**Bike-Sharing Demand Analysis**

**Objective:**  Use data to understand what factors affect the number of bike trips. Make a predictive model to predict the number of trips in a particular hour slot, depending on the environmental conditions.

**Problem Statement:**

Lyft, Inc. is a transportation network company based in San Francisco, California and operating in 640 cities in the United States and 9 cities in Canada. It develops, markets, and operates the Lyft mobile app, offering car rides, scooters, and a bicycle-sharing system. It is the second largest rideshare company in the world, second to only Uber.

Lyft’s bike-sharing service is also among the largest in the USA. Being able to anticipate demand is extremely important for planning of bicycles, stations, and the personnel required to maintain these. This demand is sensitive to a lot of factors like season, humidity, rain, weekdays, holidays, and more. To enable this planning, Lyft needs to rightly predict the demand according to these factors.

**Domain:** General

**Analysis to be done:** Rightly predict the bike demand

**Content:** Dataset: Lyft bike-sharing data (hour.csv)

Fields in the data:

**- instant:** record index

**- dteday:** date

**- season:** season (1:spring, 2:summer, 3:fall, 4:winter)

**- yr: year** (0: 2011, 1: 2012)

**- mnth:** month (1 to 12)

**- hr:** hour (0 to 23)

**- holiday :** whether the day is a holiday or not

**- weekday :** day of the week

**- workingday :** if the day is neither weekend nor a holiday is 1, otherwise is 0

**- weathersit :**

- 1: Clear, Few clouds, Partly cloudy

- 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

- 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds

- 4: Heavy Rain + Ice Pellets + Thunderstorm + Mist, Snow + Fog

**- temp :** normalized temperature in Celsius; the values are divided to 41 (max)

**- atemp:** normalized temperature felt in Celsius; the values are divided to 50 (max)

**- hum:** normalized humidity; the values are divided to 100 (max)

**- windspeed:** normalized wind speed; the values are divided to 67 (max)

**- casual:** count of casual users

**- registered:** count of registered users

**- cnt:** count of total rental bikes including both casual and registered

**Steps to perform:**

As the first step, look at the null values in the file. A sanity check, to ensure that you have clean records and the data is good to go ahead, is very important. Then, you’ll do univariate and bivariate analyses to identify the patterns in the data and the nature of the individual features. This is a very important step as this helps to not only identify features which could be interesting for the predictive model later, but also helps understand what’s going on in the data. The EDA will help identify the need to apply transformations on the features before building the model. Finally, you will make a predictive model using linear regression.