

IBM Data Science Professional Certificate

Applied Data Science Capstone

Evaluation of the environmental quality of a borough

Case study with the city of Paris, France

F.MARTIN - April 2020

INTRODUCTION

Today, more than 55% of human beings are living in an urban area. Based on current predictions, by 2050 about 64% of the developing world and 86% of the developed world will be urbanized.

Environmental issues are crucial in cities. Indeed, population well-being depends on the quality of their environment. The urban environment is subject to many pollutions and has few natural and environmental benefits. As a result, the well-being produced by the environment varies greatly from one neighbourhood to another in the same city.

For example, if you live near a park, you might have better air quality during the year and freshness in summer. Some indicators, publications and articles evaluate the quality of a district according to different criteria. Usually, it does not take the quality of the environment into account. However, highlighting this quality might be a way to reveal a “green district value” in property transaction. Revealing this type of value is a way to internalize the environmental externality and develop investments in this business.

To this end, we will create an indicator of « environmental well-being » to help people who are looking for housing to evaluate their district. This indicator will aggregate information found on Foursquare. People seem to be more and more concerned about environment but I think they do not consider it when looking for housing because they lack of information. Creating an indicator will facilitate awareness about the importance of this criterion.

The measurement will be based on the access to three types of environmental facilities:

- “Green” areas (gardens, pedestrian area and bike trails) ;
- Water facilities (fountains and bathing areas) ;

We will apply this analysis to the city of Paris.

Some publication already showed that Paris is not one of the greenest city in the world even if the municipality try to improve it. Only 4 out of 10 Parisians have an access to a green area while green areas are considered an important criterion for housing for 85% of people according to a French study.

DATA

All venues that will be analysed are available on FourSquare. The list of the venues references is detailed below :

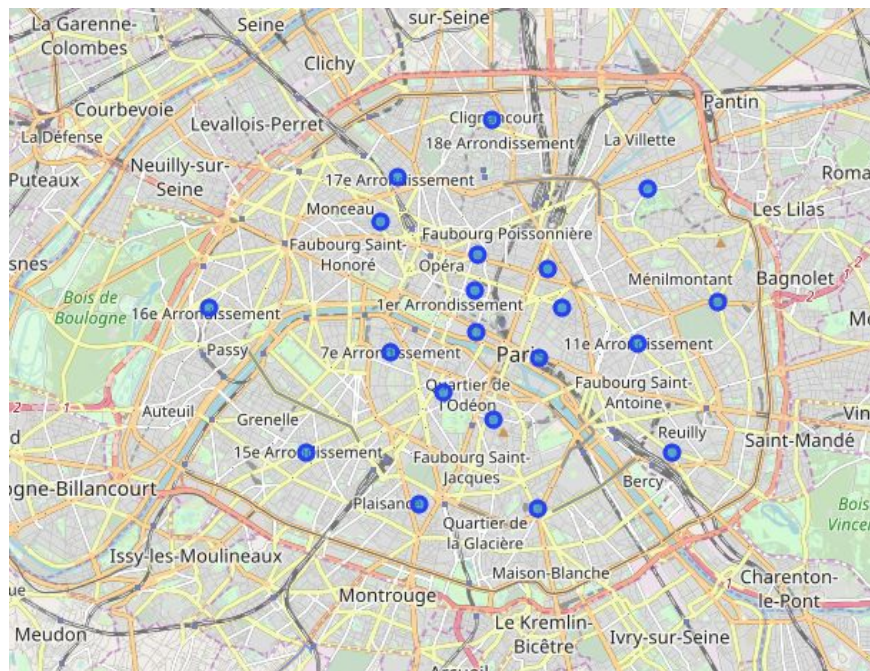
- Gardens (4bf58dd8d48988d15a941735)
- Bathing areas (52e81612bcb57f1066b7a28)
- Fountains (56aa371be4b08b9a8d573547)
- Pedestrian areas (52e81612bcb57f1066b7a25)
- Bike trails (56aa371be4b08b9a8d57355e)

METHODOLOGY SECTION

1) Getting the location of the boroughs

The objective is to create a relative indicator of environmental facilities in Paris. For each criterion, we will compare the twenty boroughs of the city of Paris. To do so, we have to find the location of the borough. The reference will be the location of the borough hall. We use the FourSquare API to do so and the category “Townhall” (4bf58dd8d48988d129941735).

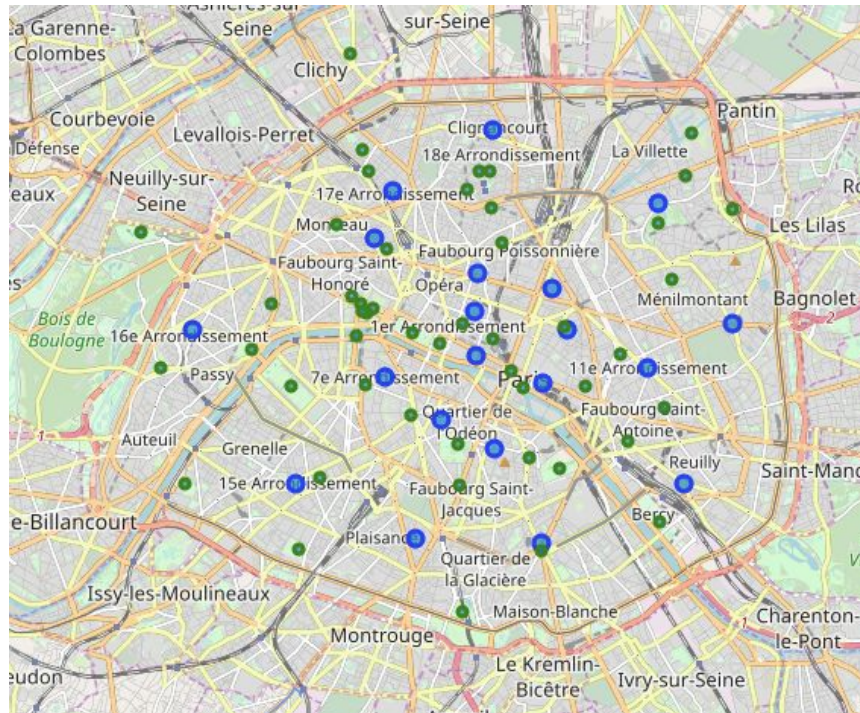
The result contain the twenty borough halls but also other result. We have to clean the data by selecting the entries which name start with “Mairie du”. We only keep the name of the borough and the location of the borough hall and create a map of Paris with Folium to show the twenty locations :



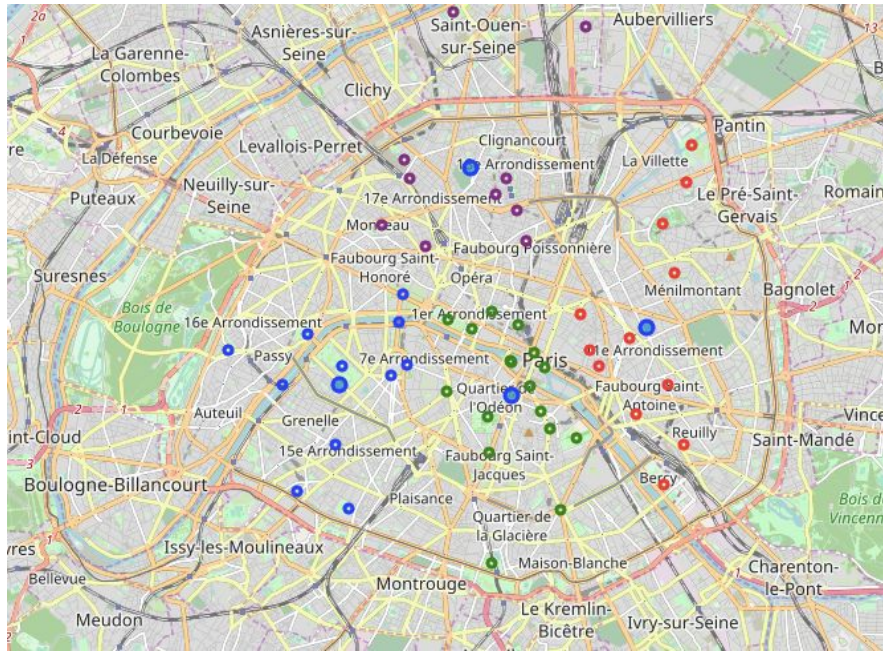
2) Exploring the parks of Paris.

We want to visualize the location of the parks and gardens in Paris to have an idea of their distribution and the accessibility from the different boroughs. We use the Foursquare API to get the maximum datas from the center of Paris. Then we use Folium to add the parks on the previous map.

We can see that the accessibility of a park can be very different from one borough to another.



We try to make clusters of the parks using K-means clustering. The relevant number of clusters here is 4. Then, we print a map of the clusters and their centroids with different color on the map of Paris, using Folium. We do not keep the borough halls on the map to make it easier to read. This map does not show any relevant difference between the cluster. The red one seems to have a lower density but it cannot allow us to conclude. We will try another type of analysis to help us decide which boroughs are the best.



3) Getting the distance for each borough and each facility

For each of the twenty borough, we search for the distance between the borough and the nearest facility for the following facilities :

- Green areas
 - Gardens
 - Pedestrian areas
 - Bike trails
- Water facilities
 - Bathing areas
 - Fountains

Unfortunately, FourSquare does not always find the nearest venue when you limit the research to one venue. Furthermore, there is a risk to have no venue found if the radius is too low.

As a result, we create a function named “Closest_venue”. The method is to :

- Define a radius large enough to always have results. In Paris, we use a radius of 4500 meters, to be sure.
- Ask for the maximum venues to be confident to have the closest inside.
- Find the closest venue using the “location.distance”.

Then, we create another function “Distance_matrix” calling “Closest_venue” on the different borough locations and the different facilities to create a Dataframe of the nearest venues for each location and each type of venues.

```
1 df_test.head()
```

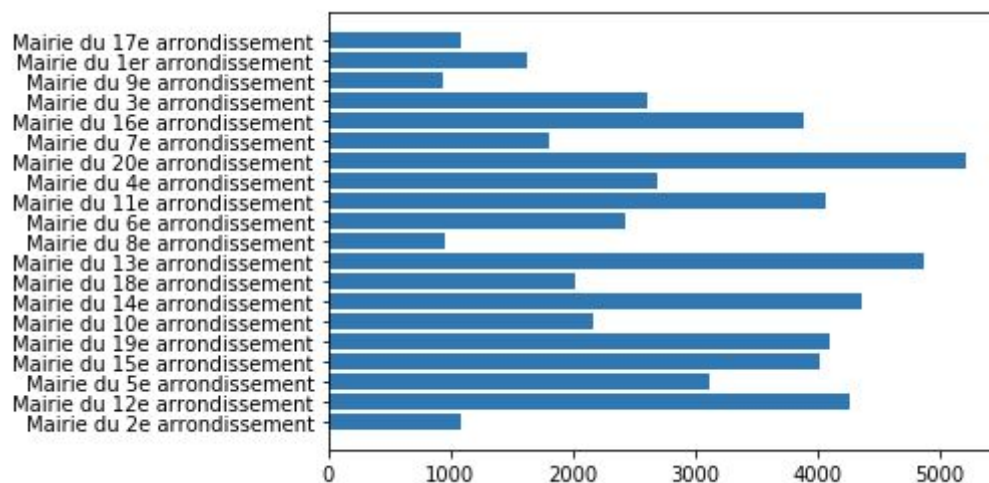
	4bf58dd8d48988d15a941735	56aa371be4b08b9a8d57355e	52e81612bcbc57f1066b7a25	56aa371be4b08b9a8d573547	52e81612bcbc57f1066b7a28
name					
Mairie du 2e arrondissement	284	1081	384	241	2934
Mairie du 11e arrondissement	500	4072	790	831	1552
Mairie du 9e arrondissement	836	942	77	504	2416
Mairie du 12e arrondissement	129	4271	885	67	2512
Mairie du 14e arrondissement	931	4369	584	1070	1653

4) Testing different indicators

Finally, we can classify the boroughs with different indicators.

a) Absolute indicator

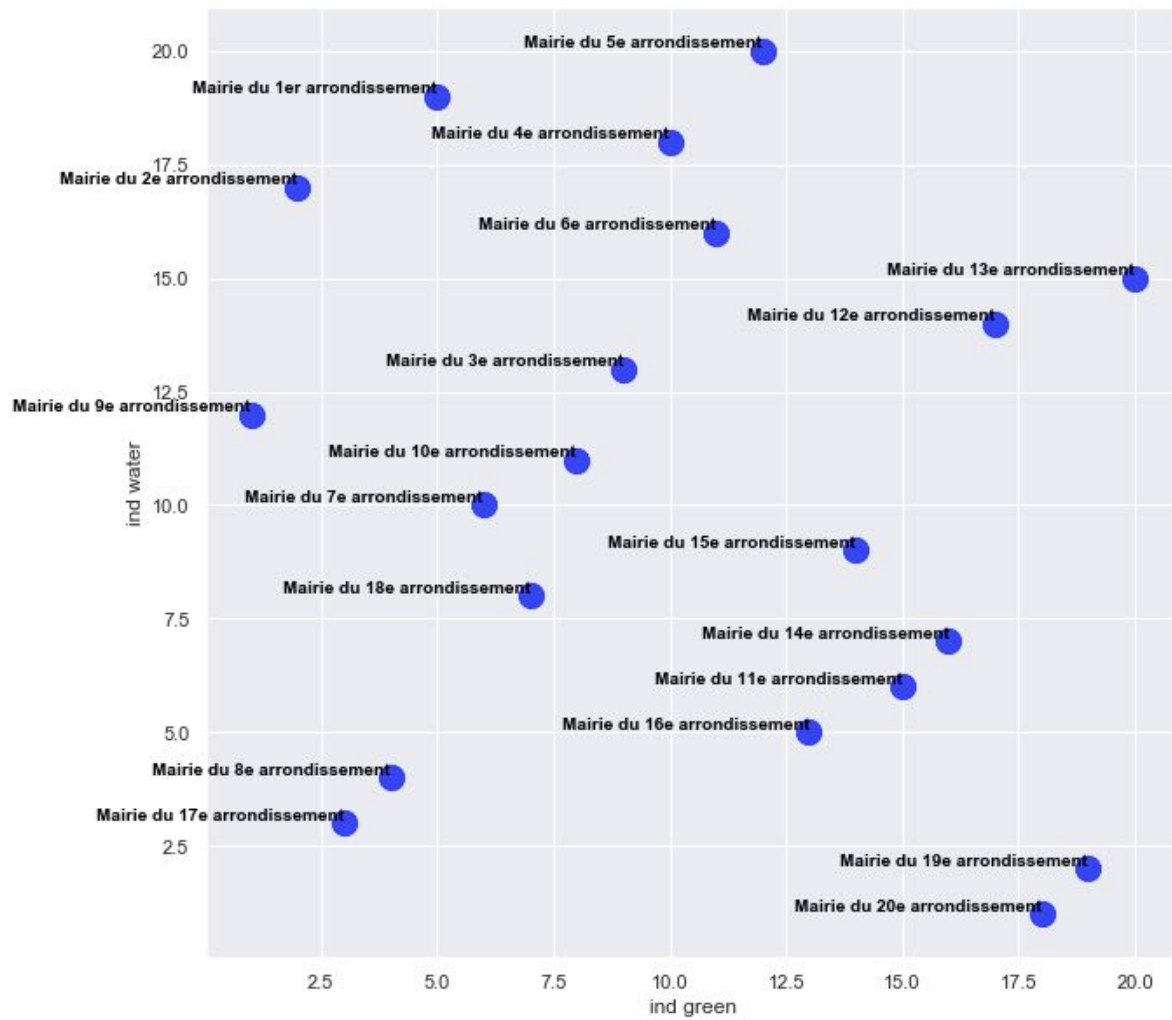
We try first an absolute indicator based on the average distance from the borough to the venues. The shortest distance will be the best and the longest will be the worst.



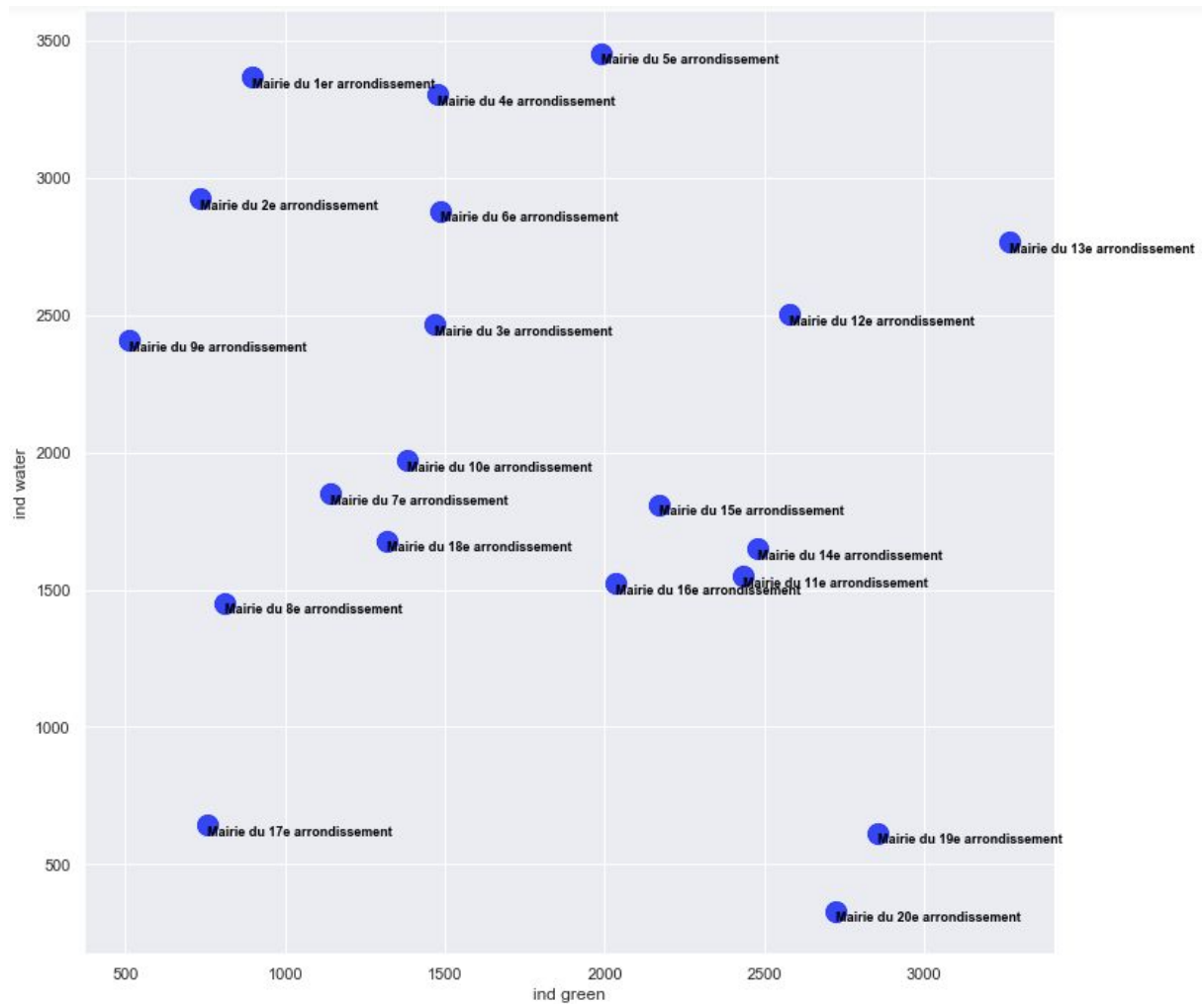
b) Relative indicators

Then we try two relative indicators to rank the boroughs based on the two aspect : Green venues and water facilities.

The first is a based on the rank of the borough (from the closest to the farrest) from 1 to 20 for each aspect. We plot the results.



The second is based on the average distance from the boroughs to the different types of venues. It is more precise because it shows the gap distance between two boroughs for each of the category venues. We plot the result.



5) Classification

Finally, based on the last indicator, we try to classify the boroughs to determine typologies. We use k-means clustering method. The best number of cluster is 4.



RESULTS SECTION

There are some hypothesis that can highly influence on the results :

Each borough is evaluated on the situation of one only location, the location of the borough hall. Some of the borough are very large and the results might be very different with another location.

Other facilities could be used and give different results. Green facilities and water facilities seemed to be good to analyse but some might find other ones useful too (like access to an hospital).

Of course, environment is much more than access to facilities. We could add analysis based on pollution measures (air pollution, noise pollution) and aggregate theses information to the previous indicators.

DISCUSSION SECTION

where you discuss any observations you noted and any recommendations you can make based on the results.

Based on the results, we can recommend to invest in the following boroughs for people looking for an environmental-friendly district :

17ème arrondissement, 8ème arrondissement, 18ème arrondissement, 7ème arrondissement, 10ème arrondissement, 9ème arrondissement.

People looking for green venues can also invest in the :

1ème arrondissement, 2ème arrondissement, 3ème arrondissement, 4ème arrondissement, 5ème arrondissement, 6ème arrondissement

People looking for water facilities can also invest in the :

19ème arrondissement, 20ème arrondissement

The following districts must be avoid, based on this criterion :

11ème arrondissement, 12ème arrondissement, 13ème arrondissement, 14ème arrondissement, 15ème arrondissement, 16ème arrondissement

CONCLUSION

This simple analysis shows that the ease of access to environmental benefits is very different from one district to another in a big city like Paris. This information could be used by real estate agents and owners to value real estate according to environmental criteria and thus encourage city politicians to make efforts for the environment.