Open a Indian restaurant in Manhattan, NYC

1. Introduction:

The aim of this project is to find the most promising location to open an authentic restaurant in Manhattan , NYC
New York is home to more than 3.2 million residents born outside the United States.New York City is home to approximately 315,000 people from the Indian subcontinent.
This diversity in culture also affects the diversity of food. You would find chinese , thai , italian ,Indian restaurants all over New York City.

2. Target Audience

This project is aimed towards Entrepreneurs , business owners who want to open a new authentic Indian restaurant or expand their business in Manhattan .

3. Methodology

- We segment all neighborhoods in manhattan on types of venues present in each neighborhood. For example if we take two neighborhoods Washington Heights and Manhattan ville , the most common venues in these neighborhoods are cafes, bookstores and gyms so they might fall under the same cluster.
- So we find the cluster with the maximum number of indian restaurants , and the idea is if indian restaurants work well in most of the neighborhoods in this cluster then they will be profitable in other neighborhoods (which do not have any indian restaurant) of this cluster.

4. Data section

1. List of neighborhoods in manhattan

- Data source which contains list of neighborhoods for each borough- https://cocl.us/new_york_dataset
- We will extract all the neighborhoods with borough Manhattan
-

```
Out[5]: {'type': 'FeatureCollection',
         'totalFeatures': 306,
         'features': [{'type': 'Feature',
           'id': 'nyu_2451_34572.1',
           'geometry': {'type': 'Point',
            'coordinates': [-73.84720052054902, 40.89470517661]},
           'geometry_name': 'geom',
           'properties': {'name': 'Wakefield',
            'stacked': 1,
            'annoline1': 'Wakefield',
            'annoline2': None,
            'annoline3': None,
            'annoangle': 0.0,
            'borough': 'Bronx',
            'bbox': [-73.84720052054902,
             40.89470517661,
             -73.84720052054902,
             40.89470517661]}},
          {'type': 'Feature',
           'id': 'nyu_2451_34572.2',
```

- After extracting neighborhood names and location coordinates we convert it into a dataframe.

2 . List of venues in each neighborhoods

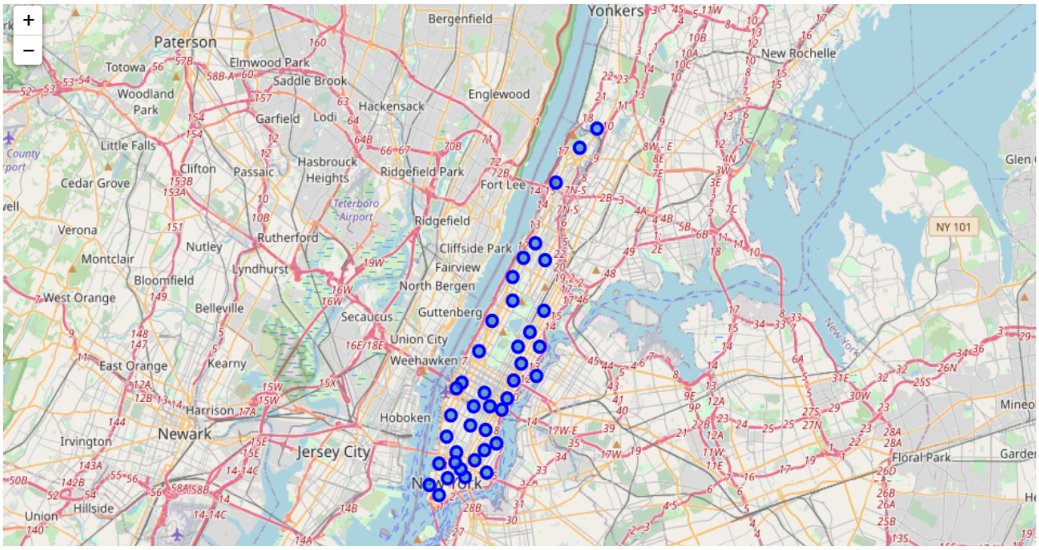
- Data source - Foursquare API
- With help foursquare API we get a json file which contains list of all venues and description of venues for all the neighborhoods

4. Data processing

1. Extract all the neighborhoods and their location in Manhattan from nyu_geojson.json and convert it into dataframe

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

- Blue dots represents all the neighborhoods in manhattan



2. Get list of venues and information about them using foursquare

- Url to get information about all the venues neighborhoods
  url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format
- I wrote functions to extract venue name , venue location and and venue categories to get the following dataframe

```
print(manhattan_venues.shape)
manhattan_venues.head(10)
```

```
(3240, 7)
```

]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
5	Marble Hill	40.876551	-73.91066	Blink Fitness	40.877271	-73.905595	Gym
6	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym
7	Marble Hill	40.876551	-73.91066	T.J. Maxx	40.877232	-73.905042	Department Store
8	Marble Hill	40.876551	-73.91066	Starbucks	40.873755	-73.908613	Coffee Shop
9	Marble Hill	40.876551	-73.91066	TCR The Club of Riverdale	40.878628	-73.914568	Tennis Stadium

5. EDA

5.1 One-hot encoding of venue categories

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athl & S
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	
1	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	
2	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	
3	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	
4	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	

ɨ let's examine the new dataframe size.

```
manhattan_onehot.shape
```

```
(3240, 332)
```

5.2 Manhattan_grouped

- Then we group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

```
manhattan_grouped = manhattan_onehot.groupby('Neighborhood').mean().reset_index()
manhattan_grouped
```

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Battery Park City	0.000000	0.000000	0.00	0.000000	0.012987	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Carnegie Hill	0.000000	0.000000	0.00	0.000000	0.010526	0.00	0.000000	0.010526	0.000000	0.021053	0.000000	0.000000
2	Central Harlem	0.000000	0.000000	0.00	0.066667	0.044444	0.00	0.000000	0.000000	0.022222	0.000000	0.000000	0.000000
3	Chelsea	0.000000	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.050000	0.000000	0.000000	0.010000
4	Chinatown	0.000000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.020000
5	Civic Center	0.000000	0.000000	0.00	0.000000	0.040000	0.01	0.000000	0.000000	0.000000	0.000000	0.000000	0.010000
6	Clinton	0.000000	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	East Harlem	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	East Village	0.000000	0.000000	0.00	0.000000	0.020000	0.00	0.000000	0.010000	0.010000	0.000000	0.010000	0.000000
9	Financial District	0.000000	0.000000	0.00	0.000000	0.050000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	Flatiron	0.000000	0.000000	0.00	0.000000	0.050000	0.00	0.000000	0.000000	0.010000	0.000000	0.010000	0.000000
11	Gramercy	0.000000	0.000000	0.00	0.000000	0.043478	0.00	0.000000	0.000000	0.010870	0.000000	0.000000	0.000000
12	Greenwich Village	0.010000	0.000000	0.00	0.000000	0.020000	0.00	0.000000	0.000000	0.010000	0.000000	0.000000	0.000000

- **Manhattan_grouped data will be used further to group neighborhoods in to cluster on bases of similar venues**

Manhattan neighborhoods + mean of frequency of indian restaurant in that neighborhood

	Borough	Neighborhood	Latitude	Longitude	Indian Restaurant
0	Manhattan	Marble Hill	40.876551	-73.910660	0.000000
1	Manhattan	Chinatown	40.715618	-73.994279	0.000000
2	Manhattan	Washington Heights	40.851903	-73.936900	0.012048
3	Manhattan	Inwood	40.867684	-73.921210	0.000000
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0.032258
5	Manhattan	Manhattanville	40.816934	-73.957385	0.022222
6	Manhattan	Central Harlem	40.815976	-73.943211	0.000000
7	Manhattan	East Harlem	40.792249	-73.944182	0.000000

Most common venues in each neighborhood

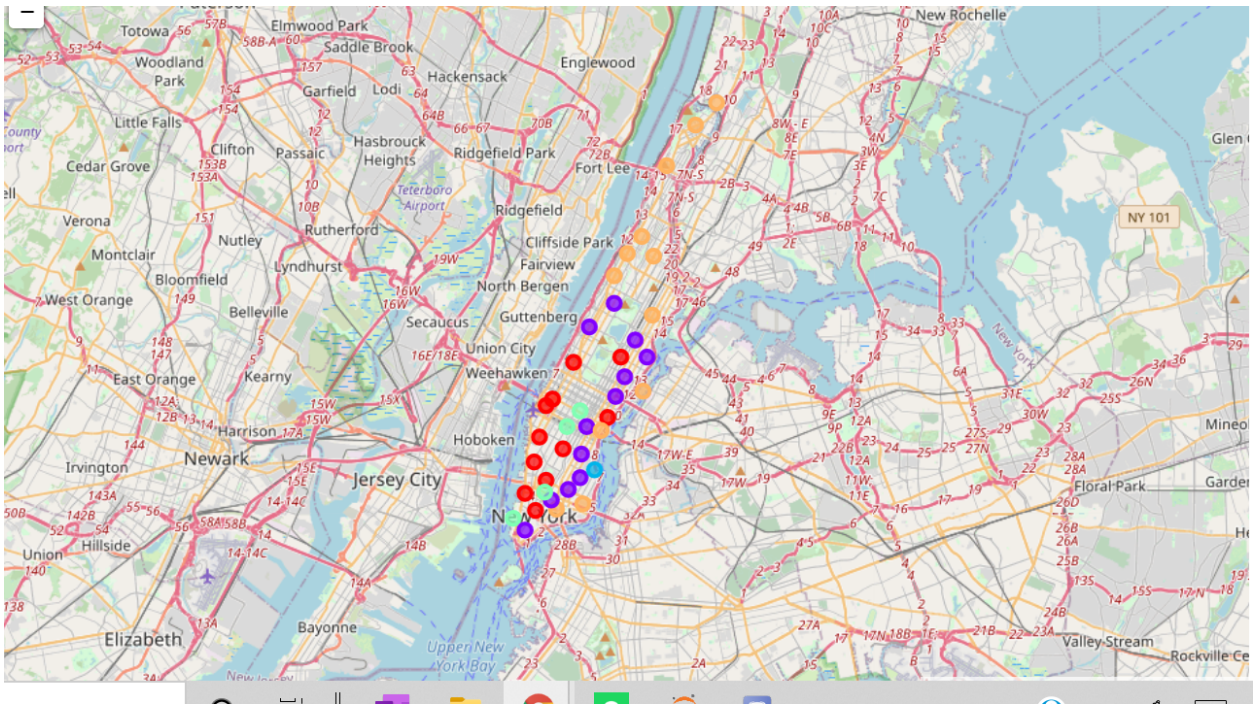
- We sort every row in Manhattan grouped and get top 10 categories foreach neighborhood

:

	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	Park	Coffee Shop	Gym	Hotel	Memorial Site	Clothing Store	BBQ Joint	Gourmet Shop	Food Court	Pizza Place
1	Carnegie Hill	Coffee Shop	Café	Bar	Yoga Studio	Gym	Bookstore	Cosmetics Shop	French Restaurant	Italian Restaurant	Wine Shop
2	Central Harlem	African Restaurant	Gym / Fitness Center	French Restaurant	Seafood Restaurant	Cosmetics Shop	Art Gallery	Bar	American Restaurant	Chinese Restaurant	Boutique
3	Chelsea	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Art Gallery	Ice Cream Shop	French Restaurant	Seafood Restaurant	Wine Shop	Bar
4	Chinatown	Chinese Restaurant	Bakery	Hotpot Restaurant	Cocktail Bar	Dessert Shop	American Restaurant	Spa	Salon / Barbershop	Shanghai Restaurant	Ice Cream Shop

6. Clustering

- We segment are neighborhoods into cluster on the basis of most common venues in each neighborhood



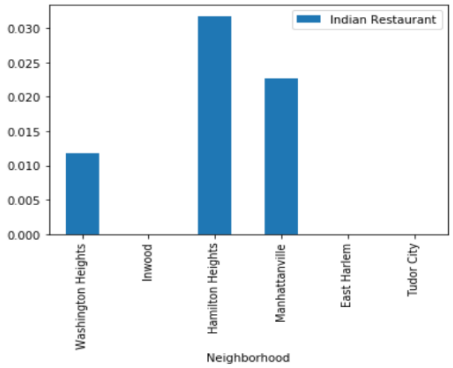
- All neighborhoods under cluster label 0 are marked red , 1 are marked purple , 2 are marked blue , 3 are marked green and 4 are marked red orange

7. Analysis The clusters / result

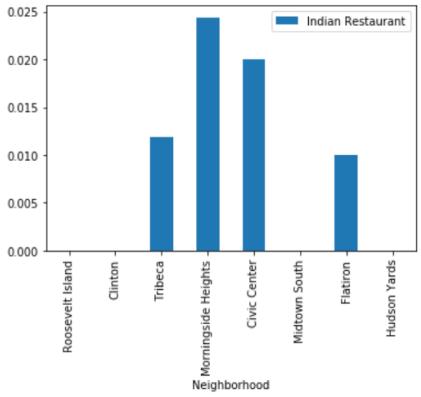
- Indian restaurants per clusters

Fig 1

- Cluster 0

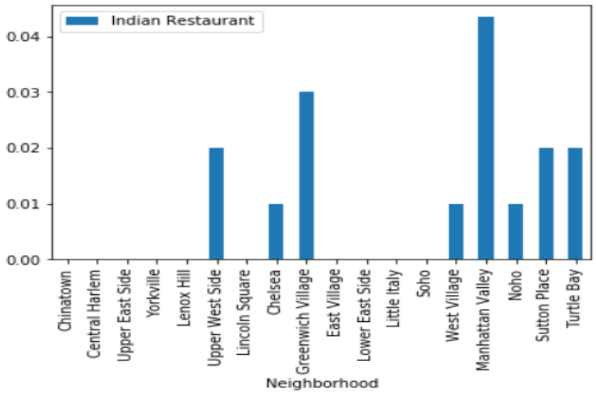


cluster 2



-

- Cluster 4



8. Discussion

- We have clustered neighborhoods based on venues in each neighborhood ,
  For example if we take two neighborhoods from the cluster 2 Washington Heights and Manhattan ville , people in both these neighborhoods enjoy cafes , bookstores and gyms as these are the most common venues in both these neighborhoods.
- From fig 1 , we can see clusters with maximum number of indian restaurants are in cluster 0 ,cluster 3 and cluster 4.
- For cluster 0 and 3 most of the neighborhoods have indian restaurants , if we open restaurants in neighborhoods from these clusters which its most probable that they are going to be profitable.
- From fig 23, you will observe that cluster forur has 18 neighborhoods , but only 8 neighborhoods have indian restaurants which is not even half , so we cannot be sure that it will be profitable to open restaurants in neighborhoods from cluster 4 .

9. Conclusion

- We will our indian restaurants neighborhoods which no indian restaurant but are part of cluster 0 or cluster 3
- If we are opening in cluster 0 - Ianwood , East Harlem ,Tudor city
- If we are opening in cluster 3- clinton , midtown south , hudson Yards.