Curtis Arnold

Capstone Project

Introduction/Business Problem

After examining New York City and Toronto in the previous week, it would be interesting to compare the most popular venues of the two cities to see if their relatively close distance provides more similarities or differences between them. This would prove helpful for any potential business owners looking to open up a venue in the most ideal location in the region. Additionally, it would provide some insight into how much impact the culture of the US has on New York and similarly with Canada and Toronto.

Data

Using the data provided for the neighborhoods of New York and Toronto (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for the Toronto data, and https://cocl.us/new_york_dataset for the New York data) the data containing information about the neighborhoods that form New York and Toronto can be gathered and placed into two separate dataframes. Additionally, the necessary latitude and longitude for Toronto can be gathered by using the existing dataset (https://cocl.us/Geospatial_data) and then merged with the existing Toronto dataframe to generate the necessary data for Foursquare in order to determine the most popular venues of New York and Toronto.

Methodology

For the New York data, the dataset was downloaded and opened as a json file. I then built a dataframe to sort the borough, neighborhood, latitude and

longitude columns from the data. Having done that, I was able to use the “shape” command to find the number of boroughs and neighborhoods of New York, as well as get the latitude and longitude coordinates to map the data. Using a folium map, I was able to get the coordinates for all of the neighborhoods of New York. With this data, I was able to then prepare to run the data through Foursquare to get the venue data.

Using Foursquare, I started by gathering the latitude and longitude for the “Wakefield” neighborhood. Next, I was able to gather the nearby venues for the neighborhood, which allowed me to then use to loop the process to find the venues for the remaining neighborhoods in New York. Using the venue data gathered by Foursquare, I was able to group and rank the top ten most frequented venues in each neighborhood. Finally, I clustered and mapped the neighborhoods of New York, allowing me to collect the most popular venues of the five clusters.

Similar to the New York data, I was able to scrape the Toronto data from the Wikipedia page above using the BeautifulSoup package. I then created a dataframe to compile the Toronto data, removing any missing values from the data. I then merged the dataframe containing the latitude and longitude values with the original dataframe, allowing me to then map the neighborhoods of Toronto with a folium map. Using the same method as with the New York data, I was able to use Foursquare to gather, group and rank the popularity of the venues in each

neighborhood. This then allowed me to cluster and map the most popular venues of each neighborhood.

Results

After clustering the venues of both cities, I was able to find some interesting results about the clusters of both cities. Looking at the results for New York, pizza places, bars, and donut shops appear to compose much of the top ten venues for cluster 0(red). Conversely, Clusters 1(purple) and 2(blue) seem to lean more favorably towards parks. Clusters 3(green) and 4(orange), while being located on the outer parts of the city like 1 and 2, appear to lean more towards similar venues as Cluster 0. Based on the results of the Toronto data, pizza venues appear to be a standout in Cluster 1(red), with Italian, Chinese and other foreign restaurants also appearing prominently in the cluster. This appears to coincide with the "Downtown" borough being the most heavily clustered part of the city. Clusters 1(purple), 2(blue), 3(green) and 4(orange), while more spread out than parts of Cluster 0, also lean favorably towards pizza places.

Discussion

New York appears to lean favorably all over towards pizza places, with bars and donut shops also thriving throughout the city. Conversely, venues such as parks appear to be more favorable when located in the outer regions of the city,

Drawing on this, the more congregated parts of Toronto seem to be hotspots for foreign restaurants or pizza places to thrive in, while smaller places such as donut shops appear to be more popular in the more sparsely packed regions of Toronto. Based on this, a larger restaurant would likely do well around the "Downtown" borough, while a deli/bodega or non-culinary venue (such as a bus station) would probably benefit slightly more from being further out in the city. Similarly, a pizza, bar or donut shop business in New York would do well in almost any part of New York. In contrast, a playground, park or bus stop would find more popularity in the outer sections of New York.

Conclusion

Despite being a part of two different countries, New York and Toronto seem to share a similar composition of the most popular venues, with pizza places thriving in both. Similarly, both cities appear to be hotspots for foreign restaurants to do well in, with Toronto being the most ideal. In addition to this, New York and Toronto also appear to have increased popularity for smaller venues on the outer sections of both cities. As such, businesses would find success in either city with a foreign restaurant or pizza place in the central parts of the city, and a smaller venue (like a bus station) in the outer parts.