**IE 434 - Deep Dive 2 (First Milestone)**

**Team Members:**

Our group members are:

1. Deepanshu Malhotra
2. Prannoy Kathiresan
3. Selvamani Subramaniam
4. Vishal Dayalan

**Problem Explanation:**

The problem involves predicting the most popular Lyft Bike destination stations in New York City based on various features related to bike rentals. This prediction can help optimize bike placement and distribution, making it more convenient for users and potentially increasing the efficiency of the bike-sharing system.

The features that will be studied are as follows:

Start Station name, Rideable Type (classic bike, electric bike, docked bike), Date (Month, Day of the week), Start time, End time, Trip Duration (seconds), Start latitude/longitude, End latitude/longitude, Weather, User Type (customer or member), Year of Birth, Gender.

**Approach:**

For our goal of predicting popular Lyft Bike destination stations, we will be using machine learning techniques to build a predictive model. We will use various features as described above to predict the end station, which represents the destination.

We are using the datasets for the year 2021 and 2022. In order to perform the prediction, we have to split the data into training (working) and testing (debugging) sets to evaluate the model's performance. The goal is to train a model that can predict the most likely end station based on given data, and then use that model to suggest where more Lyft Bikes should be placed at different times of the day.

**License:**

The dataset we are using can be found [here](https://s3.amazonaws.com/tripdata/index.html). We will be abiding by the ‘[Citi bike data use policy](https://ride.citibikenyc.com/data-sharing-policy)’ which states the “Data License agreement”.