

Analysis of neighborhood

Problem and
solution proposal

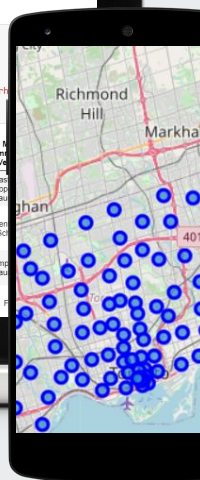
```
In [66]: # using k-means to cluster neighborhood into 3 clusters
Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=3, random_state=0).fit(Scarborough_grouped_clustering)
kmeans.labels_

Out[66]: array([[2, 2, 1, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0,
                2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2,
                2, 2, 2, 1, 2, 2, 0, 2, 2, 1, 1, 1, 2, 2, 2, 2, 2, 2, 1, 2, 2,
                1, 2, 0, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1,
                2, 2, 2, 2, 2, 2, 0, 2, 2, 1])

In [67]: neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
Scarborough_merged = df_2.iloc[:,1:]

# 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[67]:													
	Postalcode	Borough	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
0	M1B	Scarborough	Malvern, Rouge	43.81153	-79.19552	0	Zoo Exhibit	Financial or Legal Service	Business Service	Fast Food Restaurant	Falafel Restaurant	East European Restaurant	Elementary School
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871	0	Construction & Landscaping	Bar	Farm	Eastern European Restaurant	Electronics Store	Elementary School	
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.79575	-79.17520	1	Park	Athletics & Sports	Gym / Fitness Center	Dog Run	Donut Shop	Dump	
3	M1G	Scarborough	Woburn	43.76820	-79.21761	0	Park	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Filipino Restaurant		



Outline

The Problem

Data Execution and cleaning

Clustering Approach

Results

Libraries Used

Next Steps

The Problem

There are so many migrants in different states of Canada who had to settle themselves for which they had to do a lot of research. The research here comprises of good prices for housing, good schools, best places to buy grocery etc. This project is for those people who are looking for the best places and the best availabilities within an area. Other applications may include shopping malls, super stores, schools, cafe, medical shops, hospital etc. This project helps in creating an enhanced analysis and study of features for the migrants of Scarborough in order to search for the best neighborhood.

The features include the median house price and better schools on the basis of the number of ratings and scores. We will also try to focus on the crime rates of the particular area, roads, weather and availability of resources. It will help people be aware of their area and the neighborhood before actually moving to that city.



Problem statement

Sorted list of house and schools in terms of housing prices and school rates in the neighborhood.

This project is for those people who are looking for the best places and the best availabilities within an area.



Data Execution and cleaning

We have used beautifulsoup for creating parse tree to extract data from HTML.

In our case, the data is extracted from the Wikipedia Page. Link:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Clustering Approach

We segmented and combined the neighborhoods into clusters for comparing the cities.

Using K-Means to cluster neighborhood into 3 clusters

```
Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=3, random_state=0).fit(Scarborough_grouped_clustering)
kmeans.labels_
```

```
[[2, 2, 1, 1, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0,
 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2, 2,
 2, 2, 2, 1, 2, 2, 0, 2, 2, 1, 1, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2,
 1, 2, 0, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1,
 2, 2, 2, 2, 2, 2, 0, 2, 2, 1]]
```

```
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

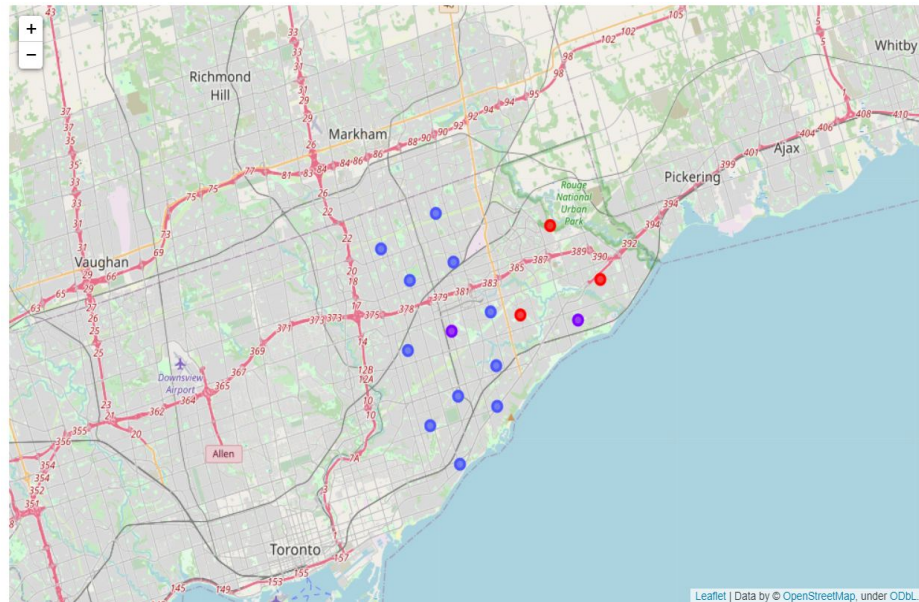
```
Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')
```

```
Scarborough_merged.head()# check the last columns!
```

Postalcode	Borough	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
M1B	Scarborough	Malvern, Rouge	43.81153	-79.19552	0	Zoo Exhibit	Financial or Legal Service	Business Service	Fast Food Restaurant	Falafel Restaurant	Eastern European Restaurant
M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871	0	Construction & Landscaping	Bar	Farm	Eastern European Restaurant	Electronics Store	Elementary School
M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17520	1	Park	Athletics & Sports	Gym / Fitness Center	Dog Run	Donut Shop	Dumpling Restaurant
M1G	Scarborough	Woburn	43.76820	-79.21761	0	Park	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Filipino Restaurant	Field

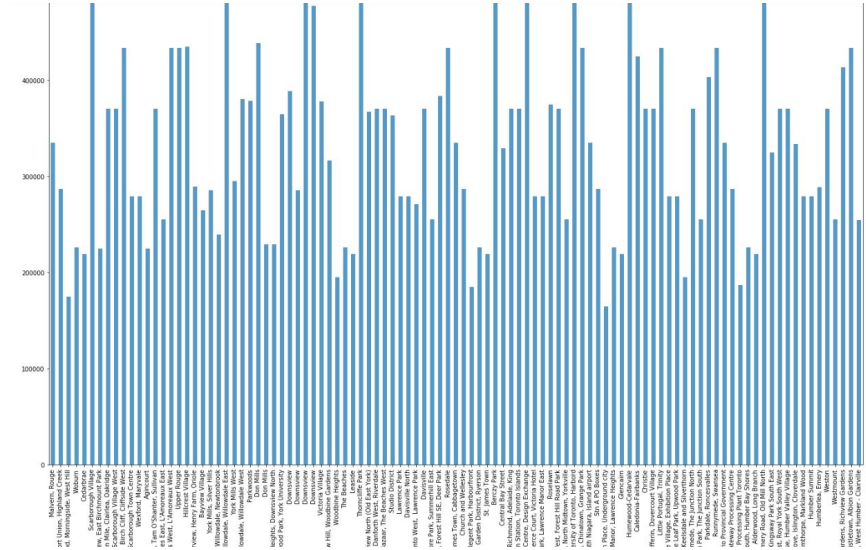
Map of Scarborough

Explain why you're focusing on a particular part of the problem or a particular subset of users.

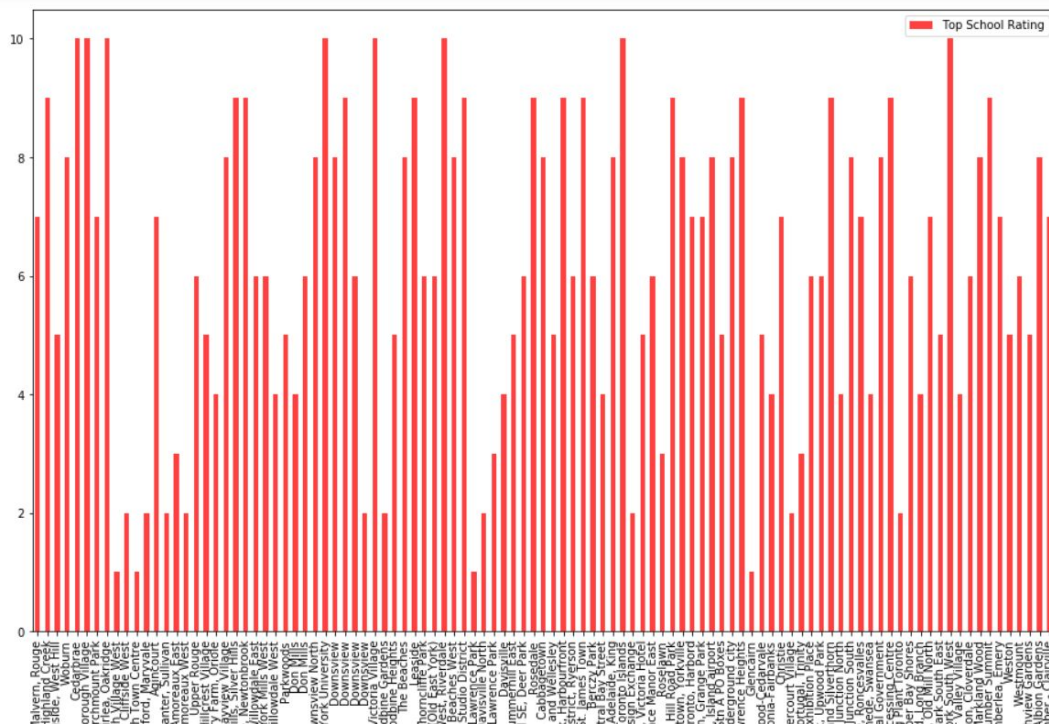


Results

Visualization of schools and housing prices



Avg. housing rates in the area



Libraries used

- Pandas
- Folium
- Scikit Learn
- JSON
- XML
- Geocoder
- Beautiful Soup
- Matplotlib



Solution Proposal

This is just a beginning, as a lot of more features can be integrated to this project. Not just avg. housing prices and school rates, we can also add other broad categories as well, some of them are:

Popular areas for nightlife. Nearest government hospitals. Popular areas for car washing.

and the list goes on.



Conclusion

There are so many migrants in different states of Canada who had to settle themselves for which they had to do a lot of research. The research here comprises of good prices for housing, good schools, best places to buy grocery etc. This project is for those people who are looking for the best places and the best availabilities within an area. Other applications may include shopping malls, super stores, schools, cafe, medical shops, hospital etc. This project helps in creating an enhanced analysis and study of features for the migrants of Scarborough in order to search for the best neighborhood.

Using the K-means algorithm I separated the neighborhood into ten different clusters. Using the charts given, we can clearly conclude the right rating for avg school rating and the house prices.

I feel that completing all of the courses helped me learn a lot of new things which I didn't even had the clue about. This project has helped me to learn about various applications of data science in practical life. Tools such as Folium, numpy, pandas and nominatim are extremely useful to make the task less complex.



Thank You