

Recommended location for opening a fashion store in Paris

Report

1. Introduction

Fashion has become synonymous with geographic locations, sometimes referred to as fashion capitals. Historically in Europe the earliest centers of fashion were defined through the royal courts with the reigning monarch or high nobility dictating the accepted dress styles and protocols of the day. Status and displays of wealth were central to early expressions of what passed for fashion, while sumptuary laws maintained and reinforced social distinctions.

Fashion design operates as part of a structured, international fashion industry. Industrial and technological advances have directly contributed to increasingly efficient supply chains, with each business fulfilling a price category to serve a fashion retail segment. In the fashion industry, the method of production usually provides an early indicator to the type of business and target market. When conceptualized as a hierarchical model, the fashion industry can be viewed as follows:

1. Couture (haute couture): unique designs made to couture standards from leading names including Chanel and Christian Dior;
2. Designer: top designer ready to wear lines, such as Dries Van Noten and Prada;
3. Bridge or diffusion: more moderately priced lines, such as Marc Jacobs and DKNY;

4. Upper high street: good-quality branded labels and private labels sold through a variety of chains and retail outlet labels, including Karen Millen and Jones New York;
5. Mid-high street: contemporary and private labels with competitively priced fashion-forward merchandise, such as Gap and Next;
6. Lower high street: volume selling merchandise that is moderately priced and might include adaptations of higher priced merchandise categories, such as H&M and New Look;
7. Budget: mass produced goods at very affordable price points, such as Primark and Cherokee.

This model should be considered when selecting the most suitable locations for the new stores as the areas depend on the target audience.

2. Business Problem

A fashion retailer decided to open a new store in Paris, France, and needs to do some research to make data-driven decisions on the new locations that are most suitable for their new store.

The fashion retailer is positioned as the lower high street brand label. As such, they do not seek stores in the premium up-market strips like Avenue Montaigne, but rather, in high traffic areas where consumers go for shopping, restaurants and entertainment. The best locations to open new fashion retail stores may not only be where other clothing stores are grouped. While it is always good to have some other fashion stores in the area, it is strongly suggested that the best places are in fact areas that have an abundance of French Restaurants, Cafés and Wine Bars. Parisians are very social and tend to frequent these places often, so opening new stores in these locations is becoming popular, especially if the stores offer something experiential.

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 neighborhood, or on the ground searching for specific sites by the company's personnel.

3. Data

The data regarding the districts in Paris needs to be researched and a suitable usable source identified. If it is found, but is not in a usable form, data wrangling and cleaning will have to be performed.

The cleansed data will then be used alongside Foursquare geolocation data, which is readily available using the API. 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.

Paris is divided into 20 administrative districts (Arrondissements Municipaux). They are normally referenced by the arrondissement number rather than a name. Data for the arrondissements is necessary to be able to select the most suitable of these areas for new store.

Initially looking to get this data by scraping the relevant [Wikipedia page](#), fortunately, after much research, this data is available on the web from [Open|DATA France](#) as a csv or json file and can be manipulated and cleansed to provide a meaningful dataset to use.

CAR		NAME	NSQAR	CAR.1	CARINSEE	LAR	NSQCO	SURFACE	PERIMETRE	Geometry_X	Geometry_Y
0	3	Temple	750000003	3	3eme Ardt	750001537	1170882828	4519264	48.862872	2.360001	
1	19	Buttes-Chaumont	750000019	19	19eme Ardt	750001537	6792651129	11253182	48.887076	2.384821	
2	14	Observatoire	750000014	14	14eme Ardt	750001537	5614677309	10317483	48.829245	2.326542	
3	10	Entrepot	750000010	10	10eme Ardt	750001537	2891739442	6739375	48.876130	2.360728	
4	12	Reuilly	750000012	12	12eme Ardt	750001537	16314782637	24089666	48.834974	2.421325	
5	16	Passy	750000016	16	16eme Ardt	750001537	16372542129	17416110	48.860392	2.261971	
6	11	Popincourt	750000011	11	11eme Ardt	750001537	3665441552	8282012	48.859059	2.380058	
7	2	Bourse	750000002	2	2eme Ardt	750001537	991153745	4554104	48.868279	2.342803	
8	4	Hotel-de-Ville	750000004	4	4eme Ardt	750001537	1600585632	5420908	48.854341	2.357630	
9	17	Batignolles-Monceau	750000017	17	17eme Ardt	750001537	5668834504	10775580	48.887327	2.306777	
10	18	Buttes-Montmartre	750000018	18	18eme Ardt	750001537	5996051308	9916464	48.892569	2.348161	
11	1	Louvre	750000001	1	1er Ardt	750001537	1824612860	6054937	48.862563	2.336443	
12	5	Panthéon	750000005	5	5eme Ardt	750001537	2539374623	6239195	48.844443	2.350715	
13	7	Palais-Bourbon	750000007	7	7eme Ardt	750001537	4090057185	8099425	48.856174	2.312188	
14	20	Menilmontant	750000020	20	20eme Ardt	750001537	5983446037	10704940	48.863461	2.401188	
15	8	elysee	750000008	8	8eme Ardt	750001537	3880036397	7880533	48.872721	2.312554	
16	9	Opera	750000009	9	9eme Ardt	750001537	2178303275	6471588	48.877164	2.337458	
17	13	Gobelins	750000013	13	13eme Ardt	750001537	7149311091	11546547	48.828388	2.362272	
18	15	Vaugirard	750000015	15	15eme Ardt	750001537	8494994081	13678798	48.840085	2.292826	
19	6	Luxembourg	750000006	6	6eme Ardt	750001537	2153095586	6483687	48.849130	2.332898	

The data was imported into the notebook from the source, but it is not in the right structure for our purposes. So some data exploring, wrangling and cleaning is needed. After renaming and dropping unnecessary columns, the dataframe was in a structure that could be used for the analysis.

Arrondissement_Num		Neighborhood	French_Name	Latitude	Longitude
0	3	Temple	3eme Ardt	48.862872	2.360001
1	19	Buttes-Chaumont	19eme Ardt	48.887076	2.384821
2	14	Observatoire	14eme Ardt	48.829245	2.326542
3	10	Entrepot	10eme Ardt	48.876130	2.360728
4	12	Reuilly	12eme Ardt	48.834974	2.421325
5	16	Passy	16eme Ardt	48.860392	2.261971
6	11	Popincourt	11eme Ardt	48.859059	2.380058
7	2	Bourse	2eme Ardt	48.868279	2.342803
8	4	Hotel-de-Ville	4eme Ardt	48.854341	2.357630
9	17	Batignolles-Monceau	17eme Ardt	48.887327	2.306777
10	18	Buttes-Montmartre	18eme Ardt	48.892569	2.348161
11	1	Louvre	1er Ardt	48.862563	2.336443
12	5	Panthéon	5eme Ardt	48.844443	2.350715
13	7	Palais-Bourbon	7eme Ardt	48.856174	2.312188
14	20	Menilmontant	20eme Ardt	48.863461	2.401188
15	8	elysee	8eme Ardt	48.872721	2.312554
16	9	Opera	9eme Ardt	48.877164	2.337458
17	13	Gobelins	13eme Ardt	48.828388	2.362272
18	15	Vaugirard	15eme Ardt	48.840085	2.292826
19	6	Luxembourg	6eme Ardt	48.849130	2.332898

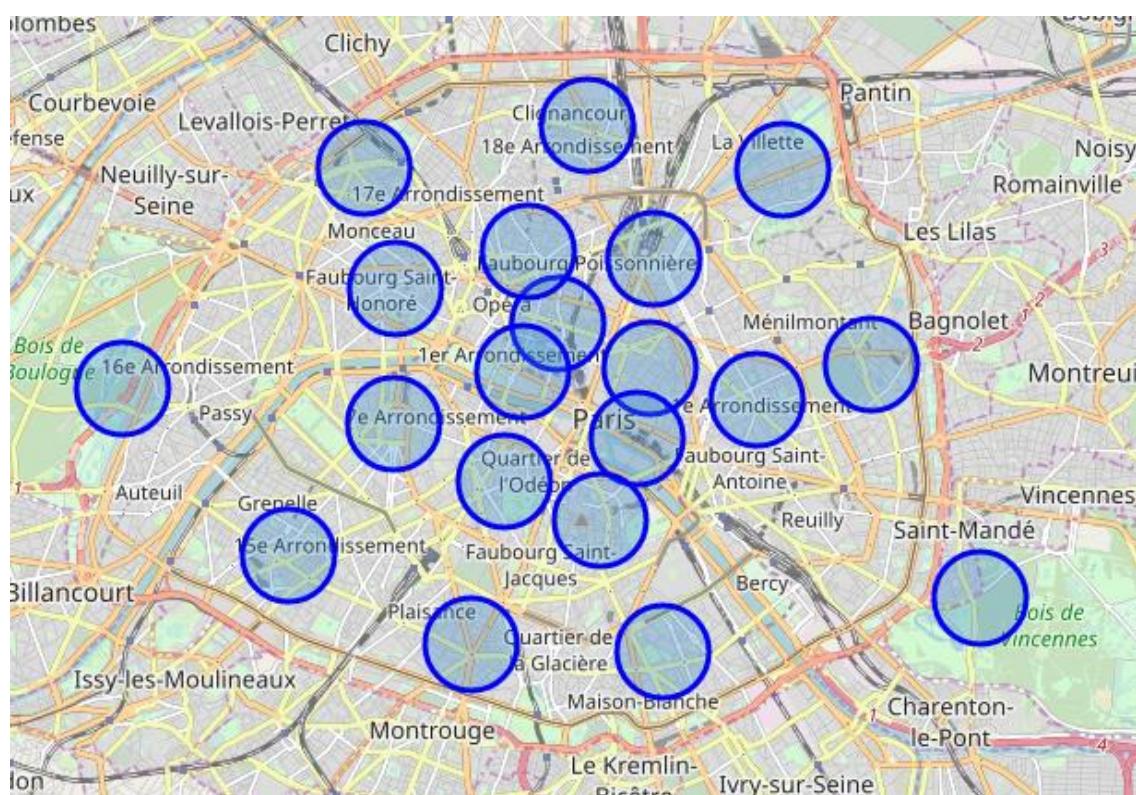
So we now have located and imported the relevant data for the districts of Paris, and have constructed a dataframe.

Next up, we will leverage Foursquare location data to obtain data on high traffic areas - where consumers go for shopping, restaurants and entertainment - in all of the 20 districts. And after this, the decision on the best location for opening a fashion store will be made.

4. Methodology

The GeoPy library for Python is used to geocode location data and get location coordinates of Paris: 48.8566969, 2.3514616.

With the geospatial data, using Folium – which is a powerful data visualization Python library – a map of Paris was created with the location of the 20 neighborhoods superimposed. When viewed within the notebook, the map can be zoomed and hover labels are included to display the name of the neighborhoods.



The neighborhoods are located in a circle around central Paris. The amount of neighborhoods is quite large, so our task is to reduce the amount and target only a few of the neighborhoods.

After setting up the Foursquare API, we can explore the geolocation data. Exploratory data analysis allows us to look at what we are dealing with, and in this case, the first district in our dataframe is explored to become familiar with the data (using the French descriptive arrondissement name).

We determine that the first neighborhood is 3eme Ardt and we were able to easily retrieve its latitude and longitude: 48.86287238, 2.3600009859999997.

Passing the search parameters and location details for this neighborhood into the API, we're able to 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	Marché des Enfants Rouges	Farmers Market	48.862806	2.361996
2	Square du Temple	Park	48.864475	2.360816
3	Chez Alain Miam Miam	Sandwich Place	48.862369	2.361950
4	Chez Alain Miam Miam	Sandwich Place	48.862781	2.362064
5	Hôtel Jules & Jim	Hotel	48.863496	2.357395
6	Fromagerie Jouannault	Cheese Shop	48.862947	2.362530
7	Les Enfants Rouges	Wine Bar	48.863013	2.361260
8	Gramme	Restaurant	48.863121	2.360375
9	Strada Café	Café	48.862224	2.357379
10	Bontemps	Dessert Shop	48.863956	2.360725
11	Paris New York	Burger Joint	48.863843	2.362661
12	SoMa	Japanese Restaurant	48.861511	2.362146
13	Les Enfants Du Marché	French Restaurant	48.862746	2.361950
14	Traiteur Marocain	Moroccan Restaurant	48.862923	2.361933
15	Hank Burger	Burger Joint	48.861340	2.358304
16	The Broken Arm	Boutique	48.864653	2.361433
17	Café Charlot	Bistro	48.862940	2.362580
18	Biglove Caffè	Italian Restaurant	48.862063	2.363557
19	Little Red Door	Speakeasy	48.863703	2.363514

So now we did the same for all of the areas – creating a nearby venues function for all the neighborhoods in Paris. And we’re able to generate a new dataframe with all of the nearby venues for all of the neighborhoods.

	French_Name	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	3eme Ardt	48.862872	2.360001	Mmmozza	48.863910	2.360591	Sandwich Place
1	3eme Ardt	48.862872	2.360001	Marché des Enfants Rouges	48.862806	2.361996	Farmers Market
2	3eme Ardt	48.862872	2.360001	Square du Temple	48.864475	2.360816	Park
3	3eme Ardt	48.862872	2.360001	Chez Alain Miam Miam	48.862369	2.361950	Sandwich Place
4	3eme Ardt	48.862872	2.360001	Chez Alain Miam Miam	48.862781	2.362064	Sandwich Place
5	3eme Ardt	48.862872	2.360001	Hôtel Jules & Jim	48.863496	2.357395	Hotel
6	3eme Ardt	48.862872	2.360001	Fromagerie Jouannault	48.862947	2.362530	Cheese Shop
7	3eme Ardt	48.862872	2.360001	Les Enfants Rouges	48.863013	2.361260	Wine Bar
8	3eme Ardt	48.862872	2.360001	Gramme	48.863121	2.360375	Restaurant
9	3eme Ardt	48.862872	2.360001	Strada Café	48.862224	2.357379	Café
10	3eme Ardt	48.862872	2.360001	Bontemps	48.863956	2.360725	Dessert Shop
...
245	10eme Ardt	48.876130	2.360728	Pomm'	48.877768	2.365382	Diner
246	10eme Ardt	48.876130	2.360728	Sheezan	48.873783	2.355220	Indian Restaurant
247	10eme Ardt	48.876130	2.360728	Giacomo	48.879784	2.362198	Italian Restaurant
248	10eme Ardt	48.876130	2.360728	East Side Café	48.876008	2.359309	Lounge
249	10eme Ardt	48.876130	2.360728	Les Caves Saint Martin	48.879651	2.363401	Wine Shop

With the new dataframe it is possible to check how many venues were returned for each neighborhood, and also calculate how many unique venue categories there are. This is a useful statistic in itself.

French_Name	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
10eme Ardt	100	100	100	100	100	100
11eme Ardt	67	67	67	67	67	67
12eme Ardt	6	6	6	6	6	6
13eme Ardt	60	60	60	60	60	60
14eme Ardt	29	29	29	29	29	29
15eme Ardt	64	64	64	64	64	64
16eme Ardt	12	12	12	12	12	12
17eme Ardt	60	60	60	60	60	60
18eme Ardt	45	45	45	45	45	45
19eme Ardt	43	43	43	43	43	43
1er Ardt	100	100	100	100	100	100
20eme Ardt	48	48	48	48	48	48
2eme Ardt	100	100	100	100	100	100
3eme Ardt	100	100	100	100	100	100
4eme Ardt	100	100	100	100	100	100
5eme Ardt	94	94	94	94	94	94
6eme Ardt	73	73	73	73	73	73
7eme Ardt	100	100	100	100	100	100
8eme Ardt	65	65	65	65	65	65
9eme Ardt	100	100	100	100	100	100

Now it is possible to analyze each of the neighborhoods, using the results, displaying how many venues of each category were in each neighborhood.

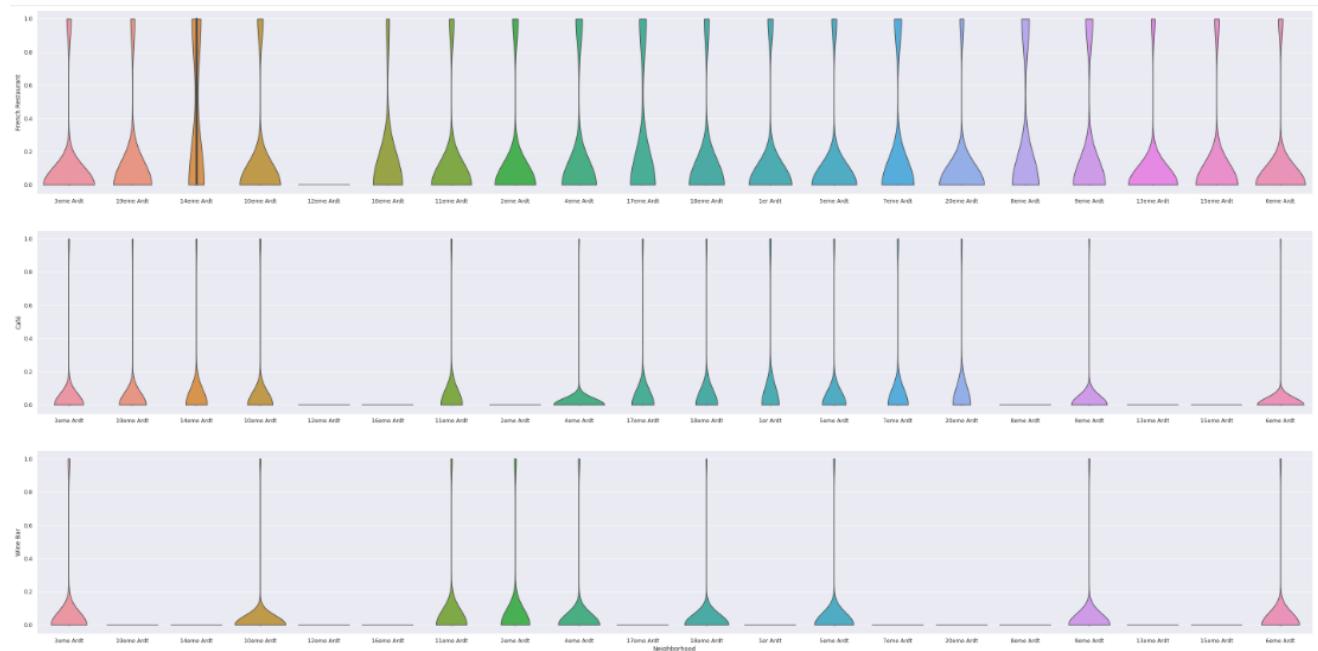
Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar	Wine Shop	Women's Store	Yoga Studio	Zoo	Zoo Exhibit
0	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
6	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
7	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	1	0	0	0	0	0
8	3eme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...	
1359	Geme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1360	Geme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1361	Geme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1362	Geme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1363	Geme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1364	Geme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1365	Geme Ardt	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

Then, an important part of the process was the frequency distribution of the categories – grouping rows by neighborhood and taking the mean of the frequency of occurrence of each category. This will be used in narrowing down the suitable neighborhoods for new stores.

Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar	Wine Shop	Women's Store	Yoga Studio	Zoo	Zoo Exhibit
0	10eme Ardt	0.000000	0.020000	0.000000	0.00	0.00	0.000000	0.000000	0.020000	...	0.010000	0.00	0.000000	0.010000	0.02	0.000000	0.00	0.000000	0.000000	
1	11eme Ardt	0.014625	0.000000	0.000000	0.00	0.00	0.000000	0.014925	0.000000	0.014925	...	0.014925	0.00	0.014925	0.014625	0.044776	0.00	0.014925	0.00	0.000000
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	0.000000	0.000000	0.00	0.000000	0.00	0.166667	0.166667
3	13eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.183333	...	0.000000	0.00	0.000000	0.216667	0.000000	0.00	0.000000	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	...	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000
5	15eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.016625	0.016625	...	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.016625	0.00	0.000000
6	16eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.083333	0.000000	0.000000	...	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000
7	17eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.016667	0.000000	0.016667	...	0.000000	0.00	0.000000	0.000000	0.000000	0.00	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	...	0.000000	0.00	0.000000	0.022222	0.022222	0.00	0.000000	0.00	0.000000
9	19eme Ardt	0.000000	0.023256	0.023256	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0.023256	0.000000	0.00	0.000000	0.00	0.000000
10	1er Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.030000	0.000000	0.000000	...	0.000000	0.00	0.000000	0.000000	0.020000	0.01	0.000000	0.00	0.000000
11	20eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0.020833	0.00	0.000000	0.00	0.000000	0.000000
12	2eme Ardt	0.000000	0.000000	0.010000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0.010000	0.050000	0.01	0.020000	0.00	0.000000
13	3eme Ardt	0.000000	0.000000	0.000000	0.00	0.01	0.020000	0.010000	0.010000	0.010000	...	0.010000	0.00	0.000000	0.020000	0.030000	0.00	0.000000	0.01	0.000000
14	4eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.020000	0.010000	0.010000	0.010000	...	0.000000	0.00	0.000000	0.000000	0.030000	0.00	0.000000	0.00	0.000000
15	5eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.010638	0.010638	...	0.000000	0.00	0.000000	0.021277	0.031915	0.00	0.000000	0.00	0.000000
16	6eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.013699	0.000000	0.000000	...	0.000000	0.00	0.000000	0.000000	0.013699	0.00	0.000000	0.00	0.000000
17	7eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.030769	0.015385	0.000000	0.000000	...	0.010000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000
18	8eme Ardt	0.000000	0.000000	0.000000	0.00	0.00	0.030769	0.015385	0.000000	0.000000	...	0.000000	0.00	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000
19	9eme Ardt	0.000000	0.010000	0.000000	0.01	0.00	0.000000	0.000000	0.000000	0.000000	...	0.020000	0.01	0.000000	0.010000	0.030000	0.00	0.000000	0.00	0.000000

Referring back to the original task, the business types criteria are ‘French Restaurants’, ‘Cafés’ and ‘Wine Bars’. These are the venue types that should be an abundant density of in the ideal store locations.

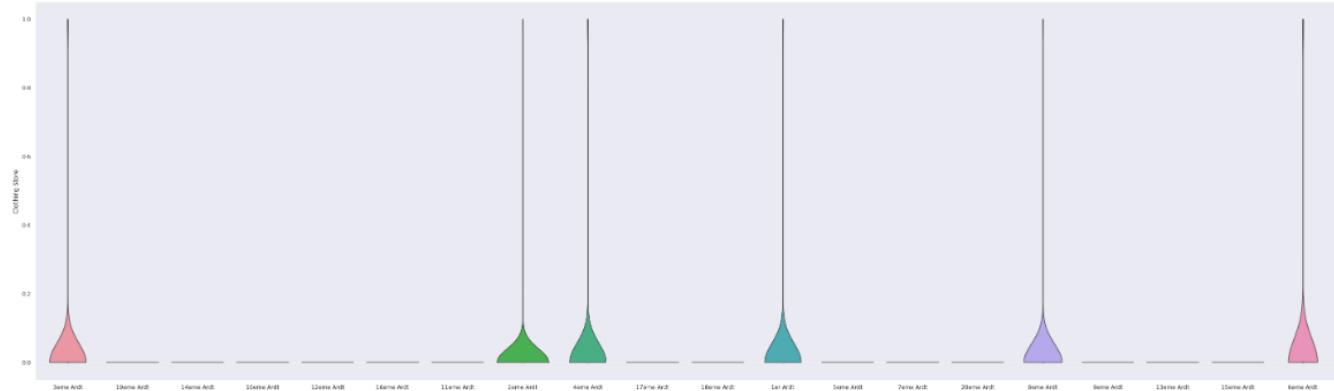
Let's look at their frequency of occurrence for all the Paris neighborhoods, isolating the categorical venues, using a violin plot from the Seaborn library for Python – it's a great way to visualize frequency distribution datasets, they display a density estimation of the underlying distribution.



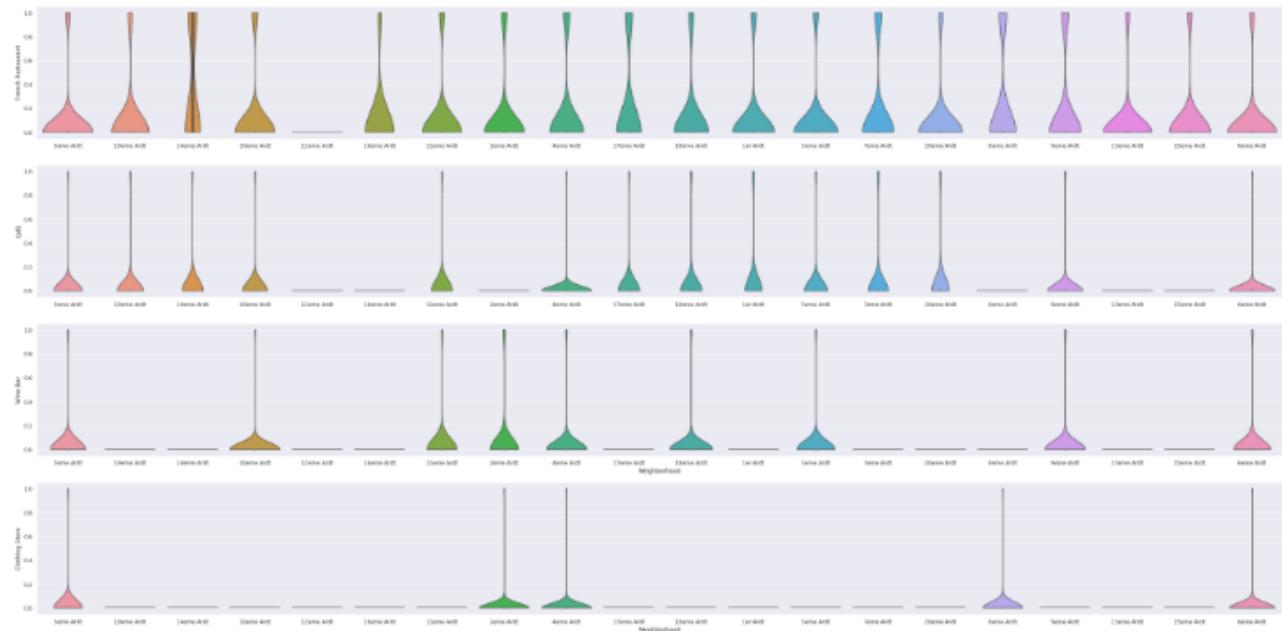
So as we can see from the analysis, there are 8 neighborhoods to open new stores, according to the criteria that there are 3 specified venues in a great frequency (French Restaurants, Cafés and Wine Bars). They are as follows:

1. 3eme Ardt
2. 10eme Ardt
3. 11eme Ardt
4. 4eme Ardt
5. 18eme Ardt
6. 5eme Ardt
7. 9eme Ardt
8. 6eme Ardt

If we include 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.



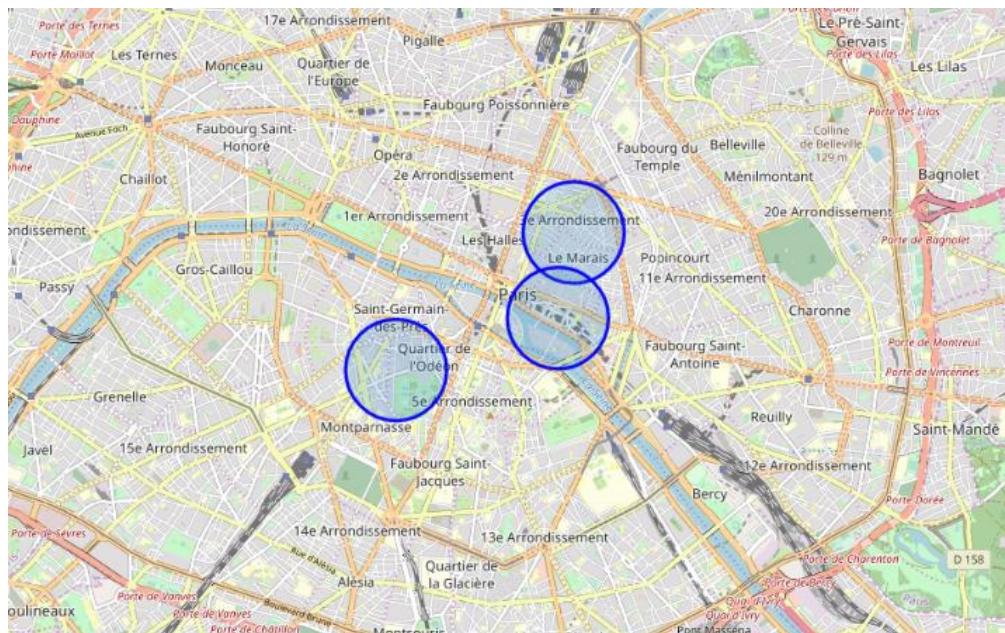
There are 5 neighborhoods that have a significant frequency density of clothing stores. Let's add this category to the analysis, with the other 3 specified categories, and visually represent the results as seen below.



So through this further analysis we're able to narrow it down to just 3 neighborhoods:

	Arrondissement_Num	Neighborhood	French_Name	Latitude	Longitude
0	3	Temple	3eme Ardt	48.862872	2.360001
1	4	Hotel-de-Ville	4eme Ardt	48.854341	2.357630
2	6	Luxembourg	6eme Ardt	48.849130	2.332898

Visualized on a Folium map of Paris, the neighborhoods are all in fact quite close to the city center.



5. Results

Inferential analysis using the data, as well as domain knowledge of retail and marketing, allowed narrowing the search area down significantly from 20 potential districts to 3 that should suit the retail business. Adding Clothing Stores to the main 3 criteria – Restaurants, Cafés and Wine Bars – had a significant impact, as having some stores of the same category in the same area, especially in fashion retail, is very desirable for 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.

The final 3 prospective neighborhoods for new store locations are:

1. 3eme Ardt: Arrondissement 3, Temple
2. 4eme Ardt: Arrondissement 4, Hotel-de-Ville
3. 6eme Ardt: Arrondissement 6, Luxembourg

6. Discussion

The main goal of data analysis here is to steer a course for the location selection of new store:

- to meet the criteria of being in the neighborhoods with lots of abundant leisure venues;
- to narrow the search down to just a few of the main areas that are best suited to match the criteria.

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.

7. Conclusion

There are many ways this analysis could have been performed based on different methodology and perhaps different data sources. Violin plots were selected as it was a straight forward way to narrow down the options, meeting the criteria for the surrounding venues and domain knowledge.

Without leveraging data, the process could have been drawn out and resulted in new store being opened in sub-standard areas. Data has helped to provide a better strategy and way forward, these data-driven decisions will lead to a better solution in the end.