

# Peer-graded Assignment: Capstone

## Project - The Battle of Neighborhoods

Is a real estate sell price good, fair, or above average?

---

### Introduction

When you want to buy a car, you go to [AutoTrader](#) or [CarGuru](#). One sweet thing about using such websites is that they tell you the car you are interested has a below-market price or above market price based on their own algorithms.

However, when you looking for either renting or buying a real estate from websites such as [Centris](#) or [DuProprio](#), you won't be able to find neither if similar real estate exists somewhere else in the city nor if it's price is below market (a.k.a., a good deal). There is not a standard to identify "similar" real estate, and price estimator is not available.

Is such thing should be done? I certainly think so. One of my friends recently wants to buy a condominium in Montreal. He told me it is exhausting to find an ideal condo to buy. He literally has to view all the available condos and compare them by heart to determine which one is the best.

My friend's experience has triggered me to start this project: find similar for sale real estate by using classification and city segmentation; analyze price within certain cluster to see if the sale price is above or below the market. Once it's done, it can help to fast locate other similar real estates once you have on in mind. It also sets a reference on price to help you determine which one is worth buying.

# Data

How am I going to do it? The first and foremost step is to find the for sale real estates. I intend to use web crawler to get data from [DuProprio](#). The real estate data shall have addresses, sale prices, and areas. Then I will use Google Geolocation APIs to get latitude and longitude for each real estate. At last, I will use the latitude and longitude to extract the neighbourhood information from [Foursquare](#), which will return a list of venues within a predefined range from that latitude and longitude.

# Methodology

## Web Crawler

Web Crawlers are used to extract data from DuProprio. “Beautiful Soup” is applied. HTML tags have to be identified for the information that need be extracted when we use “Beautiful Soup”. Below figure presents the “address” from the web page. It is clear to see that we are looking for a ‘div’ tag, with a ‘class’ attribution that has the value of “listing-location\_group-address”.

```
▼ <div class="listing-location_group-address" property="address" typeof="PostalAddress">
    <meta property="addressCountry" content="Canada">
    <meta property="addressRegion" content="QC">
    <meta property="streetAddress" content="Le MaryRobert – unité/unit 1556">
    <meta property="addressLocality" content="Griffintown">
    <meta property="postalCode" content="H3C 0M1">
▶ <div class="listing-location_address">...</div>
```

Similar processes are done to other information. Eventually, I was able to get more than 1000 for sale real estates, including Condominium, Duplex, Bungalow, Semi-detached, and other types. A few samples are presented below.

	address	areas	backyard	bathrooms	bedrooms	category	floor_if_condo	municipal	postalcode	price	year
0	5985 Boyer, Rosemont / La Petite Patrie, QC	2,000 ft <sup>2</sup> (185.81 m <sup>2</sup> )	East	2.5	3.0	Condominium	1.0	394000.0	H2S 2H8	599000	1928
1	1008-3581 boulevard Gouin Est, Montréal-Nord, QC	1,100 ft <sup>2</sup> (102.19 m <sup>2</sup> )	None	1.5	2.0	Condominium	10.0	316600.0	H1H 0A1	349000	2006
2	16107 rue Forsyth, Pointe-Aux-Trembles / Montr...	911 ft <sup>2</sup> (84.63 m <sup>2</sup> )	North-West	1.0	2.0	Condominium	3.0	207200.0	H1A 5R8	200000	1999
3	711-680 rue de Courcelles, Le Sud-Ouest, QC	950 ft <sup>2</sup> (88.26 m <sup>2</sup> )	South	1.0	2.0	Condominium	7.0	339500.0	H4C 0B8	415000	2011
4	1-5230 rue Resther, Le Plateau-Mont-Royal, QC	106.4 m <sup>2</sup> (145.28 ft <sup>2</sup> )	None	2.5	2.0	Condominium	1.0	350700.0	H2J 2W3	415000	2005

## Google GeoLocation API

The package “googlemaps” is used to call the geo-location APIs. A “Google API key” is required to use Google services. Use “address” and “postal code” as inputs, and the output will be latitude and longitude for that address.

## Foursquare API

The Venue explore API is applied here since we want to know what type of venues and how many are existed nearby a certain real estate. This API requires latitude and longitude as well as a “radius” parameter to specify the range. A radius of 500 meters is used. We define such range is walk-distance.

## Data cleaning

There are 517 condominiums for sale. We limit the real estate type to condominiums for further analysis; “Areas” has to be processed as numeric, “backyard”, “category” will be dropped; “floor\_if\_condo” and “municipal” are also dropped since half of them are None; get “borough” name based on the first three postal code; combine the data with the one from google api. Later, merge venue information into the data for city segmentation.

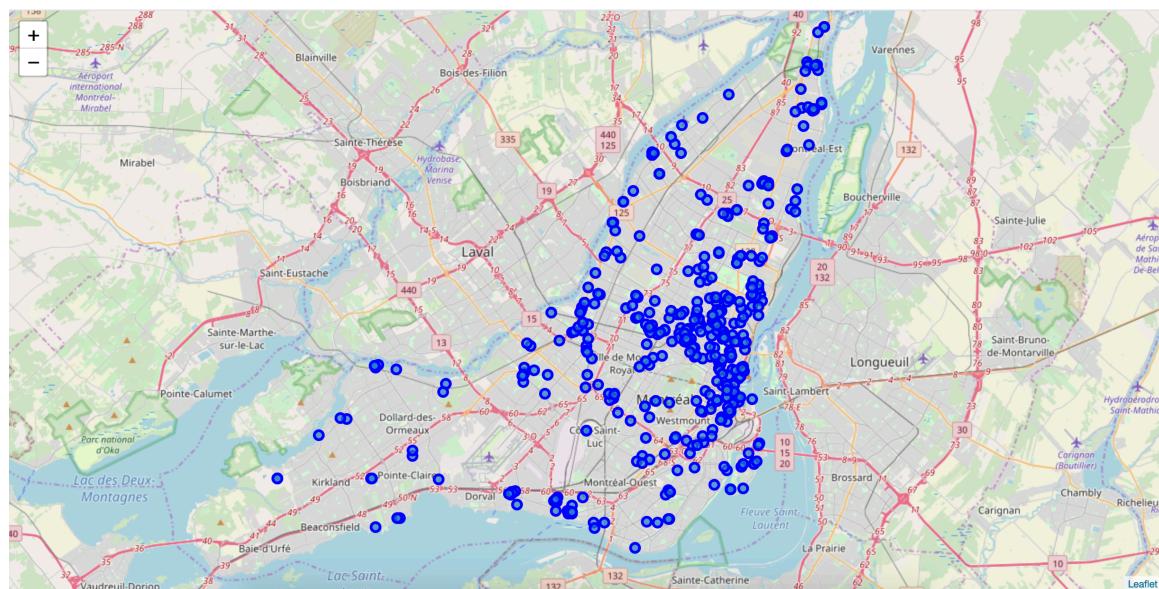
	address	areas	bathrooms	bedrooms	postalcode	price	year	borough	lat	lng
0	5985 Boyer, Rosemont / La Petite Patrie, QC	185.81	2.5	3.0	H2S 2H8	599000.0	1928	Petite-Patrie	45.535446	-73.598175
1	1008-3581 boulevard Gouin Est, Montréal-Nord, QC	102.19	1.5	2.0	H1H 0A1	349000.0	2006	Montréal-Nord	45.591632	-73.650183
2	16107 rue Forsyth, Pointe-Aux-Trembles / Montr...	84.63	1.0	2.0	H1A 5R8	200000.0	1999	Pointe-aux-Trembles	45.675129	-73.501151
3	711-680 rue de Courcelles, Le Sud-Ouest, QC	88.26	1.0	2.0	H4C 0B8	415000.0	2011	Saint-Henri	45.475026	-73.593663
4	1-5230 rue Resther, Le Plateau-Mont-Royal, QC	106.40	2.5	2.0	H2J 2W3	415000.0	2005	Plateau Mont-Royal	45.529412	-73.590276

## Classification

K-mean method is used for clustering and city segmentation. The venue information are used to divide all condominiums to several clusters.

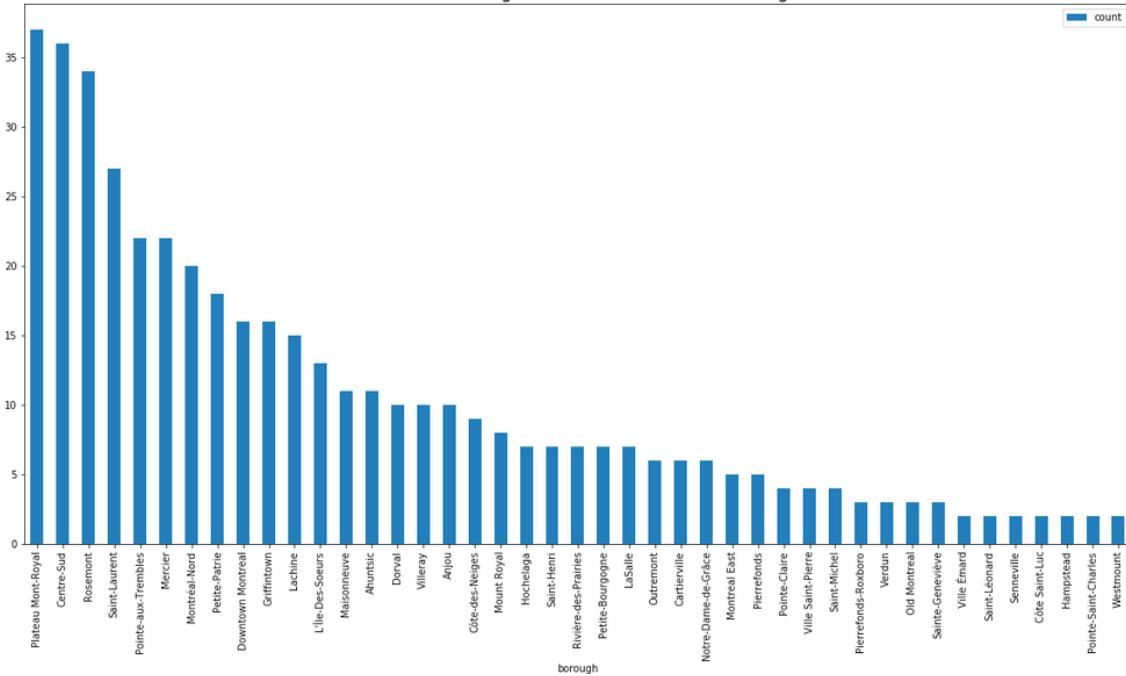
## Results and Discussion

Let's first see the locations of those condominiums on the map. It appears the most condos for sale exist in the middle part.

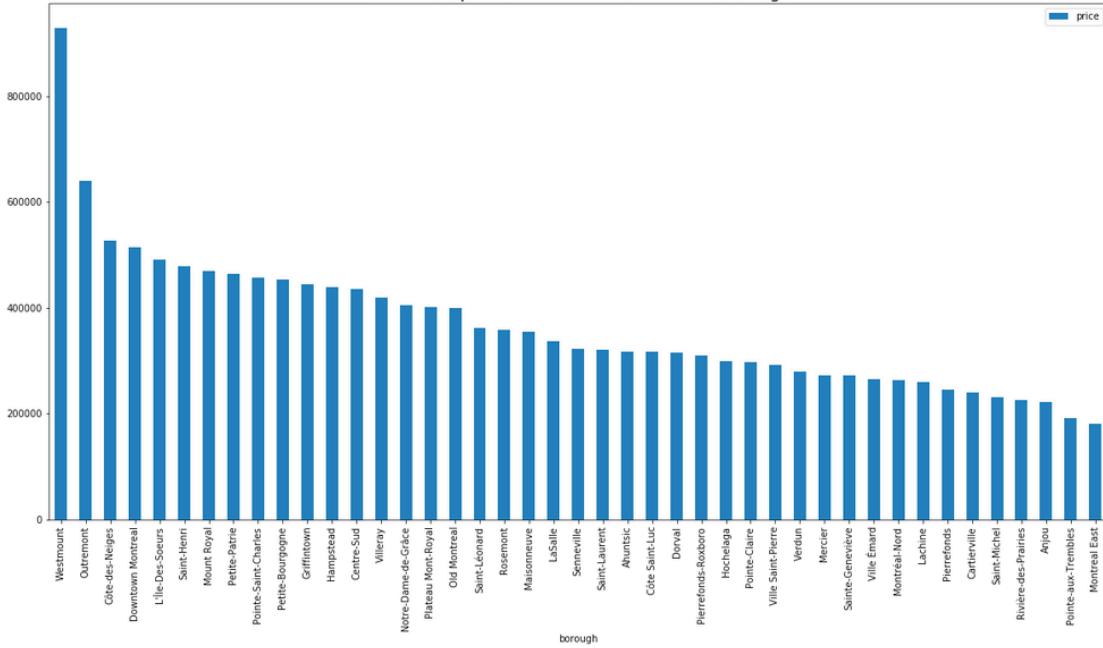


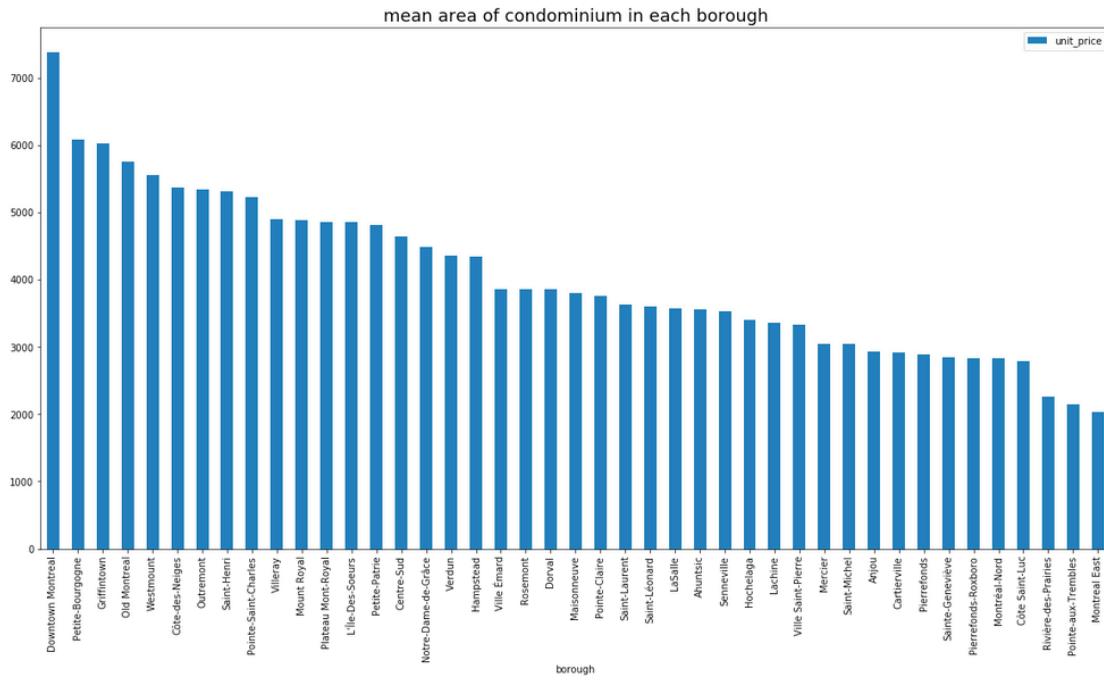
Let's further see the exact number of condominiums for each borough, and mean sale price/price per unit area.

number of selling condominiums in each borough

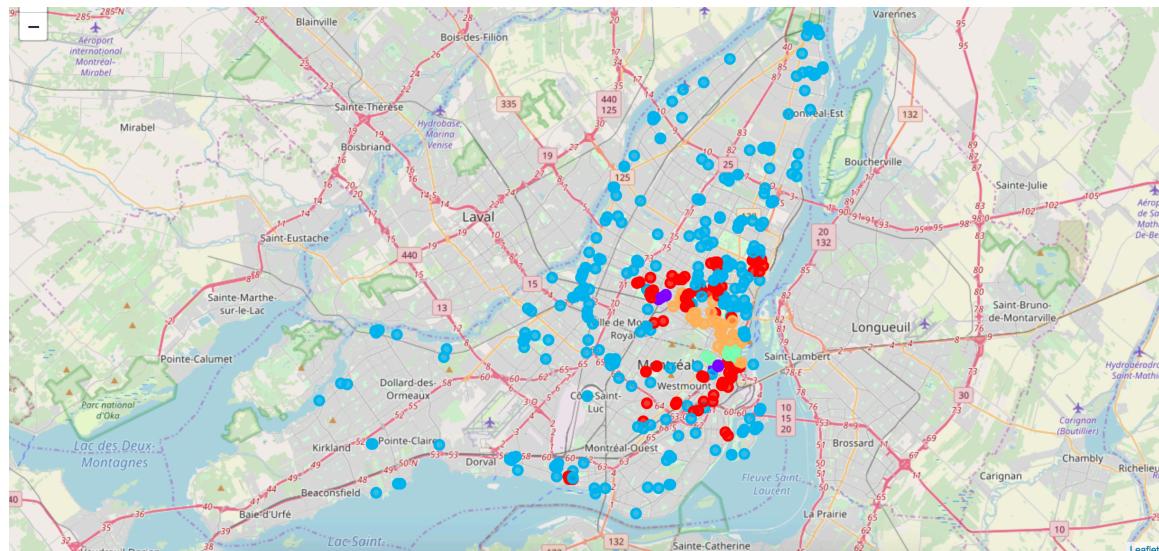


mean sale price of condominium in each borough





A few things can be seen from those figures. Montreal has 123 neighbourhoods in total and 43 of them have condos for sale; **Plateau Mont-Royal**, **Centre-Sud**, and **Rosemont** are top three neighbourhoods in the number of condos for sale; though **Westmount** has only two selling condos, their average sale price is much higher than the price of condos in other neighbourhoods; **Downtown Montreal** has the most expensive condo for sale in terms of average price per square meter. Based on mean price per unit area, it might be proper to divide all the condos into 4~5 clusters.

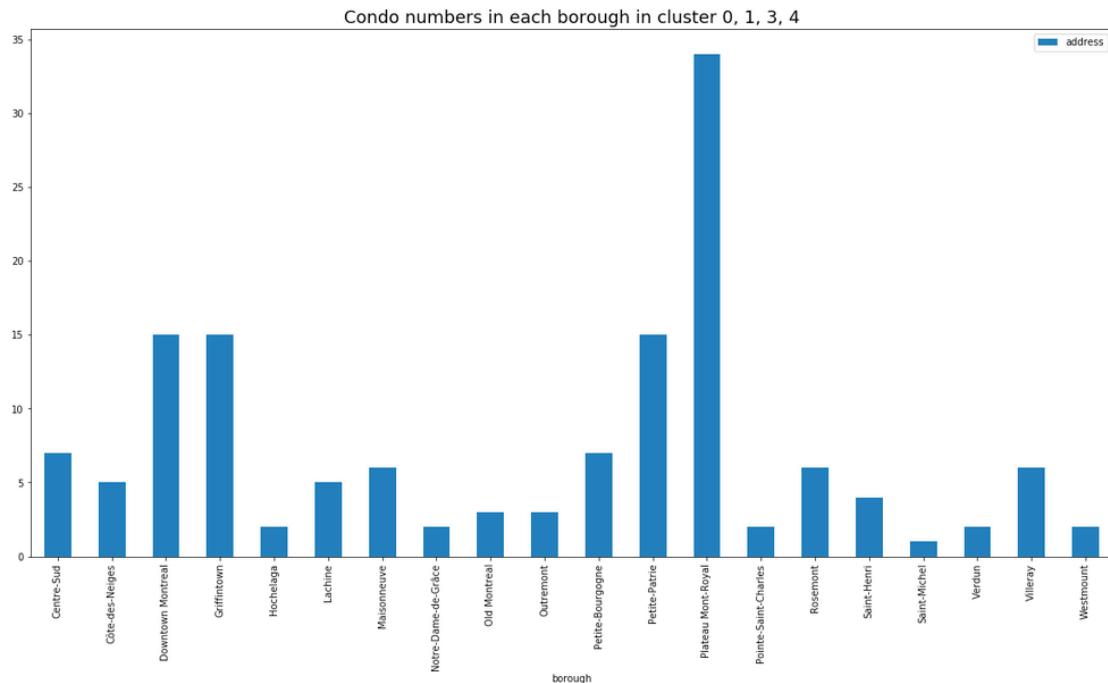


Let's divide the condos into 5 clusters. They are presented in different colors on the map. Light blue dots represent Cluster\_2, red dots represent Cluster\_0, orange represents Cluster\_4, purple represents Cluster\_1, and mint green represents Cluster\_3. We can see from the above map that Cluster\_2 has the most condos, followed by Cluster\_0 and Cluster\_4, Cluster\_1 and Cluster\_3 are about the same.

Surprisingly, more than half of for-sale condos fall into the same cluster, which may indicate Montreal has similar neighborhood all cross the island. All these condos are marked with light-blue markers on the map, and these condos locate evenly across the Montreal Island. The number of condos in each cluster is listed below.

Cluster	0	1	2	3	4
Condos	90	10	295	9	33

Cluster\_2 exists everywhere, but what about the rest four clusters? From the following figure we find out that the rest 4 clusters of condos mainly locate at **Downtown Montreal, Griffin Town, La Plateau, and Petite-Patrie (Little Italy)**.



Let's further see the mean price and area in each cluster:

Cluster	0	1	2	3	4
Price	427216	558900	329389	489256	420664
Areas	86.36	78.18	90.75	74.80	83.91

Cluster\_1 has the most expensive condos, followed by Cluster\_3; Cluster\_0 and Cluster\_4 are about the same in price; Cluster\_2 has the lowest mean price. Meanwhile, Cluster\_1 and Cluster\_3 have the smallest mean areas, and Cluster\_2 has the biggest areas, Cluster\_0 and Cluster\_4 are still about the same in area.

Let's analyze the borough and top venues each cluster has:

- Cluster 0 has reasonable price, it exists from Little Italy down to Downtown Montreal, easterned through Grinffintown to Saint-Henri and Lachine, the top nearby venues are: *bakery, cafe, all kinds restaurant, pharmacy*;
- Cluster 1 has the highest mean price, it exists in Little Italy and Downtown Montreal, the top nearby venues are: *Italian restaurant, hotel, cafe shop, and bakery store*;
- Cluster 2 has the lowest mean price, it is widely spread across the Montreal Island, and tops nearby venues are: *park, pharmacy, gym, and restaurant*.
- Cluster 3 has the second highest mean price, it exists in Griffintown, Downtown Montreal, Little Burgundy, Saint-Henry, and Old Montreal, the top nearby venues are: *hotel, cafe, French restaurant*;
- Cluster 4 is similar to cluster 0 in price, it exists in La Plateau, Little Italy, and some Downtown area, the top nearby venues are: *cafe, bar, restaurant*.

It seems that Cluster\_1 and Cluster\_3 are more like travelling zones and downtown area; Cluster\_2 is living zone; Cluster\_0 and Cluster\_4 are somehow in transition.

My friend wants to buy a condo at Rue Peel and Wellington Street, Griffintown. It is within Cluster\_0. The mean price in this cluster is \$ 427,216 and mean price per unit area is \$ 5,069 /m<sup>2</sup>. However, the price in Griffintown is \$ 475,275 and \$ 6,459 /m<sup>2</sup>. We reveal that condominiums in Griffintown have the second highest price per unit

area. It may not be an ideal condo to buy for living in, but a good investment. Within Cluster\_0 there are alternative borough, such as Little Burgundy, Cote-des-Neiges, Little Italy, and La Plateau, which have more reasonable condo prices. I would suggest my friend to consider these neighbourhoods as well.

## Conclusion

This project reveals the following things about Montreal real estate market:

- Condo is the most popular type of real estate on the market, more than half of for-sale properties are condos;
- There are 123 boroughs in Montreal, 43 among them have condos for sale;
- La Plateau has the most condos for sale; Westmount has the highest mean price; Downtown Montreal has the highest mean price per area;
- Most condos for sale on the Montreal Island has similar walking-distance neighbourhood. Within the range there are parks, pharmacies, gyms, different restaurants.
- More expensive condos exist in Downtown and surrounding areas. Within its neighbourhood there are more likely to have Italian or French restaurants, cafe shops, bakeries, and hotels.

Based on the findings, I would recommend my friend to re-consider buying a condo in Griffintown because its price is higher than similar neighbourhoods. I can also provide some alternative neighbourhoods to him, such as **Little Burgundy, Cote-des-Neiges, Little Italy, or La Plateau**. However, if my friend wants to investigate into the condo, Griffintown may be a wise choice.