

# IDEAL NEIGHBORHOODS FOR PARKS IN YORK, CANADA

By Rithwik

# Introduction /Business Problem

## **Objectives :**

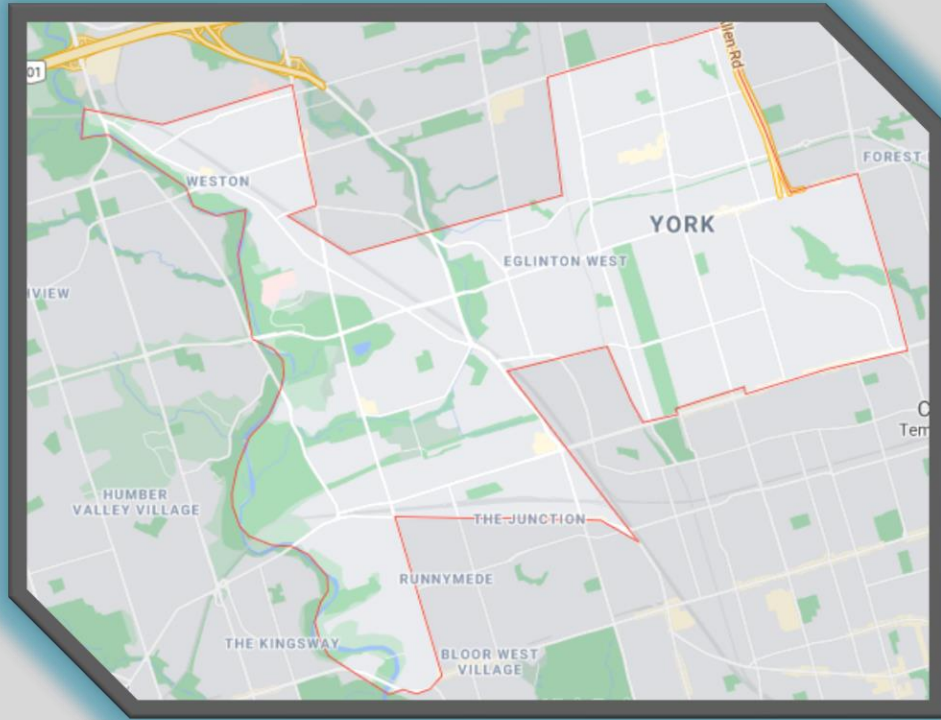
- To find the neighborhoods that **do not have** parks in York, Canada .
- To find the neighborhoods that **already have** parks in York, Canada.

# Purpose

## **This will be beneficial to:**

- Those planning to construct a new park where they don't already exist
- People seeking residence / Office space in York and prefer to have a park in their neighborhood.
- People seeking parks as a mode of recreation
- Even people who are trying to avoid parks due to health reasons ,professional reasons or just plain personal choices

# Location- York, Canada



- Located northwest of Old Toronto, southwest of North York and east of Etobicoke, where it is bounded by the Humber River.
- It was one of six municipalities that amalgamated in 1998 to form the current city of Toronto.
- The original York Township was formed in 1793, and it was once called Dublin.

Source : Google Maps

# Data Collection

- **Postal Codes of Canada-** The data which consists of the Postal codes of Canada starting with the letter "m" is taken from the Wikipedia page :  
[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- **Geospatial data :** The Geospatial data is available in the form of a CSV file in [https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data). It contains the Latitude and longitude data against the postal codes.
- **Foursquare API :** Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. Additionally, Foursquare allows developers to build audience segments for analysis and measurement.

# Machine Learning Algorithms Used

- **ONE HOT ENCODING** : integer encoded variable is removed and a new binary variable is added for each unique integer value. we use one-hot encoding of the data using the `get_dummies()` method. Then we group the new one-hot encoded dataset by neighborhood. This is done to find out the common venues in each neighborhood by their means. This helps to find out which neighborhoods have pre-built parks (mean > 0) and which neighborhoods do not.
- **K- Means Clustering**: The objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (k) of clusters in a dataset.” A cluster refers to a collection of data points aggregated together because of certain similarities. Here ,we use k -means to cluster the neighborhoods with and without parks.

# Libraries Used

- **Pandas , NumPy**: for data manipulation and efficient handling
- **Matplotlib, Seaborn**: Data visualization and plotting libraries
- **Folium: Python** library used for visualizing geospatial data.
- **Geocoder**: Python library to convert an address into latitude and longitude values
- **JSON**: Library to handle and parse JSON files
- **requests**: Library to handle requests

# Procedure

- Load Libraries
- Introduce Data Source
- Read data source into panda data frame

Data successfully read into Pandas dataframe. Here are the first five rows:

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront



# Procedure (contd..)

- Clean data by removing not assigned values

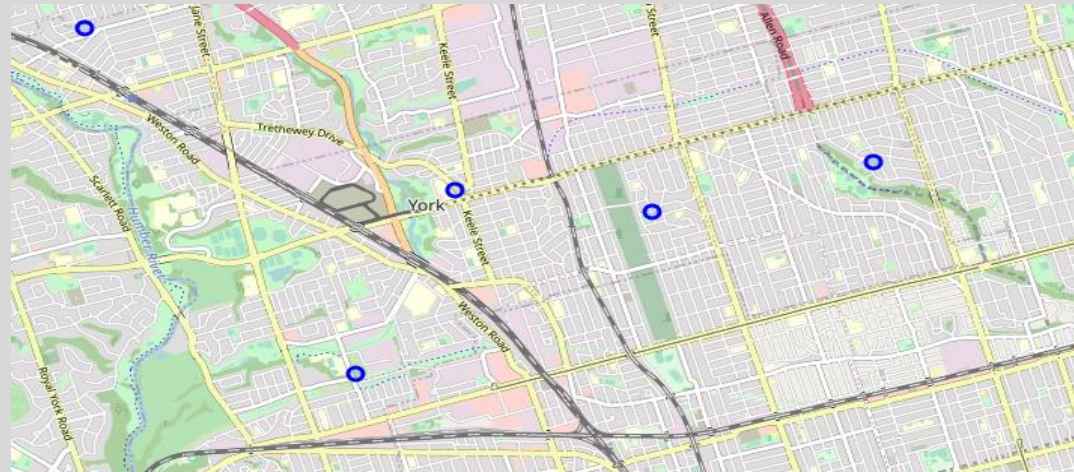
	Postal Code	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

- Read Geospatial data from CSV file and merge with Postal code data from earlier

	Postal Code	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	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
...	...	...	...	...	...

# Procedure (contd..)

- Cluster the neighborhoods and visualize on a map
- Introduce Foursquare API to gather venues.



	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Humewood-Cedarvale	43.693781	-79.428191	Cedarvale Park	43.692535	-79.428705	Field
1	Humewood-Cedarvale	43.693781	-79.428191	Cedarvale Ravine	43.690188	-79.426106	Trail
2	Humewood-Cedarvale	43.693781	-79.428191	Glen Cedar Park	43.695399	-79.429253	Playground
3	Humewood-Cedarvale	43.693781	-79.428191	Phil White Arena	43.691303	-79.431761	Hockey Arena
4	Caledonia-Fairbanks	43.689026	-79.453512	Nairn Park	43.690654	-79.456300	Park

# Procedure (contd..)

- Analyze each neighborhood using one hot encoding

	Neighborhood	Bar	Breakfast Spot	Brewery	Bus Line	Convenience Store	Discount Store	Field
0	Humewood-Cedarvale	0	0	0	0	0	0	1
1	Humewood-Cedarvale	0	0	0	0	0	0	0
2	Humewood-Cedarvale	0	0	0	0	0	0	0
3	Humewood-Cedarvale	0	0	0	0	0	0	0
4	Caledonia-Fairbanks	0	0	0	0	0	0	0

- Group rows by neighborhood and by taking the mean of the frequency of occurrence of each category.

	Neighborhood	Bar	Breakfast Spot	Brewery	Bus Line	Convenience Store	Discount Store	Field	Hockey Arena	Park	Pl
0	Caledonia-Fairbanks	0.25	0.00	0.00	0.00	0.00	0.0	0.00	0.00	0.5	
1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	0.20	0.00	0.00	0.00	0.20	0.2	0.00	0.00	0.0	
2	Humewood-Cedarvale	0.00	0.00	0.00	0.00	0.00	0.0	0.25	0.25	0.0	
3	Runnymede, The Junction North	0.00	0.25	0.25	0.25	0.25	0.0	0.00	0.00	0.0	
4	Weston	0.00	0.00	0.00	0.00	0.00	0.0	0.00	0.00	1.0	

# Procedure (contd..)

➤ When the park mean value equals 0

	Neighborhood	Park
1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	0.0
2	Humewood-Cedarvale	0.0
3	Runnymede, The Junction North	0.0

➤ When the park mean value is greater than 0.1

	Neighborhood	Park
0	Caledonia-Fairbanks	0.5
4	Weston	1.0

# Procedure (contd..)

➤ When the park mean value equals 0 (means no park)

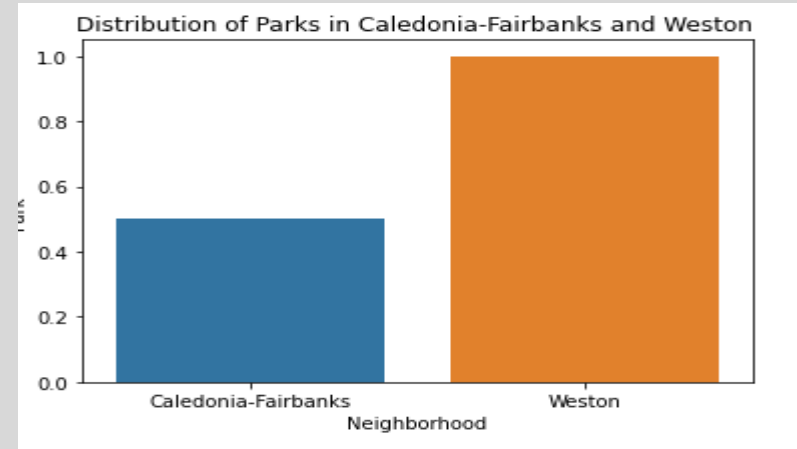
	Neighborhood	Park
1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	0.0
2	Humewood-Cedarvale	0.0
3	Runnymede, The Junction North	0.0

➤ When the park mean value is greater than 0.1 (means there are parks)

	Neighborhood	Park
0	Caledonia-Fairbanks	0.5
4	Weston	1.0

# Procedure (contd..)

- Plot a graph to visualize the distribution of parks in the neighborhoods
- Cluster the neighborhoods without a park on the map

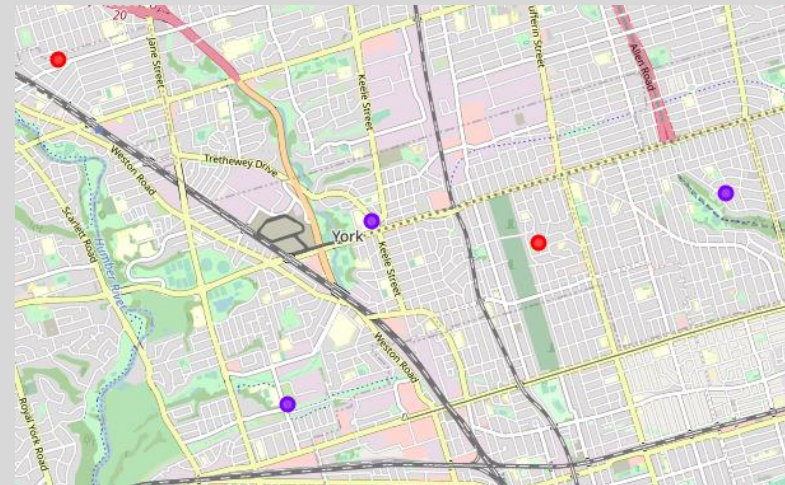


# Procedure (contd..)

➤ Apply K means cluster algorithm

	Neighborhood	Latitude	Longitude	Cluster_Labels
0	Caledonia-Fairbanks	43.689026	-79.453512	0
1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	43.691116	-79.476013	1
2	Humewood-Cedarvale	43.693781	-79.428191	1
3	Runnymede, The Junction North	43.673185	-79.487262	1
4	Weston	43.706876	-79.518188	0

➤ Visualize the clusters on the map



# Observations

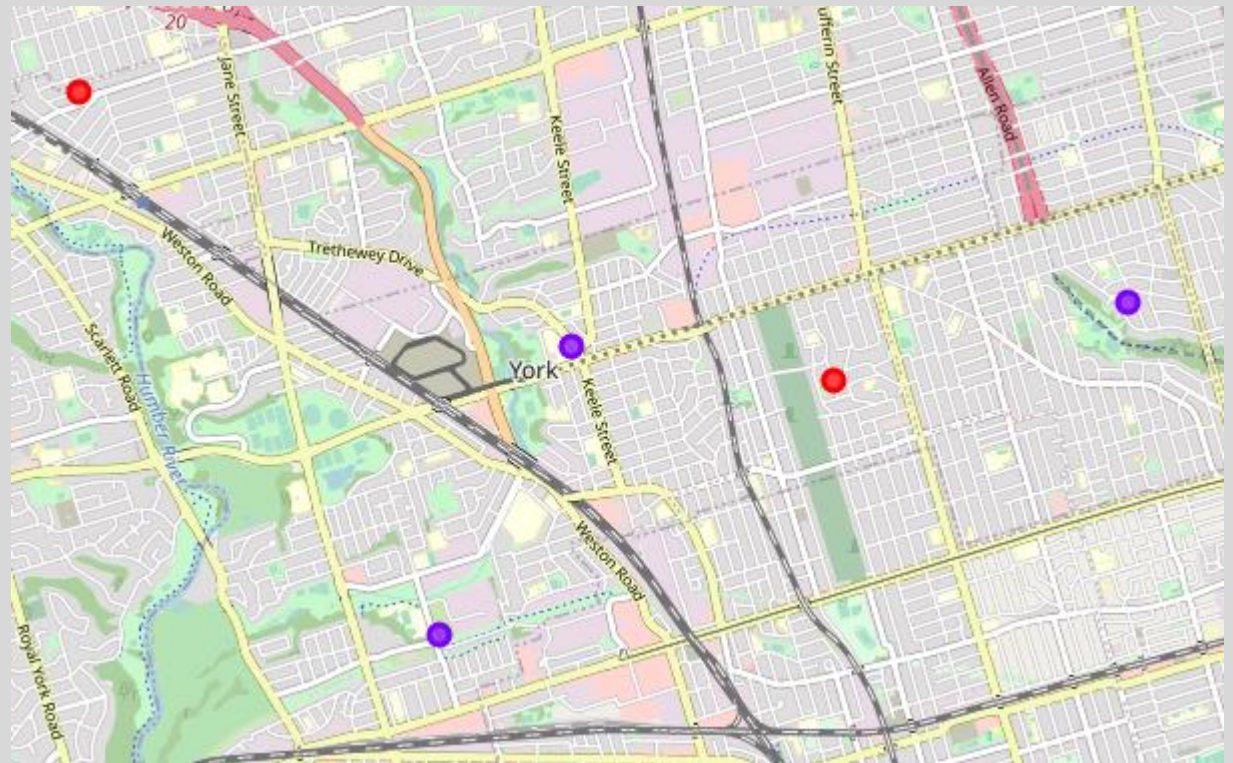
- Using  $k=2$ , we see 2 clusters
- Cluster 0 represents those neighborhoods with parks prebuilt
- Cluster 1 represents those neighborhoods without park

	Neighborhood	Latitude	Longitude	Cluster_Labels
0	Caledonia-Fairbanks	43.689026	-79.453512	0
1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	43.691116	-79.476013	1
2	Humewood-Cedarvale	43.693781	-79.428191	1
3	Runnymede, The Junction North	43.673185	-79.487262	1
4	Weston	43.706876	-79.518188	0



# Observations (contd..)

- Cluster 0 is marked in red clusters
- Cluster 1 is marked in violet clusters



# Conclusion

**Cluster 0** - Contains - **Caledonia-Fairbanks and Weston** - These are the neighborhoods that **have parks**

**Cluster 1** - Contains - **Del Ray, Mount Dennis, Keelsdale and Silverthorn, Humewood-Cedarvale , Runnymede, The Junction North** - These are the neighborhoods that **do NOT have parks**

# Conclusion(contd..)

- If a person seeks to move to a neighborhood with a park then they should consider cluster 0 neighborhoods
- If a person does not seek to move to a neighborhood with a park then they should consider cluster 1 neighborhoods

*FIN.*