The-Battle-of-the-Neighborhoods

Introduction/Business Problem

Every city is built on different circumstances and environment . I am going to compare the neighborhoods of of 2 major cities and how can they be compared with respect to places to eat, better connectivity to several useful regions and how are they distributed around both cities. The places I will be considering are airports, metro, coffee shop, restaurants, schools, college, general stores, hospitals etc. The audience will be tourists who consider travelling through financial capital of the 2 countries analyzed and search for better neighborhoods suited for their needs.

Data Section

The data used to calculate is available for New york and the other half i.e. for toronto has been been scraped for wikipedia. Both of the dataset consist of the boroughs, neighborhoods and the locations of them. The foursqaure api will be used to analyze the places nearby these neighborhoods and see the proximity of important places from the corresponding neighborhoods.

Out[23]:

	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 Most Common Venue	10th Most Common Venue
0	Battery Park City	Coffee Shop	Hotel	Park	Gym	Memorial Site	Clothing Store	Gourmet Shop	Beer Garden	Food Court	Shopping Mall
1	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Gym	Cosmetics Shop	Pizza Place	Bookstore	French Restaurant	Wine Shop	Bakery
2	Central Harlem	African Restaurant	Chinese Restaurant	Cosmetics Shop	Seafood Restaurant	Gym / Fitness Center	Bar	Art Gallery	American Restaurant	French Restaurant	Library
3	Chelsea	Coffee Shop	Art Gallery	Bakery	American Restaurant	French Restaurant	Ice Cream Shop	Wine Shop	Seafood Restaurant	Market	Bookstore
4	Chinatown	Bakery	Chinese Restaurant	Cocktail Bar	Dessert Shop	Hotpot Restaurant	Spa	American Restaurant	Bubble Tea Shop	Salon / Barbershop	Ice Cream Shop

Methodology Section

For analyzing the neighborhoods for toronto as well as manhattan first we find the neighborhoods popular places using the explore section of thr foursquare api and then for each neighborhood for both of these cities we get the top 10 most common places and try to visualize then to get the similarities among them. Then, from the scikit learn library we use the k-means algorithms to cluster the places and see similarities for different neighborhoods for each city and find similarities among them. The visulization for all both cities is done through the folium library. To find similarities a unsupervised learning is used which helps in clustering places and is quite effective.

```
In [1]: #importing modules
        import numpy as np # library to handle data in a vectorized manner

        import pandas as pd # library for data analsysis
        pd.set_option('display.max_columns', None)
        pd.set_option('display.max_rows', None)

        import json # library to handle JSON files

        !conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the Foursquare API lab
        from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

        import requests # library to handle requests
        from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

        # Matplotlib and associated plotting modules
        import matplotlib.cm as cm
        import matplotlib.colors as colors

        # import k-means from clustering stage
        from sklearn.cluster import KMeans

        !conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't completed the Foursquare API lab
        import folium # map rendering library
        #importing modules
        import pandas as pd
        import numpy as np
        from bs4 import BeautifulSoup
        import requests
        !conda install -c conda-forge folium=0.5.0
        import folium

        print('Libraries imported.')
```

Out[27]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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 M Com Venu
0	Manhattan	Marble Hill	40.876551	-73.910660	3	Gym	Discount Store	Sandwich Place	Coffee Shop	Yoga Studio	Pizza Place	Steakhouse	Shop Mall
1	Manhattan	Chinatown	40.715618	-73.994279	0	Bakery	Chinese Restaurant	Cocktail Bar	Dessert Shop	Hotpot Restaurant	Spa	American Restaurant	Bubb Shop
2	Manhattan	Washington Heights	40.851903	-73.936900	0	Café	Bakery	Mobile Phone Shop	Bank	Grocery Store	Deli / Bodega	Gym	Latin Amer Rest:
3	Manhattan	Inwood	40.867684	-73.921210	0	Mexican Restaurant	Lounge	Restaurant	Café	Caribbean Restaurant	Bakery	Chinese Restaurant	Park
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Deli / Bodega	Bakery	Park	Cock Bar
5	Manhattan	Manhattanville	40.816934	-73.957385	1	Coffee Shop	Seafood Restaurant	Deli / Bodega	Sushi Restaurant	Italian Restaurant	Chinese Restaurant	Mexican Restaurant	Bus S
6	Manhattan	Central Harlem	40.815976	-73.943211	1	African Restaurant	Chinese Restaurant	Cosmetics Shop	Seafood Restaurant	Gym / Fitness Center	Bar	Art Gallery	Amer Rest:
7	Manhattan	East Harlem	40.792249	-73.944182	3	Mexican Restaurant	Bakery	Deli / Bodega	Thai Restaurant	Sandwich Place	Spa	Latin American Restaurant	Rest:
8	Manhattan	Upper East Side	40.775639	-73.960508	0	Coffee Shop	Italian Restaurant	Bakery	Exhibit	Gym / Fitness Center	American Restaurant	Yoga Studio	Spa
9	Manhattan	Yorkville	40.775930	-73.947118	1	Italian Restaurant	Coffee Shop	Gym	Deli / Bodega	Sushi Restaurant	Bar	Wine Shop	Japa Rest: Gym

Result Section

The results which were found out after visualization the neighborhoods of both cities and applying unsupervised learning algorithm i.e. K-means clustering. The results obtained after applying the K-means algorithm on both the dataset was there was a similarity for both cities first few common places in the 1st cluster. The common places for both the cities had coffee shops, cafe's and restaurants for certain clusters and others for different clusters. The analysis for different clusters are shown in the end of the Battle of Neighborhoods notebook. Both cites common places are compared among different clusters and some similarities in distributions is apparent in both the cities.

Discussion Section

Certain similarities are observed which is the top ten common place similarity is calculated through K-means and according to those 10 most common places does the clusters form. For example, Cluster 0 for New York and Cluster 3 for toronto has first two places pretty much same as coffee shops and cafe's yet they are not classified in the same cluster as the remaining places differ very extremely.

Conclusion Section

After analyzing both the cities most common places by applying machine learning algorithm (K-means clustering), I realized that many common places are present in both the cities but are distributed differently for both them and it's the nature and circumstances as well as arrangement which leads to these different distribution in both toronto as well as New York.

Out[31]:

	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 Most Common Venue	10th Most Common Venue
1	Chinatown	Bakery	Chinese Restaurant	Cocktail Bar	Dessert Shop	Hotpot Restaurant	Spa	American Restaurant	Bubble Tea Shop	Salon / Barbershop	Ice Cream Shop
2	Washington Heights	Café	Bakery	Mobile Phone Shop	Bank	Grocery Store	Deli / Bodega	Gym	Latin American Restaurant	Tapas Restaurant	Italian Restaurant
3	Inwood	Mexican Restaurant	Lounge	Restaurant	Café	Caribbean Restaurant	Bakery	Chinese Restaurant	Park	Pizza Place	Wine Bar
4	Hamilton Heights	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Deli / Bodega	Bakery	Park	Cocktail Bar	Sandwich Place	Chinese Restaurant
8	Upper East Side	Coffee Shop	Italian Restaurant	Bakery	Exhibit	Gym / Fitness Center	American Restaurant	Yoga Studio	Spa	French Restaurant	Juice Bar
11	Roosevelt Island	Park	Playground	Outdoors & Recreation	School	Liquor Store	Supermarket	Dry Cleaner	Soccer Field	Coffee Shop	Greek Restaurant
12	Upper West Side	Italian Restaurant	Café	Bar	Bakery	Indian Restaurant	Wine Bar	Sushi Restaurant	Vegetarian / Vegan Restaurant	Breakfast Spot	Ice Cream Shop
13	Lincoln Square	Plaza	Theater	Concert Hall	Performing Arts Venue	Café	Gym / Fitness Center	Indie Movie Theater	Wine Shop	Cycle Studio	Cosmetics Shop
15	Midtown	Hotel	Coffee Shop	Steakhouse	Theater	Sporting Goods Shop	Clothing Store	Bakery	Gym	Bookstore	American Restaurant
18	Greenwich Village	Italian Restaurant	Clothing Store	Sushi Restaurant	Café	Gym	Indian Restaurant	Boutique	American Restaurant	Seafood Restaurant	Cosmetics Shop

Data extracted:

```python
#Dataframe extraction for New York
!wget -q -O 'newyork_data.json' https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json
print('Data downloaded!')
with open('newyork_data.json') as json_data:
```

```python
toronto = df[df['Borough'].str.contains('Toronto')].reset_index(drop=True)
toronto.shape
url = 'https://maps.googleapis.com/maps/api/geocode/json?key={}&address={}'.format(API_key, 'Toronto, Canada')
response = requests.get(url).json() # get response
geographical_data = response['results'][0]['geometry']['location'] # get geographical coordinates
```