## **Bike Sharing Demand Prediction**

**Introduction & Problem Statement**

A bike sharing service is the one in which a person lends bikes for other people on rent. This can sometimes be a shop where the lender receives payment for providing bikes for use. Nowadays we have an unmanned system in which bikes are available at docks which the user can go to & rent a bike & pay digitally online. In this project we are going to analyse the given dataset for valuable insights that might help the company improve their profits by analyzing various factors given in the dataset. We are also going to deploy a regression model to predict the count of rented bikes which may include a few more steps.

Bike sharing is an important means of travelling in today's world , so it's quite necessary for the bike rental services to provide their services on time. We need to create a system which can predict Bike rental demand based on users behaviour.

We will approach our project in the following steps :-

Data Discovery - In this step we import data & understand the data, its variables, values etc

Data Wrangling - In this step we modify the data so that it becomes ready for next procedures

Exploratory Data Analysis - in this step we create a bunch of charts & graphs to gain insights on the dataset that can help us making better decisions for our rental business.

Model Building - in this step we will try multiple machine learning models & test it to choose the best one. We will also do hyperparameter tuning to find the best parameters for our selected model

**Dataset Discovery**

****

**Steps Involved**

**1] Data Discovery -** this step involves importing the dataset, & gaining some preliminary information. Used functions like describe() , info() to gain insights on the variables. Also found unique values per variable.

2] **Data Cleaning** - this step involves cleaning the data so that it becomes fit for analysis. Includes imputing missing values , removing useless columns. Here we splitted the date column into 3 columns - date,month,year.

3] **Exploratory data analysis** - This step involves creating graphs & statistics to gain insights from the dataset. Various graphs such as bar graphs, pie charts, scatterplots, histograms have been used in this project. The entire process is divided into 2 phases -

3.1] **Univariate Analysis** - Univariate analysis explores each variable in a data set, separately. It was done on neighborhood group & neighborhood, Count of room type , Latitude , longitude, minimum nights, availability 365.

3.2] **Bivariate Analysis** - Avg Price per neighborhood group & neighborhood, Price vs latitude,Price vs longitude,Avg price per room\_type,Price vs minimum nights, avg Availability per room type, Availability per neighborhood group [& neighborhood].

3.3] **Correlation Heatmap -** Correlation heatmap is the best for analyzing correlations between all possible variables.Theres not much correlation between these variables , i.e. they are independent.

**3.4]**  **Pair Plot** - Pair plot draws all possible graphs for given variables.

4] Model Building - Here we do all steps necessary for building the model

4.1] Dummy encoding - here we convert all categorical values into numerical ones since our machine learning models cannot interpret the categorical values.

4.2] Model training & valuation - Here we import models , train those models and test those models for evaluation. Here we used R2 score as performance metric since it is easier to interpret. We found that xgboost has the best performance.

4.3] hyperparameter tuning - Here we use certain algorithms for selecting best parameters for our selected model. Here we used Randomized Search cv.