IBM Data Science Capstone Project

THE BATTLE FOR RESTAURANTS IN THE PARIS NEIGHBORHOODS

Philippe BOTTIER | Applied Data Science Capstone by IBM | May 19, 2021

Introduction

DESCRIPTION AND DISCUSSION OF THE CONTEXT

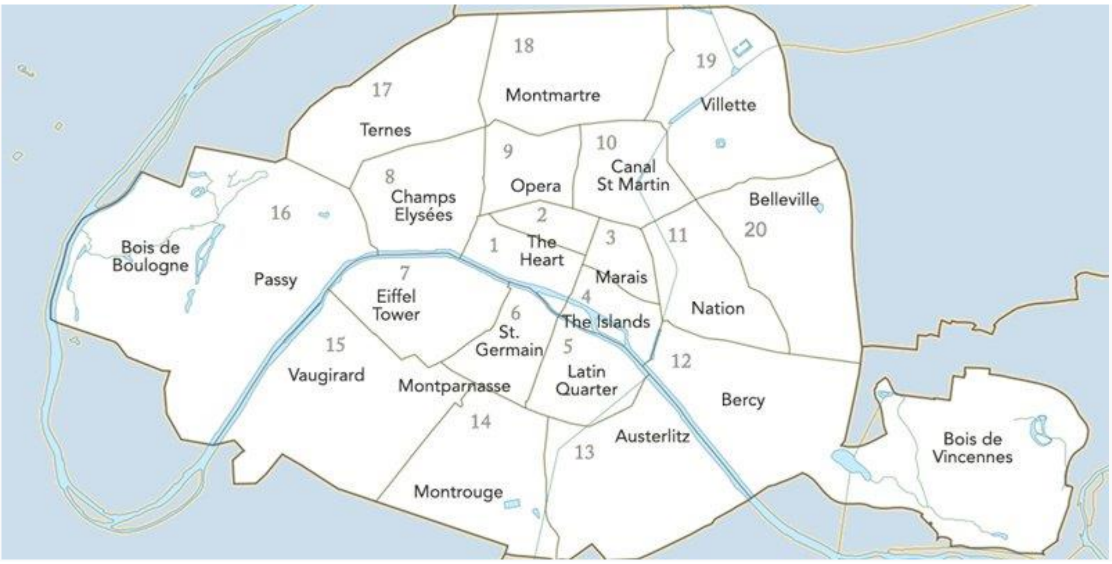
Paris is the capital and most populous city of France, with an estimated population of 2,175,601 residents as of 2018, in an area of more than 105 square kilometres (41 square miles). Since the 17th century, Paris has been one of Europe's major centres of finance, diplomacy, commerce, fashion, science, arts and gastronomy

Paris received 38 million visitors in 2019, measured by hotel stays, with the largest numbers of foreign visitors coming from the United States, the United Kingdom, Germany, and China. It was ranked as the second most visited travel destination in the world in 2019, after Bangkok and just ahead of London.

The city of Paris is divided into twenty administrative districts called arrondissements. The twenty arrondissements are arranged in the form of a clockwise spiral (often likened to a snail shell), starting from the middle of the city, with the first on the Right Bank (north bank) of the Seine.

In French, notably on street signs, the number is often given in Roman numerals. For example, the Eiffel Tower belongs to the VIIe arrondissement while Gare de l'Est is in the Xe arrondissement.

Each Parisian arrondissement has four neighborhoods, so a total of 80, which constitute the highest level of public administration in Paris.



As a resident in a suburb near the capital, I chose the city of Paris to lead my project.

BUSINESS PROBLEM

The problem to be solved is totally imaginary.

My company, PhB Data Consulting, specialized in data analysis, was contacted by the Tourist Office of the city of Paris to provide them with an analysis that would allow them to advise tourists on the types of restaurants they could find during their visits to the Paris neighborhoods.

This analysis could be carried out using an unsupervised learning model that would reproduce on the map of Paris the groupings of neighborhoods according to the types of restaurants that are most represented there.

DATA GATHERING

To carry out this project, I needed the following data :

The list of Parisian districts (arrondissements in French) as well as the list of Parisian neighborhoods (quartiers). These lists were imported from the Open Data site of the city of Paris :

- https://opendata.paris.fr/explore/dataset/arrondissements/export/?disjunctive.c_ar&disjunctive.c_arinsee&disjunctive.l_ar
- https://opendata.paris.fr/explore/dataset/quartier_paris/export/

The Geo-coordinates of the districts in Paris, obtained with the help of the geocoder tool in the notebook.

The Top venues data of neighborhoods, obtained from Foursquare through an API

After cleaning, preparing and merging the data, here is the dataframe that was used for the rest of the project. In this dataframe, we therefore have the 20 arrondissements of Paris with for each of them, their name, their geographical coordinates as well as the four neighbourhoods which depend of the arrondissement.

Merge boroughDF with neighbourhoodDF => borough_neighbourhoodDF

```
borough_neighbourhoodDF= bouroughDF.set_index('BoroughNumber').join(neighbourhoodDF.set_index('BoroughNumber')).reset_index()
borough_neighbourhoodDF.head()
```

	BoroughNumber	Borough	Latitude	Longitude	BoroughPerimetre	NeighNumber	Neighbourhood	NeighLatitude	NeighLongitude	NeighGeoloc	NeighPerimetre
0	1	Louvre	48.862563	2.336443	6054.936862	3	Palais-Royal	48.864660	2.336309	[48.8646599781, 2.33630891897]	2166.839239
1	1	Louvre	48.862563	2.336443	6054.936862	1	Saint-Germain-l'Auxerrois	48.860650	2.334910	[48.8606501352, 2.33491032928]	5057.549475
2	1	Louvre	48.862563	2.336443	6054.936862	2	Halles	48.862289	2.344899	[48.8622891081, 2.34489885831]	2606.417128
3	1	Louvre	48.862563	2.336443	6054.936862	4	Place-Vendôme	48.867019	2.328582	[48.8670185906, 2.32858166493]	2147.817602
4	2	Bourse	48.868279	2.342803	4554.104360	6	Vivienne	48.869100	2.339461	[48.8691001998, 2.33946074375]	2058.472959

```
borough_neighbourhoodDF.shape
```

```
(80, 11)
```

Methodology

To conduct this study, I list below the different steps of the methodology that I followed :

- Step 1 : Data Acquisition with the JSON library
- Step 2 : Preparing, Cleaning and Merging Data with the PANDA library
- Step 3 : Venues Acquisiton with the Foursquare API
- Step 4 : One Hot Encoding with the PANDA library
- Step 5 : Clustering K-means with sklearn.cluster library

- Step 6 : Cluster Analysis
- Step 7 : Creation of a clusters map with the Folium library

STEP 1 - DATA ACQUISITION WITH THE JSON LIBRARY

As I said before, I have retrieved the data from the Open Data site of the city of Paris (https://opendata.paris.fr/page/home/) :

- list of Paris arrondissements : https://opendata.paris.fr/explore/dataset/arrondissements/export/

- list of Paris neighbourhoods : https://opendata.paris.fr/explore/dataset/quartier_paris/export/

These files are in GeoJSON format. To be able to use the data, I have saved the files in the same directory as my Python notebook.

To import the files, I used the json library :

```
import json
geo_borough = json.load(open("arrondissements.geojson")) # Paris arrondissements
geo_neighbourhood = json.load(open("quartier_paris.geojson")) # Paris neighborhoods
```

The GeoJSON format is used to represent data of a geographic type. In this object, transformed into a dictionary under python, there are two elements: the type and the information (named features).

In each object of the features list, we also have different objects types :

```
geo_borough["features"][0].keys()
```

```
: dict_keys(['type', 'geometry', 'properties'])
```

In the properties field, there are various useful information, including the longitude and latitude coordinates of the center of the borough (or of the neighbourhood).

```
geo_borough["features"][0]['properties']
```

```
: {'n_sq_co': 750001537,
   'perimetre': 4519.26364836,
   'l_ar': '3ème Ardt',
   'surface': 1170882.82818778,
   'geom_x_y': [48.86287238, 2.3600009859],
   'n_sq_ar': 750000003,
   'l_aroff': 'Temple',
   'c_arinsee': 75103,
   'c_ar': 3}
```

```
geo_neighbourhood["features"][0]['properties']
```

```
: {'n_sq_qu': 750000015,
   'n_sq_ar': 750000004,
   'geom_x_y': [48.851585175, 2.36476795387],
   'c_qu': 15,
   'surface': 487264.93707154,
   'l_qu': 'Arsenal',
   'perimetre': 2878.55965556,
   'c_quinsee': 7510403,
   'c_ar': 4}
```

For the rest of the project, with the Panda library imported previously, I have created 2 DataFrames with the useful information.

```
# Borough Dataframe
bouroughDF = pd.DataFrame({
     "BoroughNumber"      : [bor["properties"]["c_ar"] for bor in geo_borough["features"]]
    ,"Borough"            : [bor["properties"]["l_aroff"] for bor in geo_borough["features"]]
    ,"Latitude"    : [bor["properties"]["geom_x_y"][0] for bor in geo_borough["features"]]
    ,"Longitude" : [bor["properties"]["geom_x_y"][1] for bor in geo_borough["features"]]
    ,"BoroughPerimetre" : [bor["properties"]["perimetre"] for bor in geo_borough["features"]]
})
```

	BoroughNumber	Borough	Latitude	Longitude	BoroughPerimetre
0	3	Temple	48.862872	2.360001	4519.263648
1	7	Palais-Bourbon	48.856174	2.312188	8099.424883
2	13	Gobelins	48.828388	2.362272	11546.546526
3	17	Batignolles-Monceau	48.887327	2.306777	10775.579516
4	20	Ménilmontant	48.863461	2.401188	10704.940486

```
# Neighbourhood Dataframe
neighbourhoodDF = pd.DataFrame({
    "BoroughNumber"   : [neigh["properties"]["c_ar"] for neigh in geo_neighbourhood["features"]]
    ,"NeighNumber"    : [neigh["properties"]['c_qu'] for neigh in geo_neighbourhood["features"]]
    ,"Neighbourhood" : [neigh["properties"]["l_qu"] for neigh in geo_neighbourhood["features"]]
    ,"NeighLatitude" : [neigh["properties"]["geom_x_y"][0] for neigh in geo_neighbourhood["features"]]
    ,"NeighLongitude": [neigh["properties"]["geom_x_y"][1] for neigh in geo_neighbourhood["features"]]
    ,"NeighGeoloc"    : [neigh["properties"]["geom_x_y"] for neigh in geo_neighbourhood["features"]]
    ,"NeighPerimetre": [neigh["properties"]["perimetre"] for neigh in geo_neighbourhood["features"]]
})
```

	BoroughNumber	NeighNumber	Neighbourhood	NeighLatitude	NeighLongitude	NeighGeoloc	NeighPerimetre
0	4	15	Arsenal	48.851585	2.364768	[48.851585175, 2.36476795387]	2878.559656
1	5	18	Jardin-des-Plantes	48.841940	2.356894	[48.8419401934, 2.35689388962]	4052.729521
2	10	39	Porte-Saint-Martin	48.871245	2.361504	[48.8712446509, 2.36150364735]	3245.891413
3	11	43	Roquette	48.857064	2.380364	[48.8570640408, 2.38036406173]	4973.010557
4	12	46	Picpus	48.830359	2.428827	[48.8303592424, 2.42882681508]	18261.910318
...	...	...	...	...	...	...	...

STEP 2 - PREPARING, CLEANING AND MERGING DATA WITH THE PANDA LIBRARY

After cleaning, merging, combining the values of the data frame, I obtain the following data frame which will allow me to continue the study :

```
borough_neighbourhoodDF= bouroughDF.set_index('BoroughNumber').join(neighbourhoodDF.set_index('BoroughNumber')).reset_index()
borough_neighbourhoodDF.head()
```

	BoroughNumber	Borough	Latitude	Longitude	BoroughPerimetre	NeighNumber	Neighbourhood	NeighLatitude	NeighLongitude	NeighGeoloc	NeighPerimetre
0	1	Louvre	48.862563	2.336443	6054.936862	3	Palais-Royal	48.864660	2.336309	[48.8646599781, 2.33630891897]	2166.839239
1	1	Louvre	48.862563	2.336443	6054.936862	1	Saint-Germain-l'Auxerrois	48.860650	2.334910	[48.8606501352, 2.33491032928]	5057.549475
2	1	Louvre	48.862563	2.336443	6054.936862	2	Halles	48.862289	2.344899	[48.8622891081, 2.34489885831]	2606.417128
3	1	Louvre	48.862563	2.336443	6054.936862	4	Place-Vendôme	48.867019	2.328582	[48.8670185906, 2.32858166493]	2147.817602
4	2	Bourse	48.868279	2.342803	4554.104360	6	Vivienne	48.869100	2.339461	[48.8691001998, 2.33946074375]	2058.472959

```
borough_neighbourhoodDF.shape
```

```
(80, 11)
```

In this dataframe we have 80 Paris Neighbourhoods.
I used this dataframe for the rest of the project. I didn't need to clean the data.

STEP 3 - VENUES ACQUISITON WITH THE FOURSQUARE API

I have used Foursquare API to get venues suggestions for each neighbourhood in Paris. Data from Foursquare API is received in json format.

The parameters passed during the call to the Foursquare API made it possible to retrieve 100 suggestions of venues for each neighbourhood within a radius of 500 meters from the coordinate point of the neighbourhood.

The created and arranged data frame looks like this :

```
paris_venues.head()
```

	Neighbourhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Palais-Royal	48.86466	2.336309	Jardin du Palais Royal	48.864941	2.337728	Garden
1	Palais-Royal	48.86466	2.336309	Palais Royal	48.863236	2.337127	Historic Site
2	Palais-Royal	48.86466	2.336309	Comédie-Française	48.863088	2.336612	Theater
3	Palais-Royal	48.86466	2.336309	Udon Bistro Kunitoraya (Kunitoraya)	48.865884	2.336782	Udon Restaurant
4	Palais-Royal	48.86466	2.336309	Sanukiya	48.864713	2.333805	Udon Restaurant

```
paris_venues.shape
```

```
(5061, 7)
```

Since our client, the Tourist Office of the City of Paris, wants an analysis to advise tourists on the types of restaurants they might find during their visits to Parisian neighborhoods, we have decided to filter the places that deal with only food and drink.

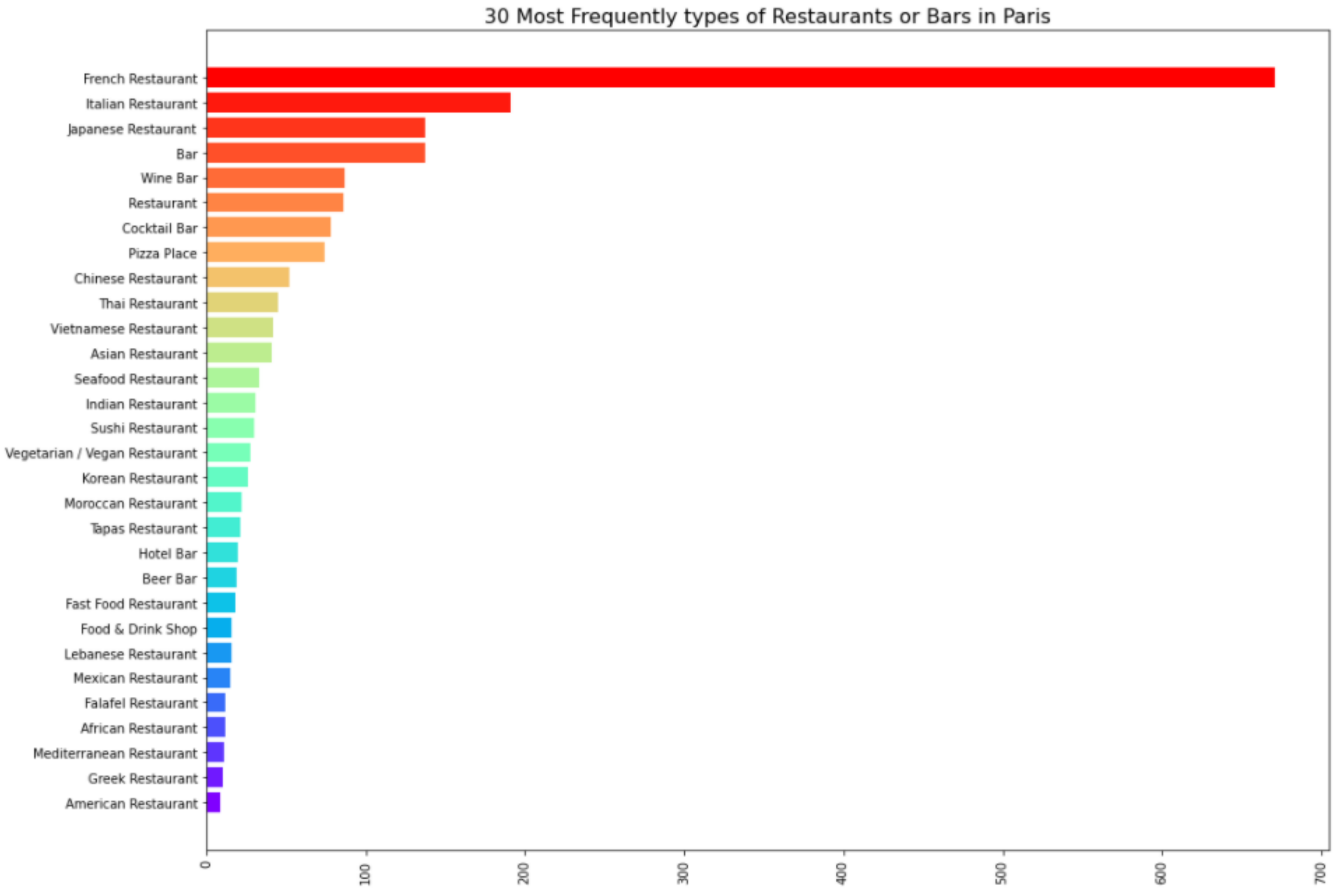
The new data frame created looks like this :

```
df_food
```

	Neighbourhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Palais-Royal	48.86466	2.336309	Udon Bistro Kunitoraya (Kunitoraya)	48.865884	2.336782	Udon Restaurant
2	Palais-Royal	48.86466	2.336309	Sanukiya	48.864713	2.333805	Udon Restaurant
3	Palais-Royal	48.86466	2.336309	Brasserie Réjane	48.865486	2.334824	Restaurant
4	Palais-Royal	48.86466	2.336309	Verjus Bar à Vins	48.866306	2.337471	Wine Bar
5	Palais-Royal	48.86466	2.336309	Restaurant Kunitoraya	48.866116	2.336467	Japanese Restaurant
...	...	...	...	...	...	...	...
2232	Charonne	48.85476	2.407430	Le Magnolia	48.858992	2.405873	French Restaurant
2233	Charonne	48.85476	2.407430	Domino's Pizza	48.852447	2.403890	Pizza Place
2234	Charonne	48.85476	2.407430	McDonald's	48.853252	2.410679	Fast Food Restaurant
2235	Charonne	48.85476	2.407430	Royal Kebab	48.853461	2.409690	Kebab Restaurant
2236	Charonne	48.85476	2.407430	Pizzeria du glacier de Venise	48.853790	2.411600	Pizza Place

2236 rows × 7 columns

By exploring the data, we realize that French restaurants are widely represented in Paris. Indeed :



STEP 4 - PERFORMING ONE-HOT ENCODING TO ANALYZE NEIGHBOURHOODS WITH THE PANDA LIBRARY

One-Hot Encoding is a technique which ensures that machine learning algorithms can process data. Namely, it converts categorical variables into the binary Boolean ones.

Below, an extration of the data frame created after applying the One-Hot Encoding function :

```
df_encoded = pd.get_dummies(df_food[['Venue Category']], prefix="", prefix_sep="")
df_encoded
```

	Afghan Restaurant	African Restaurant	Alsatian Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Auvergne Restaurant	Bar	Basque Restaurant	...	Tapas Restaurant	Thai Restaurant	Theme Restaurant	Tibetan Restaurant	Turkish Restaurant	Udon Restaurant	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Vietnamese Restaurant
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	1	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2232	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0
2233	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
2234	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
2235	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0
2236	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0

2236 rows × 102 columns

Then, I have used this data frame to create a new one grouping the categories of restaurants in each neighborhood of the city. Below, an extraction of this new data frame:

```
df_grouped = df_encoded.groupby('Neighbourhood').mean().reset_index()
df_grouped.head()
```

	Neighbourhood	Afghan Restaurant	African Restaurant	Alsatian Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Auvergne Restaurant	Bar	...	Tapas Restaurant	Thai Restaurant	Theme Restaurant	Tibetan Restaurant	Turkish Restaurant	Udon Restaurant	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Vietname Restaura
0	Amérique	0.0	0.0	0.0	0.0	0.0	0.000000	0.00000	0.0	0.000000	...	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0000
1	Archives	0.0	0.0	0.0	0.0	0.0	0.000000	0.02381	0.0	0.095238	...	0.071429	0.047619	0.0	0.0	0.0	0.0	0.023810	0.0	0.0000
2	Arsenal	0.0	0.0	0.0	0.0	0.0	0.000000	0.00000	0.0	0.032258	...	0.064516	0.064516	0.0	0.0	0.0	0.0	0.064516	0.0	0.0000
3	Arts-et-Métiers	0.0	0.0	0.0	0.0	0.0	0.018182	0.00000	0.0	0.054545	...	0.018182	0.018182	0.0	0.0	0.0	0.0	0.018182	0.0	0.0545
4	Auteuil	0.0	0.0	0.0	0.0	0.0	0.000000	0.00000	0.0	0.000000	...	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0000

5 rows × 103 columns

Finally, I have used the previously created data frame to create a new one with the 10 most common types of restaurants for each neighborhood in the city.

	Neighbourhood	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	Amérique	French Restaurant	Health Food Store	Wine Bar	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Fondue Restaurant
1	Archives	French Restaurant	Italian Restaurant	Bar	Japanese Restaurant	Cocktail Bar	Tapas Restaurant	Falafel Restaurant	Thai Restaurant	Pizza Place	Cajun / Creole Restaurant
2	Arsenal	French Restaurant	Vegetarian / Vegan Restaurant	Cocktail Bar	Italian Restaurant	Thai Restaurant	Tapas Restaurant	Wine Bar	Southwestern French Restaurant	Bar	Brazilian Restaurant
3	Arts-et-Métiers	French Restaurant	Wine Bar	Cocktail Bar	Chinese Restaurant	Italian Restaurant	Restaurant	Bar	Vietnamese Restaurant	Japanese Restaurant	Moroccan Restaurant
4	Auteuil	French Restaurant	Wine Bar	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Fondue Restaurant	Food
...	...	...	...	...	...	...	...	...	...	...	...
74	Sorbonne	French Restaurant	Bar	Wine Bar	Italian Restaurant	Lebanese Restaurant	Seafood Restaurant	Asian Restaurant	Chinese Restaurant	Cocktail Bar	Ethiopian Restaurant
75	Ternes	French Restaurant	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Hotel Bar	Moroccan Restaurant	American Restaurant	Asian Restaurant	Breton Restaurant	Cocktail Bar
76	Val-de-Grâce	French Restaurant	Bar	Asian Restaurant	Wine Bar	Pizza Place	Chinese Restaurant	Mexican Restaurant	Lebanese Restaurant	Beer Bar	Korean Restaurant
77	Villette	Bar	French Restaurant	Food Truck	Italian Restaurant	Fast Food Restaurant	Asian Restaurant	Beer Bar	Seafood Restaurant	Japanese Restaurant	Middle Eastern Restaurant
78	Vivienne	French Restaurant	Japanese Restaurant	Wine Bar	Cocktail Bar	Udon Restaurant	Korean Restaurant	Food & Drink Shop	Israeli Restaurant	English Restaurant	Taiwanese Restaurant

79 rows × 11 columns

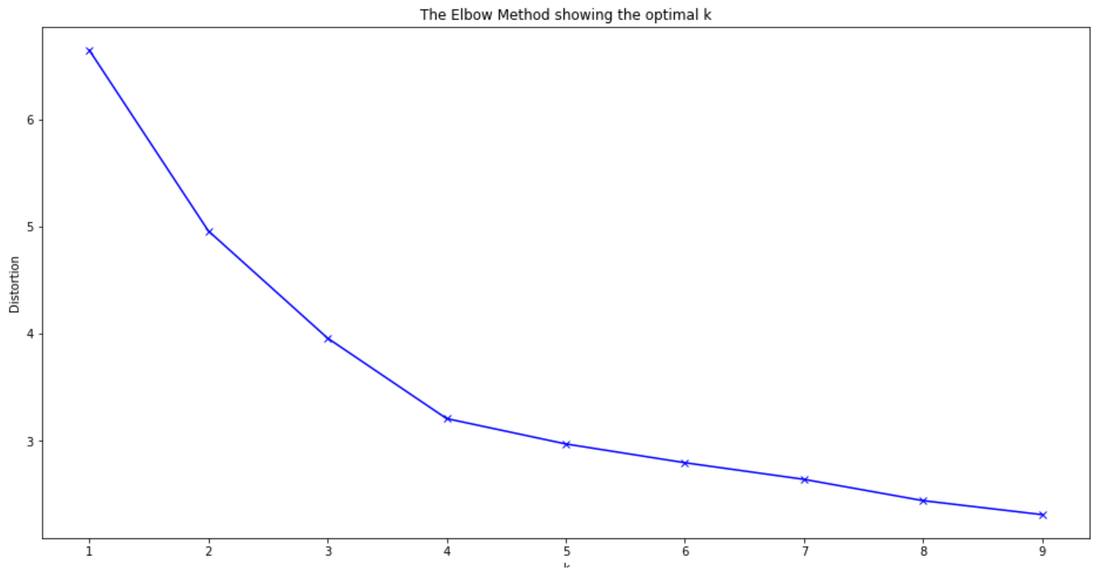
This last data frame allowed me to run an unsupervised machine learning algorithm, more precisely, a k-means clustering algorithm from the scikit-learn package.

STEP 05 - CLUSTERING K-MEANS WITH SKLEARN CLUSTER LIBRARY - TO DO

But before I could run the k-means clustering algorithm, I have used the ellbow method to set the value of the optimum k :

```
plt.figure(figsize=(16,8))
plt.plot(K, distortions, 'bx-')
plt.xlabel('k')
plt.ylabel('Distortion')
plt.title('The Elbow Method showing the optimal k')
```

```
Text(0.5, 1.0, 'The Elbow Method showing the optimal k')
```



The graph of the elbow method, shows that the optimal value of k is 4. So I chose k as being 4, to run the k-means clustering algorithm.

```
from sklearn.cluster import KMeans

k_clusters = 4

#drop the Neighbourhood column to work with numerical values only
df_k_clustering = df_grouped.drop('Neighbourhood', 1)

KM = KMeans(n_clusters=k_clusters, random_state=0)
```

After adding the label of the cluster to the data frame and after having merged it with the data frame containing the borough information, I get this data frame which will allow me to analyze the clusters that have been created :

```
df_venues_final = df_venues_sorted.join(borough_neighbourhoodDF.set_index('Neighbourhood'), on='Neighbourhood')
df_venues_final
```

	Cluster Labels	Neighbourhood	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	...	BoroughNumber	Borough	Latitude	Longitude	BoroughPerimetre	NeighNumber	NeighLatitude	NeighLong
0	2	Amérique	French Restaurant	Health Food Store	Wine Bar	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	...	19	Buttes-Chaumont	48.887076	2.384821	11253.182479	75	48.881638	2.39
1	3	Archives	French Restaurant	Italian Restaurant	Bar	Japanese Restaurant	Cocktail Bar	Tapas Restaurant	Falafel Restaurant	Thai Restaurant	...	3	Temple	48.862872	2.360001	4519.263648	11	48.859192	2.36
2	0	Arsenal	French Restaurant	Vegetarian / Vegan Restaurant	Cocktail Bar	Italian Restaurant	Thai Restaurant	Tapas Restaurant	Wine Bar	Southwestern French Restaurant	...	4	Hôtel-de-Ville	48.854341	2.357630	5420.908434	15	48.851585	2.36
3	3	Arts-et-Métiers	French Restaurant	Wine Bar	Cocktail Bar	Chinese Restaurant	Italian Restaurant	Restaurant	Bar	Vietnamese Restaurant	...	3	Temple	48.862872	2.360001	4519.263648	9	48.866470	2.35
4	2	Auteuil	French Restaurant	Wine Bar	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	...	16	Passy	48.860392	2.261971	17416.109657	61	48.850622	2.25
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
74	0	Sorbonne	French Restaurant	Bar	Wine Bar	Italian Restaurant	Lebanese Restaurant	Seafood Restaurant	Asian Restaurant	Chinese Restaurant	...	5	Panthéon	48.844443	2.350715	6239.195396	20	48.849045	2.34
75	0	Ternes	French Restaurant	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Hotel Bar	Moroccan Restaurant	American Restaurant	Asian Restaurant	...	17	Batignolles-Monceau	48.887327	2.306777	10775.579516	65	48.881178	2.28
76	3	Val-de-Grâce	French Restaurant	Bar	Asian Restaurant	Wine Bar	Pizza Place	Chinese Restaurant	Mexican Restaurant	Lebanese Restaurant	...	5	Panthéon	48.844443	2.350715	6239.195396	19	48.841684	2.34
77	3	Villette	Bar	French Restaurant	Food Truck	Italian Restaurant	Fast Food Restaurant	Asian Restaurant	Beer Bar	Seafood Restaurant	...	19	Buttes-Chaumont	48.887076	2.384821	11253.182479	73	48.887661	2.37
78	3	Vivienne	French Restaurant	Japanese Restaurant	Wine Bar	Cocktail Bar	Udon Restaurant	Korean Restaurant	Food & Drink Shop	Israeli Restaurant	...	2	Bourse	48.868279	2.342803	4554.104360	6	48.869100	2.33

79 rows × 22 columns

STEP 06 - CLUSTER ANALYSIS

So four groups of concentrations of types of restaurants were created by the K-Means model for the city of Paris. I have named these groups according to the frequency of the types of restaurants that appear the most among the first 3 most common venues.

Cluster 1 (0) - French and Italian Restaurants

```
cluster = 0
cluster_1 = df_venues_final.loc[df_venues_final['Cluster Labels'] == cluster, df_venues_final.columns[[1] + list(range(2, df_venues_final.shape[1]))]]
```

```
cluster_1_1stMostCommonVenue = cluster_1['1st Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
cluster_1_2ndMostCommonVenue = cluster_1['2nd Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
cluster_1_3rdMostCommonVenue = cluster_1['3rd Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
print(cluster_1_1stMostCommonVenue);print(cluster_1_2ndMostCommonVenue);print(cluster_1_3rdMostCommonVenue)
```

```
                   frequency
French Restaurant         36
                              frequency
Italian Restaurant                   17
Bar                                   4
Japanese Restaurant                   3
Wine Bar                              2
Pizza Place                           2
Vegetarian / Vegan Restaurant         2
Chinese Restaurant                    1
Food Truck                            1
Portuguese Restaurant                 1
Greek Restaurant                      1
Vietnamese Restaurant                 1
Corsican Restaurant                   1
                              frequency
Japanese Restaurant                   6
Italian Restaurant                    5
Wine Bar                              4
Pizza Place                           4
Bar                                   3
Vietnamese Restaurant                 2
Thai Restaurant                       2
Hotel Bar                             2
Restaurant                            2
Cocktail Bar                          2
Seafood Restaurant                    1
Moroccan Restaurant                   1
Vegetarian / Vegan Restaurant         1
Fast Food Restaurant                  1
```

```
cluster_1
```

	Neighbourhood	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	...	BoroughNumber	Borough	Latitude	Longitude	BoroughPerimetre	NeighNumber	NeighLatitu
2	Arsenal	French Restaurant	Vegetarian / Vegan Restaurant	Cocktail Bar	Italian Restaurant	Thai Restaurant	Tapas Restaurant	Wine Bar	Southwestern French Restaurant	Bar	...	4	Hôtel-de-Ville	48.854341	2.357630	5420.908434	15	48.8515
5	Batignolles	French Restaurant	Italian Restaurant	Bar	Restaurant	Pizza Place	Japanese Restaurant	Turkish Restaurant	Chinese Restaurant	Cocktail Bar	...	17	Batignolles-Monceau	48.887327	2.306777	10775.579516	67	48.8884
10	Chaillot	French Restaurant	Italian Restaurant	Japanese Restaurant	Chinese Restaurant	Seafood Restaurant	Bar	Cambodian Restaurant	Cantonese Restaurant	Food & Drink Shop	...	16	Passy	48.860392	2.261971	17416.109657	64	48.8684
11	Champs-Elysées	French Restaurant	Italian Restaurant	Japanese Restaurant	Peruvian Restaurant	Hotel Bar	Thai Restaurant	Cocktail Bar	Seafood Restaurant	Alsatian Restaurant	...	8	Élysée	48.872721	2.312554	7880.533268	29	48.8670
13	Chaussée-d'Antin	French Restaurant	Italian Restaurant	Vietnamese Restaurant	Chinese Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	Seafood Restaurant	Fondue Restaurant	...	9	Opéra	48.877164	2.337458	6471.588290	34	48.8735
15	Combat	French Restaurant	Italian Restaurant	Restaurant	Wine Bar	Thai Restaurant	Bar	Latin American Restaurant	Moroccan Restaurant	Japanese Restaurant	...	19	Buttes-Chaumont	48.887076	2.384821	11253.182479	76	48.8786
20	Europe	French Restaurant	Pizza Place	Thai Restaurant	Wine Bar	Sushi Restaurant	Italian Restaurant	Gluten-free Restaurant	Korean Restaurant	Lebanese Restaurant	...	8	Élysée	48.872721	2.312554	7880.533268	32	48.8781
21	Faubourg-Montmartre	French Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Restaurant	Cocktail Bar	Doner Restaurant	Pizza Place	Japanese Restaurant	...	9	Opéra	48.877164	2.337458	6471.588290	35	48.8739
22	Faubourg-du-Roule	French Restaurant	Italian Restaurant	Japanese Restaurant	Asian Restaurant	Hotel Bar	Scandinavian Restaurant	Seafood Restaurant	Fast Food Restaurant	Food	...	8	Élysée	48.872721	2.312554	7880.533268	30	48.8741
64	Saint-Germain-l'Auxerrois	French Restaurant	Italian Restaurant	Japanese Restaurant	Udon Restaurant	Bar	Food Truck	Eastern European Restaurant	Empanada Restaurant	English Restaurant	...	1	Louvre	48.862563	2.336443	6054.936862	1	48.8606
65	Saint-Gervais	French Restaurant	Italian Restaurant	Wine Bar	Falafel Restaurant	Thai Restaurant	Tapas Restaurant	Scandinavian Restaurant	Burgundian Restaurant	Israeli Restaurant	...	4	Hôtel-de-Ville	48.854341	2.357630	5420.908434	14	48.8557
66	Saint-Lambert	French Restaurant	Italian Restaurant	Thai Restaurant	Japanese Restaurant	Sports Bar	Vietnamese Restaurant	Korean Restaurant	Lebanese Restaurant	Middle Eastern Restaurant	...	15	Vaugirard	48.840085	2.292826	13678.798315	57	48.8342
67	Saint-Merri	French Restaurant	Portuguese Restaurant	Italian Restaurant	Szechuan Restaurant	Restaurant	Cocktail Bar	Salon / Barbershop	Seafood Restaurant	Snack Place	...	4	Hôtel-de-Ville	48.854341	2.357630	5420.908434	13	48.8585
68	Saint-Thomas-d'Aquin	French Restaurant	Italian Restaurant	Bar	Vietnamese Restaurant	Hotel Bar	Peruvian Restaurant	Pizza Place	Restaurant	Cocktail Bar	...	7	Palais-Bourbon	48.856174	2.312188	8099.424883	25	48.8552
69	Saint-Victor	French Restaurant	Wine Bar	Italian Restaurant	Japanese Restaurant	Falafel Restaurant	Bar	Portuguese Restaurant	Turkish Restaurant	Peruvian Restaurant	...	5	Panthéon	48.844443	2.350715	6239.195396	17	48.8476
72	Sainte-Marguerite	French Restaurant	Italian Restaurant	Bar	Wine Bar	Thai Restaurant	Ethiopian Restaurant	Korean Restaurant	Beer Bar	Restaurant	...	11	Popincourt	48.859059	2.380058	8282.011886	44	48.8520
74	Sorbonne	French Restaurant	Bar	Wine Bar	Italian Restaurant	Lebanese Restaurant	Seafood Restaurant	Asian Restaurant	Chinese Restaurant	Cocktail Bar	...	5	Panthéon	48.844443	2.350715	6239.195396	20	48.8490
75	Ternes	French Restaurant	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Hotel Bar	Moroccan Restaurant	American Restaurant	Asian Restaurant	Breton Restaurant	...	17	Batignolles-Monceau	48.887327	2.306777	10775.579516	65	48.8811

36 rows × 21 columns

Cluster 2 (1) - Pizza Place and International Cuisine

```
cluster_2_1stMostCommonVenue = cluster_2['1st Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
cluster_2_2ndMostCommonVenue = cluster_2['2nd Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
cluster_2_3rdMostCommonVenue = cluster_2['3rd Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
print(cluster_2_1stMostCommonVenue);print(cluster_2_2ndMostCommonVenue);print(cluster_2_3rdMostCommonVenue)
```

```
             frequency
Pizza Place          1
          frequency
Wine Bar          1
            frequency
Food Truck          1
```

```
cluster_2
```

	Neighbourhood	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	...	BoroughNumber	Borough	Latitude	Longitude	BoroughPerimetre	NeighNumber	NeighLatitude	NeighLongitude
53	Porte-Dauphine	Pizza Place	Wine Bar	Food Truck	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	...	16	Passy	48.860392	2.261971	17416.109657	63	48.871766	2.271918

1 rows × 21 columns

Cluster 3 (2) - French Restaurants and Wine Bars

```
cluster_3_1stMostCommonVenue = cluster_3['1st Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
cluster_3_2ndMostCommonVenue = cluster_3['2nd Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
cluster_3_3rdMostCommonVenue = cluster_3['3rd Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
print(cluster_3_1stMostCommonVenue);print(cluster_3_2ndMostCommonVenue);print(cluster_3_3rdMostCommonVenue)
```

```
                   frequency
French Restaurant          6
                    frequency
Wine Bar                    2
Health Food Store           1
Italian Restaurant          1
Snack Place                 1
Asian Restaurant            1
                             frequency
Wine Bar                             3
Eastern European Restaurant          2
Japanese Restaurant                  1
```

```
cluster_3
```

	Neighbourhood	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	...	BoroughNumber	Borough	Latitude	Longitude	BoroughPerimetre	NeighNumber	NeighLatitude
0	Amérique	French Restaurant	Health Food Store	Wine Bar	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	...	19	Buttes-Chaumont	48.887076	2.384821	11253.182479	75	48.881638
4	Auteuil	French Restaurant	Wine Bar	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Fondue Restaurant	...	16	Passy	48.860392	2.261971	17416.109657	61	48.850622
6	Bel-Air	French Restaurant	Wine Bar	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Fondue Restaurant	...	12	Reuilly	48.834974	2.421325	24089.666298	45	48.837996
17	Ecole-Militaire	French Restaurant	Asian Restaurant	Wine Bar	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	...	7	Palais-Bourbon	48.856174	2.312188	8099.424883	27	48.850359
32	Invalides	French Restaurant	Italian Restaurant	Japanese Restaurant	Cocktail Bar	Restaurant	Vegetarian / Vegan Restaurant	Food	Doner Restaurant	Eastern European Restaurant	...	7	Palais-Bourbon	48.856174	2.312188	8099.424883	26	48.858515
41	Muette	French Restaurant	Snack Place	Wine Bar	Food & Drink Shop	Doner Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	...	16	Passy	48.860392	2.261971	17416.109657	62	48.863275

6 rows × 21 columns

Cluster 4 (3) - French Restaurants

```
cluster_4_1stMostCommonVenue = cluster_4['1st Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
cluster_4_2ndMostCommonVenue = cluster_4['2nd Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
cluster_4_3rdMostCommonVenue = cluster_4['3rd Most Common Venue'].value_counts()[0:topn].to_frame(name='frequency')
print(cluster_4_1stMostCommonVenue);print(cluster_4_2ndMostCommonVenue);print(cluster_4_3rdMostCommonVenue)
```

```
                     frequency
French Restaurant           23
Bar                          5
Japanese Restaurant          3
Indian Restaurant            2
Cocktail Bar                 2
Italian Restaurant           1
                     frequency
French Restaurant            8
Bar                          7
Japanese Restaurant          7
Wine Bar                     3
Italian Restaurant           3
Restaurant                   2
Asian Restaurant             2
Chinese Restaurant           1
Cocktail Bar                 1
Sushi Restaurant             1
Thai Restaurant              1
                               frequency
Italian Restaurant                     7
Wine Bar                               7
Japanese Restaurant                    5
Cocktail Bar                           4
French Restaurant                      3
Chinese Restaurant                     2
Restaurant                             2
Food Truck                             1
Portuguese Restaurant                  1
Vegetarian / Vegan Restaurant          1
Asian Restaurant                       1
Bar                                    1
African Restaurant                     1
```

```
cluster_4
```

	Neighbourhood	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	...	BoroughNumber	Borough	Latitude	Longitude	BoroughPerimetre
1	Archives	French Restaurant	Italian Restaurant	Bar	Japanese Restaurant	Cocktail Bar	Tapas Restaurant	Falafel Restaurant	Thai Restaurant	Pizza Place	...	3	Temple	48.862872	2.360001	4519.263648
3	Arts-et-Métiers	French Restaurant	Wine Bar	Cocktail Bar	Chinese Restaurant	Italian Restaurant	Restaurant	Bar	Vietnamese Restaurant	Japanese Restaurant	...	3	Temple	48.862872	2.360001	4519.263648
7	Belleville	Bar	French Restaurant	Japanese Restaurant	Italian Restaurant	Pizza Place	African Restaurant	Chinese Restaurant	Cocktail Bar	Indian Restaurant	...	20	Ménilmontant	48.863461	2.401188	10704.940486
8	Bercy	French Restaurant	Italian Restaurant	Japanese Restaurant	Restaurant	Bar	Beer Bar	Cambodian Restaurant	Chinese Restaurant	Doner Restaurant	...	12	Reuilly	48.834974	2.421325	24089.666298
9	Bonne-Nouvelle	Cocktail Bar	French Restaurant	Wine Bar	Japanese Restaurant	Restaurant	Italian Restaurant	Chinese Restaurant	Bar	Thai Restaurant	...	2	Bourse	48.868279	2.342803	4554.104360
12	Charonne	Bar	Japanese Restaurant	Portuguese Restaurant	Pizza Place	Fast Food Restaurant	French Restaurant	Brazilian Restaurant	Indian Restaurant	Hawaiian Restaurant	...	20	Ménilmontant	48.863461	2.401188	10704.940486
14	Clignancourt	French Restaurant	Bar	Italian Restaurant	Pizza Place	Restaurant	Wine Bar	Seafood Restaurant	Arepa Restaurant	Asian Restaurant	...	18	Buttes-Montmartre	48.892569	2.348161	9916.464176
16	Croulebarbe	French Restaurant	Sushi Restaurant	Italian Restaurant	Bar	Ramen Restaurant	Restaurant	Indian Restaurant	Cocktail Bar	Ethiopian Restaurant	...	13	Gobelins	48.828388	2.362272	11546.546526
18	Enfants-Rouges	French Restaurant	Wine Bar	Japanese Restaurant	Italian Restaurant	Vietnamese Restaurant	Cocktail Bar	Bar	Korean Restaurant	Restaurant	...	3	Temple	48.862872	2.360001	4519.263648
19	Epinettes	French Restaurant	Restaurant	Japanese Restaurant	Turkish Restaurant	Pizza Place	Bar	Ethiopian Restaurant	Sushi Restaurant	Hotel Bar	...	17	Batignolles-Monceau	48.887327	2.306777	10775.579516
23	Folie-Méricourt	French Restaurant	Bar	Restaurant	Wine Bar	Pizza Place	African Restaurant	Chinese Restaurant	Italian Restaurant	Juice Bar	...	11	Popincourt	48.859059	2.380058	8282.011886
24	Gaillon	Japanese Restaurant	French Restaurant	Wine Bar	Italian Restaurant	Ramen Restaurant	Korean Restaurant	Udon Restaurant	Asian Restaurant	Cocktail Bar	...	2	Bourse	48.868279	2.342803	4554.104360

70	Saint-Vincent-de-Paul	Indian Restaurant	French Restaurant	African Restaurant	Italian Restaurant	Japanese Restaurant	Sports Bar	Food & Drink Shop	Breton Restaurant	Israeli Restaurant	...	10	Entrepôt	48.876130	2.360728	6739.375055
71	Sainte-Avoie	French Restaurant	Restaurant	Chinese Restaurant	Wine Bar	Italian Restaurant	Vietnamese Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Health Food Store	...	3	Temple	48.862872	2.360001	4519.263648
73	Salpêtrière	Indian Restaurant	French Restaurant	Italian Restaurant	Corsican Restaurant	Chinese Restaurant	Bar	Mediterranean Restaurant	Sushi Restaurant	Wine Bar	...	13	Gobelins	48.828388	2.362272	11546.546526
76	Val-de-Grâce	French Restaurant	Bar	Asian Restaurant	Wine Bar	Pizza Place	Chinese Restaurant	Mexican Restaurant	Lebanese Restaurant	Beer Bar	...	5	Panthéon	48.844443	2.350715	6239.195396
77	Villette	Bar	French Restaurant	Food Truck	Italian Restaurant	Fast Food Restaurant	Asian Restaurant	Beer Bar	Seafood Restaurant	Japanese Restaurant	...	19	Buttes-Chaumont	48.887076	2.384821	11253.182479
78	Vivienne	French Restaurant	Japanese Restaurant	Wine Bar	Cocktail Bar	Udon Restaurant	Korean Restaurant	Food & Drink Shop	Israeli Restaurant	English Restaurant	...	2	Bourse	48.868279	2.342803	4554.104360

36 rows × 21 columns

STEP 07 - CREATION OF A CLUSTERS MAP WITH THE FOLIUM LIBRARY

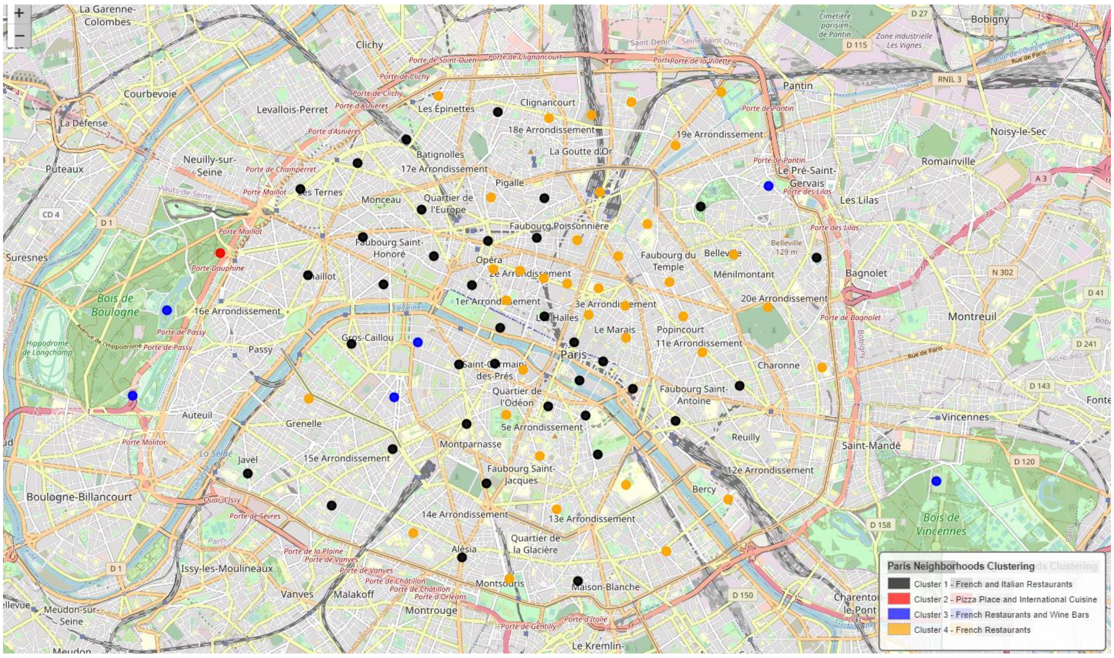
The geographical coordinates of Paris were extracted using GeoPy libray in Python.

```
address = 'Paris'

geolocator = Nominatim(user_agent="Paris_Explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print(f'The geograpical coordinates of Paris are {latitude}, {longitude}.')
```

```
The geograpical coordinates of Paris are 48.8566969, 2.3514616.
```

The geographic coordinates will be used to draw the map of Paris with arrondissements colored according to the types of restaurants that are present.



Discussion

As I indicated previously, French restaurants are widely represented in Paris. Indeed, there are nearly 680.

In 3 out of 4 clusters reproduced by the K-Means model, there is a large number of French restaurants in the first most frequent place of each of the districts.

Only group 2 (Pizza Place and International Cuisine) does not have a French restaurant. This cluster is made up solely of the Porte Dauphine district located in the 16th arrondissement of Paris.

We also note that only the Picpus district, located in the 12th arrondissement of Paris, does not have a restaurant.

Conclusion

The objective of the project presented at the beginning of this report has been achieved. Indeed, my company has delivered as planned to the Tourist Office of the city of Paris a tool which now allows them to advise tourists on the types of restaurants they could find during their visits to the Parisian districts.

Acknowledgement, References & Links

ACKNOWLEDGEMENT

This report refers to the lab projects of the IBM Data Science Professional Certificate course on Coursera.

To carry out this project, I recovered a large part of the Python code used in the different labs of Course 9.

I was also inspired by the many Notebooks published by students who had thought about the subject before me.

REFERENCES

Source : https://en.wikipedia.org/wiki/Arrondissements_of_Paris

Source : https://www.parisinsidersguide.com/paris-neighborhoods.html

Source : https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris

LINKS

The notebook with the code for this project, as well as the report, can be found in my github repository.