RECOMMENDING A LOCATION FOR A NEW IT COMPANY OFFICE IN PARIS

Applied Data Science Capstone

IBM Data Science Professional Certificate

November 2018

Performed by MRABAH Yassine

Contents:

1- Introduction to the problem

2- Data scraping

Opendata.paris.fr

- Stations data & locations
- Restaurants locations
- Public parks, sport centers locations

A real estate agency

- The price of a m^2 in each arrondissement in Paris

The foursquare API

- Coordinates of Paris

We're going to collect data from these different sources, and then we will create a new data frame where we will combine the most important "columns" extracted from each source

3- Used methodology

1

Data from multiple sources as described in the previous chapter

Weights

M² price affordability	4
Accessibility to public transport	2
Accessibility to various restaurants	2
Accessibility to public parks	1
Accessibility to sport centers	1

Sum = 10

2

We define the settlement index (a value between 1 and 10), we will select the neighborhood that has a value closer to 10

Comparison

Perform clustering analysis to find similar neighborhoods without calculating the settlement index

3

Visualization of the results on a map

3- Used methodology

X 2 | X 1 | X 1 | X 2 | X 4

	Neighborhood	Nb restaurants	Nb sport centers	Nb parks	Nb stations	price per squared m
0	1	0.335470	0.000000	0.055556	0.306569	3.457760e-01
1	2	0.341880	0.095238	0.055556	0.031630	5.383104e-01
2	3	0.000000	0.238095	0.111111	0.000000	3.555992e-01
3	4	0.042735	0.142857	0.055556	0.306569	2.750491e-01
4	5	0.628205	0.238095	0.000000	0.381995	2.612967e-01
5	6	0.337607	0.190476	0.000000	0.467153	-2.220446e-16
6	7	0.032051	0.190476	0.000000	0.450122	1.964637e-01
7	8	0.833333	0.285714	0.055556	0.919708	3.516699e-01
8	9	1.000000	0.333333	0.000000	0.454988	5.599214e-01
9	10	0.664530	0.047619	0.277778	0.566910	8.271120e-01
10	11	0.869658	0.857143	0.333333	0.257908	6.994106e-01
11	12	0.431624	0.476190	0.333333	0.907543	8.271120e-01
12	13	0.517094	0.190476	0.444444	0.776156	7.956778e-01
13	14	0.395299	0.380952	0.444444	0.744526	6.601179e-01
14	15	0.871795	1.000000	0.166667	1.000000	6.306483e-01
15	16	0.303419	0.571429	0.055556	0.880779	5.108055e-01
16	17	0.841880	0.857143	0.111111	0.732360	6.974460e-01
17	18	0.820513	0.285714	0.611111	0.802920	9.135560e-01
18	19	0.209402	0.333333	0.888889	0.386861	1.000000e+00
19	20	0.177350	0.476190	1.000000	0.537713	9.469548e-01

	Neighborhood	Nb restaurants	Nb sport centers	Nb parks	Nb stations	price per squared m
0	1	0.670940	0.000000	0.055556	0.613139	1.383104e+00
1	2	0.683761	0.095238	0.055556	0.063260	2.153242e+00
2	3	0.000000	0.238095	0.111111	0.000000	1.422397e+00
3	4	0.085470	0.142857	0.055556	0.613139	1.100196e+00
4	5	1.256410	0.238095	0.000000	0.763990	1.045187e+00
5	6	0.675214	0.190476	0.000000	0.934307	-8.881784e-16
6	7	0.064103	0.190476	0.000000	0.900243	7.858546e-01
7	8	1.666667	0.285714	0.055556	1.839416	1.406680e+00
8	9	2.000000	0.333333	0.000000	0.909976	2.239686e+00
9	10	1.329060	0.047619	0.277778	1.133820	3.308448e+00
10	11	1.739316	0.857143	0.333333	0.515815	2.797642e+00
11	12	0.863248	0.476190	0.333333	1.815085	3.308448e+00
12	13	1.034188	0.190476	0.444444	1.552311	3.182711e+00
13	14	0.790598	0.380952	0.444444	1.489051	2.640472e+00
14	15	1.743590	1.000000	0.166667	2.000000	2.522593e+00
15	16	0.606838	0.571429	0.055556	1.761557	2.043222e+00
16	17	1.683761	0.857143	0.111111	1.464720	2.789784e+00
17	18	1.641026	0.285714	0.611111	1.605839	3.654224e+00
18	19	0.418803	0.333333	0.888889	0.773723	4.000000e+00
19	20	0.354701	0.476190	1.000000	1.075426	3.787819e+00

Data before and after applying the weights

4- Results

Arrondissement = Neighborhood

	Neighborhood	Nb restaurants	Nb sport centers	Nb parks	Nb stations	price per squared m	latitude	longitude	settlement index
0	3	201	6	2	93	9900	48.86287238	2.3600009859	1.77
1	6	359	5	0	285	11710	48.8491303586	2.33289799905	1.80
2	7	216	5	0	278	10710	48.8561744288	2.31218769148	1.94
3	4	221	4	1	219	10310	48.8543414263	2.35762962032	2.00
4	1	358	1	1	219	9950	48.8625627018	2.33644336205	2.72
5	2	361	3	1	106	8970	48.8682792225	2.34280254689	3.05
6	5	495	6	0	250	10380	48.8444431505	2.35071460958	3.30
7	16	343	13	1	455	9110	48.8603921054	2.26197078836	5.04
8	8	591	7	1	471	9920	48.8727208374	2.3125540224	5.25
9	9	669	8	0	280	8860	48.8771635173	2.33745754348	5.48
10	14	386	9	8	399	8350	48.8292445005	2.3265420442	5.75
11	10	512	2	5	326	7500	48.8761300365	2.36072848785	6.10
12	11	608	19	6	199	8150	48.8590592213	2.3800583082	6.24
13	13	443	5	8	412	7660	48.8283880317	2.36227244042	6.40
14	19	299	8	16	252	6620	48.8870759966	2.38482096015	6.41
15	20	284	11	18	314	6890	48.8634605789	2.40118812928	6.69
16	12	403	11	6	466	7500	48.8349743815	2.42132490078	6.80
17	17	595	19	2	394	8160	48.887326522	2.30677699057	6.91
18	15	609	22	3	504	8500	48.8400853759	2.29282582242	7.43
19	18	585	7	11	423	7060	48.892569268	2.34816051956	7.80

Index < 4

4 < Index < 6

Index > 6

Arrondissements sorted by the settlement index

4- Results

	Neighborhood	Nb restaurants	Nb sport centers	Nb parks	Nb stations	price per squared m	latitude	longitude	settlement index	cluster label
0	1	0.335470	0.000000	0.055556	0.306569	3.457760e-01	48.862563	2.336443	2.72	0
1	2	0.341880	0.095238	0.055556	0.031630	5.383104e-01	48.868279	2.342803	3.05	0
2	3	0.000000	0.238095	0.111111	0.000000	3.555992e-01	48.862872	2.360001	1.77	0
3	4	0.042735	0.142857	0.055556	0.306569	2.750491e-01	48.854341	2.357630	2.00	0
4	5	0.628205	0.238095	0.000000	0.381995	2.612967e-01	48.844443	2.350715	3.30	0
5	6	0.337607	0.190476	0.000000	0.467153	-2.220446e-16	48.849130	2.332898	1.80	0
6	7	0.032051	0.190476	0.000000	0.450122	1.964637e-01	48.856174	2.312188	1.94	0
7	8	0.833333	0.285714	0.055556	0.919708	3.516699e-01	48.872721	2.312554	5.25	2
8	9	1.000000	0.333333	0.000000	0.454988	5.599214e-01	48.877164	2.337458	5.48	2
9	10	0.664530	0.047619	0.277778	0.566910	8.271120e-01	48.876130	2.360728	6.10	1
10	11	0.869658	0.857143	0.333333	0.257908	6.994106e-01	48.859059	2.380058	6.24	2
11	12	0.431624	0.476190	0.333333	0.907543	8.271120e-01	48.834974	2.421325	6.80	1
12	13	0.517094	0.190476	0.444444	0.776156	7.956778e-01	48.828388	2.362272	6.40	1
13	14	0.395299	0.380952	0.444444	0.744526	6.601179e-01	48.829245	2.326542	5.75	1
14	15	0.871795	1.000000	0.166667	1.000000	6.306483e-01	48.840085	2.292826	7.43	2
15	16	0.303419	0.571429	0.055556	0.880779	5.108055e-01	48.860392	2.261971	5.04	2
16	17	0.841880	0.857143	0.111111	0.732360	6.974460e-01	48.887327	2.306777	6.91	2
17	18	0.820513	0.285714	0.611111	0.802920	9.135560e-01	48.892569	2.348161	7.80	1
18	19	0.209402	0.333333	0.888889	0.386861	1.000000e+00	48.887076	2.384821	6.41	1
19	20	0.177350	0.476190	1.000000	0.537713	9.469548e-01	48.863461	2.401188	6.69	1

Label = 0 corresponds very well to settlement index < 4

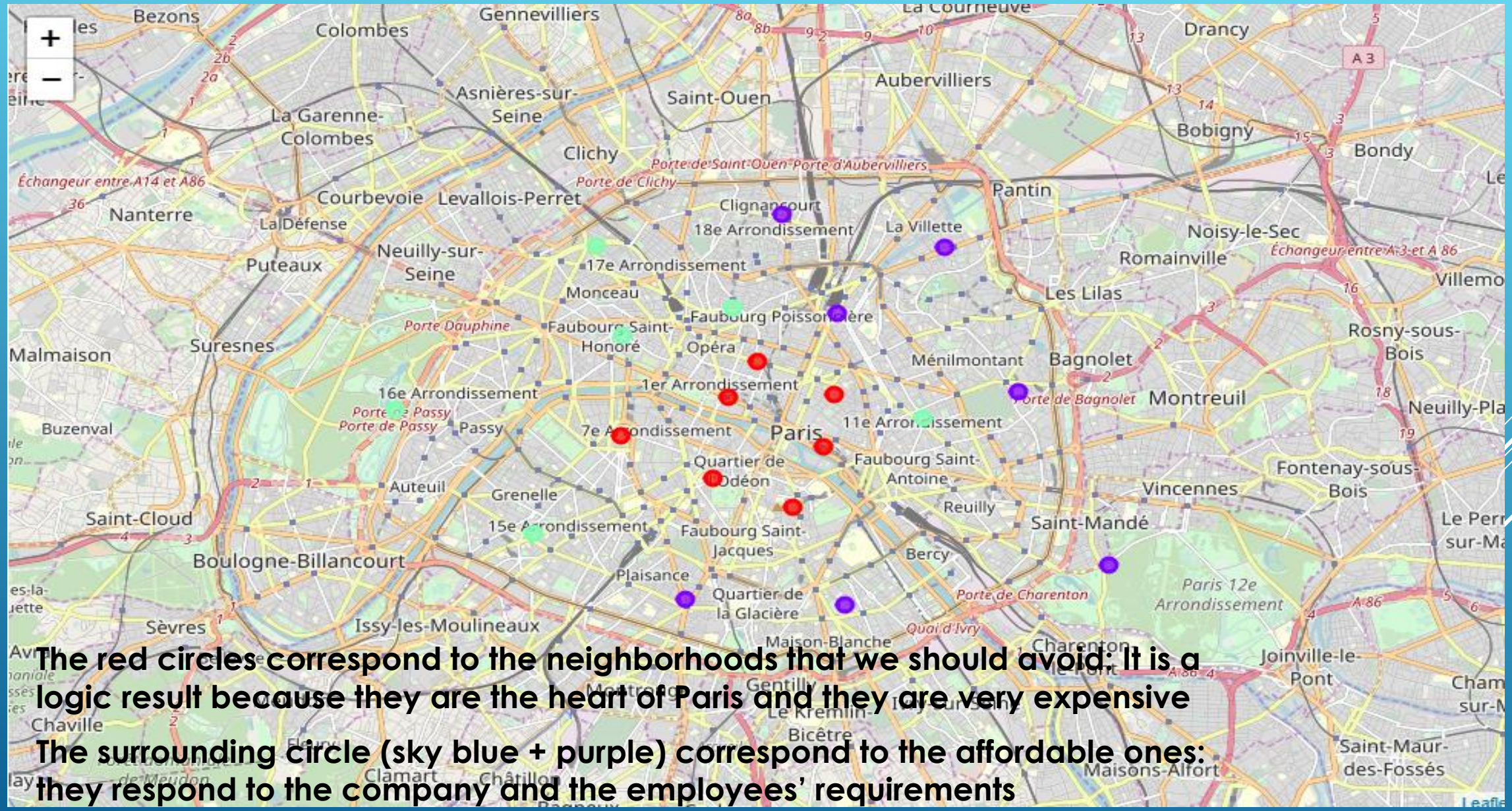
For label = 1 and 2, settlement index > 4

It is mainly = 5,6 and 7

An acceptable choice for a location starts from an index > 4

Results after performing clustering analysis for n = 3

4- Results



The red circles correspond to the neighborhoods that we should avoid: It is a logic result because they are the heart of Paris and they are very expensive

The surrounding circle (sky blue + purple) correspond to the affordable ones: they respond to the company and the employees' requirements

5- Conclusions

After calculating the settlement index, we were able to sort the neighborhoods, and provide a suitable choice to the company, and we were able to ensure that by performing clustering analysis

if settlement index > 6 : The company is advised to establish its new office in the corresponding neighborhoods

if settlement index is between 4 and 6 : It can be a solution

if settlement index < 4 : The company should absolutely avoid these choices

We found that : arrondissements : 1 to 7 should be avoided (< 4)

arrondissements : 16, 8, 9, 14 can be a solution (4< index < 6)

arrondissements : 10, 11, 13, 19, 20, 12, 17, 15, 18 are a good solution (index > 6)

As a suggestion, the company should establish the new office at the 18th arrondissement as it has the highest settlement index