Introduction

London is the capital and largest city of England and the United Kingdom, with the largest municipal population in the European Union, London is considered to be one of the world's most important global cities 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, thus we want to study the idea of starting a business

1) Discussion and Background of the Business Problem:

Problem Statement: prospects of a restaurant in a London:

Opening a new restaurant can be intimidating hard. There are so many things and aspects to consider, for example: from choosing the right location to finding financing to selecting the right name, and the category of food served, in addition to the group of people it will target, in this notebook will discuss the trending kind of venues in each neighborhood of London city and comparing these neighborhoods to together in terms of venues popularity

We will go through each step of this project and address them separately

I will first outline the initial data preparation and describe future steps to start the battle of neighborhoods in London

Targeted audience:

- Business personnel who wants to invest or open a restaurant in London. This analysis will be a comprehensive guide to start or expand restaurant
- Data Scientists, who want to implement some of the most used exploratory Data Analysis techniques to obtain necessary data, analyze it, and, finally be able to tell a story out of it

2) Data:

For this project, we need the following data:

- Data containing the Neighborhoods,Boroughs,and Post Codes of London
    - Data source : https://en.wikipedia.org/wiki/List_of_areas_of_London
    - The idea to scrape the table using Beautiful soup and writing into a csv file

- Dataset of the latitude and longitude of each neighborhood
  - Data source: obtaining the data using geolocator
  - Data source: https://www.freemaptools.com/download/outcode-postcodes/postcode-outcodes.csv.
- Data containg the venues of each neighborhood:
  - Data source : Foursquare API calls
  - Description: By using this API we will get all the venues in each neighborhood. We can filter these venues to get the common restaurants and venues types.

3) Approach:

- Scraping the neighborhoods of London from Wikipedia using beautiful soup.
- Obtaining the location of each neighborhood using geolocator.
- Plotting the neighborhoods using folium.
- Obtaining the venues of each neighborhood using FoursquareAPI
- finding the trending venues and trending restaurants type and which neighborhoods has more restaurants
- Using machine learning to cluster the neighborhoods to find the which neighborhoods are similar

4) Implementation :

First by scraping the Following page, we get this data frame:

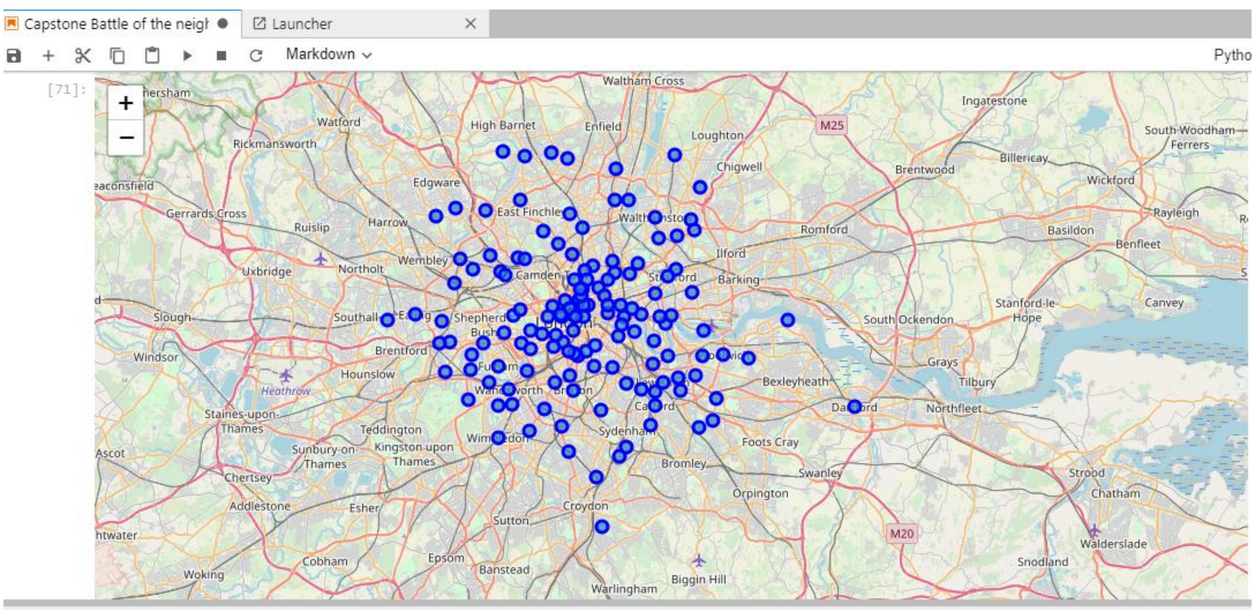
[4]:

	Neighborhood	Borough	Post_town	postcode	Dial Code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [2]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[3]	LONDON	W3, W4	020	TQ205805
6	Aldgate	City[5]	LONDON	EC3	020	TQ334813
7	Aldwych	Westminster[5]	LONDON	WC2	020	TQ307810
9	Anerley	Bromley[6]	LONDON	SE20	020	TQ345695

Now the next step would be to get the longitude and latitude of each neighborhood using GeoLocator:

	Neighborhood	Borough	Post_town	postcode	Dial Code	OS grid ref	latitude	longitude
0	Acton	Ealing, Hammersmith and Fulham[3]	LONDON	W3, W4	020	TQ205805	51.508140	-0.273261
1	Aldgate	City[5]	LONDON	EC3	020	TQ334813	51.514248	-0.075719
2	Aldwych	Westminster[5]	LONDON	WC2	020	TQ307810	51.512625	-0.118568
3	Anerley	Bromley[6]	LONDON	SE20	020	TQ345695	51.407599	-0.061939
4	Angel	Islington[3]	LONDON	EC1, N1	020	TQ345665	51.531946	-0.106106

After that we Plot the Map of London that shows every neighborhood:



Next we use the Foursquare API to get the venues of each neighborhood in addition to their category and location so that we can plot them later and the data frame looks like this :

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acton	51.50814	-0.273261	London Star Hotel	51.509624	-0.272456	Hotel
1	Acton	51.50814	-0.273261	The Aeronaut	51.508376	-0.275216	Pub
2	Acton	51.50814	-0.273261	Dragonfly Brewery at George & Dragon	51.507378	-0.271702	Brewery
3	Acton	51.50814	-0.273261	Bake Me	51.508452	-0.268543	Creperie
4	Acton	51.50814	-0.273261	Everyone Active	51.506608	-0.266878	Gym / Fitness Center

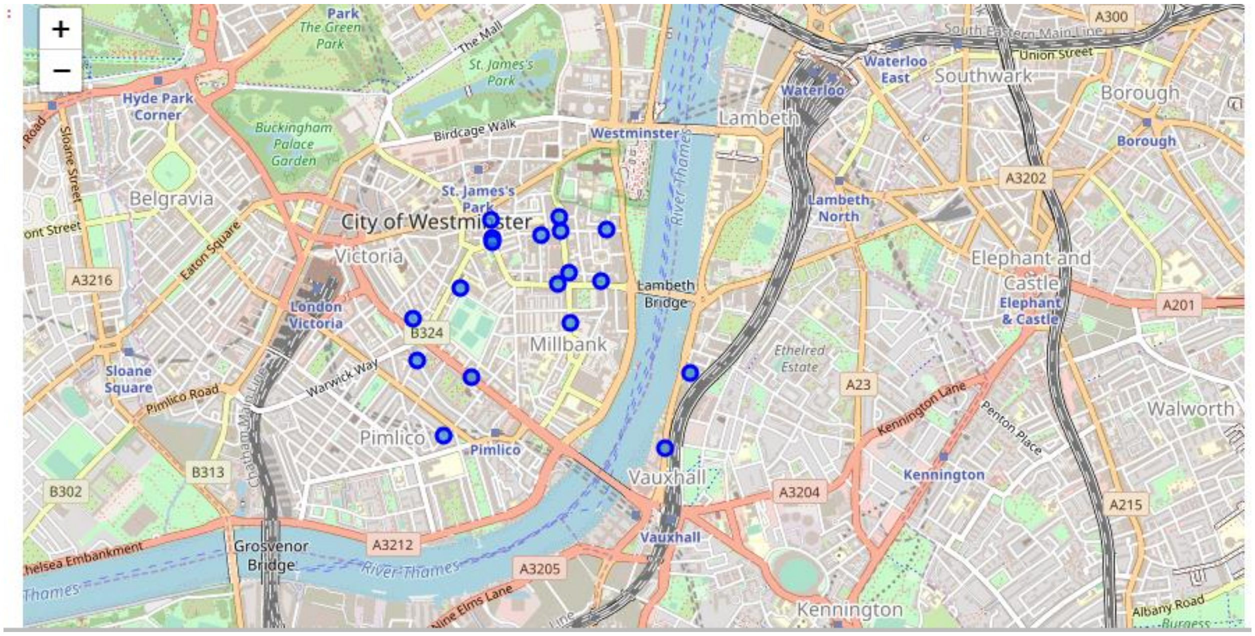
The data frame has 376 unique venue category:

Now we create a data frame that contains only the restaurants:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
9	Acton	51.508140	-0.273261	Amigo's Peri Peri	51.508396	-0.274561	Fast Food Restaurant
12	Acton	51.508140	-0.273261	Ting Tong Thai	51.508363	-0.277755	Thai Restaurant
13	Acton	51.508140	-0.273261	North China Restaurant	51.508251	-0.277435	Chinese Restaurant
18	Acton	51.508140	-0.273261	Persian Nights	51.508529	-0.282383	Middle Eastern Restaurant
37	Aldgate	51.514248	-0.075719	The Japanese Canteen	51.513775	-0.079079	Japanese Restaurant

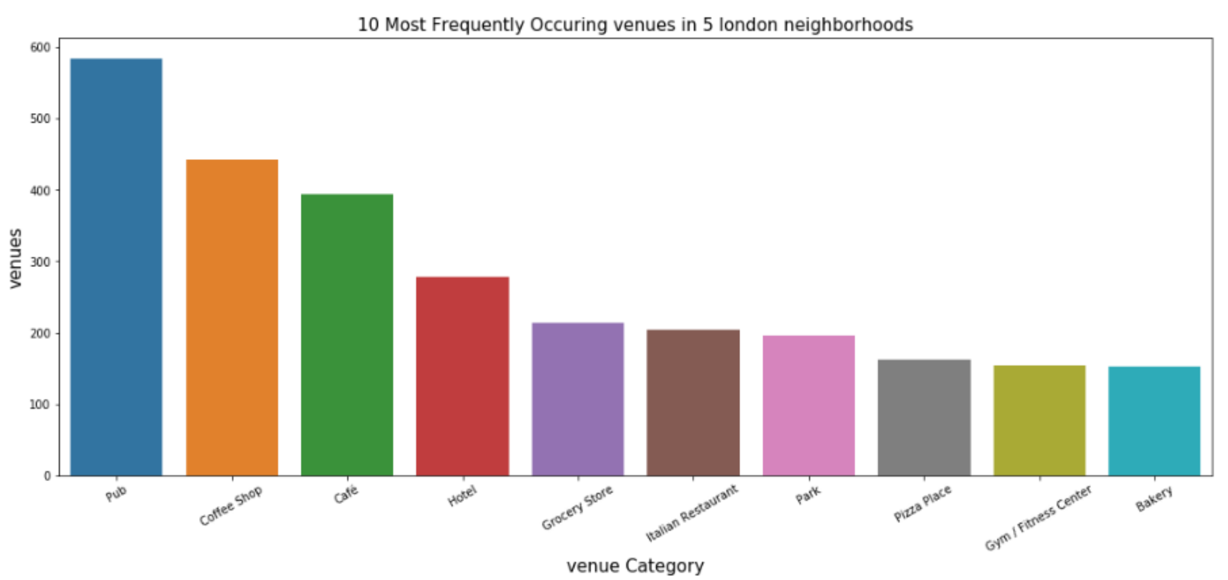
visualize restaurants in each neighborhood

next we visulisze the restaurants for example in Millbank neighborhood:



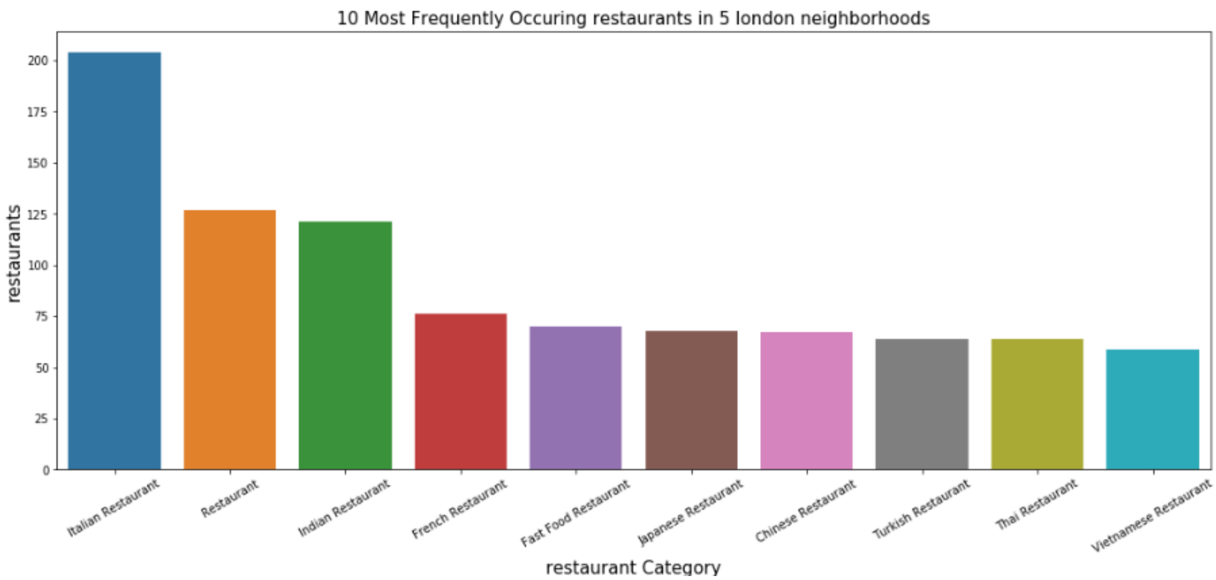
Neighborhood Analysis:

aCreate a Data-frame with the 10 Most Frequently Occuring Venue_Category:



as seen from the bar chart Pups, Coffee Shops, Cafes and Hotels are the most common Venues

let's see which restaurant category is most frequent :



It appears that Italian Resturants and indian Restaurants are the most common type of restaurants ,next we are going to check the top 5 in each neighborhood

The result will appera in a data frame that looks like this :

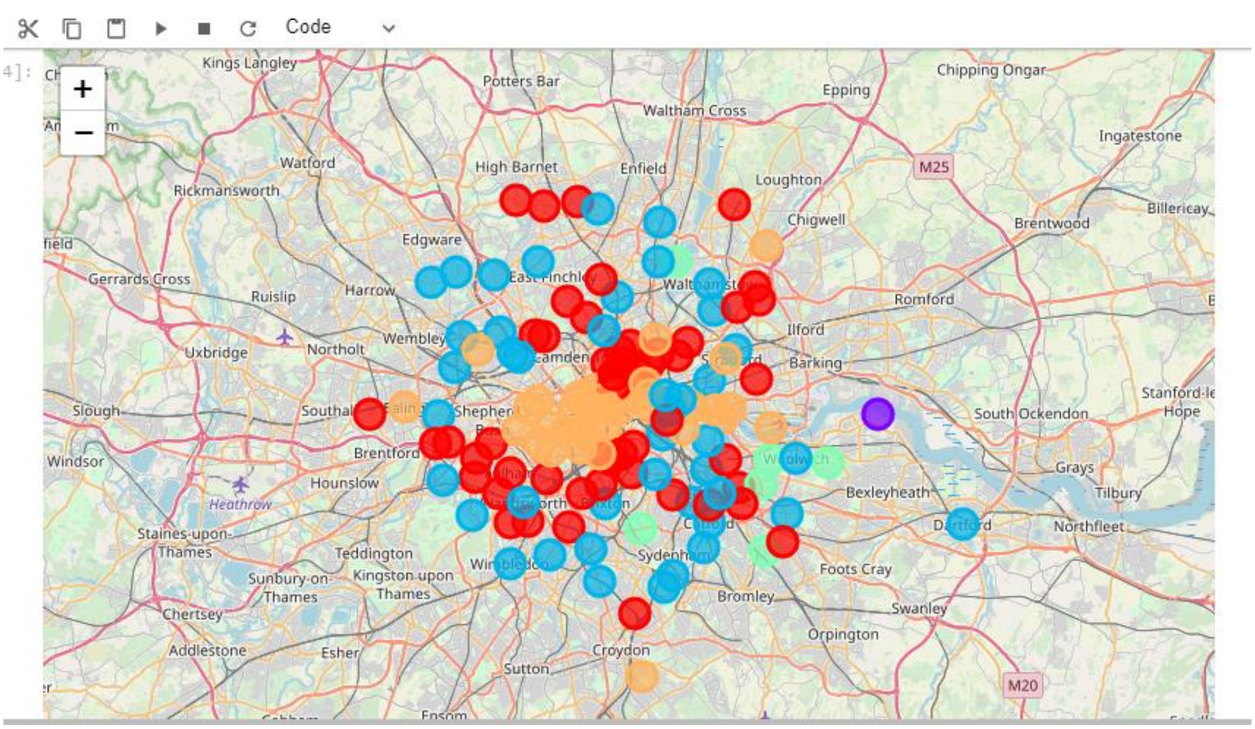
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Acton	Pub	Gym / Fitness Center	Grocery Store	Hotel	Coffee Shop
1	Aldgate	Coffee Shop	Hotel	Cocktail Bar	Salad Place	Gym / Fitness Center
2	Aldwych	Theater	Burger Joint	Coffee Shop	Dessert Shop	Tea Room
3	Anerley	Grocery Store	Train Station	Convenience Store	Hardware Store	Gastropub
4	Angel	Pub	Coffee Shop	Café	Mediterranean Restaurant	Pizza Place

Then we check the neighborhood which have an italian restaurant as most common restaurant:

	Neighborhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant
5	Archway	Italian Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Kebab Restaurant	Asian Restaurant
7	Barbican	Italian Restaurant	Vietnamese Restaurant	Sushi Restaurant	French Restaurant	English Restaurant
8	Barnes	Italian Restaurant	Thai Restaurant	Restaurant	Yakitori Restaurant	Greek Restaurant
10	Battersea	Italian Restaurant	Thai Restaurant	Chinese Restaurant	Portuguese Restaurant	Restaurant
12	Belgravia	Italian Restaurant	Restaurant	Indian Restaurant	French Restaurant	Seafood Restaurant
15	Blackfriars	Italian Restaurant	Restaurant	Falafel Restaurant	Seafood Restaurant	Japanese Restaurant
18	Bloomsbury	Italian Restaurant	Japanese Restaurant	Falafel Restaurant	Turkish Restaurant	Okonomiyaki Restaurant
22	Brompton	Italian Restaurant	Japanese Restaurant	French Restaurant	Tapas Restaurant	Mediterranean Restaurant
26	Castelnau	Italian Restaurant	French Restaurant	Indian Restaurant	Chinese Restaurant	Yakitori Restaurant
27	Catford	Italian Restaurant	Greek Restaurant	Japanese Restaurant	Portuguese Restaurant	Fast Food Restaurant
29	Chelsea	Italian Restaurant	English Restaurant	Japanese Restaurant	French Restaurant	Restaurant
30	Chinatown	Italian Restaurant	Seafood Restaurant	Indian Restaurant	Tapas Restaurant	Restaurant
31	Chingford	Italian Restaurant	Yakitori Restaurant	Grilled Meat Restaurant	Falafel Restaurant	Fast Food Restaurant

Clutsering Neighborhoods :

Using Kmean clustering on neighborhoods we get the following map :



Then we explore each cluster at a time :

Frist cluster :

```
london_merged.loc[london_merged['Cluster Labels'] ==0, london_merged.columns[[0] + list(range(5, london_merged.shape[1]))]]
```

	Neighborhood	OS grid ref	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Angel	TQ345665	51.531946	-0.106106	0	Pub	Coffee Shop	Café	Mediterranean Restaurant	Pizza Place
5	Archway	TQ285875	51.565437	-0.134998	0	Coffee Shop	Pub	Grocery Store	Pizza Place	Café
6	Balham	TQ285735	51.442828	-0.151443	0	Coffee Shop	Pub	Pizza Place	Bakery	Indian Restaurant
9	Barnes	TQ225765	51.471896	-0.238744	0	Park	Pub	Food & Drink Shop	Platform	Track
10	Barnsbury	TQ305845	51.538935	-0.114735	0	Pub	Café	Gastropub	Park	Grocery Store
11	Battersea	TQ285765	51.470793	-0.172214	0	Pub	Bakery	Bus Stop	Hotel	Italian Restaurant
17	Blackheath	TQ395765	51.466318	0.008562	0	Pub	Café	Bakery	Pizza Place	Indian

Second cluster:

```
london_merged.loc[london_merged['Cluster Labels'] ==1, london_merged.columns[[0] + list(range(5, london_merged.shape[1]))]]
```

	Neighborhood	OS grid ref	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
39	Crossness	TQ480800	51.509037	0.138411	1	Mobile Phone Shop	History Museum	Ethiopian Restaurant	Event Space	Exhibit

Third cluster

```
london_merged.loc[london_merged['Cluster Labels'] ==2, london_merged.columns[[0] + list(range(5, london_merged.shape[1]))]]
```

	Neighborhood	OS grid ref	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Acton	TQ205805	51.508140	-0.273261	2	Pub	Gym / Fitness Center	Grocery Store	Hotel	Coffee Shop
3	Anerley	TQ345695	51.407599	-0.061939	2	Grocery Store	Train Station	Convenience Store	Hardware Store	Gastropub
14	Bellingham	TQ375715	51.431081	-0.024515	2	Grocery Store	Café	Clothing Store	Gym	Supermarket
15	Bermondsey	TQ335795	51.497012	-0.063268	2	Brewery	Grocery Store	Pub	Park	Café
20	Bow	TQ365825	51.528309	-0.019482	2	Pub	Grocery Store	Metro Station	Bus Stop	Art Gallery
21	Brixton	TQ315755	51.456804	-0.116796	2	Coffee Shop	Caribbean Restaurant	Pub	Grocery Store	Vegetarian / Vegan Restaurant
22	Brockley	TQ365745	51.457833	-0.036087	2	Grocery Store	Convenience Store	Coffee Shop	Beer Store	Fish & Chips Shop
24	Brondesbury	TQ245845	51.544990	-0.202791	2	Pub	Indian Restaurant	Coffee Shop	Park	Grocery Store

And soo on..

5) Results:

The results of the exploratory data analysis and clustering are summarized below:

- Pups and coffee shops are on of top the charts of most common venues in the all neighborhoods
- Italian and Indian restaurants are the most common restaurants if the neighborhoods of London
- As the clustering is based on the venues of each neighborhood we can see that the first cluster is the largest
- Restaurant location can be chosen by finding the neighborhood which have a successful restaurant as common venue then see which cluster it belongs to and explore the cluster to choose the appropriate location

6)Discussion:

Recommending a restaurant location was discussed by only one feature which is how common the restaurant and its type , this feature alone is not enough because there are other features that can by analyzed such as foot and car traffic and the census of the area in addition to the kind of people inhabiting the area .

This project can be expanded and developed by investigating the more about the features and using different machine learning clustering methods or even enhance

The we k-mean clustering works by finding the write number of clusters

This project is could be just a start to a bigger more realistic study.

7) Conclusion :

Finally, to conclude this project, We have got a small glimpse of how real life data-science projects look like. I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the neighborhoods of London and saw the results of segmentation of districts using Folium leaflet map. Potential for this kind of analysis in a real life business problem is discussed in detail. In addition, some of the drawbacks and chance for improvements to represent pictures that are even more realistic are mentioned.