Capstone Project
Paris: Restaurants and Movies

Author: zhanikey

i. Introduction

WHY

As part of my Data Science Capstone Project in these sometimes dark and uncertain times I have decided to consider a case of exploring Paris venues and help them improve their customer experience by the means of the following:

When I used to live in Paris to do my studies I used to have that idea to visit all the cool places in the city that were filmed in my favourite movies to make photos after watching "Midnight in Paris" by Woody Allen. That is how I have come up with the initial idea for my project.

Who might be interested by this project?

Venues (mainly restaurants in our case) owners

Marketing agencies

City tourism department

Customers

Let's also not forget that this project is a pilot and in case of our idea viability it might be scaled up to any city or venue type. You can find a link to the Project Description notebook published on my GitHub repository.

ii. Data Usage

The data we will be using in the Project are:

- French government open dataset to get neighbourhoods and their locations
- Foursquare open API for fetching the exact location and addresses of the venues
- Kaggle open Dataset providing us with the list of the movies that were filmed in Paris with their exact location
- Additional data from open sources for movies list extending

iii. Methodology

In this project, we will use the Foursquare API to explore neighborhoods in Paris. We will use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. We will use the k-means clustering algorithm to complete this task. Also, we will use the Folium library to visualize the neighborhoods in Paris and their emerging clusters.

PHASE 1

Locating main neighbourhood clusters in order to find out what are the most visited restaurants in the area to help us find out what people from here do like

PHASE 2

Exploring the neighbourhoods in Paris

PHASE 3

Analyzing each neihbourhood that we have found

PHASE 4

Cluster the neighbourhoods with attempt to identify the patterns

PHASE 3

Creating a map of the above-mentioned clusters

iv. Results

GENERAL OVERVIEW

- We have fetched the Open Data dataset for Paris neihborhoods
- We have fetched the movies filmed in Paris dataset
- We have transformed all our data to dataframes
- We have created superimposed map of Paris with neighborhoods marked on it
- We have used Foursquare API to categorize the venue for each neighborhood on the basis of 100 venues within the radius of 500 meters
- We used one-hot encoding to explore the categories of the venue by calculating the mean of the frequency of occurrence of each category
- We have also calcuated the frequency for each neighborshood's venue category
- We have obtained 5 clusters for our neighborhoods and top 10 venues using k-means
- We have examined each cluster
- We have used movies dataframe to create a map and joined the map layer to our existing Paris clusters map

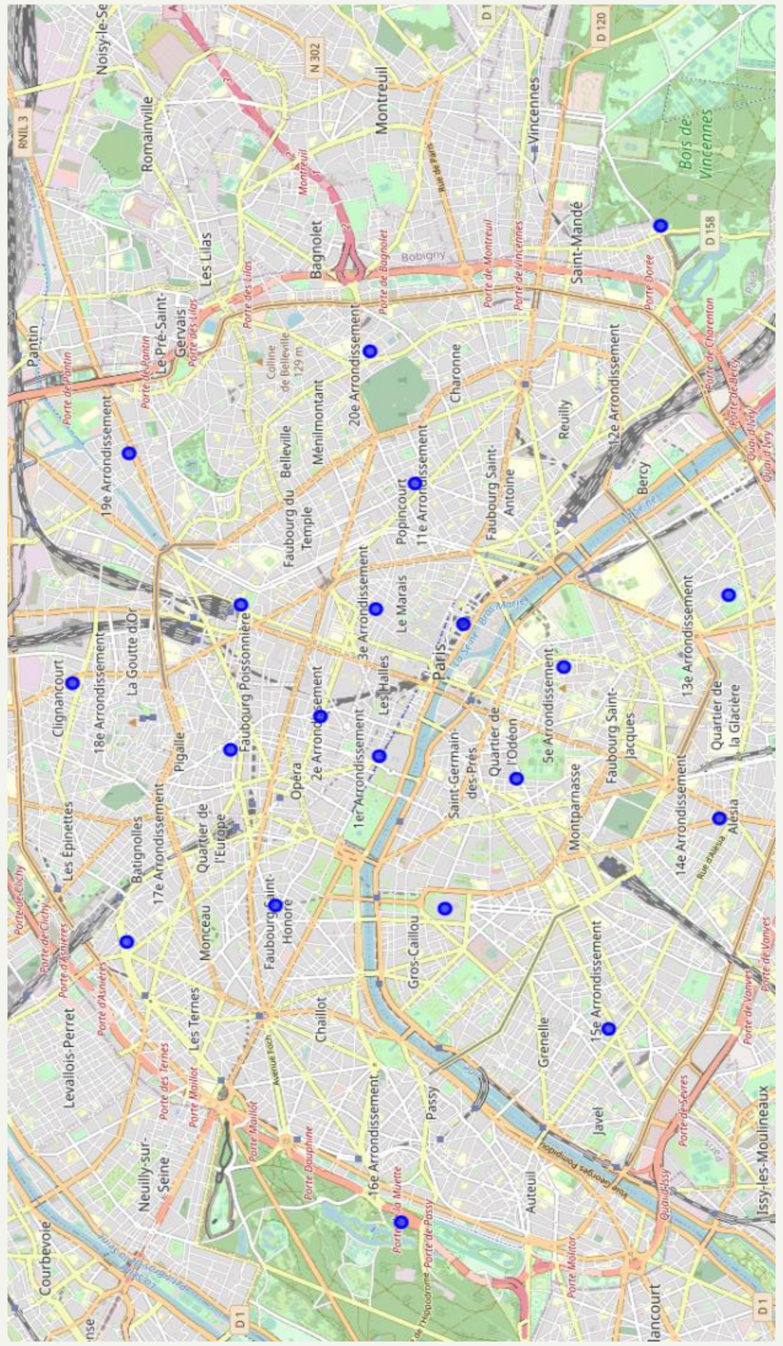
iv. Results

ARRONDISSEMENTS

We have found all the neighborhoods coordinates and names

	C_AR	C_ARINSEE	L_AR	L_AROFF	LATITUDE	LONGITUDE
0	11	75111	11ème Ardt	Popincourt	48.859059	2.380058
1	13	75113	13ème Ardt	Gobelins	48.828388	2.362272
2	4	75104	4ème Ardt	Hôtel-de-Ville	48.854341	2.357630
3	8	75108	8ème Ardt	Élysée	48.872721	2.312554
4	18	75118	18ème Ardt	Buttes-Montmartre	48.892569	2.348161
5	15	75115	15ème Ardt	Vaugirard	48.840085	2.292826
6	3	75103	3ème Ardt	Temple	48.862872	2.360001
7	2	75102	2ème Ardt	Bourse	48.868279	2.342803
8	17	75117	17ème Ardt	Batignolles-Monceau	48.887327	2.306777
9	5	75105	5ème Ardt	Panthéon	48.844443	2.350715
10	6	75106	6ème Ardt	Luxembourg	48.849130	2.332898
11	12	75112	12ème Ardt	Reuilly	48.834974	2.421325
12	9	75109	9ème Ardt	Opéra	48.877164	2.337458
13	19	75119	19ème Ardt	Buttes-Chaumont	48.887076	2.384821
14	7	75107	7ème Ardt	Palais-Bourbon	48.856174	2.312188
15	14	75114	14ème Ardt	Observatoire	48.829245	2.326542
16	20	75120	20ème Ardt	Ménilmontant	48.863461	2.401188
17	1	75101	1er Ardt	Louvre	48.862563	2.336443
18	10	75110	10ème Ardt	Entrepôt	48.876130	2.360728
19	16	75116	16ème Ardt	Passy	48.860392	2.261971

Also we created a map of the neighborhoods



iv. Results

USING FOURSQUARE

Foursquare API helped us to fetch 1246 different venues within the limit of 100 entries with a radius of 500 meters. Grouping venues by neighborhood helps us to better understand how many venues do we have in every arrondissement.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Batignolles-Monceau	59	59	59	59	59	59
Bourse	100	100	100	100	100	100
Buttes-Chaumont	43	43	43	43	43	43
Buttes-Montmartre	43	43	43	43	43	43
Entrepôt	100	100	100	100	100	100
Gobelins	61	61	61	61	61	61
Hôtel-de-Ville	100	100	100	100	100	100
Louvre	73	73	73	73	73	73
Luxembourg	38	38	38	38	38	38
Ménilmontant	47	47	47	47	47	47
Observatoire	29	29	29	29	29	29
Opéra	100	100	100	100	100	100
Palais-Bourbon	100	100	100	100	100	100
Panthéon	86	86	86	86	86	86
Passy	11	11	11	11	11	11
Popincourt	69	69	69	69	69	69
Reuilly	4	4	4	4	4	4
Temple	82	82	82	82	82	82
Vaugirard	65	65	65	65	65	65
Élysée	36	36	36	36	36	36

iv. Results

ONE-HOT ENCODING

Using *get_dummies* helps us to group our neighborhoods and explore each one based on the categories of the venues.

MEANS

We also transformed the dataframe to calculate the frequency of occurrence of each category.

CATEGORIES

We obtain 198 columns containing each venue category

iv. Results

```
----Bourse----
                venue  freq
0  French Restaurant  0.13
1        Cocktail Bar  0.06
2            Wine Bar  0.06
3               Hotel  0.05
4              Bakery  0.04
```

Top 5

Using grouping we obtain each neighborhood table with its top 5 venues (example left).

TOP 10

Next, we are creating a dataframe with 10 most common venues for the neighborhoods.

	Neighborhood	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	Batignolles-Monceau	French Restaurant	Hotel	Italian Restaurant	Café	Plaza	Japanese Restaurant	Bakery	Wine Shop	Bistro	Pizza Place
1	Bourse	French Restaurant	Wine Bar	Cocktail Bar	Hotel	Bakery	Bistro	Creperie	Ice Cream Shop	Thai Restaurant	Concert Hall
2	Buttes-Chaumont	French Restaurant	Bar	Supermarket	Hotel	Seafood Restaurant	Beer Bar	Bistro	Music Store	Coffee Shop	Steakhouse
3	Buttes-Montmartre	Bar	French Restaurant	Hotel	Restaurant	Coffee Shop	Pizza Place	Convenience Store	Deli / Bodega	Seafood Restaurant	Café
4	Entrepôt	French Restaurant	Hotel	Coffee Shop	Café	Indian Restaurant	Bistro	Pizza Place	Japanese Restaurant	Bakery	Seafood Restaurant

iv. Results

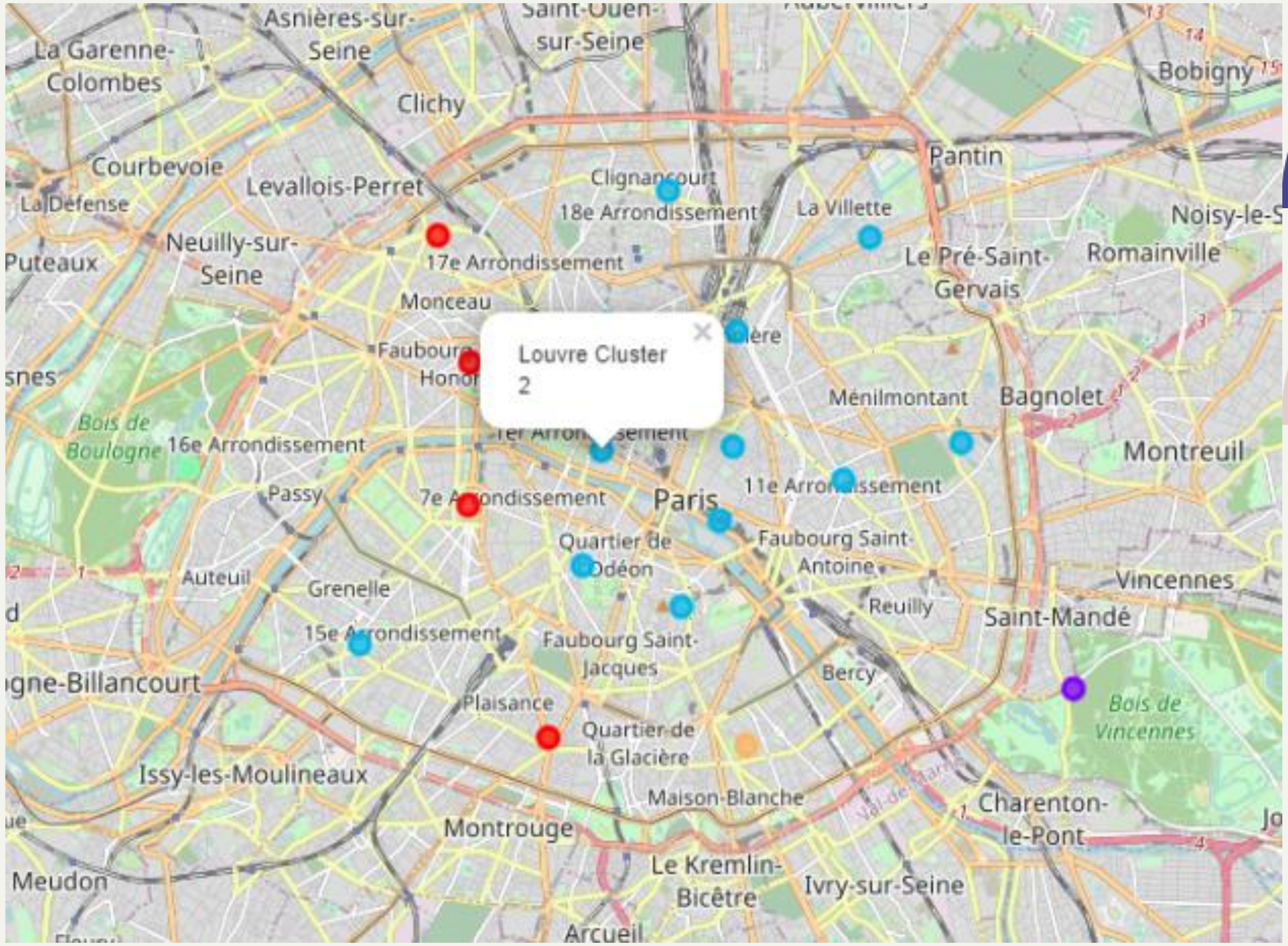
iv. Results

K-Means

Finally, we are running k-means clustering to divide our data to 5 clusters.

	C_AR	C_ARINSEE	L_AR	L_AROFF	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	9th Most Common Venue	10th Most Common Venue
0	11	75111	11ème Ardt	Popincourt	48.859059	2.380058	2	French Restaurant	Café	Supermarket	Restaurant	Wine Bar	Pastry Shop	Italian Restaurant	Cocktail Bar	Bar	Bakery
1	13	75113	13ème Ardt	Gobelins	48.828388	2.362272	4	Vietnamese Restaurant	Asian Restaurant	French Restaurant	Chinese Restaurant	Thai Restaurant	Juice Bar	Coffee Shop	Park	Cambodian Restaurant	Cosmetics Shop
2	4	75104	4ème Ardt	Hôtel-de-Ville	48.854341	2.357630	2	French Restaurant	Ice Cream Shop	Hotel	Plaza	Italian Restaurant	Clothing Store	Pedestrian Plaza	Garden	Wine Bar	Art Gallery
3	8	75108	8ème Ardt	Élysée	48.872721	2.312554	0	French Restaurant	Hotel	Art Gallery	Spa	Theater	Plaza	Cocktail Bar	Park	Resort	Modern European Restaurant
4	18	75118	18ème Ardt	Buttes-Montmartre	48.892569	2.348161	2	Bar	French Restaurant	Hotel	Restaurant	Coffee Shop	Pizza Place	Convenience Store	Deli / Bodega	Seafood Restaurant	Café

iv. Results



We have created cluster map of Paris

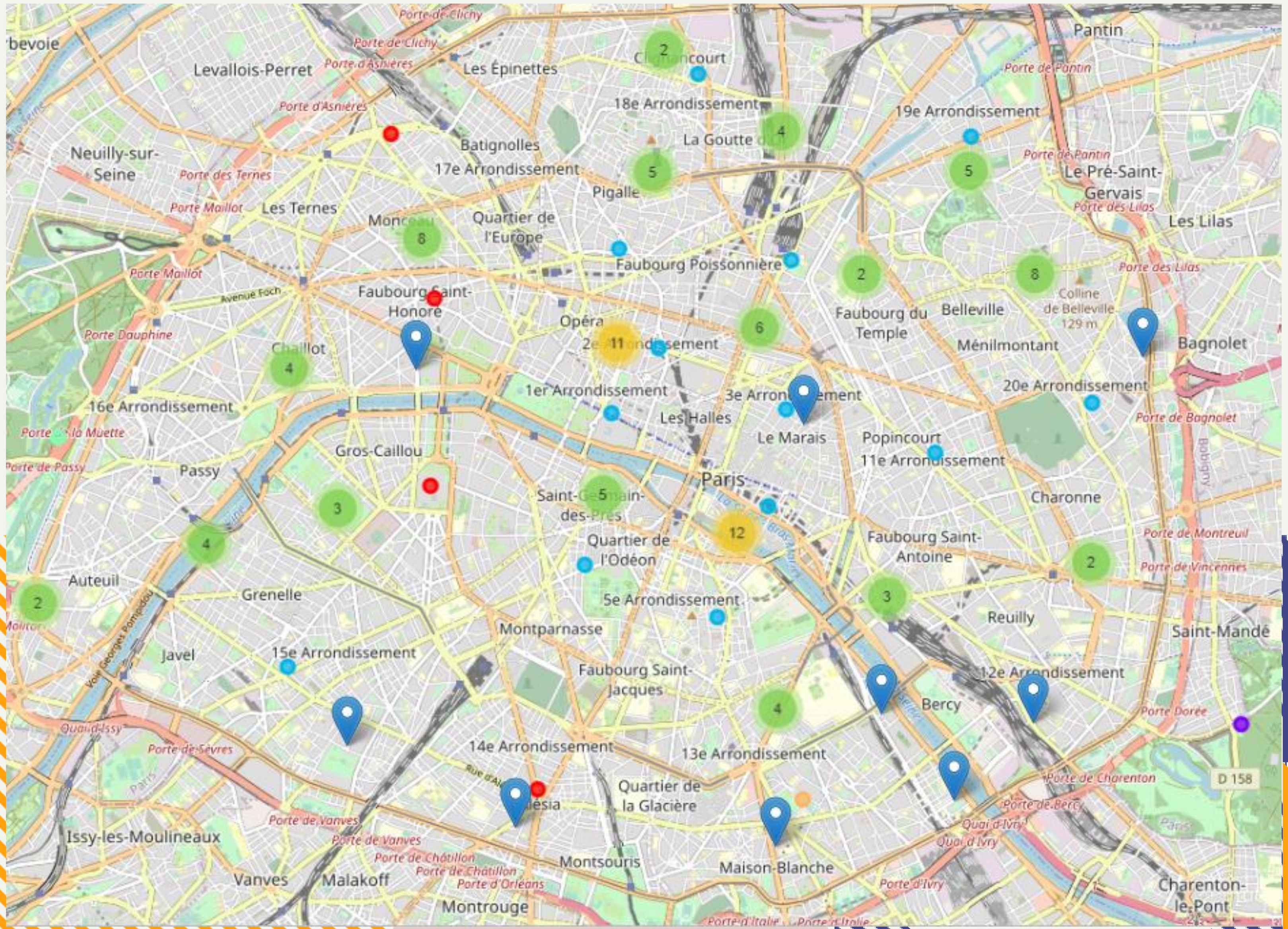
Clusters

We have examined each of the clusters

CLUSTER 0	4 Neighborhoods
CLUSTER 1	1 Neighborhood
CLUSTER 2	13 Neighborhoods
CLUSTER 3	1 Neighborhood
CLUSTER 4	1 Neighborhood

Movies part

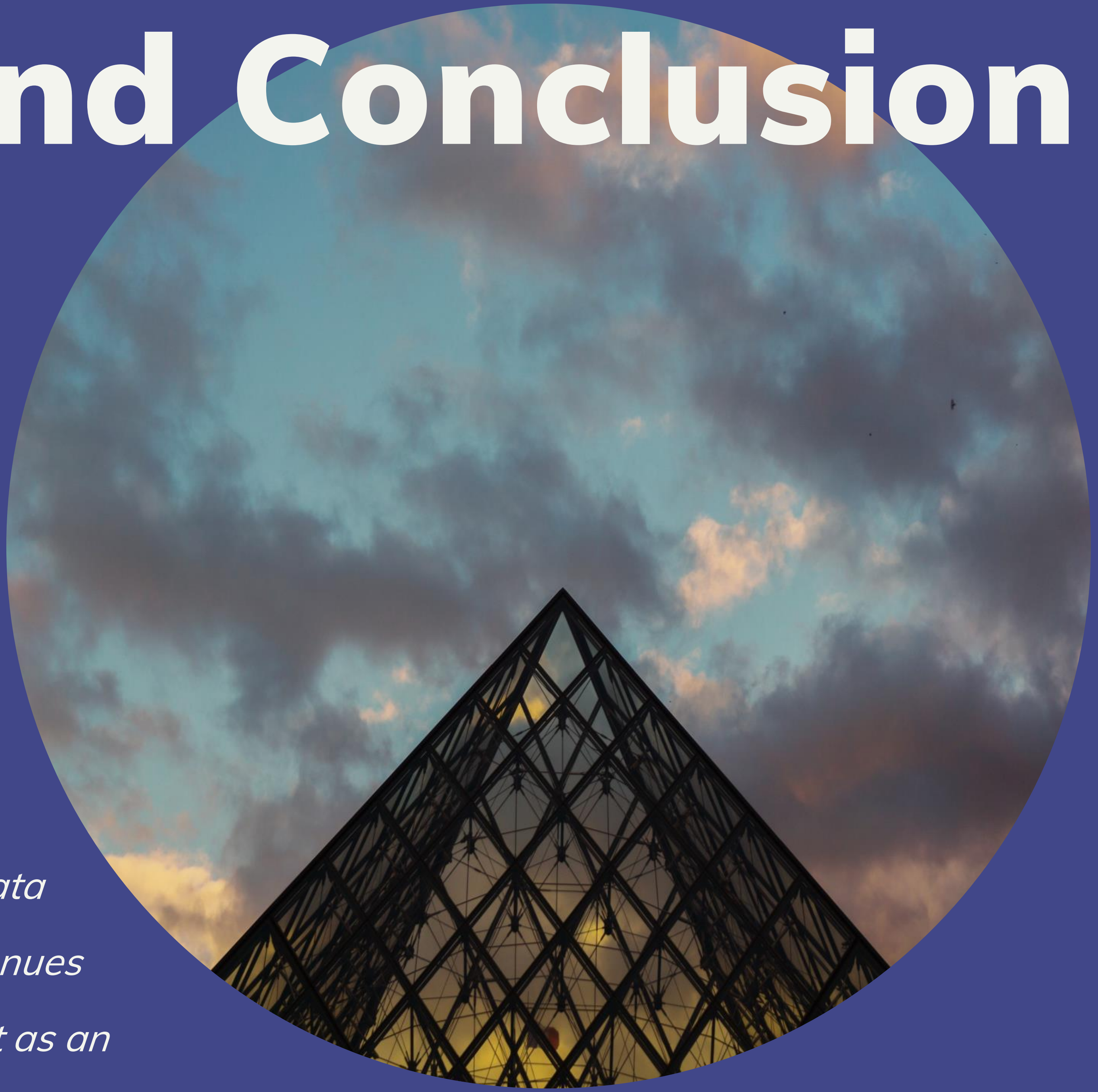
GENERAL OVERVIEW



KAGGLE

We loaded the Kaggle dataset to fetch the movies filmed in Paris data, created the dataframe and cleaned it. We have limited the number of the locations that we are going to show to 100 for demonstration. Finally, we have superimposed our newly created map of the movie sets to our existing cluster map.

v+vi. Observations and Conclusion



WHAT

Whether you are deciding to open a restaraunt our analysis helps us to know what cuisine will be more popular where. We can see that many of the clusters have French Restaurant as their 1st most popular venue category, except for the 4th cluster. We have limited our movies dataframe to 100 but we can clearly see that a lot of movies were filmed very closely to our cluster points.

This project demonstrates the capabilities of combining any dataframe with geographical data using Python. We have used folium to build our maps, and Foursquare API enabled using venues data for our analysis. As the data might not be always precise, I was considering this project as an opportunity to enhance my skills and apply them directly via this practical task. When extending new skills further (which I hope I will be able to do) I will continue to create notebooks using more advanced techniques and statistical methods.