Capstone Project – Battle of Neighbourhoods

Leverage the FourSquare location data to explore or compare neighbourhoods or cities of your choice

Project Title: Determine best locations to establish new stores for a fashion retailer in Paris

INTRODUCTION

Paris is home to all the top luxury brands and the chicest fashion in the world. Shopping in Paris is the most extraordinary experience to have on a Parisian holiday, especially if you're a Shopaholic. From elegant bags to classy shoes and sassy dresses to gorgeous accessories, you will find everything that'll make you look your best from head to toe. The city does not offer ordinary shops rather a whole world of your dream brands and favourite items and the experience is totally lavish.

So now, a renowned fashion retailer, with a substantial e-commerce footprint, has begun the rollout of their new stores, which begins in Paris. As they are new to Paris, they would need some help in determining which locations would be best to setup their new stores that can eventually attract a lot of people. So, as part of this project, we are going to make data-driven decisions on the new locations/ districts that are more suitable for their new stores in Paris.

Target audience:

This project is going to help any fashion retailer planning to set up their stores in Paris, which is the hub of fashion, explore which location in Paris would be perfect to establish their stores.

Criteria:

This retailer brand is not what can be considered as high-end, but they are positioned in the mid-range of the fashion industry. So they do not seek to establish their stores in premium/ plush localities, but rather, in high traffic areas where consumers go for shopping, visit restaurants or go for entertainment. We will be using Foursquare data to make decisions about the best of the areas.

Observing the data from the other retail stores suggests that the best locations to open new stores may not only be where there are other fashion outlets, but in fact, the best areas are near French restaurants, cafes, wine bars. Since

Parisians are very social in nature and visit such places quite often, opening new stores in such locations is getting popular.

So, the analysis will focus on districts in Paris with such establishments – narrowing down to the best district options, and then further research can be done.

Goal

- Provide recommendations and results based on the data analysis which will help the fashion retailer setup new stores that yield high profits

- Discussion of any limitations and how the results can be used, and any conclusions that can be drawn.

DATA ACQUISITION AND CLEANING

Data Source:

There are 20 main districts in Paris. The data regarding these 20 districts needs to be researched.
Then, the cleansed data will be used alongside the FourSquare data, which is readily available. FourSquare data will be used to explore or compare the districts in Paris, identifying the most populous areas where most of the fashion brands are set up.

Data for the districts should be collected - to select the most suitable of these areas for new stores. Obtained this data from:
https://opendata.paris.fr/explore/dataset/arrondissements/table/?dataChart

Downloaded the .csv file from the website and uploaded it to the GitHub repository.

Data Cleaning

Initially the downloaded data had many invalid characters and numbers in exponential values, which had to be replaced by correct ones.
Secondly, the column names had to be replaced with a more meaningful names for better understanding.
Thirdly, some of the data which were not necessary in the analysis had to be removed.

Exploring, Wrangling and Cleaning the Data ¶

```
# Rename the necessary columns

paris.rename(columns={'NAME': 'Neighborhood ', 'CAR': 'Dist_Num', 'Geometry_X ': 'Latitude', 'Geometry_Y': 'Longitude',  'LAR
paris
```

	Dist_Num	Neighborhood	NSQAR	CARINSEE	French_Name	NSQCO	SURFACE	PERIMETRE	Latitude	Longitude
0	3	Temple	750000003	75103	3eme Ardt	750001537	1.170883e+06	4519.263648	48.862872	2.360001
1	2	Bourse	750000002	75102	2eme Ardt	750001537	9.911537e+05	4554.104360	48.868279	2.342803
2	17	Batignolles-Monceau	750000017	75117	17eme Ardt	750001537	5.668835e+06	10775.579520	48.887327	2.306777
3	7	Palais-Bourbon	750000007	75107	7eme Ardt	750001537	4.090057e+06	8099.424883	48.856174	2.312188
4	4	Hotel-de-Ville	750000004	75104	4eme Ardt	750001537	1.600586e+06	5420.908434	48.854341	2.357630

```
# Clean up the dataset to remove unnecessary columns

paris.drop(['NSQAR','CARINSEE','NSQCO','SURFACE', 'PERIMETRE' ], axis=1, inplace=True)
paris
```

	Dist_Num	Neighborhood	French_Name	Latitude	Longitude
0	3	Temple	3eme Ardt	48.862872	2.360001
1	2	Bourse	2eme Ardt	48.868279	2.342803
2	17	Batignolles-Monceau	17eme Ardt	48.887327	2.306777
3	7	Palais-Bourbon	7eme Ardt	48.856174	2.312188
4	4	Hotel-de-Ville	4eme Ardt	48.854341	2.357630
5	8	elysee	8eme Ardt	48.872721	2.312554
6	18	Buttes-Montmartre	18eme Ardt	48.892569	2.348161

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

Our business objective, strategy and methods to achieve our goal have been laid out.

Next, we will leverage Foursquare location data to obtain data on high traffic areas - where consumers go for shopping, restaurants and entertainment - in all the 20 districts.

Methodology and Exploratory Data Analysis

We will use Foursquare location data to explore or compare all the 20 districts around Paris. And then use data manipulation and analysis to derive subsets of the initial data.
We will identify the high traffic areas using data visualisation and statistical analysis.

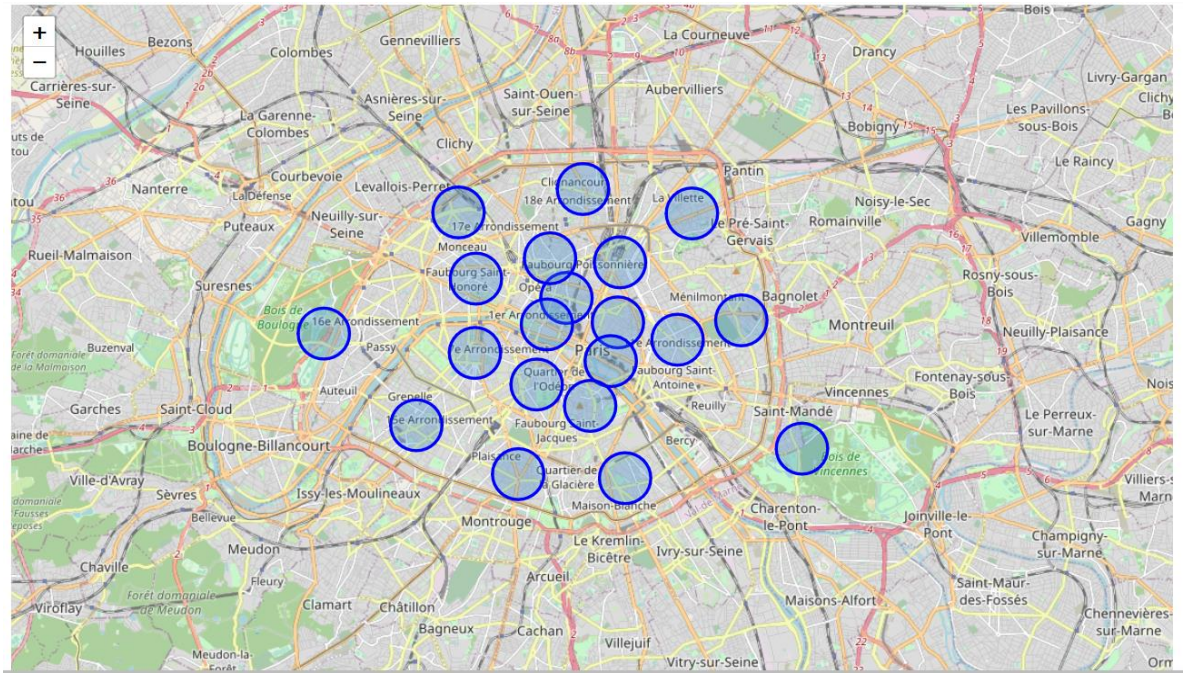
Visualization:
Analysis and plotting visualizations
Data visualization using various mapping libraries

After the cleaning the data, we now have 5 columns for each of the 20 districts

```
# Check the shape of the dataframe
paris.shape
```

```
(20, 5)
```

Let's create a map of Paris with all 20 districts superimposed



We will use the Foursquare API to explore the Districts of Paris (Neighbourhoods)

Let's begin with exploring the first district - `3eme Ardt`, by getting the top 100 venues that are in this neighbourhood within a radius of 500mt.

	name	categories	lat	lng
0	Mmmozza	Sandwich Place	48.863910	2.360591
1	Chez Alain Miam Miam	Sandwich Place	48.862369	2.361950
2	Fromagerie Jouannault	Cheese Shop	48.862947	2.362530
3	Square du Temple	Park	48.864475	2.360816
4	Marché des Enfants Rouges	Farmers Market	48.862806	2.361996
5	Chez Alain Miam Miam	Sandwich Place	48.862781	2.362064
6	Okomusu	Okonomiyaki Restaurant	48.861453	2.360879
7	Le Burger Fermier des Enfants Rouges	Burger Joint	48.862831	2.362073
8	Hôtel Jules & Jim	Hotel	48.863496	2.357395
9	SoMa	Japanese Restaurant	48.861511	2.362146
10	Les Enfants Rouges	Wine Bar	48.863013	2.361260
11	Bontemps	Dessert Shop	48.863956	2.360725
12	Chez Taeko	Japanese Restaurant	48.862734	2.362136
13	Bar de l'Hôtel Jules et Jim	Hotel Bar	48.863463	2.357393
14	Paris New York	Burger Joint	48.863843	2.362661

```
print('There are {} venues in 3eme Ardt returned by Foursquare'.format(nearby_venues.shape[0]))
```

```
There are 83 venues in 3eme Ardt returned by Foursquare
```

Similarly, we will check how many venues are available for the rest of the neighbourhoods

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	4	4	4	4	4	4
13eme Ardt	62	62	62	62	62	62
14eme Ardt	25	25	25	25	25	25
15eme Ardt	61	61	61	61	61	61
16eme Ardt	12	12	12	12	12	12
17eme Ardt	56	56	56	56	56	56
18eme Ardt	40	40	40	40	40	40
19eme Ardt	42	42	42	42	42	42
1er Ardt	70	70	70	70	70	70
20eme Ardt	46	46	46	46	46	46
2eme Ardt	100	100	100	100	100	100
3eme Ardt	83	83	83	83	83	83
4eme Ardt	100	100	100	100	100	100
5eme Ardt	88	88	88	88	88	88
6eme Ardt	39	39	39	39	39	39
7eme Ardt	100	100	100	100	100	100
8eme Ardt	33	33	33	33	33	33
9eme Ardt	100	100	100	100	100	100

Let’s now analyse each neighbourhood (one hot encoding) and see the most common top 10 venues for each of them.

This is table here, is a very useful result, that can provide important details about the venues for all the districts, at a glance. Even once any conclusions are drawn further into the data workflow, we can refer to this table for

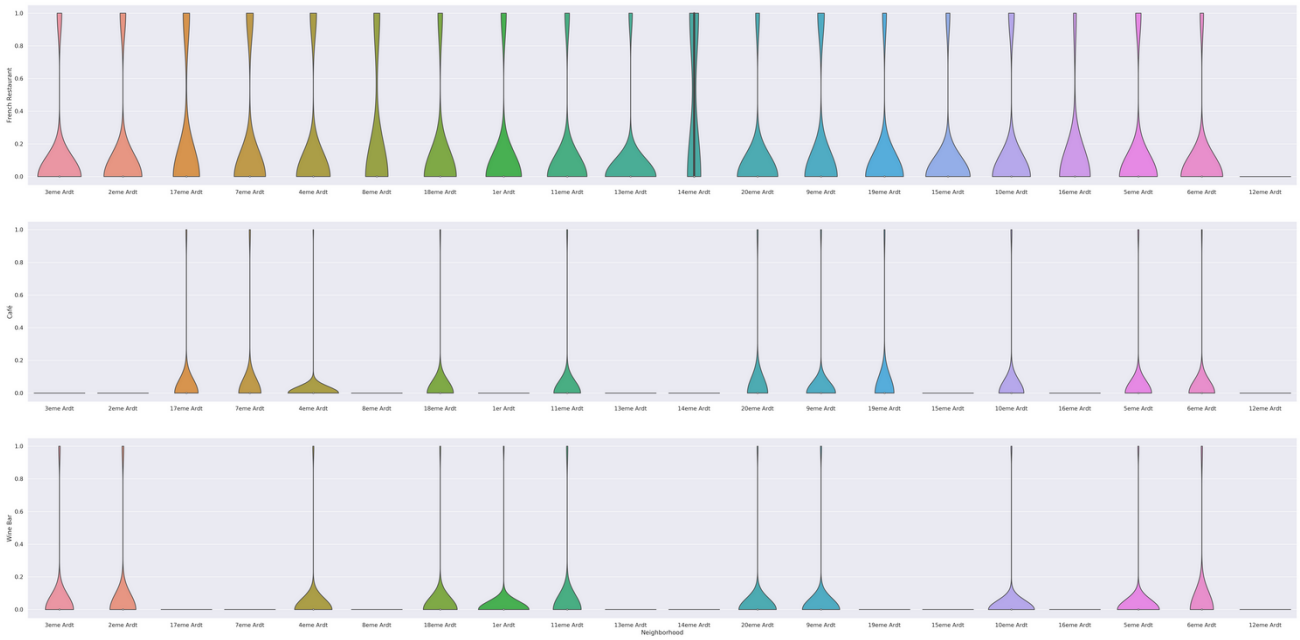
meaningful insights about the top categories of businesses in all the neighbourhoods.

	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	Bistro	Coffee Shop	Indian Restaurant	Café	Hotel	Japanese Restaurant	Pizza Place	Mediterranean Restaurant	Italian Restaurant
1	11eme Ardt	French Restaurant	Restaurant	Supermarket	Wine Bar	Pastry Shop	Japanese Restaurant	Pizza Place	Cocktail Bar	Theater	Bar
2	12eme Ardt	Zoo Exhibit	Supermarket	Monument / Landmark	Zoo	Antique Shop	Diner	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant
3	13eme Ardt	Vietnamese Restaurant	Asian Restaurant	Chinese Restaurant	Thai Restaurant	French Restaurant	Juice Bar	Coffee Shop	Food & Drink Shop	Bookstore	Cosmetics Shop
4	14eme Ardt	French Restaurant	Hotel	Bakery	Brasserie	Food & Drink Shop	Pizza Place	Bistro	Supermarket	Sushi Restaurant	Tea Room
5	15eme Ardt	Italian Restaurant	Hotel	French Restaurant	Coffee Shop	Bakery	Japanese Restaurant	Indian Restaurant	Thai Restaurant	Brasserie	Supermarket
6	16eme Ardt	Plaza	Park	Lake	Pool	French Restaurant	Art Museum	Boat or Ferry	Bus Station	Bus Stop	Diner
7	17eme Ardt	French Restaurant	Hotel	Italian Restaurant	Bakery	Japanese Restaurant	Plaza	Bistro	Restaurant	Café	Burger Joint
8	18eme Ardt	Bar	French Restaurant	Convenience Store	Coffee Shop	Restaurant	Seafood Restaurant	Supermarket	Bistro	Fast Food Restaurant	Middle Eastern Restaurant
9	19eme Ardt	Bar	French Restaurant	Beer Bar	Bistro	Supermarket	Seafood Restaurant	Café	Hotel	Concert Hall	Creperie
10	1er Ardt	French Restaurant	Japanese Restaurant	Plaza	Italian Restaurant	Hotel	Art Museum	Cheese Shop	Garden	Thai Restaurant	Candy Store
11	20eme Ardt	French Restaurant	Bakery	Italian Restaurant	Japanese Restaurant	Plaza	Bar	Bistro	Café	Restaurant	Pizza Place
12	2eme Ardt	French Restaurant	Cocktail Bar	Wine Bar	Hotel	Bakery	Italian Restaurant	Creperie	Bistro	Coffee Shop	Thai Restaurant
13	3eme Ardt	French Restaurant	Japanese Restaurant	Art Gallery	Wine Bar	Gourmet Shop	Coffee Shop	Italian Restaurant	Clothing Store	Sandwich Place	Cocktail Bar
14	4eme Ardt	French Restaurant	Ice Cream Shop	Hotel	Clothing Store	Garden	Italian Restaurant	Plaza	Wine Bar	Pedestrian Plaza	Creperie
15	5eme Ardt	French Restaurant	Italian Restaurant	Bakery	Hotel	Pub	Café	Plaza	Museum	Bar	Coffee Shop
16	6eme Ardt	Plaza	French Restaurant	Wine Bar	Dessert Shop	Fountain	Bookstore	Café	Supermarket	Playground	Bistro
17	7eme Ardt	Hotel	French Restaurant	Italian Restaurant	Café	Plaza	History Museum	Bistro	Cocktail Bar	Historic Site	Gourmet Shop
18	8eme Ardt	French Restaurant	Hotel	Corsican Restaurant	Art Gallery	Spa	Thai Restaurant	Park	Cocktail Bar	Resort	Modern European Restaurant
19	9eme Ardt	French Restaurant	Hotel	Cocktail Bar	Bakery	Bistro	Wine Bar	Café	Lounge	Japanese Restaurant	Turkish Restaurant

The business criteria specified by the client to setup stores where there are *'French Restaurants', 'Cafés' and 'Wine Bars'*

Let's look at their 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. The violin plot from the seaborn library is used here as it is a great way to visualise frequency distribution datasets, they display a density estimation of the underlying distribution.

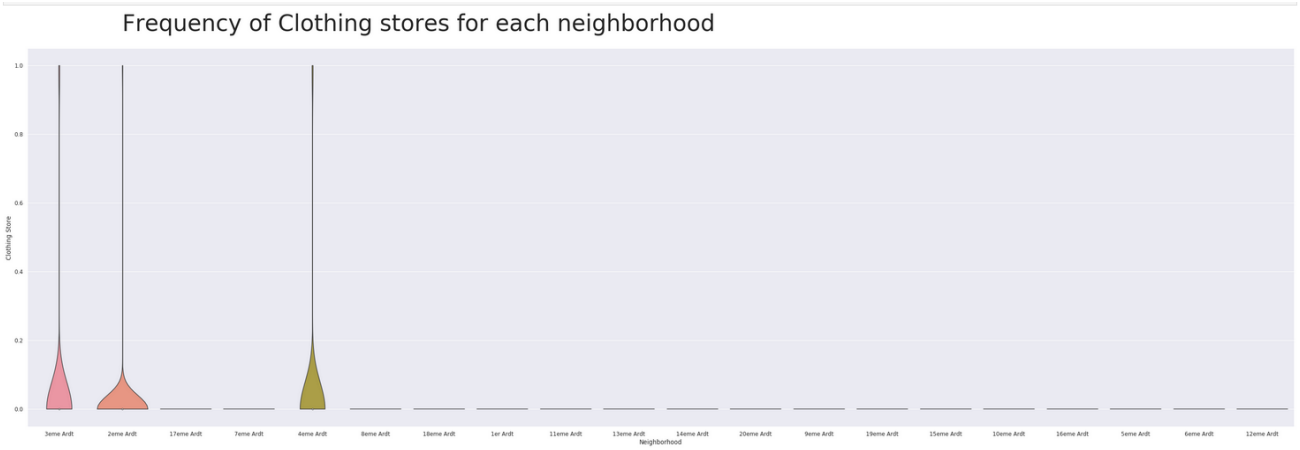


So, as we can see from the above plots, there are 8 neighbourhoods where the client can open their new stores, as the 3 specified venues (French Restaurants, Cafes and Wine Bars) are in a great frequency there. The neighbourhoods are:

- 4eme Ardt
- 5eme Ardt
- 6eme Ardt
- 9eme Ardt
- 10eme Ardt
- 11eme Ardt
- 18eme Ardt
- 20eme Ardt

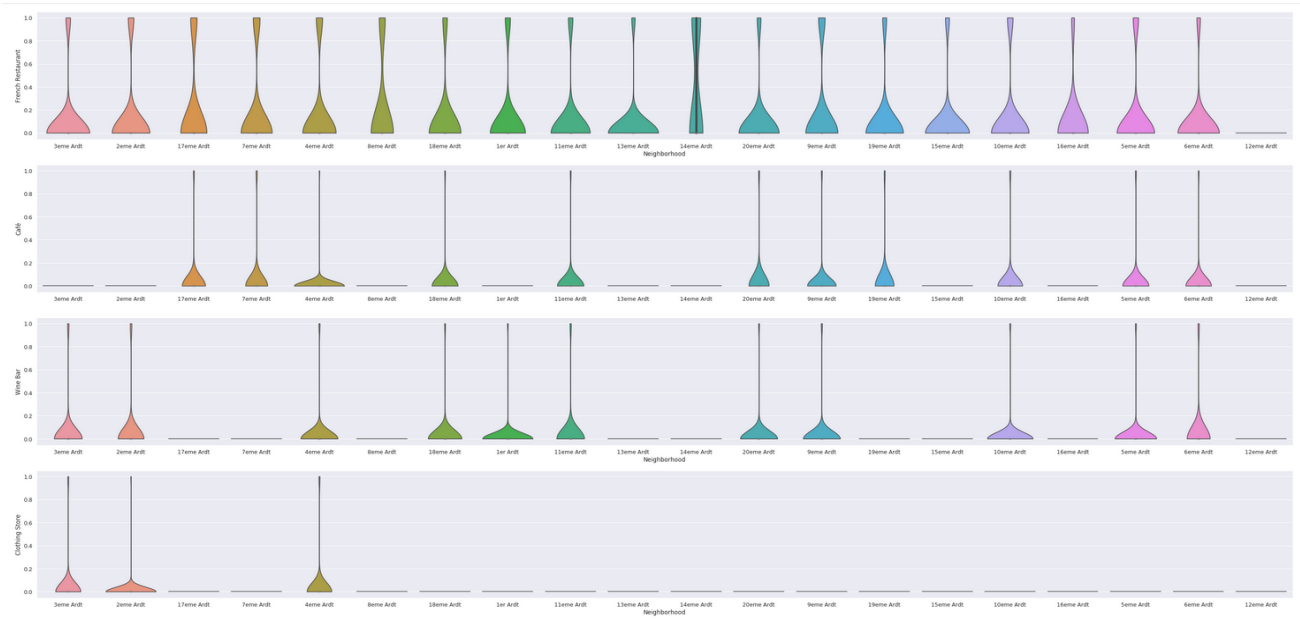
Let's take this a step further with some exploration and inferential analysis

We have the 8 neighbourhoods that all include the venue category criteria, but if we include the "Clothing_Store" venue 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.

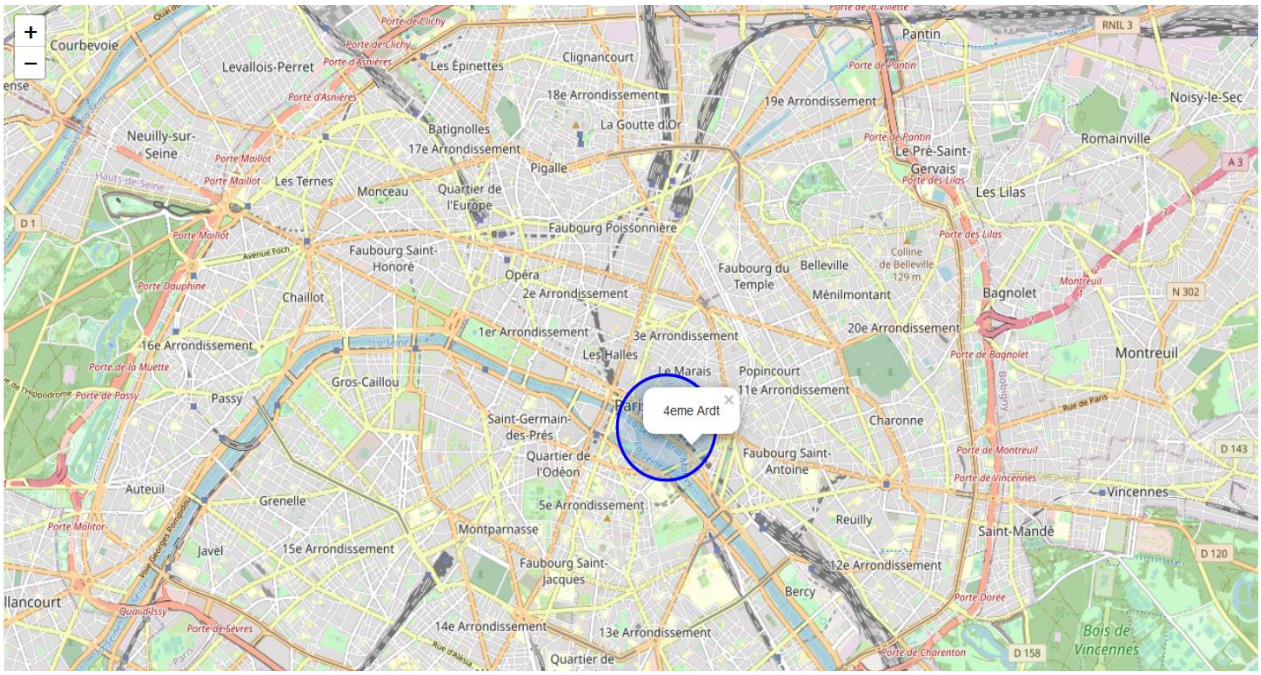


Only 3 districts have clothing stores – 3eme Ardt, 2eme Ardt and 4eme Ardt

Adding the above Clothing_Store plot to the plot with 3 venues, we see that there is only 1 neighbourhood filtered out of 8 – ***4eme Ardt***
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 for a fashion retailer.



Let’s visualize the chosen neighbourhood on the map of Paris



Observations:

It's not a surprise that the district is centrally located in the circular arrangement of Paris's arrondissements (districts). Locations fitting the criteria for popular venues would generally be in central locations in many cities of the world.

From this visualisation, on a practical level, with no data to base our decisions on, the circle of the 20 districts is very large, and researching and them visiting them all would be a tedious task. So, we have narrowed the search area down significantly from 20 potential districts to just 1 that should best 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 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.

Conclusions

There are many ways this analysis could have been performed based on different methodologies and perhaps different data sources. I chose the method I selected as it was a straightforward way to narrow down the options, not complicating what is simple in many ways – meeting the criteria for the surrounding venues. 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.

Thank You!!