

Suitable New Store Locations in Paris for a Fashion Retailer



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

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 data frame.

1. 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/>

2. Exploring, Wrangling and Cleaning the Data:

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values from earlier seasons, because of lack of record keeping.

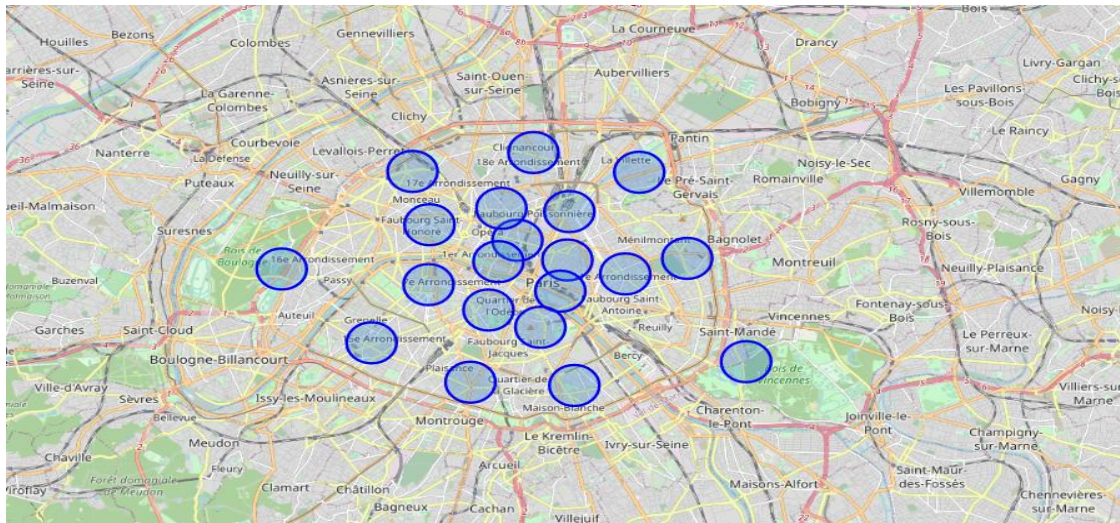
	Arrondissement_Num	Neighborhood	NSQAR	CAR.1	CARINSEE	French_Name	NSQCO	SURFACE	PERIMETRE	Latitude	Longi
0	3	Temple	750000003	3	3	3eme Ardt	750001537	1170882828	4519264	48.862872	2.3600
1	19	Buttes-Chaumont	750000019	19	19	19eme Ardt	750001537	6792651129	11253182	48.887076	2.3848
2	14	Observatoire	750000014	14	14	14eme Ardt	750001537	5614877309	10317483	48.829245	2.3265
3	10	Entrepot	750000010	10	10	10eme Ardt	750001537	2891739442	6739375	48.876130	2.3607
4	12	Reuilly	750000012	12	12	12eme Ardt	750001537	16314782637	24089666	48.834974	2.4213
5	16	Passy	750000016	16	16	16eme Ardt	750001537	16372542129	17416110	48.860392	2.2619
6	11	Popincourt	750000011	11	11	11eme Ardt	750001537	3665441552	8282012	48.859059	2.3800
7	2	Bourse	750000002	2	2	2eme Ardt	750001537	991153745	4554104	48.868279	2.3428
8	4	Hotel-de-Ville	750000004	4	4	4eme Ardt	750001537	1600585632	5420908	48.854341	2.3576
9	17	Batignolles-Monceau	750000017	17	17	17eme Ardt	750001537	5668834504	10775580	48.887327	2.3067
10	18	Buttes-Montmartre	750000018	18	18	18eme Ardt	750001537	5996051308	9916464	48.892569	2.3481
11	1	Louvre	750000001	1	1	1er Ardt	750001537	1824612860	6054937	48.862563	2.3364

Data Analysis

3. Methodology and Exploratory Data Analysis

- **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.

Map of Paris with districts superimposed



4. Exploratory data analysis:

Explore the first district in our data frame 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

Send the GET request and examine the results.

1. Create a nearby venues function for all the neighborhoods in Paris
2. Create a new data frame called for the venues of Paris called *Paris-venues*
3. Check how many venues were returned for each neighborhood
4. Calculate how many unique venue categories there are
5. Analyze each of the Neighborhoods: Do the one hot encoding
6. Group rows by neighborhood and take the mean of the frequency of occurrence of each category
7. 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

```

Put that data into a pandas data frame and sort the venues in descending order

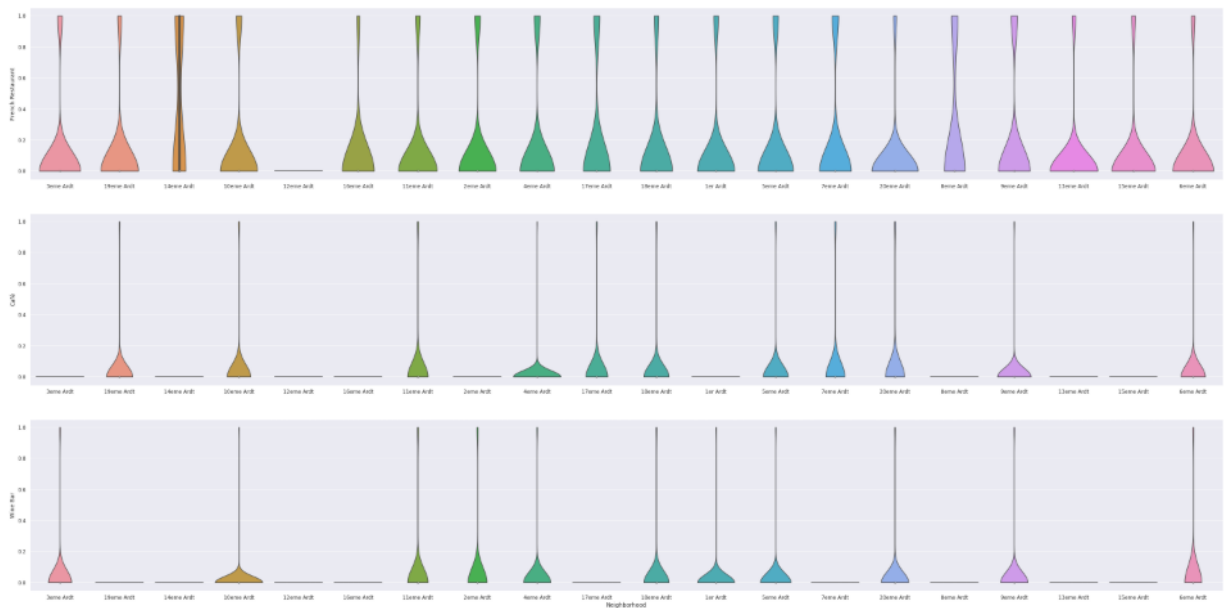
The business types criteria specified by the client! 'French Restaurants', 'Cafés' and 'Wine Bars'

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 visualize frequency distribution datasets, they display a density estimation of the underlying distribution.

Use **Violin plots** from seaborn library – it is a great way to visualize frequency distribution datasets, then they display the density estimation of the underlying distribution.

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



(Violin Plot)

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:

3eme Ardt

10eme Ardt

11eme Ardt

4eme Ardt

18eme Ardt

18eme Ardt

5eme Ardt

9eme Ardt

6eme Ardt

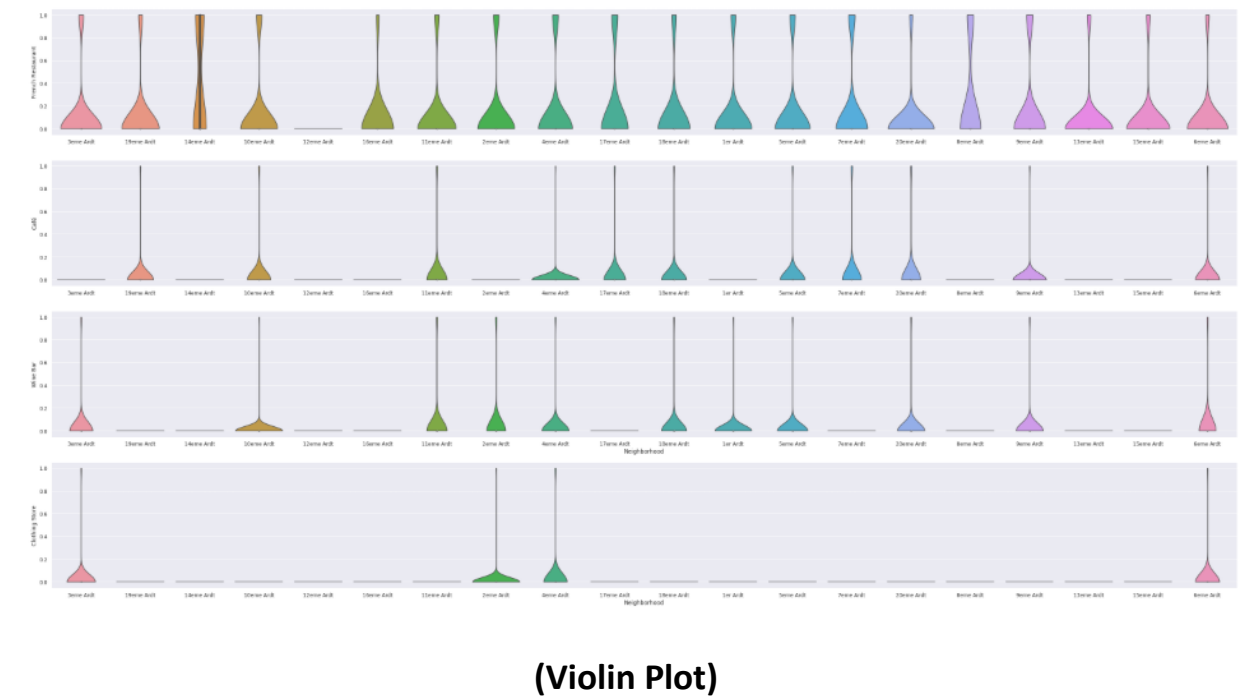
Let's take this further with some exploration and Inferential Analysis:

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":



Let's add this to the analysis with the other 3 specified categories as below.
Frequency distribution for the top 3 venue categories for each neighborhood (includes clothing)



5. Inferences and Discussion

Chosen Neighborhoods - Results Inferential analysis using the data, as well as domain knowledge of retail and marketing, allow the list to be focused to just 3 neighborhoods from the previous 8.

The reasoning being that if the 3 criteria have been met - identifying neighborhoods 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.

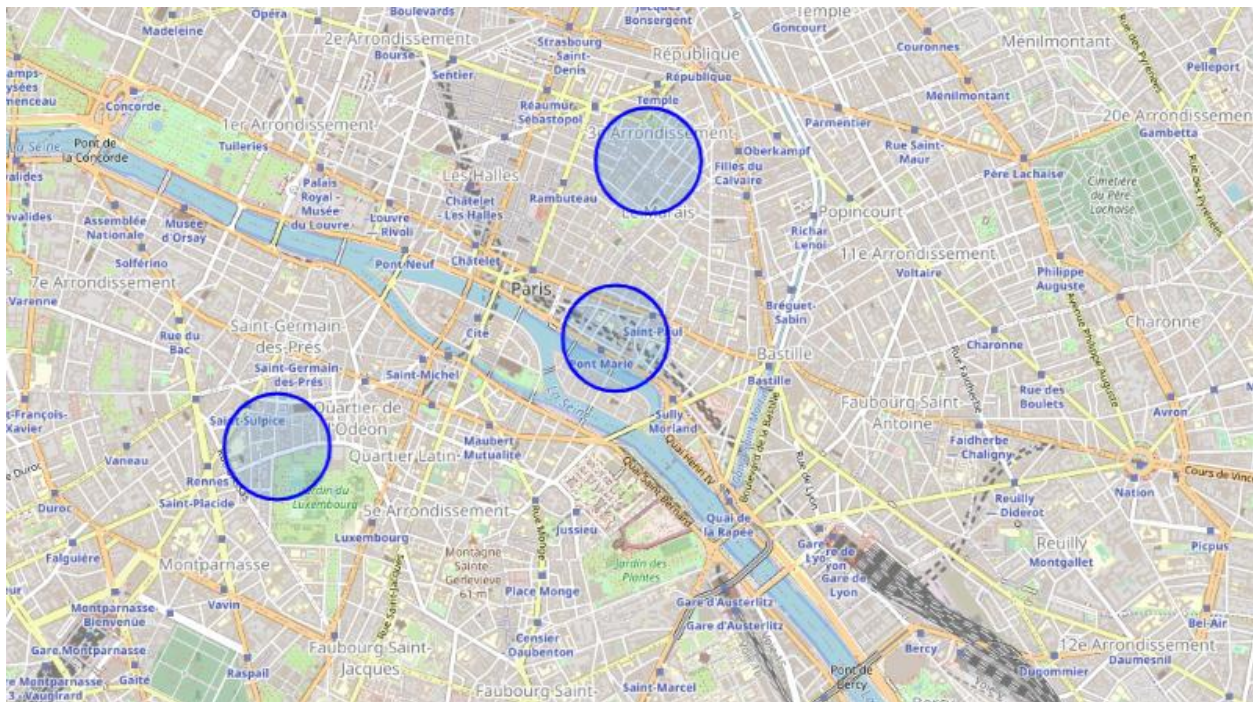
So the final 3 prospective neighborhoods for new store locations are where 4 criteria are met:

- 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



6. 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.

7. 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.

8. 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 focused 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.