

Explore the Neighborhood and Venues of Montreal

Peng Seng Kuok

October 23, 2020

1. Introduction

The city of Montreal is the second most populous city in Canada, just behind the city of Toronto. It is the most important city in the province of Quebec. Montreal is a place that is warm, dynamic, relaxed, innovative, cosmopolitan, modern and historic. In the city, we live there in bilingualism and multiculturalism. It is undoubtedly the most bilingual city in North America where more than 50% of the “Montrealers” are bilingual. Less pretentious than Paris, less busy than New York, more creative than Toronto, Montreal has kept its own identity without influencing by the international standard. The unique art of living that combines good humor, accessibility, cosmopolitanism and culture is a perfect mix that inevitably makes you want to spend at least few days in this city.

2. Business Problem

Similar to various part in the world, due to the COVID-19 pandemic, most of the people in the world are restricted to travelling. The objective of the project is to use data science skills to allow the readers to explore the beautiful city of Montreal virtually. With the hope that, the readers will gain interest of the city throughout the project and will explore the city of Montreal in person when the pandemic end (hopefully, very soon).

In the end, a K-Means Clustering analysis is performed to group the neighborhoods into clusters. This will be particularly useful to determine how much days the readers should spend in Montreal if one decides travel to Montreal (i.e. one can spend 1 day per cluster).

3. Data Description

The following data is used for the analysis:

- List of FSA with the corresponding neighborhood in the city of Montreal [1].
- Forsquare API to get the most common venues of given Neighborhood of Montreal [2].
- Geopy package to get the latitude and the longitude coordinates of each neighborhood.

4. Methodology

4.1 Data Cleaning and Dataset Exploration

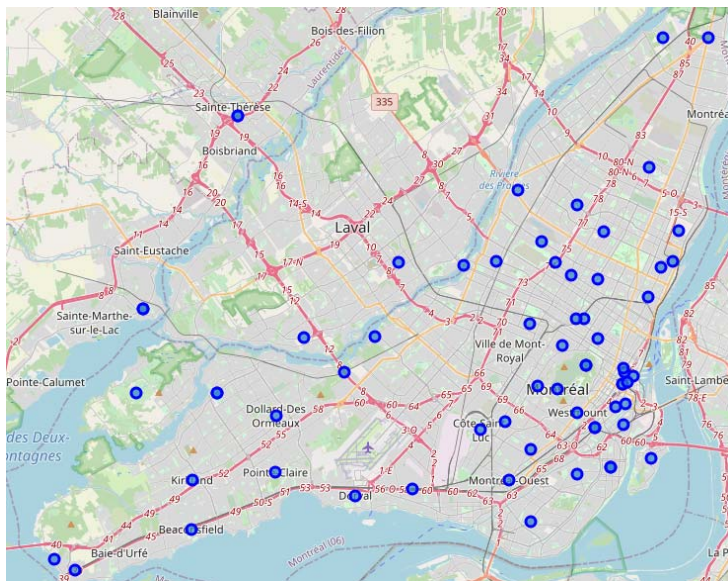
As a starting point, a list of Montreal's neighborhoods with the corresponding postal code is extracted from Wikipedia web page with the web scrapping technic from BeautifulSoup package.

The head of the postal code table is as follow:

	Postcode	Neighborhood
1	H1A	Pointe-aux-Trembles
2	H2A	Saint-Michel
3	H3A	Downtown Montreal
4	H4A	Notre-Dame-de-Grâce
5	H5A	Place Bonaventure

Once the list of postal code with the corresponding neighborhood is extracted from Wikipedia, the longitude and latitude of each neighborhood is obtained using the Geopy package. Then, a map of Montreal with neighborhoods superimposed on top is created using Folium package.

The resulting map is as follow:



Then, Foursquare API is used to explore the neighborhood of Downtown Montreal by setting limit of venues to 100 and a radius of 500 meters.

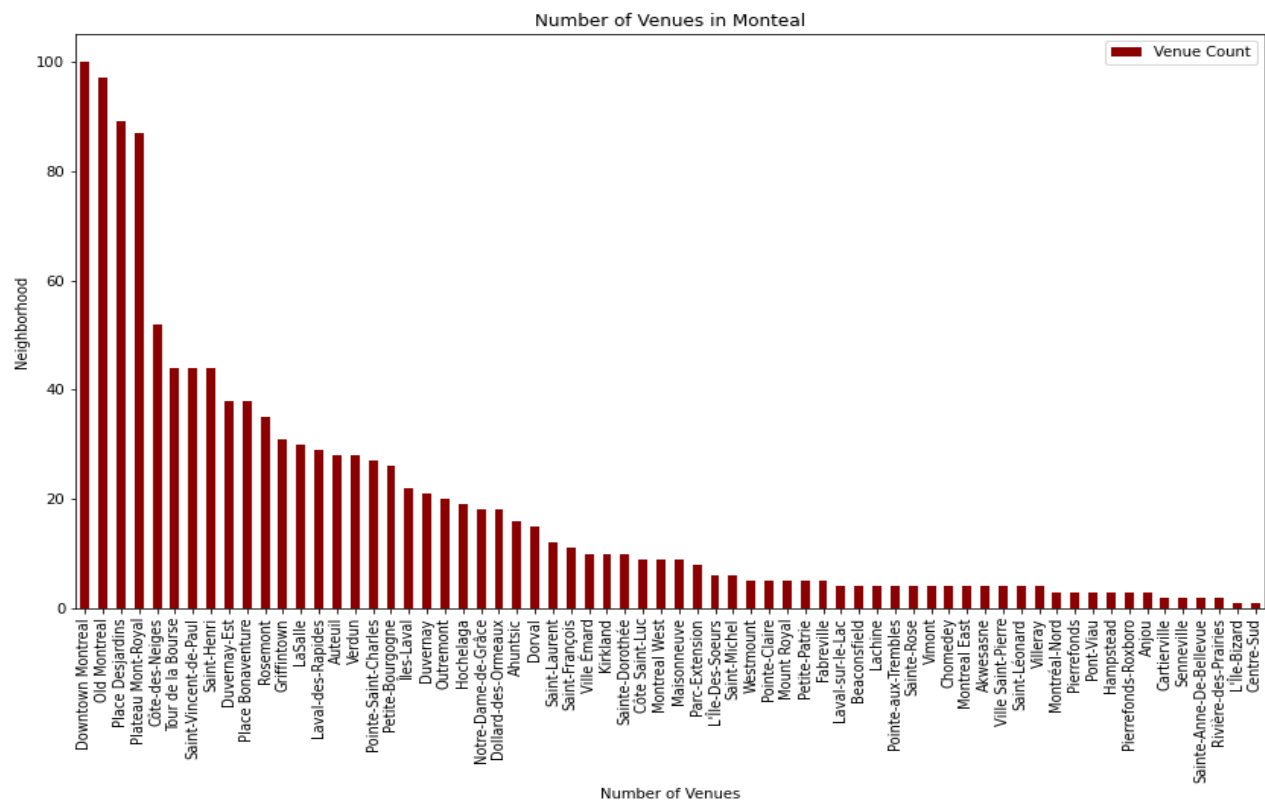
The head of the venues in Downtown Montreal table is as follow:

	name	categories	lat	Ing
0	La Finca – café & bureau	Café	45.504520	-73.563116
1	Courtyard Montreal Downtown	Hotel	45.505269	-73.564198
2	NYKS Bistro Pub	French Restaurant	45.506324	-73.566382
3	Monsieur Restaurant + Bar	Restaurant	45.504703	-73.563613
4	Tiradito	Peruvian Restaurant	45.504544	-73.563405

4.2 Explore Neighborhoods in Montreal

The same exercise is done for all the neighborhood in Montreal. Downtown Montreal is the only neighborhood where it reached the limit of 100 venues. On the other hand, there are a significant amount of neighborhoods (35 neighborhoods) where there are only a few venues (less than 10).

The number of the venues in Montreal's neighborhood are shown in the following bar chart:



As shown from the bar graph, 35 neighborhoods where the number of venues are low which indicate that those neighborhoods are suburb areas. For the purpose of the analysis, those neighborhoods will be ignored.

In total, Foursquare API have extracted 206 uniques categories in the city of the Montreal.

4.3 Analyze Each Neighborhood

The 10 most common venues are extracted by neighborhood. This will be useful for tourists to plan for their trip in Montreal.

The head of the 10 most common venues by neighborhood table is as follow:

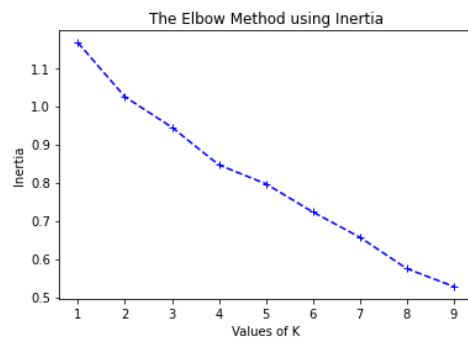
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ahuntsic	Restaurant	Café	Park	Breakfast Spot	Coffee Shop	Fast Food Restaurant	Tea Room	Ice Cream Shop	Pharmacy	Gastropub
1	Auteuil	Breakfast Spot	Grocery Store	Indian Restaurant	Pizza Place	Food	Tea Room	Sandwich Place	Fish & Chips Shop	Café	French Restaurant
2	Côte-des-Neiges	Coffee Shop	Middle Eastern Restaurant	Fast Food Restaurant	Bookstore	Vietnamese Restaurant	Farmers Market	Restaurant	Bakery	Café	Bar
3	Dollard-des-Ormeaux	Skating Rink	Italian Restaurant	Thai Restaurant	Fast Food Restaurant	Shopping Mall	Food & Drink Shop	Restaurant	Burger Joint	Park	Mexican Restaurant
4	Dorval	Hotel	Gas Station	Bank	Sandwich Place	Rental Car Location	Recreation Center	Sushi Restaurant	Coffee Shop	Bar	Fast Food Restaurant

5. Results

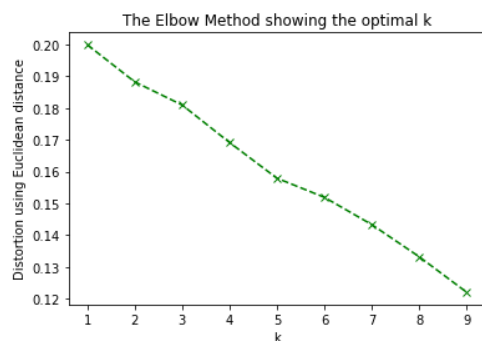
In this section, the neighborhoods in Montreal will be classified into groups using K-Means Clustering. Before, building the model, the optimum k will be determined. Elbow method is one of the most popular method to determine the optimal k (number of clusters). There are two metrics used in the Elbow method:

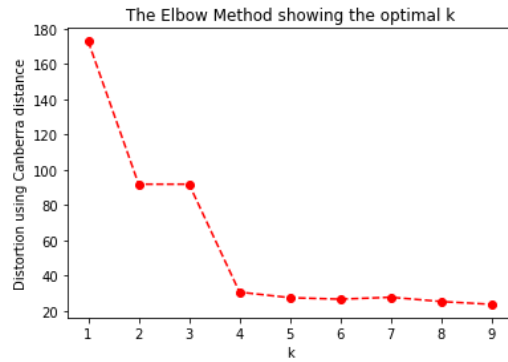
- Distortion: The average of the squared distances from the cluster centers of the respective clusters. Typically, the Euclidean distance metric is used.
- Inertia: It is the sum of squared distances of data to their closest cluster center.

The graph “The Elbow Method using Inertia” is as follow:



The graphs “The Elbow Method showing the optimal k” are as follow:





To determine the optimal number of clusters, we have to select the value of k at the “elbow” (i.e. the point after which the distortion/inertia starts decreasing in a linear fashion). From the graphs above, using Distortion with Euclidean distance and Inertia metrics, the optimal k is not obvious. However, using Distortion with Canberra distance, we see a clear evidence that the optimal number of clusters for the data is four.

The head of cluster table using optimum k (4 clusters):

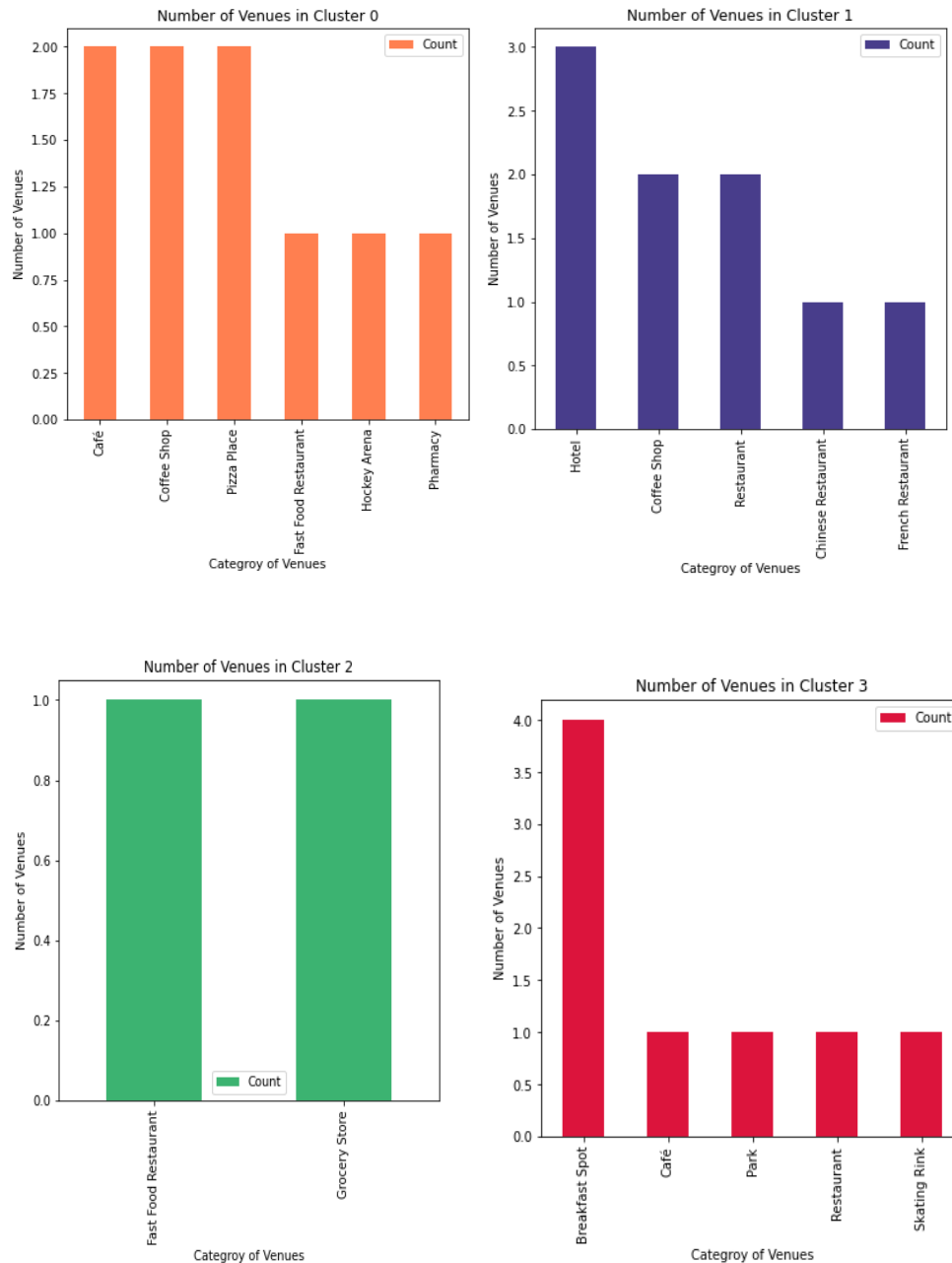
	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Downtown Montreal	45.505290	-73.564076	1	Hotel	Café	Asian Restaurant	Plaza	French Restaurant	Performing Arts Venue	Japanese Restaurant	Restaurant	Art Museum	Hawaiian Restaurant
1	Noire-Dame-de-Grâce	45.487967	-73.629922	3	Café	Grocery Store	Pizza Place	Mexican Restaurant	Mac & Cheese Joint	Sporting Goods Shop	Post Office	Sandwich Place	Tea Room	Park
2	Place Bonaventure	45.499593	-73.564917	1	Coffee Shop	Hotel	French Restaurant	Restaurant	Plaza	Café	Breakfast Spot	Building	Burger Joint	Japanese Restaurant
3	Duvernay-Est	45.499592	-73.564917	1	Coffee Shop	Hotel	French Restaurant	Restaurant	Plaza	Café	Breakfast Spot	Building	Burger Joint	Japanese Restaurant
4	Dollard-des-Ormeaux	45.494230	-73.806455	3	Skating Rink	Italian Restaurant	Thai Restaurant	Fast Food Restaurant	Shopping Mall	Food & Drink Shop	Restaurant	Burger Joint	Park	Mexican Restaurant

The resulting map with clusters is as follow:



For each cluster, the number of 1st most common venues by cluster is calculated. This will allow to assign a name for each cluster.

The number of the venues by cluster are shown in the following bar chart:



When examining graphs above, a name is assigned to each cluster:

Cluster	Name	
0	0	Café/Coffee Shop/Pizza Place Venues
1	1	Hotel Venues
2	2	Fast Food Restaurant/Grocery Store Venues
3	3	Breakfast Spot Venues

6. Discussion

When comparing to New York City and the city of Toronto, Montreal is less dense as city. More than 30 neighborhoods where the number of venues are very low. This indicates that those neighborhoods are not worth exploring when visiting Montreal. Furthermore, only Downtown Montreal has more than 100 venues from Foursquare API. This indicates that Downtown Montreal have many venues to visit.

7. Conclusion

Although, Montreal is less dense city as New York City and Toronto, it is definitively a city to visit. From the bar chart of the number of the venues by cluster, most of the 1st common venues is all the clusters are restaurants. This indicates that Montreal is a city to go for good restaurants.

As mentioned earlier, many neighborhoods are suburb area. The readers should considering ignoring those neighborhoods if they decide to visit Montreal.

In addition, the readers can decide to plan a four days trip since the K-Means Clustering analysis has suggested grouping Montreal neighborhoods into four clusters.

8. References

[1] [List of postal codes of Canada: H](#)

[2] [Forsquare API](#)