**Final Project Documentation**

**Bike sharing**

**Project :**  I was provided hourly rental data spanning two years. For this competition, the training set is comprised of the first 19 days of each month, while the test set is the 20th to the end of the month. You must predict the total count of bikes rented during each hour covered by the test set, using only information available prior to the rental period.

**Data :**

**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

**Machine learning :**  I used cross validation to calculate the score which turned out to be with an accuracy of .82. And I have built my model using Gradient Boosting Regressor with 200 n\_estimators and stored the whole log information in the submission file which is saved in the same folder as the python notebook is running



