Capstone Project
Bike Sharing Demand Prediction

Sharath Diwakar

Contents

Introduction

- Currently rental bikes are introduced in many urban cities for the enhancement of mobility comfort. The purpose of this movement is to **modernize** cities and encourage people to head to **a green world**. Let's take the examples of Paris in 2007, where "velibs" were introduced and Amsterdam, where there are more bikes than cars. The goal is to **facilitate** the commute in the Seoul and reduce the amount of cars and the pollution. Indeed, the development of the way to commute reduced the use of cars to go to work and visit the city.

- It is important to make the rental bike available and accessible to the public, as it provides many alternatives to commuters in metropolises. There are a lot of **advantages** to bike rents, it is **convenient** because it permits people not to keep the bike all day long, whether it is at work or at school. Furthermore it is the healthiest way to travel and it has many environmental benefits.

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 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:

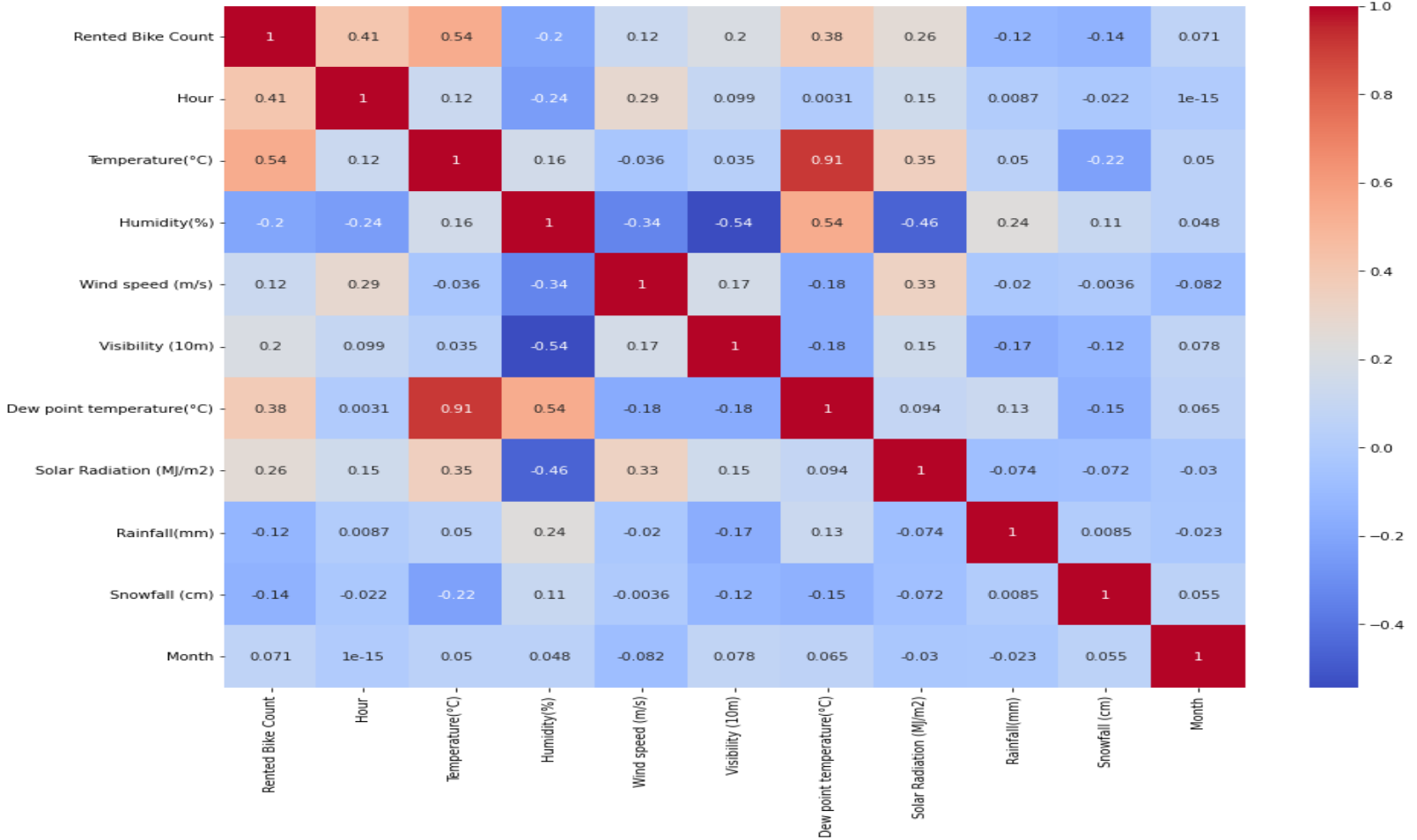
- Rented Bike count - Count of bikes rented at each hour

Independent variables:

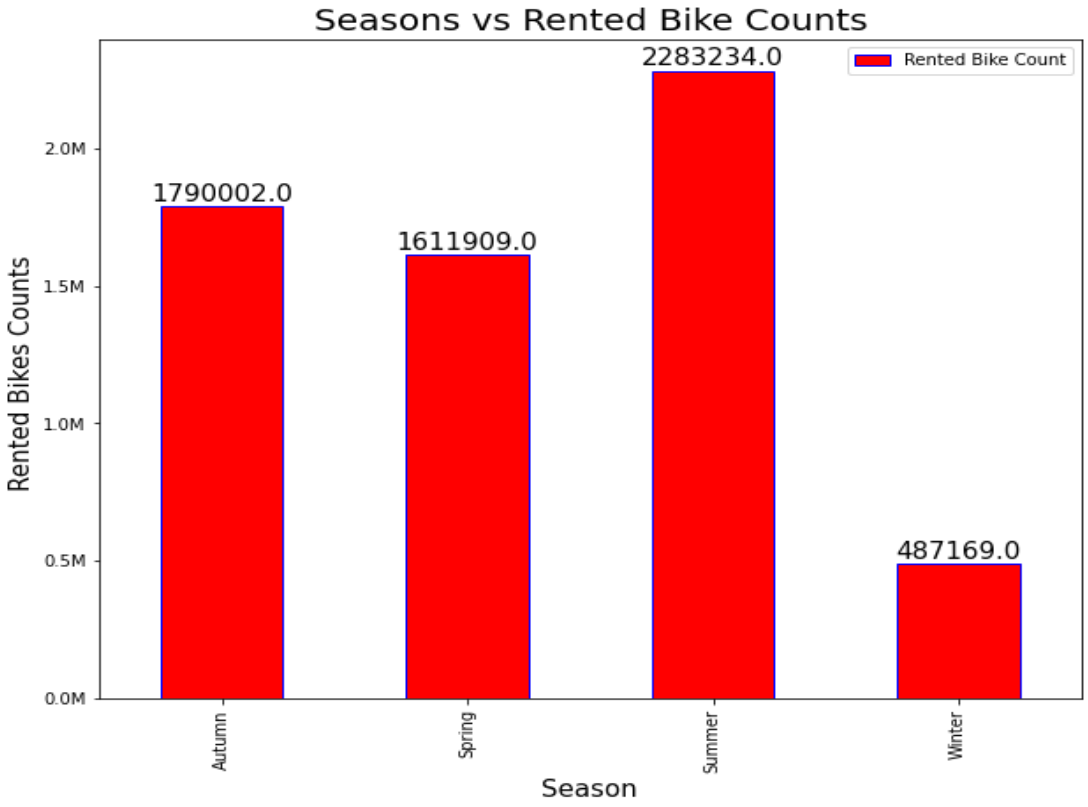
- Date : year-month-day
- Hour - Hour of he day
- Temperature-Temperature in Celsius
- Humidity - %
- Windspeed - m/s
- Visibility - 10 m
- Dew point temperature - Celsius
- Solar radiation - MJ/m2
- Rainfall - mm
- Snowfall - cm
- Seasons - Winter, Spring, Summer, Autumn
- Holiday - Holiday/No holiday
- Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

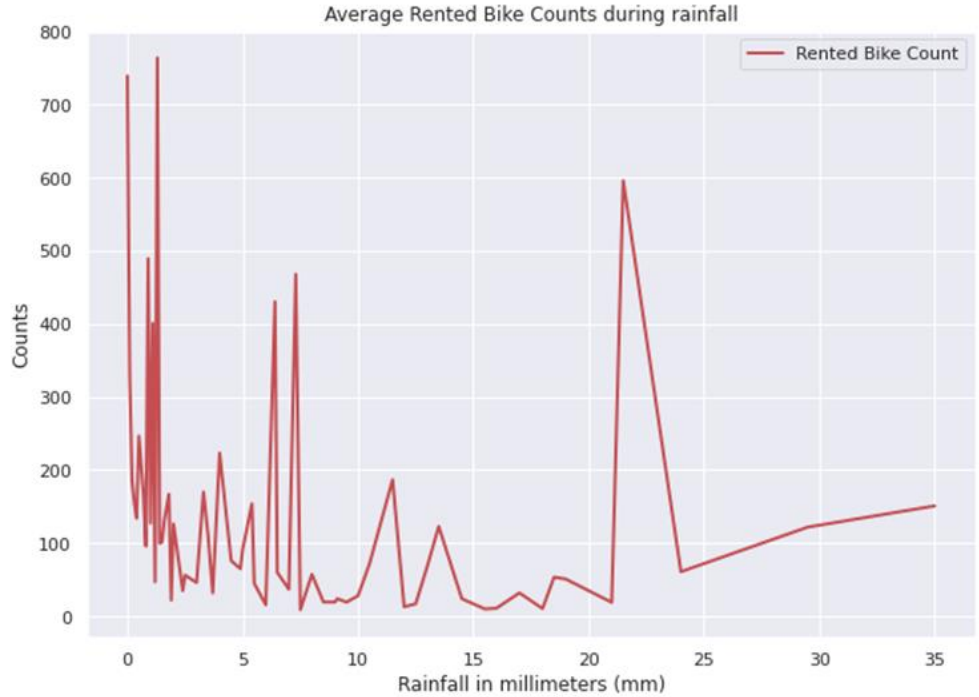
Exploratory Data Analysis (EDA)

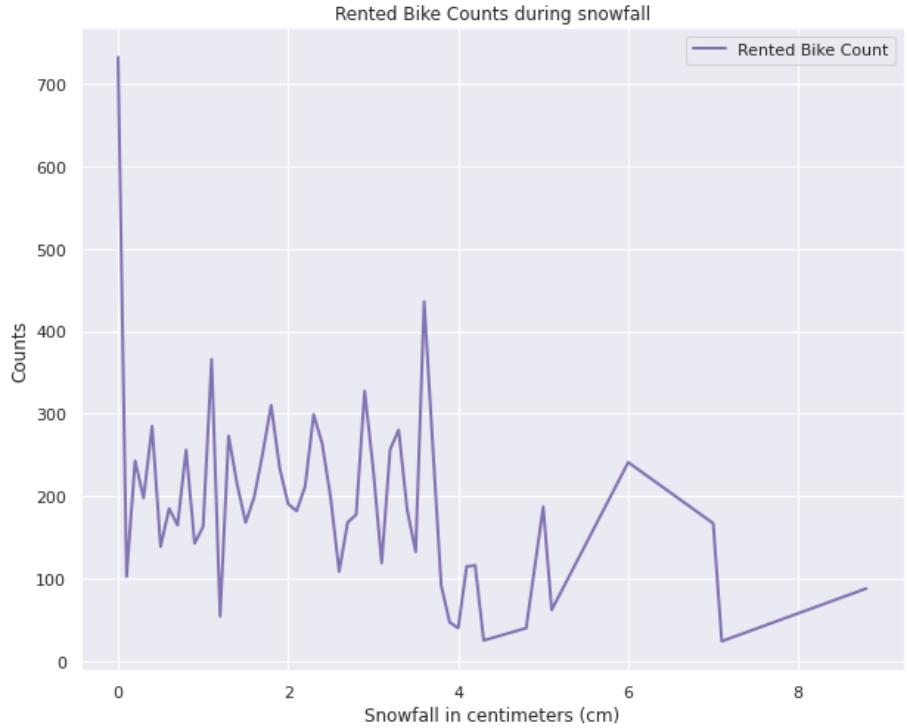
Feature Correlation

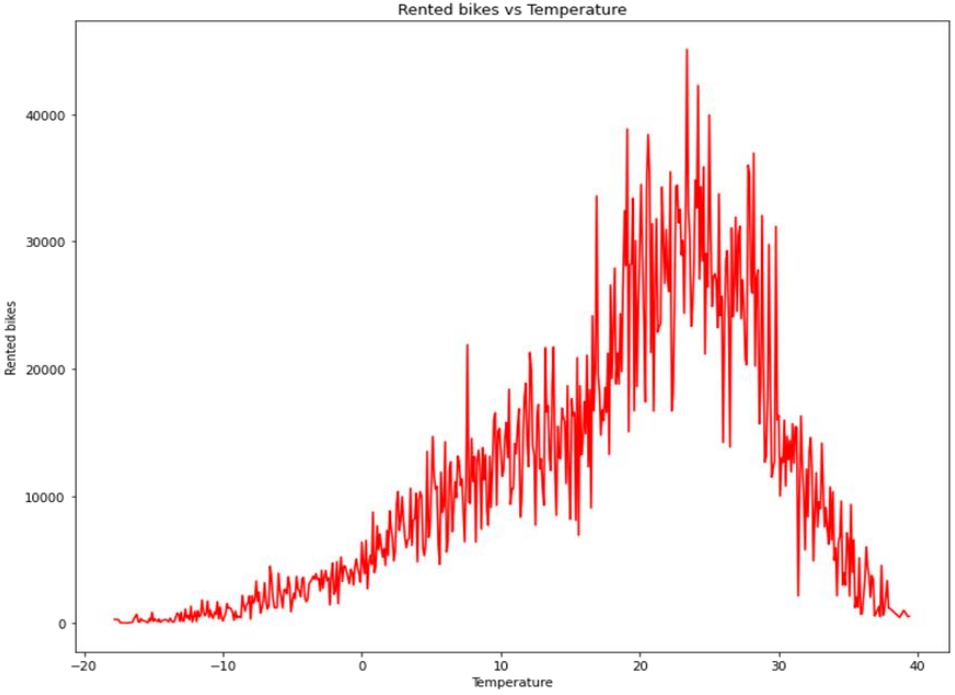


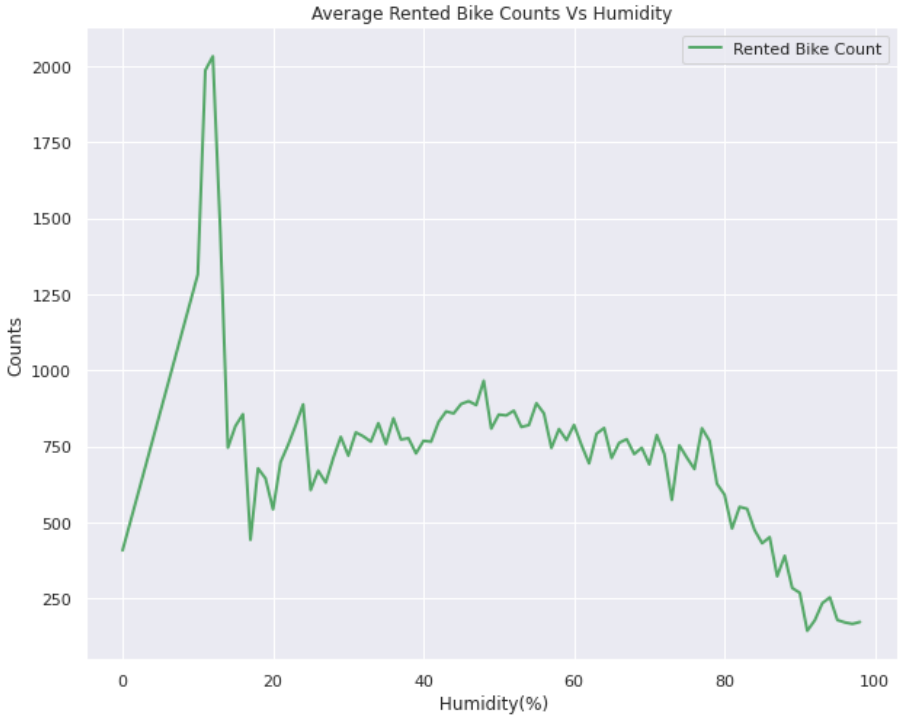
Effect of seasons & temperature on the number of rented bikes


Seasons vs Rented Bike Counts
Rented Bike Count
2283234.0
1790002.0
1611909.0
487169.0
2.0M
1.5M
1.0M
0.5M
0.0M
Rented Bikes Counts
Autumn
Spring
Summer
Winter
Season


Average Rented Bike Counts during rainfall
Rented Bike Count
Counts
Rainfall in millimeters (mm)
0
100
200
300
400
500
600
700
800
0
5
10
15
20
25
30
35

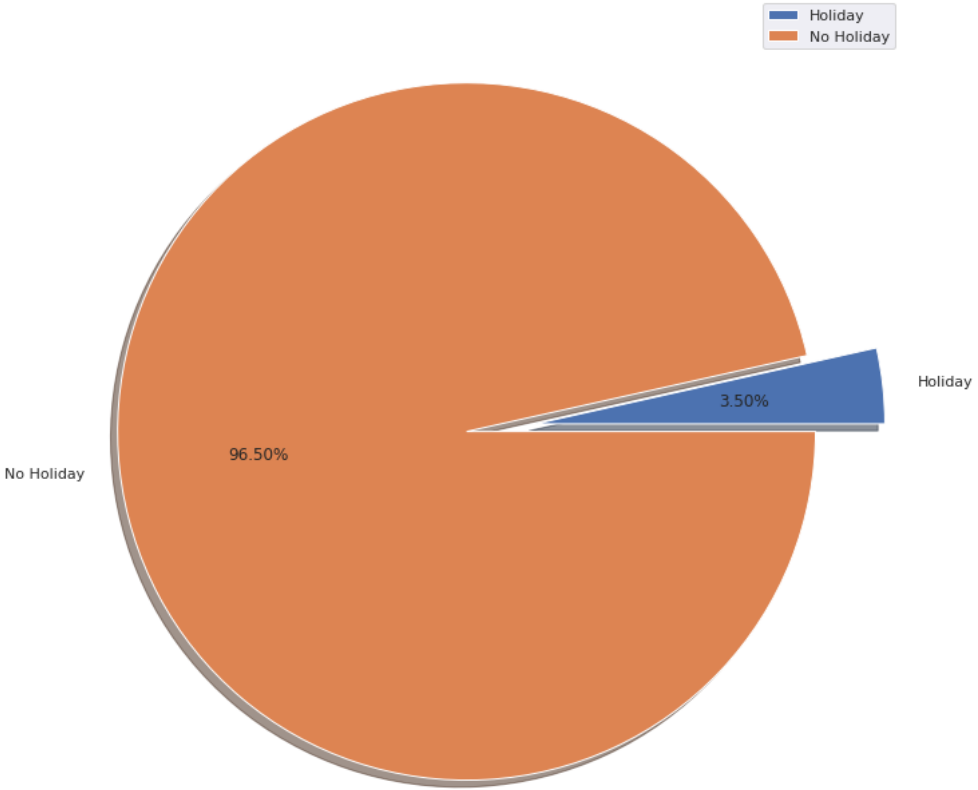

Rented Bike Counts during snowfall
Rented Bike Count
Counts
Snowfall in centimeters (cm)
0
100
200
300
400
500
600
700
0
2
4
6
8


Rented bikes vs Temperature
Rented bikes
40000
30000
20000
10000
0
−20
−10
0
10
20
30
40
Temperature

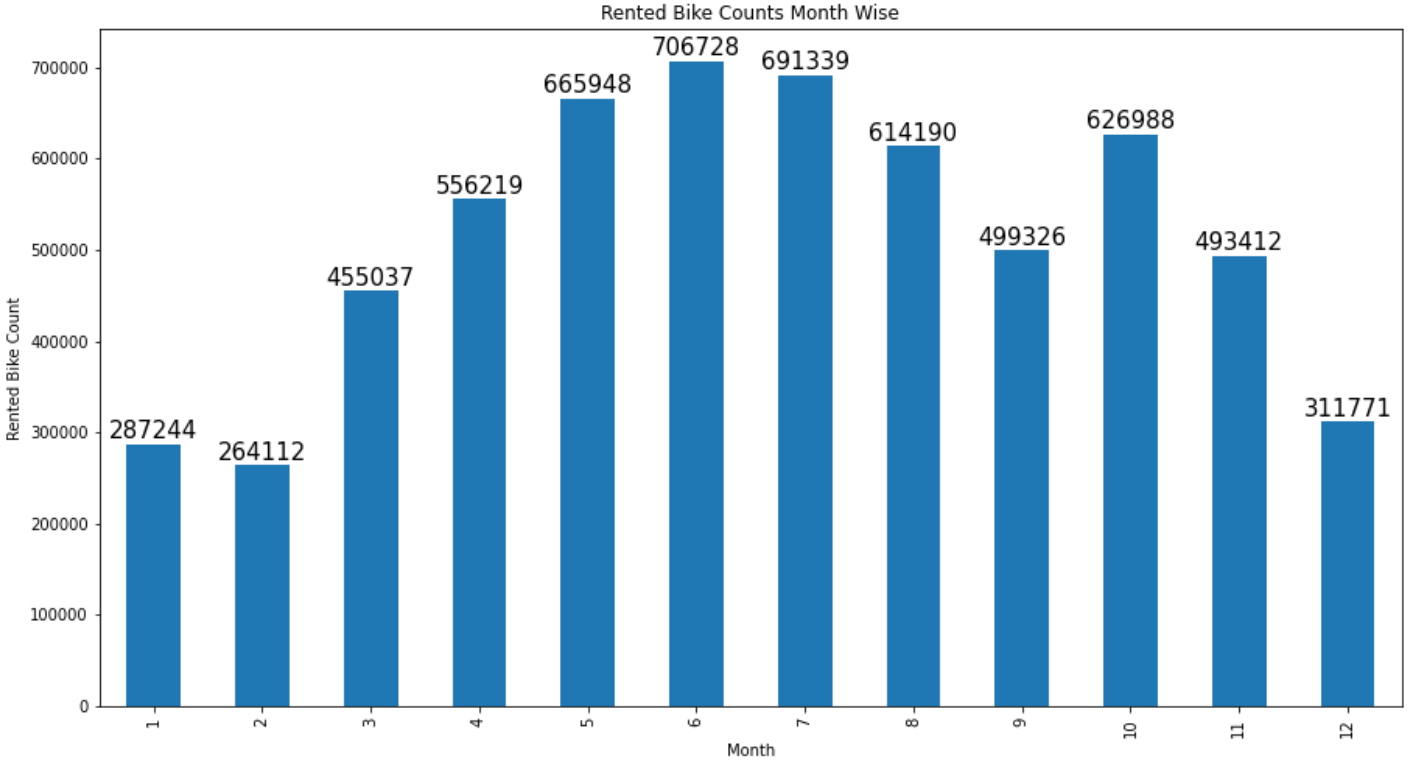

Average Rented Bike Counts Vs Humidity
Rented Bike Count
Counts
2000
1750
1500
1250
1000
750
500
250
0
20
40
60
80
100
Humidity(%)

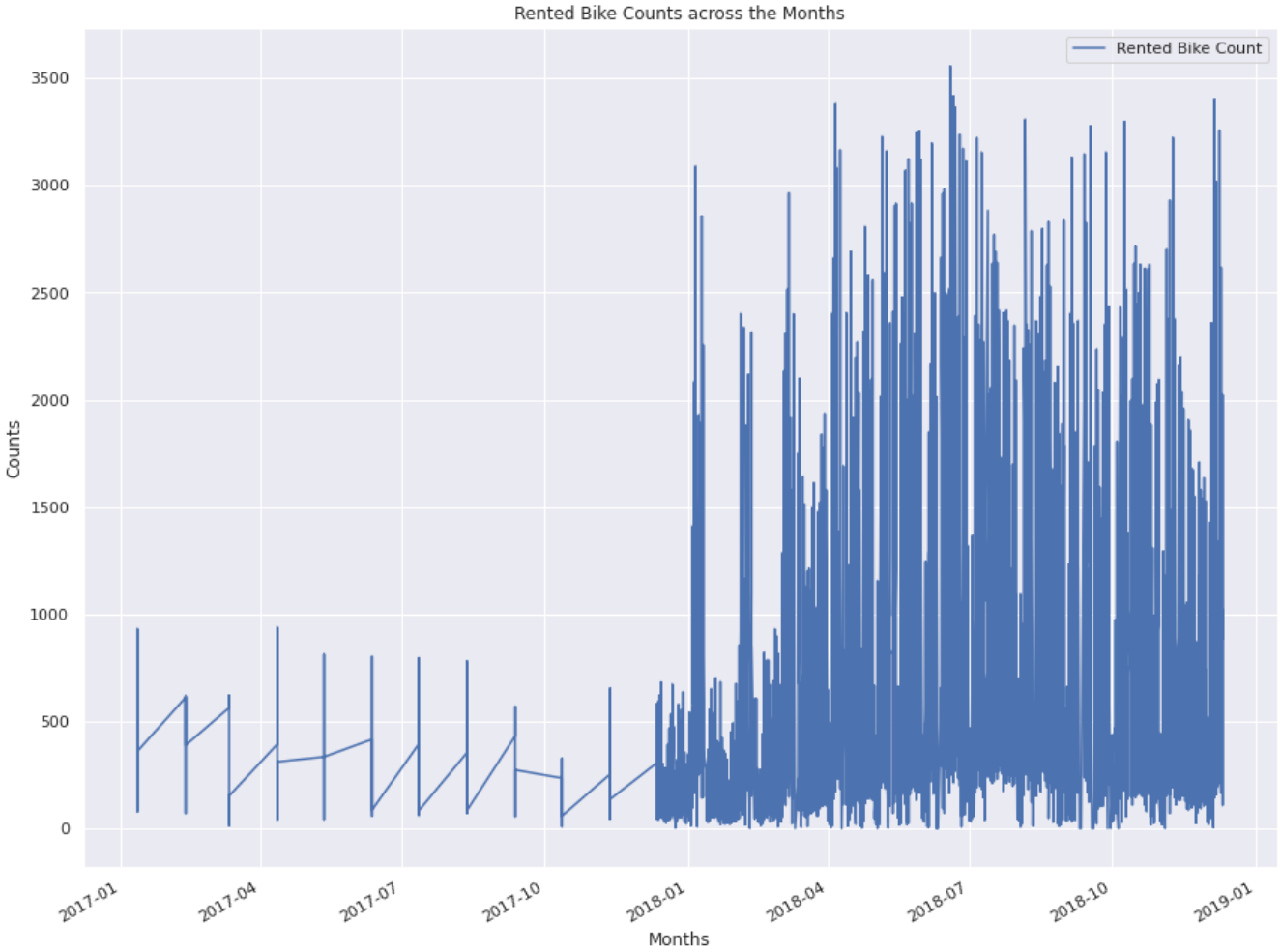
Bike renting characteristics on various days

Bike Renting on holidays

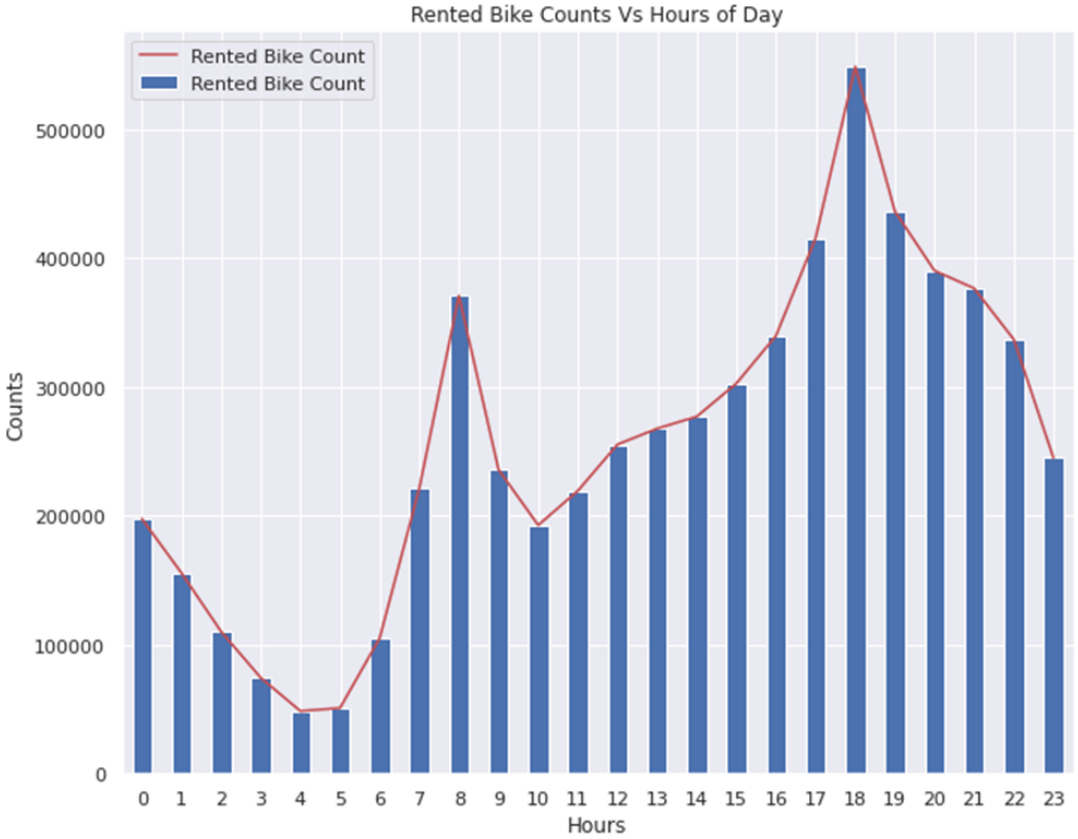


Month-wise rented bikes count

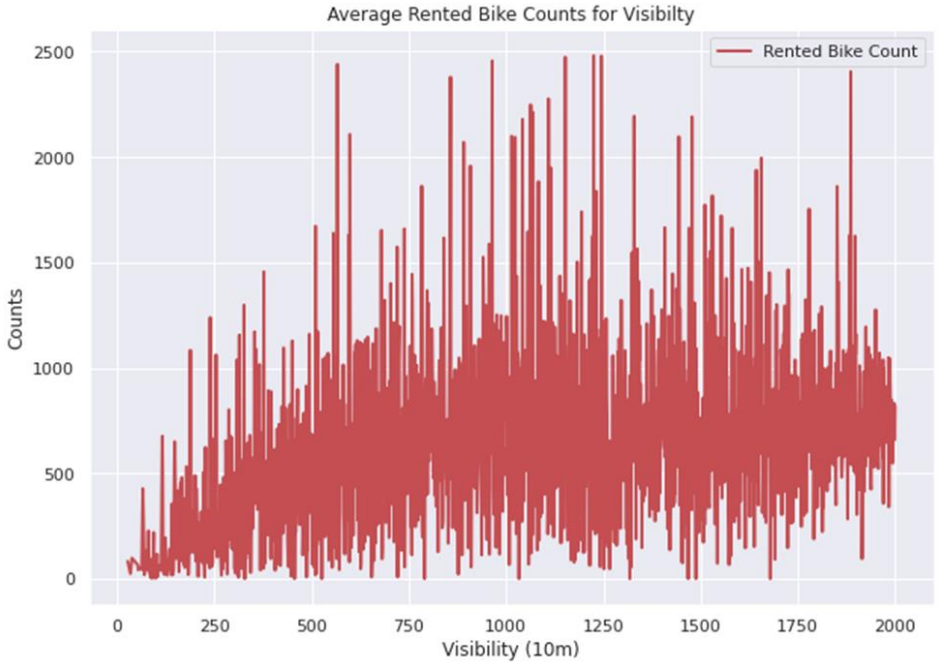


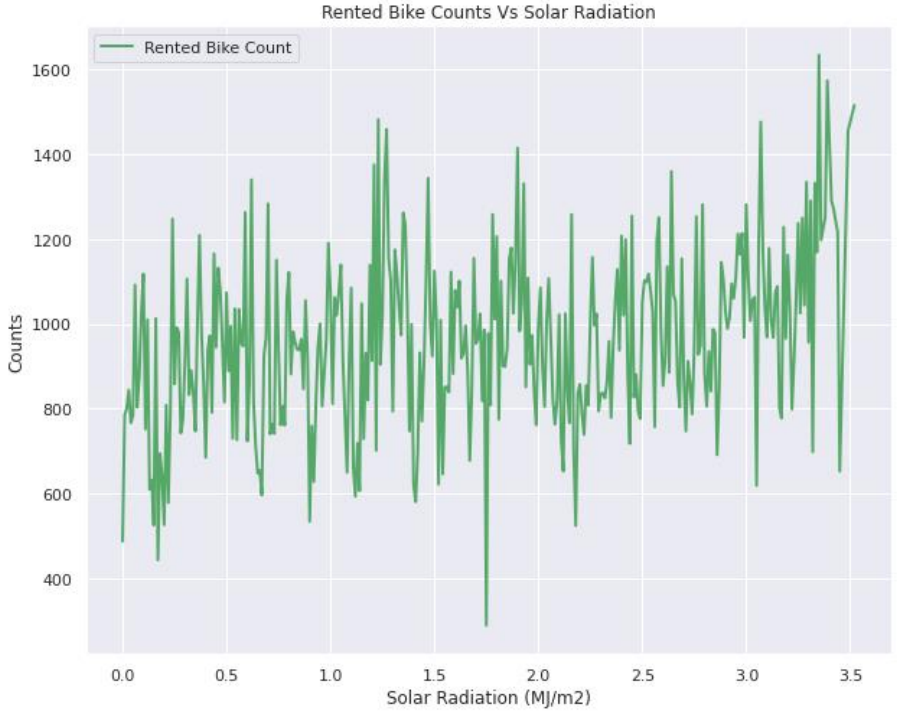

Rented Bike Counts across the Months
Rented Bike Count
Counts
Months
0
500
1000
1500
2000
2500
3000
3500
2017-01
2017-04
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10
2019-01

Bike renting at different Times of the day

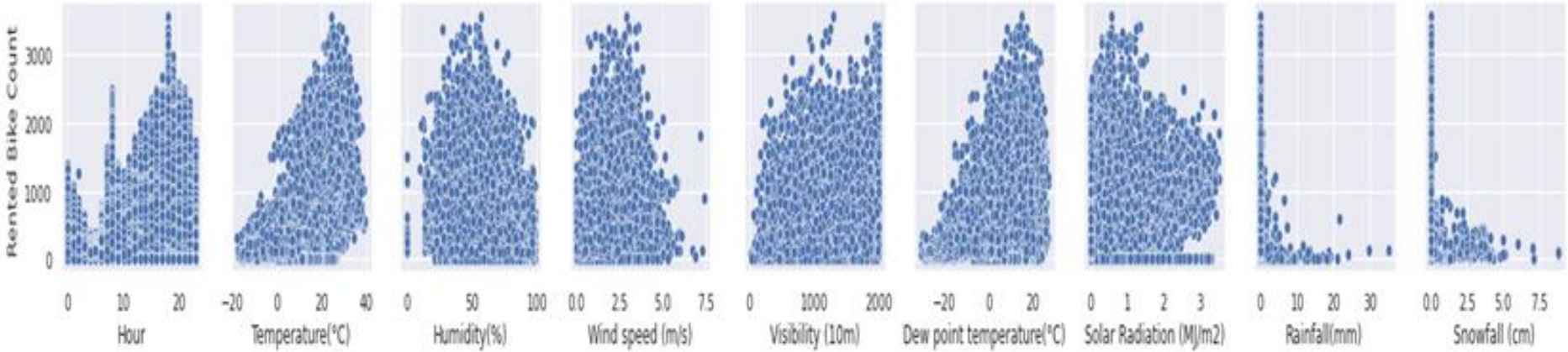


Effect of visibility and solar radiation on bike riders





Distribution between the dependent and independent variables



Models Performed

- Linear Regression
- Ridge & Lasso Regression
- Decision Tree
- Random Forest

Models Evaluation Matrices

Model_Name	Training_score	MSE	RMSE	R-squared	Adjusted R-Squared
Linear Regression	0.554335	194765.876142	441.322871	0.532539	0.528500
Lasso Regression	0.554391	194675.698674	441.220692	0.532755	0.528718
Ridge Regression	0.554440	194759.747522	441.315927	0.532554	0.528515
Decision Tree	0.819249	91621.186310	302.689918	0.780098	0.778198
Random Forest	0.830656	83900.809029	289.656364	0.798628	0.796888

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

•The temperature, hours, and solar radiation features were found to be more relevant for the bike count required at each hour for the stable supply of rental bikes.

•As we have analyzed the various features, we have seen that people prefer to take bikes on rent when temperature is near about 25 degrees Celsius.

•Other factors such as rainfall and snowfall also have an impact on the requirement of bikes for rent. Because in heavy rainfall and snowfall bike riding sometime becomes dangerous.

•As we have analyzed that the rental bike demands are high in the evening and in the morning. So bikes should be available at that time to fulfill the bike demands.

•The Bike demand increases with an increase in visibility and decreases with an increase with humidity.

•We tried adding possible columns to make the model a bit more complex but for Linear Regression model it is still too general.

•We have to make our model more complex for better discretion or move to tree based and ensembling algorithm for better results.

•Random forest gives predictions better than a decision tree model. Predictions made by Random Forest is better than all the models used. The value of the Adjusted R-squared for the Random Forest is 0.796, which is good.

THANK YOU