Capstone Project -2

Bike Sharing Demand Prediction

(Supervised ML-Regression)

By : Ajahar Daroga

Problem Statement:

- 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.

Aim of Project

- Aim of this project is minimize the waiting time or make rental bike available at the right time.

- Developing a model which can predict the number of rental bikes required per hours (i.e. during a whole day)

Methodology

Data Understanding

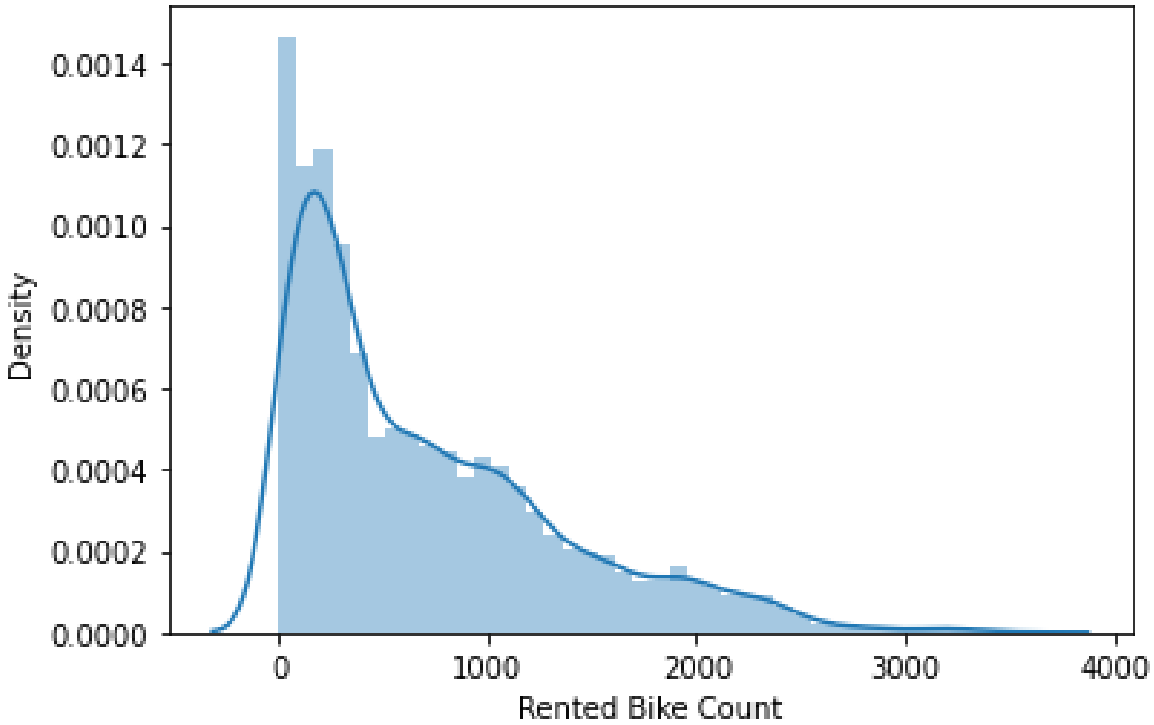
- **Date** - year-month-day.
- **Rented Bike count** - Count of bikes rented at each hour**. (Dependent Variable)**
- **Hour** - Hour of the day. (Time)
- **Temperature**-Temperature in Celsius.
- **Humidity** - %
- **Windspeed** - m/s.
- **Visibility** - 10m.
- **Dew point temperature** – Celsius.
-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

Data Preparation

- Records = 8760 & Features = 14.
- No data is missing in given Dataset.
- Date feature convert to DateTime initially its str object.
- Seasons, Holiday, functioning Day this features are categorical in nature need to converting by using one hot encoding or label encoding.
- Combine the dataset after converting numerical and categorical in proper format so model accept.
- Scaling the data or standardize the data.

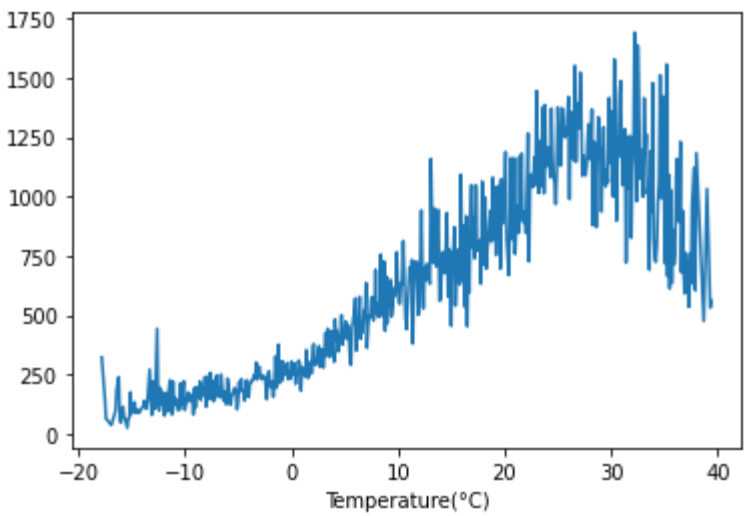
Dependent Variable

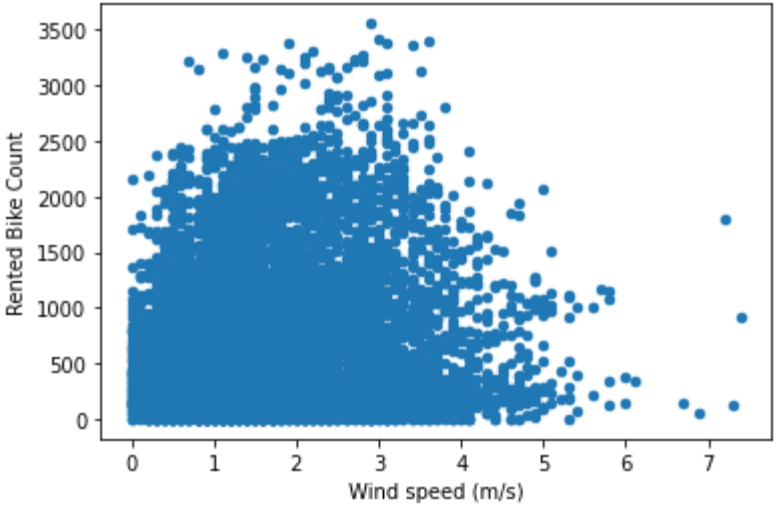
- Rented Bike Count
- positively skewed data need to transform using log or sqrt transformation.

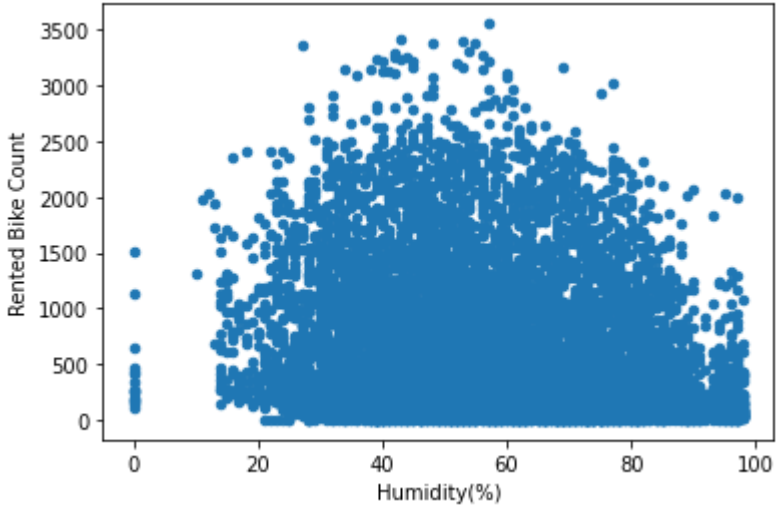


Relationship with independent Numerical Variable

- Higher the temp the requirement of bikes rented also high.
- Humidity range between 20 to 80% then count of bikes rented is high.
- Less the Wind speed more demand of bikes.

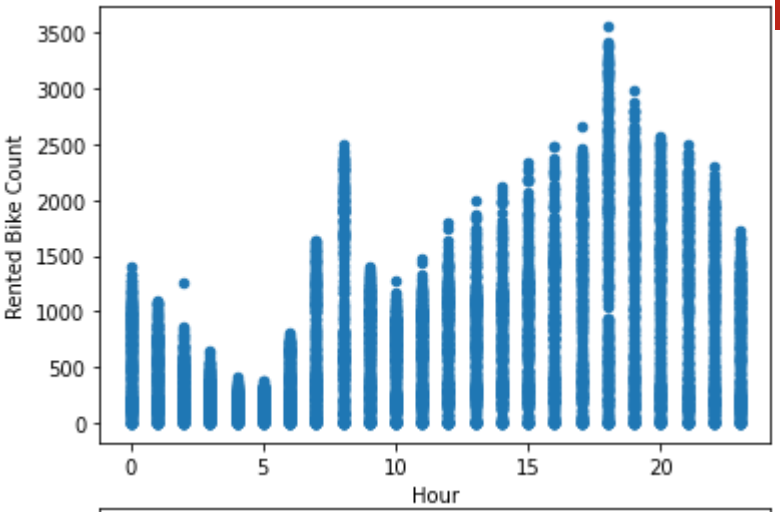


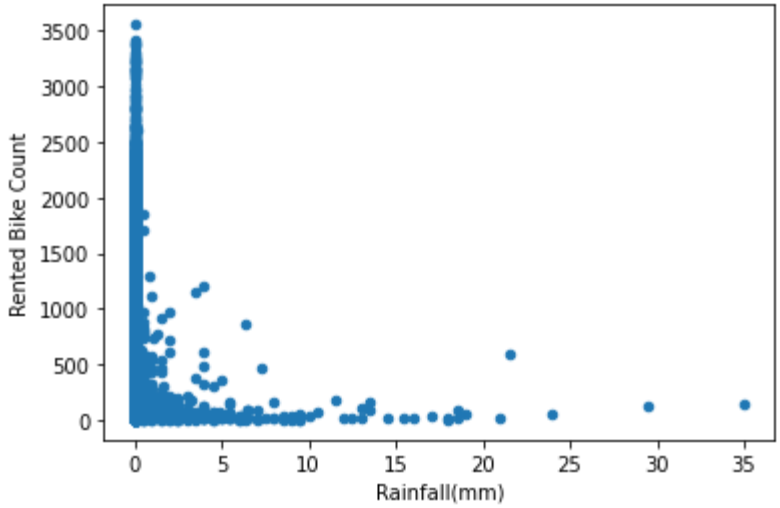


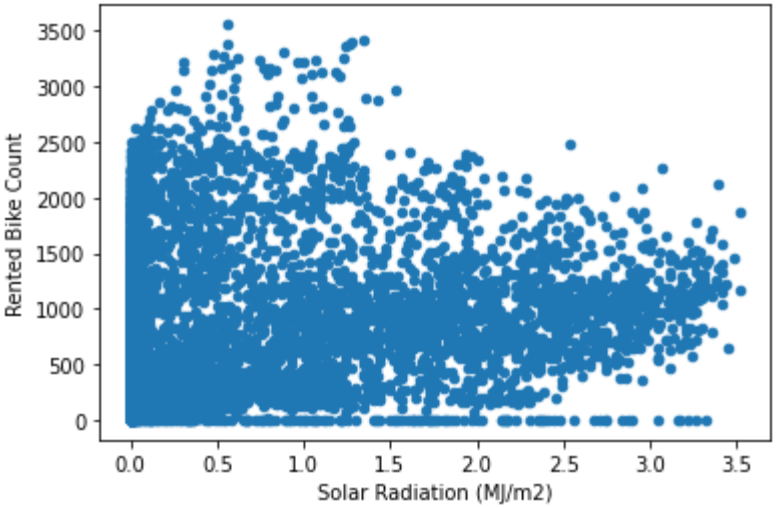


Relationship with independent Numerical Variable

- High demand on 8 am in morning and 3 to 8 pm.
- we can conclude that if no rainfall then number of bike count is more.
- Solar radiation work same as the temperature relation with dependent variable.

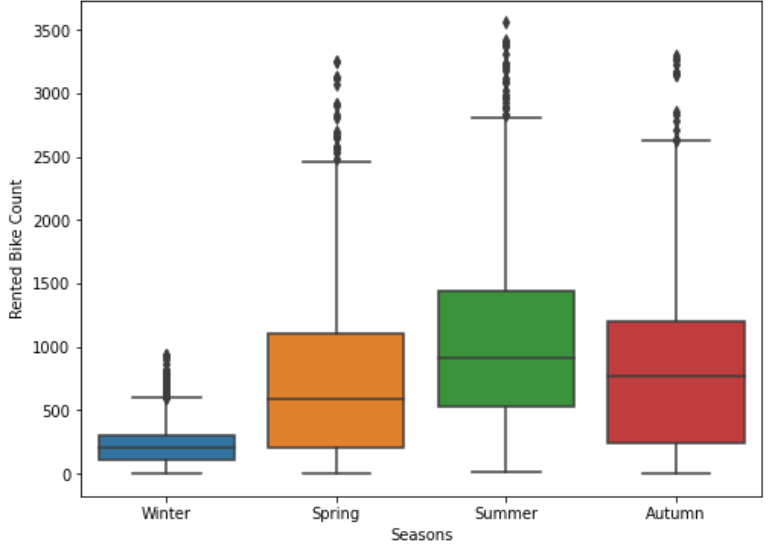


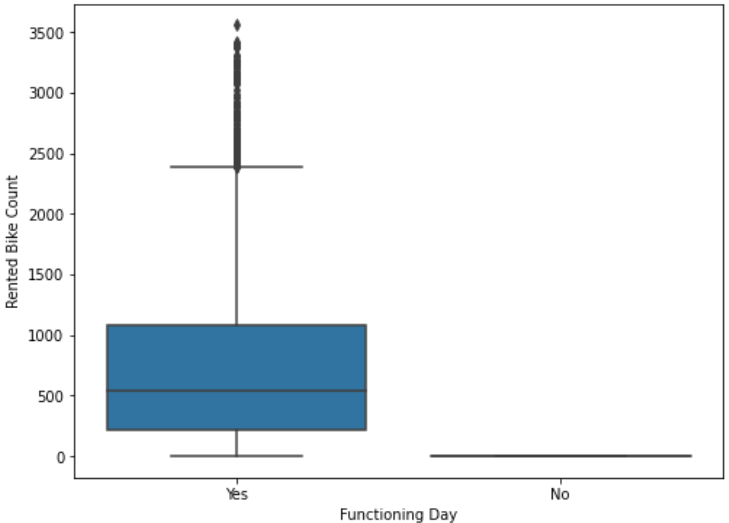


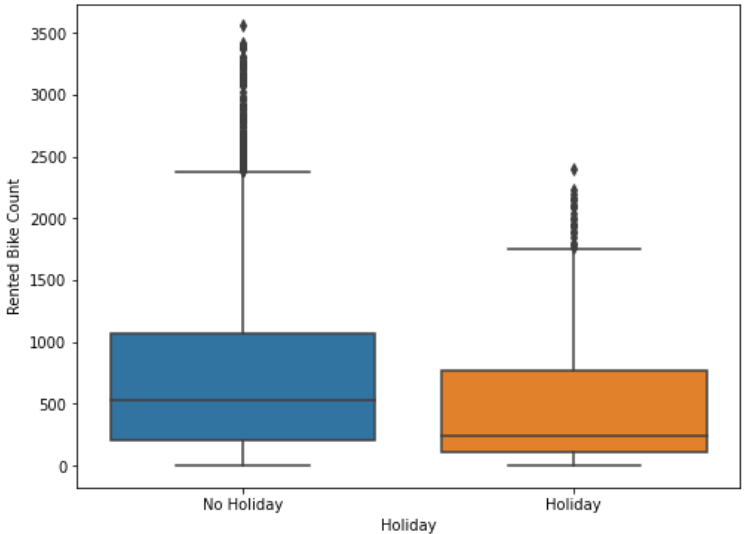


Relationship with independent Categorical Variable

- Less demand on winter seasons and High demand on Summer seasons.
- High demand on No Holidays compare to Holidays.
- On non functioning day there is no demand.



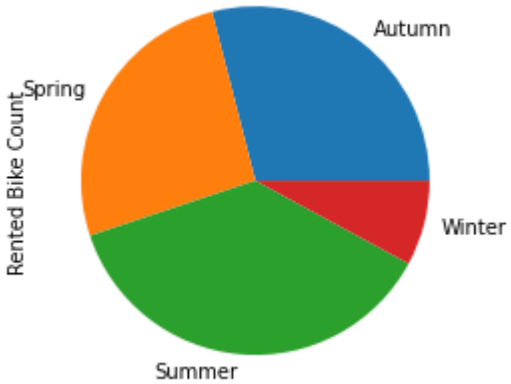


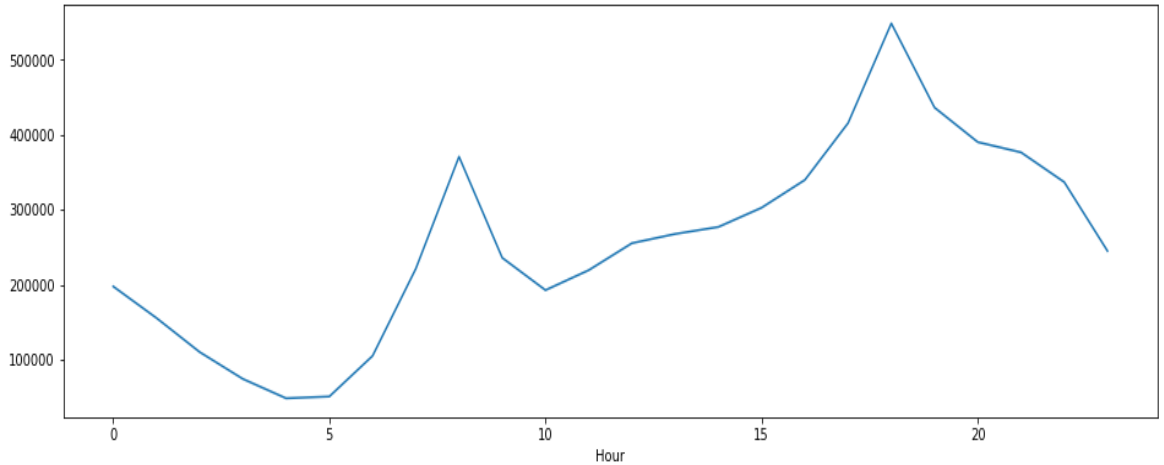


Relationship with Date, Hour, Seasons

- The average high demand of bike is on Friday and low on Sunday.
- The average high demand of bike at 8 am and 6pm we can conclude from graph.
- Summer season has the high dement of bike compare to all.

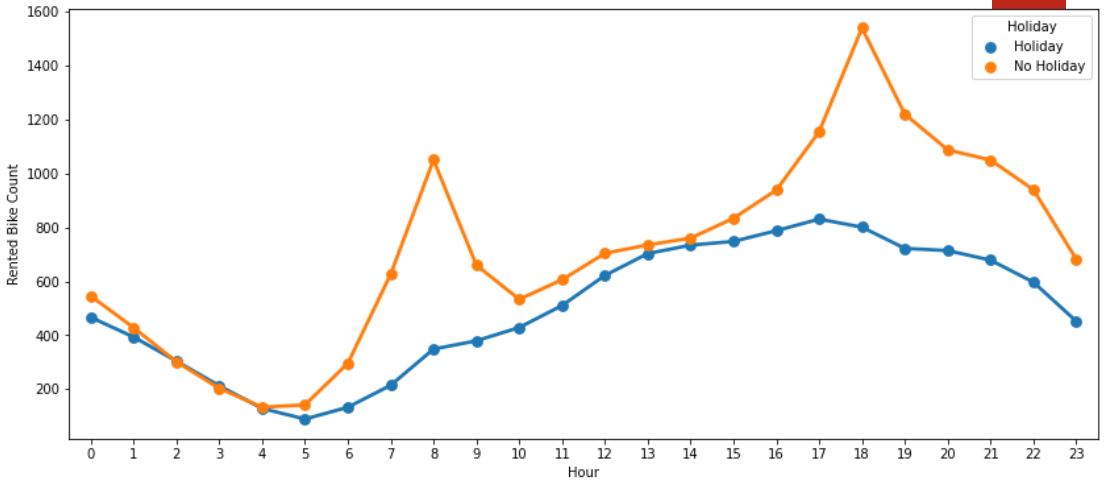
Weekdays	Rented Bike Count
Friday	950334
Wednesday	923956
Monday	911743
Saturday	885492
Thursday	861999
Tuesday	858596
Sunday	780194

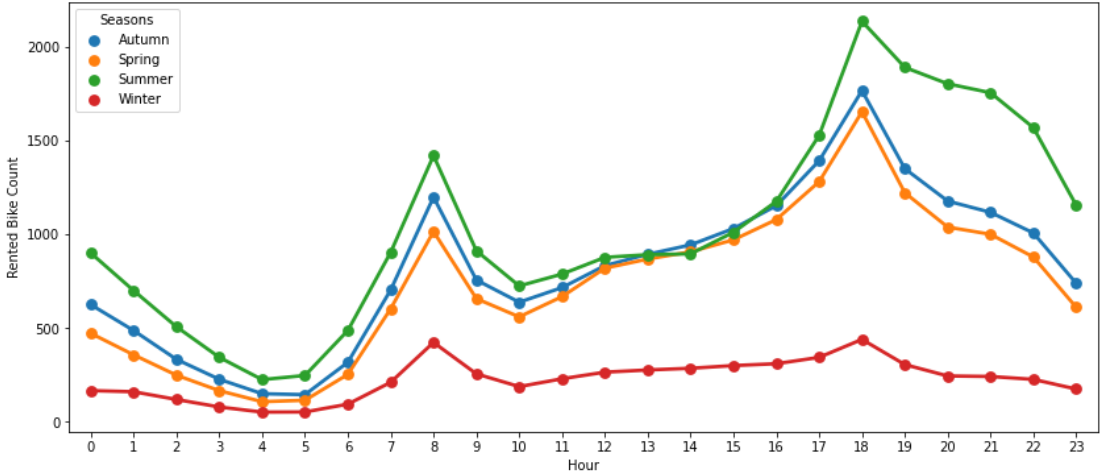




Relationship with Holiday, Hour, Bike Count

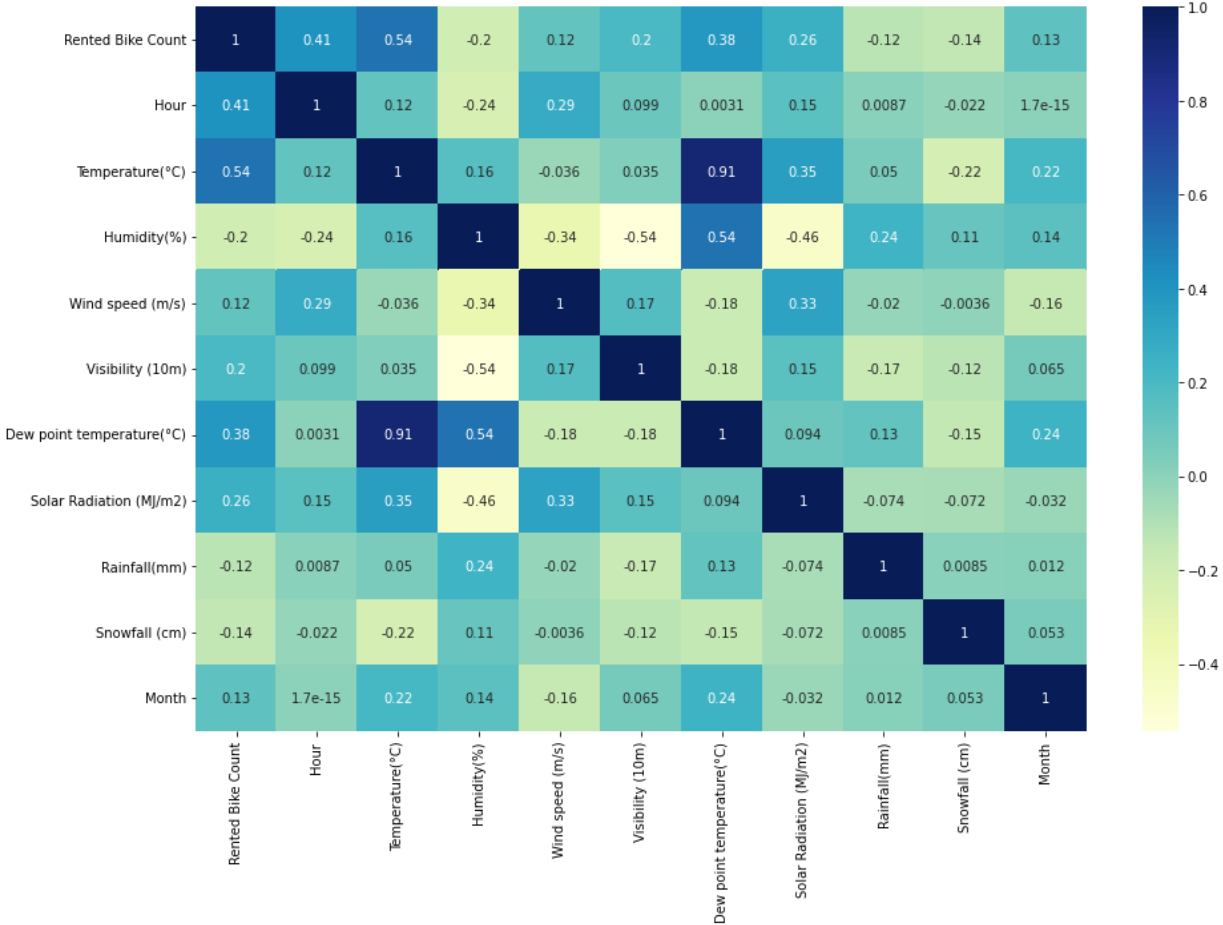
- The average bike count during the holidays is more or less constant.
- On no holidays demand spike two times a day at 8am and 6pm.
- More bike count is during the no holidays day.
- The average bike count during the season more is summer.





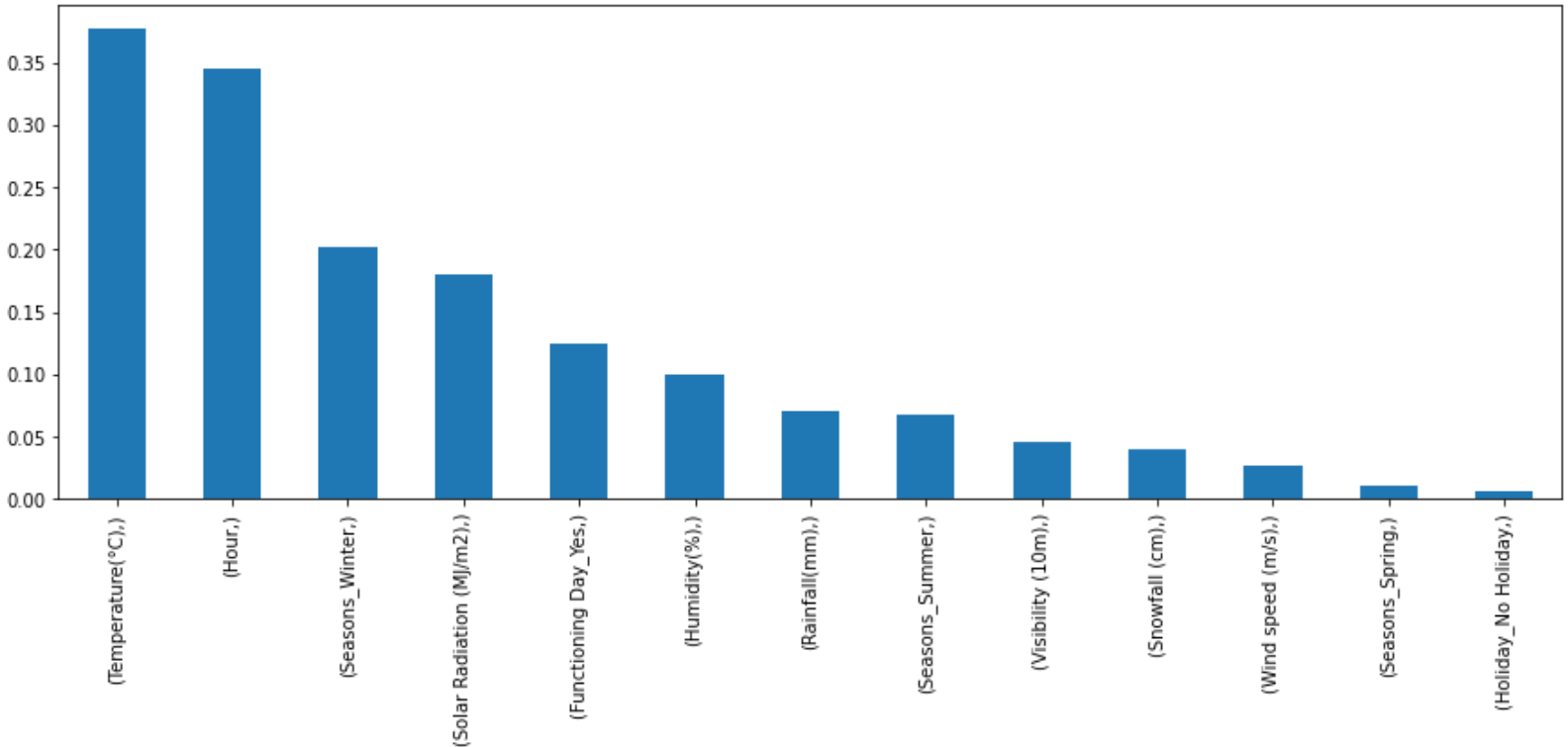
Correlation between Features

- Correlation heatmap between the features.
- Temperature(°C) and Dew point temperature(°C) have very strong relation so need to check the multicollinearity issue.
- Dropping the Dew point temperature features from dataset.



Feature Engineering

- Most important features calculated based on Information gain.
- Using mutual information for regression library.
- Temperature and hour of the day are most influence variable in hourly rented bike demand prediction.



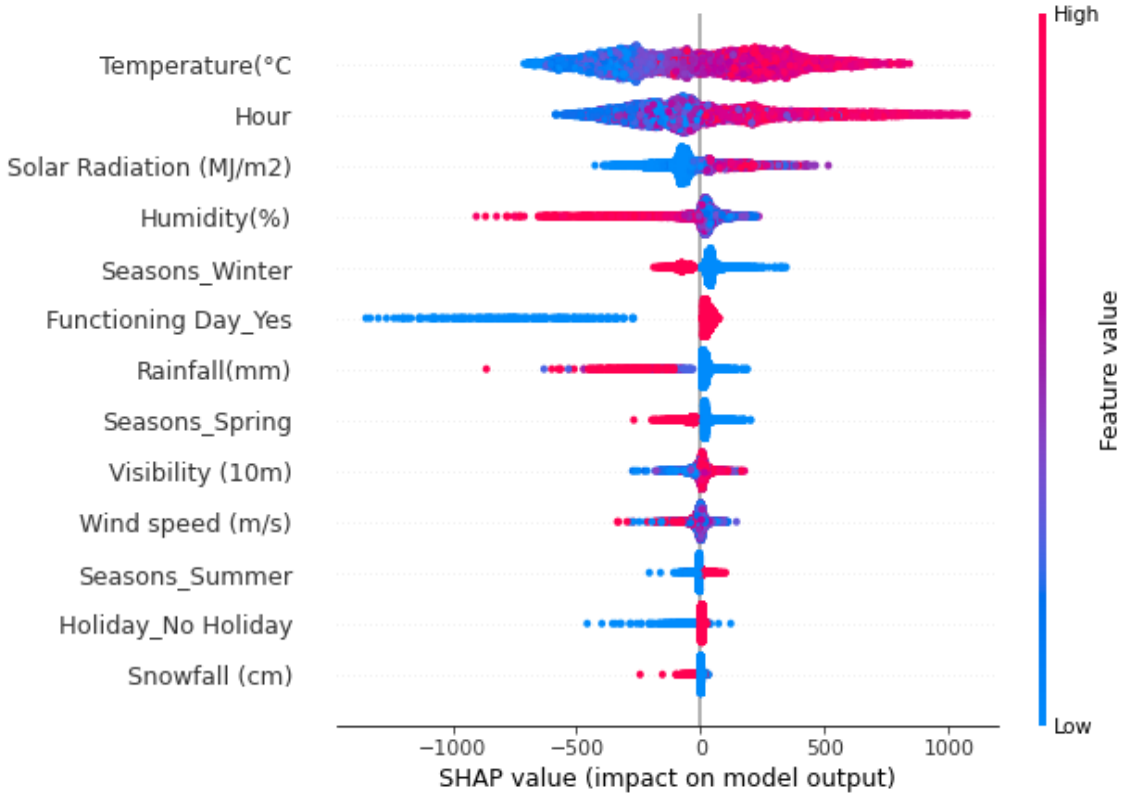
Model Building & Evaluation

Models	MLR_model	Ridge_model	Decision_Tree_model	KNN_model	xgb_model
RMSE_value	436.543950	436.585791	322.037962	301.047653	229.050470
R2_Score	0.544659	0.544571	0.752203	0.783453	0.874644

- Using Grid search method for hyperparameter tunning the model.
- XGBoost model gives more accuracy comparing to other models and low root mean squired error.
- XGBoost model have low bias & low variance.

Model Explainability

- Model explain ability by using SHARPLY method.
- blue, we have negative Shap values and red have positive Shap values.
- All the values on the left represent the observations that shift the predicted value in the negative direction while the points on the right contribute to shifting the prediction in a positive direction. All the features are on the left y-axis.



Conclusion

- Higher the temp the requirement of bikes rented also high.
- Humidity range between 20 to 80% then count of bikes rented is high.
- Important feature during model building is Temperature and Hour.
- Less the Wind speed more demand of bikes.
- High demand on 8 am in morning and 3 to 8 pm.
- we can conclude that if no rainfall then number of bike count is more.
- Less demand on winter seasons and High demand on Summer seasons.
- High demand on No Holidays compare to Holidays.
- On non functioning day there is no demand.
- The average high demand of bike is on Friday.
- The average high demand of bike at 8 am and 6pm we can conclude from graph.
- Summer season has the high dement of bike compare to all.
- The average bike count during the holidays is more or less constant and on no holidays demand spike two times a day at 8am and 6pm.
- XGBoost model gives more accuracy comparing to other models and low room mean squired error.

THANK YOU