CAPSTONE PROJECT

REALTY INVESTMENT IN PARIS AREA (FRANCE)

Author: Dimitri BONDARENKO

INTRODUCTION

The goal of my capstone project is to guide a given young family in their first investment project to become a realty owner in Paris suburb area. It is not an easy decision that often conducts to an engagement with bank for multiple years. With this project I am going to reveal several criteria to facilitate decision making.

Paris is a worldwide known city and attract people from all over the world because of its historical sightseeing, French cuisine and fashion industry. Alongside with that Paris itself has an early mediaeval planification with dense building and very few places for new constructions. All these factors contributed to a significant increase in real-estate prices. Hence, young people are forced to focus on the purchase of real estate in Paris suburbs in Ile-de-France region.

ASSUMPTIONS

In this study I am going to make the following assumptions:

- The area of interest is limited with the closest suburbs around Paris;
- The study covers French professional real-estate market only;
- The legal side and French laws in realty sector (ex. Carrez law, Pinel law, etc.) remain outside the presented analysis.

All data used for analysis were taken from open sources.

DATA DESCRIPTION

Within this project I decided to limit the area of interest around the Paris with the closest communes (i.e. towns). It gives me three districts as follows (https://kelcodepostal.fr/):

- District Hauts-de-Seine with 36 communes;
- District Seine-Saint-Denis with 40 communes;
- District Val-de-Marne with 47 communes;

In total, there is a list with 123 communes with their names and postal codes to consider in the analysis. The question is, how to differentiate all these communes in order to provide additional information for decision making while working on investment project. Surfing the web, I found the following site https://www.ville-ideale.fr/ where users can evaluate and rate theirs communes using criteria:

- Environment
- Transport
- Security

- Healthcare
- Sports & leisures
- Culture
- Education
- Commerce
- Quality of life

Next, via Google Chrome + Python with Selenium, BeautifulSoup and Pandas libraries I managed to scrap this website and create a dataframe containing rates for interesting communes in Hauts-de-Seine, Seine-Saint-Denis and Val-de-Marne districts. Here is five first lines of the dataframe:

	Code INSEE	Environment	Transport	Security	Healthcare	Sports & leisures	Culture	Education	Commerce	Quality of life
0	92002	7.63	7.30	6.97	8.10	8.04	7.67	7.89	7.38	7.97
1	92019	7.07	5.54	5.15	5.84	6.93	7.43	5.89	5.16	6.20
2	92060	8.23	5.72	8.46	7.74	7.81	7.43	7.12	7.33	8.12
3	92071	8.47	6.86	7.51	7.57	7.80	7.86	8.89	7.79	8.41
4	92014	6.88	8.55	7.88	7.67	7.12	6.81	7.97	7.43	8.33

It should be outlined, that in France each commune has its unique identification code which is INSEE code (look at the first column in the above dataframe). Furthermore, my web scrapping Python algorithm extracted rates only for 102 communes from 123 introduced above.

Another web site https://www.meilleursagents.com/prix-immobilier/ was used to scrap mean prices per squared meter for apartment and houses for all of 102 communes. I used INSEE code to identify each commune. Here is five first rows of the dataframe:

	Code INSEE	Mean price m2 (apartment)	Mean price m2 (house)
0	92002	4963	5400
1	92019	4390	5701
2	92060	5084	5878
3	92071	6131	7810
4	92014	5542	6653

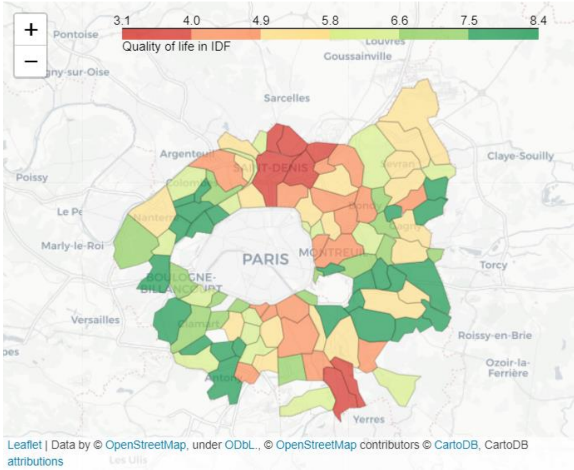
Now, to visualize all the collected data with rates and prices on a map we need longitude & latitude coordinates of each commune. This data was downloaded and extracted from https://sql.sh/736-base-donnees-villes-francaises and converted to a dataframe as follows:

	Code dept	Commune	Postal code	Code INSEE	Surface	Longitude	Latitude
0	92	NEUILLY-SUR-SEINE	92200	92051	3.73	2.26667	48.8833
1	92	CHATILLON	92320	92020	2.92	2.28333	48.8000
2	92	BOIS-COLOMBES	92270	92009	1.92	2.26667	48.9167
3	92	PUTEAUX	92800	92062	3.19	2.23333	48.8667
4	92	CLAMART	92140	92023	8.77	2.26667	48.8000

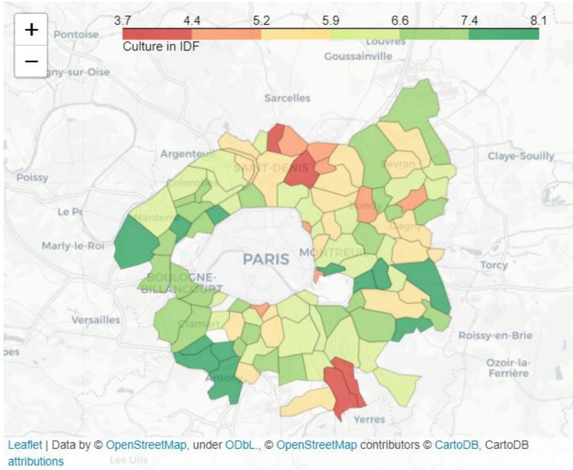
Where the column "Code dept" stands for the district's identification code: 92 for Hauts-de-Seine, 93 for Seine-Saint-Denis and 94 for Val-de-Marne. The column "Surface" stands for the surface of commune in squared kilometres.

The map data in geojson format is available for free on French government website https://www.data.gouv.fr/fr/datasets/geofla-departements-idf/.

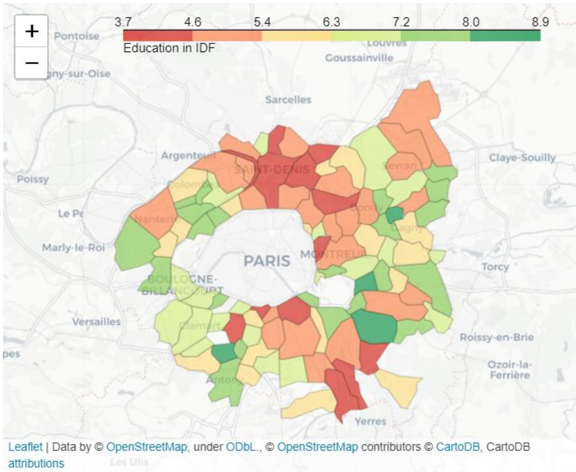
Now we are ready to visualize rates and prices per commune on the choropleth geographic map using Python with Folium & Leaflet libraries:


3.1 4.0 4.9 5.8 6.6 7.5 8.4
Quality of life in IDF
PARIS
Leaflet | Data by © OpenStreetMap, under ODbL., © OpenStreetMap contributors © CartoDB, CartoDB attributions

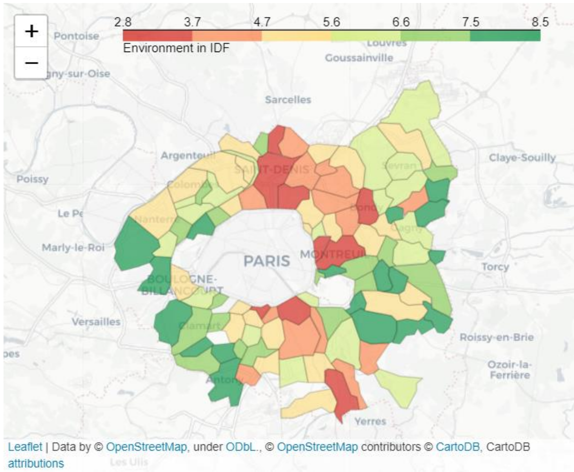
(a)


3.7 4.4 5.2 5.9 6.6 7.4 8.1
Culture in IDF
PARIS
Leaflet | Data by © OpenStreetMap, under ODbL., © OpenStreetMap contributors © CartoDB, CartoDB attributions

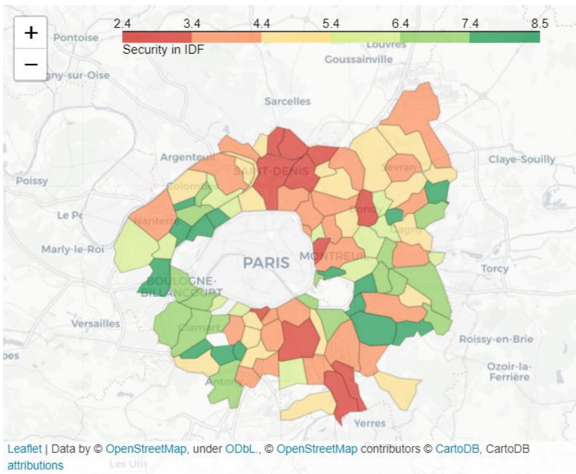
(b)


3.7 4.6 5.4 6.3 7.2 8.0 8.9
Education in IDF
PARIS
Leaflet | Data by © OpenStreetMap, under ODbL., © OpenStreetMap contributors © CartoDB, CartoDB attributions

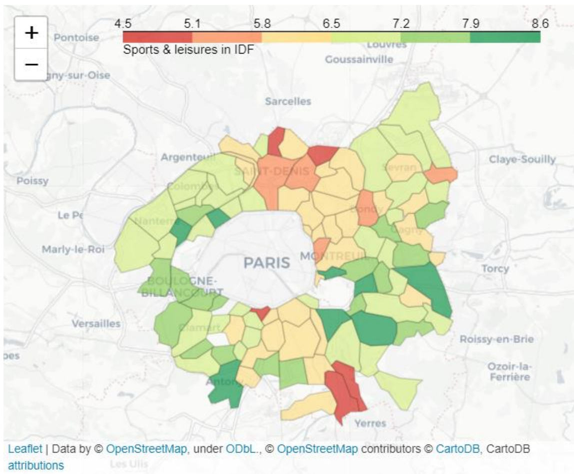
(c)


2.8 3.7 4.7 5.6 6.6 7.5 8.5
Environment in IDF
PARIS
Leaflet | Data by © OpenStreetMap, under ODbL., © OpenStreetMap contributors © CartoDB, CartoDB attributions

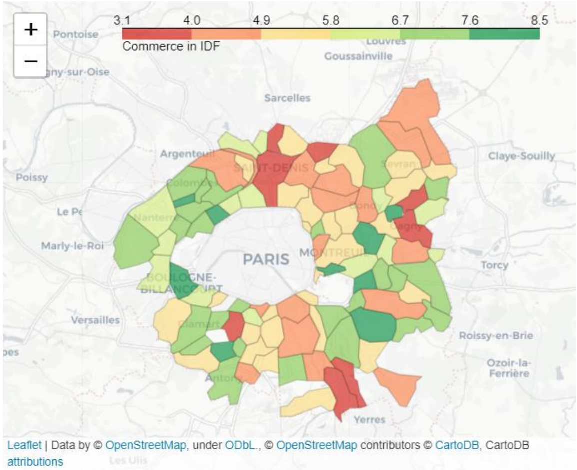
(d)


2.4 3.4 4.4 5.4 6.4 7.4 8.5
Security in IDF
PARIS
Leaflet | Data by © OpenStreetMap, under ODbL., © OpenStreetMap contributors © CartoDB, CartoDB attributions

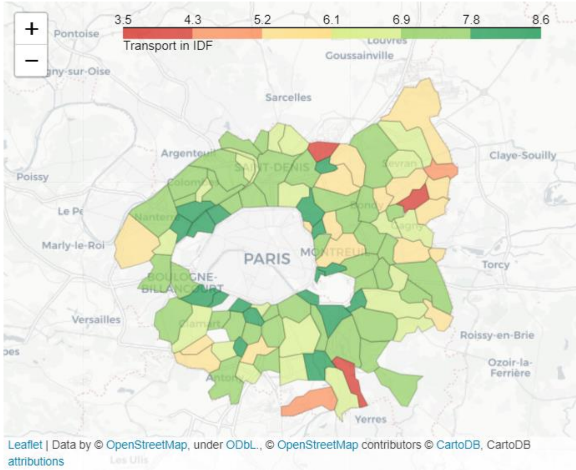
(e)


4.5 5.1 5.8 6.5 7.2 7.9 8.6
Sports & leisures in IDF
PARIS
Leaflet | Data by © OpenStreetMap, under ODbL., © OpenStreetMap contributors © CartoDB, CartoDB attributions

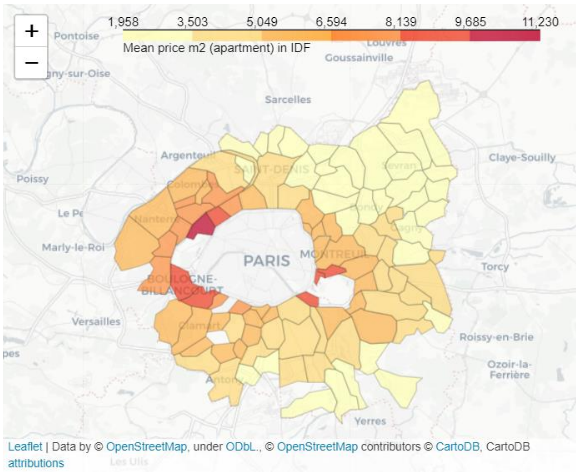
(f)



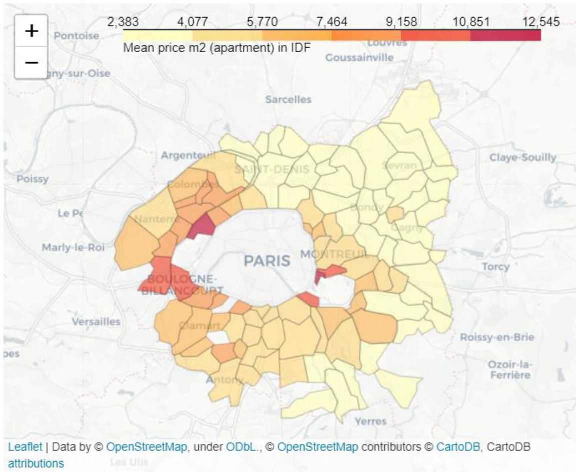
(g)



(h)



(i)



(j)

Figure 1: Different indicators per commune in Ile-de-France region, (a) – Quality of life; (b) – Culture rating; (c) – Education rating; (d) – Environment rating; (e) – Security level; (f) – Sport & leisure rating; (g) – Commerce rating; (h) – Transport network; (i) – Mean price in € per square meter for apartments; (j) – Mean price in € per square meter for houses.

Looking at all these choropleth maps its not so obvious to make an optimum choice of commune for investment in both apartment and house. Nevertheless, some observations could be outlined for the given dataset:

- East side and west side closest to Paris communes are most expensive;
- Cheapest realty can be found in north-east communes;
- Different indicators have their own distribution between different communes, except few communes in the north where most indicators are in red.

In the next section I will apply machine learning techniques to classify communes and to enhance the decision making in our realty investment project. An addition, in the next section I will evaluate the accuracy of discovered ratings (Quality of life, Security, Education, etc.) through distribution of user votes.

Now, lets image we made a decision and chose a commune to invest based on mean price and different ratings introduced hereabove. The next question we could ask is what types of venues are the most visited and rated in the selected commune. Foursquare.com database based on user ratings is a good tool to scrap and to examine different venues. Let’s take an example here and look at venues in BOULOGNE-BILLANCOURT. Using Foursquare.com API with Python + Request library we reveal Top10 venues for this commune as follows:

```
----BOULOGNE-BILLANCOURT----
                venue  freq
0  Japanese Restaurant  0.13
1    French Restaurant  0.11
2   Italian Restaurant  0.11
3               Bakery  0.07
4                Hotel  0.06
5          Supermarket  0.06
6     Asian Restaurant  0.06
7          Pizza Place  0.04
8               Bistro  0.04
9                 Café  0.04
```

From my point of view, it is not surprised to see different kinds of restaurant as the most visited places in France :)