Estimating Travel Time

Lyft, Inc.

Task

Given training data on trip locations and durations we would like to estimate travel times between two specified locations at a given departure time. Quality of solution would be assessed using mean absolute error of predicted versus actual durations.

Input

The input data consists of two comma-separated values ("csv") files, both with headers.

The first file contains training data with ride start and end locations (specified as latitudes and longitudes), start timestamps in epoch seconds (also known as unix time), and trip duration in seconds. Each line is a trip and has the following format:

```
row_id,start_lng,start_lat,end_lng,end_lat,start_timestamp,duration
```

The second file contains test data in the exact same format as the training data, except durations are missing. Each line is a trip for which we want to estimate travel times, and has the following format:

```
row_id,start_lng,start_lat,end_lng,end_lat,start_timestamp
```

Output

Your (ideally compressed) output file, titled `duration.csv`, should have two columns: `row_id` and estimated duration in seconds for each line in the test file, where the `row_id` column refers to the corresponding row in the test file for which you are making the prediction. It should also have the following header specified as the first line of the file:

```
row_id,duration
```

Explanations and source code

Please include a document explaining your approach, what you considered and why, potential improvements, ideas for a real-time implementation, or anything else that you think is relevant to understanding your solution. Also, please attach the source code that you used to fit and evaluate your model. In an effort to reduce unconscious bias in our interview pipeline, **please do not put identifying information (e.g., your name) in any file or filename that you submit**.

We encourage you to use any language you are comfortable with. Note that we often find that candidates using R on a laptop run into issues related to the size of the dataset. These issues shouldn't stop you — think creatively.

Good luck, and please let us know if you have further questions.