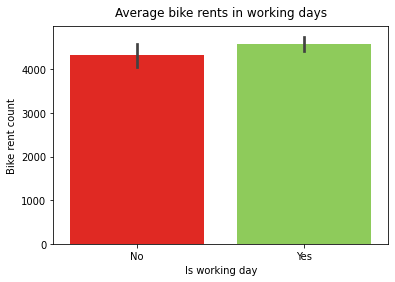
After visualizing our dependent variable “Count” we can see in the plot there is no outlier for the count variable in our dataset.

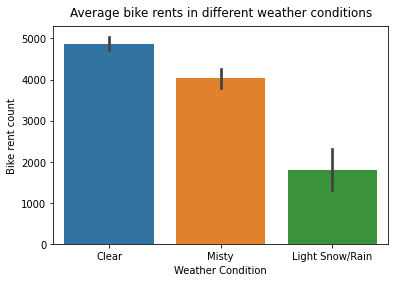


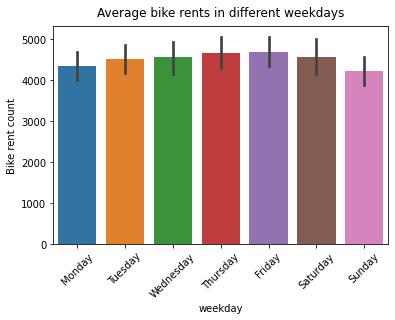


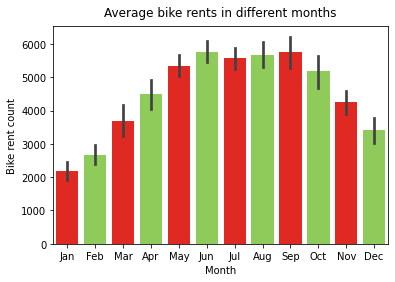
In the above figure, we can see how our dependent variable “count “ changes over time in different upward and downward trends.

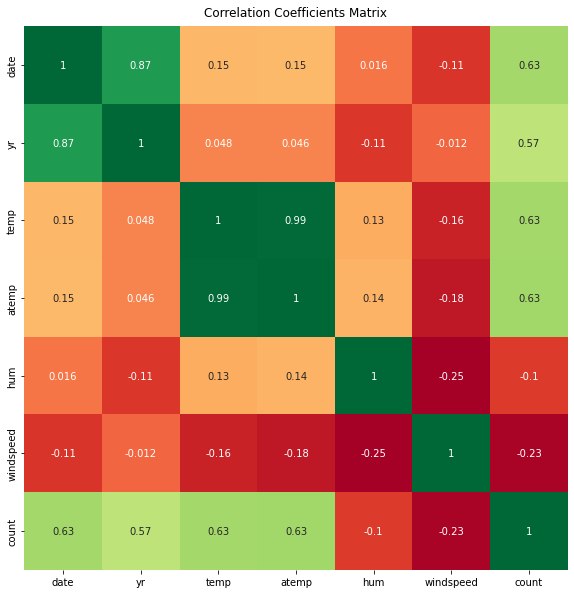
If we barplot our dependent variable vs various independent variables, then we can interpret how our dependent variable relates with various independent variables.



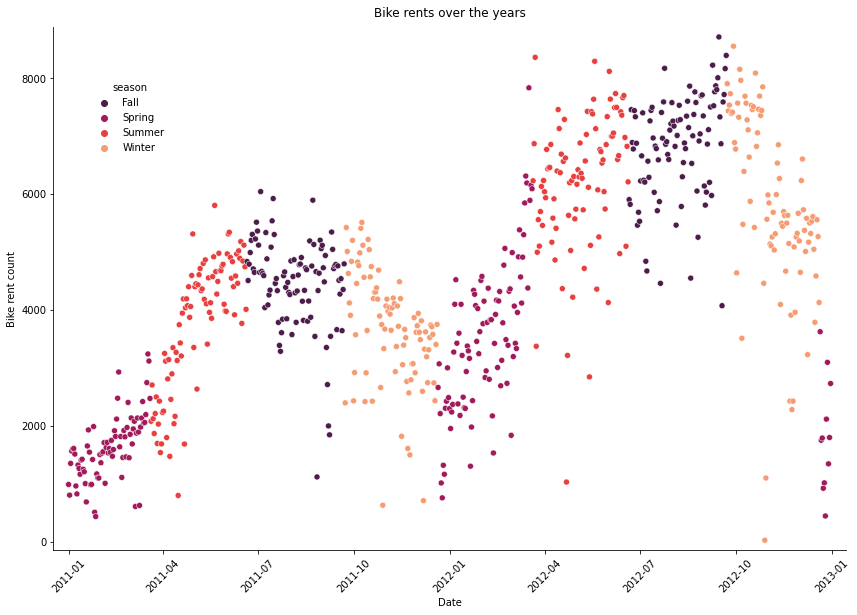
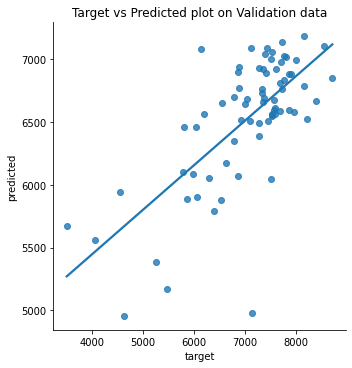








In the Correlation Coefficients Matrix above we can see Normalized temperature and Normalized feeling temperature variables are highly correlated. As they will provide the same information, we have dropped out the Normalized temperature variable.



By plotting “Target” vs “Prediction” of validation data, we can determine from the above figure(1)

1. The baseline model has quite a high variance on validation data.
2. The models perform well on days with high demand and poorly on low demand of bike rental.

By plotting bike rents over the year in the above figure(2), we can determine that the bike-rental demand over the years is non-linear.

For these reasons we have decided to use a non-linear model like Decision Tree Regressor.

But After training on Decision Tree Regressor, we have a negative 𝑟2 value on the validation dataset.

So we have reached to a conclusion that our Decision Tree Regressor fits poorly on the data.

So we want to try out the following modeling techniques to improve our model performance:

1. Non-linear Regression Model
2. Ensemble Methods:Gradient Boosted Decision Trees