Clustering Montreal Neighbourhoods to Improve Real Estate Agent Efficiency

A report by Vishnu Kumaraswami

Submitted as part of the Data Science Capstone Project on Coursera

Date: 24th June 2021

Introduction:

You are a real estate agent in the city of Montreal, and resources are stretched. You have many enquiries coming in, and not enough time to handle them all efficiently. There are young adults looking for their first apartment, middle aged couples with children in school, and even the occasional retiree who just wants to relax and pass the rest of their days in a peaceful neighbourhood. But time is money, and it's very limited right now.

What you need is a way to efficiently cluster the neighbourhoods in Montreal, so that you can recommend specific locations based on your clients' profiles, instead of going on a wild goose chase around the city. Further, by clustering an existing dataset of client profiles, every new client can be assigned a cluster based on their profile, simplifying your task further.

What we will aim to achieve with this project is to:

1) Cluster the neighbourhoods in Montreal based on the facilities and amenities available in each location.
2) Cluster an existing client database to fit multiple profiles
3) Assign client profiles to neighbourhood clusters based on assumptions regarding popular activities

---

Data:

The relevant data for this project will be as follows:

1. Foursquare location data for Montreal – using the Foursquare API

   This data will help provide a list of points of interest across multiple locations within the greater Montreal metropolitan area.

2. List of postal codes in Montreal – available at the link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H

   We will need to retrieve coordinate data for each neighbourhood using the postal codes available at this address, in order to facilitate accurate use of the Foursquare API.

3. A pre-existing database of customer profiles from the course, available at the link: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%204/data/Cust_Segmentation.csv

   This data contains details such as age, income and education levels, all of which may be relevant to client clustering.

Methodology:

The starting point of the analysis was to scrape the webpage containing postal codes and neighbourhood names for the Canadian city of Montreal, available at:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H

H0A *Not assigned*	**H1A** Pointe-aux-Trembles	**H2A** Saint-Michel, East	**H3A** Downtown Montreal North (McGill University)	**H4A** Notre-Dame-de-Grâce Northeast	**H5A** Place Bonaventure	**H7A** Duvernay-Est
H0B *Not assigned*	**H1B** Montreal East	**H2B** Ahuntsic North	**H3B** Downtown Montreal East	**H4B** Notre-Dame-de-Grâce Southwest	**H5B** Place Desjardins	**H7B** Saint-François
H0C *Not assigned*	**H1C** Rivière-des-Prairies Northeast	**H2C** Ahuntsic Central	**H3C** Griffintown (Includes Île Notre-Dame & Île Sainte-Hélène) (Université de Montréal)	**H4C** Saint-Henri	**H5C** *Not assigned*	**H7C** Saint-Vincent-de-Paul
H0E *Not assigned*	**H1E** Rivière-des-Prairies Southwest	**H2E** Villeray Northeast	**H3E** L'Île-Des-Soeurs	**H4E** Ville Émard	**H5E** *Not assigned*	**H7E** Duvernay

Unassigned postal codes were bypassed, and the remaining were mapped into a dataframe:

	Postal Code	Neighborhood
0	H1A	Pointe-aux-Trembles
1	H2A	Saint-Michel,East
2	H3A	Downtown Montreal North
3	H4A	Notre-Dame-de-GrâceNortheast
4	H5A	Place Bonaventure
5	H7A	Duvernay-Est
6	H9A	Dollard-des-OrmeauxNorthwest
7	H1B	Montreal East
8	H2B	AhuntsicNorth
9	H3B	Downtown MontrealEast
10	H4B	Notre-Dame-de-GrâceSouthwest

There are a total of 124 neighbourhoods considered for the purposes of segmentation and clustering

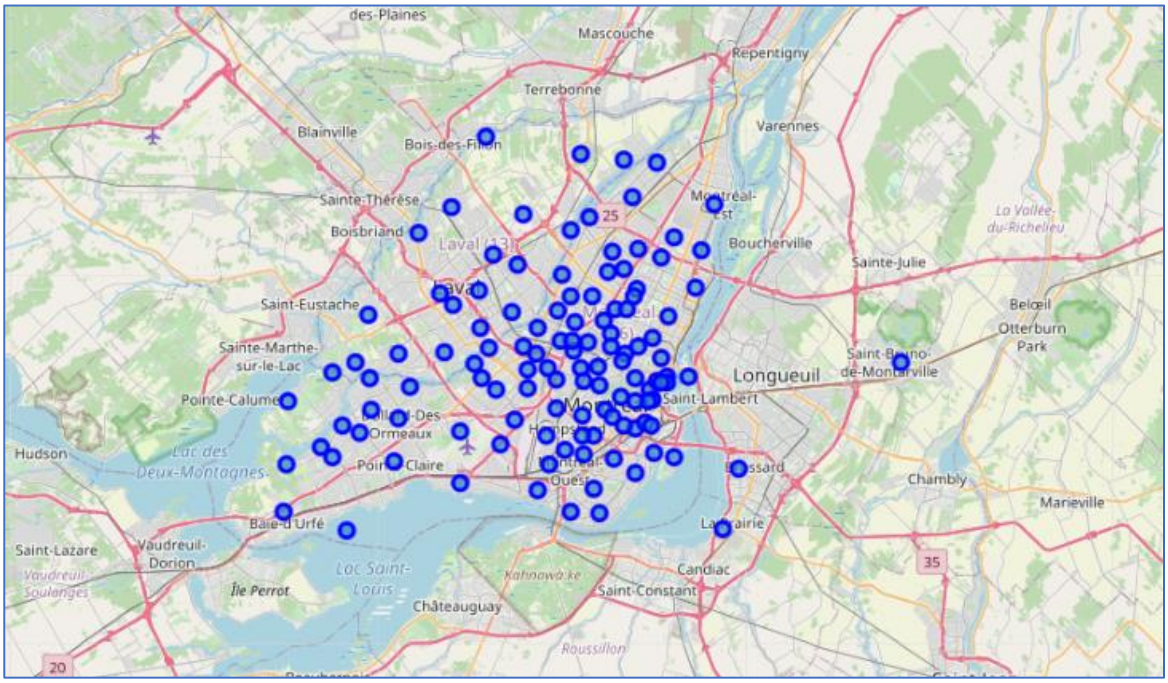
Geographical coordinates were sourced from the internet in order to provide reference points for the Foursquare API. These were mapped against the postal codes and a comma separated value file (.csv) was uploaded to the cloud. A pandas dataframe was created using this information:

```
In [17]: df_data_1.head()
```

Out[17]:

	Unnamed: 0	Postal Code	Neighborhood	Latitude	Longitude
0	0	H1A	Pointe-aux-Trembles	45.454300	-73.483774
1	1	H2A	Saint-Michel East	45.561567	-73.601288
2	2	H3A	Downtown Montreal North	45.506992	-73.568941
3	3	H4A	Notre-Dame-de-Grâce Northeast	45.476496	-73.622168
4	4	H5A	Place Bonaventure	45.500795	-73.565230

Using Folium, all the locations were plotted on a map for visual reference:



Following this, the Foursquare API was used to map all the venues relevant to the locations of each of the neighbourhoods:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Saint-Michel East	45.561567	-73.601288	STM Station Saint-Michel	45.559425	-73.599749	Metro Station
1	Saint-Michel East	45.561567	-73.601288	Marché Aux Puces Saint-Michel	45.562502	-73.605079	Flea Market
2	Saint-Michel East	45.561567	-73.601288	Petro-Canada	45.560984	-73.602396	Gas Station
3	Saint-Michel East	45.561567	-73.601288	Restaurant Kim Hour	45.561836	-73.605112	Chinese Restaurant
4	Saint-Michel East	45.561567	-73.601288	Poissonnerie Mediterraniénne	45.562328	-73.605625	Fish & Chips Shop
...	...	...	...	...	...	...	...
869	Tour de la Bourse	45.515558	-73.531910	Complexe aquatique de l'Île	45.513078	-73.534298	Pool
870	Tour de la Bourse	45.515558	-73.531910	Biosphère	45.513979	-73.531582	Science Museum
871	Tour de la Bourse	45.515558	-73.531910	Tour de Lévis	45.517164	-73.533460	Historic Site
872	Tour de la Bourse	45.515558	-73.531910	STM Ligne 767 La Ronde	45.512670	-73.530872	Bus Stop
873	Pierrefonds	45.515428	-73.836390	B W N	45.513529	-73.839649	Construction & Landscaping

874 rows × 7 columns

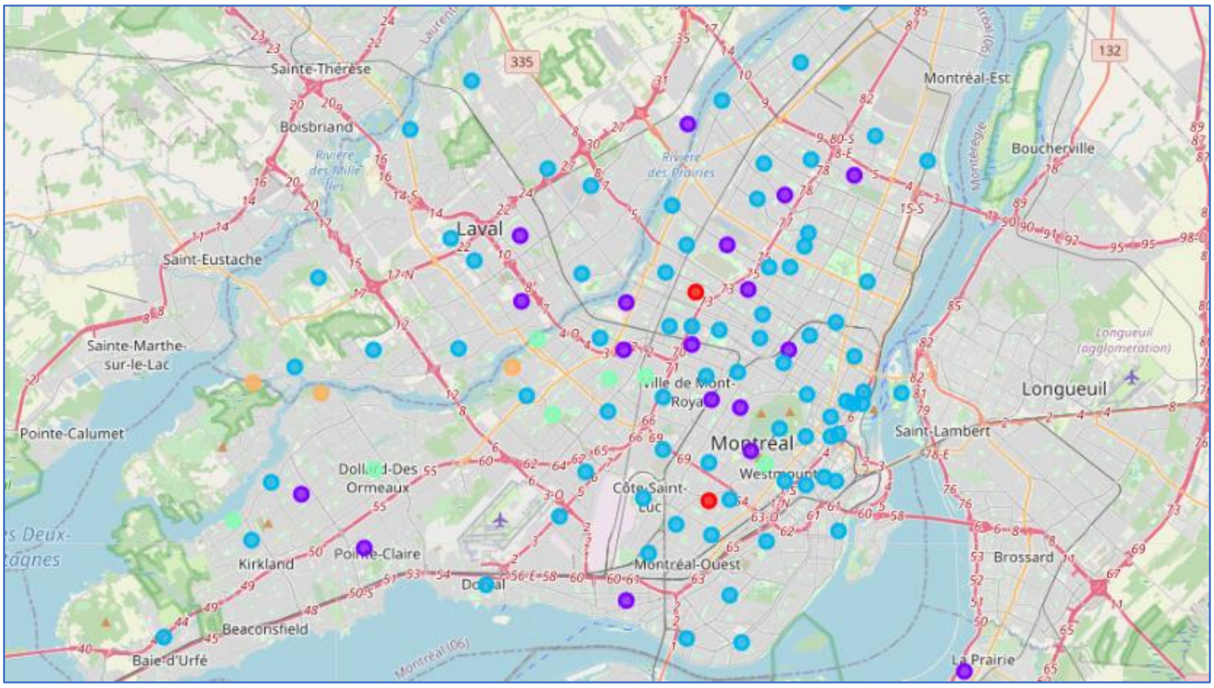
190 unique categories of venue were found in the city of Montreal.

The most common venues relevant to each neighbourhood were then mapped, in order to find points of similarity to help cluster them.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Ahuntsic Central	Breakfast Spot	Ice Cream Shop	Coffee Shop	Pharmacy	Restaurant
1	Ahuntsic East	Playground	Athletics & Sports	Yoga Studio	Discount Store	Farm
2	Ahuntsic North	Soccer Field	Yoga Studio	College Science Building	Farm	Factory
3	Ahuntsic Southeast	Italian Restaurant	Women's Store	Clothing Store	Sandwich Place	Bank
4	Ahuntsic Southwest	Home Service	Ice Cream Shop	Pizza Place	Bakery	Park

Based on factors of commonality, k-means clustering was used to find groups of neighbourhoods with similar groups of venues nearby. For the purposes of this exercise, k was set to 5, due to the variety of unique venues in Montreal.

The clusters were mapped using Folium as below:



Once the neighbourhoods were clustered, attention turned to the database of clients, to see if any patterns could be found there.

Please note that a pre-existing database of clients was used for the purpose of analysis, as indicated in the 'Data' section of this report.

Using k-means, clusters were created from the database of clients, as follows:

	Customer Id	Age	Edu	Years Employed	Income
Clus_km					
0	402.295082	41.333333	1.956284	15.256831	83.928962
1	432.468413	32.964561	1.614792	6.374422	31.164869
2	410.166667	45.388889	2.666667	19.555556	227.166667

RESULTS:

Clusters were successfully generated both for the neighbourhoods of Montreal, as well as for the client database which was on file.

For Neighbourhoods: Based on the most common venues available in each of the clusters, it would be possible to extrapolate the kind of lifestyle one would be able to become accustomed to.

Cluster	Most Common Venues	Assumed Client Profile
0	Playground, Stores, Yoga Studios	Young families with children below 5
1	Parks, Gym/Yoga, Convenience Stores	Middle Aged, Upper middle income, families
2	Pubs, Restaurants, Entertainment	Young adults, disposable income
3	Park, Yoga Studio, Farm	Older, High Income, Possibly retired
4	Construction, Yoga, Fast food	Non-residential area

For Clients: From the clusters above, it would be possible to classify the customers as follows:

Type A (Cluster 0): Middle Aged, Upper Middle Income

Type B (Cluster 1): Young, Upwardly Mobile

Type C (Cluster 2): Older, High Income

Accordingly, we can allocate neighbourhood clusters to client profiles:

Client Profile	Neighbourhood Cluster(s)
Type A	1,0
Type B	2,1
Type C	3

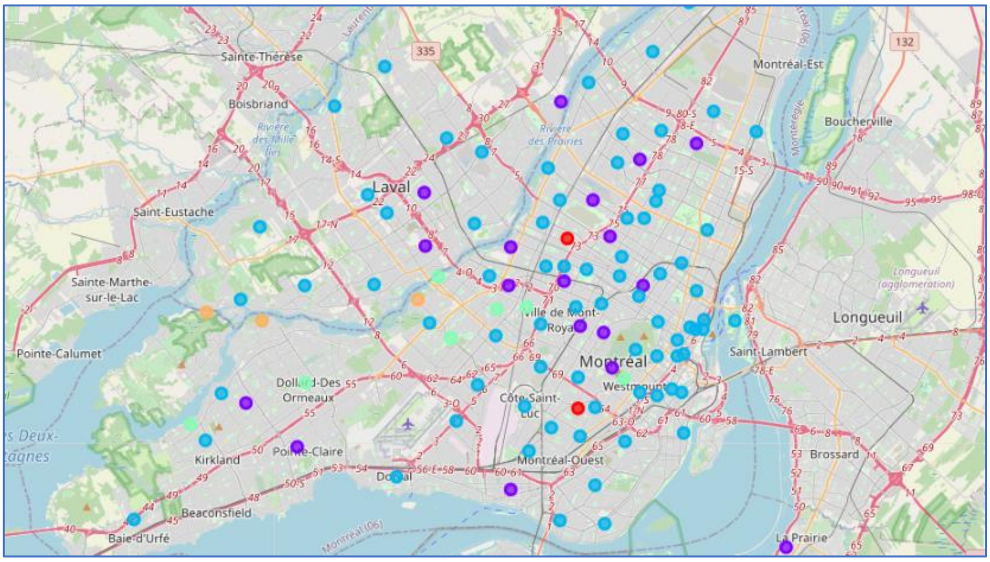
DISCUSSION:

The clustering of neighbourhoods in Montreal can act as a guide for real estate agents when assessing potential clients, helping save time and improve efficiency.

A prospective client's details would need to be updated in the database and then the program would need to be run. Accordingly, they would be assigned a cluster and provided with a suitable set of real estate options (either to sell or to rent depending on their requirement).

Additional Observations:

- The neighbourhoods in Montreal are not really clustered geographically. It is evident that any part of the city has at least 2 clusters to choose from, making a real estate agent's job that little bit easier.



- Montreal is clearly a very *fit* city – the presence of yoga studios in practically every neighbourhood shows a great demand for this ancient fitness regime.

81	H9R	1	Convenience Store	Yoga Studio	Dog Run	Fast Food Restaurant	Farm	Factory	Event Space
84	H3S	1	Gym / Fitness Center	Train Station	Park	Yoga Studio	Discount Store	Farm	Factory
86	H7S	1	Home Service	Yoga Studio	Dog Run	Fast Food Restaurant	Farm	Factory	Event Space
87	H8S	1	Park	Sandwich Place	Skate Park	Diner	Event Space	Escape Room	English Restaurant
91	H3T	1	Park	Diner	College Gym	Sandwich Place	Factory	Event Space	Escape Room
97	H3V	1	Convenience Store	Home Service	Memorial Site	Chinese Restaurant	Caribbean Restaurant	Park	Yoga Studio
99	H7V	1	Park	Asian Restaurant	Donut Shop	Yoga Studio	Farm	Factory	Event Space
115	H4Y	1	Spa	Skating Rink	Park	Yoga Studio	Discount Store	Event Space	Escape Room

CONCLUSION:

Montreal as a city appears to have something for everyone, so it is vital for a real estate agent to find each client’s niche.

The clustering of neighbourhoods within the city should serve as a guide to connecting clients with their dream home as quickly as possible.

Thank you for your time spent reading this report.