
Determining optimal location for new coffee shop in Toronto city

Edwin Simbaña

edwinsimb@gmail.com

09/01/2020



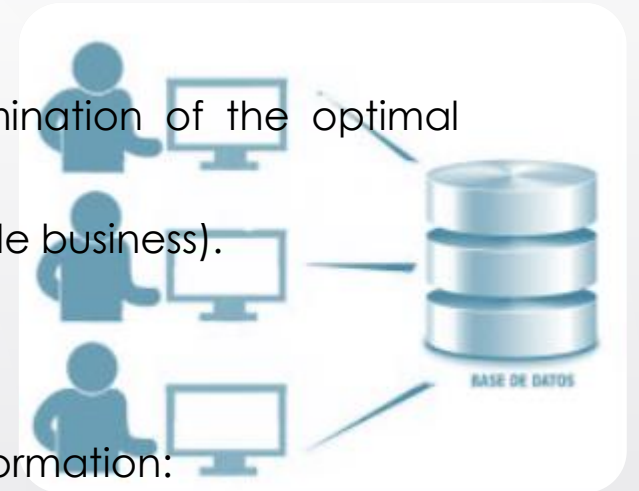


Introduction: Business Problem

- The success in **opening a new branch** is often determined by its location. A **suitable location** allows to acquire higher profits and thus it allows **further expansion of the brand**. For this reason, this project will try to find the **optimal location** to open a **new cafeteria** in a neighborhood in the city of Toronto.
- Since there are many coffee shops in Toronto, attempts will be made to identify locations that contain **few coffee shops** and whose **rating is low**, in order to open a cafeteria with better features to be able to capture the largest number of customers.
- To carry out this project, cluster techniques will be used, such as the algorithm of k-means and descriptive statistics on the information that will be detailed below. In addition, we will assume that there is a cafeteria with good income in the **Wychwood Park neighborhood**, so the **optimal location should be similar to that neighborhood**. This project may be of great interest to big coffee food industries such as: **Starbucks**.

Data

- Based on the problem definition, the factors that will influence the determination of the optimal location are:
- Number of coffee shops, restaurants, schools, etc in neighborhoods (all available business).
- Ratings of coffee shops in neighborhoods.
- Sociodemographic information of the neighborhoods.
- The following data sources will be needed to extract/generate the required information:
- The list of existing neighborhoods in the city of Toronto will be obtained from the **Neighborhood Profiles** from the link <https://www.toronto.ca/>. In addition, this database contains population and population density information that will be used in the analysis.
- The coordinates of the neighborhood centers of the city of Toronto will be obtained using the **geopy library** to obtain the latitude and longitude of the corresponding coordinates.
- The number of restaurants, coffee shops, educational centers, etc; with their respective location in each neighborhood and rating for available coffee shops will be obtained using the **Foursquare API**.



Collecting and gathering data

<https://www.toronto.ca/>

| | POPULATION_2016 | POPULATION_DENSITY | YOUTH_PEOPLE | WORKING_PEOPLE | PRERETIREMENT_PEOPLE | MANAGEMENT_OCCUPATION | BUSINE |
|------------------------------|-----------------|--------------------|--------------|----------------|----------------------|-----------------------|---------|
| Agincourt North | 29.113 | 3.929 | 3.705 | 11.305 | 4.230 | | 960.000 |
| Agincourt South-Malvern West | 23.757 | 3.034 | 3.360 | 9.965 | 3.265 | | 960.000 |
| Alderwood | 12.054 | 2.435 | 1.235 | 5.220 | 1.825 | | 840.000 |
| Annex | 30.526 | 10.863 | 3.750 | 15.040 | 3.480 | | 2.645 |
| Banbury-Don Mills | 27.695 | 2.775 | 2.730 | 10.810 | 3.555 | | 2.120 |

Foursquare API

| | ID | NAME | CATEGORY | NEIGHBORHOOD |
|---|--------------------------|--|--------------------|-----------------|
| 0 | 51fe9e21498ecbf4fef45c1 | Menchie's | Ice Cream Shop | Agincourt North |
| 1 | 4b93d4a7f964a520eb5334e3 | Saravanaa Bhavan South Indian Restaurant | Indian Restaurant | Agincourt North |
| 2 | 4aee2557f964a52080d221e3 | Samosa King - Embassy Restaurant | Indian Restaurant | Agincourt North |
| 3 | 4b9d48b2f964a52011a036e3 | Shoppers Drug Mart | Pharmacy | Agincourt North |
| 4 | 4dacc7855da32d679da9ee55 | Congee Town 太皇名粥 | Chinese Restaurant | Agincourt North |

Foursquare API

| | LIKES | RATING | TIPS |
|------------------------------|-------|----------|------|
| NEIGHBORHOOD | | | |
| Agincourt North | 5.5 | 6.879975 | 34.5 |
| Agincourt South-Malvern West | 21.0 | 8.155157 | 32.0 |
| Alderwood | 16.0 | 3.073492 | 26.0 |
| Annex | 26.0 | 5.932260 | 19.8 |
| Banbury-Don Mills | 17.0 | 4.661452 | 43.0 |

| | POPULATION_2016 | POPULATION_DENSITY | YOUTH_PEOPLE | WORKING_PEOPLE | PRERETIREMENT_PEOPLE | MANAGEMENT_OCCUPATION | BUSINE |
|------------------------------|-----------------|--------------------|--------------|----------------|----------------------|-----------------------|---------|
| Agincourt North | 29.113 | 3.929 | 3.705 | 11.305 | 4.230 | | 960.000 |
| Agincourt South-Malvern West | 23.757 | 3.034 | 3.360 | 9.965 | 3.265 | | 960.000 |
| Alderwood | 12.054 | 2.435 | 1.235 | 5.220 | 1.825 | | 840.000 |
| Annex | 30.526 | 10.863 | 3.750 | 15.040 | 3.480 | | 2.645 |
| Banbury-Don Mills | 27.695 | 2.775 | 2.730 | 10.810 | 3.555 | | 2.120 |

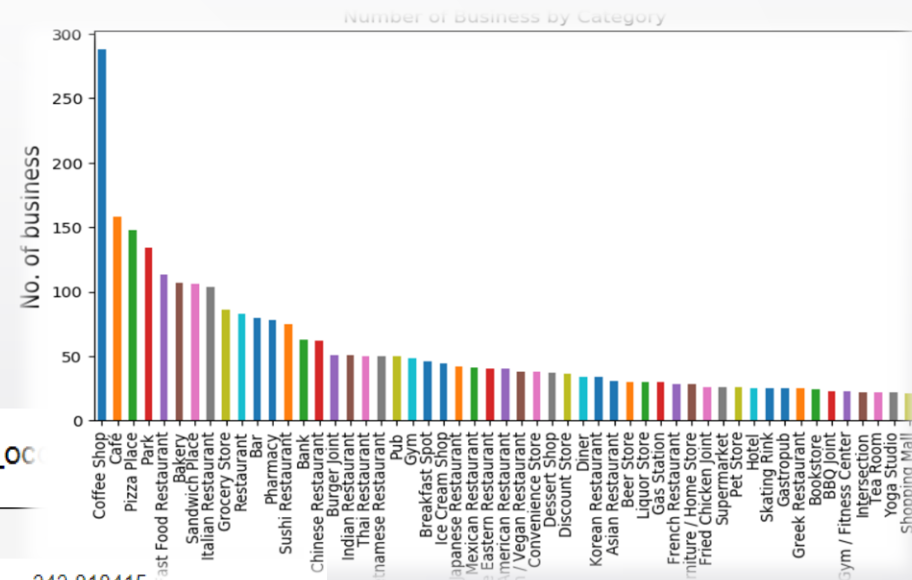
Showing neighborhood on map

Available venues are displayed on map for Toronto City



Statistical table

- 82 available rows
- 67 numeric columns

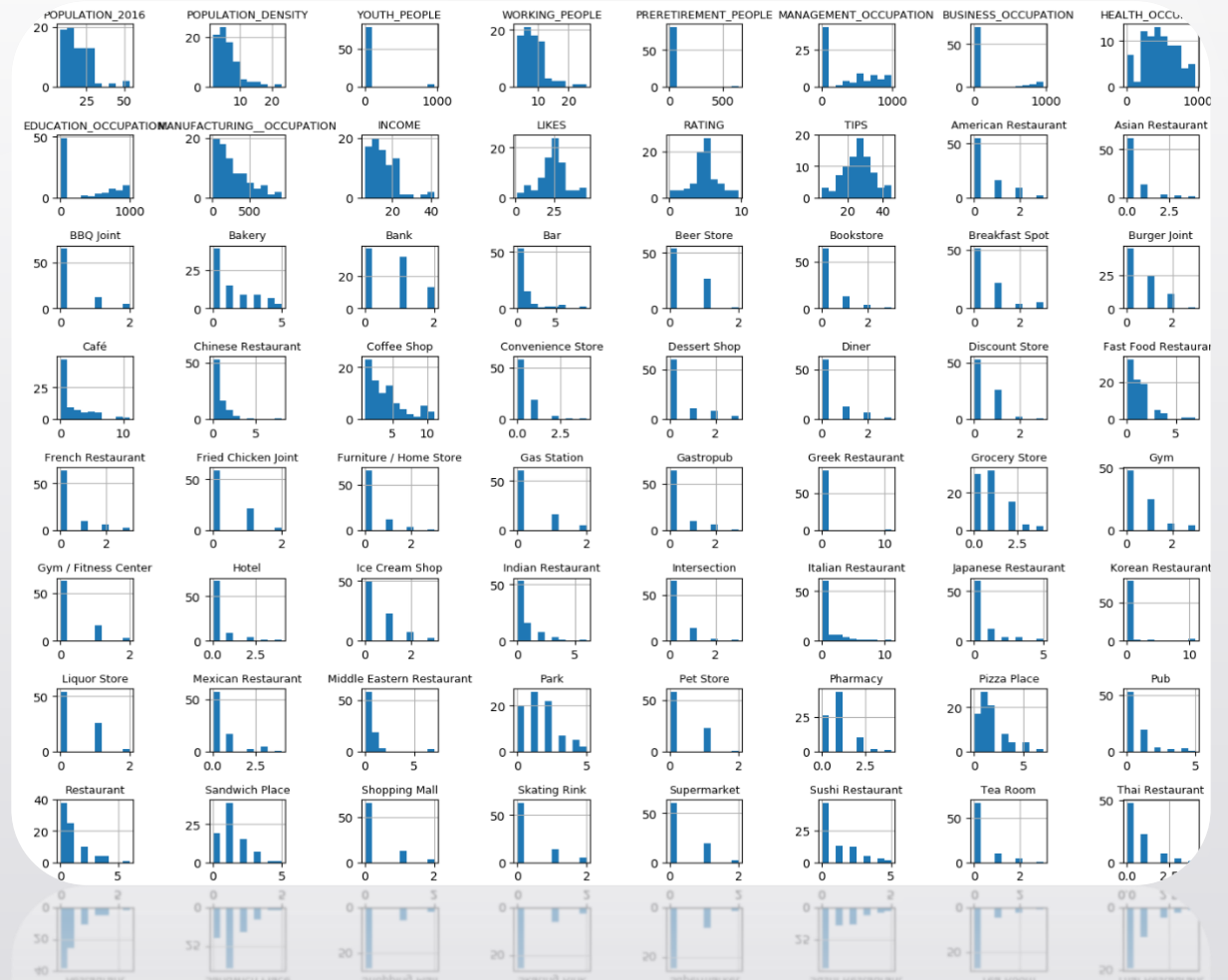


| | POPULATION_2016 | POPULATION_DENSITY | YOUTH_PEOPLE | WORKING_PEOPLE | PRERETIREMENT_PEOPLE | MANAGEMENT_OC |
|-------|-----------------|--------------------|--------------|----------------|----------------------|---------------|
| count | 82.000000 | 82.00000 | 82.000000 | 82.000000 | 82.000000 | |
| mean | 19.306000 | 6.32322 | 36.495244 | 8.569878 | 10.296951 | 343.818415 |
| std | 8.710025 | 3.75520 | 175.564397 | 4.048433 | 71.523426 | 364.293723 |
| min | 7.607000 | 1.57000 | 1.040000 | 3.245000 | 1.100000 | 1.020000 |
| 25% | 13.002750 | 3.72750 | 1.511250 | 5.860000 | 1.596250 | 1.413750 |
| 50% | 16.935000 | 5.41850 | 2.180000 | 7.850000 | 2.145000 | 310.000000 |
| 75% | 23.901500 | 7.46950 | 3.217500 | 10.636250 | 3.117500 | 697.500000 |
| max | 53.485000 | 23.04400 | 960.000000 | 25.850000 | 650.000000 | 985.000000 |



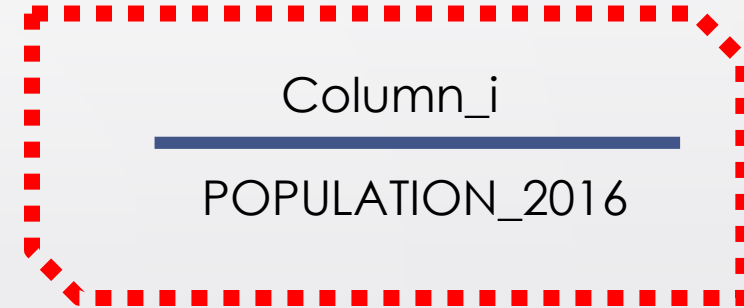
Histogram by column

- Many columns are good distributed.
- Some columns have low values. Then, plots look separated.



Data is standardized

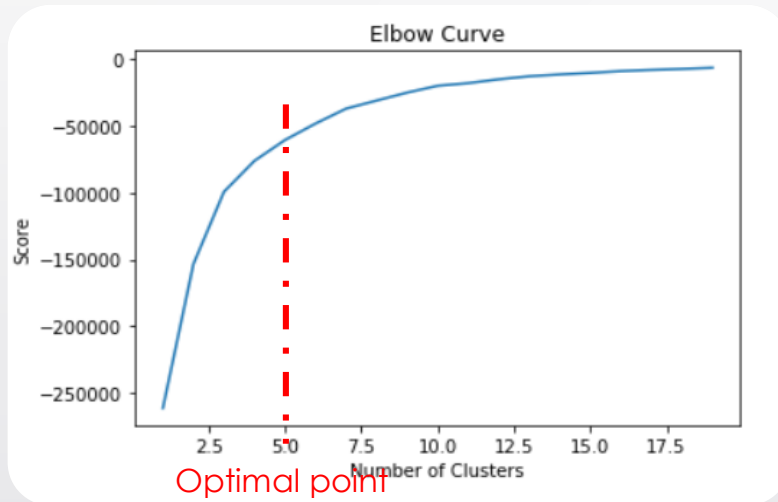
| | POPULATION_2016 | POPULATION_DENSITY | YOUTH_PEOPLE | WORKING_PEOPLE | PRERETIREMENT_PEOPLE | MANAGEMENT_OCCUPATION | BUSINE |
|------------------------------|-----------------|--------------------|--------------|----------------|----------------------|-----------------------|---------|
| Agincourt North | 29.113 | 3.929 | 3.705 | 11.305 | 4.230 | | 960.000 |
| Agincourt South-Malvern West | 23.757 | 3.034 | 3.360 | 9.965 | 3.265 | | 960.000 |
| Alderwood | 12.054 | 2.435 | 1.235 | 5.220 | 1.825 | | 840.000 |
| Annex | 30.526 | 10.863 | 3.750 | 15.040 | 3.480 | | 2.645 |
| Banbury-Don Mills | 27.695 | 2.775 | 2.730 | 10.810 | 3.555 | | 2.120 |



To standardize we use
POPULATION 2016 for each
neighborhood

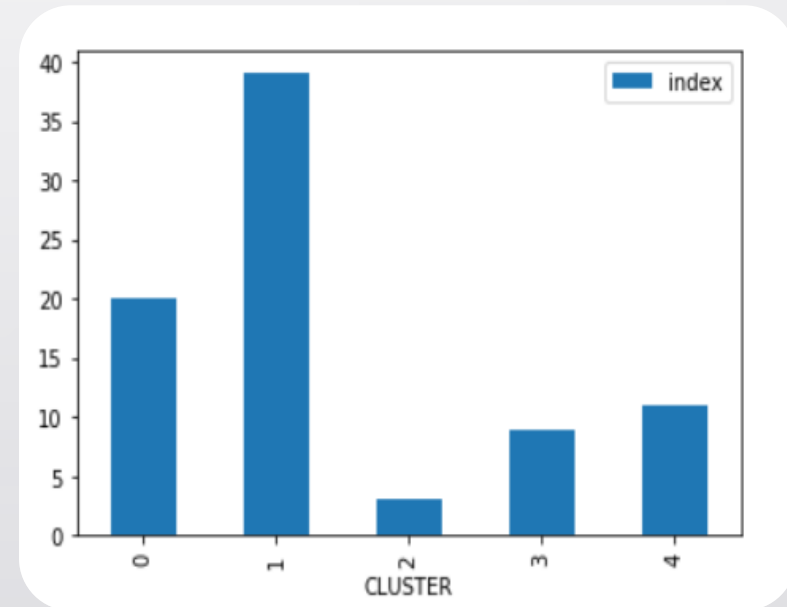
| | POPULATION_2016 | POPULATION_DENSITY | YOUTH_PEOPLE | WORKING_PEOPLE | PRERETIREMENT_PEOPLE | MANAGEMENT_OCCUPATION | BUSINE |
|------------------------------|-----------------|--------------------|--------------|----------------|----------------------|-----------------------|---------|
| Agincourt North | 29.113 | 3.929 | 3.705 | 11.305 | 4.230 | | 960.000 |
| Agincourt South-Malvern West | 23.757 | 3.034 | 3.360 | 9.965 | 3.265 | | 960.000 |
| Alderwood | 12.054 | 2.435 | 1.235 | 5.220 | 1.825 | | 840.000 |
| Annex | 30.526 | 10.863 | 3.750 | 15.040 | 3.480 | | 2.645 |
| Banbury-Don Mills | 27.695 | 2.775 | 2.730 | 10.810 | 3.555 | | 2.120 |

Clustering analysis



Wychwood is located in cluster 1. Then, we have to find the optimal location within that cluster.

- We tried with k from 2 to 20
- Optimal number of cluster of **5**



Selecting optimal location in cluster 1

| | POPULATION_2016 | POPULATION_DENSITY | LATITUDE | LONGITUDE | LIKES | Coffee Shop | RATING | TIPS | CLUSTER |
|------------------------|-----------------|--------------------|-----------|------------|-----------|-------------|----------|-----------|---------|
| Glenfield-Jane Heights | 30.491 | 5.864 | 47.540336 | -52.842015 | 45.000000 | 1.0 | 0.065178 | 15.000000 | 1 |
| Rosedale-Moore Park | 20.923 | 4.500 | 43.690388 | -79.383297 | 23.428571 | 5.0 | 1.605968 | 26.428571 | 1 |
| Bendale | 29.960 | 4.011 | 43.753520 | -79.255336 | 38.000000 | 1.0 | 2.123747 | 26.000000 | 1 |
| Woburn | 53.485 | 4.345 | 43.759824 | -79.225291 | 31.500000 | 2.0 | 2.385849 | 18.500000 | 1 |
| O'Connor-Parkview | 18.675 | 3.780 | 43.705537 | -79.312718 | 13.000000 | 1.0 | 2.787105 | 5.000000 | 1 |
| Leaside-Bennington | 16.828 | 3.596 | 43.702062 | -79.378044 | 24.500000 | 5.0 | 3.170992 | 19.000000 | 1 |
| Stonegate-Queensway | 25.051 | 3.199 | 43.621950 | -79.523499 | 22.000000 | 4.0 | 3.212015 | 23.000000 | 1 |
| Moss Park | 20.506 | 14.753 | 43.654644 | -79.369728 | 27.300000 | 9.0 | 3.996222 | 21.000000 | 1 |
| Bayview Village | 21.396 | 4.195 | 43.769197 | -79.376662 | 22.500000 | 4.0 | 4.002400 | 29.250000 | 1 |
| Little Portugal | 15.559 | 12.859 | 43.647413 | -79.431116 | 18.000000 | 4.0 | 4.066465 | 35.000000 | 1 |
| Church-Yonge Corridor | 31.340 | 23.044 | 43.679919 | -79.388689 | 18.750000 | 3.0 | 4.272050 | 20.750000 | 1 |
| Don Valley Village | 27.051 | 6.441 | 43.792673 | -79.354722 | 28.500000 | 4.0 | 4.535776 | 39.750000 | 1 |
| Wychwood | 14.349 | 8.541 | 43.682094 | -79.423855 | 27.500000 | 3.0 | 4.605463 | 27.000000 | 1 |
| Willowdale East | 50.434 | 10.087 | 43.761510 | -79.410923 | 23.954545 | 11.0 | 4.621152 | 23.545455 | 1 |
| Willowdale West | 16.936 | 5.820 | 43.761510 | -79.410923 | 23.954545 | 11.0 | 4.621152 | 23.545455 | 1 |
| Banbury-Don Mills | 27.695 | 2.775 | 43.734804 | -79.357243 | 17.000000 | 1.0 | 4.661452 | 43.000000 | 1 |
| North Riverdale | 11.916 | 6.770 | 43.665470 | -79.352594 | 31.210526 | 7.0 | 4.730109 | 28.000000 | 1 |
| South Riverdale | 27.876 | 3.136 | 43.665470 | -79.352594 | 31.210526 | 7.0 | 4.730109 | 28.000000 | 1 |
| Junction Area | 14.366 | 5.442 | 43.665478 | -79.470352 | 30.666667 | 6.0 | 4.868062 | 31.500000 | 1 |
| Bay Street Corridor | 25.797 | 14.097 | 43.664951 | -79.387395 | 26.555556 | 9.0 | 5.001478 | 22.888889 | 1 |
| High Park North | 22.162 | 11.726 | 43.654709 | -79.460071 | 7.000000 | 2.0 | 5.058805 | 20.250000 | 1 |
| Forest Hill South | 10.732 | 4.380 | 43.693559 | -79.413902 | 20.000000 | 1.0 | 5.080857 | 24.000000 | 1 |
| Danforth East York | 17.180 | 7.881 | 43.686433 | -79.300355 | 13.400000 | 5.0 | 5.192319 | 29.300000 | 1 |
| L'Amoreaux | 43.993 | 6.144 | 43.799003 | -79.305967 | 9.000000 | 2.0 | 5.286417 | 13.000000 | 1 |
| Oakwood Village | 21.210 | 9.511 | 43.682725 | -79.438055 | 23.428571 | 4.0 | 5.347171 | 27.571429 | 1 |
| Roncesvalles | 14.974 | 9.851 | 43.651443 | -79.451038 | 20.545455 | 9.0 | 5.500015 | 23.727273 | 1 |
| East End-Danforth | 21.381 | 8.038 | 43.668440 | -79.330670 | 41.500000 | 4.0 | 5.509504 | 19.500000 | 1 |
| Trinity-Spadina | 15.179 | 4.685 | 43.720109 | -79.402352 | 20.333333 | 3.0 | 5.722982 | 26.333333 | 1 |
| East End-Danforth | 51.381 | 8.038 | 43.668440 | -79.330670 | 41.500000 | 4.0 | 5.509504 | 19.500000 | 1 |
| Trinity-Spadina | 15.179 | 4.685 | 43.720109 | -79.402352 | 20.333333 | 3.0 | 5.722982 | 26.333333 | 1 |
| Oakwood Village | 21.210 | 9.511 | 43.682725 | -79.438055 | 23.428571 | 4.0 | 5.347171 | 27.571429 | 1 |
| L'Amoreaux | 43.993 | 6.144 | 43.799003 | -79.305967 | 9.000000 | 2.0 | 5.286417 | 13.000000 | 1 |
| Danforth East York | 17.180 | 7.881 | 43.686433 | -79.300355 | 13.400000 | 5.0 | 5.192319 | 29.300000 | 1 |
| Forest Hill South | 10.732 | 4.380 | 43.693559 | -79.413902 | 20.000000 | 1.0 | 5.080857 | 24.000000 | 1 |

- Optimal location is at **Glenfield-Jane Heights**.
- Low average rating.
- Population over average

- Statistical table for cluster 1

| | POPULATION_2016 | POPULATION_DENSITY | LATITUDE | LONGITUDE | LIKES | Coffee Shop | RATING | TIPS | CLUSTER |
|-------|-----------------|--------------------|-----------|------------|-----------|-------------|-----------|-----------|---------|
| count | 39.000000 | 39.000000 | 39.000000 | 39.000000 | 39.000000 | 39.000000 | 39.000000 | 39.000000 | 39.0 |
| mean | 23.148923 | 7.174359 | 43.799907 | -78.698773 | 24.758817 | 4.512821 | 4.841883 | 24.280167 | 1.0 |
| std | 9.782530 | 4.109420 | 0.616625 | 4.249827 | 8.251248 | 2.845789 | 1.642675 | 7.086892 | 0.0 |
| min | 10.732000 | 2.724000 | 43.621950 | -79.523499 | 7.000000 | 1.000000 | 0.065178 | 5.000000 | 1.0 |
| 25% | 16.057500 | 4.362500 | 43.665470 | -79.413892 | 20.439394 | 2.500000 | 4.169257 | 20.125000 | 1.0 |
| 50% | 21.396000 | 5.864000 | 43.686433 | -79.387395 | 24.500000 | 4.000000 | 5.001478 | 24.000000 | 1.0 |
| 75% | 27.570500 | 9.540500 | 43.751221 | -79.322967 | 29.194444 | 6.000000 | 5.736806 | 28.000000 | 1.0 |
| max | 53.485000 | 23.044000 | 47.540336 | -52.842015 | 45.000000 | 11.000000 | 8.661944 | 43.000000 | 1.0 |

Conclusion and future directions

- Our analysis reported 38 potential places to open a new coffee shop. This was done by collecting information from various sources to obtain a better result and use clustering algorithms such as k-means to group neighborhoods with similar characteristics.
- Selection of the best neighborhood within cluster 1 was chosen based on the rating. In this case, one with the lowest value, because, by locating a coffee shop in this neighborhood, it is guaranteed to capture a large part of the dissatisfied consumers, which will generate greater profits for the brand.
- This project can generate greater value if information about time series of profitability would be available, With this, it could also predict the possible profitability that would be obtained by locating the new coffee shop in that place.
- The purpose of this project was to identify similar neighborhoods to Wychwood, in order to locate a new coffee shop in optimal location. Using information of different types and advanced clustering techniques. It is concluded that the best place to locate a new coffee shop is **Glenfield-Jane Heights** neighborhood.

