Capstone Project
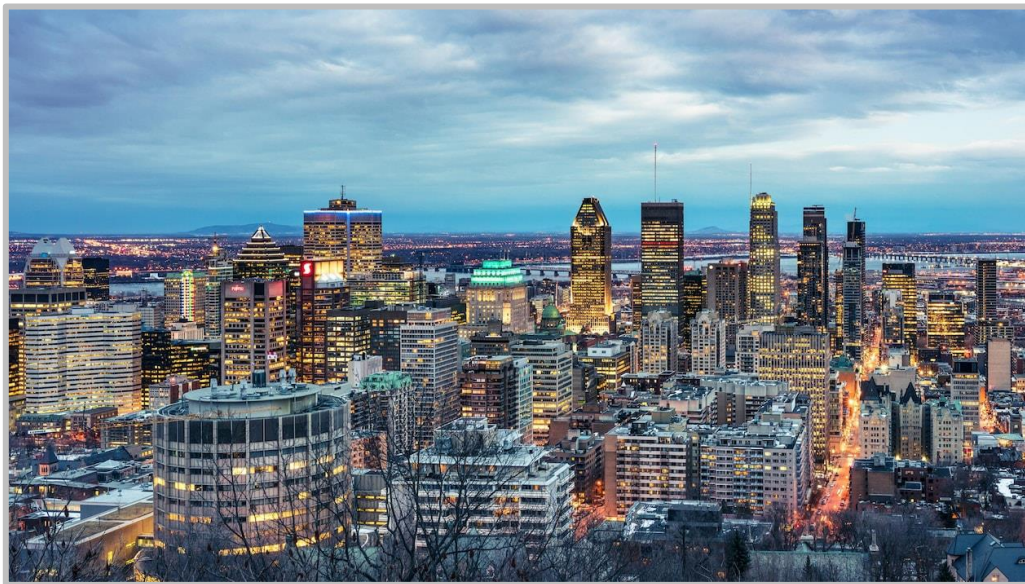
Battle of neighborhoods

Find the right area to open a new French Restaurant
in **Montreal**, Canada

Jérémy B.

July 2020

Table of Contents

Introduction/Business Problem section

Introduction: my personal experience with Montreal

I came in 2014 in Montreal during 2 weeks for holydays.

As a European guy, It was the first time (and the only) I had the opportunity to visit Canada, and especially this beautiful city.

Montreal represents for me a real multicultural city, modern and dynamic, with a lot of spots for arts, restaurants and tourism. The openness of this city and its people made me think many times about the advisability of starting a business in this place. I will take the opportunity of this Capstone project to evaluate where this kind of project could be done in Montreal.

Business Problem section

Be able to find the right place to open a new business is not an easy thing. Competitors and locations are important variables which need to be taken account before to start such a project.

Data Science offers the possibility to make analysis and studies at the beginning of the project building; Open Data and algorithms can now help us to take the best decision. Our problematic is the following:

- *In which area of Montreal could be opened a French Restaurant?*

Thanks to Foursquare API, to Open Data and to a clustering algorithm, the following study will propose acceptable areas to open a new burger restaurant (or American Restaurant) in Montreal

Target Audience

The target of this analysis is potential investors or people who want to start a new Business. Due to my nationality and the French culture of Montreal, the choice was done on "French Restaurant".

Banks involved in the business plan of such a project could also be interested by this Data Science analysis.

Data section

Data Available

Open Data used come from the website https://donnees.montreal.ca

The geographical coordinates used from this website concern the Police posts in Montreal. **These Data allow us to identify the main geographical areas of this city** and give us a representative view of the split of this city in term of human activities: these coordinates will give us the opportunity to check the venues around all over the city.

In a second steps, thanks to the Foursquare API, we will associate the venues related to these areas; thanks to these Data from Foursquare, we will be able to do some assumptions and proposal.

Data Structure

Data from https://donnees.montreal.ca[1] have the following structure:

	NO_CIV_LIE	PREFIX_TEM	NOM_TEMP	DIR_TEMP	MUN_TEMP	DESC_LIEU	Longitude	Latitude
0	4139	BOUL	SOURCES DES	NaN	DDO	POSTE DE QUARTIER 4	-73.807304	45.494380
1	8930	BOUL	PIE-IX	NaN	MTL	POSTE DE QUARTIER 30	-73.617433	45.577744
2	4555	RUE	HOCHELAGA	NaN	MTL	POSTE DE QUARTIER 23	-73.545915	45.556866
3	11756	BOUL	O'BRIEN	NaN	MTL	POSTE DE QUARTIER 10	-73.701029	45.528543
4	1435	AVEN	VAN-HORNE	NaN	OUT	POSTE DE QUARTIER 24	-73.615612	45.520110
5	6100	BOUL	HENRI-BOURASSA	NaN	MTN	POSTE DE QUARTIER 39	-73.615625	45.607724

For this analysis, these 3 fields will be used:

- NOM_TEMP ➔ Provide a global description of the area where is the Police post ("Will be renamed in "Area")
- Longitude ➔ Longitude of the area
- Latitude ➔ Latitude of the area

---

[1]These detailed Data are available through the following link: https://donnees.montreal.ca/ville-de-montreal/carte-postes-quartier

Methodology section

Algorithm used

To find the right places where to open a French Restaurant in Montreal, a clustering analysis has been chosen. Thanks to the Foursquare Data available, we can easily plot on a map or in a table the existing French Restaurant in Montreal.

We can now find for each geographical areas (from https://donnees.montreal.ca) the frequency of the existing French Restaurants (from Foursquare).

I will use **K-Means clustering** to categorize the identified geographical areas linked with the French Restaurant to find the best places to open such a restaurant. K-Means allows to divide the geographical areas into non-overlapping subsets (cluster), without any cluster-internal structure. Within each cluster, a strong similarity will be present; we need to find a cluster with a little number of French Restaurants (and therefore very little competition).

Steps followed

- Import Open Data (https://donnees.montreal.ca) in a Dataframe
- Data Cleaning (Column NOM_TEMP becomes Area)
- Storage of Montreal Coordinates
- Creation of the map of Montreal to display the geographical areas
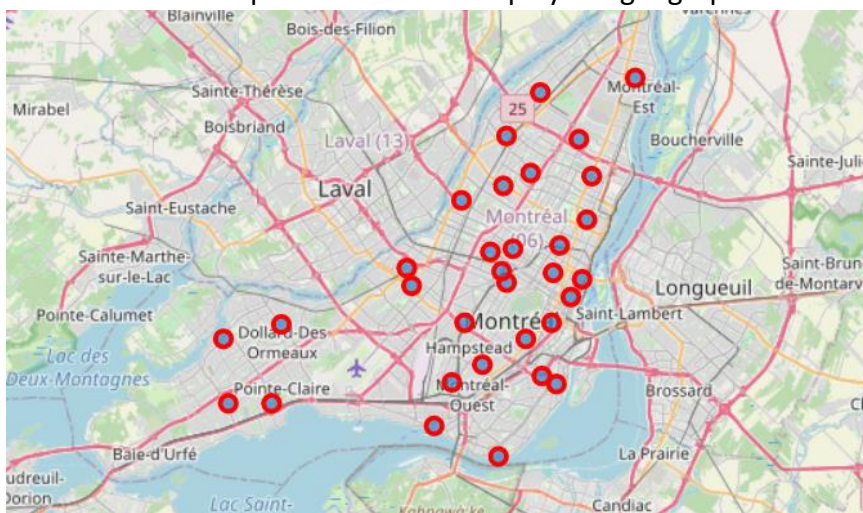
- Foursquare Credentials
- Creation of a function to explore all the areas in Montreal through Foursquare API
- Use of this function on each neighborhood and create a new Dataframe
- Analyze Each Neighborhood, display the existing Categories (venues)
- Grouping of rows by neighborhood and by taking the mean of the frequency of occurrence of each category
- Identification of the "French Restaurant" category

- In the current Dataframe, we keep only the "French Restaurant" category
- **k-means**: Find the right value of "K" (Elbow method)
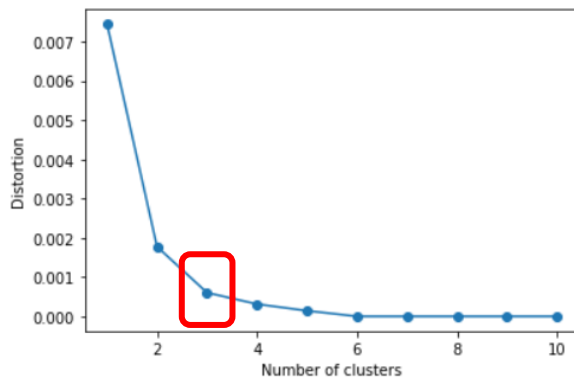
- Run k-means to cluster the neighborhood into 3 clusters
- Create a new Dataframe that includes the cluster as well as the coordinates for each neighborhood / area
- Data Visualization on a map of these 3 Clusters
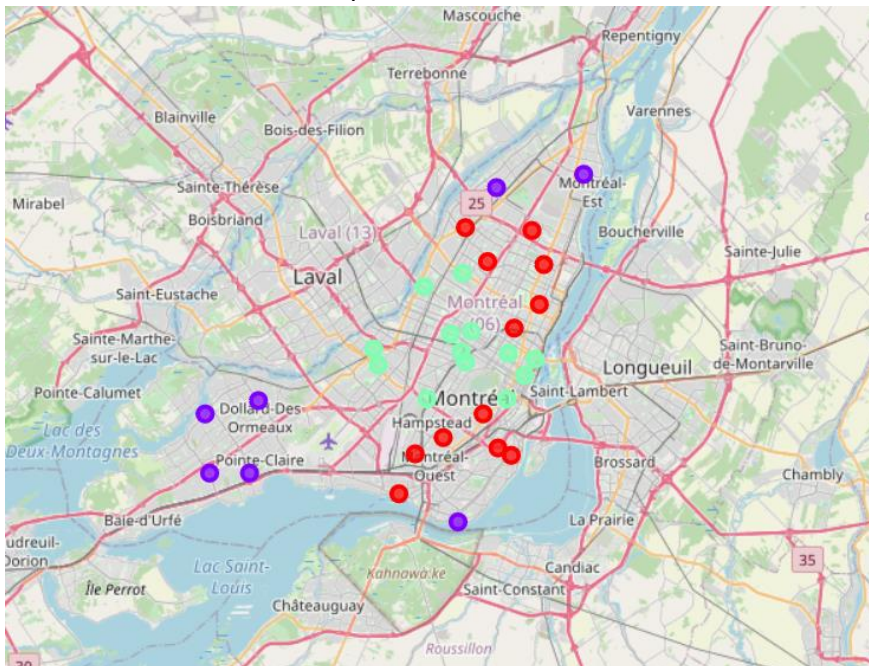
- Cluster Analysis (Ranking and detail of each Cluster)

Result section

We can see below the different clusters calculated by k-means algorithm, with some information to interpret these clusters, like the frequency of French Restaurants in the corresponding area:

Cluster Labels	Neighborhood (Area)	French Restaurant Frequency
0	15E	0.02
0	STANTON	0.03
0	SOMERLED	0.03
0	SHERBROOKE	0.03
0	WESTMINSTER	0.03
0	MOLSON	0.03
0	LACORDAIRE	0.03
0	JOSEPH-RENAUD	0.02
0	WILLIBRORD	0.02
0	HENRI-BOURASSA	0.02
0	EGLISE DE L'	0.02
0	HOCHELAGA	0.03
1	LASALLE	0.01
1	MAURICE-DUPLESSIS	0.01
1	PIERREFONDS DE	0.0
1	ST-JEAN-BAPTISTE	0.0
1	SOURCES DES	0.0
1	ST-CHARLES	0.0
1	ST-JEAN	0.0
2	VAN-HORNE	0.04
2	STE-ELISABETH	0.04
2	STE-CATHERINE	0.04
2	BEAUMONT	0.05
2	ST-LAURENT	0.04
2	BELANGER	0.04
2	PIE-IX	0.04
2	FLEURY	0.05
2	O'BRIEN	0.04
2	MOUNTAIN SIGHTS	0.04
2	GRENET	0.04
2	RACHEL	0.04
2	PAPINEAU	0.04

The following analysis can be done regarding these 3 clusters:

- "Mint" Cluster (n°2): contains frequencies between **0.04 and 0.05**, which means a high concentration of French Restaurants in the areas of this cluster
- "Red" Cluster (n°0): contains frequencies between **0.02 and 0.03**, which a middle concentration of French Restaurants in the areas in this cluster
- "Purple" Cluster (n°1): contains frequencies between **0.0 and 0.01**, which a low concentration of French Restaurants in the areas in this cluster

Based on the Data available for this analysis and on these results, the following map could help us to find the places with the low competition to open a French Restaurant in Montreal. In this framework, the "Purple" Cluster shows the interesting areas where to launch such a project:
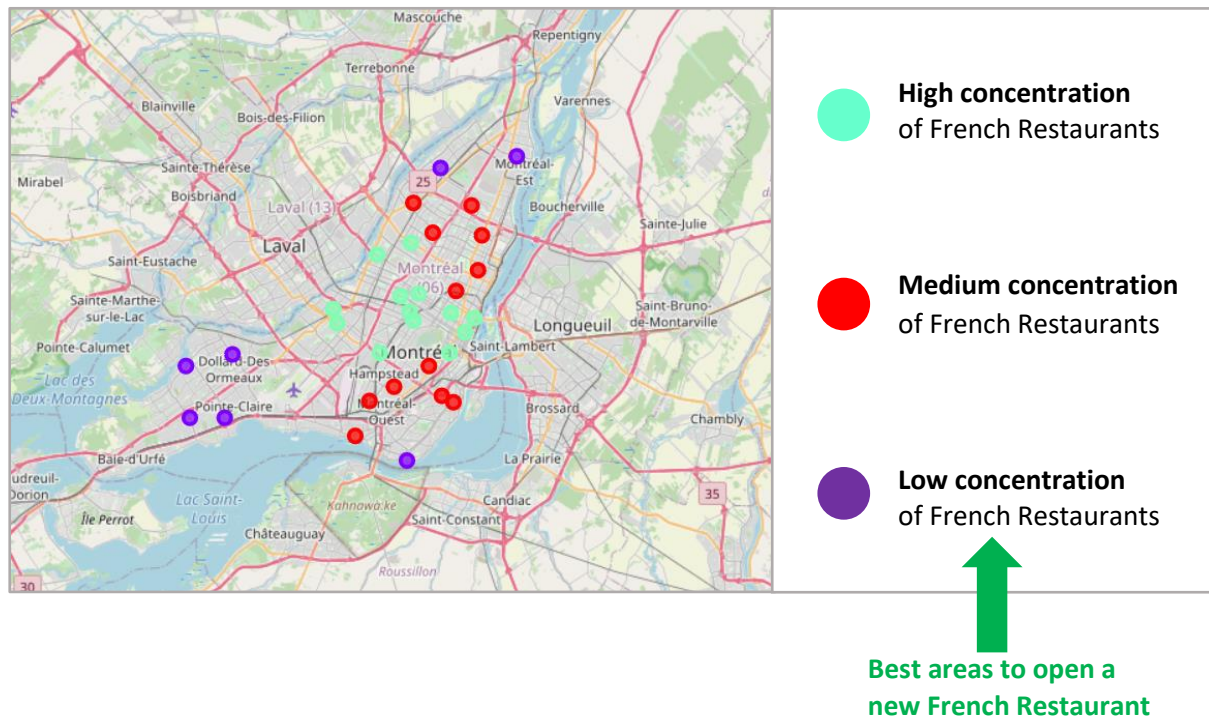
The results of this analysis highlight the areas where there is a little of French Restaurants. These areas allow to avoid strong competition with other French Restaurants, and to find easily new customers; **the 6 areas in Montreal of the purple Cluster** could be good places to start this kind of business.

Discussion section

Even if these results could provide a first view and some clues regarding the places where to open a French Restaurant in Montreal, some other variables must be taken in consideration.

Indeed, the "competitors" aspect must be completed by other information to target more precisely the better areas:

- **Demographical Data for each areas:**
    - number of habitants
    - age

- **Economic Data for each areas:**
    - number of active companies
    - average basket for a dinner in a restaurant

- **Sociological Data:**
    - habits in term of lunch for the people who works closely in these areas
    - habits in term of lunch for the people who lives in these areas

All these variables are not exhaustive; a lot of others Data could enrich k-means algorithm.

Conclusion Section

In this project some important steps of a Data Science project have been completed: definition of a business problem, specification of the target, search and acquisition of the Data needed, Data cleaning, choice of the appropriate algorithm and analysis of the results.

The results provided, a list of 6 potentials areas in Montreal where a French restaurant could be opened with a limited competition.

This analysis provides a first overview of Data Science (and this case, of k-means algorithm) possibilities ; however to be totally complete and efficient, some new Data set could be added and used.