

THE BATTLE OF NEIGHBORHOODS

APPLIED DATA SCIENCE CAPSTONE PROJECT

PATRICIO LOZA

INTRODUCTION

- THE PURPOSE OF THIS CAPSTONE PROJECT IS TO HELP PEOPLE IN EXPLORING BETTER FACILITIES AROUND THEIR NEIGHBORHOOD. IT WILL HELP PEOPLE MAKING SMART AND EFFICIENT DECISION ON SELECTING GREAT NEIGHBORHOOD OUT OF NUMBERS OF OTHER NEIGHBORHOODS IN SCARBOROUGH, TORONTO.
- LOTS OF PEOPLE ARE MIGRATING TO VARIOUS STATES OF CANADA AND NEEDED LOTS OF RESEARCH FOR GOOD HOUSING PRICES AND REPUTATED SCHOOLS FOR THEIR CHILDREN. THIS PROJECT IS FOR THOSE PEOPLE WHO ARE LOOKING FOR BETTER NEIGHBORHOODS. FOR EASE OF ACCESSING TO CAFE, SCHOOL, SUPER MARKET, MEDICAL SHOPS, GROCERY SHOPS, MALL, THEATRE, HOSPITAL, LIKE MINDED PEOPLE, ETC.
- THIS CAPSTONE PROJECT AIM TO CREATE AN ANALYSIS OF FEATURES FOR A PEOPLE MIGRATING TO SCARBOROUGH TO SEARCH A BEST NEIGHBORHOOD AS A COMPARATIVE ANALYSIS BETWEEN NEIGHBORHOODS. THE FEATURES INCLUDE MEDIAN HOUSING PRICE AND BETTER SCHOOL ACCORDING TO RATINGS, CRIME RATES OF THAT PARTICULAR AREA, ROAD CONNECTIVITY, WEATHER CONDITIONS, GOOD MANAGEMENT FOR EMERGENCY, WATER RESOURCES BOTH FRESH AND WASTE WATER AND EXCREMENT CONVEYED IN SEWERS AND RECREATIONAL FACILITIES.
- IT WILL HELP PEOPLE TO GET AWARENESS OF THE AREA AND NEIGHBORHOOD BEFORE MOVING TO A NEW CITY, STATE, COUNTRY OR PLACE FOR THEIR WORK OR TO START A NEW FRESH LIFE.

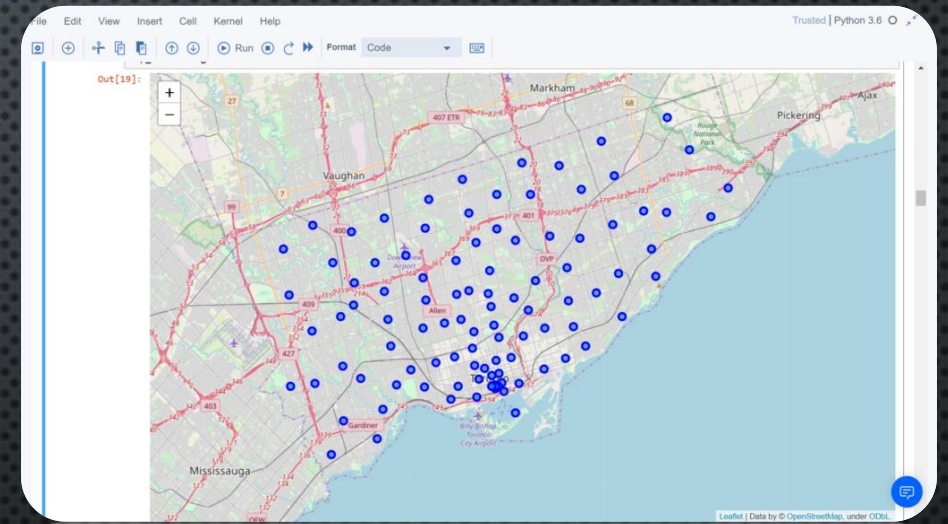
DATA SECTION

- DATA LINK:
[HTTPS://EN.WIKIPEDIA.ORG/WIKI/LIST_OF_POSTAL_CODES_OF_CANADA:_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

WILL USE SCARBOROUGH DATASET WHICH WE SCRAPPED FROM WIKIPEDIA ON WEEK 3. DATASET CONSISTING OF LATITUDE AND LONGITUDE, ZIP CODES.

- FOURSQUARE API DATA:

WE WILL NEED DATA ABOUT DIFFERENT VENUES IN DIFFERENT NEIGHBORHOODS OF THAT SPECIFIC BOROUGH.



3. METHODOLOGY SECTION

- CLUSTERING APPROACH:

TO COMPARE THE SIMILARITIES OF TWO CITIES, WE DECIDED TO EXPLORE NEIGHBORHOODS, SEGMENT THEM, AND GROUP THEM INTO CLUSTERS TO FIND SIMILAR NEIGHBORHOODS 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.

```
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
Scarborough_merged = df_2.iloc[:16,:]
# merge toronto_grouped with toronto_data to add Latitude/Longitude for each neighborhood
Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')
Scarborough_merged.head()# check the last columns!
```

Out[36]:

orough	Neighborhood	Latitude	Longitude	Cluster Labels	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
orough	Rouge, Malvern	43.811525	-79.195517	0	Zoo Exhibit	Financial or Legal Service	Fast Food Restaurant	Construction & Landscaping	Fish & Chips Shop	Filipino Restaurant	Field	Fish Market	Farmers Market	Doner Restaurant
orough	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725	0	Bar	Falafel Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant	Event Space	Yoga Studio
orough	Guildwood, Morningside, West Hill	43.765815	-79.175193	2	Park	Gym / Fitness Center	Pool	Fried Chicken Joint	Indian Restaurant	Athletics & Sports	Ethiopian Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant
orough	Woburn	43.768369	-79.217590	0	Coffee Shop	Fast Food Restaurant	Business Service	Park	Yoga Studio	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
orough	Cedarbrae	43.769688	-79.239440	0	Flower Shop	Athletics & Sports	Thai Restaurant	Bank	Bakery	Caribbean Restaurant	Hakka Restaurant	Indian Restaurant	Eastern European Restaurant	Electronics Store

Map of Clusters

```
In [37]: kclusters = 10
```

3. METHODOLOGY SECTION

- WORK FLOW:

USING CREDENTIALS OF FOURSQUARE API FEATURES OF NEAR-BY PLACES OF THE NEIGHBORHOODS WOULD BE MINED. DUE TO HTTP REQUEST LIMITATIONS THE NUMBER OF PLACES PER NEIGHBORHOOD PARAMETER WOULD REASONABLY BE SET TO 100 AND THE RADIUS PARAMETER WOULD BE SET TO 500.

```
In [34]: 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'] = Scarborough_grouped['Neighborhood']

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

neighborhoods_venues_sorted.head()
```

Out[34]:

	Neighborhood	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
0	Adelaide, King, Richmond	Coffee Shop	Cafe	Hotel	Gastropub	Burger Joint	Asian Restaurant	Bar	Restaurant	American Restaurant	Steakhouse
1	Agincourt	Chinese Restaurant	Shopping Mall	Pizza Place	Supermarket	Sushi Restaurant	Breakfast Spot	Print Shop	Mediterranean Restaurant	Coffee Shop	Pool
2	Agincourt North, L'Amoreaux East, Milliken, St...	Pharmacy	Sandwich Place	Sushi Restaurant	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Ethiopian Restaurant
3	Altton Gardens, Beaumont Heights, Humbergate, ...	Grocery Store	Park	Sandwich Place	Discount Store	Japanese Restaurant	Fried Chicken Joint	Beer Store	Hardware Store	Pizza Place	Fast Food Restaurant
4	Alderwood, Long Branch	Convenience Store	Pub	Sandwich Place	Coffee Shop	Gas Station	Dance Studio	Gym	Pharmacy	Pizza Place	Falafel Restaurant

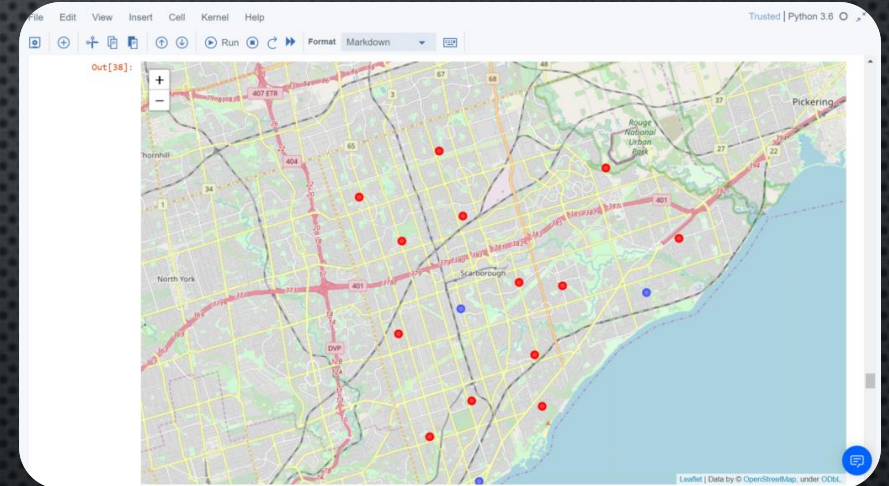
RESULTS SECTION

- THE LOCATION:

SCARBOROUGH IS A POPULAR DESTINATION FOR NEW IMMIGRANTS IN CANADA TO RESIDE. AS A RESULT, IT IS ONE OF THE MOST DIVERSE AND MULTICULTURAL AREAS IN THE GREATER TORONTO AREA, BEING HOME TO VARIOUS RELIGIOUS GROUPS AND PLACES OF WORSHIP. ALTHOUGH IMMIGRATION HAS BECOME A HOT TOPIC OVER THE PAST FEW YEARS WITH MORE GOVERNMENTS SEEKING MORE RESTRICTIONS ON IMMIGRANTS AND REFUGEES, THE GENERAL TREND OF IMMIGRATION INTO CANADA HAS BEEN ONE OF ON THE RISE.

- FOURSQUARE API:

THIS CAPSTONE PROJECT HAVE USED FOUR-SQUARE API AS ITS PRIME DATA GATHERING SOURCE AS IT HAS A DATABASE OF MILLIONS OF PLACES, ESPECIALLY THEIR PLACES API WHICH PROVIDES THE ABILITY TO PERFORM LOCATION SEARCH, LOCATION SHARING AND DETAILS ABOUT A BUSINESS.



DISCUSSION SECTION

- PROBLEM WHICH TRIED TO SOLVE:
- THE MAJOR PURPOSE OF THIS PROJECT, IS TO SUGGEST A BETTER NEIGHBORHOOD IN A NEW CITY FOR THE PERSON WHO ARE SHIFFTING THERE. SOCIAL PRESENCE IN SOCIETY IN TERMS OF LIKE MINDED PEOPLE. CONNECTIVITY TO THE AIRPORT, BUS STAND, CITY CENTER, MARKETS AND OTHER DAILY NEEDS THINGS NEARBY.
- 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

6. CONCLUSION SECTION

- IN THIS CAPSTONE PROJECT, USING K-MEANS CLUSTER ALGORITHM I SEPARATED THE NEIGHBORHOOD INTO 10(TEN) DIFFERENT CLUSTERS AND FOR 103 DIFFERENT LATITUDE AND LOGITUDE FROM DATASET, WHICH HAVE VERY-SIMILAR NEIGHBORHOODS AROUND THEM. USING THE CHARTS ABOVE RESULTS PRESENTED TO A PARTICULAR NEIGHBORHOOD BASED ON AVERAGE HOUSE PRICES AND SCHOOL RATING HAVE BEEN MADE.
- I FEEL REWARDED WITH THE EFFORTS AND BELIEVE THIS COURSE WITH ALL THE TOPICS COVERED IS WELL WORTHY OF APPRECIATION.
THIS PROJECT HAS SHOWN ME A PRACTICAL APPLICATION TO RESOLVE A REAL SITUATION THAT HAS IMPACTING PERSONAL AND FINANCIAL IMPACT USING DATA SCIENCE TOOLS.
THE MAPPING WITH FOLIUM IS A VERY POWERFUL TECHNIQUE TO CONSOLIDATE INFORMATION AND MAKE THE ANALYSIS AND DECISION BETTER WITH CONFIDENCE.