

Taxi Trip Time Prediction

NAME - RAJENDRA SINGH

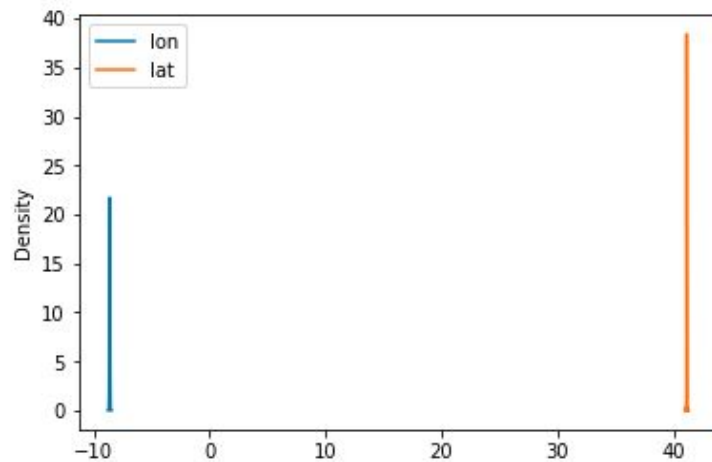
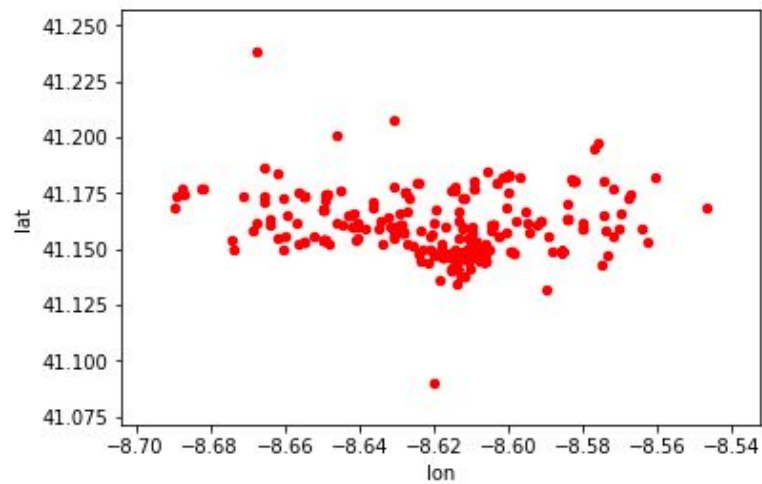
ROLL NO. - 111601017

Introduction

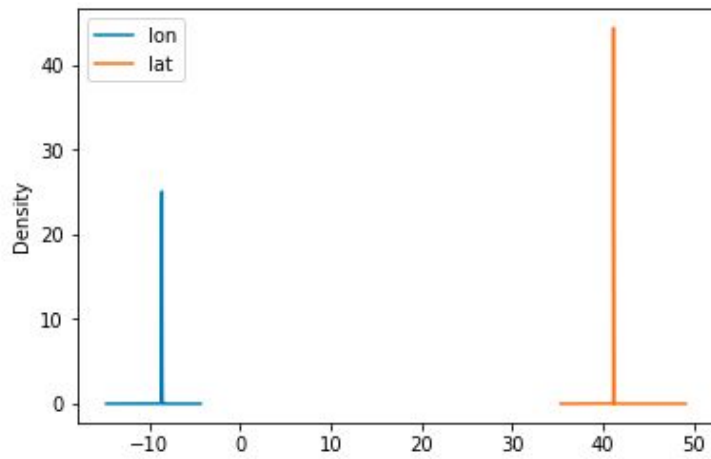
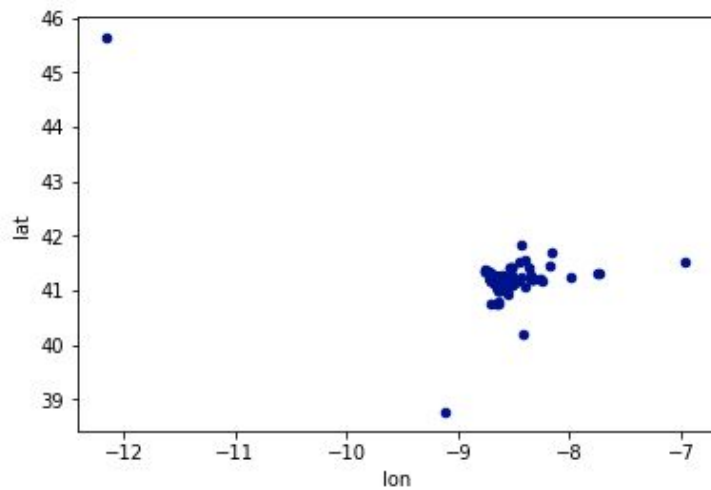
The goal of this project is to build a predictive framework that is able to predict the total travelling time of 320 taxi rides based on their (initial) partial trajectories. The framework can then be used to improve the efficiency of electronic taxi dispatching systems in Porto.

ABOUT DATA:

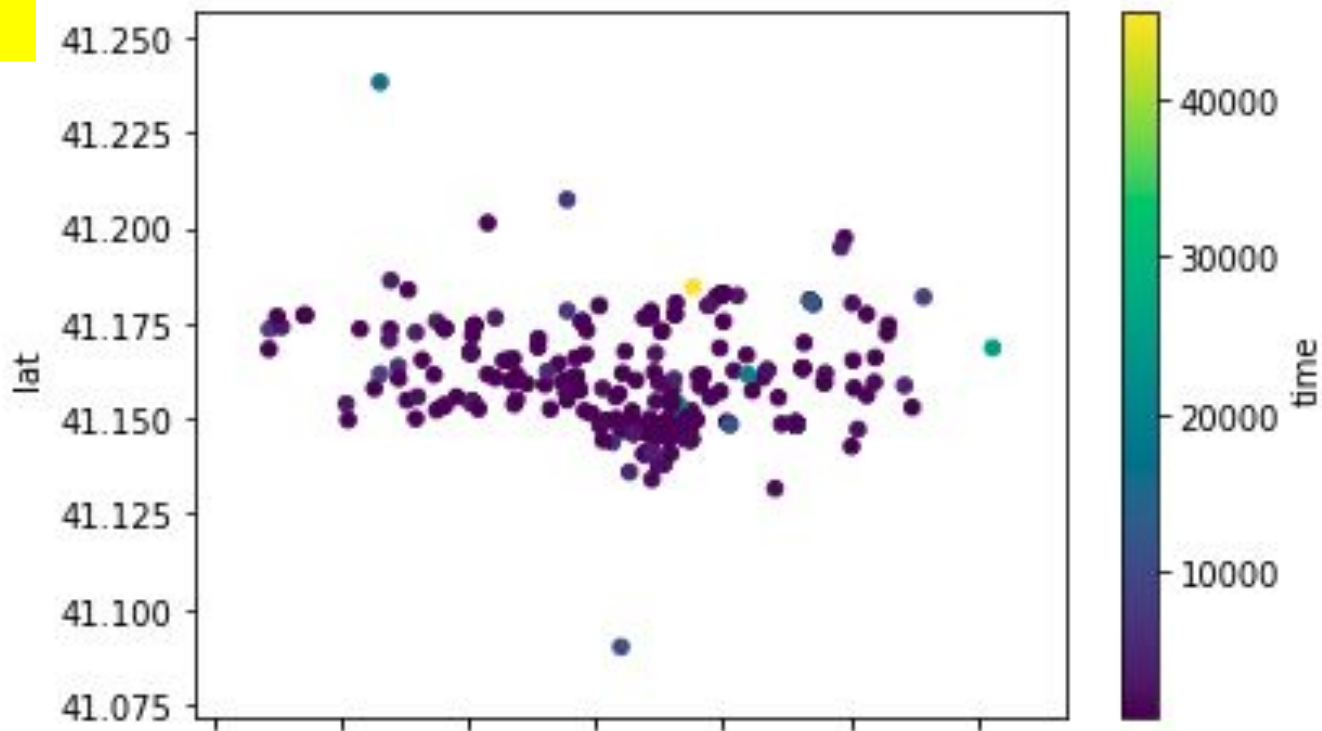
Test data



Train data



Test data vs Time

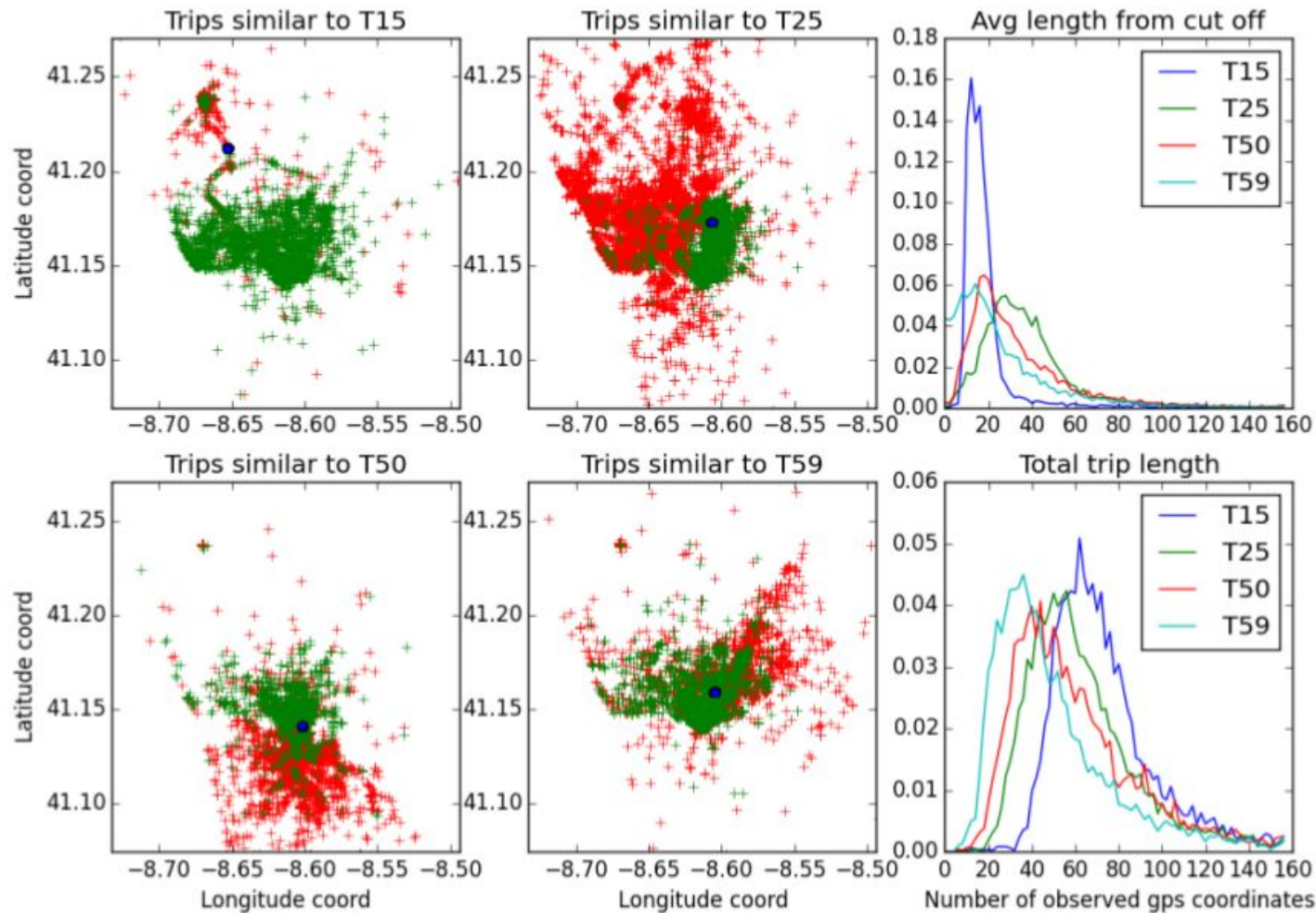


Data Observation

One of the first observations I made was that the position and also the remaining driving time depends very strongly on the cut-off position of the trip. For certain positions it was very clear where the taxi trip is heading to (for example T15 in below Fig), and making a precise prediction is feasible. For other trips only some parts

of the city could be excluded. But looking at all trips in the training set it is possible to identify these parts. This can be done by collecting all trips which were close to the cut-off position of the test trip. Below figure shows the start (green)

and end (red) position of the collected trips for four trips in the test set. The blue dot shows the cut-off position of the corresponding test trip. We can see that for the trips T25, T50, and T59 the latitude coordinate has a strong influence on the distribution of the end points and thus also for the remaining driving time.



ABOUT MODEL AND CONCEPT

After observing the data, i did follow:

Get the trips in train with the closest starting location.

Use the weighted average of train trips to estimate trip duration

Result we get by this model is as good as by any other ML trained model..

Challenges:

- 1) Since data was huge, it took long time to process data.**
- 2) As it was a research problem, there wasn't much resource for this.**
- 3) Data was noisy.**

Additional comments and observations

- I tried to identify some key points in the map (e.g. crossings of main routes) and to extract some extra features based on the last visited key point of the car. But it only gave very little improvement.

Conclusion

I had a lot of fun working on this problem and I learned a lot. Given data was very self-descriptive. Many thanks my friends Kaushal and Suchit for clearing doubts.