Battle of Neighborhoods

ANCHALA BALARAJ

INTRODUCTION

This report is part of the capstone Project for IBM's Applied Data Science Professional Certificate offered by Coursera, part of the 10 series course.

The purpose of this project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decisions on selecting great neighborhoods out of numbers of other neighborhoods in Scarborough, Toronto.

We will be using data visualization techniques, Foursquare API to retrieve location data for the state of Toronto and use this to perform data analysis.

BACKGROUND

In growing regions - it is a challenge to know where it is appropriate to start a business, let alone what business to start. In this report we will be exploring the need for additional restaurants as in Toronto.

The goal is to inform potential business owners who are looking to expand in new areas. After pulling the data for these respective cities - we will have an ability to provide a recommendation. We will be using several data visualization techniques, in particular, we will be making use of Foursquare API to retrieve location data for the state of Toronto in Canada and use this data to perform data analysis.

PROBLEM

Prior launching any restaurant, it's important to know if the business as a good opportunity. In order to do so, this report will try to gather data about other restaurant localization, competitors and best localization.

For an entrepreneur to open a new restaurant in the neighborhood in Toronto then what neighborhood should he/she choose based on the type of Cuisine. And if they have specific type in mind then which location would be better ensuring a good amount of customer traffic but also keeping in mind the nearby tourist spots they can visit which makes it easier for business person to choose the right neighborhood for their restaurant.

INTEREST

This analysis will be useful for those who want to start a new business in the state of Toronto .It also aims on helping those who want to travel to Toronto and want to visit global Cuisine restaurants in Toronto. Clusters tell us what neighborhoods are fairly similar to each other so the person can skip travelling to many of the same neighborhoods.

DATA ACQUISTION AND CLEANING :

In order to gain information ,we will use 'Foursquare API' locational information. Foursquare is location data provider with the information about all manner of venues and events within an are of interest. Such information includes venue names and location and photo.

 The foursquare location platform will be used as the sole data source all the stated required information can be obtained through the API.

We connect the Foursquare API to gather information about the various venues inside each and every neighborhood. The data retrieved from Foursquare contained information of venues within a specified distance of longitude and latitude of the postcodes. We acquire the data about various neighborhoods , boroughs and postal codes from Wikipedia page using Beautiful Soup ,we put it into a data frame. Pandas is used to read the longitudes and latitudes in the CSV file. We make Foursquare API calls using our credentials to acquire the location data.

	Postal Code	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Data from Wikipedia

- Cleaning Data:

We can combine rows that have same postal code and make them into a single row and neighborhoods that are associated with that postal code would be put into the same row separated by commas. We sort data buy the postal codes.

	index	Postal Code	Borough	Neighborhood
0	6	M1B	Scarborough	Malvern, Rouge
1	12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	18	M1E	Scarborough	Guildwood, Morningside, West Hill
3	22	M1G	Scarborough	Woburn
4	26	M1H	Scarborough	Cedarbrae

Sorted the data by Postal Code

- Feature Selection:

We focus on the neighborhood in Toronto. So drop all rows that have Borough outside of Toronto. And our dataframe is ready to use.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Data only for Neighborhoods in Toronto

- Exploratory Data Analysis :

We can visualize the various neighborhoods in Canada by drawing a map and plotting the neighborhoods on top. This allows us to see what we are dealing with and they are scattered. We an now find various venues present in the neighborhood in Toronto by exploring any one neighborhood .

Similarly ,we can explore other neighborhoods and this information can be very useful for potential business owner wanting to start a new business in any of the neighborhoods.

Neighbourhoods in Toronto

METHODLOGY

- Presence of Vast data & Classification :

The dataset will allow us to group the neighborhoods together according to the similarity in the type of venues in each neighborhood.
If two neighborhoods are very popular for their beaches and cafes then they can be put into the same group.
we can start to group the neighborhoods based on the similarity of their topmost venues. If two neighborhoods have the same top few venues then they can be groups together in the same row. We make use of the mean of the frequency of the occurrence of each category and combine

the neighborhoods with similar venues.

	Postal Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Trail	Health Food Store	Pub	Women's Store	Cupcake Shop	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Restaurant	Bubble Tea Shop	Bakery	Pub	Pizza Place
2	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	0	Park	Board Shop	Food & Drink Shop	Sandwich Place	Light Rail Station	Italian Restaurant	Burrito Place	Liquor Store	Restaurant	Ice Cream Shop
3	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Café	Coffee Shop	Gastropub	Bakery	Brewery	American Restaurant	Yoga Studio	Comfort Food Restaurant	Seafood Restaurant	Sandwich Place
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	2	Park	Swim School	Bus Line	Electronics Store	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run	Distribution Center	Discount Store

Clustered Neighbourhoods

- Clustering

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 .We will cluster the data using KMeans Clustering.
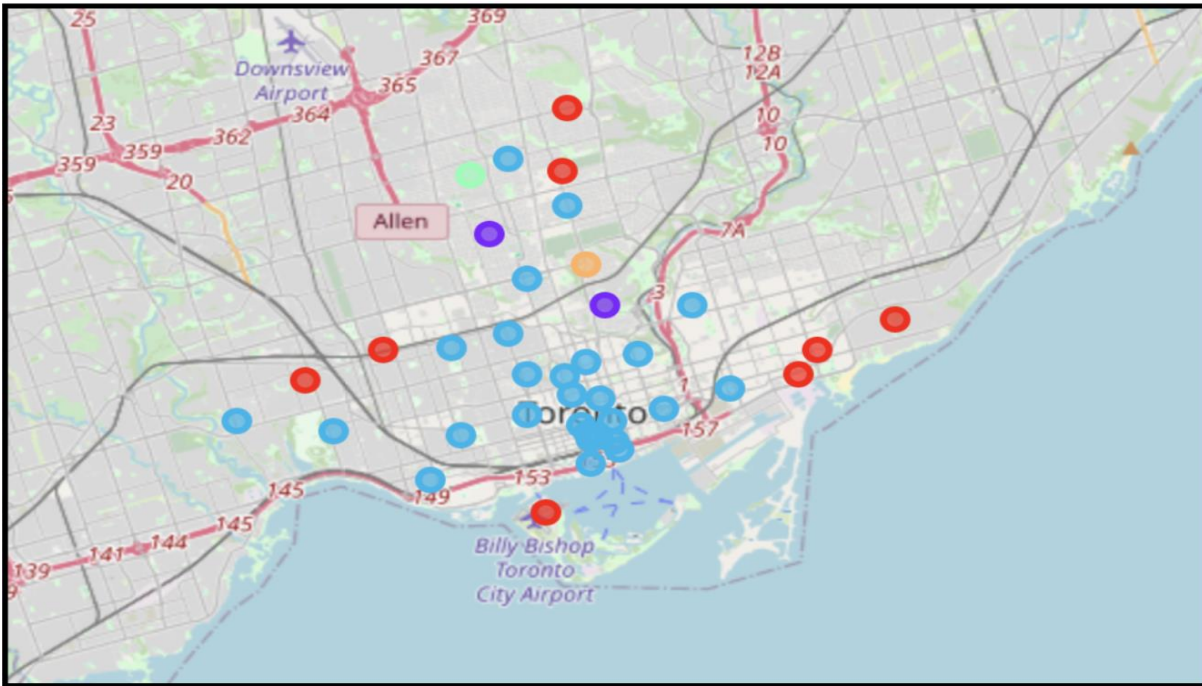
Clustered Neighborhoods in Toronto

RESULT

Scarborough is a popular destination. It is one of the most diverse and multicultural areas in the Greater Toronto Area. The neighborhoods in the Red cluster are mainly on the outskirts of Toronto.

The neighborhoods in the Red cluster are mainly on the outskirts of Toronto.

This project has used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing, and details about a business.

DISCUSSION

The major purpose of this project is to suggest a better neighborhood in a new city for the person who is shifting there. Social presence in society in terms of like-minded people. Connectivity to the airport, bus stand, city center, markets, and other tourist spots nearby. Our aim was to help potential business owners and tourists in picking our the right neighborhood to travel or open a business.

But they should also keep in mind the interest of the people. People in a particular neighborhood should be interested in the entrepreneur's business.

This also gives them an idea about how they can scale profitably and open more branches in different clusters.

CONCLUSION

In this project, using the k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters, this allows interested people to understand how similar neighborhoods are in Toronto.
The similarity in neighborhoods allows tourists to decide which places they should add on the to-visit list without making redundant choices. It also tells them what are the most popular places and each neighborhood that they must visit.