Paris: a world cuisine capital ?

Coursera Applied Data Science Capstone Project by IBM

A.MARTAN, 25/01/2020

Paris: a world cuisine capital ?

- Paris : world famous city for its French cuisine but also highly multicultural
- Idea :
  - Clustering Parisian neighborhoods based on the cuisine served in their restaurants
- Stakeholders:
  - Restaurant owners : opening a restaurant in the best spot
  - Tourists : wandering in an area with regional cuisine
  - Public institutions : highlighting boroughs with key characteristics

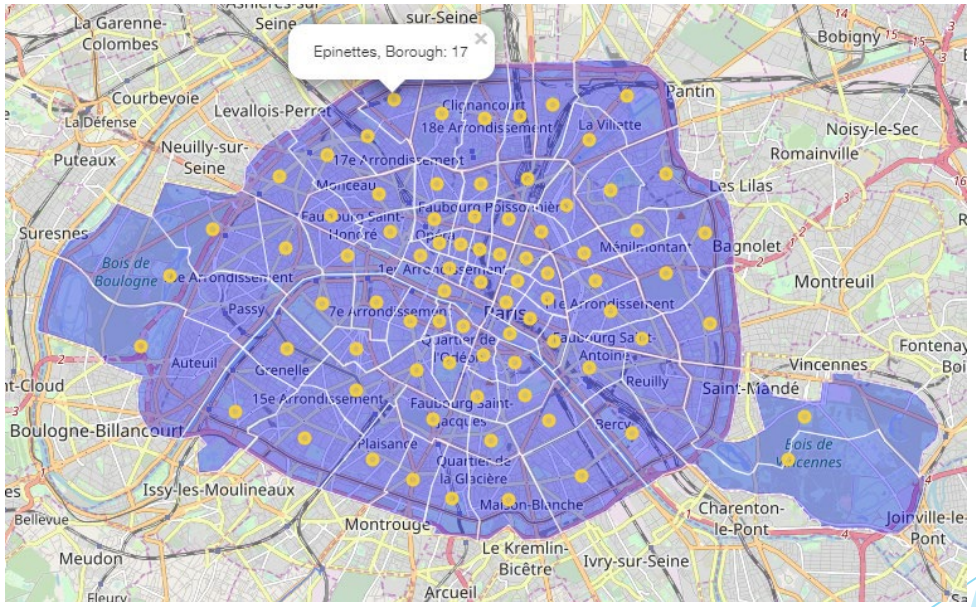
Data sources

- **Paris Data**
  - Link : https://opendata.paris.fr/explore/dataset/quartier_paris/informatio
  - License : Open Database License (ODbL)
  - Providing the administrative districts of Paris, files available in CSV and GeoJSON

- **Foursquare Places API**
  - Link : https://developer.foursquare.com/docs
  - Location data provider based on REST API: providing venues nearby a defined place

Data cleaning

- Extracting main information from retrieved dataset:
  - Name of the neighborhood, localization and shape
  - Category of venues
- Removing irrelevant information:
  - Surfaces, national ID…
  - Broad categories
- Renaming detailed categories



Data selection

Retrieved data from sources:

```
Venue Category
French Restaurant       1831
Italian Restaurant       558
Bakery                   388
Café                     380
Japanese Restaurant      363
                         ...
English Restaurant         1
Fondue Restaurant          1
Provençal Restaurant       1
German Restaurant          1
Cuban Restaurant           1
Name: Venue, Length: 123, dtype: int64
```

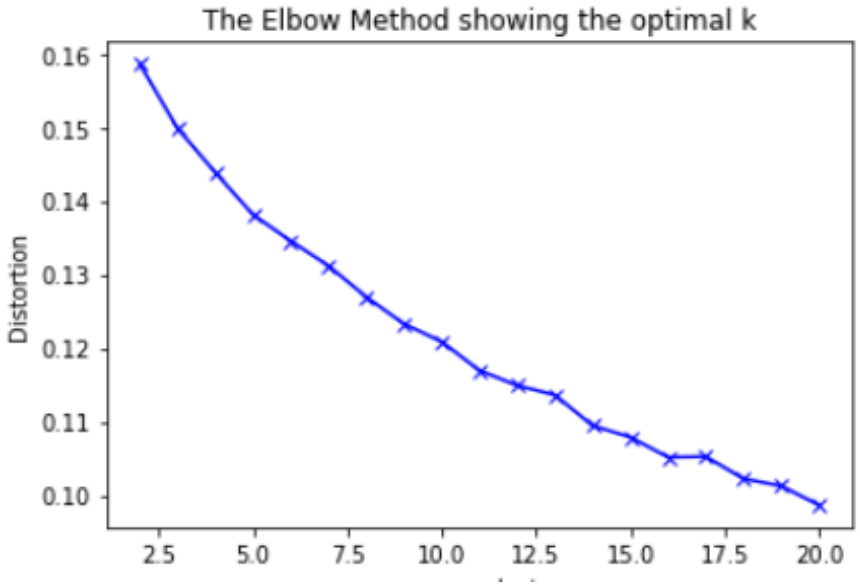
Data after cleaning:

```
Venue Category
Italian Restaurant       562
Japanese Restaurant      475
Brasserie                391
Bakery                   388
Pizza Place              220
                         ...
Cuban Restaurant           1
German Restaurant          1
Fondue Restaurant          1
Venezuelan Restaurant      1
Provençal Restaurant       1
Name: Venue, Length: 81, dtype: int64
```

Before : 6991 restaurants into 123 categories

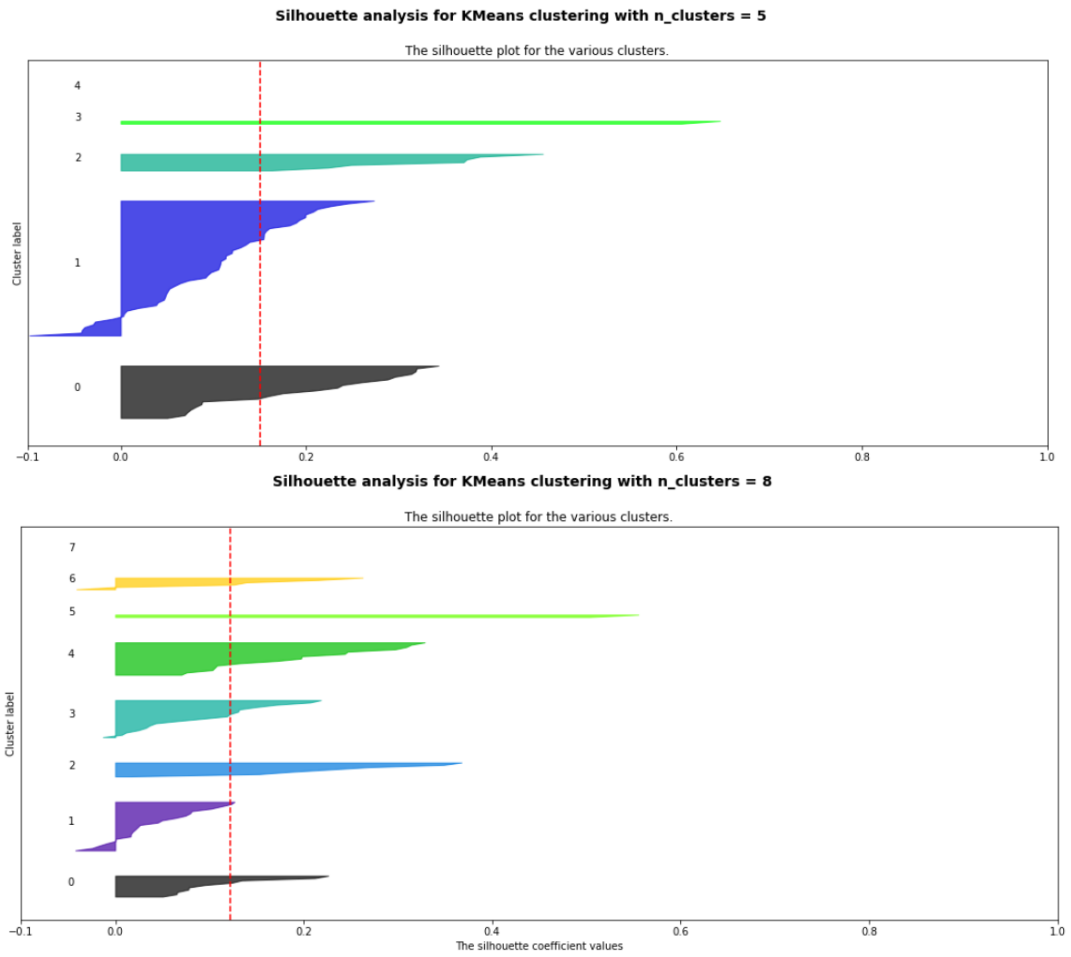
After : 4008 restaurants into 81 categories

K-means algorithm – Finding K
The Elbow Method

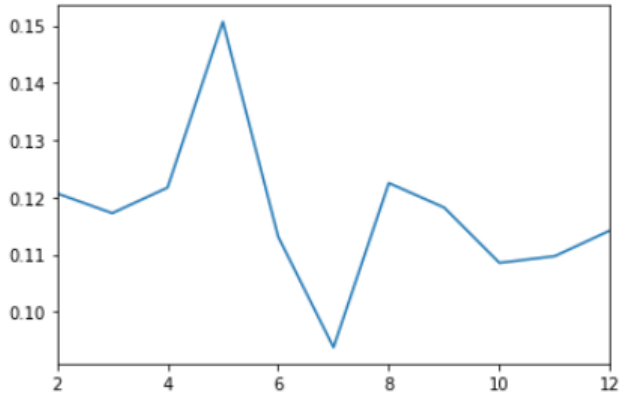


- Expected : plotting a graph with clear disruption
- Obtained : slowly decreasing graph
- Results : Method not reliable

K-means algorithm – Finding K
The Silhouette Method

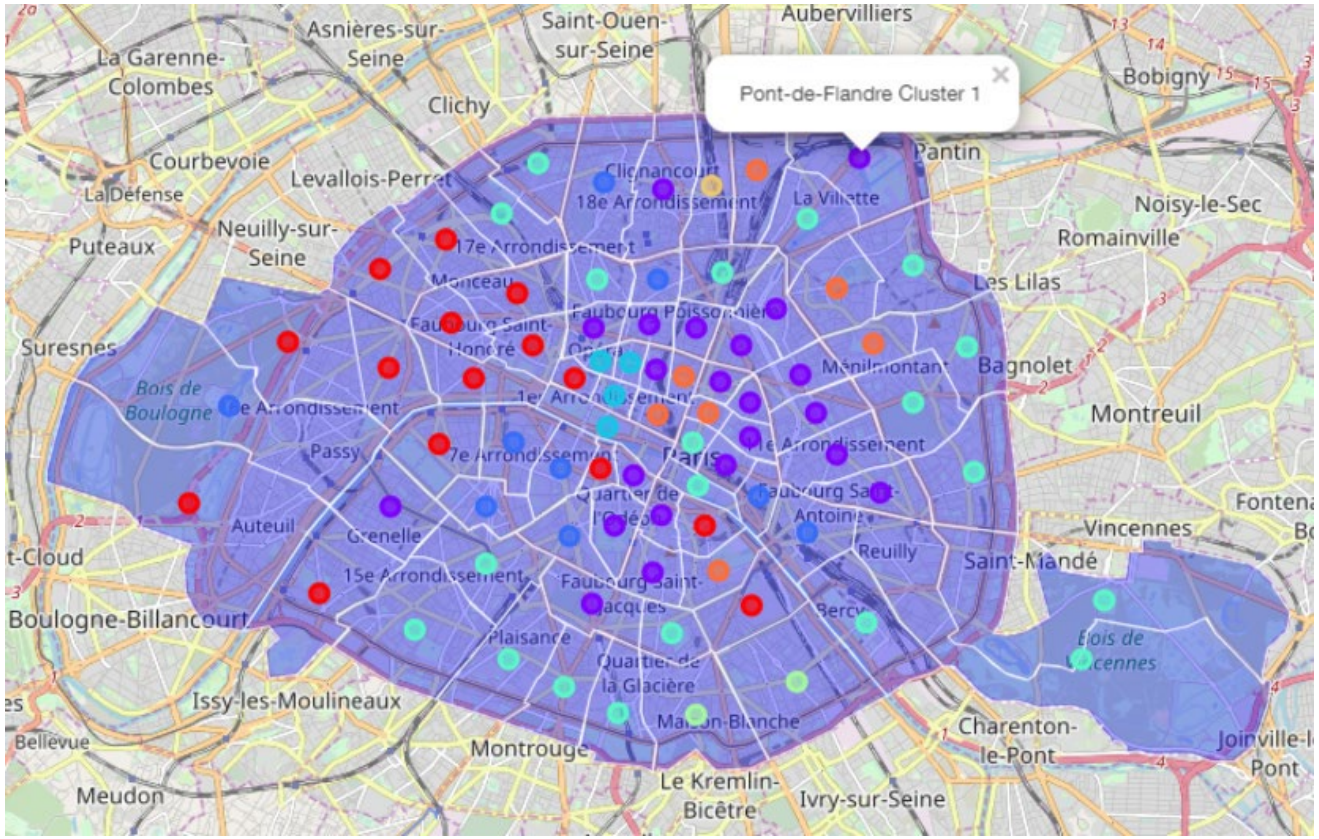


- Average score :



- Compromise between:
  - Incorrect assignment
  - Similar cluster size
  - Good average silhouette score

Map of Paris into 8 clusters



Cluster description

Cluster	Main region of cuisine
0	Italian – Japanese
1	French
2	Italian – French
3	Japanese
4	Japanese – French
5	Asian
6	Fast-Food
7	Mix

Conclusion and future direction

- Mapped the 80 Parisian neighborhoods into only 8 clusters
- Accuracy to be improved:
  - Retrieving data specialized in Restaurants with more detailed categories
  - Restricting venues inside neighborhoods, not nearby (avoid collision)
- Ideas to investigate:
  - Adding socio-economical data (average income, real-estate price…)
  - Comparing other city around the world and compare them