Predicting the Optimal Brewery Location

Alexa Lodise

March 29, 2019

Introduction

Breweries have become a popular choice for where people in the cities hang out after work and on the weekends. That being said, in the right city, opening a brewery can be a lucrative business. However, it is difficult to tell what city and where in the city to open one. Using data analysis and clustering, I will determine whether someone should open their new brewery in Toronto or New York City as well as where in the resulting city they should open it.

Data

In order to solve this problem, I will need to gather coordinates for various neighborhoods in both New York City and Toronto. The New York City data is from NYU and the Toronto data is from Wikipedia. Once this data is converted into clean data frames, Foursquare will be utilized to find where breweries are in each city. I will choose the city based on which one has the least number of breweries. Once the city is picked, I will determine which neighborhood has the least number of breweries using clustering and will then recommend that area for a new business.

Methodology

The analysis of where a business person should place a brewery will began with analyzing the number of breweries that are in each city as well as visualizing how close together they are. Once that is complete, I can decide which city I will run a cluster analysis on. This cluster analysis will determine which neighborhood to place a brewery in.

Coordinates were gathered for NYC neighborhoods by first loading data into a dataframe from an NYU dataset. The coordinates for New York City itself were found by using a geolocator package. Data for neighborhoods Toronto was discovered by loading a Wikipedia article into a dataframe with the Beautiful Soup package. This data had to be cleaned by filtering out “not assigned” data points and by combining duplicate zip codes. Coordinates for each neighborhood were loaded from an excel sheet and then merged into the current Toronto dataframe. The geolocator was also utilized for the coordinates of Toronto.

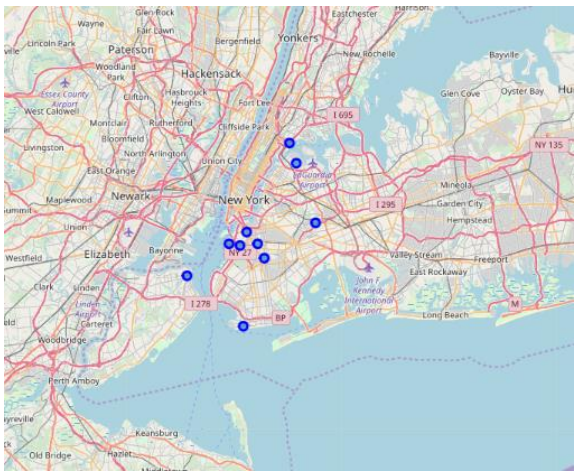
Venue data for each neighborhood was gathered from Foursquare for both NYC and Toronto. These dataframes were then filtered out for only brewery venues. Once this was complete, the number of breweries were found and the locations were mapped for both cities using Folium.

It was decided to go with NYC for clustering, so this venue data was clustered into 7 clusters. These clusters were then mapped with Folium and the breweries in each cluster were analyzed.

Results

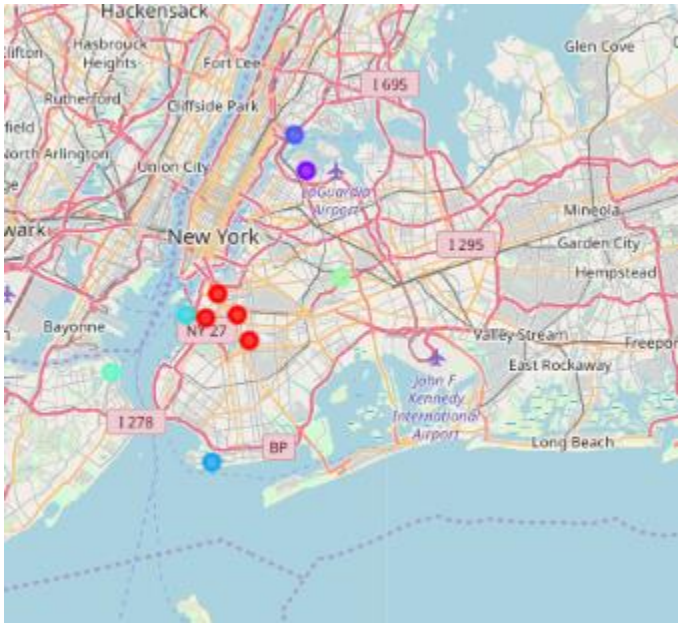
Toronto was found to have 8 breweries and NYC was found to have 10 breweries.

NYC breweries:



Toronto breweries:

Clustered NYC breweries:



Discussion

It was chosen to perform a cluster analysis on NYC, even though it had more total breweries, because they were more spread out and it would mean more areas to put a new brewery in. The cluster analysis on NYC breweries only showed 1 large cluster with 4 breweries. The other 6 clusters only had one brewery which means that there are lots of options to place a brewery except those 4 neighborhoods which included Prospect Heights, Gowanus, Downtown, or Prospect Lefferts Gardens.

Conclusion

The main purpose of this project was to recommend a city to place a brewery in as well as where in the city to put it. This analysis concluded that placing a brewery in New York in any neighborhood, besides

Prospect Heights, Gowanus, Downtown, or Prospect Lefferts Gardens, will provide a great business opportunity for a new brewery. This is a great outcome because it gives a new business owner the opportunity to pick based on other criteria such as cost of location, taxes, and other venues surrounding the neighborhood. This report recommends that further analysis on other neighborhood attributes should be completed on the other 6 brewery locations.