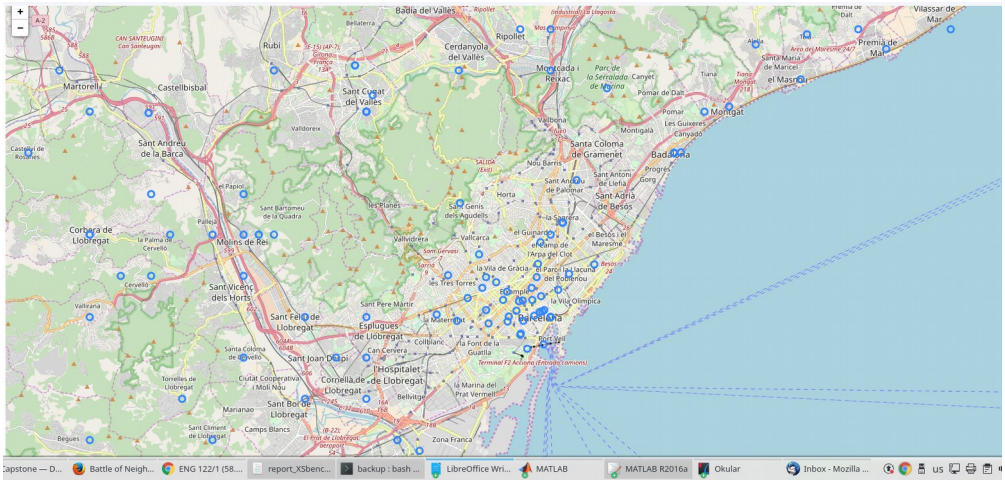
Methodology

Once, the data has been obtained from various sources, it is first needed to be cleaned and organized. The methodology is defined through the following points.

- All, the coordinate points obtained in the data frame is uploaded into the folium map plotter, as such;



- After, plotting the data, we filter out the coordinates which are not inside the main city of Barcelona
- The filtered data is then used for further analysis. The name of each district is added manually (because the district names weren't given beforehand) owing to small number of total districts in the Dataframe (45 total)
- After renaming the districts, we login to Foursquare API using our client credentials and obtain the results in json
- We use the respective latitudes and longitudes to get the corresponding nearby venues for all the neighborhoods

1. We generate one-hot vectors for the venue types in all the neighborhoods and generate most common venues.

- We use the list of most common venues to get the top10 venues
- Then, K-means clustering is used to cluster the neighborhoods based on the similarity of the venue types.
- Further the data of rent of house per square km is added for final recommendation system