



Inferring statistical information on the trajectories a user takes in their travel schedule: Each such trajectory can be represented as two GPS pairs, one for the starting point (e.g. the residence of the user) and the other pair for the destination (e.g. the work place). I wrote a code (https://github.com/fkamiab/data_incubator_project/blob/master/Finding_Trajectories.py) that takes these two data points, and finds all trajectories in the Taxi data whose start and end points are one block away from the start and end points of the input trajectory. These are all therefore similar trajectories to the one of the resident, and can be used for giving statistical information on the resident's travel time. This information is plotted as histograms.

One of these histograms (on the left) shows the distribution of the Taxi trajectories similar to the user's based on the hour in the day. We clearly see that there are more trajectories in busy times of the day, such as 9AM or 6PM. The other histogram (on the right) shows the distribution of travel time durations for Taxi trajectories similar to the one of the user. We see they peak at about 10 to 11 minutes which can be used as an average prediction for the time of travel of the user.