Rapport

Coursera capstone

Realised by :

* Amine mirate
* Abdelhamid Belodi
* Mouad cherkaoui

**Summary** :

1. General definitions
2. Introduction
3. Data Section
4. List\_of\_London\_boroughs
5. Foursquare location data
6. Methodology Section
7. Data visualization
8. Data clustering
9. Checking collected data about London boroughs in map
10. Calculate the sum of venues for each borough or neighborhood and get the frequencies of venues in neighborhoods
11. Get the top venues for each neighborhood
12. Results
13. Discuss Section
14. Conclusion

1. General definition

**Deep Learning** : Deep learning is an [artificial intelligence](https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp) function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of [machine learning](https://www.investopedia.com/terms/m/machine-learning.asp) in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network

**Overfitting :** Overfitting is a modeling error that occurs when a function is too closely fit to a limited set of data points. Overfitting the model generally takes the form of making an overly complex model to explain idiosyncrasies in the data under study.

**Hyperparameter optimization :** In [machine learning](https://en.wikipedia.org/wiki/Machine_learning), hyperparameter optimization or tuning is the problem of choosing a set of optimal [hyperparameters](https://en.wikipedia.org/wiki/Hyperparameter_(machine_learning)) for a learning algorithm. A hyperparameter is a [parameter](https://en.wikipedia.org/wiki/Parameter) whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are learned.

2. Introduction

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.

3. Data Section

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

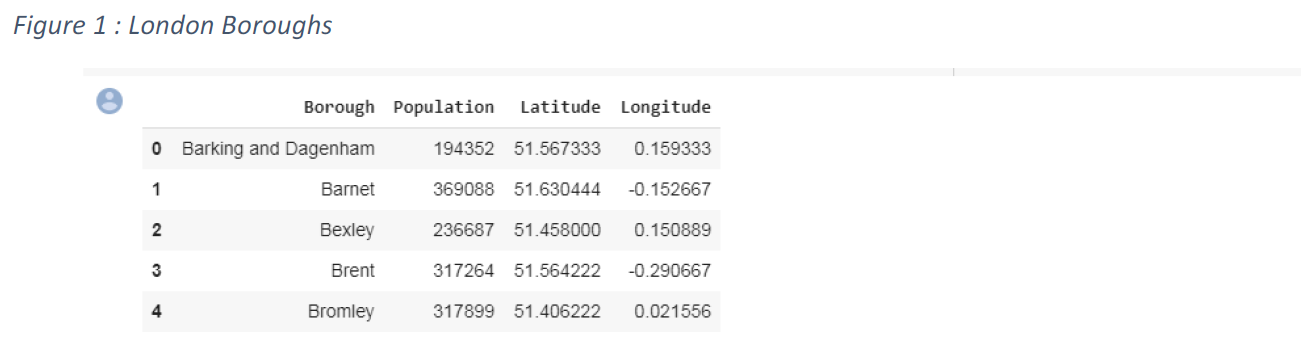
List\_of\_London\_boroughs

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 .



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.

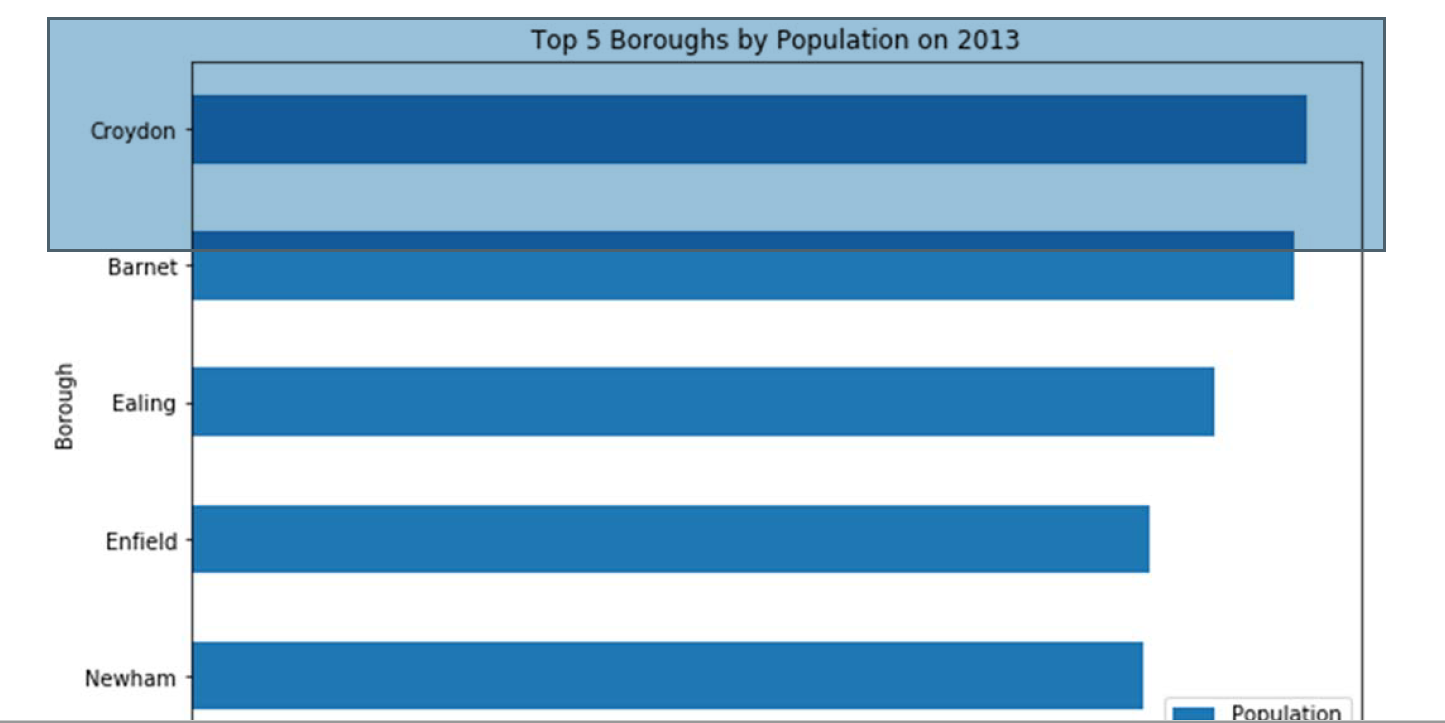
4. Methodology Section

In the project we will mainly use two methodologies in order to have enough

information has to make a decision:

Data visualization :

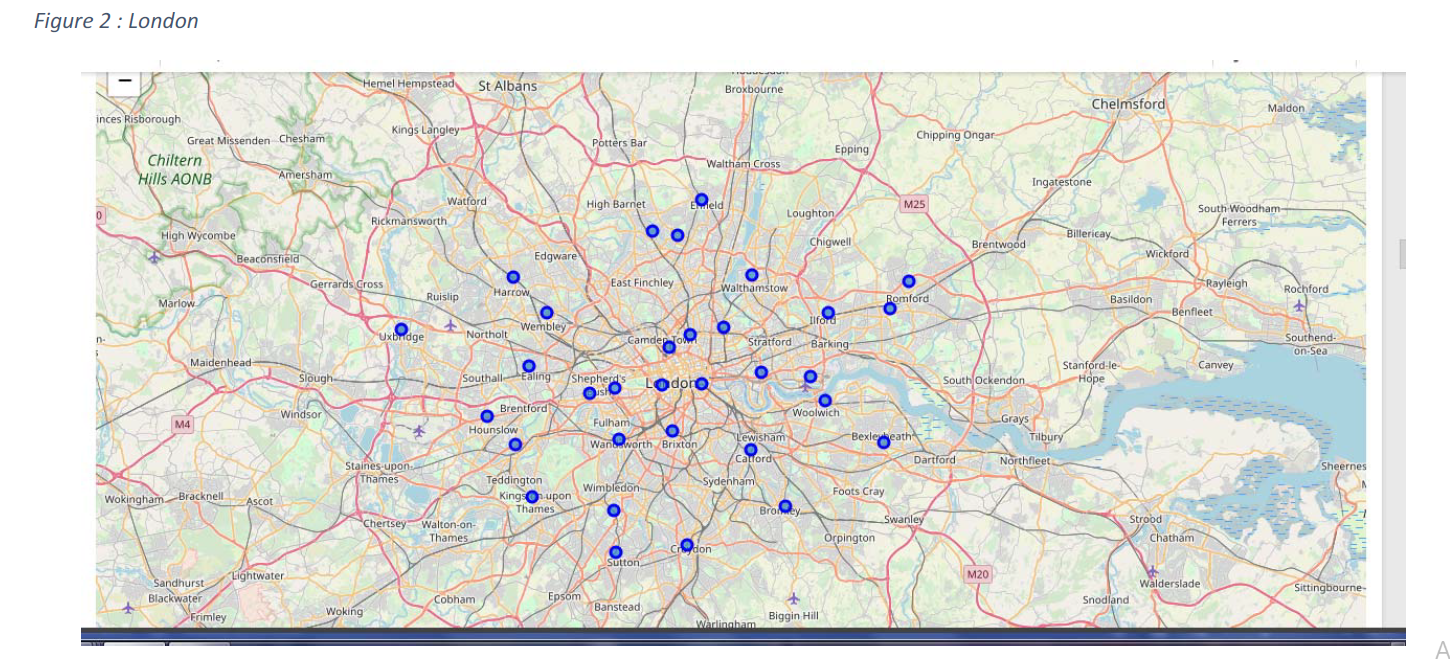
We will visualize first our data in order to get information about population and the top boroughs according to population:



Data clustering :

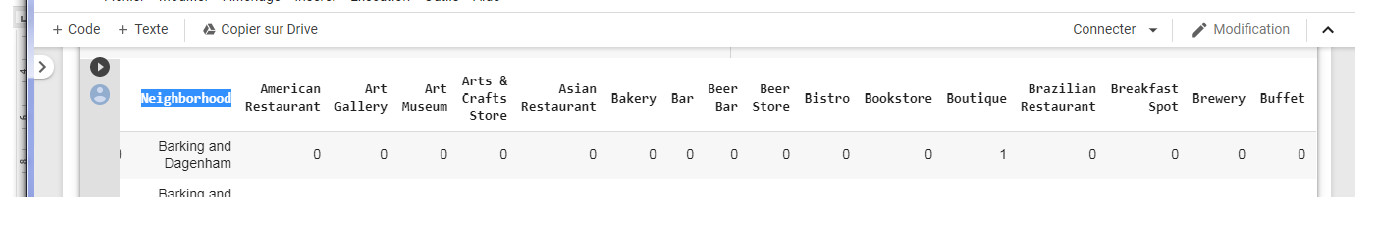
Because the first result is not enough the help us to make a decision and have a solution for our business problem , for that we need to use also foursquare api, folium in order to get more information about venues and make clusters for London boroughs

Checking collected data about London boroughs in map



Calculate the sum of venues for each borough or neighborhood and get

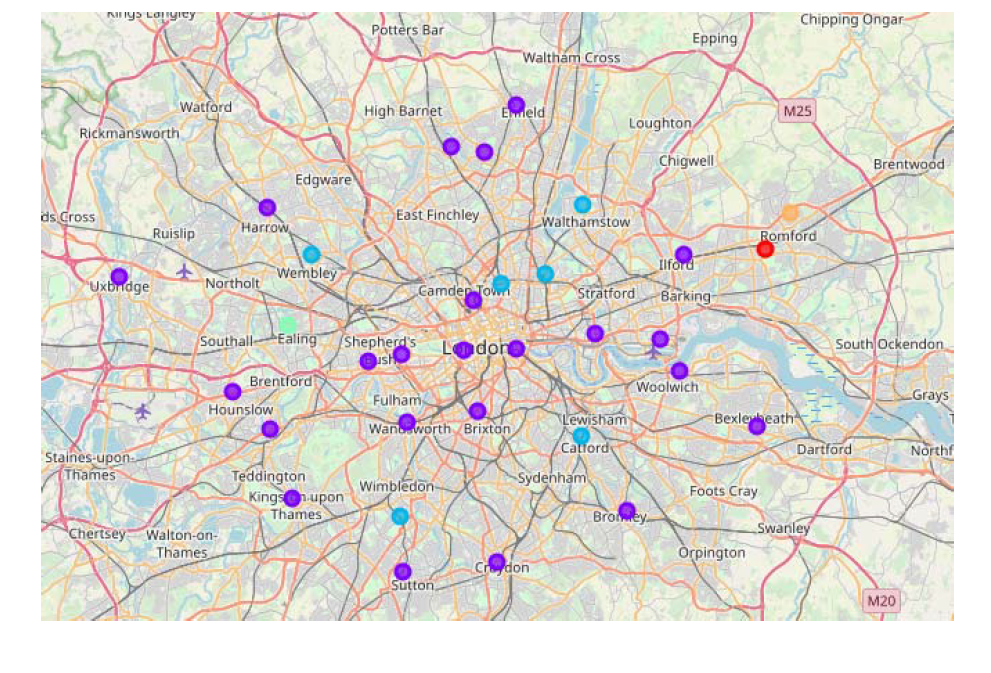
the frequencies of venues in neighborhoods





Get the top venues for each neighborhood



**London maps with clusters:**

5. 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)

6.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**

7.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.