Battle of Neighborhoods

Live in Montreal - by Adil Ezzaam

Selecting the right place to live in Montreal

- Live near venues that fit our preferences. The variety of venues tend to likely be where the population is **more dense**.

- Find a peaceful place to live, with the least criminality. **Least** populated neighborhoods have **less** criminality.

- Be able to afford where you want to live. Westmount and surroundings are **more likely** to be **expensive** while being in the middle of everything(rich amount of restaurants and other venues). Spots in the west island are also **pricey** but **less** populated and have **less** accessibility to services.

Which data

- The set of data required for this analysis covers boroughs details pulled from **Wikipedia** and the **jmacman12.carto.com** website(geojson).
- Venues per neighborhoods are gathered from **Foursquare API**. We will request the data within a predefined radius for each neighborhood.
- Criminality data from **donnees.montreal.ca/ville-de-montreal/actes-criminels** which we will append neighborhoods.
- Census data of Montreal which includes Real Estate price, Population and density and income data. This dataset be used cautiously since it's outdated(2016).

Which methods

- We will use **Python** on Jupyter notebook, on which we will perform all analysis. There's many libraries involved but we will focus on the essential ones.
- **Pandas** library will be used to do all datasets manipulation and cleaning.
- Using **BeautifulSoup** library helps on pulling data from HTML pages.
- **Folium** library will permit us to place our data on a map covering the Montreal island.
- We will perform **Clustering** as a Machine Learning method. **KMeans** is used for clustering through **Sklearn** library.

Venues density
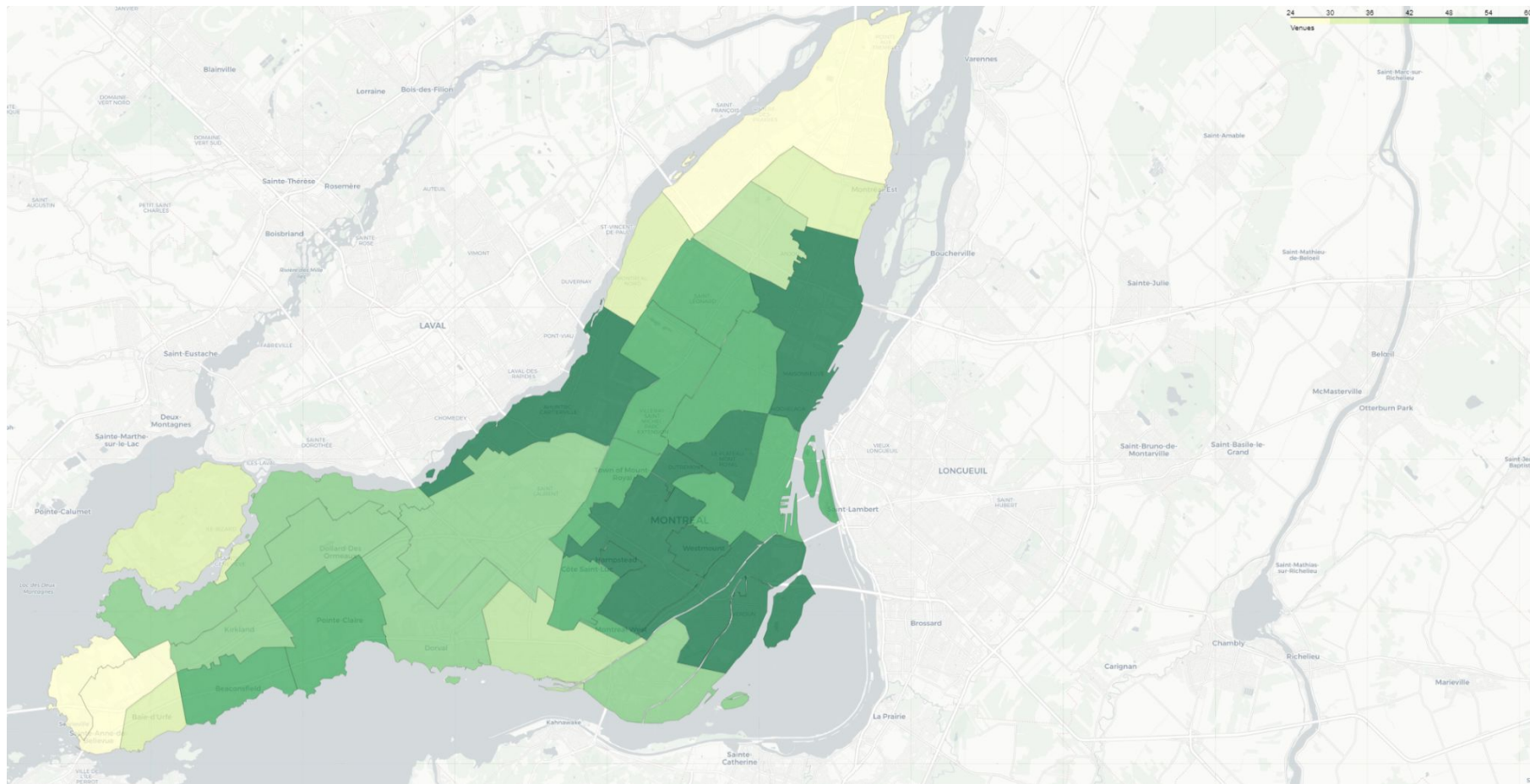
Venues density per Neighborhoods

Neighborhoods with most venues

Neighborhood	Total
westmount	62
verdun	58
hampstead	57
ville-marie	57
mont-royal	57

Neighborhoods with least venues

Neighborhood	Total
senneville	18
sainte-anne-de-bellevue	20
baie d'urfe	20
riviere-des-prairies-pointe-aux-trembles	24
l'ile-bizard-sainte-genevieve	26

While Neighborhoods with **most** venues are part of or **near the downtown area of Montreal**.

Although for the **least**, the neighborhoods are on the **edges of the city**. This **correlates** a lot with the **population and density**.

Population

# Densite

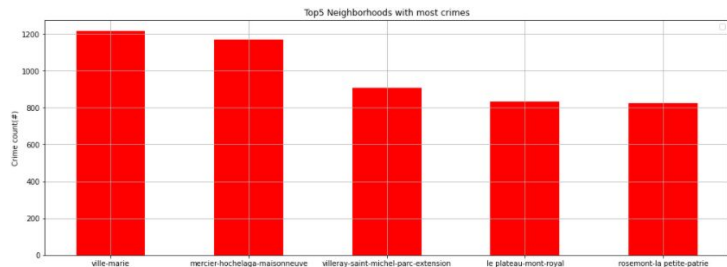# Crimes

Real estate prices

Observation

While correlations could be observed through the maps visualization, here's is a correlation matrix that relates the several features of our datasets. Important to note a **high**(**0.8**) correlation between **population** and **crimes**. The least but also interesting to note is the **density** correlates with **venues**(although at only **0.59**).
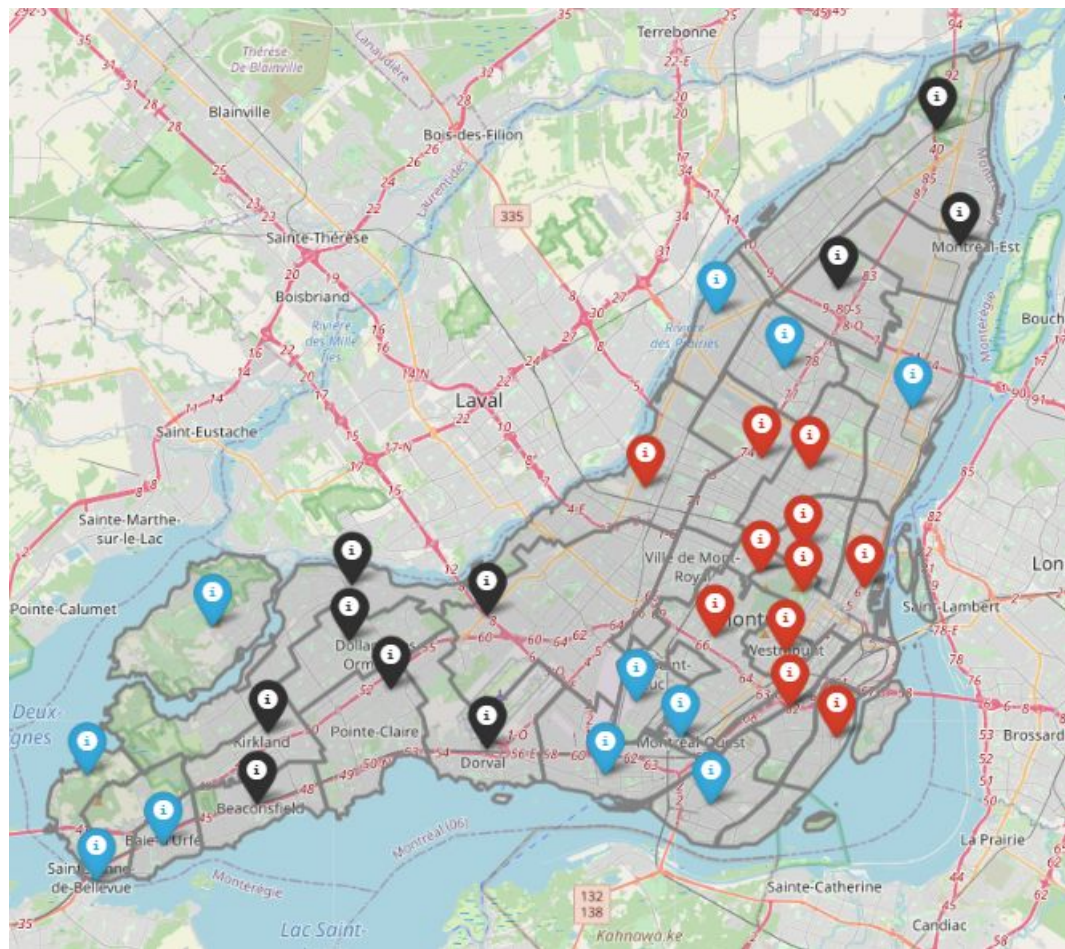
Criminality
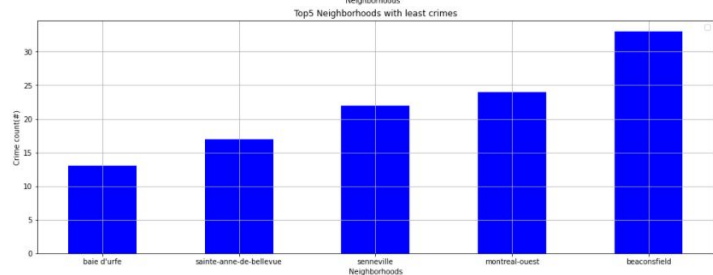
For a better understanding, the crime dataset covers the last **6 months** of **year 2020**. The data offers 5 type of criminal activities as seen below. We've took the 5 top and bottom active neighborhoods. So far, the correlation with population is **strong**. The same was observed with venues and population.
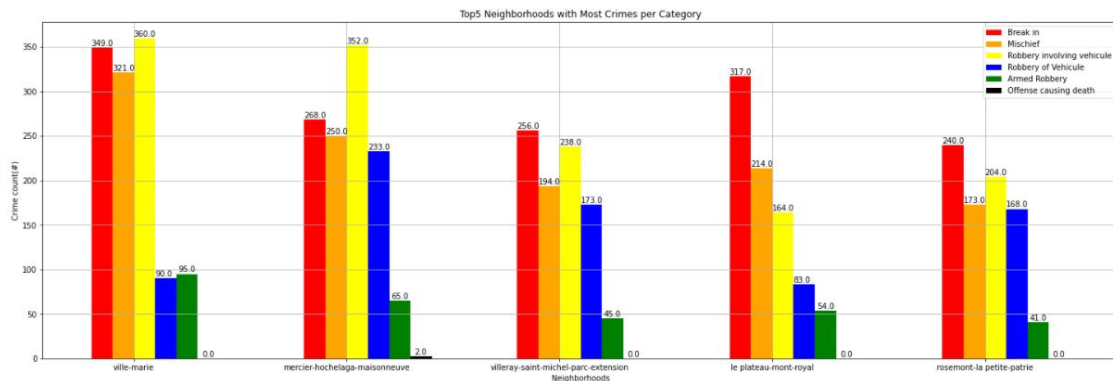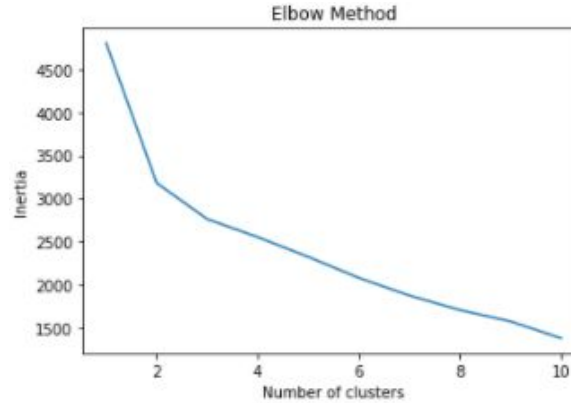
Clustering

Neighborhood	Cluster
ahuntsic-cartierville	0
anjou	2
baie d'urfe	1
beaconsfield	2
cote-des-neiges-notre-dame-de-grace	0
cote-saint-luc	1
dollard-des-ormeaux	2
dorval	2
hampstead	1
kirkland	2
l'ile-bizard-sainte-genevieve	1
lachine	1
lasalle	1
le plateau-mont-royal	0
le sud-ouest	0
mercier-hochelaga-maisonneuve	1
mont-royal	0
montreal-est	2
montreal-nord	1
montreal-ouest	1
outremont	0
pierrefonds-roxboro	2
pointe-claire	2
riviere-des-prairies-pointe-aux-trembles	2
rosemont-la petite-patrie	0
saint-laurent	2
saint-leonard	1
sainte-anne-de-bellevue	1
senneville	1
verdun	0
ville-marie	0
villeray-saint-michel-parc-extension	0
westmount	0

Cluster 0 = Red
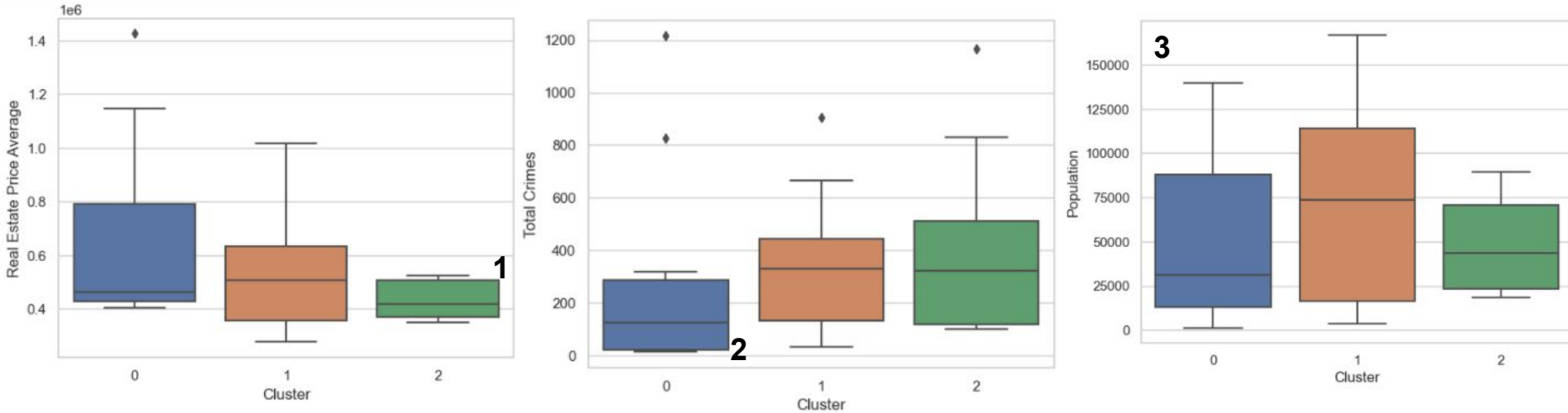Cluster 1 = Blue
Cluster 2 = Black

Clustering

As part of the clustering, the optimal # of cluster was 3. The methods was only done on venues data. We can observe similarities between the clustering process and the crime/population features.

Method for cluster optimization with Inertia value.

Clusters against features

Clusters are based on venues repartition and occurrences. We've done so to have it as a baseline of our analysis. Here, we see the boxplots our clusters, covering real estate price average, total crimes and population count. **Observation#1** points on a **low and close repartition** of real estate prices within the **cluster 2**. **Observation#2** points out on a low and fairly spreaded crimes counts across **cluster1**(a few exceptions). **Observation#3** analyze the correlation between population and crimes which are seeing less through our clustering.

Conclusion

The analysis we've done observed multiple correlations and possible trade-offs required when selecting a neighborhood or even a cluster(of neighborhoods).

We've observed a clustering excessively relative to the proximity of the neighborhoods. This is explained by the same venues listed per Neighborhoods since they are within a certain radius. This is a good thing to leverage the proximity as part of our clustering. Definitely applicable for our **Cluster#0**, the other clusters are segmented either in 2 or 3 groups.

As low cost option, anywhere within **Cluster#2** is cheap and well surrounded by various venues. While this same cluster occupies a range of criminal activities, Beaconsfield is the neighborhood with the least.

Westmount identified as part of **Cluster#0** makes the difference with an average real estate cost near 1.5 million. The cluster has a median barely reaching the semi million. The criminality within the cluster is also low for having a median below 100 criminal activities within 6 months (note with a few exceptional neighborhoods).

What about **Cluster#1** ? Well... It is a cluster which obviously regroups neighborhoods with varied venues(Café/Coffee Shops, restaurants and other convenience stores). However, the neighborhoods from the west island are not sharing the same criminal records compared to the others. As an example, Baie d'Urfe, Senneville, Sainte-Anne de Bellevue are the neighborhoods with the least criminal activities.

As a result, our clustering process helped on grouping neighborhoods similar in terms of venues and we've used the boxploting on verifying the pattern with our other datasets. The decision is yours and depends on how much you can afford and how much you rate the risk of criminality within the neighborhood.