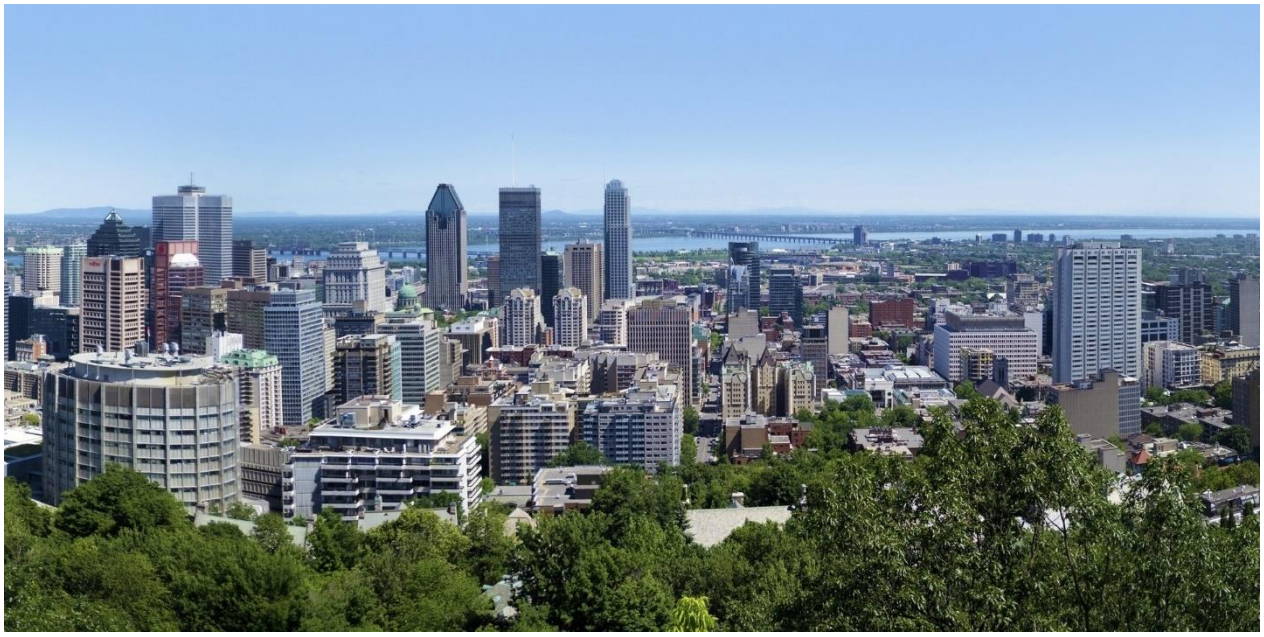
Where are the restaurants in Montréal ?



By : Thierry Lévesque

On : July 21, 2021

As part of : IBM Data Science Certification – Capstone Project

Introduction

Whether you are visiting Montréal or are interested in opening a restaurant, it is good to know where the current restaurants are. Montréal has a lot to offer in terms of nightlife, cafés and restaurant. But where to start if you don't know the city ?

As part of the IBM Data Science Certification, students like me have to do a capstone project leveraging the Foursquare API. The API makes it possible to know a lot of information about venues around a point of interest. We can therefore make a list of the venues in the Montréal neighborhoods and make maps and charts to show where to go to find what you are looking for.

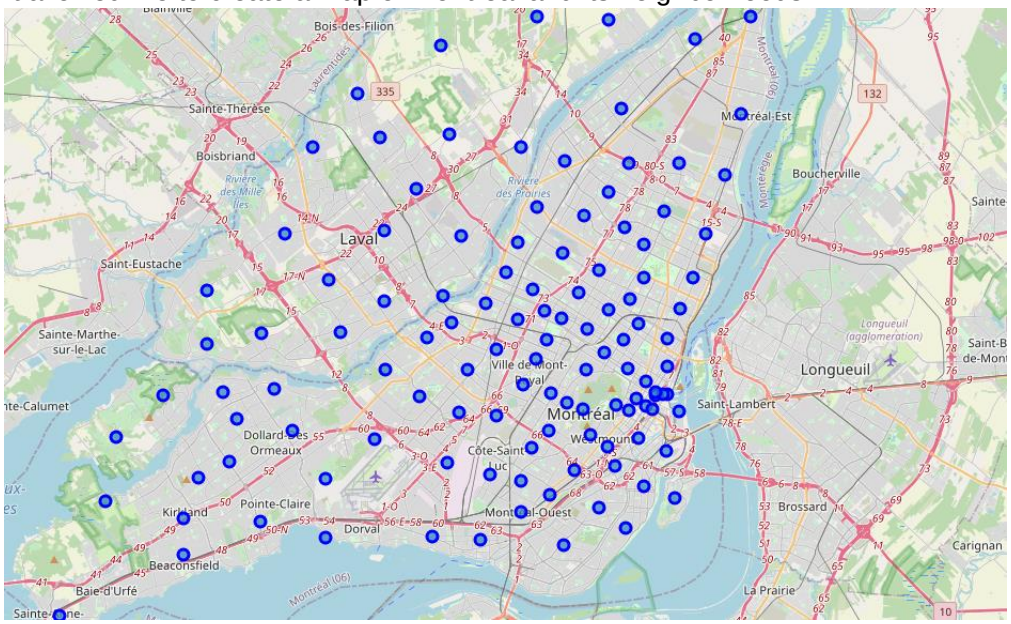
Data Gathering

First of all, I used a list of postal codes and the associated neighborhood found on the Wikipedia page : https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H. After a little bit of data manipulation, I've been able to properly format the data for my analysis.

With those informations, I was able to find the coordinates of each neighborhood using "pgeocode", a Python librairy built for finding informations from postal codes, such as coordinates, municipality name and region name :

	Neighborhood	PostalCode	Latitude	Longitude
0	Akwesasne	H0M	45.6986	-73.5025
1	Pointe-aux-Trembles	H1A	45.6753	-73.5016
2	Montreal	H1B	45.6320	-73.5075
3	Rivière-des-Prairies	H1C	45.6656	-73.5367
4	Rivière-des-Prairies	H1E	45.6342	-73.5842
5	Montréal-Nord	H1G	45.6109	-73.6211
6	Montréal-Nord	H1H	45.5899	-73.6389
7	Anjou	H1J	45.6097	-73.5794
8	Anjou	H1K	45.6097	-73.5472
9	Mercier	H1L	45.6043	-73.5178

It allowed me to create a map of Montréal and its neighborhoods :



Having a database of the neighborhoods and of the associated coordinates allows me to work with the Foursquare API. The free version of the API makes it possible to know the venues around a point of interest according to its coordinates. I've then been able to built the following dataset :

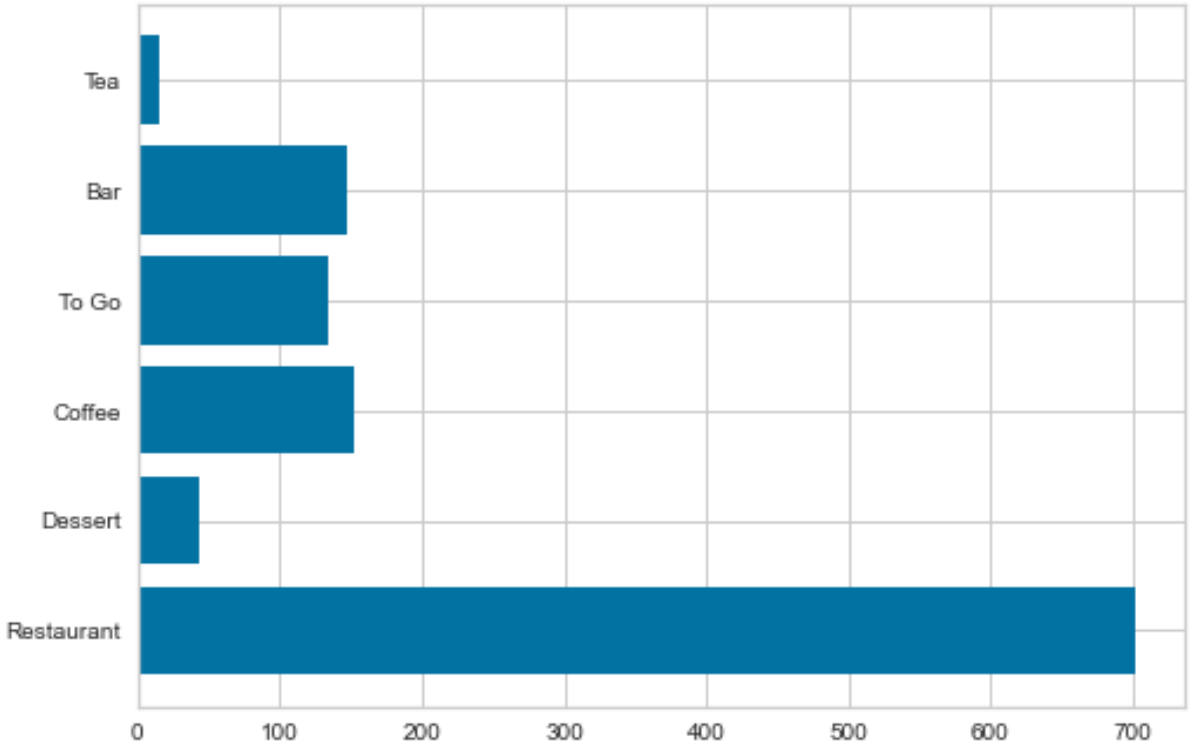
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Akwesasne	45.6986	-73.5025	Site Historique National De La Bataille De La ...	45.699499	-73.505149	Historic Site
1	Pointe-aux-Trembles	45.6753	-73.5016	Parc-nature de la Pointe-aux-Prairies	45.678834	-73.501162	Park
2	Pointe-aux-Trembles	45.6753	-73.5016	Uniprix	45.677521	-73.502226	Pharmacy
3	Pointe-aux-Trembles	45.6753	-73.5016	AMT Gare Pointe-aux-Trembles	45.674882	-73.504908	Train Station
4	Pointe-aux-Trembles	45.6753	-73.5016	Parc Yves-Thériault	45.678675	-73.502037	Park
...	...	...	...	...	...	...	...
1919	Sainte-Anne-De-Bellevue	45.4062	-73.9456	Tandoori Bellevue	45.403280	-73.950303	Indian Restaurant
1920	Sainte-Anne-De-Bellevue	45.4062	-73.9456	Munch Box John Abbott College	45.405680	-73.941787	Fast Food Restaurant
1921	Sainte-Anne-De-Bellevue	45.4062	-73.9456	AMT Gare Sainte-Anne-de-Bellevue	45.407761	-73.950959	Train Station
1922	Sainte-Anne-De-Bellevue	45.4062	-73.9456	Resto Pub Bord'Eaux	45.403015	-73.949254	Gastropub
1923	Sainte-Anne-De-Bellevue	45.4062	-73.9456	Couche-Tard	45.402770	-73.949529	Convenience Store

Data Manipulation

I was interested in grouping the venues by their categories. The problem I had with the dataset is that it contained 260 unique categories and I found it to be too much for a simple analysis such as mine. I thought that I could group some categories together to have all the restaurants in the same category, instead of having a separate one for 'Indian Restaurant' and 'Fast Food Restaurant' for example.

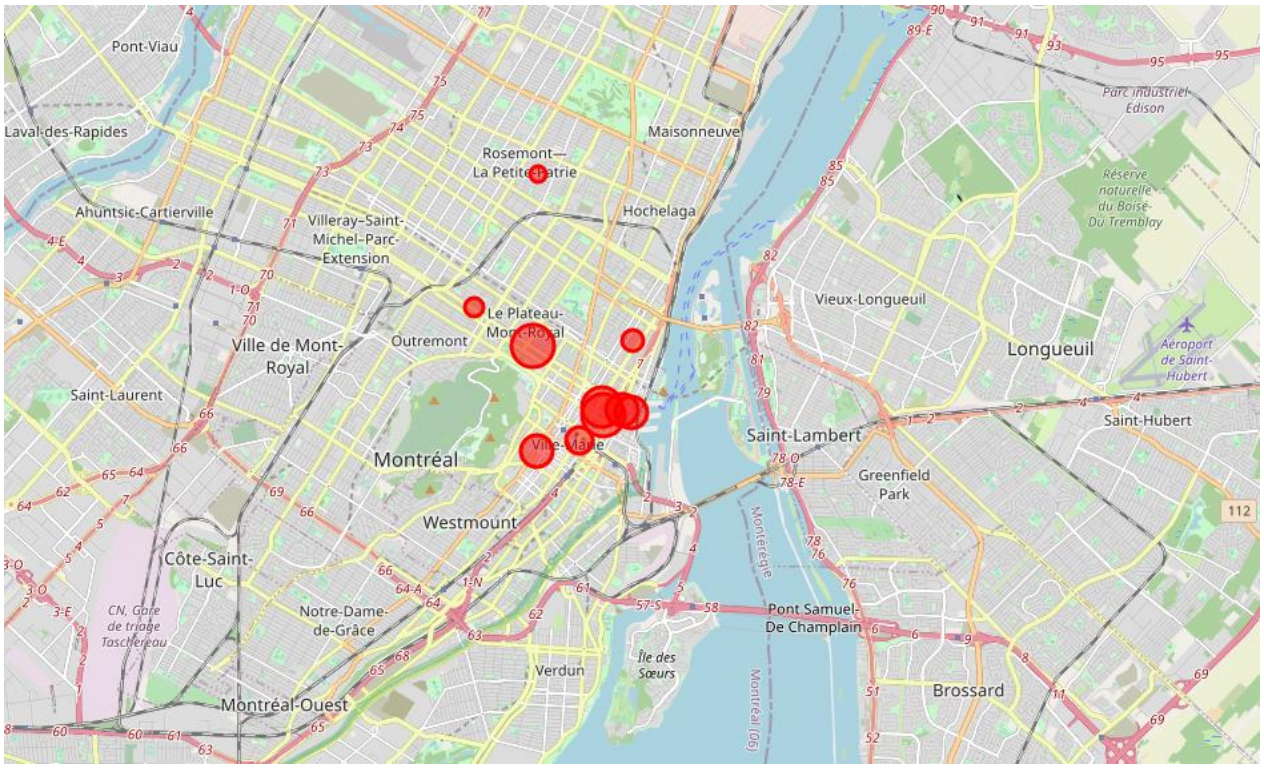
I decided to make a list of venue categories to group together to make my own categories. I went with 6 different categories : restaurants, dessert places, coffee shops, places where you buy food to go, bars and tea places.

The following graphs shows the size of each categories :

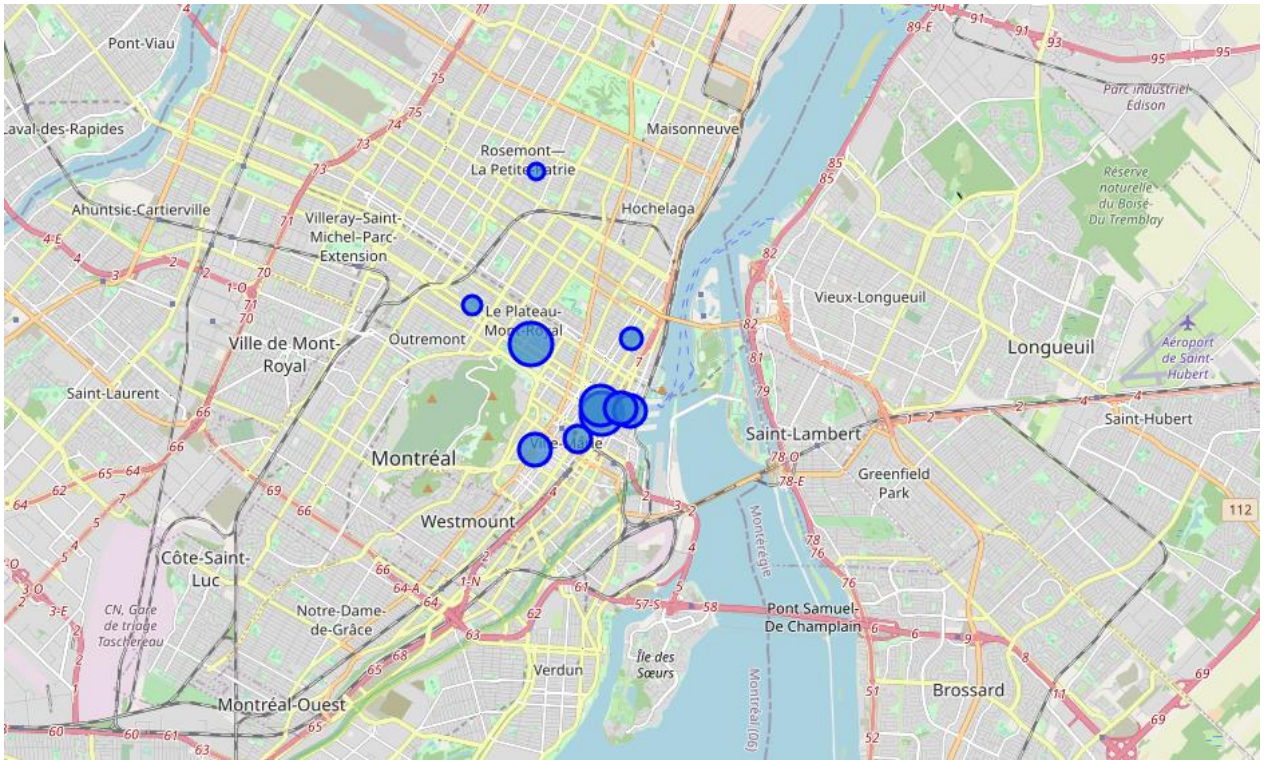


Mapping the top 10 neighborhoods for each categories

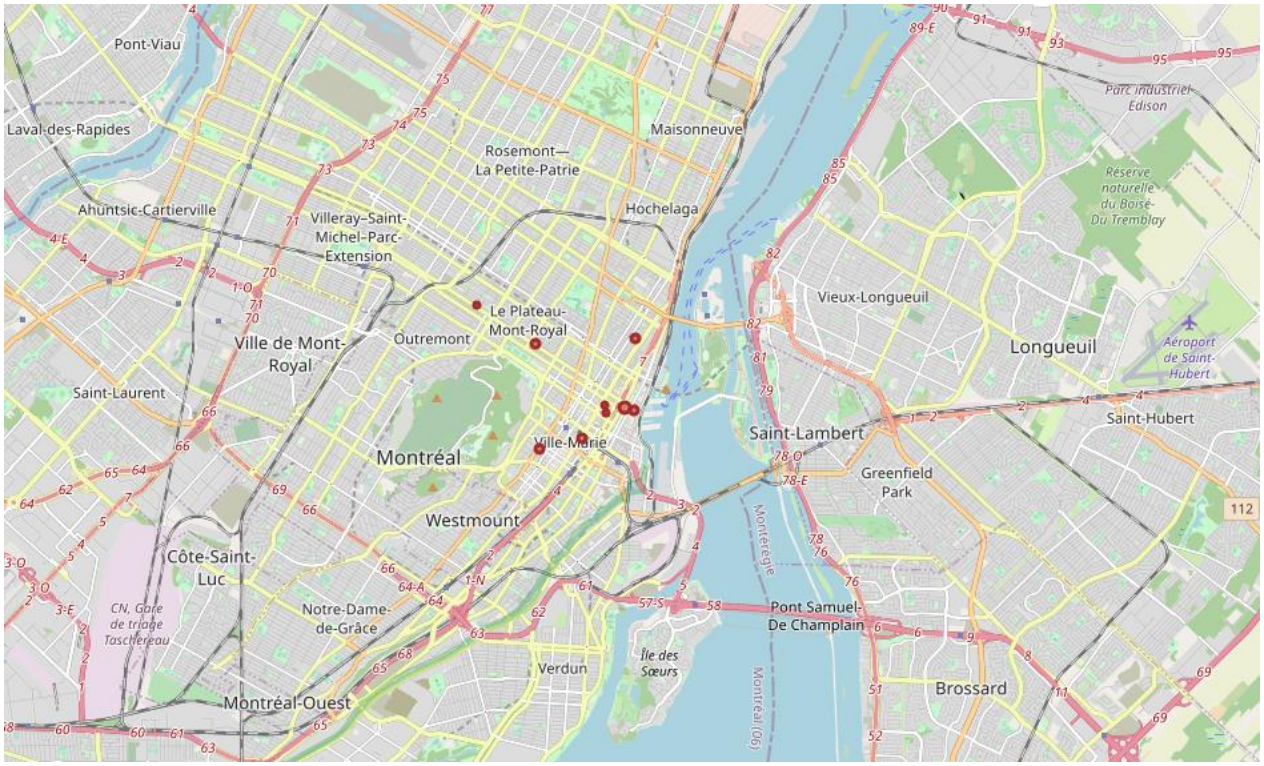
Restaurants



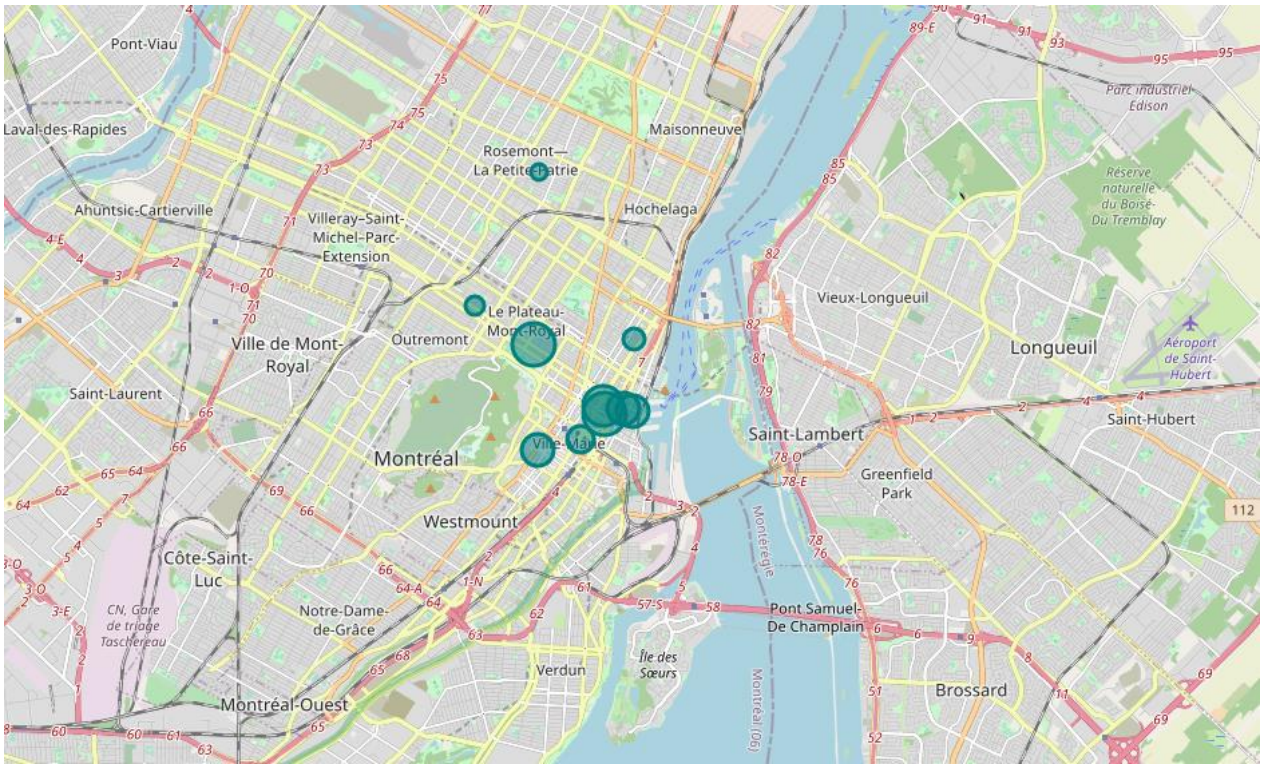
Dessert places



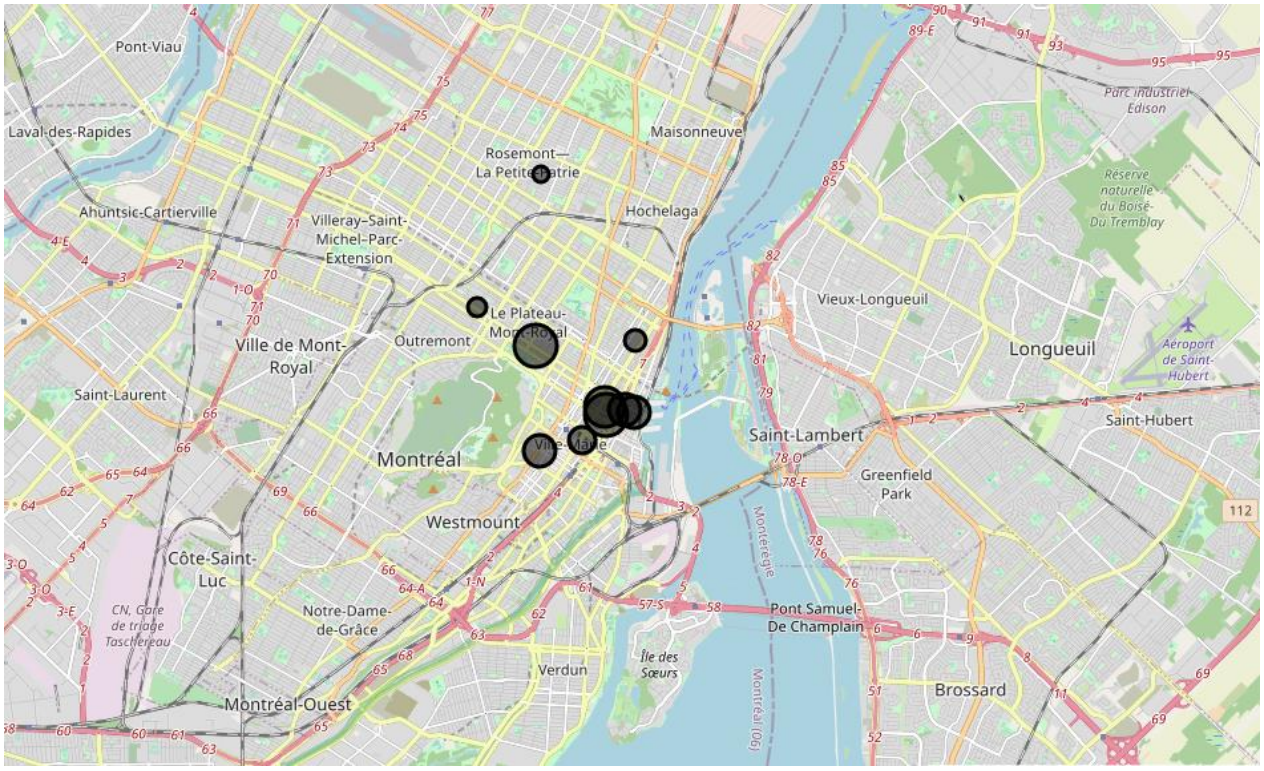
Coffee shops



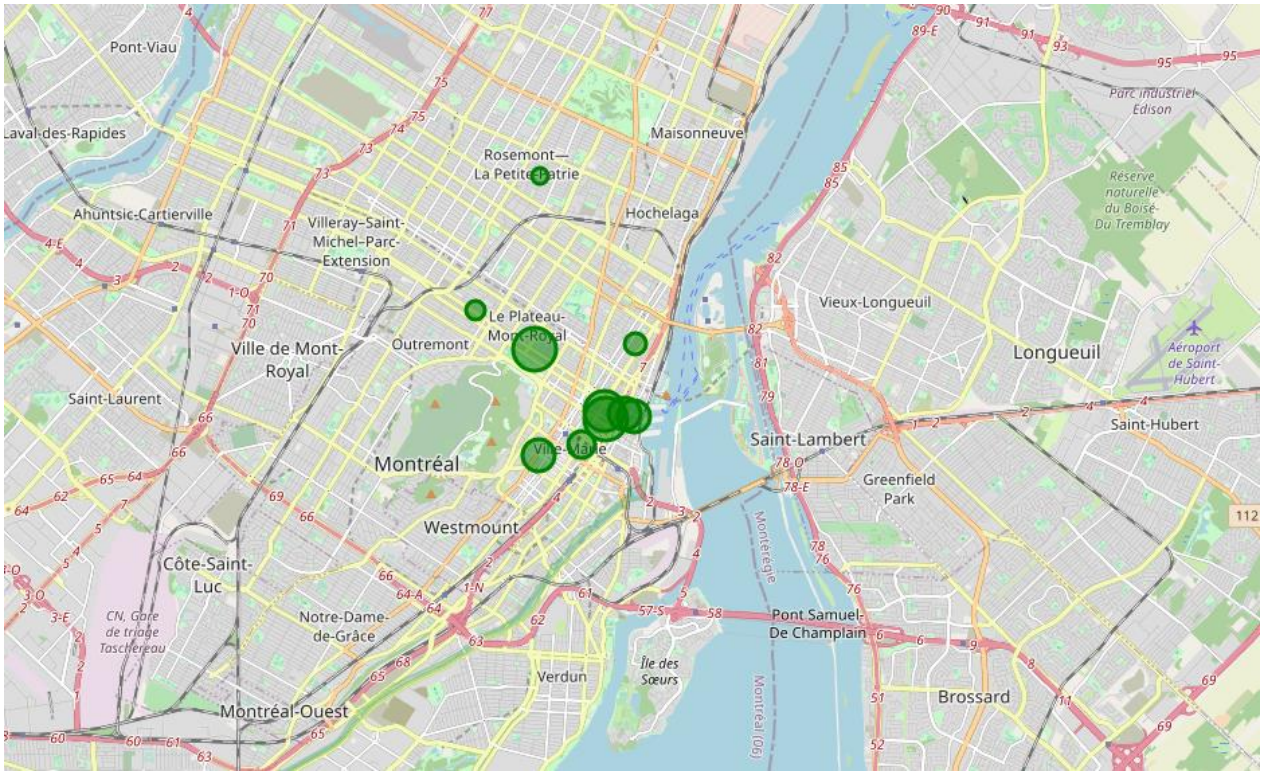
Food to go



Bars



Tea places



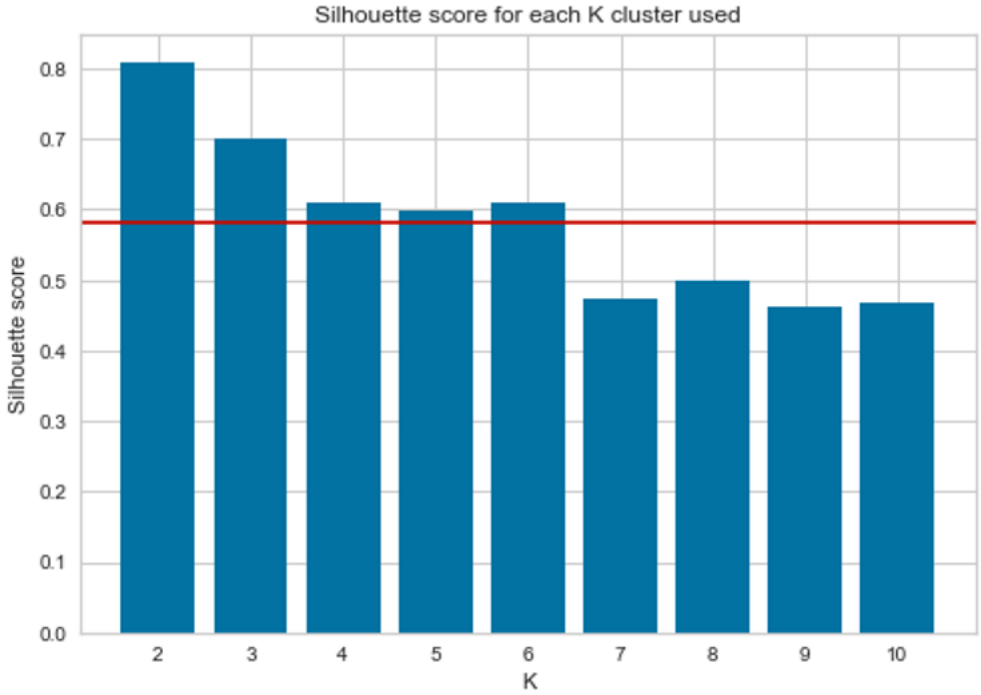
It seems that the best neighborhoods for each categories are always around the Plateau Mont-Royal and Downtown Montréal.

Training of the clustering algorithm

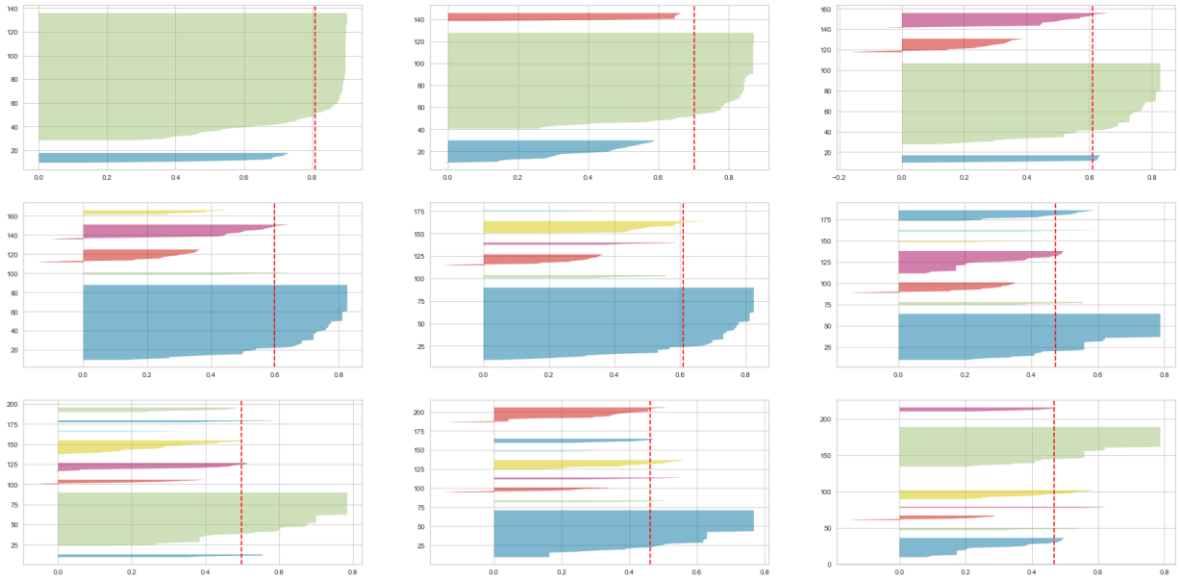
It's useful to cluster the neighborhoods because it helps us understand which neighborhoods are alike and which are not. It's the kind of information a business person would like to know in order to open a restaurant, or any other kind of venue, at the right location because it helps to know how many neighborhoods could be considered and which is it.

The clustering algorithm that I chose is KMeans clustering, a popular algorithm that is easy to use and which gives good results without too much hyperparametering. I trained the algorithm with the 6 venue categories that I previously had (restaurant, dessert place, coffee shop, food to go, bar and tea place) but the result was not good enough because of how few tea places and dessert places there is in Montréal. I then dropped these two categories and had much better clusters.

In order to find the best amount of cluster to use for the algorithm, I went with the silhouette score. The graph below shows with the blue bars the score for each K used and the red line shows the average of the scores :



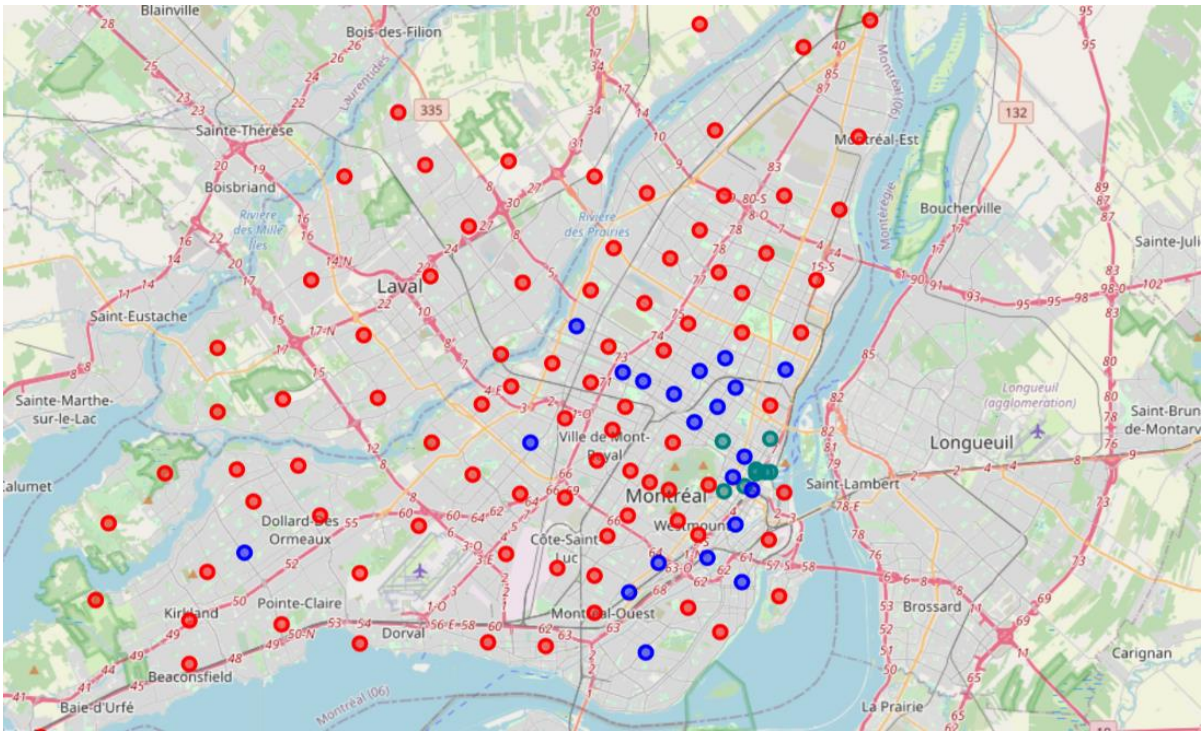
To have a good amount of K, we should use a K value that has a silhouette score at least above average, which eliminated a K value above 6. Next, we look a the plots below to find a K value where there is the least negative scores and where the clusters seems to give the most information :



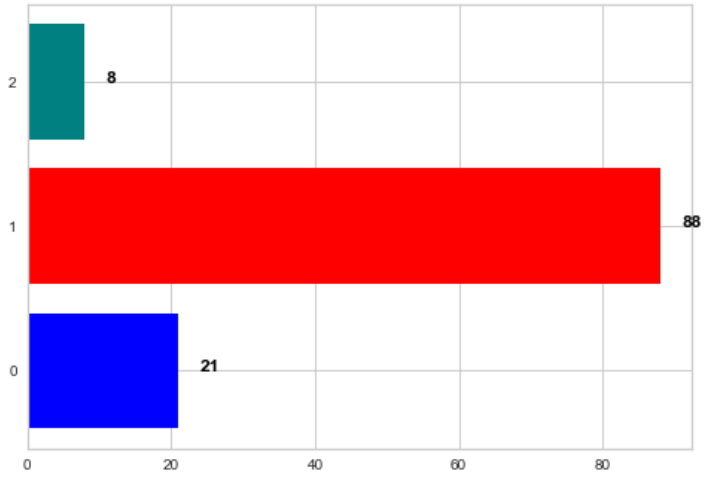
I went with 3 clusters but I could have went with 6. What made me decide between those 2 is that I wanted the result to be simple and easy to understand. If I've wanted more details and/or aggregate some clusters together, I could have went with 6.

Results

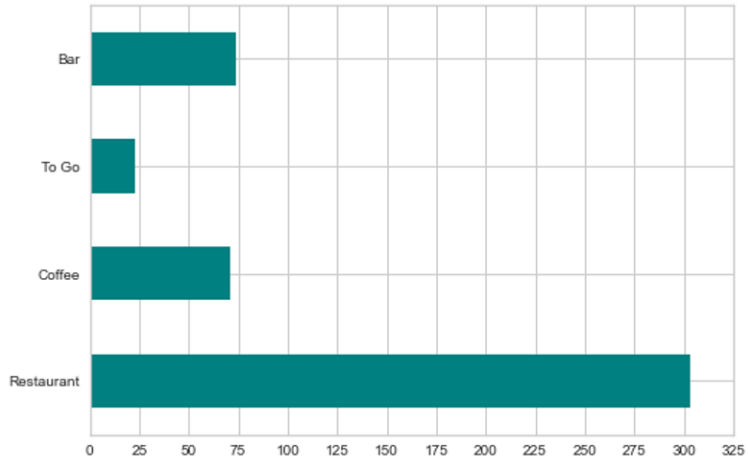
How the neighborhoods have been clustered by the algorithm.

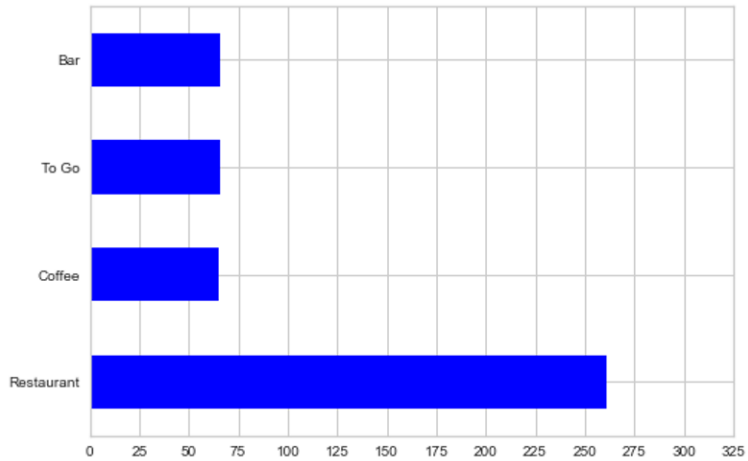


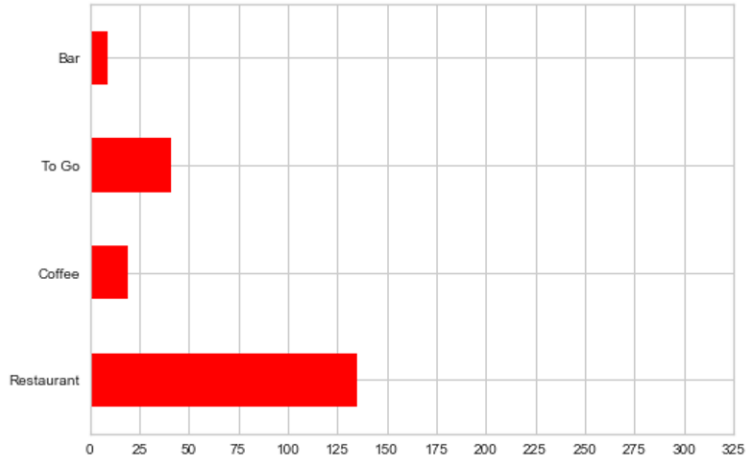
The amount of neighborhoods in each clusters :



How many venues of each categories in each clusters :







Discussion

I liked the results of the analysis but I could have changed a few things. First of all, the neighborhoods where divided by the postal codes and not by their names. It helped me get the coordinates right using the geolocator but it may be confusing for someone who looks at the results without knowing the details of the methodology. In the conclusion section, there is Downtown Montréal 3 times in the list of neighborhoods because I treated them as being 3 different neighborhoods. The result shown is good, but may be confusing.

Another thing that I could have looked at in more details is what is the best radius to use when querying the Foursquare API. I wonder if I have all the venues in Montréal in my dataset and if some may be there more than once. Maybe I could have went differently about that, but I sticked with the same methodology that we used with the New-York and Toronto examples.

Conclusion

Out of 117 neighborhoods, the 8 that are in the green cluster have nearly half of the restaurants of Montréal. This cluster have the most restaurants, bars and coffee shop of the three ! The map also shows all the neighborhoods of the cluster are very close to one another. If you are looking for a restaurant in Montréal, the list below should guide you very well !

	Restaurant	Coffee	To Go	Bar
Neighborhood				
Centre-Sud	25	9	3	15
Downtown Montreal	31	10	1	5
Downtown Montreal	47	7	2	6
Downtown Montreal (Concordia University)	37	9	3	13
Old Montreal	36	9	4	9
Place Desjardins	44	6	3	5
Plateau Mont-Royal Central	47	9	4	13
Tour de la Bourse	36	12	3	8