Analysis of Neighborhoods in New York to open a fusion type restaurant based on different cuisines or categories in a suitable Neighborhood

Rishabh Sharma

Introduction:

USA is a large and ethnically diverse country. Its largest city New York has a long history of international immigration. Asians, Europeans, Africans, etc. make up a large amount of this population and New York is at the center of it. Cultural diversity brings along a difference of preferences, opinions, background and tastes. Our goal here is analyzing the diversity and popularity of restaurants in the neighborhoods of New York to come up with a business strategy and the most suitable neighborhood to establish the business among all the given neighborhoods. We will use Foursquare places API to find out venues and explore the neighborhoods and scrape out details in a given radius. Our goal here is finding the popular cuisines of different neighborhoods to establish a fusion style restaurant catering to different tastes of multi ethnic population in a diverse neighborhood.

Data:

We will require the neighborhood data of New York along with their location in latitudes and longitudes. The Neighborhoods in NY has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

We will use this freely available dataset on the internet: https://geo.nyu.edu/catalog/nyu_2451_34572

The dataset contains details of location of different neighborhoods under the given boroughs in json format. We transform the json to a panda's data frame. We now have a panda's data frame containing the location data of all the neighborhoods. We define a Foursquare API URI containing the credentials to make requests for the given neighborhoods using the locations of the given neighborhoods. We define and transform the data frames to select the features of our requirements. We find the top 3 popular cuisines of each neighborhoods and then combine the dataset to find the most popular cuisines overall for a fusion restaurant.

For example, while making an API call to explore the surroundings of a given neighborhood, we will collect the top venues and segregate them based upon the cuisines in a given radius around the neighborhood and make a similar dataset for all the neighborhoods.

The downloaded data will be in json format:

```
In [4]: neighborhoods_data[0]

Out[4]: {'geometry': {'coordinates': [-73.84720052054902, 40.89470517661],
   'type': 'Point'},
  'geometry_name': 'geom',
  'id': 'nyu_2451_34572.1',
  'properties': {'annoangle': 0.0,
   'annoline1': 'Wakefield',
   'annoline2': None,
   'annoline3': None,
   'bbox': [-73.84720052054902,
    40.89470517661,
    -73.84720052054902,
    40.89470517661],
   'borough': 'Bronx',
   'name': 'Wakefield',
   'stacked': 1},
  'type': 'Feature'}
```

Dataset was transformed into a pandas dataframe:

Out[6]:

	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

After gaining the data of neighborhoods and their locations, Foursquare Places API was used to gather information about the neighborhoods. Information was refined categorically in the API calls so that results only pertain to food categories.

In [126]:
```
LIMIT = 200 # limit of number of venues returned by Foursquare API

radius = 1000 # define radius

# create URL
url = 'https://api.foursquare.com/v2/venues/search?categoryId=4d4b7105d754a06374d81259&client_id={}&client_secret={}&v
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    latitude,
    longitude,
    radius,
    LIMIT)
url
```

Out[126]:
```
'https://api.foursquare.com/v2/venues/search?categoryId=4d4b7105d754a06374d81259&client_id=&client_secret=&v=20200
```

In [0]:
```
results = requests.get(url).json()
```

We make get requests and get the results in json format for the corresponding location provided. We can filter out the results by defining a limit of responses and the radius in which to search from the given location.

Analysis of the json response was done until a suitable format was found to parse the data. The corresponding json json response was transformed to:

Out[16]:

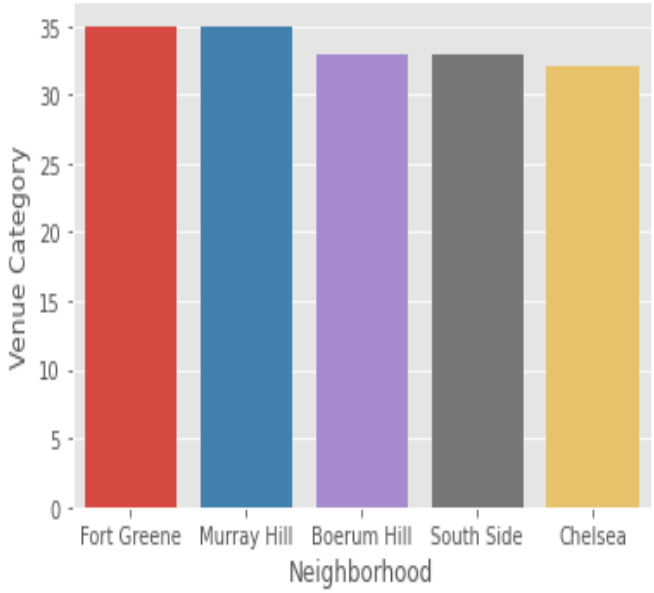
	id	name	categories	lat	lng	place categories
0	5d5f24ec09484500079aee00	Los Tacos No. 1	Taco Place	40.714267	-74.008756	4bf58dd8d48988d151941735
1	5c6f03f30802d4002c16884c	Joe's Pizza	Pizza Place	40.710318	-74.007694	4bf58dd8d48988d1ca941735
2	49ccd495f964a52091591fe3	Kaffe 1668	Coffee Shop	40.715045	-74.011509	4bf58dd8d48988d1e0931735
3	5d39f3d5052ad00008304116	Taco Bell Cantina	Taco Place	40.708428	-74.004985	4bf58dd8d48988d151941735
4	4a56a32ff964a52090b51fe3	Jubilee Marketplace	Grocery Store	40.708241	-74.006468	4bf58dd8d48988d118951735
5	5d754979ea4bf40007c080c4	Tonii's Fresh Rice Noodle	Chinese Restaurant	40.715671	-73.999074	4bf58dd8d48988d145941735
6	4b0c402af964a520c33923e3	Stage Door Delicatessen	Sandwich Place	40.711796	-74.010027	4bf58dd8d48988d1c5941735
7	4ea0afbf9adf1e334e4cc0e6	Laughing Man Coffee & Tea	Coffee Shop	40.717394	-74.010103	4bf58dd8d48988d1e0931735
8	5ded51eaf492de00080966ed	Joe's Shanghai 鹿鳴春	Shanghai Restaurant	40.715661	-73.996693	52af3b593cf9994f4e043c00
9	4b7de017f964a52049d82fe3	Starbucks	Coffee Shop	40.710922	-74.010284	4bf58dd8d48988d1e0931735
10	4afd9156f964a520a82822e3	Burger King	Fast Food Restaurant	40.709677	-74.011887	4bf58dd8d48988d16e941735
11	4c43890cfb6eb713c1304e4a	55 Fulton Market	Grocery Store	40.708678	-74.004905	4bf58dd8d48988d118951735
12	5dba113267232c000803cb37	For Five Coffee Roasters	Coffee Shop	40.712240	-74.014915	4bf58dd8d48988d1e0931735
13	4a59f31df964a52095b91fe3	Leo's Bagels	Bagel Shop	40.704935	-74.009786	4bf58dd8d48988d179941735
14	58ab3a6a53f5bb4b6b267ad1	City Acres Market	Food Court	40.706261	-74.007730	4bf58dd8d48988d120951735

When appropriate format was defined, all the locations of our neighborhoods were used to make a similar request to proceed with our data collection of eateries around New York. A data set was made having all the neighborhoods, their corresponding eateries, the category of food served along with their location.

The data was analyzed to find the most diverse location for our multi ethnic target environment.

```
In [59]: sns.barplot(x="Neighborhood", y="Venue Category", data=top5)

Out[59]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9de2d39c18>
```



We find that Fort Greene and Murray Hill offer the most diversity of all the given neighborhoods with each having 35 varieties of unique food categories in their vicinity. This environment will serve our purpose well for our Fusion Styled restaurant.

We then proceed with clustering of neighborhoods based on food categories. One hot encoding was performed on the dataset so we could standardize the operations for categorical data.

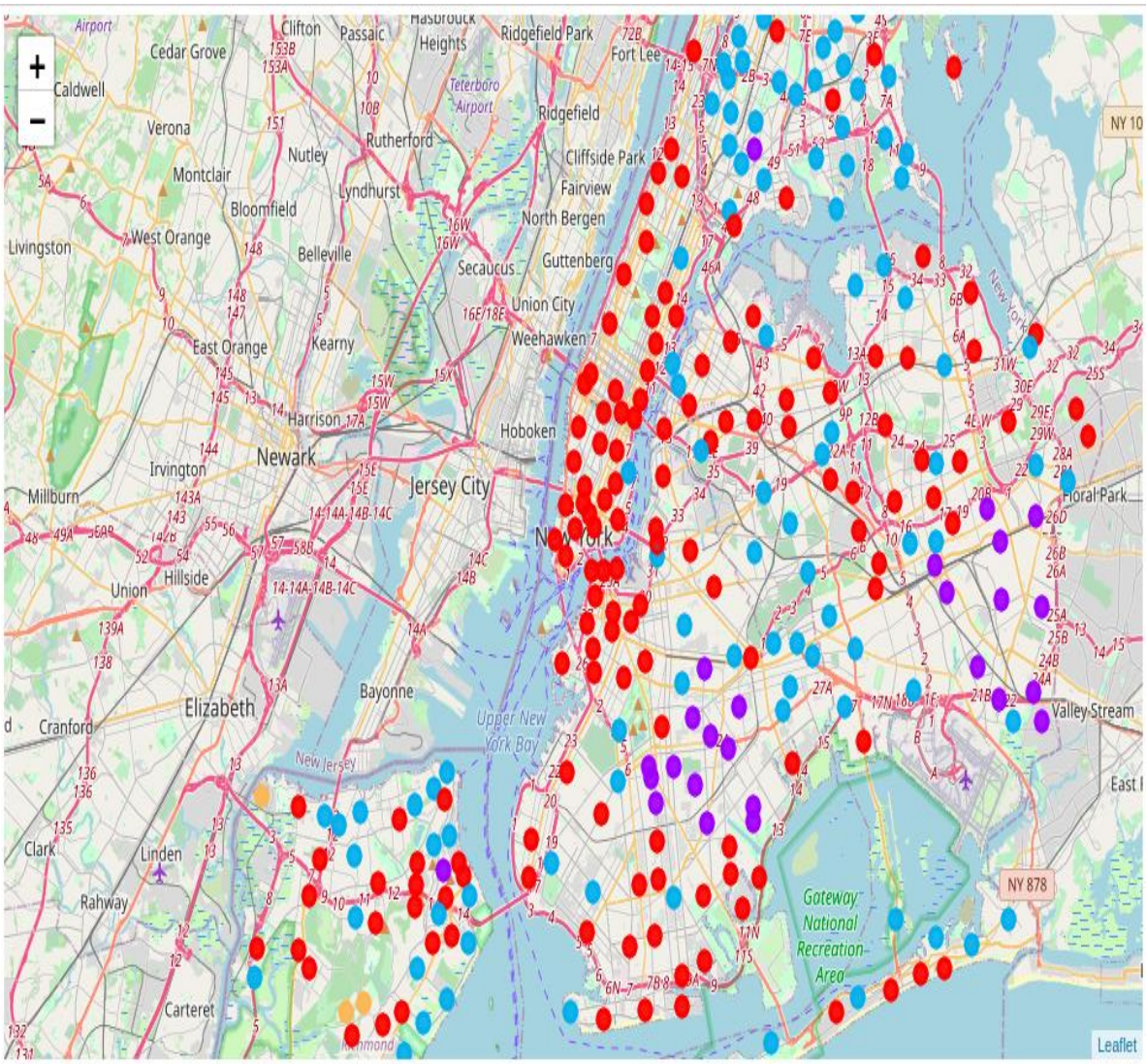
Using our dataset after performing One hot encoding, we proceed to find the top 5 categories of each neighborhood:

```
----Allerton----
                venue  freq
0       Deli / Bodega  0.20
1         Pizza Place  0.16
2  Chinese Restaurant  0.09
3                Food  0.07
4  Mexican Restaurant  0.07


----Annadale----
                 venue  freq
0          Pizza Place  0.32
1  American Restaurant  0.16
2           Restaurant  0.11
3               Bakery  0.11
4           Bagel Shop  0.11
```

After that clustering of the data was given and the clusters were plotted on the map using Folium. Five clusters were made for the dataset and based upon the top 5 venues of each neighborhoods, clustering was done.

Out[51]:



A folium map of New York City showing the clusters made from our datasets.

Further analysis was done for each of the clusters to find the top most categories in each cluster.

In [111]:
```
cl1=ny_merged.loc[ny_merged['Cluster Labels'] == 0, ny_merged.columns[[1] + list(range(5, ny_merged.shape[1]))]]
cl1.head()
```

Out[111]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	Fieldston	Fast Food Restaurant	German Restaurant	Wings Joint	Eastern European Restaurant	Fish & Chips Shop
4	Riverdale	Pizza Place	Italian Restaurant	Deli / Bodega	Japanese Restaurant	Mexican Restaurant
5	Kingsbridge	Pizza Place	Deli / Bodega	Donut Shop	Bakery	Fast Food Restaurant
6	Marble Hill	Deli / Bodega	Donut Shop	American Restaurant	Bakery	Food
8	Norwood	Pizza Place	Chinese Restaurant	Bakery	Fried Chicken Joint	Latin American Restaurant

Similar analysis was done for all the clusters and most frequent of those categories was added to our final dataset.

Out[103]:

	0	1	2	3	4
1st Most Common Venue	Coffee Shop	Caribbean Restaurant	Deli / Bodega	BBQ Joint	Italian Restaurant
2nd Most Common Venue	Pizza Place	Deli / Bodega	Pizza Place	BBQ Joint	American Restaurant
3rd Most Common Venue	Bakery	Chinese Restaurant	Chinese Restaurant	Egyptian Restaurant	Deli / Bodega
4th Most Common Venue	Pizza Place	Food	Pizza Place	Egyptian Restaurant	Diner
5th Most Common Venue	Donut Shop	Food	Food	Fish & Chips Shop	Asian Restaurant

Results:

We collected a lot of data and made API calls using Foursquare API. There were 306 neighborhoods in New York and only categorically API calls were made so that we could refine our data set. More than 9000 results were found which goes on to show that for a single city that is a lot of opportunity in this space. Our previous analysis showed the top neighborhoods in terms of food category diversity. Then we go on to make a dataset containing the location of venues along with their categories. However, to refine our results further, only the top most 5 popular places of each location was chosen and one-hot encoding was done. We cluster our datasets and then plot it on a folium map. After all the collection of venues around the neighborhoods of New York and subsequent cleanings and clustering, here is our top five common venues among the five clusters we designed.

Conclusion:

Our final results show that there is a clear diversity in tastes resulting due to multi ethnicity. Our analysis before showed that Fort Greene and Murray Hill are the most diverse categorically, so ideally, they can be the best place for a diverse group of people to offer a fusion restaurant. Next, we do an exploratory analysis to see in which clusters our neighborhoods lie.

```
In [123]: nei=cl1[cl1['Neighborhood']=='Murray Hill']

          nei
```

Out[123]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
115	Murray Hill	Korean Restaurant	Coffee Shop	Deli / Bodega	Donut Shop	Pizza Place
180	Murray Hill	Korean Restaurant	Coffee Shop	Deli / Bodega	Donut Shop	Pizza Place

```
In [124]: nei1=cl1[cl1['Neighborhood']=='Fort Greene']
          nei1
```

Out[124]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
69	Fort Greene	Italian Restaurant	Deli / Bodega	Coffee Shop	Café	Pizza Place

We can see that both the neighborhoods lie in cluster 1. So, we now have a dataset of most common food's categories across all the neighborhoods in New York and we know the neighborhoods which are delved in diversity the most.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.

The notebook can be viewed at:
https://nbviewer.jupyter.org/github/rishabhprashr/Coursera_Capstone/blob/master/battle_of_neighborhoods.ipynb