Market Research
Opening New Mexican Restaurant in Paris, France

Karl Bayer

Table of Contents

1. Introduction

1.1. Overview

Paris is the **French capital** and most interesting city of France. There are 2 140 526 residents (official estimate, 1 January 2019) in an area of 105 square kilometers (41 square miles).

In addition, Paris Region, also called as Île-de-France, includes the City of Paris which is the center and seat of government. Île-de-France is located in the north-central region of France. It is the most populous of the 18 regions of France. It covers about 12 012 square kilometers (about 4638 square miles), about 2% of French territory . However, its population is about 12 213 364 (about 18.2% of French population) and it contributes nearly 30% of GDP in France .

According to the Economist Intelligence Unit Worldwide Cost of Living Survey in 2018, Paris was the second most expensive city in the world, after Singapore, and ahead of Zürich, Hong Kong, Oslo and Geneva. Another source ranked Paris as most expensive, on a par with Singapore and Hong Kong, in 2018.

Normally, we call *the City of Paris as Paris* and *Paris Region as Île-de-France*.

1.2. Business problem

In this assignment, we focus on **finding the good districts from 20 districts to open a Mexican Restaurant in Paris, France.**

Indeed, in Paris, it is not limited in the Parisian cuisine to beef and escargots bourguignon. There are a concentration of Latin American people in the French capital where the tourists could sample several Mexican specialties.

As you see from the above figures, Paris is one of the most interesting city to open Mexican restaurant.

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

If we think about the venues restaurant, they could want to choose the neighborhood district where some Mexican restaurants are opened.

1.3. Main Contributions

After presenting the goals, we can emphasize on the following contributions:

- Collecting the information of neighborhood data and their coordinates of the districts in Paris using library **geopy.**

- Obtaining the information of the most famous venues in Paris using **Foursquare API.**

- Preprocessing and Analyzing the data corpus (to clean, to normalize).

- Exploring the preprocessed data.
- Clustering the districts based on the top 10 venues for each neighborhood district.
- Illustrating the data corpus in a map where the districts of Paris are clustered according to the
  venue density.

2. Data Description

To explore our problem, we need build Paris neighborhood data and their coordinates.

- Concerning to Paris neighborhood data, we use the following references:
    - Paris Arrondissements & Neighborhoods Map
    - Arrondissements in Paris, France
- Concerning to relative coordinates (latitude, longitude) of each district in Paris
  Using package geopy to convert an address into latitude and longitude values.
- Concerning to the most famous venues and the relative locations ○ Using Foursquare API.

3. Methodology

- Firstly, we build the Paris neighborhood data (Postcode, Neighborhood).

- Secondly, we build the coordinates of all districts in Paris, France. We use python **geopy** package to to convert an address into latitude and longitude values.

- Thirdly, we use Foursquare API to generate to the most famous venues and the relative locations.

- Then, we need to explore, segment and using **K-Means** as part of this clustering study 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.

- Finally, we use python package **folium** to visualize geographic details of Paris and its district neighborhood which are superimposed on top. We then analyze the clustering result and then propose some suggestions about the good locations (districts) to open Mexican Restaurant in Paris. And, we give some perspectives to enhance the performances.

4. Data Settings

In order to prepare the corpus for our problem, we have to build the neighborhood data for the City of Paris and their coordinates.

However, it is difficult to have the Parisian neighborhood data. Thus, I built it manually based on the information in [10, 11]. Then, we have the neighborhood corpus that is described in Table 1.

In addition, we use python **geopy** package to convert an address into the latitude and the longitude values. Then, we could build the coordinates of all districts in Paris, France that is described in Table 2.

Table 1: Neighborhood Data in Paris, France

Table 2: The coordinates of 20-district in Paris, France

	postcode	neighbourhood
0	75001	75002,75003,75004,75005,75006,75007,75008,75009
1	75002	75001,75003,75009,75010
2	75003	75001,75002,75004,75010,75011
3	75004	75001,75003,75005,75006,75011,75012
4	75005	75001,75004,75006,75012,75013,75014
5	75006	75001,75004,75005,75007,75014,75015
6	75007	75001,75006,75008,75015,75016
7	75008	75001,75007,75009,75016,75017,75018
8	75009	75001,75002,75008,75010,75017,75018
9	75010	75002,75003,75009,75011,75018,75019,75020
10	75011	75003,75004,75010,75012,75019,75020
11	75012	75004,75005,75011,75013,75020
12	75013	75005,75012,75014
13	75014	75005,75006,75013,75015
14	75015	75006,75007,75014,75016
15	75016	75007,75008,75015,75017
16	75017	75008,75009,75016,75018
17	75018	75008,75009,75010,75017,75019
18	75019	75010,75011,75018,75020
19	75020	75010,75011,75012,75019

	postcode	latitude	longitude
0	75001	48.863554	2.338856
1	75002	48.867418	2.344256
2	75003	48.862607	2.360211
3	75004	48.856004	2.357028
4	75005	48.852752	2.346343
5	75006	48.853537	2.343370
6	75007	48.855913	2.313839
7	75008	48.872385	2.312707
8	75009	48.877355	2.336856
9	75010	48.879201	2.354391
10	75011	48.855630	2.370806
11	75012	48.839734	2.380054
12	75013	48.826997	2.353396
13	75014	48.828590	2.307541
14	75015	48.838461	2.315728
15	75016	48.855031	2.273958
16	75017	48.883508	2.304923
17	75018	48.893074	2.343881
18	75019	48.878076	2.376198
19	75020	48.857126	2.409257

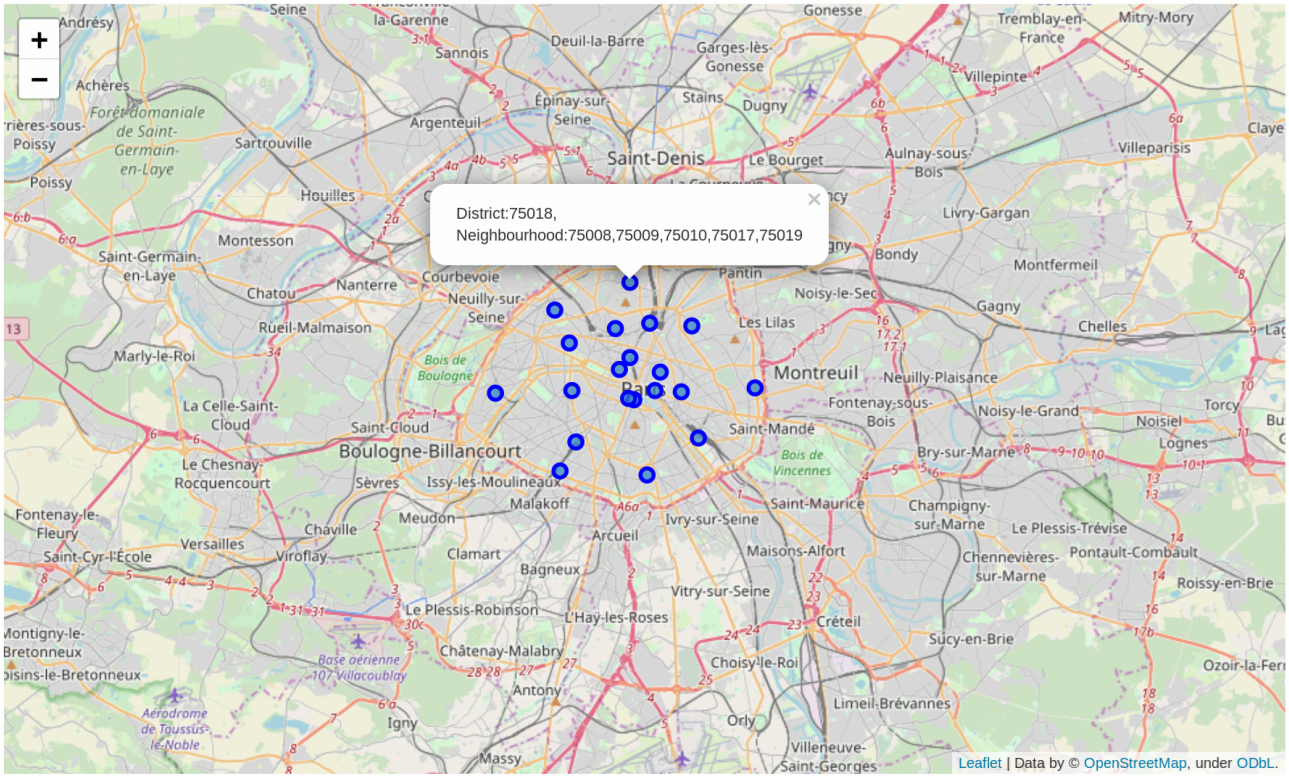
We use **Foursquare API** to generate the information of the top-100 venues within the given 500 metres radius from the centre of each district in the City of Paris, for instance, the relative locations, categories. For example, the relative information of some districts are describes in the Table 3.

	neighbourhood	neighbourhood latitude	neighbourhood longitude	Venue	Venue latitude	Venue longitude	Venue Category
0	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Jardin du Palais Royal	48.864941	2.337728	Garden
1	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Palais Royal	48.863236	2.337127	Historic Site
2	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Comédie-Française	48.863088	2.336612	Theater
3	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Place du Palais Royal	48.862523	2.336688	Plaza
4	75002,75003,75004,75005,75006,75007,75008,75009	48.863554	2.338856	Christian Louboutin	48.862697	2.340757	Shoe Store

Table 3: Example of the top-100 venues in radius 500 metres from the centre of each district.

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

Moreover, we use **folium package** to visualize geographic details of Paris and its district neighborhood which are superimposed on top.



5. Results

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.

5.1. Determining the optimal number of clusters for K-Means clustering

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

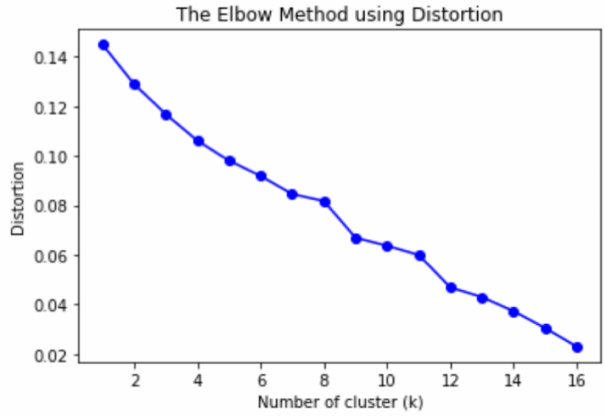
+ values for k on the horizontal axis

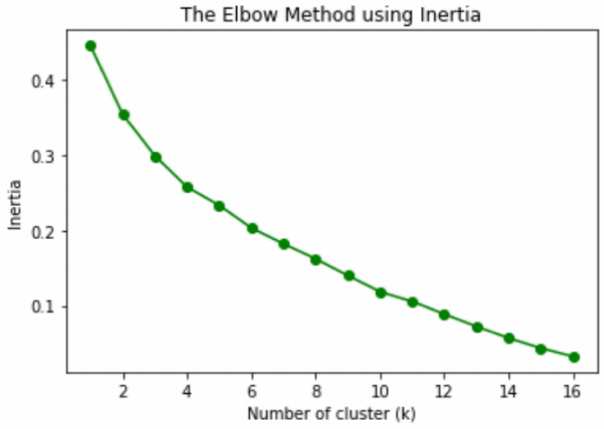
+ the distortion or the inertia on the Y axis which described the values calculated by the cost function.

Then, 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 experiments, we have the following charts of Elbow technique:

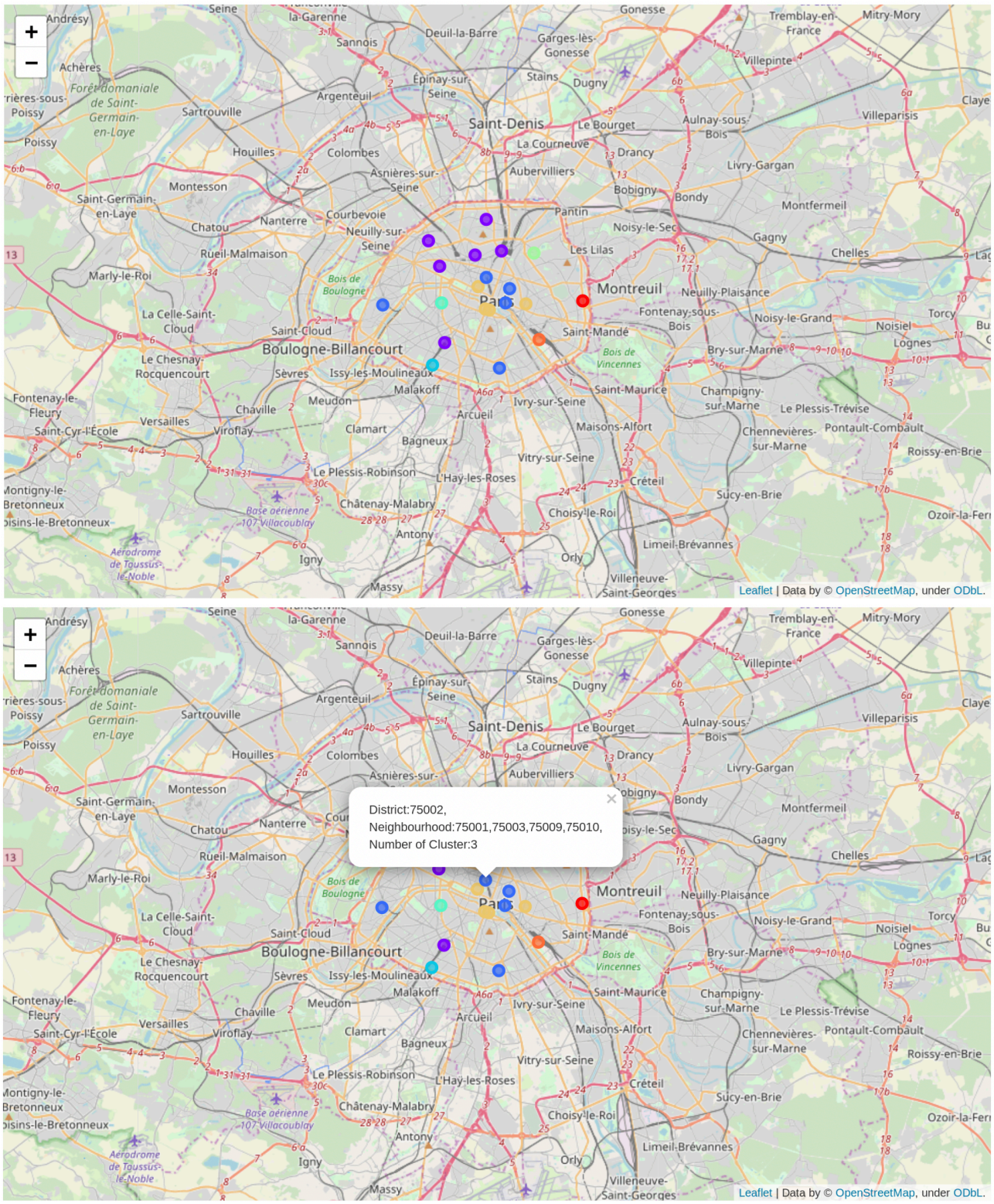
As you see above charts of the Elbow Method using the distortion or the inertia, it is difficult to determine the Elbow point in graph using the inertia. Indeed, using the chart of Elbow method applying the distortion with the number of clusters, we could choose the optimal number of clusters for our data corpus is **8**.





5.2. Clustering Neighborhoods of Paris, France

Using the optimal value k, we launch K-Means technique to cluster the neighborhood into 8 clusters. We, then, have the result as described in the following picture:



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

+ Cluster 1:

	neighbourhood	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
19	75010,75011,75012,75019	0	Hotel	Supermarket	Tram Station	French Restaurant	Pizza Place	Brazilian Restaurant	Fast Food Restaurant	Stadium	Music Venue	Discount Store

+ Cluster 2:

neighbourhood	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
75001,75007,75009,75016,75017,75018	1	French Restaurant	Hotel	Bakery	Thai Restaurant	Theater	Art Gallery	Spa	Insurance Office	Creperie	Cocktail Bar
75001,75002,75008,75010,75017,75018	1	French Restaurant	Hotel	Bakery	Cocktail Bar	Bistro	Italian Restaurant	Bar	Lounge	Japanese Restaurant	Café
75002,75003,75009,75011,75018,75019,75020	1	French Restaurant	Hotel	Indian Restaurant	Restaurant	Japanese Restaurant	Coffee Shop	Bakery	African Restaurant	Gastropub	Sports Bar
75006,75007,75014,75016	1	Hotel	French Restaurant	Pizza Place	Dessert Shop	Coffee Shop	Italian Restaurant	Japanese Restaurant	Seafood Restaurant	Bakery	Lebanese Restaurant
75008,75009,75016,75018	1	Hotel	French Restaurant	Italian Restaurant	Bar	Bistro	Bakery	Sushi Restaurant	Yoga Studio	Café	Chinese Restaurant
75008,75009,75010,75017,75019	1	French Restaurant	Bar	Hotel	Café	Pizza Place	Bistro	Italian Restaurant	Theater	Deli / Bodega	Wine Bar

+ Cluster 3:

	neighbourhood	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	75001,75003,75009,75010	2	French Restaurant	Bakery	Cocktail Bar	Wine Bar	Bistro	Italian Restaurant	Hotel	Thai Restaurant	Pastry Shop	Vietnamese Restaurant
2	75001,75002,75004,75010,75011	2	French Restaurant	Coffee Shop	Burger Joint	Café	Bistro	Boutique	Clothing Store	Bakery	Gourmet Shop	Pizza Place
3	75001,75003,75005,75006,75011,75012	2	French Restaurant	Pastry Shop	Hotel	Wine Bar	Gourmet Shop	Art Gallery	Clothing Store	Bakery	Ice Cream Shop	Tea Room
12	75005,75012,75014	2	French Restaurant	Vietnamese Restaurant	Hotel	Bar	Bakery	Thai Restaurant	Bistro	Fast Food Restaurant	Café	Japanese Restaurant
15	75007,75008,75015,75017	2	French Restaurant	Bakery	Italian Restaurant	Japanese Restaurant	Plaza	Coffee Shop	Bar	Train Station	Seafood Restaurant	Garden

+ Cluster 4:

	neighbourhood	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
13	75005,75006,75013,75015	3	Café	Bakery	Supermarket	Hotel	Grocery Store	Plaza	French Restaurant	Flea Market	Farmers Market	Bistro

+ Cluster 5:

	neighbourhood	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
6	75001,75006,75008,75015,75016	4	French Restaurant	Hotel	Café	Plaza	Italian Restaurant	History Museum	Garden	Art Museum	Historic Site	Basque Restaurant

+ Cluster 6:

	neighbourhood	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
18	75010,75011,75018,75020	5	French Restaurant	Bar	Park	Café	Restaurant	Bistro	Diner	Moroccan Restaurant	Bus Stop	Metro Station

+ Cluster 7:

neighbourhood	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
'5002,75003,75004,75005,75006,75007,75008,75009	6	French Restaurant	Hotel	Café	Plaza	Japanese Restaurant	Bistro	Coffee Shop	Wine Bar	Bakery	Restaurant
75001,75004,75006,75012,75013,75014	6	French Restaurant	Indie Movie Theater	Hotel	Café	Bookstore	Plaza	Bar	Bakery	Coffee Shop	Ice Cream Shop
75001,75004,75005,75007,75014,75015	6	French Restaurant	Bookstore	Café	Creperie	Hotel	Bar	Seafood Restaurant	Plaza	Garden	Japanese Restaurant
75003,75004,75010,75012,75019,75020	6	French Restaurant	Bar	Pizza Place	Italian Restaurant	Bistro	Coffee Shop	Hotel	Pub	Cocktail Bar	Vegetarian / Vegan Restaurant

+ Cluster 8:

	neighbourhood	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	75004,75005,75011,75013,75020	7	French Restaurant	Hotel	Beer Bar	Coffee Shop	Restaurant	Music Venue	Museum	Skate Park	Brasserie	Steakhouse

6. Discussion

In above result, we have the clustering result of the various districts based on top 10 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 **Mexican restaurant** in cluster 3.

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

7. Conclusion and Perspectives

7.1. Conclusion

The objective of this assessment was mainly to study the information of the districts and the neighborhood data in Paris, France to proposed the good districts to open a Mexican 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.

In addition to the provision of some directions for future research, this assignment has made many contributes to the literature on market research in Paris, France.

- First, we built the neighborhood data of 20 districts in Paris

- Second, we use python package **geopy** to convert the addresses that were built in the first step into the latitude values and the longitude values.

- Third, to retrieve the important information of the top-10 famous venues for each district, we use **Foursquare API**, such as, their types, their coordinates.

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

- Fifth, we use **Elbow method** to determine the optimal value of the number of clusters for K-Means clustering.

- Sixth, we launch K-Means technique to cluster the neighborhood into 8 clusters. Then, we reuse tool **folium** to illustrate the data corpus in an information map where the districts of Paris are clustered according to the venue density.

- Finally, in our investigation of the results of experiments about market research in the capital Paris, we could propose some good districts to the investors depending on their requirements and conditions. If they would like to open new Vietnamese restaurant in the district that have already had many Vietnamese restaurants, they should open in **District 13**. Also, they can open new one in **District 2**, because this district is also good community for opening Mexican restaurant. Moreover, if the investors would like to open new one in the districts that are *similar to District 13 and District 2*, they could locate it in the districts that are clustered in **Cluster 3** such as **District 3, 4, 16 in Paris**.

7.2. Perspectives

Concerning to enhance the features of district, we should add more relevant features for each district such as:

- the transport info (public transport, parking, etc.),

- the information of Mexican communities,

- the information of major tourist venues

Concerning to clustering methods and enhancing the performances, we could do some experiments with other algorithms, for instance,

- DBSCAN: Density-based clustering
- Fuzzy c-means method
- Hierarchical K-Means Clustering
- HCPC: Hierarchical clustering on principal components
- Deep Learning Models.

References

- INSEE official estimated population by department and region as of 1 January 2019". 22 January 2019. Archived from the original on 21 April 2017. Retrieved 4 April 2017.

- Île-de-France, https://en.wikipedia.org/wiki/%C3%8Ele-de-France

- Paris, https://en.wikipedia.org/wiki/Paris

- 5 top Mexicans restaurants in Paris : https://www.mariefrance.fr/cuisine/mes-bonnes-adresses/paris-restaurants-mexicains-a-tester-absolument-439188.html#item=1 restaurants-paris/

  - Telegraph, 16 March 2018 Archived 30 March 2019 at the Wayback Machine

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

- Paris Arrondissements & Neighborhoods Map, https://parismap360.com/paris-Housing

- Sales Prices & Venues Data Analysis of Istanbul, https://www.linkedin.com/pulse/housing- sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/ arrondissement- map#.XfVpqtEo91l

- Arrondissements in Paris, France, https://francetravelplanner.com/go/paris/areas/arrondismt.html

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