Capstone Project

Bike Sharing Demand Prediction

BY – AYUSH SHARMA

CONTENT

- Problem Statement
- Data Description
- EDA
- Models Used
- Models Evaluation metrics
- Model Selection
- Conclusion

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.

DATA DESCRIPTION

- The Dataset that was provided consisted of 8760 rows and 15 columns.
- Apart from four columns all the columns were of numeric data type.
- Thankfully there were not a single null value in the dataset.

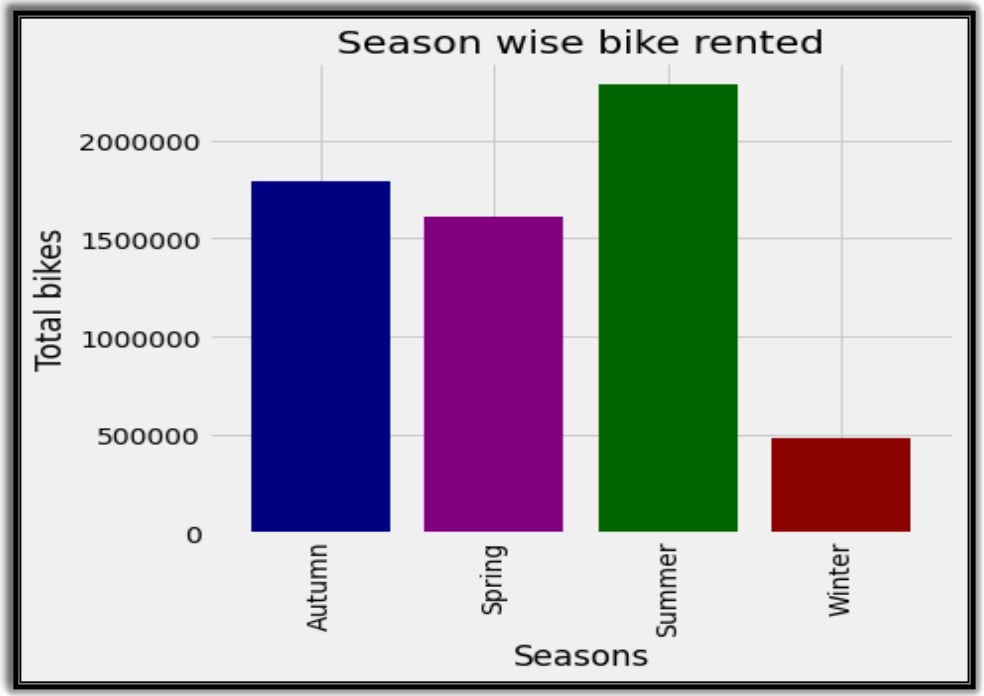
Dependent Variable - Rented Bike Count

Independent Variables

- Date
- Hour
- Temperature
- Humidity
- Windspeed
- Visibility
- Dew Point Temperature
- Solar Radiation
- Rainfall
- Snowfall
- Seasons
- Holiday
- Functional Day

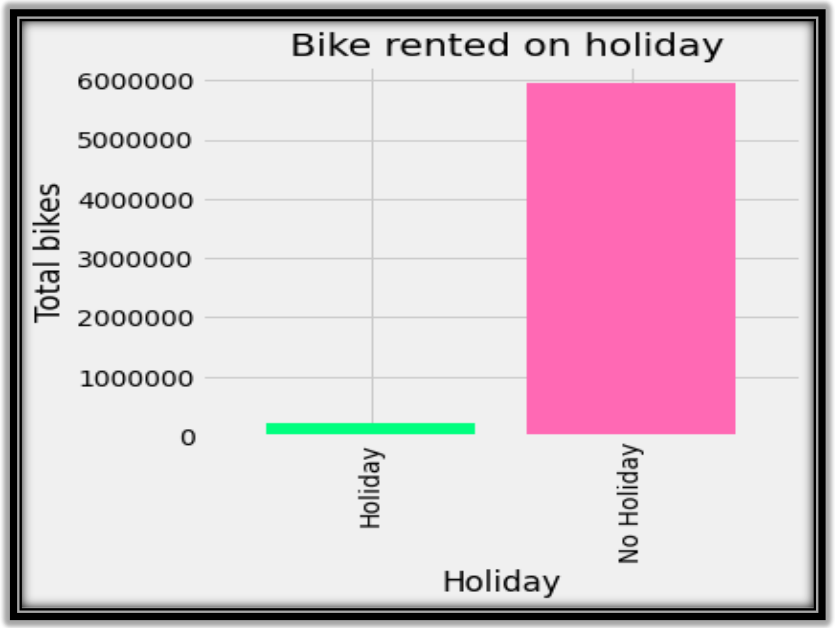
EXPLORATORY DATA ANALYSIS (EDA)

Season Wise Bikes Rented



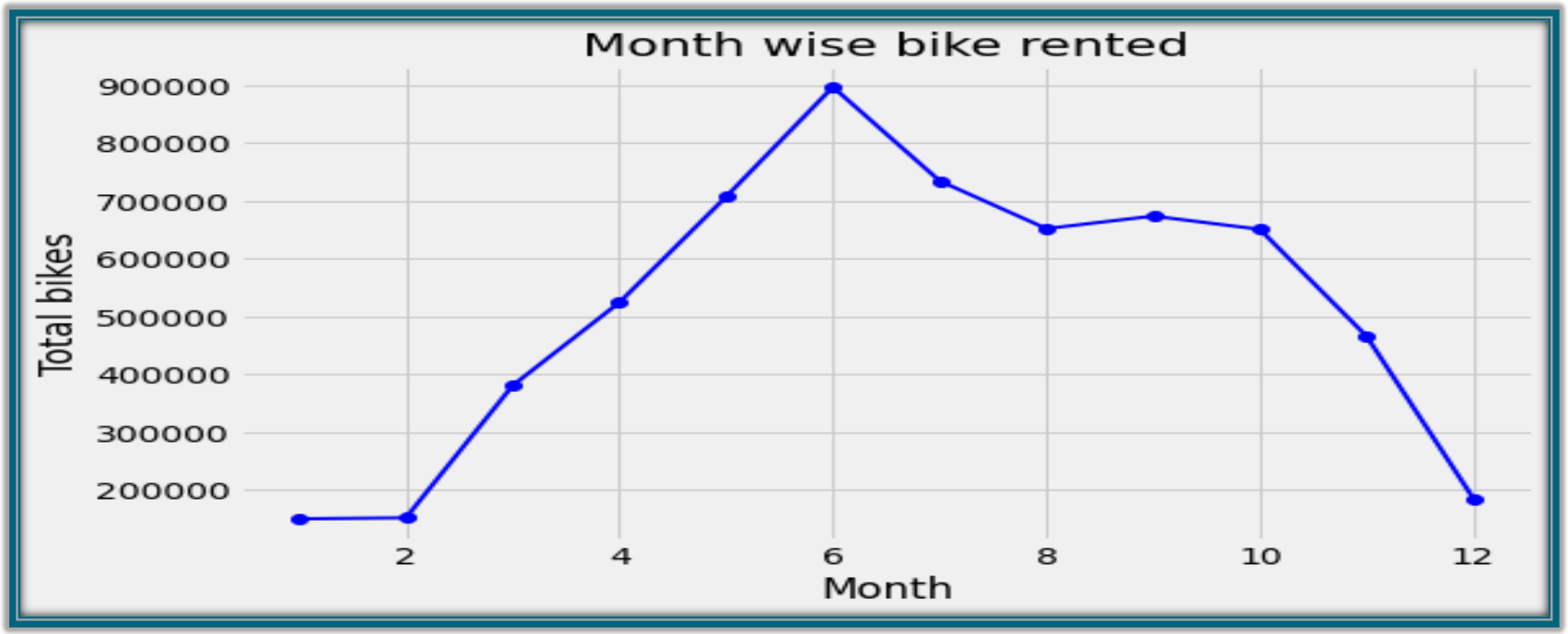
We can clearly see that most numbers of bikes were rented in Summer season followed by Autumn and Spring.

Bikes Rented On Holidays



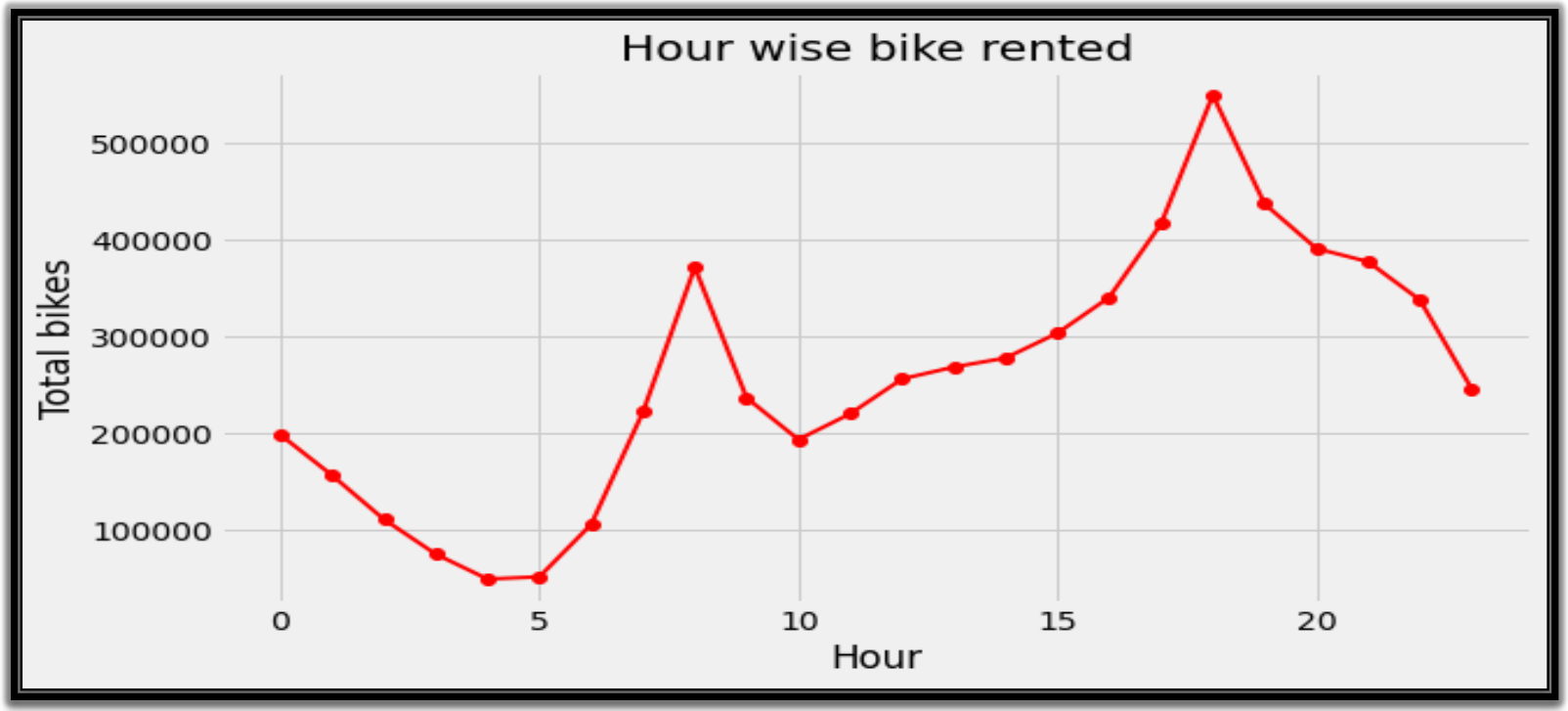
We can clearly see that maximum number of bikes were rented on days with no holidays

Bikes Rented Month Wise



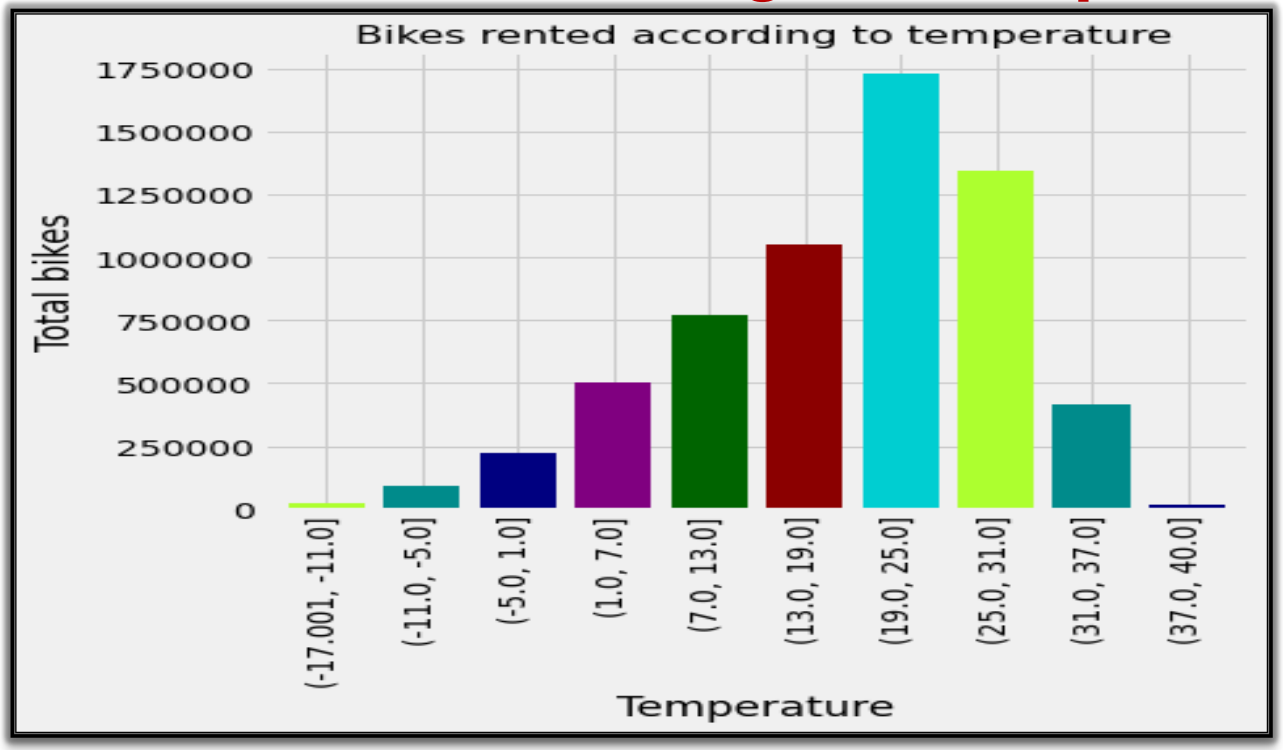
Most number bikes were rented in the month of June followed by July and may

Hourly Trend of Bikes Rented



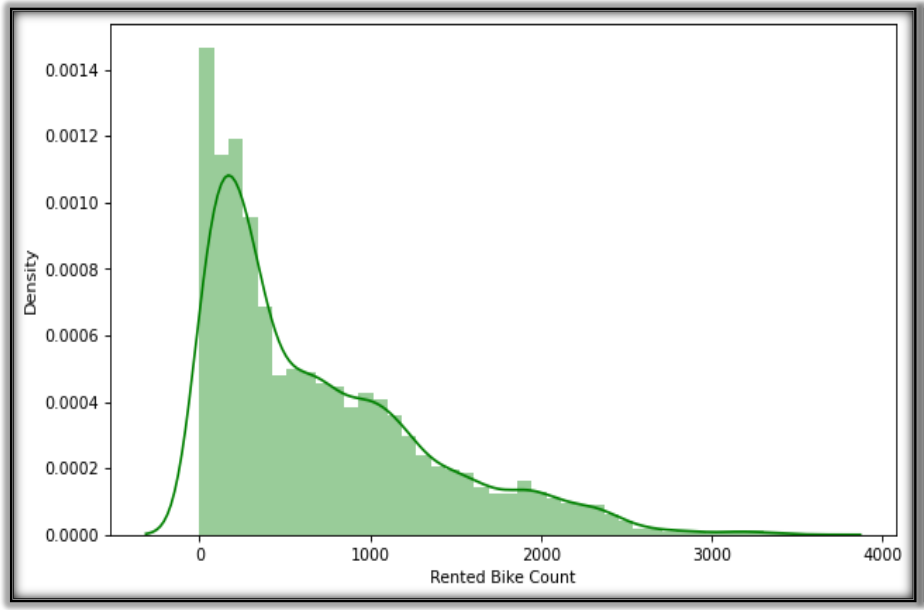
We can see sudden surge in bikes rented at 8.am and 6.pm

Bikes Rented According To Temperature

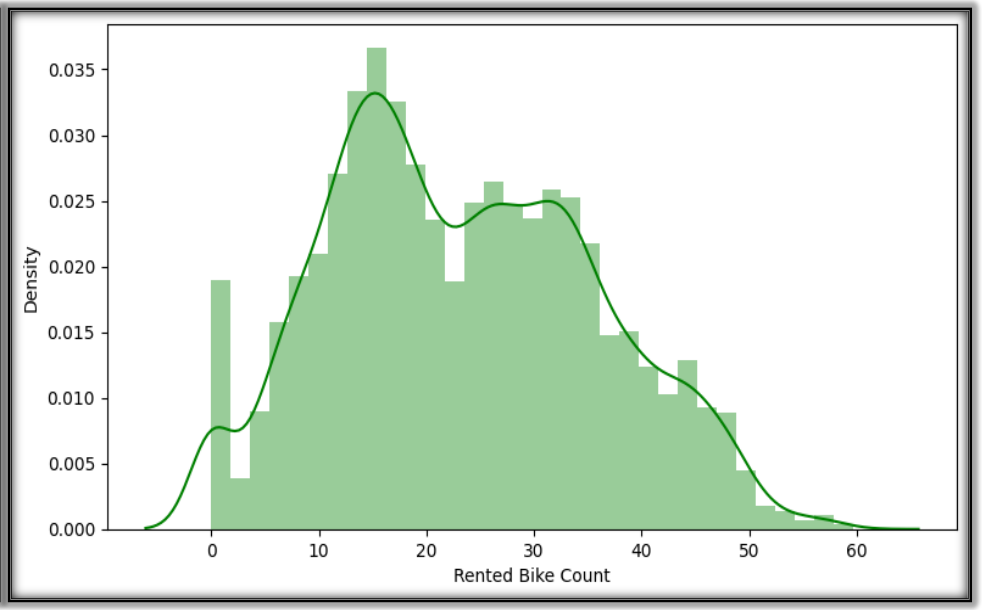


We can clearly see that people tends to love riding bikes when the temperature is between 19 to 25 °C

Dependent Variable

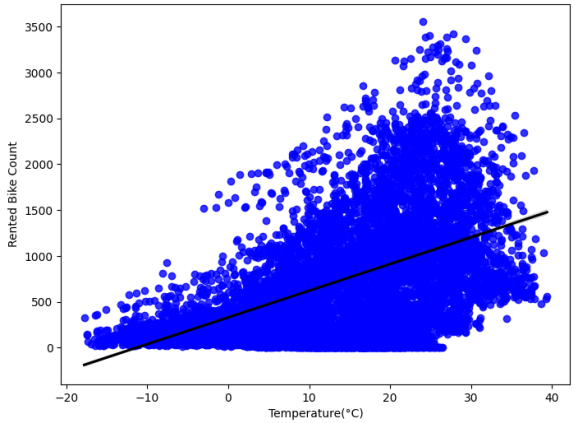


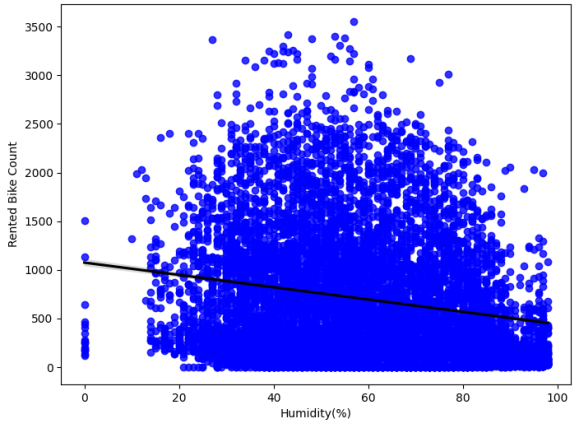
Distribution Plot Of Rented Bike Count

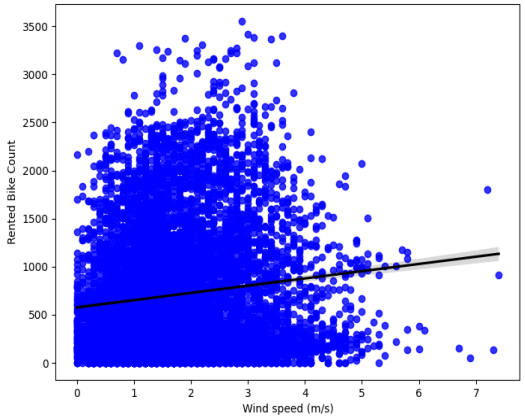


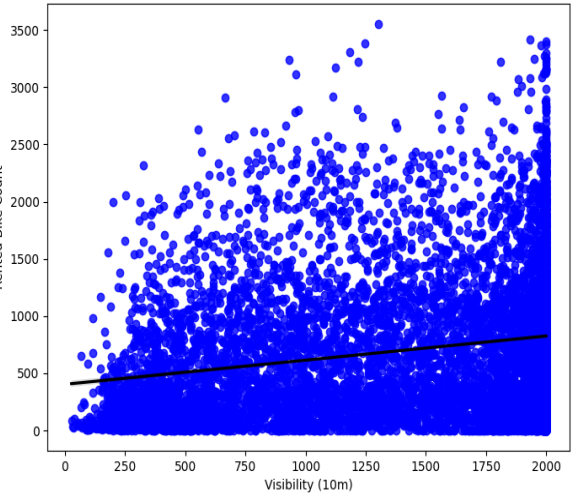
After Square Root Transformation

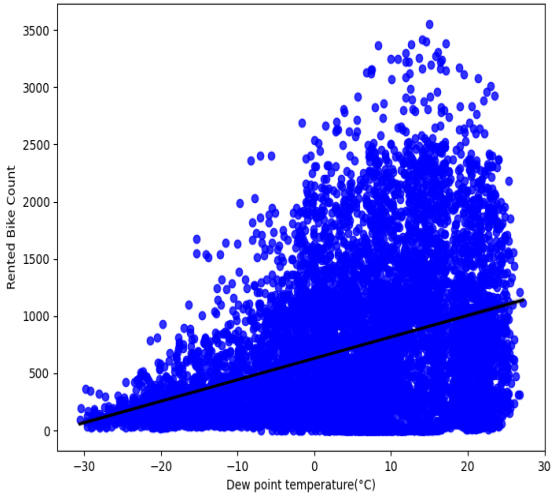
Regression Plot

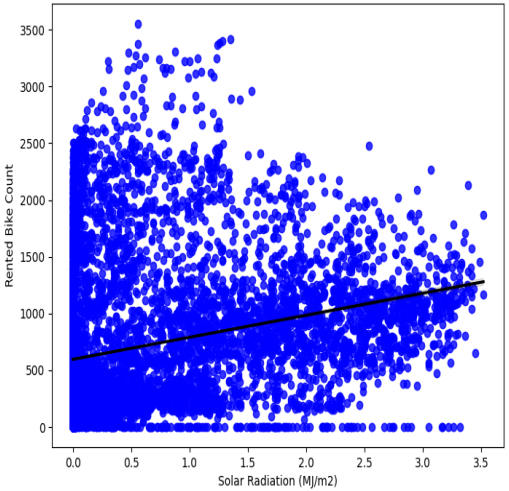




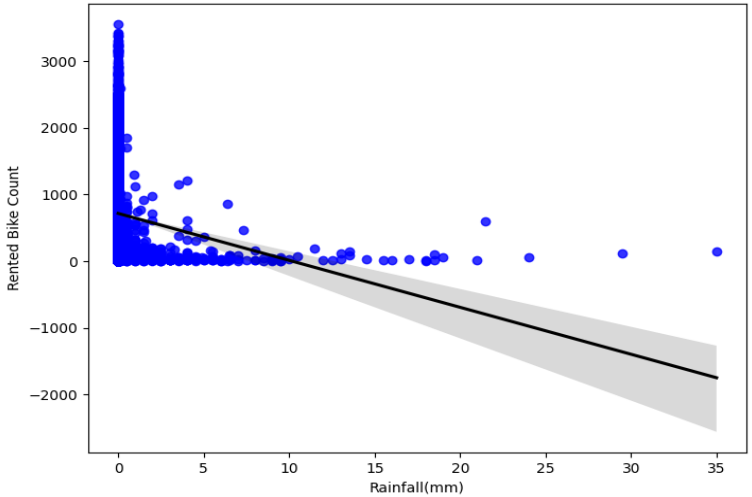


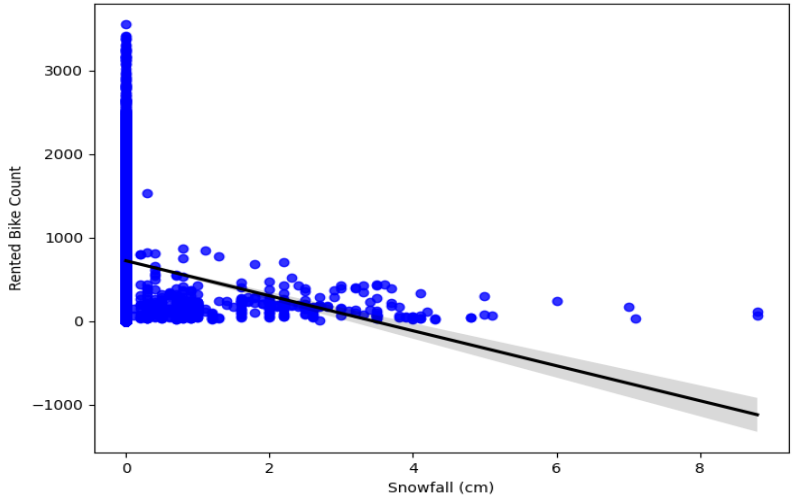






Regression Plot (condt…)





We can clearly find out that variables temperature, dew point temperature, wind speed , solar radiation, visibility has a positive relation with the dependent variable.

Whereas variables rainfall, snowfall, humidity has negative relation with the dependent variable.

Models Used

- Linear Regression
- LASSO Regression
- RIDGE Regression
- Elasticnet
- Decision Tree
- Random Forest

Models Evaluation Metrics

Models	MSE	RMSE	R2	Adjusted R2
Linear Regression	34.34	5.86	0.783	0.772
LASSO Regression	34.14	5.84	0.783	0.773
RIDGE regression	34.35	5.86	0.781	0.771
Elasticnet	59.93	7.74	0.619	0.602
Decision Tree	100.92	10.04	0.359	0.329
Random Forest	17.37	4.16	0.889	0.884

Model Selection

- By looking at the Evaluation metrics of each the model we can clearly see that the best performing model is the Random Forest model.
- Whereas the worst performing model was the Decision Tree model.
- The performance of LASSO ,RIDGE ,Linear Regression is quite similar.

Conclusion

We can evidently say that the Random Forest model is the best model for Bike sharing demand prediction . Since the evaluation metrics MSE and RMSE are the lowest among other and the R2 and Adjusted R2 score is highest among the other models that were used.

THANK YOU