Moving to Montreal: Where to buy a house?

Andres Leon Salas, 28-08-2021

1. Introduction

Moving to a new city is never easy. You find yourself in front of a very big decision. Where in this city I don't really know should I buy a house?

How could I know what makes each part of the city different or similar to each other? What can I expect to find in each borough of this new city?

It would be a lot easier if you only knew a bunch of things, like:

- How many parks, sport fields, schools, or any other kind of point of interest are there in a given borough.
- Is it easy to move around without a car? Public transit nearby? Bike paths? This benefits a lot big children and teenagers who can't drive.
- Is a certain borough more of a residential area or more of a commercial area?
- What's the price range for different types of housing for each borough?
- Which boroughs relates to each other, which ones really differ?

So many questions... So let's put Data Analysis to the rescue!

By clearly defining and characterizing each borough by answering all the questions above, anybody moving to Montreal will find in this analysis a great tool for making the big decision of choosing a specific neighborhood which satisfies its needs.

The primary objective of all this exercise is to compare the boroughs of the Montreal City in order to chose one where you can buy a new home, if you’re just moving to Montreal and you don’t know the city very well yet.

2. The Data

Fortunately, we can find a lot of the data needed on the City of Montreal *“Données ouvertes”* (*Open Data*) website: https://donnees.montreal.ca/

Here, we can find data about the boroughs names and limits, points of interest, green spaces, exterior installations (like sport fields, kids parks), bike paths, and Free Montreal City Wi-Fi. All this data is available as CSV databases and Geospatial data as GeoJson files.

We had to find some other important data from some other non-governmental websites, as the Public transport, acquired from the "Société de transport de Montréal" (metro and bus company) https://www.stm.info/fr/a-propos/developpeurs , and the Housing Prices for each borough, acquired from the Real estate company “Centris” Website https://www.centris.ca/en/tools/real-estate-statistics/montreal-island?uc=2 .

Lastly, If we want to know what’s the *“character”* of each borough, we will need to know what kind of venues we can find in each one of them. For this, we will be making use of the available FourSquare data.

2.1 Data description

Data from the Montreal Open Data website:

- **liste-arrondissements.csv:** contains the names of the 19 boroughs under the Montreal City municipality.
- **limadmin.geojson:** Geographical polygons for the administrative limits of the 19 boroughs and 15 independent cities inside the Montreal Island.
- **lieux_d_interet.csv:** Names, Categories, Family, Coordinates, Zip codes and Boroughs of Municipal points of interest, such as Museums, Markets, Schools, Universities, Cultural and religious establishments, Community Centers, Libraries, Pools, Hotels, Lodging, Community usage equipment, etc.
- **lieux_d_interet.geojson:** Geographical polygons, Names, Categories, Boroughs for the Municipal points of interest.
- **espace_vert.csv:** Names, Categories, Family, Types, Coordinates, Address, Zip codes and Boroughs of Municipal Green Spaces, such as Nature parks, Metropolitan parks, Urban parks, School parks, District parks, Neighborhoods parks, Institutional spaces, Cemetery, Road green islands.
- **espace_vert.json:** Geographical polygons, Names, Categories, Types, Address, Zip codes and Boroughs of Municipal Green Spaces
- **terrain_sport_ext.csv:** Name, Borough and Type of exterior sports and games fields, such as Kiddie parks, Soccer, Basket, Volleyball, Baseball and Tennis fields, Exterior hockey rinks, Exterior pools, Kiddie pools, etc.
- **terrain_sport_ext.json:** Geographical polygons, Name, Borough and Type of exterior sports and games fields
- **mtlwifi_bornes.csv:** Placement (address) and coordinates of Montreal City free Wi-Fi access points.
- **mtlwifi_bornes.geojson:** Geographical coordinates points and placement of Montreal City free Wi-Fi access points.

- **reseau_cyclable.geojson:** Geographical polygons, Bi-directional, 4 seasons availability, 4 seasons protection for the Bicycle Paths
- **reseau-express-velo.geojson:** Geographical polygons, Bi-directional, 4 seasons availability, 4 seasons protection for the Express Bicycle Paths

Data from other sources:

- **Centris.ca:** Real estate company website to be scrapped for Single-family, Condominium and Plex (2 to 5 units) pricing for the 2nd Quarter 2021 and the Last 4 Quarters. Exemple page: https://www.centris.ca/en/tools/real-estate-statistics/montreal-island?uc=2
- **Stm_sig.zip:** From the "Société de transport de Montréal" (metro and bus company) website https://www.stm.info/fr/a-propos/developpeurs . This is a Shapefile containing names, coordinates, route number and much more info about all the bus stops and Metro Stations for the Montreal Island.
- **Foursquare API:** We will obtain from the Foursquare database the top 100 venues (if existing) for each borough.

3. Methodology

3.1 Data acquisition

All the data from the Montreal Open Data website was acquired by feeding one list of the csv URLs to a routine that reads each URL to a Pandas Dataframe, and one list of the geojson URLs to a routine that saves each json file with the help of the request library.

This way, for the csv files we obtained the following initial dataframes:

- `boroughs_csv : shape(19, 7)`
- `points_of_interest_csv : shape(3231, 16)`
- `green_spaces_csv : shape(2206, 11)`
- `exterior_installations_csv : shape(3469, 6)`
- `mtl_free_wifi_csv : shape(450, 9)`

And for the geojson files, files we obtained the following initial json dictionaries:

- `boroughs_json`
- `points_of_interest_json`
- `green_spaces_json`

- `exterior_installations_json`
- `mtl_free_wifi_json`
- `reseau_cyclable_json`
- `reseau_express_velo_json`

The public transport data from the "Société de transport de Montréal" was in Shapefile format. As this format is incompatible with Folium maps and is not very easily converted to a GeoJson format without the use heavy libraries, the compressed shapefile stm_sig.zip was downloaded and then converted on-line to a GeoJson format with the help of the "Free Online SHAPEFILE to JSON Converter", https://products.aspose.app/gis/conversion/shapefile-to-json . Then this geojson file, `stm_arrets_sig.geojson`, was loaded to our data as `metro_stations_json` with the help of the json library and into the Pandas dataframe `metro_stations_df`.

The Housing prices from the Centris.ca site were collected by first parsing the Wikipedia page listing all the 34 boroughs/cities in the Montreal Island https://en.wikipedia.org/wiki/Urban_agglomeration_of_Montreal#Subdivisions into a Beautiful Soup object and scrapping a list with the 34 names of the boroughs/cities. Then this names were refined/adapted into another list, used to generate the appropriate Centris.ca URLs for parsing into Beautiful Soup objects, one for each webpage listing the different categories of Housing prices for each of the 34 boroughs/cities. The 34 Beautiful Soup objects were then appended together, creating a list called `median_housing_prices` to be used later to add the pricings to some other Dataframes and GeoJsons.

The top 100 venues (if existing) for each borough were retrieved from the FourSquare database by calling the Foursquare API once for each one of the 19 Montreal City boroughs. Before using the Foursquare API, a list of the 19 boroughs was created by fixing a few borough names that were giving bad or null results when feed to the API call.

3.2 Data wrangling

Here’s where the biggest part of the work is executed.

Each initial dataframe was corrected/improved, originating another dataframe after being purged from non-very-relevant data columns, new columns were created, data was checked for inconsistencies, for example, the spelling of each borough/city name was normalized so it is the same on each instance.

For the json data, new jsons were derived from the existing ones, new data was inserted, borough/city names were normalized and big chunks of data not used were deleted.

Dataframes wrangling:

- **boroughs_csv:** shape(19, 7) :
  - **Original columns:** 'Nom officiel', 'Nom abrégé', 'Acronyme', 'Code 3L', 'ID-uadm', 'No-arro-élection', 'Code REM'
  - **Dropped columns:** 'Nom abrégé', 'Acronyme', 'Code 3L', 'ID-uadm', 'No-arro-élection', 'Code REM'
  - **Renamed columns:** 'Nom officiel'→'Borough'
  - **New dataframe name: boroughs_19_df**: shape(19, 1)

	Borough
0	Ahuntsic-Cartierville
1	Anjou
2	Côte-des-Neiges-Notre-Dame-de-Grâce

- **points_of_interest_csv:** shape(3231, 16):
  - **Original columns:** 'ID', 'Famille', 'Catégorie', 'Nom français', 'Nom court', 'Type', 'Numéro', 'rue', 'Étage', 'Bureau', 'Ville', 'Code postal', 'Arrondissement', 'Classification', 'Longitude', 'Latitude'
  - **Dropped columns:** 'ID', 'Famille', 'Catégorie', 'Nom court', 'Numéro', 'rue', 'Étage', 'Bureau', 'Ville', 'Code postal', 'Classification'
  - **Renamed columns:** 'Nom français'→'Name', 'Arrondissement'→'Borough'
  - **New columns:** 'Address' = 'Numéro' + 'rue' + 'Code postal'
  - **New columns:** 'Category' = filled with the string 'Point of Interest'
  - **Borough corrections**: 'Ville Mont-Royal'→'Mont-Royal', 'Ville de Westmount'→'Westmount'
  - **New dataframe name: points_of_interest_df** : shape(3231, 7):

	Name	Address	Latitude	Longitude	Borough	Type	Category
0	Maison du Meunier	10897 Rue du Pont, H2B 1Z4	45.574862	-73.661102	Ahuntsic-Cartierville	Musée et centre d'interprétation / d'exposition	Point of Interest
1	Maison du Pressoir	10865 Rue du Pressoir, H2B 2L1	45.575970	-73.659553	Ahuntsic-Cartierville	Musée et centre d'interprétation / d'exposition	Point of Interest
2	Marché Central	9187 Boulevard de l'Acadie, H4N 3K1	45.534023	-73.654956	Ahuntsic-Cartierville	Centre commercial	Point of Interest

- **green_spaces_csv:** shape(2206, 11):
    - **Original columns**: 'OBJECTID', 'Type', 'Lien', 'Nom', 'NUM_INDEX', 'SUPERFICIE', 'PROPRIETE', 'GESTION', 'COMPETENCE', 'TYPO1', 'TYPO2'
    - **Dropped columns:** 'OBJECTID', 'Type', 'Lien', 'Nom', 'NUM_INDEX', 'PROPRIETE', 'COMPETENCE', 'TYPO1'
    - **Renamed columns:** 'TYPO2'→'Type', 'SUPERFICIE'→'Area'
    - **New columns:** 'Name' = ' Type' + 'Lien' + 'Nom'
    - **New columns:** 'Category' = filled with the string 'Green Space'
    - **New columns:** 'Borough' = filled temporarily with the string 'To be corrected'
    - **Modified columns:** 'Borough' was filled with a list containing the boroughs names where each green space is located. This list was generated with the values of the key 'GESTION'(borough name) from green_spaces_json
    - **New dataframe name: green_spaces_df** : shape(2206, 5)

	Name	Borough	Area	Type	Category
0	Université Concordia - Campus Loyola	Côte-des-Neiges-Notre-Dame-de-Grâce	5.054930	Espace sportif	Green Space
1	Av. du Boisé	Côte-des-Neiges-Notre-Dame-de-Grâce	0.032364	Îlot de voirie	Green Space
2	Parc Maurice-Cullen	Côte-des-Neiges-Notre-Dame-de-Grâce	0.542681	Parc de quartier	Green Space

- **exterior_installations_csv:** shape(3469, 6):
    - **Original columns:** 'OBJECTID', 'ID', 'INDEX_PARC', 'NOM', 'ARROND', 'TYPE'
    - **Dropped columns:** 'OBJECTID', 'ID', 'INDEX_PARC'
    - **Renamed columns:** 'ARROND'→'Borough', 'NOM'→'Name', 'TYPE'→'Type'
    - **New columns:** 'Category' = filled with the string 'Exterior Installation'
    - **Borough corrections**: 'Verdum'→'Verdun'
    - **New dataframe name: exterior_installations_df** : shape(3469, 4)

	Name	Borough	Type	Category
0	Aire de jeu 18 mois à 5 ans	Le Plateau-Mont-Royal	Aire de jeu - enfant	Exterior Installation
1	Aire de jeu 18 mois à 5 ans	Le Plateau-Mont-Royal	Aire de jeu - enfant	Exterior Installation
2	Arbri soleil	Le Plateau-Mont-Royal	Plein air	Exterior Installation

- **mtl_free_wifi_csv:** shape(450, 9):
    - **Original columns:** 'ID', 'Lieu', 'Latitude', 'Longitude', 'X', 'Y', 'Type', 'Arrondissement', 'Zone active'
    - **Dropped columns:** 'ID', 'X', 'Y', 'Type', 'Zone active'

- **Renamed columns:** 'Lieu'→'Address', 'Arrondissement'→'Borough'
- **New columns:** 'Name' = str('MtlWifi') + 'Lieu'
- **New columns:** 'Category' = filled with the string 'Mtl Free Wifi '
- **New dataframe name: mtl_free_wifi_df** : shape(450, 6)

	Name	Address	Latitude	Longitude	Borough	Category
0	MtlWifi Rue Saint Jacques / Place d'Armes	Rue Saint Jacques / Place d'Armes	45.504720	-73.557920	Ville-Marie	Mtl Free Wifi
1	MtlWifi Rue Notre-Dames/ Place d'Armes	Rue Notre-Dames/ Place d'Armes	45.504860	-73.556600	Ville-Marie	Mtl Free Wifi
2	MtlWifi Rue Le Royer / boul. Saint-Laurent	Rue Le Royer / boul. Saint-Laurent	45.505780	-73.554670	Ville-Marie	Mtl Free Wifi

- **metro_stations_df:** shape(68, 4):
  - The dataframe Metro_stations_df was created from the metro_stations_json, by reading its keys 'stop_name', 'borough' and 'route_id' into the dataframe columns 'Name', 'Borough', 'Metro line'
  - **Original columns:** 'Name', 'Borough', 'Metro line', 'Category'
  - **New columns:** 'Category' = filled with the string 'Metro station'

	Name	Borough	Metro line	Category
0	Station Angrignon	Le Sud-Ouest	Green line	Metro station
1	Station Monk	Le Sud-Ouest	Green line	Metro station
2	Station Jolicoeur	Le Sud-Ouest	Green line	Metro station

- **boroughs_19_df:** shape(19, 7) :
  - Re-created by adding the Housing prices data from boroughs_19_json
  - **Original columns:** 'Borough', 'single Family 2nd Quarter 2021', 'Single Family last 4 Quarters', 'Condo 2nd Quarter 2021', 'Condo last 4 Quarters', 'Plex 2nd Quarter 2021', 'Plex last 4 Quarters'

	Borough	single Family 2nd Quarter 2021	single Family last 4 Quarters	Condo 2nd Quarter 2021	Condo last 4 Quarters	Plex 2nd Quarter 2021	Plex last 4 Quarters
0	Outremont	**	$2,180,000	$690,000	$671,500	**	**
1	LaSalle	**	$595,000	$380,500	$372,040	$756,500	$690,000
2	Ville-Marie	$1,530,000	$1,149,500	$449,000	$445,000	$819,000	$847,000

- **boroughs_15_df:** shape(15, 7) :
  - Created with the Borough Name and the Housing prices data from boroughs_15_json
  - **Original columns:** 'Borough', 'single Family 2nd Quarter 2021', 'Single Family last 4 Quarters', 'Condo 2nd Quarter 2021', 'Condo last 4 Quarters', 'Plex 2nd Quarter 2021', 'Plex last 4 Quarters'

	Borough	single Family 2nd Quarter 2021	single Family last 4 Quarters	Condo 2nd Quarter 2021	Condo last 4 Quarters	Plex 2nd Quarter 2021	Plex last 4 Quarters
0	Mont-Royal	$1,649,000	$1,661,500	$555,000	$526,900	**	**
1	Hampstead	**	$1,830,000	**	**	**	**
2	Dorval	$649,000	$575,000	$383,750	$360,000	**	**

- **boroughs_df:** shape(34, 7) :
  - Created by concatenating boroughs_19_df and boroughs_15_df. To be used as a reference dataframe, where we can compare the pricing over the whole Montreal Island in just one dataframe
  - **Original columns:** 'Borough', 'single Family 2nd Quarter 2021', 'Single Family last 4 Quarters', 'Condo 2nd Quarter 2021', 'Condo last 4 Quarters', 'Plex 2nd Quarter 2021', 'Plex last 4 Quarters'

	Borough	single Family 2nd Quarter 2021	single Family last 4 Quarters	Condo 2nd Quarter 2021	Condo last 4 Quarters	Plex 2nd Quarter 2021	Plex last 4 Quarters
0	Outremont	**	$2,180,000	$690,000	$671,500	**	**
1	LaSalle	**	$595,000	$380,500	$372,040	$756,500	$690,000
2	Ville-Marie	$1,530,000	$1,149,500	$449,000	$445,000	$819,000	$847,000

- **montreal_venues_df:** shape(1620, 6) :
  - Created by calling the Foursquare API database to obtain the top 100 venues (if existing) for each borough.
  - **Original columns:** 'Borough', 'Venue', 'Venue Latitude', 'Venue Longitude', 'Venue Address', 'Venue Category'

	Borough	Venue	Venue Latitude	Venue Longitude	Venue Address	Venue Category
0	Outremont	Mamie Clafoutis	45.521466	-73.613998	[1291 ave. Van Horne (coin av. Outremont), Mon...	Bakery
1	Outremont	Le Glacier Bilboquet	45.518993	-73.609737	[1311 ave. Bernard Ouest (coin Outremont), Mon...	Ice Cream Shop
2	Outremont	Damas	45.522596	-73.613112	[1201 Van Horne, Montréal QC H2V 1K4, Canada]	Mediterranean Restaurant

- **venues_per_neigh:** shape(19, 2) :
  - Exploratory dataframe created for checking how many venues per borough were foudd.
  - **Original columns:** 'Borough', 'Num of Venues'

	Borough	Num of Venues
0	Ahuntsic-Cartierville	100
1	Mercier-Hochelaga-Maisonneuve	100
2	Verdun	100

- **montreal_island_features_df:** shape(11026, 2) :
    - Exploratory dataframe created for listing all the venues found.
    - **Original columns:** 'Borough', ' Feature Category'

	Borough	Feature Category
0	Outremont	Bakery
1	Outremont	Ice Cream Shop
2	Outremont	Mediterranean Restaurant

- **montreal_onehot_grouped:** shape(19, 274) :
    - Exploratory dataframe where we can see as columns all the different Feature Categories we obtained. This dataframe will we feed to the K-means clustering algorithm to separate the 19 boroughs into 6 clusters
    - **Original columns:** 'Borough', 'x273 different Feature Categories

	Borough	Agence internationale	Aire de jeu - enfant	American Restaurant	Art Gallery	Art public	Arts & Crafts Store	Artère commerciale	Asian Restaurant	Athletics & Sports	...	Warehouse Store
0	Ahuntsic-Cartierville	0	61	0	0	1	1	2	1	1	...	0
1	Anjou	0	26	0	0	0	0	0	1	1	...	0
2	Côte-des-Neiges-Notre-Dame-de-Grâce	0	50	0	1	2	0	0	1	0	...	0

- **boroughs_features_sorted:** shape(19, 12) :
    - Dataframe created for identifying the 10 most common feature categories for the 19 Montreal City boroughs and the cluster they belong to
    - **Original columns:** 'Borough', '1st Most Common Feature', '2nd Most Common Feature', ….. , '10th Most Common Feature'
    - **New columns:** 'Cluster'. This column was inserted after using the Kmeans Clustering machine learning algorithm from the library Sklearn for classifying the 19 Montreal City boroughs into 6 clusters with similar common features.

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
0	Ahuntsic-Cartierville	5	Sportif	Récréatif	Aire de jeu - enfant	Parc de quartier	Parc	Loisirs et détente	Îlot de voirie	École primaire	Mtl Free Wifi	Organisme communautaire et centre de loisirs
1	Anjou	2	Sportif	Aire de jeu - enfant	Récréatif	Parc de quartier	Parc	Loisirs et détente	Green space	Pharmacy	École primaire	Restaurant
2	Côte-des-Neiges-Notre-Dame-de-Grâce	1	Sportif	Récréatif	Aire de jeu - enfant	Organisme communautaire et centre de loisirs	Parc	Parc de quartier	École primaire	École secondaire et professionnelle	Mtl Free Wifi	Loisirs et détente

- **Cluster[1] to Cluster[5]:** shape(X, 12) :
  - The 6 Cluster dataframes are contained in the list Cluster
  - Each Cluster[x] dataframe shows the Boroughs making part of the cluster and their 10 Most Common Featuers

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
6	Le Plateau-Mont-Royal	0	Aire de jeu - enfant	Parc	Récréatif	Parc de voisinage	Théâtre / salle de spectacle / petit lieu de d...	Loisirs et détente	Mtl Free Wifi	Îlot de verdure	Organisme communautaire et centre de loisirs	Sportif
7	Le Sud-Ouest	0	Parc	Aire de jeu - enfant	Mtl Free Wifi	Organisme communautaire et centre de loisirs	Récréatif	Îlot de verdure	Loisirs et détente	Parc de quartier	Sportif	Parc de voisinage
16	Verdun	0	Aire de jeu - enfant	Parc	Récréatif	Sportif	Green space	Parc de quartier	Parc de voisinage	Loisirs et détente	Espace vert	Plein air

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
2	Côte-des-Neiges-Notre-Dame-de-Grâce	1	Sportif	Récréatif	Aire de jeu - enfant	Organisme communautaire et centre de loisirs	Parc	Parc de quartier	École primaire	École secondaire et professionnelle	Mtl Free Wifi	Loisirs et détente
4	LaSalle	1	Aire de jeu - enfant	Récréatif	Sportif	Parc	Parc urbain	Parc de voisinage	Parc de quartier	Loisirs et détente	École primaire	Green space
13	Rosemont-La Petite-Patrie	1	Récréatif	Aire de jeu - enfant	Parc	Loisirs et détente	Sportif	Parc de quartier	École primaire	Mtl Free Wifi	Îlot de verdure	École secondaire et professionnelle
	Saint-		Aire de			Loisirs et				Parc de	École	

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
1	Anjou	2	Sportif	Aire de jeu - enfant	Récréatif	Parc de quartier	Parc	Loisirs et détente	Green space	Pharmacy	École primaire	Restaurant
3	L'Île-Bizard-Sainte-Geneviève	2	Parc	Aire de jeu - enfant	Sportif	Plein air	Parc de voisinage	Green space	Récréatif	Parc-nature	Parc de quartier	Îlot de verdure
5	Lachine	2	Sportif	Plein air	Récréatif	Parc	Aire de jeu - enfant	Mtl Free Wifi	Parc de quartier	Parc de voisinage	Loisirs et détente	Place publique
	Montréal-		Aire de jeu		Green			Parc de	Loisirs et	Mtl Free	École	

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
17	Ville-Marie	3	Mtl Free Wifi	Hôtel	Place publique	Poste d'attente de taxis	Art public	Parc	Aire de jeu - enfant	Autre catégorie d'établissement d'hébergement	Service consulaire	Bâtiment historique / patrimonial

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
11	Pierrefonds-Roxboro	4	Aire de jeu - enfant	Green space	Parc	Sportif	Parc de quartier	Parc de voisinage	Plein air	Récréatif	Parc-nature	Loisirs et détente

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
0	Ahuntsic-Cartierville	5	Sportif	Récréatif	Aire de jeu - enfant	Parc de quartier	Parc	Loisirs et détente	Îlot de voirie	École primaire	Mtl Free Wifi	Organisme communautaire et centre de loisirs
8	Mercier-Hochelaga-Maisonneuve	5	Aire de jeu - enfant	Récréatif	Sportif	Parc de quartier	Parc	Loisirs et détente	Organisme communautaire et centre de loisirs	Plein air	Mtl Free Wifi	École primaire
12	Rivière-des-Prairies-Pointe-aux-Trembles	5	Aire de jeu - enfant	Sportif	Parc	Parc de quartier	Récréatif	Îlot de voirie	Loisirs et détente	Mtl Free Wifi	Parc de voisinage	Îlot de verdure

Jsons wrangling:

- **boroughs_json:**
  - **Relevant Keys:** 'NOM', 'geometry'
  - **Borough corrections**: 'Ville Mont-Royal'→'Mont-Royal', 'Ville de Westmount'→'Westmount'
  - **New jsons derived:** boroughs_json was splitted into **boroughs_19_json** with the 19 Montreal city boroughs and into **boroughs_15_json** with the 15 independent cities

- **points_of_interest_json**
  - **Relevant Keys:** 'geometry', 'Nom français', 'Type', 'Catégorie', 'Arrondissement'

- **green_spaces_json:**
  - **Relevant Keys:** 'geometry', 'Nom', 'SUPERFICIE', 'TYPO2', 'GESTION'
  - **Modified Keys:** 'Nom' = ' Type' + 'Lien' + 'Nom'
  - **Modified Keys:** If 'Nom' = empty, then 'Nom' = 'No name'
  - **Modified Keys:** If 'TYPO2' = empty, then 'TYPO2'= 'Green space'
  - **Modified Keys:** 'GESTION' was filled with the value of the key 'NOM'(borough name) from boroughs_json corresponding to the Geographical intersection between the value of the key 'geometry' (geographical limits polygon) from green_spaces_json and 'geometry' from boroughs_json. This was done in order to find out and then assign to each green space the borough where it is located. The Turfpy library was used for this task.

- **exterior_installations_json:**
  - **Relevant Keys:** 'geometry', 'NOM', 'TYPE', 'ARROND'

- **mtl_free_wifi_json**
  - **Relevant Keys:** 'geometry', 'Lieu'

- **reseau_cyclable_json**
  - **Relevant Keys:** 'geometry', 'NBR_VOIE', 'SAISONS4', 'PROTEGE_4S', 'NOM_ARR_VI'

- **reseau_express_velo_json:**
  - **Relevant Keys:** 'geometry', 'NBR_VOIE', 'SAISONS4', 'PROTEGE_4S', 'NOM_ARR_VI'

- **metro_stations_json:**
    - **Relevant Keys:** 'geometry', 'stop_name', 'route_id', 'borough'
    - **Removed Features/Keys:** All the features with keys other than ('loc_type' == 0 and 'route_id' in ['1', '2', '4', '5', '1,2']) were removed. This was in order to remove all the bus stops, just keeping the Metro stations. If the bus stops were to be kept, the Folium map would be really overcrowded without providing any useful info as the bus stops are distributed very uniformly across the whole island.
    - **Modified Keys:** 'route_id':['1', '2', '4', '5', '1,2'] → 'route_id':['Green line', 'Orange line', 'Yellow line', 'Blue line', 'Green and Orange lines']
    - **New Keys:** 'borough' was filled with the value of the key 'NOM'(borough name) from boroughs_json corresponding to the Geographical intersection between the value of the key 'geometry' (geographical limits polygon) from metro_stations_json and 'geometry' from boroughs_json. This was done in order to find out and then assign to Metro Station the borough where it is located. The Turfpy library was used for this task.

- **boroughs_19_json:**
    - Created by splitting boroughs_json
    - **Relevant Keys:** 'geometry', 'NOM', 'singleFamily_2nd_Q_2021', 'singleFamily_last_4Qs', 'condo_2nd_Q_2021', 'condo_last_4Qs', 'plex_2nd_Q_2021', 'plex_last_4Qs'
    - **New Keys:** 'singleFamily_2nd_Q_2021', 'singleFamily_last_4Qs', 'condo_2nd_Q_2021', 'condo_last_4Qs', 'plex_2nd_Q_2021', 'plex_last_4Qs'. All the Housing Prices Values were added to its respective borough after creating a list of dictionaries each one containing the borough name and the 6 different housing prices. This list was created by webscrapping the Real Estate website Centris.ca.

- **boroughs_15_json:**
    - Created by splitting boroughs_json
    - **Relevant Keys:** 'geometry', 'NOM', 'singleFamily_2nd_Q_2021', 'singleFamily_last_4Qs', 'condo_2nd_Q_2021', 'condo_last_4Qs', 'plex_2nd_Q_2021', 'plex_last_4Qs'
    - **New Keys:** 'singleFamily_2nd_Q_2021', 'singleFamily_last_4Qs', 'condo_2nd_Q_2021', 'condo_last_4Qs', 'plex_2nd_Q_2021', 'plex_last_4Qs'. All the Housing Prices Values were added to its respective city after creating a list of dictionaries each one containing the city name and the 6 different housing prices. This list was created by webscrapping the Real Estate website Centris.ca.

3.3 Machine learning – *Clustering*

As stated on the introduction, the primary objective of all this exercise is to compare the boroughs of the Montreal City in order to chose one where you can buy a new home, if you're just moving to Montreal and you don't know the city very well yet.

One great way of identifying which boroughs are similar to each other, is by using Machine Learning techniques. One technique, or more precisely, one unsupervised machine learning algorithm that's up to the task is Kmeans Clustering, which is designed to group together entities with similar characteristics

By applying Kmeans Clustering we'll be able to classify the 19 Montreal City boroughs into 6 clusters with similar common features.

We can execute the clustering after obtaining a "one-hot" dataframe with all the different types/categories of features as columns for each one of the 19 boroughs. This dataframe was created and named as montreal_onehot_grouped.

Then, for the actual clustering, the montreal_onehot_grouped dataframe with a number of 6 clusters was passed to the KMeans function of the Sklearn machine learning library.

This cluster assignment information was then inserted as a column in the dataframe boroughs_features_sorted, for easy reference and identification.

4. Results

After all the data acquisition and wrangling, we can take a look and analyze the most relevant dataframes, which will help us make an educated decision on where to buy our new house.

Then, we can visualize everything together in a Folium Map, where we can turn on and off the main feature categories being displayed. This is great for visualizing spatial data such as the density of the features in your target borough or the proximity to certain feature you would like close to home.

boroughs_features_sorted: shape(19, 12) : Dataframe for identifying the 10 most common feature categories for the 19 Montreal City boroughs and the cluster they belong to. Not a huge difference between boroughs, except for Ville-Marie (downtown).

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
0	Ahuntsic-Cartierville	5	Sportif	Récréatif	Aire de jeu - enfant	Parc de quartier	Parc	Loisirs et détente	Îlot de voirie	École primaire	Mtl Free Wifi	Organisme communautaire et centre de loisirs
1	Anjou	2	Sportif	Aire de jeu - enfant	Récréatif	Parc de quartier	Parc	Loisirs et détente	Green space	Pharmacy	École primaire	Restaurant
2	Côte-des-Neiges-Notre-Dame-de-Grâce	1	Sportif	Récréatif	Aire de jeu - enfant	Organisme communautaire et centre de loisirs	Parc	Parc de quartier	École primaire	École secondaire et professionnelle	Mtl Free Wifi	Loisirs et détente
3	L'Île-Bizard-Sainte-Geneviève	2	Parc	Aire de jeu - enfant	Sportif	Plein air	Parc de voisinage	Green space	Récréatif	Parc-nature	Parc de quartier	Îlot de verdure
4	LaSalle	1	Aire de jeu - enfant	Récréatif	Sportif	Parc	Parc urbain	Parc de voisinage	Parc de quartier	Loisirs et détente	École primaire	Green space
5	Lachine	2	Sportif	Plein air	Récréatif	Parc	Aire de jeu - enfant	Mtl Free Wifi	Parc de quartier	Parc de voisinage	Loisirs et détente	Place publique
6	Le Plateau-Mont-Royal	0	Aire de jeu - enfant	Parc	Récréatif	Parc de voisinage	Théâtre / salle de spectacle / petit lieu de d...	Loisirs et détente	Mtl Free Wifi	Îlot de verdure	Organisme communautaire et centre de loisirs	Sportif
7	Le Sud-Ouest	0	Parc	Aire de jeu - enfant	Mtl Free Wifi	Organisme communautaire et centre de loisirs	Récréatif	Îlot de verdure	Loisirs et détente	Parc de quartier	Sportif	Parc de voisinage
8	Mercier-Hochelaga-Maisonneuve	5	Aire de jeu - enfant	Récréatif	Sportif	Parc de quartier	Parc	Loisirs et détente	Organisme communautaire et centre de loisirs	Plein air	Mtl Free Wifi	École primaire
9	Montréal-Nord	2	Aire de jeu - enfant	Récréatif	Green space	Parc	Sportif	Parc de quartier	Loisirs et détente	Mtl Free Wifi	École primaire	Plein air
10	Outremont	2	Parc	Aire de jeu - enfant	Sportif	Bâtiment historique / patrimonial	Récréatif	Café	Parc de quartier	Parc de voisinage	Loisirs et détente	École secondaire et professionnelle
11	Pierrefonds-Roxboro	4	Aire de jeu - enfant	Green space	Parc	Sportif	Parc de quartier	Parc de voisinage	Plein air	Récréatif	Parc-nature	Loisirs et détente
12	Rivière-des-Prairies-Pointe-aux-Trembles	5	Aire de jeu - enfant	Sportif	Parc	Parc de quartier	Récréatif	Îlot de voirie	Loisirs et détente	Mtl Free Wifi	Parc de voisinage	Îlot de verdure
13	Rosemont-La Petite-Patrie	1	Récréatif	Aire de jeu - enfant	Parc	Loisirs et détente	Sportif	Parc de quartier	École primaire	Mtl Free Wifi	Îlot de verdure	École secondaire et professionnelle
14	Saint-Laurent	1	Aire de jeu - enfant	Récréatif	Sportif	Loisirs et détente	Plein air	Parc	Green space	Parc de quartier	École primaire	Square
15	Saint-Léonard	2	Sportif	Loisirs et détente	Aire de jeu - enfant	Récréatif	Îlot de voirie	Mtl Free Wifi	École primaire	Parc de quartier	Parc de voisinage	Parc
16	Verdun	0	Aire de jeu - enfant	Parc	Récréatif	Sportif	Green space	Parc de quartier	Parc de voisinage	Loisirs et détente	Espace vert	Plein air
17	Ville-Marie	3	Mtl Free Wifi	Hôtel	Place publique	Poste d'attente de taxis	Art public	Parc	Aire de jeu - enfant	Autre catégorie d'établissement d'hébergement	Service consulaire	Bâtiment historique / patrimonial
18	Villeray-Saint-Michel-Parc-Extension	1	Aire de jeu - enfant	Récréatif	Parc	Sportif	École primaire	Loisirs et détente	Organisme communautaire et centre de loisirs	Parc de quartier	Parc de voisinage	Îlot de voirie

boroughs_19_df: shape(19, 7) : Dataframe created for comparing in just one glance the different types of Housing prices from borough to borough. Here, we can easily see big differences in prices from region to region.

	Borough	single Family 2nd Quarter 2021	single Family last 4 Quarters	Condo 2nd Quarter 2021	Condo last 4 Quarters	Plex 2nd Quarter 2021	Plex last 4 Quarters
0	Outremont	**	$2,180,000	$690,000	$671,500	**	**
1	LaSalle	**	$595,000	$380,500	$372,040	$756,500	$690,000
2	Ville-Marie	$1,530,000	$1,149,500	$449,000	$445,000	$819,000	$847,000
3	Le Plateau-Mont-Royal	$1,180,000	$1,221,000	$505,000	$495,000	$1,050,000	$1,050,000
4	Le Sud-Ouest	$805,000	$777,000	$430,000	$420,000	$735,500	$712,000
5	Rivière-des-Prairies-Pointe-aux-Trembles	$455,000	$420,000	$280,000	$242,000	$597,500	$550,000
6	Lachine	$590,000	$575,000	$410,000	$341,000	$642,500	$580,000
7	Montréal-Nord	$403,000	$396,000	$285,000	$270,000	$615,000	$555,000
8	L'Île-Bizard-Sainte-Geneviève	$595,000	$550,000	**	$246,000	**	**
9	Ahuntsic-Cartierville	$708,500	$700,000	$357,250	$330,700	$808,000	$777,500
10	Saint-Léonard	**	$613,500	$350,000	$340,000	$884,000	$840,000
11	Mercier-Hochelaga-Maisonneuve	$480,000	$511,500	$360,100	$349,000	$685,000	$635,000
12	Côte-des-Neiges-Notre-Dame-de-Grâce	$1,240,000	$1,114,000	$440,000	$440,000	$900,000	$847,500
13	Rosemont-La Petite-Patrie	**	$749,000	$460,000	$440,000	$835,500	$785,000
14	Saint-Laurent	$832,500	$795,125	$435,000	$412,000	**	$755,000
15	Villeray-Saint-Michel-Parc-Extension	$645,000	$580,000	$388,875	$385,500	$735,000	$690,000
16	Anjou	**	$530,000	$322,000	$293,500	$712,500	$683,500
17	Pierrefonds-Roxboro	$585,000	$527,000	$326,250	$300,000	**	**
18	Verdun	$1,050,000	$941,000	$500,500	$489,900	$801,250	$787,500

** *“Insufficient number of transactions to produce reliable statistics”*, according to Centris.ca (source of data).

We can separate the clusters for easier viewing. Once again, not a huge difference between boroughs, except for Ville-Marie (downtown).

Cluster[0] : shape(3, 12) :

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
6	Le Plateau-Mont-Royal	0	Aire de jeu - enfant	Parc	Récréatif	Parc de voisinage	Théâtre / salle de spectacle / petit lieu de d...	Loisirs et détente	Mtl Free Wifi	Îlot de verdure	Organisme communautaire et centre de loisirs	Sportif
7	Le Sud-Ouest	0	Parc	Aire de jeu - enfant	Mtl Free Wifi	Organisme communautaire et centre de loisirs	Récréatif	Îlot de verdure	Loisirs et détente	Parc de quartier	Sportif	Parc de voisinage
16	Verdun	0	Aire de jeu - enfant	Parc	Récréatif	Sportif	Green space	Parc de quartier	Parc de voisinage	Loisirs et détente	Espace vert	Plein air

Cluster[1] : shape(5, 12) :

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
2	Côte-des-Neiges-Notre-Dame-de-Grâce	1	Sportif	Récréatif	Aire de jeu - enfant	Organisme communautaire et centre de loisirs	Parc	Parc de quartier	École primaire	École secondaire et professionnelle	Mtl Free Wifi	Loisirs et détente
4	LaSalle	1	Aire de jeu - enfant	Récréatif	Sportif	Parc	Parc urbain	Parc de voisinage	Parc de quartier	Loisirs et détente	École primaire	Green space
13	Rosemont-La Petite-Patrie	1	Récréatif	Aire de jeu - enfant	Parc	Loisirs et détente	Sportif	Parc de quartier	École primaire	Mtl Free Wifi	Îlot de verdure	École secondaire et professionnelle
14	Saint-Laurent	1	Aire de jeu - enfant	Récréatif	Sportif	Loisirs et détente	Plein air	Parc	Green space	Parc de quartier	École primaire	Square
18	Villeray-Saint-Michel-Parc-Extension	1	Aire de jeu - enfant	Récréatif	Parc	Sportif	École primaire	Loisirs et détente	Organisme communautaire et centre de loisirs	Parc de quartier	Parc de voisinage	Îlot de voirie

Cluster[2] : shape(6, 12) :

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
1	Anjou	2	Sportif	Aire de jeu - enfant	Récréatif	Parc de quartier	Parc	Loisirs et détente	Green space	Pharmacy	École primaire	Restaurant
3	L'Île-Bizard-Sainte-Geneviève	2	Parc	Aire de jeu - enfant	Sportif	Plein air	Parc de voisinage	Green space	Récréatif	Parc-nature	Parc de quartier	Îlot de verdure
5	Lachine	2	Sportif	Plein air	Récréatif	Parc	Aire de jeu - enfant	Mtl Free Wifi	Parc de quartier	Parc de voisinage	Loisirs et détente	Place publique
9	Montréal-Nord	2	Aire de jeu - enfant	Récréatif	Green space	Parc	Sportif	Parc de quartier	Loisirs et détente	Mtl Free Wifi	École primaire	Plein air
10	Outremont	2	Parc	Aire de jeu - enfant	Sportif	Bâtiment historique / patrimonial	Récréatif	Café	Parc de quartier	Parc de voisinage	Loisirs et détente	École secondaire et professionnelle
15	Saint-Léonard	2	Sportif	Loisirs et détente	Aire de jeu - enfant	Récréatif	Îlot de voirie	Mtl Free Wifi	École primaire	Parc de quartier	Parc de voisinage	Parc

Cluster[3] : shape(1, 12) :

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
17	Ville-Marie	3	Mtl Free Wifi	Hôtel	Place publique	Poste d'attente de taxis	Art public	Parc	Aire de jeu - enfant	Autre catégorie d'établissement d'hébergement	Service consulaire	Bâtiment historique / patrimonial

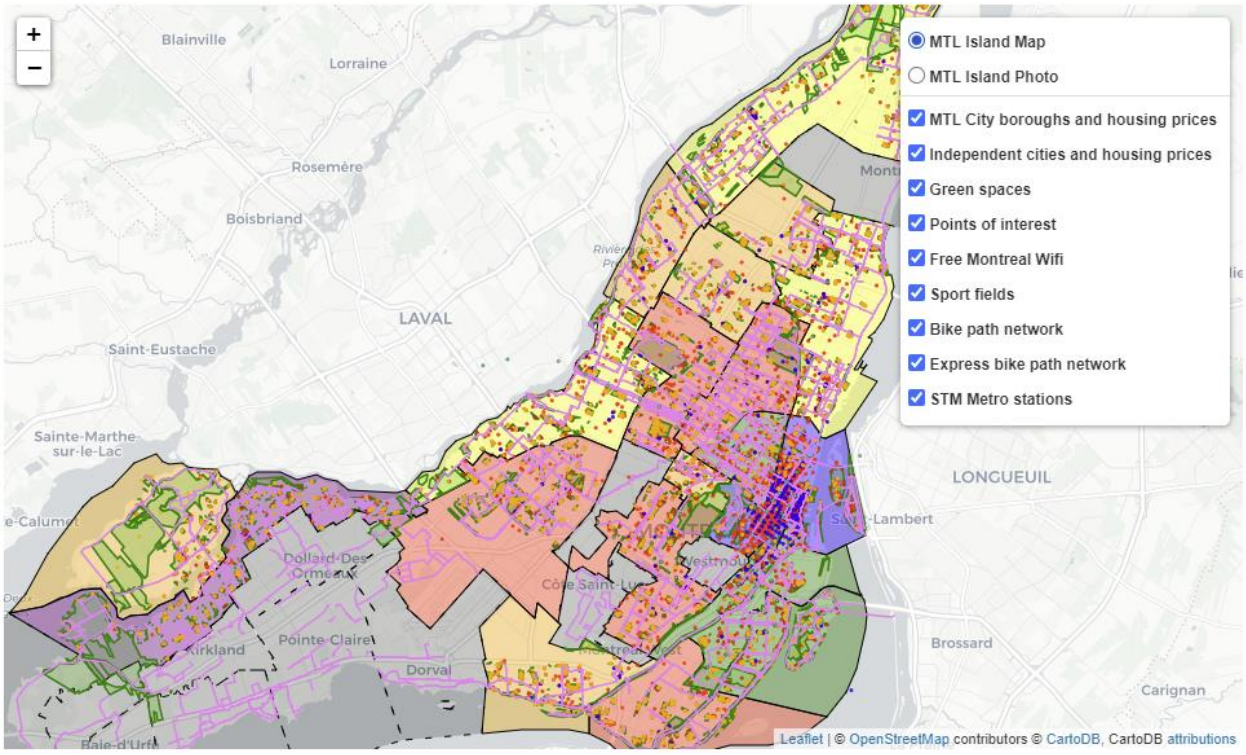
Cluster[4] : shape(1, 12) :

	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
11	Pierrefonds-Roxboro	4	Aire de jeu - enfant	Green space	Parc	Sportif	Parc de quartier	Parc de voisinage	Plein air	Récréatif	Parc-nature	Loisirs et détente

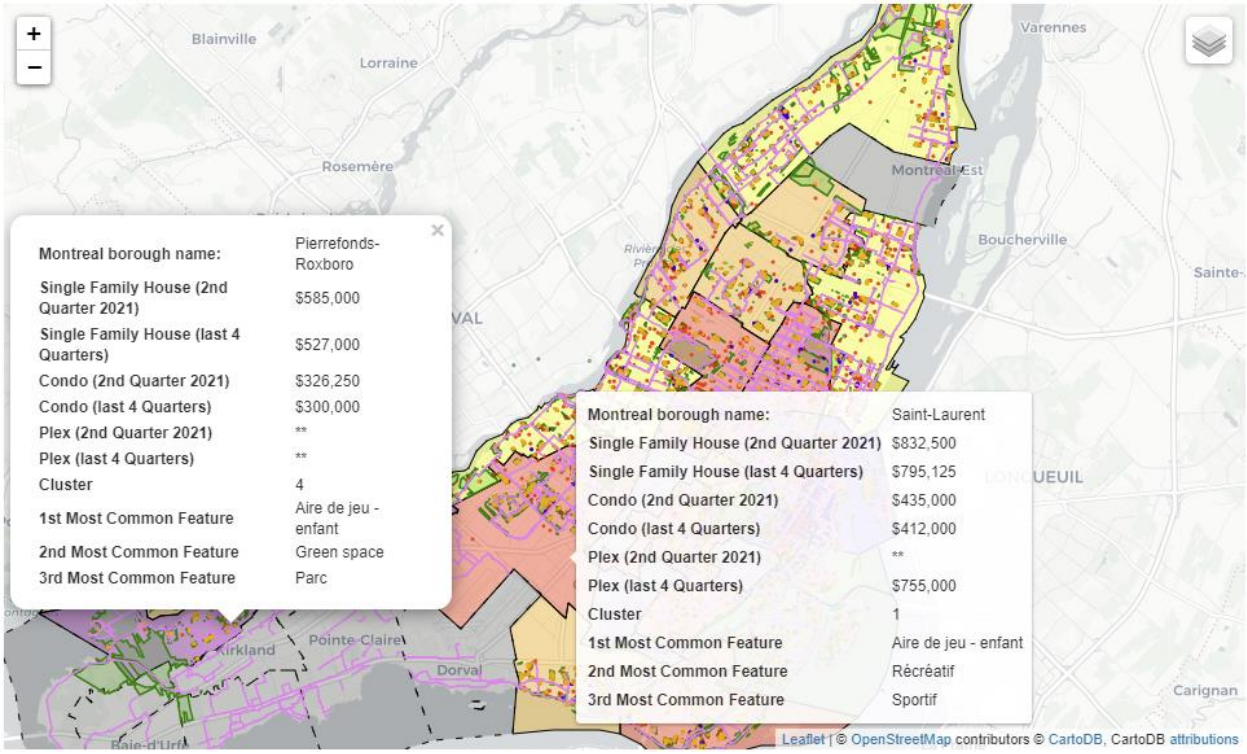
Cluster[5] : shape(3, 12) :

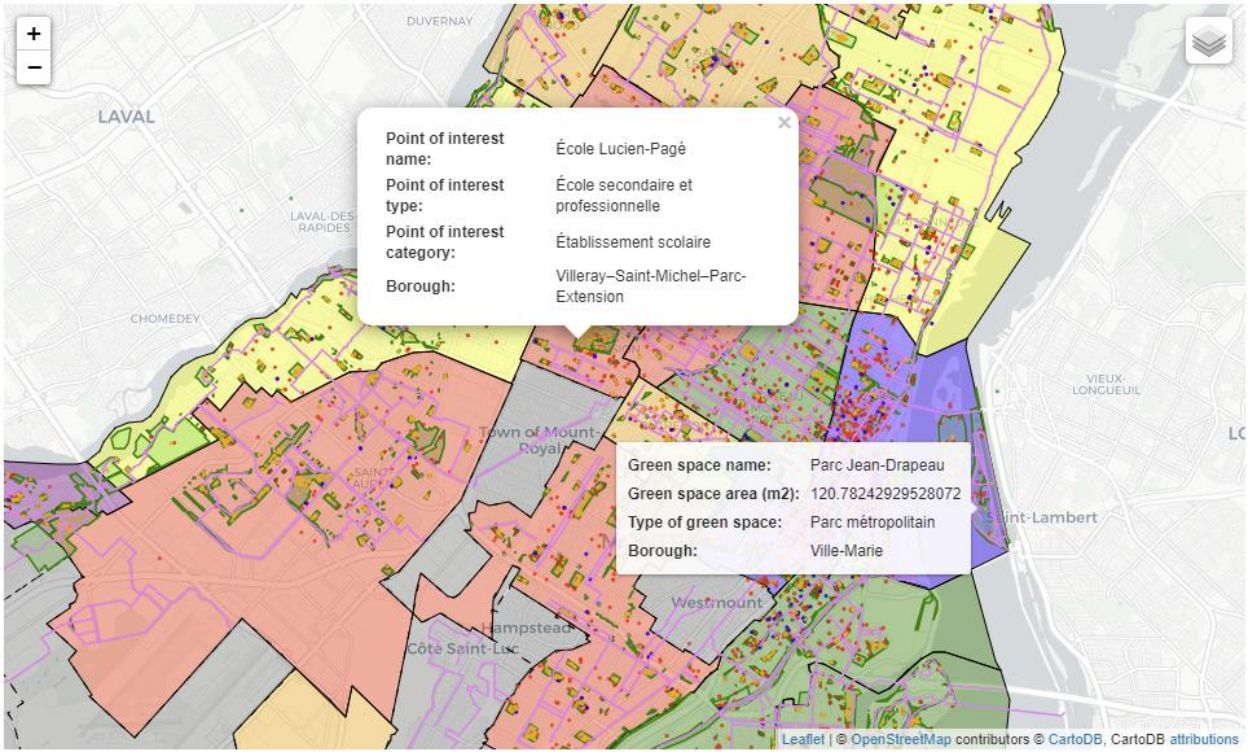
	Borough	Cluster	1st Most Common Feature	2nd Most Common Feature	3rd Most Common Feature	4th Most Common Feature	5th Most Common Feature	6th Most Common Feature	7th Most Common Feature	8th Most Common Feature	9th Most Common Feature	10th Most Common Feature
0	Ahuntsic-Cartierville	5	Sportif	Récréatif	Aire de jeu - enfant	Parc de quartier	Parc	Loisirs et détente	Îlot de voirie	École primaire	Mtl Free Wifi	Organisme communautaire et centre de loisirs
8	Mercier-Hochelaga-Maisonneuve	5	Aire de jeu - enfant	Récréatif	Sportif	Parc de quartier	Parc	Loisirs et détente	Organisme communautaire et centre de loisirs	Plein air	Mtl Free Wifi	École primaire
12	Rivière-des-Prairies-Pointe-aux-Trembles	5	Aire de jeu - enfant	Sportif	Parc	Parc de quartier	Récréatif	Îlot de voirie	Loisirs et détente	Mtl Free Wifi	Parc de voisinage	Îlot de verdure

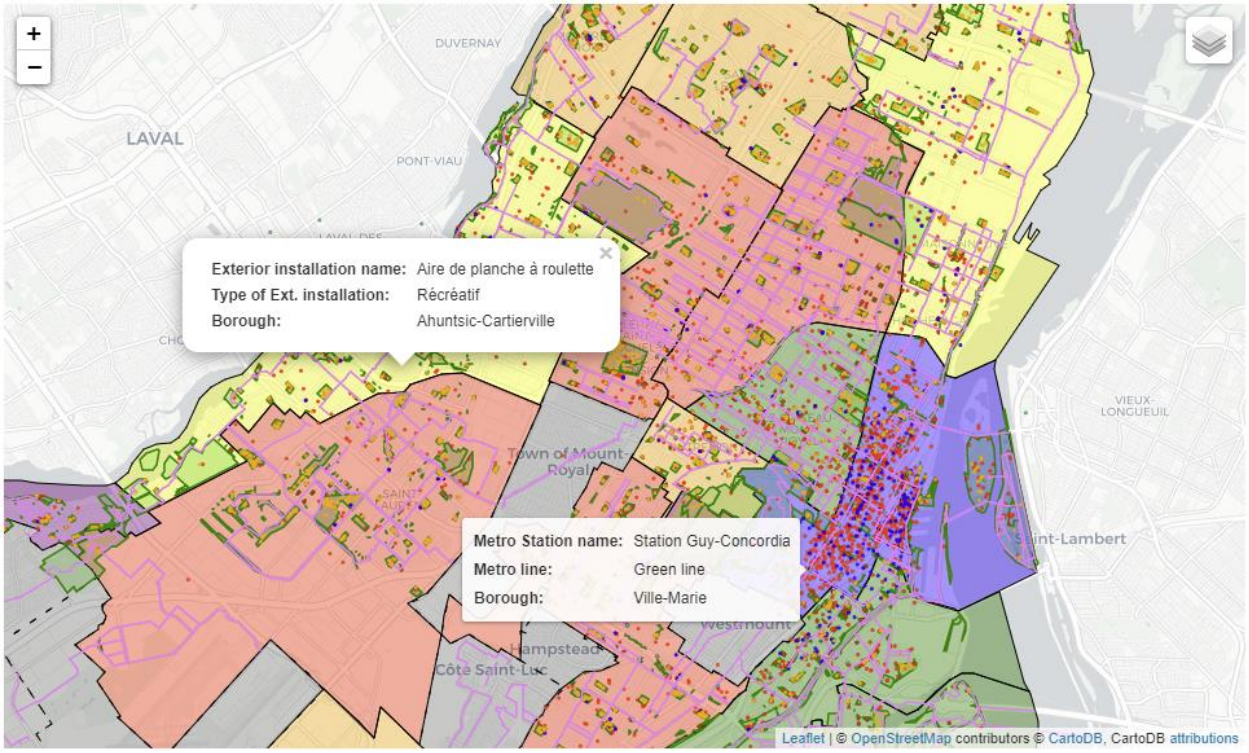
Now we're going to visualize everything together in a Folium Map, where we can turn on and off the main feature categories being displayed. A Folium Map is very convenient for visualizing spatial data such as the density of the features in your target borough or the proximity to certain feature you would like close to home.



We can click on or just hover over any feature on the map to get a pop-up with the associated data. We can use this for comparing data between 2 different boroughs, like for example, their Housing Prices and their 3 Most Common Features.







5. Conclusions and recommendations

Now we can see that based on the most common features of almost all the boroughs, they seem fairly similar and homogeneous, all good to adopt as your future home location. The more subtle differences between them would come on the less common features, which are less numerous and therefore don't make part of the 10 Most Common Features and influence a lot less the clustering algorithm.

The final choice of one borough over the other will probably be influenced a lot more by the House pricing that by the features existing in the borough, and maybe a bit because of the proximity to certain desired less-common features as proximity to work or venues oriented to certain ethnicities.

The one big exception is the borough "Ville-Marie", which is pretty much the Downtown, not the most residential area and completely dominated by features other than those of the rest of the boroughs.

For expanding the scope of this Data Analysis, is would be recommended to also consider the pricing for Rental Housing, rush-hour traffic, and getting a secondary clustering based only on the venue categories for finding out the "Vibe" of the boroughs.