BATTLE OF NEIGHBORHOODS

NEWYORK VS TORONTO

OSAZEMEN ORIAKHI M.

JUNE 2020

CONTENT

- INTRODUCTION
- BUSINESS PROBLEM
- INTERESTS
- DATA ACQUISITION AND CLEANING
- METHODOLOGY
- RESULTS
- OBSERVATIONS AND RECOMMENDATION
- CONCLUSION

INTRODUCTION

Today Tourism is one of the pillars of the economy and the people most often visits those countries who are rich in heritage and developed enough from a foreign prospective, like friendly environment. Every city is unique in their own way and give something new. However, people have their different interests and desire a level of satisfaction from a leisure vacation. Most times, people return dissatisfied at their experience as it wasn't up to expectation. And now the information is so common regarding location of every place around the world on your fingertips which make it easier to explore. Therefore, tourists always eager to travel to different places on the basis of available information, and the comparison between various cities.

BUSINESS PROBLEM

Toronto and New York are the famous places in the world. They are diverse in many ways. Both are multicultural as well as the financial hubs of their respective countries. The purpose of this exploration is to determine which of them is the best choice for a vacation. We would achieve this by comparing between their respective boroughs; Manhattan and Central Toronto using the Foursquare API.

INTERESTS

This analysis would catch the interests of the following:

- Tourists: they would be able to get information about the city they are to travel to and they would know in advance the satisfaction to expect
- Couples: newly weds desire a nice and comfy location for their honey moon, this project would be beneficial to them in choosing a good location.
- Firms and Organizations: most firms organize vacations for their employees and would love to get the best satisfaction, hence they would benefit from this project.

DATA ACQUISITION AND CLEANING

- For Toronto case, we have extracted table of Toronto's Borough (Central Toronto) from Wikipedia page (https://www.en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- We arrange the data according to our requirements. In the arrangement phase, which applied multiple steps including but not limited to, eliminating "Not assigned" values.
- we also get the list of coordinates for each of the locations from a csv file, from this page, (http://cocl.us/Geospatial_data) after which we would sort each coordinates with their respective locations from the previous data.

DATA ACQUISITION AND CLEANING CONTD.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
1	M5N	Central Toronto	Roselawn	43.711695	-79.416936
2	M4P	Central Toronto	Davisville North	43.712751	-79.390197
3	M5P	Central Toronto	Forest Hill North & West, Forest Hill Road Park	43.696948	-79.411307
4	M4R	Central Toronto	North Toronto West, Lawrence Park	43.715383	-79.405678
5	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678
6	M4S	Central Toronto	Davisville	43.704324	-79.388790
7	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160
8	M4V	Central Toronto	Summerhill West, Rathnelly, South Hill, Forest...	43.686412	-79.400049

Fig. 1. The resulting Data-frame for Central Toronto

DATA ACQUISITION AND CLEANING CONTD.

- Also, for Manhattan, being a city in New York, we extracted its data from a JSON file, (newyork_data.json) after which we filtered out the details of Neighborhoods under Manhattan and converted it into a data frame.

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Fig. 2. The resulting Data-frame for Manhattan

DATA ACQUISITION AND CLEANING CONTD.

For data verification and further exploration, we use Foursquare API to get the coordinates of Toronto and explore its neighborhoods (for both cities). The neighborhoods are further characterized as venues and venue categories as shown below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lawrence Park	43.728020	-79.388790	Lawrence Park Ravine	43.726963	-79.394382	Park
1	Lawrence Park	43.728020	-79.388790	Zodiac Swim School	43.728532	-79.382860	Swim School
2	Lawrence Park	43.728020	-79.388790	TTC Bus #162 - Lawrence-Donway	43.728026	-79.382805	Bus Line
3	Roselawn	43.711695	-79.416936	Ceiling Champions	43.713891	-79.420702	Home Service
4	Roselawn	43.711695	-79.416936	Rosalind's Garden Oasis	43.712189	-79.411978	Garden

Fig. 3. Foursquare API data classification.

METHODOLOGY

As we have selected two cities to explore their neighborhoods. The data exploration, analysis and visualization for both cities are done in the same way but separately. The steps are summarized as follows:

- we started by visualizing both cities, using the folium library (for map generation) as well as the geopy library (to get the longitude and latitude coordinates) with the data collected from both data sources.
- we utilized the Foursquare API in exploring the neighborhoods of both cities and segmenting them. We defined a function to get the near by venues around the neighborhoods of both cities and classify them into different categories.

METHODOLOGY CONTD.

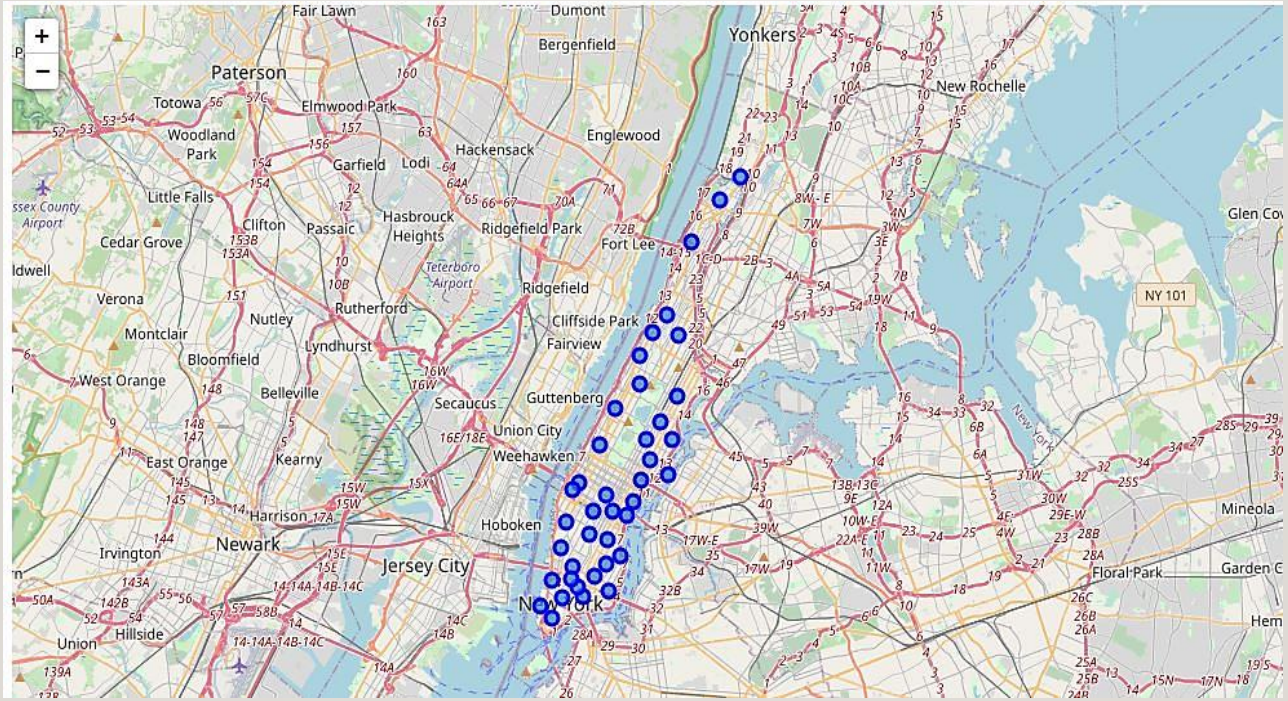

Fig. 4 Map visualization of Manhattan before clustering

METHODOLOGY CONTD.

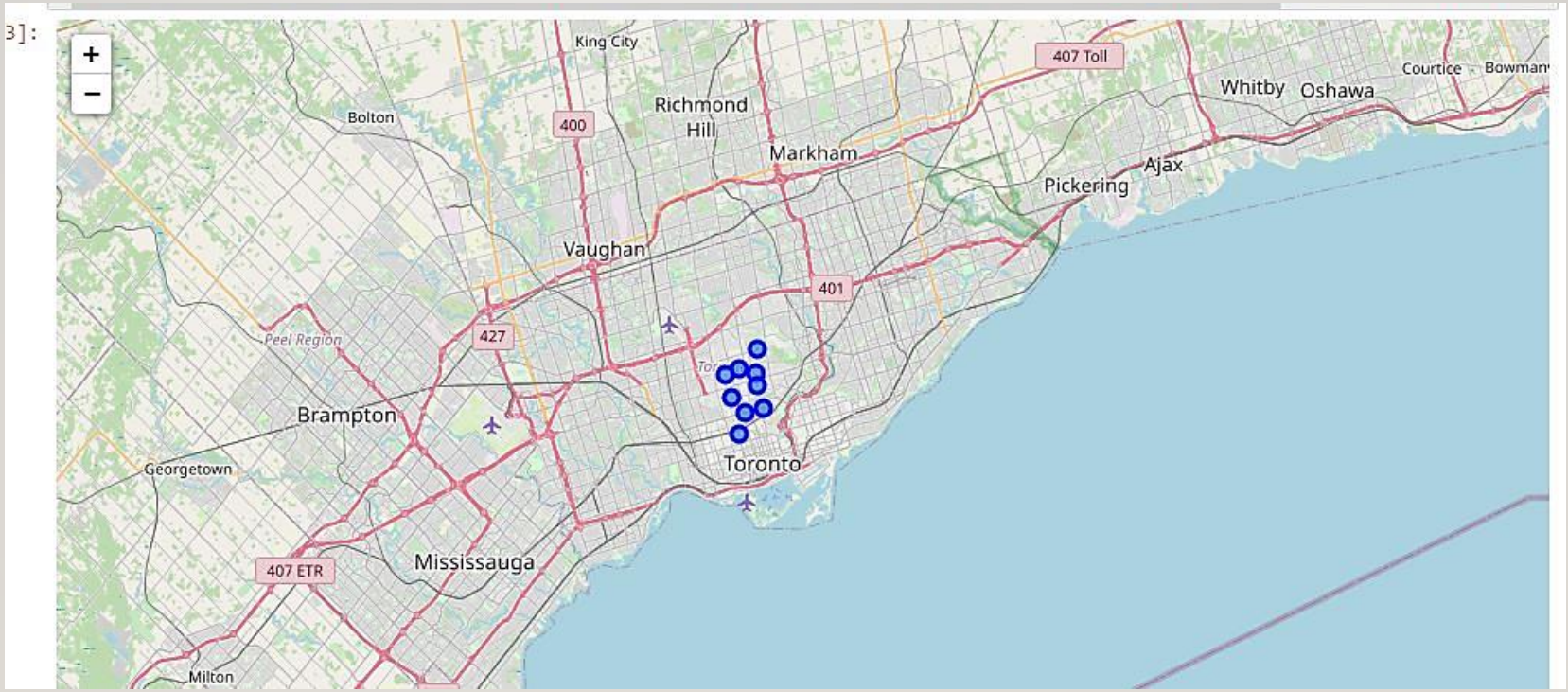

Fig. 5 Map visualization of Central Toronto before clustering

METHODOLOGY CONTD.

- We then observed that there were 64 unique categories for Central Toronto and 329 unique categories for Manhattan.
- Next, we analyzed both neighborhoods through one hot encoding (giving '1' if a venue category is there, and '0' in case of venue category is not there).
- We then calculate mean of the frequency of occurrence of each category and grouped them per neighborhood after which we picked top ten venues on that basis for each neighborhood

METHODOLOGY CONTD.

- Next, we apply the K-means Clustering machine learning algorithm to the data for proper segmentation. We set the number of clusters 'k' to 5, then we generated labels for each cluster using numbers (0 – 4). Each neighborhoods as well as its category were classified into their respective clusters and the result was displayed as a data frame, ready for visualization.

METHODOLOGY CONTD.

	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
0	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	0	Bus Line	Park	Swim School	Yoga Studio	Food & Drink Shop	Fried Chicken Joint	Garden	Gas Station
1	M5N	Central Toronto	Roselawn	43.711695	-79.416936	3	Home Service	Music Venue	Garden	Yoga Studio	Ice Cream Shop	History Museum	Gym / Fitness Center	Gym
2	M4P	Central Toronto	Davisville North	43.712751	-79.390197	1	Pizza Place	Hotel	Gym / Fitness Center	Gym	Breakfast Spot	Sandwich Place	Food & Drink Shop	Department Store
3	M5P	Central Toronto	Forest Hill North & West, Forest Hill Road Park	43.696948	-79.411307	4	Mexican Restaurant	Trail	Jewelry Store	Sushi Restaurant	Yoga Studio	Fast Food Restaurant	Food & Drink Shop	Fried Chicken Joint
4	M4R	Central Toronto	North Toronto West, Lawrence Park	43.715383	-79.405678	1	Coffee Shop	Clothing Store	Yoga Studio	Ice Cream Shop	Gift Shop	Italian Restaurant	Metro Station	Mexican Restaurant

Fig. 6. Resulting Data frame after Clustering

RESULTS

We used the folium library once more to visualize the resulting clustered data for both Central Toronto and Manhattan as shown in the figures below:

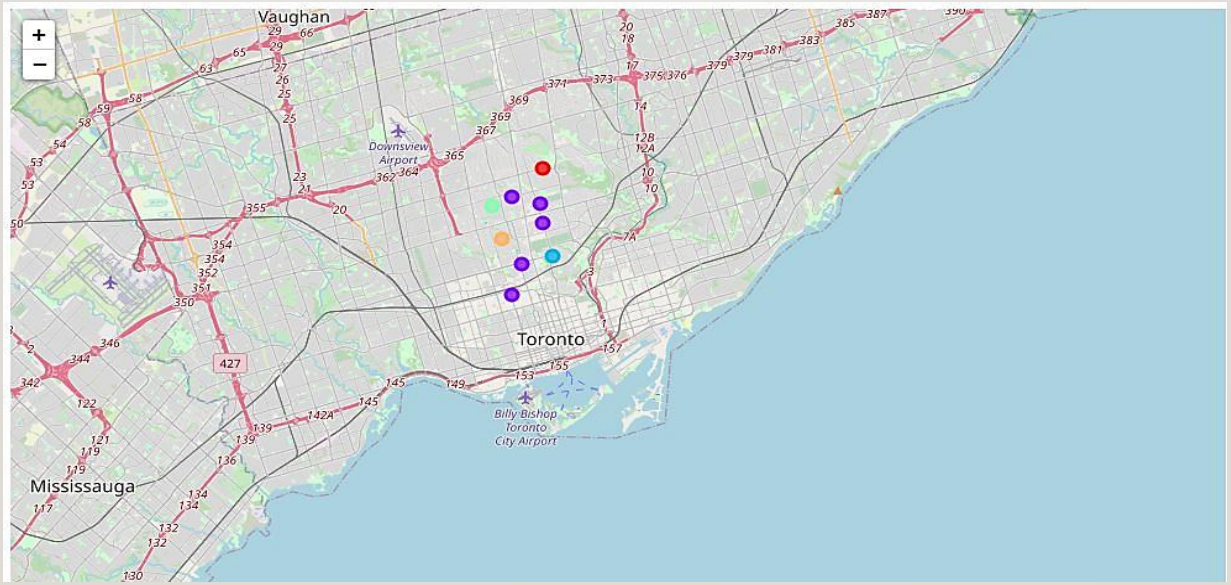

Fig. 7. Map showing the clustered results for Central Toronto. Each color represents a cluster label.

RESULTS CONTD.

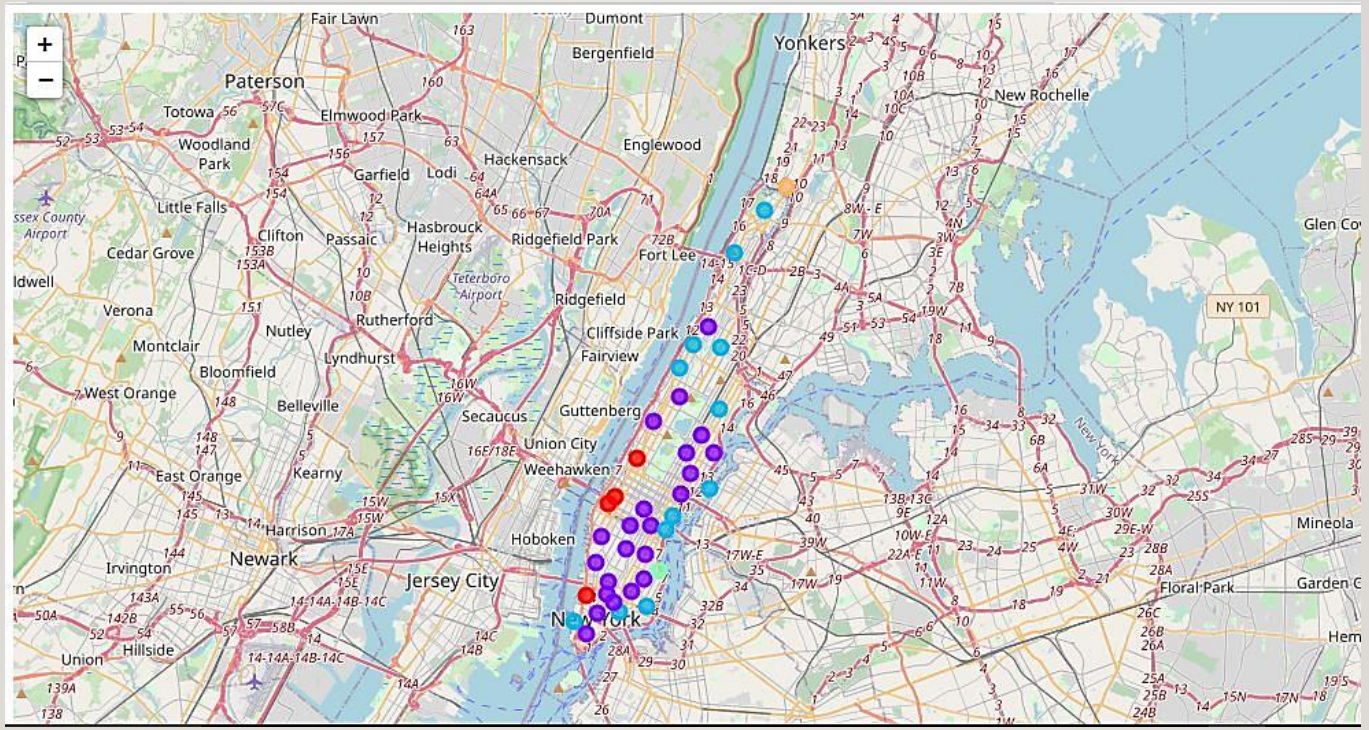

Fig. 8. Map showing the clustered results for Manhattan. Each color represents a cluster label

RESULTS CONTD.

- To summarize the results, for Central Toronto, Cluster 1 consisted of parks, Cluster 2 consisted of restaurants and joints, Cluster 3 consisted of playgrounds and leisure places, Cluster 4 consisted of snacks and fitness joints, Cluster 5 consisted of shopping hubs.
- For Manhattan, Cluster 1 consisted of restaurants and fitness joints, Cluster 2 consisted of tourist centres and hubs, Cluster 3 consisted of Travels and leisure places, Cluster 4 consisted of public places and stations, Cluster 5 consisted of Pharmacy and health centres.
- For more details, visit:

https://github.com/osazee25/Coursera_Capstone2/blob/master/BATTLE%20OF%20NEIGHBORHOODS.ipynb

OBSERVATION AND RECOMMENDATION

After clustering the data of the respective neighborhoods, it can be seen that though both cities are fun to visit. The cities are similar in that they both have restaurants, gym centres, wine shops, hotels, spa etc., but Manhattan differs more in that it has a harbour, heliport, boat or ferry (cluster3), which is good for tourism and also, movie theatres which I personally consider a perquisite. Also, Central Toronto has a bank, which is very ok for cash deposits but this rarely occurs during vacation as money carried is meant to be spent.

Based on this analysis, I would recommend Manhattan a better choice for vacation as it has the required facilities and venues for an enjoyable visit.

CONCLUSION

The Central Toronto and Manhattan neighborhoods are great venues. As we know that every place is unique in its own way, hence the appreciated function of the Foursquare API. This code can be modified for future comparisons with other Boroughs in various other cities with their location data.

THANK YOU

FOR VIEWING