Suitable New Store Locations in Paris for a Fashion Retailer

The problem

In a hypothetical situation, I play a role as a data scientist to support a fashion retailer to open new stores in busy traffic areas in Paris, France.

I have been given the interesting task of helping them make data-driven decisions on the new locations that are most suitable for their new stores in Paris. This will be a main element of their decision-making process, the other being on the ground qualitative analysis of districts once when the report for this data is reviewed and studied.

This fashion retailer is positioned in the upper end of the fast fashion market. Therefore, they only focus to 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.

The goal of this problem is to identify the best districts to open new stores as part of the company's plan. The results will be translated into management in a simple form that will convey the data-driven analysis of the best locations to open stores.

The discussion for the background

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 places often, so opening new stores in these locations are 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.

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 data will provide better answers and better solutions to their task at hand.

The description of the data and how it will be used to solve the problem

The main districts in Paris are divided into 20 administrative districts, shortened to arrondissements. This data is available on the web and can be manipulated and cleansed to provide a meaningful dataset to use.

- Wikipedia page: https://en.wikipedia.org/wiki/Arrondissements_of_Paris
- Open|DATA France: https://opendata.paris.fr/explore/dataset/arrondissements/table/?data Chart
- Opendatasoft: https://data.opendatasoft.com/explore/dataset/arrondissements %40parisdata/export/

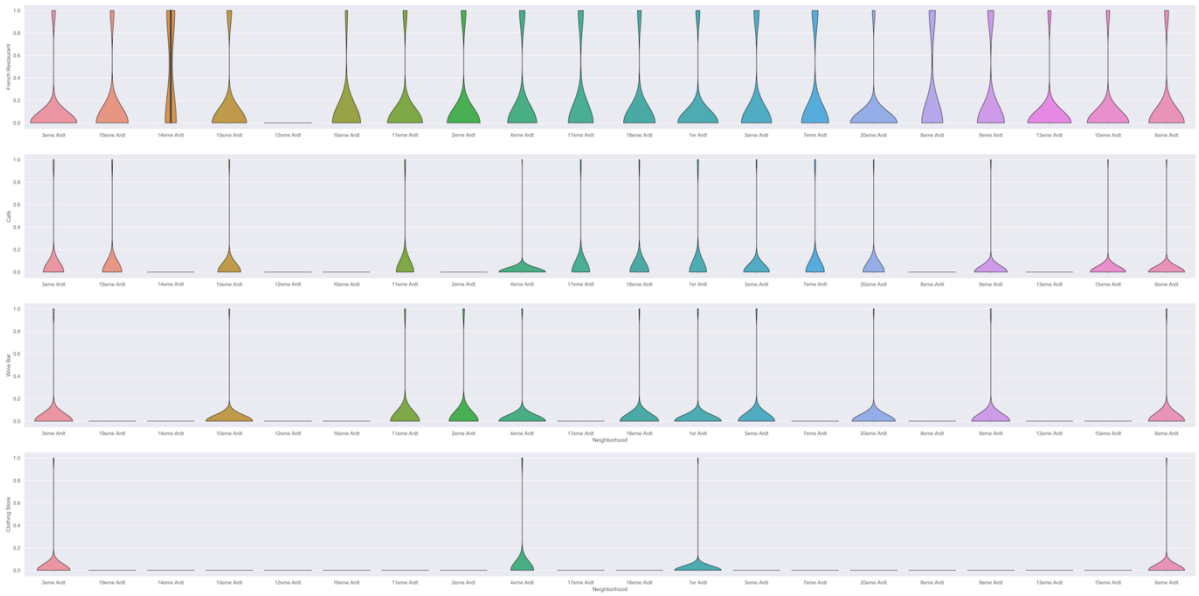
The data regarding the districts in Paris need to be researched and a suitable useable source identified. If it is found but is not in a useable form, data wrangling and cleaning will have to be performed.
The cleansed data will then be used alongside Foursquare data, which is readily available. Foursquare 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 is most interested in opening new stores.

- **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 dataframe.
- **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:**
  - Recomendations 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.

Result and Visualization
These are the venue types that the client wants to have an abundant density of in the ideal store locations. I have used a violin plot from the seaborn library, a great way to visualise frequency distribution datasets, to display a density estimation of the underlying distribution.
I choose specified venues in a great frequency (French Restaurants, Cafés and Wine Bars)

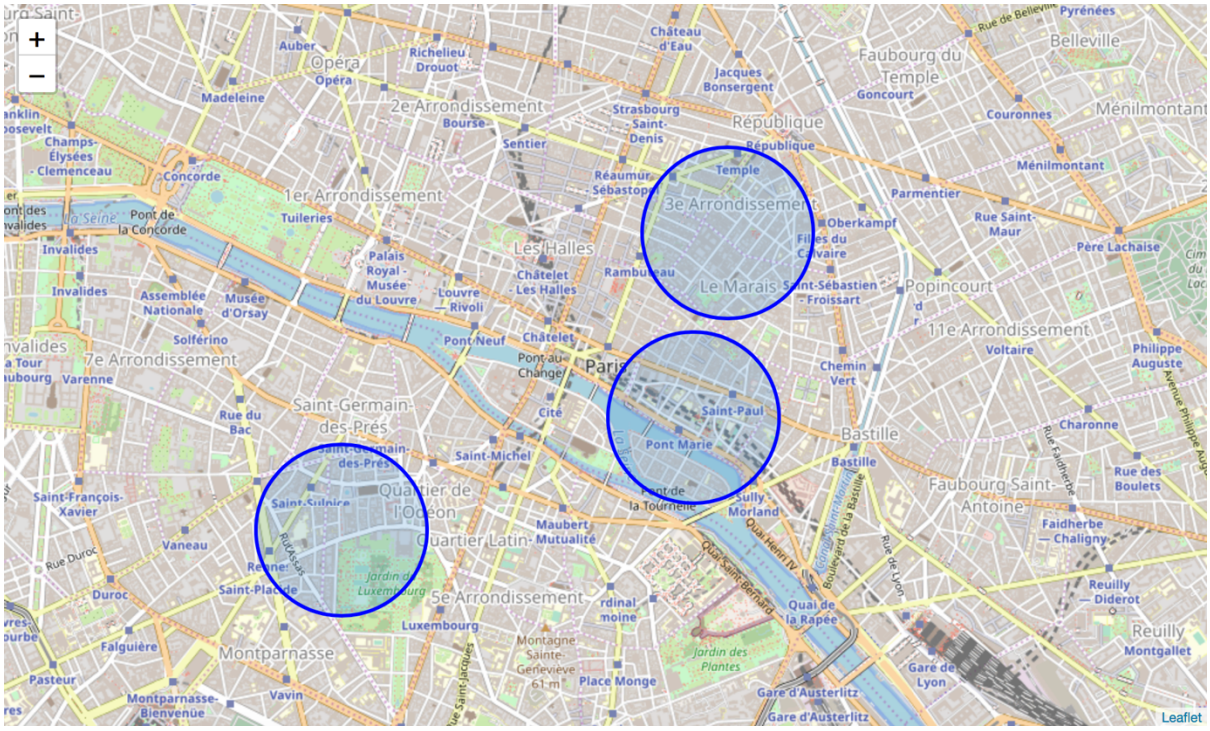
So from the analysis, there are 8 neighborhoods to open new stores, according to the criteria that they have the 3 specified venues in a greater frequency (French Restaurants, Cafés, and Wine Bars). They include as follows:

- 3eme Arrondissement
- 10eme Arrondissement
- 11eme Arrondissement
- 4eme Arrondissement
- 18eme Arrondissement
- 5eme Arrondissement
- 9eme Arrondissement
- 6eme Arrondissement



Inferential analysis using the data, as well as domain knowledge of retail and marketing, allows the list to be focussed to just 3 neighbourhoods from the previous 8 neighbourhoods. So, the final 3 prospective neighborhoods for new fashion retail store locations are where 4 criteria, Restaurants, Cafés, Wine Bars, and Clothing Stores, are met:

1. 3eme Arrondissement
2. 4eme Arrondissement
3. 6eme Arrondissement



Discussion

I guess it is 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.

For 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. 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 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 straight forward way to narrow down the options, not complicating what is actually simple in many ways – meeting 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 the 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 a way forward, these data-driven decisions will lead to a better solution in the end.