# Report: Predict Bike Sharing Demand with AutoGluon Solution
#### Khaled Yaseen

## Initial Training
### What did you realize when you tried to submit your predictions? What changes were needed to the output of the predictor to submit your results?
I have repeated problems with launching cloud on udacity so I couldn't use Sagemaker studio so I used Local Machine.
"'kaggle' is not recognized as an internal or external command" error was raised when I wanted to submit my solution, After a lot of searchingand asking tutors I could use Kaggle CLI and establish .kaggle folder and submit the solution by commands in jupyter notebook. Despite not having predictions less than 0, I converted the possible negative values to 0 and submitted the results.

```
In [35]: !kaggle competitions submit -c bike-sharing-demand -f submission_new_hpo.csv -m "new features with hyperparameters"

         Successfully submitted to Bike Sharing Demand


           0%|          | 0.00/194k [00:00<?, ?B/s]
           4%|4         | 8.00k/194k [00:00<00:04, 43.7kB/s]
          45%|####5     | 88.0k/194k [00:00<00:00, 340kB/s]
          66%|######5   | 128k/194k [00:00<00:00, 191kB/s]
          82%|########2 | 160k/194k [00:00<00:00, 163kB/s]
          95%|#########4| 184k/194k [00:01<00:00, 150kB/s]
         100%|##########| 194k/194k [00:02<00:00, 66.8kB/s]
```

### What was the top ranked model that performed?
Weighted Ensemble L3 is the model that performed well. Upon adding additional features of date, month, year, and the hyperparameter tuning, the performance improved drastically.

## Exploratory data analysis and feature creation
### What did the exploratory analysis find and how did you add additional features?
Upon Exploratory Data Analysis, I inferred the following:
1- The features - Season and Weather are categorical.
2- The features - holiday and working day are binary values.
3- Datetime is a DateTime object, and it is transformed into different features - day, month, year, hour

### How much better did your model preform after adding additional features and why do you think that is?
After adding additional features, the accuracy of the model is increased significantly. This is because the new features are strongly correlated to the target.

## Hyper parameter tuning
### How much better did your model preform after trying different hyper parameters?
After, trying out different parameters, I did notice that the performance of the model is changed. The model with the hyperperameters and new features performs more or less the same as the one without hyperperameter configurations.

### If you were given more time with this dataset, where do you think you would spend more time?
I could spend some time more on EDA and see if I could derive new features to improve the model's prediction.

### Create a table with the models you ran, the hyperparameters modified, and the kaggle score.

### Create a line plot showing the top model score for the three (or more) training runs during the project.

### Create a line plot showing the top kaggle score for the three (or more) prediction submissions during the project.

## Summary
With this assignment I aimed solving a regression problem to predict the bike sharing demand, given the historical data. Framework used - AutoGluon The initial model was built with the default predictor and the model doesn't perform great. With Exploratory Data Analysis, assigned right data types for the features and with feature engineering, created new features based on the date time. With the new features, the performance of the model is drastically increased. Hyper parameter tuning is also done to improve the performance but that did not yield better results as of now.