**SEOUL BIKE SHARING DEMAND PREDICTION SUMMARY**

**Sundar K**

**Data science trainee**

**Alma better**

**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**.**

**Abstract**

The bike sharing system has brought wide convenience to residents in the city and serves as important tools to transport from one place to another place. For the bike sharing companies, they need to know the total users of bike, so they can release suitable number of bikes into the marketThis paper uses visualization technology to visualize data and figure out the possible factors which can impact the total number of users. After completing the data analyzing, this paper figures out the season, weather, temperature, humanity and wind speed are the main factors which can have impacts on the total number of users. the second stages, this paper uses regression model, to predict the possible number of bike users. The input factors are season, weather, temperature, humanity and wind speed.

**Exploratory Data Analysis**

After loading the dataset we performed this method by comparing our target variable that is Rented\_bike\_Count with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables

Our dataset contains a large number of null values which might tend to disturb our accuracy hence we replaced it with zero made them at the beginning of our project inorder to get a better result. We used One Hot Encoding to produce binary integers of 0 and 1 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format.

**Conclusion**

Starting with loading the data so far we have done EDA , null values treatment, encoding of categorical columns, feature selection and then model building. It is quite evident from the results that Linear regression is the best model that can be used for the Bike Sharing Demand Prediction since the performance metrics (MSE,RMSE) shows lower and (r2,adjusted\_r2) show a higher value