

Market Research for Opening New Vietnamese Restaurant in Paris, France

Author: Tien LE

Course: Applied Data Science Capstone - IBM

Paris, 16/12/2019

Outline

- Introduction
- Data Description
- Methodology
- Results
- Discussion
- Conclusion and Perspectives
- References

Introduction

- Paris is the French capital and most interesting city of France
- 2 140 526 residents in 105 square kilometers [1]
- Over 300 000 Vietnamese people in Paris [7]
- **Business problem**
 - Finding the good districts from 20 districts in Paris to open a Vietnamese restaurant.

Data Description

- Neighborhood data of 20 districts in Paris, France
 - Paris Arrondissements & Neighborhoods Map [10]
 - Arrondissements in Paris, France [11]
- Coordinates (latitude, longitude) of each district
 - Using tool **geopy**
- The most famous venues and their relative locations in the radius 500 meters from the center of each
 - Using **Foursquare API**

Methodology (1)

- Building Paris neighborhood data (Postcode, Neighborhood)
- Building the coordinates of all districts in Paris, France using tool **geopy**.
- Using **Foursquare API** to generate to the most famous venues and the relative locations.
- Analyzing, exploring the data and finding the optimal number of clusters for K-Means clustering using the **Elbow** method

Methodology (2)

- Using **K-Means** to cluster the neighborhoods in the city of Paris **based on the top 10 venues** for each neighborhood district.
- Visualizing geographic details of Paris and its district neighborhoods which are superimposed on top using tool **folium**
- Analyzing the clustering result and then proposing some suggestions about the good districts to open Vietnamese Restaurant in Paris
- Giving some perspectives to enhance the performances

Results

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

Results

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

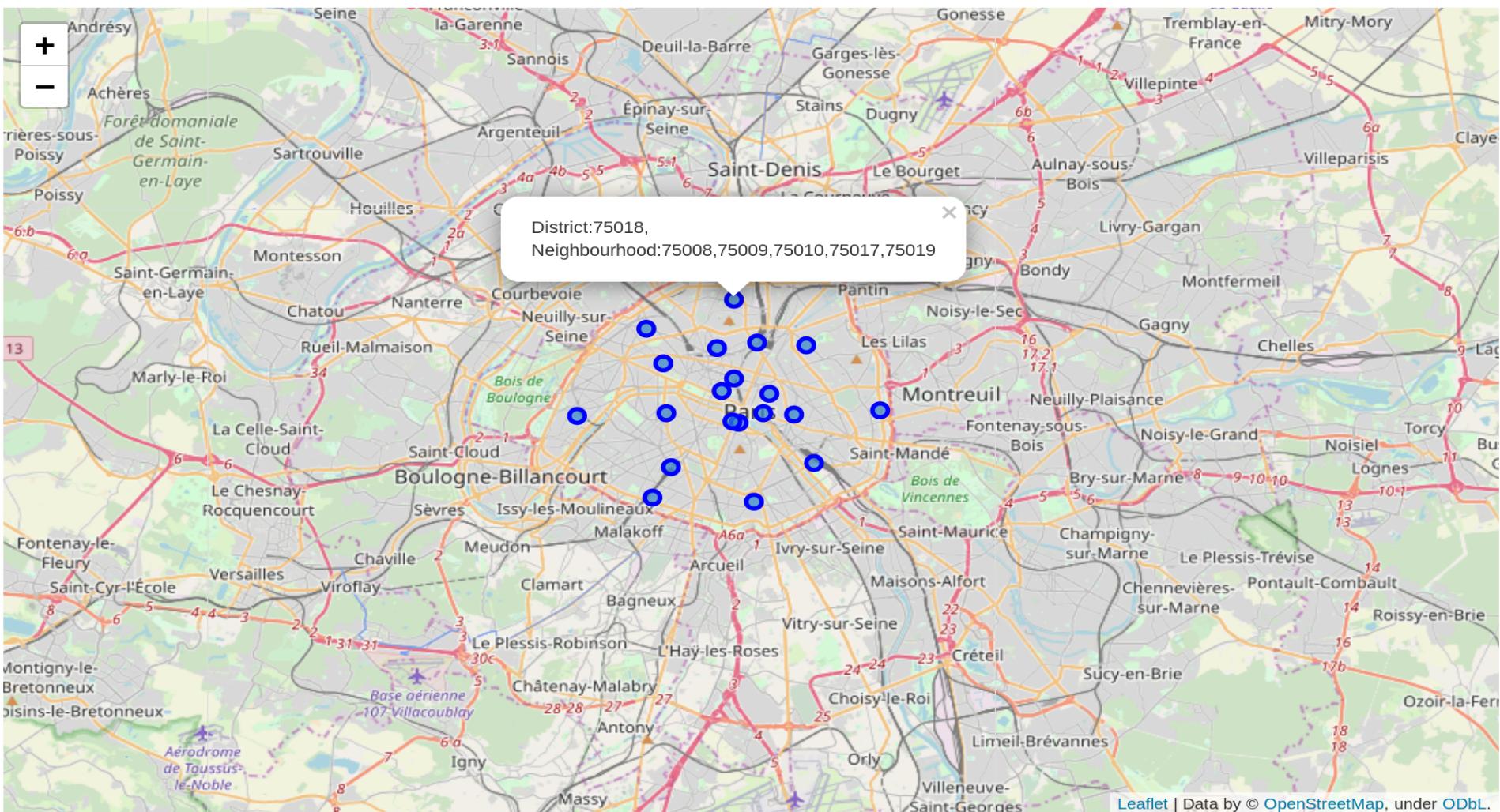
	postcode	latitude	longitude
0	75001	48.863554	2.338856
1	75002	48.867418	2.344256
2	75003	48.862607	2.360211
3	75004	48.856004	2.357028
4	75005	48.852752	2.346343
5	75006	48.853537	2.343370
6	75007	48.855913	2.313839
7	75008	48.872385	2.312707
8	75009	48.877355	2.336856
9	75010	48.879201	2.354391
10	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

Results

	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

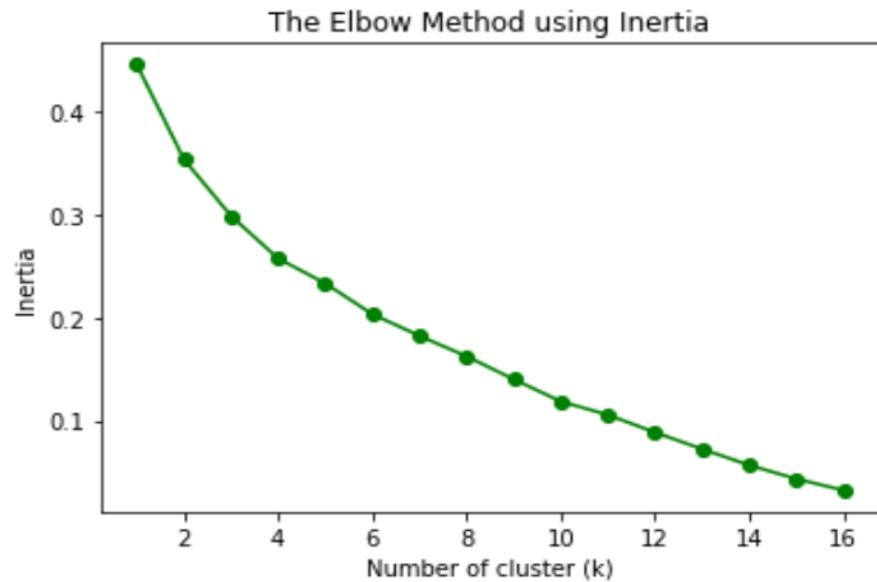
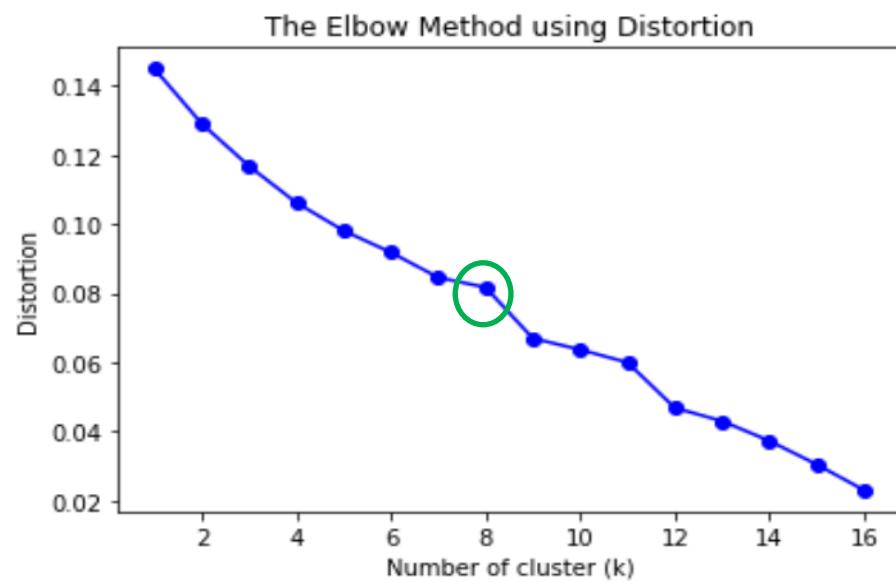
Table 3: Example of the top-100 venues in radius 500 metres from the centre of each district

Results



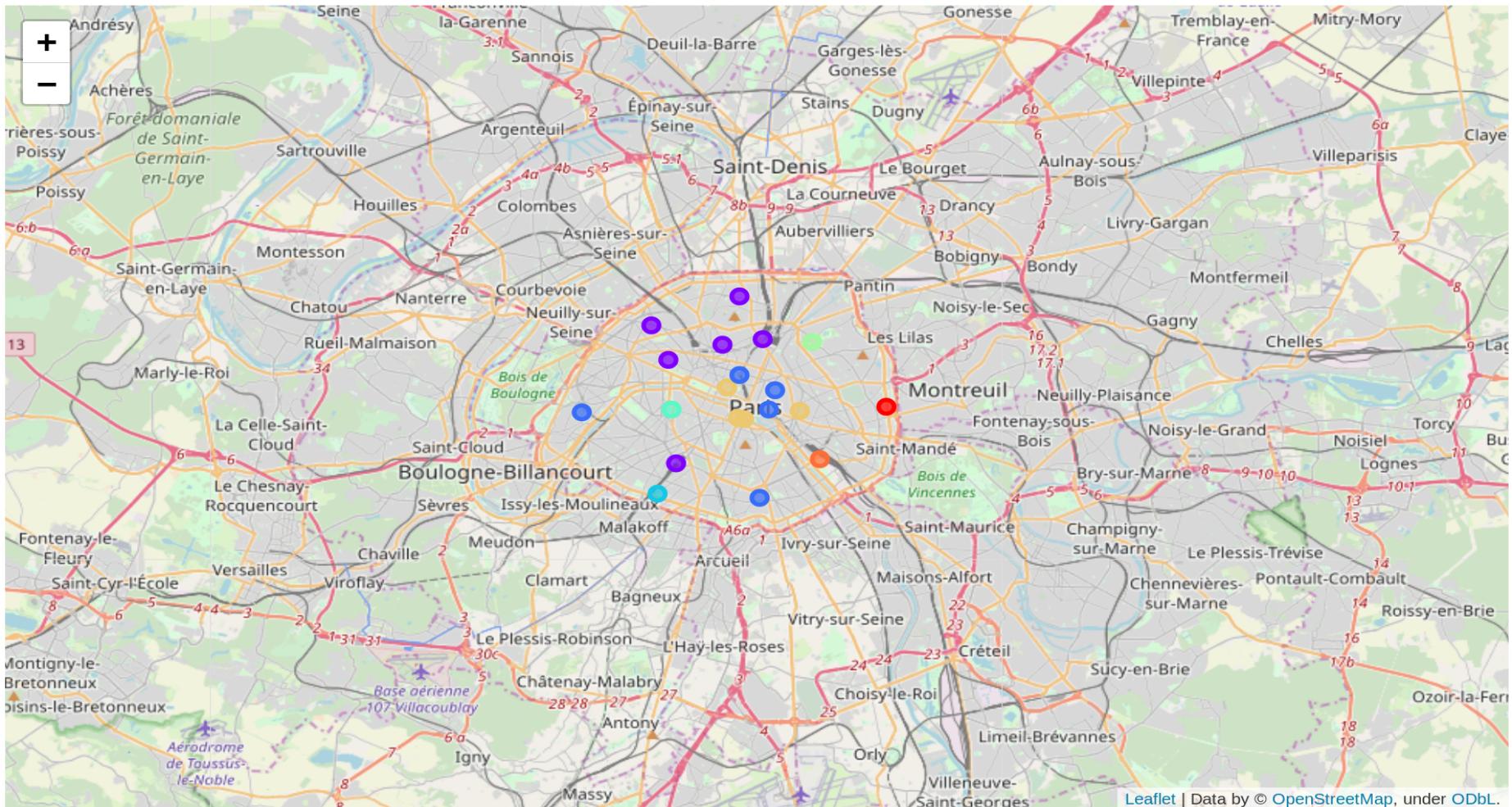
Visualization of geographic details of Paris
and its district neighborhood which are superimposed on top

Determining the optimal value k



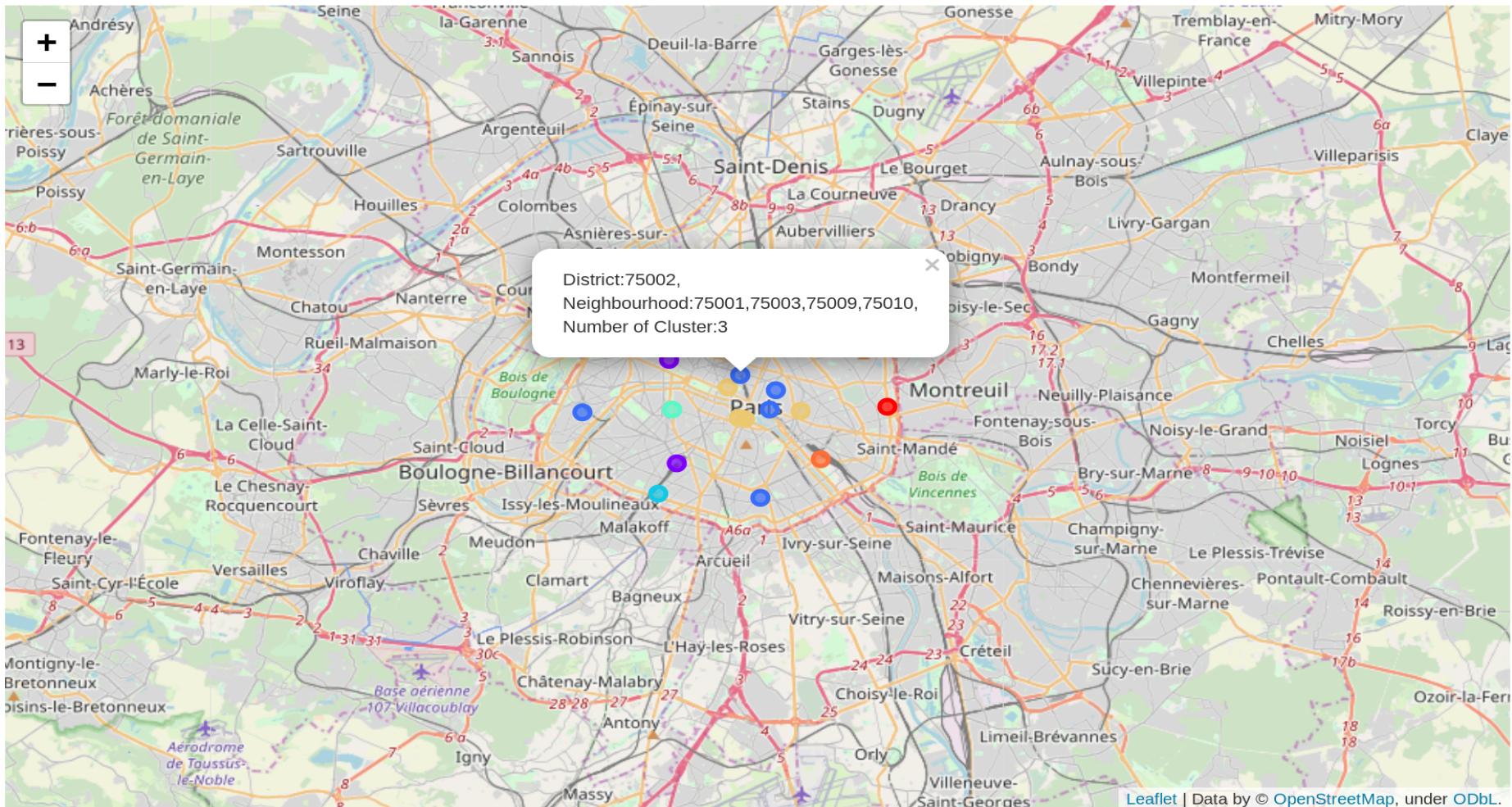
- The charts of the Elbow Method using the distortion or the inertia
- Difficult to determine the Elbow point using the inertia.
- Applying the distortion with the number of clusters, we choose the optimal number of clusters for K-Means is **8**.

Clustering Neighborhoods of Paris



Using the optimal value k, we launch K-Means technique to cluster the neighborhood into **8 clusters**

Clustering Neighborhoods of Paris



Using the optimal value k, we launch K-Means technique to cluster the neighborhood into **8 clusters**

Result - 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

Obtained Result - Cluster 3

Discussion (1)

- From above result, we obtain **8 clusters** that contain of the various districts **based on top-10 venues** for each neighborhood.
- **French Restaurant** is the **first most common venue** in most of districts in Paris.
- When reviewing the clusters, we could see that the Vietnamese restaurant in **cluster 3. Vietnamese Restaurant** is the **2nd** and the **10th** most common venues in **District 13** and **District 2**, respectively.

Discussion (2)

- **Districts in cluster 3**, could be **good potential candidates**.
- 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**.

Conclusion (1)

- First, we built the **neighborhood data of 20 districts in Paris** based on the information in [10, 11].
- Second, we use tool **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.

Conclusion (2)

- Fourth, **to visualize geographic details** of Paris and its district neighborhoods, we use tool **folium**
- 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 an information map where the districts in Paris are clustered according to the venue density.

Conclusion (3)

- Finally, we proposed some good districts to the investors **depending on their requirements and conditions**
 - District which have had top-10 Vietnamese restaurants → **District 13 or District 2**
 - Other good potential districts: **District 3, 4 and 16**

Perspectives (1)

- To enhance the features of district, we could **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, etc.

Perspectives (2)

- To **improve the performance** of clustering result, we could do with other algorithms, for instance,
 - **Fuzzy c-means** method
 - **DBSCAN**: Density-based clustering Hierarchical K-Means Clustering
 - **HCPC**: Hierarchical clustering on principal components
 - **Deep Learning Models** To see more detail, please see in "A Survey of Clustering With Deep Learning: From the Perspective of Network Architecture" (2018) -
<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8412085>

References

- [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] "Île-de-France- Portrait of the Region- Key figures (in French)". Regional Council of the Île-de-France. Retrieved 24 November 2018.
- [7] 10 Top Vietnamese Restaurants in Paris, <https://www.travelmag.com/articles/vietnamese-restaurants-paris/>
- [8] The tutorials in course "Applied Data Science Capstone",
<https://www.coursera.org/learn/applied-data-science-capstone/>
- [9] 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/>
- [10] Paris Arrondissements & Neighborhoods Map, <https://parismap360.com/paris-arrondissement-map#.XfVpqtEo91I>
- [11] Arrondissements in Paris, France,
<https://francetravelplanner.com/go/paris/areas/arrondismt.html>
- [12] Elbow Method for optimal value of k in KMeans, <https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/>

Thank You For Your Attention!