Analyzing Median House Prices and School Ratings for Scarborough Canada for Immigrants

Applied Data Science Capstone

IBM Data Science Professional Certificate

Part 1: Problem Description

 Many people migrating to various states of Canada require search of a good housing prices as well as good rating schools for their children. The projects aim to create an analysis of features for a neighborhood as a comparative analysis between neighborhoods. The features include median house price and school ratings and recreational facilities. This would help people to get awareness of the places before moving to a new country, state, city or place for their work or to start a new life

Part 2: Data We Need

- Longitude and Latitude Data:
- We will need geo-locational information about that specific borough and the neighborhoods in that borough. It is "Scarborough" in Toronto.
- Dataset comprising latitude and longitude, zip codes is already available through the previous notebook. The location of Scarborough would be filtered using the same:
- Part 2 Python Notebook Link Github

Scarborough / Coordinates
43.7764° N, 79.2318° W

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.811650	-79.195561
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785605	-79. <mark>15870</mark> 1
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765690	-79.175299
3	M1G	Scarborough	Woburn	43.768216	-79.217610
4	M1H	Scarborough	Cedarbrae	43.769608	-79.239440
5	M1J	Scarborough	Scarborough Village	43.743085	-79.23 <mark>217</mark> 2
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.726260	-79.263670
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.713213	-79.284910
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.723575	-79.234976
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.696690	-79.260069

Foursquare API

- Connecting to Foursquare and Retrieving Locational Data for Each Venue in Every Neighborhood
- After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

Data Preprocessing

- Processing the Retrieved Data and Creating a DataFrome for All the Venues inside the Scarborough.
- When the data is completely gathered, we will perform processing on that raw data to find our desirable features for each venue. Our main feature is the category of that venue. After this stage, the column "Venue's Category" will be One-hot encoded and different venues will have different feature-columns.

Creating a DataFrome

 Processing the Retrieved Data and Creating a DataFrome for All the Venues inside the Scarborough

Neighborhood	African Restaurant	Athletics & Sports	Auto Garage	Bakery	Bank	Bar	Bistro	Breakfast Spot	Brewery	Bubble Tea Shop	Burger Joint	
Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0
Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0
Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0
Rouge, Malvern	0	0	0	0	0	0	0	0	0	0	0	0
Highland Creek, Rouge Hill, Port Union	0	0	0	0	0	0	0	0	0	0	0	0

Main Article

• Part 4: Applying one of Machine Learning Techniques (K-Means Clustering)

```
# import k-means from clustering stage
from sklearn.cluster import KMeans

# run k-means clustering
kmeans = KMeans(n_clusters = 5, random_state = 0).fit(scarborough_onehot)
```

	Bakery	Breakfast Spot	Diner	Fish Market	Food & Drink Shop	Fruit & Vegetable Store	Grocery Store	Noodle House	Pizza Place	Sandwich Place	Total Restaurants	Total Joints	Total Sum
G5	2.000000	1.000000	0.000000	0.0	0.000000	0.00	0.000000	1.000000	1.000000	2.000000	21.000000	0.000000	28.000000
G1	1.333333	0.000000	0.000000	0.0	0.000000	0.00	0.333333	0.666667	1.666667	1.000000	13.333333	2.000000	20.333333
G4	0.000000	1.000000	0.000000	1.0	0.000000	0.00	3.000000	0.000000	3.000000	0.000000	8.000000	1.000000	17.000000
G3	1.500000	0.250000	0.000000	0.0	0.000000	0.25	1.000000	0.000000	0.750000	0.750000	6.750000	1.250000	12.500000
G2	0.285714	0.142857	0.285714	0.0	0.142857	0.00	0.142857	0.000000	0.857143	0.428571	2.000000	0.714286	5.000000

Finding Most Common Venues

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
Agincourt	Shopping Mall	Chinese Restaurant	Supermarket	Pool	Breakfast Spot	Malay Restaurant	Mediterranean Restaurant	Hong Kong Restaurant
Agincourt North, L'Amoreaux East, Milliken, St	Coffee Shop	Pharmacy	Sandwich Place	Zoo Exhibit	Construction & Landscaping	Convenience Store	Deli / Bodega	Department Store
Birch Cliff, Cliffside West	Park	Gym	Gym Pool	General Entertainment	Café	Skating Rink	College Stadium	Discount Store
Cedarbrae	Flower Shop	Athletics & Sports	Bakery	Thai Restaurant	Bank	Hakka Restaurant	Caribbean Restaurant	Indian Restaurant
Clairlea, Golden Mile, Oakridge	Coffee Shop	Bus Line	Diner	General Entertainment	Ice Cream Shop	Intersection	Metro Station	Convenience Store

Decision Making and Reporting Results

Now, we focus on the centers of clusters and compare them for their "Total Restaurants" and their "Total Joints". The group which its center has the highest "Total Sum" will be our best recommendation to the contractor.

{Note: Total Sum = Total Restaurants + Total Joints.} This algorithm although is pretty straightforward yet is strongly powerful.

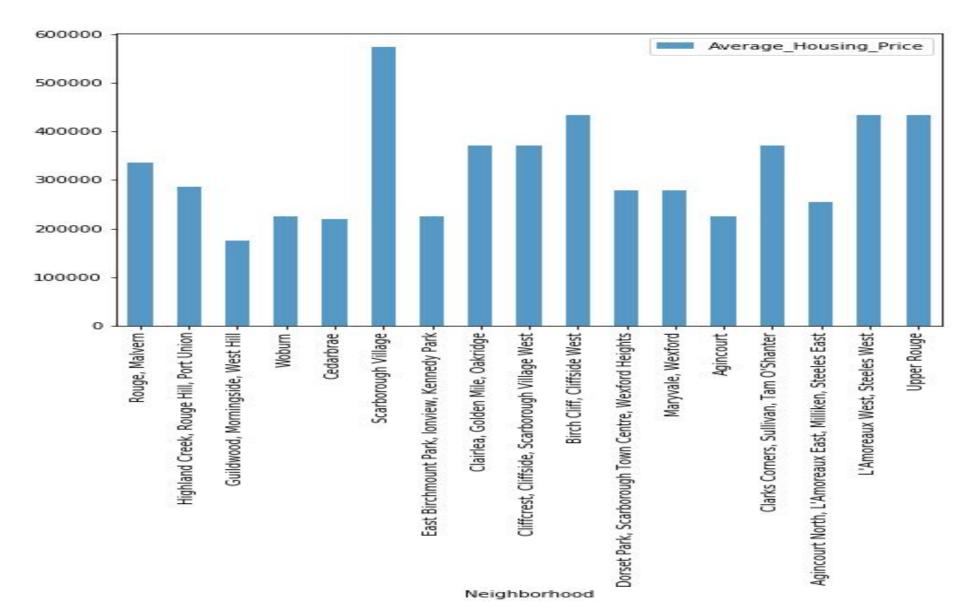
Clustering

```
# Using k-means to cluster the neighborhood into 3 clusters.

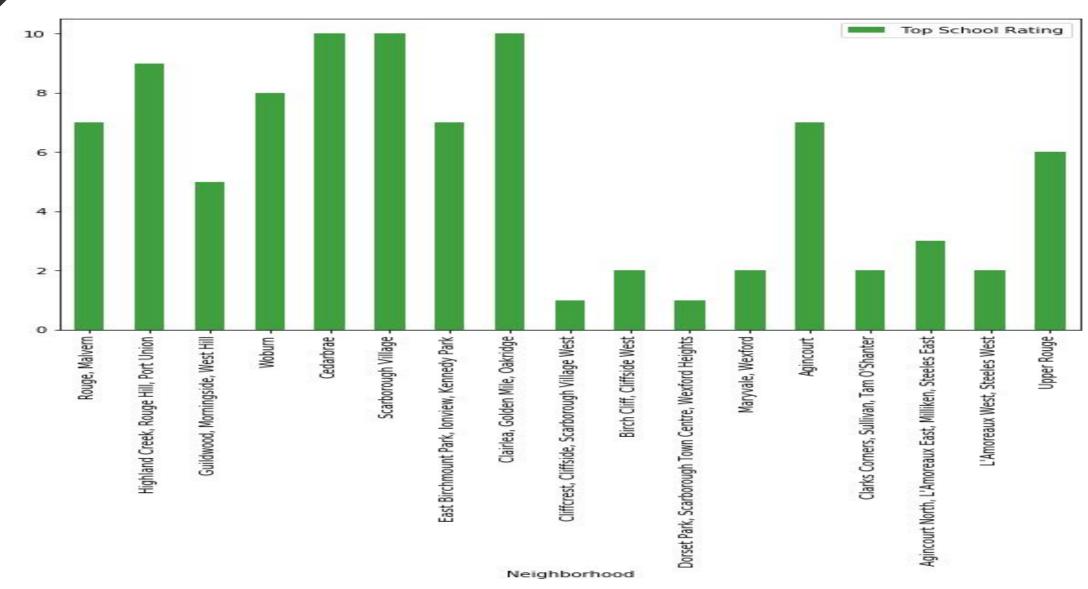
Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)
#Scarborough_grouped_clustering = Scarborough_grouped

# run k-means clustering
kmeans = KMeans(n_clusters=3, random_state=0).fit(Scarborough_grouped_clustering)
```

NEIGHBORHOOD MEDIAN HOUSING PRICES



NEIGHBORHOOD SCHOOL RATINGS



Conclusion:

• In this project, through a k-means cluster algorithm we separate the neighborhood into 03 clusters, which have similar neighborhoods around them. Using the charts above decision leading to a particular neighborhood based on average house prices and school rating can be made

THANKS

Author: Hrishikesh Jadhav