The Battle of Neighborhood – Final Assignment
Suitable New Store Locations in Paris for a Fashion Retailer

Ankush Lokhande

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 Municipal (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.

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 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.

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 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/

	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

Clean up the dataset to remove unnecessary columns

	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	Pantheon	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

Week 1:

Discussion of the Business Objective and Problem / The Data Workflow

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

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

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.

The Battle of Neighborhoods continues in the next section.

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:

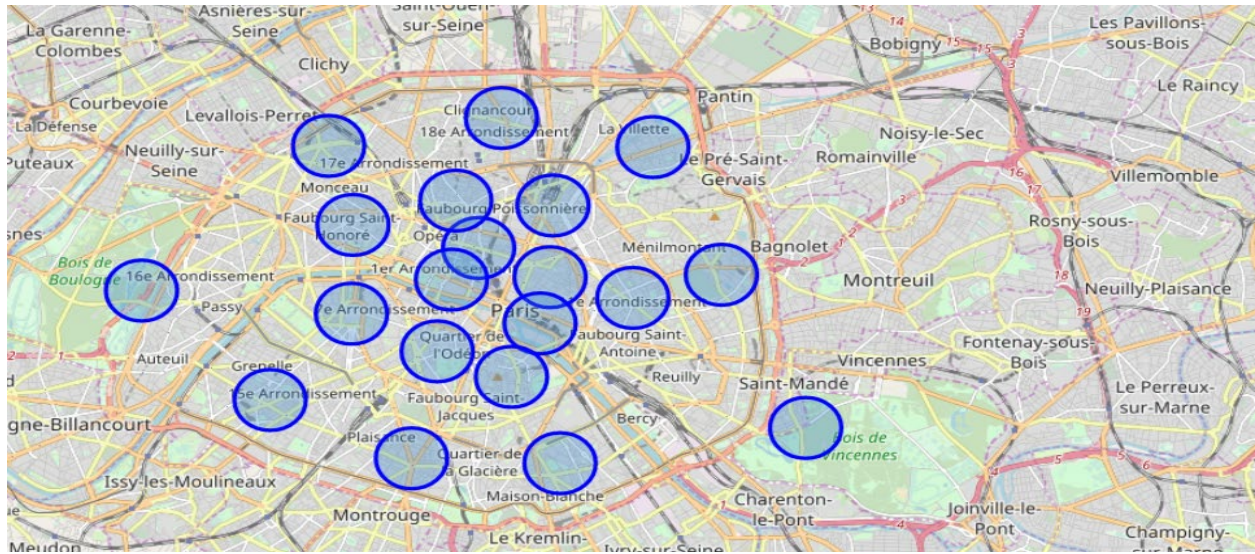Analysis and plotting visualizations.

Data visualization using various mapping libraries.

Discussion and Conclusions:

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.

Use the geography library to get the latitude and longitude values of Paris
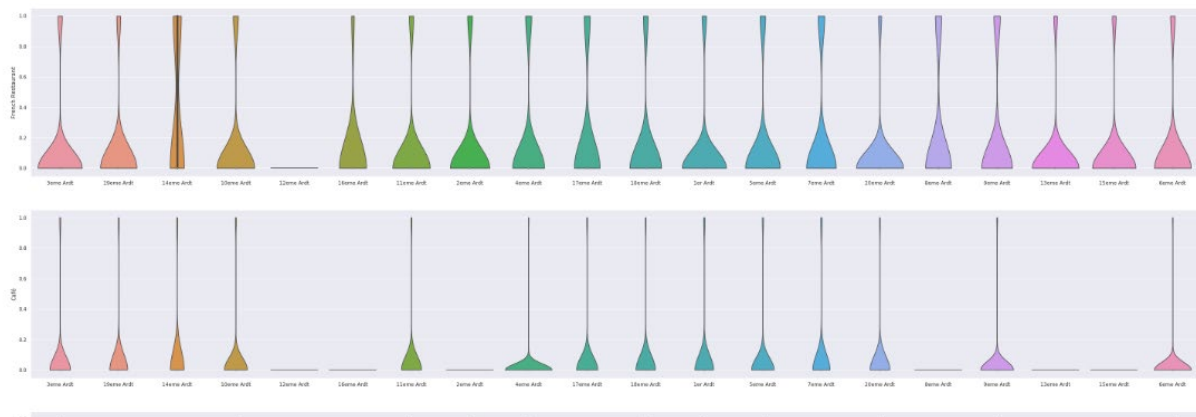
4. Results

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 meaning ful insights about the top categories of businesses in all the neighborhoods. 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	Bistro	Indian Restaurant	Café	Japanese Restaurant	Pizza Place	Seafood Restaurant	Italian Restaurant
1	11eme Ardt	French Restaurant	Restaurant	Pastry Shop	Café	Wine Bar	Supermarket	Japanese Restaurant	Bistro	Bakery	Pizza Place
2	12eme Ardt	Zoo Exhibit	Park	Supermarket	Monument / Landmark	Zoo	Cosmetics Shop	Coworking Space	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
3	13eme Ardt	Vietnamese Restaurant	Asian Restaurant	Chinese Restaurant	Thai Restaurant	French Restaurant	Juice Bar	Bakery	Sandwich Place	Furniture / Home Store	Cosmetics Shop
4	14eme Ardt	French Restaurant	Hotel	Café	Pizza Place	Bistro	Supermarket	Bakery	Plaza	Fast Food Restaurant	Food & Drink Shop
5	15eme Ardt	Hotel	Italian Restaurant	French Restaurant	Japanese Restaurant	Supermarket	Coffee Shop	Thai Restaurant	Lebanese Restaurant	Park	Bakery
6	16eme Ardt	Plaza	Lake	Pool	French Restaurant	Boat or Ferry	Art Museum	Bus Station	Park	Bus Stop	Dive Bar
7		French	Hotel	Italian	Bistro	Dub	Japanese	Dub Stop	Dla	Destaurant	Café

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

4 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:
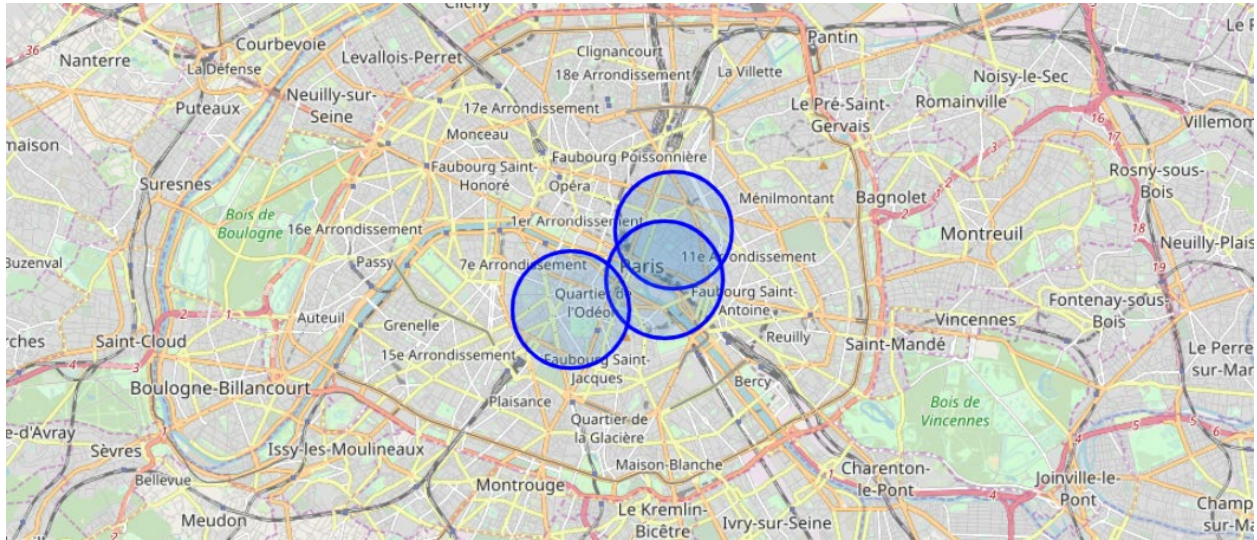

- 3eme Ardt : Arrondissement 3, Temple

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

- 6eme Ardt : Arrondissement 6, Luxembourg

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

	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

Map

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.

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 neighborhoods 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 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 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.

Thanks for taking part in my Data Science journey!