Paris Venues and real estate prices analysis

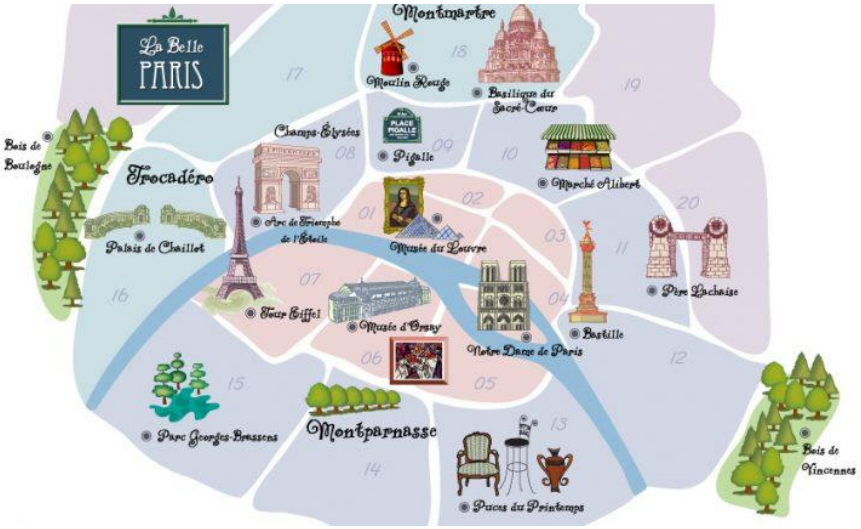
The report was written in relation with the IBM Data science Capstone project. The objective of the report is to apply data science concepts learned in the course to real data and present it in a formal report. For my report I choose to implement the different techniques as k-means clustering to the real estate market and venues of Paris.

We can found important steps of a data science analysis in my report :

- The problem statement
- Data gathering
- Visualization
- Clustering
- Conclusion

Problem statement

Paris is an ancient city dating back to the Roman's period, still growing which reached 2.2 million inhabitants only "intra-muros" in 2019 and is around 10 millions inhabitants with the metropolitan area. In addition to its famous places to visit, Paris is also known for its twenty boroughs numbered from the centre to the border as a snail as we can see below:



Paris is made of a large diversity venue in different boroughs and it is sometimes difficult to know where to buy a property. The prices of properties considerably increased the last decade and represent a huge investment. That is why we want to know, considering our interests and the amount we can spend, the best place to buy a property.

The aim of the project is to find out what are the most popular venues in the different 20 boroughs and to cluster boroughs in order to find similarities. I also analysed the average price per square meter per borough for properties sold in 2019 until now. The analysis could serve as a help to decide for people that would decide to move to Paris and choose the best place corresponding to their interests and the amount they could spend for a property.

Data gathering and pre-processing

I found the prices data on the DVF https://www.data.gouv.fr/fr/datasets/demandes-de-valeurs-foncieres/. Data could be downloaded in a csv format and regrouped all real estate transactions that occurred during 2019. I highly preprocessed it in a separate file because of the important number of steps needed to keep only apartments and houses transactions and to clean transaction in duplicate. In addition to this pre-processing, I cleaned the data to keep only Paris intra-muros transaction, represented by the region number 75, and eliminated abnormal transactions as below 100k or more than 15M which could be considered as outliers for Paris market according to multiple sources found on the internet (https://www.globalpropertyguide.com/Europe/France/Price-History),(https://parispropertygroup.com/blog/2018/5-year-review-of-paris-real-estate-price-evolution/)

Data second pre-processing in order to have clean data to use for clustering:

	Date mutation	Valeur fonciere	No voie	Type de voie	Voie	Code postal	Commune	Code type local	Type local	Surface reelle bati	Nombre pieces principales	Price/sqm2
0	02/01/2019	180000.0	66	RUE	DU FBOURG MONTMARTRE	75009	PARIS 09	2	Appartement	25	2	7200.00
1	02/01/2019	180000.0	9	PAS	SAINT MICHEL	75017	PARIS 17	2	Appartement	25	2	7200.00
2	02/01/2019	99000.0	98	RUE	SAUSSURE	75017	PARIS 17	2	Appartement	13	1	7615.38
3	02/01/2019	440000.0	36	AV	D ITALIE	75013	PARIS 13	2	Appartement	57	2	7719.30
4	02/01/2019	635000.0	16	RUE	ANTOINE ROUCHER	75016	PARIS 16	2	Appartement	78	4	8141.03

Data after cleaning, and renaming the columns in English :

```
1 df_paris.head()
```

	Date	price	adress_num	street_type	street	postcode	borough	estate_type_code	estate_type	surface	room_number	price_per_sqm
0	02/01/2019	180000.00	66	RUE	DU FBOURG MONTMARTRE	75009	PARIS 09	2	Appartement	25	2	7200.00
1	02/01/2019	180000.00	9	PAS	SAINT MICHEL	75017	PARIS 17	2	Appartement	25	2	7200.00
2	02/01/2019	99000.00	98	RUE	SAUSSURE	75017	PARIS 17	2	Appartement	13	1	7615.38
3	02/01/2019	440000.00	36	AV	D ITALIE	75013	PARIS 13	2	Appartement	57	2	7719.30
4	02/01/2019	635000.00	16	RUE	ANTOINE ROUCHER	75016	PARIS 16	2	Appartement	78	4	8141.03

I grouped the transactions by borough, and calculated the price per square meter with the function groupby, and mean.

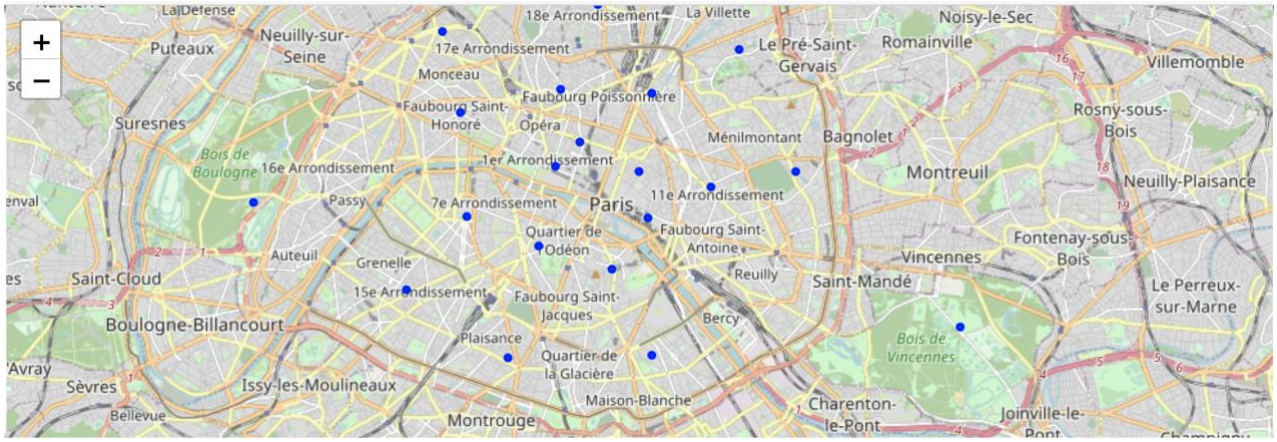
	borough	postcode	price_per_sqm
0	PARIS 01	75001	51743.16
1	PARIS 02	75002	17455.89
2	PARIS 03	75003	46959.52
3	PARIS 04	75004	21889.82
4	PARIS 05	75005	16580.28
5	PARIS 06	75006	19264.15
6	PARIS 07	75007	20559.21
7	PARIS 08	75008	24198.68
8	PARIS 09	75009	20859.98
9	PARIS 10	75010	22931.74
10	PARIS 11	75011	15371.60
11	PARIS 12	75012	11780.43
12	PARIS 13	75013	14728.84
13	PARIS 14	75014	16035.64
14	PARIS 15	75015	17322.84
15	PARIS 16	75016	18104.15
16	PARIS 17	75017	15631.00
17	PARIS 18	75018	25410.04
18	PARIS 19	75019	36809.44
19	PARIS 20	75020	27052.85

Then I join the table above with geographical coordinates in order to visualize the points on a Paris map :

```
df_borough[['Latitude','Longitude']]=pd.DataFrame(coords_list,columns=['Latitude', 'Longitude'])
df_borough.head()
```

	borough	postcode	price_per_sqm	Latitude	Longitude
0	PARIS 01	75001	51743.16	48.86	2.34
1	PARIS 02	75002	17455.89	48.87	2.34
2	PARIS 03	75003	46959.52	48.86	2.36
3	PARIS 04	75004	21889.82	48.85	2.36
4	PARIS 05	75005	16580.28	48.85	2.35

I then used the Folium library as we have seen during the courses to visualize the different points where we find the twenty Paris boroughs :



Following to the visualization, the aim of the project was to explore and cluster the different boroughs that we have in Paris. In order to explore the most visited venues in each borough I used the Foursquare API. I choose as parameters to explore the 100 most visited places in 500meters from the point of interest as we know that in term of dimension Paris intra muros is only 12x9 km.

For example, we can see below the venue exploration for the PARIS 01 borough, a famous borough known for its tourism places to visit.

	name	categories	lat	lng
0	Jardin du Palais Royal	Garden	48.86	2.34
1	Palais Royal	Historic Site	48.86	2.34
2	Place du Palais Royal	Plaza	48.86	2.34
3	Comédie-Française	Theater	48.86	2.34
4	Place Colette	Plaza	48.86	2.34

The same procedure was done for the other Paris boroughs and put into a dataframe, we have 20 boroughs in total :

	borough	borough Latitude	borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	PARIS 01	48.86	2.34	Jardin du Palais Royal	48.86	2.34	Garden
1	PARIS 01	48.86	2.34	Palais Royal	48.86	2.34	Historic Site
2	PARIS 01	48.86	2.34	Place du Palais Royal	48.86	2.34	Plaza
3	PARIS 01	48.86	2.34	Comédie-Française	48.86	2.34	Theater
4	PARIS 01	48.86	2.34	Place Colette	48.86	2.34	Plaza
5	PARIS 01	48.86	2.34	La Clef Louvre Paris	48.86	2.34	Hotel

I then transformed the dataframe and created a dataframe with the 100 most visited venues in a perimeter of 500meters

	borough	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
0	PARIS 01	French Restaurant	Japanese Restaurant	Hotel	Plaza	Coffee Shop	Italian Restaurant	Café	Historic Site	Bar	Art Museum
1	PARIS 02	French Restaurant	Cocktail Bar	Wine Bar	Italian Restaurant	Bistro	Salad Place	Hotel	Thai Restaurant	Bakery	Furniture / Home Store
2	PARIS 03	Coffee Shop	Bakery	Bistro	Burger Joint	French Restaurant	Art Gallery	Clothing Store	Café	Cocktail Bar	Sandwich Place
3	PARIS 04	French Restaurant	Hotel	Clothing Store	Italian Restaurant	Ice Cream Shop	Bakery	Plaza	Wine Bar	Tapas Restaurant	Coffee Shop
4	PARIS 05	French Restaurant	Bar	Hotel	Italian Restaurant	Coffee Shop	Bakery	Creperie	Café	Wine Bar	Pub

Clustering

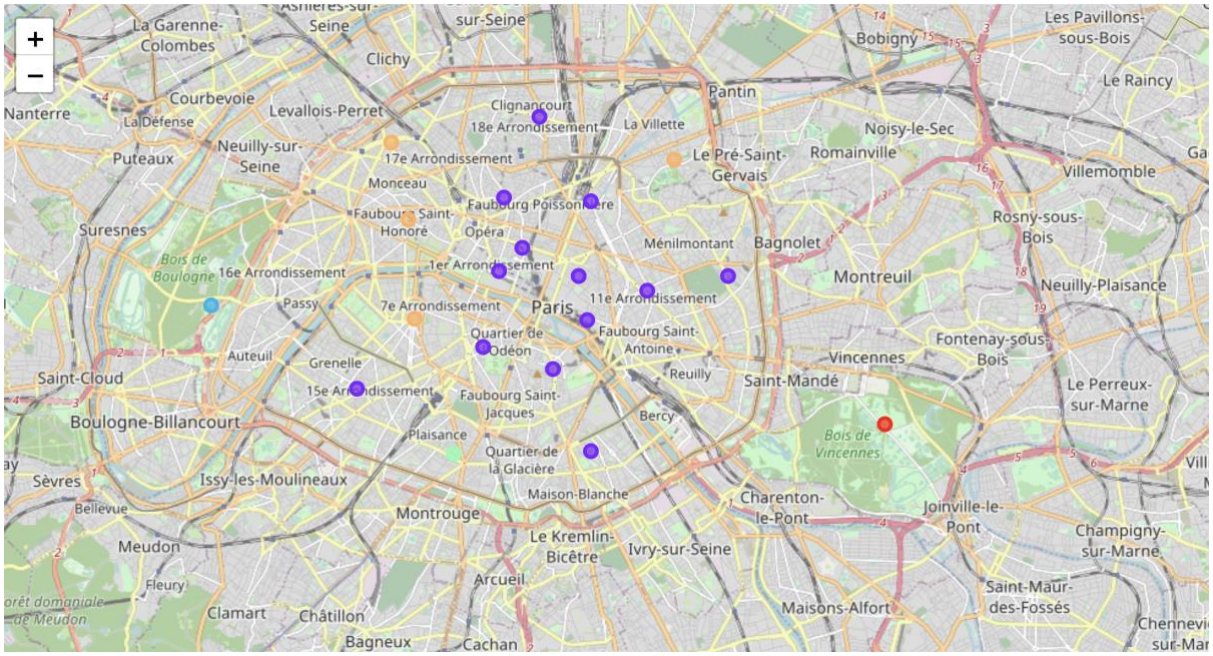
Finally, I used the Machine Learning concepts we have seen in the course to cluster the Paris borough. I used one hot encoding to understand the most visited categories of venue per borough in order to find similarities to cluster them.

	borough	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
0	PARIS 01	French Restaurant	Japanese Restaurant	Hotel	Plaza	Coffee Shop	Italian Restaurant	Café	Historic Site	Bar	Art Museum
1	PARIS 02	French Restaurant	Cocktail Bar	Wine Bar	Italian Restaurant	Bistro	Salad Place	Hotel	Thai Restaurant	Bakery	Furniture / Home Store
2	PARIS 03	Coffee Shop	Bakery	Bistro	Burger Joint	French Restaurant	Art Gallery	Clothing Store	Café	Cocktail Bar	Sandwich Place
3	PARIS 04	French Restaurant	Hotel	Clothing Store	Italian Restaurant	Ice Cream Shop	Bakery	Plaza	Wine Bar	Tapas Restaurant	Coffee Shop
4	PARIS 05	French Restaurant	Bar	Hotel	Italian Restaurant	Coffee Shop	Bakery	Creperie	Café	Wine Bar	Pub

After calculating the optimal k , I found that the best suiting k was k=5 so 5 clusters, I will then generate cluster labels and include it in my dataframe, we can note in the illustration below the addition of the column Cluster Labels :

	borough	postcode	price_per_sqm	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
0	PARIS 01	75001	51743.16	48.86	2.34	1	French Restaurant	Japanese Restaurant	Hotel	Plaza	Coffee Shop	Italian Restaurant	Café	Historic Site
1	PARIS 02	75002	17455.89	48.87	2.34	1	French Restaurant	Cocktail Bar	Wine Bar	Italian Restaurant	Bistro	Salad Place	Hotel	Thai Restaurant
2	PARIS 03	75003	46959.52	48.86	2.36	1	Coffee Shop	Bakery	Bistro	Burger Joint	French Restaurant	Art Gallery	Clothing Store	Café
3	PARIS 04	75004	21889.82	48.85	2.36	1	French Restaurant	Hotel	Clothing Store	Italian Restaurant	Ice Cream Shop	Bakery	Plaza	Wine Bar
4	PARIS 05	75005	16580.28	48.85	2.35	1	French Restaurant	Bar	Hotel	Italian Restaurant	Coffee Shop	Bakery	Creperie	Café

I then used the concepts on Folium learned in the IBM course to create a map of the clustered boroughs :



Finally I displayed the different clusters in my notebook and giving them a category. In addition I included the price per square meter for each borough.

I found the different categories below :

- Red Cluster : Sport/Leisure oriented borough

	postcode	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
11	75012	0	Theater	Stadium	Performing Arts Venue	Playground	Fast Food Restaurant	Bike Trail	Recreation Center	Botanical Garden	Restaurant	Comedy Club

```
price(12)
```

```
The average price per sqm in PARIS 12 is : 11780.43 €
```

- Purple Cluster : Tourism,night life and entertainment

	postcode	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
0	75001	1	French Restaurant	Japanese Restaurant	tel	Plaza	Coffee Shop	Italian Restaurant	Café	Historic Site	Bar	Art Museum
1	75002	1	French Restaurant	Cocktail Bar	Wine Bar	Italian Restaurant	Bistro	Salad Place	Hotel	Thai Restaurant	Bakery	Furniture / Home Store
2	75003	1	Coffee Shop	Bakery	Bistro	Burger Joint	French Restaurant	Art Gallery	Clothing Store	Café	Cocktail Bar	Sandwich Place
3	75004	1	French Restaurant	Hotel	Clothing Store	Italian Restaurant	Ice Cream Shop	Bakery	Plaza	Wine Bar	Tapas Restaurant	Coffee Shop
4	75005	1	French Restaurant	Bar	Hotel	Italian Restaurant	Coffee Shop	Bakery	Creperie	Café	Wine Bar	Pub
5	75006	1	French Restaurant	Chocolate Shop	Hotel	Italian Restaurant	Restaurant	Plaza	Fountain	Pub	Tea Room	Athletics & Sports
8	75009	1	French	Hotel	Bakery	Japanese	Bistro	Wine Bar	Lounge	Cocktail Bar	Restaurant	Gym / Fitness

```
The average price per sqm in PARIS 1 is : 51743.16 €
The average price per sqm in PARIS 2 is : 17455.89 €
The average price per sqm in PARIS 3 is : 59.52 €
The average price per sqm in PARIS 4 is : 21889.82 €
The average price per sqm in PARIS 5 is : 16580.28 €
The average price per sqm in PARIS 6 is : 19264.15 €
The average price per sqm in PARIS 9 is : 20859.98 €
The average price per sqm in PARIS 10 is : 22931.74 €
The average price per sqm in PARIS 11 is : 15371.6 €
The average price per sqm in PARIS 13 is : 14728.84 €
The average price per sqm in PARIS 15 is : 17322.84 €
The average price per sqm in PARIS 18 is : 25410.04 €
The average price per sqm in PARIS 20 is : 27052.85 €
```

- Blue Cluster Family and leisure borough

	postcode	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
15	75016	2	Pool	Lake	Bike Rental / Bike Share	Circus	Yoga Studio	Falafel Restaurant	Frozen Yogurt Shop	French Restaurant	Fountain	Food & Drink Shop

```
1 price(16)
```

```
The average price per sqm in PARIS 16 is : 18104.15 €
```

- Green Cluster International, tourism and restaurants

	postcode	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
13	75014	3	French Restaurant	Hotel	Italian Restaurant	Thai Restaurant	Bistro	Food & Drink Shop	Café	Fast Food Restaurant	Tea Room	Bakery

```
1 price(14)
```

```
The average price per sqm in PARIS 14 is : 16035.64 €
```

- Orange Cluster :

	postcode	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
6	75007	4	Hotel	French Restaurant	Café	Plaza	History Museum	Italian Restaurant	Garden	Coffee Shop	Historic Site	Shop & Service
7	75008	4	French Restaurant	Hotel	Art Gallery	Spa	Japanese Restaurant	Cocktail Bar	Sporting Goods Shop	Department Store	Cosmetics Shop	Corsican Restaurant
16	75017	4	French Restaurant	Hotel	Italian Restaurant	Japanese Restaurant	Bakery	Café	Plaza	Bistro	Pastry Shop	Restaurant
18	75019	4	French Restaurant	Supermarket	Bakery	Hotel	Beer Garden	Concert Hall	Bike Rental / Bike Share	Scenic Lookout	Historic Site	Bar

```
price(7)
price(8)
price(17)
price(19)
```

```
The average price per sqm in PARIS 7 is : 20559.21 €
The average price per sqm in PARIS 8 is : 24198.68 €
The average price per sqm in PARIS 17 is : 15631.0 €
The average price per sqm in PARIS 19 is : 36809.44 €
```

Conclusion

The Paris real estate prices are very high and did not stop to rise this last decade. In addition to expensive prices, Paris is made of different and historically diversified borough boroughs. When moving in Paris and looking for the best place to live for our budget, it is really difficult to decide knowing that there are multiple factors and information to consider.

We can see from this analysis that the clusters are quite different but in some, we find multiple borough and in others we find only one. In conclusion this analysis will help to take a decision, a user can choose the best cluster that would correspond to its need. The analysis allows to do “arbitrage” when deciding where to live, because we can see that in some borough we find exactly the same type of venues as well as the same activity so we can choose the cheapest borough in term of price/ sqm2 to live. The analysis also highlight the type of venue, activities that we can find in each part of Paris.