Data

The data acquired from Wikipedia pages, [town list](https://www.townslist.co.uk/) UK and restructure to csv file for easier manipulation and reading.

Another aspect to consider for this project is the Foursquare data. I believe that the data as good

as provided, meaning although we are using Foursquare data for segmentation and clustering,

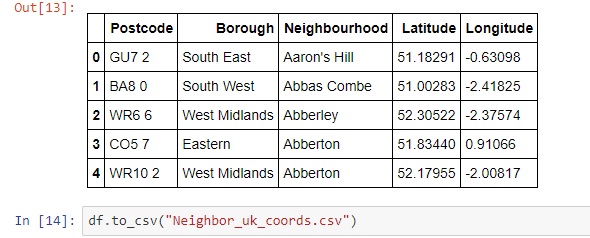
the amount and accuracy of data captured can't 100% determine correct classification in real

world.

To start, let's get and look at the data. I've already downloaded it, so let's read it (from local

drive) and load it to data frame: Using geocoder, we able to get Latitude and longitude for each

area.



Methodology

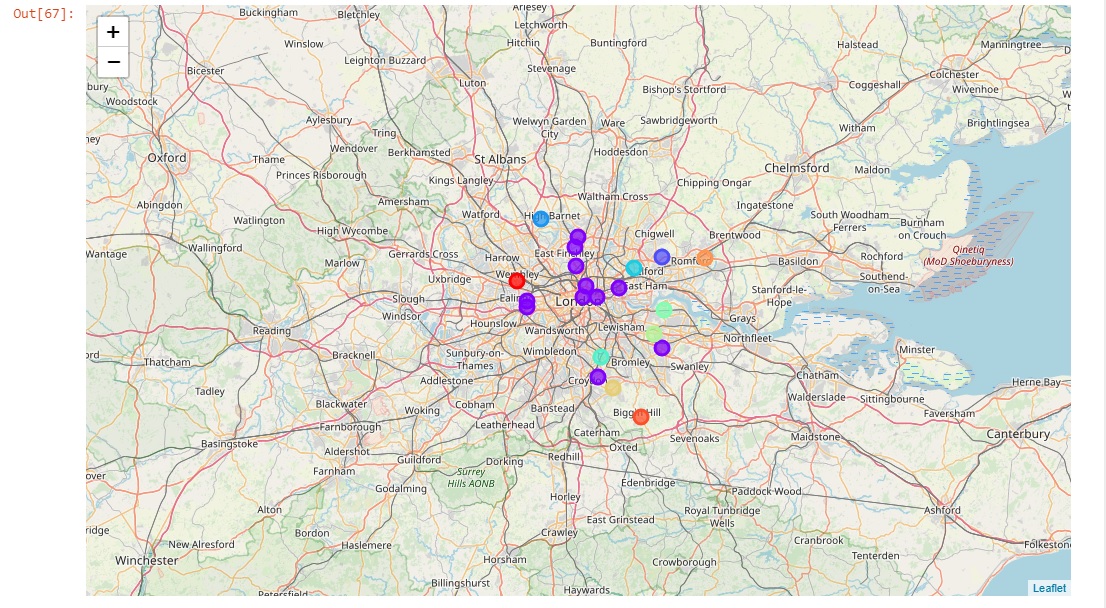
* In this project, I will use the basic methodology as taught in Week 3 lab.
* Above, we have done convert addresses into their equivalent latitude and longitude values.
* Then we will use the Foursquare API to explore neighbourhoods in United Kingdom.
* After that, explore function to get the most common venue categories in each

neighbourhood, and then use this feature to group the neighbourhoods into clusters.

* K-means clustering algorithm will be using to complete this task. And the Folium library to

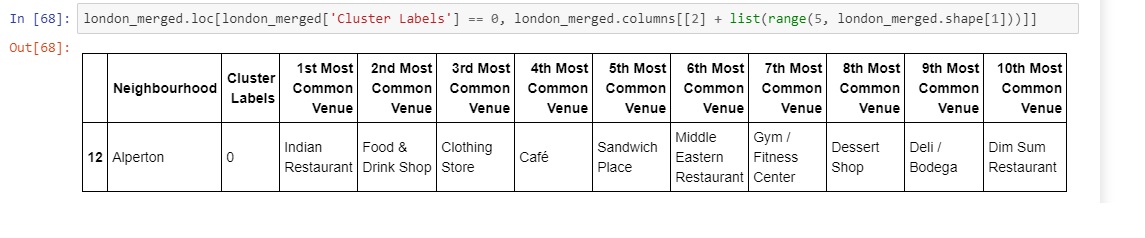
visualize the neighbourhoods in London and identify their clusters.

K-means for London

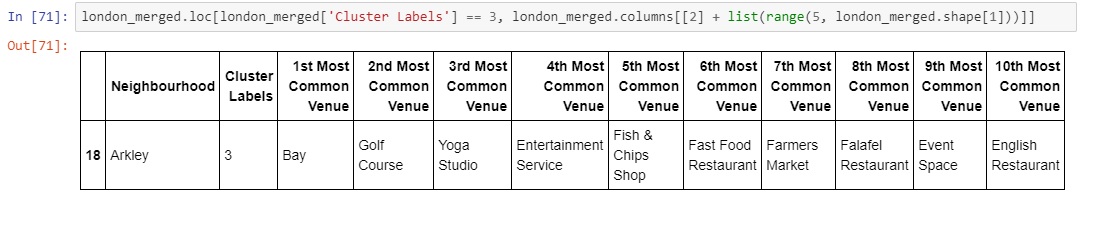


Results

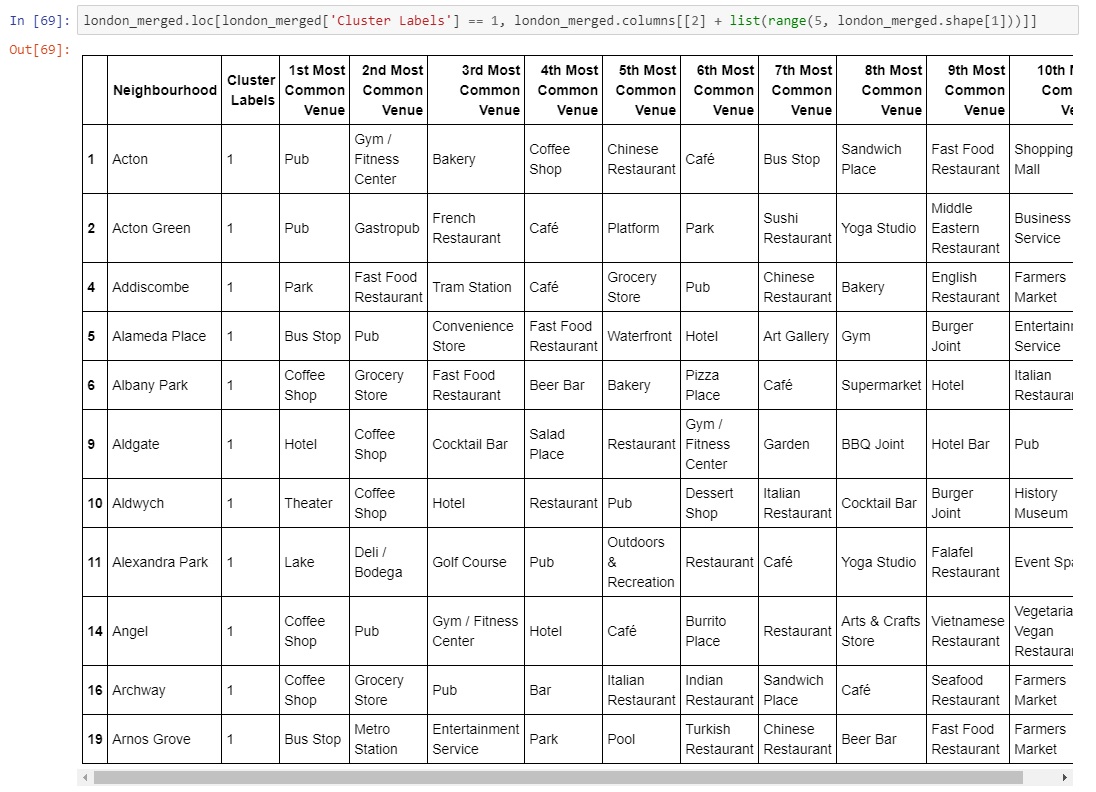
Cluster 1 –



Cluster 2 --



Cluster 3 –

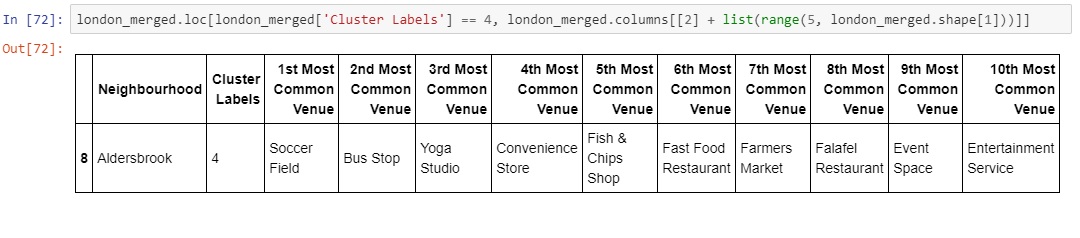


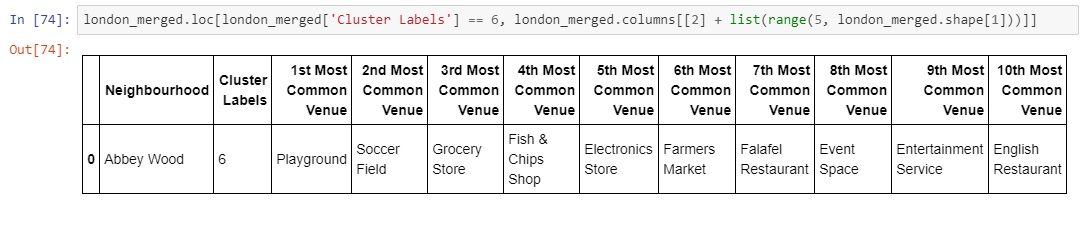
Cluster 4 –





Cluster 5 –





Discussion

Based on cluster for each city above, we believe that classification for each cluster can be done better with calculation of venues categories (most common) in each locations. Referring to each cluster, we can't determine clearly what represent in each cluster by using Foursquare – Most Common Venue data.

However, for the sake of this project we assumed each cluster preference for people moving to London as follow:

Cluster 1: Middle east and South Asian Country

Cluster 2: Business man, Investor, especially bachelors as pub, entertainment dominates

Cluster 3: Tourist and recreation

Cluster 4: East Asians – especially Chinese and Japanese

Cluster 5: Children, youngster looking for growth and training in sports

What is lacking at this point is a systematic, quantitative way to identify and distinguish

different locations and to describe the correlation most common venues as recorded in

Foursquare. The reality is however more complex: similar cities might have or might not have similar common venues. A further step in this classification would be to find a method to extract these common venues and integrate the spatial correlations between different of areas or district.

We believe that the classification we propose is an encouraging step towards a quantitative and systematic comparison of the different cities. Further studies are indeed needed in order to relate the data acquired, then observe it to more meaningful and objective results.

Conclusion

Using Foursquare API, we can captured data of common places all around the world. Using it, we refer back to our main objectives, which is to determine; The similarity or dissimilarity of both cities classification of area located inside the city whether it is residential, tourism places, or others In conclusion, London is the centre of attraction among people around the world. However, to declare locations of London on common venues visited is quite difficult. All cities are similar in some venues also dissimilar in certain venues. And for classification based on common venues, again we must have more systematic or quantitative way to identify and declare this. Comparison can be made, but no such method or quantitative data to determine this. We hope in the future, a method to determine it can be establish and explore for references.