The Battle of the Neighbourhoods - Report

Suitable New Store Locations in Paris for a Fashion Retailer

This Report contains multiple parts:

A description of the problem and a discussion of the background - Week 1

A description of the data and how it will be used to solve the problem - Week 1

Methodology and Exploratory Data Analysis - Week 2

Inferences and Discussion - Week 2

1 Introduction and Discussion of the Business Objective and Problem

Locations for New Fashion Stores in High Traffic Areas in Paris France

The Task At Hand

A digitally native vertical fashion retailer, with a substantial e-commerce footprint, has begun the rollout of brick and mortar stores as part of their omnichannel retail strategy. After rolling out stores in a few select cities by guessing where the best locations were to open, as part of their store expansion for Paris they've decided to be more informed and selective, and take the time to do some research.

I've been given the exciting task of assisting them to make data-driven decisions on the new locations that are most suitable for their new stores in Paris. This will be a major part of their decision-making process, the other being on the ground qualitative analysis of districts once this data and report are reviewed and studied.

The fashion brand is not what is considered high-end, they are positioned in upper end of the fast fashion market. As such, they do not seek stores in the premium upmarket strips like Avenue Montaigne, but rather, in high traffic areas where consumers go for shopping, restaurants and entertainment. Foursquare data will be very helpful in making data-driven decisions about the best of those areas.

Criteria

Qualitative data from another retailer that they know, suggests that the best locations to open new fashion retail stores may not only be where other clothing is located. This data strongly suggests that the best places are in fact areas that are near French Restaurants, Cafés and Wine Bars. Parisians are very social people that frequent these place often, so opening new stores in these locations is becoming popular.

The analysis and recommendations for new store locations will focus on general districts with these establishments, not on specific store addresses. Narrowing down the best district options derived from analysis allows for either further research to be conducted, advising agents of the chosen district, or on the ground searching for specific sites by the company's personnel.

Why Data?

Without leveraging data to make decisions about new store locations, the company could spend countless hours walking around districts, consulting many real estate agents with their own district biases, and end up opening in yet another location that is not ideal.

Data will provide better answers and better solutions to their task at hand.

Outcomes

The goal is to identify the best districts - Arrondissements - to open new stores as part of the company's plan. The results will be translated to management in a simple form that will convey the data-driven analysis for the best locations to open stores.

2 The Data Science Workflow

Data Requirements

The main districts in Paris are divided into 20 Arrondissements Municipaux (administrative districts), shortened to arrondissements.

The data regarding the districts in Paris needs to be researched and a suitable useable source identified. If it is found but is not in a useable form, data wrangling and cleaning will have to be performed.
The cleansed data will then be used alongside Foursquare data, which is readily available. Foursquare location data will be leveraged to explore or compare districts around Paris, identifying the high traffic areas where consumers go for shopping, dining and entertainment - the areas where the fashion brand are most interested in opening new stores.
The Data Science Workflow for Part 1 & 2 includes the following:
Outline the initial data that is required:
District data for Paris including names, location data if available, and any other details required.

Obtain the Data:
Research and find suitable sources for the district data for Paris.
Access and explore the data to determine if it can be manipulated for our purposes.

Initial Data Wrangling and Cleaning:
Clean the data and convert to a useable form as a dataframe.
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 visualisation and tatistical nalysis.
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.

2 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 and 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 relevent 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/

Data looks like below.

	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

3 Methodology and Exploratory Data Analysis

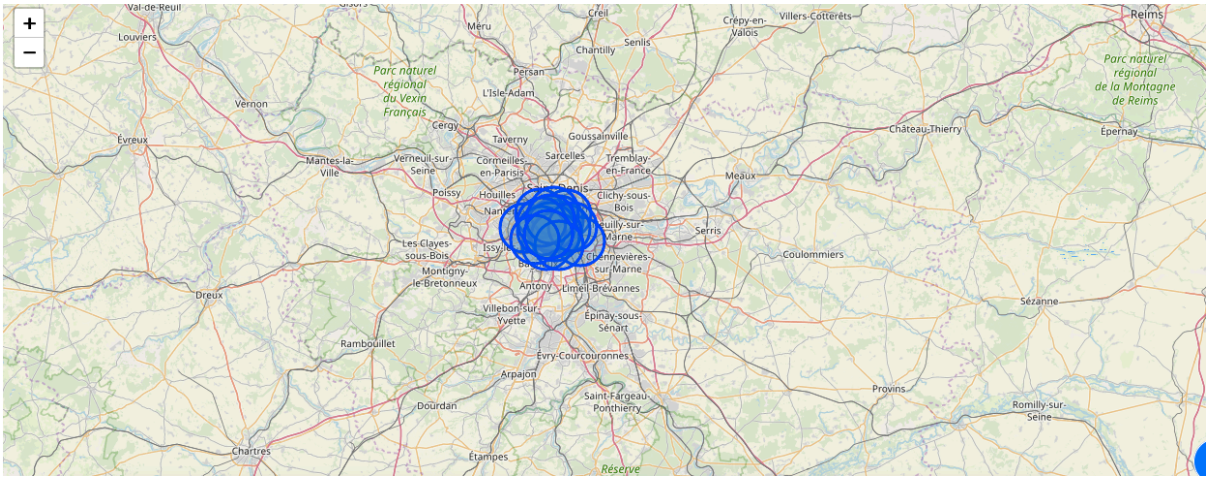
The Data Science Workflow for parts 3 & 4 includes:

- **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 visualisation and tatistical nalysis.

- **Visualization:**
  - Analysis and plotting visualizations.
  - Data visualization using various mapping libraries.

- **Discussion and Conclusions:**
  - Recomendations 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.

Create a map of Paris with districts superimposed



Use the Foursquare API to explore the Arrondissements of Paris (Neighborhoods)
Exploratory data analysis

Explore the first district in our dataframe to become familiar with the data (use the French descriptive arrondissement name)

The first arrondissement is identified as 3eme Ardt

Get the top 100 venues that are in the neighborhood 3eme Ardt within a radius of 500 meters

	name	categories	lat	lng
0	Mmmozza	Sandwich Place	48.863910	2.360591
1	Square du Temple	Park	48.864475	2.360816
2	Marché des Enfants Rouges	Farmers Market	48.862806	2.361996
3	Chez Alain Miam Miam	Sandwich Place	48.862781	2.362064
4	Chez Alain Miam Miam	Sandwich Place	48.862369	2.361950
5	Fromagerie Jouannault	Cheese Shop	48.862947	2.362530
6	Les Enfants Rouges	Wine Bar	48.863013	2.361260
7	Okomusu	Okonomiyaki Restaurant	48.861453	2.360879
8	Hôtel Jules & Jim	Hotel	48.863496	2.357395
9	Musée de la Chasse et de la Nature	Museum	48.861507	2.358624
10	Bontemps	Dessert Shop	48.863956	2.360725
11	Strada Café	Café	48.862224	2.357379
12	Comme un Roman	Bookstore	48.863095	2.362208
13	Paris New York	Burger Joint	48.863843	2.362661
14	SoMa	Japanese Restaurant	48.861511	2.362146
15	Les Enfants Du Marché	French Restaurant	48.862746	2.361950
16	Hank Burger	Burger Joint	48.861340	2.358304
17	Nanashi	Japanese Restaurant	48.863702	2.363025
18	The Broken Arm	Boutique	48.864653	2.361433
19	Biglove Caffè	Italian Restaurant	48.862063	2.363557

Calculate how many unique venue categories there are

Analyze each of the Neighbourhoods

Group rows by neighborhood and take the mean of the frequency of occurrence of each category

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Udon Restaurant	Vegetarian / Vegan Restaurant
0	10eme Ardt	0.000000	0.020000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.020000	...	0.00	0.010000
1	11eme Ardt	0.015152	0.015152	0.000000	0.00	0.00	0.000000	0.015152	0.000000	0.015152	...	0.00	0.015152
2	12eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000
3	13eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.206897	...	0.00	0.000000
4	14eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000
5	15eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.015873	0.000000	...	0.00	0.000000
6	16eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.090909	0.000000	0.000000	...	0.00	0.000000
7	17eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.019608	0.000000	0.000000	...	0.00	0.000000
8	18eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.020000
9	19eme Ardt	0.000000	0.000000	0.024390	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000
10	1er Ardt	0.000000	0.000000	0.000000	0.01	0.00	0.000000	0.030000	0.000000	0.000000	...	0.02	0.000000
11	20eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000
12	2eme Ardt	0.000000	0.000000	0.010000	0.00	0.00	0.000000	0.000000	0.000000	0.010000	...	0.00	0.000000
13	3eme Ardt	0.000000	0.000000	0.000000	0.00	0.01	0.020000	0.010000	0.000000	0.000000	...	0.00	0.020000
14	4eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.020000	0.010000	0.010000	0.000000	...	0.00	0.000000
15	5eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.00	0.000000
16	6eme Ardt	0.000000	0.000000	0.011364	0.00	0.00	0.000000	0.011364	0.000000	0.000000	...	0.00	0.000000
17	7eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.030000	0.000000	0.000000	...	0.00	0.010000
18	8eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.013699	0.013699	0.000000	0.000000	...	0.00	0.000000
19	9eme Ardt	0.000000	0.010000	0.000000	0.00	0.01	0.000000	0.000000	0.000000	0.000000	...	0.00	0.020000

Print each neighborhood with it's top 10 most common venues

```
----10eme Ardt----
                  venue  freq
```

```
0    French Restaurant  0.12
1                Hotel  0.05
2              Bistro  0.04
3                Café  0.04
4          Coffee Shop  0.04
5          Pizza Place  0.03
6  Japanese Restaurant  0.03
7    Indian Restaurant  0.03
8   Italian Restaurant  0.03
9             Wine Bar  0.02


----11eme Ardt----
                venue  freq
0    French Restaurant  0.11
1           Restaurant  0.06
2             Wine Bar  0.05
3                Café  0.05
4          Pastry Shop  0.05
5         Cocktail Bar  0.03
6  Japanese Restaurant  0.03
7          Pizza Place  0.03
8                  Bar  0.03
9              Bistro  0.03


----12eme Ardt----
                      venue  freq
0                       Zoo  0.25
1               Supermarket  0.25
2  Bike Rental / Bike Share  0.25
3       Monument / Landmark  0.25
4              Music Store  0.00
5                Nightclub  0.00
6              Noodle House  0.00
7                    Office  0.00
8    Okonomiyaki Restaurant  0.00
9             Optical Shop  0.00


----13eme Ardt----
                   venue  freq
0       Asian Restaurant  0.21
1  Vietnamese Restaurant  0.19
2     Chinese Restaurant  0.09
3        Thai Restaurant  0.09
4      French Restaurant  0.07
5              Juice Bar  0.05
6    Japanese Restaurant  0.03
7         Cosmetics Shop  0.02
8         Sandwich Place  0.02
9           Gourmet Shop  0.02


----14eme Ardt----
                venue  freq
0     French Restaurant  0.37
1                Hotel  0.10
2           Supermarket  0.07
3               Bistro  0.07
4    Italian Restaurant  0.03
```

```
5     Convenience Store  0.03
6           Pizza Place  0.03
7  Fast Food Restaurant  0.03
8     Food & Drink Shop  0.03
9                 Café  0.03


----15eme Ardt----
                venue  freq
0               Hotel  0.13
1   Italian Restaurant  0.10
2    French Restaurant  0.10
3         Coffee Shop  0.05
4    Indian Restaurant  0.03
5              Bistro  0.03
6      Thai Restaurant  0.03
7  Japanese Restaurant  0.03
8                Park  0.03
9              Bakery  0.03


----16eme Ardt----
               venue  freq
0               Lake  0.18
1              Plaza  0.18
2  French Restaurant  0.09
3        Art Museum  0.09
4               Pool  0.09
5      Boat or Ferry  0.09
6        Bus Station  0.09
7           Bus Stop  0.09
8               Park  0.09
9  Afghan Restaurant  0.00

----17eme Ardt----
                venue  freq
0    French Restaurant  0.22
1               Hotel  0.16
2   Italian Restaurant  0.08
3              Bakery  0.08
4                Café  0.04
5              Plaza  0.04
6              Bistro  0.04
7  Japanese Restaurant  0.04
8                 Bar  0.04
9        Grocery Store  0.02


----18eme Ardt----
                  venue  freq
0                   Bar  0.16
1      French Restaurant  0.12
2           Pizza Place  0.06
3           Coffee Shop  0.04
4      Convenience Store  0.04
5                 Café  0.04
6               Theater  0.04
7                 Hotel  0.04
8  Vietnamese Restaurant  0.04
9            Restaurant  0.04
```

```
----19eme Ardt----
                venue  freq
0                 Bar  0.10
1    French Restaurant  0.10
2          Restaurant  0.07
3  Japanese Restaurant  0.05
4         Supermarket  0.05
5   Seafood Restaurant  0.05
6               Hotel  0.05
7              Bistro  0.05
8           Beer Bar  0.05
9               Café  0.02


----1er Ardt----
                venue  freq
0    French Restaurant  0.11
1               Hotel  0.08
2                Café  0.08
3  Japanese Restaurant  0.06
4               Plaza  0.06
5             Exhibit  0.04
6          Coffee Shop  0.04
7           Restaurant  0.03
8                 Bar  0.03
9           Art Museum  0.03


----20eme Ardt----
                venue  freq
0              Bakery  0.08
1                 Bar  0.08
2   French Restaurant  0.08
3               Café  0.06
4              Plaza  0.06
5             Bistro  0.06
6  Italian Restaurant  0.06
7    Sushi Restaurant  0.04
8   Indian Restaurant  0.04
9  Mexican Restaurant  0.02


----2eme Ardt----
                venue  freq
0   French Restaurant  0.12
1           Wine Bar  0.06
2             Bistro  0.05
3       Cocktail Bar  0.05
4  Italian Restaurant  0.03
5        Salad Place  0.03
6              Bakery  0.03
7               Hotel  0.03
8      Ice Cream Shop  0.02
9         Coffee Shop  0.02


----3eme Ardt----
                venue  freq
0    French Restaurant  0.08
1            Wine Bar  0.04
```

```
2          Coffee Shop  0.04
3   Italian Restaurant  0.04
4               Bistro  0.03
5                 Café  0.03
6         Burger Joint  0.03
7  Moroccan Restaurant  0.03
8       Sandwich Place  0.03
9  Japanese Restaurant  0.03
```

The top 10 venue categories for each neighborhood

This is a very useful results table that can provide at a glance information for all of the districts. Even once any conclusions are drawn further into the data workflow, we can refer back to this table for meaaningful insights about the top categories of businesses in all the neighbourhoods. Even without actual counts and numbers, it makes a great reference table for the client

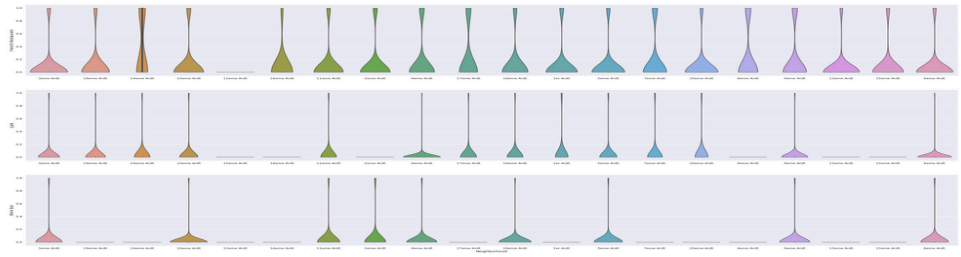
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	10eme Ardt	French Restaurant	Hotel	Coffee Shop	Café	Bistro	Pizza Place	Japanese Restaurant	Indian Restaurant	Italian Restaurant	Bar
1	11eme Ardt	French Restaurant	Restaurant	Wine Bar	Pastry Shop	Café	Japanese Restaurant	Pizza Place	Bar	Bistro	Theater
2	12eme Ardt	Zoo	Monument / Landmark	Supermarket	Bike Rental / Bike Share	Furniture / Home Store	Fountain	Food & Drink Shop	Flower Shop	Fish & Chips Shop	Fast Food Restaurant
3	13eme Ardt	Asian Restaurant	Vietnamese Restaurant	Thai Restaurant	Chinese Restaurant	French Restaurant	Juice Bar	Japanese Restaurant	Cambodian Restaurant	Bookstore	Convenience Store
4	14eme Ardt	French Restaurant	Hotel	Bistro	Supermarket	Italian Restaurant	Bakery	Pizza Place	Café	Fast Food Restaurant	Food & Drink Shop
5	15eme Ardt	Hotel	Italian Restaurant	French Restaurant	Coffee Shop	Japanese Restaurant	Bakery	Park	Bistro	Lebanese Restaurant	Thai Restaurant
6	16eme Ardt	Lake	Plaza	Pool	Boat or Ferry	Bus Station	Bus Stop	Park	French Restaurant	Art Museum	Cupcake Shop
7	17eme Ardt	French Restaurant	Hotel	Italian Restaurant	Bakery	Plaza	Bar	Japanese Restaurant	Bistro	Café	Flower Shop
8	18eme Ardt	Bar	French Restaurant	Pizza Place	Convenience Store	Restaurant	Theater	Café	Vietnamese Restaurant	Coffee Shop	Hotel

The business types criteria specified by the client! *'French Restaurants', 'Cafés' and 'Wine Bars'*¶

Let's look at their frequency of occurance 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 enlage)



The Neighborhoods

So as we can see from the analysis there are 8 neighborhoods to open new stores - according to the criteria that they have the 3 specified venues in a great frequency (*French Restaurants, Cafés and Wine Bars).* They are as follows:

Neighborhoods

- **3eme Ardt**
- **10eme Ardt**
- **11eme Ardt**
- **4eme Ardt**
- **18eme Ardt**
- **18eme Ardt**
- **5eme Ardt**
- **9eme Ardt**
- **6eme Ardt**

Let's take this further with some exploration and Inferential Analysis

We have the 8 neighborhoods that all include the venue category criteria.
But if we included the 'Clothing_Store" venue category into the analysis, then we might be able to make some inferences based on the data, and domain knowledge of marketing and the industry, to focus the list.

Let's look at the venue category - *"Clothing Store*'

So there are 5 neighborhoods that have a significant frequency density of clothing stores.

4 Inferences and Discussion

Chosen Neighbourhoods - Results

Inferential analysis using the data, as well as domain knowledge of retail and marketing, allow the list to be focussed to just 3 neighbourhoods from the previous 8.

The reasoning being that if the 3 criteria have been met - identifying neighbourhoods that are lively with Restaurants, Cafés and Wine Bars - adding Clothing Stores into the mix of stores in the area is a significant bonus. Having some of the same category of stores in the same area - especially in fashion retail - is very desirable as a retailer.

So we can increase the criteria to include *Restaurants, Cafés, Wine Bars and Clothing Stores* - which narrows down and focuses the suggested districts for new stores to be located, and at the same time provides better locations for the brand.

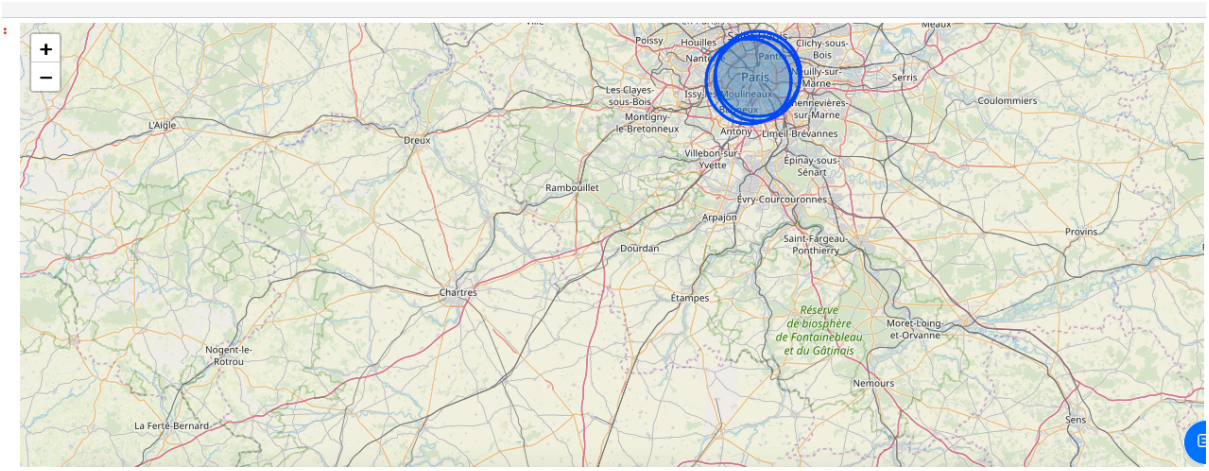
So the final 3 prospective neighborhoods for new store locations are where 4 criteria are met:

- 3eme Ardt : Arrondissement 3, Temple

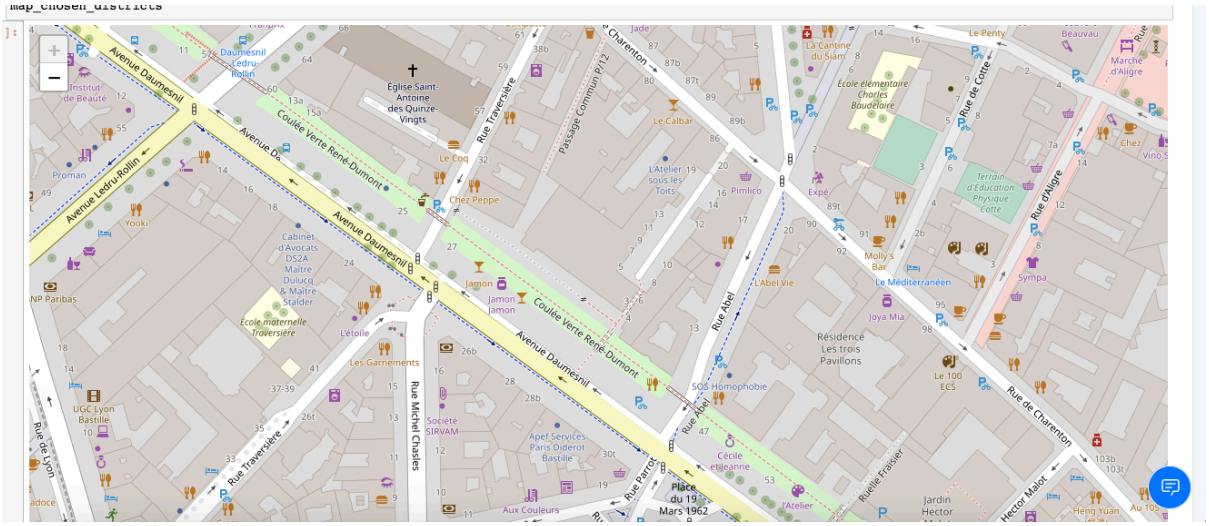
- 4eme Ardt : Arrondissement 4, Hotel-de-Ville

- 6eme Ardt : Arrondissement 6, Luxembourg

Let's look at the 3 districts on a Paris map



Or for a closer view of Central Paris



5 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 visualisation it is clear that on a practical level, with no data to base decisions on, the circle of the 20 districs 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.

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 neighbourhoods 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.

6 Conclusions

There are many ways this analysis could have been performed based on different methodolgy 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 the critera for the surrounding venues, and in my case, domain knowledge I have on the subject. I originally intended to use the clustering algorythms 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.

Thanks for taking part in my Data Science journey!