Market Survey
for Open New Asian Restaurant
in Paris, France

Ngoc Tan LE

06 March 2020

Table of contents

1.Business Problem

In this work, we would like to find the good district from 20 districts to recommend Asian Restaurants in Paris, France.

2.Proposed Approach

First of all, we create a dataframe with the Paris neighborhood dataset. This dataset contains Postcode and Neighborhood information.

Second, we create the coordinates of all districts in Paris, France.

Third, we experiment by exploring, segmenting and clustering all the neighborhoods in the city of Paris based on the most common venues.

For evaluations, we analyze the clustered results and then propose appropriated districts to recommend Asian Restaurants in Paris. Finally, we conclude with some perspectives to enhance the performances of our approach.

3.Data configuration

In this section, we create the dataframe from Paris neighborhood dataset with their coordinates.

- Paris Arrondissements & Neighborhoods Map: https://parismap360.com/paris-arrondissement-map#.XfVpqtEo91I
  The neighborhood corpus is described in **Table 1**.
- Arrondissements in Paris, France: https://francetravelplanner.com/go/paris/areas/arrondismt.html
- Using package **geopy** to convert an address into latitude and longitude values. The coordinates of Neighborhoods with geopy package is describe in **Table 2**.
- Using **Foursquare API** to deal with the most famous venues and the relative locations

	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

Table 1: Neighborhoods grouped by postal code with commas

	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

Table 2: Coordinates of Neighborhoods with geopy package

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

Table 3: Coordinates of venue categories with Foursquare API

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 python **folium** package to visualize geographic details of Paris and its district neighborhood which are superimposed on top (see **Figure 1**).

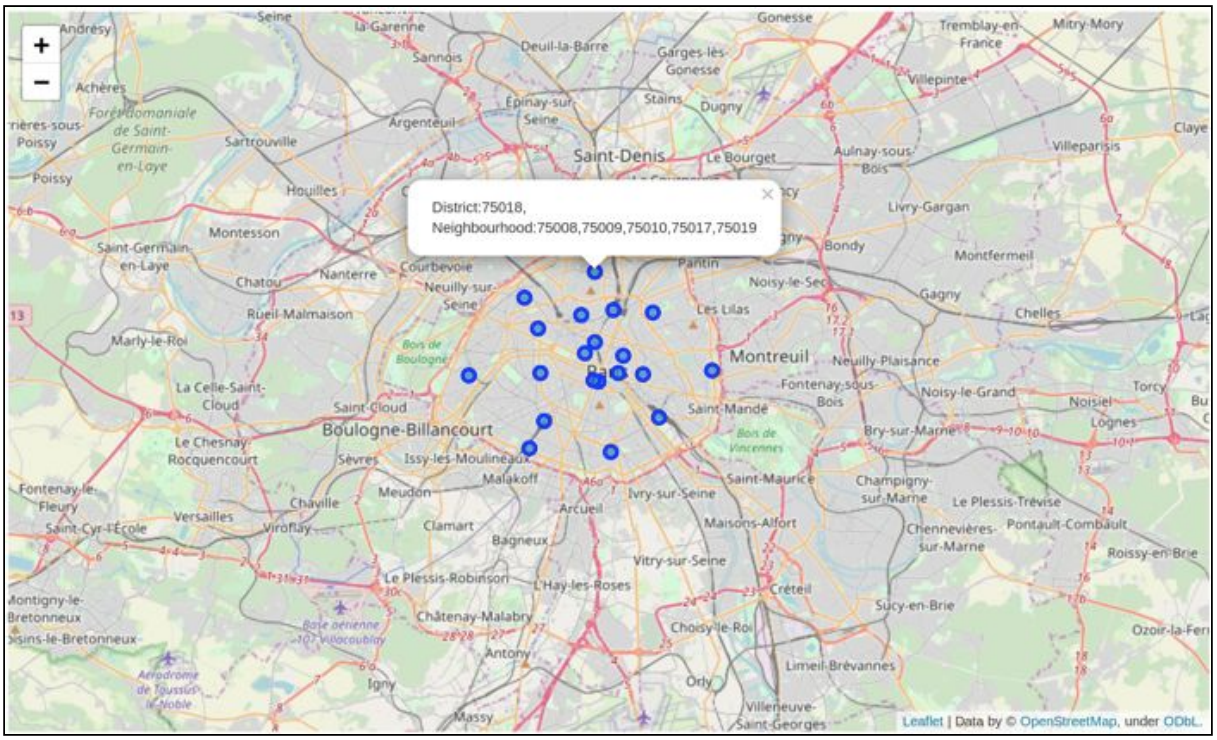


Figure 1: Map to visualize geographic details of Paris and its district neighborhood with folium package

According to the statistics, there are 199 distinct venue categories. For example:

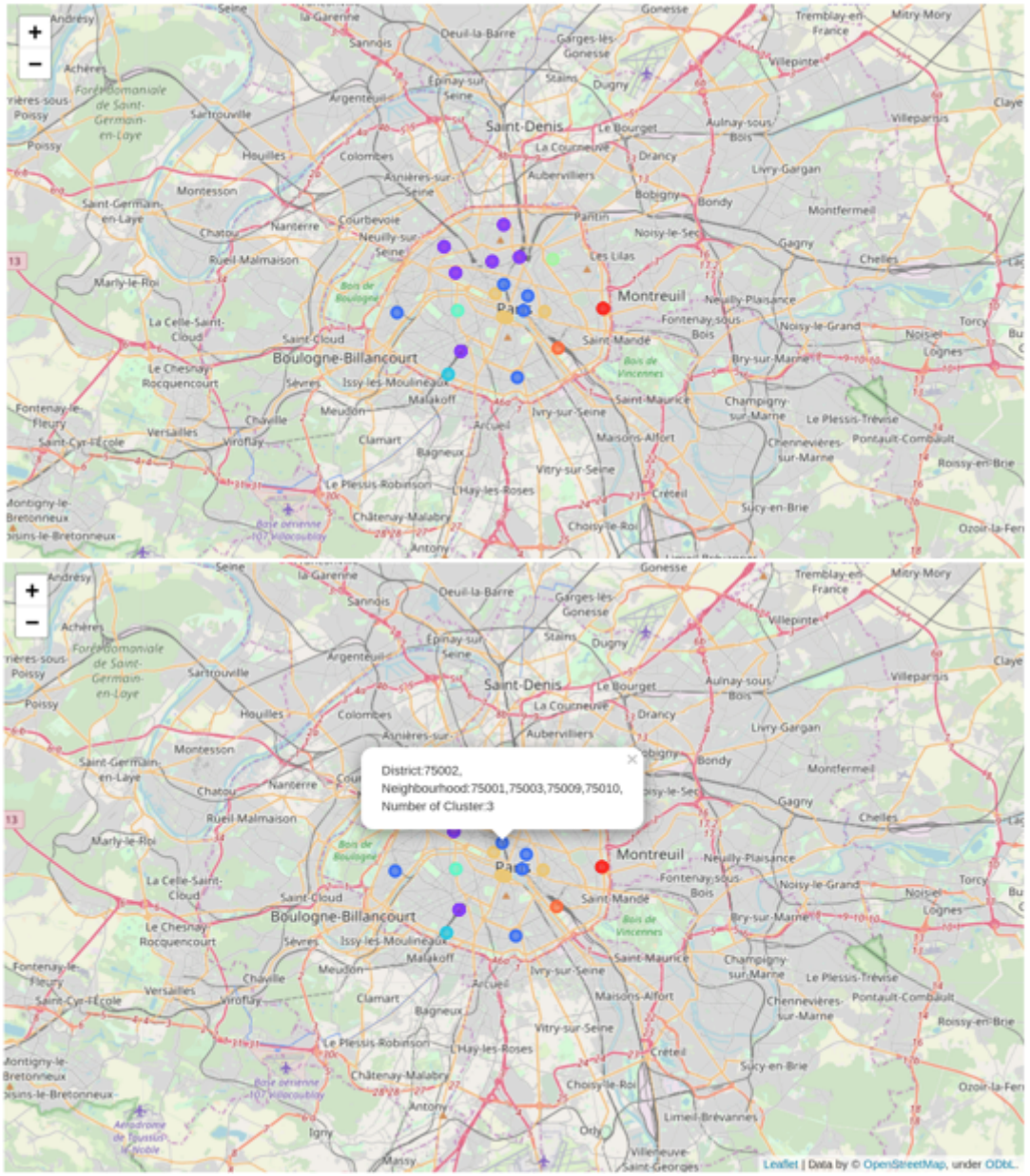
'Garden', 'Historic Site', 'Theater', 'Plaza', 'Hotel',
'Shoe Store', 'French Restaurant', 'Cheese Shop', 'Restaurant',
'Bar', 'Smoke Shop', 'Café', 'Spa', 'Breakfast Spot',
'Sculpture Garden', 'Coffee Shop', 'Pizza Place',
'Ramen Restaurant', 'Bistro', 'Bakery', 'Wine Shop',
'Udon Restaurant', 'Sandwich Place', 'Wine Bar', 'Art Museum',
'Pedestrian Plaza', 'Japanese Restaurant', 'Chinese Restaurant',
'Korean Restaurant', 'Brasserie', 'Clothing Store', 'Cocktail Bar',
'Tea Room', 'Cosmetics Shop', 'Italian Restaurant', 'Exhibit',
'Furniture / Home Store', 'General College & University',
'Bubble Tea Shop', 'Shopping Mall', 'Perfume Shop',
'Grocery Store', 'Gift Shop', 'Vietnamese Restaurant', 'Bookstore',
'Beer Bar', 'Nightclub', 'Souvlaki Shop', etc.

```
print('There are {} distinct categories.'.format(
    len(venues_neighbourhoods[COL_NAME_VENUE_CATEGORY].unique())))
```

```
There are 199 distinct categories.
```

4.Experiments and Evaluation

In order to cluster the districts of the City in Paris, we use KMeans method. Using the optimal value k, we launch KMeans technique to cluster the neighborhood into 6 clusters.
We, then, have the result as described in the following picture:



We display each neighborhood along with the top 5 most common venues:

```
----75001,75002,75004,75010,75011----
                 Venue  freq
0    French Restaurant  0.07
1                Café  0.05
2          Coffee Shop  0.05
3  Japanese Restaurant  0.04
4              Bakery  0.04


----75001,75002,75008,75010,75017,75018----
                 Venue  freq
0   French Restaurant  0.17
1               Hotel  0.12
2        Cocktail Bar  0.05
3  Italian Restaurant  0.04
4              Bakery  0.04


----75001,75003,75005,75006,75011,75012----
                Venue  freq
0  French Restaurant  0.10
1     Clothing Store  0.05
2        Pastry Shop  0.04
3              Hotel  0.03
4             Bakery  0.03


----75001,75003,75009,75010----
                Venue  freq
0  French Restaurant  0.10
1       Cocktail Bar  0.06
2           Wine Bar  0.05
3             Bakery  0.05
4              Hotel  0.04


----75001,75004,75005,75007,75014,75015----
                Venue  freq
0  French Restaurant  0.12
1          Bookstore  0.05
2              Hotel  0.05
3                Bar  0.04
4              Plaza  0.04


----75001,75004,75006,75012,75013,75014----
                Venue  freq
0  French Restaurant  0.11
1               Café  0.05
2        Coffee Shop  0.04
```

```
3                Bar  0.04
4              Plaza  0.04


----75001,75006,75008,75015,75016----
              Venue  freq
0  French Restaurant  0.23
1              Hotel  0.18
2               Café  0.07
3     History Museum  0.05
4              Plaza  0.05


----75001,75007,75009,75016,75017,75018----
                Venue  freq
0    French Restaurant  0.21
1                Hotel  0.11
2               Bakery  0.05
3  Japanese Restaurant  0.03
4                  Spa  0.03


----75002,75003,75004,75005,75006,75007,75008,75009----
                Venue  freq
0    French Restaurant  0.14
1                Hotel  0.09
2  Japanese Restaurant  0.06
3                Café  0.06
4              Plaza  0.05


----75002,75003,75009,75011,75018,75019,75020----
                Venue  freq
0    French Restaurant  0.18
1                Hotel  0.14
2    Indian Restaurant  0.07
3  Japanese Restaurant  0.05
4          Coffee Shop  0.05


----75003,75004,75010,75012,75019,75020----
               Venue  freq
0   French Restaurant  0.11
1         Coffee Shop  0.05
2         Pizza Place  0.05
3                 Bar  0.05
4  Italian Restaurant  0.04


----75004,75005,75011,75013,75020----
              Venue  freq
0  French Restaurant  0.23
1              Hotel  0.14
```

```
2        Beer Garden  0.05
3        Coffee Shop  0.05
4         Steakhouse  0.05


----75005,75006,75013,75015----
                Venue  freq
0          Supermarket  0.15
1              Bakery  0.15
2               Hotel  0.07
3  Japanese Restaurant  0.04
4              Bistro  0.04


----75005,75012,75014----
                  Venue  freq
0      French Restaurant  0.13
1  Vietnamese Restaurant  0.08
2                   Bar  0.08
3        Thai Restaurant  0.07
4                Bakery  0.07


----75006,75007,75014,75016----
                Venue  freq
0   French Restaurant  0.16
1               Hotel  0.16
2        Dessert Shop  0.05
3  Seafood Restaurant  0.05
4         Coffee Shop  0.05


----75007,75008,75015,75017----
                 Venue  freq
0   Italian Restaurant  0.14
1    French Restaurant  0.14
2               Bakery  0.08
3  Japanese Restaurant  0.06
4                Plaza  0.05


----75008,75009,75010,75017,75019----
               Venue  freq
0  French Restaurant  0.14
1               Bar  0.12
2             Hotel  0.07
3        Pizza Place  0.05
4              Café  0.03


----75008,75009,75016,75018----
                Venue  freq
0   French Restaurant  0.14
```

```
1   Italian Restaurant  0.14
2                Hotel  0.12
3              Bakery  0.09
4              Bistro  0.05


----75010,75011,75012,75019----
                 Venue  freq
0                Hotel  0.14
1          Supermarket  0.10
2    French Restaurant  0.10
3         Tram Station  0.10
4  Japanese Restaurant  0.05


----75010,75011,75018,75020----
                Venue  freq
0  French Restaurant  0.13
1                Bar  0.08
2              Café  0.05
3               Park  0.05
4         Restaurant  0.05
```

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:

	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	0	French Restaurant	Hotel	Beer Garden	Museum	Garden	Skate Park	Chinese Restaurant	Steakhouse	Coffee Shop	Convenience 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
1	75001,75003,75009,75010	1	French Restaurant	Cocktail Bar	Bakery	Wine Bar	Italian Restaurant	Hotel	Bistro	Coffee Shop	Thai Restaurant	Restaurant
2	75001,75002,75004,75010,75011	1	French Restaurant	Café	Coffee Shop	Japanese Restaurant	Burger Joint	Bistro	Bakery	Gourmet Shop	Restaurant	Sandwich Place
3	75001,75003,75005,75006,75011,75012	1	French Restaurant	Clothing Store	Pastry Shop	Bakery	Ice Cream Shop	Wine Bar	Gourmet Shop	Hotel	Garden	Furniture / Home Store
4	75001,75004,75006,75012,75013,75014	1	French Restaurant	Café	Bar	Hotel	Coffee Shop	Bookstore	Plaza	Bakery	Creperie	Bistro
5	75001,75004,75005,75007,75014,75015	1	French Restaurant	Hotel	Bookstore	Plaza	Bar	Seafood Restaurant	Café	Creperie	Coffee Shop	Lebanese Restaurant
10	75003,75004,75010,75012,75019,75020	1	French Restaurant	Pizza Place	Coffee Shop	Bar	Italian Restaurant	Hotel	Bookstore	Bistro	Pub	Cocktail Bar
12	75005,75012,75014	1	French Restaurant	Vietnamese Restaurant	Bar	Bakery	Thai Restaurant	Hotel	Bistro	Japanese Restaurant	Italian Restaurant	Juice Bar
17	75008,75009,75010,75017,75019	1	French Restaurant	Bar	Hotel	Pizza Place	Bistro	Gastropub	Café	Italian Restaurant	Supermarket	Restaurant
18	75010,75011,75018,75020	1	French Restaurant	Bar	Park	Bistro	Pool	Restaurant	Café	Moroccan Restaurant	Greek Restaurant	Bus Stop

+ 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
13	75005,75006,75013,75015	2	Bakery	Supermarket	Hotel	Japanese Restaurant	Theater	Café	Flea Market	Plaza	Fast Food Restaurant	Stadium

+ 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
0	75002,75003,75004,75005,75006,75007,75008,75009	3	French Restaurant	Hotel	Japanese Restaurant	Café	Plaza	Coffee Shop	Historic Site	Bakery	Udon Restaurant	Bistro
6	75001,75006,75008,75015,75016	3	French Restaurant	Hotel	Café	Plaza	Italian Restaurant	History Museum	Art Museum	Garden	Park	Historic Site
7	75001,75007,75009,75016,75017,75018	3	French Restaurant	Hotel	Bakery	Cocktail Bar	Spa	Theater	Japanese Restaurant	Art Gallery	Cycle Studio	Brewery
8	75001,75002,75008,75010,75017,75018	3	French Restaurant	Hotel	Cocktail Bar	Bistro	Italian Restaurant	Bar	Bakery	Japanese Restaurant	Lounge	Theater
9	75002,75003,75009,75011,75018,75019,75020	3	French Restaurant	Hotel	Indian Restaurant	Coffee Shop	Japanese Restaurant	Restaurant	Bakery	Breton Restaurant	Café	Record Shop
14	75006,75007,75014,75016	3	French Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Coffee Shop	Seafood Restaurant	Dessert Shop	Beer Store	Pizza Place	Garden

+ 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
19	75010,75011,75012,75019	4	Hotel	French Restaurant	Tram Station	Supermarket	Japanese Restaurant	Tennis Court	Music Venue	Fast Food Restaurant	Discount Store	Pharmacy

+ 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
15	75007,75008,75015,75017	5	French Restaurant	Italian Restaurant	Bakery	Japanese Restaurant	Plaza	Bar	Seafood Restaurant	Train Station	Grocery Store	Sandwich Place
16	75008,75009,75016,75018	5	French Restaurant	Italian Restaurant	Hotel	Bakery	Bistro	Sushi Restaurant	Japanese Restaurant	Bar	Restaurant	Diner

5.Conclusion

In this work, we have clustered and examined all six clusters based on top 10 most common venues for each neighborhood.

We observed the French restaurant is omnipresent in clusters 1, 2, 4 and 5 with the first most common venue in most of districts in Paris.

However, we notice that the recommendations for Asian restaurants are not sufficient and not relevant as well. For example, in cluster 1, Chinese restaurant is in the seventh most common venue. In cluster 2, Vietnamese restaurant is in the second most common venue; Thai restaurant is in the fifth and ninth most common venues depending on several districts and so on.

In perspective, we should add more relevant features for each district such as the transport info (public transport, parking, etc.), the information of asian communities, the information of major tourist venues as well.

We could experiment more algorithms: Fuzzy c-means method, Density-based clustering (DBSCAN), Hierarchical K-Means Clustering or Deep Learning Models.

References

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

2. Paris Arrondissements & Neighborhoods Map:
https://parismap360.com/paris-arrondissement-map#.XfVpqtEo91I

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