HELPING TO INTERNATIONAL STUDENT SELECTING NEIGHBOURHOOD IN TORONTO

Jesús Romero Palacio

Capstone Project - The Battle of Neighborhoods (Week 2)

INTRODUCTION: BUSINESS PROBLEM

- Toronto offers various types of lifestyles zones according to different preferences. 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*
    3. *Near to different public transportation options (bus, train, metro…)*
    4. *Near to food places*
    5. *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.

DATA

- The rent price around CAD 1.700 (1 bedroom): From Zumper (home rental website) [9] and Canada Mortgage and Housing Corporation, the average rent.
- Secure leves: From Toronto Police Service, the Major Crime Indicator - MCI (2014 to 2019).
- Neighbourhoods fit the desired rental value and securest: From the previous data.
- The coordinates of neighbourhoods fit the desired rental value and securest: From Toronto Police Service, the Major Crime Indicator - MCI.
- Venues in neighborhoods: Foursquare API to get the most common venues from the Toronto selected neighbourhoods.
- General information about neighbourhoods: From the previous data.

METHODOLOGY

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.

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.

Data Science methodology:

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

METHODOLOGY

Neighbourhoods fit the desired rental value and securest:

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

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.

Venues in neighbourhoods:

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

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

METHODOLOGY

Venues in neighbourhoods:

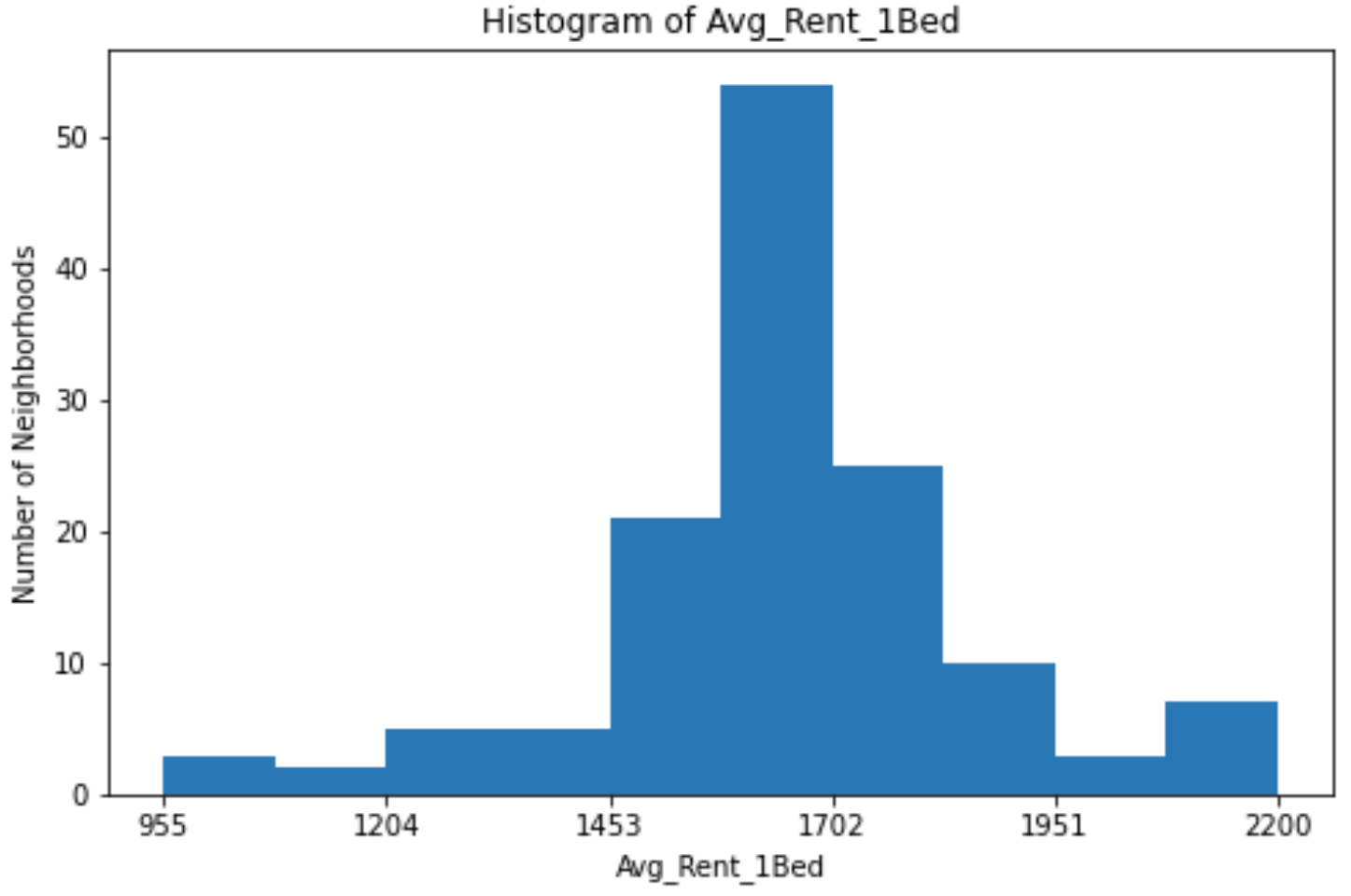
- It classified by venue category and group by neighbourhood.

- 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.

The average rent 1 bedroom apartment)



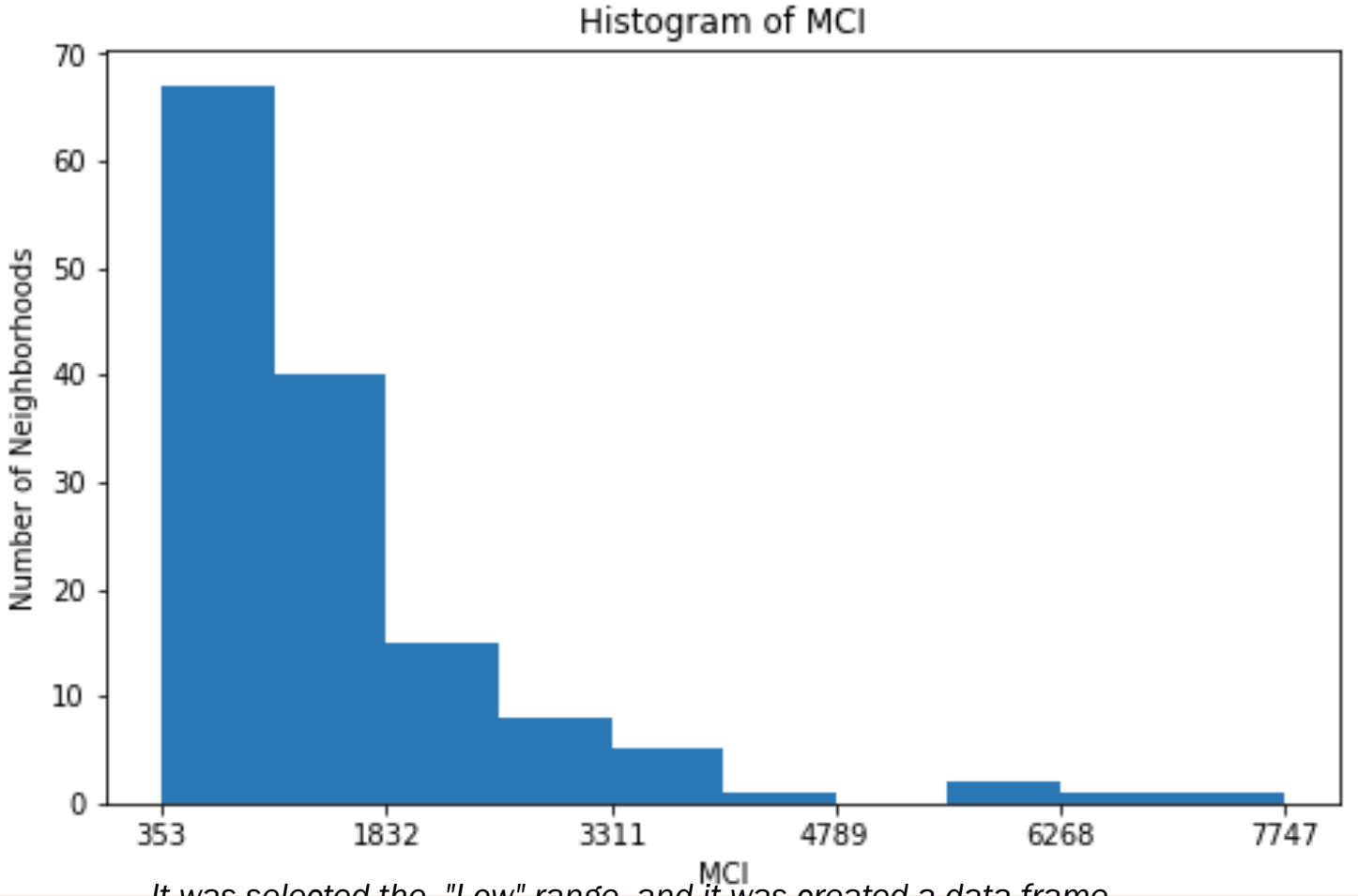
It was selected the "Low", "Mid Low" and "Mid High" ranges, and it was created a dataframe (desire_rent) with the neigbourhoods and its average rent.

ANALYSIS

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

Secure Levels. Major Crime Indicator – MCI

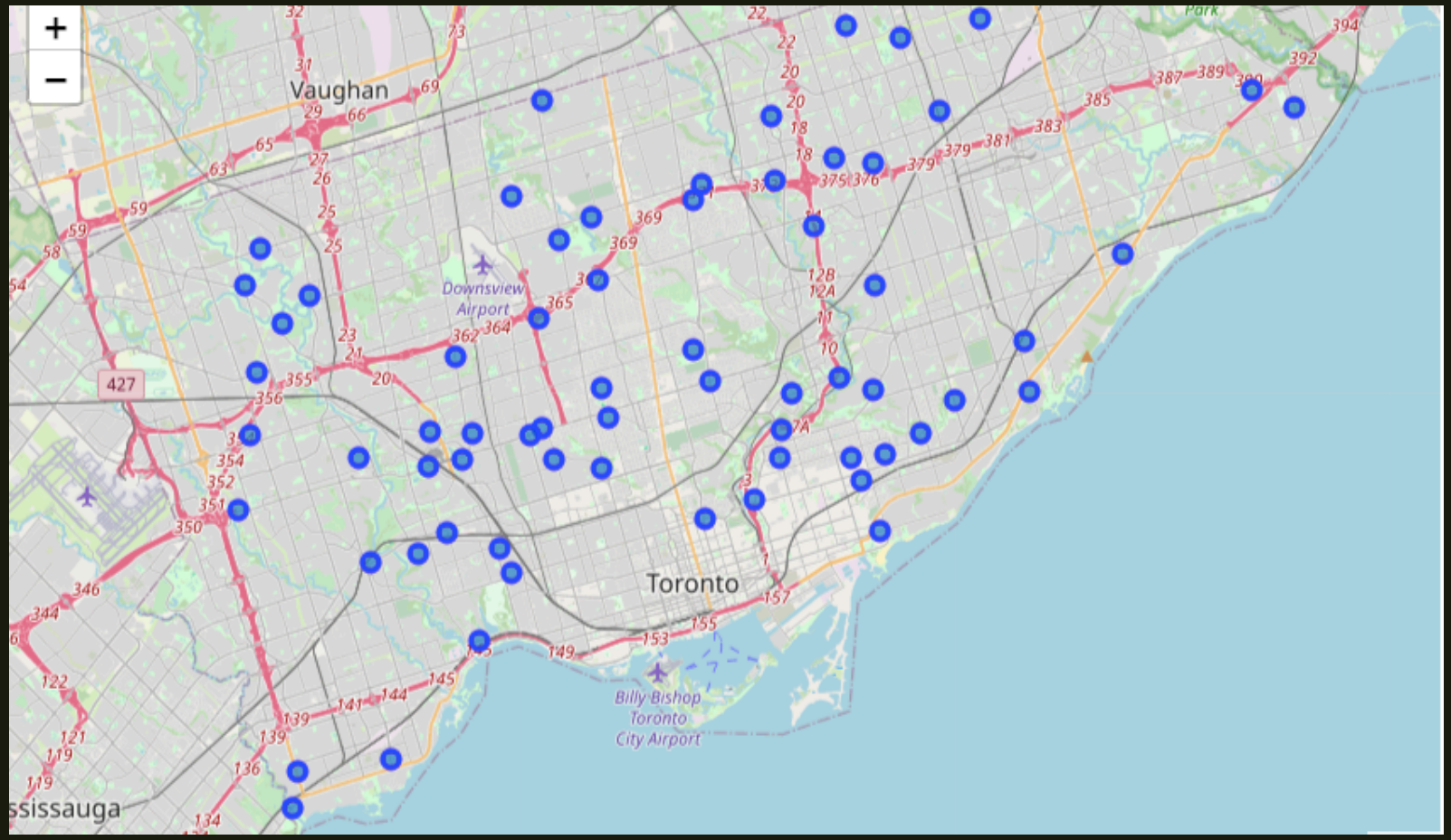


It was selected the "Low" range, and it was created a data frame (desire_secure) with the neighbourhoods and their coordinates.

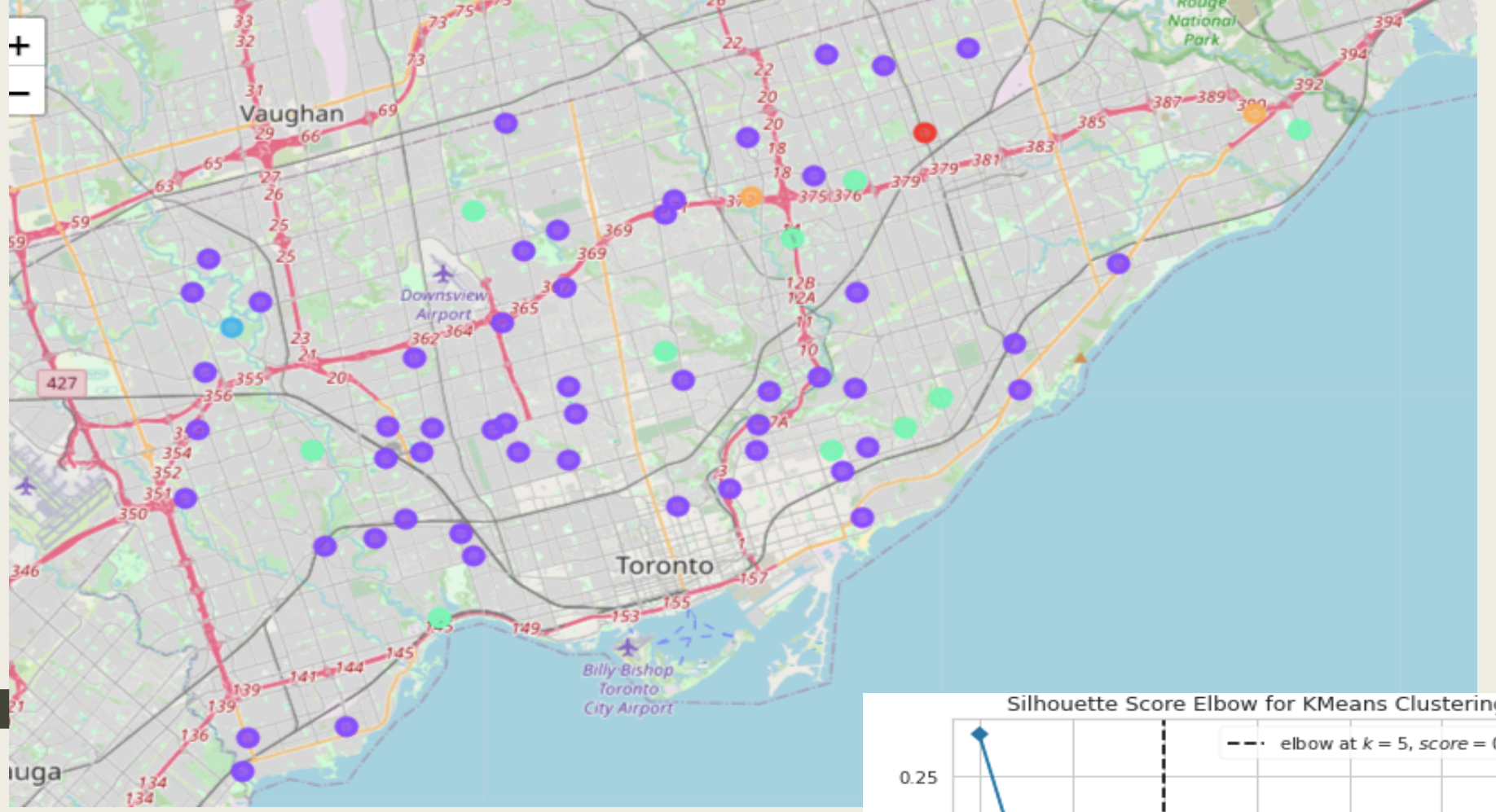
ANALYSIS

Range of Major Crime Indicator – MCI:

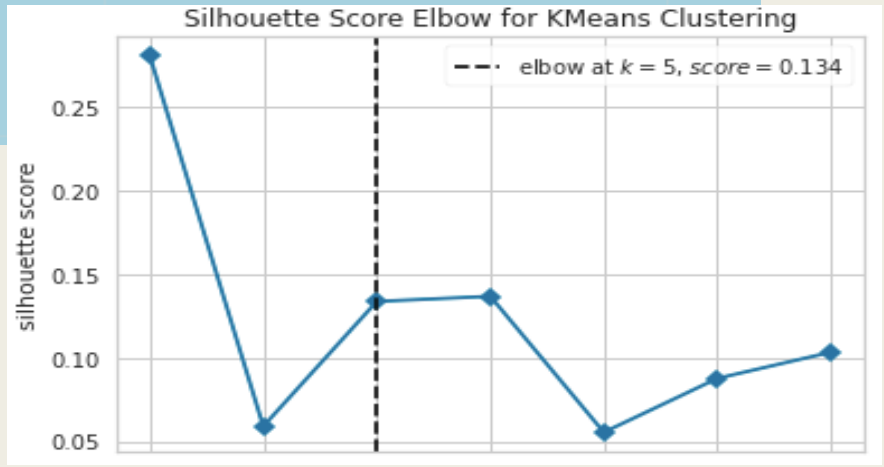
- **Low:** 353 - 1832
- **Mid Low:** 1832 - 3311
- **Mid High:** 3311 - 4790
- **High:** 4790 - 6269
- **Upper High:** 6269 - 7747

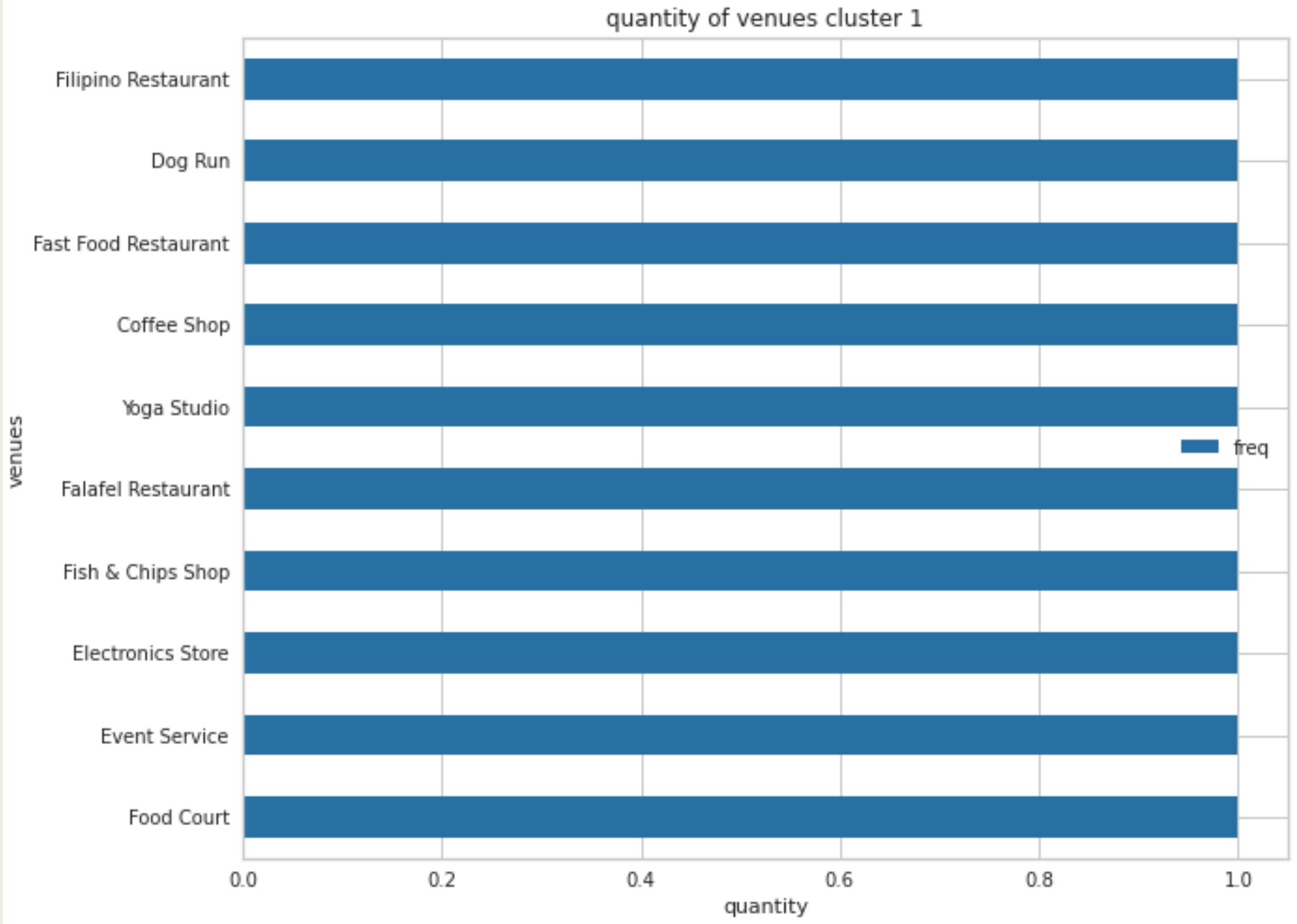


NEIGHBOURHOODS THAT FIT THE DESIRED RENTAL VALUE SECURE LEVELS AND A MAP WITH THEIR GEOLOCATIONS



Geolocation of Clusters (5)

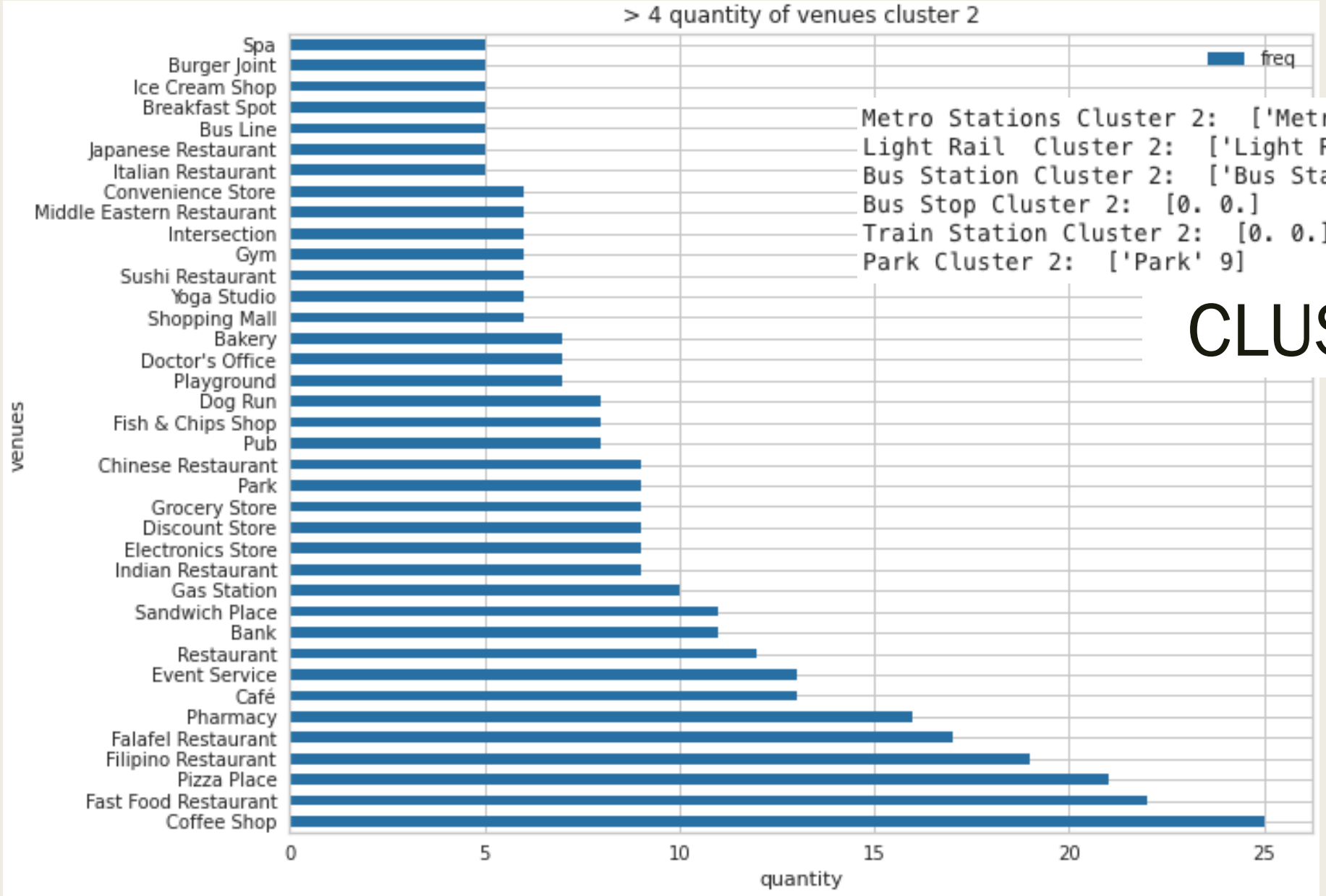




CLUSTER 1

```
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.]
```

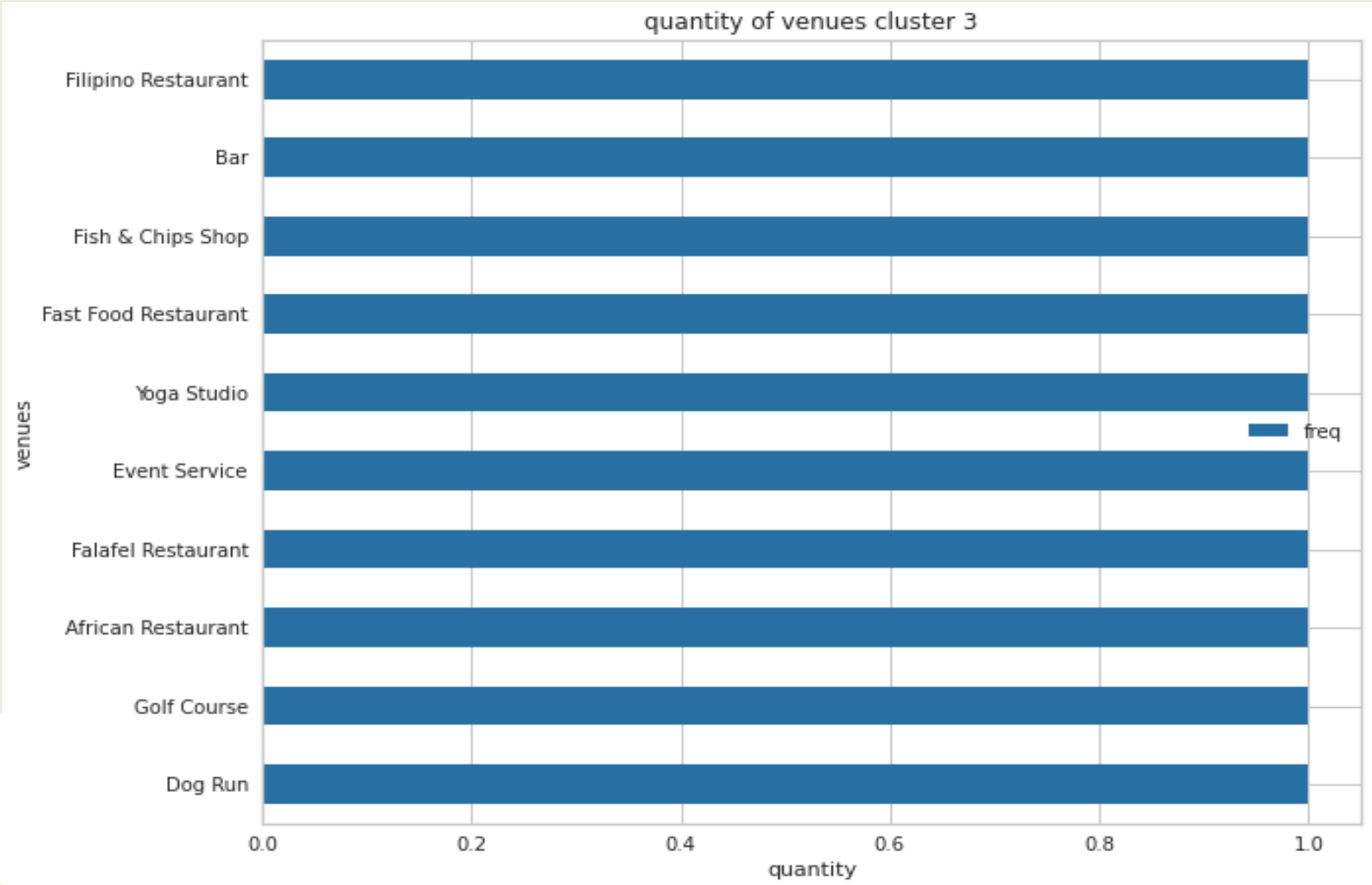
One neighbourhood, Agincourt South-Malvern West,
10 different categories of venues



52 Neighbourhoods, the biggest cluster

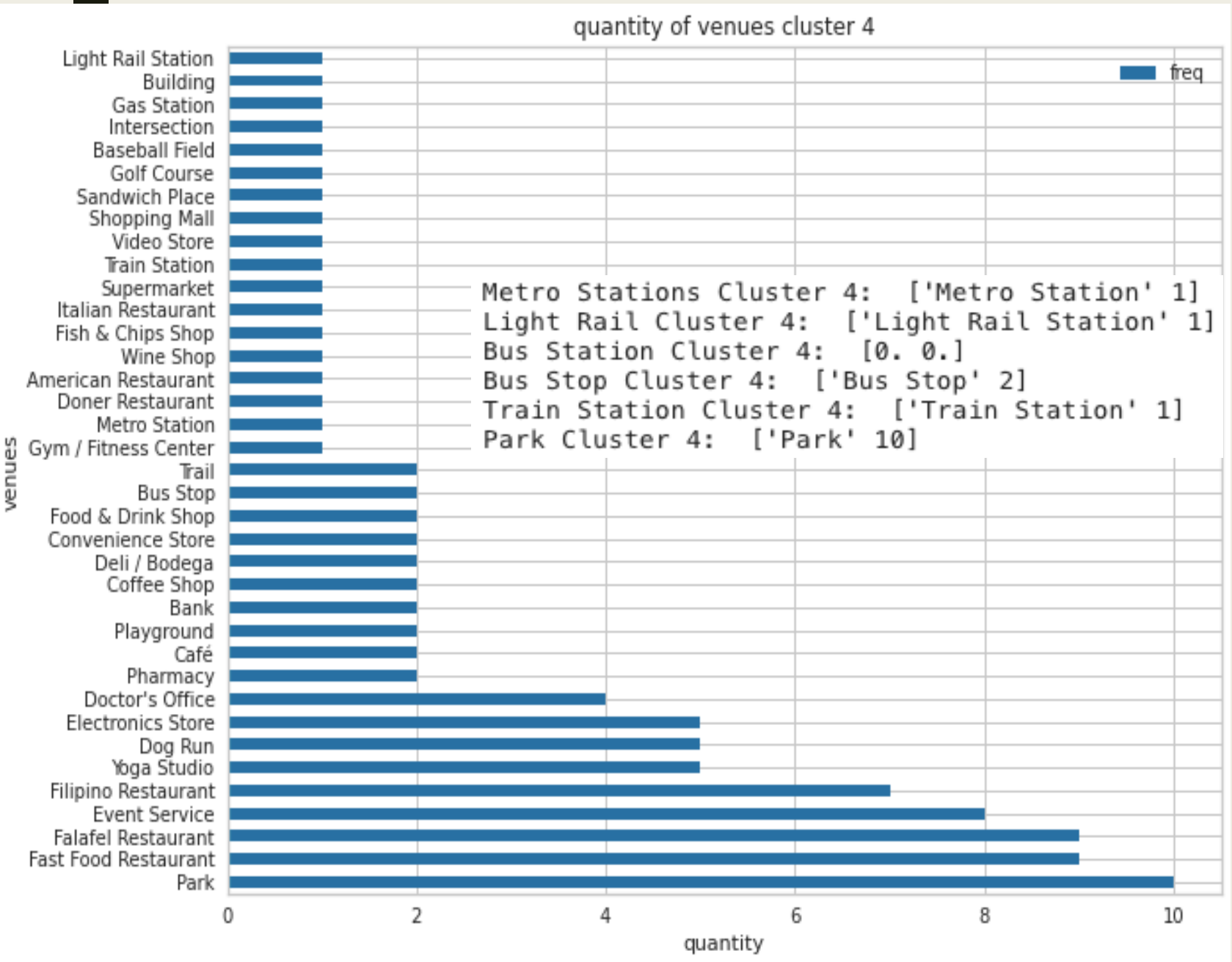
CLUSTER 3

One neighbourhood, Elms-Old Rexdale



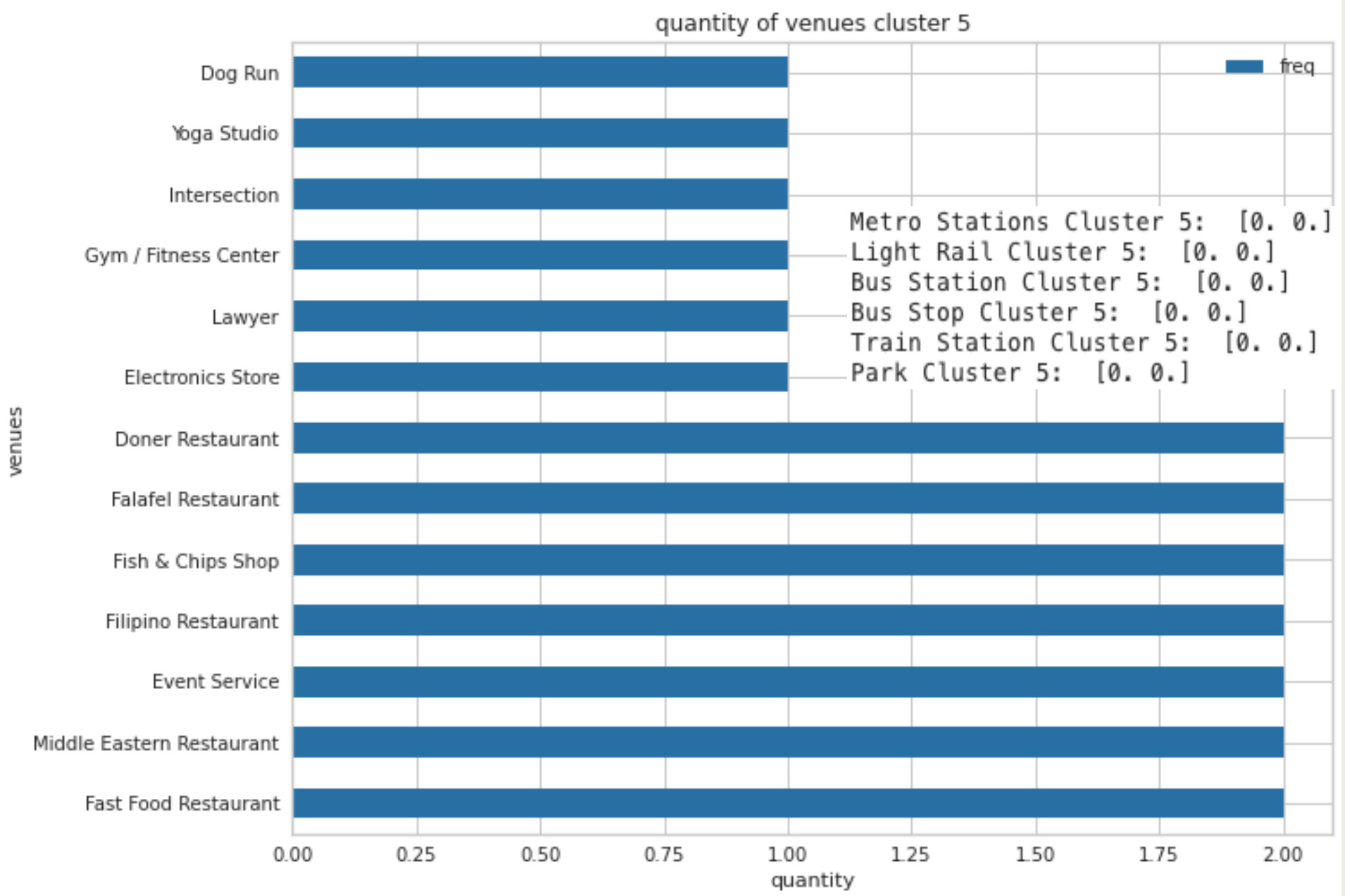

```
Metro Stations Cluster 3:  [0. 0.]
Light Rail Cluster 3:  [0. 0.]
Bus Station Cluster 3:  [0. 0.]
Bus Stop Cluster 3:  [0. 0.]
Train Station Cluster 3:  [0. 0.]
Park Cluster 3:  [0. 0.]
```

10 neighbourhoods, the second biggest



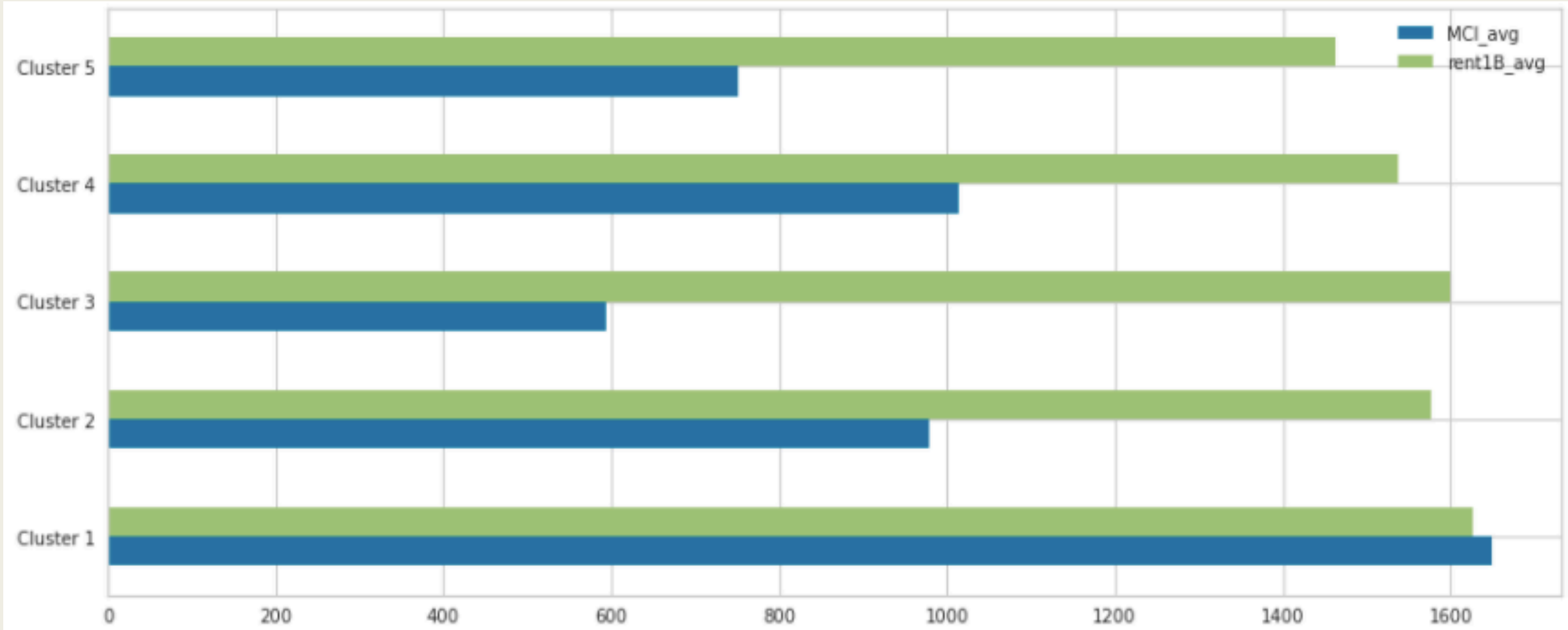
CLUSTER 4

2 neighbourhoods, Henry Farm and Highland Creek



CLUSTER 5

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
```

Conclusion

- 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.