Introduction/Business Problem

Two of the world's most popular tourist locations are Toronto and New York City. In many respects, they are diverse, with each neighborhood within those cities presenting different types of food and culture. The two cities are multicultural in nature, and they are their respective countries ' financial hubs as well. Our desire is to explore how similar or different these two cities are in terms of food, lodging, beautiful locations/attractions, and several other tourist criteria.

Today, tourism is one of the economic pillars of the world for most countries. People most frequently visit countries that are rich in heritage and well established from an international viewpoint, either economically or environmentally. Every city is unique and offers something different and exciting, especially for tourists. And nowadays, data pertaining to the sights and locations of everything on the planet is just a few clicks of a button away, making it easier and more accessible to explore than ever before. This crucial travel data is one of the most powerful tools that a tourist can have when it comes deciding where they would like to travel, as well as comparing multiple different locations to each other in order to see the unique qualities of each destination.

But the data must be gathered and compiled first if the tourists hope to use it one day. Stakeholders would be interested in this project as it is a great basis for developing a program that can be used to compare different travel destinations and weigh their pros and cons as well as determining what is unique about what each destination offers. By being able to efficiently and effectively gather and analyze geographic data from Foursquare, one may be able to achieve this. My target audience should therefore be tourists and tourist/travel agencies. The more information a tourist is given about a location, the more likely they might plan to actually go there someday.

Data

We will use the Foursquare API service to gather and compile data about Toronto and New York. We will group this data in terms of each of the neighborhoods within the cities. This data will include information about the places around each neighborhood such as restaurants, hotels, coffee shops, parks, theaters, art galleries, museums, etc. Within each city, we will select one borough from each city to narrow down the area even more, and then we will move on to analyzing the neighborhoods within that borough. I have chosen to go with Manhattan from New York and Downtown Toronto from Toronto. We will use the machine learning technique known as "Clustering" to cluster the neighborhoods with similar amenities based on the neighborhood data we gather. These amenities will be given priority on the basis of amount of foot traffic (activity) in their respective neighborhoods. This will help to locate the tourist's areas and hubs and where people tend to go the most. Then we can judge the similarity or dissimilarity between the two cities based on that.

An example of the data that we might collect, and compile is a Pandas dataframe of the different types of venues within a particular neighborhood of one of the boroughs we chose, as well as the frequency at which those venues appear. By using that, we can easily determine what the most popular and apparent venues are in that region, and then give that information to the tourist so that they can make an educated decision as to where they would want to go.

Methodology

Although I have decided to explore and analyze both New York City and Toronto, the data exploration, analysis, and visualization for both boroughs will be done in relatively the same manner, just in separate sections. There is no need to change up the analysis techniques if all the data from both cities are in the same format for the most part. This helps keeps the project more consistent as well.

For the case of Downtown Toronto, we have extracted a table from the Wikipedia page of Toronto. Then we arranged the data according to our goal of grouping them by neighborhoods in order to suit our needs. We then cleaned up the data to make it more useable and efficient, which included eliminating "Not assigned" values in the Neighborhood column to prevent null values. I also combined neighborhoods that had the same geographical coordinates and sorted against the borough they were in. For increased data verification and further in-depth exploration of the neighborhoods, I used Foursquare API service to get the geographic coordinates of Downtown Toronto and its neighborhoods. I then used this geographic data to further explore and document each neighborhood and its venues and the categories of venues that were present.

For Manhattan, I used a publicly available JSON file that used the foursquare API service to extract all the necessary information about the neighborhoods of New York that I needed, and then I sorted against the borough that the neighborhoods were in as well. I then used this geographic data to further explore and document each neighborhood in Manhattan and its venues and the categories of venues that were present as well.

I continually visualized the data in dataframes/tables as I progressively worked with the data n order to always have an idea of what I was working with. I also created two maps of each city to further my understanding and to give a better visualization of the data in general. The first map of each city was just created using the geographic data of each neighborhood so I could get a good idea of the layout of each neighborhood in relation to one another, as well as an idea of the layout of the borough as a whole. The second map for each city added cluster names and marks so I could better understand how I was going to apply clustering and where these clusters would be located and how they would relate to each other. Later on after I perform the clustering, I create a third map of each city that displays which cluster each neighborhood is in.

I also analyzed each neighborhood by using one hot encoding (which is basically giving a '1' if a venue category is there, or a '0' if a venue category is not there). Based on one hot encoding, we then calculate the mean of the frequency of occurrences in each category. After that, I found the

top ten venues for each neighborhood based on that data. The rating of the top venues are based upon the amount of foot traffic they receive (how much they are visited).

After successfully clustering the neighborhoods of each borough base upon similar venues, I then examined each cluster and looked at their discriminating differences/defining venue categories. Using that information, I gave each cluster a name.

Results

After looking at and analyzing the data of all the clusters of both Toronto and New York, one can easily ascertain that both city boroughs are diverse tourist regions that have a wide variety of tourist venues and attractions. Most of the neighborhoods are fairly similar in regard to what they generally offer such as theaters, restaurants, etc. The only real sources of dissimilarity between the two cities are the unique locations like historical locations, monuments, or landmarks.

Observations and Recommendations

When comparing the two cities, it is clear that both of them offer a very wide variety of differing attractions and venues that could appeal to an increasingly wider variety of tourists. However in spite of that, there are some minute differences between the two cities that tourists should take into account before planning to travel to either city.

When one looks at the clusters, one can observe that Manhattan seems to have an abundance of restaurants and good eats in general, so if someone prioritized the food aspect of their traveling experience, Manhattan could be recommended more than Toronto in that case.

Toronto on the other hand has easier direct access to the airport, as well as a greater abundance of cafes, gastropubs, and other amenities of that sort. Therefore, it is relatively easier to travel there as well as relax there with an assortment of drinks, etc

So all in all, I would recommend that the tourists looking for more intense and involved travel would turn their eyes towards Manhattan, while the people looking for more relaxed travel might consider leaning towards Toronto.

Conclusion

After considering everything that has been given to us, all in all, one could say that both Toronto and Manhattan are great cities filled with a plethora of interesting and appealing venues for

tourists of all kinds. The only differences that really exist between the two cities are minute differences such as assorted historical places, etc. If a tourist were to travel to either of these cities in search of fun and exciting travel, they would be far from disappointed.