Capstone Project -2

Bike Sharing Demand Prediction

Individual Project

Kalyani Nikam

Problem Description

- Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes.

Data Description

- **Date** - year-month-day
- **Rented Bike count** - Count of bikes rented at each hour
- **Hour** - Hour of he day
- **Temperature**-Temperature in Celsius
- **Humidity** - %
- **Windspeed** - m/s
- **Visibility** - 10m
- **Dew point temperature** - Celsius
- **Solar radiation** - MJ/m2
- **Rainfall** - mm
- **Snowfall** - cm
- **Seasons** - Winter, Spring, Summer, Autumn
- **Holiday** - Holiday/No holiday
- **Functional Day** – No (Non Functional Hours), Yes (Functional hours)

Data Preparation

- **Handling missing values**
- No Missing Values In Our Dataset

- **Handling duplicate values**
- No Missing Values In Our Dataset

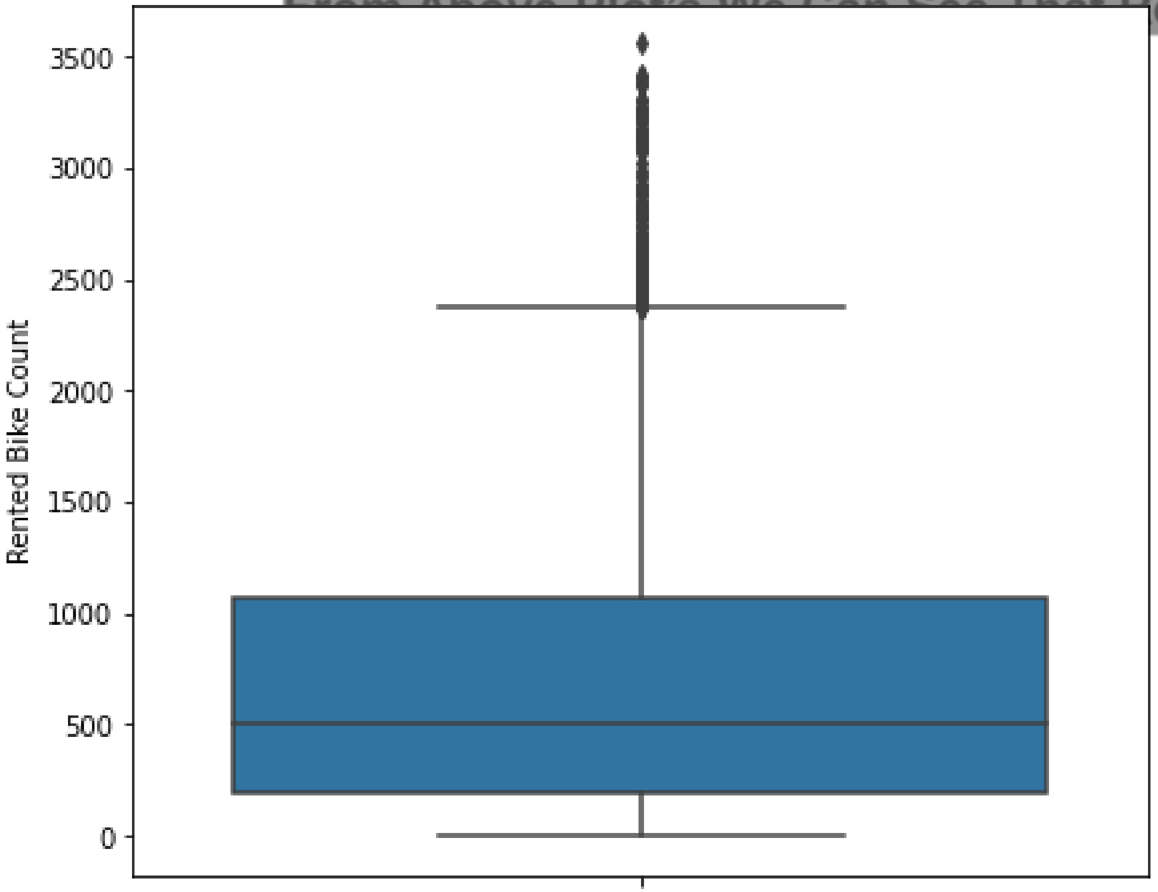
- **Making each column values in proper format** • Making Date in proper format from its string type.
- Creating new columns ‘Month’, ‘Year’, ‘Day’ from Date column

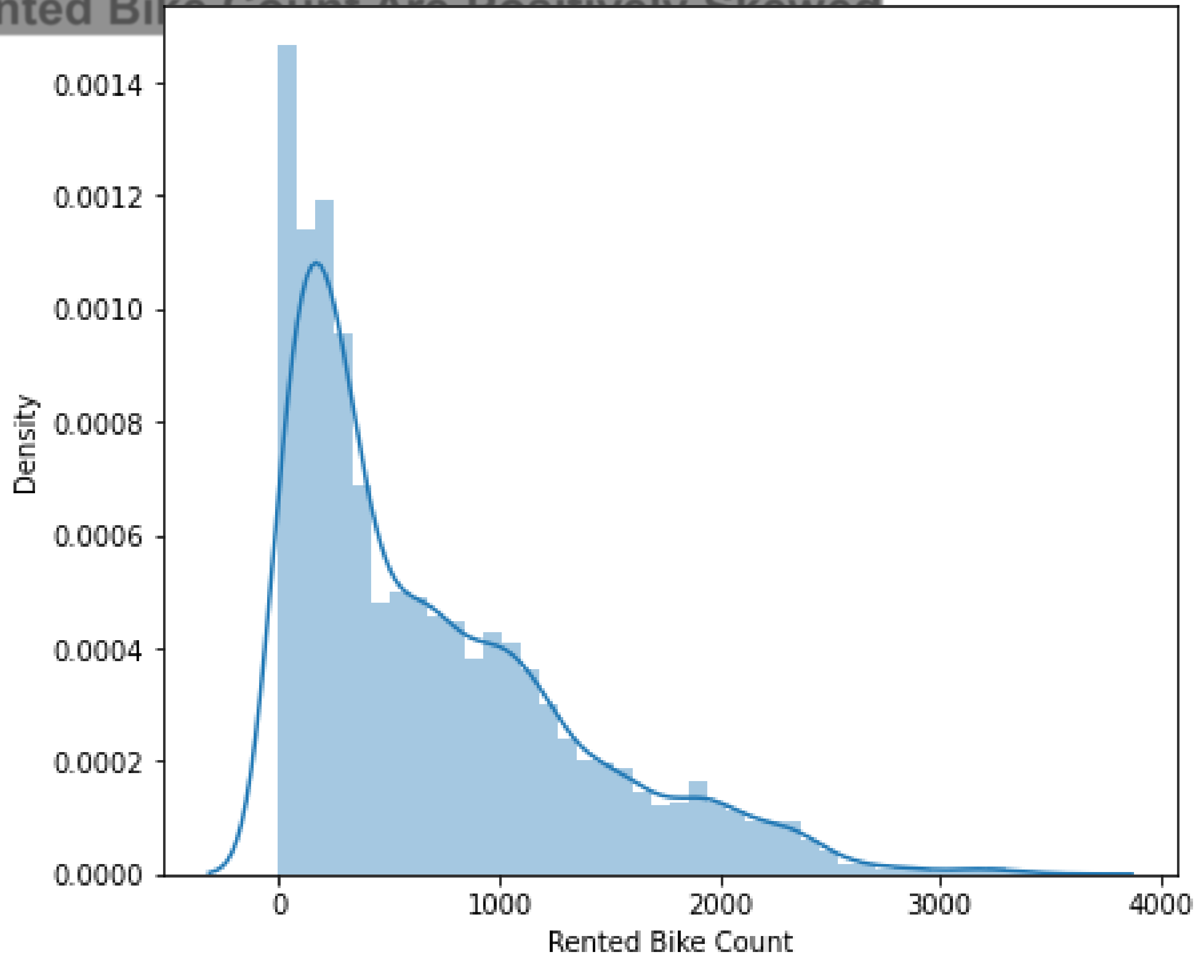
Insights Into Data

	Date	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)	Seasons	Holiday	Functioning Day
0	01/12/2017	254	0	-5.2	37	2.2	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
1	01/12/2017	204	1	-5.5	38	0.8	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
2	01/12/2017	173	2	-6.0	39	1.0	2000	-17.7	0.0	0.0	0.0	Winter	No Holiday	Yes
3	01/12/2017	107	3	-6.2	40	0.9	2000	-17.6	0.0	0.0	0.0	Winter	No Holiday	Yes
4	01/12/2017	78	4	-6.0	36	2.3	2000	-18.6	0.0	0.0	0.0	Winter	No Holiday	Yes

Range Index: 8760 entries, 0 to 8759 Data columns (total 14 columns) *9 Numerical columns, 1 Date Column and 3 Categorical columns*

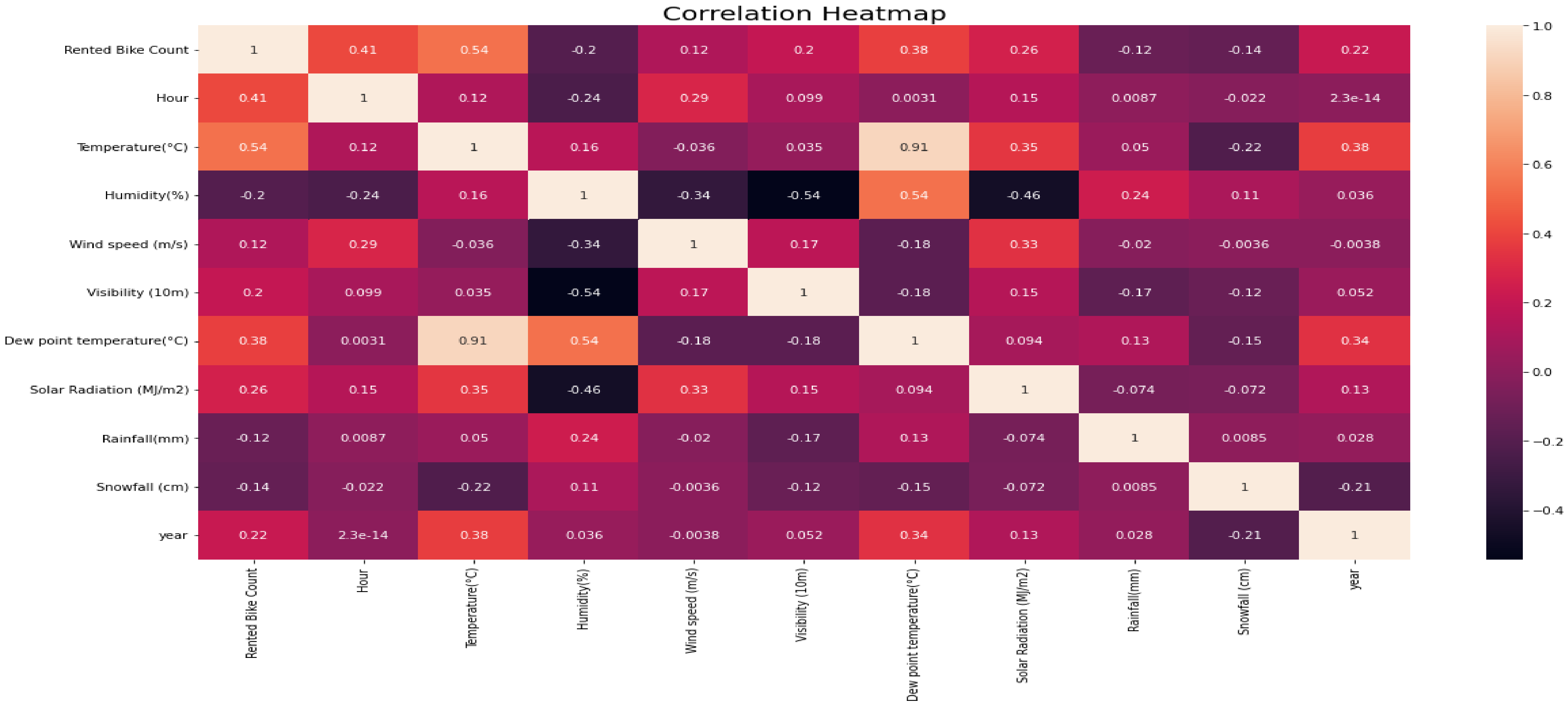
EDA- Boxplot And Distribution Plot Of Rented Bike Count





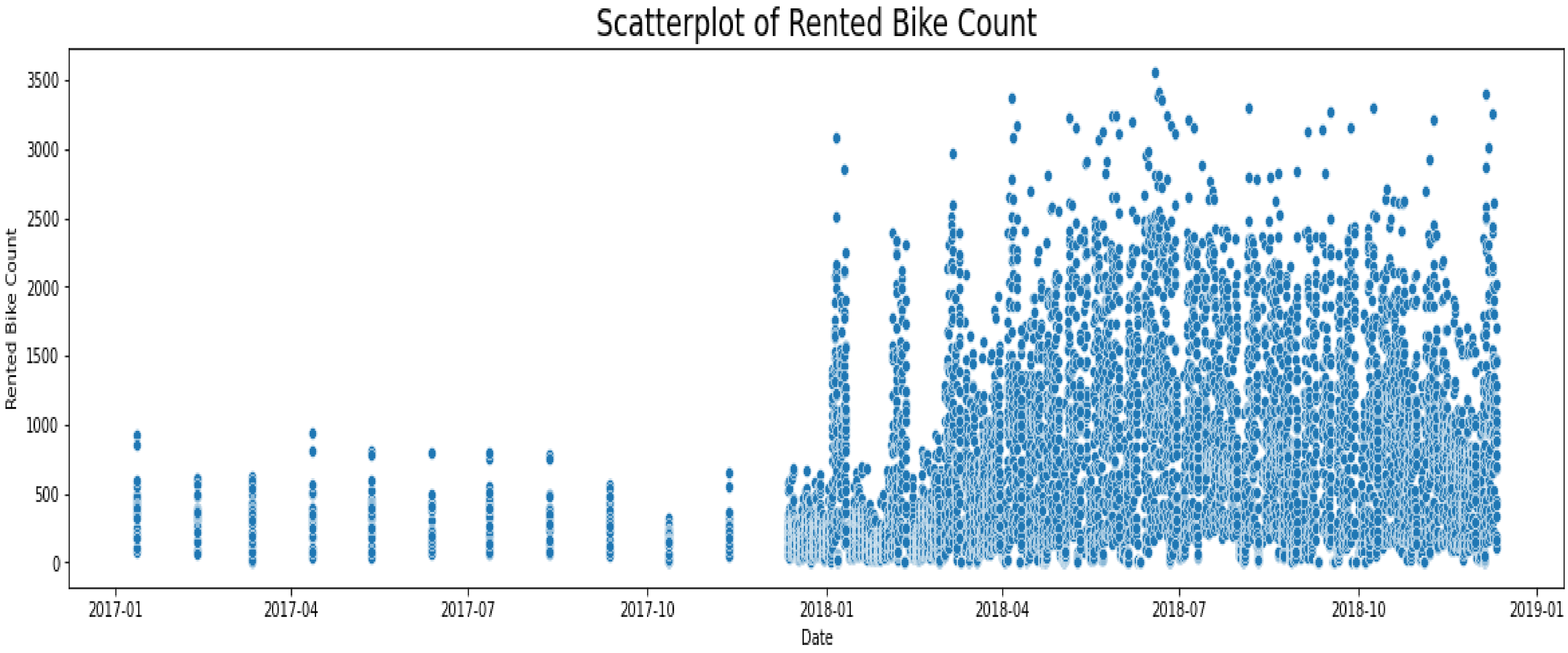
From Above Plot's We Can See That Rented Bike Count Are Positively Skewed

Correlation Heatmap



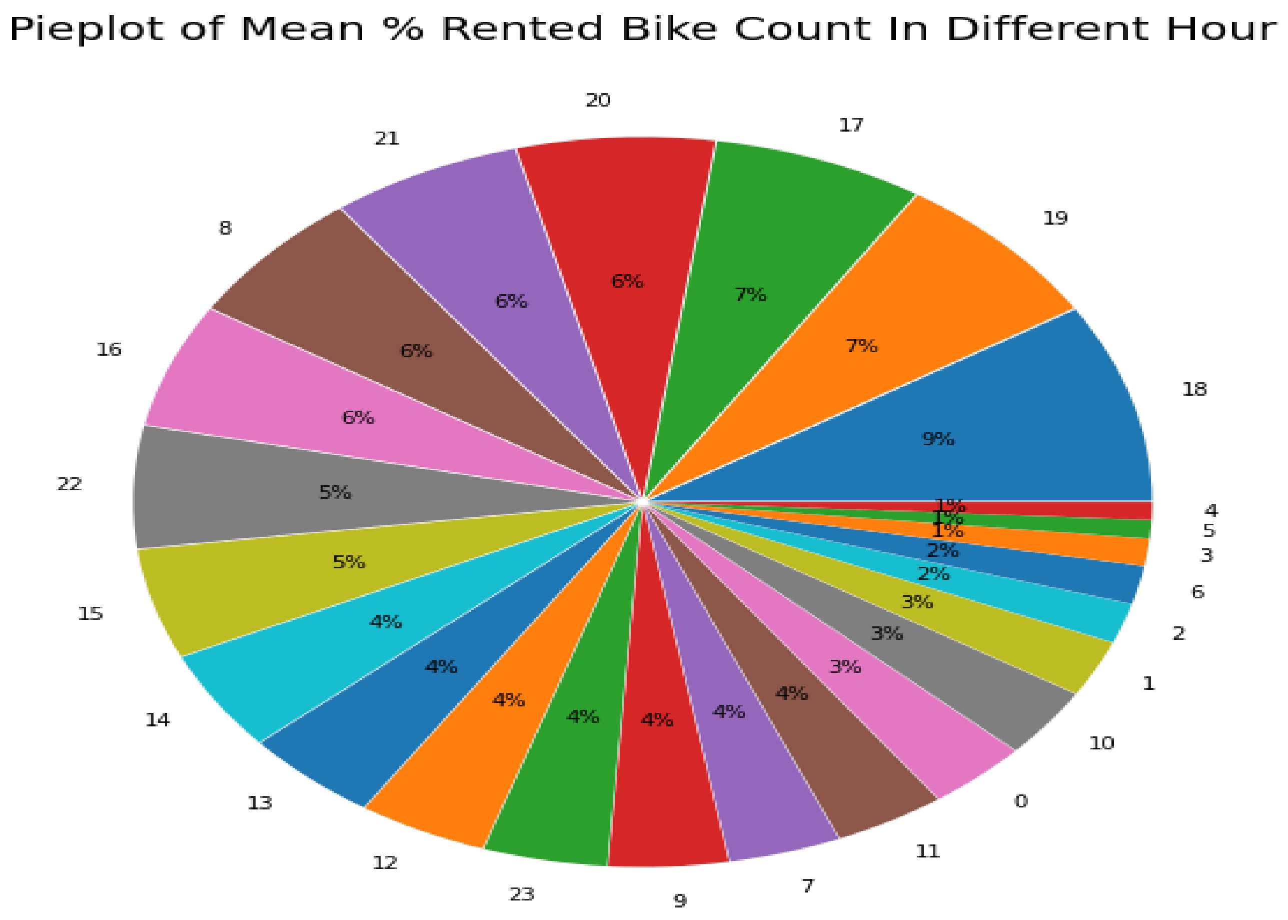
Temperature is positively correlated with Dew Point Temperature with correlation value of 0.91

Scatterplot Of Rented Bike Count Vs Date



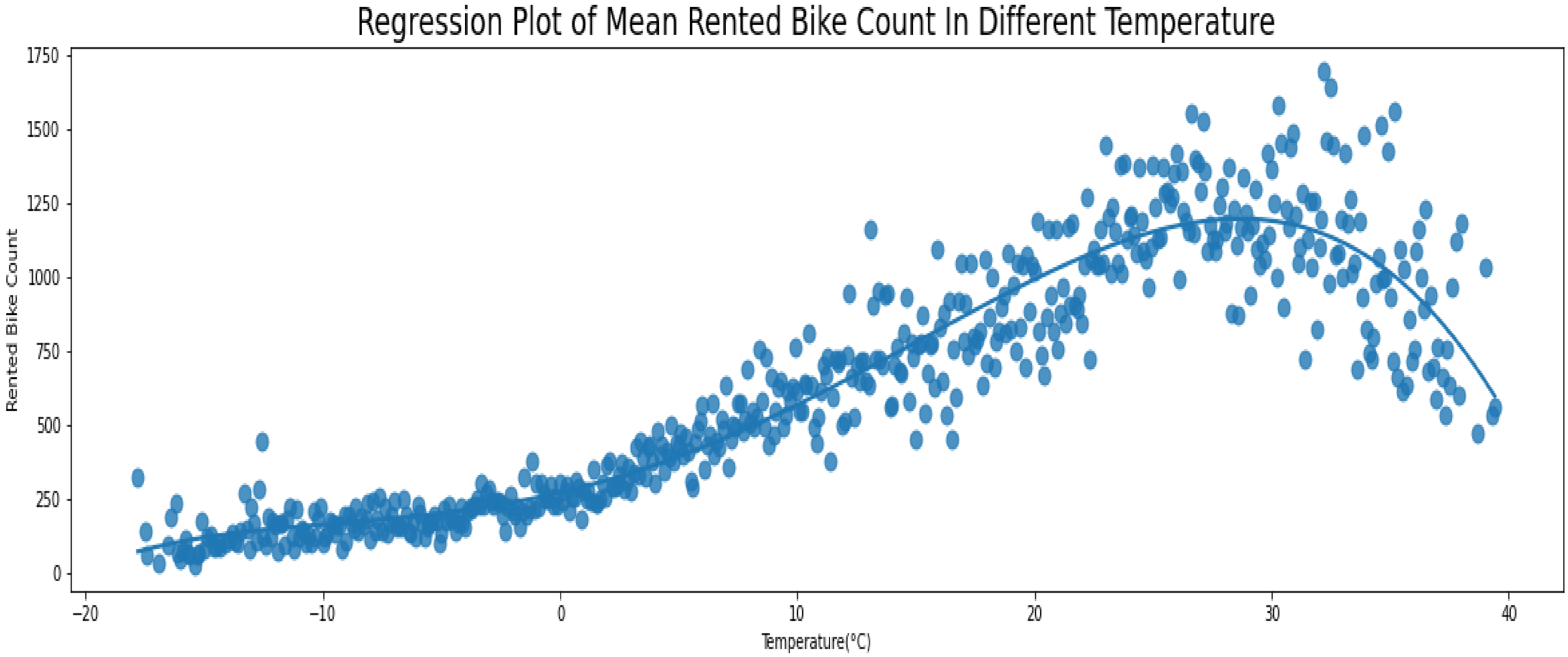
We can see that there is high demand of Rented bike in year 2018 when compare with year 2017

Pie Plot Of Mean % Of Rented Bike Count In Different Hour



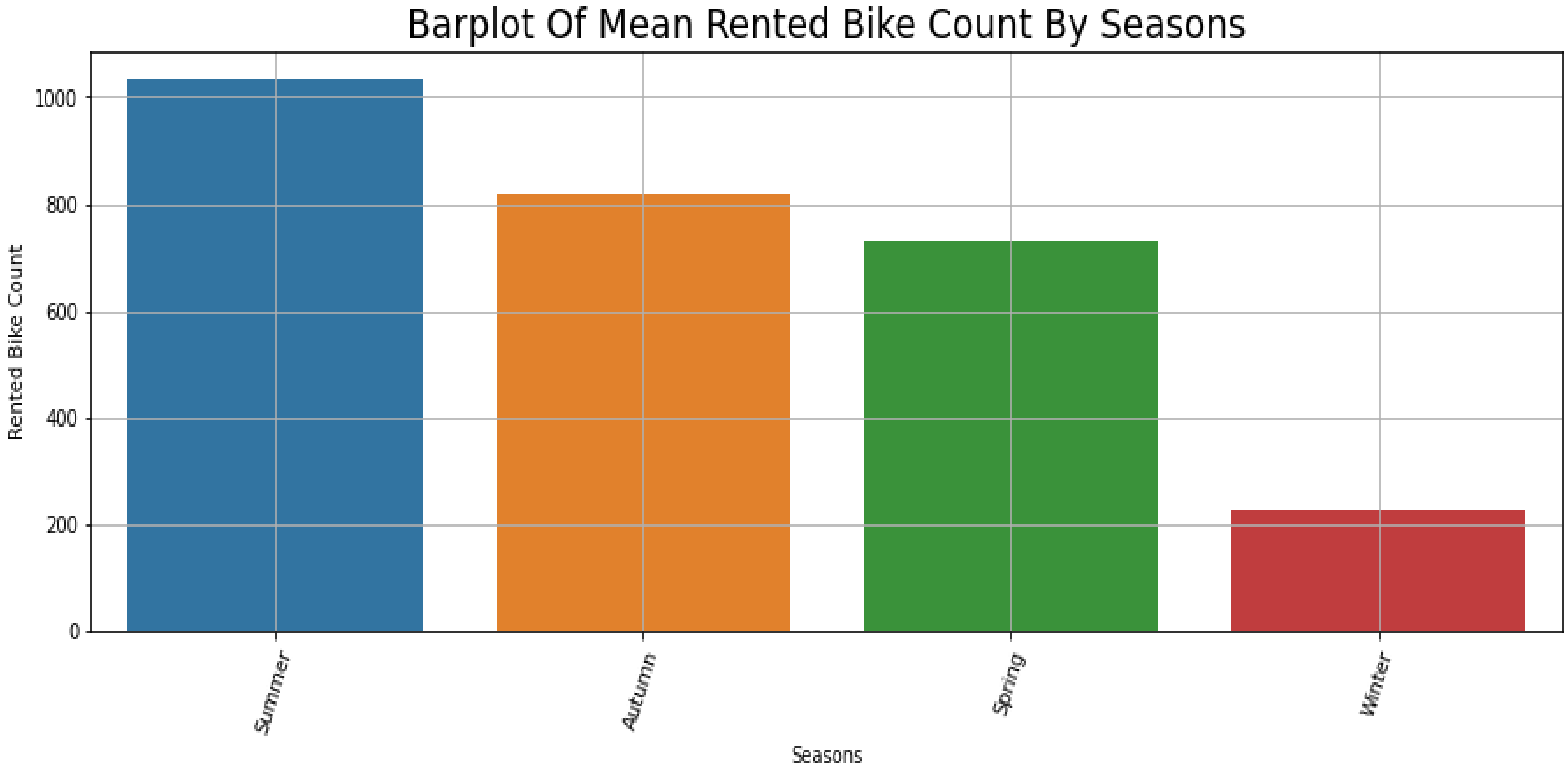
We can conclude from above pie plot that demand of rented bike is large in Hour of 18

Regression Plot Of Rented Bike Count In Different Temperature



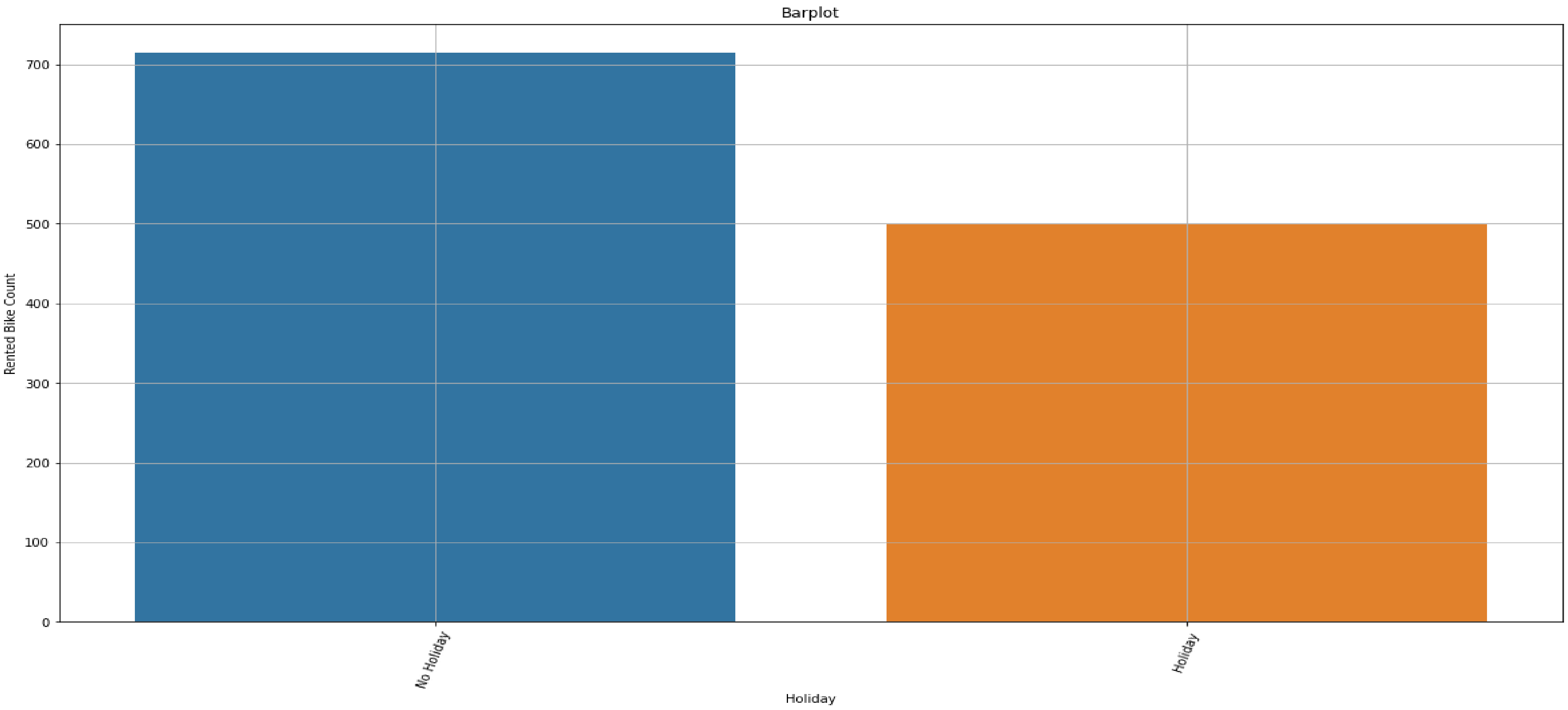
From above plot we can see that as temperature increases count of rented bike also increases

Bar Plot Of Mean Rented Bike Count In Different Seasons



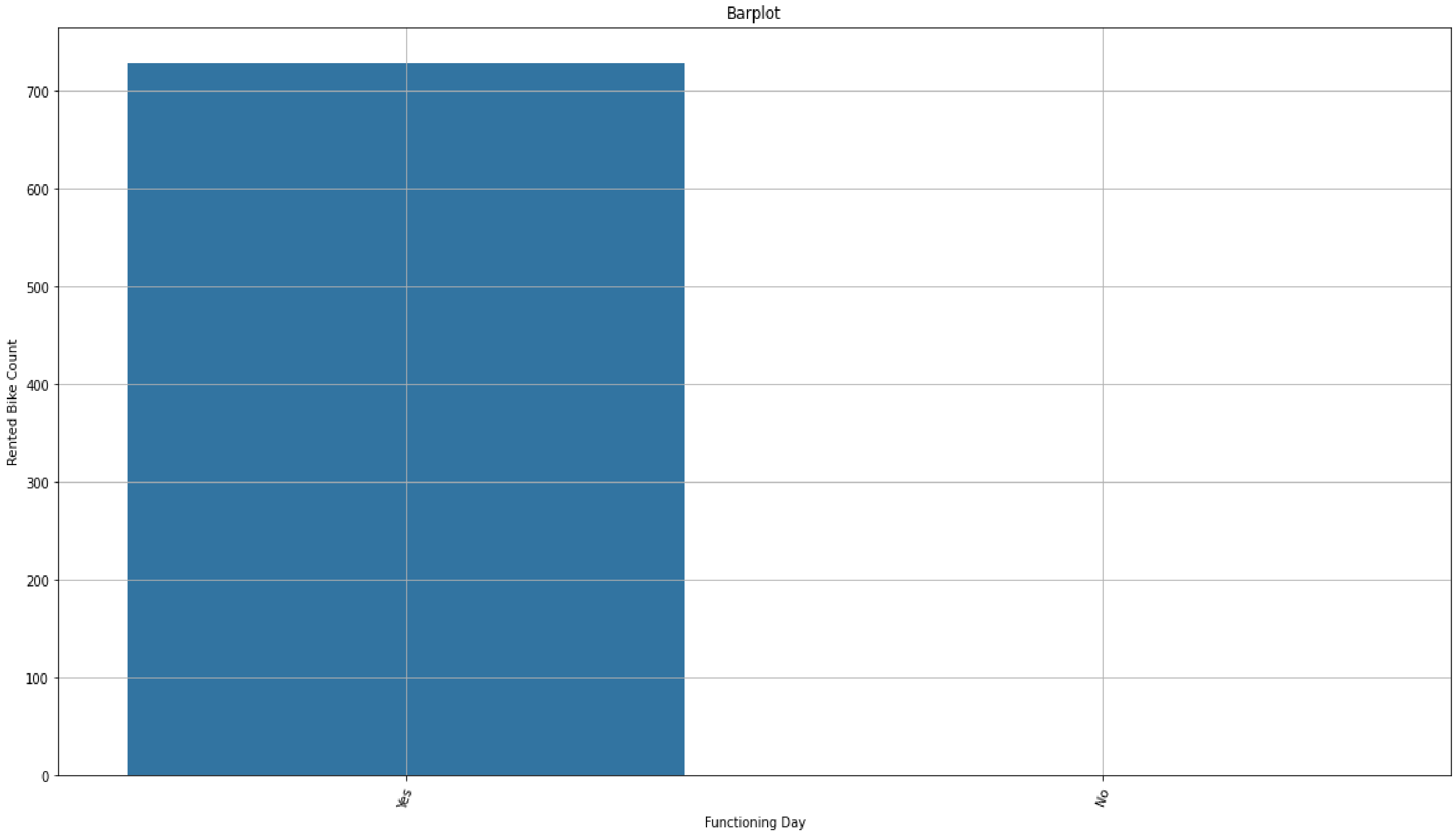
Count Of Rented Bike Is High In Summer

Bar Plot Of Mean Rented Bike Count In Holidays And Non Holidays



Count Of Rented Bike Is High In Non Holidays

Bar Plot Of Mean Rented Bike Count In Functioning & Non Functioning Days



In Non Functioning Days There Is No Demand Of Rented Bike

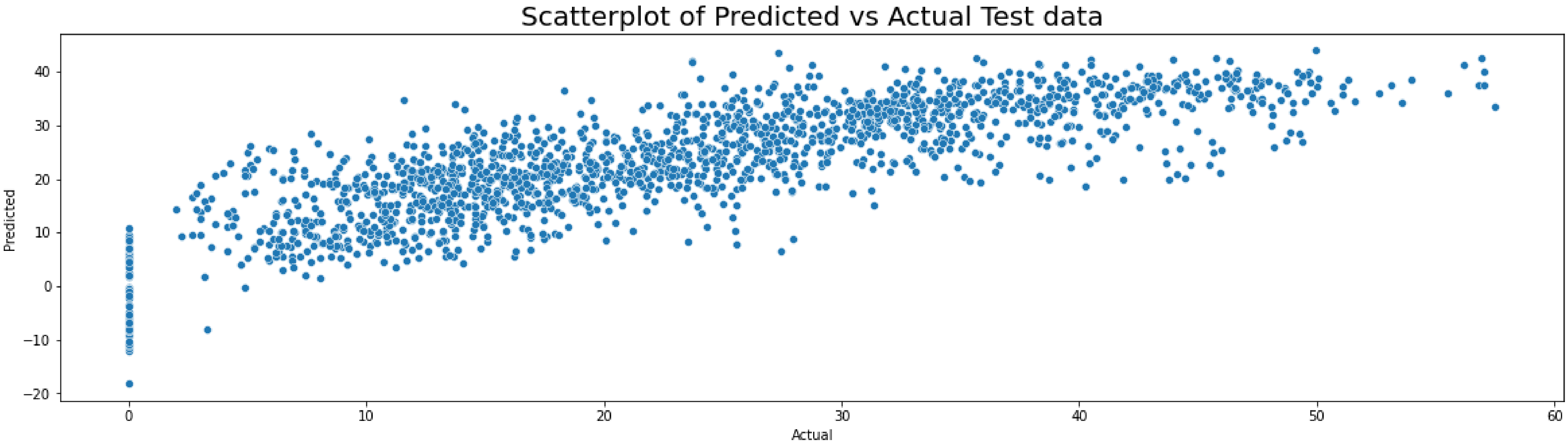
Model's Used

- Linear Regression
- Decision Tree Regression
- Decision Tree Regression With Hyperparameter Tuning
- Random Forest Regression
- Random Forest Regression With Hyperparameter Tuning
- Gradient Boosting
- Gradient Boosting With Hyperparameter Tuning
- XG Boost
- XG Boost With Hyperparameter Tuning

All Models With Evaluation

	Linear Regression	Decision Tree	Decision Tree Using GridSearchCV	Random Forest	Random Forest Using GridSearchCV	Gradient Boosting	Gradient Boosting using GridSearchCV	XG Boost	XG Boost Using GridSearchCV
Mean Squared Error Train	51.631967	0.000000	23.018425	1.765854	19.618511	16.739420	37.352121	16.651344	4.841590
Mean Squared Error Test	53.377437	26.043439	30.217535	12.994437	23.348786	19.363921	41.007791	18.978952	11.610330
r2 score Train	0.665451	1.000000	0.852302	0.988558	0.872882	0.891537	0.757977	0.892108	0.968629
r2 score Test	0.661066	0.830109	0.802880	0.917488	0.851741	0.877044	0.739610	0.879488	0.926277
Adjusted r2 score Train	0.663675	1.000000	0.851365	0.988497	0.872207	0.890961	0.756693	0.891535	0.968462
Adjusted r2 score Test	0.653750	0.827906	0.800323	0.915707	0.848540	0.874389	0.733989	0.876887	0.924686

Scatterplot Of Predicted Vs Actual Test Data Of Linear Regression Model

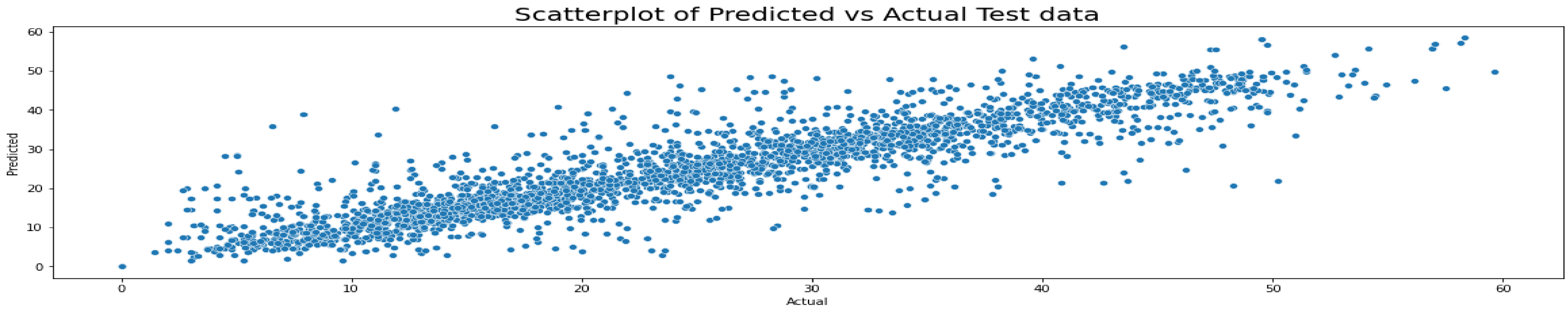


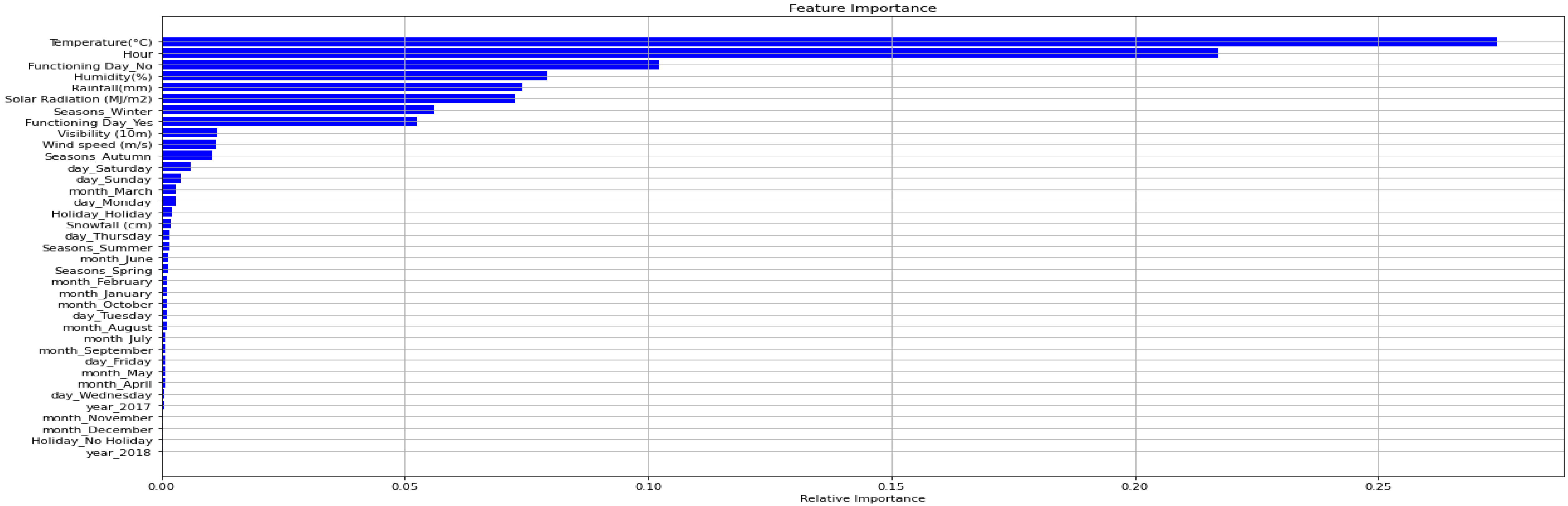
mean_squared_error_linear_test: 53.377437083669705

r2_score_linear_test: 0.6610660635389255

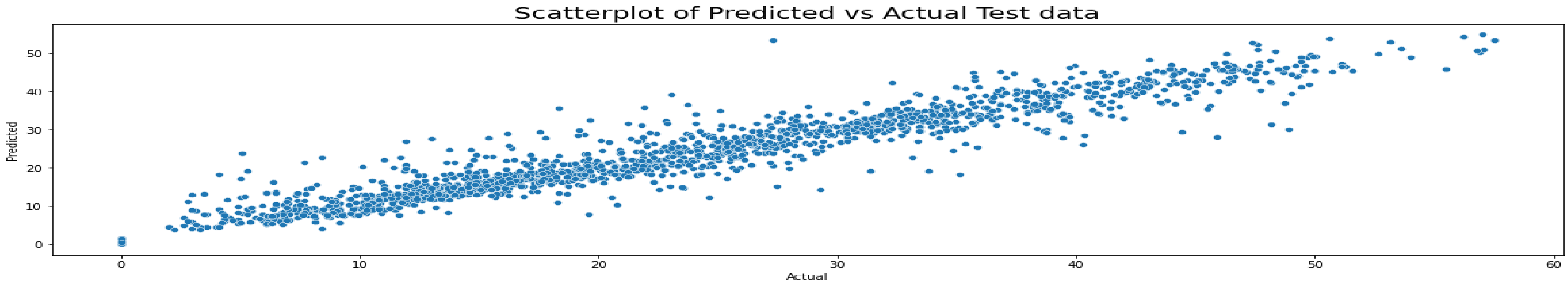
adjusted_r2_score_linear_test: 0.6537495199863819

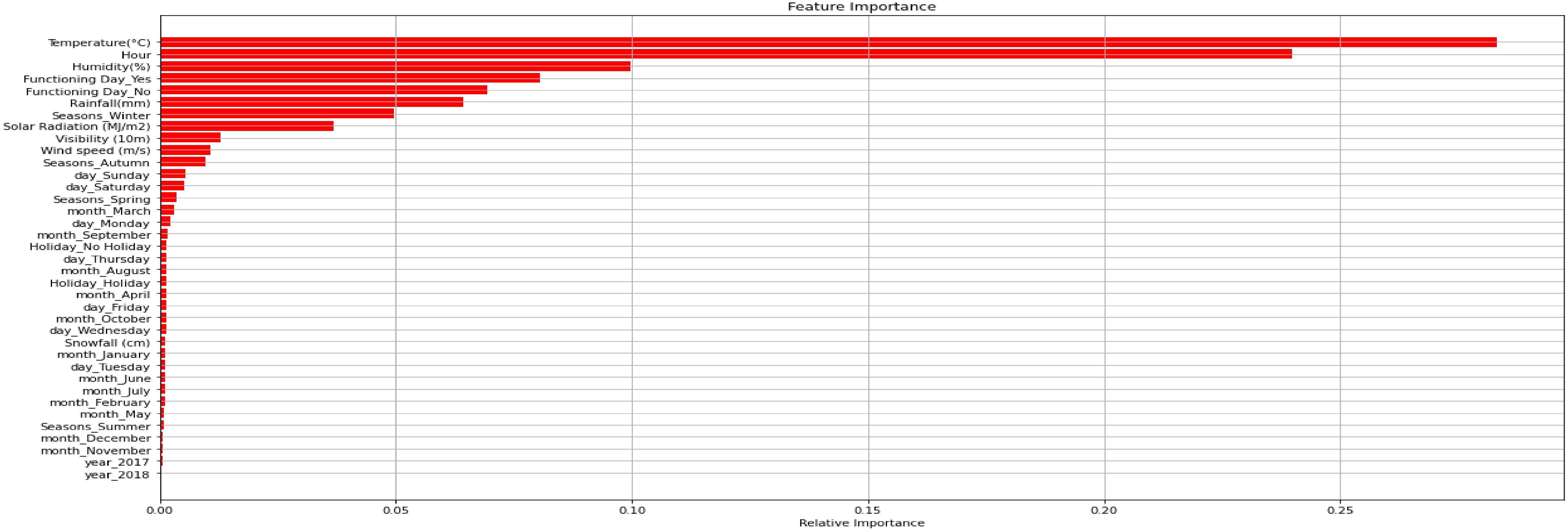
Scatter Plot And Features Importance Of Decision Tree Model



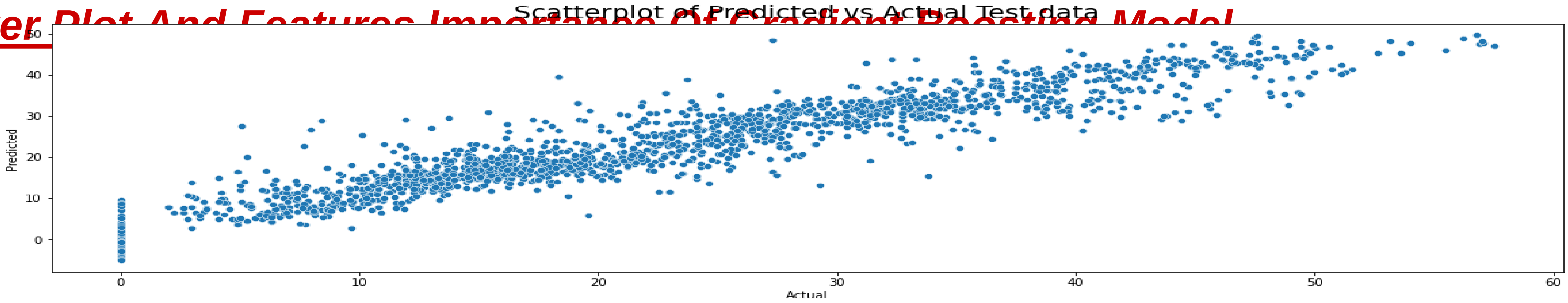


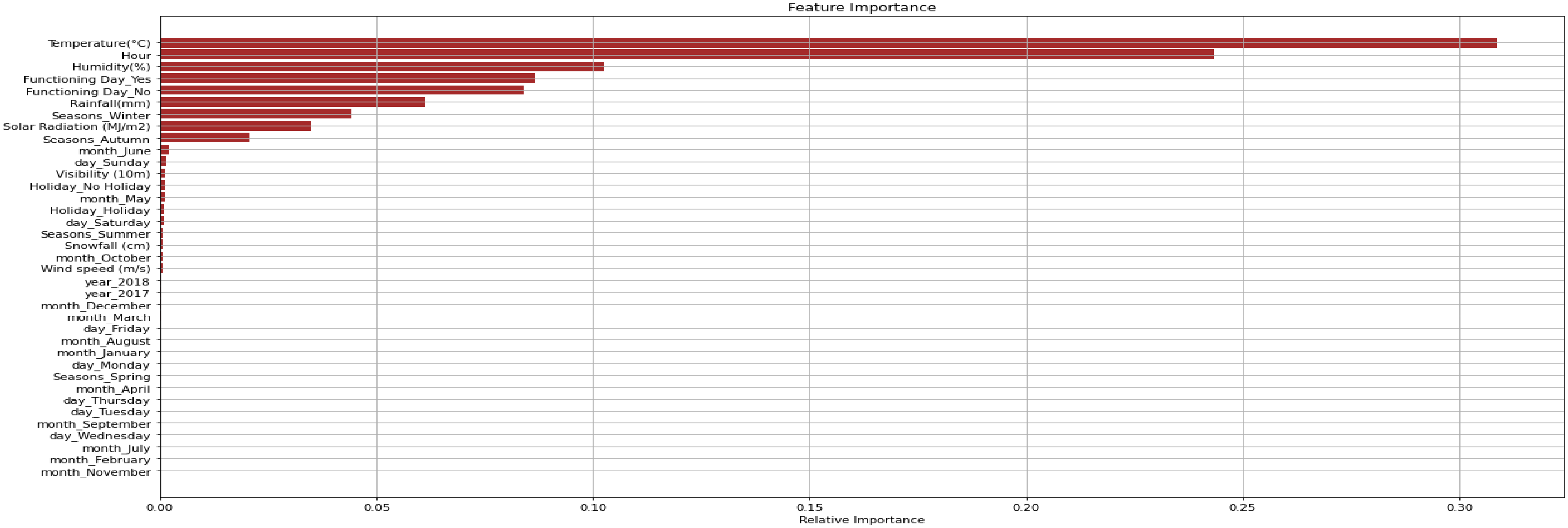
Scatter Plot And Features Importance Of Random Forest Model

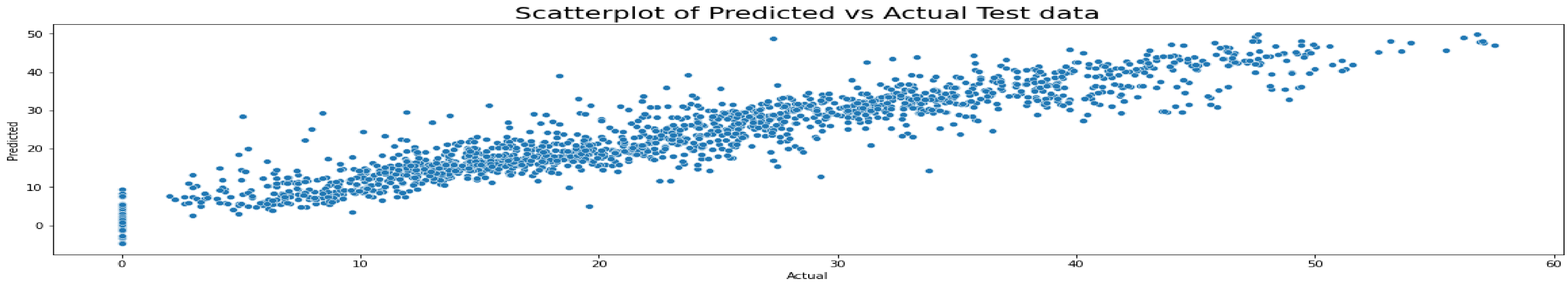


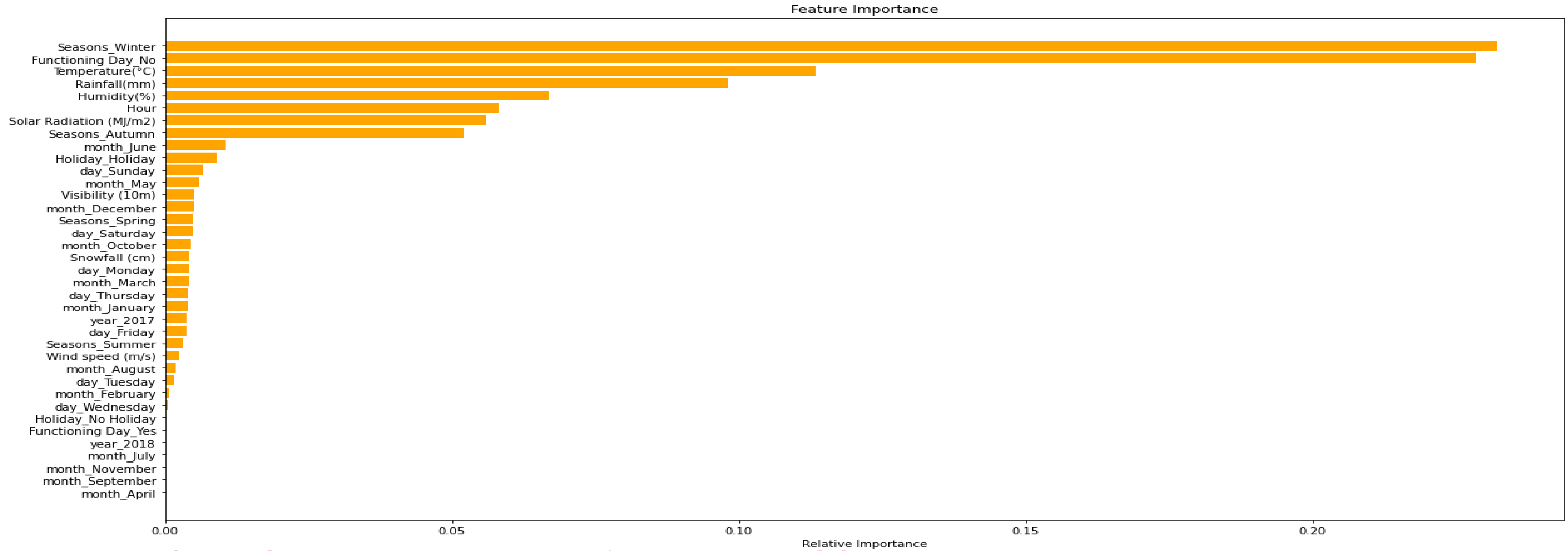


Scatter Plot And Features Importance Of Gradient Boosting Model









Scatter Plot And Feature Importance Of XG Boost Model

Conclusion’s

- Linear Regression Model Has Accuracy Of 65%.
- Decision Tree Has Accuracy Of 82%.
- Decision Tree With Hyperparameter Tuning Has Accuracy Of 80%.
- Random Forest Has Accuracy Of 91%.
- Random Forest With Hyperparameter Tuning Has Accuracy Of 84%
- Gradient Boosting Has Accuracy Of 87%.
- Gradient Boosting With Hyperparameter Tuning Has Accuracy Of 73%.
- XG Boost Has Accuracy Of 87%.
- XG Boost With Hyperparameter Tuning Has Accuracy Of 92%.
- **From Above We Can Conclude That XG Boost With Hyperparameter Tuning Is The Best Fitted Model To Our Data.**
- **XG Boost With Hyperparameter Tuning Gives About 97% Accuracy In Training Data And 92% Accuracy In Test Data. Also XG Boost Has Lowest Mean Squared Error In Test Data.**
- Seasons, Temperature, Hour, Functioning Day, Humidity are the most important features which affects our Target variable.


Thank you