

CAPSTONE PROJECT

THE BATTLE OF NEIGHBOURHOODS

**FINDING A BETTER PLACE IN NORTH YORK,
TORONTO**

CONTENTS

- **INTRODUCTION**
- **DATA SELECTION**
- **METHODOLOGY**
- **RESULTS**
- **DISCUSSION**
- **CONCLUSION**

INTRODUCTION

- Purpose is to help people in exploring better facilities around their neighbourhood
- Helps people making smart and efficient decision on selecting great neighbourhood out of numbers of other neighbourhoods in North York, Toronto
- This project is for those people who are looking for better neighbourhoods
- For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

DATA SELECTION

DATA LINK:

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

THE LOCATION:

- North York is a popular destination for new immigrants in Canada to reside
- One of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship

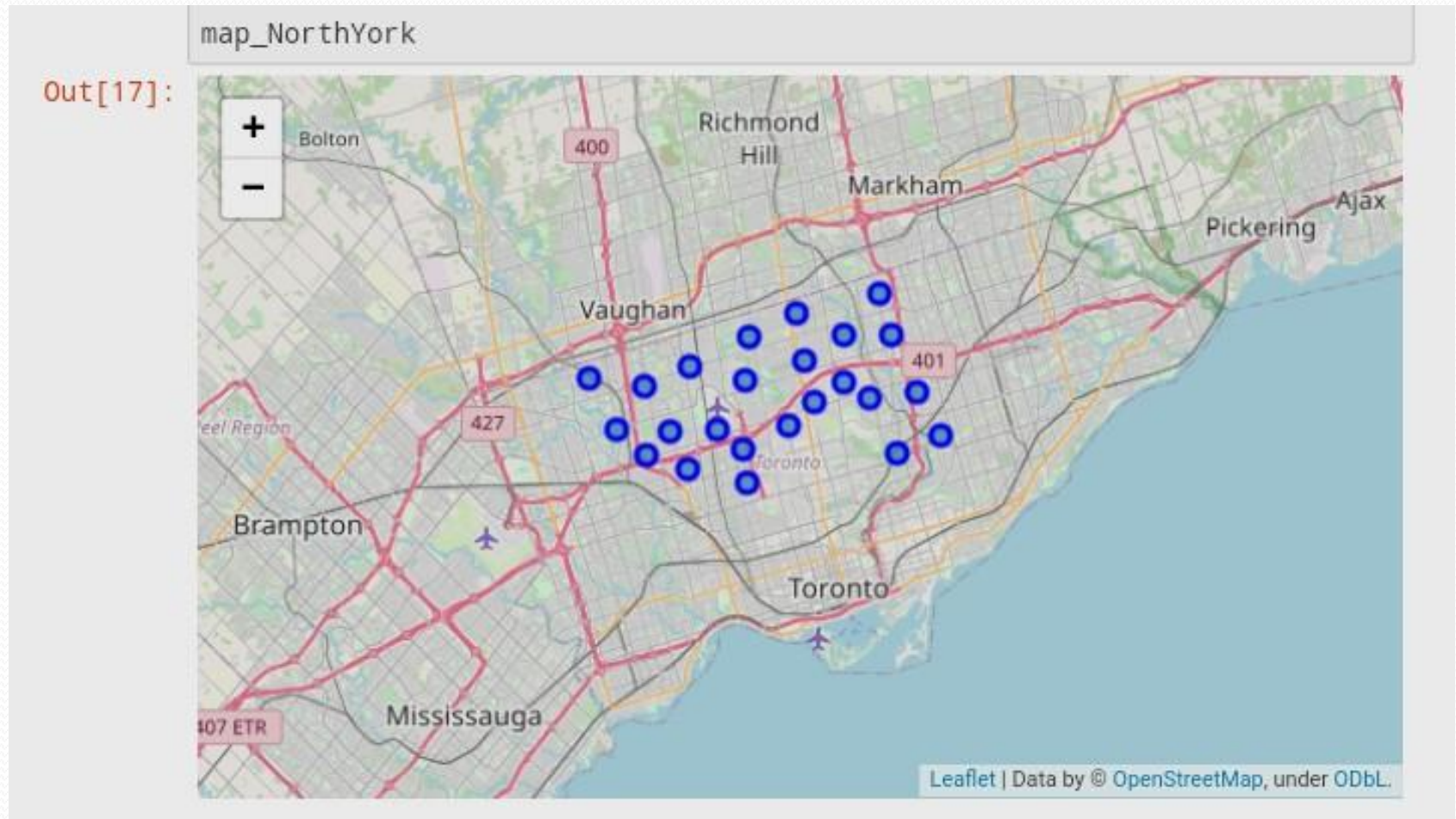
FOURSQUARE API DATA:

- Location data provider with information about all manner of venues and events within an area of interest

Data retrieved are :

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

MAP OF NORTH YORK:



METHODOLOGY

CLUSTERING APPROACH:

- To compare the similarities of two cities, we decided to explore neighbourhoods, segment them, and group them into clusters to find similar neighbourhoods in a big city like New York and Toronto
- To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

USING K-MEANS CLUSTERING APPROACH:

```
File Edit View Insert Cell Kernel Help Not Trusted | Python 3.7

Format
+ - * / < > Run Code

8. K-Means Clustering Approach

In [33]: # Using K-Means to cluster neighborhood into 3 clusters
NorthYork_grouped_clustering = NorthYork_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=3, random_state=0).fit(NorthYork_grouped_clustering)
kmeans.labels_

Out[33]: array([2, 2, 2, 2, 2, 2, 2, 1, 0, 2, 2, 0, 2, 1, 2, 2, 2, 2, 1],
              dtype=int32)

In [34]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
NorthYork_merged = df_2.iloc[:,16:]

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
NorthYork_merged = NorthYork_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'))

NorthYork_merged.head() # check the last columns!

Out[34]:
```

uster abels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Park	Bakery	Chinese Restaurant	Fast Food Restaurant	Residential Building (Apartment / Condo)	Pharmacy	Department Store	Deli / Bodega	Food & Drink Shop	Dessert Shop
2	Clothing Store	Coffee Shop	Fast Food Restaurant	Convenience Store	Restaurant	Women's Store	Toy / Game Store	Electronics Store	Baseball Field	Juice Bar
2	Park	Asian Restaurant	Trail	Gas Station	Yoga Studio	Discount Store	Food & Drink Shop	Fast Food Restaurant	Falafel Restaurant	Electronics Store
1	Concert Hall	Park	Food Truck	Cosmetics Shop	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
2	Korean Restaurant	Middle Eastern Restaurant	Coffee Shop	Café	Pizza Place	Ice Cream Shop	Shopping Mall	Grocery Store	Hookah Bar	Greek Restaurant

MOST COMMON VENUES NEAR NEIGHBORHOOD:

File Edit View Insert Cell Kernel Help Not Trusted | Python 3.7

Format

Code

7. Most Common Venues Near Neighborhood

```
In [32]: import numpy as np
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = NorthYork_grouped['Neighborhood']

for ind in np.arange(NorthYork_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(NorthYork_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[32]:

borhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Bathurst, Wilson Heights, wmsview North	Pizza Place	Park	Coffee Shop	Mobile Phone Shop	Sandwich Place	Restaurant	Fried Chicken Joint	Sushi Restaurant	Middle Eastern Restaurant	Intersection
Bayview Village	Park	Asian Restaurant	Trail	Gas Station	Yoga Studio	Discount Store	Food & Drink Shop	Fast Food Restaurant	Falafel Restaurant	Electronics Store
3rd Park, Lawrence North East	Restaurant	Coffee Shop	Italian Restaurant	Pizza Place	Sandwich Place	Butcher	Pub	Café	Pharmacy	Grocery Store
Don Mills	Coffee Shop	Restaurant	Supermarket	Italian Restaurant	Japanese Restaurant	Sporting Goods Shop	Bank	Gym	Dim Sum Restaurant	Movie Theater
wmsview	Coffee Shop	Vietnamese Restaurant	Grocery Store	Pizza Place	Discount Store	Fast Food Restaurant	Hotel	Park	Beer Store	Pharmacy

WORK FLOW:

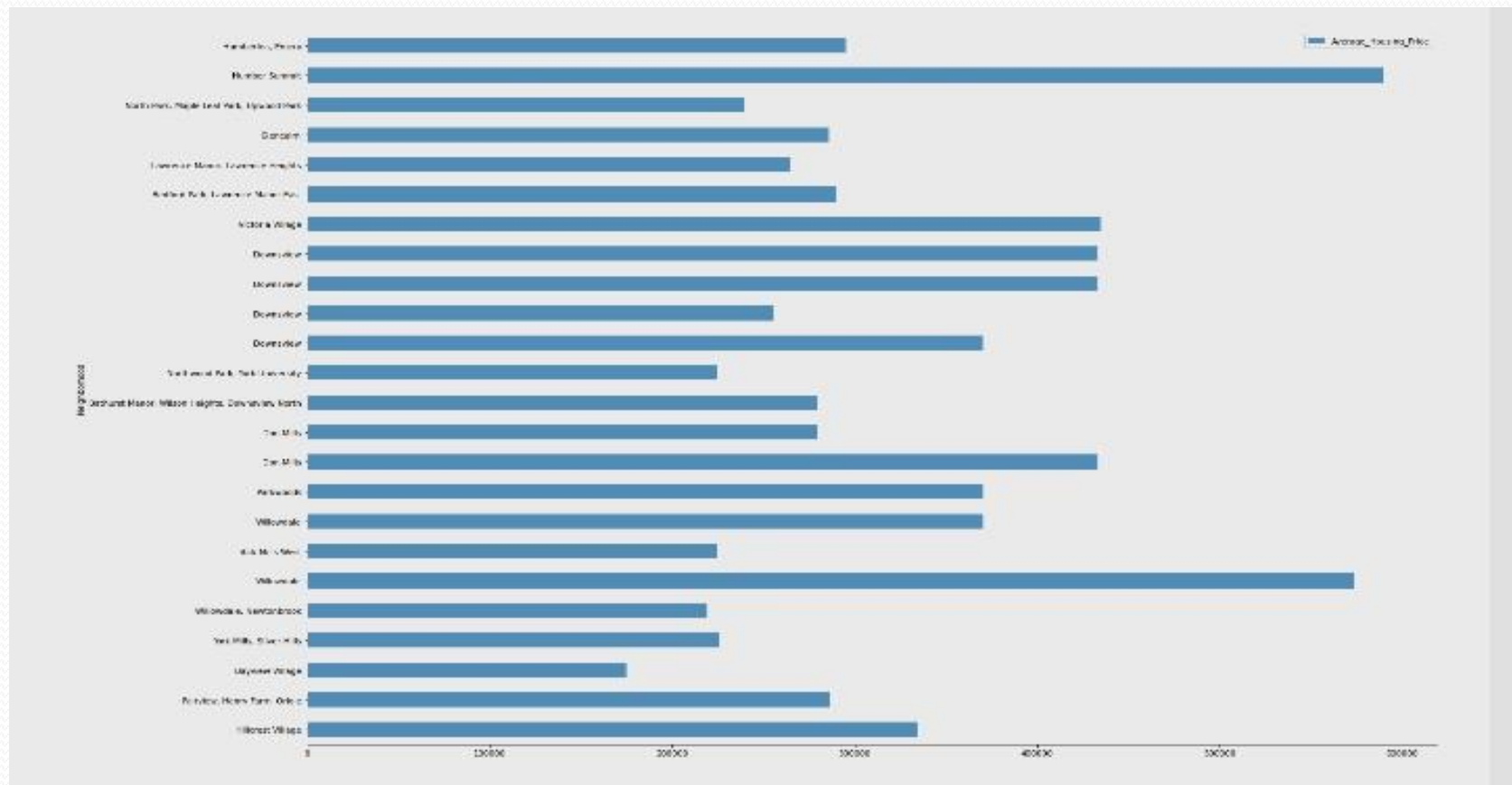
- Using credentials of Foursquare API features of near-by places of the neighbourhoods would be mined
- Due to http request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 500

RESULTS

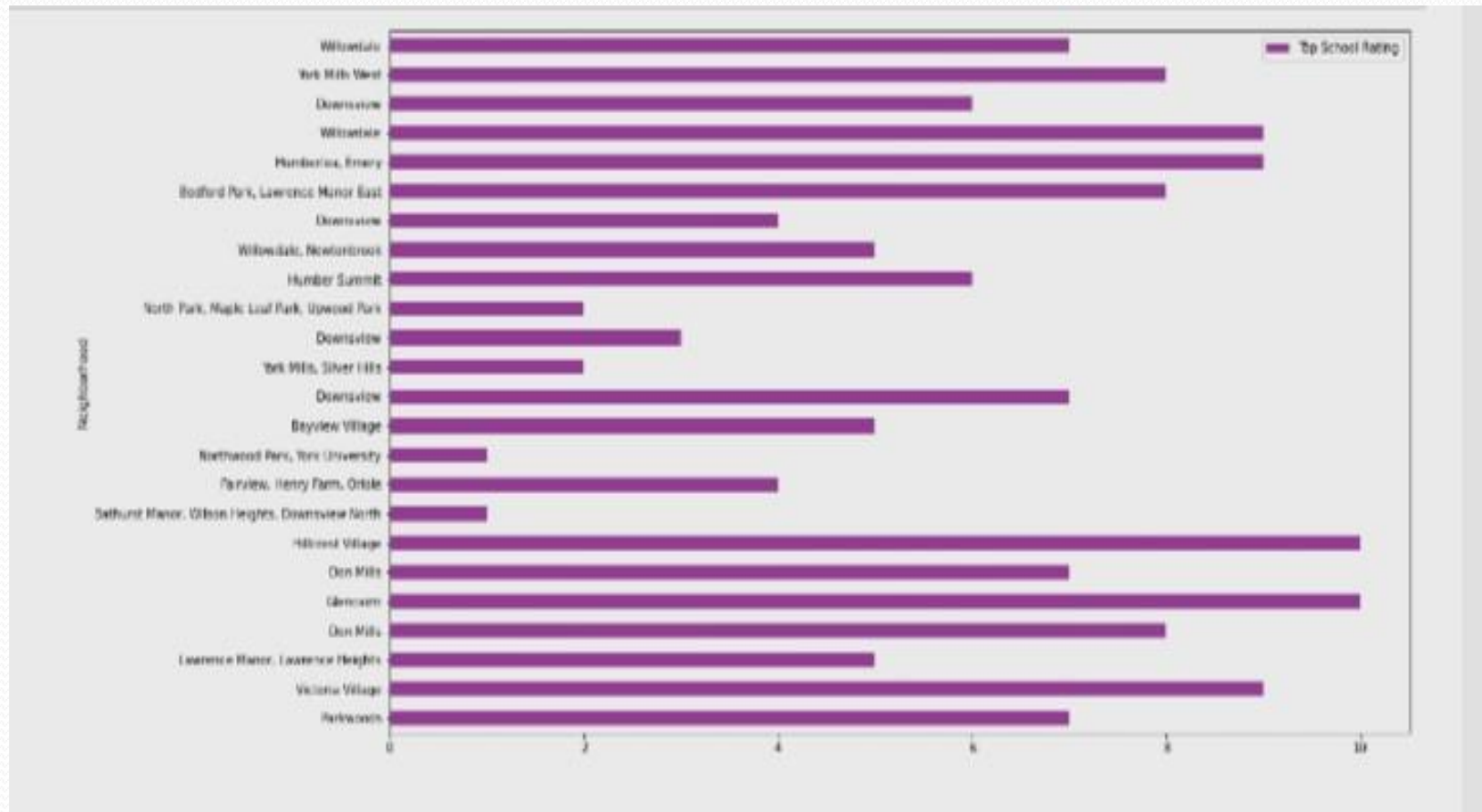
MAP OF CLUSTERS IN NORTH YORK:



AVERAGE HOUSING PRICE BY CLUSTERS IN NORTH YORK:



SCHOOL RATINGS BY CLUSTERS IN NORTH YORK:



DISCUSSION

PROBLEM WHICH TRIED TO SOLVE:

- The major purpose of this project, is to suggest a better neighbourhood in a new city for the person who are shifting there. Social presence in society in terms of like minded people
- Sorted list of house in terms of housing prices in a ascending or descending order
- Sorted list of schools in terms of location, fees, rating and reviews

CONCLUSION

- Using k-means cluster algorithm separated the neighbourhood into 10(Ten) different clusters and for 24 different latitude and longitude from dataset, which have very-similar neighbourhoods around them
- Using the charts above results presented to a particular neighbourhood based on average house prices and school rating have been made
- Mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence



FUTURE WORKS:

This Capstone project can be continued for making it more precise in terms to find best house in North York. Best means on the basis of all required things(daily needs or things we need to live a better life) around and also in terms of cost effective



THANK YOU