

**Capstone Project – The Battle of
Neighborhoods
Report**

1. Introduction:

The City of Paris, is the most populous city in the United States. It is diverse and is the financial capital of France. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States.

This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

2. Problem Description:

A restaurant is a business which prepares and serves food and drink to customers in return for money, either paid before the meal, after the meal, or with an open account. The City of Paris is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city's immigrant history.

1. Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens
2. Italian immigrants - New York-style pizza and Italian cuisine
3. Jewish immigrants and Irish immigrants - pastrami and corned beef
4. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city
5. mobile food vendors - Some 4,000 licensed by the city
6. Middle Eastern foods such as falafel and kebabs examples of modern New York street food
7. It is famous for not just Pizzerias, Cafe's but also for fine dining Michelin starred restaurants.
8. The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin.

So it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the Location such as :

1. Paris Population
2. Paris City Demographics
3. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?

4. Are there any venues like Gyms, Entertainment zones, Parks etc nearby where floating population is high etc
5. Who are the competitors in that location?
6. Cuisine served / Menu of the competitors
7. Segmentation of the Borough
8. Untapped markets
9. Saturated markets etc

Even though well funded XYZ Company Ltd. need to choose the correct location to start its first venture. If this is successful they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

3. Target Audience:

To recommend the correct location, XYZ Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighborhood of Paris city will be best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made.

This would interest anyone who wants to start a new restaurant in Paris city.

4. Success Criteria:

The success criteria of the project will be a good recommendation of borough/Neighborhood choice to XYZ Company Ltd based on Lack of such restaurants in that location and nearest suppliers of ingredients.

5. Data :

One city will be analyzed in this project : **Paris City.**

We will be using the below datasets for analyzing Paris city

Data 1 : Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. Link to the dataset is : https://geo.nyu.edu/catalog/nyu_2451_34572

Data 2 : Second data which will be used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets.

<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2>
Website-<https://www.grownyc.org/greenmarketco/foodbox>

GrowNYC's Fresh Food Box Program is a food access initiative that enables under-served communities to purchase fresh, healthy, and primarily regionally grown produce well below traditional retail prices.

A farmers' market is often defined as a public site used by two or more local or regional producers for the direct sale of farm products to consumers. In addition to fresh fruits and vegetables, markets may sell dairy products, fish, meat, baked goods, and other minimally processed foods.

Data 3 : For the below analysis we will get data from wikipedia as given below :

1. Paris Population
2. Paris City Demographics
3. Cuisine of Paris city

Data 4 : Paris city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in Paris City. The below is image of the Foursquare API data.

Data Research and Preparation

Import the Paris District Data

Arrondissements Municipaux for Paris CSV (administrative districts)

Paris is divided into 20 Arrondissements Municipaux (or administrative districts), shortened to just arrondissements. They are normally referenced by the arrondissement number rather than a name.

Data for the arrondissements is necessary to select the most suitable of these areas for new stores.

Initially looking to get this data by scraping the relevant Wikipedia page (https://en.wikipedia.org/wiki/Arrondissements_of_Paris), fortunately, after much research, this data is available on the web and can be manipulated and cleansed to provide a meaningful dataset to use.

Data from Open|DATA

France: <https://opendata.paris.fr/explore/dataset/arrondissements/table/?dataChart>

Also available from

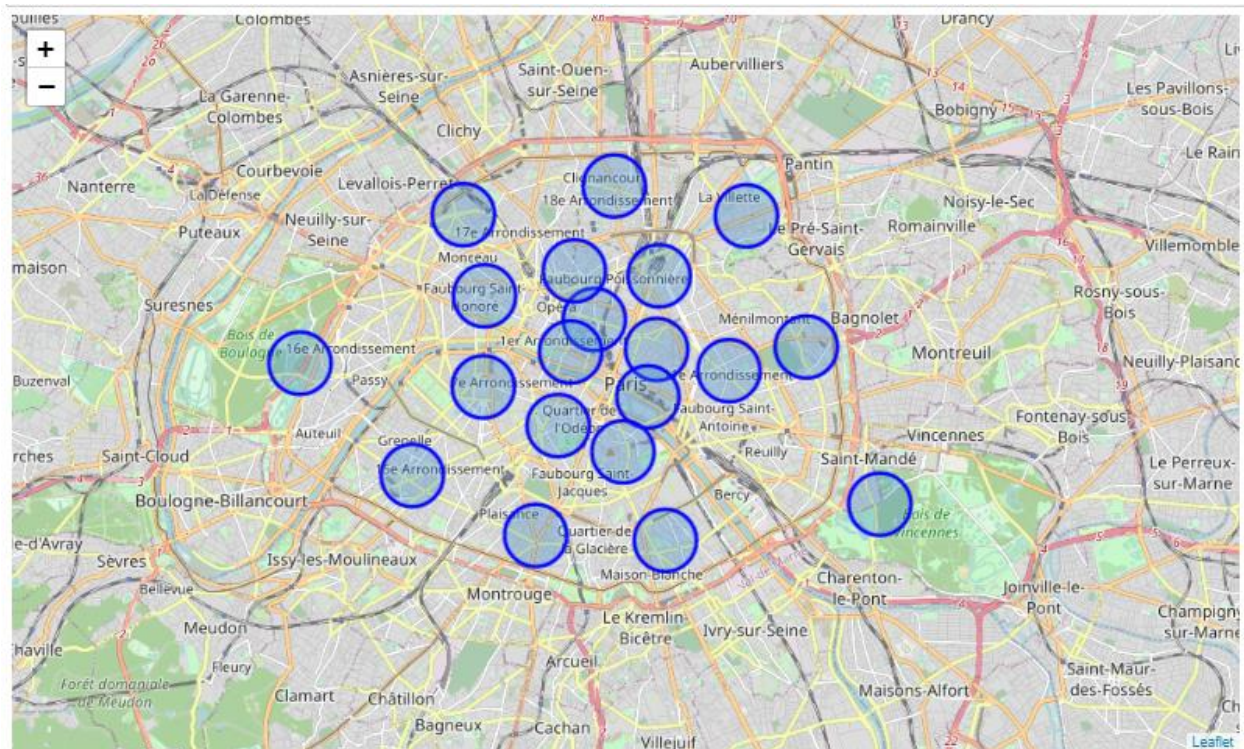
Opendatasoft: <https://data.opendatasoft.com/explore/dataset/arrondissements%40parisdata/export/>

	CAR	NAME	NSQAR	CAR.1	CARINSEE	LAR	NSQCO	SURFACE	PERIMETRE	Geometry_X	Geometry_Y
0	3	Temple	750000003	3	3	3eme Ardt	750001537	1170882828	4519264	48.862872	2.360001
1	19	Buttes-Chaumont	750000019	19	19	19eme Ardt	750001537	6792651129	11253182	48.887076	2.384821
2	14	Observatoire	750000014	14	14	14eme Ardt	750001537	5614877309	10317483	48.829245	2.326542
3	10	Entrepot	750000010	10	10	10eme Ardt	750001537	2891739442	6739375	48.876130	2.360728
4	12	Reuilly	750000012	12	12	12eme Ardt	750001537	16314782637	24089666	48.834974	2.421325
5	16	Passy	750000016	16	16	16eme Ardt	750001537	16372542129	17416110	48.860392	2.261971
6	11	Popincourt	750000011	11	11	11eme Ardt	750001537	3665441552	8282012	48.859059	2.380058
7	2	Bourse	750000002	2	2	2eme Ardt	750001537	991153745	4554104	48.868279	2.342803
8	4	Hotel-de-Ville	750000004	4	4	4eme Ardt	750001537	1600585632	5420908	48.854341	2.357630
9	17	Batignolles-Monceau	750000017	17	17	17eme Ardt	750001537	5668834504	10775580	48.887327	2.306777
10	18	Buttes-Montmartre	750000018	18	18	18eme Ardt	750001537	5996051308	9916464	48.892569	2.348161
11	1	Louvre	750000001	1	1	1er Ardt	750001537	1824612860	6054937	48.862563	2.336443
12	5	Pantheon	750000005	5	5	5eme Ardt	750001537	2539374623	6239195	48.844443	2.350715
13	7	Palais-Bourbon	750000007	7	7	7eme Ardt	750001537	4090057185	8099425	48.856174	2.312188
14	20	Menilmontant	750000020	20	20	20eme Ardt	750001537	5983446037	10704940	48.863461	2.401188
15	8	elysee	750000008	8	8	8eme Ardt	750001537	3880036397	7880533	48.872721	2.312554
16	9	Opera	750000009	9	9	9eme Ardt	750001537	2178303275	6471588	48.877164	2.337458
17	13	Gobelins	750000013	13	13	13eme Ardt	750001537	7149311091	11546547	48.828388	2.362272
18	15	Vaugirard	750000015	15	15	15eme Ardt	750001537	8494994081	13678798	48.840085	2.292826
19	6	Luxembourg	750000006	6	6	6eme Ardt	750001537	2153095586	6483687	48.849130	2.332898

6. Data Analysis

The Data Science Workflow for parts 3 & 4 includes:

- **Data Analysis and Location Data:**
 - Foursquare location data will be leveraged to explore or compare districts around Paris.
 - Data manipulation and analysis to derive subsets of the initial data.
 - Identifying the high traffic areas using data visualization and statistical analysis.
- **Visualization:**
 - Analysis and plotting visualizations.
 - Data visualization using various mapping libraries.
- **Discussion and Conclusions:**
 - Recommendations and results based on the data analysis.
 - Discussion of any limitations and how the results can be used, and any conclusions that can be drawn.



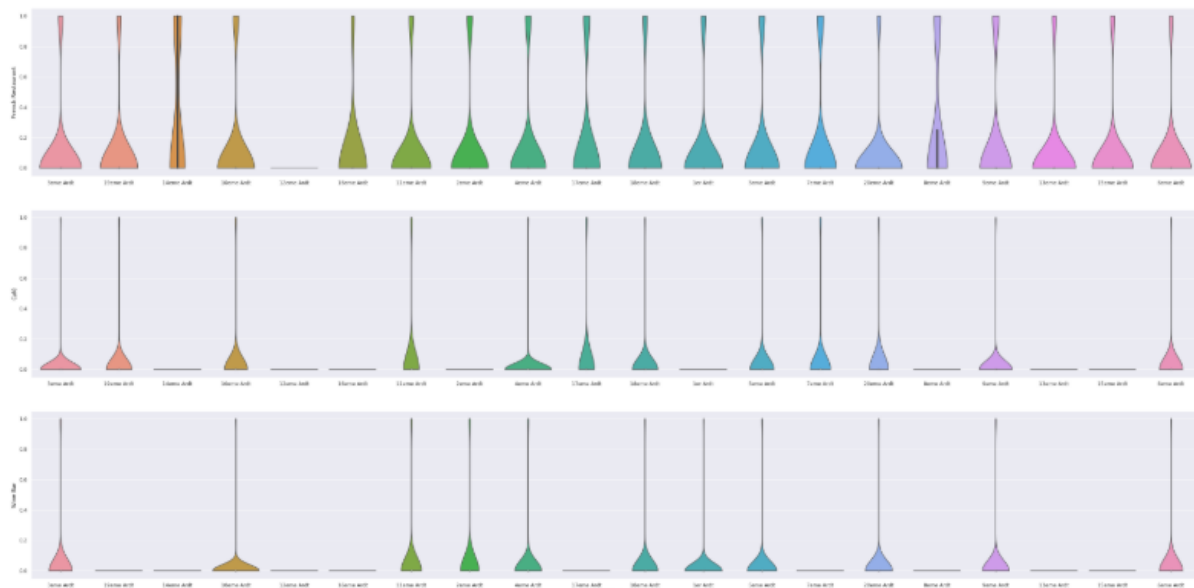
The top 10 venue categories for each neighborhood

	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	10eme Ardt	French Restaurant	Hotel	Coffee Shop	Café	Indian Restaurant	Bistro	Pizza Place	Japanese Restaurant	Bakery	Seafood Restaurant
1	11eme Ardt	French Restaurant	Café	Supermarket	Restaurant	Wine Bar	Pastry Shop	Italian Restaurant	Cocktail Bar	Bar	Bakery
2	12eme Ardt	Zoo Exhibit	Supermarket	Monument / Landmark	Zoo	Antique Shop	Argentinian Restaurant	Food & Drink Shop	Flower Shop	Fish & Chips Shop	Fast Food Restaurant
3	13eme Ardt	Vietnamese Restaurant	Asian Restaurant	French Restaurant	Chinese Restaurant	Thai Restaurant	Juice Bar	Coffee Shop	Park	Cambodian Restaurant	Cosmetics Shop
4	14eme Ardt	French Restaurant	Hotel	Bistro	Supermarket	Convenience Store	Bakery	Food & Drink Shop	Brasserie	Pizza Place	Sushi Restaurant
5	15eme Ardt	Hotel	Italian Restaurant	French Restaurant	Bakery	Coffee Shop	Lebanese Restaurant	Japanese Restaurant	Wine Shop	Park	Brasserie
6	16eme Ardt	Park	Plaza	Lake	French Restaurant	Boat or Ferry	Art Museum	Bus Station	Bus Stop	Donut Shop	Fast Food Restaurant
7	17eme Ardt	French Restaurant	Hotel	Italian Restaurant	Café	Plaza	Japanese Restaurant	Bakery	Wine Shop	Bistro	Pizza Place
8	18eme Ardt	Bar	French Restaurant	Hotel	Restaurant	Coffee Shop	Pizza Place	Convenience Store	Deli / Bodega	Seafood Restaurant	Café
9	19eme Ardt	French Restaurant	Bar	Supermarket	Hotel	Seafood Restaurant	Beer Bar	Bistro	Music Store	Coffee Shop	Steakhouse

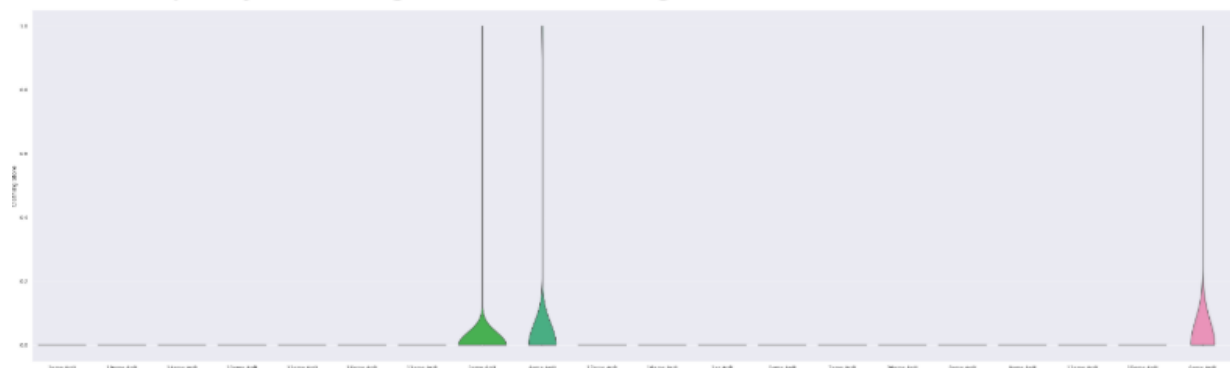
Let's look at their frequency of occurrence for all the Paris neighborhoods, isolating the categorical venues

These are the venue types that the client wants to have an abundant density of in the ideal store locations. I've used a violin plot from the seaborn library - it is a great way to visualise frequency distribution datasets, they display a density estimation of the underlying distribution.

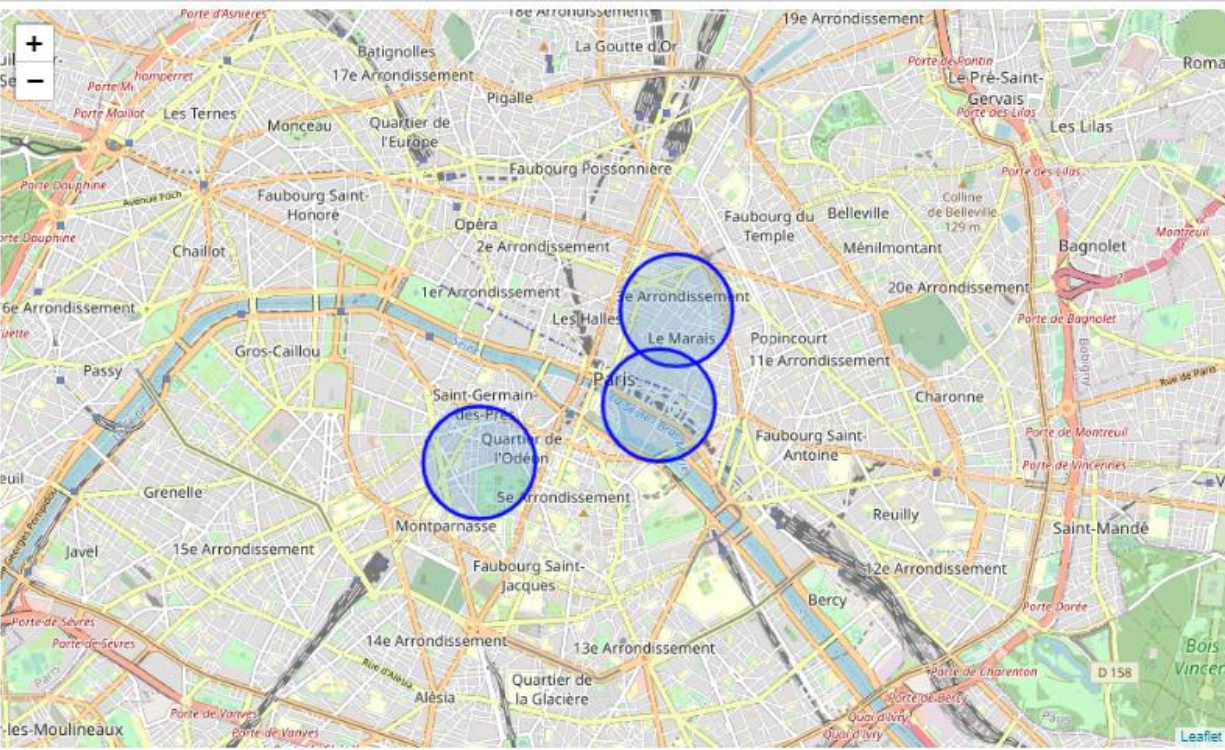
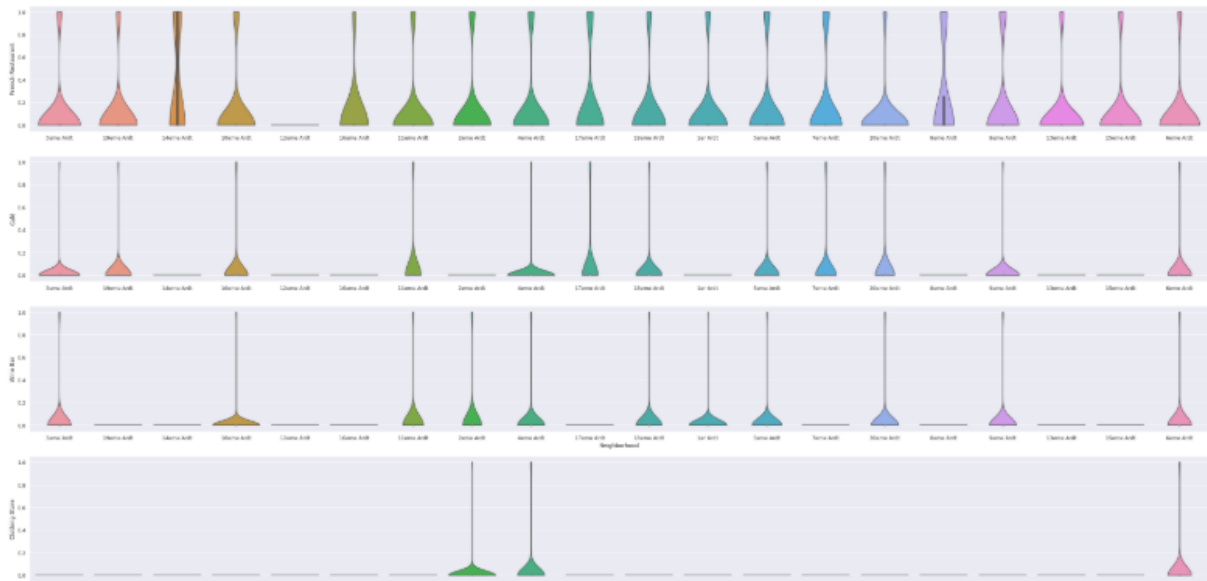
Frequency distribution for the top 3 venue categories for each neighborhood (click to enlarge)



Frequency of Clothing stores for each neighborhood



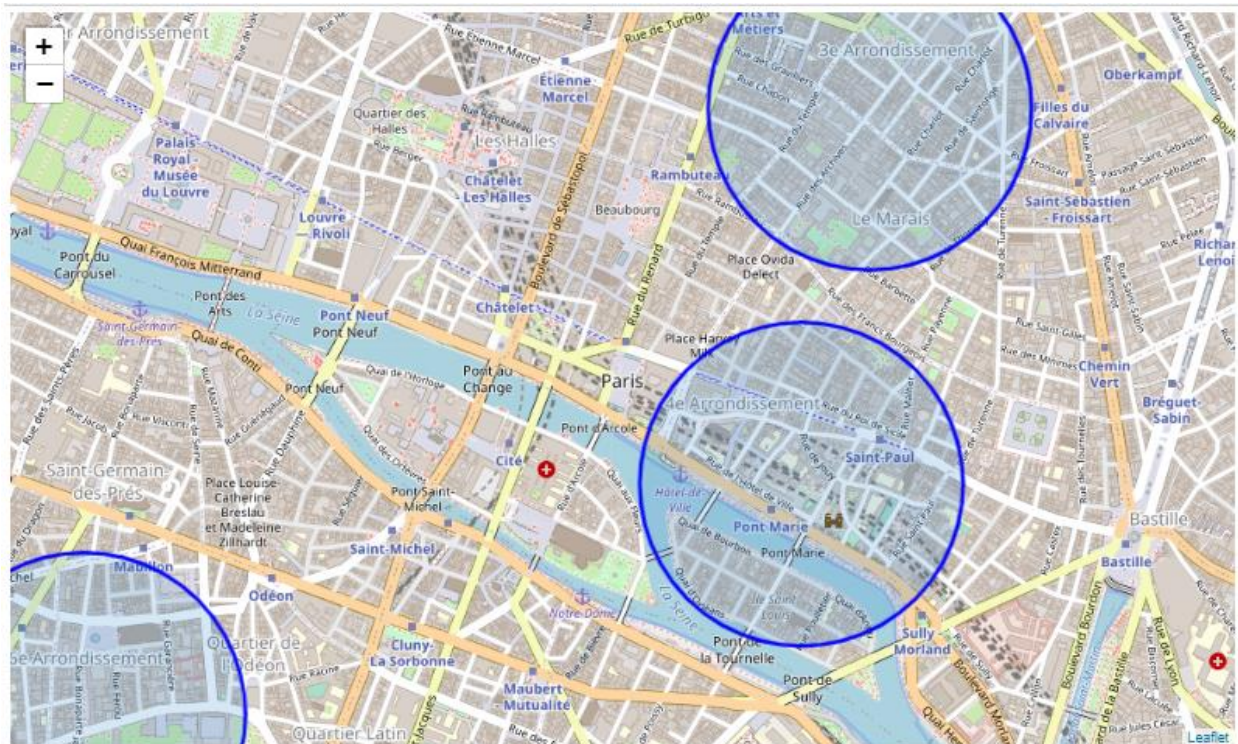
Frequency distribution for the top 3 venue categories for each neighborhood (includes clothing)



7. Observations

I guess it's not a surprise that these districts are all very centrally located in the circular arrangement of Paris's arrondissements. Locations fitting the criteria for popular venues would normally be in central locations in many cities of the world.

From this visualization it is clear that on a practical level, with no data to base decisions on, the circle of the 20 districts is very large, and researching and then visiting them all would be a daunting and time-consuming task. We have narrowed the search area down significantly from 20 potential districts to 3 that should suit the client's retail business.



8. Inferences

We have made inferences from the data in making the location recommendations, but that is exactly the point. There is no right or wrong answer or conclusion for the task at hand. The job of data analysis here is to steer a course for the location selection of new stores (i) to meet the criteria of being in neighborhoods that are lively with abundant leisure venues, and (ii) to narrow the search down to just a few of the main areas that are best suited to match the criteria.

9. Conclusions

There are many ways this analysis could have been performed based on different methodology and perhaps different data sources. I chose the method I selected as it was a straight forward way to narrow down the options, not complicating what is actually simple in many ways – meeting the criteria for the surrounding venues, and in my case, domain knowledge I have on the subject. I originally intended to use the clustering algorithms to cluster the data, but as it progressed it became obvious that this only complicated the task at hand. The analysis and results are not an end point, but rather a starting point that will guide the next part of the process to find specific store locations. The next part will involve domain knowledge of the industry, and perhaps, of the city itself. But the data analysis and resulting recommendations have greatly narrowed down the best district options based on data and what we can infer from it.

Without leveraging data to make focussed decisions, the process could have been drawn out and resulted in new stores opening in sub-standard areas for this retailer. Data has helped to provide a better strategy and way forward, these data-driven decisions will lead to a better solution in the end.