Clustering the boroughs of Montreal city

Abdesamad Sammah

July 27, 2020

A. Introduction

A1. Discussion of the background

Quebec, the Canadian province, is receiving yearly more than 40 000 new residents, and more than 70% of them (approximately 28 000) choose to stay in Montreal region.

The main reason for these immigrants is that Montreal, as a big and multicultural city, is considered a very good starting point, especially for people who don't know yet where the want to settle permanently.

One of the biggest preoccupations for the new immigrants is to minimize their spending during their job search on first 2-3 months, specially, the rent and the transportation bills. Any apartment with a lower cost of rent isn't the ideal solution, but there must be a compromise between the rent and the proximity of the services related to the chosen apartment.

Depending on their budgets, their profiles (individual, couple or family), thus, their needs, deciding on the installation area could easily become a headache for the new immigrants, and the objective of this project is to help them decide in which neighborhood of Montreal they should settle in, according to their budget for the rent and the services they need to be close to them.

A2. Data description

To help giving a solution to the problem, below the data we will be using:

- Rental price by borough of Montreal city, found at the Open-Data Portal of the city of Montreal (official reference in French)

- Geolocation of the boroughs of the city of Montreal, found at the Open-Data Portal of the Quebec government

- Foursquare to get the most common venues of given boroughs of Montreal

B. Methodology

B.1 Data gathering

B1.1 Data gathering

First, we gathered the information about the average rent cost for each borough in Montreal city, the data is available in a text format on the Montreal city portal, so we casted them into a data frame to make them easier for use and manipulation.

Then we got the locations for each of the 34 boroughs and related cities within Montreal island. Then, based on these locations, we extracted -via Foursquare API- the venues within each of these boroughs.

Some limits were defined for the venues data extraction in terms of:

- Maximum number of venues: 100
- Radius: 1000 meters

To help us handling this operation, a function was defined that insure the following tasks:

- Create the URL used to push the venues request to the Foursquare API
- Getting the request results
- Cast the results into a data frame to make the data manipulation easier

B1.2 Data preprocessing

All the findings were merged in the same data frame, the foursquare returns 382 rows, each row corresponds to a venue. A venue categories synthesis was done in order to get more relevant and more useful data for our study.

The foursquare returns NaN as the value of venues number for the borough “Mercier-Hochelaga-Maisonneuve”, we replaced this cell value by 0, idem for the Venues_category_count value about the same borough.

B.2 Data exploration

B2.1: General statistics:

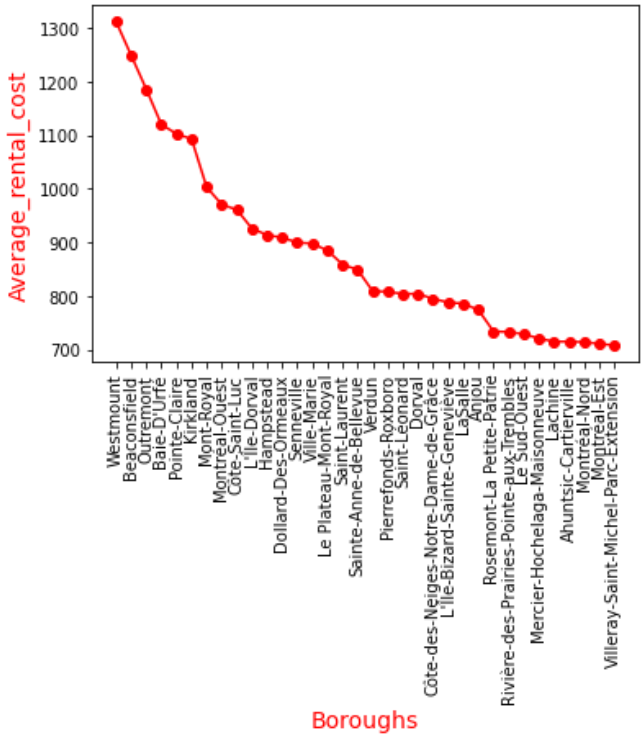
Regarding to our specifications *(100 venues within a radius of 1000 metres for each borough)*, we got:

- 382 venues in total
- 134 unique venues categories

B2.2: Top 5 boroughs with the highest rental cost

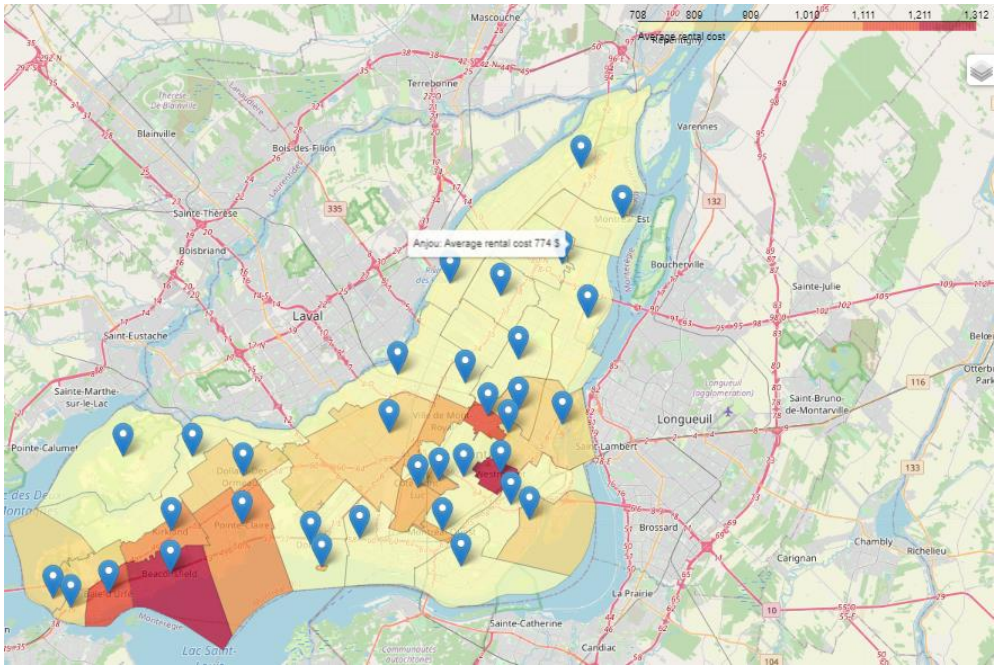
Based on the average rental cost, we found that **the top 5 boroughs with the highest rental cost** are**:**

- Westmount
- Beaconsfield
- Outremont
- Baie-D'Urfé
- Pointe-Claire



Visualizing the rental cost data on a choropleth map:

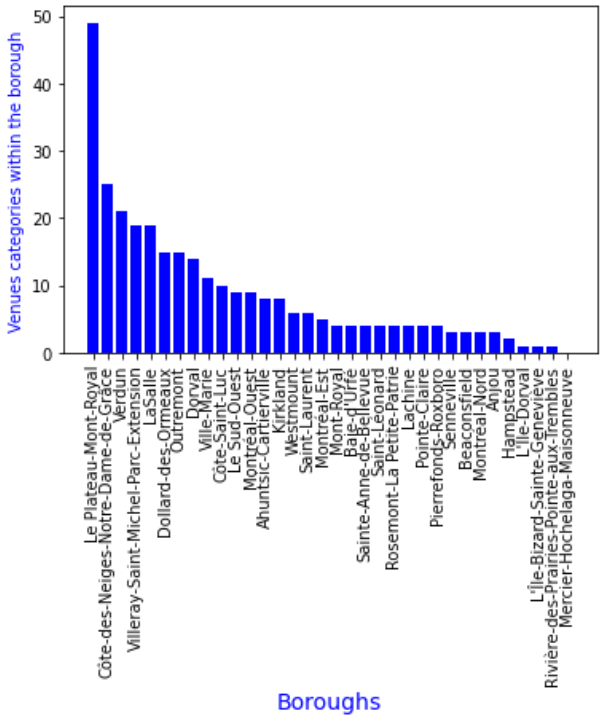
Most of the boroughs have an average rental cost under **950$**.



B2.3: Top 5 varied districts

In terms of variety of venues categories, the top 5 boroughs are:

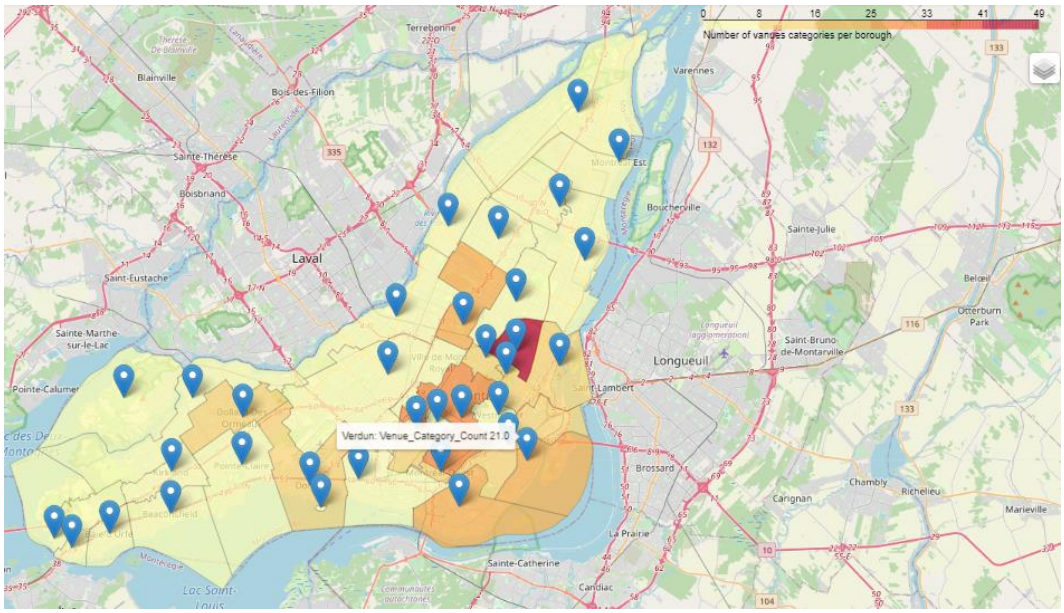
- Le Plateau-Mont-Royal
- Côte-des-Neiges-Notre-Dame-de-Grâce
- Verdun
- Villeray-Saint-Michel-Parc-Extension and
- LaSalle



These boroughs are the same boroughs with higher numbers of venues, whatever the categories of the venues.

Visualizing Venue_category_count on a choropleth map:

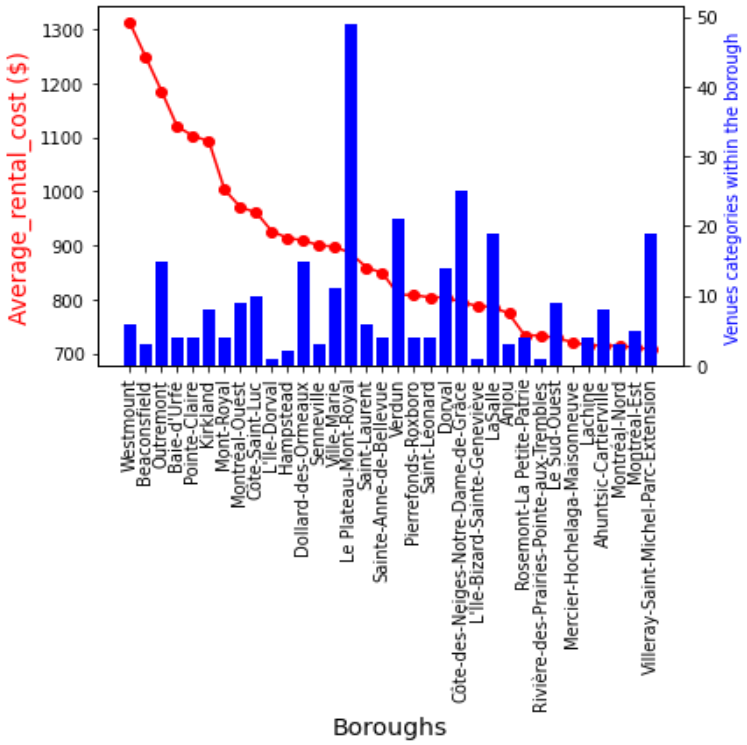
Boroughs with higher number of venues categories are concentrated in the **center** and the **south** of the city.



B2.4: Combining both information

Let's plot all the information available so far in the same graphic and see how the boroughs compare

From the graphic, we can see that most of boroughs having a high number of venues are roughly in the middle of the range of the rental cost.



B.3 Clustering Montreal Boroughs

To more help immigrants planning to settle in Montreal, a global comparison of the boroughs should be done, based on their average rental cost and the information about the venues they encompass. Therefore, clustering the boroughs based on these data is a good avenue.

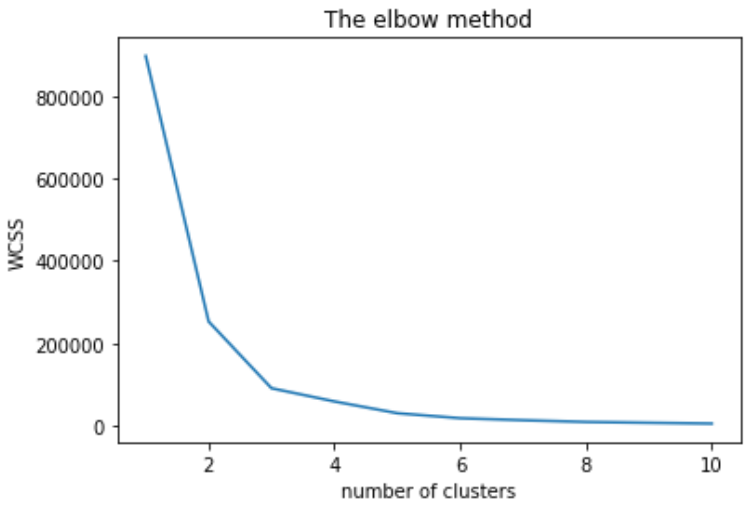
For this purpose, we will be applying the k-means algorithm, as it's a clustering algorithm that will help us group boroughs in different clusters so we can define how similar boroughs are, according to the cluster they belong to.

B3.1: Defining the number of clusters K

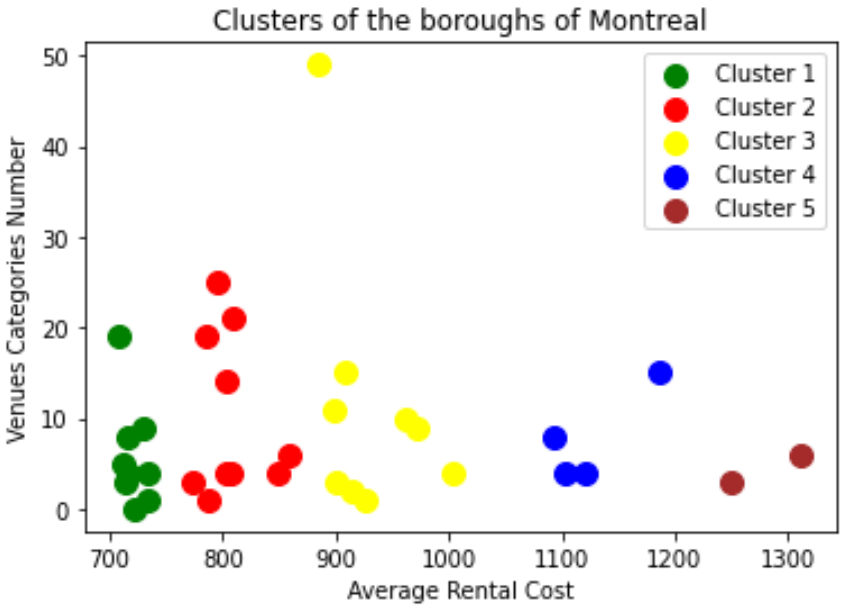
Using the elbow method to find the optimal number of clusters K

Where **WCSS** refers to: within-Cluster-Sum-of-Squares

From the graphic, we conclude that the optimal value for K is **5.**



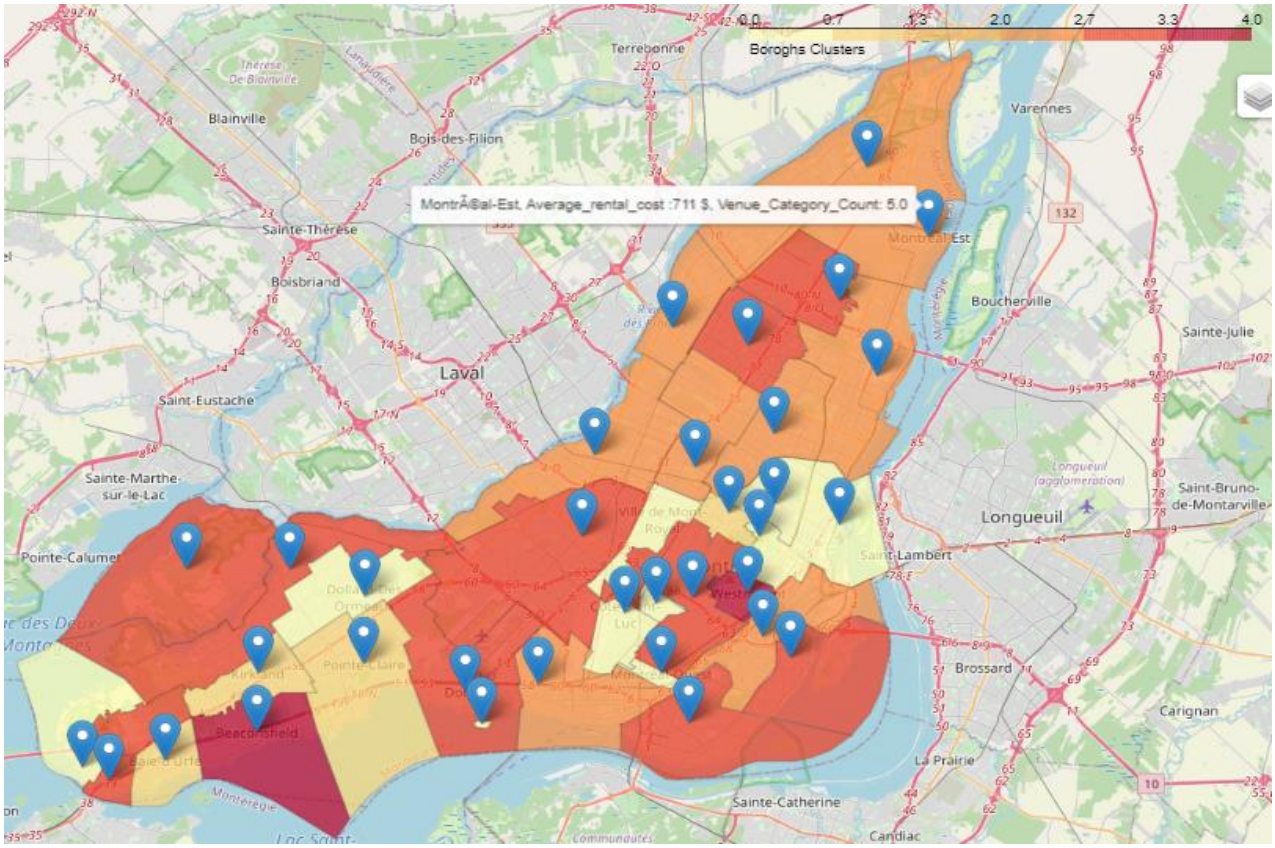
B3.2: Boroughs clusters



The table below present the different variable ranges for each cluster:

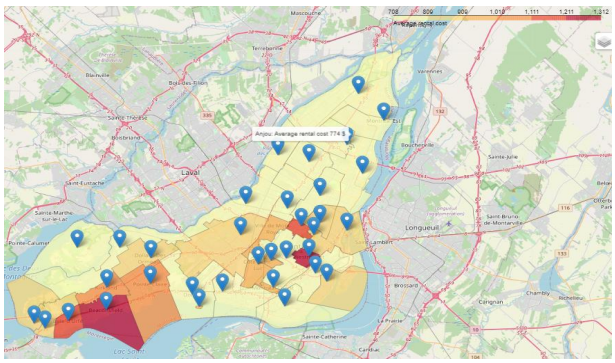
Cluster	Venues_Categories	Average_Rental_Cost ($)
1	0 - 19	708 – 734
2	1 - 25	774 – 859
3	1 - 49	885 – 1003
4	4 - 15	1093 – 1186
5	3 - 6	1249 – 1312

B3.2: Clusters choropleth map

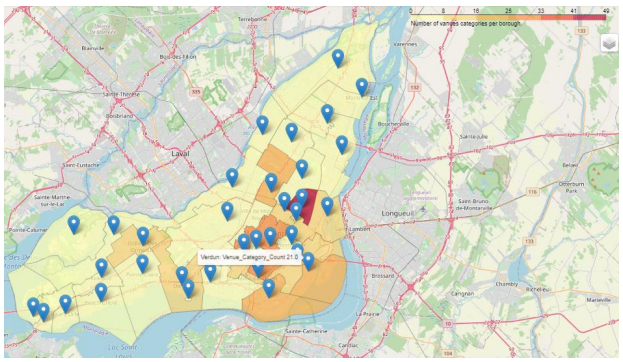


Unlike the first graphics and considering both information about rental cost and venues, similar boroughs are slightly scattered all over the city.

Rental cost choropleth map



Venues categories count choropleth map



C. Conclusion and future directions

C.1 Conclusion

We can objectively say that the boroughs located on the periphery of the island still more affordable and having a fair number of venues categories.

The following table present what boroughs compose each cluster:

Cluster	Boroughs list
1	• Côte-Saint-Luc • Dollard-des-Ormeaux • Hampstead • L'Île-Dorval • Le Plateau-Mont-Royal • Mont-Royal • Montréal-Ouest • Senneville • Ville-Marie
2	• Baie-d'Urfé • Kirkland • Outremont • Pointe-Claire
3	• Ahuntsic-Cartierville • Lachine • Le Sud-Ouest • Mercier-Hochelaga-Maisonneuve • Montréal-Est • Montréal-Nord • Rivière-des-Prairies-Pointe-aux-Trembles • Rosemont-La Petite-Patrie • Villeray-Saint-Michel-Parc-Extension
4	• Anjou • Côte-des-Neiges-Notre-Dame-de-Grâce • Dorval • L'Île-Bizard-Sainte-Geneviève • LaSalle • Pierrefonds-Roxboro • Saint-Laurent • Saint-Léonard • Sainte-Anne-de-Bellevue • Verdun
5	• Beaconsfield • Westmount

C.2 Future directions

This study could be deepened further, by including the elements below:

- More details about the venues.
- Rental availability.
- Job perspectives per borough (if available).

Eventually, the model underneath this study could be generalized to be used to cluster boroughs of any city, and surely by bringing it to a web-base platform so the accessibility could not be an issue for users.