**Data Source**

<http://www.stat.ufl.edu/~winner/datasets.html>

**Data Set** - http://www.stat.ufl.edu/~winner/data/airq402.dat

[**Data Description -** http://www.stat.ufl.edu/~winner/data/airq402.txt](http://www.stat.ufl.edu/~winner/data/airq402.txt)

**Assignment Expectations/Steps -**

Part 1 :

* Load the data using Python Pandas library.
* Do Exploratory Data Analysis of the data and state your insights.

Part 2 :

* Treat “Average Fare” – 3rd Column as your Dependent Variable and Rest of the columns as Independent Variable.
* Create Scatter Plot of Independent Variable vs Dependent Variable.
* Based on Scatter Plot see if there is any transformation required for Independent Variable.
* Build Multiple Linear Regression model.
* Get the accuracy score on train and test data.

Part 3:

* Find the most important features of this dataset to predict the average fair.
* Figure out what other model can be applied to improve the model performance.
* What are other ways this data can be improved which will help in improving the prediction.
* State your final observations after this complete analysis.