**The Battle of Neighbourhoods**

Harsh Shah

2nd August 2020

**1. Introduction**

**1.1 Background**

In London, there are numerous boroughs. The city is famous for its many touristic spots and is one of the most visited cities on the planet. In addition to tourists, many students come to London in order to study at university. As well as this, many people come for short term stays for e.g. internships, as well as for graduate programs.

**1.2 Problem**

Finding a good place to stay is very important, somewhere that has local access to shops and supermarkets, as well as restaurants, bars and a number of other facilities.

HS is moving to London in order to study at University College London. In addition to university life, HS also wants to get a feel of the city's culture, food etc. He's looking to live in a neighbourhood that is culturally vibrant but not too touristy, as he feels that he can visit tourist sites in his spare time, therefore the proximity of tourist attractions is not something that matters to him.

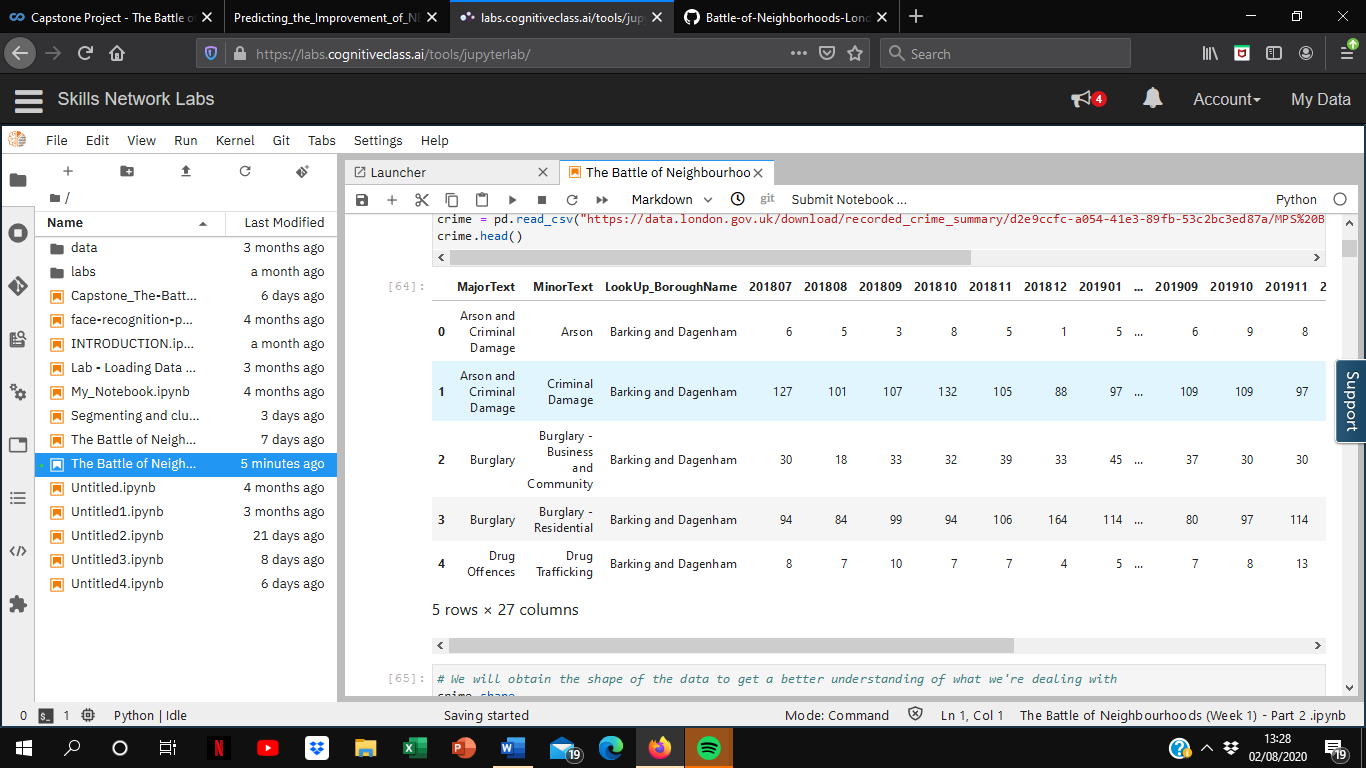
From the perspective of students, many factors are involved when searching for the best accommodation, which includes distance and rent. However, this project will focus on the general atmosphere of different areas of London as well as safety.

The research carried out is expected to be of benefit to international students looking to live in London, because despite the opportunity to explore the local culture, they would very likely want to feel at home at the same time.

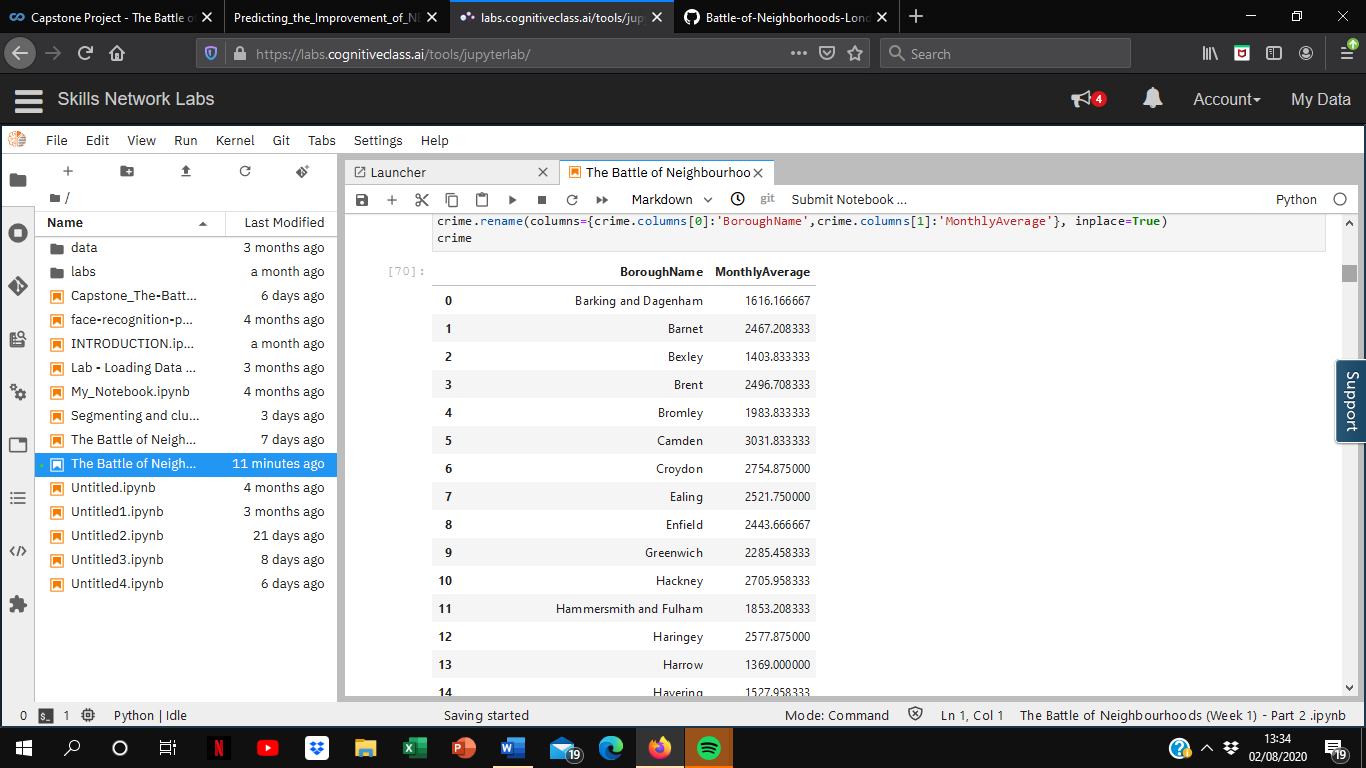
**2. Data Acquisition and Preprocessing**

In the project, I will be using the following datasets to produce an outcome - Borough Level Crime, List of London boroughs and Foursquare API. After using legitimate resources to acquire them, they will be cleansed into more useful forms so that they can be analysed further.

**2.1 Borough Level Crime**

The table above shows crime numbers per crime type at borough level over the last 24 months. The data consists of 1569 observations and 27 columns. It was obtained from the London Datastore.

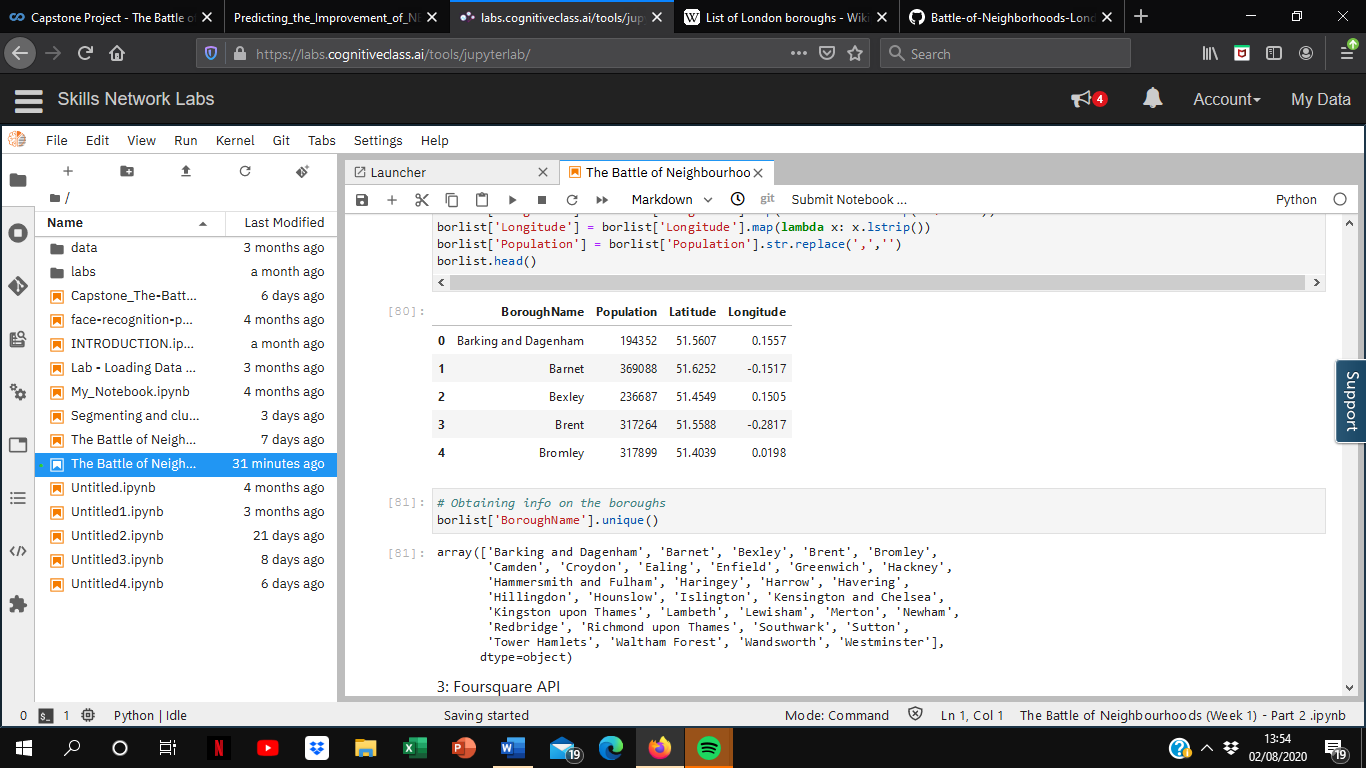
To go into further detail, the number of crimes were used to calculate monthly averages, without taking into consideration crime categories, for simplification purposes.



**2.2 List of London Boroughs**

The second dataset that I used provides information on the different London boroughs, obtained from Wikipedia

From this data, we will only be using the population data and coordinates. The population data can be used to calculate the ratio of crime to population in order to generate a better comparison, while the coordinates can be used to obtain geographical data from Foursquare. The simplified data looks like this:

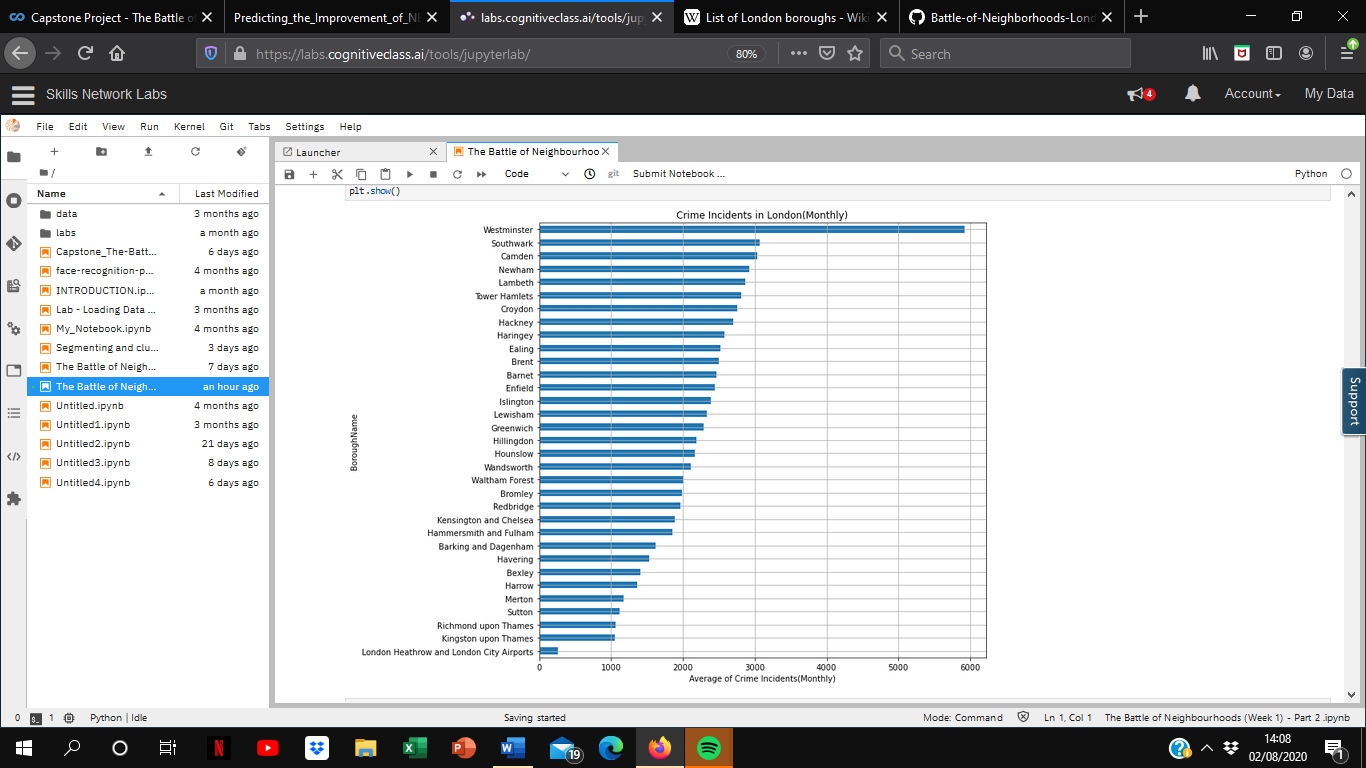


**2.3 Foursquare API**

Foursquare API was used to obtain the 50 most popular venues in each borough, which was done using the ‘explore’ function to request the URL. As a result, I was able to get data which looks like this:

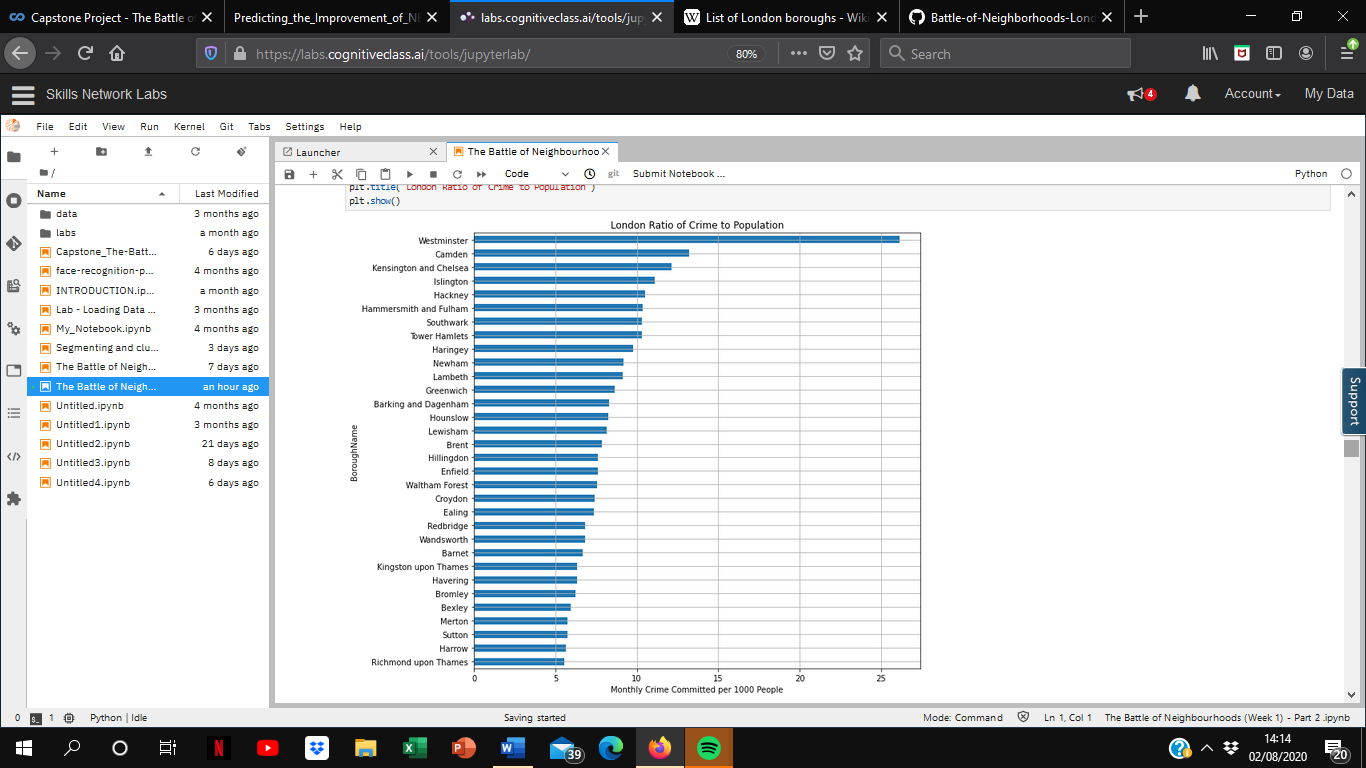
**3. Methodology**

**3.1 Exploratory Analysis**

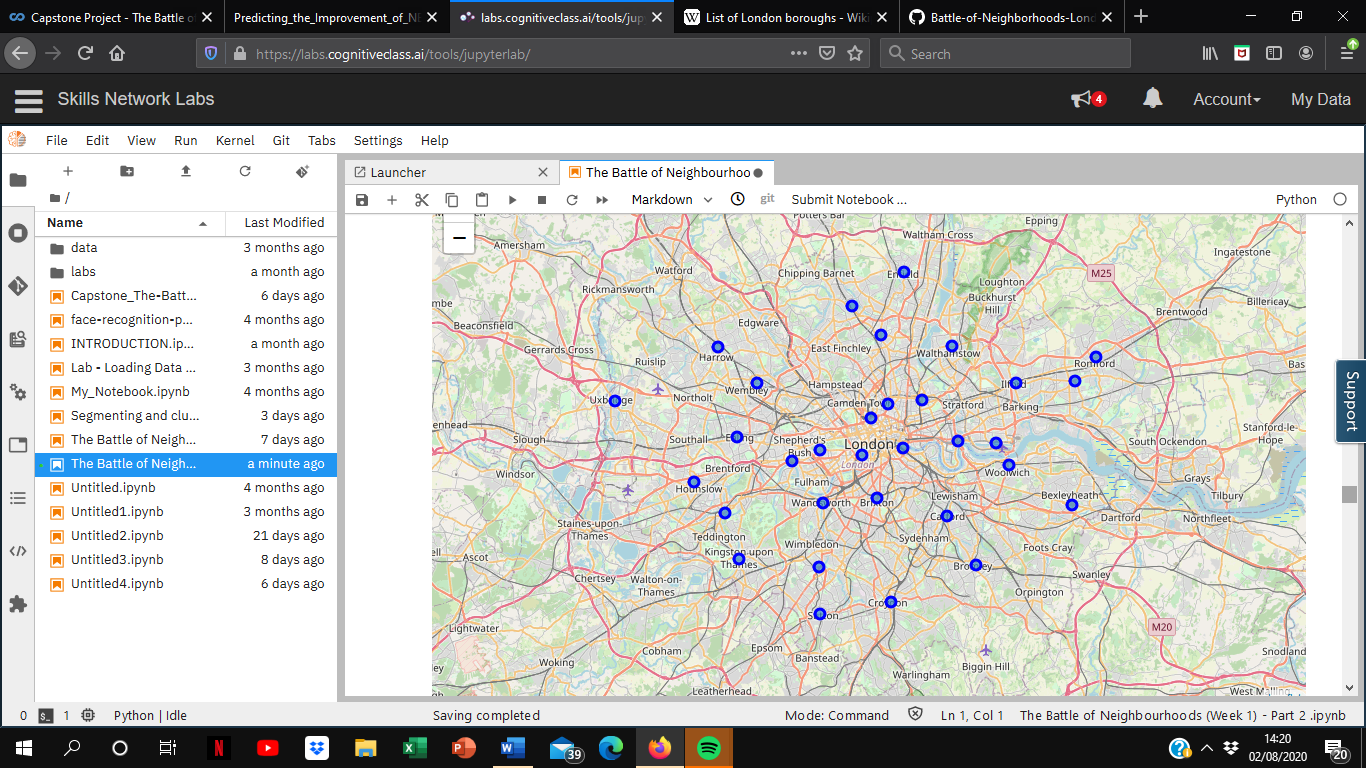
Using the cleansed data, I created visualisations to better represent the data

This bar chart shows the boroughs in descending order of monthly crime incidents. As you can see, Westminster has the highest number of recorded crimes, followed by Southwark, Camden and Newham.

Despite this, I felt that it wouldn’t be sensible to directly compare the number of recorded crimes due to differences in population. Therefore, I used the population to calculate the number of crimes per 1000 people in each borough:

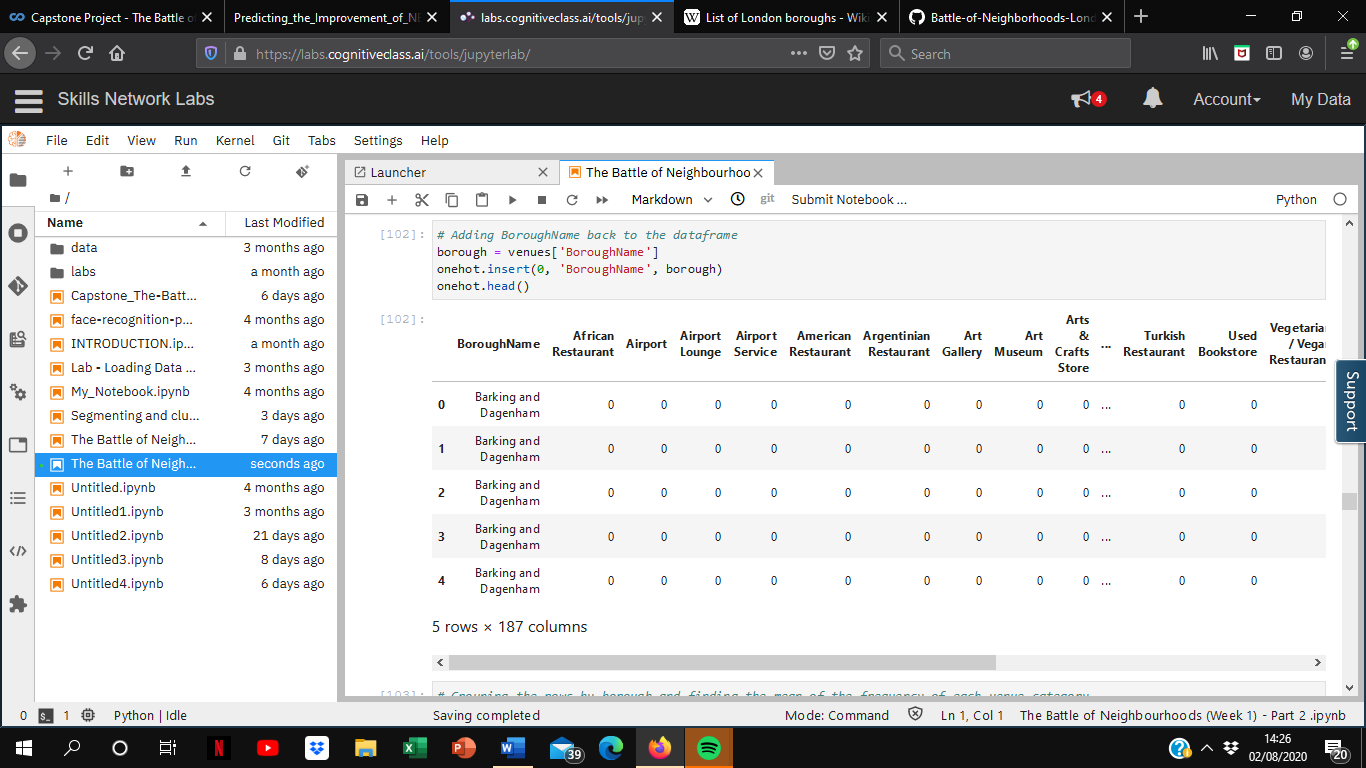


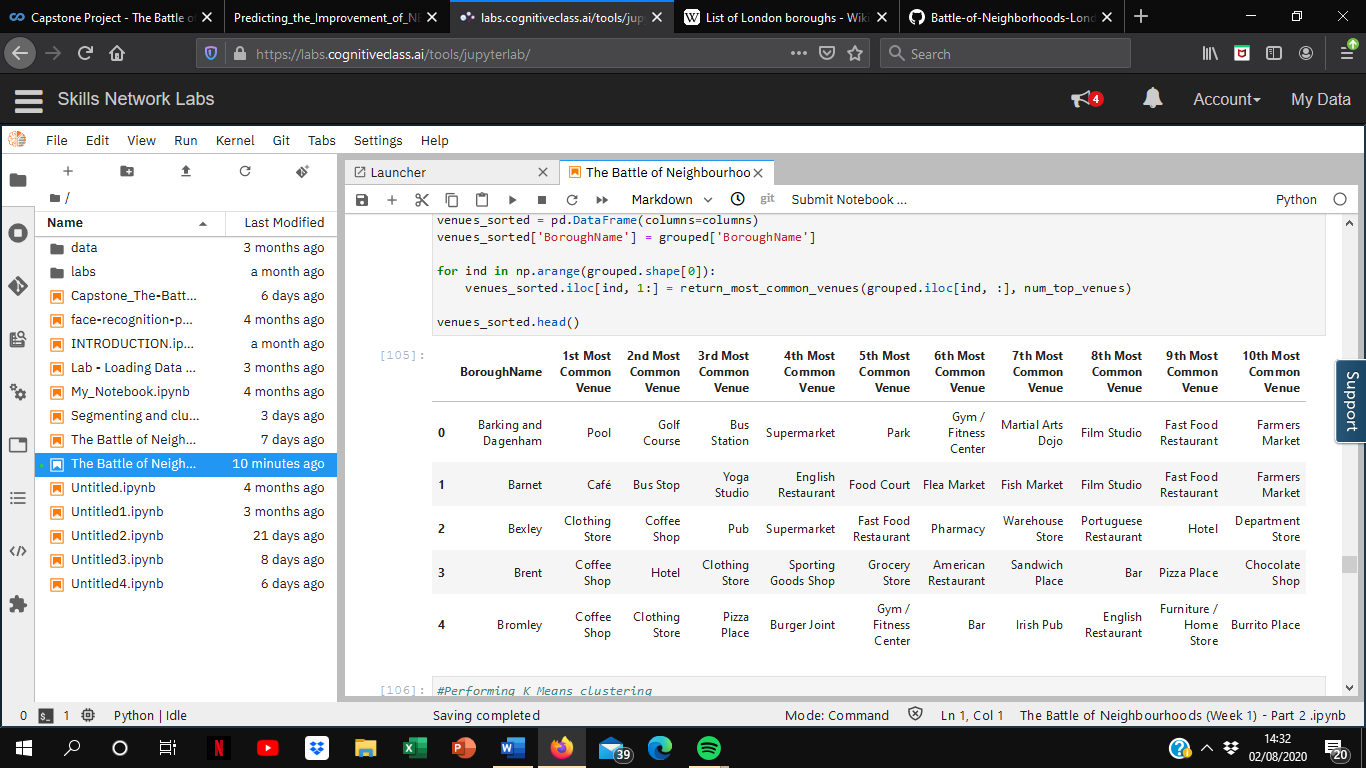
It can still be seen that Westminster and Camden still remain high up the graph, although Kensington and Chelsea, Islington and Hackney have also moved towards the top of the chart.

Also, before conducting further analysis, I decided that it would be a good idea to take a look at the different locations to get a better idea of London as a region.

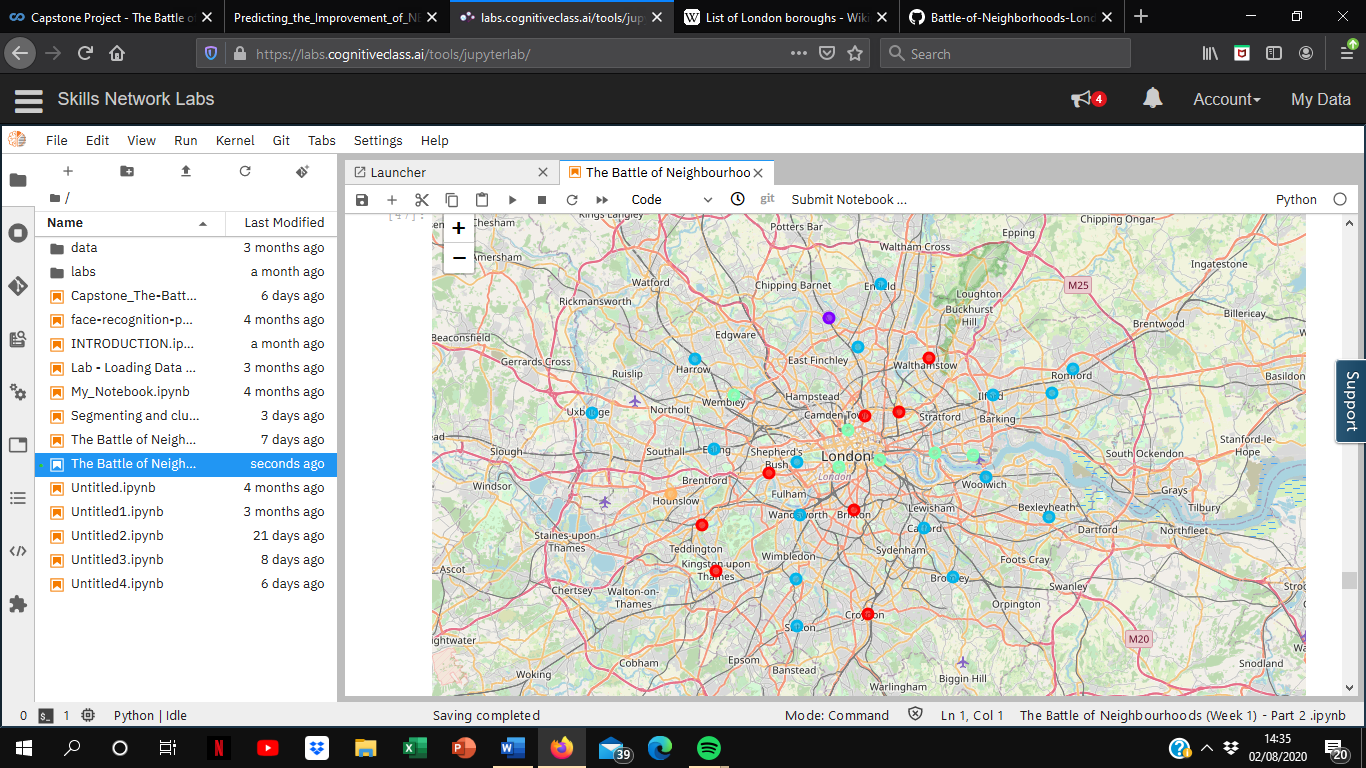
**3.2 Cluster Analysis**

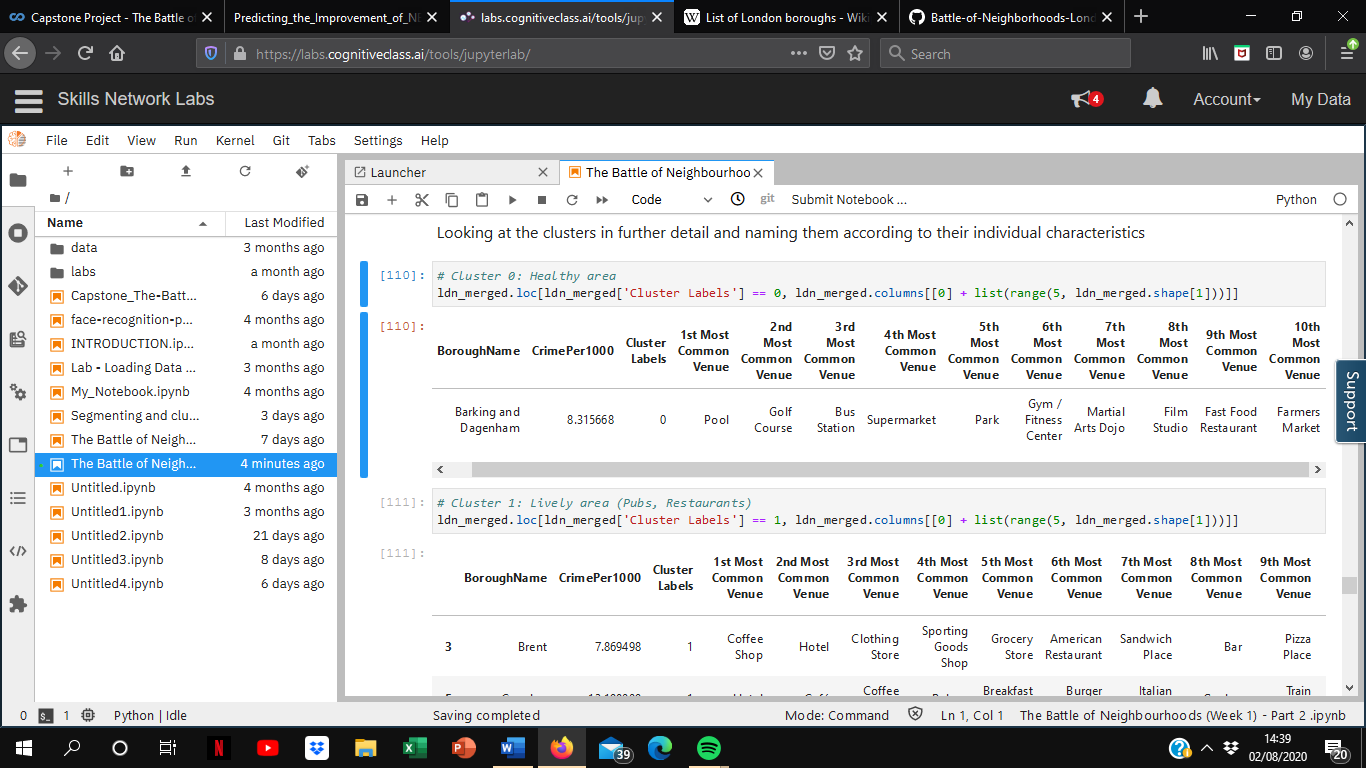
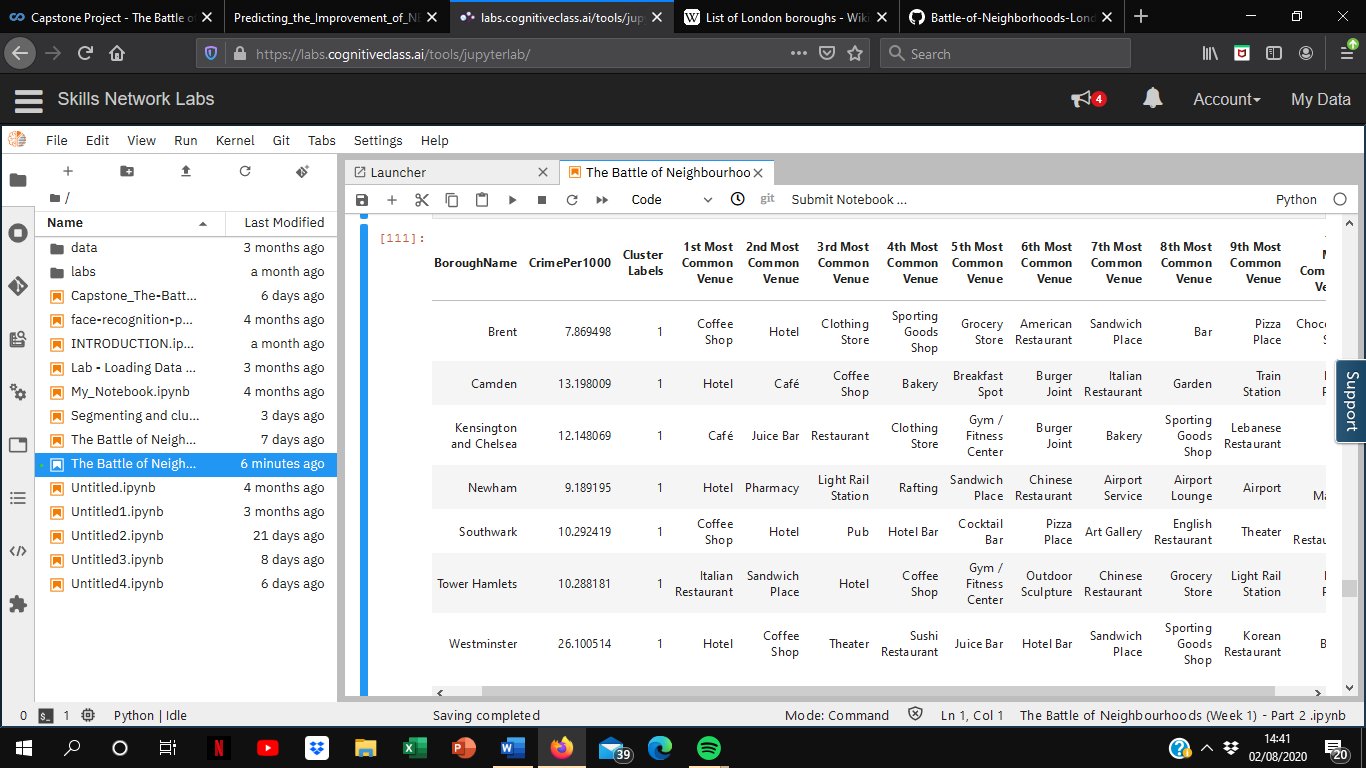
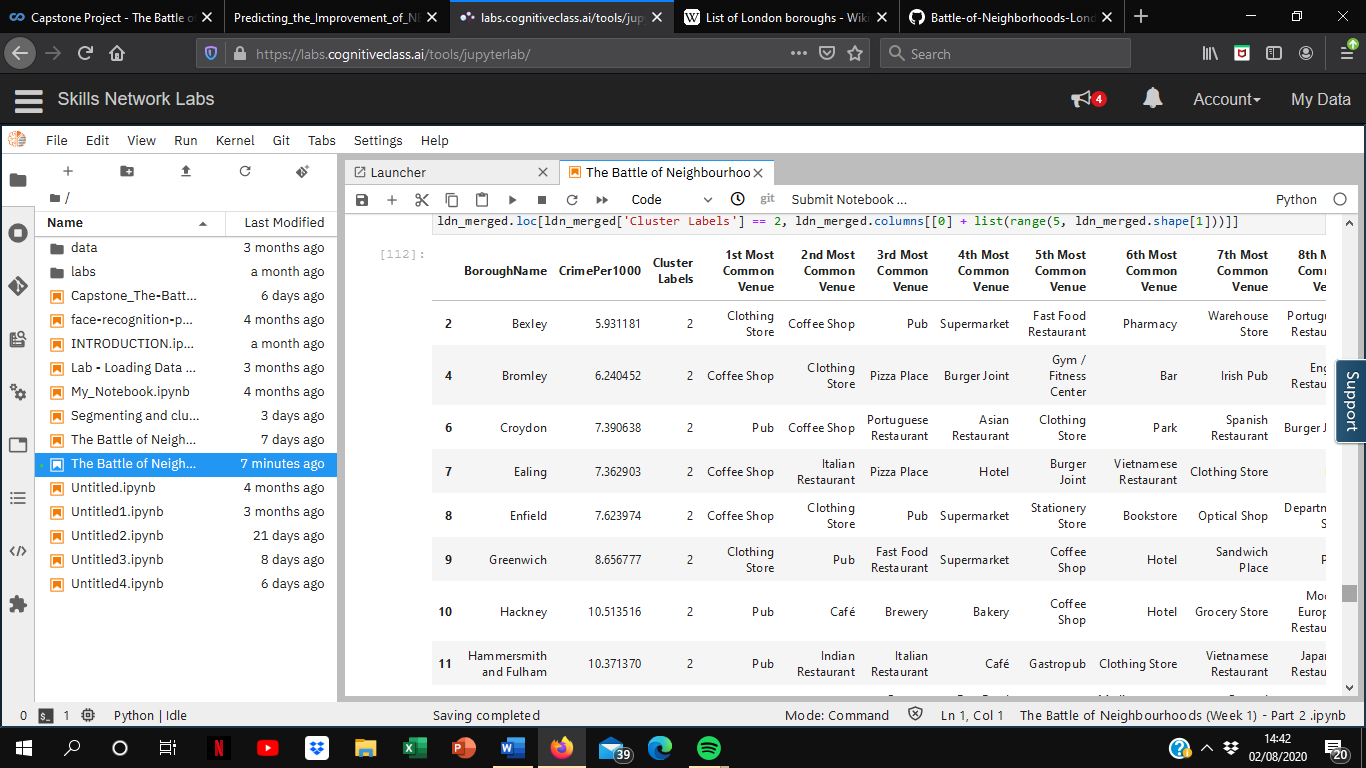
Next, K-means clustering was performed in order to group the boroughs based on the boroughs they had using the Foursquare data, to get a feel of the general atmosphere of each of the boroughs.

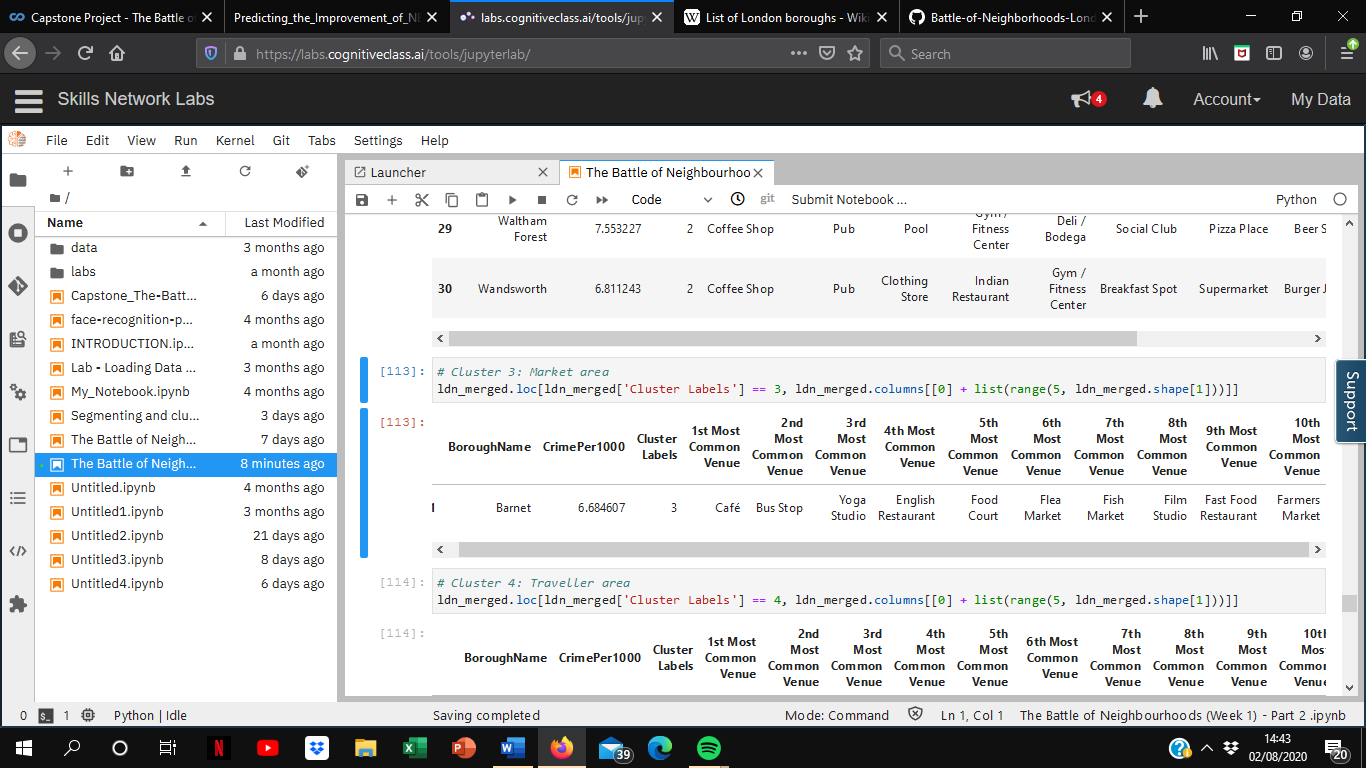
The first step that was taken was to carry out ‘one hot encoding’, in order to create binary values for each of the clusters.

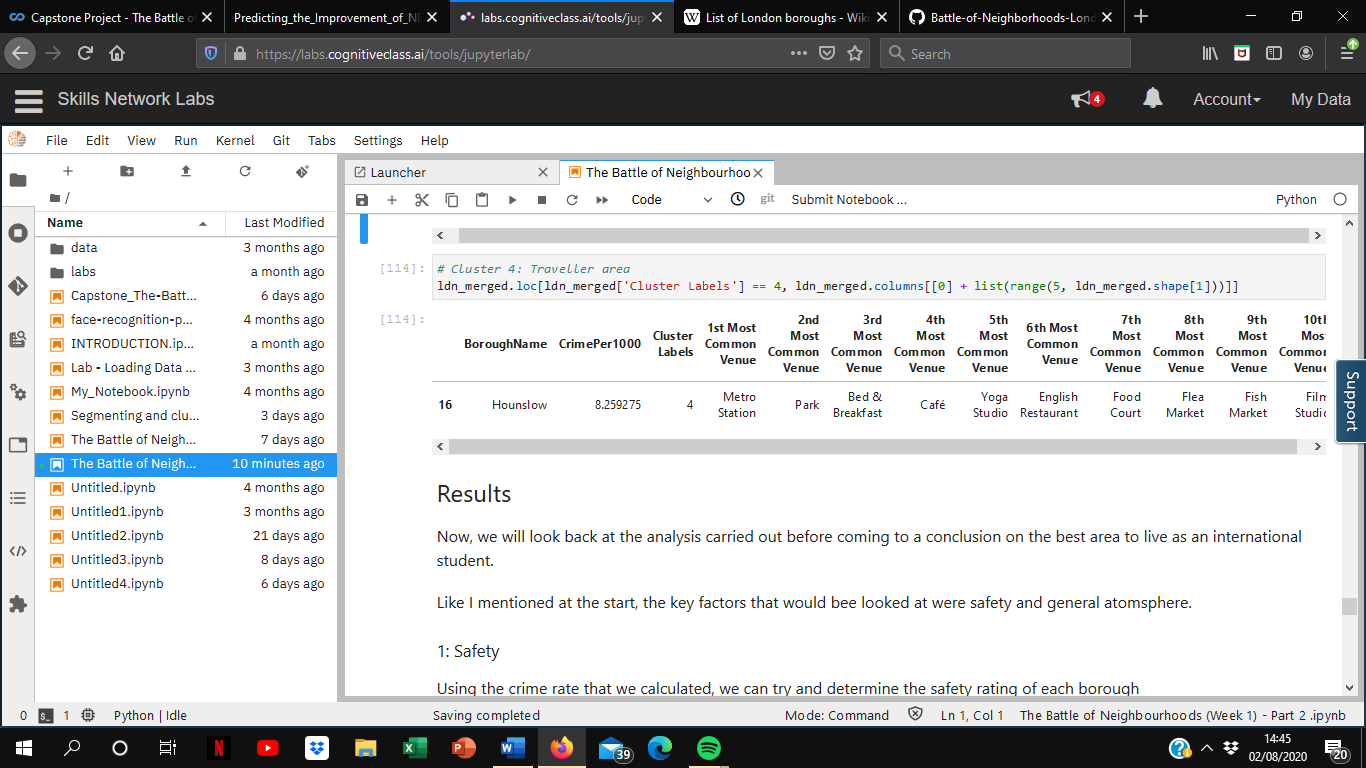
Next, the data was grouped according to borough name in order to find out the number of venues of each category that exist within the 50 most popular venues. However, due to the lack of data from Foursquare, some boroughs displayed less than 50 venues, therefore the category count was altered to the frequency of how often each category appears amongst others. Based on this frequency, we could obtain a list of the most common venue categories within each borough as follows:

Using these categories, K-means clustering was then carried out in order to group the boroughs into 5 different clusters based on similarity. The different coloured dots show different clusters:

Based on the characteristics that each of the clusters possess, we have named each of the clusters according to how their characteristics are best depicted:

* Cluster 0 – Healthy area (Gyms, pools)
* Cluster 1 – Lively area (Pubs, Restaurants)
* Cluster 2 – Shopping area (Coffee shops, clothing stores)
* Cluster 3 - Market area



* Cluster 4 – Traveller area

**4. Results**

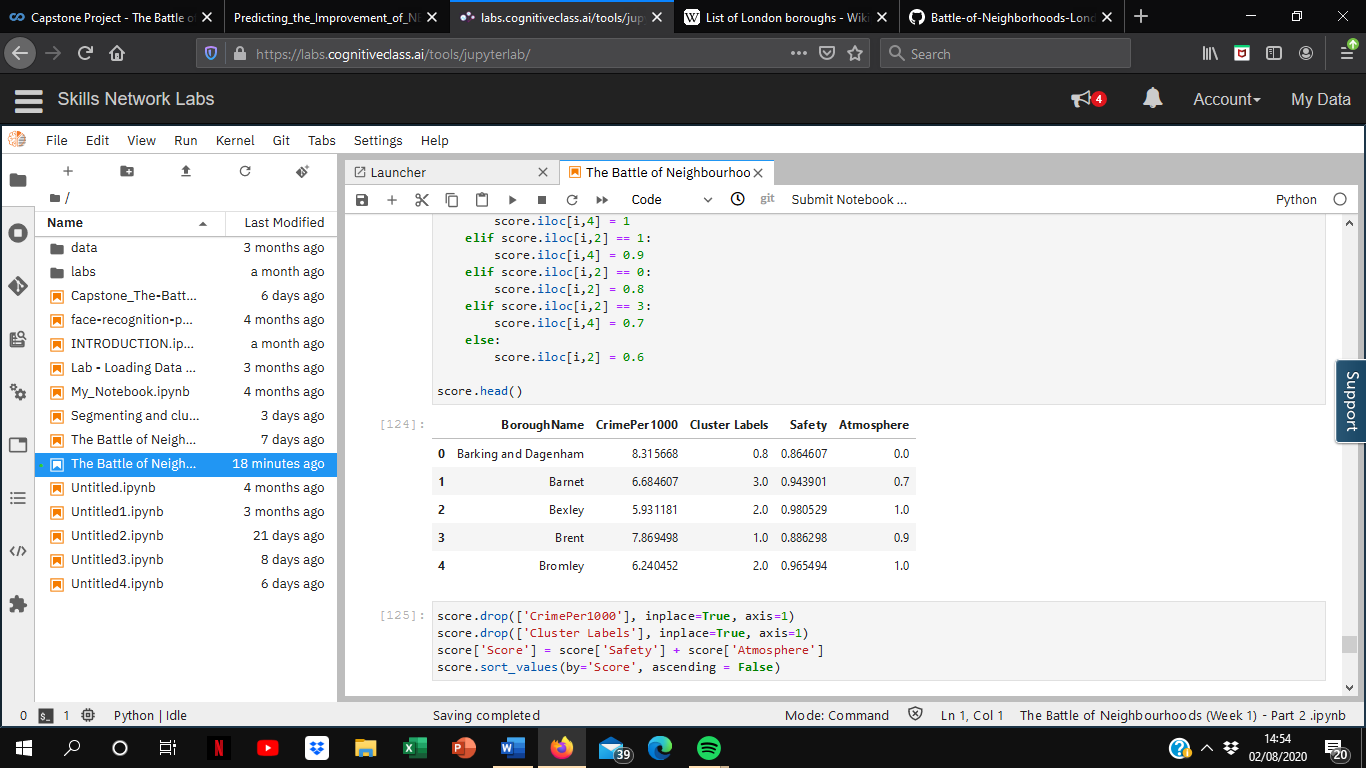
Using different analysis, I was able to discover the best areas to live based on the criteria of general atmosphere and safety. Now, we will look back at all the analysis performed in the project before making a conclusion on which area would be the best for HS to live in.

**4.1 Safety**

For safety, I decided to normalise the crime to population video and reversed the safety score so that the closes value to 1 represents the area that has the lowest crime per head.

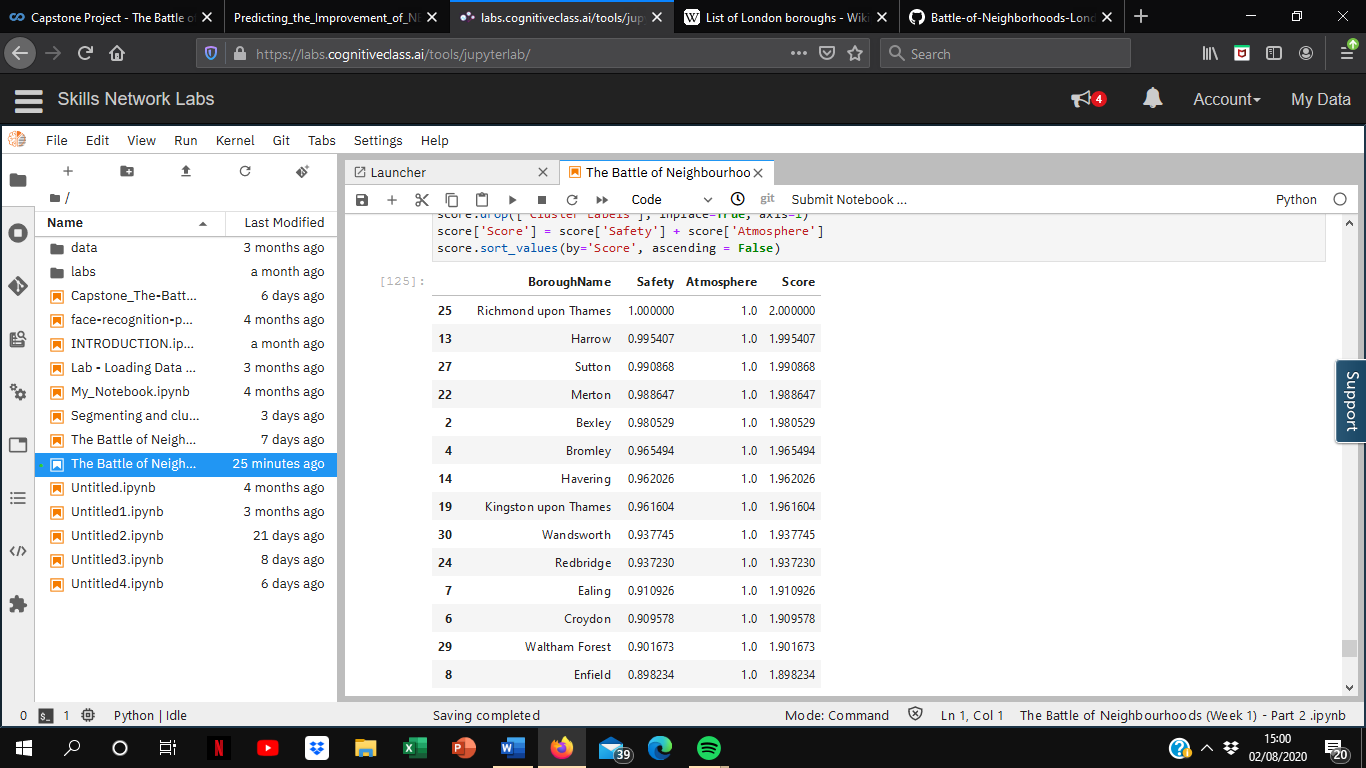
**4.2 Atmosphere**

With regards to atmosphere, I decided to give an arbitrary value to each cluster on the basis of personal preference, since preference is difficult to quantify without being subjective. The highest value was given to the Shopping Area (Cluster 2) while the lowest was given to the Traveller Area (Cluster 4).



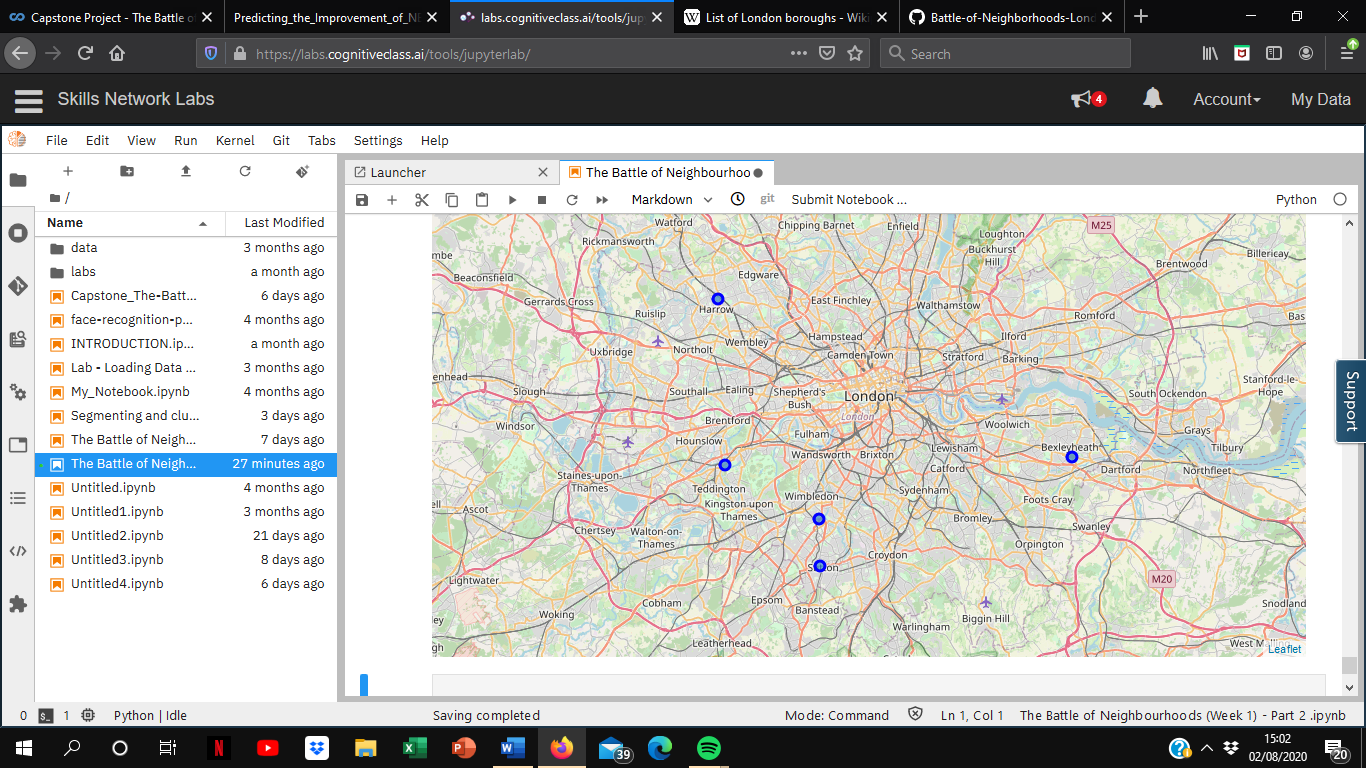
**4.3 Final Score**

Finally, by adding the Safety score to the Atmosphere value, I was able to find the best area to live in, which was Richmond upon Thames, scoring 2.0.



**5. Conclusion**

**5.1 Final result**



Based on the analysis, I have found that the 5 boroughs below would be the best places for HS to live according on the criteria that he set out, based on general atmosphere and safety. They all belong to the Shopping Area cluster, which contains many clothing shops and coffee shops, as well as a number of restaurants that serve a range of cuisines. The only thing that sets them apart is the safety score.

**5.2 Limitations**

From the map, we can see that all of the top 5 boroughs are located far out from the city, which is due to the limitations held by the research carried out.

As well as taking into account crime rates and the types of stores that exist in each area, I could’ve perhaps also taken into account the number of each type of store that exists in each borough, as well as other factors such as house/rental prices.

Nevertheless, performing this research was still incredible enjoyable as I was able to explore each borough in a lot of depth.

**References**

* **“**Borough Level Crime”, London Datastore
* “List of London Boroughs”
* Foursquare API
* IBM Data Science Professional Certificate notes, Coursera