Capstone Project Final Report

Introduction/Business Problem

I own a fictional travel agency, and a client of mine is looking for a good destination for him and his wife. They want to travel to a city close to them, and since they live in Philadelphia, their two options are New York City and Toronto. Since both cities are big metropolises with storied histories and lots of things to see and do, it is a hard decision to make. However, their main issue is food, as they are vegan and don't always find good or unique vegan food on their travels. As a result, they want to decide based on which city is more vegan friendly, and therefore would be more welcoming for their dietary restrictions.

Data

I will be using the provided NYC data from our lab, and the Toronto data that I grabbed for the previous assignment. Each dataset should have the neighborhood, the borough, and the lat-long coordinates, which I will use in conjunction with the Foursquare API to return a list of venues in each city.

Table

Description automatically generated with low confidence

Graphical user interface, text, application

Description automatically generated

**Above are snapshots of lists of each neighborhood, Borough, Latitude and Longitude in New York City and Toronto respectively.**

The data returned from the API will contain the Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude, Venue Category for each City.

A screenshot of a computer

Description automatically generated with medium confidence

Then I will reduce that down to a list of all the vegetarian/vegan restaurants in the area, and then map the locations for each venue in each city, to see not only which city has more vegan restaurants, but also which city has more variety.

Methodology

Step 1:

Grabbing the raw data and parsing it into a usable data frame.

Firstly, I will use Beautiful Soup to grab the data from the respective URLS, and then pandas to parse the raw data into the tables above. For each source the process is slightly different.

The NYC data exists as a JSON file, so we will need to parse the JSON and then load the data into a pandas data frame.

For the Toronto data, we are grabbing the neighborhood information from the Wikipedia page by finding out which unique identifier we can use to grab the table information, and then parse that into one data frame. The Lat and Long data are in another data set provided by the course instructors, and to make the complete data frame, we need to merge the two based on the postal code for each neighborhood in Toronto.

Then we will need to clean the data further to remove postal codes that are not in use or are reported incorrectly.

Step 1 should result in the following tables.

Table

Description automatically generated with low confidence

Graphical user interface, text, application

Description automatically generated

Step 2:

Using the Foursquare Places API to grab the list of venues of each city.

A screenshot of a computer

Description automatically generated with medium confidenceAfter prepping the Foursquare Credentials to grab the venue data for each city, we use the getNearbyVenues function that we used in the lab to grab the list of venues for each city.

**NOTE: if there is a key error: ‘groups’ when running the code in the notebook, it might be because the number of calls I can make with my free foursquare account is limited and as a result the quota might have been exceeded. I couldn’t post photos of my venue data nor my map data as a result.**

This will result in two data frames containing the venues for each city. This is a lot of information to deal with, and to figure out which venues are vegan/vegetarian restaurants, we need to filter the data based on the venue category parameter.

Graphical user interface, website

Description automatically generated

Map

Description automatically generatedOnce we have the filtered data, we then use the folium map package to first generate a map of each city, and then map the data points on each of the city maps.

**Map of New York City with Vegan Restaurants marked.**

Then we look at the marked points and based on those points see which city has more vegan/vegetarian restaurants.

Map

Description automatically generated

**Map of Toronto with Vegan Restaurants marked.**

Results

Based on the number of restaurants and the map visualization of this data, the clear winner in this case is New York City, and I would recommend New York City to my clients for travelling.

Discussion

When I was looking at the data, I realized that because I only focused on purely vegan/vegetarian restaurants and didn’t include restaurants which might have vegan/vegetarian options in addition to their non-vegetarian options, I might have eliminated some viable options for my clients. However, based on the sheer number of vegetarian options, I believe that even if they wanted a little more variety, there are a ton of options on the list and that NYC is very, very vegan friendly.

Conclusion

I can confidently recommend New York City as the better option for my clients to travel to. For the future and can use this approach to evaluate other cities for vegan friendliness and travel viability. I could also adjust the approach to check for restaurants that match the customers preferred cuisine and/or create food-based tours where the main focus is travelling and sampling famous local joints and eateries, thus making my package options very diverse, while offering a special theme that other places might not have.