Problem Statement

The aim of the challenge is to predict the cab fare using the given dataset. The dataset consists of 1 CSV file: TRAIN.csv

TRAIN.csv consists of 9 attributes:

- index
- time_stamp - epoch time (in seconds) when the cab was booked
- cab_provider - company (Uber/Lyft)
- source - the starting point of the cab ride
- destination - the destination of the cab ride
- distance - the distance between source and destination
- surge_multiplier - multiplier by which price increased
- cab_type - the type of cab (Uber Pool, Uber XL, etc.)
- fare - cab fare in USD (**Target Attribute**)

This data should be used to train the model; no additional data is allowed to be used for the training process.

Evaluation Basis

The project will be evaluated on the following basis:

1. The process of building the model should move from simple to complex. This means its mandatory to implement multiple linear regression and logistic regression before approaching any advance level algorithm such as Random forest etc.

2. Every model should be supported by reason of acceptance or rejection. Special emphasis on the reasons why the student has picked/dropped an algorithm.

3. The student should mention which error metric is used and why. For example, if RMSE used in place of MAPE, the reason should be clearly mentioned by the student that why a particular error metric is used. Special emphasis on goodness of fit such as AUC and confusion metrics.

4. The student should revise the concepts before starting the project work.

5. The student should be confident enough to explain every concept that is written in the project report.

7. If student unable to explain the project report or during the session, it appears that student has copied the solution then zero marks will be given to the student.

8. The code should be written keeping in mind the code file can be run from DOS prompt.

9. The instructions to run the code file should be submitted with the project report.

Deliverables from Candidate

1) Code written either Python or R.

2) Comprehensive Project Report.

3) Instruction to deploy and run code.