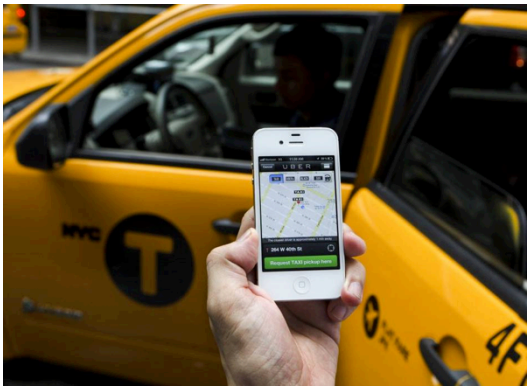
University of California
Department of Civil Engineering Fall 2018

CE88 Data Science for Smart Cities

Midterm

Taxi vs Uber



In this midterm you will work with historical data of taxi trips in the San Francisco Bay Area. You will build a taxi trip fare predictor.

General description

The dataset provided to you **(Taxi_Train.csv)** contains 50K records of taxi trips in San Francisco spanning over approximately 3 weeks of September 2012. You are required to implement an algorithm in order to predict fares for any given taxi trip in the area. You will be required to produce your predictions for a set of trips in the time period of approximately one week after that period (**Taxi_Query.csv**). The following variables are given to you:

id (int): unique trip identification number
start_time (str): start time of the trip, for example "9/1/12 14:55"
end_time (str): end time of the trip, for example "9/1/12 15:11"
fare (float): trip fare paid, $
number_pax (int): number of passengers on the trip
start_lng (float): longitude of the trip start
start_lat (float): latitude of the trip start
end_lng (float): longitude of the trip end
end_lat (float): longitude of the trip end
start_taz (int): transportation analysis zone ID of the trip start
end_taz (int): transportation analysis zone ID of the trip end

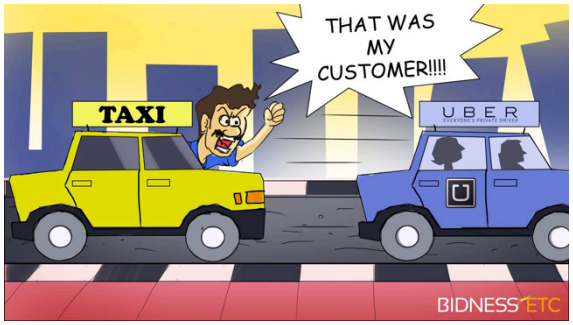
Testing dataset contains 10K 'intended trips' described by the same variables, however the arrival times and the trips fares are substituted with values of -1 (they are unknown to the passengers who intend to take a taxi). Your task is to predict the fare that was charged for each trip. The accuracy of your predictions will be evaluated by the Root Mean Squared Error (RMSE).

You are most welcome to use the starter ipython notebook to start exploring the data. It contains several useful functions to help you load the data and parse the timestamps.

Submission requirements.

1. Predictions. Kaggle website will be set up for this midterm to provide you with a submission system (there will be a separate announcement with access link). It will give you instant feedback on the performance of your model.

2. Code. You must submit an ipython notebook with the code you used to predict the taxi trip fares.



Good luck!