

Coursera Certification- Applied Data Science Capstone

Project- The battle of Neighborhoods

Opening an Italian restaurant in the city of Montreal

Prepared by Youssef Charradi

April 30, 2021

Table of Contents

1-Introduction.....	3
2-Problem	3
3-Data Exploration and Methodology	4
3-1-Boroughs with High population	4
3-2-Creating candidates' positions around the city Boroughs	5
3-3-Exploiting Foursquare API data to filter out the list of locations	6
4-Clusterification and comparison.....	9
4-1-Using Machine learning to create Clusters.....	9
4-2-Clusters Analysis 1 : Number of Restaurants per Cluster.....	11
4-3-Clusters Analysis 2 : Ratio of Restaurants per 1 million capita	12
4-4-Clusters Analysis 3 : Number of Italian Restaurants	13
4-5-Clusters Analysis 4 : Number of parks.....	13
4-6-Final results.....	15
5-Conclusion	17
6-References	17

1-Introduction

Montreal (Canada) is the second biggest French speaking city in the world after Paris (1), it is an attractive part of Canada that attracts millions of immigrants and tourists every year. Montreal has a **multinational cultural heritage** that makes its cultural landscape particularly vibrant, and this is clearly reflected through the richness and diversity of the restauration scene.

The **food industry** in Montreal is very cosmopolitan with diverse set of menus from all over the planet(1)

In addition to that, Montreal has the highest restaurants per capita in north America after New York, which makes the Restauration industry **highly competitive** sector especially for the well established cultures like the Italian culture. The city of Montreal is home for 20% of Canadians with Italian decent, so opening **an Italian restaurant business** is not a simple task.

2-Problem

In order to open a restaurant to serve Italian pizzas, choosing the **right location** in Montreal is crucial for the success of the business. It is important to locate the best available spots around the city.

The city is renewed for the reparable number of potential locations that exists: Beautiful neighborhoods, walkable areas, easy access by public transport means and bikes, closeness to museums and parks...

In this study, we will work on identifying the list of addresses for best possible locations to open **an Italian Pizza restaurant**.

3-Data Exploration and Methodology

3-1-Boroughs with High population

One of the key elements to identify the most attractive locations in Montreal, is to search for the areas with high density of population. Since more people means more visitors, we are interested in mapping out the most populated boroughs in Montreal.

The website "wikipedia" offers an overview of the boroughs in Montreal, with the size of the area and the density of the population. This can be used to draw a map of the possible spots(https://en.wikipedia.org/wiki/Boroughs_of_Montreal (2)).

We used BeautifulSoup library in order to extract data from Wikipidea page.

	Borough	Population Canada 2016 Census[1]	Area in km ²	Density per km ²
5	Le Plateau-Mont-Royal	104000	8.1	12839.5
13	Rosemont-La Petite-Patrie	139590	15.9	8779.2
18	Villeray-Saint-Michel-Parc-Extension	143853	16.5	8718.3
2	Côte-des-Neiges-Notre-Dame-de-Grâce	166520	21.4	7781.3
9	Montréal-Nord	84234	11.1	7588.6
16	Verdun	69229	9.7	7137.0
10	Outremont	23954	3.9	6142.0
15	Saint-Léonard	78305	13.5	5800.4
0	Ahuntsic-Cartierville	134245	24.2	5547.3
17	Ville-Marie	89170	16.5	5404.2
8	Mercier-Hochelaga-Maisonneuve	136024	25.4	5355.3
6	Le Sud-Ouest	78151	15.7	4977.8
4	LaSalle	76853	16.3	4714.9
1	Anjou	42796	13.7	3123.8
11	Pierrefonds-Roxboro	69297	27.1	2557.1
12	Rivière-des-Prairies-Pointe-aux-Trembles	106734	42.3	2523.2
3	Lachine	44489	17.7	2513.5
14	Saint-Laurent	98828	42.8	2309.1
7	L'Île-Bizard-Sainte-Geneviève	18413	23.6	780.2

Figure 1 – List of Boroughs sorted by their population density

For our study, we decided to work only with the top 15 broughs based on the density of population per km2.

3-2-Creating candidates' positions around the city Boroughs

Our strategy is to create potential positions around the map, places where we think it is interesting to open the Italian restaurant.

The first step is to identify the Boroughs centers' locations, so we relied on the Google Geocode API (3) to extract this information. The API provide the geographic coordinates of well-known locations, in our case the Boroughs of Montreal.

The second step is to build a list of possible positions around city Boroughs, so we used the Folium library in order to draw the map and generate th zones that contains potential spots in a hexagon shape. We chose to create locations that are 400 meters away from each other.

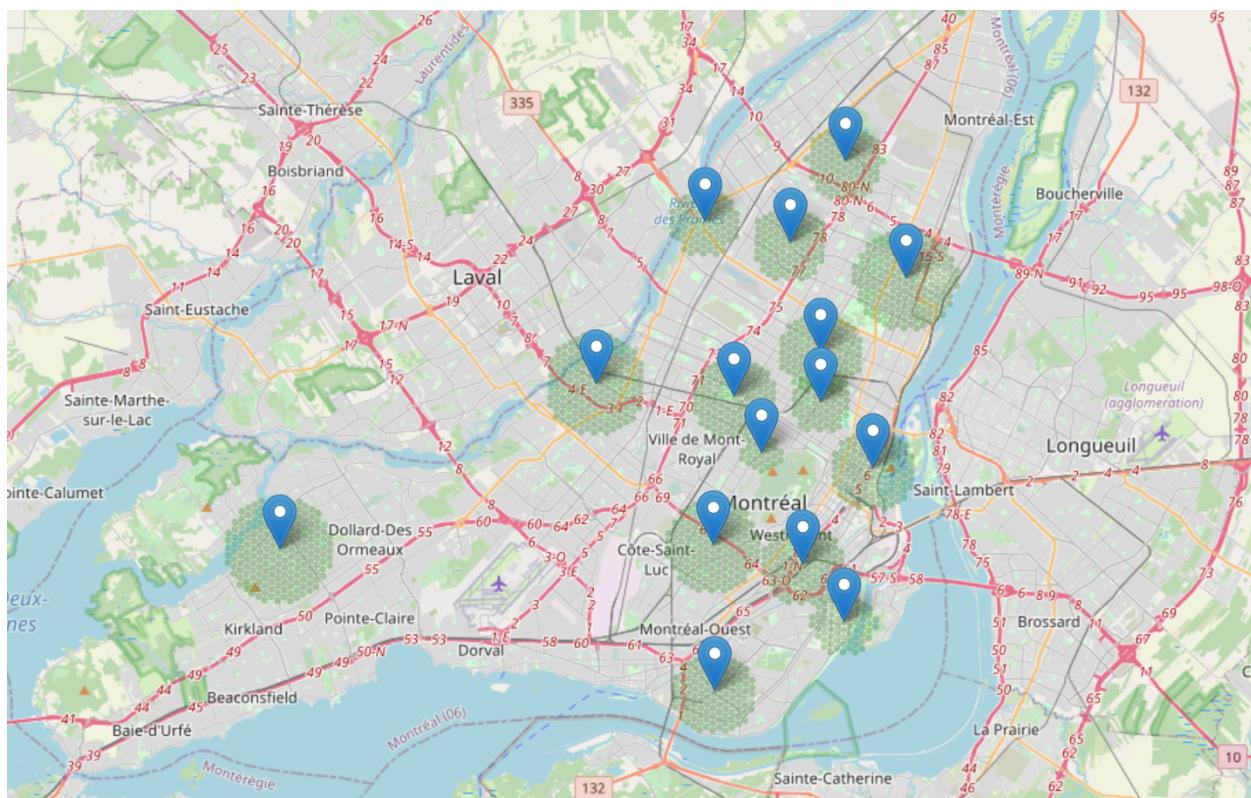


Figure 2 – Potential locations generated around the city Borough centers

We have created 1725 new locations around the city, but we are not sure that all of them are interesting and can be exploited.

1. The first challenge we see on the map if we zoom in, is that there is some locations that landed on the rivers water, or a bridge, or a shore, or even a highway, which cannot be used to open a restaurant. So we need to find a way to remove these 'fake' locations.
2. The second challenge is to remove locations that are not commercial or touristic (like residential neighborhoods, industrial areas, farms...). We want to open a business in an attractive location with high number of daily visitors, so we need to choose carefully the right spots if we want to make the business succeed.
3. The third challenge is to keep an eye on the competition, we know that Montreal has a rich Italian subculture, so we need to open a business that doesn't have a lot of Italian restaurants in its vicinity.

These three challenges can be addressed by exploiting the data provided by Foursquare API.

3-3-Exploiting Foursquare API data to filter out the list of locations

After creating the list of new candidates' positions on the map, we are interested in filtering these spots using Foursquare API (<https://foursquare.com/>)(4)

Actually Foursquare is a local search-and-discovery mobile app/website that provides valuable information of places to go near a user's specific location.

We are exploiting the Foursquare API by providing geographic coordinates (latitude and longitude) of our candidates' positions and getting in return the nearby "venues".

By choosing a radius of the search equal to 200 meters, The API shows the list of venues in that area. There is no risk of duplicating the venues found since all potential locations are 400 meters away of each other.

The Foursquare call returned a total number of 5454 venues locations.

Position number	Neighborhood	Position Latitude	Position Longitude	Position x coordinate	Position y coordinate	Venue	Venue Latitude	Venue Longitude	Venue Category
0	0 Le Plateau-Mont-Royal	45.527454	-73.568953	-5.039888e+06	9.841658e+06	Parc La Fontaine	45.526197	-73.569392	Park
1	0 Le Plateau-Mont-Royal	45.527454	-73.568953	-5.039888e+06	9.841658e+06	Tennis Parc Lafontaine	45.528093	-73.567247	Tennis Court
2	0 Le Plateau-Mont-Royal	45.527454	-73.568953	-5.039888e+06	9.841658e+06	Terrains de Pétanque - Parc Lafontaine	45.527175	-73.569026	Park
3	0 Le Plateau-Mont-Royal	45.527454	-73.568953	-5.039888e+06	9.841658e+06	Théâtre de Guignol du Parc Lafontaine	45.526340	-73.570366	Theater
4	1 Le Plateau-Mont-Royal	45.530026	-73.568826	-5.039488e+06	9.841658e+06	Centre Père Sablon	45.531192	-73.570652	Gym
...
5449	1719 Pierrefonds-Roxboro	45.490104	-73.873586	-5.046729e+06	9.874758e+06	Restaurant Laffamme	45.498918	-73.873395	Pizza Place
5450	1724 Pierrefonds-Roxboro	45.486308	-73.876902	-5.047329e+06	9.875105e+06	La Frite De L'île	45.486534	-73.879347	Burger Joint
5451	1724 Pierrefonds-Roxboro	45.486308	-73.876902	-5.047329e+06	9.875105e+06	Les Délices De L'île	45.486686	-73.878527	Ice Cream Shop
5452	1724 Pierrefonds-Roxboro	45.486308	-73.876902	-5.047329e+06	9.875105e+06	Dépanneur L'intersection	45.486631	-73.878514	Food & Drink Shop
5453	1724 Pierrefonds-Roxboro	45.486308	-73.876902	-5.047329e+06	9.875105e+06	Restaurant & Traiteur SCALO	45.486595	-73.878452	Mediterranean Restaurant

Figure 3 – List of Venues in the Candidates location vicinity

By looking at the venue's categories, we can see that Foursquare website includes any public or private location that can be visited for entertainment purpose, among them the food locations. There is 337 different venue categories in our search result.

First, we will be interested in removing potential locations from the list that has in their vicinity venues categorized under ‘Italian Restaurant’ and ‘Pizza Place’ if we want our business to survive the competition.

By applying these filters, we went from 1725 total candidate location to 947 candidate location by removing locations near an Italian/Pizza place.

Second, a candidate location which has multiple venues nearby means that there is a high chance that this location will be frequently visited by people, this means also that there is a low chance to be a ‘fake’ location (fake locations are the spots that landed on a highway, river water ...). So we will be interested in selecting only locations that has a specific number of nearby venues.

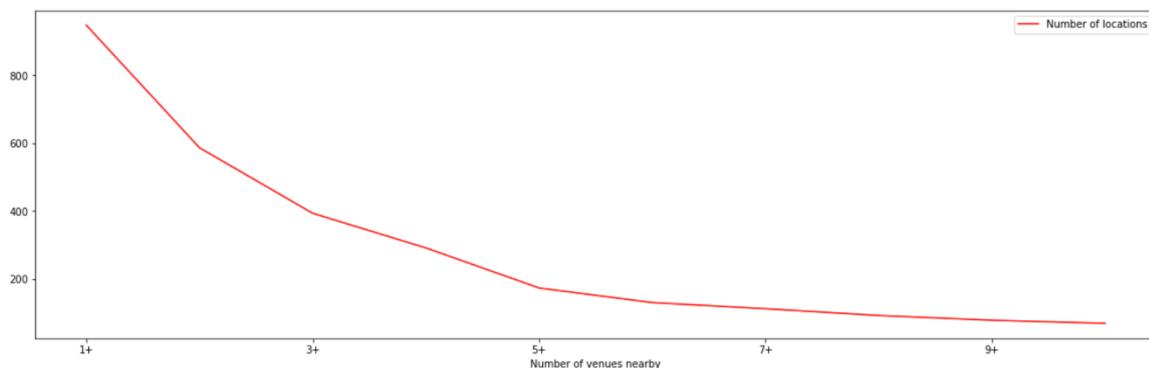


Figure 4 – Number of Candidates location based on the minimum number of nearby venues

We need to define and select the ‘right’ number of minimum nearby venues to identify the spot as attractive, but too much nearby venues has also its drawbacks, it means that the area is very crowded, and the renting price can climb up very high.

We made the decision to go with 4 as minimum number of nearby venues, after applying these filters, the final number of candidate locations went down to 291 locations.

The map below shows us the locations that remained after the filtering. We can see that we succeeded to limit the number of locations and especially remove the ‘fake’ locations.

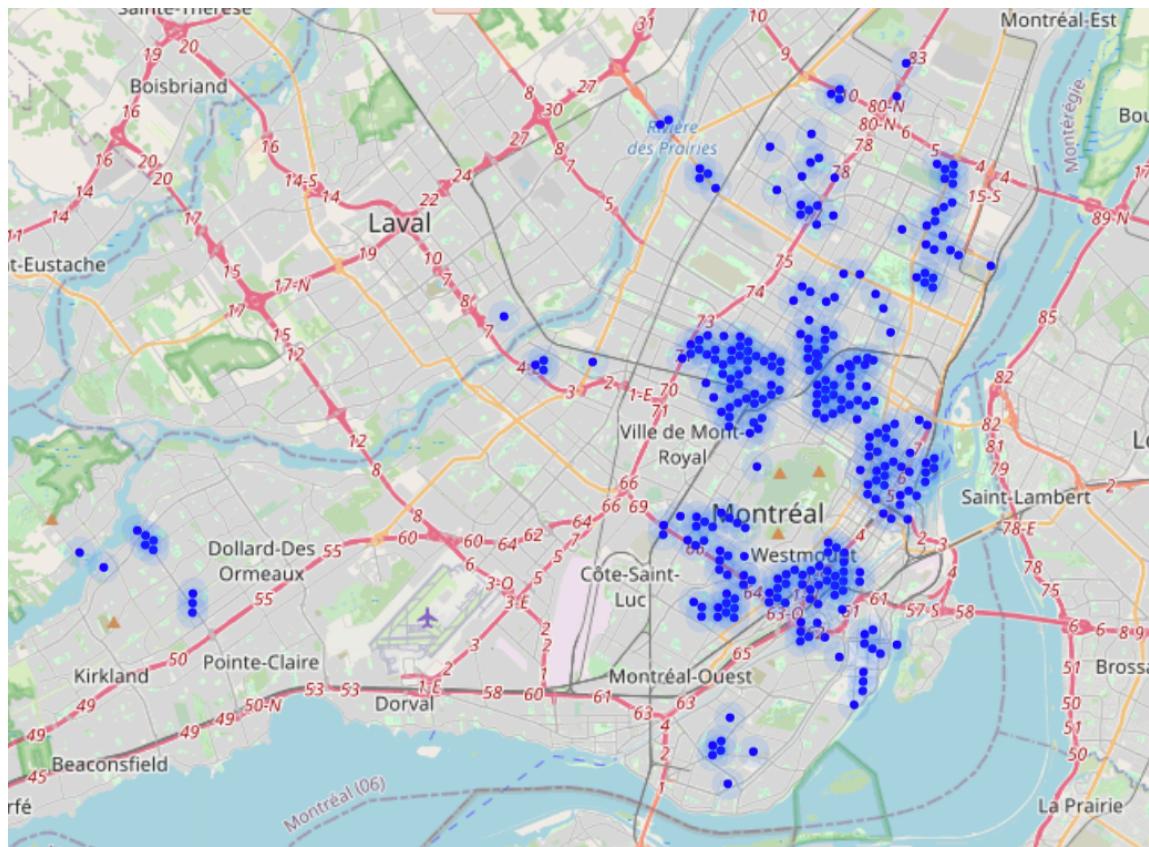


Figure 5 – Potential locations map after filtering

3-4-Returning the list of addresses

Now the list of candidates' locations is shorter, we use Google reverse Geocode(3) (by giving latitude and longitude of a location and getting an address in return) in order to identify the address of these spots on the maps, which is useful in order to visit them in the future.

Position number	Neighborhood	Position_Latitude	Position_Longitude	Position_X	Position_Y	Number_of_venues_nearby	Number_of_Italian_restaurants	Cluster Labels	address
0	Le Plateau-Mont-Royal	45.527454	-73.568953	-5.03988e+06	9.841658e+06	4	0	0	3819 Avenue Calixa-Lavallée, Montréal, QC H2L ...
5	Le Plateau-Mont-Royal	45.526245	-73.572187	-5.040088e+06	9.842005e+06	4	0	0	Unnamed Road, Montréal, QC H2L 0C7, Canada
7	Le Plateau-Mont-Royal	45.531390	-73.571932	-5.039288e+06	9.842005e+06	4	0	0	4330 Avenue Papineau, Montréal, QC H2H 1S9, Ca...
10	Le Plateau-Mont-Royal	45.539109	-73.571549	-5.038088e+06	9.842005e+06	4	0	0	4635 Rue Franchère, Montréal, QC H2H 2K6, Canada
11	Le Plateau-Mont-Royal	45.522464	-73.575547	-5.040688e+06	9.842351e+06	11	0	7	4102 Avenue De Chateaubriand, Montréal, QC H2L...

Figure 6 – List of potential locations after filtering with their respective address

Visiting 291 locations to pick up only one can be very tough, so since the number of these addresses remains high, we need another tool to compare between locations and keep only the most interesting ones.

It is hard to compare between individual locations that are scattered around the map, so we need use Clustering method (machine learning) and deep analysis in order to get more insights by comparing clusters and selecting the ones that fits our criteria's.

4-Clustering and comparison

4-1-Using Machine learning to create Clusters

We will focus on this section on creating clusters and analyzing them in order to identify the best possible spots around Montreal.

We choose to use Kmeans as clustering method with number of clusters equal to 10.

Having a smaller number of clusters won't give a representative distribution, since we want to keep the size of a cluster similar to the size of a Borough.

On the other hand, having a higher clusters number can be an over kill for the analysis.

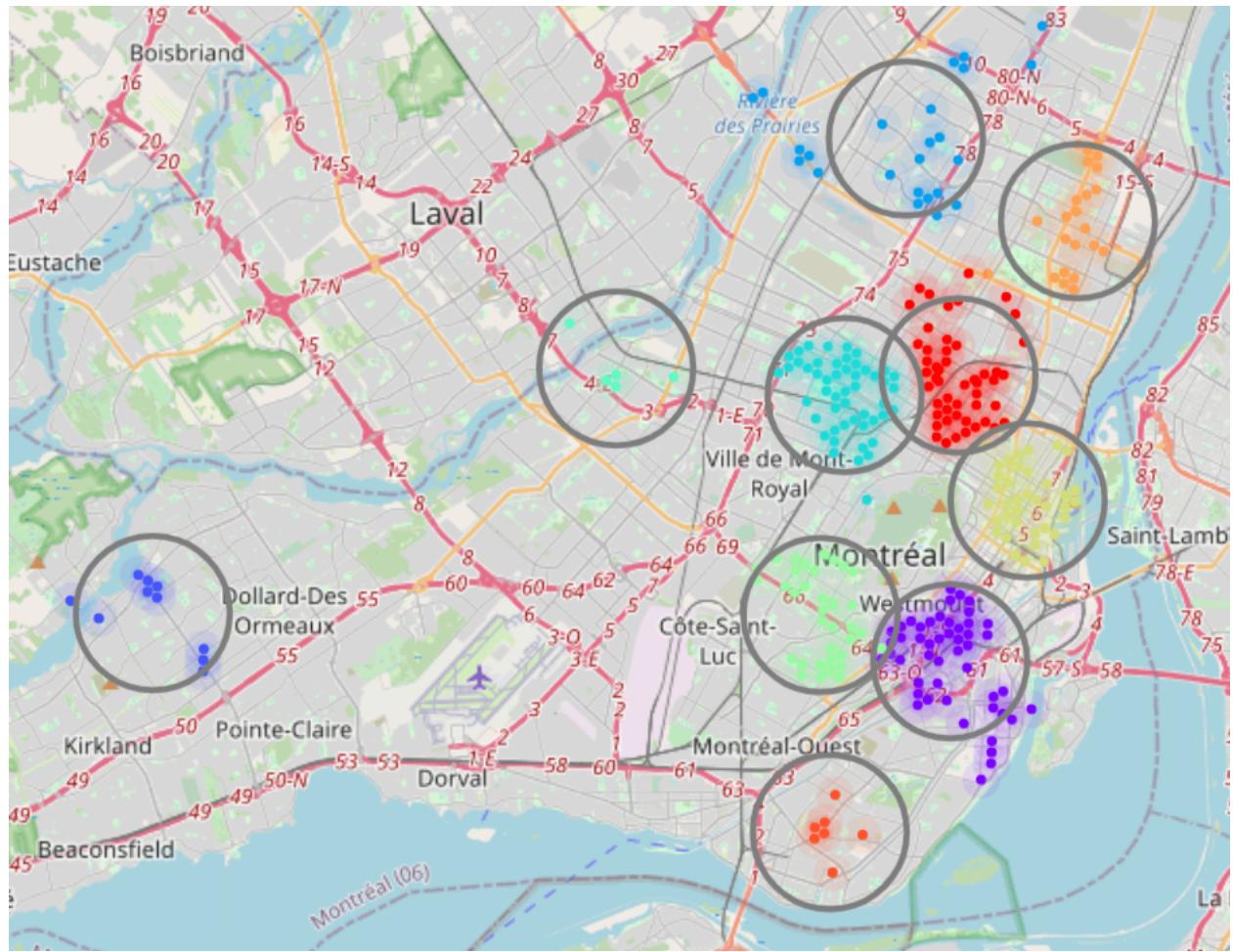


Figure 7 – Potential locations map grouped into 10 clusters

The next step is to compare between these Clusters. We will focus our analysis for the next section on 4 main aspects

- Number of Restaurants per Cluster
- Ratio of Restaurants per 1 million capita
- Number of Italian Restaurant per Cluster
- Number of Parks per Cluster

4-2-Clusters Analysis 1 : Number of Restaurants per Cluster

In this section we started by focusing on the number of restaurants that belongs to each one of the Clusters. In order to achieve that, we kept only the restaurant businesses from the list of venues provided by Foursquare API (Venues that contain key word in their category like: Food, Restaurant). Then, we assigned these restaurants to the clusters (a restaurant belongs to a cluster if it has a distance with the center of cluster less than 3Km).

The next figure shows that Cluster 8 has a high number of restaurants, while Cluster 4 and 6 has very low number of restaurants. We need further information to understand the meaning of these numbers

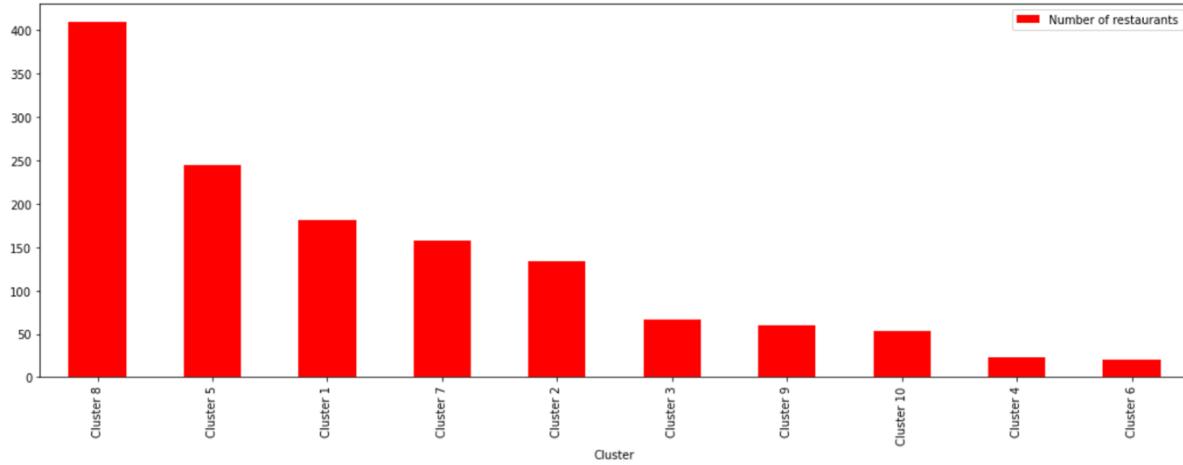


Figure 8 – Ranking of clusters based on the number of restaurants.

One possible way is to see the number of restaurants per capita for each cluster, but since we don't have such information like the population of each cluster because they are made up, we can take a look on the Borough version. Remember that we chose the number of clusters carefully so that they fit the size of Boroughs. There is a high chance that each Cluster represent a Borough on itself.

Our goal is to calculate the number of restaurants per capita for each Borough.

4-3-Clusters Analysis 2 : Ratio of Restaurants per 1 million capita

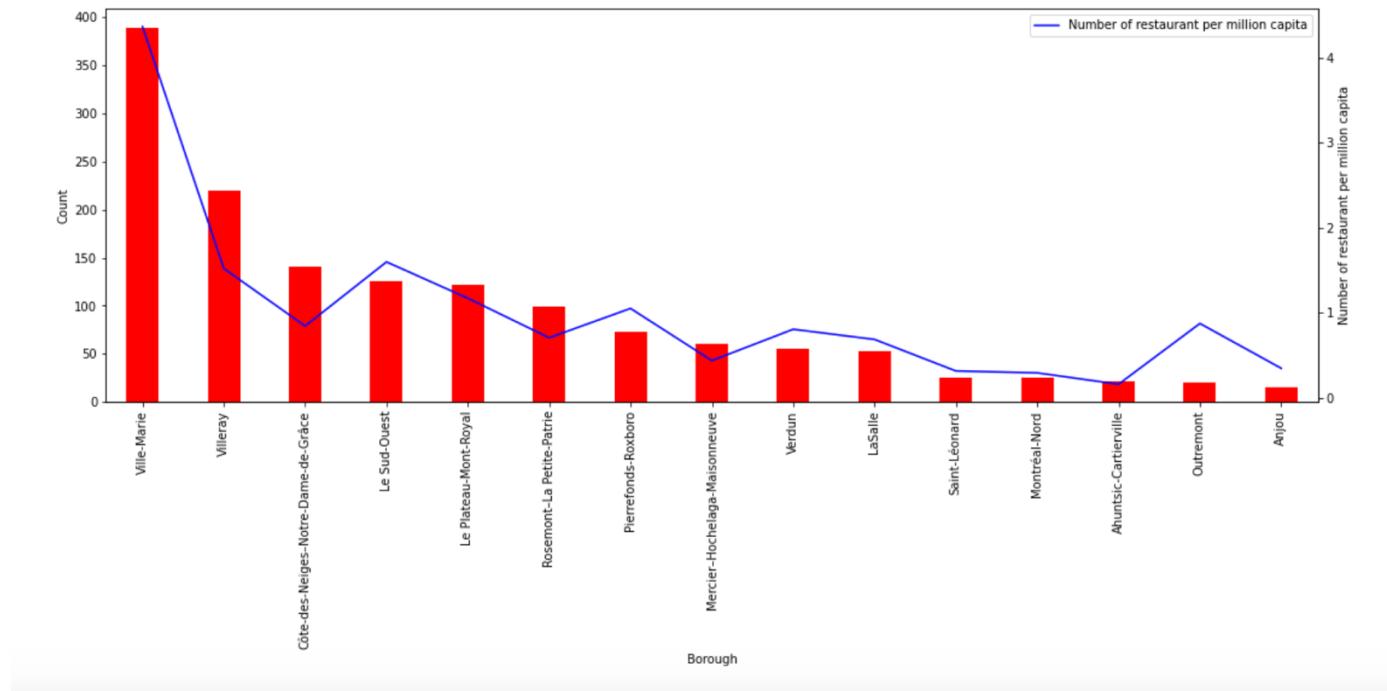


Figure 9 – Representation of Boroughs based on the number of restaurants per 1million habitant.

At a first glance, we can see that “Ville Marie” Borough is leading the group with highest number of restaurants. Actually Ville Marie is home for very important spots in Montreal, mainly Old Port, McGill University, UQAM University... This explains the huge discrepancy with the rest of the Boroughs.

In addition, the ratio in “Ville Marie” Borough is equal to 4 restaurants per million habitant which is very high compared to the rest of Boroughs (they barely surpass a ratio of 1.5).

We notice that Cluster number 8 in our study has 37 candidate locations, from which 36 belongs to Ville Marie Borough, so we recommend to stay away from Cluster 8 if we want to be successful in the future.

Our next strategy is to look for areas with low number of Italian restaurants

4-4-Clusters Analysis 3 : Number of Italian Restaurants

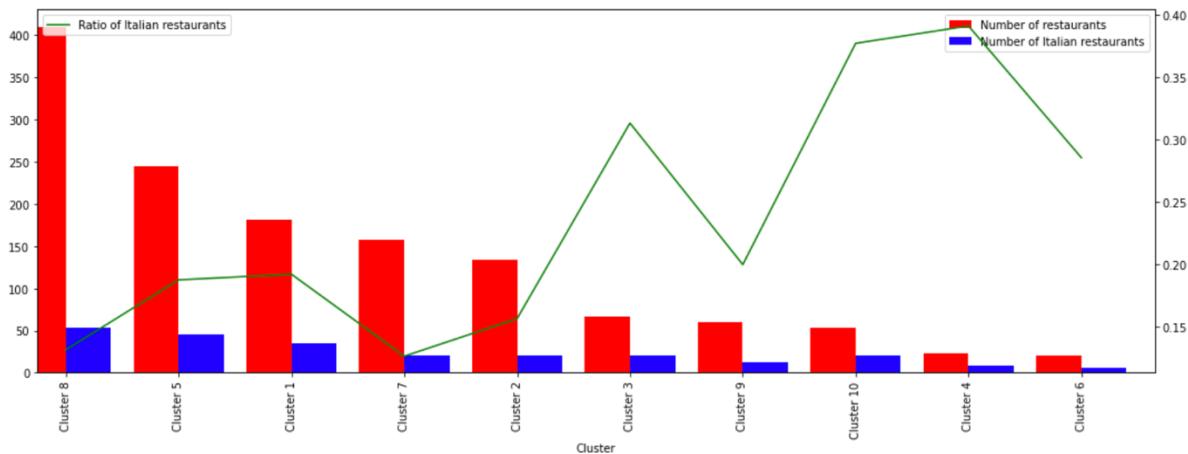


Figure 10 – Percentage of Italian restaurants in each Cluster

There are 69 different food/restaurant venue categories in Montreal based on the Foursquare API results. So we should conclude that the percentage of Italian restaurants/Pizza places should be low in a city that is known for its diverse culture. And yet surprisingly we can see that some Clusters have a very high percentage of Italian restaurants: In 5 Clusters, at least $\frac{1}{4}$ of the restaurants provide Italian food, moreover the Cluster 4 has a ratio of almost 40%.

Starting a business on those areas will be bad decision with the cut-throat competition that exists, so we choose to focus on areas with low percentage of Italian menus, we want to provide value where it is needed the most. We decide to keep our focus on Cluster 5, Cluster 1, Cluster 7 and Cluster 2.

4-5-Clusters Analysis 4 : Number of parks

One of the main activities that Montrealer's love to do is to go to the park and have picnic, so we figured out that it might be interesting to open a business in an area that has a lot of surrounding parks.

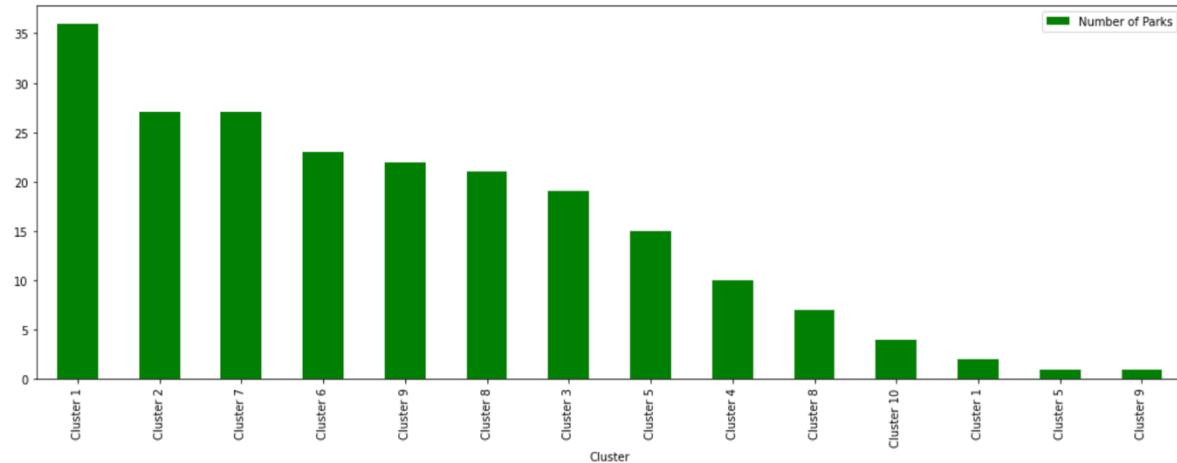


Figure 11 – Number of Parks in each Cluster

Cluster 1, Cluster 2 and Cluster 7 are leading the group with highest number of parks, we can see it more clearly through this map where the parks are represented with a dark green dots

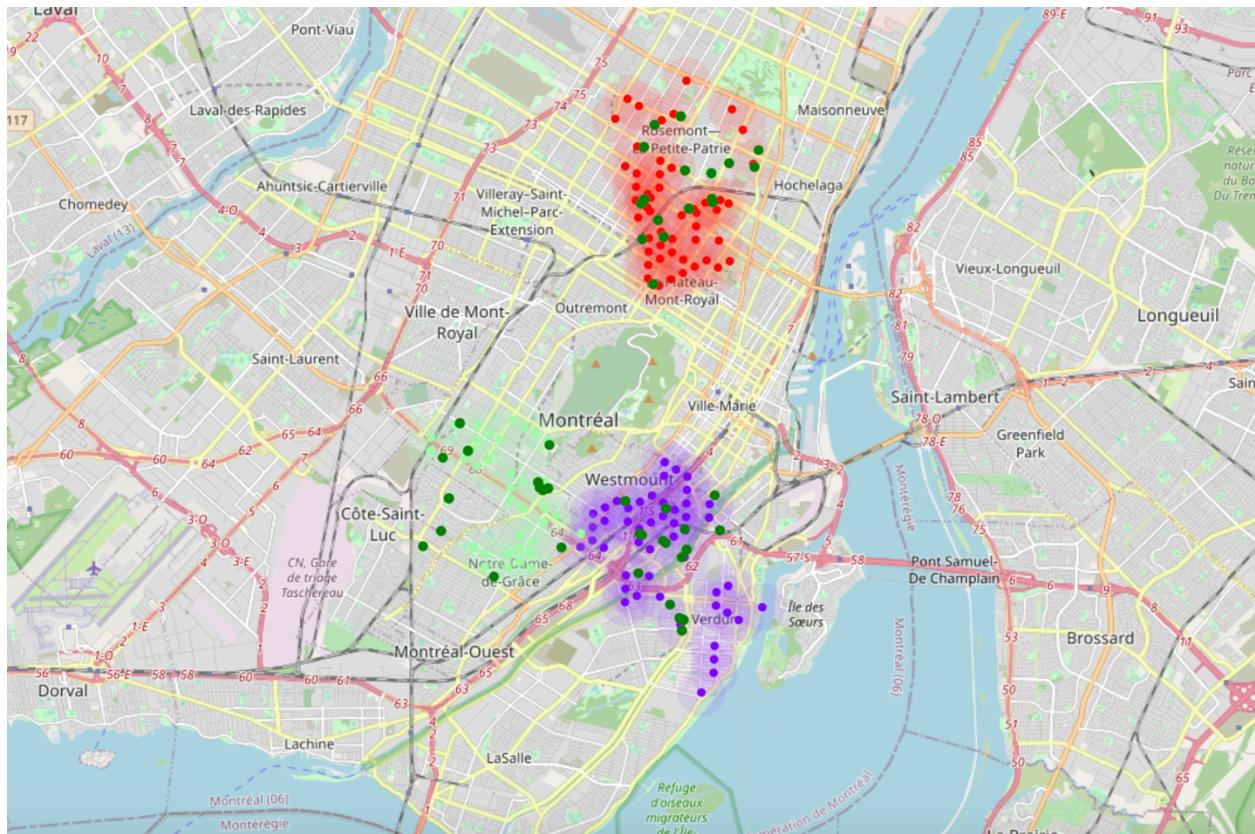


Figure 12 – Number of Parks in each Cluster

4-6-Final results

Three main clusters were identified, these areas covers mainly the Boroughs:

- Le Plateau-Mont-Royal and Rosemont–La Petite-Patrie for Cluster 1
- Le Sud-Ouest and Verdun for Cluster 2
- Côte-des-Neiges–Notre-Dame-de-Grâce for Cluster 7

These are snapshots of the list of addresses

Position number	Neighborhood	Position_Latitude	Position_Longitude	address
0	Le Plateau-Mont-Royal	45.527454	-73.568953	3819 Avenue Calixa-Lavallée, Montréal, QC H2L ...
5	Le Plateau-Mont-Royal	45.526245	-73.572187	Unnamed Road, Montréal, QC H2L 0C7, Canada
7	Le Plateau-Mont-Royal	45.531390	-73.571932	4330 Avenue Papineau, Montréal, QC H2H 1S9, Ca...
10	Le Plateau-Mont-Royal	45.539109	-73.571549	4635 Rue Franchère, Montréal, QC H2H 2K6, Canada
13	Le Plateau-Mont-Royal	45.527609	-73.575293	4333 Rue de la Roche, Montréal, QC H2J 3H8, Ca...
20	Le Plateau-Mont-Royal	45.526400	-73.578526	4379 Rue de Mentana, Montréal, QC H2J 3B4, Canada
22	Le Plateau-Mont-Royal	45.531545	-73.578272	4604 Rue Garnier, Montréal, QC H2J 3S7, Canada
23	Le Plateau-Mont-Royal	45.534118	-73.578145	1694 Rue Gilford, Montréal, QC H2J 1S5, Canada
24	Le Plateau-Mont-Royal	45.536691	-73.578018	1996 Boulevard Saint-Joseph Est, Montréal, QC ...

Figure 13 – List of candidate addresses – Cluster 1

Position number	Neighborhood	Position_Latitude	Position_Longitude	address
541	Verdun	45.461264	-73.560081	20 Allée des Brises du Fleuve, Verdun, QC H4G ...
555	Verdun	45.458850	-73.566533	205 1re Avenue, Verdun, QC H4G 2V3, Canada
564	Verdun	45.460213	-73.569631	402 Rue Rielle, Verdun, QC H4G 2S7, Canada
566	Verdun	45.465352	-73.569376	3865 Rue Gertrude, Verdun, QC H4G 1R6, Canada
568	Verdun	45.448729	-73.573366	621 Rue Richard, Verdun, QC H4H 2A5, Canada
569	Verdun	45.451298	-73.573239	686 Rue Osborne, Verdun, QC H4H 1X2, Canada
570	Verdun	45.453867	-73.573112	646 Rue Melrose, Verdun, QC H4H 1T4, Canada
573	Verdun	45.461575	-73.572730	667 Rue Gordon, Verdun, QC H4G 2R7, Canada
574	Verdun	45.464145	-73.572603	3977 Rue Joseph, Verdun, QC H4G 1J5, Canada
575	Verdun	45.444953	-73.576717	915 Rue Rolland, Verdun, QC H4H 2E9, Canada

Figure 14 – List of candidate addresses – Cluster 2

Position number	Neighborhood	Position_Latitude	Position_Longitude	address
302	Côte-des-Neiges–Notre-Dame-de-Grâce	45.478776	-73.616409	4111 Avenue Northcliffe, Montréal, QC H4A 3L2, Canada
310	Côte-des-Neiges–Notre-Dame-de-Grâce	45.477566	-73.619635	5380 Av Brodeur, Montréal, QC H4A 1J3, Canada
311	Côte-des-Neiges–Notre-Dame-de-Grâce	45.480136	-73.619512	4371 Boulevard Décarie, Montréal, QC H4A 3K4, Canada
313	Côte-des-Neiges–Notre-Dame-de-Grâce	45.485278	-73.619265	4728 Chemin Circle, Montréal, QC H3W 2C5, Canada
316	Côte-des-Neiges–Notre-Dame-de-Grâce	45.492990	-73.618896	3840 Chemin Queen Mary, Montréal, QC H3V 1A6, Canada
317	Côte-des-Neiges–Notre-Dame-de-Grâce	45.468645	-73.623229	6081 Avenue Notre-Dame-de-Grâce, Montréal, QC H4A 3L2, Canada
318	Côte-des-Neiges–Notre-Dame-de-Grâce	45.471215	-73.623106	4099 Avenue Royal, Montréal, QC H4A 2M4, Canada
319	Côte-des-Neiges–Notre-Dame-de-Grâce	45.473785	-73.622983	4220 Av Harvard, Montréal, QC H4A 2W7, Canada
327	Côte-des-Neiges–Notre-Dame-de-Grâce	45.494351	-73.622000	3800 Rue Jean-Brillant, Montréal, QC H3T 1P1, Canada
329	Côte-des-Neiges–Notre-Dame-de-Grâce	45.470005	-73.626331	4223 Avenue Hingston, Montréal, QC H4A 2J6, Canada

Figure 15 – List of candidate addresses – Cluster 7

5-Conclusion

Purpose of this project was to help business owners understand the distribution of the restaurant industry in the island of Montreal and provide the best possible spots to open a new business that serves Italian Pizzas.

We chose in our study to focus on highly attractive areas (High population density, lot of businesses, close parks...) and low risk of being run over by the competition (we selected areas that doesn't have any Italian restaurants in their surroundings).

We started the study by creating 1725 potential candidates locations around the city most dense areas, and we ended up narrowing the data until we reached a number of 134 locations grouped into 3 clusters.

Stakeholders can rely on this study and bring an additional layer of analysis to narrow down even more the list of locations using different criterias. An example would be to rely on the proximity to the metro stations and the bus stops that exists in the city. Another possible scenario of analysis is to study the distribution of the renting price of the real estate for the selected zones.

We explored during the study the information shared by the Foursquare platform, which consisted mainly of the geographic location of a venue and its category. But Foursquare offers a wider range of information that can be explored even more, like Venues rating, opening hours, menu items, prices, services and features... This begs to be exploited in a context of a Machine Learning Classification Model, by providing the list of features of a new restaurants (for example providing the list of services, prices, features ...) and see how the model would predict the success rate of these restaurant based on the visitor rating.

6-References

<https://www.geosmontreal.com/visiting-montreal/10-facts-about-montreal.htm>

https://en.wikipedia.org/wiki/Boroughs_of_Montreal

<https://developers.google.com/maps/documentation/geocoding/overview>

<https://foursquare.com/>