[**Seoul-bike-sharing-demand-prediction**](https://github.com/samarthgangurde01/Seoul-bike-sharing-demand-prediction)

Summary

Bike sharing systems have become widely popular across different cities in the world.The elements of affordability and accessibility have helped in the promotion and expansion of the concept of a bike rental system. And it definitely comes across as a win-win for just about anyone who is willing to skip the expensive cab services for a bike.Discussions around the positive effects of bike sharing often focus on environmental sustainability and the personal health benefits experienced by users. When new bike share schemes are introduced, there’s often a reference to the “greening” of the city; ideally, more bikes can equal fewer cars, which leads to a reduction in greenhouse gasses and an improvement in local air quality

we started our project by importing dataset from the kaggele by using read\_csv, further we moved toward the missing values as there were no missing values and duplicates it did half of work for preprocessing, then we rename the columns for better understanding,In preprocessing we extracted day, month and year from column date and changed the data type to categorical type

As we found the outliers in columns like rented bike count and wind speed by using box plot we removed them by finding interquartile range of 25% and 75% , we capped outliers by capping them to upper and lower limit of the values

In Exploratory Data Analysis on all the features of our dataset. first we analyzed our dependent variable, 'Rented Bike Count' & then moved toward independent variables like month, weekend , hours, season, holliday, wind, speed, humidity, temperature etc. in column month we could found that Month 5th,6th,7th and 10th gave us most of the Traffic of bike demand. As per dependent variable bike rental count showed us higher density during working days than non working days. For environmental feature like temperature and season it showed us people are generally prefer to bike at high temperatures, when little windy and in Autumn & Summer seasons is higher ,

Because of snow fall bike rental counts were lowest in the winter season.

In heat map and scatter plot we observed negative relation between humidity and bike count as increasing in humidity lead it to decreases in the number of bike counts

we implemented 6 machine learning algorithms

Linear Regression,

Lasso Regression

Ridge Regression

Decision tree

,Random Forest

XGBoost.

None of them showed us overfitting.Random Forest Model has given accuracy upto 90% which was pretty good for prediction and choosing it for deployment

Contributerts Roles:

1.Lalit Ahirrao:

1.Data Wrangling

2.Outliers detection

3.EDA of Month variable and Hour variable

4.HeatMap

5.scaling down variables

6.Model training

2.Aniket Gajmal:

1.Data Wrangling

2.Hadeling outliers

3.EDA of Functioning Day variable and season variable

4.Model training

5.Building a Predictive System

6.web app development(streamlit)

3.Rushikesh Pawar:

1.Data Wrangling

2.Handeling missing values

3.EDA of Holiday variable and Hour variable

4.Model training

5.Creating a Pickle File

6.web app development(streamlit)

4.Prasad Ghegade:

1.Data Wrangling

2.Handling Duplicates

3.EDA solar Radiation variable and rainfall variable

4.Model training

5.web app deployment on streamlit.io

5.Samarth Gangurde:

1.Data Wrangling

2.EDA of temperature variable and humidity variable

3.EDA by printing regression plots(Temperature,

Wind\_speed,Visibility,Dew\_point\_temperature,Solar\_Radiation)

To understand positives and negatives relations between variables

4.Creating dummy variables

5.Model training

6. Comparison of models performance

7.web app deployment on streamlit.io