HELPING TO INTERNATIONAL STUDENT SELECTING NEIGHBOURHOOD IN TORONTO
Capstone Project - The Battle of Neighborhoods (Week 2)

Jesús Romero Palacio

February 6, 2021

CONTENTS

1. INTRODUCTION: BUSINESS PROBLEM

If it does not consider the atypical pandemic year 2020, the number of international students has been growing annually in Canada; 642000 immigrants arrived in 2019, nearly 50% study in Ontario [1].

There are a thousand reasons to international students prefer Toronto, Canada, which was ranked in 2019 as 11th of 50 best cities in the world for students. The study, requiring eligible cities to have a population over 250,000, and have at least two institutions featured in the QS World University Rankings, was based on five themes including university rankings, student mix, employer activity, quality of living, and affordability [2]

Toronto offers various types of lifestyles zones according to different preferences [3]. For this capstone project, it will focus on one type of international student, based on a friend (it will call her Linda) who wants to travel to complete her postgraduate studies. She wants to rent a one-bedroom apartment with the following neighbourhood features:

1- The rent price around CAD 1.700 (1 bedroom)
2- Secure leves
4- Near to different public transportation options (bus, train, metro...)
5- Near to food places
6- Near to a park

Although there are online home rental platforms, this information is not enough to obtain all information that she needs. For this reason, it will create a map and information chart with some neighbourhoods options that could fit her requirements.

2. DATA

The data used are the following:

2.1- The rent price around CAD 1.700 (1 bedroom): *From Zumper (home rental website)* [9] *and Canada Mortgage and Housing Corporation, the average rent.* [3]

2.2- Secure leves: *From Toronto Police Service, the Major Crime Indicator - MCI (2014 to 2019).* [2]

2.3- Neighbourhoods fit the desired rental value and securest: *From the previous data.*

2.4- The coordinates of neighbourhoods fit the desired rental value and securest: *From Toronto Police Service, the Major Crime Indicator - MCI.* [2]

2.5- Venues in neighborhoods: *Foursquare API to get the most common venues from the Toronto selected neighbourhoods.* [1]

2.6- General information about neighbourhoods: *From the previous data.*

3. METHODOLOGY

In this project it worked with the data before with the Data Science methodology:

Analytic Approach – Data Requirements – Data Collection – Data Understanding – Data Preparation – Modeling – Evaluation.

The data were used by following way:

3.1- The rent price around CAD 1.700 (1 bedroom):

- It used a histogram to create 5 ranges and classify the average rent by neighbourhoods, with that, it was obtained the neighbourhoods that fit requirements.

3.2- Secure levels:

- It used a histogram to create 5 ranges and classify the Major Crime Indicator - MCI I by neighbourhoods, with that, it was obtained the neighbourhoods that fit requirements.

3.3- Neighbourhoods fit the desired rental value and securest:

- It selected the neighbourhoods that fit the desired rental value and securest.

3.4- The coordinates of neighbourhoods adjusted to the desired rental value and securest:

- It used the neighbourhoods coordinates reported in the data from Toronto Police Service.

3.5- Venues in neighbourhoods:

- It used the selected neighbourhoods and their coordinates from before data.

- It selected the near venues for each neighbourhood (500 m).

- It classified by venue category and group by neighbourhood.

- It used "Elbow methodology" to select the best K to use for K-means algorithm to cluster the neighbourhoods.

- It used unsupervised learning K-means algorithm to cluster the neighbourhoods.

- It showed the most common venues categories from neighbourhoods clusters.

3.6- General information about neighbourhoods:

- It showed, in clusters, the general information about venues categories of interest (public transportation options, food places, and parks).

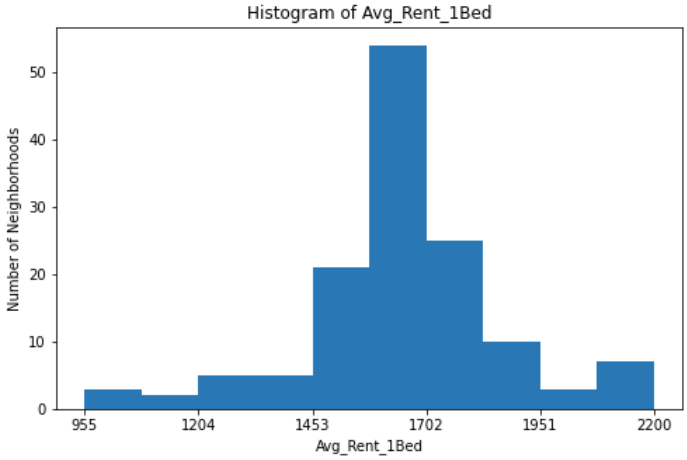
- It showed the mean of MCI and rent of one-bedroom apartments in clusters.

4. ANALYSIS

4.1- The rent price around CAD 1.700 (1 bedroom):

First, it obtained the values of one-bedroom apartments rent, in Toronto neighborhoods, the data was from Zumper (home rental website) [4] and Canada Mortgage and Housing Corporation. [5]

It was created a histogram with 5 bins, to know the behavior of average values of one-bedroom apartments in all neighbourhoods of Toronto to classify according to the rent values.



It was defined the next ranges of rent for one-bedroom apartments:

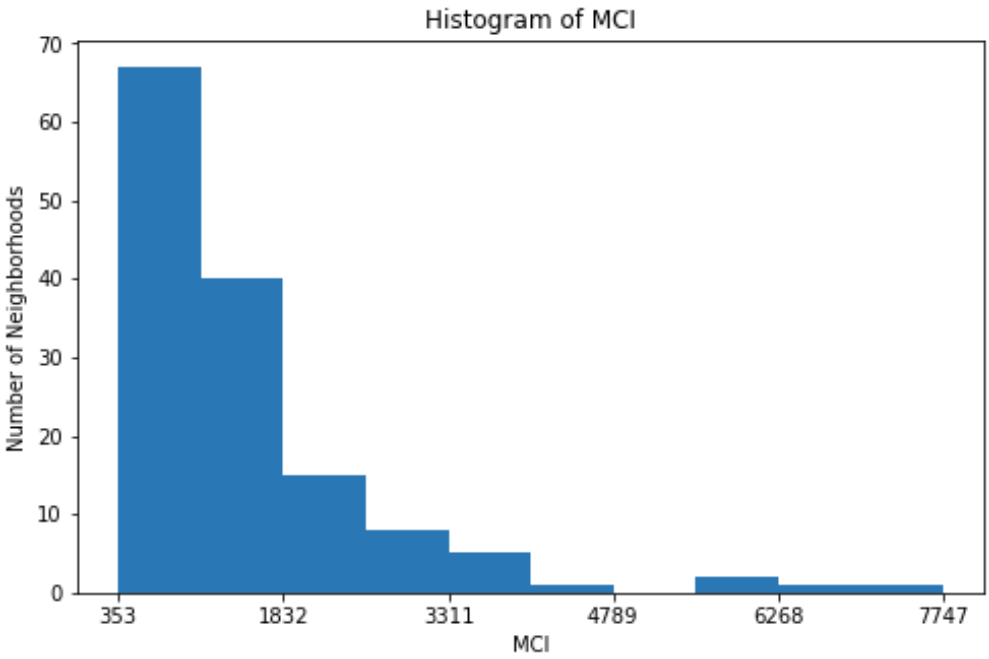
- **Low:** 955 - 1204
- **Mid Low:** 1204 - 1453
- **Mid High:** 1453 - 1702
- **High:** 1702 - 1951
- **Upper High:** 1951 – 2200

According to the desired rent for one-bedroom apartment (around CAD 1.700), it was selected the "Low", "Mid Low" and "Mid High" ranges, and it was created a dataframe (desire_rent) with the neigbourhoods and its average rent.

4.2- Secure Levels:

It was obtained the values of the Major Crime Indicator - MCI, from the Toronto Police Service (2014 to 2019). [6]

It was used a histogram to create 5 ranges and classify the MCI by neighbourhoods, with that it gets the neighbourhoods that fit requirements.



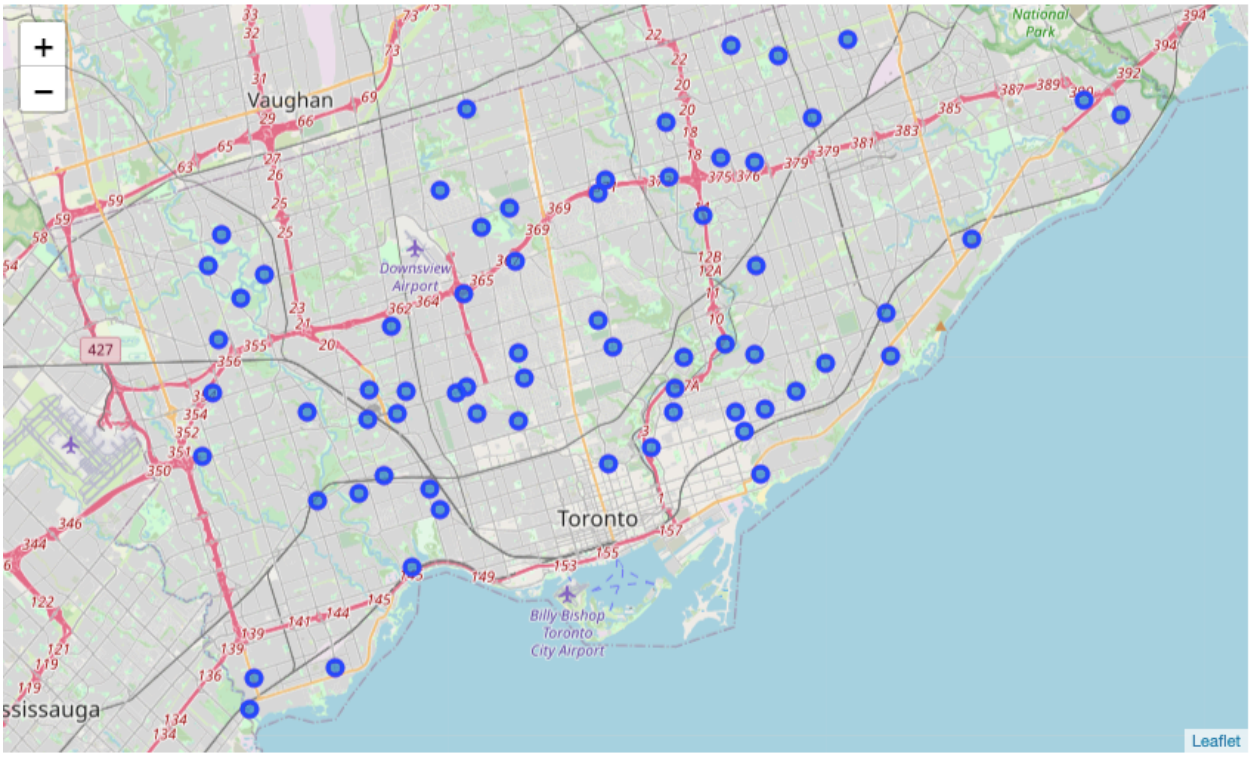
It was defined the next ranges of the crimes in the Neighbohood:

- **Low:** 353 - 1832
- **Mid Low:** 1832 - 3311
- **Mid High:** 3311 - 4790
- **High:** 4790 - 6269
- **Upper High:** 6269 - 7747

"Linda" wants a secure neighbourhood, because of that it was selected the "Low" range, and it was created a data frame (desire_secure) with the neighbourhoods and their coordinates.

4.3- Neighbourhoods fit the desired rental value and securest:

It was selected the neighbourhoods that fit the desired rental value secure levels and a map with their geolocations:



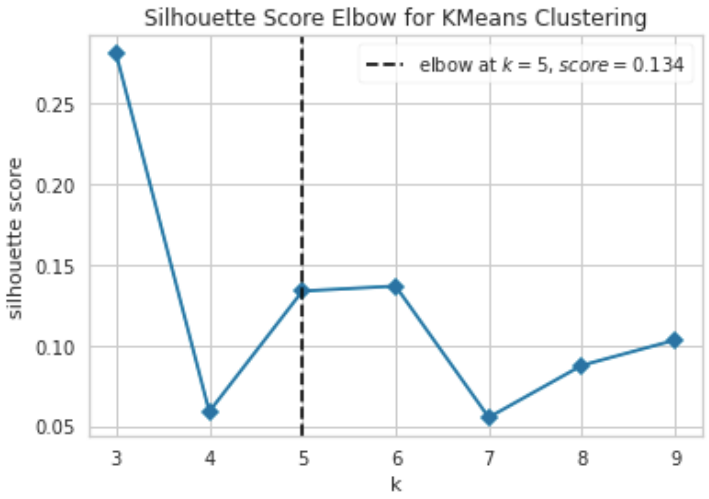
4.4- Venues in neighbourhoods:

- It was used the selected neighbourhoods before ("desire_secure_rent") and its coordinates from before data, after that it was used the Foursquare API and selected the near venues for each neighbourhood (500 m), it was 768 venues and 173 unique categories of venues.

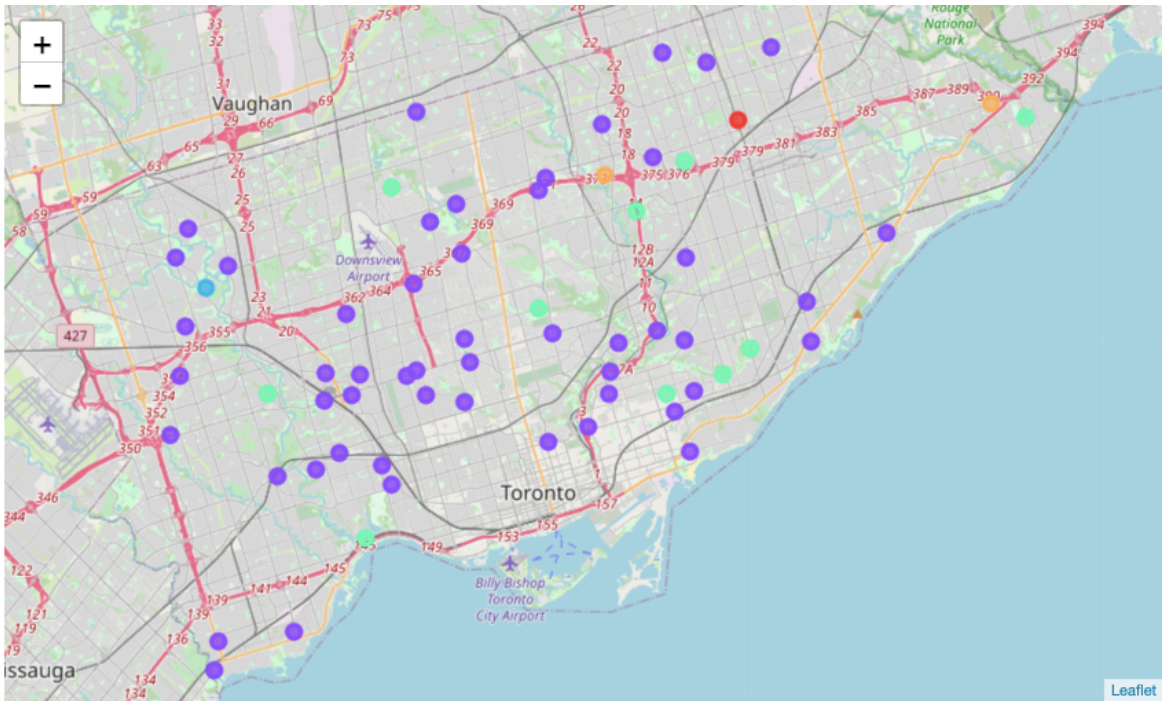
- After preparing and reorder the data, it was classified by venue category and grouped by neighbourhood and select the 10th most common venues for each neighbourhood.

It was selected K-means, because is vastly used unsupervised learning for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data, in this case, it needs to group and discover similar characteristics of neighbourhoods from its venues.

Using Elbow method, it was selected the best K to use for K-means algorithm to cluster the neighbourhoods.



The ubication of 5 clusters was located in the next map:



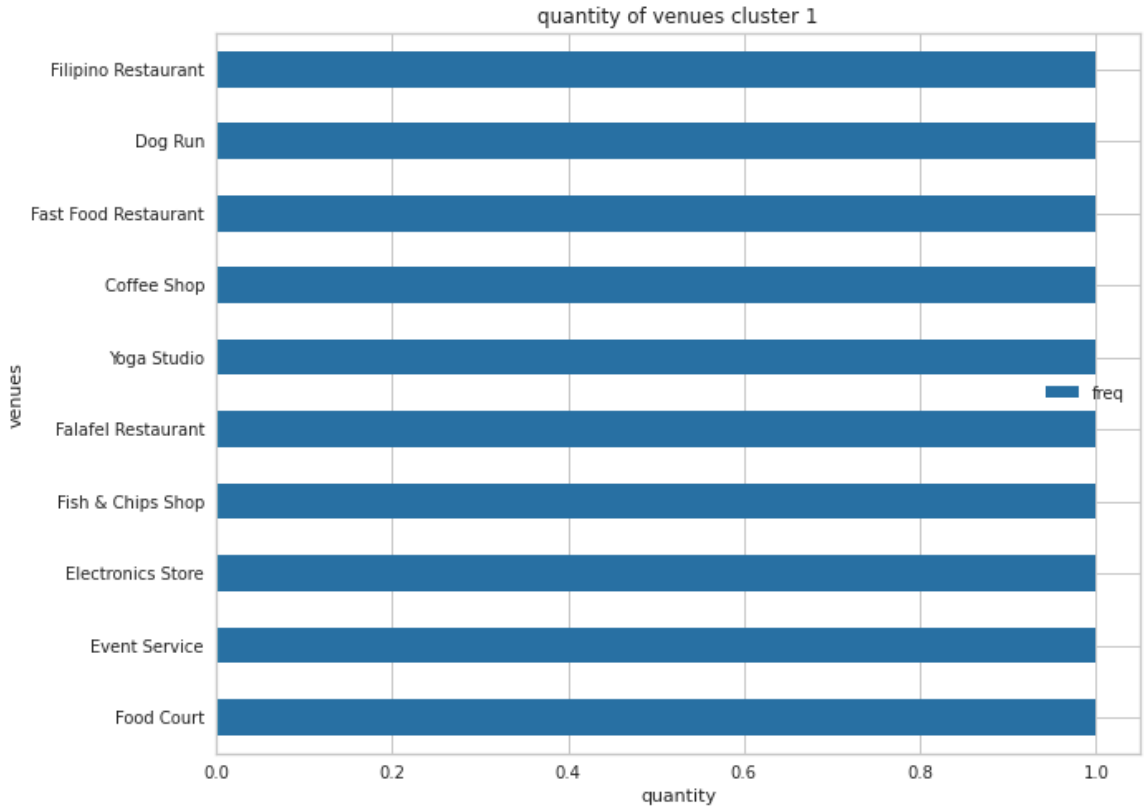
4.5- General information about neighbourhoods:

This is the general information about clusters, their venues categories of interest (public transportation options, food places, and parks), and mean of MCI and rent of one-bedroom apartments.

4.5.1- Exploring venues categories of interest:

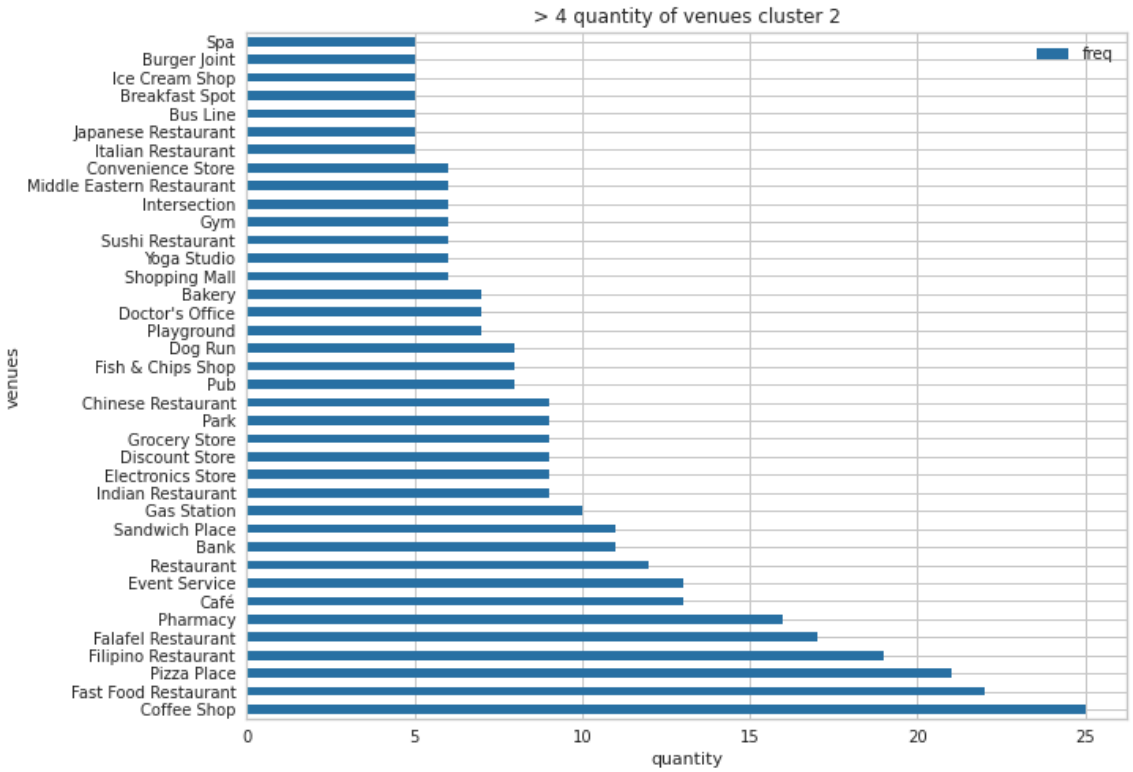
CLUSTER 1 (Red point in map)

One neighbourhood, Agincourt South-Malvern West, 10 different categories of venues:



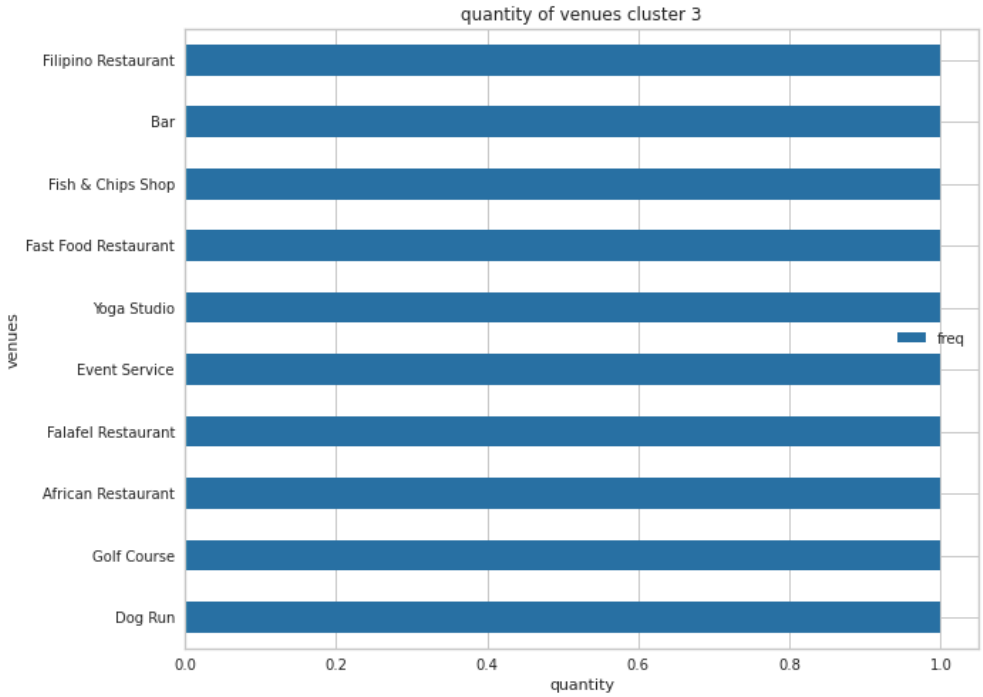
CLUSTER 2 (Purple points in map)

52 Neighbourhoods, the biggest cluster:



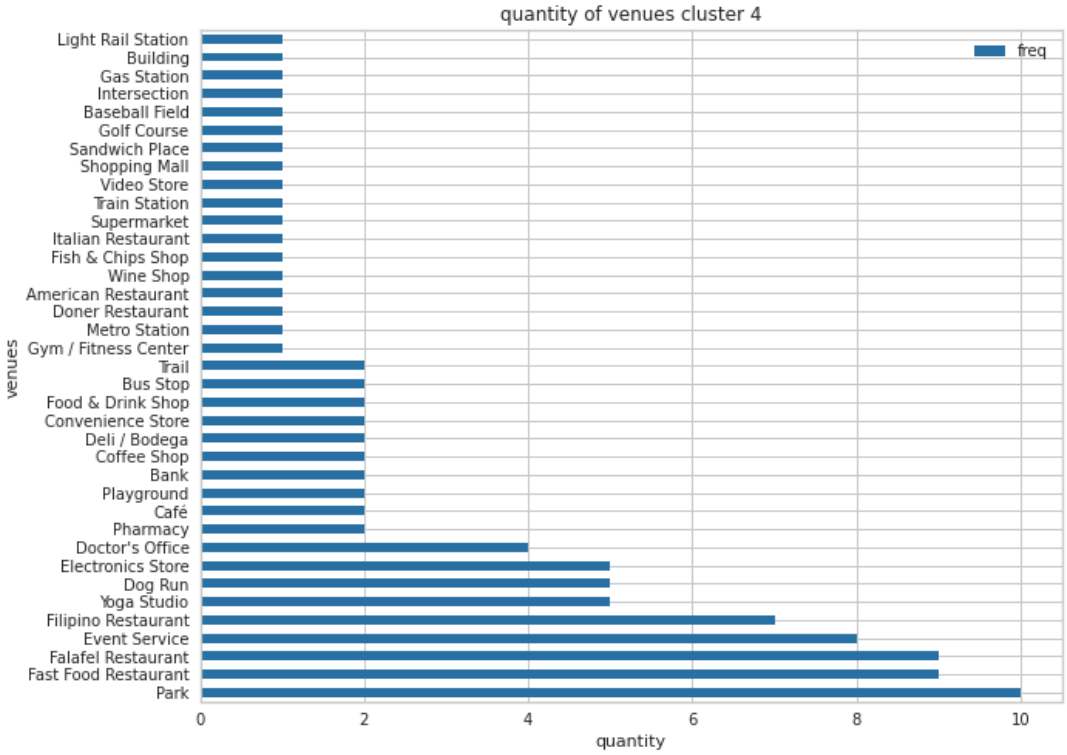
CLUSTER 3

One neighbourhood, Elms-Old Rexdale:



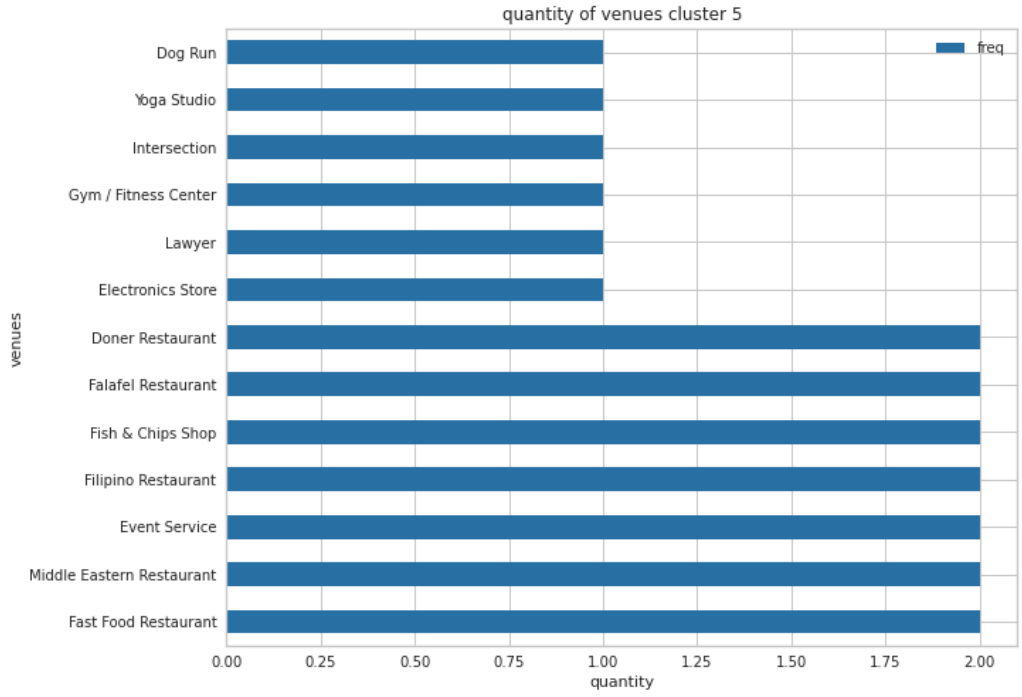
CLUSTER 4

10 neighbourhoods, the second biggest:



CLUSTER 5

2 neighbourhoods, Henry Farm and Highland Creek:



PUBLIC TRANSPORTATION AND PARK VENUES BY CLUSTER

CLUSTER 1:

```
Metro Stations Cluster 1:  [0. 0.]
Light Rail  Cluster 1:  [0. 0.]
Bus Station Cluster 1:  [0. 0.]
Bus Stop Cluster 1:  [0. 0.]
Train Station Cluster 1:  [0. 0.]
Park Cluster 1:  [0. 0.]
```

CLUSTER 2:

```
Metro Stations Cluster 2:  ['Metro Station' 1]
Light Rail  Cluster 2:  ['Light Rail Station' 1]
Bus Station Cluster 2:  ['Bus Station' 1]
Bus Stop Cluster 2:  [0. 0.]
Train Station Cluster 2:  [0. 0.]
Park Cluster 2:  ['Park' 9]
```

CLUSTER 3:

```
Metro Stations Cluster 3:  [0. 0.]
Light Rail Cluster 3:  [0. 0.]
Bus Station Cluster 3:  [0. 0.]
Bus Stop Cluster 3:  [0. 0.]
Train Station Cluster 3:  [0. 0.]
Park Cluster 3:  [0. 0.]
```

CLUSTER 4:

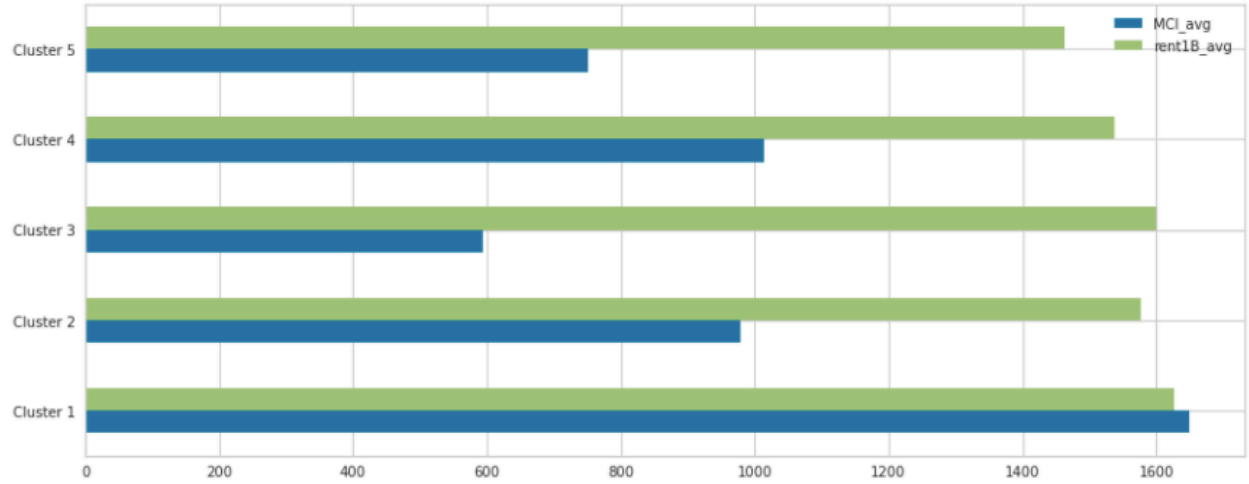
```
Metro Stations Cluster 4:  ['Metro Station' 1]
Light Rail Cluster 4:  ['Light Rail Station' 1]
Bus Station Cluster 4:  [0. 0.]
Bus Stop Cluster 4:  ['Bus Stop' 2]
Train Station Cluster 4:  ['Train Station' 1]
Park Cluster 4:  ['Park' 10]
```

CLUSTER 5:

```
Metro Stations Cluster 5:  [0. 0.]
Light Rail Cluster 5:  [0. 0.]
Bus Station Cluster 5:  [0. 0.]
Bus Stop Cluster 5:  [0. 0.]
Train Station Cluster 5:  [0. 0.]
Park Cluster 5:  [0. 0.]
```

4.5.2- Exploring One Bedroom Average Rent apartments and Average MCI:

- The average of MCI and rent of one-bedroom apartments in clusters.




```
The average MCI Cluster 1: 1650.0
The average MCI Cluster 2: 977.98
The average MCI Cluster 3: 593.0
The average MCI Cluster 4: 1013.3
The average MCI Cluster 5: 751.5
```

```
The average Rent 1 Bed. Cluster 1: 1627.0
The average Rent 1 Bed. Cluster 2: 1577.4
The average Rent 1 Bed. Cluster 3: 1600.0
The average Rent 1 Bed. Cluster 4: 1536.9
The average Rent 1 Bed. Cluster 5: 1462.0
```

5. Results and Discussion

The first evaluation and filter were based on requirements of safety and value of rent (One-bedroom apartments), these let it clean the neighborhoods that did not comply with these restrictions, and it could obtain 66 neighborhoods (of 135).

With this 66 neighborhoods, it can create 5 clusters (using K-elbow and K-means), it can observe from this 5 cluster:

- **Cluster 1:** With only one neighborhood, Agincourt South-Malvern West, its venues are diverse, between food places, Electronic Store, Yoga Studio. This Cluster doesn't have parks or public transportations options (at least among its most common). The Average Rent for one-bedroom apartments is higher (CAD 1.627), as well as it is Major Crime Indicator – MCI (1.650).

- **Cluster2:** Is the biggest cluster with 52 neighborhoods, their most commons venues are the Coffee Shop, Fast Food Restaurants (more than 20 each one), other interesting common venues are the Pharmacy and Bank. This Cluster has one Metro Station, one Light Rail Station, and one Bus Station, and nine parks. The Average Rent for one-bedroom apartments is CAD 1.577,4 below. The average of Major Crime Indicator - MCI is the third lowest (977,98).

- **Cluster3:** With only one neighborhood, Elms-Old Rexdale, its venues are diverse, between Dog Run, Golf Course, Yoga Studio, and different types of restaurants, including a bar. This Cluster doesn't have parks or public transportations options (at least among its most common). The Average Rent for one-bedroom apartments is the second higher (CAD 1.600), and its Major Crime Indicator - MCI is the lowest (593).

- **Cluster4:** With ten neighborhoods, is the second biggest cluster, its most commons venues are Park, Fast Food Restaurant, and Falafel Restaurant, other interesting common venues are Yoga Studio, Doctor's Office, Pharmacy, and Bank. This Cluster has one Metro Station, one Light Rail Station and two Bus Station,

one Train Station and ten parks. The Average Rent for one-bedroom apartments is CAD 1.536,9. The average of Major Crime Indicator - MCI is the fourth lowest (1.013,3).

- Cluster 5: It has 2 neighborhoods, Henry Farm and Highland Creek, their most commons venues are different types of restaurants (Fast Food, Middle Eastern, Falafel, etc.), other interesting venues are a Yoga studio and Gym / Fitness Center. This Cluster doesn't have parks or public transportations options (at least among its most common). The Average Rent for one-bedroom apartments is the lowest, CAD 1.462. The average of Major Crime Indicator - MCI is the second lowest (751,5)

6. Conclusion

If it reviews the features that "Linda" give it:

1- The rent price around CAD 1.700 (1 bedroom)
2- Secure Levels
4- Near to different public transportation options (bus, train, metro...)
5- Near to food places
6- Near to a park

The first filter on security levels and rent, let the clusters complied with the rent price and security level that "Linda" desires.

The only clusters that comply with all features are Cluster 2 and Cluster 4, the rest all doesn't have parks or public transportations options (at least among its most common).

Cluster 2 has one Metro Station, one Light Rail Station, and one Bus Station, and nine parks.

Cluster 4 has one Metro Station, one Light Rail Station, and two Bus Station, one Train Station and ten parks.

Cluster 4 could be the first choice because has near more parks and public transportation options, the average of MCI and rent are acceptable too, however, Cluster 2 could be an excellent choice too.

Works Cited

[1] K. El-Assal, "CIC NEWS," 2020. [Online]. Available: https://www.cicnews.com/2020/02/642000-international-students-canada-now-ranks-3rd-globally-in-foreign-student-attraction-0213763.html#gs.rgki25.

[2] Q. Q. Symonds, "Top Universities," [Online]. Available: https://www.topuniversities.com/city-rankings/2019.

[3] City of Toronto, [Online]. Available: https://www.toronto.ca/community-people/moving-to-toronto/about-toronto/.

[4] Zumper, 2021. [Online]. Available: https://www.zumper.com/rent-research/toronto-on.

[5] Canada Mortgage and Housing Corporation, 2020. [Online]. Available: https://www.cmhc-schl.gc.ca/en/data-and-research/data-tables/rental-market.

[6] Toronto Police, 2020. [Online]. Available: https://data.torontopolice.on.ca/pages/major-crime-indicators.

[7] Foursquare API , [Online]. Available: https://foursquare.com/developers/.