

Final Report

Introduction

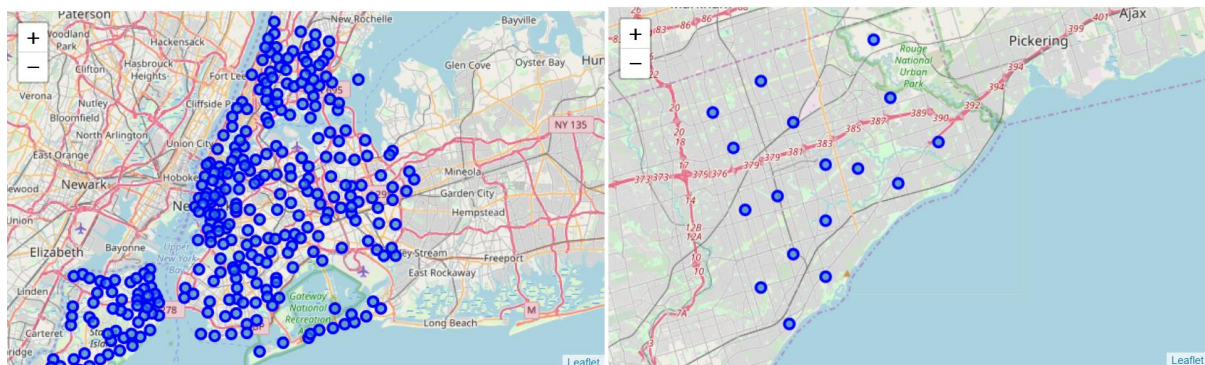
We will be exploring the neighborhood of New York City and the city of Toronto and find out if they are many similarities between them. We will explore the venues in the respective cities and check the top neighborhoods. This can be useful for those who will be visiting the cities from Canada to the USA or the other way around. Both countries are diverse, also being among the popular cities in the countries.

Data

The data I will be using is the New York City neighborhood data and the scrape the Wikipedia page for Toronto neighborhood data. Convert the HTML data into a pandas data-frame, clean it and add the latitude and longitude for each of the neighborhoods. A foursquare endpoint will be called to get the data for venues in every neighborhood for New York and Toronto.

Methodology

In the exploratory analysis visualizations of the boroughs in New York and Toronto city were generated as a starting step. These provide an understanding of the neighborhoods are distributed in the respective cities. The pictures below show the neighborhoods for the respective cities, to generate the visualization Folium visualization library was used. Each circle mark reveals the name of the neighborhood and its respective borough.



a) New York City

b) Toronto City

The Foursquare endpoint was used to explore the neighborhoods so let's first explore a neighborhood in each of the city. For illustration purposes, we have chosen Manhattan borough from the New York City and Scarborough from the Toronto city. A GET call is used to retrieve the information about the neighborhoods by calling the explore endpoint of the foursquare API. The Client credentials such as the client id and the client secret are passed in the URL and it returns a JSON response which has the results in it. The JSON is cleaned and structured into a pandas data frame. For each neighborhood in the city, it returns the list of venues.

Now we will generate the list of venues for every neighborhood in both the cities with help of the foursquare explore endpoint. The venues are merged into the data frame of the neighborhoods. Now, we can find out which neighborhoods consist are similarly based on the venue categories further we take the mean of the frequency of the categories to find out the most common venues in the neighborhood.

Now we generate a data frame consisting of the top 10 venues in both the cities. This will help the individual traveling to the city to decide which neighborhood is like the one he is coming from and plan his stay accordingly. For that we merge the original data and the venue data into a single data frame to gain more insight as shown in the figure below:

Out[53]:

	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	Coffee Shop	General Entertainment	Gas Station	Fried Chicken Joint	Fast Food Restaurant
1	Cedarbrae	Hakka Restaurant	Athletics & Sports	Bakery	Bank	Gas Station	Fried Chicken Joint	Caribbean Restaurant	Thai Restaurant	Construction & Landscaping	Gym Pool
2	Cliffcrest	American Restaurant	Motel	Coffee Shop	Grocery Store	General Entertainment	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Electronics Store	Department Store
3	Cliffside West	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Coffee Shop	Grocery Store	Gas Station	Fried Chicken Joint	Fast Food Restaurant
4	East Birchmount Park	Hobby Shop	Coffee Shop	Department Store	Chinese Restaurant	Gym Pool	Grocery Store	General Entertainment	Gas Station	Fried Chicken Joint	Fast Food Restaurant

a) Top 10 venues in the Scarborough neighborhood

Out[60]:

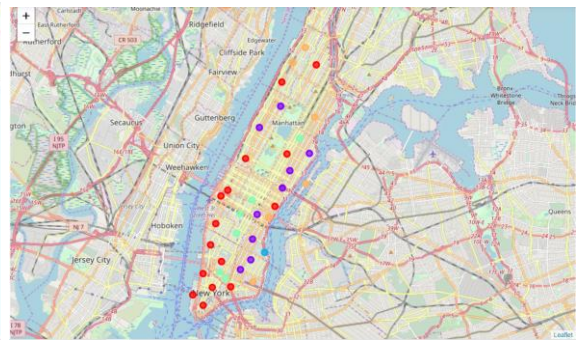
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Gym	Clothing Store	Wine Shop	Women's Store	Memorial Site	Boat or Ferry	Burger Joint
1	Carnegie Hill	Coffee Shop	Pizza Place	Cosmetics Shop	Bookstore	Bakery	Gym	Japanese Restaurant	Yoga Studio	Café	Wine Shop
2	Central Harlem	Cosmetics Shop	African Restaurant	American Restaurant	Bar	French Restaurant	Seafood Restaurant	Chinese Restaurant	Café	Tapas Restaurant	Market
3	Chelsea	Coffee Shop	Bakery	Italian Restaurant	Ice Cream Shop	American Restaurant	Nightclub	Hotel	Bookstore	Market	Sushi Restaurant
4	Chinatown	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Cocktail Bar	Salon / Barbershop	Spa	Optical Shop	Bakery	Hotpot Restaurant	Asian Restaurant

b) Top 10 venues in the Manhattan neighborhood

The clustering technique was used to group similar neighborhoods together for which the KMeans clustering method was used. This model was trained on the merged data and the cluster labels were predicted by the algorithm for each neighborhood each cluster represents a category as shown in the figure below:



a) Clusters in Scarborough



b) Clusters in Manhattan

Results

Analyzing the neighborhood data and exploring the foursquare API to generate venues for the cities we found out some interesting results which would have not to be understood easily. The analysis helped us to find out similar neighborhoods according to the different venues such as coffee shops, gyms, swimming pools, etc.

Discussion

This study can be further extended to finding out similar cities as we did for neighborhoods which would help to find out which cities have similar cultural, food, social, life, etc.

Conclusion

Thus, a successful analysis between the neighborhoods has been performed with the help of location data and the foursquare API by leveraging the data analysis and machine learning techniques.