The Battle of Neighborhoods Week 2 - Finding the right neighborhood to live in New York City

INTRODUCTION

Background
People who want to come to New York City always ask the same question. What neighborhood is best to eat in New York City? My question back to them is what type of food are you looking to eat? New York City is has a very diverse food culture. I cannot say that one neighborhood's food is better than another because it depends on the individual's preference or craving at the moment. For example, if you are craving for pizza, there are many options in each neigborhoods. It depends if you are looking for a pizza place in your neighborhood or if you are willing to travel to a different neighborhood for pizza place that have a better rating.

This capstone project is to help people find places to eat in any neighbhorhood in New York City. It will also help those who are visiting New York City to decide which neighborhood to stay depending on the surrounding restaurants and cafes nearby.

Also people who are already living in New York City will benefit from this project as well because just because you live in the city that does not mean you know all the neighborhood and what kind of venues each neighborhood has to offer. The project will people explore the surrounding areas and see what kind of venues are there so they can make the best choice that best fits their needs.

Problem
There are many expat groups in New York City and a lot more people are coming to live in New York City. Each year New York City receive a great amount of tourist all over the world looking for places to eat around them and while they are traveling. For those that live in New York City already, sometime when they change neighborhood, they become unfamiliar with their environment so it takes time for them to know where to go and eat.

The more information we can provide to help people become more aware of the surrounding venues in each neighborhood, the more prepared they will be before visiting or moving to New York City.

Interest
This project is also a personal interest to me because I used to live in New York City but I do not know all the neighborhoods and what kind of venues it entails. This is why I mentioned that just because you live in New York City, that does not mean you know what each neighborhood has to offer.

Also, I can share this information with expat communities, friends that planning to visit New York City, and others that would be interested in having this information.

DATA

The data I will collect is similar to what we have been doing from the course materials. I will use the data to convert addresses into their equivalent latitude and longitude values. Also, I will use the Foursquare API to explore neighborhoods in New York City. Then I will use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. I will use the k-means clustering algorithm to complete this task and I will use the Folium library to visualize the neighborhoods in New York City and their emerging clusters.

By using all these different methods to make the data more understandable. I can make an analysis of what type of venues each neighborhood in New York City has in order to help those who need help in deciding where to live in New York City.

METHODOLOGY

Download and explore dataset
Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, I will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.

This dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

Since the file is already on the server, I can simply run a wget command and access the data.

Once the data ran, I noticed how all the relevant data is in the *features key*, which is basically a list of the neighborhoods. So, I define a new variable that includes this data and take a look at the first item in this list.

Tranform the data into a pandas dataframe
The next task is essentially transforming this data of nested Python dictionaries into a pandas dataframe. So I started by creating an empty dataframe. Then take a look at the empty dataframe to confirm that the columns are as intended.

Once columns are correct, I then loop through the data and fill the dataframe one row at a time and examine the resulting dataframe and I made sure that the dataset has all 5 boroughs and 306 neighborhoods.

Define Foursquare Credentials and Version
After getting the Foursquare credentials, I use it to explore:
- neighhorhoods in New York City
- nearby venues in the
- get first 5 venues with latitue and longitude
- get neighborhood counts of venues.
- analyse each neighborhood
- group the neighborhoods
- get most common venues
- get top 10 most common venues in each neighborhood

K-Mean
I ran k-means to cluster the neighborhood into 5 clusters, then create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

Afterward I sorted the neighborhood venues and run the code to see each cluster.

Getting the coordinates for New York City

I used the geolocator to get the coordinates for New York City. The geograpical coordinate of New York City are 40.7127281, -74.0060152.

Creaping a map of New York City
I used folium to get the map of New York City with the colored markers for each cluster.

REPORTS

For this project I used the data from this link: https://geo.nyu.edu/catalog/nyu_2451_34572 And once the data was downloaded, I opened the json file of the date and transform it into a pandas dataframe by inputing all the necessary information to create a dataframe which consists of New York City: Borough, Neighborhood, Latitude, and Longitude.

Once I was satisfed with the dataframe, I called the Foursquare API to start preparing to get the dataset of venues in New York City's neighborhood. Then I explore each neighborhoods and venues with their latitude and longitude. Since there were many neighbords and venues, I did a count to see how many venues in each neighbhood, then I further analyse the data by each neighborhood and find the most common venues and limit that to the top 10 venues for each neighborhood.

Then I ran k-mean to cluster 5 neighborhoods, then I ran each cluster individually to see what kind of results it yield.

Lastly, I use the geolocator to get New York City's latitude and longitude, then created a map of New York City with different color markers.

DISCUSSION

Result
The data from New York University catalog that we used to get New York City's boroughs, neighborhoods, latitude, and longitude showed that we have 5 boroughs and 306 neighborhoods.

The data from Foursquare shows all the venues in each neighbhorhoods which yield many results that I had to only limit it to the top 10 venues for each neighborhoods and created 5 clusters of the neighborhoods.

The latitude and longitude of New York City is 40.7127281, -74.0060152.

The map of New York City shows the 5 markers with different color to identify the different cluster. The colors are: Red = Cluster 1 (clustser 0 on map), Purple = Cluster 2 (cluster 1 on map), Blue = Cluster 3 (cluster 2 on map), Green = Cluster 4 (cluster 3 on map), Orange = Cluster 5 (cluster 4 on map).

Observation
From the data, it shows that New York City has many neighborhoods that are enriched with many different venues.

In the map, cluster 1 (cluster 0,), cluster 2 (cluster1,) and cluster 3 (cluster 2,) has the most venues in their area.

Recommendation
If you are visiting New York City or moving to different neighborhood, the best borough to stay that have a lot of venues are within Manhattan and Brooklyn. There you cannot go wrong in finding any venues that suit your needs since the neighborhoods in those two boroughs have the most venues which you will find something that satisfy you.

CONCLUSION

New York City is a very diverse place full of culture and food. Every neighborhoods are full of different restaurants, cafes, bars, etc for you to choose from and they are open early to accommodate early birds and close at mucher later time to accommodate those who work graveyard shift or for those who are just coming out of the clubs or bars.

So no matter what neighborhood you are in at anytime of day or night, there will always be venue that is open to accept you when you are hungry.