

LOCATION OF A NEW BASEBALL STADIUM IN MONTREAL

Laurence Rioux-Fiset

December 15, 2020

1. Introduction

1.1 Background

Montreal is the most populous city in the province of Quebec in Canada. It's an island that cuts the Saint Lawrence River in half.

16 years ago, Montreal lost its baseball team, the Montreal Expos. At the end of the 1990s, Expos slowly disappeared from Montreal's sports landscape. Baseball fans lost interest in their team. That led to the relocation of the Expos. The team was moved to Washington D.C.

The major reason for this failure was the poor location of the Olympic Stadium. The stadium was far from population centres, restaurants, and bars. Basically, there was nothing to do around the stadium before and after the game. People could not eat at a good restaurant before the game, have a drink after, do some shopping, etc.

1.2 Problem

For 5 years, private investors have been looking to bring baseball back in town and to acquire a Major League Baseball (MLB) team. The condition from the MLB for a return of baseball in Montreal is a new stadium. For such a project, public investments from the government are needed. A perfect location needs to be found in order to convince both the government and the MLB that the return of baseball in Montreal is possible.

1.3 Interest

The goal of this project is to present to the Montreal City Council the optimal location for a new baseball stadium so public funds can be allocated to the project. The private investors need the help of data science to convince the Council of the optimal location.

For a location to be optimal, there must have restaurants, bars and shops around the new stadium. We will use a clustering model to cluster Montreal neighborhoods based on their similarities. We will make sure that the selected location does not share the same cluster as the old Olympic Stadium, previous home of the Montreal Expos.

2. Data

2.1 What we need

To address our problem, we will need the following information:

- **Population in each neighborhood:**

We want the stadium to be in a neighborhood with a high-density population. In Montreal, neighborhoods with big population are well served by public transportation and we want the stadium to be easily accessible. Having a lot of people living around the stadium would also evoke a sense of belonging to the team.

- **Number of restaurants in each neighborhood**
- **Number of bars in each neighborhood**
- **Number of stores / shops in each neighborhood**

These 3 venue categories are what we need around the new stadium based on Montreal's first experience with the Expos and the Olympic Stadium

2.2 Source of data

- First, we need to define the neighborhoods of Montreal. We will use the postal codes to divide the city into neighborhoods. The complete list of postal codes can be found here : [Wiki Montreal Postal Codes](#) Some postal codes are not assigned. They will be ignored. Example of a neighborhood: *H1A – Pointes-aux-Trembles*
- Second, we need the coordinates of each neighborhood. We will use Python Geocoder library. For each postal code obtained in first step, Geocoder will return its latitude and longitude.
- Third, we need the number of restaurants, bars, stores and shops in each neighborhood. We will use Foursquare API. We will explore each neighborhood in order to get nearby venues for these categories.
- Fourth, we need the population for each neighborhood. We will use Statistics Canada 2016 Census. Each postal code in Canada has its corresponding population. We will extract the population for each postal code of our dataset : [Population and dwelling counts](#). Example: *H1A – Pointes-aux-Trembles has a population of 32,516.*

3. Methodology

3.1 Master data

The main data frame was built with the data sources mentioned above. The components are *Postal Code, Neighborhood, Latitude, Longitude* and *Population*:

	PostalCode	Neighborhood	Latitude	Longitude	Population
0	H1A	Pointe-aux-Trembles	45.67474	-73.50009	32516
1	H2A	Saint-Michel,East	45.56153	-73.59985	19651
2	H3A	Downtown Montreal North	45.50533	-73.57488	3580
3	H4A	Notre-Dame-de-Grâce Northeast	45.47241	-73.61586	26453
4	H5A	Place Bonaventure	45.49878	-73.56580	0

In order to have a stadium in a populous neighborhood, we decided to sort the neighborhoods by population in decreasing order and to keep only the top 60:

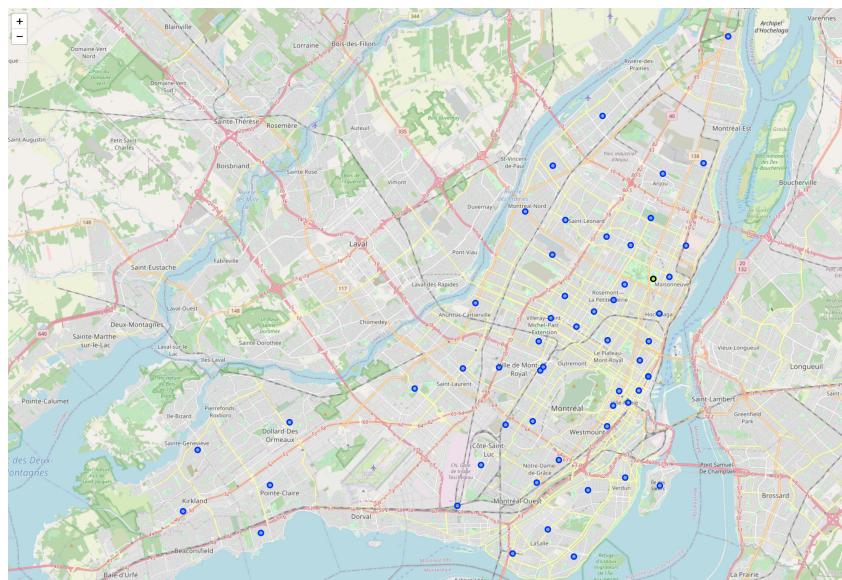
	PostalCode	Neighborhood	Latitude	Longitude	Population
26	H1G	Montréal-Nord North	45.61298	-73.61998	48556
20	H1E	Rivière-des-Prairies Southwest	45.63676	-73.58604	42420
64	H7N	Laval-des-Rapides	45.51240	-73.55469	42318
102	H7W	Chomedey South	45.51240	-73.55469	41619
54	H7L	Sainte-Rose	45.51240	-73.55469	37252

Finally we decided to add the 5 neighborhoods in Downtown Montreal. They don't have a big population but a lot of people travel to downtown everyday (workers, tourists, etc):

	PostalCode	Neighborhood	Latitude	Longitude	Population
60	H3A	Downtown Montreal North	45.50533	-73.57488	3580
61	H3B	Downtown Montreal East	45.49988	-73.56868	581
62	H3G	Downtown Montreal Southeast	45.49844	-73.57879	7457
63	H3H	Downtown Montreal Southwest	45.48846	-73.58296	19025
64	H2Z	Downtown Montreal Northeast	45.50553	-73.56133	2313

3.2 Data analysis

Using Python Folium library, we can visualize a map of Montreal with remaining neighborhoods and old Olympic Stadium (in yellow) on top of it:



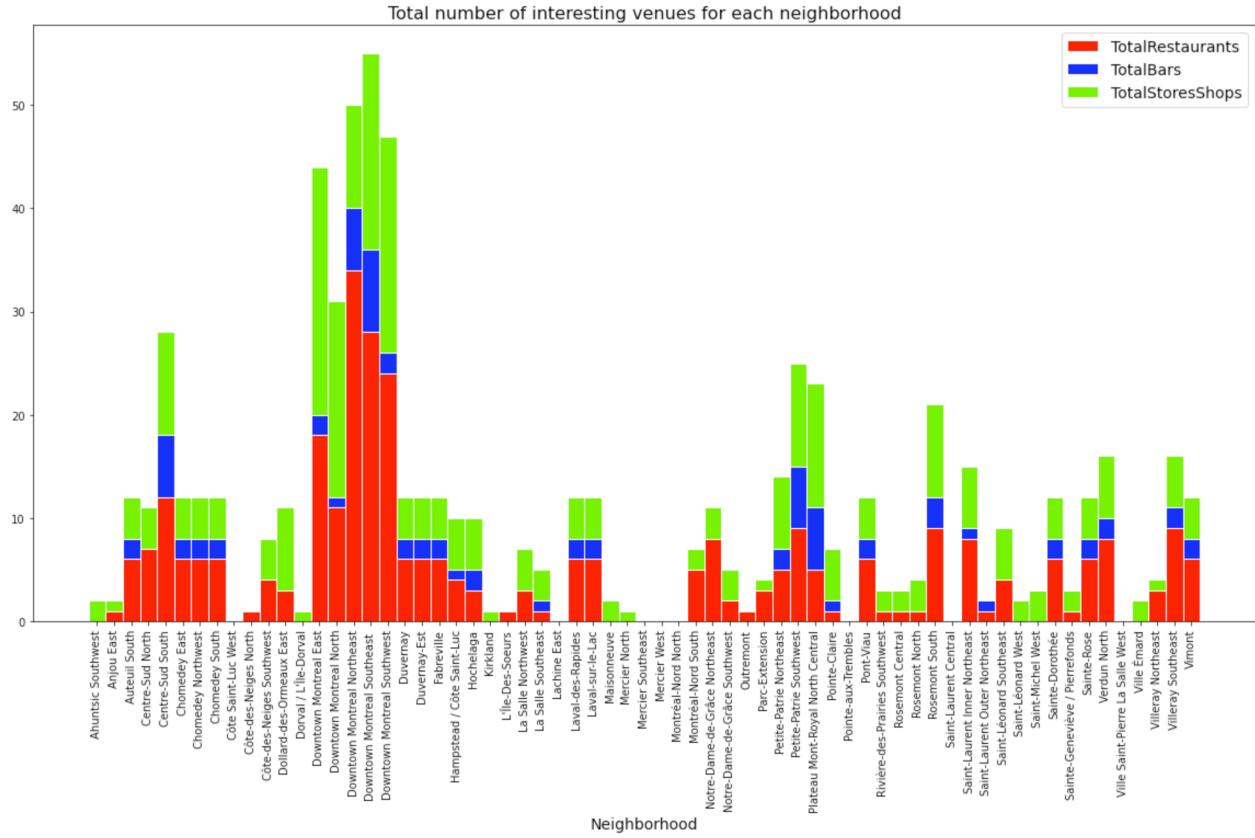
Using Foursquare API, we collected the number of restaurants, bars and stores/shops for each neighborhood. We used a radius of **500** meters and a limit of **100** venues.

Foursquare returned a total of **203** unique categories for venues.

We are only interested in 3 main categories so we grouped all categories with the word “Restaurant” under a general Restaurant category, all categories with the word “Bar” under a general Bar category and all categories with the word “Shop” or “Store” under a general Shop/Store category. **42** categories were grouped together as “Restaurant”, **12** categories were grouped together as “Bar” and **52** categories were grouped together as “Shop/Store”. For example, Centre-Sud South has a total of 12 restaurants. They can be any kind of restaurants (Indian, Thai, Mexican, etc).

	Neighborhood	Total Restaurants	Total Bars	Total Stores/Shop
0	Ahuntsic Southwest	0	0	2
1	Anjou East	1	0	1
2	Auteuil South	6	2	4
3	Centre-Sud North	7	0	4
4	Centre-Sud South	12	6	10

The bar chart in the following page allows us to compare the neighborhoods based on the total number of venues for these 3 categories. The categories are stacked so we can also compare the total number of venues. As expected, neighborhoods in downtown have the highest totals. Usually, downtown regroups a lot of bars, restaurants and shops. We can also see that some neighborhoods outside downtown Montreal have a lot of venues compared to the others: *Centre-Sud South, Petite-Patrie Southwest, Plateau-Mont-Royal North Central and Rosemont South*. We can guess that neighborhoods with no venue are strict residential areas.



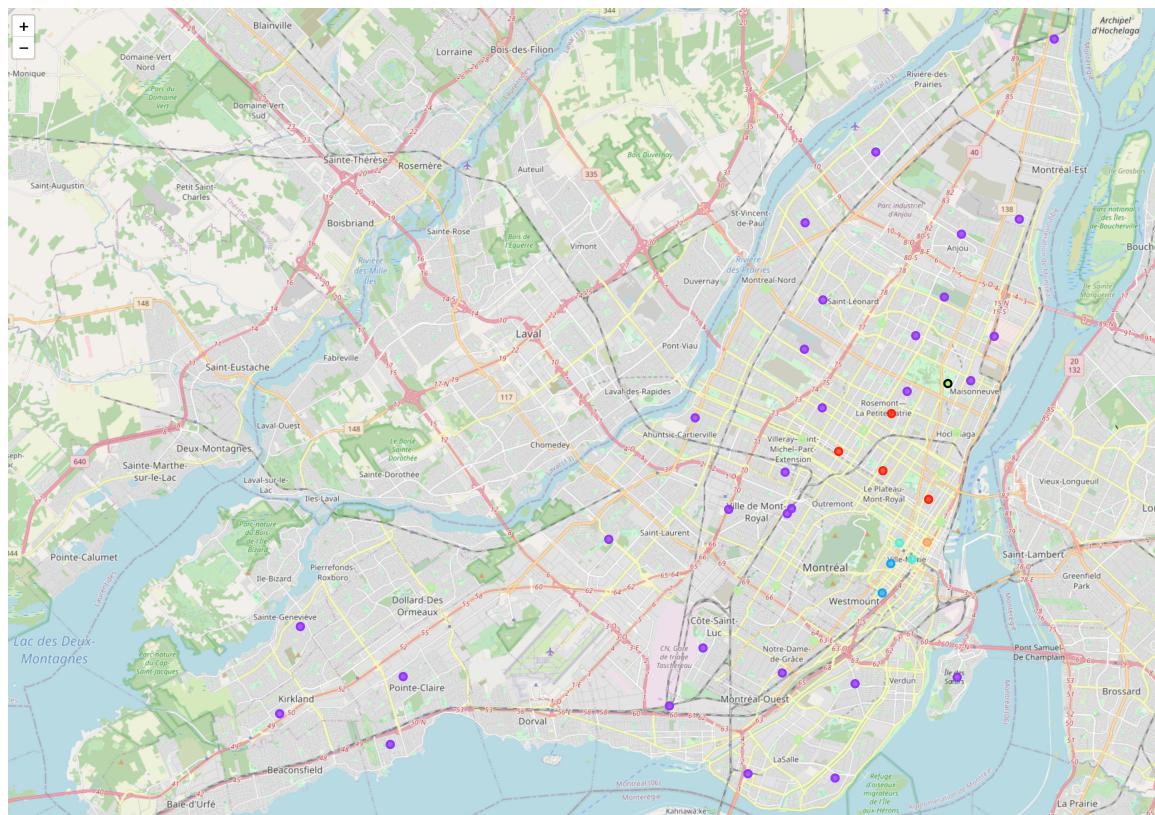
3.3 Machine learning

In this project, we want to cluster together similar neighborhoods based on the number of restaurants, bars and stores. Our goal is to select one of these clusters of neighborhoods for which the location of our new stadium is optimal. For this reason, we decided to use k-means clustering algorithm, one of the most popular algorithms for unsupervised machine learning.

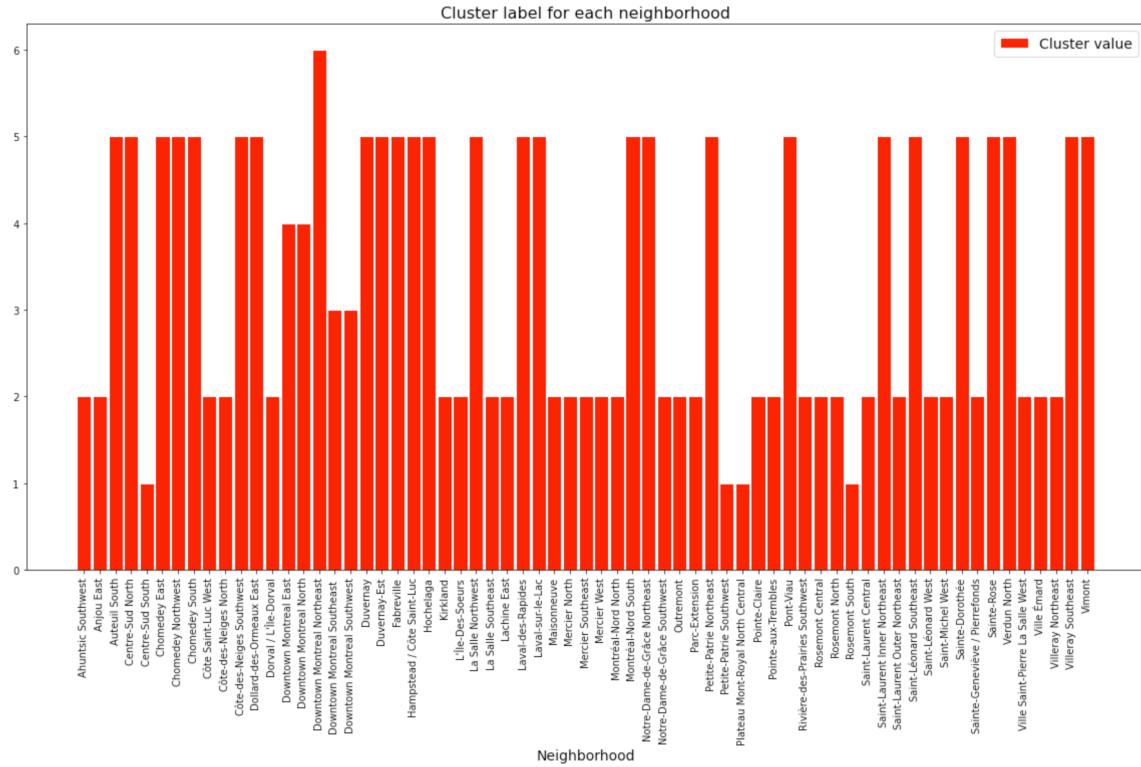
Based on the bar chart above, we decided to cluster the neighborhoods into 6 groups, meaning k=6. Here is the final table with cluster results:

	PostalCode	Neighborhood	Latitude	Longitude	Population	Cluster Labels	TotalRestaurants	TotalBars	TotalStoresShops
0	H1G	Montréal-Nord North	45.61298	-73.61998	48556	1	0	0	0
1	H1E	Rivière-des-Prairies Southwest	45.63676	-73.58604	42420	1	1	0	2
2	H7N	Laval-des-Rapides	45.51240	-73.55469	42318	4	6	2	4
3	H7W	Chomedey South	45.51240	-73.55469	41619	4	6	2	4
4	H7L	Sainte-Rose	45.51240	-73.55469	37252	4	6	2	4
...
60	H3A	Downtown Montreal North	45.50533	-73.57488	3580	3	11	1	19
61	H3B	Downtown Montreal East	45.49988	-73.56868	581	3	18	2	24
62	H3G	Downtown Montreal Southeast	45.49844	-73.57879	7457	2	28	8	19
63	H3H	Downtown Montreal Southwest	45.48846	-73.58296	19025	2	24	2	21
64	H2Z	Downtown Montreal Northeast	45.50533	-73.56133	2313	5	34	6	10

Using Python Folium library, we can visualize again a map of Montreal but this time with the 6 clusters and the old Olympic Stadium (in yellow) on top of it:



The following bar chart shows the cluster label for each neighborhood:



4. Results

Cluster 1:

	Neighborhood	Population	Cluster Labels	Total Restaurants	Total Bars	Total Stores Shops
29	Petite-Patrie Southwest	27359	0	9	6	10
30	Rosemont South	27172	0	9	3	9
32	Plateau Mont-Royal North Central	26559	0	5	6	12
41	Centre-Sud South	24222	0	12	6	10

This cluster has neighborhoods with a lot of interesting venues compared to the other clusters. Not as many as “Downtown” clusters but this is the perfect compromise if it’s not possible to build something as big as a stadium in downtown Montreal. These 4 neighborhoods have a lot of different options for each selected venue categories and the population is big enough to have a lot of people living around the stadium to evoke a sense of belonging to the team.

Cluster 2:

	Neighborhood	Population	Cluster Labels	TotalRestaurants	TotalBars	TotalStoresShops
0	Montréal-Nord North	48556	1	0	0	0
1	Rivière-des-Prairies Southwest	42420	1	1	0	2
7	Saint-Michel West	35472	1	0	0	3
9	Anjou East	34821	1	1	0	1
10	Mercier North	33468	1	0	0	1
11	Côte-des-Neiges North	33125	1	1	0	0
12	Pointe-aux-Trembles	32516	1	0	0	0
13	Rosemont North	31338	1	1	0	3
14	Outremont	31323	1	1	0	0
15	Saint-Léonard West	30903	1	0	0	2
17	Ville Émard	30513	1	0	0	2
18	Rosemont Central	30366	1	1	0	2
20	Ville Saint-Pierre La Salle West	29913	1	0	0	0
21	Saint-Laurent Outer Northeast	29755	1	1	1	0
24	Mercier West	29359	1	0	0	0
27	Parc-Extension	28277	1	3	0	1
31	Saint-Laurent Central	27065	1	0	0	0
37	Sainte-Geneviève / Pierrefonds	25449	1	1	0	2
38	Côte Saint-Luc West	24742	1	0	0	0
39	Lachine East	24407	1	0	0	0
40	Villeray Northeast	24297	1	3	0	1
43	Mercier Southeast	23870	1	0	0	0
44	Pointe-Claire	23861	1	1	1	5
46	La Salle Southeast	22706	1	1	1	3
48	Dorval / L'Île-Dorval	22580	1	0	0	1
51	Maisonneuve	22086	1	0	0	2
53	Kirkland	21728	1	0	0	1
57	Ahuntsic Southwest	20636	1	0	0	2
58	L'Île-Des-Soeurs	20461	1	1	0	0
59	Notre-Dame-de-Grâce Southwest	20368	1	2	0	3

This cluster contains residential areas with a very small number of interesting venues. A lot of people live there but there is nothing to do in terms of bars, restaurants and shops. The old Olympic Stadium is part of this cluster. Its location was the major reason why we lost the Montreal Expos so we absolutely need to avoid this cluster. We learn from our past experience and we don't want to repeat the same mistake.

Cluster 3:

	Neighborhood	Population	Cluster Labels	TotalRestaurants	TotalBars	TotalStoresShops
62	Downtown Montreal Southeast	7457	2	28	8	19
63	Downtown Montreal Southwest	19025	2	24	2	21

We can qualify this cluster as “busy” downtown. This is where we got the maximum number of venues. This would be ideal for our new stadium but finding a vacant land big enough for a stadium is almost impossible.

Cluster 4:

	Neighborhood	Population	Cluster Labels	TotalRestaurants	TotalBars	TotalStoresShops
60	Downtown Montreal North	3580	3	11	1	19
61	Downtown Montreal East	581	3	18	2	24

This cluster is the “less busy” part of downtown Montreal. We have a lot of venues but not as many as cluster 3. We have the same problem as cluster 3: we won’t find any vacant land big enough for a stadium.

Cluster 5:

	Neighborhood	Population	Cluster Labels	TotalRestaurants	TotalBars	TotalStoresShops
64	Downtown Montreal Northeast	2313	5	34	6	10

This cluster is almost identical as cluster 3 but we have a lot more restaurants and less shops. This is the best neighborhood to have dinner in a restaurant. But as mentioned before, we want to avoid downtown for our stadium.

Cluster 6:

Cluster 6 is another “residential” cluster but we have more restaurants, bars and shops than cluster 2. The number of venues is not important enough to make this cluster appropriate for our new stadium compared to cluster 2.

	Neighborhood	Population	Cluster Labels	TotalRestaurants	TotalBars	TotalStoresShops
2	Laval-des-Rapides	42318	4	6	2	4
3	Chomedey South	41619	4	6	2	4
4	Sainte-Rose	37252	4	6	2	4
5	Saint-Laurent Inner Northeast	36795	4	8	1	6
6	Montréal-Nord South	36127	4	5	0	2
8	Fabreville	35056	4	6	2	4
16	Côte-des-Neiges Southwest	30678	4	4	0	4
19	Laval-sur-le-Lac	30051	4	6	2	4
22	Verdun North	29488	4	8	2	6
23	La Salle Northwest	29465	4	3	0	4
25	Hochelaga	28889	4	3	2	5
26	Centre-Sud North	28727	4	7	0	4
28	Vimont	27898	4	6	2	4
33	Notre-Dame-de-Grâce Northeast	26453	4	8	0	3
34	Duvernay	26385	4	6	2	4
35	Sainte-Dorothée	26124	4	6	2	4
36	Saint-Léonard Southeast	25737	4	4	0	5
42	Chomedey Northwest	24155	4	6	2	4
45	Hampstead / Côte Saint-Luc	23069	4	4	1	5
47	Chomedey East	22594	4	6	2	4
49	Villeray Southeast	22436	4	9	2	5
50	Auteuil South	22259	4	6	2	4
52	Petite-Patrie Northeast	21981	4	5	2	7
54	Pont-Viau	21157	4	6	2	4
55	Dollard-des-Ormeaux East	20943	4	3	0	8
56	Duvernay-Est	20918	4	6	2	4

5. Discussion

In light of these results, the first option would be to build the new stadium in one of the neighborhoods in downtown Montreal. That's where we can find most of the restaurants, bars and shops in Montreal. But space is a problem. A baseball stadium is very big. There is no vacant land in downtown Montreal that would allow us to build a new stadium. The cost of buying adjacent properties in order to demolish them and create enough space is

too high and expropriation is not an option. All these office spaces are required by the workers.

That's why we strongly suggest one of these 4 neighborhoods: *Centre-Sud South*, *Petite-Patrie Southwest*, *Plateau-Mont-Royal North Central* and *Rosemont South*. Vacant land can be found, the number of restaurants, bars and stores is really big compared to the other neighborhoods outside downtown and a lot of people live there, potential new fans for the baseball team. The best option is *Centre-Sud South* with a total of 28 interesting venues but the 3 other options are very similar in terms of venues and population. Some of these 4 neighborhoods even have parks. A lot of MLB stadiums were built inside parks.

6. Conclusion

In this report, we used a machine learning algorithm, k-means clustering, in order to support the decision of a new location for a baseball stadium in Montreal. With the help of data science, we showed that *Centre-Sud South*, *Petite-Patrie Southwest*, *Plateau-Mont-Royal North Central* and *Rosemont South* are the 4 optimal options considering the population, the number of bars, the number of restaurants and the number of shops and stores. These neighborhoods are quite different from the neighborhood of the old Olympic Stadium based on these aspects.

Our approach could be improved by considering vacant land in Montreal big enough for a baseball stadium. We could also considerate properties and lands for sale. The acquisition of a land is an important factor in the final cost of the project.

We could also considerate neighborhood in suburban area of Montreal. Some of these neighborhoods were very developed in the past years with a lot of restaurant and bar options.