# Strategic Analytics Research about Wellness and Real Estate in the City of Paris

Student Name: Moncef Salmi

Submission Date: 19/06/19

Framework of the study

o The Olympic games for 2024 are organized in the city of Paris

o Different stakeholders are highly interested in the real estate status in Paris

o Exemple of evolved stakeholders that are interested by such study are but not limited to:

   o Investors

   o French Govermental organizations

   o International olympic committees

   o Etc
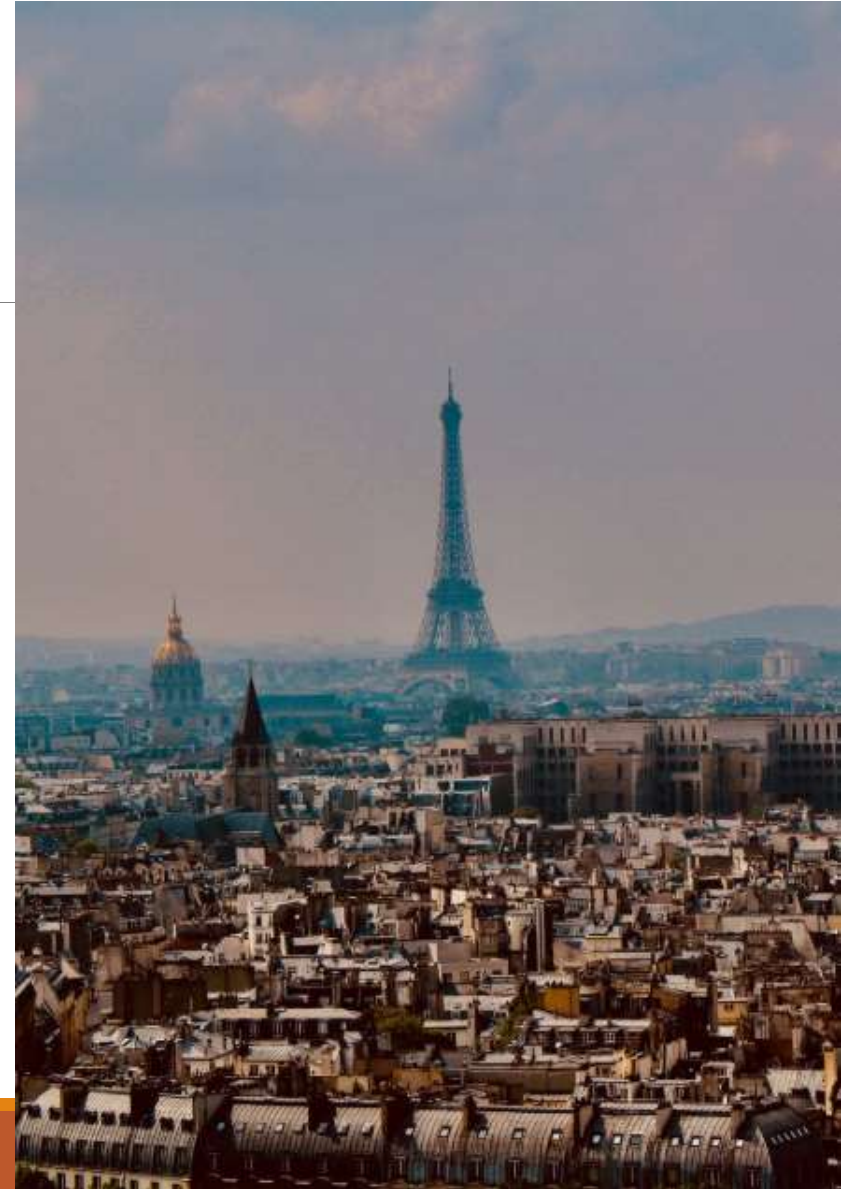
Reserach question

o Main Research Question to be answered: "Can we assess qualitatively and quantitatively the different boroughs of the city of Paris from a real estate point of view based on their Wellness score ?"

o The conducted study should not be a classical real estate assessment since it should take into account paricular constraints and KPIs when evaluating the best opportunities such as

  o Quality/Price ratio of the studied neighborhood in Paris

  o Global quality of life

  o Vicinity of Wellness accomodation in the considered neighborhood

Data Science Methodology

In order to answer to the stated research question, the folowing data science methodology is implemented

1. Analytic approach

2. Data requirements

3. Data collection

4. Data understanding and preparation

5. Modeling and evaluation

Analytic Approach

o Descriptive and disgnostic approaches are adopted

o Visualization of the collected data in order to have a first insight into i

- o Box plots
- o Scatter plot
- o Regression plots
- o Choropleth Map

o Machine learning modeling: Kmeans clustering will be used in order to assess particular patterns among the collected data

Data requirements

o Basic real estate data about the city of Paris. This data should contain the different boroughs, neighbourhoods, IDs and post codes of the city of Paris.

o Data corresponding to the house price of each borough of the city of Paris.

o Data corresponding to online crowd sourcing evaluating the score that current and former residents of the different considered borough of the city of Paris gave to them.

o Geojson file corresponding to the city of Paris and its constitutive boroughs is needed in order to be able to generate insightful Choropleth Maps,

Data collection

o Basic real estate data about the city of Paris
  o www.annuaire-administration.com

o Data corresponding to the house price of each borough of the city of Paris
  o www.meilleursagents.com

o Data corresponding to online crowd sourcing rating each Parisian borough
  o www.ville-ideale.fr

o Geojson file corresponding to the city of Paris
  o opendata.paris.fr

	A	B	C
	postCode	Borough	Neighborhood
	75001	Paris 1er Arrondi	Paris 1er Arrondi
	75002	Paris 2e Arrondis	Paris 2e Arrondis
	75003	Paris 3e Arrondis	Paris 3e Arrondis
	75004	Paris 4e Arrondis	Paris 4e Arrondis
	75005	Paris 5e Arrondis	Paris 5e Arrondis
	75006	Paris 6e Arrondis	Paris 6e Arrondis
	75007	Paris 7e Arrondis	Paris 7e Arrondis
	75008	Paris 8e Arrondis	Paris 8e Arrondis
	75009	Paris 9e Arrondis	Paris 9e Arrondis
	75010	Paris 10e Arrond	Paris 10e Arrond
	75011	Paris 11e Arrond	Paris 11e Arrond
	75012	Paris 12e Arrond	Paris 12e Arrond
	75013	Paris 13e Arrond	Paris 13e Arrond
	75014	Paris 14e Arrond	Paris 14e Arrond
	75015	Paris 15e Arrond	Paris 15e Arrond
	75016	Paris 16e Arrond	Paris 16e Arrond
	75017	Paris 17e Arrond	Paris 17e Arrond
	75018	Paris 18e Arrond	Paris 18e Arrond
	75019	Paris 19e Arrond	Paris 19e Arrond
	75020	Paris 20e Arrond	Paris 20e Arrond

Raw data

Data Understanding and Preparation

o Based on the former data sources, the initial data frame containing the basic needed data is built

o The main 2 basic features are
  o Average house price per neighborhood
  o Crowd rating per neighborhood

o Additional insightful features are added such as the mixed product of the 2 main features that assesses the quality/price ratio for each neighborhood

o geographical coordinates are added for each neighborhood

|	postCode	Borough	Neighborhood	avgHousePrice	croudRating
0	75001	Paris 1er Arrondissement	Paris 1er Arrondissement	12436	6,85
1	75002	Paris 2e Arrondissement	Paris 2e Arrondissement	11214	6,31
2	75003	Paris 3e Arrondissement	Paris 3e Arrondissement	12140	8,45
3	75004	Paris 4e Arrondissement	Paris 4e Arrondissement	12906	6,82
4	75005	Paris 5e Arrondissement	Paris 5e Arrondissement	11965	8,13

Initial data frame

|	postCode	Borough	Neighborhood	avgHousePrice	croudRating	Latitude	Longitude
0	75001	Paris 1er Arrondissement	Paris 1er Arrondissement	12436	6.85	48.863512	2.338962
1	75002	Paris 2e Arrondissement	Paris 2e Arrondissement	11214	6.31	48.865300	2.351360
2	75003	Paris 3e Arrondissement	Paris 3e Arrondissement	12140	8.45	48.862666	2.360259
3	75004	Paris 4e Arrondissement	Paris 4e Arrondissement	12906	6.82	48.860845	2.352929
4	75005	Paris 5e Arrondissement	Paris 5e Arrondissement	11965	8.13	48.845812	2.348651

Adding geographic coordinates

Exploratory Data Analysis: Box plots

o Box plot of the average house price shows that
  o House prices in Paris are relatively high
  o A high discrepancy can be seen

o Box plot of the crowd rating
  o Median value is at 7/10 rating
  o Minimum is at 4.8/10 and maximum at 8/10
 o A high discrepancy can be seen
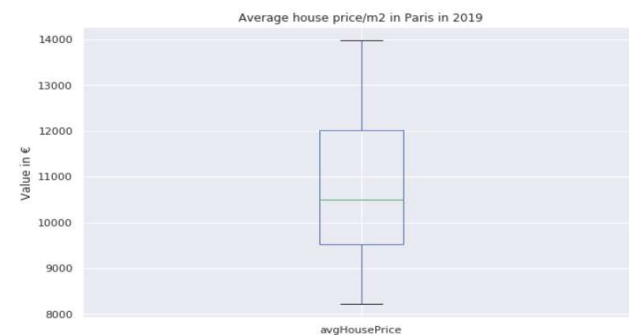
Box plot of the average house price
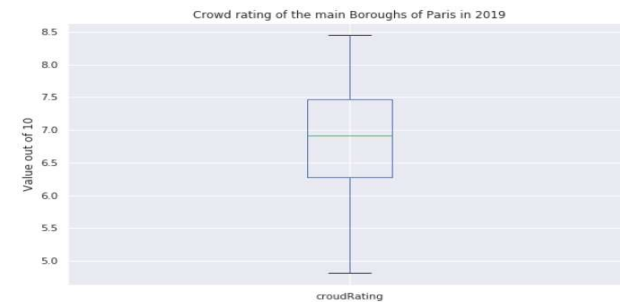
Box plot of the crowd rating

Exploratory Data Analysis: Choropleth Map

o The best opportunities correspond to configurations where
  o Crowd rating index is high
  o House price value is low

o Based on this assumption, western side of Paris can be expected to be a good real estate opportuntiy

o Let's remind that at this level the vicinity from wellness acoommodations is not yet taken into account

o Further upcoming studies will quantify these first qualitative insights
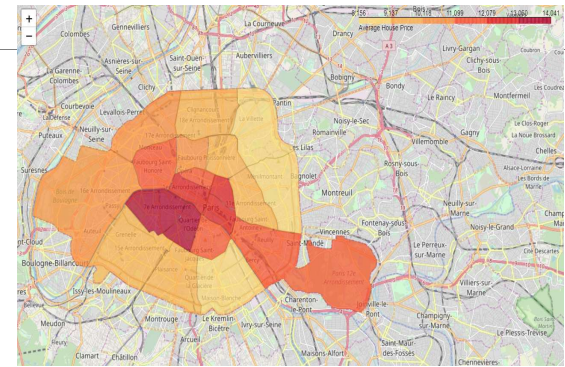
Choropleth Map of average house price
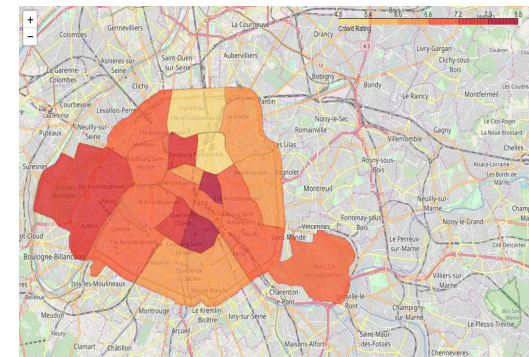
Choropleth Map of crowd rating

Exploratory Data Analysis: Correlation

o A positive and moderate correlation can be seen between crowd rating and house price

o This is consistent with intuition since people pay more for a better quality

o The moderate correlation predicts good opportunities

o In fact, relatively low correlation between these 2 features predicts that high quality houses are not necessarily the most expensive

o This is confirmed by the upper left handside zone of the scatter plot: high crowd rating Vs relative low prices


Regression plot of the main features

Exploratory Data Analysis: Adding insightful features

o The initial data is post processed and a new feature reporting the quality/price ratio is computed

o The Choropleth Map of the new feature is displayed

o This new feature shows that the best real estate opportunities are in the western, southern and north eastern sides of Paris

o Let's remind that at this level the vicinity from wellness acoommodations is not yet taken into account

o This quantitative result confirms en enriches the previous qualitative outcome

	postCode	Borough	Neighborhood	avgHousePrice	croudRating	Latitude	Longitude	avgHousePrice_binned	boroughCustomScore
11	75012	Paris 12e Arrondissement	Paris 12e Arrondissement	9109	7.52	48.839734	2.380054	8	6.016
18	75019	Paris 19e Arrondissement	Paris 19e Arrondissement	8214	6.17	48.876829	2.394105	9	5.553
19	75020	Paris 20e Arrondissement	Paris 20e Arrondissement	8537	6.03	48.857126	2.409257	9	5.427
12	75013	Paris 13e Arrondissement	Paris 13e Arrondissement	8874	6.52	48.829357	2.362456	8	5.216
14	75015	Paris 15e Arrondissement	Paris 15e Arrondissement	9627	7.16	48.842884	2.277391	7	5.012

Ranking based on mixed score
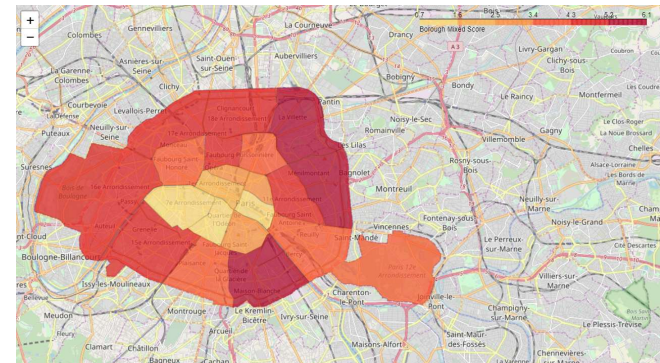
Choropleth Map based on mixed score

Wellness venues in Paris: Building data

o Based on Foursquare API, the wellness venues in Paris are built for each Parisian neighbourhood

o The resulting data frame contains the Venue names and venue categories for each neighborhood

o This is the raw data that will be leveraged in order to add the vicinity to Wellness accommodations in the performed study

```
paris_venues_wellness.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Paris 1er Arrondissement	48.863512	2.338962	Jardin du Palais Royal	48.864941	2.337728	Garden
1	Paris 1er Arrondissement	48.863512	2.338962	Thémaé	48.863589	2.339756	Spa
2	Paris 1er Arrondissement	48.863512	2.338962	Colonnes de Buren	48.863618	2.336917	Sculpture Garden
3	Paris 1er Arrondissement	48.863512	2.338962	Lémoni	48.864795	2.341292	Salad Place
4	Paris 1er Arrondissement	48.863512	2.338962	Cour Carrée du Louvre	48.860360	2.338543	Pedestrian Plaza

Paris venues for Wellness

Wellness venues in Paris: Final Score

o Based on the Wellness venues, a new score called « Wellness Score » is computed for each neighborhood

o Based on « Wellness Score » and the former mixed « Quality/Price » score, a «Final Score» is computed for each neighborhood

o The corresponding ranking w.r.t. «Final Score» feature shows that the top 3 Parisian neighborhoods are the 13th, 14th and 11th neighborhoods

```
df.sort_values(by=['finalScore'], inplace=False, ascending=False).head(5)
```

|	postCode	Borough	Neighborhood	avgHousePrice	croudRating	Latitude	Longitude	avgHousePrice_binned	boroughCustomScore	wellnessScore	finalScore
12	75013	Paris 13e Arrondissement	Paris 13e Arrondissement	8874	6.52	48.829357	2.362456	8	5.216	10.00	5.216000
13	75014	Paris 14e Arrondissement	Paris 14e Arrondissement	9912	6.06	48.828590	2.307541	7	4.242	10.00	4.242000
10	75011	Paris 11e Arrondissement	Paris 11e Arrondissement	10021	6.63	48.855630	2.370806	7	4.641	7.50	3.480750
1	75002	Paris 2e Arrondissement	Paris 2e Arrondissement	11214	6.31	48.865300	2.351360	5	3.155	8.75	2.760625
17	75018	Paris 18e Arrondissement	Paris 18e Arrondissement	9188	5.38	48.896511	2.334311	8	4.304	6.25	2.690000

Building the final score column and corresponding ranking

Wellness venues in Paris: Choropleth Maps

o The Choropleth Maps for « Wellness Score » and «Final Score» are reported

o The adopted scoring method or rating criterion has a direct impact on the final assessment of the real estate opportunities

o In fact, the center of Paris is very well rated when taking into account « Wellness Score » but its rating drops drastically when taking into account the «Final Score»

o This is because «Final Score» is enriched by the information regarding the house price of the considered zone which is a major weak point for the center of Paris since it's on of the most expensive Parisian areas
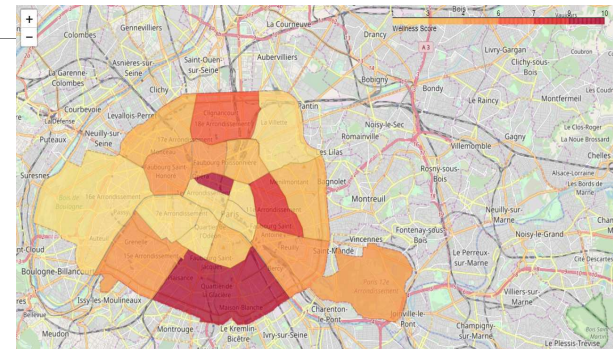

Choropleth Map of Wellness Score
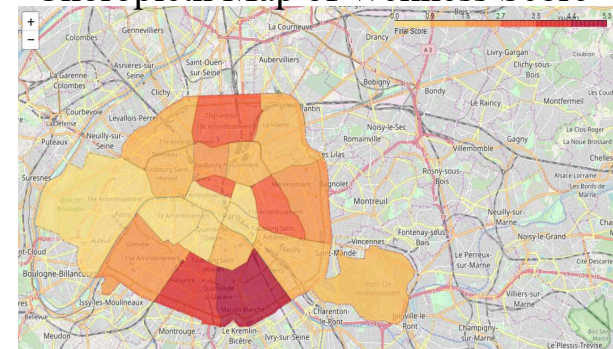
Choropleth Map of Final Score
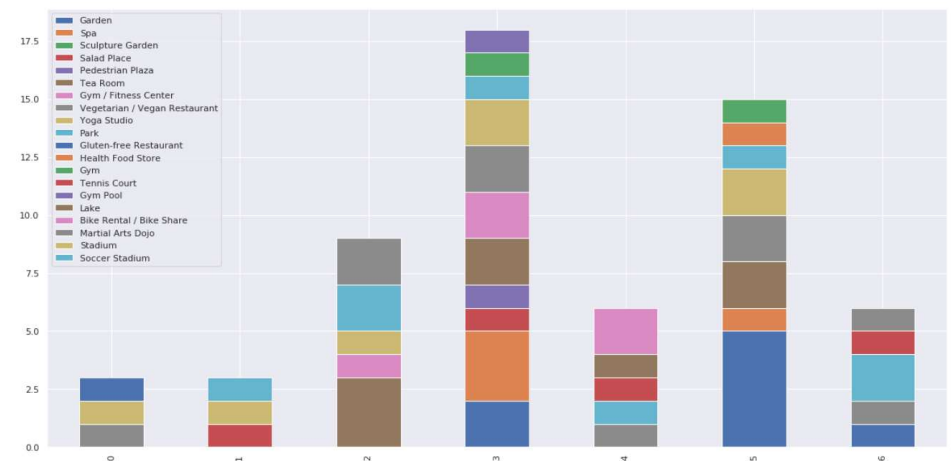
Predictive modeling: Kmeans Clustering

o An Elbow analysis is performed in order to define the best K value for clustering the wellness venues in Paris

o The value of K=7 is selected

o The selected value corresponds to a good trade-off between minimizing the error and avoiding an exaggerated number of clusters
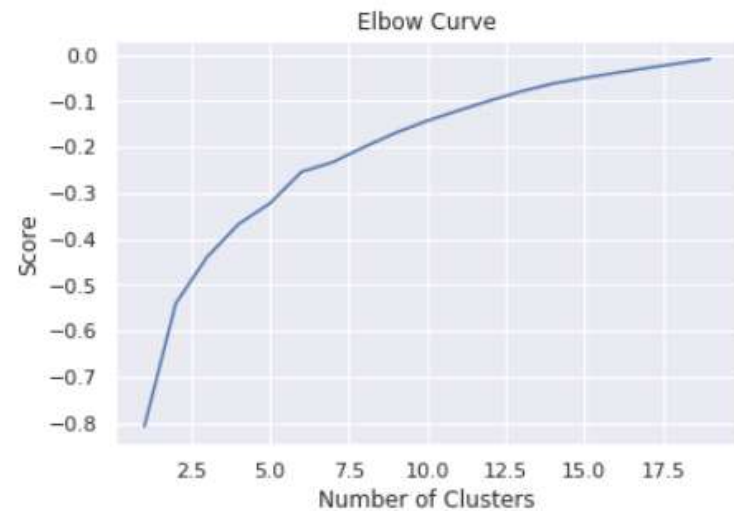
Elbow analysis

Predictive modeling: Clusters interpretation

o In order to be able to better interpret the predicted clusters by the Kmeans algorithm, a bar chart is reported

o Based on the bar chart results, the different formed clusters can be interpreted as follows:

   o Cluster 0: Asian wellness

   o Cluster 1: Urban wellness

   o Cluster 2: Healthy food

   o Cluster 3: Wellness Paradise

   o Cluster 4: Sport and healthy activity

   o Cluster 5: Spiritual Wellness

   o Cluster 6: Natural resourcing

Bar chart of the built Kmeans clusters

Predictive modeling: Kmeans clusters displayed on the Parisian Map

o The built clusters are reported on a Folium map.

o The final score for each neighbourhood is also reported as label.

o This makes it possible for the stakeholders of the conducted study to have access to a high level, yet deeply insightful, analysis within a simple click
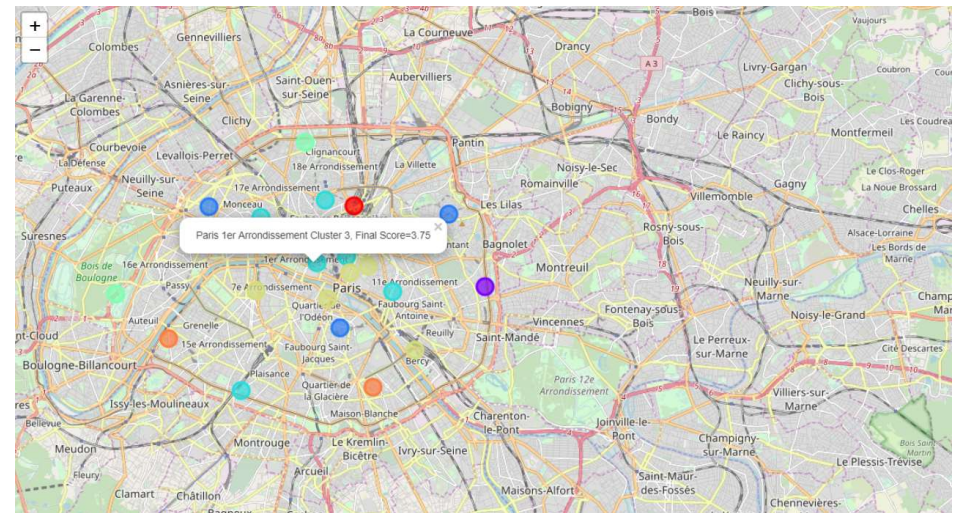
Folium Map reporting the built Kmeans clusters

Conclusions and future perspectives

o The main answers based on the conducted study are the following:

- o The south of Paris is definitively the best place to be when considering all the 3 constraints of price, quality of life and wellness accommodation in the nearby

- o The top 3 Parisian neighborhoods are the 13th, 14th and 11th neighborhoods as reported previously

- o The conducted study can be used in a more general way by real estate stakeholders that do not have a specific interest for wellness accommodations.

- o The ranking of the neighborhoods changes considerably based on the considered criterion. This is where Data science has a high added value since it brings accurate and quantitative evaluation of a given question like the one asked at the beginning of the current study

- o Based "quality/price" analysis and its corresponding Choropleth Map the following answers are obtained:

  - o Southern and Eastern zones are the best in Paris

  - o The top 3 Parisian neighborhoods are the 12th, 19th and 20th neighborhoods

o The implemented tools can be used for further investigations for other similar cities and have more insight at the real estate opportunities given several constraints like the ones considered in the performed study

References

o https://www.annuaire-administration.com/code-postal/region/ile-de-france.html

o https://www.meilleursagents.com

o https://www.ville-ideale.fr/paris-1er-arrondissement_75101

o https://opendata.paris.fr

o https://www.businessinsider.fr/us/worlds-expensive-richest-real-estate-markets-hong-kong-london-2018-12

o https://fr.foursquare.com/

o www.pexels.com is the source for the non scientific pictures of the current report