Report Coursera capstone

By : Mounir Ourekouch

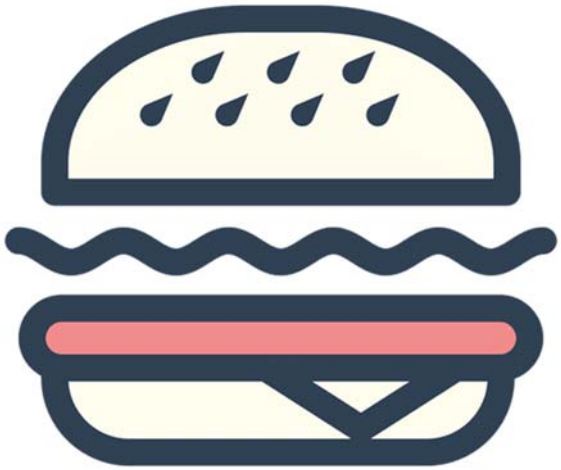
Software Engineer Student, ENSIAS ,Morocco

1. Introduction/Business Problem :

A contractor want to start a new business in fast food in London. Unfortunately, he has no idea about the right area for this project.

Therefore, he decided to rely on the science of data analysis in order to find the appropriate area for this new project,

Especially the population density in various neighborhoods of London, as well as the distribution of different venues and facilities in the city of London.



2. Data Section :

In order to solve our business problem, we will use two datasets:

- **List_of_London_boroughs :**

  This dataset is constructed by scraping the Wikipedia page : "https://en.wikipedia.org/wiki/List_of_London_boroughs", the final dataset

will help us to get boroughs Coordinates and also the populationof each borough .

Figure 1 : London Boroughs

	Borough	Population	Latitude	Longitude
0	Barking and Dagenham	194352	51.567333	0.159333
1	Barnet	369088	51.630444	-0.152667
2	Bexley	236687	51.458000	0.150889
3	Brent	317264	51.564222	-0.290667
4	Bromley	317899	51.406222	0.021556

- **Foursquare location data :**

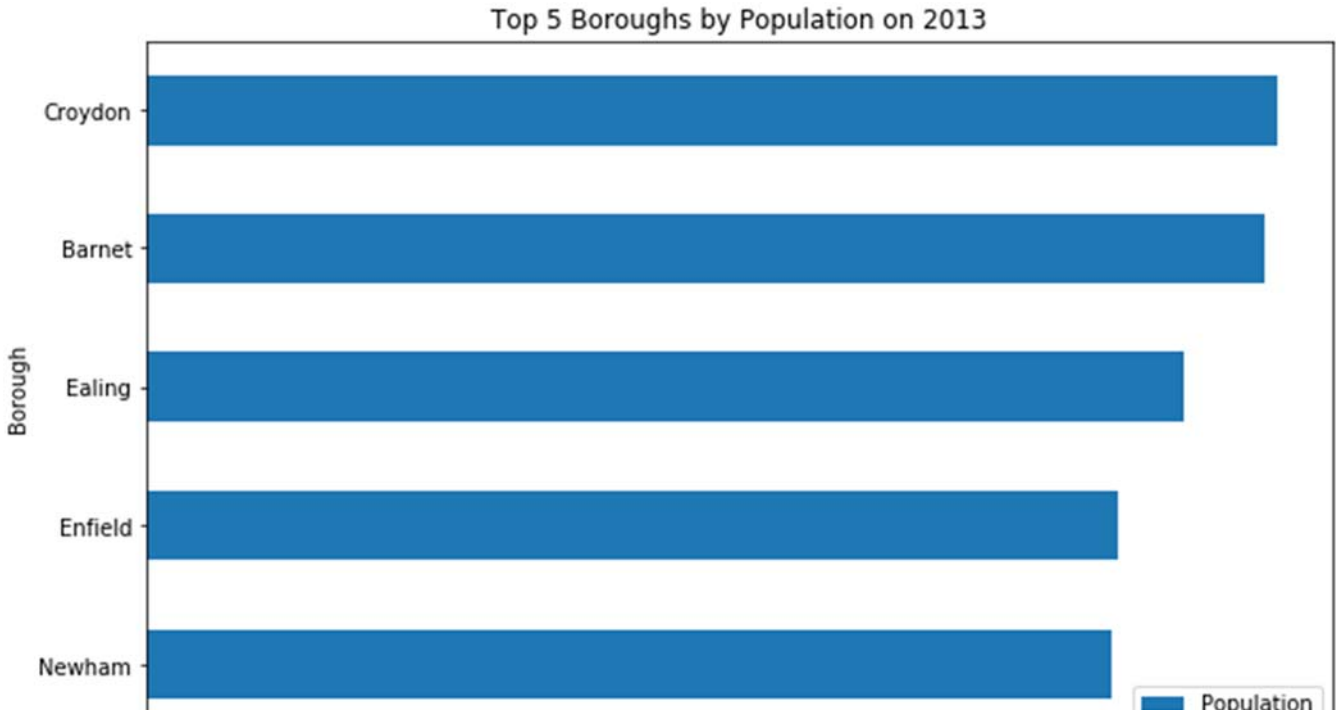
After getting information's about boroughs coordinates, we will use this coordinates to discover venues in each borough and cluster all boroughs according the most frequent venues in each borough that will help us to make the final decision.

3. Methodology Section :

In the project we will mainly use two methodologies in order to have enough information has to make a decision:

- Data visualization :

We will visualize first our data in order to get information about population and the top boroughs according to population:

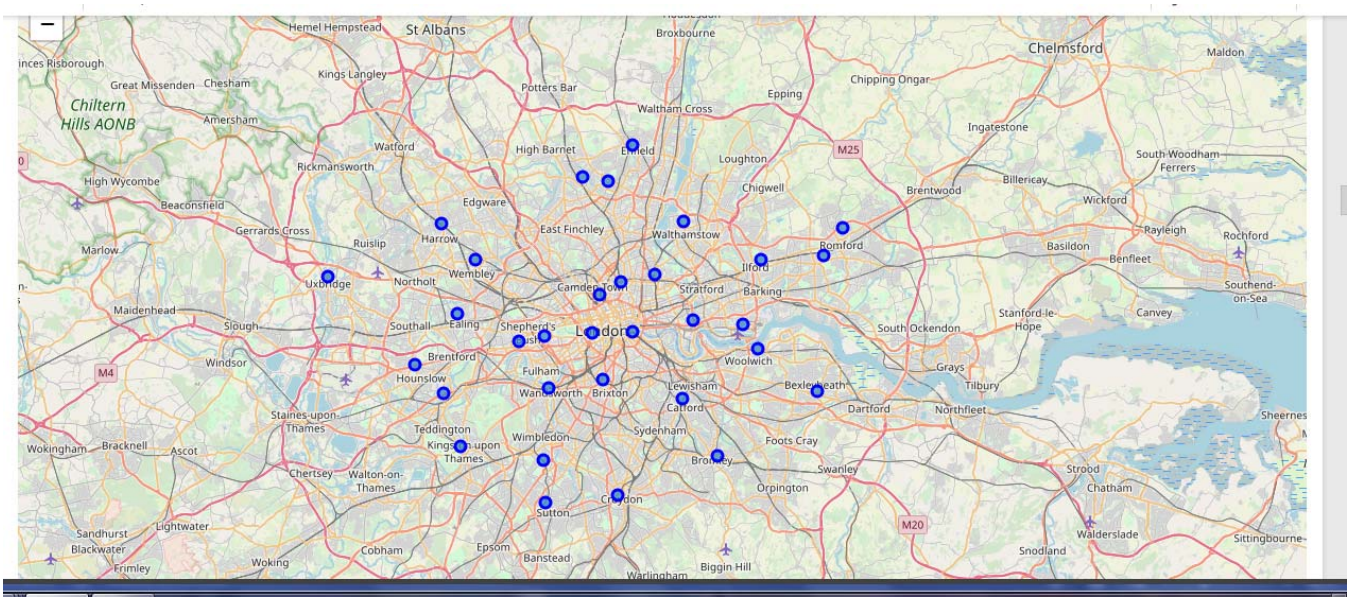


- Data clustering :

Because the first result is not enough the help us to make a decision and have a solution for our business problem , for that we need to use also foursquare api , folium in order to get more information about venues and make clusters for London boroughs

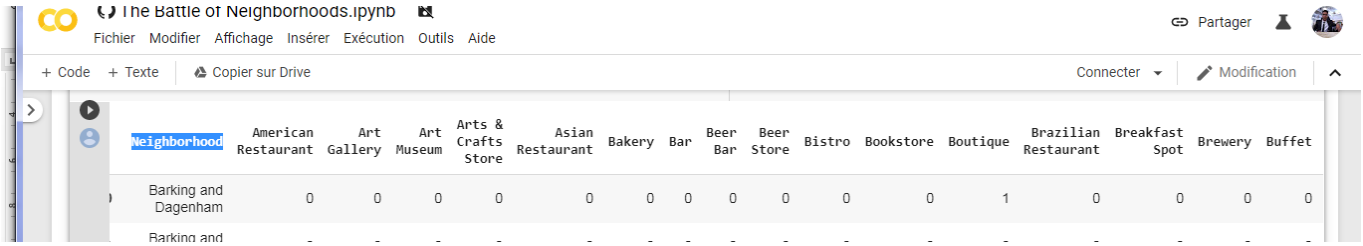
- Checking collected data about London boroughs in map :

Figure 2 : London



- Calculate the sum of venues for each borough or neighborhood and get the frequencies of venues in neighborhoods :

-



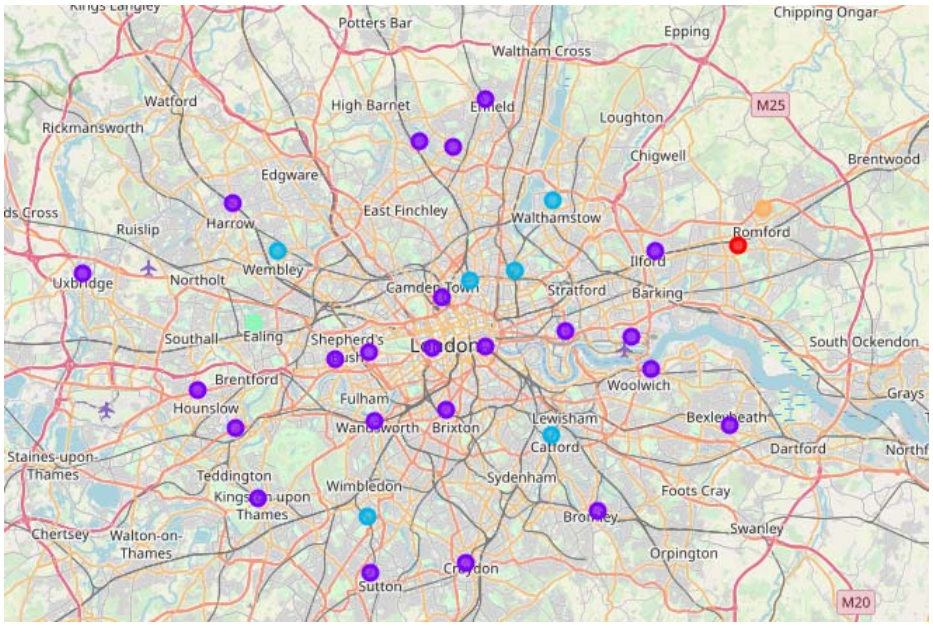
Neighborhood	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Bakery	Bar	Beer Bar	Beer Store	Bistro	Bookstore	Boutique	Brazilian Restaurant	Breakfast Spot	Brewery	Buffet
Barking and Dagenham	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Barking and																

```
[ ]     print('\n')

----Barking and Dagenham ----
          venue  freq
0  Soccer Field  0.50
1   Sports Club  0.25
2      Boutique  0.25
3   Opera House  0.00
4        Museum  0.00


----Barnet----
                venue  freq
0           Bus Stop  0.25
1                Pub  0.25
2      Event Service  0.25
3  Fish & Chips Shop  0.25
4  Outdoor Sculpture  0.00


----Bexley----
                  venue  freq
0                   Pub  0.11
1          Coffee Shop  0.11
2          Supermarket  0.07
3  Fast Food Restaurant  0.07
4    Italian Restaurant  0.07
```

- Get the top venues for each neighborhood :

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barking and Dagenham	Soccer Field	Sports Club	Boutique	French Restaurant	Food Stand
1	Barnet	Bus Stop	Fish & Chips Shop	Pub	Event Service	English Restaurant
2	Bexley	Coffee Shop	Pub	Supermarket	Clothing Store	Fast Food Restaurant
3	Brent	Café	Cosmetics Shop	Women's Store	Event Service	Food Stand
4	Bromley	Coffee Shop	Pizza Place	Pub	Clothing Store	Burger Joint

- London maps with clusters :



4. Results :

After the first visualization, we get that the first 5 boroughs according to population are:

- Croydon (372 K)
- Barnet (369 K)
- Ealing (342 K)
- Enfield (320 K)
- Newham (318 K)

After analyzing and clustering boroughs we get that :

----Croydon----

Venue freq:

- Coffee Shop 0.17
- Clothing Store 0.17
- Pub 0.10
- Bookstore 0.07
- Women's Store 0.03

----Barnet----

Venue freq:

- Bus Stop 0.25
- Pub 0.25
- Event Service 0.25
- Fish & Chips Shop 0.25
- Outdoor Sculpture 0.00

----Ealing----

Venue freq:

- Park 0.75
- Pharmacy 0.25
- American Restaurant 0.00
- Outdoor Sculpture 0.00
- Museum 0.00

----Enfield----

Venue freq

- Pub 0.5
- Indian Restaurant 0.1
- Grocery Store 0.1
- Coffee Shop 0.1
- Sandwich Place 0.1

----Newham----

Venue freq

- Light Rail Station 0.25
- Supermarket 0.25
- Gym / Fitness Center 0.12
- Bus Station 0.12
- Pub 0.12

So in order to make the final decision between those 5 boroughs who has the max of population we will add another score (Competitive rate) :

Competitive rate :

The sum of frequencies of venues in boroughs multicoated by +1 (if the venue will be benefit for our business for example: +1 for schools , stadium ,...) , by -1(if the venue make a danger for our business for example another restaurant)

Borough	rate
Croydon	-0.17+0.17+0.10+.07+0.03=**0.2**
Barnet	0.25+0.25+0.25+0.25=**1**
Ealing	0.75+0.25=**1**
Enfield	0.5-0.1+0.1-0.1-0.1=**0.3**
Newham	0.25+0.25+0.12+0.12+0.12=**0.86**

5. Discuss Section :

In the end, the best borough for starting a fast food restaurant in London with the big frequencies of population, and less competitive venues are:

1- **BARNET OR EALING**
2- NEWHAM

6. Conclusion :

Thanks to foursquare api and our collected data we had succeeded to get good results that helped us to make a decision and have a solution for our business problem.