


CAPSTONE PROJECT

The Battle of Neighborhoods

Finding a Better Place in Scarborough, Toronto

► Finding a Better Place

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.

Several thin, white, parallel diagonal lines are positioned in the bottom right corner of the slide, extending from the right edge towards the center.

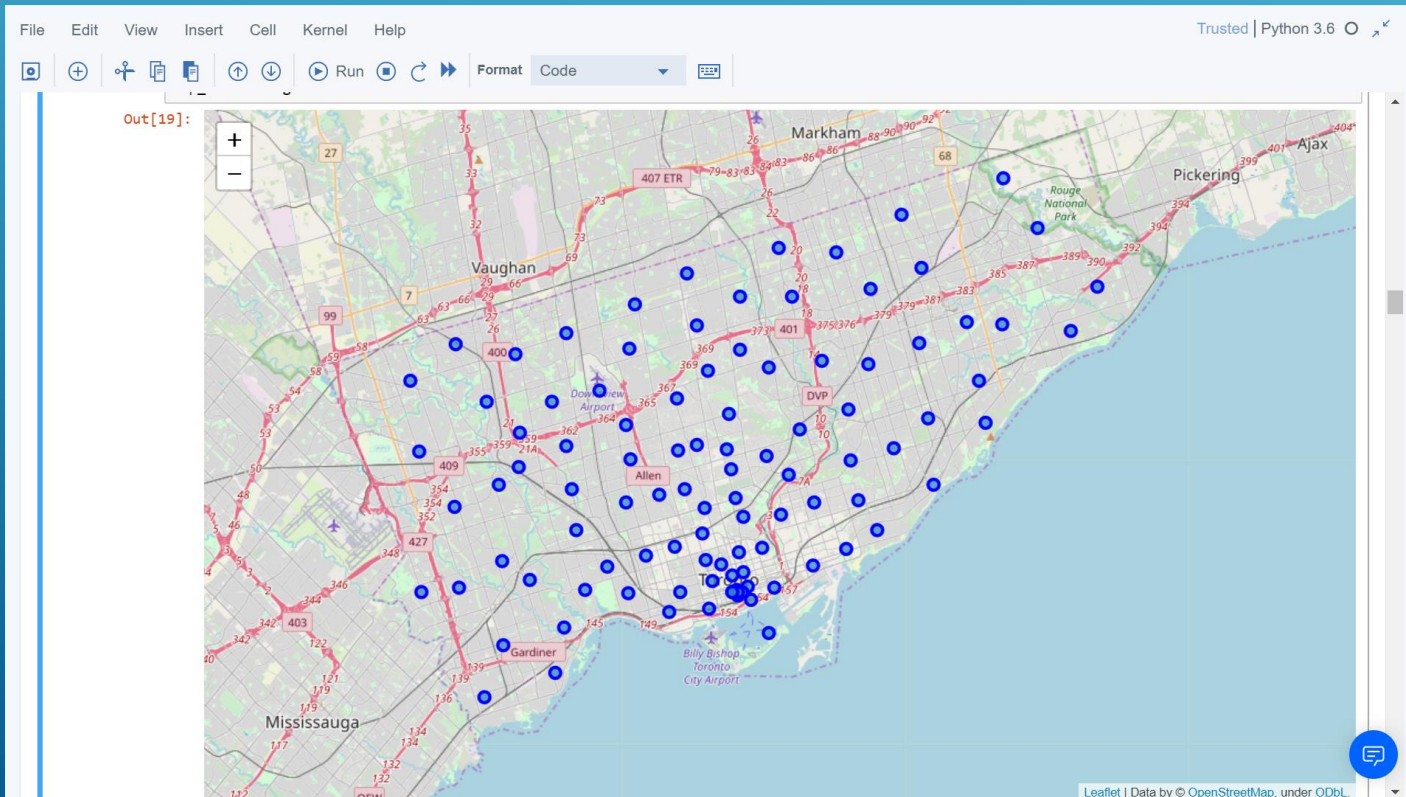
► Data Section

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.

► Data Section

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. Neighborhood Latitude
3. Neighborhood 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

► 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.

Several white lines of varying lengths and angles are positioned on the right side of the slide, creating a modern, abstract graphic element.

► Methodology Section

USING K-MEANS CLUSTERING APPROACH | MOST COMMON VENUE

File Edit View Insert Cell Kernel Help Trusted | Python 3.6

Format Markdown

```
In [36]: 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	Done
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
```

► Methodology Section

MOST COMMON VENUES NEAR NEIGHBORHOOD | USING CLUSTERING

File Edit View Insert Cell Kernel Help Trusted | Python 3.6

+

✂

↶

↷

↺

↻

Format

Markdown

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)

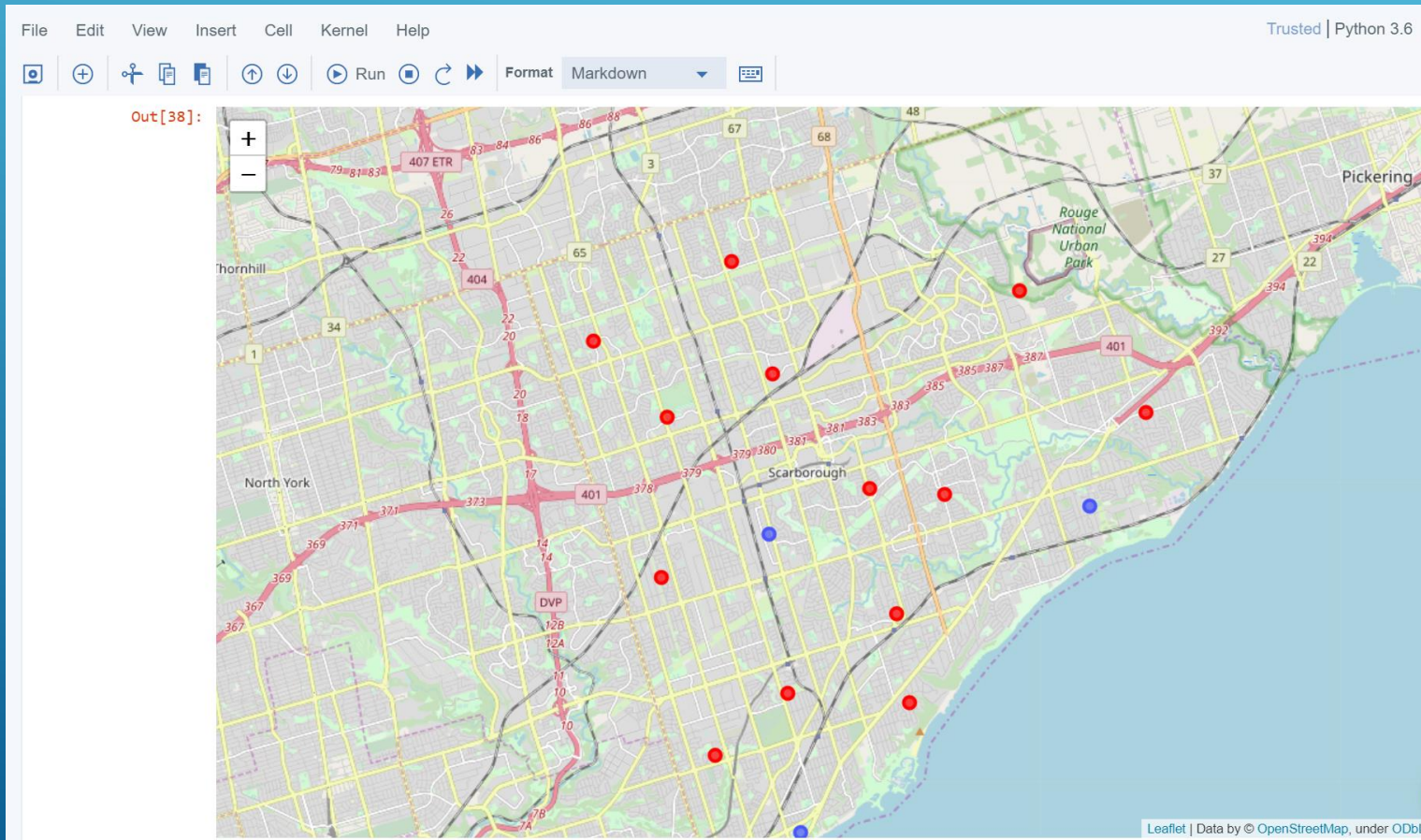
neighborhoods_venues_sorted.head()
```

Out[34]:

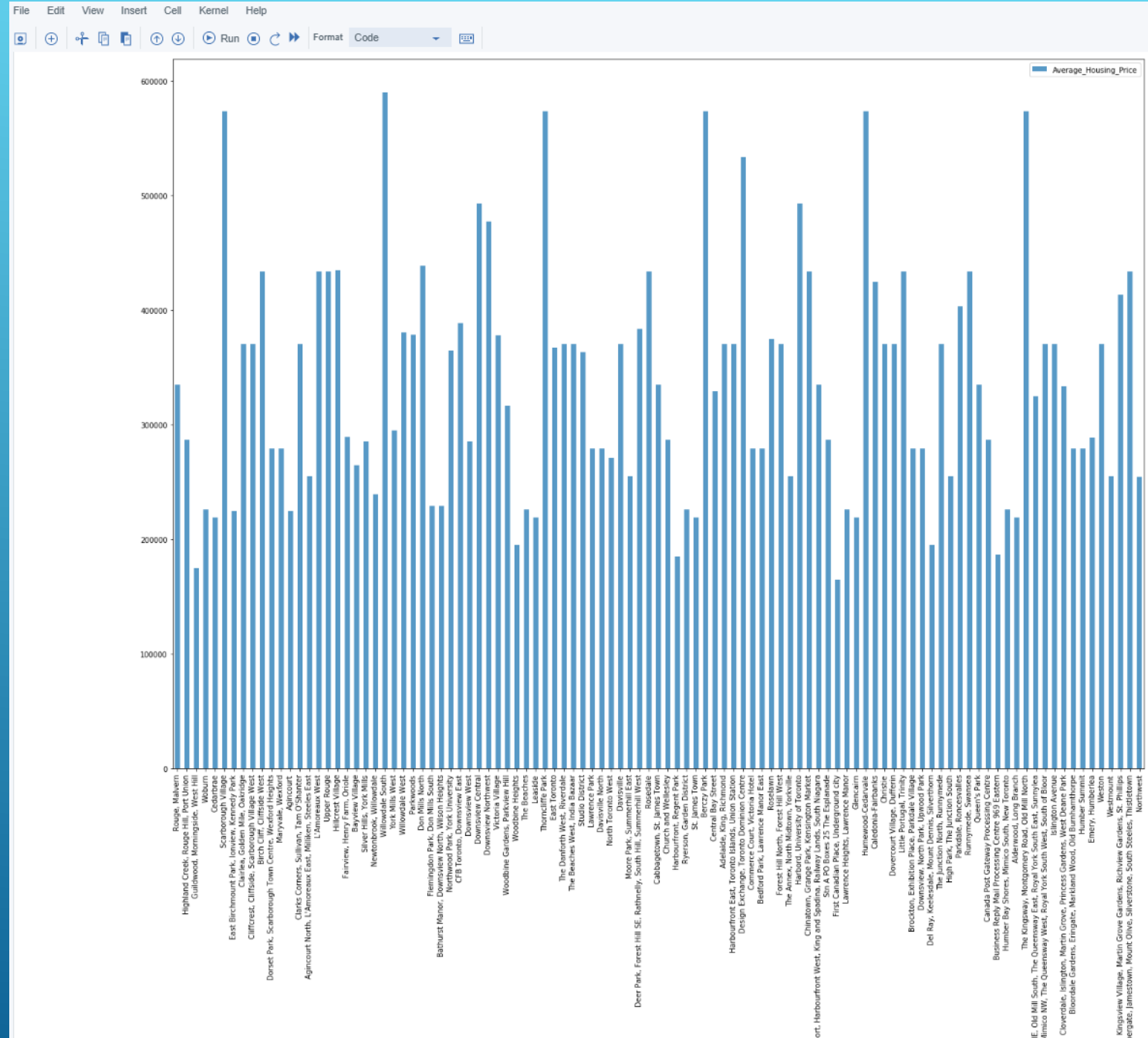
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adelaide, King, Richmond	Coffee Shop	Café	Hotel	Gastropub	Burger 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	Albion 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

MAP OF CLUSTERS IN SCARBOROUGH

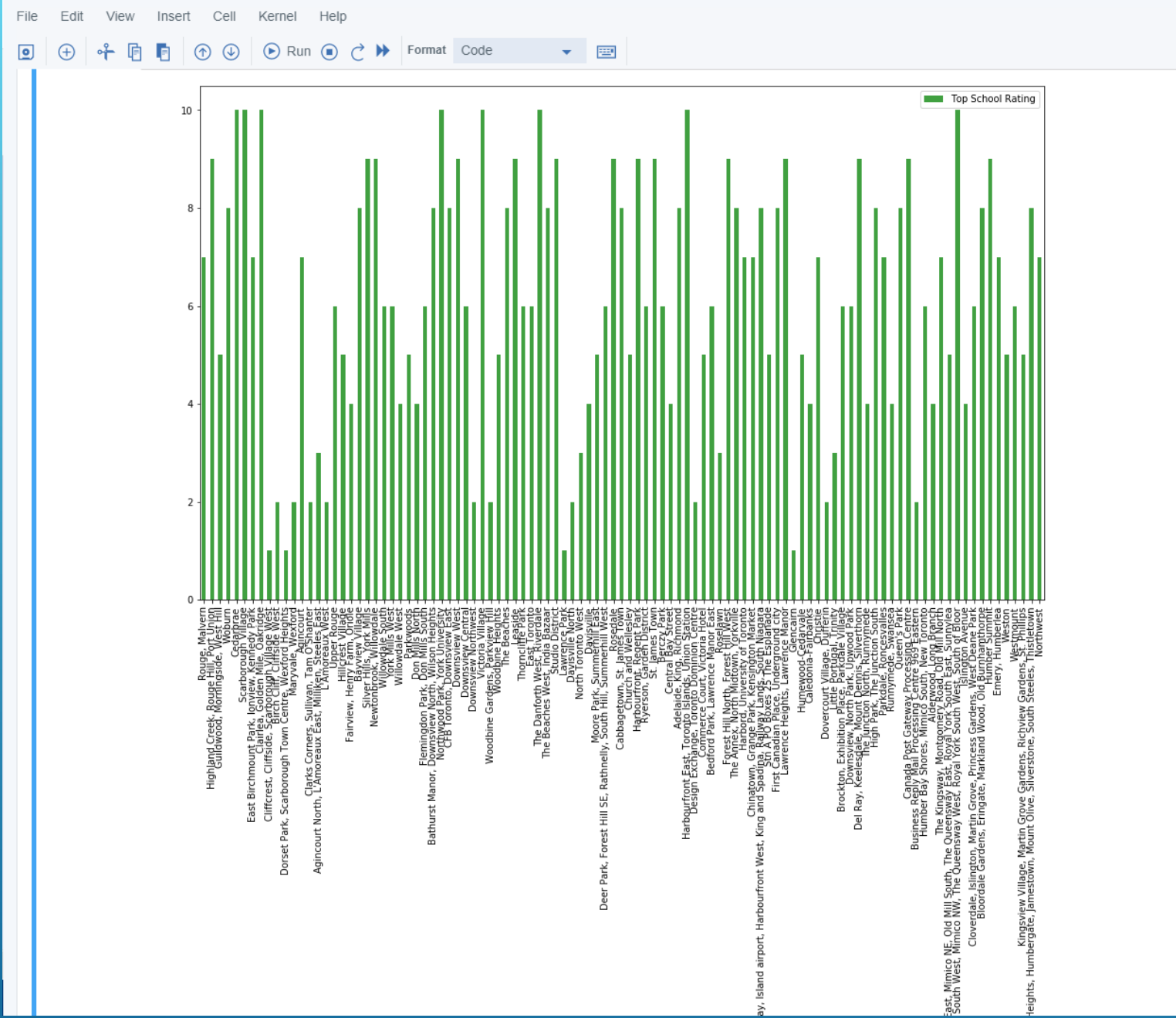


AVERAGE HOUSING PRICE BY CLUSTERS IN SCARBOROUGH



► Results Section


SCHOOL RATINGS BY CLUSTERS IN SCARBOROUGH



► 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.




► Discussion Section

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

► 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.

Several white lines of varying lengths and slopes are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

► Conclusion Section

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.

Several white lines of varying lengths and angles are positioned on the right side of the slide, creating a modern, abstract graphic element.

► Conclusion Section

FUTURE WORKS:

THIS CAPSTONE PROJECT CAN BE CONTINUED FOR MAKING IT MORE PRECISE IN TERMS TO FIND BEST HOUSE IN SCARBOROUGH. 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.

LIBRARIES WHICH ARE USED TO DEVELOPE THE PROJECT:

PANDAS: FOR CREATING AND MANIPULATING DATAFRAMES.

FOLIUM: PYTHON VISUALIZATION LIBRARY WOULD BE USED TO VISUALIZE THE NEIGHBORHOODS CLUSTER DISTRIBUTION OF USING INTERACTIVE LEAFLET MAP.

SCIKIT LEARN: FOR IMPORTING K-MEANS CLUSTERING.

JSON: LIBRARY TO HANDLE JSON FILES.

XML: TO SEPARATE DATA FROM PRESENTATION AND XML STORES DATA IN PLAIN TEXT FORMAT.

GEOCODER: TO RETRIEVE LOCATION DATA.

BEAUTIFUL SOUP AND REQUESTS: TO SCRAP AND LIBRARY TO HANDLE HTTP REQUESTS.

MATPLOTLIB: PYTHON PLOTTING MODULE.