Capstone Project - The Battle of Neighborhoods

Created with IBM Watson Studio

Suitable New Store Locations in Paris for a Fashion Retailer

This notebook contains multiple parts:

1. A description of the problem and a discussion of the background
2. A description of the data and how it will be used to solve the problem
3. Methodology and Exploratory Data Analysis
4. Inferences and Discussion

Introduction and Discussion of the Business Objective and Problem

<u>Locations for New Fashion Stores in High Traffic Areas in Paris France</u>

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 up market 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.

The Data Science Workflow

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.

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 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.

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/*

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 dataframe.
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.

Data Analysis

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 visualization and statistical analysis.

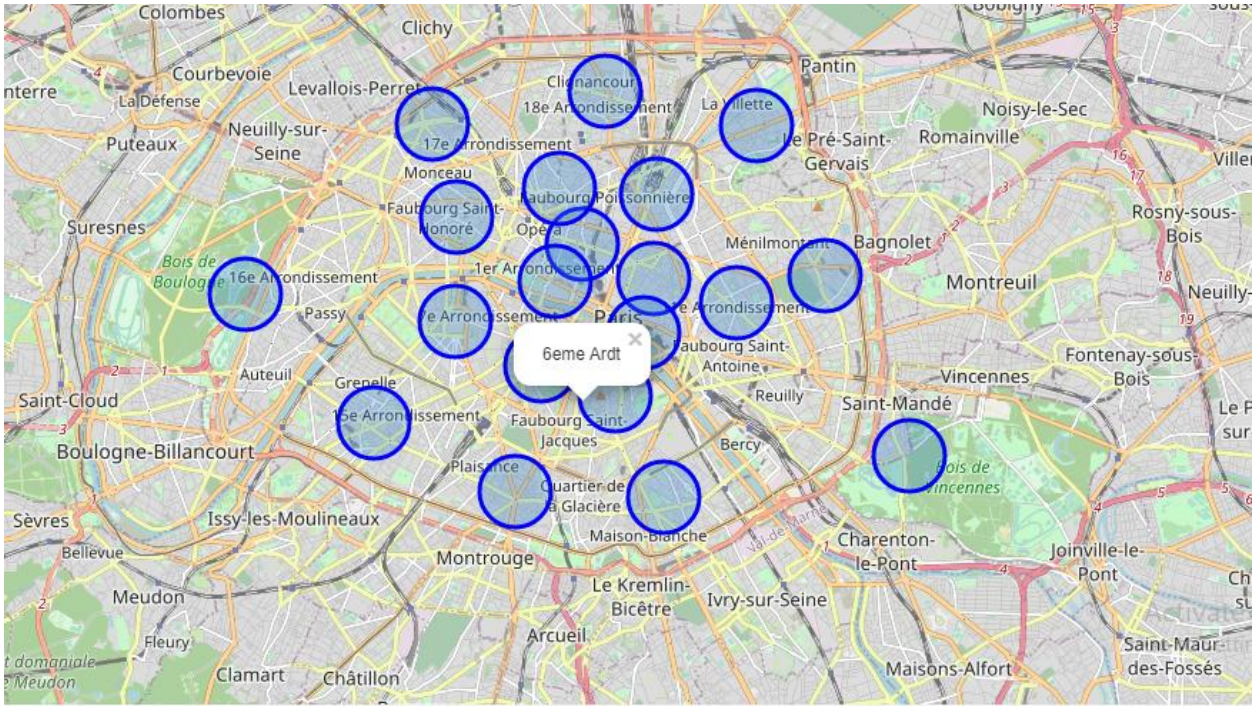
Visualization:

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.

Create a map of Paris with districts superimposed

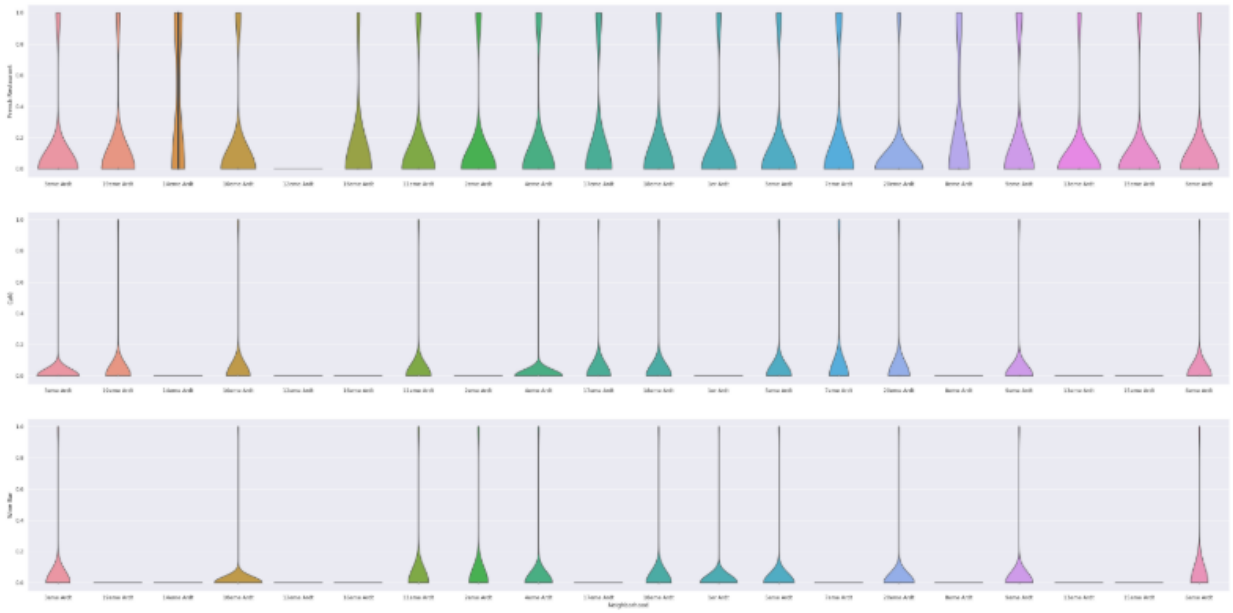


The business types criteria specified by the client! *'French Restaurants', 'Cafés' and 'Wine Bars'*

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.

Frequency distribution of top 3 venue categories for each neighbourhood using violinplot



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:

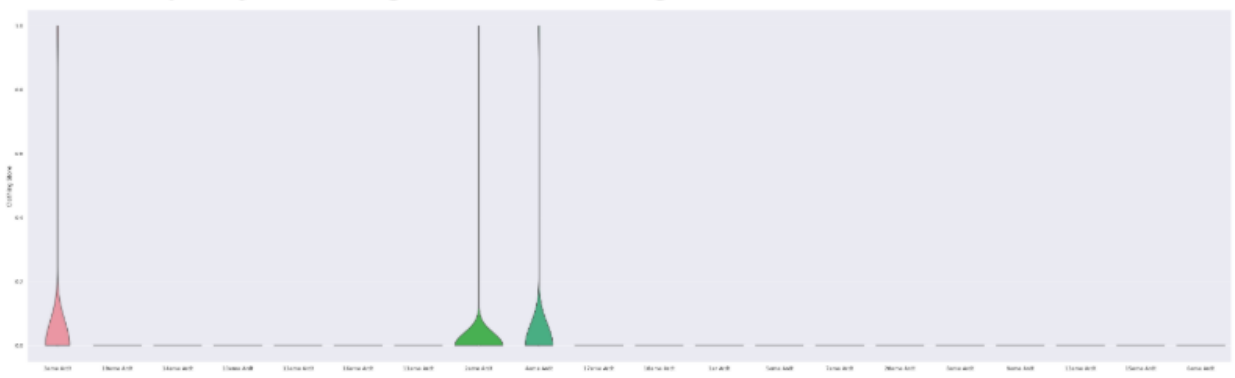
Neighborhoods

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

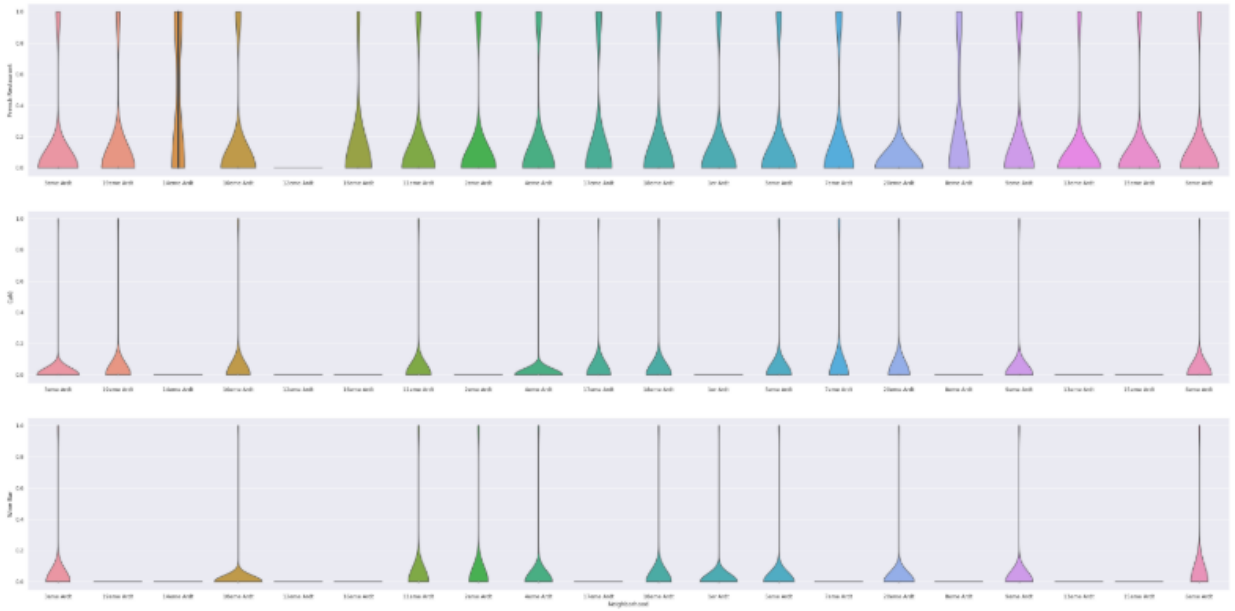
Let's take this further with some exploration and Inferential Analysis

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.

Frequency of clothing stores for each neighbourhood



Frequency distribution of top 3 venue categories for each neighbourhood (include clothing's)



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

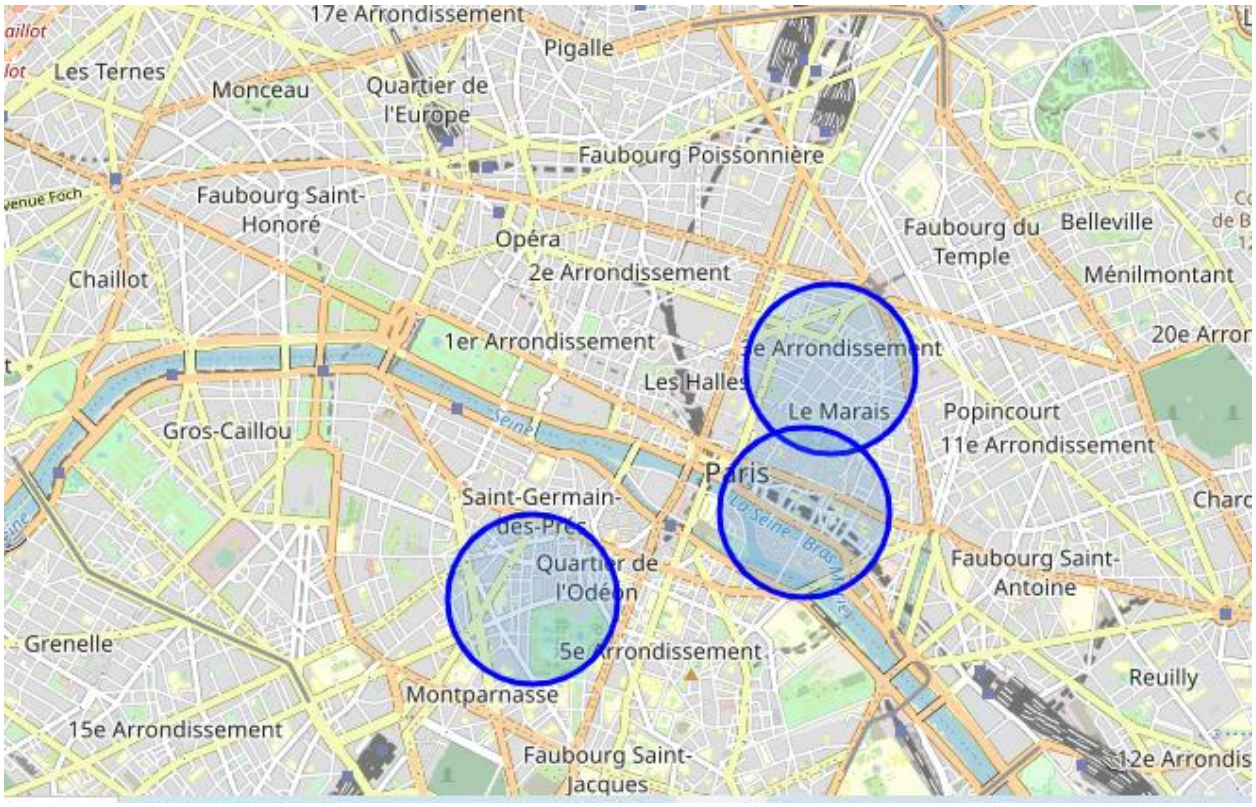
- 4eme Ardt : Arrondissement 4, Hotel-de-Ville

- 6eme Ardt : Arrondissement 6, Luxembourg

Chosen Neighborhoods – Results

	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

Visualize result on a map of Paris



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!

Report by Rishi Ranjan (Techno Engineering college Banipur)