

# APPLIED DATASCIENCE CAPSTONE

## FINAL ASSIGNMENT REPORT

### Business problem

For this capstone project, I came up with an idea to leverage the Foursquare location data to explore the city of Paris, where I live. Paris is a big city and there is so much things to do and to visit that tourists maybe lost at first sight. Where to stay ? What are the most famous places to visit ? Where are the best restaurants to eat ? These are the questions most people ask when arriving in Paris. To answer these questions, we have to know precisely what people are looking for : historical places, trending places, shopping malls,... so that we can recommend the best borough to go depending on what they search. As you guess, the target here is the tourists visiting Paris. In 2018, they were about 50 millions.

So, here is the problem I would like to solve :

If your friends are coming to Paris, which places would you recommend them to visit ? What are the most common venues in each of the 20 boroughs of Paris ?

Answering this problem would allow people to gain time and make sure that they do not miss famous and interesting places in Paris !

### Data

To solve this problem, we will need different kinds of data.

First, we will rely on a dataset present on a Wikipedia page ([https://fr.wikipedia.org/wiki/Liste\\_des\\_quartiers\\_administratifs\\_de\\_Paris](https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris)). In this HTML page, we can find a table containing 5 columns : Borough, Neighbourhood, Population, Area and Density. Here, we will only focus on two of the five columns : all the neighbourhoods of Paris and their corresponding borough. We will work on the data to group the neighbourhoods by borough so that we can easily find the most common venues in each borough. For example, the borough « 1<sup>er</sup> arrondissement » contains the following neighbourhoods : Palais-Royal, Place Vendôme, Halles, Saint Germain l'Auxerrois.

In combination to the data mentionned above, we will use Foursquare location data as it will allow us to explore the city of Paris and access the venues located in the different boroughs. Foursquare is important because it will help us to find the trending places of the capital and all kinds of venues (restaurants, cultural places, hotels,...) that we will then recommend to the tourists.

## Methodology

Let me remind you our problem : 'If your friends are coming to Paris, which places would you recommend them to visit ? What are the most common venues in each of the 20 boroughs of Paris ?'.

To do this, we will first collect and analyze data dealing with Paris and its boroughs and neighbourhoods. I have found a table with many categories like the density and population of each borough, the neighbourhoods and the boroughs they belong to,... but let's drop columns that are not necessary for our study. First, we get a table with 2 columns (borough and neighbourhood) and 80 rows (the total number of neighbourhood in Paris). Then, we group the neighbourhoods according to their borough. The dataframe obtained contains 20 rows corresponding to the 20 neighbourhoods and will be more adequate for the continuation of our study. Here is the beginning of our dataframe.

	Borough	Neighbourhood
0	10e arrondissementit « de l'Entrepôt »	Saint-Vincent-de-Paul, Porte-Saint-Denis, Port...
1	11e arrondissementit « de Popincourt »	Folie-Méricourt, Saint-Ambroise, Roquette, Sai...
2	12e arrondissementit « de Reuilly »	Bel-Air, Picpus, Bercy, Quinze-Vingts
3	13e arrondissementit « des Gobelins »	Salpêtrière, Gare, Maison-Blanche, Croulebarbe
4	14e arrondissementit « de l'Observatoire »	Montparnasse, Parc-de-Montsouris, Petit-Montro...

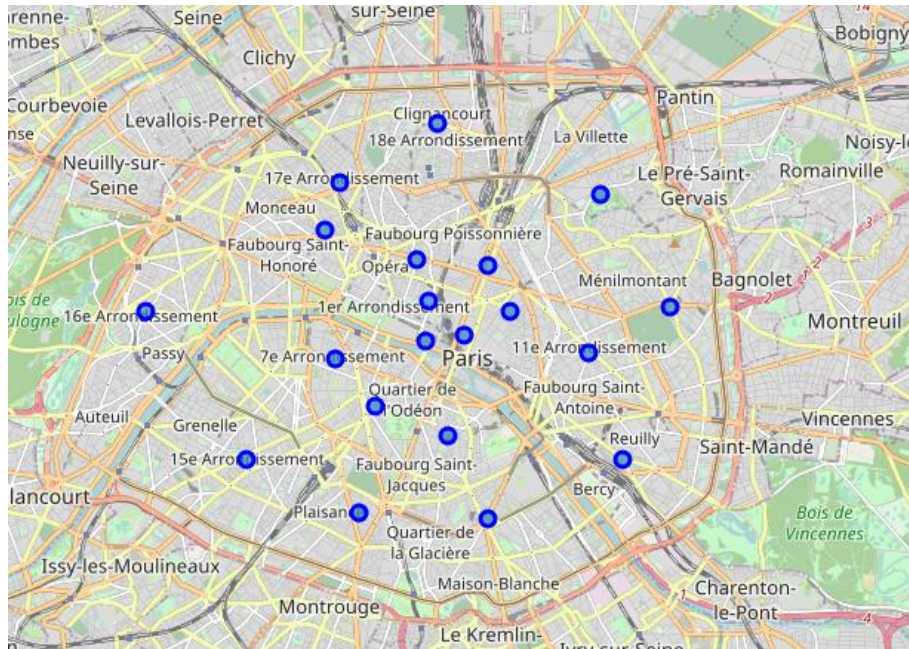
In addition, I searched for the geographical coordinates of each Parisian neighborhood to be able to add two columns to our table: latitude and longitude. These coordinates will be useful later on. Here is what we get :

	Borough	Neighbourhood	Latitude	Longitude
0	10e arrondissementit « de l'Entrepôt »	Saint-Vincent-de-Paul, Porte-Saint-Denis, Port...	48.8709	2.3561
1	11e arrondissementit « de Popincourt »	Folie-Méricourt, Saint-Ambroise, Roquette, Sai...	48.8574	2.3795
2	12e arrondissementit « de Reuilly »	Bel-Air, Picpus, Bercy, Quinze-Vingts	48.8412	2.3876
3	13e arrondissementit « des Gobelins »	Salpêtrière, Gare, Maison-Blanche, Croulebarbe	48.8322	2.3561
4	14e arrondissementit « de l'Observatoire »	Montparnasse, Parc-de-Montsouris, Petit-Montro...	48.8331	2.3264

We also use :

- the Geopy library to determine the geographic coordinates of a place;
- the Folium library to display a map of the place concerned;

Here, on the map of Paris, markers are added to indicate the different boroughs, as you can see below.

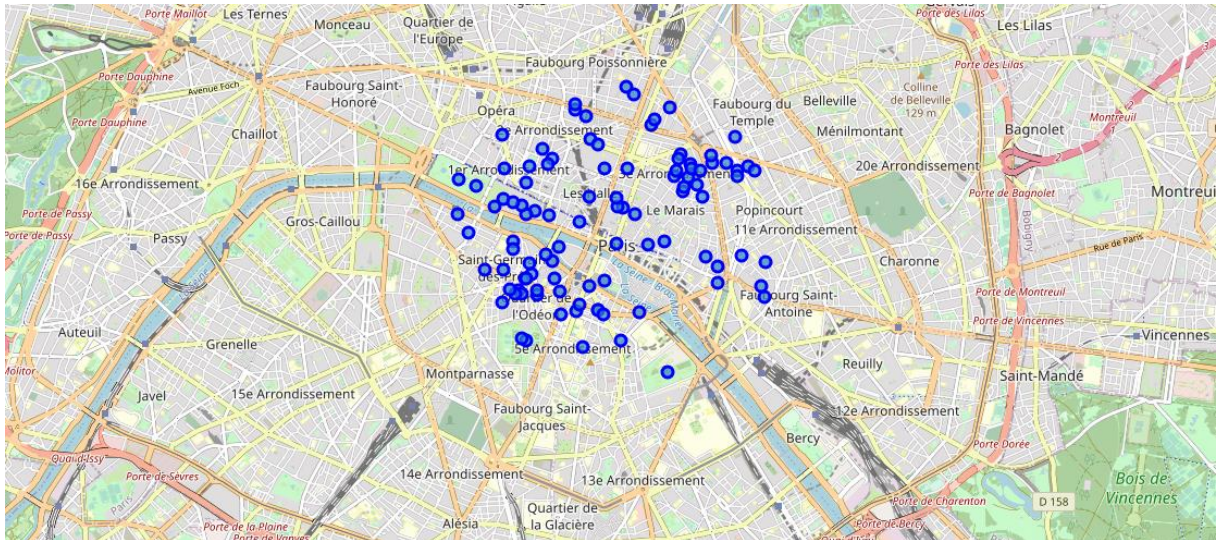


Afterwards, we use the Foursquare API to explore Paris and find the busiest places in the city that are likely to please our tourists. We send a request to Foursquare API to find 200 places to visit within a radius of 5 km from the center of Paris. This returns a json file that we will process to store the information in a table. Here is the table :



	name	categories	address	lat	lng	distance	postalCode	cc	city	state	country	formattedAddress	crossStreet	labeledLatLngs
0	Place de l'Hôtel de Ville – Esplanade de la Li...	Plaza	Place de l'Hôtel de Ville	48.856925	2.351412	25	75004	FR	Paris	Île-de-France	France	[Place de l'Hôtel de Ville, 75004 Paris, France]	NaN	NaN
1	Cathédrale Notre-Dame de Paris	Church	6 parvis Notre-Dame	48.853124	2.349561	421	75004	FR	Paris	Île-de-France	France	[6 parvis Notre-Dame (Place Jean-Paul II), 750...	Place Jean-Paul II	[{'label': 'display', 'lat': 48.85312443201169...
2	Boutique du Centre Pompidou	Souvenir Shop	Place Georges Pompidou	48.860706	2.352362	451	75004	FR	Paris	Île-de-France	France	[Place Georges Pompidou, 75004 Paris, France]	NaN	[{'label': 'display', 'lat': 48.86070574305312...

Using the Folium library, we create a map of Paris showing the 200 places to see.



Finally, we are going to look at the establishments present in each district of the city. To do this, we send a request to API Foursquare to ask for the places to be visited, specifying the neighborhood where they are located. We then group all the places according to their district. We get 20 establishments per neighborhood as a result of our request. To meet the needs and desires of our tourists, we will indicate the category of the establishment. We thus have 111 categories among which: restaurants, toy stores, art galleries, bookshops, etc.

	Borough	Alsatian Restaurant	American Restaurant	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Bakery	Bar	Basque Restaurant	Beer Bar	Beer Garden	Beer Store	Bi:
0	10e arrondissementdit « de l'Entrepôt »	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.05	0.00	0.0	0.0	0.0	0.0	
1	11e arrondissementdit « de Popincourt »	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.05	0.05	0.00	0.0	0.0	0.0	0.0	
2	12e arrondissementdit « de Reuilly »	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.05	0.00	0.0	0.0	0.0	0.0	
3	13e arrondissementdit « des Gobelins »	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.05	0.05	0.05	0.0	0.0	0.0	0.0	
4	14e arrondissementdit « de l'Observatoire »	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.10	0.0	0.0	0.0	0.0	

5 rows x 111 columns

Finally, a list of the top 10 most popular and must-see establishments is compiled for each of the 20 neighbourhoods. Thus, depending on where the tourist is, he will be able to access the different shops and cultural places around him. Here is the summary table.



	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	10e arrondissement « de l'Entrepôt »	Pizza Place	Coffee Shop	Cocktail Bar	Bakery	Hotel	Cheese Shop	Music Venue	Café	Sandwich Place	Burger Joint
1	11e arrondissement « de Popincourt »	Cocktail Bar	French Restaurant	Pastry Shop	Performing Arts Venue	Seafood Restaurant	Sandwich Place	Mediterranean Restaurant	Korean Restaurant	Bakery	Asian Restaurant
2	12e arrondissement « de Reuilly »	Chinese Restaurant	Bistro	Japanese Restaurant	Tapas Restaurant	Soup Place	French Restaurant	Bakery	Furniture / Home Store	Garden	Middle Eastern Restaurant
3	13e arrondissement « des Gobelins »	Hotel	Vietnamese Restaurant	Indian Restaurant	Asian Restaurant	Dessert Shop	Movie Theater	Museum	Pub	Bar	Sushi Restaurant

## Results

Following our study, if friends were to visit Paris, we would advise them to go mainly to the center of the capital, where the most popular tourist spots are located: for example, the City Hall, Notre-Dame Cathedral or the Georges Pompidou Center as mentioned in one of the previous tables. A map of Paris showing these different places is also proposed to them in order to better situate them in relation to each other. In addition, we have highlighted the tourist points in each district of Paris so that everyone can find the must-see points around their home.

## Discussion

Because each neighborhood has its specificities, we offer our friends a list of the 10 most frequented points in each neighborhood. It's up to them to see if they prefer to meet in a pizzeria or a cocktail bar. So, if our friends would like to taste French gastronomy, we will recommend the "Luxembourg" district (5th arrondissement) or the "Palais-Bourbon" district (7th arrondissement) because French restaurants are the most popular establishments there.

## Conclusion

To conclude, this study was an opportunity to highlight the must-see points of Paris and help tourists prepare their visits without wasting time. Thanks to the Foursquare API and the data available on the Wikipedia website, we were able to establish by neighborhood the most frequented places and thus present, for each of them, the top 10 of the most trendy places.