Coursera Capstone Project: Opening a new Coffee shop in London

Ameni KACEM

1. INTRODUCTION

London is the capital and largest city of England and the United Kingdom. Standing on the River Thames in the south-east of England, at the head of its 50-mile (80 km) estuary leading to the North Sea, London has been a major settlement for two millennia. Londinium was founded by the Romans.

The City of London, London's ancient core − an area of just 1.12 square miles (2.9 km2) and colloquially known as the Square Mile − retains boundaries that closely follow its medieval limits. The City of Westminster is also an Inner London borough holding city status. London is governed by the Mayor of London and the London Assembly.

For opening a new restaurant or a coffee shop, having a good menu and professional staff is important to success. However, these are not the only things to consider. Having a good location can give the business another push toward success.

In this context, a client would like to open a new Coffee Shop in London city. The main task is to find the best location in London that fits the client's need and criteria. As the client is French and the coffee shop as well, one of the criteria that the location will be around french restaurant. This will lead to have people that are interested in French culture and cuisine.

Moreover, the location should also be close to pubs and hotels and do not contain too much coffee shops.

2. DATA

We used, in this project, the list of London's main streets provided by https://www.latlong.net/category/streets-235-17.html. Thus, we can have the longitude and latitude for each street.

Then, we can use Foursquare API to collect relevant information about all venues of the selected streets.

For each street, we have the name, latitude and longitude.

Thanks to Foursquare API, we will have for each streets, the top-N venues.

We will focus on the client's criteria. So, we will focus on the following venues:

- Coffee shops
- French Restaurants
- Hotels
- Pubs

The streets that will be selected to answer this business need, will have as much as possible from those criteria.

3. METHODOLOGY

From the data, we will analyze the venues described for each street.

We will collect the venues, display them on the map, check their frequencies and distribution over streets.

After that, we can have insights on which street to choose.

Then, we will perform clustering of the streets according to specific criteria provided by the client.

Finally, the relevant streets will be selected and provided to the client to make the final decision.

The tools and methods used in this project include (but are no limited to):

Scraping: Web scraping allows accurately and quickly extract data from a Web page.

Foursquare API Places: Enable location discovery, venue search, and more. Access firmographic and rich, user-generated content such as photos, ratings, and reviews.

K-means Clustering: It is a method that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.

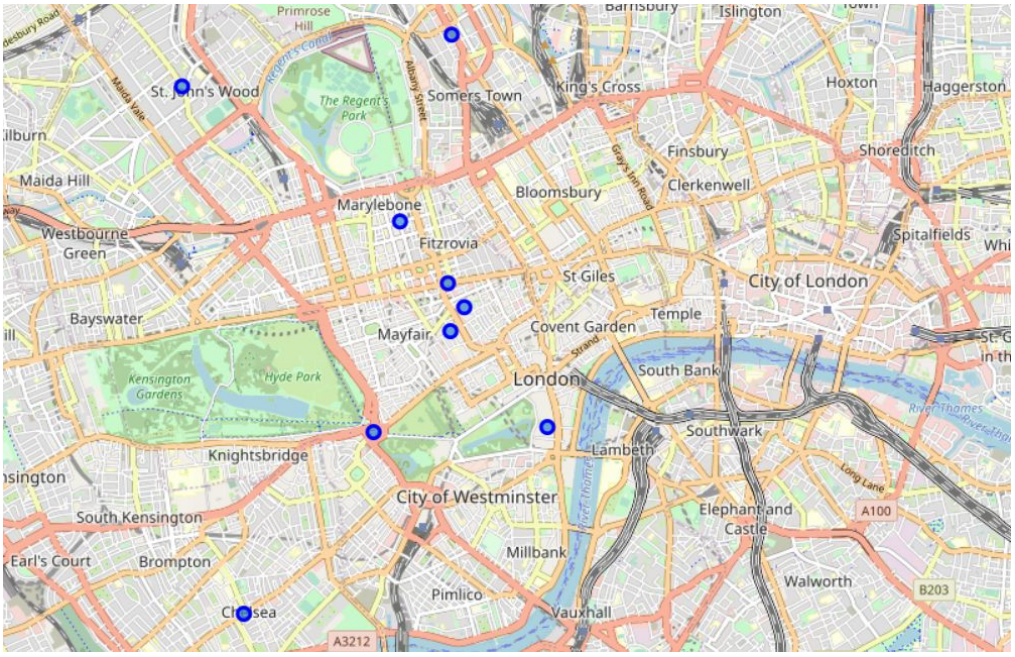
4. RESULTS

The data are collected from the following link:
https://www.latlong.net/category/streets-235-17.html.

Python and BeautifulSoup were used to scrape the page and present the data in a dataframe containing the street name, latitude and longitude.

	street	longitude	latitude
0	Harley Street, London, the UK	51.520599	-0.147700
1	Savile Row, London, the UK	51.511284	-0.140709
2	Antrim Coast Road, Northern Ireland, the UK	54.681999	-5.880000
3	Camden High Street, London, the UK	51.536388	-0.140556
4	Hyde Park Corner, London, the UK	51.502777	-0.151250

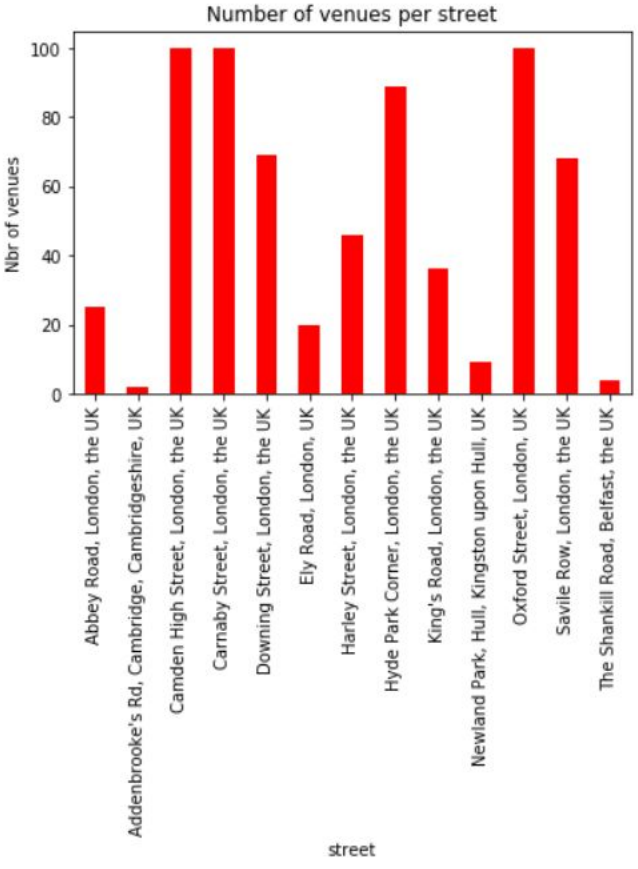
Folium was used to display London city and the streets collected:

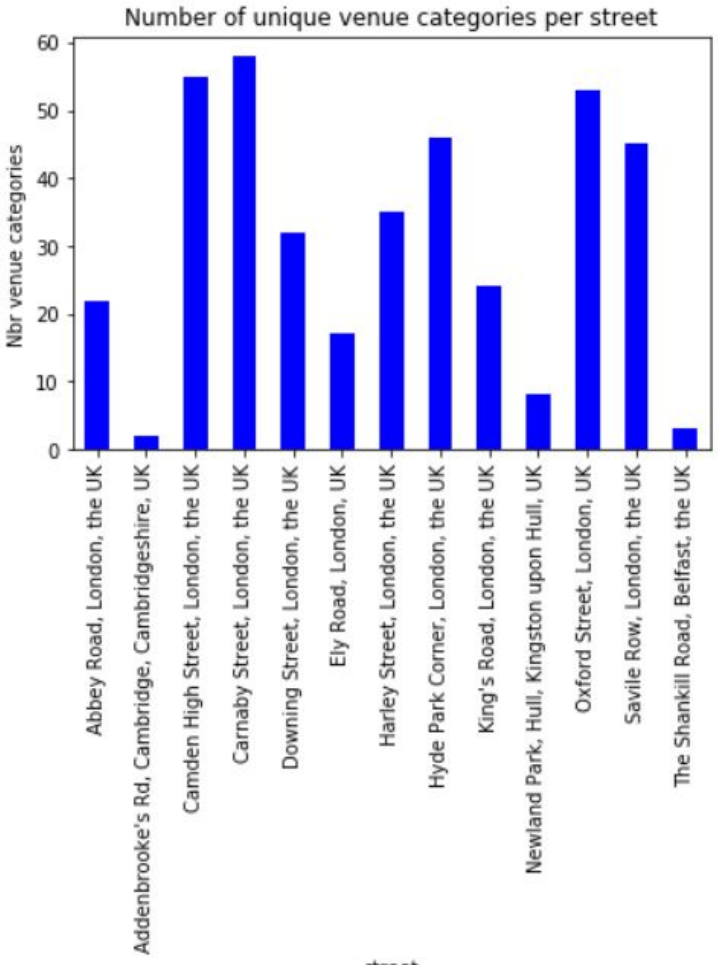


Using Foursquare API, the venues for each street are collected and added to the dataset:

	street	latitude	longitude	venue	venue latitude	venue longitude	venue category
0	Harley Street, London, the UK	-0.147700	51.520599	Royal Institute of British Architects	51.521108	-0.145014	Art Gallery
1	Harley Street, London, the UK	-0.147700	51.520599	Daunt Books	51.520433	-0.151824	Bookstore
2	Harley Street, London, the UK	-0.147700	51.520599	Pierre Marcolini Chocolatier	51.520653	-0.151909	Dessert Shop
3	Harley Street, London, the UK	-0.147700	51.520599	La Fromagerie	51.519858	-0.152238	Cheese Shop
4	Harley Street, London, the UK	-0.147700	51.520599	Jardin Du Jasmin	51.521534	-0.143434	Café
...	...	...	...	...	...	...	...
663	Oxford Street, London, UK	-0.141099	51.515419	Waitrose & Partners JL Foodhall	51.515093	-0.145093	Supermarket
664	Oxford Street, London, UK	-0.141099	51.515419	Disrepute	51.512121	-0.138179	Cocktail Bar
665	Oxford Street, London, UK	-0.141099	51.515419	Beast	51.515356	-0.146896	Steakhouse
666	Oxford Street, London, UK	-0.141099	51.515419	Fresh Healthy Eating Cafe	51.513228	-0.136394	Juice Bar
667	Oxford Street, London, UK	-0.141099	51.515419	Sanderson Hotel	51.517422	-0.137273	Hotel

Moreover, the number of venues and unique venues' categories are shown in bar charts:

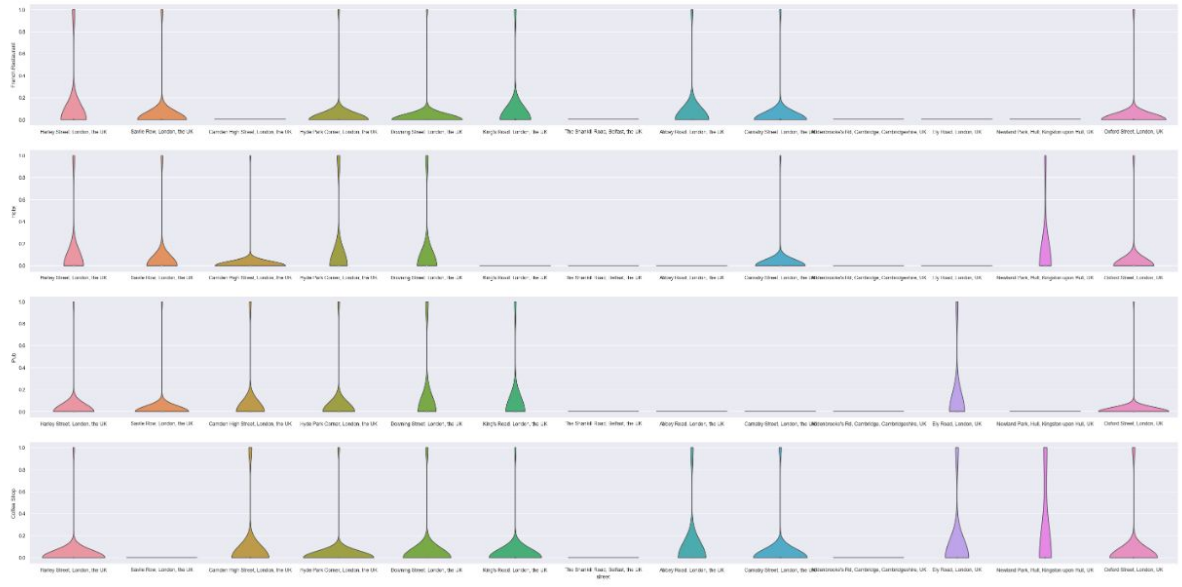




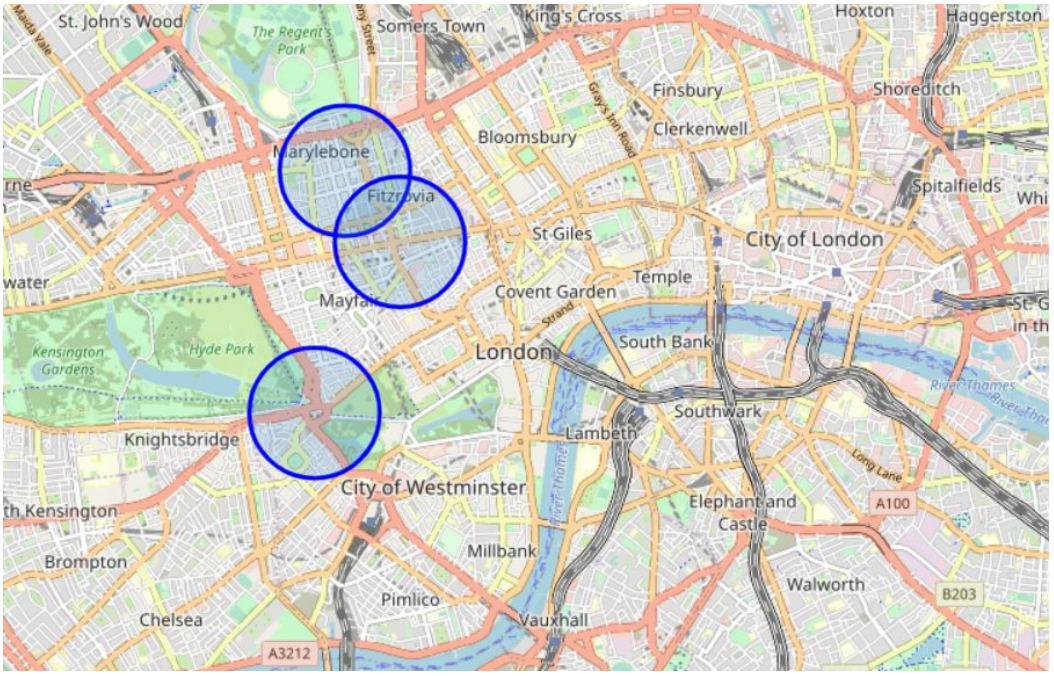
After that, the top-3 most common venues were selected for each street as well as their frequency distribution:

	street	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Abbey Road, London, the UK	Cricket Ground	Coffee Shop	Yoga Studio
1	Addenbrooke's Rd, Cambridge, Cambridgeshire, UK	Playground	Campground	Yoga Studio
2	Camden High Street, London, the UK	Coffee Shop	Pub	Greek Restaurant
3	Carnaby Street, London, the UK	Clothing Store	Boutique	Coffee Shop
4	Downing Street, London, the UK	Pub	Hotel	Monument / Landmark
5	Ely Road, London, UK	Supermarket	Coffee Shop	Pub
6	Harley Street, London, the UK	French Restaurant	Hotel	Indian Restaurant
7	Hyde Park Corner, London, the UK	Hotel	Lounge	Italian Restaurant
8	King's Road, London, the UK	Bakery	Pub	English Restaurant
9	Newland Park, Hull, Kingston upon Hull, UK	Coffee Shop	Library	Bookstore
10	Oxford Street, London, UK	Clothing Store	Coffee Shop	Italian Restaurant
11	Savile Row, London, the UK	Clothing Store	Art Gallery	Indian Restaurant
12	The Shankill Road, Belfast, the UK	Grocery Store	Garden	Sandwich Place

Frequency distribution for the top 3 venue categories for each street (including coffee shops)



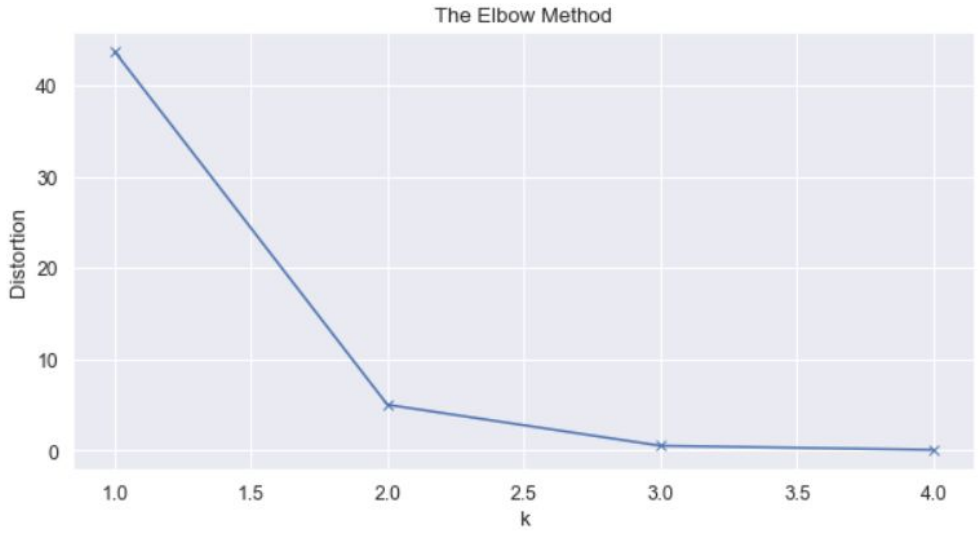
From this preliminary analysis, three streets were selected and displayed in a map:



To ensure the choice of the streets selected, I performed clustering based on K-means. To do so, the data were first prepared:

	street	longitude	latitude	Nbr Pub	Nbr Hotel	Nbr French Restaurant	Nbr Coffee Shop
0	Harley Street, London, the UK	51.520599	-0.147700	0.021739	0.086957	0.086957	0.021739
1	Savile Row, London, the UK	51.511284	-0.140709	0.014706	0.044118	0.029412	0.000000
2	Camden High Street, London, the UK	51.536388	-0.140556	0.060000	0.010000	0.000000	0.080000
3	Hyde Park Corner, London, the UK	51.502777	-0.151250	0.044944	0.134831	0.022472	0.022472
4	Downing Street, London, the UK	51.503223	-0.127500	0.115942	0.101449	0.014493	0.043478
5	King's Road, London, the UK	51.487370	-0.168874	0.083333	0.000000	0.055556	0.027778
6	The Shankill Road, Belfast, the UK	54.604301	-5.953000	0.000000	0.000000	0.000000	0.000000
7	Abbey Road, London, the UK	51.532005	-0.177331	0.000000	0.000000	0.040000	0.080000
8	Carnaby Street, London, the UK	51.513332	-0.138889	0.000000	0.020000	0.030000	0.040000
9	Addenbrooke's Rd, Cambridge, Cambridgeshire, UK	52.166824	0.125377	0.000000	0.000000	0.000000	0.000000
10	Ely Road, London, UK	51.573174	-0.007290	0.100000	0.000000	0.000000	0.100000
11	Newland Park, Hull, Kingston upon Hull, UK	53.767735	-0.368377	0.000000	0.111111	0.000000	0.222222
12	Oxford Street, London, UK	51.515419	-0.141099	0.010000	0.030000	0.020000	0.050000

Then, the number of k clusters was selected using the Elbow methodology which is a heuristic used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of clusters, and picking the elbow of the curve as the number of clusters to use.



From the results of Elbow chart, k = 2 was chosen to get the streets' clusters:

	street	longitude	latitude	Nbr Pub	Nbr Hotel	Nbr French Restaurant	Nbr Coffee Shop	cluster
0	9	51.520599	-0.147700	0.021739	0.086957	0.086957	0.021739	0
1	3	51.511284	-0.140709	0.014706	0.044118	0.029412	0.000000	1
2	5	51.536388	-0.140556	0.060000	0.010000	0.000000	0.080000	1
3	10	51.502777	-0.151250	0.044944	0.134831	0.022472	0.022472	0
4	7	51.503223	-0.127500	0.115942	0.101449	0.014493	0.043478	0
5	11	51.487370	-0.168874	0.083333	0.000000	0.055556	0.027778	0
6	4	54.604301	-5.953000	0.000000	0.000000	0.000000	0.000000	1
7	0	51.532005	-0.177331	0.000000	0.000000	0.040000	0.080000	1
8	6	51.513332	-0.138889	0.000000	0.020000	0.030000	0.040000	0
9	1	52.166824	0.125377	0.000000	0.000000	0.000000	0.000000	1
10	8	51.573174	-0.007290	0.100000	0.000000	0.000000	0.100000	0
11	12	53.767735	-0.368377	0.000000	0.111111	0.000000	0.222222	0
12	2	51.515419	-0.141099	0.010000	0.030000	0.020000	0.050000	1

5. Discussion and Conclusion

These results support our choice of two out of the three streets selected previously based only on first analysis of the data without clustering. Consequently, the final choice included the two streets: Hyde Park Corner and Harley Street.

Another thing that should be taking into consideration is the presence or not of coffee shops around the streets selected.

Based on that, the selection is narrowed to the Harley Street where only one coffee shop is present and there are the most important number of French Restaurants, or Hyde Park Corner where there are only 2 coffee shops and 2 French restaurants but 12 hotels and 4 pubs.

	street_x	longitude	latitude_x	Nbr Pub	Nbr Hotel	Nbr French Restaurant	Nbr Coffee Shop	cluster	street_y
0	6	51.520599	-0.147700	0.021739	0.086957	0.086957	0.021739	0	Harley Street, London, the UK
2	4	51.503223	-0.127500	0.115942	0.101449	0.014493	0.043478	0	Downing Street, London, the UK
4	3	51.513332	-0.138889	0.000000	0.020000	0.030000	0.040000	0	Carnaby Street, London, the UK