# BATTLE OF NEIGHBORHOODS

Finding a better place in North York Canada

# IBM DATA SCIENCE CAPSTONE PROJECT

syed sulaiman

29 june 2021

# BATTLE OF NEIGHBORHOODS

## OUTLINE:

- Executive Summary
- Introduction
- Methodology
- Results
    - Visualization – Charts
    - Clusters
- Discussion
    - Findings & Implications
- Conclusion

# EXECUTIVE SUMMARY:

The aim of this project is to help people in finding better alternatives for their neighbourhood's based on the number of various facilities available there.

It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in North York, Toronto Canada.

# INTRODUCTION:

According to Aristotle "man is a social animal, who live in a society, so what is society? in simple words society is nothing but the surrounding environment in which a person lives his life. during his life cycle a person is not only influenced by the people's but also by the surrounding environment.

these surrounding environments in other words neighborhoods having adequate venues and facilities plays a vital role in easing the life of a person.

The aim of this project is to help people in finding better alternatives for their neighbourhood's based on the number of various facilities available their.

It will help people making smart and efficient decision on selecting great neighbourhoods out of numbers of other neighbourhood's in North York, Toronto Canada.

As a relatively young nation when compared to countries in Europe, Asia and other regions, Canada is very much a country built with the help of immigration. Canada is also a country that continues to need newcomers in order to grow, thrive and remain a nation of inclusion.

New comers coming to Canada are important for the future of the country. Not only does it allow for more skilled and qualified workers to assist with the employment needs of businesses and thereby support the economic growth of the country.

In this project we are exploring the neighbourhoods in North York, Toronto in order to help these newcomers in selecting proper neighbourhood according to their needs and one which is similar to their native place. The features include are housing price and better school according to ratings, crime rates of that particular area, transportation, emergency services, water resources, sewers and recreational facilities

# METHODOLOGY:

- Data Sources:

- Data Link Used to fetch data :
  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- GeospatialCoordinates.csv

- Data Exploration and Cleaning

  Foursquare API
  - Python Libraries Pandas

- Data Visualization
  - Matplotlib Library
  - Folium Library

## Problem Which Tried to Solve:

The prime objective of the project is to help in decision making by suggesting the best neighborhood available in a new city to the persons who are willing to relocate. suggestions will be based on the number of factors such as

Hotels, malls, transport services, connectivity to airport, markets theaters, gyms etc.

Sorted list of accommodations in terms of housing prices

Sorted list of schools in terms of location, fees, rating and reviews

## Some important libraries used in this project are as follows:

Pandas: For creating and manipulating data frames.

Folium: Python visualization library would be used to visualize the

neighbourhood's cluster distribution of using interactive leaflet map.

Scikit Learn: For Machine learning clustering techniques such as k-means

clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text

format

Beautiful Soup and Requests: To handle http requests, and web scraping.

Matplotlib: Python Plotting Module.

# Data Description:

Data Link Used to fetch data such as Postal code, Borough and Neighborhood: [https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

```
In [2]: url = 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
        data = requests.get(url).text
```

Creating Soup object to extract data

```
In [3]: soup = BeautifulSoup(data, "html5lib")
```

Data Scraped:

```
In [5]: table_contents
```

```
Out[5]: [{'PostalCode': 'M3A', 'Borough': 'North York', 'Neighborhood': 'Parkwoods'},
         {'PostalCode': 'M4A',
          'Borough': 'North York',
          'Neighborhood': 'Victoria Village'},
         {'PostalCode': 'M5A',
          'Borough': 'Downtown Toronto',
          'Neighborhood': 'Regent Park, Harbourfront'},
         {'PostalCode': 'M6A',
          'Borough': 'North York',
          'Neighborhood': 'Lawrence Manor, Lawrence Heights'},
         {'PostalCode': 'M7A',
          'Borough': "Queen's Park",
          'Neighborhood': 'Ontario Provincial Government'},
         {'PostalCode': 'M9A',
          'Borough': 'Etobicoke',
          'Neighborhood': 'Islington Avenue'},
         {'PostalCode': 'M1B',
          'Borough': 'Scarborough',
          'Neighborhood': 'Malvern, Rouge'},
```

Converting list to pandas dataframe

```
In [6]: df = pd.DataFrame(table_contents)
```

```
In [64]: df
```

Out[64]:

| | PostalCode | Borough | Neighborhood |
|---|---|---|---|
| 0 | M3A | North York | Parkwoods |
| 1 | M4A | North York | Victoria Village |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights |
| 4 | M7A | Queen's Park | Ontario Provincial Government |
| ... | ... | ... | ... |
| 98 | M8X | Etobicoke | The Kingsway, Montgomery Road, Old Mill North |
| 99 | M4Y | Downtown Toronto | Church and Wellesley |
| 100 | M7Y | East TorontoBusiness reply mail Processing Cen... | Enclave of M4L |
| 101 | M8Y | Etobicoke | Old Mill South, King's Mill Park, Sunnylea, Hu... |
| 102 | M8Z | Etobicoke | Mimico NW, The Queensway West, South of Bloor,... |

103 rows × 3 columns

Using GeospatialCoordinates CSV data file which contains list of Postal Code Latitude and longitude for North York.

Database After Cleaning:

Out[16]:

| | Postal Code | Latitude | Longitude |
|---|---|---|---|
| 0 | M1B | 43.806686 | -79.194353 |
| 1 | M1C | 43.784535 | -79.160497 |
| 2 | M1E | 43.763573 | -79.188711 |
| 3 | M1G | 43.770992 | -79.216917 |
| 4 | M1H | 43.773136 | -79.239476 |
| ... | ... | ... | ... |
| 98 | M9N | 43.706876 | -79.518188 |
| 99 | M9P | 43.696319 | -79.532242 |
| 100 | M9R | 43.688905 | -79.554724 |
| 101 | M9V | 43.739416 | -79.588437 |
| 102 | M9W | 43.706748 | -79.594054 |

103 rows × 3 columns

Final database:

In [22]:
```
df = df.join(df_geo.set_index('Postal Code'), on='PostalCode')
df
```

Out[22]:

| | PostalCode | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M3A | North York | Parkwoods | 43.753259 | -79.329656 |
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront | 43.654260 | -79.360636 |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights | 43.718518 | -79.464763 |
| 4 | M7A | Queen's Park | Ontario Provincial Government | 43.662301 | -79.389494 |
| ... | ... | ... | ... | ... | ... |
| 98 | M8X | Etobicoke | The Kingsway, Montgomery Road, Old Mill North | 43.653654 | -79.506944 |
| 99 | M4Y | Downtown Toronto | Church and Wellesley | 43.665860 | -79.383160 |
| 100 | M7Y | East Toronto Business | Enclave of M4L | 43.662744 | -79.321558 |
| 101 | M8Y | Etobicoke | Old Mill South, King's Mill Park, Sunnylea, Hu... | 43.636258 | -79.498509 |
| 102 | M8Z | Etobicoke | Mimico NW, The Queensway West, South of Bloor,... | 43.628841 | -79.520999 |

103 rows × 5 columns

# Foursquare API Data:

Foursquare API is a platform which provides information about nearby places in neighborhoods, using developers account here we are exploring venues in surroundings and help us in finding data about different venues in different neighborhoods of that specific borough. In order to gain that information, we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus, ratings, reviews and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through their API.

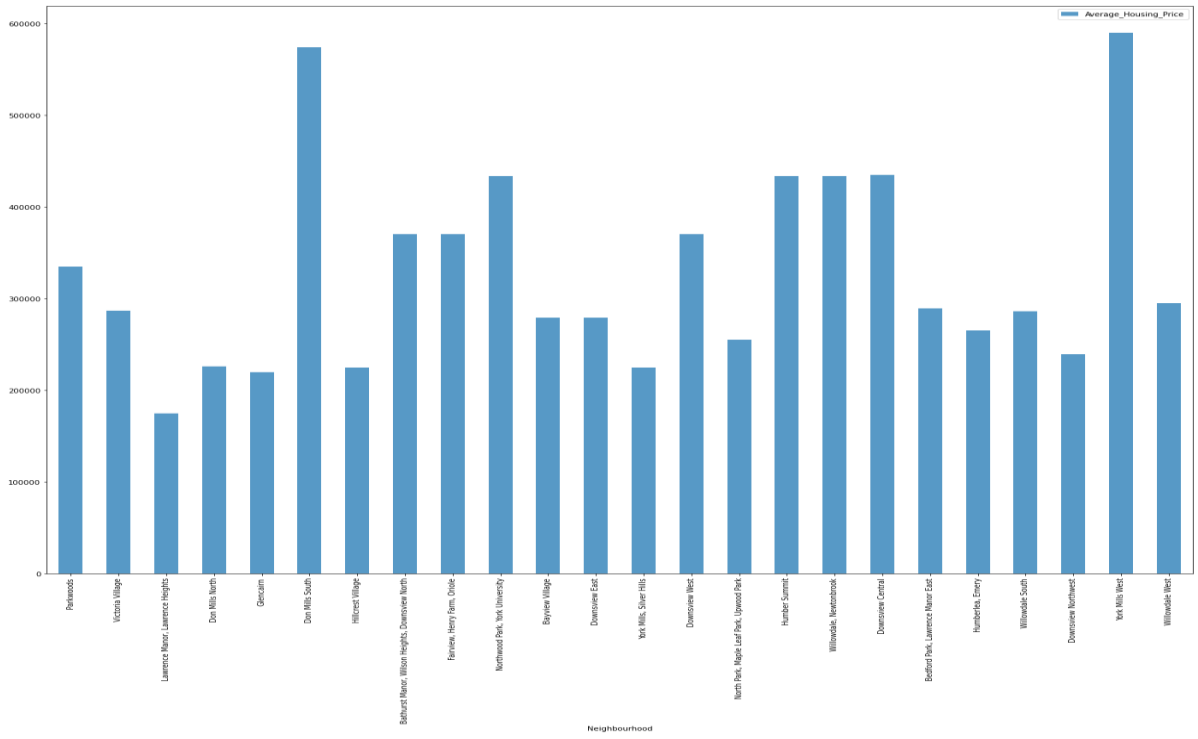
The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. venue
3. Neighborhood Longitude
4. Neighborhood Latitude
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

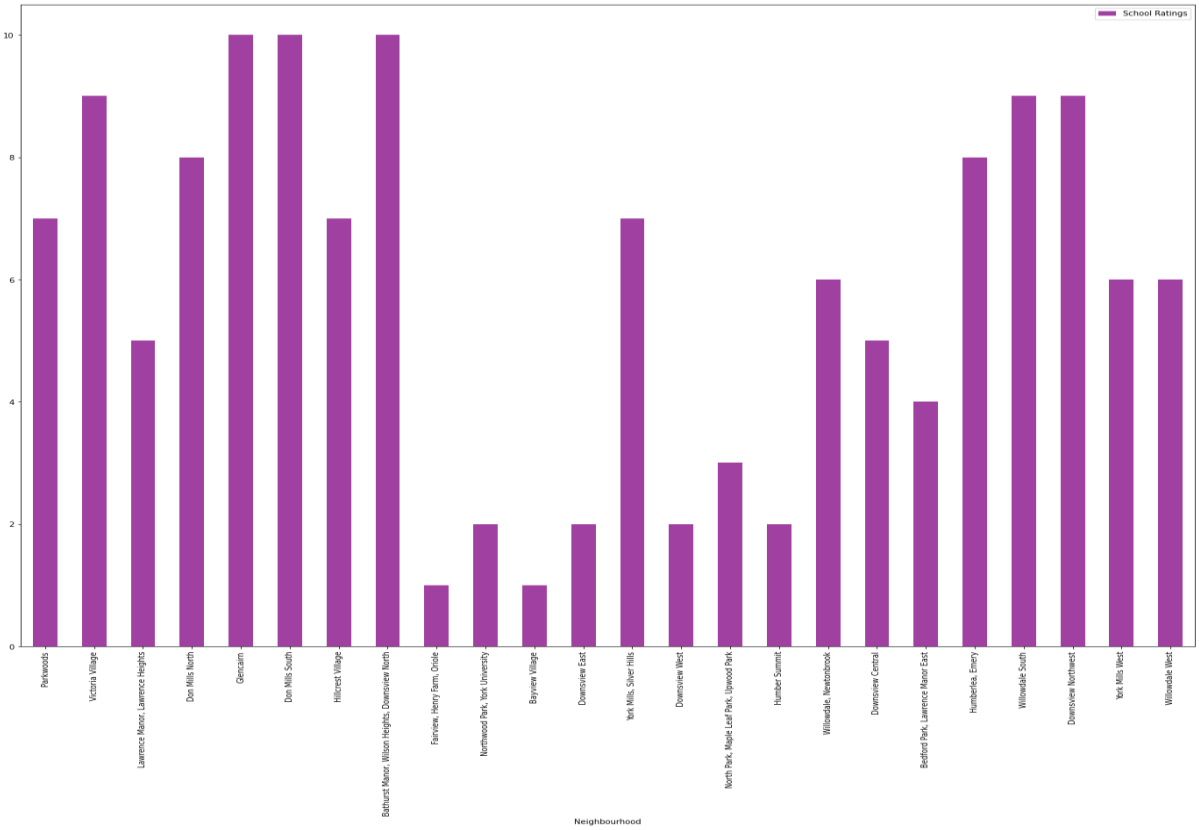
# DISCUSSION:

In this project we are exploring the neighbourhood's in North York Toronto in order to help these newcomers in selecting proper neighbourhood according to their needs and one which is similar to their native place. The features include are housing price and better school according to ratings, crime rates of that particular area, transportation, emergency services, water resources, sewers and recreational facilities.

# Avg Housing price in North York Neighbors:



# Top School Ratings in North York neighbors:




# CONCLUSION:

Conclusion: In this project, using k-means cluster algorithm I separated the neighbourhood into 10(Ten) different clusters and for 103 different latitude and longitude from dataset, which have very-similar neighbourhood's around them. Using the charts above results presented to a particular neighbourhood based on average house prices and school rating have been made.