Capstone Project

Bike Sharing Demand Prediction

Content

- Problem Statement
- Data Summary
- Data Analysis
- Challenges
- Conclusions
- Q&A

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.

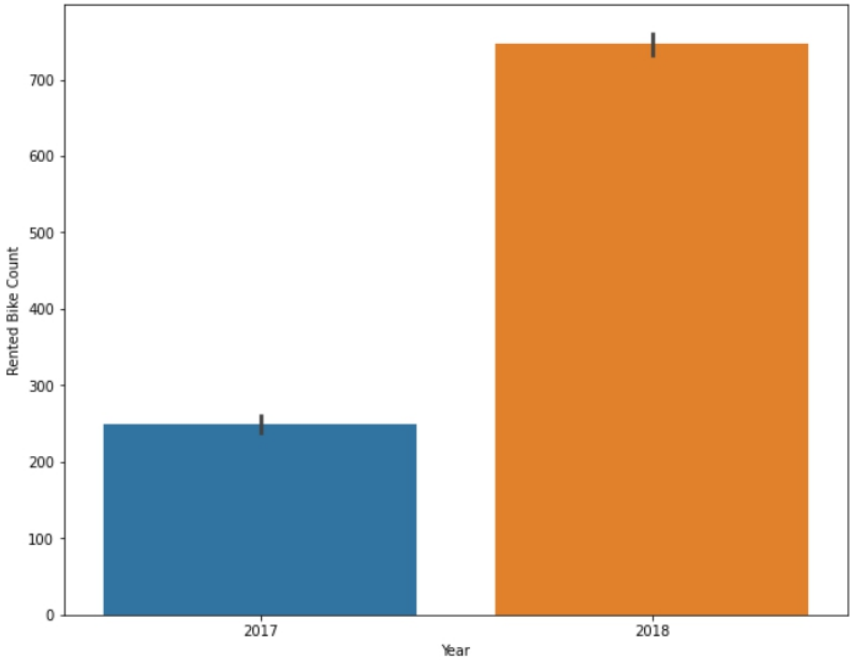
Data Summary

- Date : year-month-day
- Rented Bike count - Count of bikes rented at each hour
- Hour - Hour of the day
- Temperature-Temperature in Celsius
- Humidity - relative humidity in percentage
- Windspeed - wind speed in 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 - NoFunc(Non Functional Hours), Fun(Functional hours)

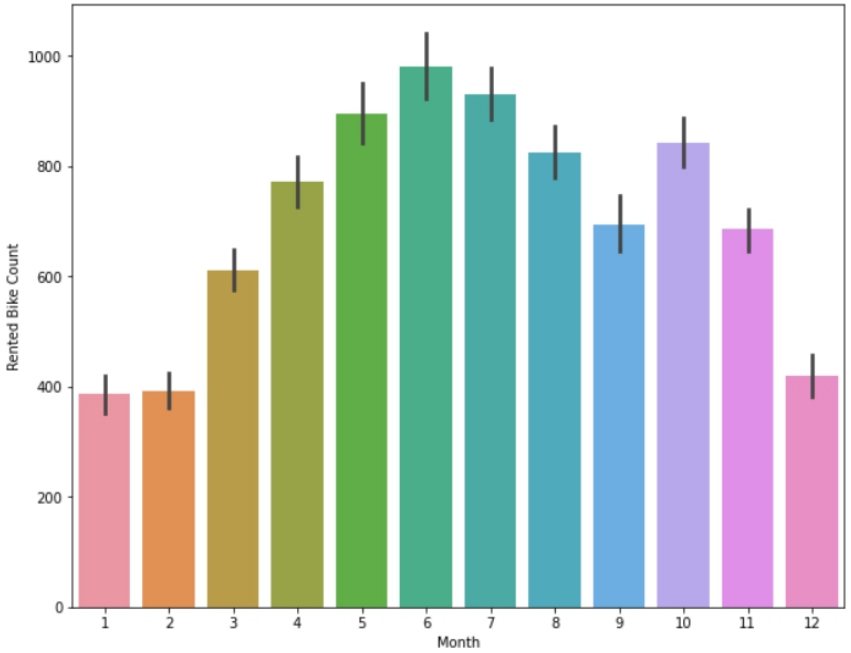
Basic Exploration

- Dataset contains 8760 rows & 14 columns
- Five categorical features ‘Date’ , ‘Hour’ , ‘Seasons’ , ‘Holiday’ & ‘Functioning Day’
- Nine continuous feature ‘Rented Bike Count’ , ‘Temperature’ , ‘Humidity’ , ‘Wind speed’ , ‘Visibility’ , ‘Dew point temperature’ , ‘Solar Radiation’ , ‘Rainfall’ & ‘Snowfall’
- Rented Bike Count is dependent variable
- No null values present in dataset

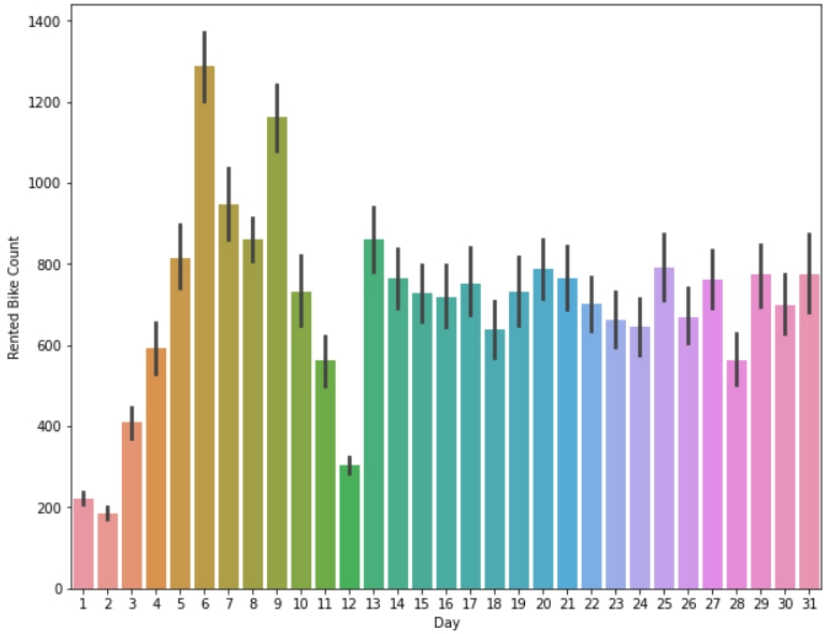
Rented Bike Count per Year



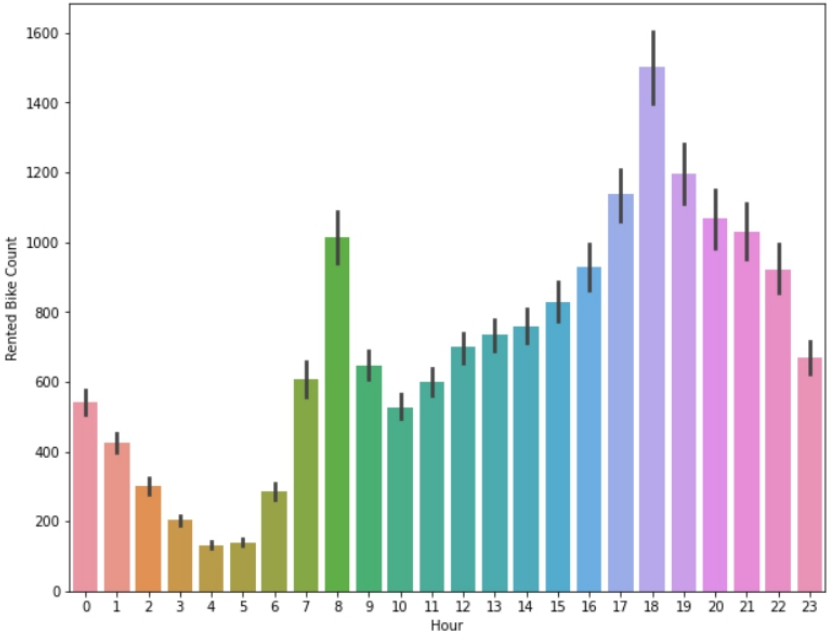
Rented Bike Count per Month



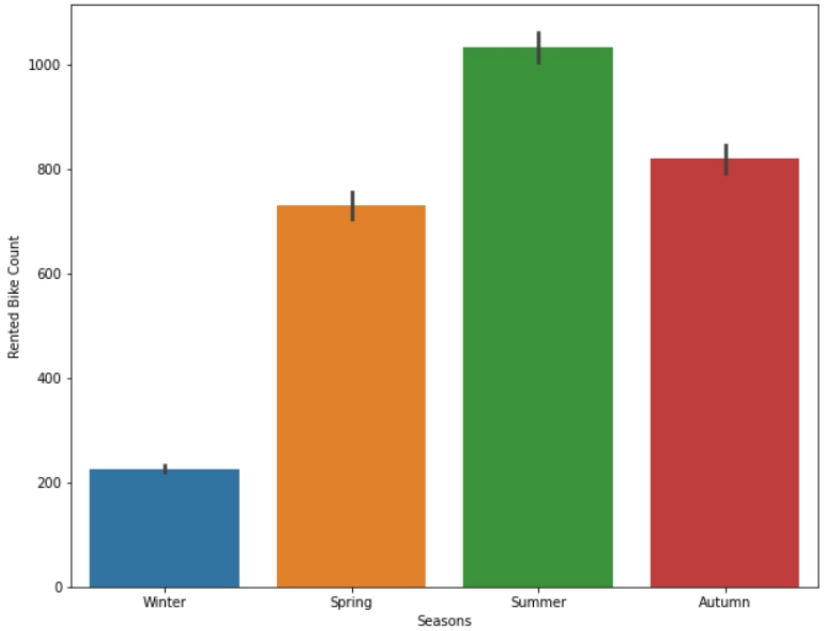
Rented Bike Count per Day



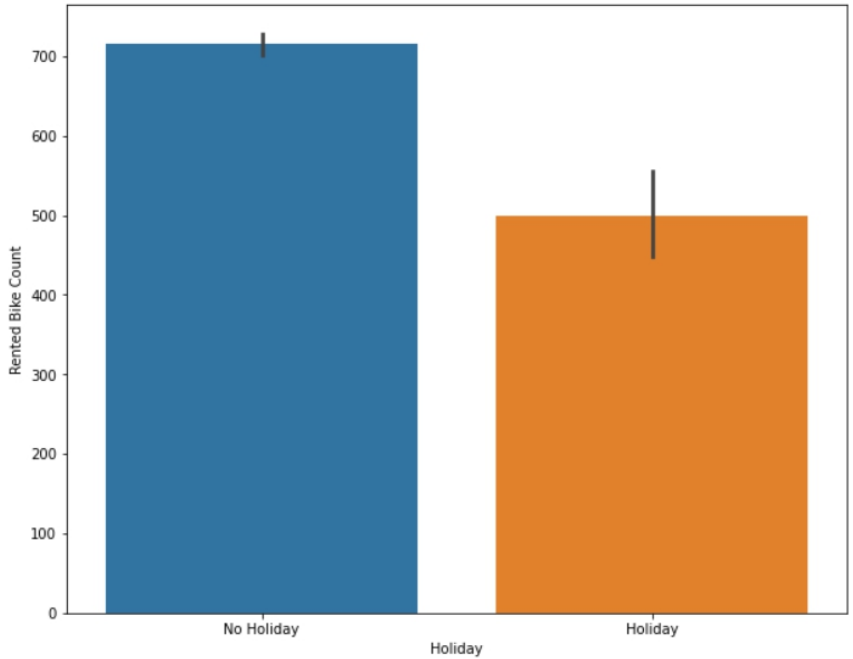
Rented Bike Count per Hour



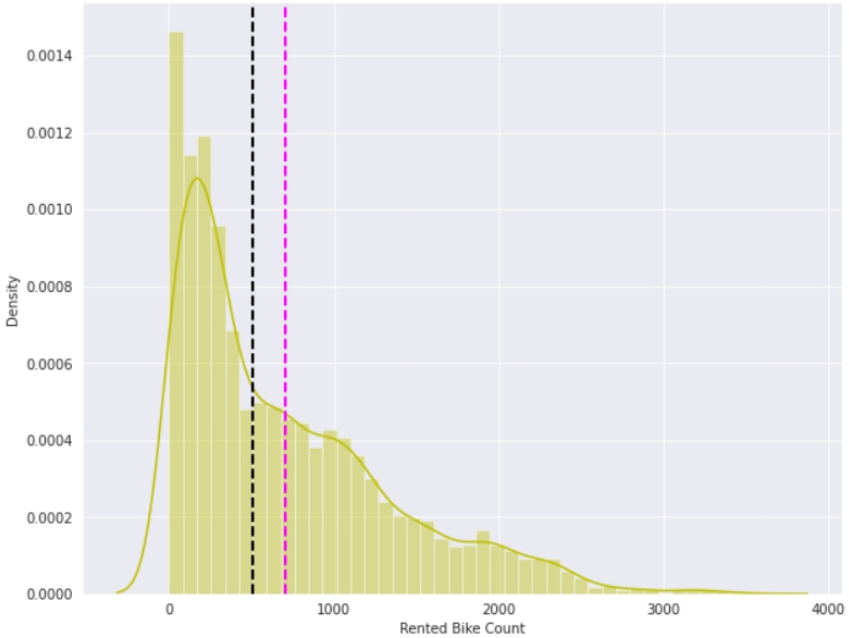
Rented Bike Count per Seasons



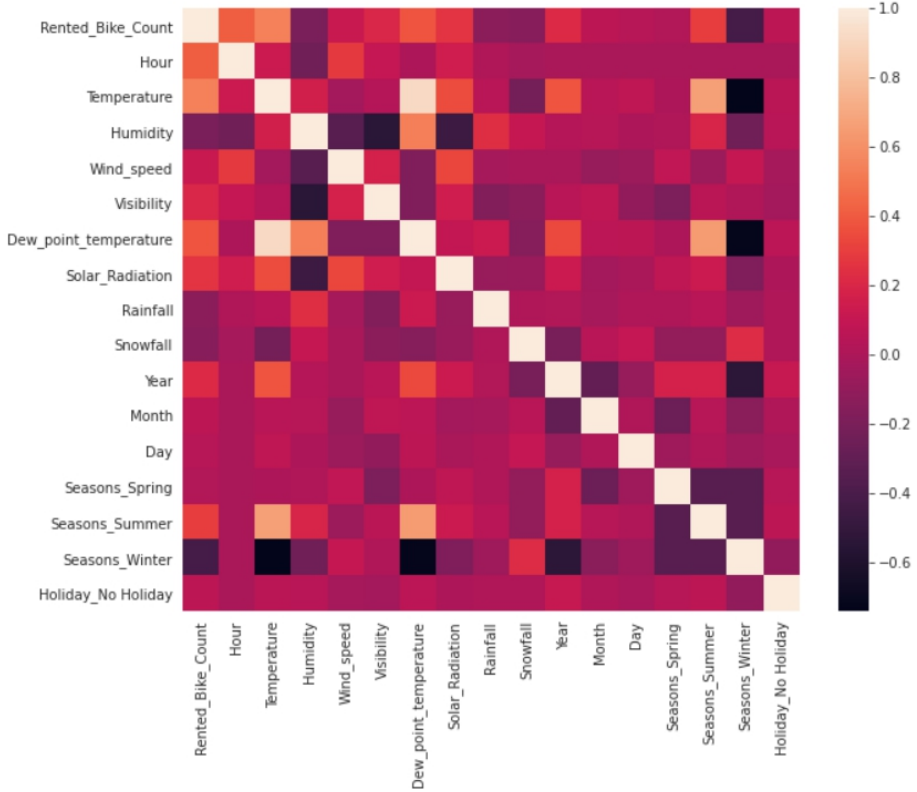
Rented Bike Count per Holiday



Distribution plot of Rented Bike Count



Correlation



Linear Regression

Train Set Metrics

R2	0.50
Adjusted R2	0.50

Test Set Metrics

R2	0.49
Adjusted R2	0.49

Polynomial Regression

Train Set Metrics

R2	0.65
Adjusted R2	0.64

Test Set Metrics

R2	0.62
Adjusted R2	0.62

Ridge Regression

Train Set Metrics

R2	0.50
Adjusted R2	0.50

Test Set Metrics

R2	0.49
Adjusted R2	0.49

Lasso Regression

Train Set Metrics

R2	0.50
Adjusted R2	0.50

Test Set Metrics

R2	0.49
Adjusted R2	0.49

Random Forest Regressor Algorithm

Train Set Metrics

R2	0.97
Adjusted R2	0.97

Test Set Metrics

R2	0.81
Adjusted R2	0.81

XGboost Algorithm

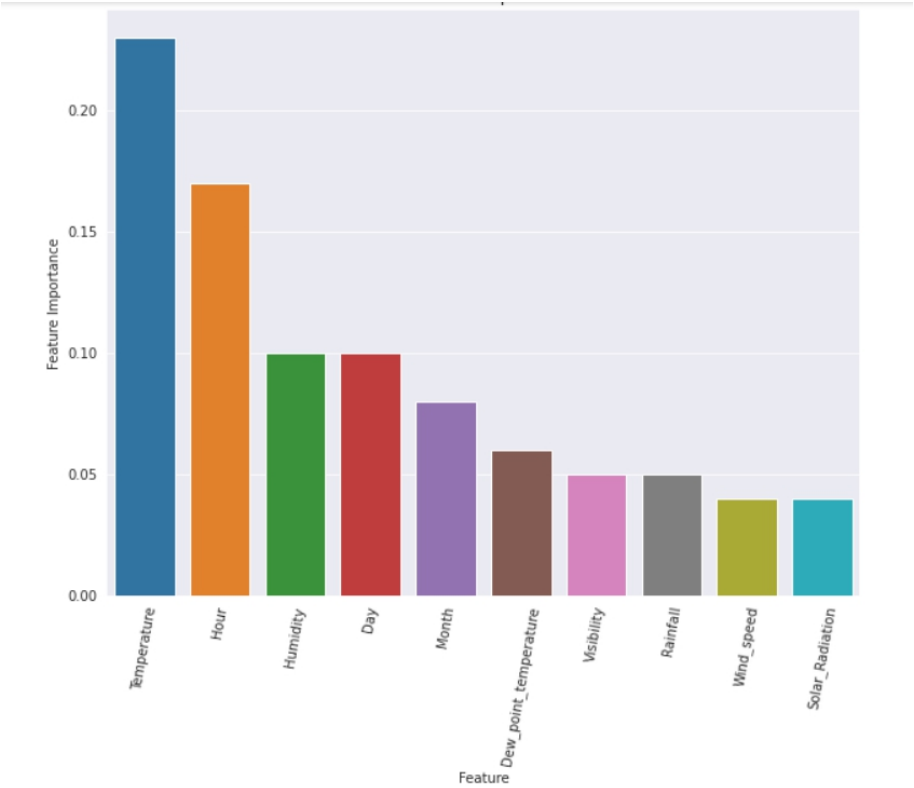
Train Set Metrics

R2	0.99
Adjusted R2	0.99

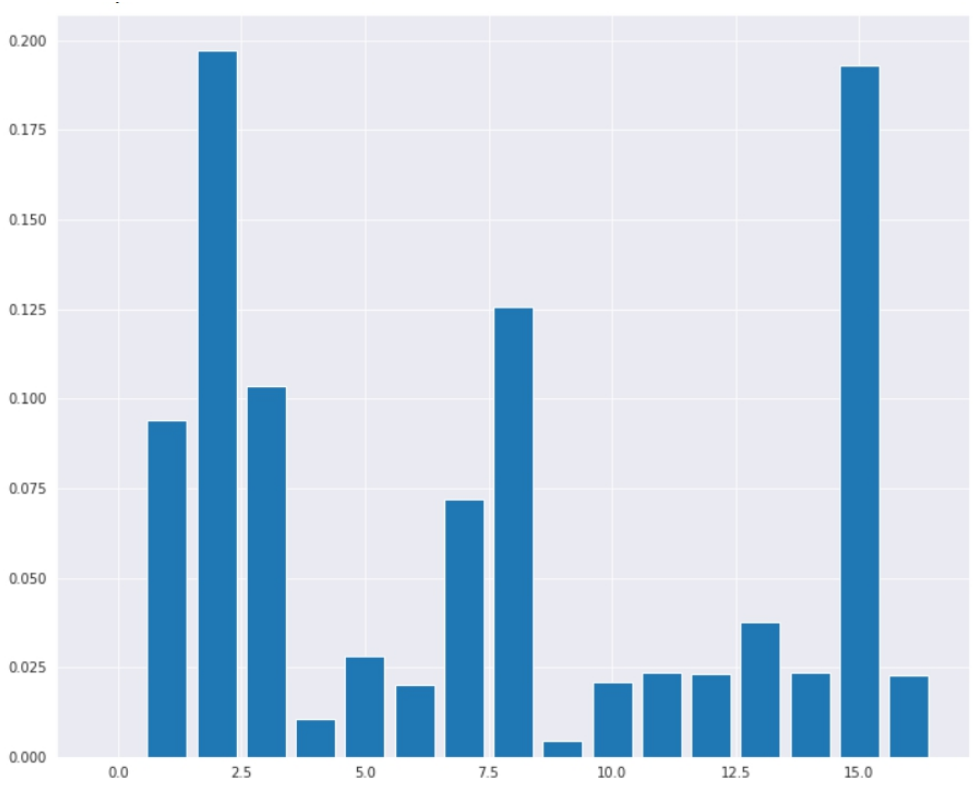
Test Set Metrics

R2	0.96
Adjusted R2	0.96

Random Forest Regressor Feature Importance



XGboost Algorithm Feature Importance



Challenges

- Feature Engineering
- Optimising the Model

Conclusion

- There is an high demand for Rented Bikes in the year 2018 .
- There is an high demand for Rented Bikes in the month of May, June and July .
- There is an high demand for Rented Bikes in the day 6 , 7 and 9 .
- There is an high demand for Rented Bikes in the hour 17 , 18 and 19 .
- Summer season has highest Demand for Rented bikes and Winter has least Demand.
- No Holiday has higher Demand compare to Holiday .
- The features hour & temperature are the features that influence the most of the bike sharing demand data
- Comparing to all algorithms XGboost algorithm has less Mean Squared error and Mean absolute error,and gives a model score upto 99% and R-Squared value is 96%, so it is concluded as optimal model.

Q & A

Thank You