**Data Source**

[http://www.stat.ufl.edu/~winner/datasets.html (Links to an external site.)Links to an external site. (Links to an external site.)Links to an external site.](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 (Links to an external site.)Links to an external site.](http://www.stat.ufl.edu/~winner/data/airq402.txt)

**Exercise Expectations/Steps -**

1. Import the data and assign appropriate column names.
2. Drop the variables which has textual information (categorical variables)
3. Convert all the variable’s datatype to Integer
4. Remove all the outliers from the data using IQR method (Retain data points within 1.5 \* IQR)
5. Print the correlation matrix.
6. Treat “Average Fare” – as your Dependent Variable and the variable which has the highest correlation with Average Fare as the Independent Variable. Drop all the other variables
7. Create Scatter Plot of Independent Variable Vs Dependent Variable.
8. Divide the data set into training and test data set and build a Linear Regression model.
9. Print the train and test accuracy of the model .
10. Print the coefficient and the intercept of the model .