LOCATION PROVIDER FOR STAY IN TORONTO

PROBLEM DESCRIPTION

Most of the people from various locations either from different city or abroad are vising Toronto for stay for many purposes. The purpose may be for job or tourist visit. As per wiki Toronto is very highly populated and people don't have much time to do research the location to get stay according to their preferences. Due to limited time and expenses to research the location they unwillingly found the stay in their unreferred locations where they don't find venues of their choices nearby.

To work on this problem we are taking the leverage of Data science and machine learning to help people find the locations of stay based on their preferences, near by venues of their choice.

By using the locations and their preferred venue, using machine learning models and data science we are helping new visitors to find the stay of their choice in very short and with very little or no expenses.

The target audience will be any one who visits Toronto for any purpose.

DATA COLLECTION

The following data will be used for data exploration and data analysis:

1) Toronto Neighborhood Data - 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada'

2) Foursquare Location Data (Venues [Restaurants, Community Centers], TOP tips, Favorites, User Experience, etc.) will be used to cluster, segment, target, and position to craft recommendations for the new visitors.

First, Toronto borough neighborhood addresses are converted into their equivalent latitude and longitude using the values retrieved from Toronto geodata sheet using Pandas library in Python. Then, the Foursquare API is used to explore neighborhoods in Toronto borough. Using the explore function, the most common venue categories within a radius of 500 meters is extracted for each Toronto neighborhood and the number of returned venues by Foursquare API is limited to 100.

Now, the venue categories feature is used to group the neighborhoods into clusters. The k-means clustering algorithm is employed to perform this task. Finally, the Folium library is utilized to visualize the neighborhoods in Toronto borough and their emerging clusters.

METHODOLOGY SECTION

This section presents the datasets used for this study and details of the methods used to perform our analysis. The data for this project is obtained from wiki 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada' and geo location data from csv 'Geospatial_Coordinates.csv ' and the Foursquare location data to solve the business problem discussed above. First, Borough contaning 'Toronto' neighborhood addresses are converted into their equivalent latitude and longitude values retrieved from Toronto geodata using Pandas library in Python. Side Table shows the first 5 examples of a dataset comprising of Borough, Neighborhood, Latitude, and Longitude of 70 Toronto borough neighborhoods.

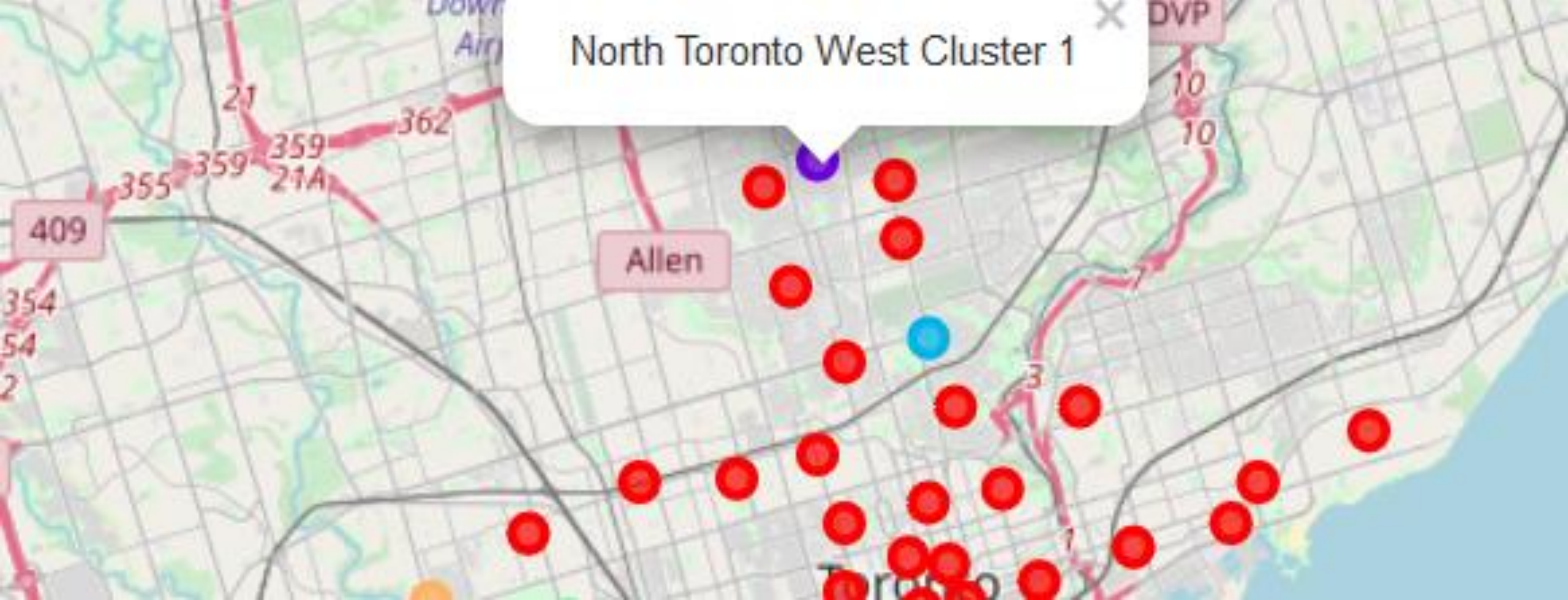|	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Harbourfront,Regent Park	43.654260	-79.360636
1	M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937
2	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
3	M4E	East Toronto	The Beaches	43.676357	-79.293031
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306

METHODOLOGY SECTION (CONTINUED)

Foursquare API is used to explore neighborhoods in Toronto borough. Using the explore function, the most common venue categories within a radius of 500 meters is extracted for each Toronto neighborhood and the number of returned venues by Foursquare API is limited to 100. Side Table presents the first 5 venues of dataset returned.

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront,Regent Park	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Harbourfront,Regent Park	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Harbourfront,Regent Park	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
3	Harbourfront,Regent Park	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot
4	Harbourfront,Regent Park	43.65426	-79.360636	Cooper Koo YMCA	43.653191	-79.357947	Gym / Fitness Center

METHODOLOGY SECTION (CONTINUED)

Foursquare API is used to explore neighborhoods in Toronto borough. Using the explore function, the most common venue categories within a radius of 500 meters is extracted for each Toronto neighborhood and the number of returned venues by Foursquare API is limited to 100. Below tableshows the first 5 venues of dataset returned.

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront,Regent Park	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Harbourfront,Regent Park	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Harbourfront,Regent Park	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
3	Harbourfront,Regent Park	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot
4	Harbourfront,Regent Park	43.65426	-79.360636	Cooper Koo YMCA	43.653191	-79.357947	Gym / Fitness Center

RESULTS

5 cluster is created consisting of top 10 most common venue represented in different colors

CONCLUSION SECTION

This report explores borough containing 'Toronto' neighborhoods using the Foursquare API and utilizes K-means machine learning algorithm to group neighborhoods into clusters. These clusters are based on the 10 most common venue categories feature extracted from Foursquare location data. This information can now be used in making selection decision for a stay in each neighborhood by the vistitors to Toronto. For all Neighborhood in each cluster venues are sorted in 10 most common venue. If there prefreed venue is Coffee Shop,Bakery,Café,Park,Mexican Restaurant,Breakfast Spot,Restaurant,Pub,Theater,Yoga Studio with coffee shop as most priortized venue then they can prefer to stay in Downtown Toronto.