

Introduction

New York is the most populous city in the United States . New York is also the most densely populated major city in the United States. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports . New York is home residents born outside the United States the largest foreign-born population of any city in the world .There are many languages are spoken in New York .This makes this city extremely diverse .With its diverse culture there are many food culture / restaurants available in City . As part of this project we will try to navigate different area of New York and determine which area has best Indian restaurants.

Data

For this project we will use Data source : https://cocl.us/new_york_dataset to collect boroughs and Neighborhoods information. We will use Foursquare API to collect restaurants *information* , *rating* .

Approach

We will first collect New York City data .

Using Foursquare API we will find venue for each area.

Filter out data only with Indian restaurants and find rating ,tips etc information .

We will use rating and sort the data .

Questions that can be asked using the above mentioned datasets

- What is best location in New York City for Indian Cuisine ?
- Which areas have potential Indian Restaurant Market ?
- Which all areas lack Indian Restaurants ?
- Which is the best place to stay if I prefer Indian Cuisine ?

Analysis

We will import the required libraries for python.

- pandas and numpy for handling data.

- request module for using FourSquare API.
- geopy to get co-ordinates of City of New York.

```
In [77]: #install necessary Libraries
import pandas as pd
import numpy as np
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
import requests
from bs4 import BeautifulSoup
import geocoder
import os
import folium # map rendering library
from geopy.geocoders import Nominatim # convert an address into latitude and lon
# Matplotlib and associated plotting modules
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
%matplotlib inline

print('Libraries imported.')
```

Now we define a function to get the geocodes i.e latitude and longitude of a given location using geopy.

```
def geo_location(address):
    # get geo location of address
    geolocator = Nominatim(user_agent="ny_explorer")
    location = geolocator.geocode(address)
    latitude = location.latitude
    longitude = location.longitude
    return latitude, longitude
```

We define a function to interact with FourSquare API and get top 100 venues within a radius of 1000 metres for a given latitude and longitude. Below function will return us the venue id , venue name and category.

```

In [79]: #function to return us the venue id , venue name and category
def get_venues(lat,lng):

    #set variables
    radius=1000
    LIMIT=100
    CLIENT_ID ='XXXXX' # your Foursquare ID
    CLIENT_SECRET ='XXXX' # your Foursquare Secret
    VERSION ='20200419' # Foursquare API version

    #url to fetch data from foursquare api
    url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_sec
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat, |
        lng,
        radius,
        LIMIT)

    # get all the data
    results = requests.get(url).json()
    venue_data=results["response"]["groups"][0]['items']
    venue_details=[]
    for row in venue_data:
        try:
            venue_id=row['venue']['id']
            venue_name=row['venue']['name']
            venue_category=row['venue']['categories'][0]['name']
            venue_details.append([venue_id,venue_name,venue_category])
        except KeyError:
            pass

    column_names=['ID', 'Name', 'Category']
    df = pd.DataFrame(venue_details,columns=column_names)
    return df

```

Now we will define a function to get venue details like like count , rating , tip counts for a given venue id. This will be used for ranking.

```

In [80]: #function to get venue details like like count , rating , tip counts for a given

def get_venue_details(venue_id):

    CLIENT_ID = 'XXX' # your Foursquare ID
    CLIENT_SECRET = 'XXX' # your Foursquare Secret
    VERSION = '20200419' # Foursquare API version

    #url to fetch data from foursquare api
    url = 'https://api.foursquare.com/v2/venues/{?}&client_id={}&client_secret={
        venue_id,
        CLIENT_ID,
        CLIENT_SECRET, |
        VERSION)

    # get all the data
    results = requests.get(url).json()
    venue_data=results['response']['venue']
    venue_details=[]
    try:
        venue_id=venue_data['id']
        venue_name=venue_data['name']
        venue_likes=venue_data['likes']['count']
        venue_rating=venue_data['rating']
        venue_tips=venue_data['tips']['count']
        venue_details.append([venue_id,venue_name,venue_likes,venue_rating,venue
    except KeyError:
        pass

    column_names=['ID','Name','Likes','Rating','Tips']
    df = pd.DataFrame(venue_details,columns=column_names)
    return df

```

Now we define a funtion to get the new york city data such as Boroughs, Neighborhoods along with their latitude and longitude.

#function to get the new york city data such as Boroughs, Neighborhoods along with

```
def get_new_york_data():
    url='https://cocl.us/new_york_dataset'
    resp=requests.get(url).json()
    # all data is present in features label
    features=resp['features']

    # define the dataframe columns
    column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']
    # instantiate the dataframe
    new_york_data = pd.DataFrame(columns=column_names)

    for data in features:
        borough = data['properties']['borough']
        neighborhood_name = data['properties']['name']

        neighborhood_latlon = data['geometry']['coordinates']
        neighborhood_lat = neighborhood_latlon[1]
        neighborhood_lon = neighborhood_latlon[0]

        new_york_data = new_york_data.append({'Borough': borough,
                                              'Neighborhood': neighborhood_name,
                                              'Latitude': neighborhood_lat,
                                              'Longitude': neighborhood_lon}, ignore

    return new_york_data
```

We will call the above function to get the new york city data.

```
# get new york data
new_york_data=get_new_york_data()
```

```
new_york_data.head()
```

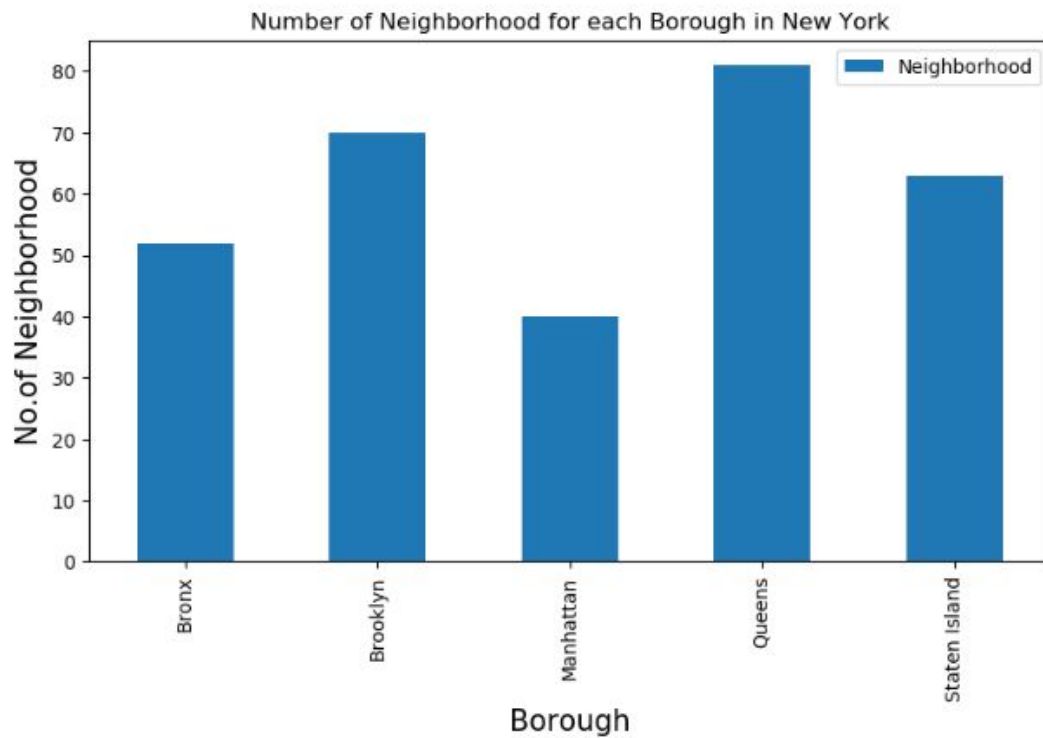
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

```
new_york_data.shape
```

```
(306, 4)
```

So there are total of 306 different Neighborhoods in New York


```
In [85]: plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Neighborhood for each Borough in New York')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Neighborhood', fontsize=15)
#giving a bar plot
new_york_data.groupby('Borough')['Neighborhood'].count().plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```



We see that Queens has highest number of neighborhoods
Now we will collect Indian restaurants for each Neighborhood

```
In [230]: # prepare neighborhood list that contains indian resturants
column_names=['Borough', 'Neighborhood', 'ID', 'Name']
indian_rest_ny=pd.DataFrame(columns=column_names)
count=1
for row in new_york_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    indian_resturants=venues[venues['Category']=='Indian Restaurant']
    print('(',count, '/',len(new_york_data),')','Indian Resturants in '+Neighborhood+', '+Borough+',str(len(indian_resturants)))
    for restaurant_detail in indian_resturants.values.tolist():
        id, name , category=restaurant_detail
        indian_rest_ny = indian_rest_ny.append({'Borough': Borough,
                                                'Neighborhood': Neighborhood,
                                                'ID': id,
                                                'Name' : name
                                                }, ignore_index=True)

    count+=1

( 1 / 306 ) Indian Resturants in Wakefield, Bronx:0
( 2 / 306 ) Indian Resturants in Co-op City, Bronx:0
( 3 / 306 ) Indian Resturants in Eastchester, Bronx:0
( 4 / 306 ) Indian Resturants in Fieldston, Bronx:0
( 5 / 306 ) Indian Resturants in Riverdale, Bronx:1
( 6 / 306 ) Indian Resturants in Kingsbridge, Bronx:1
( 7 / 306 ) Indian Resturants in Marble Hill, Manhattan:0
( 8 / 306 ) Indian Resturants in Woodlawn, Bronx:1
( 9 / 306 ) Indian Resturants in Norwood, Bronx:0
( 10 / 306 ) Indian Resturants in Williamsbridge, Bronx:0
( 11 / 306 ) Indian Resturants in Baychester, Bronx:0
( 12 / 306 ) Indian Resturants in Pelham Parkway, Bronx:0
( 13 / 306 ) Indian Resturants in City Island, Bronx:0
( 14 / 306 ) Indian Resturants in Bedford Park, Bronx:0
( 15 / 306 ) Indian Resturants in University Heights, Bronx:0
( 16 / 306 ) Indian Resturants in Morris Heights, Bronx:0
( 17 / 306 ) Indian Resturants in Fordham, Bronx:0
( 18 / 306 ) Indian Resturants in East Tremont. Bronx:0
```

```
In [87]: indian_rest_ny.head()
```

```
Out[87]:
```

	Borough	Neighborhood	ID	Name
0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot
1	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill
2	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine
3	Bronx	Concourse	551b7f75498e86c00a0ed2e1	Hungry Bird
4	Bronx	Unionport	4c194631838020a13e78e561	Melanies Roti Bar And Grill

```
In [88]: indian_rest_ny.shape
```

```
Out[88]: (137, 4)
```

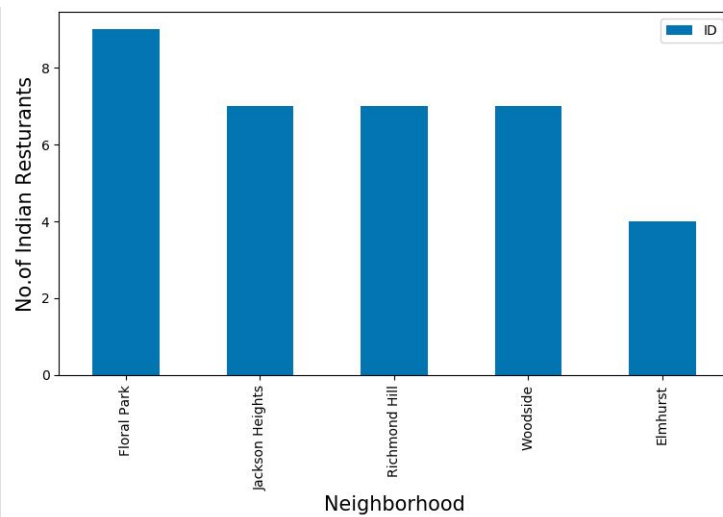
We got 137 Indian Resturants across New York City


```
In [90]: plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Indian Resturants for each Borough in New York')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Indian Resturants', fontsize=15)
#giving a bar plot
indian_rest_ny.groupby('Borough')['ID'].count().plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```



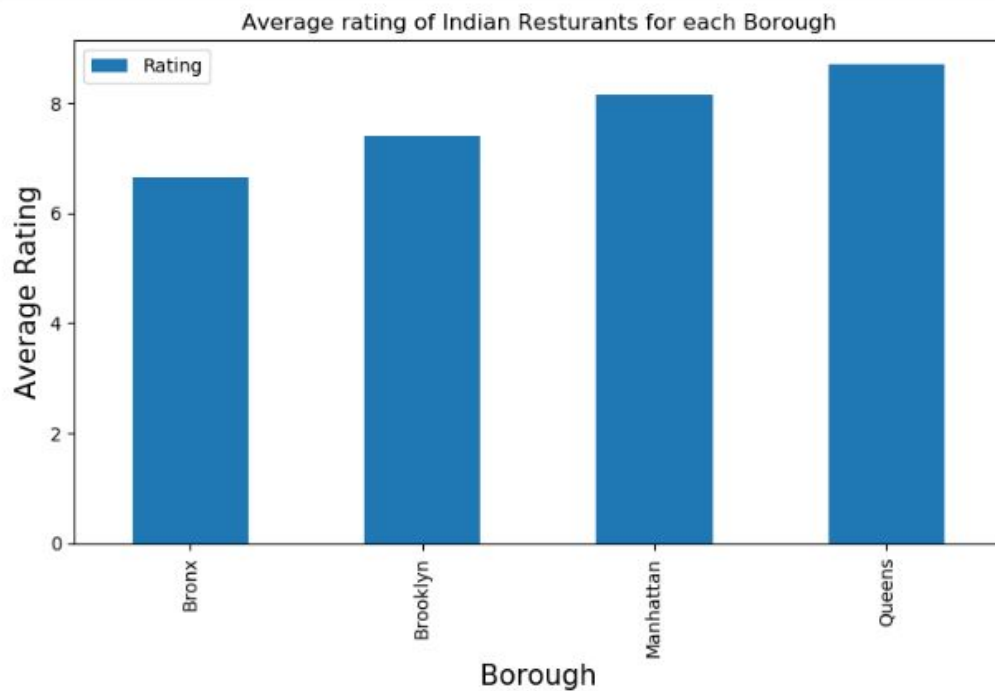
We see that Queens has the largest number of indian restaurants.

```
In [91]: plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Indian Restaurants for each Neighborhood in New York City')
#On x-axis
plt.xlabel('Neighborhood', fontsize = 15)
#On y-axis
plt.ylabel('No.of Indian Resturants', fontsize=15)
#giving a bar plot
indian_rest_ny.groupby('Neighborhood')['ID'].count().nlargest(5).plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```



So Floral Park in Queens has the highest number of Indian Resturants

```
In [57]: plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Average rating of Indian Restaurants for each Borough')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('Average Rating', fontsize=15)
#giving a bar plot
indian_rest_stats_ny.groupby('Borough').mean()['Rating'].plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```



Conclusion

- Queens has the best neighborhoods for indian cuisine.
- Manhattan has a potential Indian Restaurant Market.
- Staten Island ranks last in average rating of Indian Restaurants.