**Linear Regression Project**

### **Project must be done in Jupyter Notebook using R od Python. Save project as .html.**

Use the supplied csv in the repository. The data has the following features:

* datetime - hourly date + timestamp
* season - 1 = spring, 2 = summer, 3 = fall, 4 = winter
* holiday - whether the day is considered a holiday
* workingday - whether the day is neither a weekend nor holiday
* weather -
  + 1: Clear, Few clouds, Partly cloudy, Partly cloudy
  + 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
  + 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
  + 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
* temp - temperature in Celsius
* atemp - "feels like" temperature in Celsius
* humidity - relative humidity
* windspeed - wind speed
* casual - number of non-registered user rentals initiated
* registered - number of registered user rentals initiated
* count - number of total rentals

## Load the Data

**1. Read in bikeshare.csv file**

## Exploratory Data Analysis

**2.** **Create a scatter plot of count vs temp. Set a good alpha value.**

**3. Plot count versus datetime as a scatterplot with a color gradient based on temperature. What do you notice?**

**4.What is the correlation between temp and count?**

**5. Let's explore the season data. Create a boxplot, with the y axis indicating count and the x axis begin a box for each season.**

## Feature Engineering

**6. Create an "hour" column that takes the hour from the datetime column.**

**7. Now create a scatterplot of count versus hour, with color scale based on temp. Only use bike data where workingday==1.**

**8. Now create the same plot for non working days. What is the difference?**

## Building the Model

**9. Use Linear Regression to build a model that predicts count based solely on the temp feature**

**10. How many bike rentals would we predict if the temperature was 25 degrees Celsius?**

**11. Finally build a model that attempts to predict count based off of the following features**

* season
* holiday
* workingday
* weather
* temp
* humidity
* windspeed
* hour (factor)

**12. Get the summary of the model.** **Did the model perform well on the training data? What do you think about using a Linear Model on this data?**

**13. See how well you can predict for future data points by creating a train/test split. But instead of a random split, your split should be "future" data for test, "previous" data for train.**