Suitable New Store Locations in Paris for a Fashion Retailer

Deepashri Sane

May 1, 2020

1. Introduction

1.1 Background

Paris is regarded as the world fashion capital, and spread throughout the city are many fashion boutiques. Most of the major French fashion brands, such as Chanel, Louis Vuitton, Dior, and Lacroix, are currently headquartered here. Numerous international fashion labels also operate shops in Paris, such as Valentino, Gucci, Loewe, Escada, Bottega Veneta, and Burberry, as well as an Abercrombie & Fitch flagship store which has become a main consumer attraction.

Paris hosts a fashion week twice a year, similar to other international centers such as Milan, London, Tokyo, New York, Los Angeles and Rome.

Paris acts as the center of the fashion industry and holds the name of global fashion capital. The city is home to many prime designers, including Chanel, LouisVuitton, Givenchy, Balmain, Louboutin, Pierre Cardin, Yves Saint Laurent, Roger Vivier, Thierry Mugler, Dior, Jean Paul Gaultier, Hermès, Lanvin, Chloé, Rochas, and Céline.

The globalization of textile and garment manufacturing is changing the economics of the entire fashion system, but the couture, which really exists only in Paris, retains its prestige and helps to drive an array of luxury goods from perfume to handbags and ready-to-wear lines. Even though another city might become paramount during some seasons, Paris remains generally acknowledged as the most important fashion city.

1.2 Business Objective

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.

So the task is to assist 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.

1.3 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.

1.4 Problem

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.

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. Data Acquisition and Cleaning

2.1 Data Source

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.

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.

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table. After Combining same some of the columns are Renamed like NAME to Neighborhood , CAR to Arrondissement_Num , Geometry_X to Latitude , Geometry_Y to Longitude , LAR to French_Name.

Then Cleaning of the dataset was need in order to remove unnecessary columns like NSQAR, CAR.1, CARINSEE, NSQCO, SURFACE, PERIMETRE as they were not require for further processing So, After cleaning data 100 samples were present. Out of 11 only 5 parameter were selected which useful while preparing data for location. Selected features are shown in below Figure 1 after cleaning.

Figure 1: Selected Features

	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

3. Exploratory Data Analysis

3.1 Explore the districts

Exploratory Data Analysis is an approach to analyzing data. It’s often the first step in data analysis, implemented before any formal statistical techniques are applied. So,Exploring then to become familiar with the data , by Getting the Neighborhood's name ,latitude and longitude values.

Data visualization using various mapping libraries to create a map of Paris with districts superimposed using latitude and longitude values. The map is shown below in Figure 2.

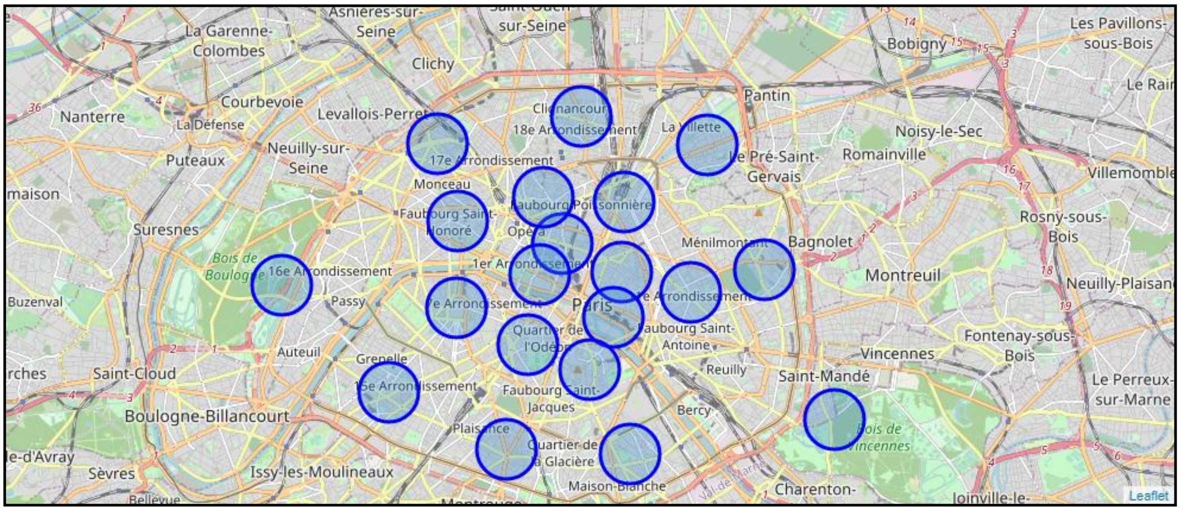

Figure 2: Map of Paris with districts

The first district in dataframe was identified as 3eme Ardt. After indentifying the longitude and latitude for 3eme Ardt. I have explore top 100 venues that are in the neighborhood *3eme Ardt* within a radius of 500 meters. So, Foursquare location data was been leveraged to explore or compare districts around Paris. Example of explored venue and there categories are shown as in below figure 3

	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

Figure 3: Venue and there categories

There was need to Create a nearby venues function for all the neighbourhoods in Paris to create a new dataframe for with merged columns with location which resulted into 250x7 samples which is as shown in figure 4

	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	Square du Temple	48.864475	2.360816	Park
2	3eme Ardt	48.862872	2.360001	Marché des Enfants Rouges	48.862806	2.361996	Farmers Market
3	3eme Ardt	48.862872	2.360001	Chez Alain Miam Miam	48.862781	2.362064	Sandwich Place
4	3eme Ardt	48.862872	2.360001	Chez Alain Miam Miam	48.862369	2.361950	Sandwich Place
5	3eme Ardt	48.862872	2.360001	Fromagerie Jouannault	48.862947	2.362530	Cheese Shop

Figure 4: Merge table with location

3.2 Calculate how many unique venue category

Applying groupby operation to combine the location with venues which will further use forget the unique value of venue category which . There are 209 unique venue categories. The groupby give the following table as shown below.

	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
French_Name						
10eme Ardt	100	100	100	100	100	100
11eme Ardt	66	66	66	66	66	66
12eme Ardt	4	4	4	4	4	4
13eme Ardt	58	58	58	58	58	58
14eme Ardt	30	30	30	30	30	30
15eme Ardt	63	63	63	63	63	63
16eme Ardt	11	11	11	11	11	11
17eme Ardt	51	51	51	51	51	51
18eme Ardt	50	50	50	50	50	50

After finding the unique venue and categories and there is need to analyze the each of the Neighborhoods. This analysis is done by using one-hot encoding transforms categorical features to a format that works better with classification and regression algorithms. It's very useful in methods where multiple types of data representation is necessary. It represented these categories in one-hot encoding; it would actually replace the rows with columns. Which result into location merged with the nearby venue like restaurant, shop, gallery etc. The following table shows the analysis of Neighbourhoods

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store
0	3eme Ardt	0	0	0	0	0	0	0	0
1	3eme Ardt	0	0	0	0	0	0	0	0
2	3eme Ardt	0	0	0	0	0	0	0	0
3	3eme Ardt	0	0	0	0	0	0	0	0
4	3eme Ardt	0	0	0	0	0	0	0	0
5	3eme Ardt	0	0	0	0	0	0	0	0
6	3eme Ardt	0	0	0	0	0	0	0	0
7	3eme Ardt	0	0	0	0	0	0	0	0
8	3eme Ardt	0	0	0	0	0	0	0	0
9	3eme Ardt	0	0	0	0	0	0	0	0
10	3eme Ardt	0	0	0	0	0	0	0	0

Figure 5: Analysis of Neighbourhoods

Taking the mean of the frequency of occurrence of each category is important as its will give how frequently the venue occur in the Neighbourhoods. This result gives 196 venue which will enable to locate the location which high footprints.

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store
0	10eme Ardt	0.000000	0.02	0.00000	0.00	0.000000	0.000000	0.000000	0.000000
1	11eme Ardt	0.014925	0.00	0.00000	0.00	0.000000	0.000000	0.014925	0.000000
2	12eme Ardt	0.000000	0.00	0.00000	0.00	0.000000	0.000000	0.000000	0.000000
3	13eme Ardt	0.000000	0.00	0.00000	0.00	0.000000	0.000000	0.000000	0.000000
4	14eme Ardt	0.000000	0.00	0.00000	0.00	0.000000	0.000000	0.000000	0.000000
5	15eme Ardt	0.000000	0.00	0.00000	0.00	0.000000	0.000000	0.000000	0.016393
6	16eme Ardt	0.000000	0.00	0.00000	0.00	0.000000	0.000000	0.090909	0.000000
7	17eme Ardt	0.000000	0.00	0.00000	0.00	0.000000	0.000000	0.016949	0.000000
8	18eme Ardt	0.000000	0.00	0.00000	0.00	0.000000	0.000000	0.000000	0.000000
9	19eme Ardt	0.000000	0.00	0.02381	0.00	0.000000	0.000000	0.000000	0.000000
10	1er Ardt	0.000000	0.00	0.00000	0.00	0.000000	0.000000	0.042254	0.000000

Figure 6: Frequency of occurrence

3.3 The top 10 venue categories for each neighbourhood

To find each neighborhood with top 10 most common venues they have to group together on the bases of neighborhoods and the frequency. Then the venues are sorted into descending order. The new dataframe has been created for which has been aligned with neighborhoods and sorted as from 1st most common to 10th most common venue as shown below in figure 7.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	10eme Ardt	French Restaurant	Bistro	Coffee Shop	Café	Hotel
1	11eme Ardt	French Restaurant	Supermarket	Restaurant	Pastry Shop	Wine Bar
2	12eme Ardt	Zoo Exhibit	Supermarket	Monument / Landmark	Zoo	Antique Shop
3	13eme Ardt	Vietnamese Restaurant	Asian Restaurant	French Restaurant	Chinese Restaurant	Thai Restaurant
4	14eme Ardt	French Restaurant	Hotel	Convenience Store	Bakery	Food & Drink Shop
5	15eme Ardt	Italian Restaurant	French Restaurant	Hotel	Coffee Shop	Thai Restaurant

Figure 7: Top 10 most common venues

The business types criteria specified by the client- 'French Restaurants', 'Cafés' and 'Wine Bars'. 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. Checking the frequency of occurrence for all the Paris neighbourhoods, isolating the categorical venues. These are the venue types that the client wants to have an abundant density of in the ideal store locations.
To plot of this data a violin plot is used which is a density estimation of the underlying distribution which is shown figure 8

A violin plot: It is a method of plotting numeric data. It is similar to a box plot, with the addition of a rotated kernel density plot on each side.

Violin plots are similar to box plots, except that they also show the probability density of the data at different values, usually smoothed by a kernel density estimator. A violin plot is more informative than a plain box plot.

While a box plot only shows summary statistics such as mean/median and interquartile ranges, the violin plot shows the full distribution of the data. The difference is particularly useful when the data distribution is multimodal (more than one peak). In this case a violin plot shows the presence of different peaks, their position and relative amplitude.

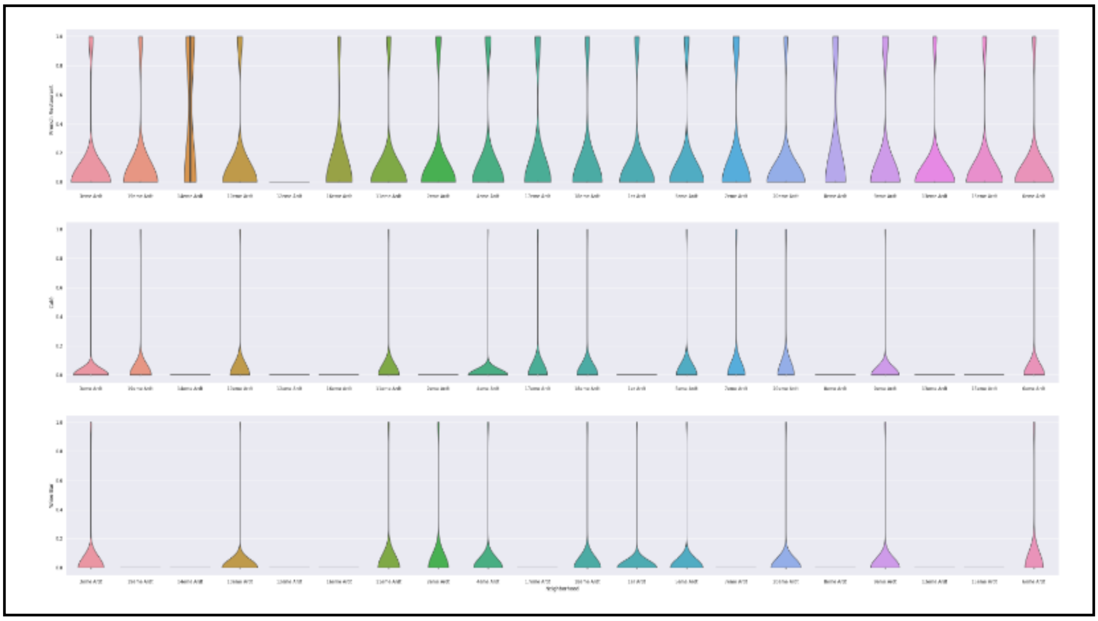

Figure 8: Frequency distribution for the top 3 venue categories for each neighbourhood

4. Inferential Analysis

4.1 The Neighbourhoods

Inferential statistics allows you to make predictions (“inferences”) from that data. With inferential statistics, you take data from samples and make generalizations about a population.

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:

1.3eme Ardt

2.10eme Ardt

3.11eme Ardt

4.4eme Ardt

5.18eme Ardt

6.6eme Ardt

7.5eme Ardt

8.9eme Ardt

Now neighbourhood need to Exploring further for venue category - ''Clothing Store' We have the 8 neighbourhoods that all include the venue category criteria.

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. The figure 9 shows frequency of clothing store

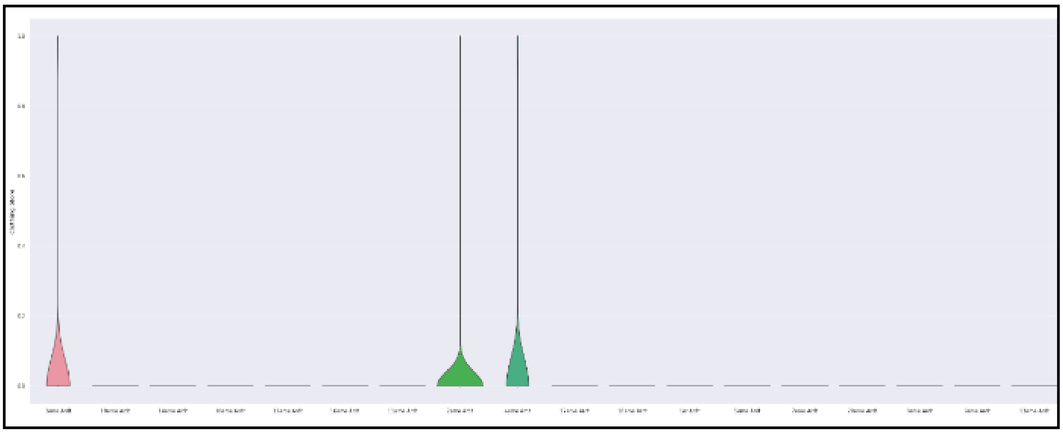

Figure 9: Frequency of Clothing stores for each neighbourhood

So, the Output shows there are 3 neighbourhoods that have a significant frequency density of clothing stores

Now the analysis with the other 3 specified categories need to check as Checking the frequency of occurrence for all the Paris neighbourhoods, isolating the categorical venues. The analysis gives Frequency distribution for the top 3 venue categories for each neighborhood (includes clothing) which show how other venues are occurring with clothing shop.

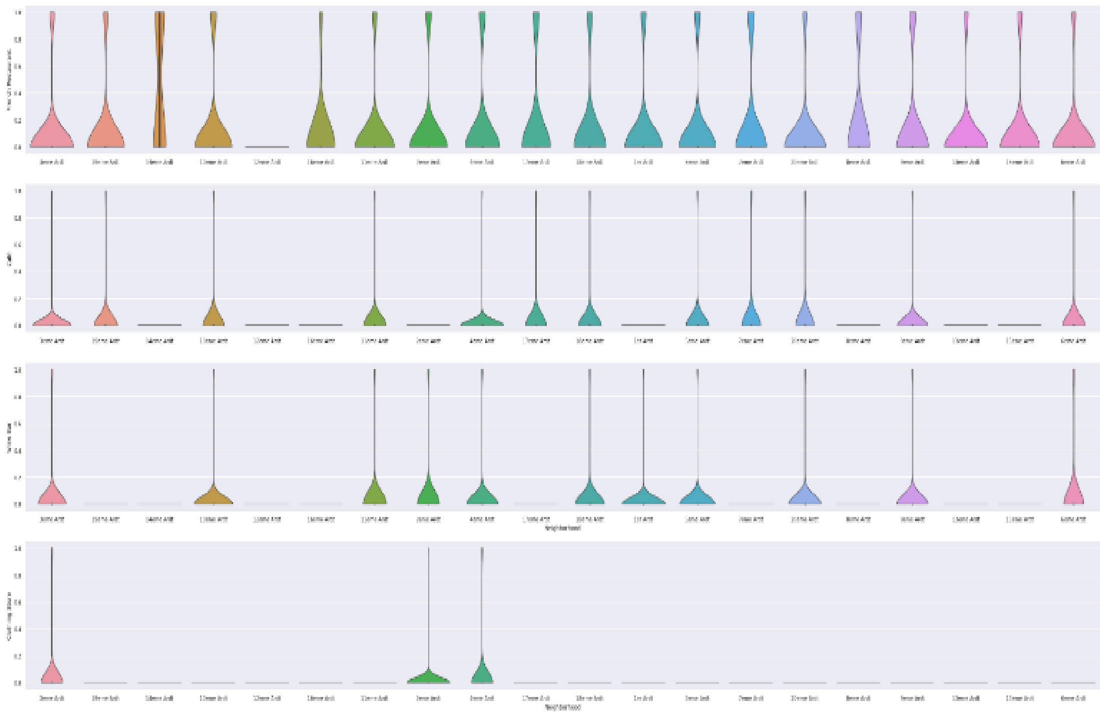

Figure 10: Frequency distribution for the top 3 venue categories for each neighbourhood

5. Result

Above 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 2 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 2 prospective neighbourhoods for new store locations

- 3eme Ardt : Arrondissement 3, Temple
- 4eme Ardt : Arrondissement 4, Hotel-de-Ville

The selected location is show below on map

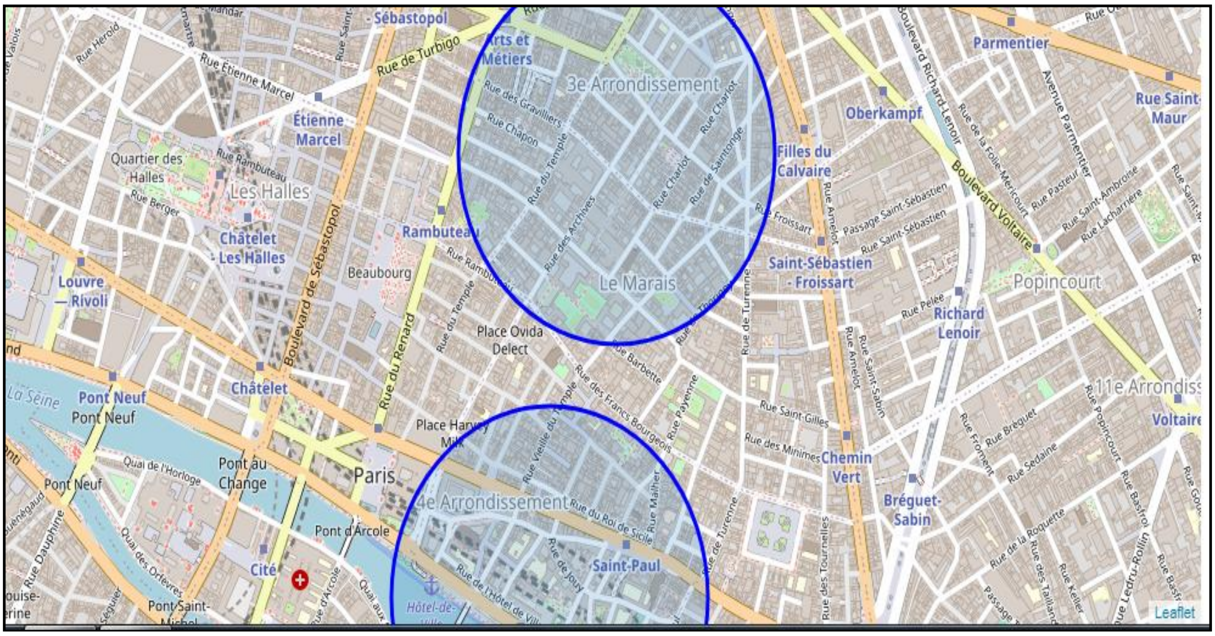

Figure 11: Final location selected

6. Observation

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 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 2 that should suit the client's retail business.

- **Inferences**

  The inferences from the data in making the location recommendations

  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.

7. Conclusion

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.