**Intoduction**

**England** is a [country](https://en.wikipedia.org/wiki/Country) that is [part](https://en.wikipedia.org/wiki/Countries_of_the_United_Kingdom) of the [United Kingdom](https://en.wikipedia.org/wiki/United_Kingdom).[[5]](https://en.wikipedia.org/wiki/England#cite_note-5)[[6]](https://en.wikipedia.org/wiki/England#cite_note-6)[[7]](https://en.wikipedia.org/wiki/England#cite_note-7) It shares land borders with [Wales](https://en.wikipedia.org/wiki/Wales) to the west and [Scotland](https://en.wikipedia.org/wiki/Scotland) to the north-northwest. The country covers five-eighths of the island of [Great Britain](https://en.wikipedia.org/wiki/Great_Britain), which lies in the [North Atlantic](https://en.wikipedia.org/wiki/Atlantic_Ocean#Northern_Atlantic), and includes [over 100 smaller islands](https://en.wikipedia.org/wiki/List_of_islands_of_England), such as the [Isles of Scilly](https://en.wikipedia.org/wiki/Isles_of_Scilly) and the [Isle of Wight](https://en.wikipedia.org/wiki/Isle_of_Wight).

**London** is the [capital](https://en.wikipedia.org/wiki/Capital_city) and largest city of both [England](https://en.wikipedia.org/wiki/England) and the [United Kingdom](https://en.wikipedia.org/wiki/United_Kingdom). London is considered to be one of the world's most important [global cities](https://en.wikipedia.org/wiki/Global_city) and has been termed the world's most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment friendly, most popular for work, and the most vegetarian friendly city in the world. London exerts a considerable impact upon the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism and transportation. London ranks 26 out of 300 major cities for economic performance. It is the most-visited city as measured by international arrivals and has the busiest [city airport system](https://en.wikipedia.org/wiki/List_of_busiest_city_airport_systems_by_passenger_traffic) as measured by passenger traffic. It is the leading [investment](https://en.wikipedia.org/wiki/Foreign_direct_investment) destination, hosting more [international retailers](https://en.wikipedia.org/wiki/Retail#Global_top_ten_retailers) and [ultra high-net-worth individuals](https://en.wikipedia.org/wiki/Ultra_high-net-worth_individual" \o "Ultra high-net-worth individual) than any other city. In [2012](https://en.wikipedia.org/wiki/2012_Summer_Olympics), London became the first city to have hosted three modern [Summer Olympic Games](https://en.wikipedia.org/wiki/Summer_Olympic_Games). London has a diverse range of people and cultures, and more than 300 languages are spoken in the region. Its estimated mid-2016 municipal population (corresponding to [Greater London](https://en.wikipedia.org/wiki/Greater_London)) was 8,787,892, the most populous of any [city in the European Union](https://en.wikipedia.org/wiki/List_of_cities_in_the_European_Union_by_population_within_city_limits) and accounting for 13.4% of the UK population. [London's urban area](https://en.wikipedia.org/wiki/Greater_London_Built-up_Area) is the second [most populous in the EU](https://en.wikipedia.org/wiki/List_of_urban_areas_in_the_European_Union), after Paris, with 9,787,426 inhabitants at the 2011 census.

**Objective**

In this project, we will study in detail the area classification using Foursquare data and machine learning segmentation and clustering. The aim of this project is to segment areas in London based on the most common Locations captured from Foursquare. Using segmentation and clustering, we hope we can determine the area located inside the city whether it is residential, tourism places, or others

**Target Audience**

Through this project we are expecting following people to benefit out of the findings.  
• Tourist.  
• People migrating city for work.  
• Business person looking for new location to start office etc.  
• Restaurants to finalise menu based on the type or people, their likings based on feedbacks etc.  
• And many more.

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,

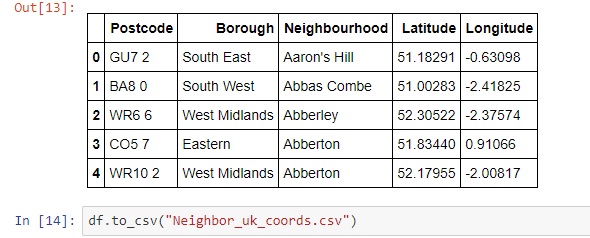
the amount and accuracy of data captured can't 100% determine correct classification in real

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

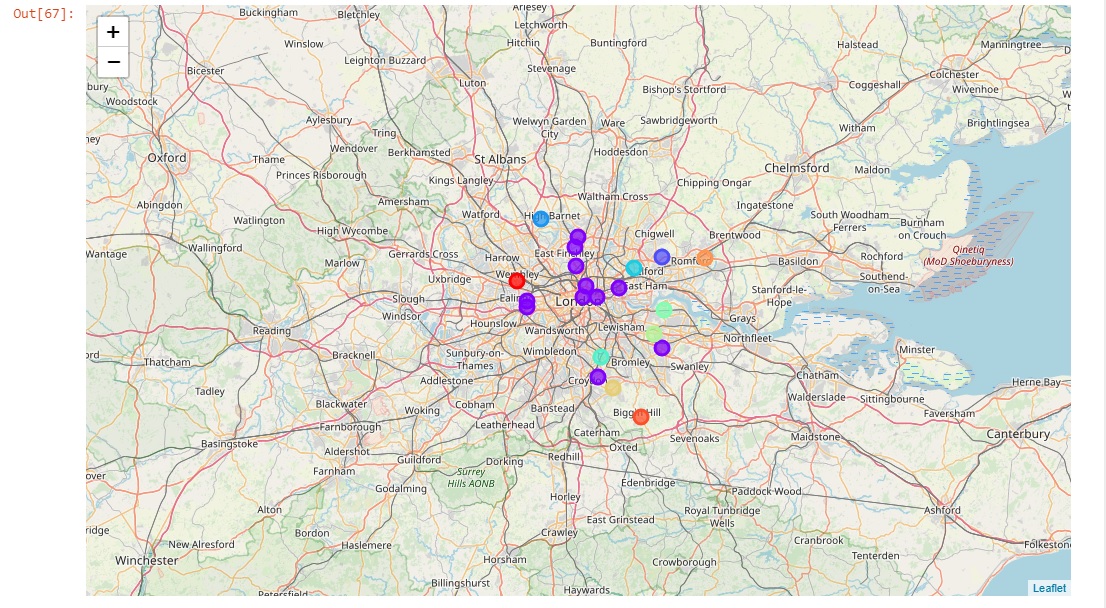
* In this project, I will use the basic methodology as taught in Week 3 lab.
* Above, we have done convert addresses into their equivalent latitude and longitude values.
* Then we will use the Foursquare API to explore neighbourhoods in United Kingdom.
* After that, explore function to get the most common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

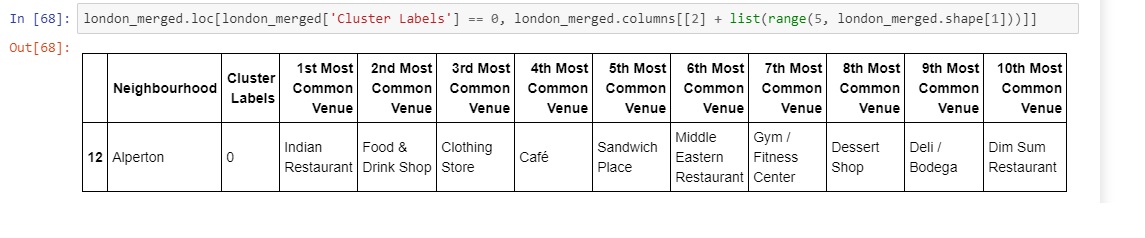
visualize the neighbourhoods in London and identify their clusters.

K-means for London

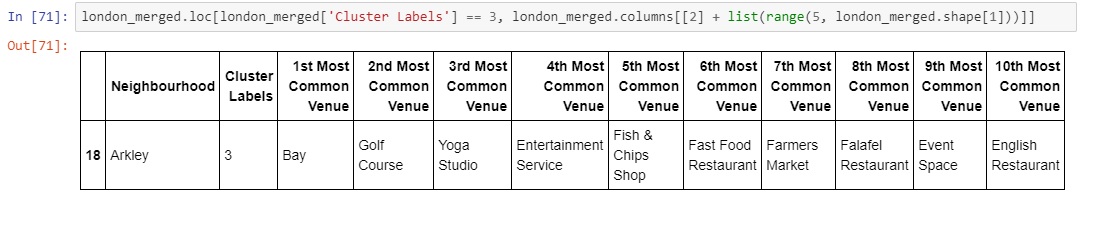


Results

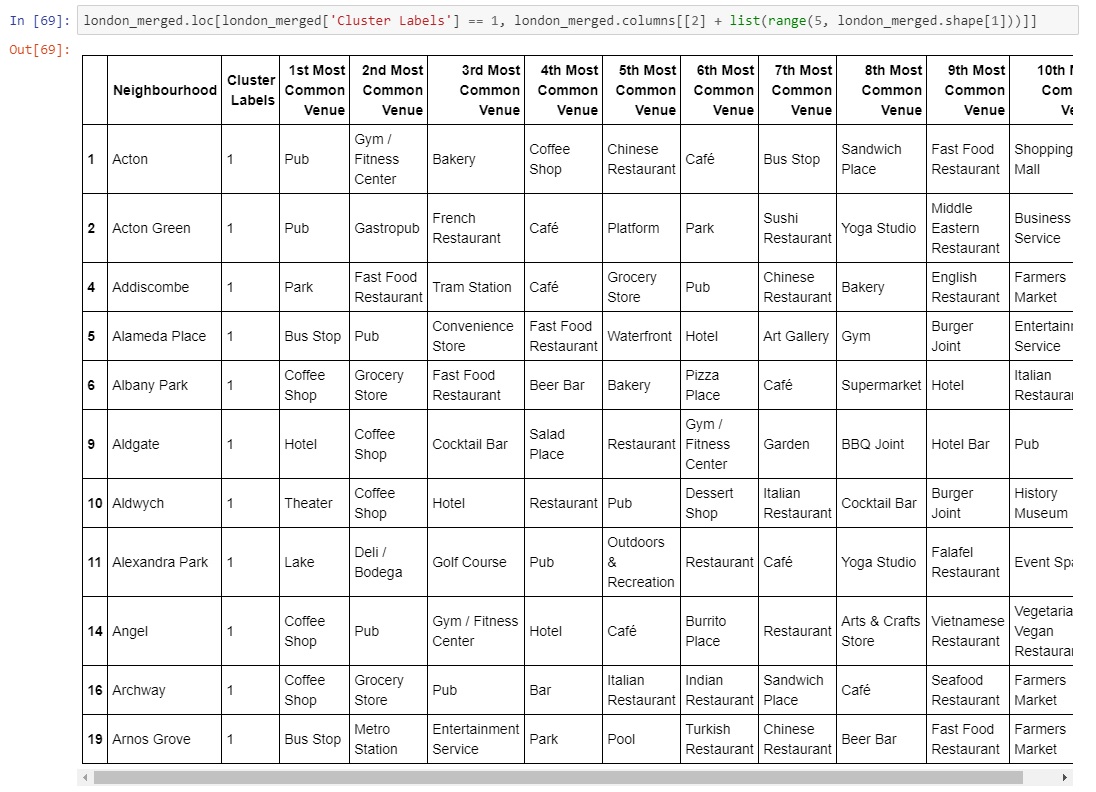
Cluster 1 –



Cluster 2 --



Cluster 3 –

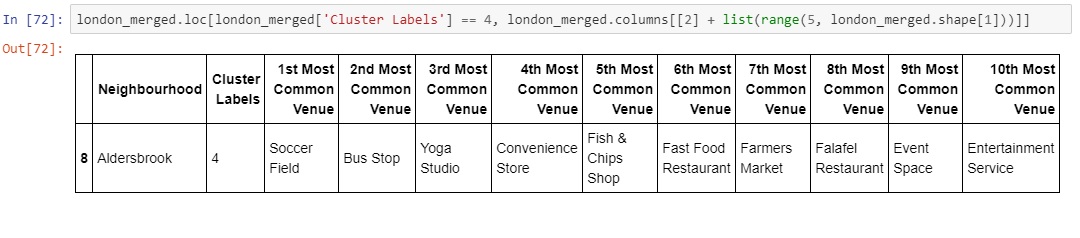


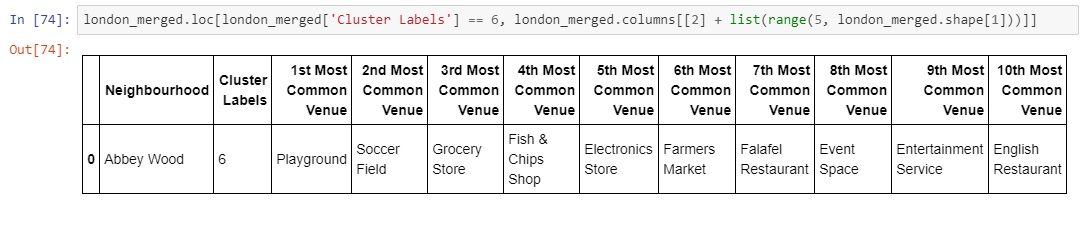
Cluster 4 –





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.