Predicting the best place to live in Paris, France

IBM Data Science Certificate Capstone

Introduction

- Paris is an attractive city to work and live, but each district has its own specificities

- When ones wants to live there, it could be difficult to make an objective choice taking into consideration ones personal criterias

- Analyzing and comparing the characteristics of the 20 districts of Paris, such as venues available, crime rate or real estate prices can be a help for people who aim to live there

- This analysis aims to be a proof of concept. It takes into consideration a restricted range of criterias only in Paris. The final goal is to make it work with a wider range of cities and criterias

Data acquisition and cleaning

The data needed are :

- Geographical coordinates of each of the 20 districts of Paris
- Venues locations & characteristics using **Foursquare**
- Real estate prices using the website https://ouinvestir.net/
- Data about crime rate using **ONDRP data** (a French observatory of delinquency)

For each of those data, we will establish a score to compare the districts between eachother

Data acquisition and cleaning
Criminality rates

	District	Violences physiques non crapuleuses /1000hab	Violences physiques crapuleuses /1000hab	Violences sexuelles /1000hab	Menaces /1000hab	Vols simples /1000hab	Cambriolages /1000hab	Destructions & degradations /1000hab
0	Paris 1	18.7	30.4	2.7	9.5	277.9	11.1	16.9
1	Paris 2	8.1	11.6	1.2	2.6	89.0	16.6	12.2
2	Paris 3	5.4	6.3	0.8	2.0	51.5	11.0	9.4
3	Paris 4	8.9	10.4	1.6	3.5	86.6	11.7	12.9
4	Paris 5	4.8	4.1	0.6	1.6	37.8	7.7	6.9
5	Paris 6	4.3	5.1	0.6	1.6	52.0	7.8	6.4
6	Paris 7	4.6	3.6	0.6	1.5	38.4	7.6	5.5

- Criminality rates collection using ONDP data
- Scoring for criminality is defined by the mean of the different crime types

	District	Criminality score
0	Paris 1	52.5
1	Paris 2	20.2
2	Paris 3	12.3
3	Paris 4	19.4
4	Paris 5	9.1
5	Paris 6	11.1
6	Paris 7	8.8

Data acquisition and cleaning
Real estate prices

	District	Price (/m2)
0	Paris 1	12184
1	Paris 2	11039
2	Paris 3	11583
3	Paris 4	13701
4	Paris 5	12037
5	Paris 6	13429
6	Paris 7	17959

Real estate prices collection using the website
https://ouinvestir.net/

Data acquisition and cleaning
Geographical coordinates and concatenation of information

	City	District	Postcode	Criminality score	Price (/m2)	latitude	longitude
0	Paris	Paris 1	75001	52.5	12184	48.8592	2.3417
1	Paris	Paris 2	75002	20.2	11039	48.8655	2.3426
2	Paris	Paris 3	75003	12.3	11583	48.8637	2.3615
3	Paris	Paris 4	75004	19.4	13701	48.8601	2.3507
4	Paris	Paris 5	75005	9.1	12037	48.8448	2.3471
5	Paris	Paris 6	75006	11.1	13429	48.8493	2.3300
6	Paris	Paris 7	75007	8.8	17959	48.8565	2.3210
7	Paris	Paris 8	75008	27.7	16537	48.8763	2.3183
8	Paris	Paris 9	75009	18.6	10135	48.8718	2.3399
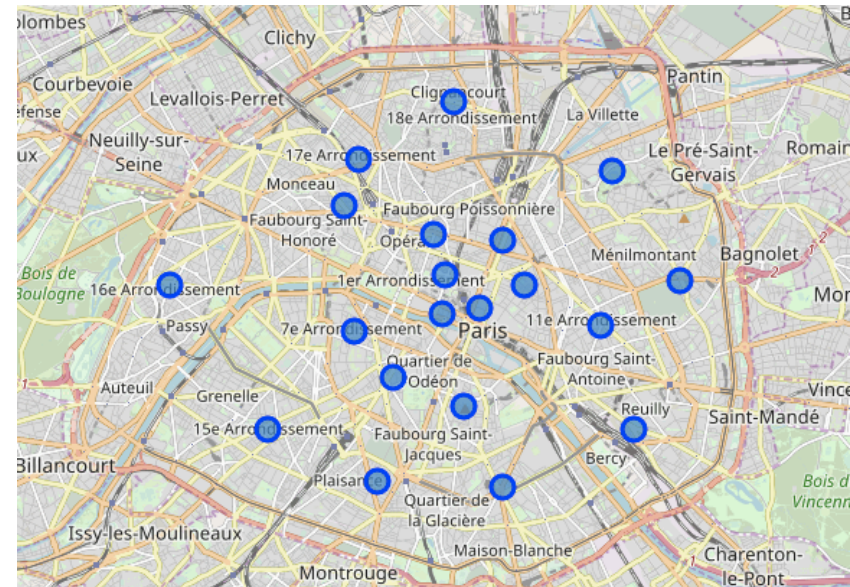
Data acquisition and cleaning
Foursquare data collection

We use Foursquare to collect venues and compare each district of Paris

	City	District	Postcode	Criminality score	Price (/m2)	latitude	longitude	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	8' C
0	Paris	Paris 1	75001	52.5	12184	48.8592	2.3417	1	French Restaurant	Café	Bakery	Hotel	Plaza	Pizza Place	Clothing Store	
1	Paris	Paris 2	75002	20.2	11039	48.8655	2.3426	1	French Restaurant	Italian Restaurant	Wine Bar	Cocktail Bar	Restaurant	Café	Bakery	
2	Paris	Paris 3	75003	12.3	11583	48.8637	2.3615	1	Bistro	Café	French Restaurant	Coffee Shop	Art Gallery	Bakery	Japanese Restaurant	
3	Paris	Paris 4	75004	19.4	13701	48.8601	2.3507	1	French Restaurant	Hotel	Bakery	Art Gallery	Plaza	Clothing Store	Pub	
4	Paris	Paris 5	75005	9.1	12037	48.8448	2.3471	2	French Restaurant	Hotel	Bar	Italian Restaurant	Bakery	Pub	Creperie	

Methodology

- Now we have scrapped the venues of each district in Paris and clustered those districts, we are able to chose the cluster that matches the best with our own preferences.

- When the best cluster is chosen, we can compare each district inside it using crime rates and real estate prices to chose the best district to live.

Analysis

5 clusters are created using Foursquare data about most common venues in each district
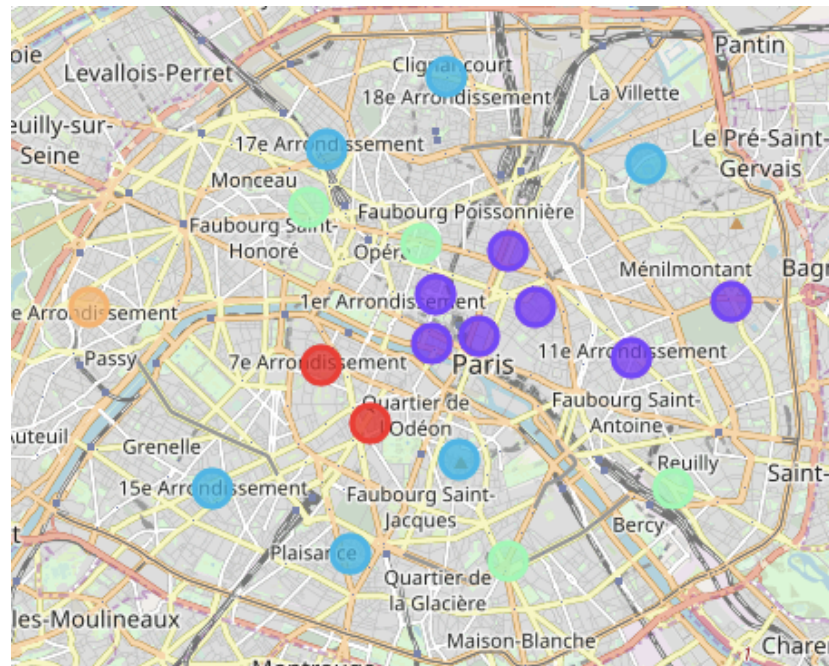
Analysis

Assuming cluster 2 is the most diversified and less touristical, we can compare the districts with the criterias we defined at the beginning

	District	latitude	longitude	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
4	Paris 5	48.8448	2.3471	2	French Restaurant	Hotel	Bar	Italian Restaurant	Bakery	Pub	Creperie	Plaza	Wine Bar	Café
13	Paris 14	48.8331	2.3264	2	French Restaurant	Bar	Bakery	Italian Restaurant	Hotel	Bistro	Vietnamese Restaurant	Latin American Restaurant	Bookstore	Plaza
14	Paris 15	48.8412	2.3003	2	French Restaurant	Hotel	Coffee Shop	Supermarket	Italian Restaurant	Bar	Lebanese Restaurant	Korean Restaurant	Park	Thai Restaurant
16	Paris 17	48.8835	2.3219	2	French Restaurant	Wine Bar	Hotel	Bar	Restaurant	Italian Restaurant	Thai Restaurant	Coffee Shop	Asian Restaurant	Bookstore
17	Paris 18	48.8925	2.3444	2	French Restaurant	Bar	Pizza Place	Deli / Bodega	Italian Restaurant	Restaurant	Gastropub	Hotel	Café	Burger Joint
18	Paris 19	48.8817	2.3822	2	French Restaurant	Bar	Italian Restaurant	Restaurant	Pool	Historic Site	Diner	Café	Scenic Lookout	Moroccan Restaurant

Results and Discussion

- This analysis shows that disricts 14 and 15 are both interesting viewing real estate prices and criminality rate.

- Assuming that we want a "day-to-day" way of life, district 15 seems the best choice because of the number of supermarkets (4th most common venue)
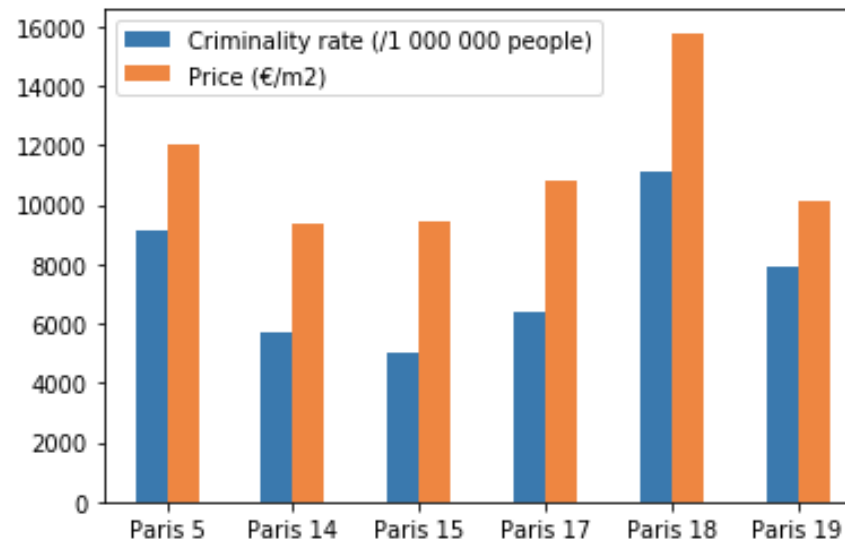
Conclusion

- This analysis helped us to chose a convenient place to live in Paris, by comparing each of the 20 disctricts in terms of most common venues, criminality rates and real estate prices

- Since the 15th district of Paris is large, the next step could be an analysis of the neighborhoods of this district

- We can also take into consideration density of population or transportation facilities to make a more precise choice