Predicting Suitable New Store Locations in Paris for a Fashion Retailer

Valuable for new business ideas

Predicting Suitable New Store Locations in Paris for a Fashion Retailer is valuable for new business ideas.

Generally, selecting new locations are valued by near existing **French Restaurants, Cafés and Wine Bars** areas.

Data acquisition and cleaning

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/

Clustering Techniques

Get Paris location data from Foursquare like Latitude and Longitude.
Cluster (Group) the Venues using French_Name(District name) .
Analyze each neighborhoods.

----10eme Ardt----

	venue	freq
0	French Restaurant	0.13
1	Hotel	0.05
2	Coffee Shop	0.05
3	Bistro	0.05
4	Café	0.04
5	Indian Restaurant	0.04
6	Pizza Place	0.03
7	Japanese Restaurant	0.03
8	Restaurant	0.02
9	African Restaurant	0.02

----11eme Ardt----

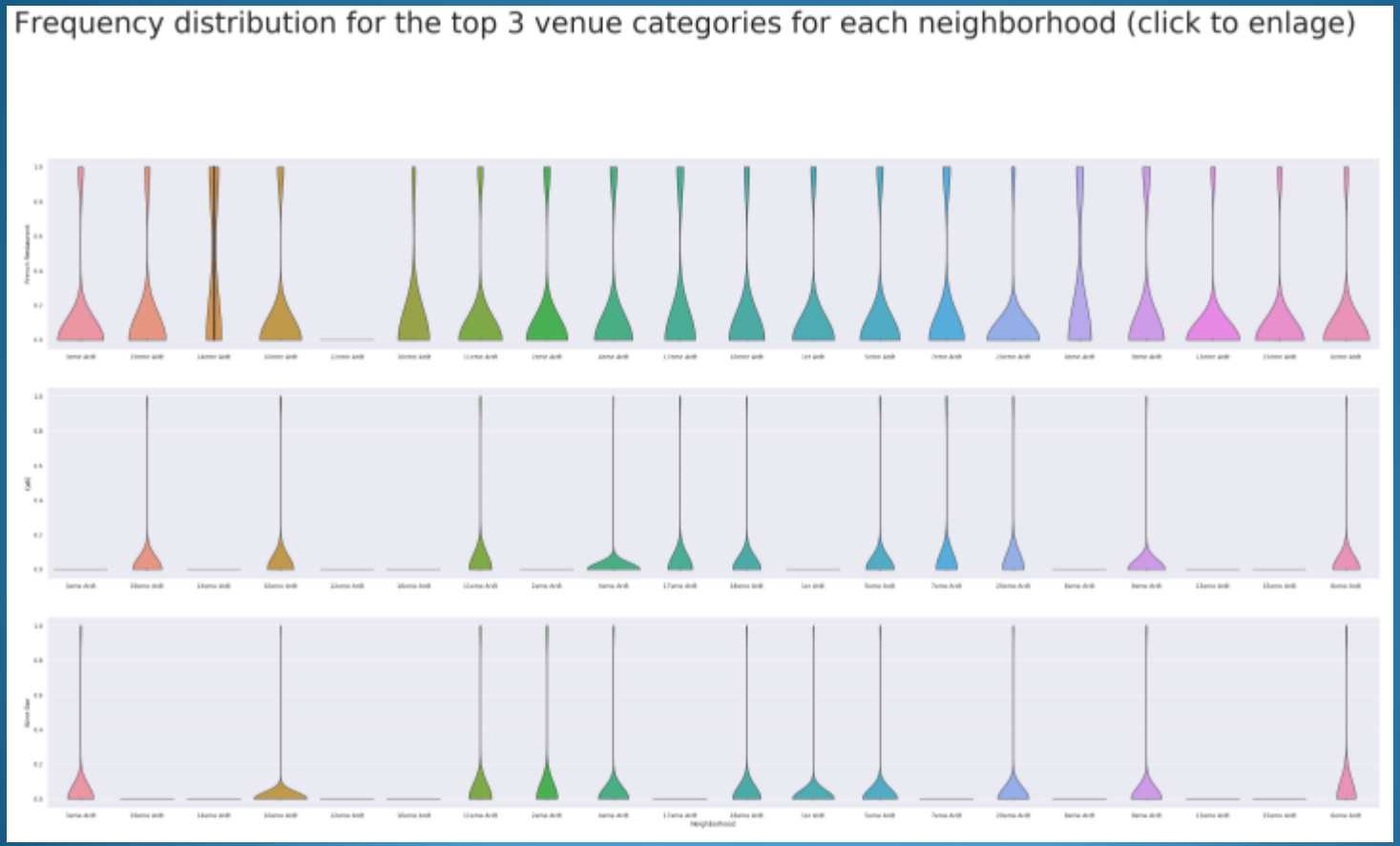
	venue	freq
0	French Restaurant	0.10
1	Supermarket	0.06
2	Restaurant	0.06
3	Café	0.04
4	Wine Bar	0.04
5	Pastry Shop	0.04
6	Bakery	0.03
7	Japanese Restaurant	0.03
8	Theater	0.03
9	Bar	0.03

----12eme Ardt----

	venue	freq
0	Zoo Exhibit	0.25
1	Supermarket	0.25
2	Monument / Landmark	0.25
3	Zoo	0.25
4	Plaza	0.00
5	Organic Grocery	0.00
6	Movie Theater	0.00
7	Museum	0.00
8	Music Store	0.00
9	New American Restaurant	0.00

Visualize frequency distribution

Let's look at their frequency of occurrence for all the Paris neighborhoods, isolating the categorical venues.



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.

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.

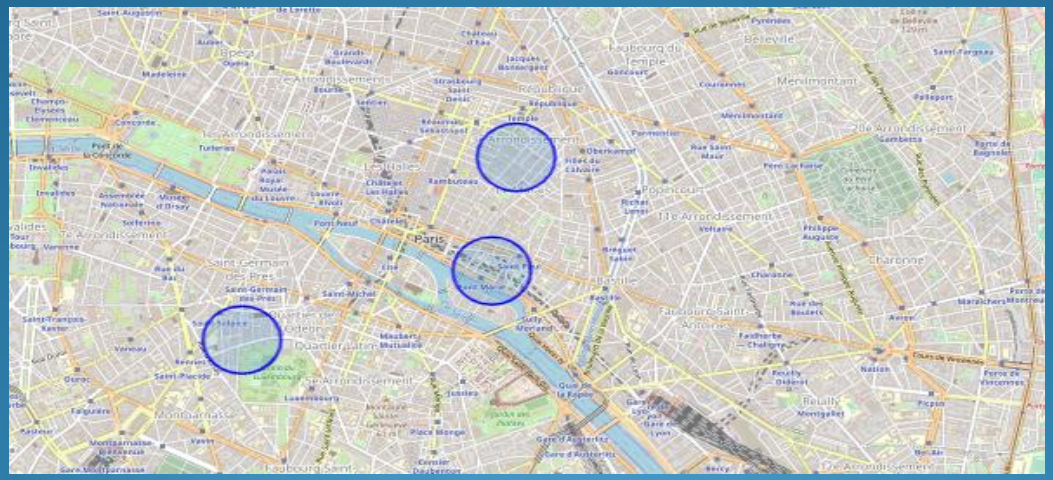
So we can increase the criteria to include Restaurants, Cafés, Wine Bars and Clothing Stores - which narrows down and focuses the suggested districts for new stores to be located, and at the same time provides better locations for the brand. So the final 3 prospective neighborhoods for new store locations are where 4 criteria are met:

- 3eme Ardt : Arrondissement 3, Temple
- 4eme Ardt : Arrondissement 4, Hotel-de-Ville
- 6eme Ardt : Arrondissement 6, Luxembourg

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.

Conclusion

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 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.