Micah Parlato

XTERN Application 2018

Analysis of user location data

After reviewing the user information provided in checkin_dataset.csv I have found two main sets of useful information. The first set of information is a mapping of the locations that specific users travel to. The second and more interesting data set I found was a list of all the locations users traveled to. Both sets of information come with meta-data such as what times they are visited and the time between visits. This information has many possible uses in improving the osXtern product.

By grouping clustered coordinates together I was able to create a list of locations each user recorded visiting. The locations are each tied to statistics about when and how often the user visited that location. A typical example of what data can be collected is presented in *figure 1* below.
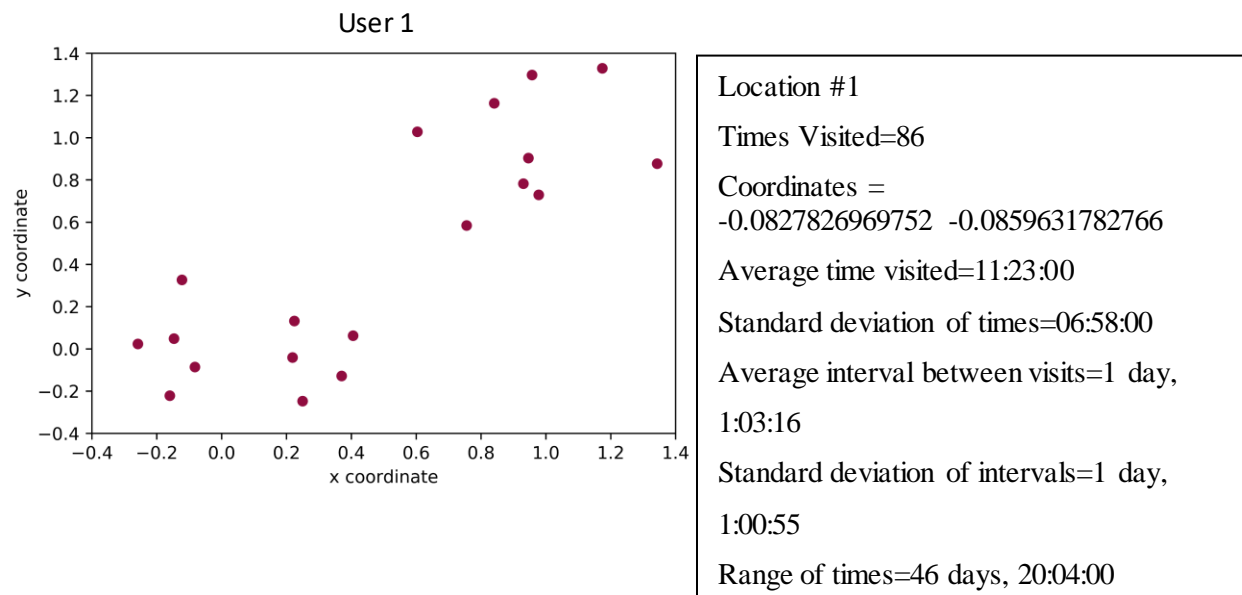

Location #1

Times Visited=86

Coordinates = -0.0827826969752 -0.0859631782766

Average time visited=11:23:00

Standard deviation of times=06:58:00

Average interval between visits=1 day, 1:03:16

Standard deviation of intervals=1 day, 1:00:55

Range of times=46 days, 20:04:00

figure 1

The simplest use of this data would be to integrate it with any sort of mapping application to keep track of "favorite" locations easily. The locations could also be used to track how far a user has traveled in a day by calculating simple distance of all marked locations at the end of the day. Surprisingly there was little to no regularity to when a user visited a location or how long it took for the user to return. I did not check whether there were multiple consistent times during the day when a user might visit a location, but considering the high standard deviation of intervals between visits that seems unlikely. It is possible that if the data is normalized for comparing weekdays to weekdays and weekends to weekends a clearer pattern would emerge, sadly that is beyond the scope of what I do at the moment. Given a more regular dataset I expect it could be used to predict when a person will be in a certain location. This would be useful for

Xbot to help remind people of places they might want to go. For example, a person that exercises every afternoon after work could get a reminder as they leave to stop by the gym along with the traffic conditions on their route. Overall the uses of this data don't bring much benefit to the users, which is why I decided to look elsewhere.

Reorganizing the data as a list of locations along with the information about who goes there and when yielded some promising resul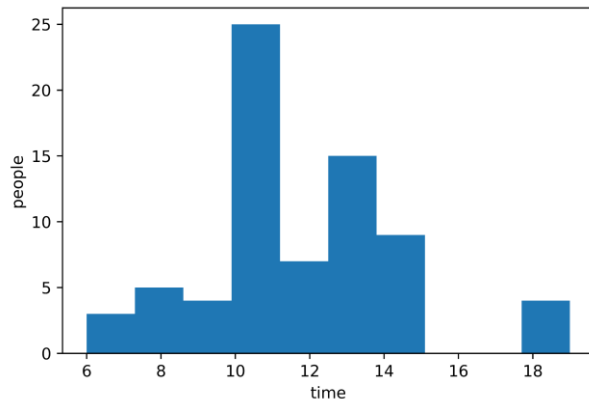ts. *Figure 2* shows a graph of the number of people using th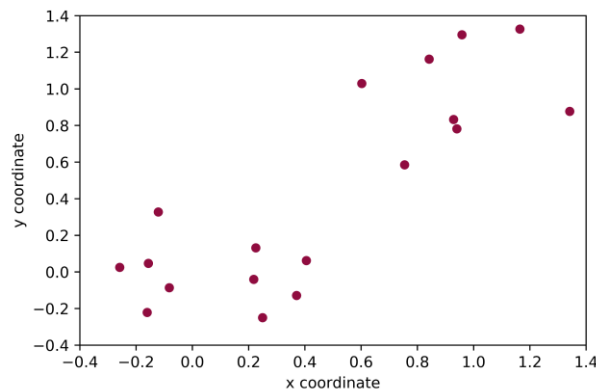e application that are at a location at any given time. Given the current data it isn't possible to tell what the name of a location is, but if the users had an option to name their "favorite" places that could easily be fixed. People using the app could then look at when there will be a lot of people using OSXtern at their favorite places. The data also contains a list of which people frequent that location. A user that enjoys swing dancing at Fountain Square for example, could use the app to find other Xtern users that also go there. The other interesting use of this information is creating a map of places of interest. Pictured in *figure 3* is a map of all the locations people visit within presumably Indianapolis. Users could provide some additional information on places such as a name, review, star rating, etc. People new to the area could go through and find places they should visit. Restaurants, gyms, stores, museums, and even parking garages could be stored and shared within the community of users.

figure 2

figure 3

For having so few data categories this dataset was surprisingly informative. The data gives personal information for a user to keep track of their own habits and preferences. It can also provide anonymous (or de-anonymized depending on your settings) information about what other people in the Xtern community are doing. The data can even help people connect over common interests. For these reasons and more I recommend that more of this type of data be collected along with the extra categories mentioned above such as location name. The rest of the plots and data I collected can be found in data.txt and the plots folder.