CLUSTER DISTRICTS IN PARIS BASED ON THE SIMILARITY OF ART VENUES USING K-MEANS

Duy Anh MAI

CONTENT

I. INTRODUCTION

- Paris is the capital of France, including 20 urban districts. There are many works of art, famous museums, and leisure activities for tourists

- Problem : travelers can't get enough information about the location offer which types of art options
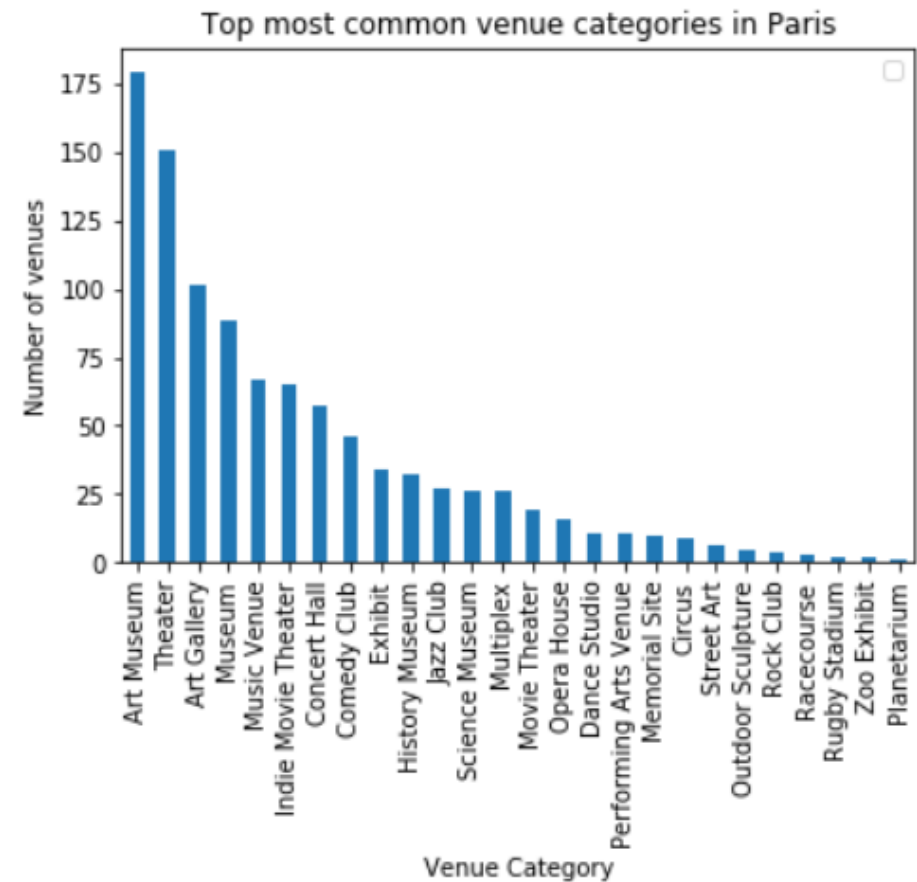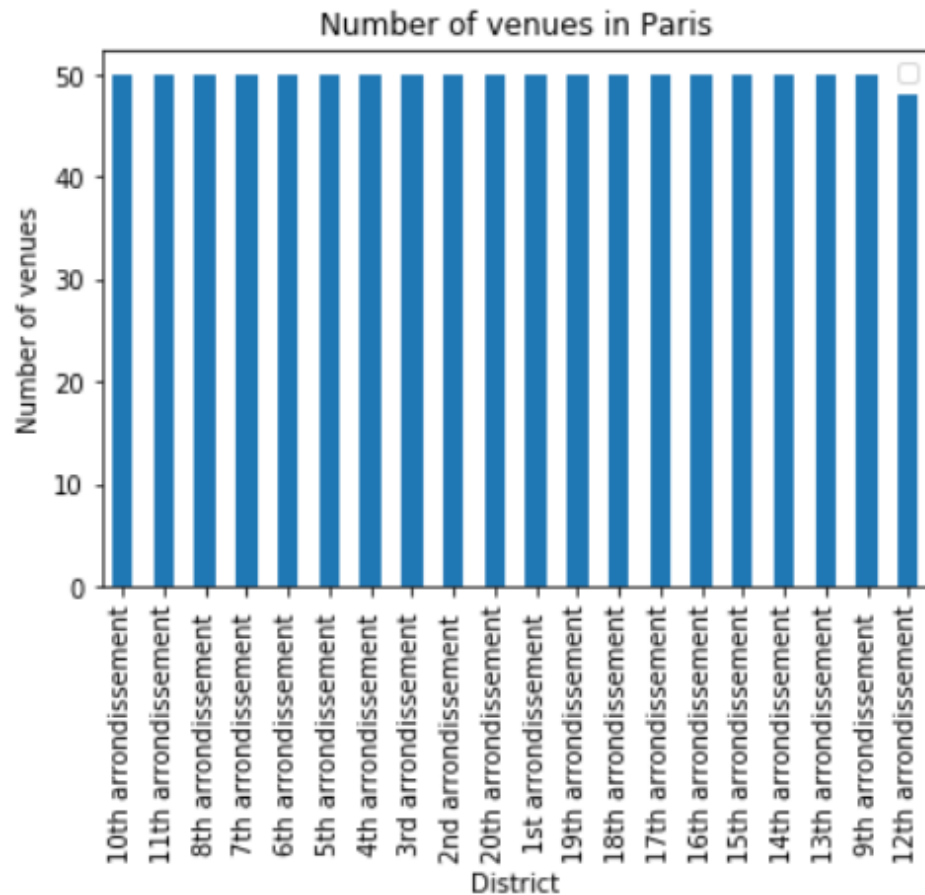
II. BUSINESS PROBLEM

- Business owners, artists or art fans and especially the tourists can find the result of this project useful in their decision-making process to visit Paris.

III. DATA

➤ The list of district names in Paris (scraped from Wikipedia using Beautiful Soup)

➤ The latitude and longitude of these neighborhoods (Geocoder package)

➤ Venues data related to the art section in each district to help find the location with more interest in art galleries (from Foursquare API)

Number of venues in Paris

Top most common venue categories in Paris

- All categories were one-hot encoded to another data frame

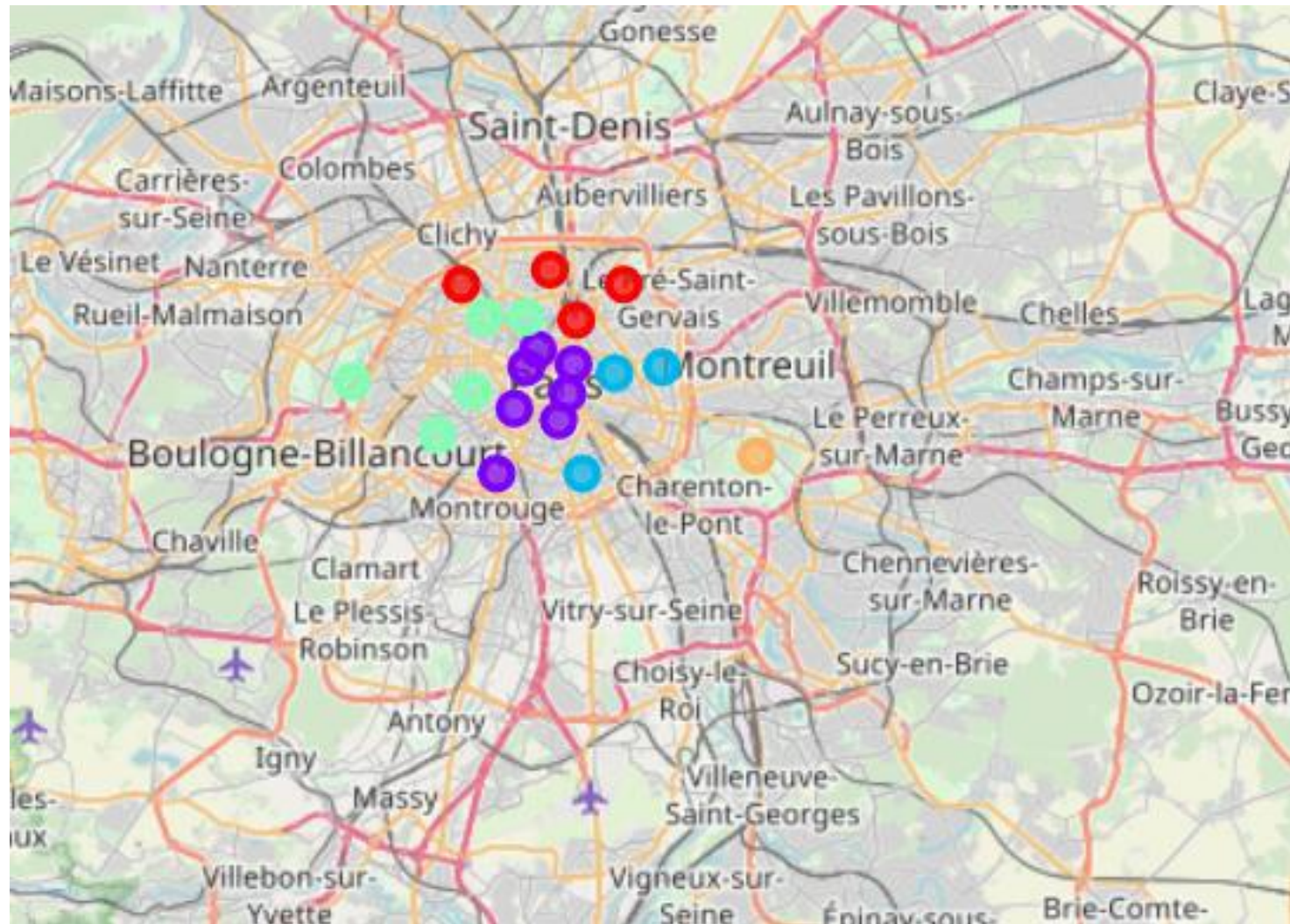|	District	Art Gallery	Art Museum	Circus	Comedy Club	Concert Hall	Dance Studio	Exhibit	History Museum	Indie Movie Theater	Jazz Club	Memorial Site	Movie Theater	Multiplex	Museum
0	1st arrondissement	0	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1st arrondissement	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1st arrondissement	0	0	0	0	0	0	1	0	0	0	0	0	0	0
3	1st arrondissement	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	1st arrondissement	0	1	0	0	0	0	0	0	0	0	0	0	0	0

- The mean of all counts for each category in a district

	District	Art Gallery	Art Museum	Circus	Comedy Club	Concert Hall	Dance Studio	Exhibit	History Museum	Indie Movie Theater	Jazz Club	Memorial Site	Movie Theater	Multiplex
0	10th arrondissement	0.1200	0.10	0.02	0.040000	0.080000	0.000000	0.02	0.020000	0.1000	0.04	0.000000	0.000000	0.040000
1	11th arrondissement	0.1200	0.10	0.02	0.060000	0.060000	0.020000	0.02	0.020000	0.0800	0.04	0.020000	0.000000	0.020000
2	12th arrondissement	0.0625	0.00	0.00	0.020833	0.083333	0.083333	0.00	0.020833	0.0625	0.00	0.041667	0.104167	0.041667
3	13th arrondissement	0.2000	0.14	0.00	0.020000	0.000000	0.020000	0.02	0.020000	0.0600	0.04	0.020000	0.000000	0.040000
4	14th arrondissement	0.0800	0.28	0.00	0.040000	0.000000	0.000000	0.06	0.040000	0.1000	0.06	0.000000	0.000000	0.040000
5	15th arrondissement	0.0800	0.32	0.00	0.060000	0.040000	0.000000	0.06	0.020000	0.0200	0.02	0.000000	0.020000	0.020000

- Top 6 most common venues of each district

	District	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	10th arrondissement	Theater	Art Gallery	Art Museum	Indie Movie Theater	Concert Hall	Music Venue	Museum	Comedy Club	Jazz Club	Multiplex
1	11th arrondissement	Art Gallery	Museum	Theater	Art Museum	Music Venue	Indie Movie Theater	Comedy Club	Concert Hall	Science Museum	Jazz Club
2	12th arrondissement	Theater	Music Venue	Movie Theater	Concert Hall	Dance Studio	Art Gallery	Indie Movie Theater	Museum	Memorial Site	Multiplex
3	13th arrondissement	Art Gallery	Museum	Art Museum	Theater	Music Venue	Science Museum	Indie Movie Theater	Jazz Club	Multiplex	Zoo Exhibit
4	14th arrondissement	Art Museum	Theater	Museum	Indie Movie Theater	Art Gallery	Science Museum	Exhibit	Jazz Club	Comedy Club	History Museum
5	15th arrondissement	Art Museum	Theater	Art Gallery	Museum	Comedy Club	Exhibit	Concert Hall	Music Venue	History Museum	Indie Movie Theater
6	16th arrondissement	Art Museum	Theater	Museum	Art Gallery	Music Venue	Movie Theater	Multiplex	Outdoor Sculpture	Racecourse	History Museum
7	17th arrondissement	Theater	Art Museum	Music Venue	Concert Hall	Art Gallery	Comedy Club	Museum	History Museum	Indie Movie Theater	Movie Theater
8	18th arrondissement	Theater	Music Venue	Concert Hall	Indie Movie Theater	Art Museum	Comedy Club	Multiplex	Jazz Club	History Museum	Opera House
9	19th arrondissement	Concert Hall	Theater	Music Venue	Multiplex	Art Museum	Art Gallery	Indie Movie Theater	Jazz Club	Circus	Museum

VI. CONCLUSION

- This project only focuses on clustering the districts based on their art venues. However, in reality, many other factors can contribute to deciding which district is good to open an art gallery as well. In the scope and time frame of this project, I accept the result and hope to have further research to improve the solution