IBM

APPLIED DATA SCIENCE CAPSTONE

PRESENTATION

INTRODUCTION

This is a capstone project for IBM Data Science Professional Certificate. For this project I will be focusing extensively on Scarborough Toronto Canada.

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. It includes a number of natural landmarks, including the Toronto Zoo, Rouge Park and the Scarborough Bluffs.

For this project, I would be looking at Scarborough Toronto and the exciting and common places or venues that new immigrants can visit to easily get accustomed to their new environment. This is because Scarborough is a large, multicultural area that contains the Scarborough Bluffs, huge cliffs overlooking Lake Ontario, lined with parks, beaches, and hiking trails. Inland, the sprawling Toronto Zoo features global animal pavilions, close-up encounters, and a wildlife health center.

This project is aimed to help new immigrants from different parts of the world relocating to Scarborough get accustomed to their new environment by knowing the common venues around like restaurants, beaches, parks etc.

BUSINESS PROBLEM

The objective of this capstone project is to find common venues around the Scarborough Neighborhoods. By using Wikipedia as our data source and data science tools python, pandas and beautiful soup for web scraping. By also using Foursquare API to get venue data on restaurants, parks, beaches, worship centers around the Scarborough Neighborhoods, this project is aimed at helping new immigrants get accustomed to their environment in time and also to citizens whether or not to visit Scarborough Toronto Canada.

TARGET AUDIENCE

To help new immigrants who just moved to Scarborough know their surroundings and also help citizens of other countries who are looking to visit or relocate to Scarborough know the different venues around there and make their decision

DATA SECTION

DATA SOURCE

To solve this problem, we will need below data:

● List of neighborhoods in Toronto, Canada

● Latitude and Longitude of these Scarborough Neighborhood

● Venue data related to restaurants, worship centers etc. around Scarborough.

DATA EXTRACTION

● Scrapping of Toronto neighborhoods via Wikipedia

● Getting Latitude and Longitude data of Scarborough neighborhood via Geocoder package

● Using Foursquare API to get venue data related to these Neighborhoods around Scarborough Toronto Canada.

METHODOLOGY

First, I need to get the list of neighborhoods in Toronto, Canada. This is possible by extracting the list of neighborhoods from Wikipedia:

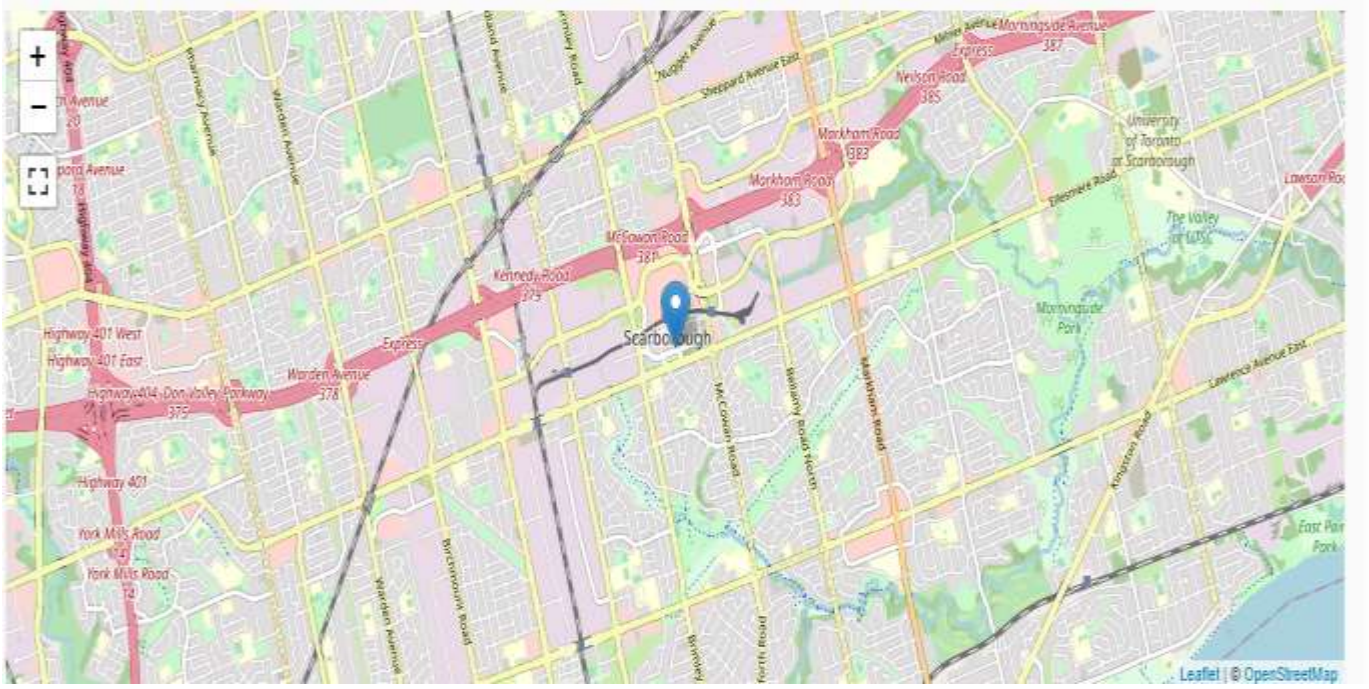
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

I performed web scraping by utilizing pandas and beautiful soup as it is easier and more convenient to pull tabular data directly from a web page into the data frame.

	PostalCode	Borough	Neighborhood
0	M1B\n	Scarborough\n	Malvern, Rouge
1	M1C\n	Scarborough\n	Rouge Hill, Port Union, Highland Creek
2	M1E\n	Scarborough\n	Guildwood, Morningside, West Hill
3	M1G\n	Scarborough\n	Woburn
4	M1H\n	Scarborough\n	Cedarbrae

However, it is only a list of neighborhood names and postal codes. I need to get their coordinates to utilize Foursquare to pull the list of venues near these neighborhoods. To get the coordinates, I tried using Geocoder Package but it was not working so I used the CSV file provided by IBM team to match the coordinates of Toronto neighborhoods. After gathering

these coordinates, I visualize the map of Scarborough Toronto using Folium package to verify whether these are correct coordinates.



I tested API key and ID connection on Foursquare by first pulling out one venue from Scarborough Neighborhood, after I did that successfully.

I now used Foursquare API to pull the list of top 100 venues within 500 meters radius. I was able to accomplish this because I created a Foursquare developer account in order to obtain account ID and API key to pull the data.

```
Malvern, Rouge
Rouge Hill, Port Union, Highland Creek
Guildwood, Morningside, West Hill
Woburn
Cedarbrae
Scarborough Village
Kennedy Park, Ionview, East Birchmount Park
Golden Mile, Clairlea, Oakridge
Cliffside, Cliffcrest, Scarborough Village West
Birch Cliff, Cliffside West
Dorset Park, Wexford Heights, Scarborough Town Centre
Wexford, Maryvale
Agincourt
Clarks Corners, Tam O'Shanter, Sullivan
Milliken, Agincourt North, Steeles East, L'Amoreaux East
Steeles West, L'Amoreaux West
Upper Rouge
```

From Foursquare, I am able to pull the names, categories, latitude, and longitude of the venues. With this data, I also checked how many unique categories that I can get from these venues.

While utilizing pandas and one hot coding I placed all venues and their categories in a new data frame describing each venue and the most common spots around the venue in Scarborough.

Then, I analyzed each neighborhood by grouping the rows by neighborhood in Scarborough and taking the mean on the frequency of occurrence of each venue category.

```
----Agincourt----
                  venue  freq
0           Clothing Store   0.2
1  Latin American Restaurant   0.2
2                   Lounge   0.2
3          Breakfast Spot   0.2
4            Skating Rink   0.2


----Birch Cliff, Cliffside West----
                  venue  freq
0       College Stadium  0.25
1          Skating Rink  0.25
2  General Entertainment  0.25
3                  Café  0.25
4   American Restaurant  0.00


----Cedarbrae----
                  venue  freq
```

After which I made a new data frame for neighborhood in Scarborough and their common

Result and Discussion:

	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	Agincourt	Clothing Store	Skating Rink	Breakfast Spot	Latin American Restaurant	Lounge	Vietnamese Restaurant	Gas Station	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
1	Birch Cliff, Cliffside West	General Entertainment	Skating Rink	College Stadium	Café	Vietnamese Restaurant	Clothing Store	Gas Station	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant
2	Cedarbrae	Hakka Restaurant	Athletics & Sports	Bakery	Bank	Gas Station	Fried Chicken Joint	Caribbean Restaurant	Thai Restaurant	College Stadium	Grocery Store
3	Clarks Corners, Tam O'Shanter, Sullivan	Pizza Place	Pharmacy	Fast Food Restaurant	Noodle House	Italian Restaurant	Rental Car Location	Gas Station	Bank	Fried Chicken Joint	Thai Restaurant
4	Cliffside, Cliffcrest, Scarborough Village West	American Restaurant	Motel	General Entertainment	Gas Station	Gaming Cafe	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Discount Store	Department Store

The result shows the different neighborhoods in Scarborough and the respective common venues around them. A neighborhood like

“Agincourt” has more clothing stores located around it will neighborhood be for immigrants who love shopping, for someone more interested in sports neighborhood “Cedarbrae” will be a better selection as it has more athletic and sports venues. This result also helps citizens from other countries know which neighborhoods of Scarborough have more of their areas of interest.

Although, the real challenge is constructing the dataset because the needed data isn’t publicly available. When combining data from multiple sources, inconsistent can happen. And lots of efforts are required to check, research and change the data before merge.

But again, on the bright side, the insight, gotten from observing the analysis results, seems logical. And the insight is awareness of the mutual places around for immigrants getting accustomed to Scarborough.

Conclusion

More research and analysis should put compiling better data on common venues and spots in Scarborough as it is one of the most diverse and multicultural areas in Toronto, and common ground for people of different ethnicity and cultures. Its also one of the most visited spots by new immigrants who relocate to Toronto Canada.