

**CAPSTON PROJECT**

**PROJECT TITLE :- BIKE RENTAL PREDICTION**

**Presented By**

**Student name :-** Pahalad Kumar

**College name :-** Tula's Institute

**Department :-** Computer Science Engg.

**Email Id:-** pahalad845401mth@gmail.com

**AICTE STUDENT ID:-**STU666e45128e4361718502674



## 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.**

# Content

- ☐ Data Pipeline
- ☐ Data Description
- ☐ Exploratory Data Analysis
- ☐ Models performed
- ☐ Model Validation & Selection
- ☐ Evaluation Matrix of All the models
- ☐ Model Explainability - SHAP
- ☐ Challenges
- ☐ Conclusion

## Data Pipeline

- Exploratory Data Analysis (EDA): In this part we have done some EDA on the features to see the trend.
- Data Processing: In this part we went through each attributes and encoded the categorical features.
- Model Creation: Finally in this part we created the various models. These various models are being analysed and we tried to study various models so as to get the best performing model for our project.

# Data Description

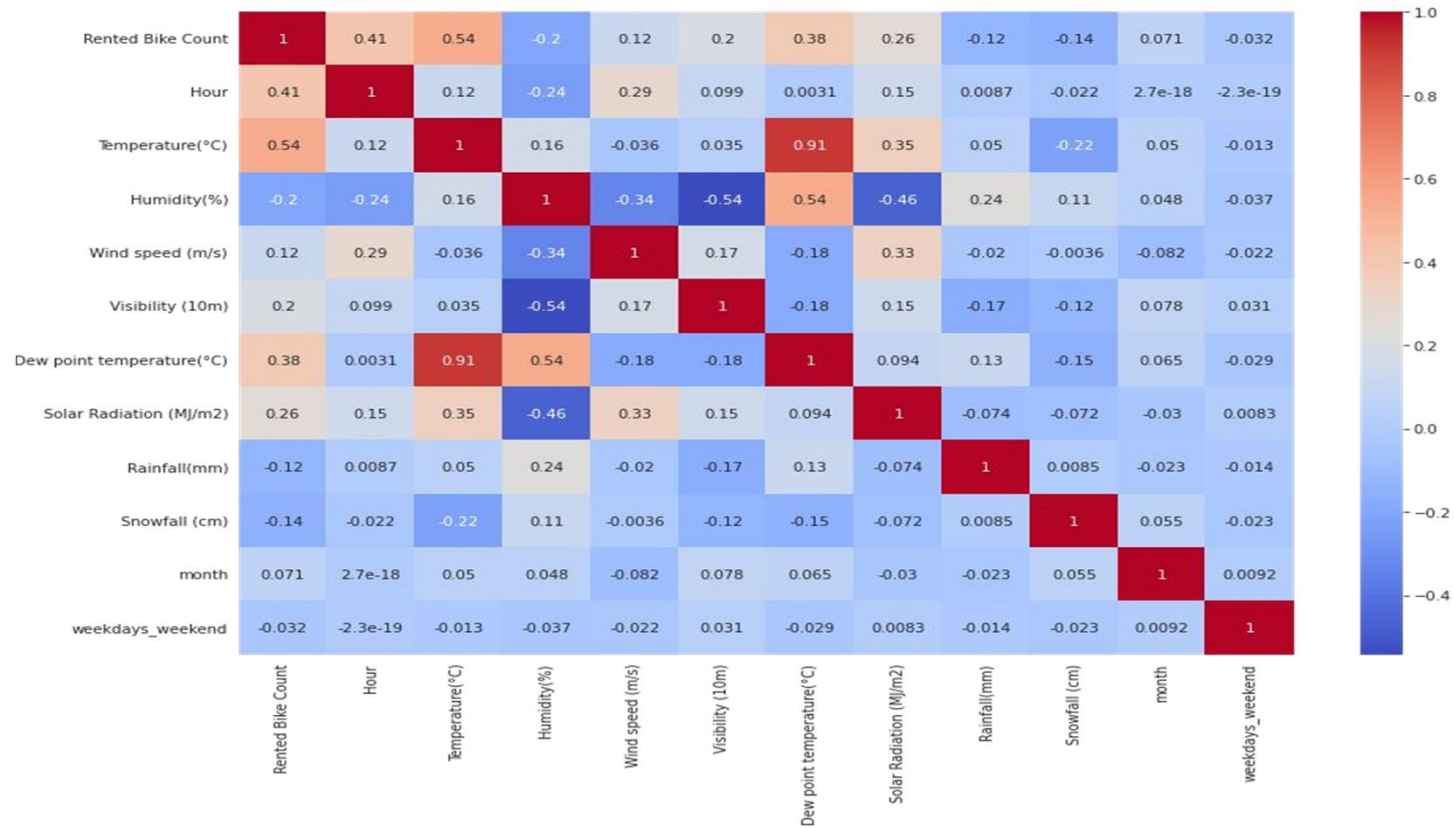
## Dependent variable:

- Rented Bike count - Count of bikes rented at each hour

## Independent variables:

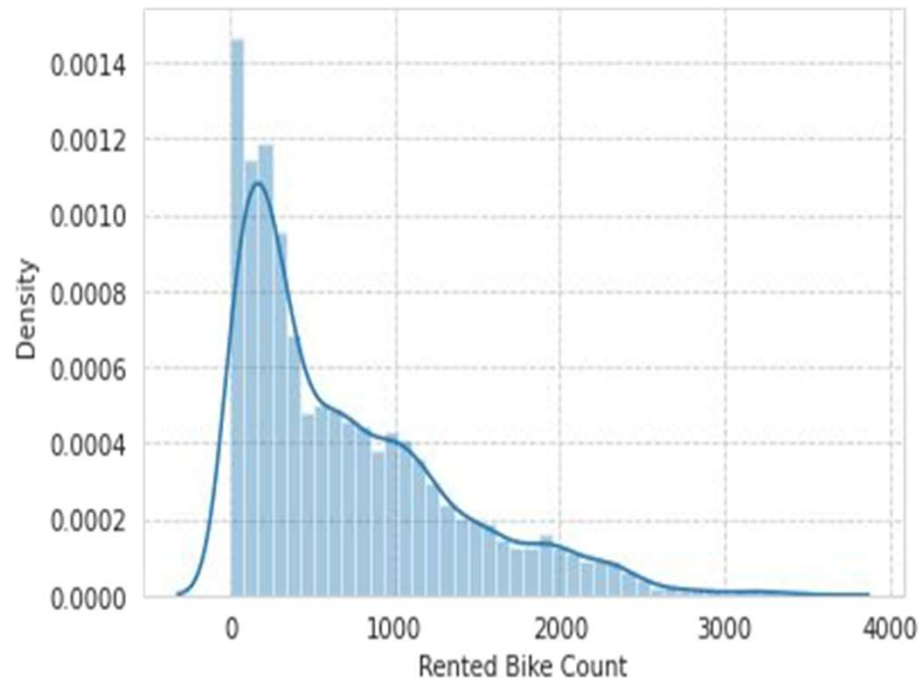
- Date : year-month-day
- Hour - Hour of the day
- Temperature-Temperature in Celsius
- Humidity - %
- Windspeed - m/s
- Visibility - 10 m
- Dew point temperature - Celsius
- Solar radiation - MJ/m<sup>2</sup>
- Rainfall - mm
- Snowfall - cm
- Seasons - Winter, Spring, Summer, Autumn
- Holiday - Holiday/No holiday
- Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

# EDA - Feature Correlation

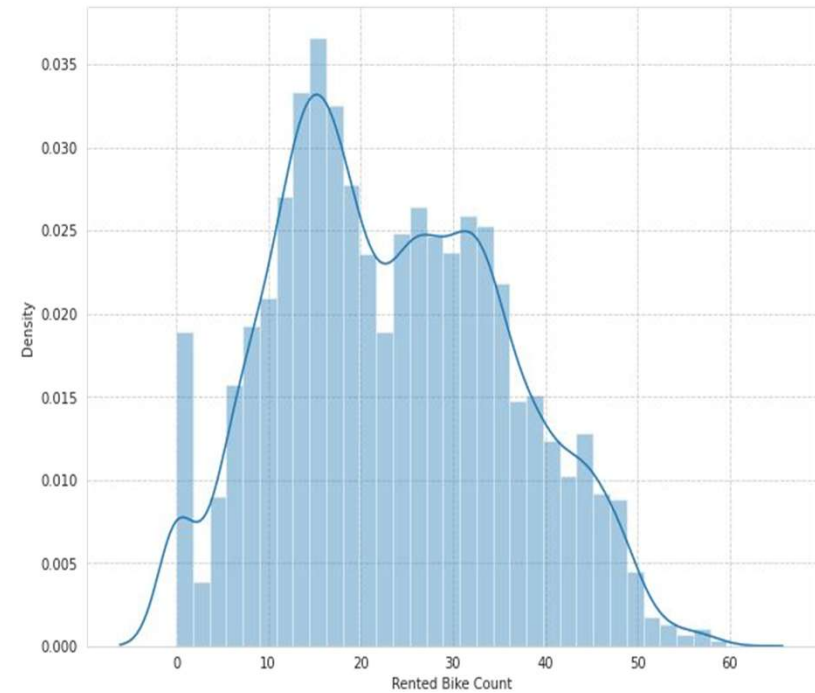


Correlation Graph

## EDA (contd...)

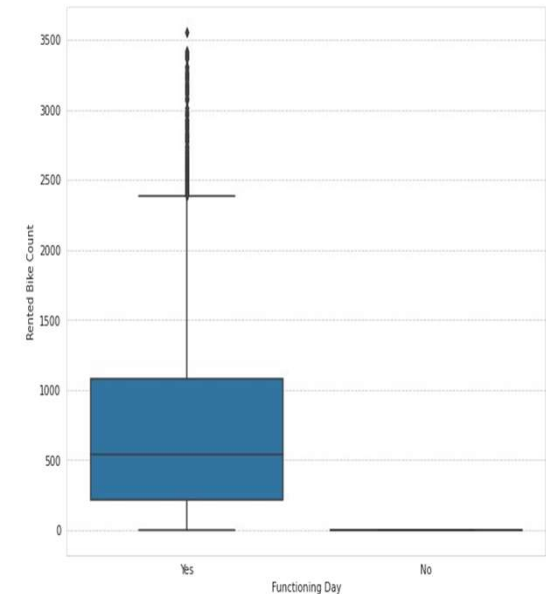
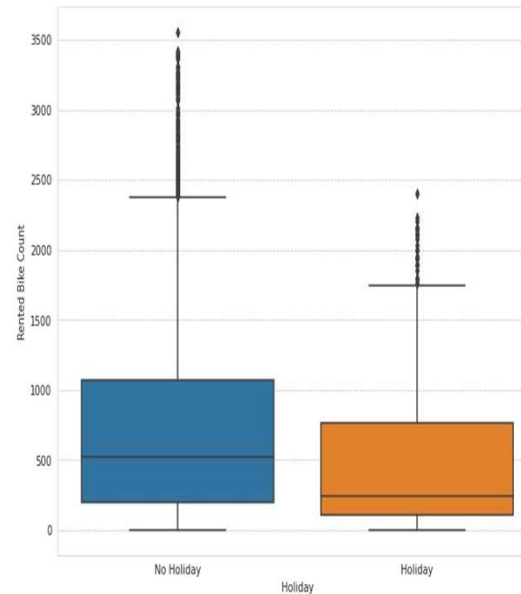
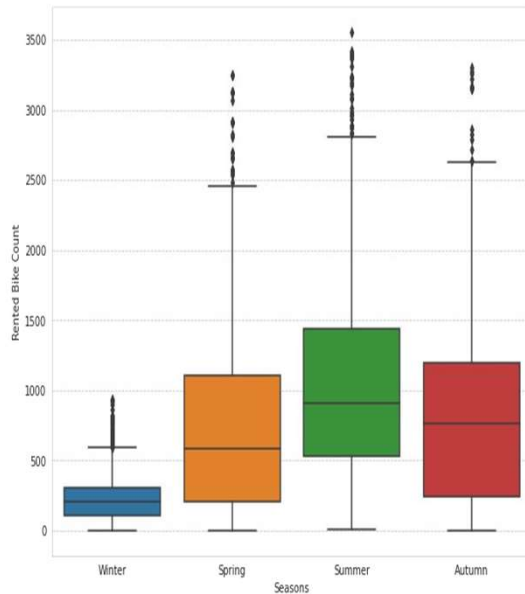


Distribution of rented bike count



Square root transformation of rented bike count

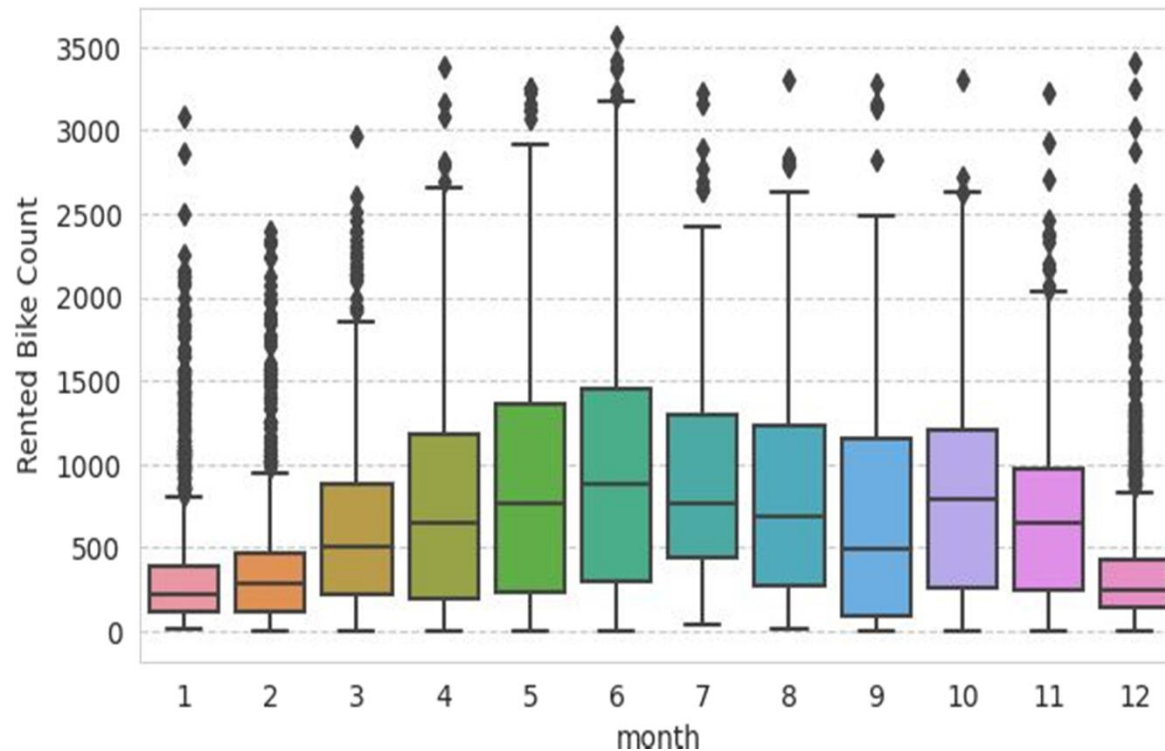
## EDA (contd...)



- Less demand on winter seasons
- Slightly Higher demand during Non holidays
- Almost no demand on non functioning day

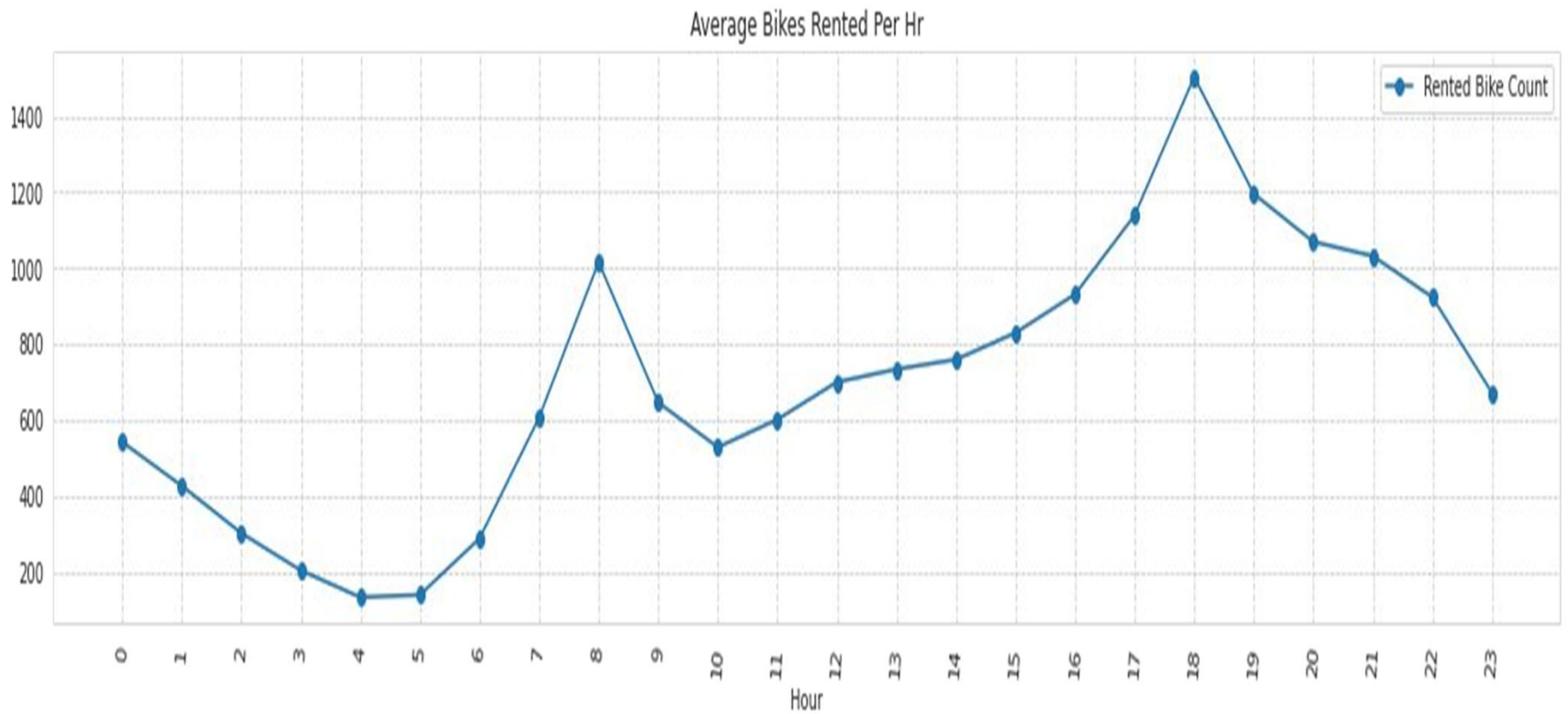


## EDA (contd...)



- We can see that there is less demand of Rented bike in the month of December, January, February i.e. during winter seasons
- Also demand of bike is maximum during May, June, July i.e. Summer seasons

## EDA (contd...)



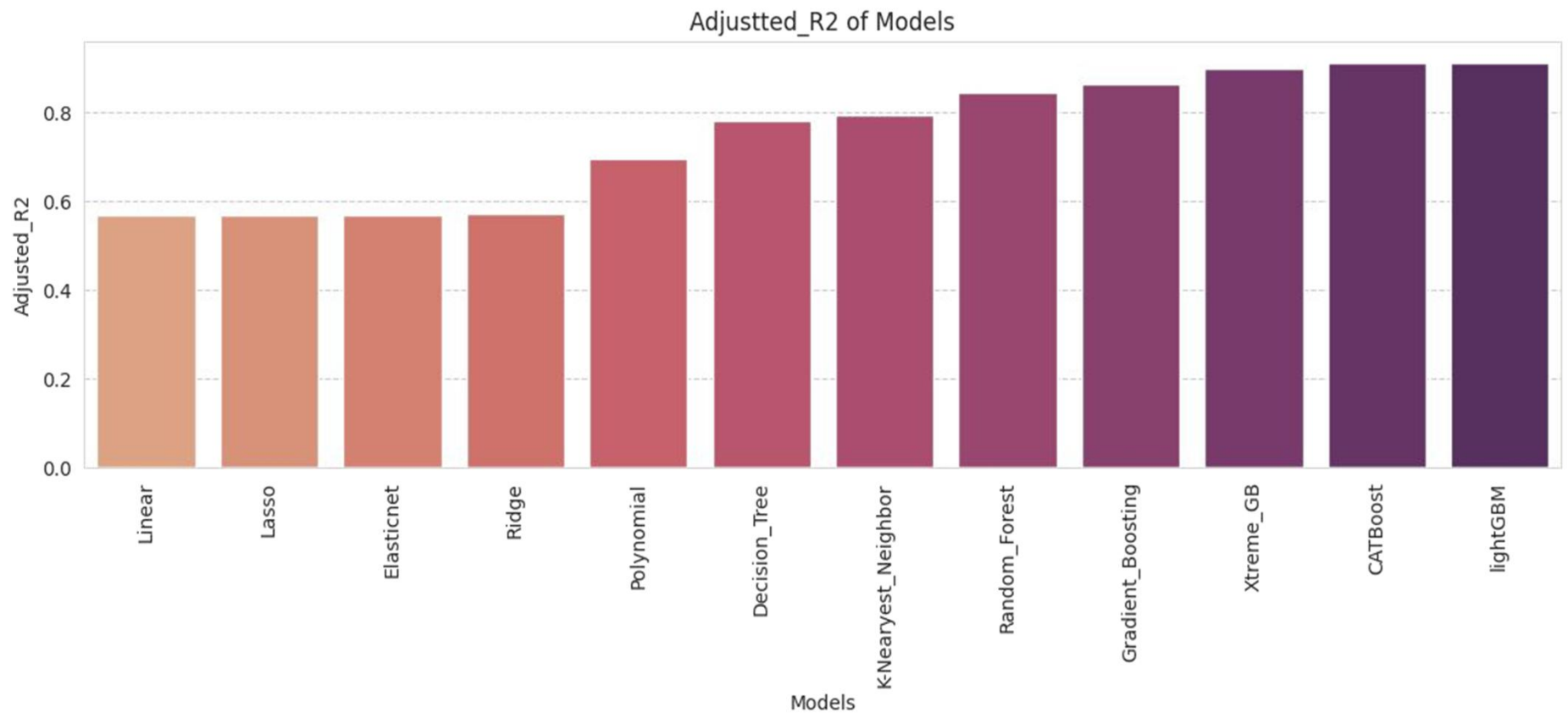
## Model's Performed

- Linear Regression with regularizations
- Polynomial Regression
- K nearest neighbours
- Decision tree
- Random forest
- Gradient Boost
- eXtreme Gradient Boost
- lightGBM
- CatBoost

# Model's Evaluation Matrices

	Models	Mean_square_error	Root_Mean_square_error	R2	Adjusted_R2
0	Linear	175590.552873	419.035264	0.572911	0.569766
1	Lasso	175560.907118	418.999889	0.572983	0.569839
2	Ridge	175248.935066	418.627442	0.573742	0.570603
3	Elasticnet	175346.867499	418.744394	0.573504	0.570363
4	Polynomial	123952.860328	352.069397	0.698509	0.696289
5	K-Nearyest_Neighbor	83411.759209	288.810940	0.796159	0.794659
6	Decision_Tree	88506.087215	297.499726	0.783710	0.782117
7	Random_Forest	62790.180423	250.579689	0.846554	0.845424
8	Gradient_Boosting	55090.172685	234.712958	0.865371	0.864380
9	Xtreme_GB	40812.801816	202.021785	0.900262	0.899528
10	CATBoost	36339.421527	190.629015	0.911194	0.910540
11	lightGBM	35410.753754	188.177453	0.913464	0.912826

# Adjusted R2 of Model's Performed

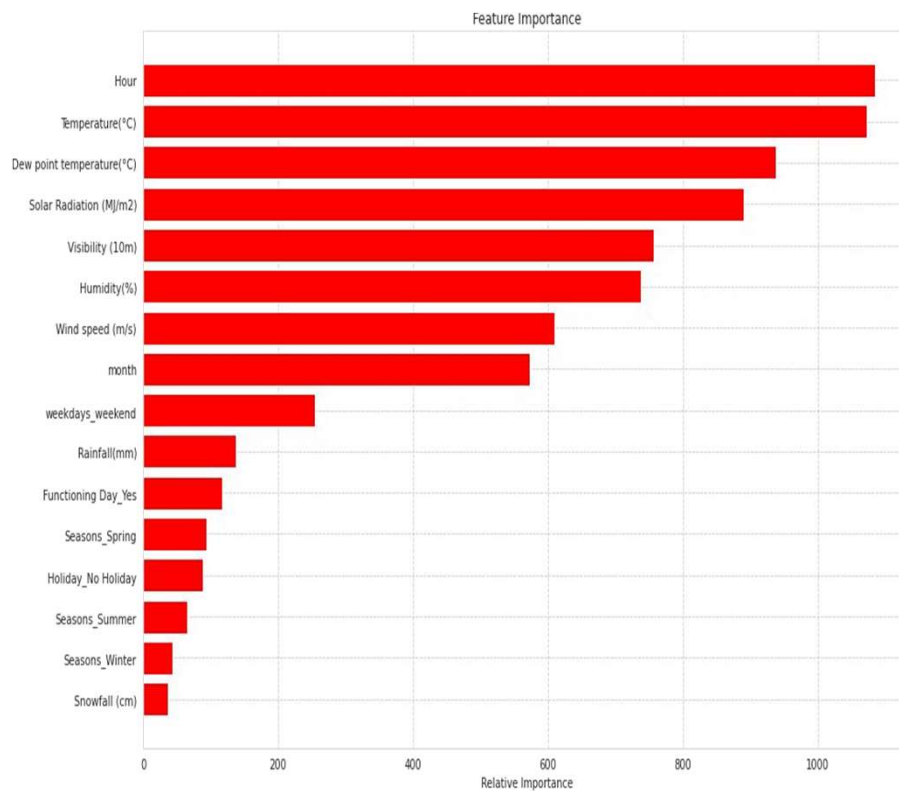


## Model Validation & Selection(continued)

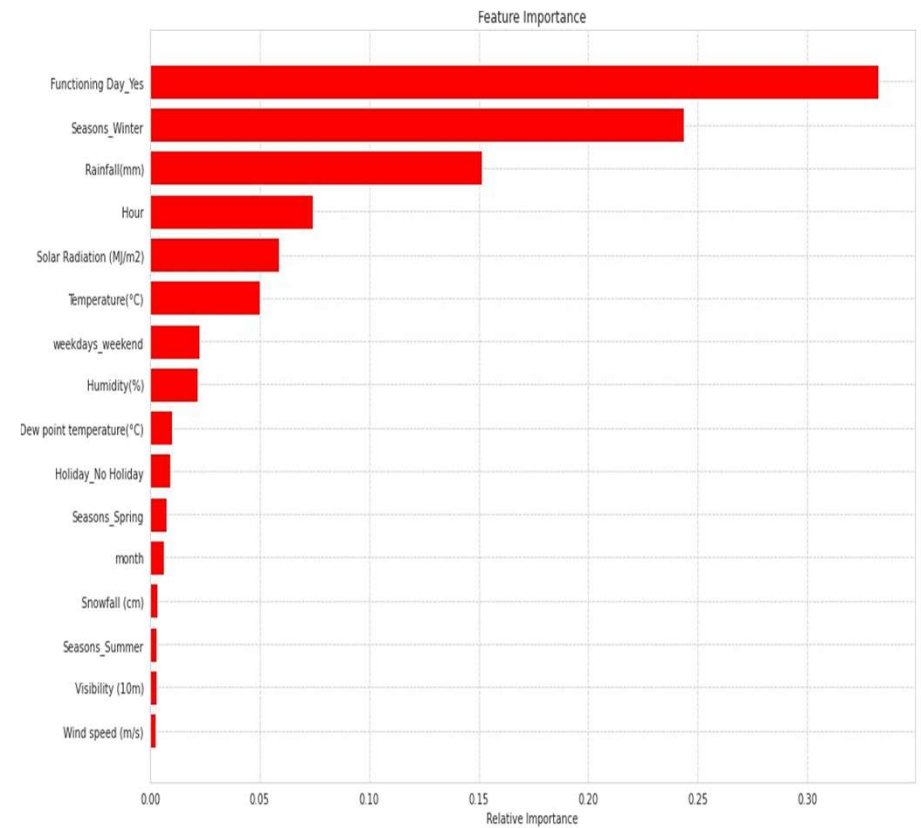
- **Observation 1:** As seen in the Model Evaluation Matrices table, Linear Regression, KNN is not giving great results.
- **Observation 2:** Random forest & GBR have performed equally good in terms of adjusted  $r^2$ .
- **Observation 3:** We are getting the best results from lightGBM and CatBoost.



# Feature Importance

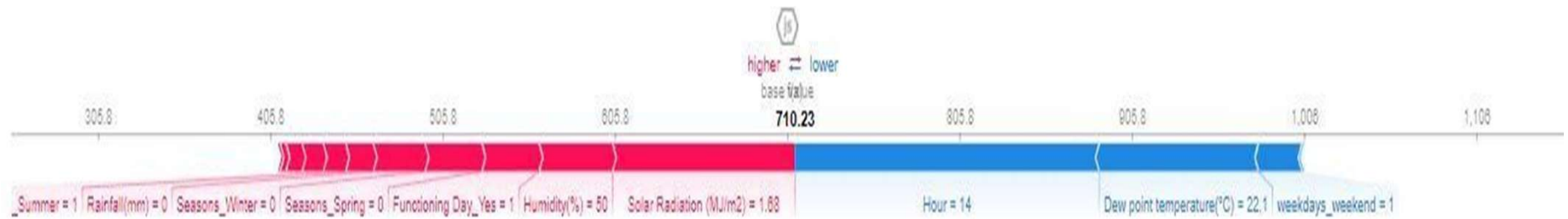


lightGBM



CatBoost

# Model Explainability - SHAP



lightGBM



CatBoost



# Challenges

- A huge amount of data needed to be dealt while doing the project which is quite an important task and also even small inferences need to be kept in mind.
- As dataset was quite big enough which led more computation time.



## Conclusion

- It is quite evident from the results that lightGBM and Catboost is the best model that can be used for the Bike Sharing Demand Prediction since the performance metrics (mse,rmse) shows lower and (r2,adjusted\_r2) show a higher value for the lightGBM and Catboost models.
- So, we can use either lightGBM or catboost model for the above problem





**THANK  
YOU**