The Battle of Neighbourhoods

Capstone Project

Applied Data Science

Seif El Kilany

March 8, 2021

Introduction

Background

Montreal is the second most populated city in Canada and the largest city in the province of Quebec (Government of Canada, 2019). Just like all large metropolises, new and young families end up leaving the downtown center looking for less congested, more wholesome neighbourhoods, in addition more affordable homes to grow their families. With working from home now a fact of life thanks to the COVID pandemic, the exodus from the downtown hastens (Hanes, 2020).

Question

The question that these young families looking to move have to answer and one that real estate agents are always trying to address is: where should they be moving to, which neighborhoods should they be looking at.

Data

To help address this question, we need to geographically identify Montreal neighbourhoods, popular venues and schools. To do so several sources were used, below is a description of each source used, along with the data wrangling performed on it. These data points are the main deciding factors for our question. An important factor that will not be discussed in this paper would be real estate prices. However, with Montreal downtown prices being some of the most expensive in the country (Century 21 Canada), we will make the assumption that prices as you leave downtown become more affordable.

Montreal Postal Codes

Data Sources

We need to look for geographical data of the city of Montreal, and identify each neighbourhood and to do so we use the postal codes. On Wikipedia we can find a list of the city's postal codes and neighbourhoods https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_H'.

Data Wrangling

After loading the Wikipedia table with Montreal neighbourhoods in to a dataframe, we can see that it is not clearly listed and that each postal code and neighbourhood name are combined into one cell.

Figure 1
Initial scrapping of the Wikipedia page.

	0	1	2	3	4	5	6	
0	H0ANot assigned	H1APointe-aux-Trembles	H2ASaint-Michel,East	H3ADowntown Montreal North(McGill University)	H4ANotre-Dame-de-GrâceNortheast	H5APlace Bonaventure	H7ADuvernay-Est	H8A
1	H0BNot assigned	H1BMontreal East	H2BAhuntsicNorth	H3BDowntown MontrealEast	H4BNotre-Dame-de-GrâceSouthwest	H5BPlace Desjardins	H7BSaint-François	H8B
2	H0CNot assigned	H1CRivière-des-PrairiesNortheast	H2CAhuntsicCentral	H3CGriffintown(Includes Île Notre-Dame & Île S...	H4CSaint-Henri	H5CNot assigned	H7CSaint-Vincent-de-Paul	H8C
3	H0ENot assigned	H1ERivière-des-PrairiesSouthwest	H2EVillerayNortheast	H3EL'Île-Des-Soeurs	H4EVille Émard	H5ENot assigned	H7EDuvernay	H8E
4	H0GNot assigned	H1GMontréal-NordNorth	H2GPetite-PatrieNortheast	H3GDowntown MontrealSoutheast (Concordia Unive...	H4GVerdunNorth	H5GNot assigned	H7GPont-Viau	H8G
5	H0HReserved0H0: Santa Claus	H1HMontréal-NordSouth	H2HPlateau Mont-RoyalNorth	H3HDowntown MontrealSouthwest	H4HVerdunSouth	H5HNot assigned	H7HAuteuilWest	H8H
6	H0JNot assigned	H1JAnjouWest	H2JPlateau Mont-RoyalNorth Central	H3JPetite-Bourgogne	H4JCartiervilleCentral	H5JNot assigned	H7JAuteuilNortheast	H8J

So first we use the pandas stack function to output a one leveled list with all the items. Then we strip the postal code and neighbourhood name into a column each from the combined cell.

Figure 2

Wikipedia scrapped table after manipulation.

	Postal Code	Neighbourhood
1	H1A	Pointe-aux-Trembles
2	H2A	Saint-Michel,East
3	H3A	Downtown Montreal North(McGill University)
4	H4A	Notre-Dame-de-GrâceNortheast
5	H5A	Place Bonaventure
6	H7A	Duvernay-Est
8	H9A	Dollard-des-OrmeauxNorthwest
10	H1B	Montreal East
11	H2B	AhuntsicNorth
12	H3B	Downtown MontrealEast
13	H4B	Notre-Dame-de-GrâceSouthwest

Lastly we drop the original combined column, and any 'not assigned' codes to arrive at a clean dataframe with postal codes and neighbourhood names with a shape of (123, 2).

Figure 3

Confirmation of shape.

```
In [11]: montreal_df.shape
Out[11]: (123, 2)
```

Montreal Coordinates

Data Sources

The Wikipedia page does not provide coordinates, we first try to get them from geocoder, but unfortunately after letting it run for over an hour with no outcome, an alternate source was necessary. We were able to find a list of all Canadian postal codes and coordinates from GeoNames http://download.geonames.org/export/zip/

Data Wrangling

At GeoNames we find coordinates for all Canadian postal codes in a text file that was in a compressed Zip folder. We use the ZipFile and BytesIO libraries to unzip the folder and read the contents: "readme.txt" and "CA.txt". We read the coordinates table into a dataframe, label the columns as per readme file and drop the unnecessary columns to arrive at a clean dataframe of: all Canadian postal codes, neighbourhood names, province and coordinates. Since we are only interested in Montreal, we merge the Montreal postal codes dataframe from earlier with this Canadian coordinates dataframe on 'Postal Code' to get a new dataframe that has all Montreal postal codes, neighbourhoods and coordinates for mapping.

Figure 4

GeoNames data before manipulation.

	0	1	2	3	4	5	6	7	8	9	10	11
0	CA	T0A	Eastern Alberta (St. Paul)	Alberta	AB	NaN	NaN	NaN	NaN	54.7660	-111.7174	6.0
1	CA	T0B	Wainwright Region (Tofield)	Alberta	AB	NaN	NaN	NaN	NaN	53.0727	-111.5816	6.0
2	CA	T0C	Central Alberta (Stettler)	Alberta	AB	NaN	NaN	NaN	NaN	52.1431	-111.6941	5.0
3	CA	T0E	Western Alberta (Jasper)	Alberta	AB	NaN	NaN	NaN	NaN	53.6758	-115.0948	5.0
4	CA	T0G	North Central Alberta (Slave Lake)	Alberta	AB	NaN	NaN	NaN	NaN	55.6993	-114.4529	6.0
5	CA	T0H	Northwestern Alberta (High Level)	Alberta	AB	NaN	NaN	NaN	NaN	57.5403	-116.9153	6.0
6	CA	T0J	Southeastern Alberta (Drumheller)	Alberta	AB	NaN	NaN	NaN	NaN	50.9944	-111.4632	6.0
7	CA	T0K	International Border Region (Cardston)	Alberta	AB	NaN	NaN	NaN	NaN	49.4721	-112.2408	6.0

Figure 5

GeoNames data after manipulation.

	Postal Code	Neighbourhood	Place Name	Province	Province Code	Latitude	Longitude
0	H1A	Pointe-aux-Trembles	Pointe-Aux-Trembles	Quebec	QC	45.6753	-73.5016
1	H2A	Saint-Michel,East	Saint-Michel East	Quebec	QC	45.5618	-73.5990
2	H3A	Downtown Montreal North(McGill University)	Downtown Montreal North	Quebec	QC	45.5040	-73.5747
3	H4A	Notre-Dame-de-GrâceNortheast	Notre-Dame-de-GrÔce Northeast	Quebec	QC	45.4717	-73.6149
4	H5A	Place Bonaventure	Place Bonaventure	Quebec	QC	45.4992	-73.5646
5	H7A	Duvernay-Est	Duvernay-Est	Quebec	QC	45.6739	-73.5924
6	H9A	Dollard-des-OrmeauxNorthwest	Dollard-Des-Ormeaux Northwest	Quebec	QC	45.4948	-73.8317
7	H1B	Montreal East	Montreal East	Quebec	QC	45.6320	-73.5075

Montreal Venues

Data Sources

To explore the neighbourhoods, we use the Foursquare API to get points of interest/venues in each Montreal neighbourhood. Since we use a free account with limitations and for simplicity's sake, we limit each neighbourhood to one hundred venues within a five hundred meter radius of the coordinates.

Data Wrangling

The retrieved data required no manipulation to use (at this point) and consists of neighbourhood name and coordinates, venue name, category and coordinates, that we load into a dataframe.

Figure 6
Foursquare Montreal venues data.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Pointe-aux-Trembles	45.6753	-73.5016	Parc-nature de la Pointe-aux-Prairies	45.678834	-73.501162	Park
1	Pointe-aux-Trembles	45.6753	-73.5016	AMT Gare Pointe-aux-Trembles	45.674882	-73.504908	Train Station
2	Pointe-aux-Trembles	45.6753	-73.5016	Parc Yves-Thériault	45.678675	-73.502037	Park
3	Saint-Michel,East	45.5618	-73.5990	Bar Zoe	45.559673	-73.597542	Karaoke Bar
4	Saint-Michel,East	45.5618	-73.5990	Marché Aux Puces Saint-Michel	45.562502	-73.605079	Flea Market
5	Saint-Michel,East	45.5618	-73.5990	STM Station Saint-Michel	45.559425	-73.599749	Metro Station
6	Saint-Michel,East	45.5618	-73.5990	Restaurant Kim Hour	45.561836	-73.605112	Chinese Restaurant
7	Saint-Michel,East	45.5618	-73.5990	Petro-Canada	45.560984	-73.602396	Gas Station

Montreal English Schools

Data Sources

Since families are a big stakeholder here, we need the location of all English elementary and high schools in each neighborhood. We can get an up to date list of elementary and high schools from the website of the English Montreal School Board https://az184419.vo.msecnd.net/emsb/emsb-website/en/docs/2020-2021/list-of-schools-20-21.pdf .

Data Wrangling

The elementary and high school data are in two separate tables in a pdf file. First we use the Tabula library to read the pdf file. This gives us a list of all the tables in the pdf document. The first scrapped table contains elementary schools without headings as they were in text in the pdf body.

Figure 7

Pdf scrapped data.

	0	1	2	3	4	5	6	7
0	BancroftPK	B	1001	4563 St. Urbain H2T 2V9	514.845.8031	514.845.4352	Dorothy Ostrowicz	NaN
1	CarlylePK	E&IB	1002	109 Carlyle, TMR H3R 1S8	514.738.1256	514.738.0373	Dina Vourdousis	NaN
2	Cedarcrest	I	1003	1505 Muir, St. Laurent H4L 4T1	514.744.2614	514.744.3310	Elena Zervas	NaN
3	CoronationPK	E&I	1045	4810 Van Horne H3W 1J3	514.733.7790	514.733.7701	Mike Talevi	NaN
4	Dalkeith\rPK	E	1004	7951 Dalkeith, Anjou H1K 3X6	514.352.6730	514.352.0243	John Wright	NaN
5	DantePK	B	1005	6090 Lachenaie, St. Léonard H1S 1P1	514.254.5941	514.254.6697	Joseph Schembri	NaN
6	Dunrae GardensPK	I	1006	235 Dunrae, TMR H3P 1T5	514.735.1916	514.735.7051	Despina Michakis	NaN
7	East HillPK	I	1007	10350 Perras, RDP H1C 2H1	514.494.3202	514.494.3153	Liboria Amato	Cynthia Canale

So we take that and put it in a dataframe, rename the column headings correctly and add a column 'Type' to label these schools as elementary for our records.

Figure 8

Elementary school dataframe after manipulation.

	School	Prog	Ext	Address	Tel	Fax	Principal	Vice Principal	Type
0	BancroftPK	B	1001	4563 St. Urbain H2T 2V9	514.845.8031	514.845.4352	Dorothy Ostrowicz	NaN	Elementary
1	CarlylePK	E&IB	1002	109 Carlyle, TMR H3R 1S8	514.738.1256	514.738.0373	Dina Vourdousis	NaN	Elementary
2	Cedarcrest	I	1003	1505 Muir, St. Laurent H4L 4T1	514.744.2614	514.744.3310	Elena Zervas	NaN	Elementary
3	CoronationPK	E&I	1045	4810 Van Horne H3W 1J3	514.733.7790	514.733.7701	Mike Talevi	NaN	Elementary
4	Dalkeith\rPK	E	1004	7951 Dalkeith, Anjou H1K 3X6	514.352.6730	514.352.0243	John Wright	NaN	Elementary
5	DantePK	B	1005	6090 Lachenaie, St. Léonard H1S 1P1	514.254.5941	514.254.6697	Joseph Schembri	NaN	Elementary
6	Dunrae GardensPK	I	1006	235 Dunrae, TMR H3P 1T5	514.735.1916	514.735.7051	Despina Michakis	NaN	Elementary
7	East HillPK	I	1007	10350 Perras, RDP H1C 2H1	514.494.3202	514.494.3153	Liboria Amato	Cynthia Canale	Elementary

We now repeat the process for high school data which is the second table in the pdf tables list.

Figure 9

High school dataframe after manipulation.

	School	Ext	Address	Tel	Fax	Principal	Vice Principal	Type
0	F.A.C.E.	1147	3449 University H3A 2A8	514.350.8899	514.350.2612	Marilyn Ramlakhan	Jennifer Harriet	High
1	James Lyng	1101	5440 Notre Dame W. H4C 1T9	514.846.8814	514.846.3006	Lino Buttino	Andrea Dillon	High
2	John F. Kennedy	1102	3030 Villeray H2A 1E7	514.374.1449	514.374.2224	Otis Delaney	Vito Campbell-\rGuerriero	High
3	John Grant	1117	5785 Parkhaven, Cote St. Luc H4W 1X8	514.484.4161	514.484.4969	Jennifer Le Huquet	NaN	High
4	L.I.N.K.S.	1109	9905 Papineau H2B 1Z9	514.723.2845	514.723.2666	Maria Calderella	NaN	High
5	Laurenhill Academy	1104	2505 Cote Vertu, St. Laurent H4R 1P3	514.331.8781	514.331.7145	Donna Manos	Rea Limperopoulos	High
6	Laurenhill Jr. Campus	5662	2355 Decelles H4M 1C2	514.331.8019	514.331.0205	Alexander Kulczyk\rMireille Tehbelian	NaN	High

To simplify our data for clustering and segmenting later on we need to clean up the schools data, so we start off by putting both school dataframes together. Since both dataframes have the same columns except for one, we just drop the extra column from the elementary dataframe and use the pandas append function to combine them.   We can see the postal codes (needed for coordinates), in the ‘Address’ field, so we use string manipulation to strip out the postal code from the ‘Address’ and put it in its own ‘Postal Code’ column. We don’t really need all the school info now, so we can drop six columns from the dataframe to leave us with just school name, type and postal code.

Figure 10

Schools dataframe after basic cleaning and combining.

	School	Prog	Type	Postal Code
0	BancroftPK	B	Elementary	H2T
1	CarlylePK	E&IB	Elementary	H3R
2	Cedarcrest	I	Elementary	H4L
3	CoronationPK	E&I	Elementary	H3W
4	Dalkeith\rPK	E	Elementary	H1K
5	DantePK	B	Elementary	H1S

To finalize a school dataframe ready for mapping later (Figure 11 below), we groupby ‘Postal Code’ to get the number of schools in a postal code, then add coordinates and neighbourhood names from the Montreal postal codes dataframe from Figure 5.

Figure 11

Schools dataframe after manipulation.

	Postal Code	School	Type	Neighbourhood	Place Name	Province	Province Code	Latitude	Longitude
0	H1C	1	1	Rivière-des-PrairiesNortheast	Rivière-des-Prairies Northeast	Quebec	QC	45.6656	-73.5367
1	H1E	2	2	Rivière-des-PrairiesSouthwest	Rivière-Des-Prairies Southwest	Quebec	QC	45.6342	-73.5842
2	H1G	2	2	Montréal-NordNorth	Montreal North North	Quebec	QC	45.6109	-73.6211
3	H1H	1	1	Montréal-NordSouth	Montreal North South	Quebec	QC	45.5899	-73.6389
4	H1K	1	1	AnjouEast	Anjou East	Quebec	QC	45.6097	-73.5472
5	H1N	1	1	MercierSoutheast	Mercier Southeast	Quebec	QC	45.5779	-73.5304
6	H1P	1	1	Saint-LéonardNorth	Saint-Léonard North	Quebec	QC	45.5966	-73.5928
7	H1R	1	1	Saint-LéonardWest	Saint-Léonard West	Quebec	QC	45.5864	-73.6082

Methodology

Data Collection

The first phase in addressing the question posed in this paper was the data collection and wrangling phase. As described in the Data section we had four sources and it involved using:

- Requests library to read data from the websites. It was used in 3 out of our four data sets: Wikipedia, GeoNames and Foursquare API.
- Tabula library to read data from a pdf file for our fourth data set, English Montreal School Board.
- The ZipFile and BytesIO libraries to import and unzip the compressed folder for our GeoNames data.
- Pandas library to load the data read from all sources into dataframes and analyze the data.
- NumPy library to further work with and analyze our arrays.

Review and Statistics

We take a brief exploratory review of the cleaned up data sets we have collected and wrangled.

- On the Wikipedia data set we take a look at how many postal codes were collected and in running the NymPy shape and unique functions we can see that we have 123 unique postal codes.
- On the finalized Montreal postal codes with coordinates data set (Figure 5), we also take a look using the shape function, to see we have 122 entries. This is reassuring as it’s the same number of entries from the Wikipedia data, 122 (122 plus the H0H reserved code for Santa Claus).

- As for the venue data from Foursquare, using the shape function we can see with have 1886 entries. Let's take a quick look to see what that initially means:
    - Neighbourhood view: in grouping the Foursquare data by 'Neighbourhood', then running a count function and panda's describe function we see that the data covers 116 neighbourhoods with an average of 16 venues per neighbourhood. With a minimum venue of 1 and max of 100 (our limit).

Figure 12
Line of code for group by Neighbourhood.

```
montreal_venues.groupby('Neighbourhood').count().describe()
```

-
    - Venues view: in grouping the Foursquare data by 'Venue Category', then running a count function and panda's describe function we see that we have 260 categories with a category having an average of 7 venues. With a minimum venue of 1 and max of 105.

Figure 13
Line of code for group by Venue Category.

```
montreal_venues.groupby('Venue Category').count().describe()
```

- On the initial all schools data set (Figure 10) we run the describe function, to see we have a total of 51 English schools, 35 of which are elementary schools.
- On the finalized school data set grouped by postal codes (Figure 11) we run a describe function and find that our English schools are in 36 postal codes out the 122 that make up the greater Montreal area. Most of these neighbourhoods have 1 school, around 25% have 2 schools and a little less than 10% have 3 schools. With H3W, H3X and H4X codes having 3 schools each (between elementary and high school).

Analysis and Machine learnings

To really help answer the question posed in this paper, we have to make it easy for our stakeholders to identify the neighbourhoods they should be targeting; thus we need to visualize the data. However, we can't just plot thousands of points on a map, it would be hard to read and over whelming, so first we need to utilize basic machine learning to cluster and segment the data.

Analyzing Neighbourhoods

Before we can go and cluster the 1886 entries, we have to prepare our categorical data in a manner that machine learning can be applied to and to do so we use panda's get_dummies function to apply the One Hot encoding process to our data. If we take a quick look at the result of running the One Hot encoding we see a vectored version of our data with the 1886 rows consisting of neighbourhoods and 262 columns one for each venue category.

Figure 13

Result of One Hot encoding.

	Neighbourhood	Accessories Store	Adult Boutique	Airport Terminal	American Restaurant	Arepa Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment
0	Pointe-aux-Trembles	0	0	0	0	0	0	0	0	0
1	Pointe-aux-Trembles	0	0	0	0	0	0	0	0	0
2	Pointe-aux-Trembles	0	0	0	0	0	0	0	0	0
3	Saint-Michel,East	0	0	0	0	0	0	0	0	0
4	Saint-Michel,East	0	0	0	0	0	0	0	0	0

check new shape

```
montreal_onehot.shape
```

```
(1886, 262)
```

Now let's group the data above by 'Neighbourhood' and average it, to get the neighbourhoods and the average number of venues in each category. Next to simply further for mapping, we sort the venues in descending order and take the top 10 venues from each neighbourhood and put the result in a new dataframe that we will be reading for clustering.

Figure 14

Encoded data ready for clustering.

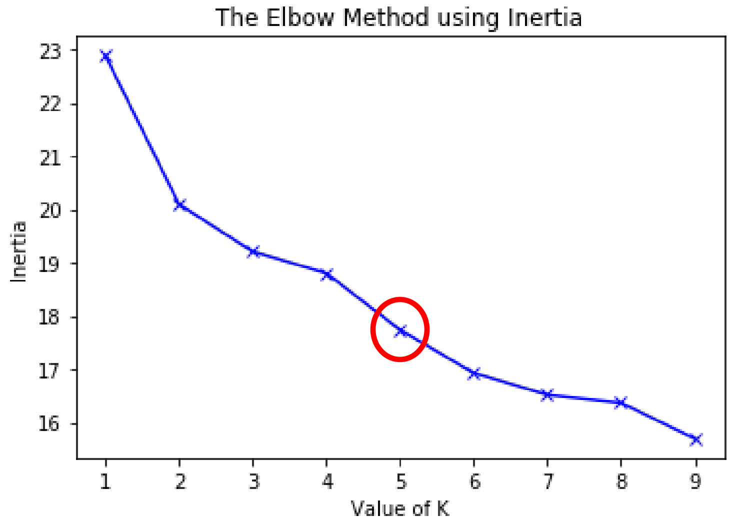
	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Laval-sur-le-Lac	BBQ Joint	Steakhouse	Golf Course	Yoga Studio	Falafel Restaurant	Empanada Restaurant	English Restaurant	Escape Room	Event Space	Farmers Market
1	AhuntsicCentral	Restaurant	Café	Plaza	Italian Restaurant	Organic Grocery	Bar	Coffee Shop	Grocery Store	Cheese Shop	Dessert Shop
2	AhuntsicEast	Park	Skating Rink	Athletics & Sports	Falafel Restaurant	Empada House	Empanada Restaurant	English Restaurant	Escape Room	Event Space	Yoga Studio
3	AhuntsicNorth	Pharmacy	Pizza Place	Italian Restaurant	Pet Store	Grocery Store	Fast Food Restaurant	Farmers Market	Fish & Chips Shop	Falafel Restaurant	Eastern European Restaurant
4	AhuntsicSoutheast	Furniture / Home Store	Sandwich Place	Bakery	Italian Restaurant	Discount Store	Electronics Store	Grocery Store	Bank	Park	Clothing Store

Clustering

To cluster the data we will use the the K-means clustering function from the Sklearn library. "A fundamental step for any unsupervised algorithm is to determine the optimal number of clusters into which the data may be clustered" (Gupta & Majumder, 2021). First we will use the Elbow Method to try to make an educated decision as to the number of clusters. We will try values of K from 1 to 9, first fitting the model then using the Kmeans inertia function to calculate. Lastly, we will use the Matplotlib library to plot the results for a visual representation.

Figure 15

Elbow method visualization.



Although the elbow isn't very clear, we will set the number of clusters to 5, initialize the method to Kmeans++ to speed up convergence, with a run count of 40 instead of the initial default of 10 (Scikit-learn developers). Now let's add the cluster number to the data from Figure 14 and merge it without Montreal coordinates data from Figure 5.

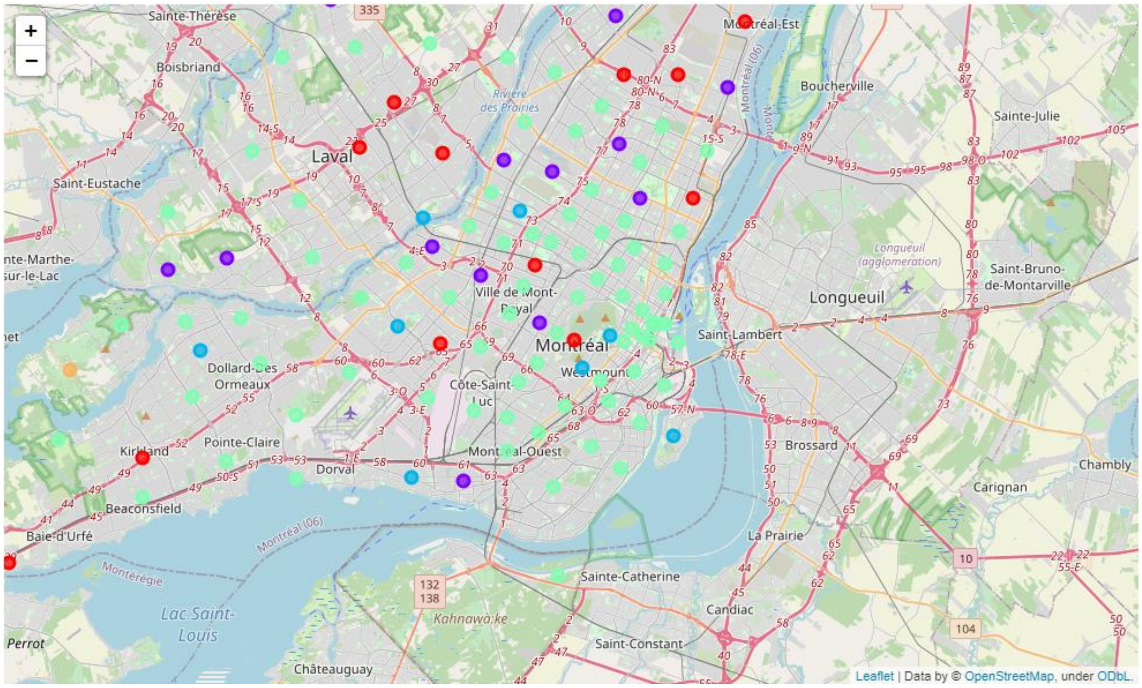
Before we can go ahead and map this data let's do a final cleaning, by checking for any NaNs using the pandas isnull function and then removing them using the pandas dropna function. Finally we are ready to map our clusters.

Mapping

To map our data we will be using the folium library. We will initialize the map to the city of Montreal and use the folium CircleMarker function to mark each neighbourhood on the map, set a label consisting of its name and the cluster number it belongs to. We will set a color scheme with a different colour for each cluster.

Figure 16

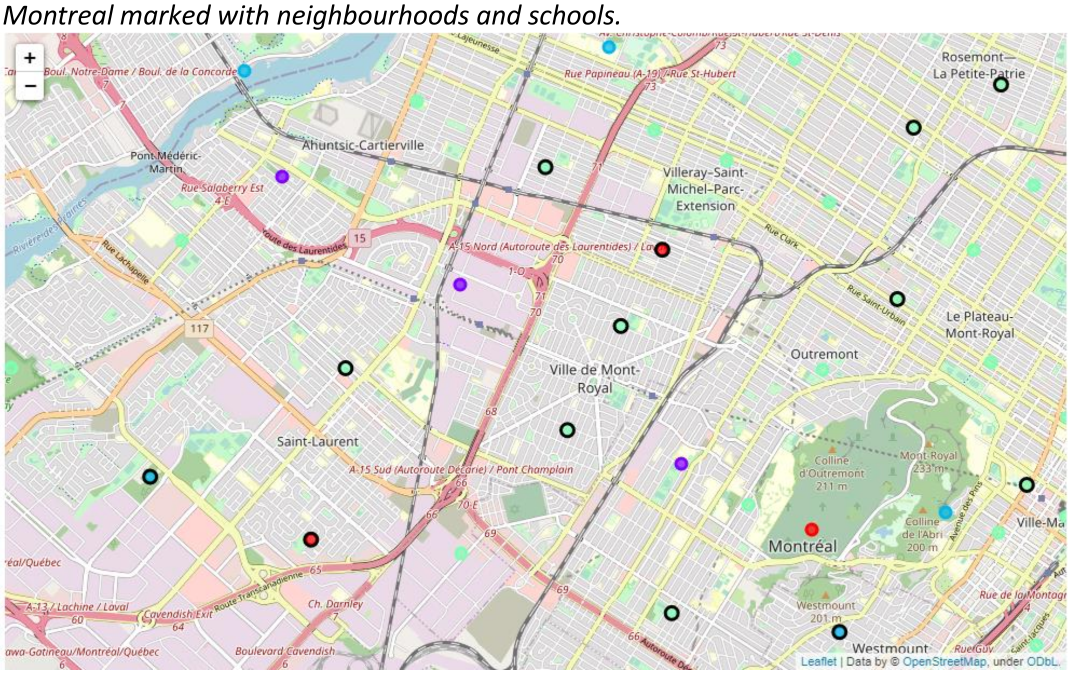
Map of Montreal marked with neighbourhoods and color codes for clustering.



Now let's add the schools to the map using the folium Circle function to draw a black circle around the neighbourhood markers with schools.

Figure 17

Map of Montreal marked with neighbourhoods and schools.



Results

When we pull up the data in each of our 5 clusters, we can quickly see the commonalities that can described as follows:

- Cluster 0 (red): mostly suburban neighbourhoods with venues like a skating rink, hockey arenas and a playground, but with their own commercial hubs due to the presence of some restaurants and business services, like landscaping, home stores, convenience stores and rental office.
- Cluster 1 (purple): these neighbourhoods are mostly residential and some being slightly rural with pharmacies, grocery stores, parks and factories showing up several times.
  They are probably the commercial zones in their respective vicinities with all the pharmacies, restaurants and retail stores.
- Cluster 2 (blue): an active residential area with parks being the most common venue in all of these neighbourhoods. Some of them are rural and further away from the city center as they have venues like 'Scenic lookouts', 'Trail', 'Mountain' and 'Factory'. Looking at the map from Figure 15, this is confirmed as neighbourhoods from this cluster are mostly on the edges of the island of Montreal
- Cluster 3 (crimson): this cluster has a lot of neighbourhoods; some are more commercial and night life neighbourhoods with their majority of venues being restaurants, coffee shops, bars and hotels. While other neighbourhoods lend to being more residential with some parks, a pool, skating rink and sports venues. This leads me to believe it is Montreal downtown and its immediate surroundings. This is confirmed if we look at the map from Figure 15.
- Cluster 4 (yellow): contains one sole neighbourhood, an island, with its most common venue being 'Golf Course'. That being said it's a residential golf community.

Taking into account that access to schools is a high priority for families, we can narrow our focus to the neighbourhoods with the highest number of schools, Table 1.

Table 1

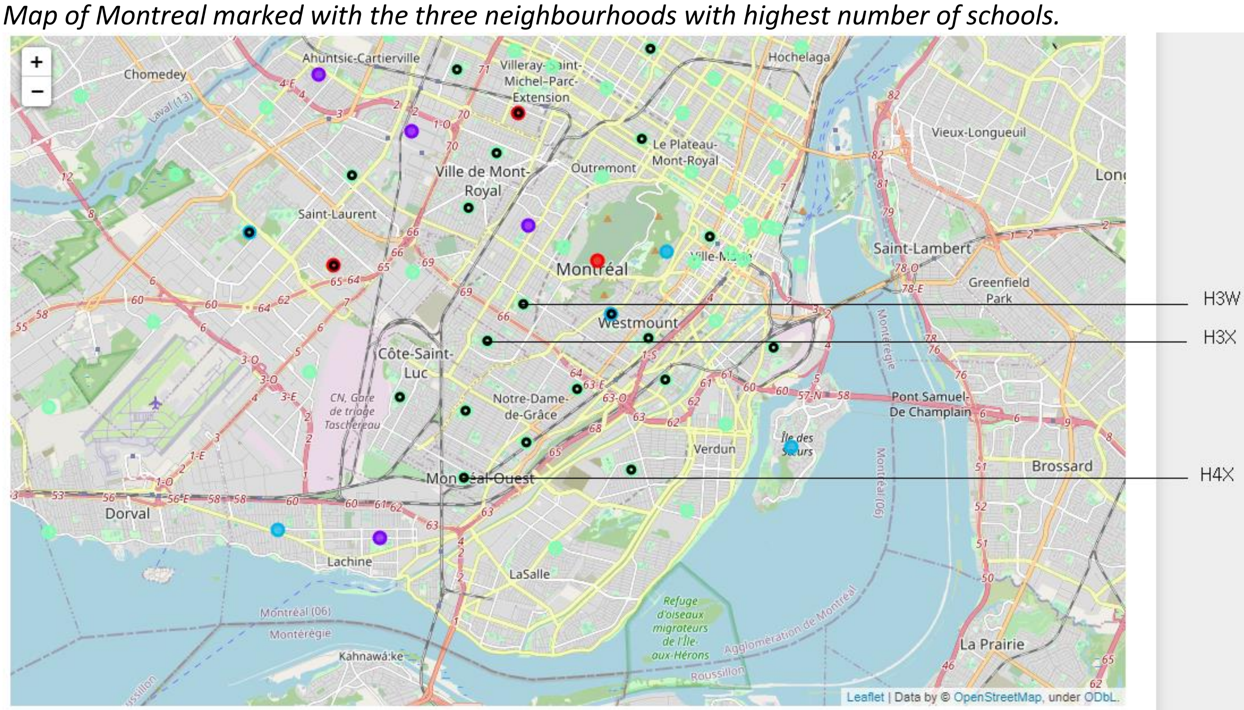
Neighbourhoods with highest number of schools.

Postal Code	Neighbourhood	Schools	Cluster #
H3W	Côte-des-Neiges Southwest	3	3
H3X	Hampstead / Côte Saint-Luc	3	3
H4X	Montreal West	3	3

We can see they all belong to the same clusters and are geographically adjacent on the map. Visually H3W (Côte-des-Neiges Southwest) is on the edge of what would be a downtown Montreal while H3X (Hampstead / Côte Saint-Luc) and H4X (Montreal West) cross into more residential areas.

Figure 18

Map of Montreal marked with the three neighbourhoods with highest number of schools.



Let's take a deeper look into school zones within the clusters. To do so, some final data wrangling is required. We merge the schools data from Figure 11 with Montreal venues data encoded for clustering, using the pandas merge function done on ‘Postal Code’ column to get: neighbourhoods , number of schools and cluster labels. Before viewing lets sort it by the number of schools in decending order.

Figure 19

Number of schools in each neighbourhood sorted in descending order.

	Postal Code	School	Neighbourhood_x	Cluster Labels
35	H4X	3	Montreal West	3
23	H3X	3	Hampstead / Côte Saint-Luc	3
22	H3W	3	Côte-des-NeigesSouthwest	3
19	H3N	2	Parc-Extension	0
31	H4M	2	Saint-LaurentEast	0
30	H4L	2	Saint-LaurentInner Northeast	3
8	H1S	2	Saint-LéonardSoutheast	1
1	H1E	2	Rivière-des-PrairiesSouthwest	1
17	H3A	2	Downtown Montreal North(McGill University)	3
16	H2T	2	Plateau Mont-RoyalWest	3
33	H4V	2	Côte Saint-LucEast	3
2	H1G	2	Montréal-NordNorth	3
34	H4W	1	Côte Saint-LucWest	3

We can see that cluster 3 has 8 out of the top 12 neighbourhoods with the highest number of schools. The fourth and fifth positions belong to cluster 0, where Saint-Laurent East would be a good option outside of Montreal.

Conclusion

With trying to address the question of which neighborhoods young families should be looking into when considering the move out of downtown Montreal, we can advise them to focus on the Côte-des-Neiges Southwest, Hampstead / Côte Saint-Luc along with the adjacent, same cluster, area of Westmount South . Real estate agents could find plenty of selling points for the neighbourhood, families would be moving out of the downtown core, away from the hustle and bustle but not too far, which is convenient for those whose work is still downtown. They would have access to key services such pharmacies and convenience stores, as well as restaurants. More importantly, there would be parks, a pool, athletic venues and several elementary and high schools; an ecosystem to allow young families to grow.

References

Century 21 Canada. (n.d.). Price per square Foot SURVEY 2020. Retrieved March 05, 2021, from http://www.century21franchise.ca/price-per-square-foot-survey-2020/

Government of Canada, S. (2019, April 03). Census in Brief: municipalities in Canada with the largest and fastest-growing populations between 2011 and 2016 CENSUS In Brief: municipalities in Canada with the largest and fastest-growing populations between 2011 and 2016. Retrieved March 02, 2021, from https://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016001/98-200-x2016001-eng.cfm

Gupta, A. (2021, February 09). Elbow method for optimal value of k in KMeans (1231919082 915109815 D. Majumder, Ed.). Retrieved March 05, 2021, from https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/

Hanes, A. (2020, May 28). Allison Hanes: Call of the suburbs gets loud during age of covid-19. Retrieved March 02, 2021, from https://montrealgazette.com/opinion/columnists/allison-hanes-call-of-the-suburbs-gets-loud-during-age-of-covid-19

Scikit-learn developers. (n.d.). Sklearn.Cluster.Kmeans. Retrieved March 04, 2021, from https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html