ANALYZING THE BOROUGHS IN PARIS FOR BETTER TOUR ADVISING

Vibhuti Singh

September 2019

1. INTRODUCTION

1.1 Description and Discussion of background

Paris , capital city of France is a bustling city full of historic monuments and interesting venues . It is the most populous city of France with an area of 105 square kilometers (41 square miles) and an official estimated population of 2,140,526 residents i.e. with 252 residents per hectare, not counting parks. Paris is divided into 20 districts called "arrondissements"(French word) and they are numbered 1 to 20 arranged in the form of a clockwise spiral), starting from the middle of the city, with the first on the Right Bank (north bank) of the Seine.

1.1.1 Problem and Interest

Paris being rich in French culture and heritage is a tourist attraction to a large set of people. Paris received 23 million visitors in 2017, measured by hotel stays. Looking at the large number figure of the visitors it's obvious the tour advisors need to keep in mind that the tourists are looking to visit some specific areas in Paris and so in order to satisfy their customers it's needed to know where to invest their money and time during a tour and to choose a Borough wisely to take accommodation in.

In order to advise a specific Borough to choose for accommodation we can cluster the similar type of Boroughs according to the venue densities in that area with the help of which tour advisor can suggest a particular set of Boroughs according to the venue interests of the visitors.

1.2 Data description

To obtain the intended results we can extract data from following sources:-

- I extracted the data about Boroughs and Quarters in Paris from the demographic data of Paris on Wikipedia page and cleaned the data removing unnecessary parameters/columns.

- I found the latitude and longitude of the Boroughs in Paris with the help of Nominatim API which is provided by the geocoder classes present in the geopy module in Python.(The module might need to be installed before you use it).

- Further the latitudes and longitudes of common venues are found with the help of Four Square API.

- I found out that the postal codes of Boroughs in Paris are made by adding the Arrondissement number of Boroughs to 75000 , so i entered this data by myself in the dataframe using codes in python. For example- The borough number for ‘Louvre ‘ is 1 so it’s postal code will be 75001.

Final Clean data frame:-

	Borough	Name	Quarters	Postal Codes	Latitude	Longitude
0	1st (Ie) R	Louvre	Saint-Germain-l'Auxerrois,Les Halles,Palais-Ro...	75001	48.861147	2.338028
1	2nd (IIe) R	Bourse	Gaillon,Vivienne,Mail,Bonne-Nouvelle	75002	48.867684	2.343126
2	3rd (IIIe) R	Temple	Arts-et-Métiers,Enfants-Rouges,Archives,Sainte...	75003	48.862683	2.358685
3	4th (IVe) R	Hôtel-de-Ville	Saint-Merri,Saint-Gervais,Arsenal,Notre-Dame	75004	48.856426	2.352528
4	5th (Ve) L	Panthéon	Quartier Saint-Victor,Jardin-des-Plantes,Val-d...	75005	48.846191	2.346079
5	6th (VIe) L	Luxembourg	Monnaie,Odéon,Notre-Dame-des-Champs,Saint-Germ...	75006	48.849392	2.332260
6	7th (VIIe) L	Palais-Bourbon	Saint-Thomas-d'Aquin,Les Invalides,École-Milit...	75007	48.861692	2.319031
7	8th (VIIIe) R	Élysée	Champs-Élysées,Faubourg-du-Roule,La Madeleine,...	75008	48.846644	2.369830
8	9th (IXe) R	Opéra	Saint-Georges,Chaussée-d'Antin,Faubourg-Montma...	75009	48.870645	2.332330
9	10th (Xe) R	Entrepôt	Saint-Vincent-de-Paul,Porte-Saint-Denis,Porte-...	75010	48.876008	2.360445

1.3 Feature Selection and column dropping

The extracted datasets had unnecessary strings and information which were no use to the purpose of clustering Boroughs with respect to venues. The Demographic data and column for Mayors of each Quarters was dropped for the ease of using the information.

2. Methodology

As a database, I used GitHub repository in my study. I used Nominatim API to obtain the geographical co-ordinates of Boroughs in Paris. Python provides the tool geopy which has various geocoder classes that help us using the APIs such as OpenStreetMap Nominatim, Google Geocoding API (V3), and many other geocoding services.

I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as **100 venue** and the radius **500 meter** for each borough from their given latitude and longitude informations. In summary of this data **1272** venues were returned by Foursquare. Here is a merged table of boroughs and venues.

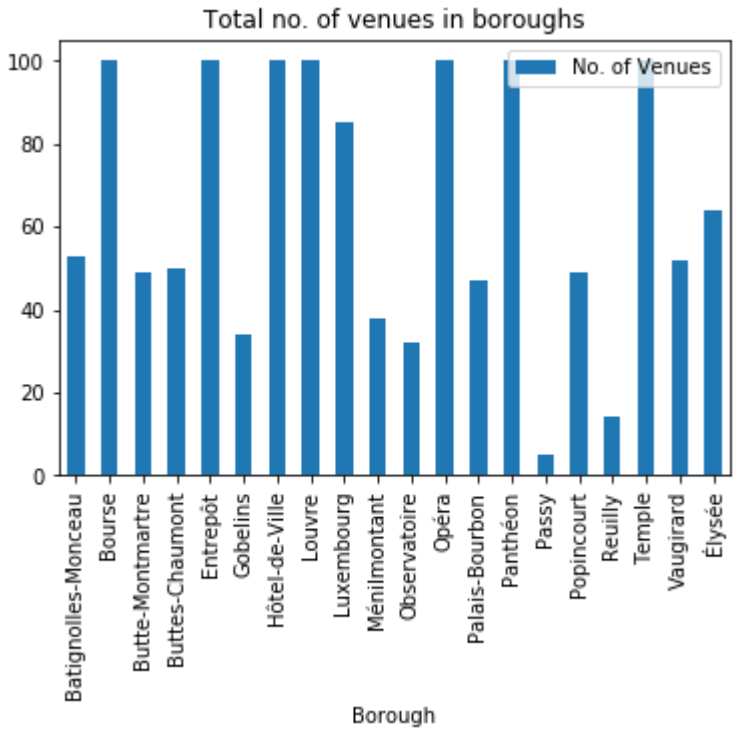
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Louvre	48.861147	2.338028	Cour Carrée du Louvre	48.860360	2.338543	Pedestrian Plaza
1	Louvre	48.861147	2.338028	Musée du Louvre	48.860847	2.336440	Art Museum
2	Louvre	48.861147	2.338028	La Vénus de Milo (Vénus de Milo)	48.859943	2.337234	Exhibit
3	Louvre	48.861147	2.338028	Place du Palais Royal	48.862523	2.336688	Plaza
4	Louvre	48.861147	2.338028	Cour Napoléon	48.861172	2.335088	Plaza
5	Louvre	48.861147	2.338028	Comédie-Française	48.863088	2.336612	Theater
6	Louvre	48.861147	2.338028	Le Fumoir	48.860341	2.340647	Cocktail Bar
7	Louvre	48.861147	2.338028	Place du Louvre	48.859841	2.340822	Plaza
8	Louvre	48.861147	2.338028	Le Café Blanc	48.862719	2.339578	Bar
9	Louvre	48.861147	2.338028	Église Saint-Germain l'Auxerrois	48.859887	2.340617	Church

.

We can see that Bourse, Entrepot, Hotel-de-Ville, Louvre and Opera how reached the **100** limit of venues. On the other hand, some venues are below 40.

Borough	No. of Venues
Batignolles-Monceau	53
Bourse	100
Butte-Montmartre	49
Buttes-Chaumont	50
Entrepôt	100
Gobelins	34
Hôtel-de-Ville	100
Louvre	100
Luxembourg	85
Ménilmontant	38
Observatoire	32
Opéra	100
Palais-Bourbon	47

The result doesn’t mean that inquiry run all the possible results in boroughs. Actually, it depends on given Latitude and Longitude informations and here is we just run single Latitude and Longitude pair for each borough. We can increase the possibilities with Neighborhood informations with more Latitude and Longitude informations.



In summary of this graph 222 unique categories were returned by Foursquare. The categorical data was hard to work with so I used one hot encoding to utilize the data for clustering ,then I created a table which shows list of top 10 venue category for each borough in below table.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Batignolles-Monceau	French Restaurant	Italian Restaurant	Hotel	Bakery
1	Bourse	French Restaurant	Wine Bar	Cocktail Bar	Hotel
2	Butte-Montmartre	Bar	French Restaurant	Pizza Place	Gastropub
3	Buttes-Chaumont	French Restaurant	Bar	Restaurant	Hotel
4	Entrepôt	French Restaurant	Hotel	Coffee Shop	Café

We have some common venue categories in boroughs. In this reason I used unsupervised learning **K-means algorithm** to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning.

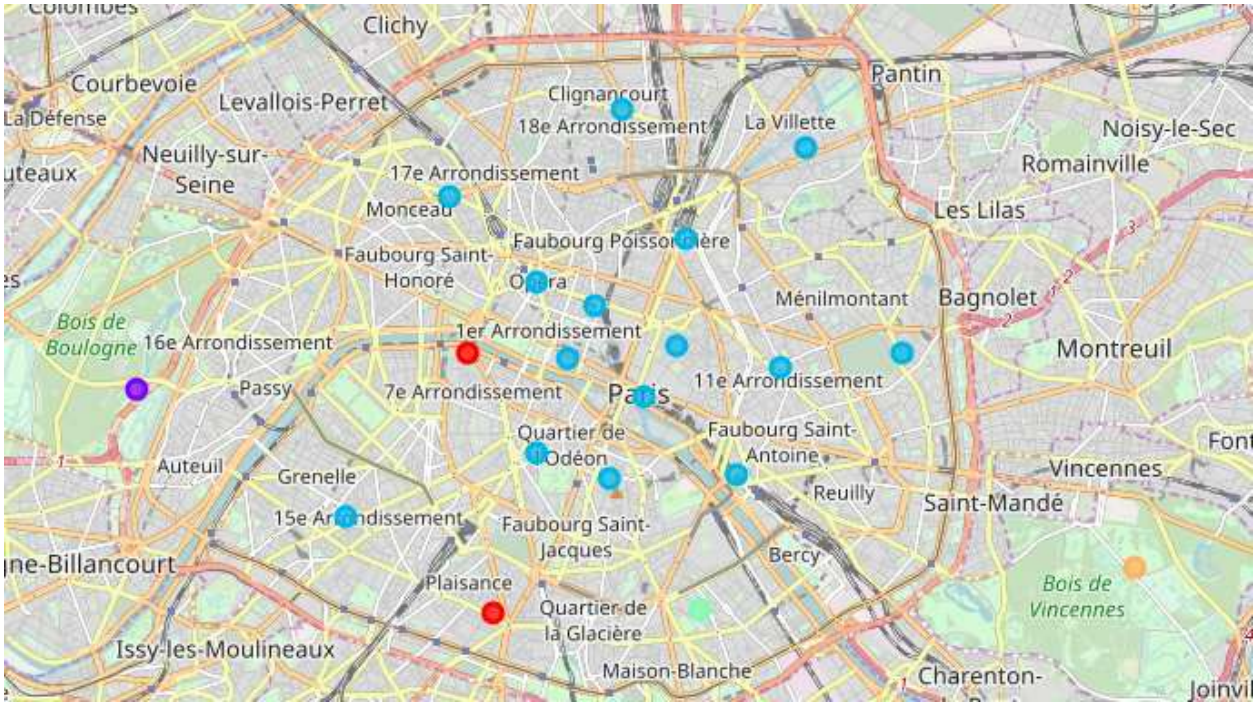
First, I will run K-Means to cluster the boroughs into **5** clusters.

Here is my merged table with cluster labels for each borough.

	Borough	Name	Quarters	Postal Codes	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	1st (Ie) R	Louvre	Saint-Germain-l'Auxerrois,Les Halles,Palais-Ro...	75001	48.861147	2.338028	2	French Restaurant	Café	Hotel	Plaza
1	2nd (IIe) R	Bourse	Gaillon,Vivienne,Mail,Bonne-Nouvelle	75002	48.867684	2.343126	2	French Restaurant	Wine Bar	Cocktail Bar	Hotel
2	3rd (IIIe) R	Temple	Arts-et-Métiers,Enfants-Rouges,Archives,Sainte...	75003	48.862683	2.358685	2	French Restaurant	Coffee Shop	Wine Bar	Italian Restaurant
3	4th (IVe) R	Hôtel-de-Ville	Saint-Merri,Saint-Gervais,Arsenal,Notre-Dame	75004	48.856426	2.352528	2	French Restaurant	Ice Cream Shop	Wine Bar	Cocktail Bar
4	5th (Ve) L	Panthéon	Quartier Saint-Victor,Jardin-des-Plantes,Val-d...	75005	48.846191	2.346079	2	French Restaurant	Bar	Hotel	Bakery

3. Results

Now, using the Cluster Labels, one hot encoded data and Borough information we visualize the clusters on the map of Paris with the help of folium.



We also analyze the clusters with respect to the most common venues:-

	Name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
6	Palais-Bourbon	0	French Restaurant	Plaza	Hotel	Café
13	Observatoire	0	French Restaurant	Hotel	Bakery	Bistro

	Name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
15	Passy	1	Pool	Bike Rental / Bike Share	Circus	Lake

	Name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Louvre	2	French Restaurant	Café	Hotel	Plaza
1	Bourse	2	French Restaurant	Wine Bar	Cocktail Bar	Hotel
2	Temple	2	French Restaurant	Coffee Shop	Wine Bar	Italian Restaurant
3	Hôtel-de-Ville	2	French Restaurant	Ice Cream Shop	Wine Bar	Cocktail Bar
4	Panthéon	2	French Restaurant	Bar	Hotel	Bakery

	Name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
12	Gobelins	3	Vietnamese Restaurant	Asian Restaurant	French Restaurant	Chinese Restaurant

	Name	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
11	Reuilly	4	Theater	Comedy Club	Bike Trail	Bike Rental / Bike Share

4. Discussion

Paris is the most populous European city and stands in top 3 for being the most expensive place and the influx of visitors to Paris is also astounding. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this metropol.

I used the Kmeans algorithm as part of this clustering study where I assumed no. of clusters to be 5. However, only 20 Arrondissement coordinates were used. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or monuments can also be drilled.

I ended the study by visualizing the data and clustering information on the Paris map.

5. Conclusion

As the tourism plays a great role in Paris , there must a cut throat competitions between tour advisors and for being the best customer satisfaction is the foremost thing to taken care of. Advising the neighborhoods for accommodation or visiting according to their interests of venues and monuments can save time and money of the visitors which would only result into good tour and satisfied customers for the tour advisors.

6. References:

- https://en.wikipedia.org/wiki/Paris
- https://en.wikipedia.org/wiki/Quarters_of_Paris
- https://en.wikipedia.org/wiki/Arrondissements_of_Paris