IBM
Applied Data Science Capstone Project

ANDI ZHANG

Hypothetical Business Problem

A French person who has been living in Japan for years noticed that there is an opportunity of introducing the business model of café-bookstores, a combination of coffee shop and bookstore, to the city center of Paris (1-9 arrondissement), as there are not many venues of the same kind in the city.

The assumption is that we can use unsupervised machine learning model to create clusters of arrondissements in Paris based on the frequency of each category of venues. Ideally, the café-bookstore should be in the cultural center of the city being close to similar spots such as bookstores, coffee shops, galleries, theatres, museums, and cinemas, etc.

Audience: People who might have the similar idea of starting his/her own business in Paris.

Data

To perform the analysis, we have collected the following information:
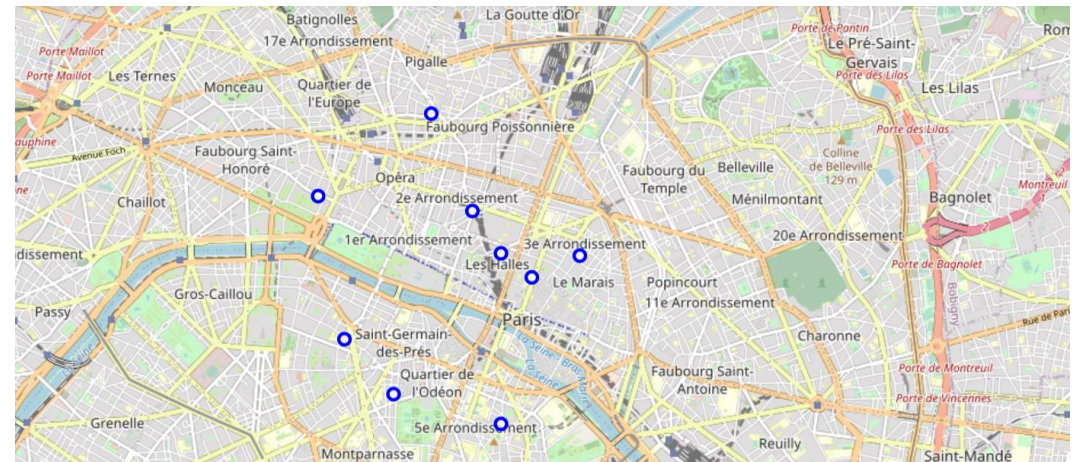
1. List of the arrondissements of Paris with postal codes and region information

2. Geo-coordinates of the arrondissements in Paris

3. Top venues of each arrondissement.

a. List of arrondissements will be obtained from Worldpostalcodes. (https://www.worldpostalcodes.org/en/france/department/list-of-postal-codes-in-paris)

b. Geo-coordinates of districts will be obtained with the geocoder tool.

c. Top venues data will be obtained from Foursquare API.

Methodology (1)

1. CLEANING THE DATA OBTAINED FROM WORLDPOSTALCODES TO GET THE ADDRESS THAT CAN BE IDENTIFIED BY GEOCODER.

2. USING GEOCODER TO FIND THE LOCATION INFORMATION OF THE FIRST NINE ARRONDISSEMENTS AND MAPPING THE ARRONDISSEMENTS USING FOLIUM.

	Postal Code	Place Name	Region	Department	Address
0	75001	1er Arrondissement Paris	Île-de-France	Paris	75001, Île-de-France
1	75002	2ème Arrondissement Paris	Île-de-France	Paris	75002, Île-de-France
2	75003	3ème Arrondissement Paris	Île-de-France	Paris	75003, Île-de-France
3	75004	4ème Arrondissement Paris	Île-de-France	Paris	75004, Île-de-France
4	75005	5ème Arrondissement Paris	Île-de-France	Paris	75005, Île-de-France
5	75006	6ème Arrondissement Paris	Île-de-France	Paris	75006, Île-de-France
6	75007	7ème Arrondissement Paris	Île-de-France	Paris	75007, Île-de-France
7	75008	8ème Arrondissement Paris	Île-de-France	Paris	75008, Île-de-France
8	75009	9ème Arrondissement Paris	Île-de-France	Paris	75009, Île-de-France

Methodology (2)

3. CREATING A FOURSQUARE API AND CREATE A QUERY FOR VENUES CLOSE TO THE CENTER OF EACH ARRONDISSEMENT.

4. PREPARING THE VENUE INFORMATION OBTAINED FROM THE QUERY AND FORMING A TABLE SHOWING THE FREQUENCY OF EACH VENUE CATEGORY IN EACH ARRONDISSEMENT.

```
Out[558]:  {'meta': {'code': 200, 'requestId': '5f03dea30bf45922e6e303c6'},
           'response': {'suggestedFilters': {'header': 'Tap to show:',
            'filters': [{'name': 'Open now', 'key': 'openNow'}]},
           'headerLocation': 'Les Halles',
           'headerFullLocation': 'Les Halles, Paris',
           'headerLocationGranularity': 'neighborhood',
           'totalResults': 245,
           'suggestedBounds': {'ne': {'lat': 48.876766413500015,
             'lng': 2.3687462611490626},
            'sw': {'lat': 48.84976638649999, 'lng': 2.327780573790696}},
           'groups': [{'type': 'Recommended Places',
             'name': 'recommended',
             'items': [{'reasons': {'count': 0,
                'items': [{'summary': 'This spot is popular',
                  'type': 'general',
                  'reasonName': 'globalInteractionReason'}]},
               'venue': {'id': '4c457e8d8c1f20a195ce3d99',
                'name': 'Passage du Grand Cerf',
                'location': {'address': 'Passage du Grand Cerf',
                 'lat': 48.86476
```

Neighborhood	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	Bakery	Bar	Basketball Court	Beer Bar	Beer Store	Bistro	Bookstore	Bout
0	1er Arrondissement Paris	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.05	0.01	0.00	0.01	0.01	0.00	0.01
1	2ème Arrondissement Paris	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.01	0.00	0.02	0.01
2	3ème Arrondissement Paris	0.01	0.00	0.00	0.01	0.05	0.02	0.00	0.03	0.00	0.00	0.01	0.00	0.04	0.03
3	4ème Arrondissement Paris	0.01	0.00	0.00	0.00	0.04	0.01	0.00	0.04	0.00	0.00	0.00	0.01	0.01	0.01
4	5ème Arrondissement Paris	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.04	0.00	0.00	0.00	0.02	0.04
5	6ème Arrondissement Paris	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.03
6	7ème Arrondissement	0.00	0.02	0.00	0.00	0.02	0.02	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.02

Methodology (3)

5. FINDING THE OPTIMAL K USING VISUALIZATION TOOLS FOR THE K-MEANS ANALYSIS.
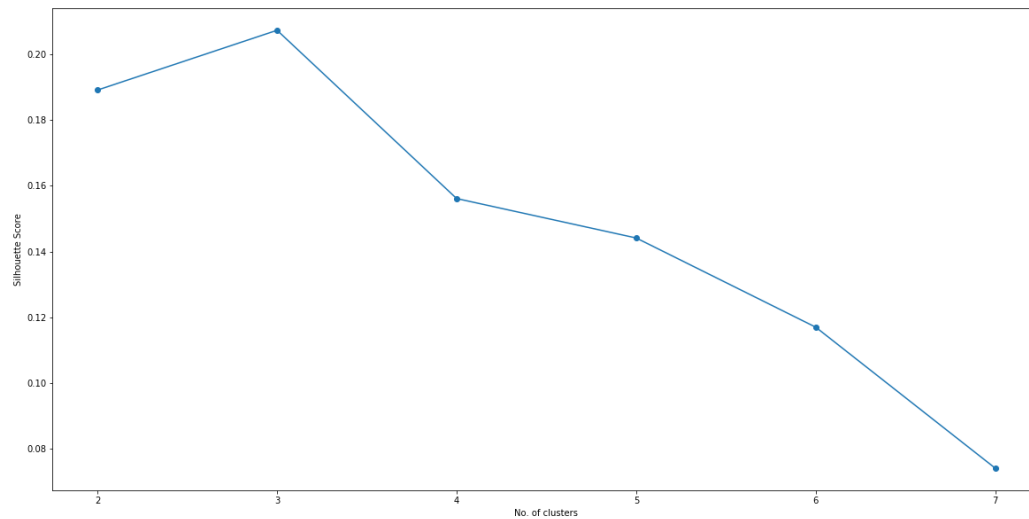
6. PERFORMING THE K-MEANS ANALYSIS AND MAPPING THE DIFFERENT CLUSTERS ON THE MAP.
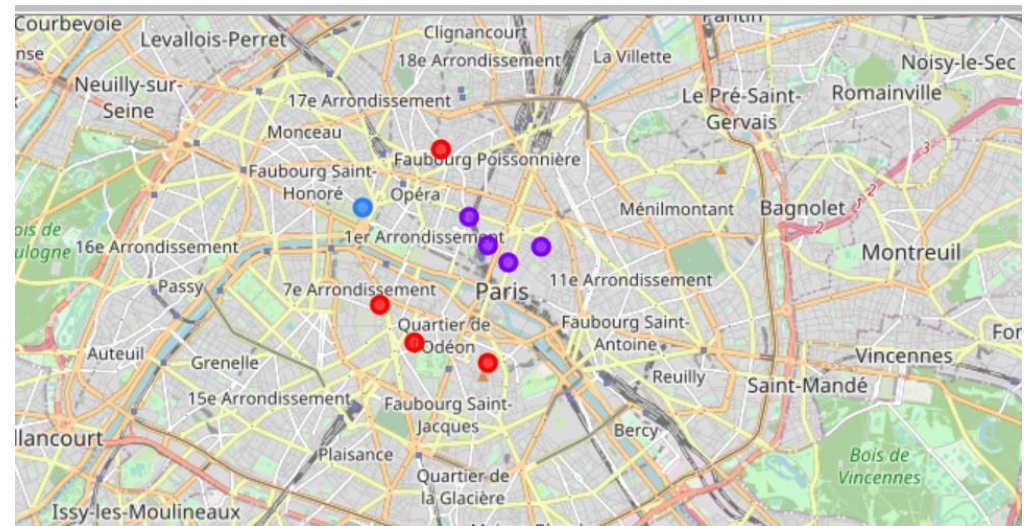
7. EXPLORING THE FEATURE OF EACH ARRONDISSEMENT AND FINDING THE PROPER LOCATION TO START THE BUSINESS.

Results (Cluster 0)

Based on the result of the analysis, there exist 3 different clusters in the city center of Paris:

Cluster 0 includes Arrondissement 5,6,7, and 8. Despite restaurants and hotels, we can observe that there exist many cultural venues such as bookstores and cinemas due to the fact that there are many schools and universities, thus many young people living in this area.

	Address	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
4	75005, Île-de-France	French Restaurant	Bakery	Italian Restaurant	Bookstore	Plaza	Bar	Coffee Shop	Indie Movie Theater	Comic Shop	Creperie
5	75006, Île-de-France	French Restaurant	Hotel	Plaza	Garden	Bistro	Restaurant	Chocolate Shop	Bookstore	Ice Cream Shop	Wine Bar
6	75007, Île-de-France	French Restaurant	Hotel	Plaza	Italian Restaurant	Coffee Shop	Garden	Bistro	Bakery	Bookstore	Japanese Restaurant
8	75009, Île-de-France	French Restaurant	Hotel	Wine Bar	Italian Restaurant	Cocktail Bar	Coffee Shop	Plaza	Pizza Place	Bar	Bistro

Results (Cluster 1)

Based on the result of the analysis, there exist 3 different clusters in the city center of Paris:

Cluster 1 represents the gastronomy center of the city, with restaurants from different countries, bars, bakeries, etc.

	Address	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	75001, Île-de-France	French Restaurant	Bakery	Coffee Shop	Cocktail Bar	Chinese Restaurant	Pizza Place	Restaurant	Ice Cream Shop	Hotel	Wine Bar
1	75002, Île-de-France	French Restaurant	Japanese Restaurant	Wine Bar	Hotel	Cocktail Bar	Italian Restaurant	Pedestrian Plaza	Restaurant	Bar	Bakery
2	75003, Île-de-France	Coffee Shop	Art Gallery	French Restaurant	Burger Joint	Restaurant	Bistro	Cocktail Bar	Clothing Store	Bookstore	Japanese Restaurant
3	75004, Île-de-France	French Restaurant	Ice Cream Shop	Burger Joint	Coffee Shop	Art Gallery	Bakery	Park	Wine Bar	Pub	Sandwich Place

Results (Cluster 2)

Based on the result of the analysis, there exist 3 different clusters in the city center of Paris:

Cluster 2 is made up of Arrondissement 8 where people can enjoy the high-quality accommodation with luxury hotels as well as shopping centers and restaurants.

	Address	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
7	75008, Île-de-France	Hotel	French Restaurant	Boutique	Garden	Women's Store	Dessert Shop	Plaza	Theater	Tea Room	Gourmet Shop

Discussion & Recommendation

Based on the analysis of the three clusters and their features, we would recommend the business owner to launch his business in Cluster 0 (Arrondissement 5/6/7/8) to have the best return possible.

However, as we have only conducted the analysis on the 1-9 Arrondissement dur to the restriction of Geopy, there might exists other optimal combinations on the map should the business owner be interested in launching his business in areas further from the city center for less competition and lower costs.
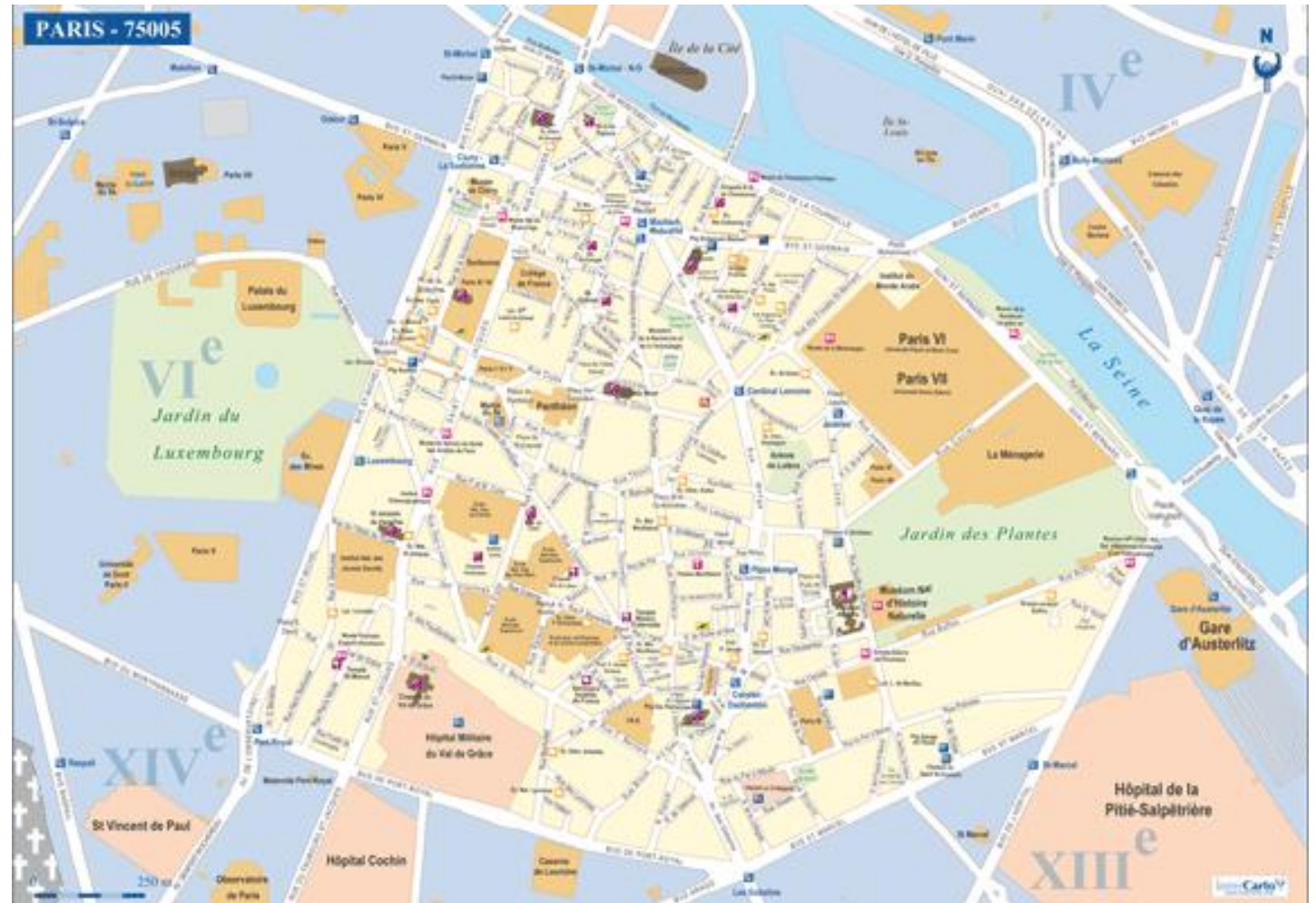
Conclusion

In this project, we have covered different topics such as data preparation, data visualization, unsupervised machine learning and the K-means machine leaning model, etc. In response to the hypothetical business problem of opening a café-bookstore in the city center of Paris, we concluded by recommending the business owner to start his business in the 5/6/7/8 Arrondissement in Paris(Cluster 0) satisfying the criteria mentioned in earlier in this report.