# Forecast Cab Booking Demand

Mid-Program Project 1

# Table of Contents

## Background

Cab booking system is the process where renting a cab is automated through an app throughout a city. Using this app people can book a cab from one location to another location. Being a cab booking app company, exploiting an understanding of cab supply and demand could increase the efficiency of their service and enhance user experience by minimizing waiting time.


Objective of this project is to combine historical usage pattern along with the open data sources like weather data to forecast cab booking demand in a city.

## Process Flow

You will be provided with hourly renting data span of two years. Data is randomly divided into train and test set. You must predict the total count of cabs booked in each hour covered by the

test set, using the information available prior to the booking period. You need to append the **train_label** dataset to **train.csv** as '**Total_booking**' column

## Dataset Description

Please find the descriptions of the columns present in the dataset as below:

- **datetime** - hourly date + timestamp
- **season** - spring, summer, autumn, winter
- **holiday** - whether the day is considered a holiday
- **workingday** - whether the day is neither a weekend nor holiday
- **weather** - Clear , Cloudy, Light Rain, Heavy temp - temperature in Celsius
- **atemp** - "feels like" temperature in Celsius
- **humidity** - relative humidity
- **windspeed** - wind speed
- **Total_booking** - number of total booking



## Target Environment

You can use Edureka's CloudLab, a cloud based Jupyter Notebook, which is pre-installed with Python and other required packages to work on this Project. It is offered by Edureka as a part of the course, where you can execute all the demos and work on the projects hassle-free.

# Tasks to be done:

1. Import the required libraries and load the training and testing dataset — Marks: 2
2. Analyze the dataset and write your observations — Marks: 6
    a. Check the shape of the training and testing set
    b. Print the data types of each column
    c. Check the missing values present in the dataset
3. Perform **Feature Engineering**: — Marks: 12
    a. Create new columns **date**, **hour**, **weekDay**, **month** from **datetime** column
    b. Coerce the datatype of **season**, **holiday**, **workingday,** and **weather** to **category**
    c. Drop the **datetime** column as we have already extracted useful features from it
4. Perform **Outlier Analysis:** — Marks: 10
    a. Plot **Box plots** across various features like **season**, **hour of the day**, **working_day**, etc to see if there are any **Outlier** and note down your inference
    b. Remove the outliers present in the dataset
5. Perform **Correlation Analysis**: — Marks: 8
    a. Plot a correlation plot between "**total booking**" and ["**temp**", "**atemp**", "**humidity**", "**windspeed**"]
    b. Write down your inference in the markdown cell
6. Perform **Data Visualization**: — Marks: 12
    a. Visualize distribution of data: **total_booking** column and plot the probability distribution plot for the column as well
    b. Visualize **total_booking** vs (**Month**, **Season**, **Hour**, **Weekday**, **Usertype**)
    c. Use **Histograms** to plot all the **continuous** variables present in the **data**

7. Convert the categorical variables into one hot vector Marks: 5
8. Split your dataset for training and testing Marks: 3
9. Fit various models (**Random Forest Regressor**, **Ada Boost Regressor**, **Bagging Regressor**, **SVR**, and **K-Neighbors Regressor)** Marks: 15
10. Display a **Factor** plot to visualize the **RMSE** values achieved by different modeling algorithm Marks: 10
11. Perform Hyper-parameter tuning on the best model using **GridSearchCV** and print the best parameters using **model.best_params_** Marks: 10
12. Perform prediction on the test set and print the mean_squared_log_error Marks: 7



## How to submit your project?

Following are the tasks, which need to be developed while executing the project:

- If you are using **colab**, please download the **IPYNB** file from file menu.
- The **IPYNB** file should have the details of each step in the markdown
- After verifying your solution, submit the **IPYNB** file on the LMS