


# The Battle of Neighborhoods

# Finding a Better Place in Scarborough, Toronto

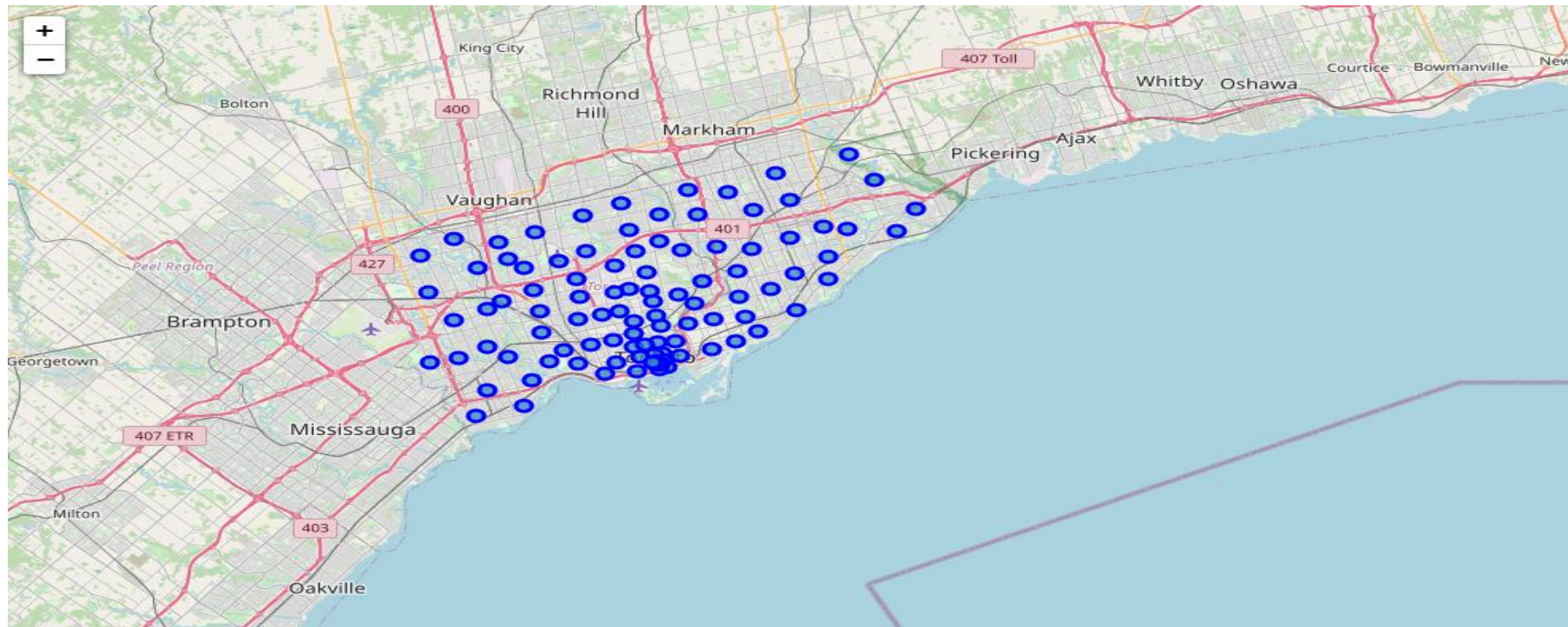
- Project aim is to create an analysis of features for a people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods
  - Features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity etc.
  - Result in Improved and Effective Decision Making.
    - Help people in exploring better facilities around their neighborhood.
    - Help people making smart and efficient decision on selecting great neighborhood.
    - Help people to get awareness of the area and neighborhood.
- 

# Data acquisition and cleaning

- 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)
- Scarborough dataset scrapped from Wikipedia. Dataset consisting of latitude and longitude, zip codes.
- Foursquare API is used to obtain information about venues and events within an area of interest including venue names, locations, menus and even photos
- Duplicate, highly similar or highly correlated features were dropped.
- Cleaned data contains 103 rows and 3 Features.



# Using Foursquare API, Map of Scarborough



# Using K-Means Clustering Approach

## K-Means Clustering Approach

```
# 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_
```

```
array([1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1,  
       1, 2, 1, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1, 2, 1, 1, 1, 1, 1,  
       1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 2, 2, 1, 1, 1, 1, 1, 2, 1, 1,  
       2, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2,  
       1, 1, 1, 1, 1, 1, 2, 1, 1, 2], dtype=int32)
```

```
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!
```

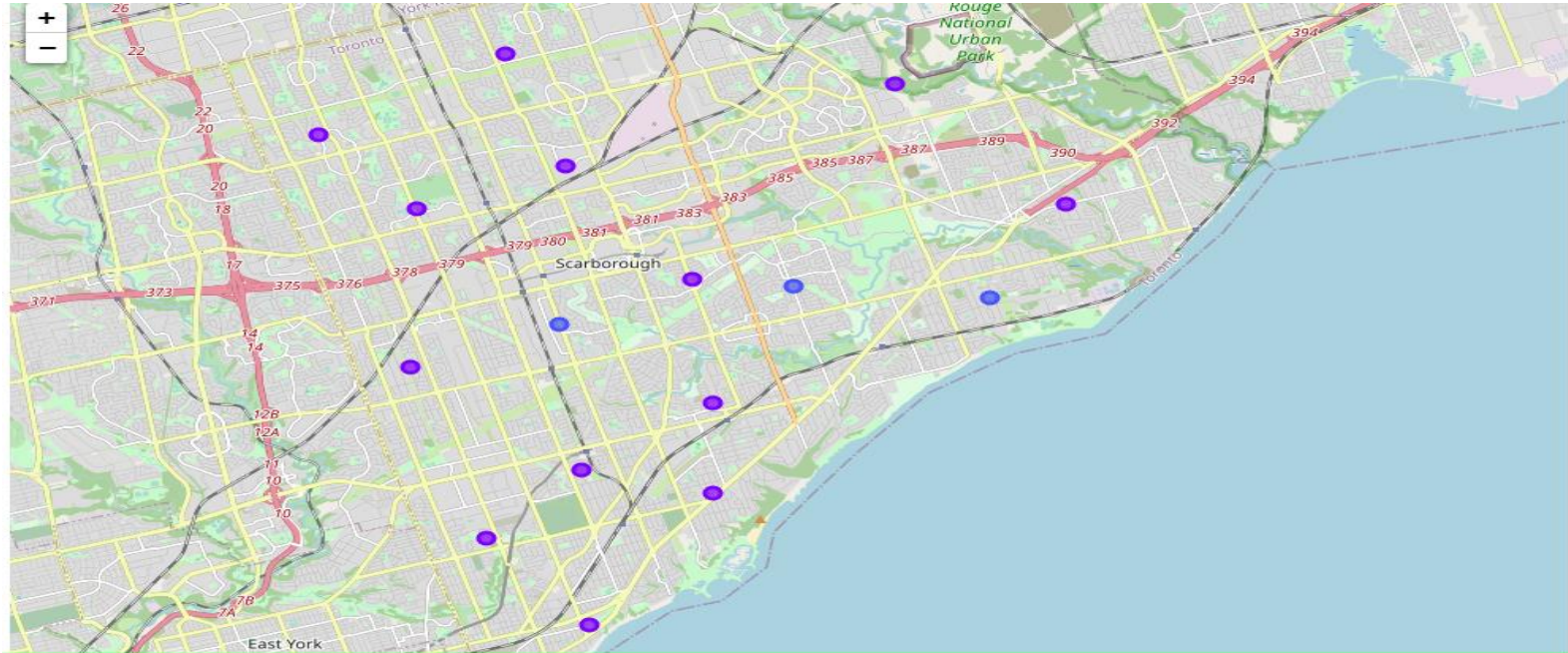
# Most Common Venues Near Neighborhoods

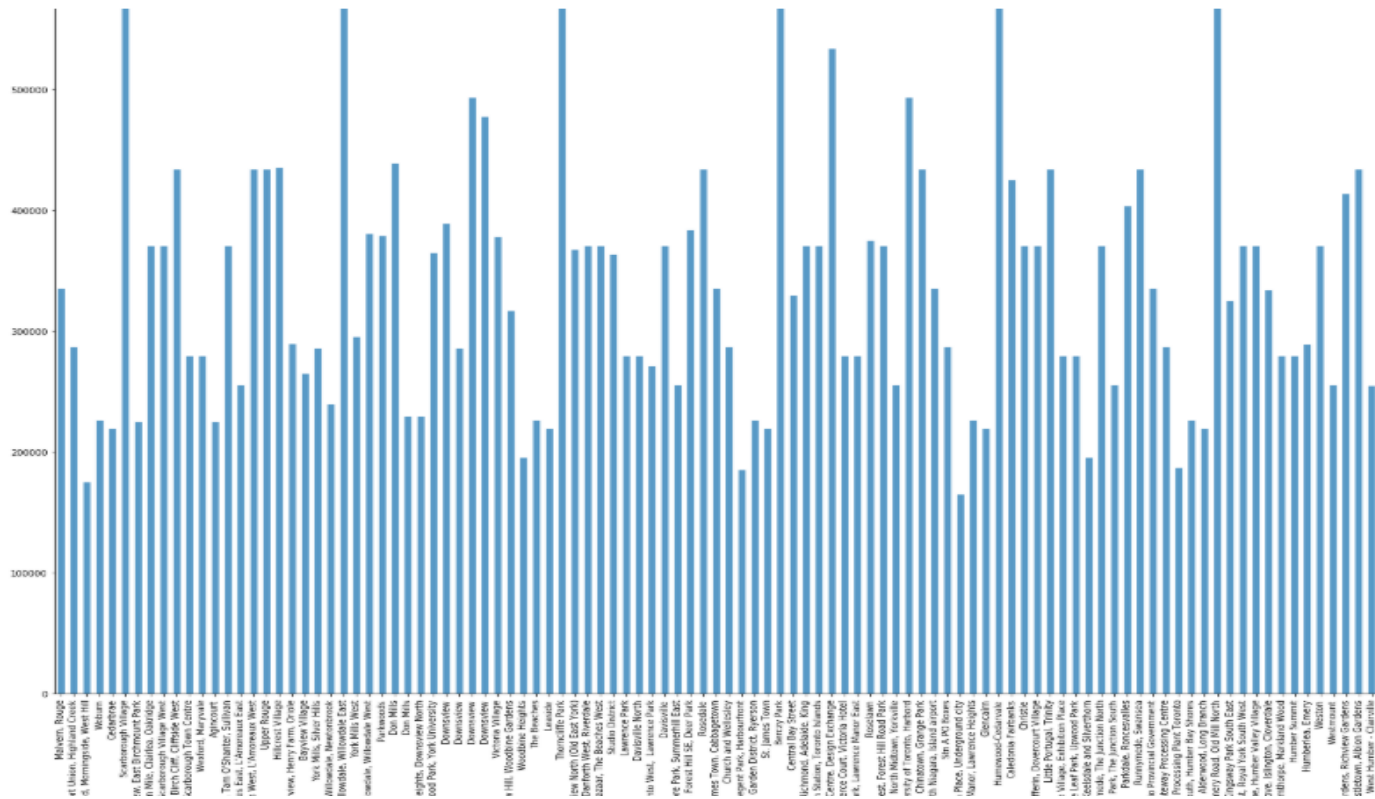
Out[45]:

talcode	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
3	Scarborough	Malvern, Rouge	43.81153	-79.19552	1	Zoo Exhibit	Fast Food Restaurant	Financial or Legal Service	Paintball Field	Cosmetics Shop	Doner Restaurant	Construction & Landscaping
2	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871	1	Home Service	Fish & Chips Shop	Moving Target	Bar	Event Space	Dumpling Restaurant	Eastern European Restaurant
1	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17520	2	Park	Athletics & Sports	Gym / Fitness Center	Yoga Studio	Dog Run	Donut Shop	Dumpling Restaurant
3	Scarborough	Woburn	43.76820	-79.21761	2	Park	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Yoga Studio	Ethiopian Restaurant	Donut Shop
1	Scarborough	Cedarbrae	43.76969	-79.23944	1	Thai Restaurant	Indian Restaurant	Bank	Gas Station	Athletics & Sports	Caribbean Restaurant	Hakka Restaurant

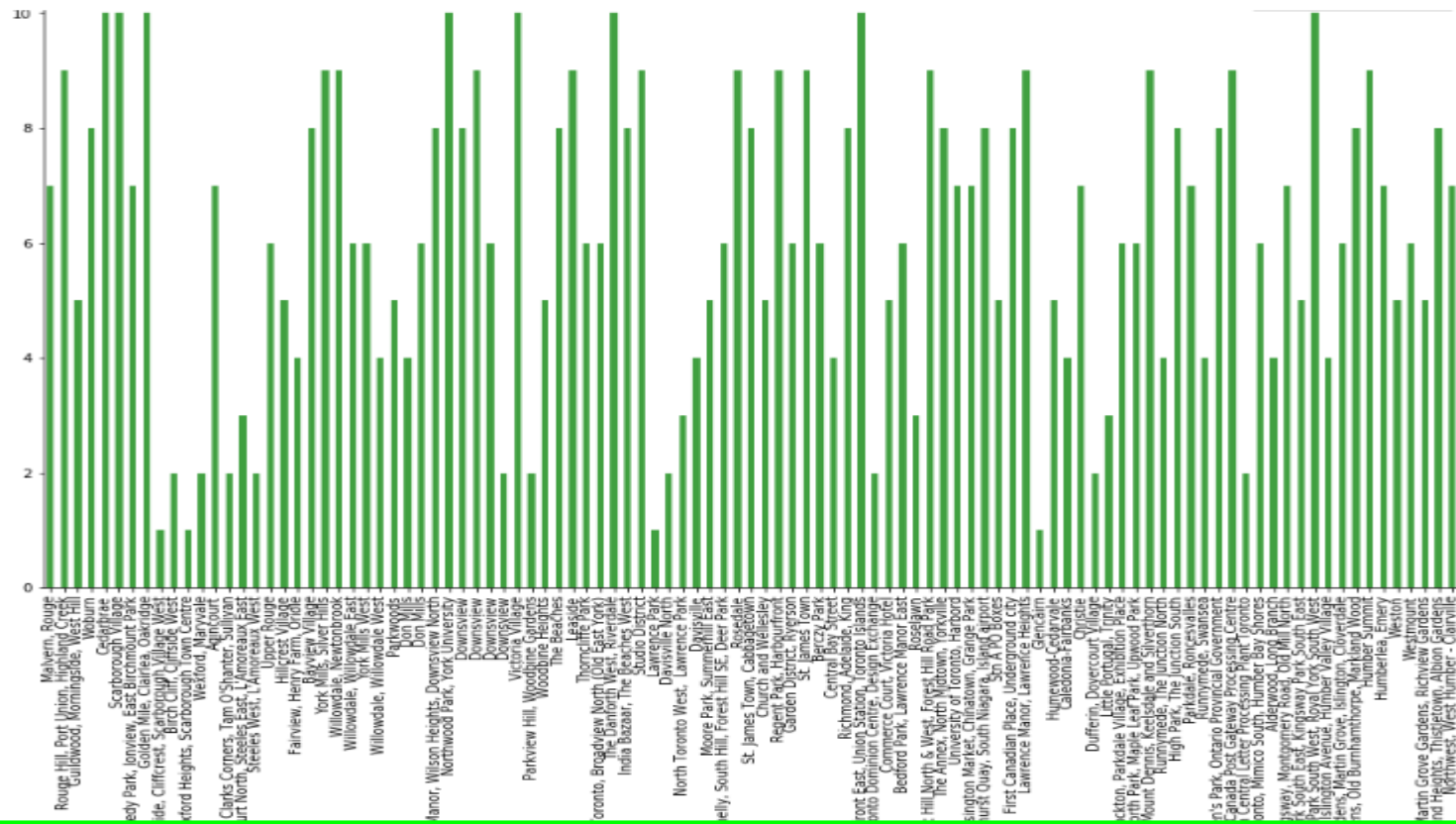
## Result Section :-

## Map of Clusters in Scarborough



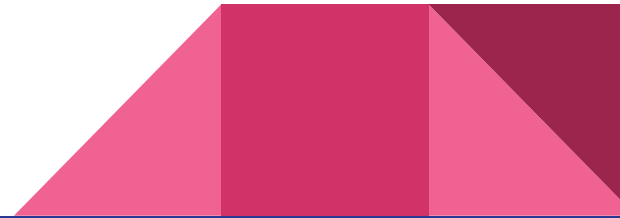






# Conclusion and future directions

- Built useful models to compare different neighborhoods.
- Separated the neighborhood into 10 different clusters and for 103 different latitude and longitude.
- Used charts to present the results of a particular neighborhood based on average house prices and school rating.



- Project can be continued by making the Model more precise in terms to find best house in Scarborough.
- Best means on the basis of all features can be calculated for better decision making and an efficient model.

