Coursera IBM Data Science Capstone Project: Exploring the common places in Bronx New York

Introduction:

For this Capstone Project, I am exploring the neighborhood of Bronx, New York City. Bronx is a borough of New York City, coterminous with Bronx County, in the United State. State of New York, the third-most densely populated county in the United States. Bronx is mainly very popular for its Bronx Zoo and Yankee Stadium.

Main Focus of the project:

The idea behind this project is to explore Bronx and see if it is a suitable place for the explorer to visit or to move to depending on the things that are available there.

Target Audience:

People who are interested in visiting Bronx or moving to Bronx are the targeted audiences.

Data:

To solve this problem, I will need below data:

- List of neighborhoods in Bronx, NY
- Latitude and Longitude of these neighborhoods.
- Venue data related to most common places in the Bronx.

Extracting the data:

- Scrapping the Bronx neighborhoods via NYU spatial data provider. **https://geo.nyu.edu/catalog/nyu_2451_34572**
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

Methodology:

First I need to get the list of neighborhoods in Bronx, NY. This is possible by extracting the list of neighborhoods from the following link available for free:

The data extraction to make a readable data frame was made by utilizing pandas html table scraping method as it is easier and more convenient to pull tabular data directly from a web page into data frame.
Foursquare website will be used to extract data about the venues available in the Bronx neighborhood. After gathering all of the required coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

Next, I used Foursquare API to extract top common venues around the neighborhoods of the Bronx Borough, NY. From Foursquare, I am able to pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. Then, I analyze each neighborhood by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare the data to perform K Mean Clustering.

Here we would look at the top 10 common places in all the neighborhoods. Using the K Mean clustering algorithm we would cluster the neighbor in five groups relating to any similarity they might have compared to each other. The clustering will make it easier for the viewer to see which place is more common for certain type of venues or entertainment places.

Results:

1) Folium map of Bronx, NY with Neighborhood

```
map_Bronx
```

Out[20]:

2) Visualizing New York with its Borough and Neighnorhoods in the map

```
Bronx_data = neighborhoods[neighborhoods['Borough'] == 'Bronx'].reset_index(drop=True)
Bronx_data.head()
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

3) Top ten common places in the Bronx as a Whole

Out[27]:

	name	categories	lat	lng
0	JungleWorld	Zoo	40.845227	-73.877181
1	African Lions	Zoo Exhibit	40.847058	-73.878024
2	Congo Gorilla Forest	Zoo	40.847774	-73.881604
3	Grizzly Corner	Zoo Exhibit	40.849023	-73.877739
4	Himalayan Highlands	Scenic Lookout	40.848404	-73.876649
5	Giraffe House	Zoo Exhibit	40.847875	-73.880127
6	Wild Asia Monorail	Zoo	40.845285	-73.875788
7	Bug Carousel	Theme Park	40.849461	-73.879659
8	World of Reptiles	Zoo Exhibit	40.849287	-73.878856
9	Zoo Store	Souvenir Shop	40.850120	-73.880204
10	River Park	Park	40.843062	-73.877520
11	Mouse House	Zoo Exhibit	40.848770	-73.879564
12	Nature Trek	Zoo Exhibit	40.850008	-73.878189
13	Pizza Italia	Pizza Place	40.845755	-73.883921
14	Polar Bear Exhibit	Zoo Exhibit	40.849218	-73.877203
15	Camel Ride	Theme Park Ride / Attraction	40.845169	-73.877202
16	Butterfly Garden	Garden	40.848829	-73.879103
17	BP	Gas Station	40.848266	-73.883594
18	Wild Asia Cafe	Fast Food Restaurant	40.845300	-73.877109
19	Wild Asia Stage	Theater	40.845294	-73.876846

4) Most common venue in Bronx with respect to its neighborhood.

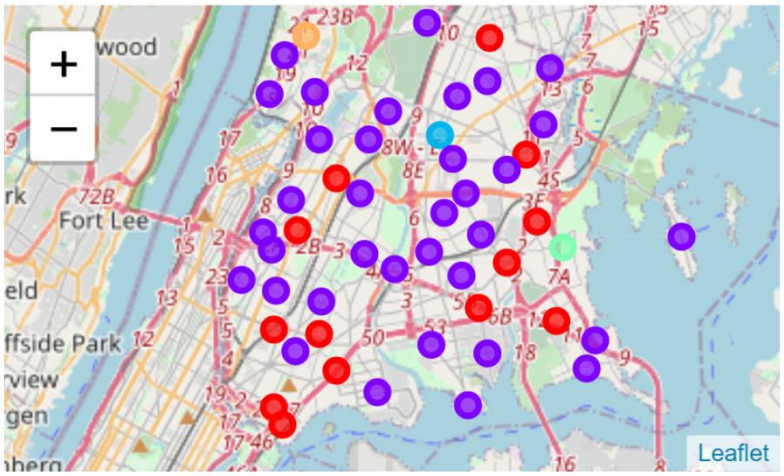
Out[41]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Mo Comm Venue
0	Allerton	Pizza Place	Deli / Bodega	Chinese Restaurant	Supermarket	Bus Station	Spanish Restaurant	Martial Arts Dojo	Pharmacy	Food
1	Baychester	Donut Shop	Pizza Place	Bank	Mattress Store	Fast Food Restaurant	Electronics Store	Spanish Restaurant	Sporting Goods Shop	Sandwi Place
2	Bedford Park	Chinese Restaurant	Diner	Pizza Place	Deli / Bodega	Mexican Restaurant	Pharmacy	Sandwich Place	Supermarket	Baseba Field
3	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Donut Shop	Dessert Shop	Bank	Grocery Store	Sandwi Place
4	Bronxdale	Pizza Place	Performing Arts Venue	Breakfast Spot	Spanish Restaurant	Eastern European Restaurant	Supermarket	Paper / Office Supplies Store	Chinese Restaurant	Bank

5) Folium map with 5 clustering of Bronx Neighborhood based on their similarity of venues available

```
map_clusters
```

Out[52]:



Conclusion:

In this project, we have gone through the process of identifying the specific propose for extracting and visualizing the data. We have extracted data, performed the machine learning by utilizing k-means clustering and providing data and visuals to people who are interested in visiting or moving to Bronx. From the above Data exploration using Four Square Website we can say that Italian Cuisine is more popular in Bronx
We can also see that some of the neighborhood in the Bronx does not have Chinese quisine as a common place. Opening a Chinese restaurant in the neighborhood might be a good investment.

Reference:

List of the data of Bronx Neighborhood:

https://geo.nyu.edu/catalog/nyu_2451_34572

Foursquare Developer Documentation: https://foursquare.com/developers/apps