# Applied Data Science Capstone Project Final Report (IBM-COURSERA)

By:

SHADI FARAHANI

# Business Problem

Data analyzing is one of the most important science that could help investors to know better areas to invest.

Play area and playgrounds are really important for family who has kid to keep them busy and entertained. Also if those areas are close to hotel and restaurants, that would be nicer for tourism who come to the city.

This project will review geographic data in Toronto city to find best neighborhoods to run a business.

# Data Description

- All data which is used in this project are listed bellow:

- ✓ List of Postal Codes of Canada

  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- ✓ Geographical Data which includes latitude and longitude of each postal code.

  https://cocl.us/Geospatial_data
- ✓ Foursquare API

  www.Foursquare.com

# Methodology

- To extract data  from a web page, it is used BeautifulSoup package as technique.
- Merging two data set tables make it a table like bellow:

| | Postal Code | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M5A | Downtown Toronto | Regent Park / Harbourfront | 43.654260 | -79.360636 |
| 1 | M7A | Downtown Toronto | Queen's Park / Ontario Provincial Government | 43.662301 | -79.389494 |
| 2 | M5B | Downtown Toronto | Garden District / Ryerson | 43.657162 | -79.378937 |
| 3 | M5C | Downtown Toronto | St. James Town | 43.651494 | -79.375418 |
| 4 | M4E | East Toronto | The Beaches | 43.676357 | -79.293031 |

- It is used Borough which has Toronto word in.
- Boroughs are Downtown Toronto, East Toronto, West Toronto and Central Toronto

# Map visualization

- Map rendering library named folium is used to create Toronto city map by using longitude and latitude of Toronto Boroughs.

# Foursquare API

- Three steps to use Foursquare data

1. Create an account by developers
2. Receiving Client_ID
3. Receiving Client_Secret
4. Sending some requests

In this project it is used 100 top venues which are in radius 1000.  After any call Foursquare will return a venue data in JSON format.

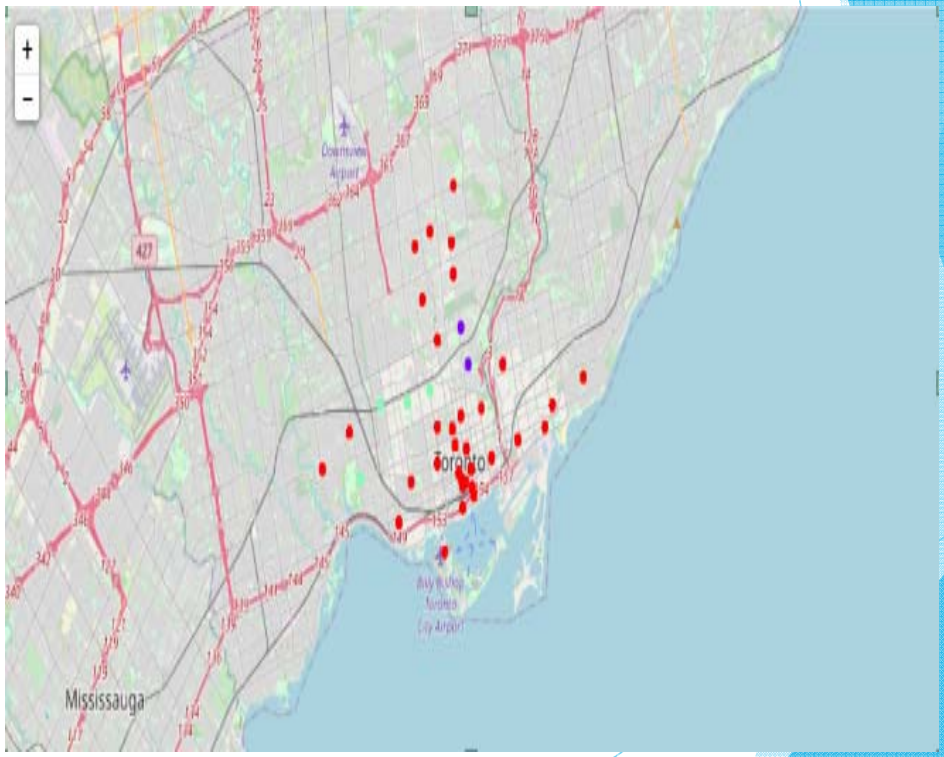
# K-means Clustering

- K-means Clustering Algorithm is an unsupervised Machine Learning Algorithm. The goal of K-Means is to partition the N samples from your data set in to K clusters where each data point belongs to the single cluster for which it is nearest to.

- In this project, it is used number 3 for K. Then number of playground are counted in each cluster. This will help to know which neighborhoods have more opportunity to open a new play area for kids.

# Result

- Cluster 0: most playground.
- Cluster 1: very low of playground
- Cluster 2: moderate concentration of playground



# Result

- As numbers show most of playground are located in Down Town and West Toronto city . And central  and East Toronto  has less playgroung in order.

- So we could say if someone wants to open a new entertainment for kids with more benefits, East Toronto would be the best choice and after that Central Toronto would be next option.

- Also cause most of the playgrounds are located in Down Town which is close to restaurants and hotels would be a good option for tourist to visit.

# Conclusion

- Solving a business problem (finding best neighborhood to open a playarea for kids)

- Extracting data from websites
- Cleaning the data and merging different table of data
- Applying K-means Clustering Algorithm
- East Toronto is the best choice to open a new business.
- West and Down Town of Toronto City are the best choice for tourism.

# References


1. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2. https://cocl.us/Geospatial_data

3. www.Foursquare.com

4. https://en.wikipedia.org/wiki/K-means_clustering