Data analytics for opening a new Moroccan restaurant in Paris, France

Author :

Naoufal NAIM

Course :

Applied Data Science Capstone - IBM

Table of Contents

1. Introduction

Paris is the French capital and most interesting city of France. There are 2,138,551 residents in an area of 105,7 square kilometres (40,7 square miles) [1].

According to the Economist Intelligence Unit Worldwide Cost of Living Survey in 2019, Paris was the fifth most expensive city in the world, after Singapore, Osaka, Hong Kong, and New York, and ahead of Zürich, Tokyo, and Los Angeles [2].

Indeed, we can found a huge number of restaurants which represents all types of cuisines in Paris. As you see from the above figures, Paris is one of the most interesting city to open a Moroccan restaurant.

Concerning to the investors, we expect from them to prefer the districts in which there is a good community of the Arab's restaurant or Moroccan restaurant and the type of business they want to install is less intense.

Our problem is that there are several phases before proposing to the investors, for instance,

- to obtain the information of the most famous venues in Paris,
- to preprocessing the data (to clean, to normalize),
- to explore the preprocessed data,
- to cluster the districts based on the top 10 venues for each district.

When we consider all above problems, we could illustrate the data in an information map where the district of Paris, the Moroccan restaurant and its neighborhood are clustered according to the venue density.

The aim of this project is to find the best borough out of 20 boroughs to open a **Moroccan restaurant in Paris, France.**

The target audience of this report is anyone who is interested in opening a Moroccan restaurant but have no idea about the location.

2. Data Description

2.1. Data Requirements

The main districts in Paris are divided into 20 administrative districts, shortened to arrondissements.

The data regarding the districts in Paris needs to be researched and a suitable useable source identified. If it is found but is not in a useable form, data wrangling and cleaning will have to be performed.

The cleansed data will then be used alongside Foursquare data, which is readily available. Foursquare location data will be leveraged to explore or compare districts around Paris, identifying the ideal district to open a Moroccan restaurant in Paris.

Data for the arrondissements is necessary to select the most suitable of these areas for new Moroccan restaurant.

Initially looking to get this data by scraping the relevant Wikipedia page (https://en.wikipedia.org/wiki/Arrondissements_of_Paris), fortunately, after much research, this data is available on the web and can be manipulated and cleansed to provide a meaningful dataset to use.

Data from Open|DATA France:
https://opendata.paris.fr/explore/dataset/arrondissements/table/?dataChart

Also available from Opendatasoft:
https://data.opendatasoft.com/explore/dataset/arrondissements%40parisdata/export/

2.2. Data Workflow

- Outline the initial data that is required:
  - District data for Paris including names, location data if available, and any other details required.
- Obtain the Data:
  - Research and find suitable sources for the district data for Paris.
  - Access and explore the data to determine if it can be manipulated for our purposes.
- Initial Data Wrangling and Cleaning:
  - Clean the data and convert to a useable form as a dataframe.

3. Methodology and Exploratory Data Analysis

- Data Analysis and Location Data:
  - Foursquare location data will be leveraged to explore or compare districts around Paris.
  - Data manipulation and analysis to derive subsets of the initial data.
  - Explore, segment and using K-Means as part of this clustering study [3] to cluster the neighborhoods in the city of Paris based on the top 10 venues for each neighborhood district. Moreover, we use the Elbow method to find out the optimal number of clusters for K-Means clustering.
- Visualization:
  - Use python folium package to visualize geographic details of Paris and its district neighborhood which are superimposed on top.
  - Analyze the clustering results and then propose the ideal districts to open a Moroccan restaurant in Paris. And, we give some perspectives to enhance the performances.

- Discussion and Conclusions:
    - Discussion of any limitations and how the results can be used, and any conclusions that can be drawn.

3.1. Data Research and Preparation

The Arrondissements dataset was downloaded from Paris|DATA:

https://opendata.paris.fr/explore/dataset/arrondissements/table/?dataChart

Then placed on the GitHub repository for the project:

https://raw.githubusercontent.com/naimnaoufal/Capstone-Project-The-Battle-of-Neighborhoods/master/arrondissements.csv

The data (table 1) was imported from the source, but as can be seen, was not in the right format.

	CAR	NAME	NSQAR	CAR.1	CARINSEE	LAR	NSQCO	SURFACE	PERIMETRE	Geometry_X	Geometry_Y
0	3	Temple	750000003	3	3	3eme Ardt	750001537	1170882828	4519264	48.862872	2.360001
1	19	Buttes-Chaumont	750000019	19	19	19eme Ardt	750001537	6792651129	11253182	48.887076	2.384821
2	14	Observatoire	750000014	14	14	14eme Ardt	750001537	5614877309	10317483	48.829245	2.326542
3	10	Entrepot	750000010	10	10	10eme Ardt	750001537	2891739442	6739375	48.876130	2.360728
4	12	Reuilly	750000012	12	12	12eme Ardt	750001537	16314782637	24089666	48.834974	2.421325
5	16	Passy	750000016	16	16	16eme Ardt	750001537	16372542129	17416110	48.860392	2.261971
6	11	Popincourt	750000011	11	11	11eme Ardt	750001537	3665441552	8282012	48.859059	2.380058

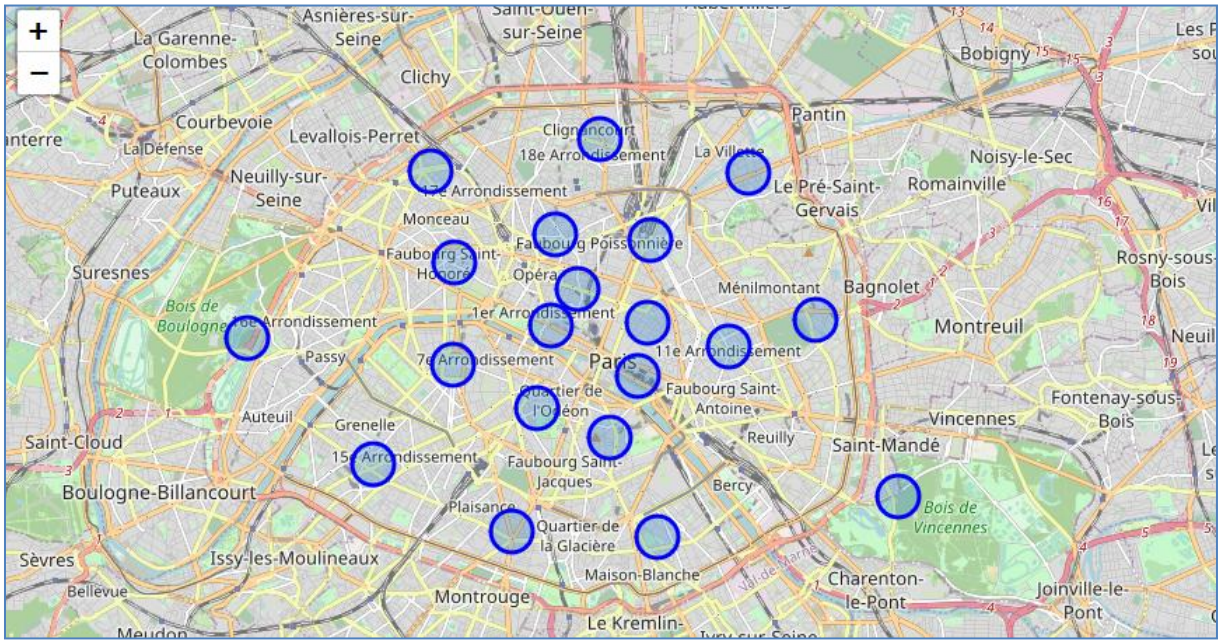
Table 1 : Imported Data

After some data wrangling and cleaning – renaming and dropping unnecessary columns - the dataframe was in a structure that could be used (table 2).

	Arrondissement_Num	Neighborhood	Arrondissement_Name	Latitude	Longitude
0	3	Temple	3eme Ardt	48.862872	2.360001
1	19	Buttes-Chaumont	19eme Ardt	48.887076	2.384821
2	14	Observatoire	14eme Ardt	48.829245	2.326542
3	10	Entrepot	10eme Ardt	48.876130	2.360728
4	12	Reuilly	12eme Ardt	48.834974	2.421325
5	16	Passy	16eme Ardt	48.860392	2.261971
6	11	Popincourt	11eme Ardt	48.859059	2.380058

Table 2 : Neighborhood Data in Paris, France

The GeoPy library is used to geocode location data and get location coordinates. It was imported here to get the latitude and longitude values of Paris. From this using folium I created a map of Paris (map 1) with the location of the 20 districts superimposed.



Map 1 : Map of neighborhoods in Paris

We use **Foursquare API** to generate the information of the top-100 venues within the given 500 metres radius from the centre of each district in the City of Paris, for instance, the relative locations, categories. For example, the relative information of some districts are describes in the Table 3.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Temple	48.862872	2.360001	Mmmozza	48.863910	2.360591	Sandwich Place
1	Temple	48.862872	2.360001	Chez Alain Miam Miam	48.862369	2.361950	Sandwich Place
2	Temple	48.862872	2.360001	Fromagerie Jouannault	48.862947	2.362530	Cheese Shop
3	Temple	48.862872	2.360001	Square du Temple	48.864475	2.360816	Park
4	Temple	48.862872	2.360001	Marché des Enfants Rouges	48.862806	2.361996	Farmers Market

Table 3 : Example of the top-100 venues

After extracting the informations based on Foursquare API, there are several types of restaurants in Paris, such as: French restaurant, Ramen restaurant, Japanese restaurant, Greek restaurant, Cambodian restaurant, Moroccan restaurant, etc. That is very important data to tackle out problem.

Then it comes the step to group rows by neighborhood and by taking the mean of the frequency of occurrence of each category, create the new dataframe and display the top 25 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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Mo Comm Veni
0	Batignolles-Monceau	French Restaurant	Hotel	Italian Restaurant	Bakery	Japanese Restaurant	Bistro	Café	Plaza	Concert Hall	Mediterranei Restaura
1	Bourse	French Restaurant	Wine Bar	Cocktail Bar	Hotel	Creperie	Bistro	Bakery	Italian Restaurant	Thai Restaurant	Cheese Sh
2	Buttes-Chaumont	French Restaurant	Bar	Pizza Place	Restaurant	Supermarket	Beer Bar	Seafood Restaurant	Bistro	Hotel	Music Sto
3	Buttes-Montmartre	Bar	French Restaurant	Convenience Store	Coffee Shop	Hotel	Restaurant	Seafood Restaurant	Supermarket	Bistro	Fast Fo Restaura
4	Entrepot	French Restaurant	Hotel	Bistro	Coffee Shop	Indian Restaurant	Café	Japanese Restaurant	Pizza Place	Italian Restaurant	Seafo Restaura
5	Gobelins	Vietnamese Restaurant	Asian Restaurant	Chinese Restaurant	Thai Restaurant	French Restaurant	Juice Bar	Plaza	Furniture / Home Store	Creperie	Pa
6	Hotel-de-Ville	French Restaurant	Ice Cream Shop	Clothing Store	Hotel	Garden	Plaza	Wine Bar	Pedestrian Plaza	Tea Room	Itali Restaura

Table 4 : Top 25 venues for each neighborhood.

3.2. Clustering neighborhoods in Paris

In order to cluster the districts of the City in Paris, we use K-Means method. Also, we use **Elbow technique** to determine the optimal value of the number of clusters for K-Means clustering.

To find out the optimal value k, we need plot the chart with the following features:

- values for K on the horizontal axis
- the distortion or inertia on the Y axis which describe the values calculated by the cost function.

In order to determine the optimal number of clusters, we select the value of number of cluster at the "Elbow" in chart. Indeed, we choose the point whose distortion or inertia starts decreases in a linear fashion.

In our case, we have the following charts of Elbow technique:

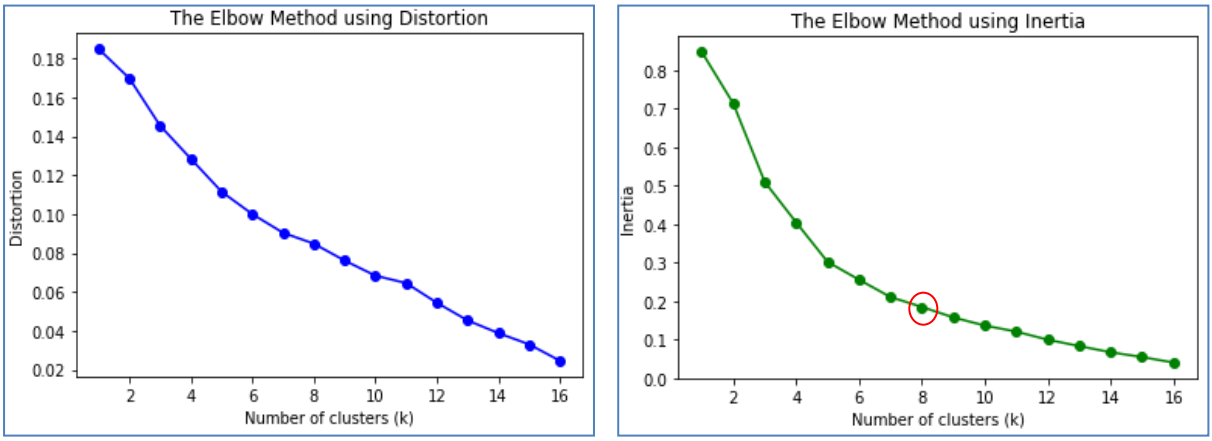


Figure 1 : The distortion and inertia in fonction of number of clusters (k)

As you see from the above charts of the Elbow Method using the distortion or the inertia, the optimal number of clusters for our data is 8.

Now, it is time to run *k*-means to cluster the neighborhood into 8 clusters and see the results.

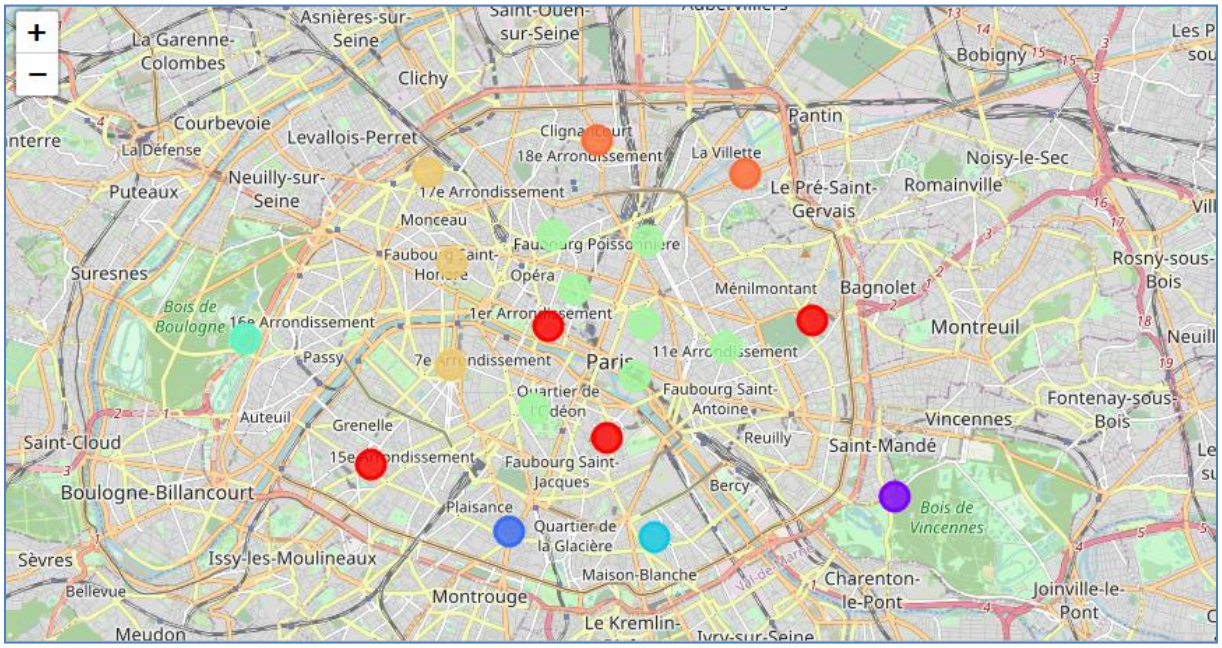
4. Results

The table 5 shows a part of a new dataframe that includes the cluster labels as well as the top 25 venues for each neighborhood.

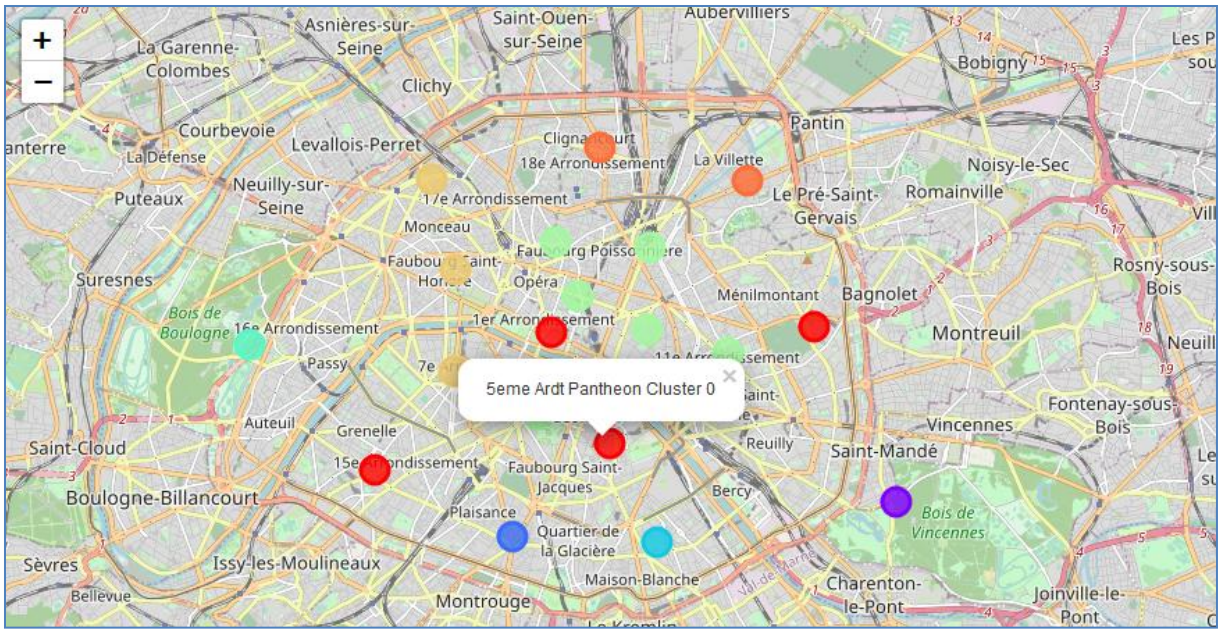
	Arrondissement_Num	Neighborhood	Arrondissement_Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	3	Temple	3eme Ardt	48.862872	2.360001	5	French Restaurant	Japanese Restaurant	Gourmet Shop	Cocktail Bar
1	19	Buttes-Chaumont	19eme Ardt	48.887076	2.384821	7	French Restaurant	Bar	Pizza Place	Restaurant
2	14	Observatoire	14eme Ardt	48.829245	2.326542	2	French Restaurant	Hotel	Bakery	Brasserie
3	10	Entrepot	10eme Ardt	48.876130	2.360728	5	French Restaurant	Hotel	Bistro	Coffee Shop
4	12	Reuilly	12eme Ardt	48.834974	2.421325	1	Zoo Exhibit	Bike Rental / Bike Share	Monument / Landmark	Supermarket
5	16	Passy	16eme Ardt	48.860392	2.261971	4	Plaza	Park	Lake	Pool
6	11	Popincourt	11eme Ardt	48.859059	2.380058	5	French Restaurant	Supermarket	Restaurant	Wine Bar

Table 5 : Cluster labels and the top venues for each neighborhood

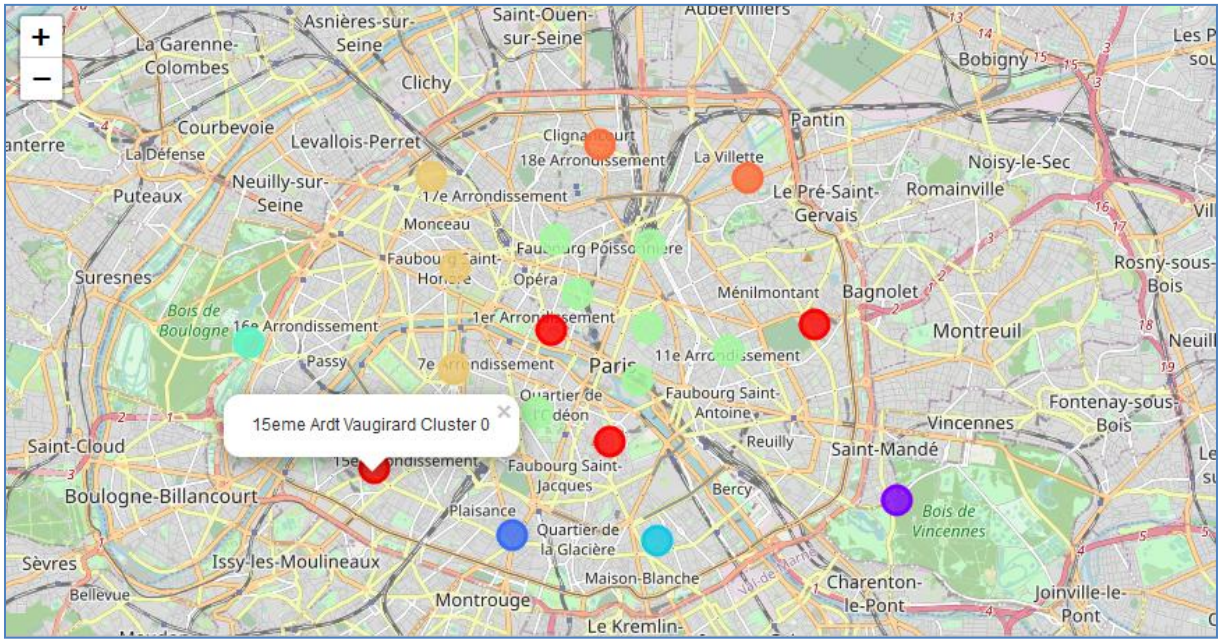
We, then, have the results as described in the following pictures:



Map 2 : The clusters



Map 3 : The 5th Arrondissement where there is Moroccan Restaurants



Map 4 : The 15th Arrondissement where there is Moroccan Restaurants

Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

Based on the defining categories, we can then assign a name to each cluster from the first cluster to the 8th cluster as illustrated as following:

- Cluster 1 :

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Louvre	0	French Restaurant	Japanese Restaurant	Plaza	Hotel	Italian Restaurant	Art Museum	Historic Site	Theater	Candy Store	Cheese Shop
12	Pantheon	0	French Restaurant	Italian Restaurant	Bakery	Hotel	Plaza	Pub	Coffee Shop	Science Museum	Café	Vietnamese Restaurant
14	Menilmontant	0	Japanese Restaurant	Bakery	French Restaurant	Plaza	Bistro	Italian Restaurant	Bar	Café	Photography Lab	Lounge
18	Vaugirard	0	Italian Restaurant	Hotel	French Restaurant	Coffee Shop	Indian Restaurant	Supermarket	Bakery	Thai Restaurant	Japanese Restaurant	Park

11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue	16th Most Common Venue	17th Most Common Venue	18th Most Common Venue	19th Most Common Venue	20th Most Common Venue	21th Most Common Venue	22th Most Common Venue	23th Most Common Venue
Coffee Shop	Thai Restaurant	Udon Restaurant	Garden	Dessert Shop	Exhibit	Boutique	Museum	Cosmetics Shop	Pedestrian Plaza	Perfume Shop	Furniture / Home Store	General College & University
Greek Restaurant	Garden	Pizza Place	Bistro	Historic Site	Bar	Wine Bar	Creperie	Museum	Hostel	Moroccan Restaurant	Portuguese Restaurant	Chocolate Shop
Bookstore	Latin American Restaurant	Fast Food Restaurant	Street Art	Market	Mexican Restaurant	Roof Deck	Restaurant	Diner	Movie Theater	Pizza Place	Korean Restaurant	Brasserie
Brasserie	Lebanese Restaurant	Cocktail Bar	Sushi Restaurant	Bistro	Persian Restaurant	Sports Bar	Moroccan Restaurant	Shoe Store	Scandinavian Restaurant	Irish Pub	Plaza	Basque Restaurant

- Cluster 2 :

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
4	Reuilly	1	Zoo Exhibit	Bike Rental / Bike Share	Monument / Landmark	Supermarket	Zoo	Argentinian Restaurant	Dive Bar	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant

- Cluster 3 :

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
2	Observatoire	2	French Restaurant	Hotel	Bakery	Brasserie	Food & Drink Shop	Pizza Place	Bistro	Supermarket	Sushi Restaurant	EV Charging Station	Tea Room

- Cluster 4 :

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
17	Gobelins	3	Vietnamese Restaurant	Asian Restaurant	Chinese Restaurant	Thai Restaurant	French Restaurant	Juice Bar	Plaza	Furniture / Home Store	Creperie	Park	Coffee Shop

- Cluster 5 :

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
5	Passy	4	Plaza	Park	Lake	Pool	Bus Station	French Restaurant	Art Museum	Bus Stop	Boat or Ferry	Bike Rental / Bike Share	Donut Shop

- Cluster 6 :

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Temple	5	French Restaurant	Japanese Restaurant	Gourmet Shop	Cocktail Bar	Art Gallery	Italian Restaurant	Burger Joint	Sandwich Place	Wine Bar	Coffee Shop
3	Entrepot	5	French Restaurant	Hotel	Bistro	Coffee Shop	Indian Restaurant	Café	Japanese Restaurant	Pizza Place	Italian Restaurant	Seafood Restaurant
6	Popincourt	5	French Restaurant	Supermarket	Restaurant	Wine Bar	Pastry Shop	Café	Pizza Place	Italian Restaurant	Bar	Bakery
7	Bourse	5	French Restaurant	Wine Bar	Cocktail Bar	Hotel	Creperie	Bistro	Bakery	Italian Restaurant	Thai Restaurant	Cheese Shop
8	Hotel-de-Ville	5	French Restaurant	Ice Cream Shop	Clothing Store	Hotel	Garden	Plaza	Wine Bar	Pedestrian Plaza	Tea Room	Italian Restaurant
16	Opera	5	French Restaurant	Hotel	Cocktail Bar	Bakery	Bistro	Wine Bar	Lounge	Japanese Restaurant	Coffee Shop	Pizza Place
19	Luxembourg	5	French Restaurant	Fountain	Bistro	Wine Bar	Italian Restaurant	Miscellaneous Shop	Lebanese Restaurant	Boutique	Outdoor Sculpture	Market

- Cluster 7 :

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
9	Batignolles-Monceau	6	French Restaurant	Hotel	Italian Restaurant	Bakery	Japanese Restaurant	Bistro	Café	Plaza	Concert Hall	Mediterranean Restaurant
13	Palais-Bourbon	6	Hotel	French Restaurant	Café	Italian Restaurant	Plaza	History Museum	Coffee Shop	Cocktail Bar	Historic Site	Garden
15	elysee	6	French Restaurant	Hotel	Spa	Art Gallery	Theater	Mediterranean Restaurant	Corsican Restaurant	Sporting Goods Shop	Resort	Plaza

- Cluster 8 :

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Buttes-Chaumont	7	French Restaurant	Bar	Pizza Place	Restaurant	Supermarket	Beer Bar	Seafood Restaurant	Bistro	Hotel	Music Store
10	Buttes-Montmartre	7	Bar	French Restaurant	Convenience Store	Coffee Shop	Hotel	Restaurant	Seafood Restaurant	Supermarket	Bistro	Fast Food Restaurant

5. Discussion

In the above results, we have the clustering of the various districts based on top 25 venues for each neighborhood.

As you see, the French Restaurant is the first most common venue in most of districts in Paris.

When reviewing the clusters, we could see that the **Moroccan restaurant** is in cluster 1. Thus, the districts, which are clustered in **cluster 1**, are potential location candidates.

In summary, our proposed districts could be considered as one of the interesting starting points. In order to analyze more in details, we need to review other important factors and conditions.

6. Conclusion and perspectives

6.1. Conclusion

The objective of this assessment was mainly to study the information of the districts and the neighborhood data in Paris, to propose the ideal districts to open a Moroccan restaurant.

Indeed, we applied several techniques to obtain the data such as the neighborhood data, the coordinates, the most famous venues in the French capital, Paris as well as combined all of the obtained features.

There are many ways this analysis could have been performed based on different methodolgy and perhaps different data sources. I chose the method I selected as it was a straight forward way to narrow down the options, not complicating what is actually simple in many ways.

Firstly, we imported the neighborhood data of 20 districts in Paris.

Secondly, we used python package geopy to convert the addresses into the latitude and the longitude values. And, to retrieve the important information of the top 25 famous venues for each district, such as, their types, their coordinates using Foursquare API.

Thridly, in order to visualize geographic details of Paris and its district neighborhoods, we used python package folium to generate a map in which the district neighborhoods are superimposed on top.

Finally, we used Elbow method to determine the optimal value of the number of clusters for K-Means clustering. And, we launched K-Means technique to cluster the neighborhood into 8 clusters. Then, we reused tool folium to illustrate the data in an information map where the districts of Paris are clustered according to the venue density.

In conclusion, depending on the several requirements of the investors, if they would like to open new Moroccan restaurant in the district that have already had many Moroccan restaurants, they should open in **District 15 : Vaugirard**. In addition, they can open new one in **District 5 : Pantheon**, because there is also a good community for opening Moroccan restaurant in this district. Moreover, if the investors would like to open a new restaurant in the

districts that are *similar to District 5 and 15*, they could locate it in the districts that are clustered in **Cluster 1** such as **District 1 (Louvre) and District 20 (Menilmontant) in Paris**.

6.2.Perspectives :

To go further in the analysis, we could enhance the features of district by adding more relevant features for each district such as:

- the transport info (public transport, parking, etc.),
- the information of Moroccan and Arab communities,
- the information of major tourist venues.

Concerning the clustering methods and enhancing the performances, we could use some other algorithms, for instance:

- FCM : Fuzzy c-means method
- DBSCAN: Density-Based Spatial Clustering of Applications with Noise
- Hierarchical K-Means Clustering
- HCPC: Hierarchical clustering on principal components
- Deep Learning Models. To see more details, please check "Unsupervised deep embedding for clustering analysis"

7. References

[1] Paris population, https://www.worldometers.info/world-population/france-population/

[2] The Economist Intelligence Unit, https://www.eiu.com/n/campaigns/worldwide-cost-of-living-2020

[3] The tutorials in course "Applied Data Science Capstone", https://www.coursera.org/learn/applied-data-science-capstone/

[4] Elbow Method for optimal value of k in KMeans, https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/