BEST PLACE WITH GOOD INFRASTRUCTURE IN TORONTO

Introduction:

Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario's north-western shore. It is a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. Toronto also has many green spaces, from the orderly oval of Queen's Park to 400-acre High Park and its trails, sports facilities, and zoo.

Toronto is a prominent centre for music, theatre, motion picture production, and television production, and is home to the headquarters of Canada's major national broadcast networks and media outlets. Its varied cultural institutions, which include numerous museums and galleries, festivals and public events, entertainment districts, national historic sites, and sports activities,[38] attract over 43 million tourists each year. Toronto is known for its many skyscrapers and high-rise buildings, in particular the tallest free-standing structure in the Western Hemisphere, the CN Tower.

The city is home to the Toronto Stock Exchange, the headquarters of Canada's five largest banks, and the headquarters of many large Canadian and multinational corporations. Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, aerospace, environmental innovation, food services, and tourism.

Business Problem:

1. List and visualize all major parts of Toronto City with top existing infrastructure.

2. List the best locations in Toronto City as per infrastructure.

3. List of all areas lacking infrastructure facilities.

4. Best place to stay in Toronto City by Clustering.

Target Audience:

The purpose of this project is to help people in exploring better facilities around place in Toronto city. It will help people making a smart and efficient decision on selecting great neighbourhoods out of numbers of other postal areas in Toronto. This project is for those people who are looking for better neighbourhoods and businesses. It will help people to get the awareness of the area and neighbourhood before moving to a new city, state, country, or place for their work or to start a new fresh life.

Data Description:

This project uses the following data:

- o List of postal codes of Canada: M

    - o Source:
      https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

    - o Description: This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

- o Geospatial data of Toronto

    - o Source: Lab week 3

    - o Description: Toronto geospatial data obtained from the third week of the Lab, this data is in accordance with the data obtained from the wiki page.

Methodology:

Data Exploration: We have to do web scraping using Python requests to extract the list of neighbourhood data and use read_html. Merge the data from wiki page with geospatial data.

Data Visualization: Visualize the neighbourhoods in a map using Folium package, this allows us to perform a sanity check to make sure the geographical coordinate's data returned by Geocoder are correctly plotted in the city of Toronto.

Infrastructure Exploration: We use Foursquare API to get the top 300 venues that are within a radius of 500 meters. We need to register a Foursquare Developer Account in order to obtain access to Foursquare Developer. Foursquare will return the venue data in JSON format and we will extract the venue name, category, latitude, and longitude. From this data we can check how many venues were returned for each neighbourhood.

Data Clustering: In the end we will perform clustering on the data using K-means clustering. k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centres or cluster centroid), serving as a prototype of the cluster. The data will be cluster into 3 clusters based on their frequency of occurrence for number of existing infrastructures. The result will show which neighbourhoods have higher, medium, and lower concentration of infrastructures.
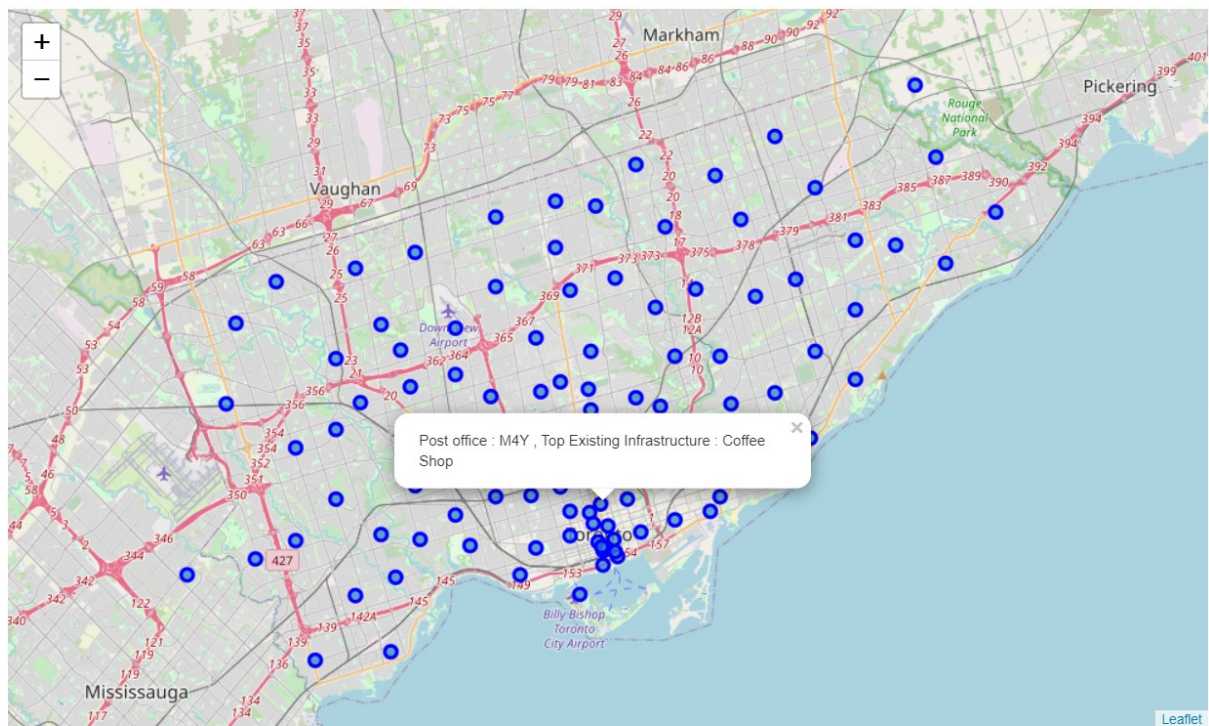
Result:

Figure 1 Display the top existing infrastructure for each Postal Code in Toronto

	50	62
Postal Code	M5H	M5X
Borough	Downtown Toronto	Downtown Toronto
Neighbourhood	Richmond, Adelaide, King	First Canadian Place, Underground city
Airport Terminal	0	0
Bank	0	0
Bus Station	0	0
Business Service	0	0
Café	9	5
College Auditorium	0	0
Electronics Store	1	1
Farmers Market	0	0
Garden	0	1
Gas Station	0	0
Gym / Fitness Center	1	1
Hospital	0	0
Hotel	4	9
Indie Movie Theater	0	0
Light Rail Station	0	0
Lounge	1	0
Market	0	0
Metro Station	0	0
Monument / Landmark	1	1
Office	1	1
Park	0	1
Pharmacy	0	0
Playground	0	0
Restaurant	5	5
School	0	0
Shopping Mall	1	0
Theater	3	2
Train Station	0	1
University	1	0
Total infrastructure	28	28

Figure 2 The best locations in Toronto as per Infrastructure

|	Postal Code	Borough	Neighbourhood				
1	M1G	Scarborough	Woburn	23	M3J	North York	Northwood Park, York University
5	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	24	M3K	North York	Downsview
6	M1N	Scarborough	Birch Cliff, Cliffside West	29	M4E	East Toronto	The Beaches
7	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	56	M5P	Central Toronto	Forest Hill North & West, Forest Hill Road Park
8	M1R	Scarborough	Wexford, Maryvale	63	M6B	North York	Glencairn
				64	M6C	York	Humewood-Cedarvale
11	M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...	70	M6L	North York	North Park, Maple Leaf Park, Upwood Park
15	M2M	North York	Willowdale, Newtonbrook	79	M8W	Etobicoke	Alderwood, Long Branch
18	M2R	North York	Willowdale, Willowdale West	80	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North
19	M3A	North York	Parkwoods	86	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...
20	M3B	North York	Don Mills				

Figure 3 Areas with Lack the Infrastructure Facilities

Clustering Based on Total Infrastructure: The results from the K-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for number of existing infrastructures:

Cluster 1: Neighbourhoods with a high number of infrastructures.

Cluster 2: Neighbourhoods with a low number to no existence of infrastructures.

Cluster 3: Neighbourhoods with a moderate number of infrastructures.
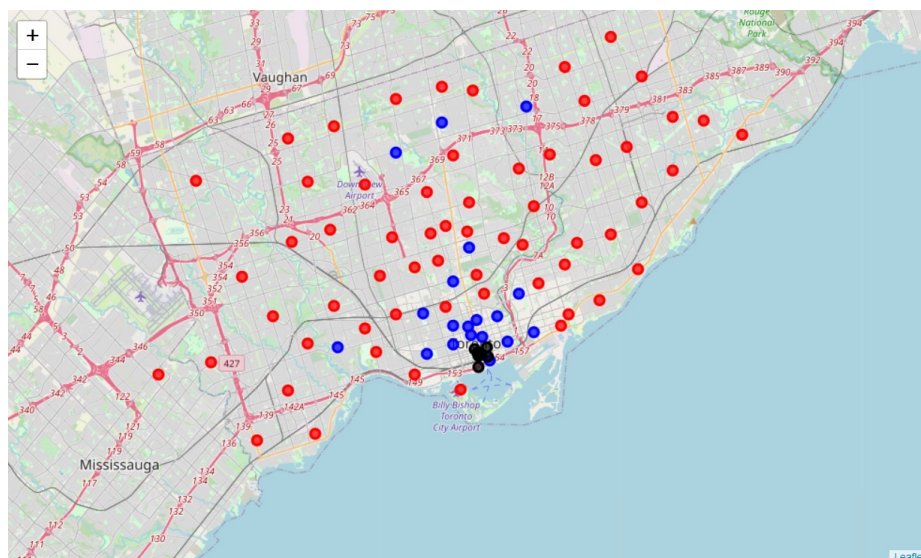
Discussion:

Most of the infrastructures are spread in Toronto, with the highest number in cluster 1 and moderate number in cluster 3. On the other hand, cluster 2 has a very low number of infrastructures in the neighbourhoods. This represents a great opportunity and high potential areas to open new infrastructures as it is very little to no competition from existing varied infrastructures. Meanwhile, one can specifically check the infrastructure of choice against the postal office choice area. Infrastructures in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of already established Infrastructures.

Conclusion:

In this project, I have gone through the process of identifying the business problems, specifying the data required, extracting, and preparing the data, visualizing the results, performing machine learning to build cluster, tackling and reaching to a definitive solution to business problems. The project is providing recommendations to relevant stakeholders i.e., property business regarding the best locations to open new infrastructure. This project also provides visitors and immigrants to the city regarding Postal Code areas for growth and living prosperously.