

# Capstone Project - The Battle of Neighborhoods

David Alejandro Hincapié Tapias

Staying in a place that is similar to your home when you are traveling is something that many would like.

- Alejandro is on a business trip to Paris, France. This trip will last one month in this city. During his stay in paris he will also carry out some tourist activities.
- His hobbies are the gym and reading books. His house in Colombia is located in a neighborhood where he has easy access to a gym and a bookstore.
- As he has to travel to Paris, he wants to find a place to stay that has the places mentioned above nearby and that offers the possibility of finding a Colombian food restaurant or any Latin American food also nearby.

# The ideal property must have the following characteristics.

- Alejandro will use the Airbnb application to find a place that he likes and that fits his budget of minimum 100 euros and maximum 200 per night
- It is important that the place of lodging is as close as possible to the central neighborhoods of paris because these are safer and are close to the main tourist destinations of the city.
- The place must be available at least 360 days of the year, it must have at least 20 customer ratings regardless of the average rating it has, a private room would be sufficient and that the last rating would have been made minimally in the year 2020.

# Data Description

- For this project, some Airbnb datasets will be used that are published in the link: [Data](#). The specific datasets that will be used are related to the city of Paris, which will contain the properties available for rent in the city with all their respective information and the geographical location of the Paris neighborhoods. The data used from the dataset with the help of Foursquare will allow the project to be carried out. The datasets used are available under the Creative Commons CCO 1.0 Universal license.

# Airbnb and Foursquare



# Analysis and cleaning of data set imported from Airbnb

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights
0	2577	Loft for 4 by Canal Saint Martin	2827	Karine	NaN	Entrepôt	48.86957	2.36127	Entire home/apt	125	3
1	3109	zen and calm	3631	Anne	NaN	Observatoire	48.83191	2.31870	Entire home/apt	60	2
2	5396	Explore the heart of old Paris	7903	Borzou	NaN	Hôtel-de-Ville	48.85247	2.35835	Entire home/apt	47	1
3	7397	MARAIS - 2ROOMS APT - 2/4 PEOPLE	2626	Franck	NaN	Hôtel-de-Ville	48.85909	2.35315	Entire home/apt	90	10
4	7964	Large & sunny flat with balcony !	22155	Anaïs	NaN	Opéra	48.87417	2.34245	Entire home/apt	130	6

Example of the logs

# Analysis and cleaning of data set imported from Airbnb

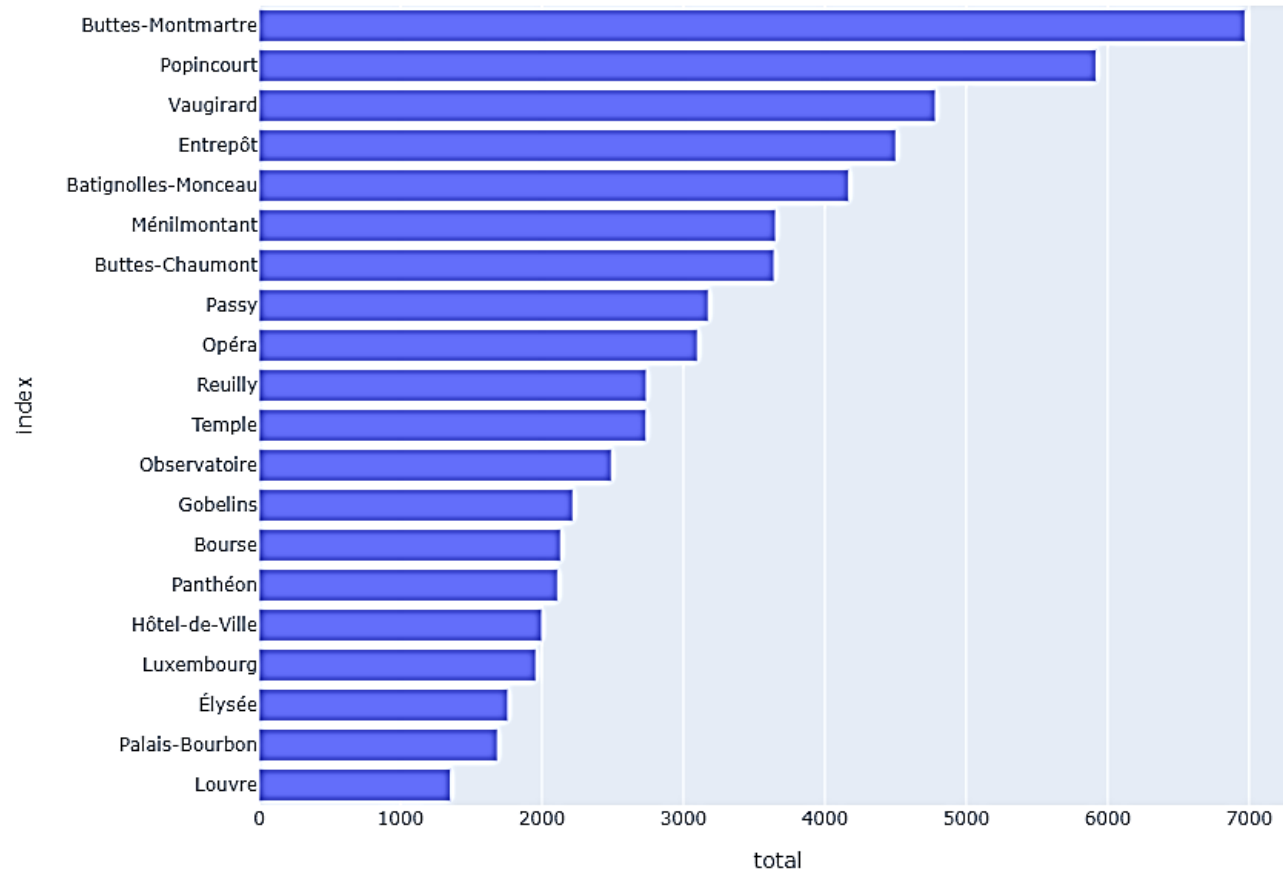
## Column data type

```
id                int64
name              object
host_id           int64
host_name         object
neighbourhood_group float64
neighbourhood     object
latitude          float64
longitude         float64
room_type         object
price            int64
minimum_nights    int64
number_of_reviews int64
last_review       object
reviews_per_month float64
calculated_host_listings_count int64
availability_365  int64
dtype: object
```

	neighbourhood	neighbourhood_group	geometry
0	Batignolles-Monceau	None	MULTIPOLYGON (((2.29517 48.87396, 2.29504 48.8...
17	Bourse	None	MULTIPOLYGON (((2.35152 48.86443, 2.35095 48.8...
2	Buttes-Chaumont	None	MULTIPOLYGON (((2.38943 48.90122, 2.39014 48.9...
18	Buttes-Montmartre	None	MULTIPOLYGON (((2.36580 48.88554, 2.36469 48.8...
4	Entrepôt	None	MULTIPOLYGON (((2.36469 48.88437, 2.36485 48.8...

# Analysis and cleaning of data set imported from Airbnb

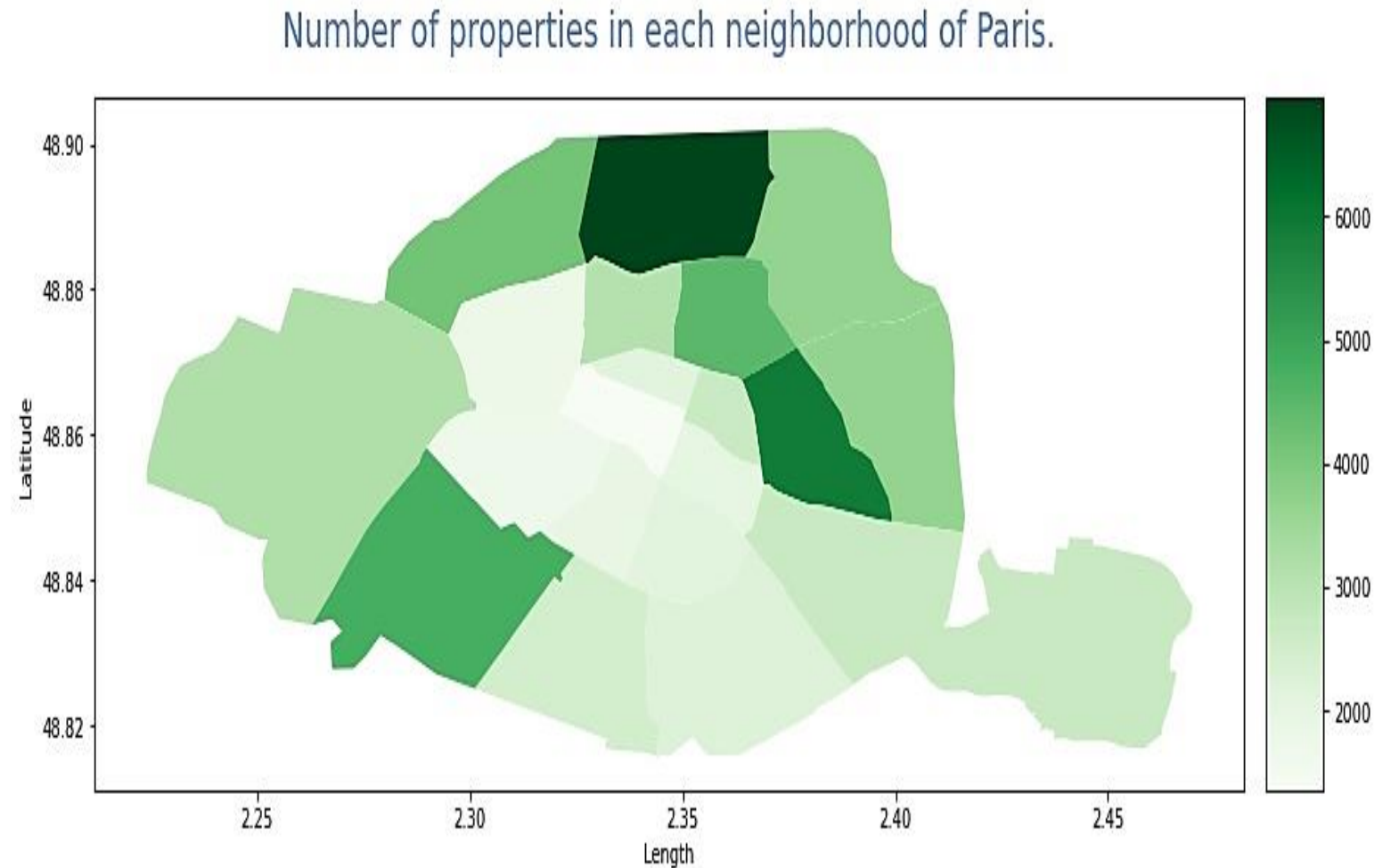
The number of properties by district of paris



Horizontal bar graph



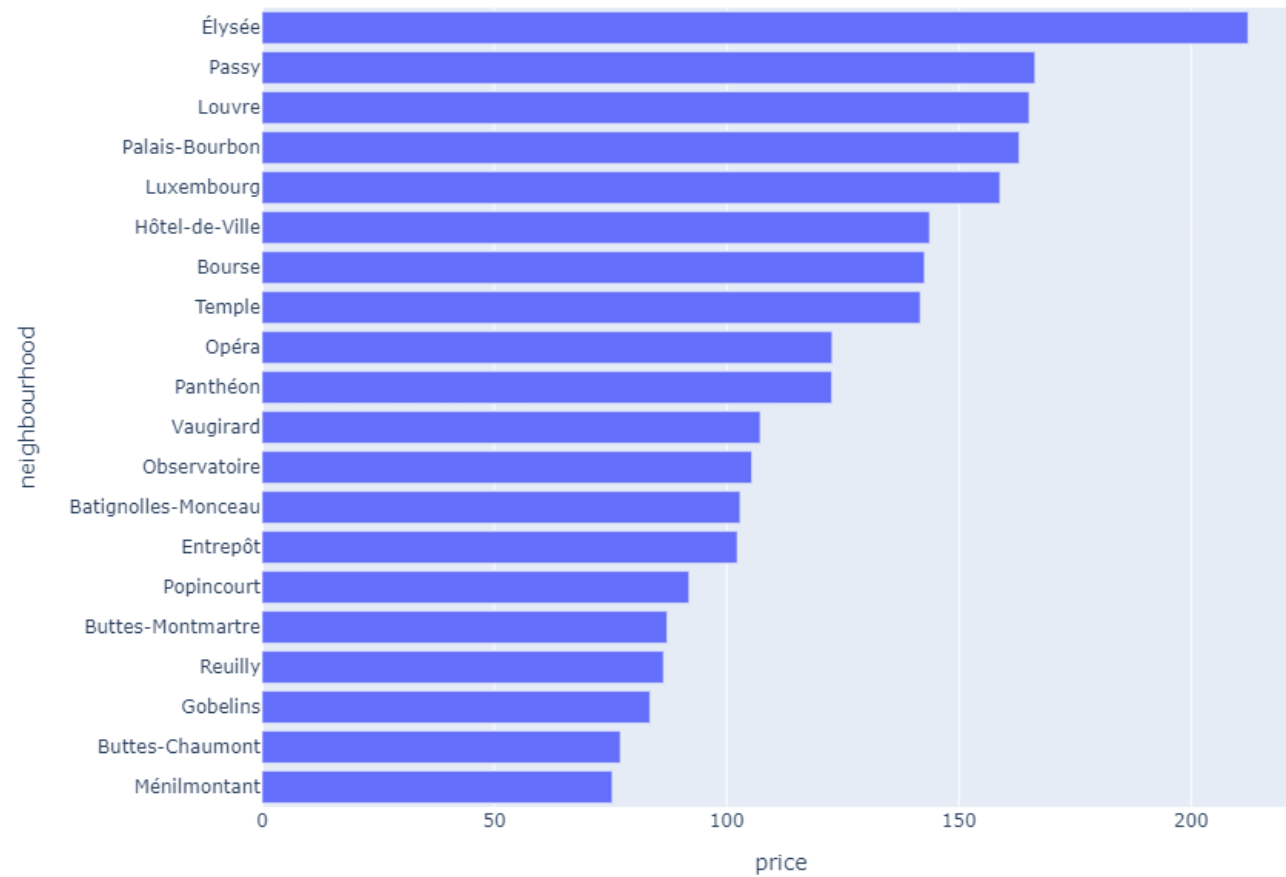
# Analysis and cleaning of data set imported from Airbnb



The darkest parts represent the neighborhoods where there are more properties

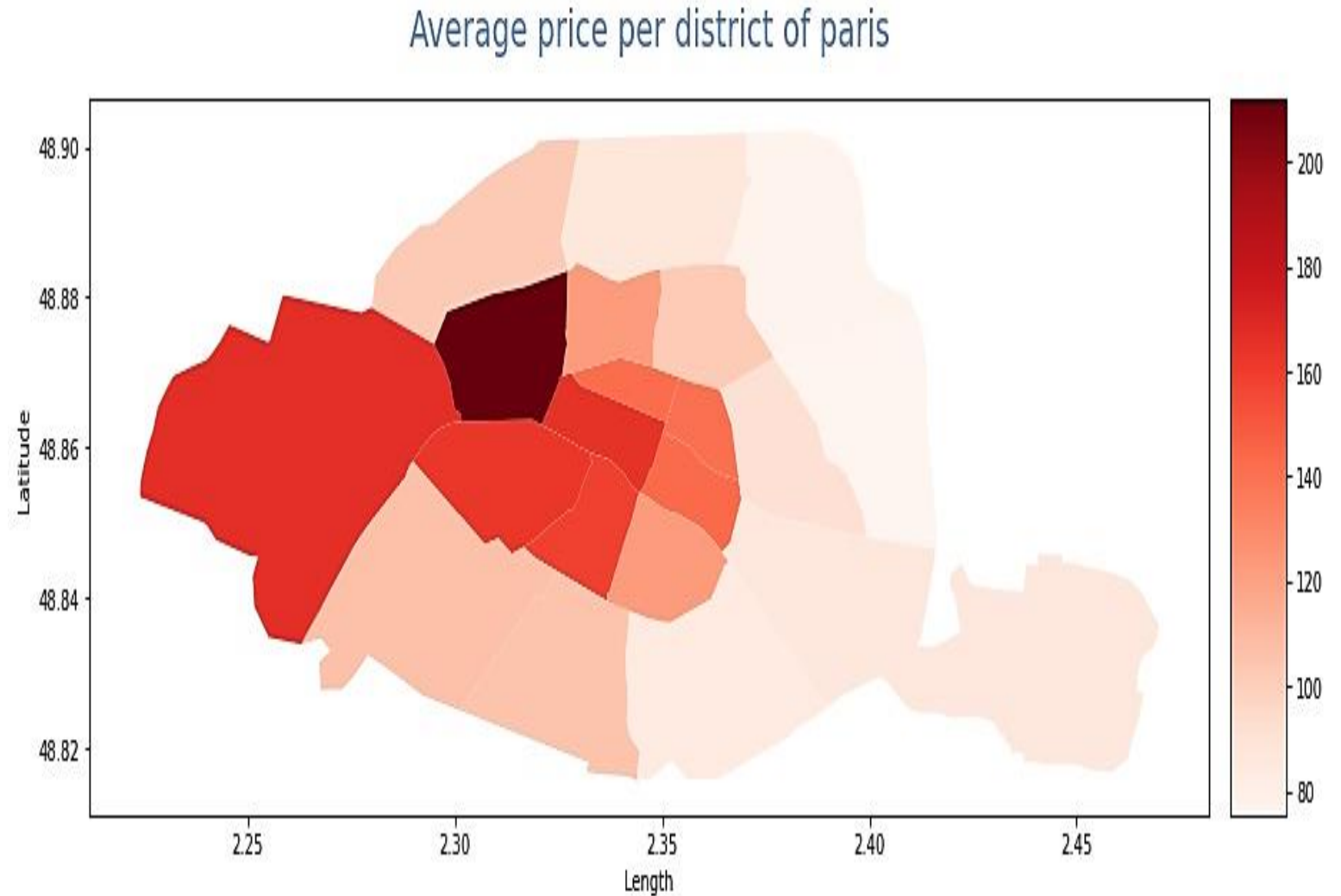
# Analysis and cleaning of data set imported from Airbnb

average property prices in each neighborhood



Horizontal bar graph

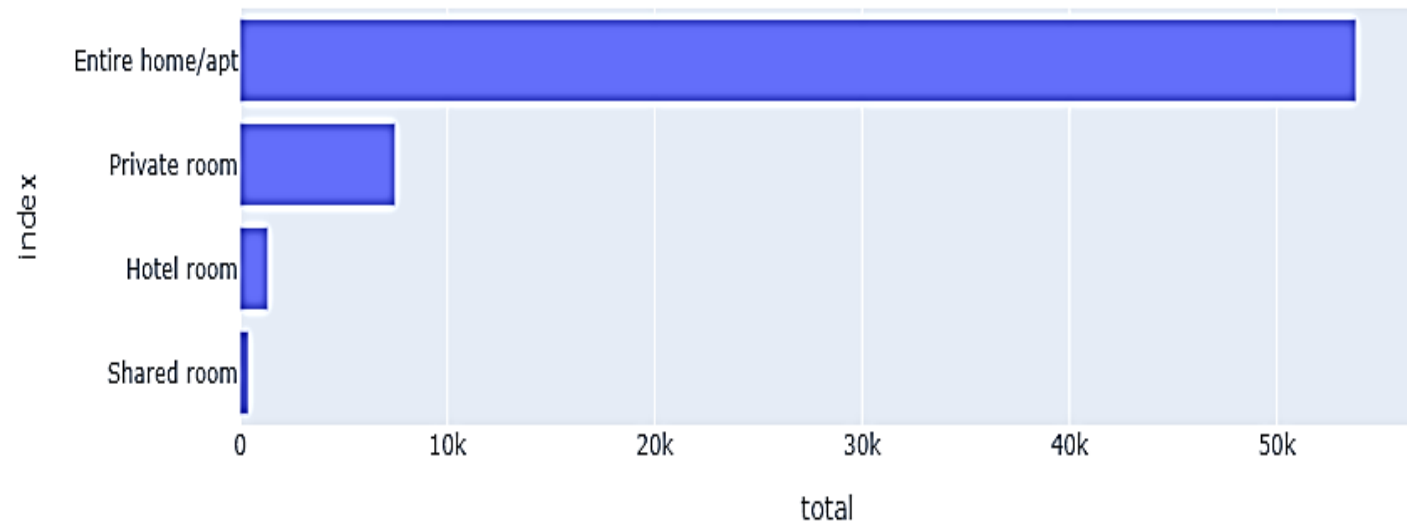
# Analysis and cleaning of data set imported from Airbnb



The darkest parts represent the neighborhoods with the highest average price.

# Analysis and cleaning of data set imported from Airbnb

The number of properties by type located in paris



Horizontal bar graph

# Analysis and cleaning of data set imported from Airbnb

Place airbn	
Double room 2, 3 or 4 beds, Saint-Germain des Prés	44
Hotel Suite with 2 rooms for 3 people in Odeon	44
Premium double room with bed Queensize in Odeon	44
Triple beds hotel room in Saint-Germain des Pres	44
Room in the Center - Louvre, & direct to Airports!	41
Double Room with 4 beds and 2 baths in Odeon	40
Cosy Flat In the heart of paris!	36
Beautiful renovation in The Marais	27
Room and terrace in Le Marais	26
Le Home St Germain	20
French Theory - The Classmate room	19
The Quintessential Parisian Apt	19

Number of places near  
each property.

# Analysis and cleaning of data set imported from Airbnb

----Beautiful renovation in The Marais----

	venue	freq
0	Bistro	0.11
1	Café	0.07
2	French Restaurant	0.07
3	Comedy Club	0.04
4	Diner	0.04
5	Hotel	0.04
6	Japanese Restaurant	0.04
7	Peruvian Restaurant	0.04
8	Plaza	0.04
9	Italian Restaurant	0.04

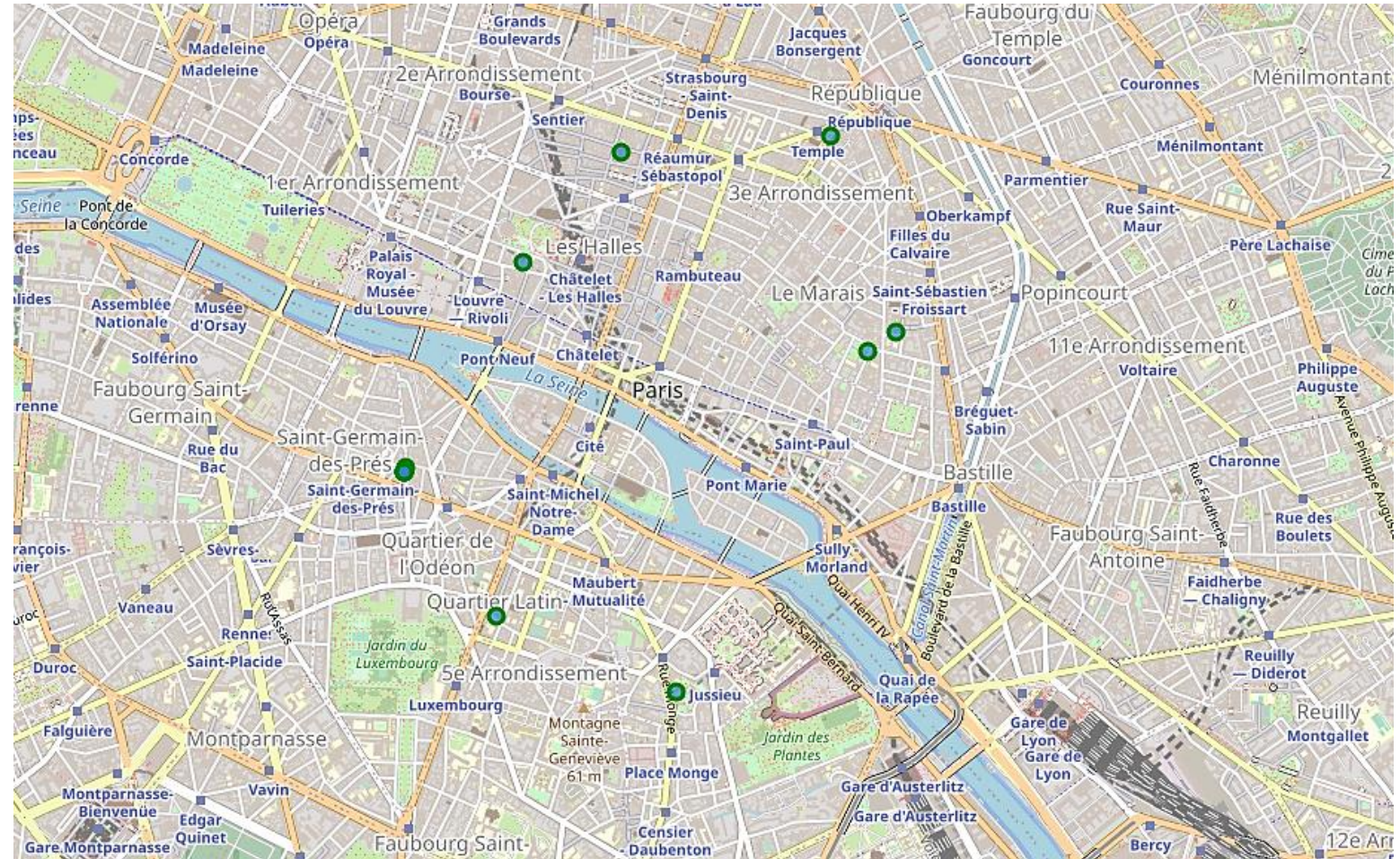
----Cosy Flat In the heart of paris!----

	venue	freq
0	Italian Restaurant	0.08
1	Wine Bar	0.08
2	Pedestrian Plaza	0.06
3	French Restaurant	0.06
4	Tea Room	0.06
5	Pizza Place	0.06
6	Cocktail Bar	0.06
7	Pastry Shop	0.03
8	Japanese Restaurant	0.03
9	Ice Cream Shop	0.03

Number of places near  
each property.



# Analysis and cleaning of data set imported from Airbnb



Represented by green circles are  
the properties.

Feed k means model to classify properties.

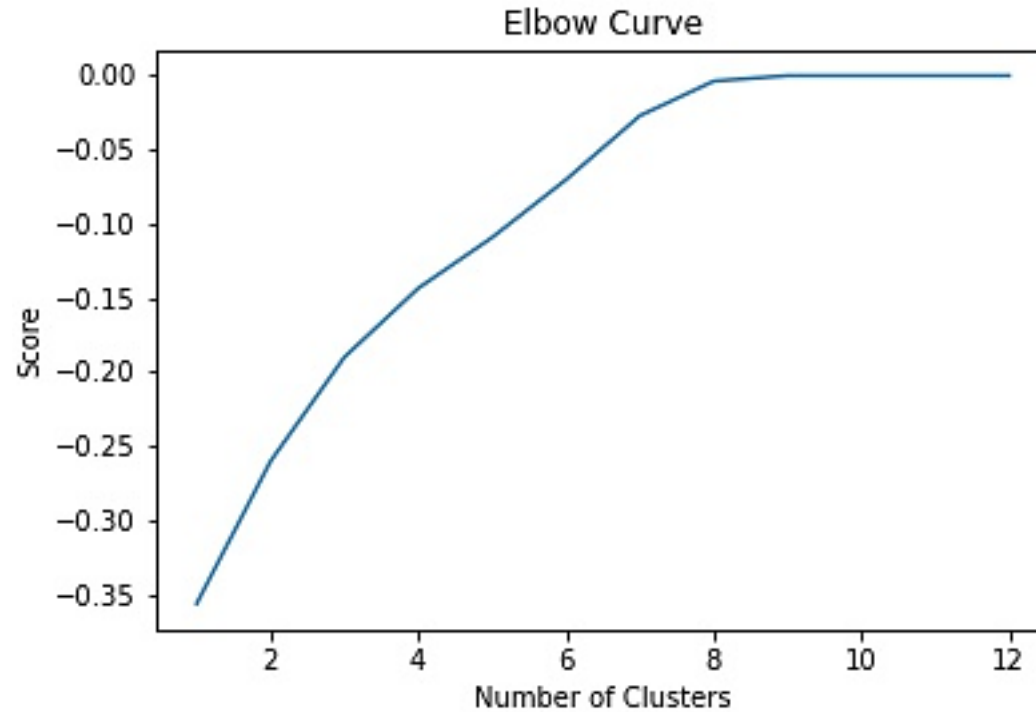
- Based on the data obtained from Airbnb and Foursquare, the k-means model is fed

	Place airbn	African Restaurant	Art Gallery	Art Museum	Asian Restaurant	Bagel Shop	Bakery	Bar	Beer Bar	Belgian Restaurant	...	Sushi Restaurant	Taco Place
0	Beautiful renovation in The Marais	0.000000	0.000000	0.000000	0.000000	0.000000	0.037037	0.000000	0.037037	0.000000	...	0.000000	0.000000
1	Cosy Flat In the heart of paris!	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.027778	0.000000
2	Double Room with 4 beds and 2 baths in Odeon	0.000000	0.000000	0.000000	0.025000	0.000000	0.000000	0.025000	0.000000	0.025000	...	0.000000	0.000000
3	Double room 2, 3 or 4 beds, Saint-Germain des ...	0.000000	0.000000	0.000000	0.022727	0.000000	0.000000	0.022727	0.000000	0.022727	...	0.000000	0.000000
4	French Theory - The Classmate room	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.052632	0.000000	0.000000	...	0.000000	0.000000

This data frame will be used to train the model.



Evaluate  
better  
parameters  
for the model.



Graph of the elbow  
method

# Evaluate better parameters for the model.

- This model will be trained with a data frame that will contain several characteristics that represent the types of places that exist near the properties. The k centroids of the model will contain a value of 8 that has been selected thanks to the elbow method, therefore it will allow the properties to be classified into 8 clusters

# Conclution

Nro cluster	Description	Qualification	%approval
0	This cluster contains properties that are too close to each other since it is evident that the nearby places they have are very similar. Of the 20 closest places to this property, only one adapts to the requirements. Although it has the advantage that having so many properties nearby in such a short distance it would make it much easier to change to another property.	1/3	33.3
1	The property that is in this cluster is located near several cultural places which is a great advantage. But it does not meet any of the requirements.	0/3	0.0
2	The property that is in this cluster is located near several food places. One of these adapts to the requirements.	1/3	33.3
3	The property that is in this cluster is located near several cultural places. Only one of these places fits the requirements.	1/3	33.3
4	The property that is in this cluster is located near several food establishments. Only one of these places fits the requirements.	1/3	33.3
5	The property that is in this cluster is located near several food establishments and a bookstore. Two of these places adapt to the requirements.	2/3	66.6
6	The property that is in this cluster is located near several food, clothing and cultural establishments. None of these places suit the requirements.	0/3	0.0
7	The property that is in this cluster is located near several food establishments and gyms. TWO of these locations accommodate requirements.	2/3	66.6

# Conclution

- It can be seen that two of the clusters have 2 of the 3 specific places mentioned in the requirements with an approval of 66.6%. This means that the properties located in these two clusters are the ones that have the most option to be Alejandro's Home in Paris.
- The two properties are listed below:

	id	name	neighbourhood	latitude	longitude	room_type	price	number_of_reviews	last_review	availability_365
8527	6360244	Cosy Flat In the heart of paris!	Bourse	48.86572	2.34933	Private room	127	222	2021-03-07	359
15235	11229350	Beautiful renovation in The Marais	Temple	48.86637	2.36145	Private room	129	147	2020-08-20	360