**Problem Statement:**

The provided dataset contains data on demand for airlines in Q4 of 2002 in the US. The dataset consists of features like average fare, to and from city destination, the market share of leading airline etc. The objective of this study is to build a linear regression model that will be able to predict the average fare for a given route for any airline.

**Data Description:**

The data is in fixed width text format. Use pandas fwf reader to read the file. The column names and widths are:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Column Width** | **Description** |
| City1 | 1-3 | City of origin |
| City2 | 5-7 | Destination city |
| Average Fare | 11-17 | Average fare on the route |
| Distance | 20-23 | Distance between the two cities |
| Average weekly passengers | 26-33 | Average no of weekly passenger |
| market leading airline | 36-37 | The market leading airline |
| market share | 40-45 | The market share of leading airline |
| Average fare | 48-54 | Average fare of leading airline |
| Low price airline | 57-58 | The lowest priced airline on the route |
| market share | 61-66 | Market share of low priced airline |
| price | 69-75 | Fare on low priced airline |

Assignment Expectations/Steps -

* Import Data and understand data.
* Exploratory Data Analysis of the data above.
* 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.
* Check for Multi-Collinearity.
* Ensure all final variables in the model are significant at alpha value of 0.01