Market Research for Opening New italian Restaurant in Paris, France

How to find the best place using data-driven approach?

Raphaël Leclercq
01/04/2021

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) [1].

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 [2]. However, its population is about 12 213 364 (about 18.2% of French population) and it contributes nearly 30% of GDP in France [6].

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 [3]. Another source ranked Paris as most expensive, on a par with Singapore and Hong Kong, in 2018 [4, 5]. As an investor, it means you have to choose the right place to minimize the risk as your rent will be one of the most expensive in the world.

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 Vietnamese Restaurant in Paris, France.

Indeed, in Paris, it is not limited in the Parisian cuisine to beef and escargots bourguignon. There are over 160 000 Italian people in the French capital where the tourists could sample several italian specialties.

Moreover, the New York times has published an article intitled "*Paris Might Be the Best City for Italian Food (Outside Italy)"* describing the love for Italian food in France [7].

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

Concerning to the investors, we expect from them to prefer the districts in which there are the good community of the Italian 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 Italian 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 exploration

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 [10]
  - Arrondissements in Paris, France [11]
- 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 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 and we build manually the neighborhood corpus that is described in Table 1.

Table 1 : Neighborhood Data 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

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.

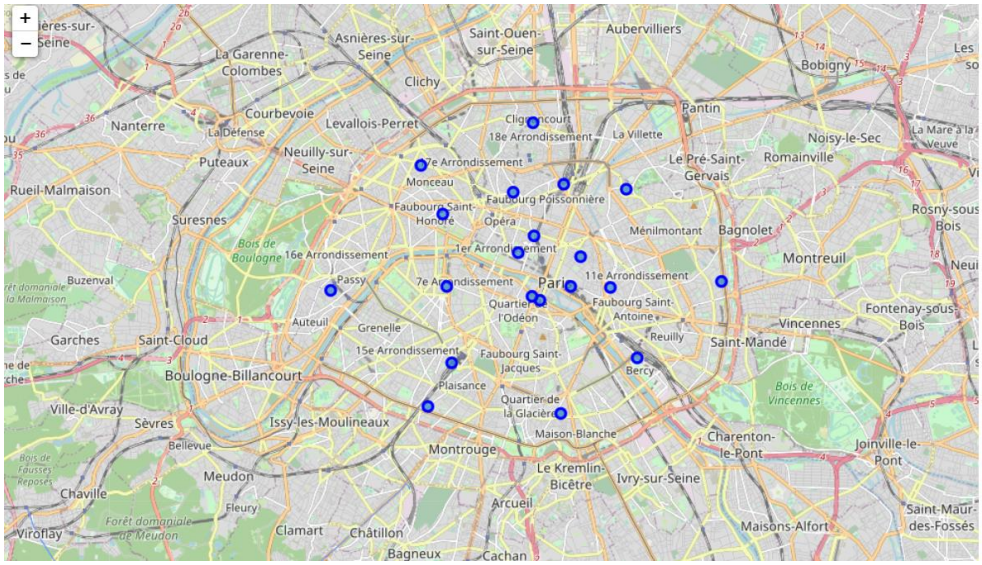
Table 2: The coordinates of 20-district in Paris, France (5 first ones)

	postcode	neighbourhood	address	latitude	longitude
0	75001	75002,75003,75004,75005,75006,75007,75008,75009	75001, FR	48.863554	2.338856
1	75002	75001,75003,75009,75010	75002, FR	48.867418	2.344256
2	75003	75001,75002,75004,75010,75011	75003, FR	48.862607	2.360211
3	75004	75001,75003,75005,75006,75011,75012	75004, FR	48.856004	2.357028
4	75005	75001,75004,75006,75012,75013,75014	75005, FR	48.852752	2.346343

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.

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.



4 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.

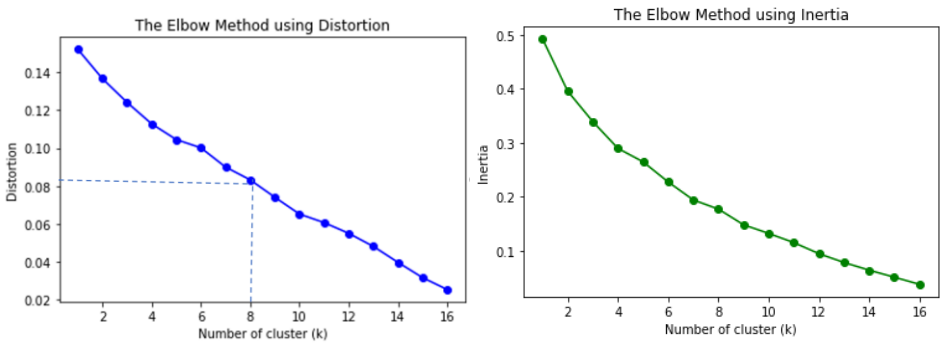
4.1 Optimal number of clusters for K-Means clustering

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.

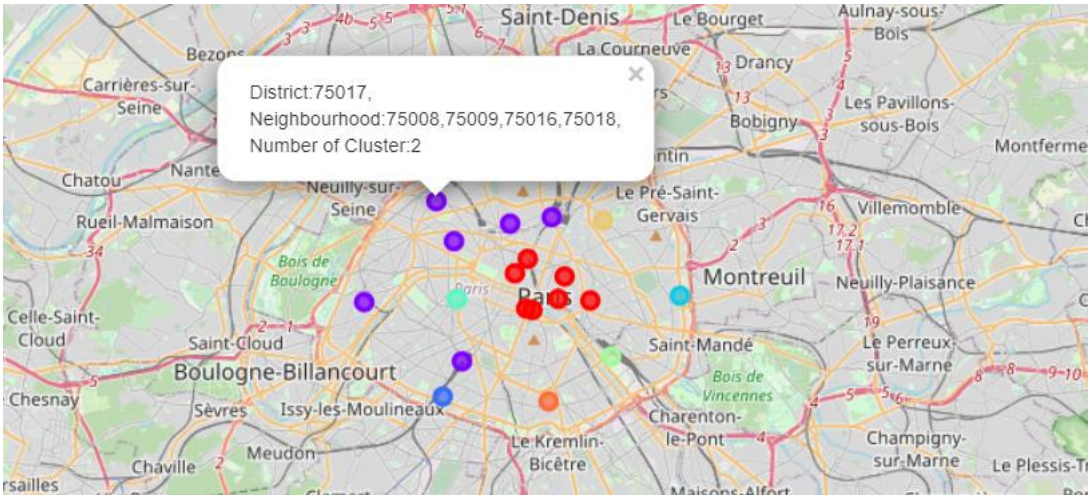
Then, we select the value of number of cluster at the "Elbow" in chart. Indeed, we choose the point whose distortion or inertia starts decreases in a linear fashion. In our experiments, we have the following charts of Elbow technique:



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.

4.2 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, we can then assign a name to each cluster from the first cluster to the 8th cluster as illustrated as following:

```
# Cluster 1
df_merged_paris.loc[df_merged_paris[COL_NAME_CLUSTER_LABELS] == 0,
                    df_merged_paris.columns[[1] + list(range(5, df_merged_paris.shape[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 M Comn Ve
0	75002,75003,75004,75005,75006,75007,75008,75009	0	French Restaurant	Japanese Restaurant	Plaza	Italian Restaurant	Historic Site	Hotel	Café	Thai Restaurant	Gar
1	75001,75003,75009,75010	0	French Restaurant	Cocktail Bar	Bakery	Wine Bar	Salad Place	Coffee Shop	Hotel	Thai Restaurant	Clot S
2	75001,75002,75004,75010,75011	0	French Restaurant	Coffee Shop	Japanese Restaurant	Gourmet Shop	Bookstore	Bistro	Sandwich Place	Italian Restaurant	Bu J
3	75001,75003,75005,75006,75011,75012	0	French Restaurant	Clothing Store	Plaza	Hotel	Wine Bar	Ice Cream Shop	Gourmet Shop	Pastry Shop	Ita Restau
4	75001,75004,75006,75012,75013,75014	0	French Restaurant	Hotel	Coffee Shop	Indie Movie Theater	Plaza	Bookstore	Japanese Restaurant	Bar	Restau
5	75001,75004,75005,75007,75014,75015	0	French Restaurant	Bookstore	Japanese Restaurant	Plaza	Indie Movie Theater	Creperie	Tea Room	Hotel	Seaf Restau
10	75003,75004,75010,75012,75019,75020	0	French Restaurant	Bar	Coffee Shop	Bistro	Bookstore	Italian Restaurant	Hotel	Pizza Place	Coc

```
# Cluster 2
df_merged_paris.loc[df_merged_paris[COL_NAME_CLUSTER_LABELS] == 1,
                    df_merged_paris.columns[[1] + list(range(5, df_merged_paris.shape[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
7	75001,75007,75009,75016,75017,75018	1	French Restaurant	Hotel	Bakery	Art Gallery	Corsican Restaurant	Spa	Hotel Bar	Grocery Store	Plaza
8	75001,75002,75008,75010,75017,75018	1	French Restaurant	Hotel	Cocktail Bar	Bakery	Bistro	Bar	Japanese Restaurant	Café	Lounge
9	75002,75003,75009,75011,75018,75019,75020	1	French Restaurant	Hotel	Restaurant	Japanese Restaurant	Coffee Shop	Indian Restaurant	Bakery	Gluten-free Restaurant	Bookstore
14	75006,75007,75014,75016	1	French Restaurant	Hotel	Bakery	Seafood Restaurant	Japanese Restaurant	Dessert Shop	Farmers Market	Coffee Shop	Indian Restaurant
15	75007,75008,75015,75017	1	French Restaurant	Italian Restaurant	Bakery	Japanese Restaurant	Plaza	Seafood Restaurant	Train Station	Bar	Clothing Store
16	75008,75009,75016,75018	1	Hotel	French Restaurant	Italian Restaurant	Bakery	Bar	Japanese Restaurant	Sushi Restaurant	Restaurant	Bistro

```
# Cluster 3
df_merged_paris.loc[df_merged_paris[COL_NAME_CLUSTER_LABELS] == 2,
                    df_merged_paris.columns[[1] + list(range(5, df_merged_paris.shape[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
13	75005,75006,75013,75015	2	Supermarket	Bakery	Hotel	Café	Japanese Restaurant	Gym Pool	Diner	Farmers Market	Fast Food Restaurant	Flea Market

```
# Cluster 4
df_merged_paris.loc[df_merged_paris[COL_NAME_CLUSTER_LABELS] == 3,
                    df_merged_paris.columns[[1] + list(range(5, df_merged_paris.shape[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	3	Hotel	Music Venue	French Restaurant	Tram Station	Indian Restaurant	Tennis Court	Discount Store	Fast Food Restaurant	Brazilian Restaurant	Stadium

```
# Cluster 5
df_merged_paris.loc[df_merged_paris[COL_NAME_CLUSTER_LABELS] == 4,
                    df_merged_paris.columns[[1] + list(range(5, df_merged_paris.shape[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
6	75001,75006,75008,75015,75016	4	French Restaurant	Hotel	Plaza	Café	Italian Restaurant	History Museum	Garden	Art Museum	Bistro	Historic Site

```
# Cluster 6
df_merged_paris.loc[df_merged_paris[COL_NAME_CLUSTER_LABELS] == 5,
                    df_merged_paris.columns[[1] + list(range(5, df_merged_paris.shape[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
11	75004,75005,75011,75013,75020	5	French Restaurant	Hotel	Beer Garden	Skate Park	Brasserie	Coffee Shop	Lounge	Boat or Ferry	Beer Bar	Road

```
# Cluster 7
df_merged_paris.loc[df_merged_paris[COL_NAME_CLUSTER_LABELS] == 6,
                    df_merged_paris.columns[[1] + list(range(5, df_merged_paris.shape[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
18	75010,75011,75018,75020	6	French Restaurant	Bistro	Bar	Pool	Park	Restaurant	Café	Gas Station	Brasserie	Metro Station

```
# Cluster 8
df_merged_paris.loc[df_merged_paris[COL_NAME_CLUSTER_LABELS] == 7,
                    df_merged_paris.columns[[1] + list(range(5, df_merged_paris.shape[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
12	75005,75012,75014	7	French Restaurant	Vietnamese Restaurant	Bar	Bakery	Hotel	Thai Restaurant	Bistro	Japanese Restaurant	Italian Restaurant	Asian Restaurant
17	75008,75009,75010,75017,75019	7	French Restaurant	Bar	Hotel	Café	Restaurant	Deli / Bodega	Supermarket	Italian Restaurant	Pizza Place	Gastropub

5 Conclusion and Perspective

5.1 Conclusion

In above result, we have the clustering result of the various districts based on top 10 venues for each neighborhood.

Thus, 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 Italian restaurant is present in several cluster: 1,2, 5 and 8. It is rated second in cluster 2, especially district 17.

So, depending on the several requirements of the investors, if we would like to open new Italian restaurant in the district that have already had many Italian restaurant, we should open in District 15.

5.2 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 italian communities,
- the information of major tourist venues
- the number of families as italian restaurant are a typical menu for french family

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

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

7 Références

[1] "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.

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

[3] Economist Intelligence Unit Worldwide Cost of Living Survey, 2018, cited in the London Telegraph, 16 March 2018 Archived 30 March 2019 at the Wayback Machine

[4] Singapour, Hong Kong, Paris : le trio des villes les plus chères du monde Archived 27 March 2019 at the Wayback Machine, International mail.com. 20 March 2019.

[5] Paris, https://en.wikipedia.org/wiki/Paris

[6] "Ile-de-France- Portrait of the Region- Key figures (in French)". Regional Council of the Ile-de-France. Retrieved 24 November 2018.

[7] New York Times, Paris Might Be the Best City for Italian Food (Outside Italy), https://www.nytimes.com/2019/12/04/travel/Italian-food-in-Paris.html#:~:text=Others%20make%20culinary%20pilgrimages%20from,figure%20to%20be%20double%20that.

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