

Capstone Project Submission

Team Member's Name, Email, and Contribution:

Name: Shahfaissal I Dharwad
email id: shahfaissal21@gmail.com

Contribution:

- EDA on the dataset
- Application of regression models
- Drawing useful insights from the predicted and actual values.

Name: Ajinkya Satish Jumde
email id: jumde.ajinkya@gmail.com

Contribution:

- Data Analysis on EDA insights
- Handling Missing values and outliers
- Data Filtering
- Feature Selection

Name: Neha Gupta
email id: nehagupta5994@gmail.com

Contribution:

- Data Cleaning
- EDA on the dataset
- Visualization of multiple relationships.

Please paste the GitHub Repo link.

GitHub Link:- https://github.com/shahfaissal/capstone_project2

Please write a summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

Seoul Bike Sharing Demand Prediction

- **Business Problem:** The objective of this capstone is to predict the Bike demand based on the various numeric and non-numeric features in the given dataset.

- **Understanding the problem statement:**

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.

- **Know your Data:**

Dealing with a huge data set is a time-consuming part. To understand the dataset we are working on, initially, we find the important columns and the data using "head()" and "info()". To minimize the workload and efforts we must have to distribute data and analyze the content first.

- **Handling missing values and outliers:**

Removing the unnecessary part while handling missing values and data cleaning. During the analysis, we found that there are no null values in the dataset. Whereas few of the features were having outliers which result in the deviation in the performance of machine learning model. In order to improve the accuracy, we tried to filter out the outliers.

- **Univariate and Multivariate analysis:**

Based on the univariate and multivariate analysis carried out on the given dataset, a set features were selected and dropped based on their relationship with the dependent variable and their impact.

- **Regression Models:**

- 1) **OLS Model:** Based on the OLS model the evaluation matrix was analyzed and the predicted values were very close to the actual values of the given dataset. And hence we can conclude that the given model can be used for determining the demand for bikes based on the selected feature data.
- 2) **L1 Regression Model:** Based on the L1 Regression Model the evaluation matrix was analyzed and the predicted values were very close to the actual values of the given dataset. And hence we can conclude that the given model can be used for determining the demand for bikes based on the selected

feature data.

- 3) **L2 Regression Model:** Based on the L2 Regression Model the evaluation matrix was analyzed and the predicted values were very close to the actual values of the given dataset. And hence we can conclude that the given model can be used for determining the demand for bikes based on the selected feature data.

- **Conclusion:**

After the analysis we conclude that, this analysis will be helpful for the supplying entities in Seoul city to predict the demand for the rental bikes based on the primary and atmospheric data of the city. The predicted values were highly accurate and so the companies can rely on the model in order to predict the demand for the bikes on a particular date based on temperature, humidity, wind speed, solar radiation, type of day (Holiday/No holiday), hour and month.

