

+ Code + Text

✓ RAM

Disk

Editing



## Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods (Week 2)

Sharookh Ali

### REPORT

#### Table of Content

1. Introduction
2. Data Extraction
3. Methodology
4. Observation & Discussion
5. Conclusion

#### 1. Introduction: A description of the problem and a discussion of the background.

1. In Module 3, we explored New York City and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries.
2. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are.
3. In this assignment we will compare the neighborhoods of the two Indian cities namely Hyderabad and Mumbai
4. Is Hyderabad more like Mumbai in terms of Shops, Restaurants , Cafes etc?
5. In Hyderabad or Mumbai, if someone is looking to open a restaurant, where would we recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office? and other important points will be discussed
6. If a person wants to shift from Hyderabad to Mumbai or vice versa, then how likely he will be adjusted to these cities will be discussed.

#### 2. Data Extraction: Extracting the data of HYDERABAD & MUMBAI

The data extracted may include:

1. List of Cafes and foodstalls of Hyderabad with their latitude and longitude
2. List of Hotels, Book stores, Restaurant, Bakeries, and all relevant locations
3. All data is limited within 10km radius

#### 3. Methodology

1. First extracting the geographical coordinates of Hyderabad & Mumbai through geolocator is to be done.
2. Then FOURSQUARE credentials are specified for the analysis of maps.
3. Then the data is gathered with FOURSQUARE API url with a limit of 10 kms radius and maximum count of 200 in both Mumbai and Hyderabad
4. After the data is gathered, the venues are sorted according to their categories and printed by including both latitude and longitude.
5. Then the maps of Mumbai and Hyderabad are represented with the help of Folium.
6. Then the value count with respect to each category is done and printed for both Hyderabad and Mumbai individually
7. The corresponding value counts are represented as bar graphs and the observations, intuitions are further discussed in "OBSERVATION & DISCUSSION" section.

```
[1]: import json
import requests
import matplotlib.pyplot as plt
from pandas.io.json import json_normalize
import numpy as np
import time
import pandas as pd
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
from geopy.geocoders import Nominatim
import folium
print('Libraries imported.')
```

Libraries imported.

#### Extracting the geographical coordinates of Hyderabad through geolocator

```

✓ [2] address = 'Hyderabad'
geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geographical coordinates of', address, 'are {}, {}'.format(latitude, longitude))
neighborhood_latitude=latitude
neighborhood_longitude=longitude

/usr/local/lib/python3.7/dist-packages/geopy/geocoders/osm.py:143: UserWarning: Using Nominatim with the default "geopy/1.17.0" `user_agent` is strongly discouraged
  UserWarning
The geographical coordinates of Hyderabad are 17.360589, 78.4740613.

CLIENT_ID = '#####' # your Foursquare ID
CLIENT_SECRET = '#####' # your Foursquare Secret
VERSION = '#####' # your version

✓ [4] radius = 10000
limit = 200
# importing the data with URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&l={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    limit)
results = requests.get(url).json()

```

#### ▼ Determining the categories of each location in data

```

✓ [5] def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']
    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

```

#### ▼ Showing the list of venues

```

✓ [6] venues = results['response'][0]['groups'][0]['items']
borough = results['response'][0]['groups'][0]['items']
HYD_venues = json_normalize(venues) # flatten JSON
HYD_borough= json_normalize(borough)

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
HYD_venues = HYD_venues.loc[:, filtered_columns]
# filter the category for each row
HYD_venues['venue.categories'] = HYD_venues.apply(get_category_type, axis=1)
# clean columns
HYD_venues.columns = [col.split(".")[-1] for col in HYD_venues.columns]
HYD_venues.head(100)

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: FutureWarning: pandas.io.json.json_normalize is deprecated, use pandas.json_normalize instead
  This is separate from the ipykernel package so we can avoid doing imports until
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:4: FutureWarning: pandas.io.json.json_normalize is deprecated, use pandas.json_normalize instead
  after removing the cwd from sys.path.

```

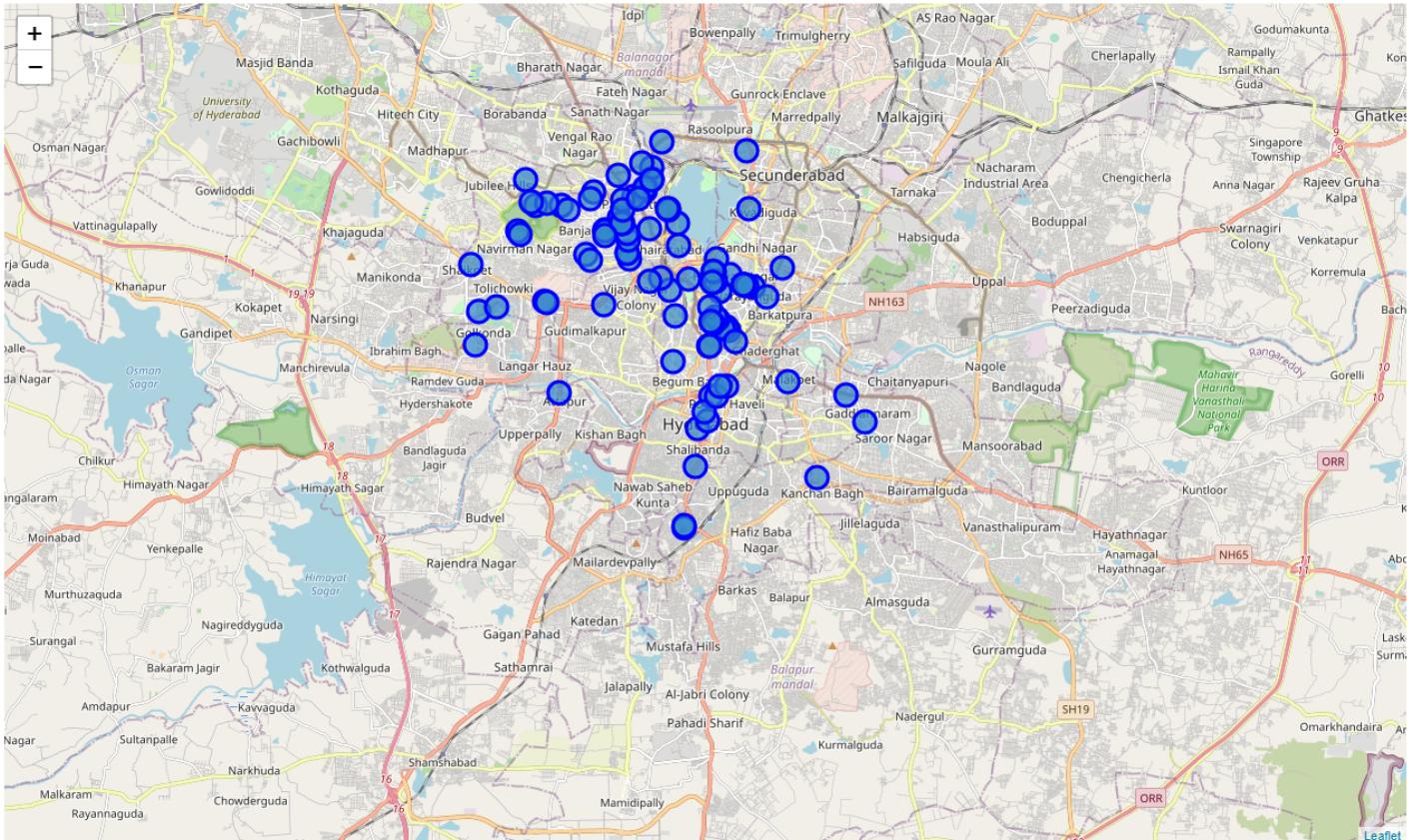
	name	categories	lat	lng
0	Chowmahala Palace	History Museum	17.359300	78.471617
1	Shadaab	Diner	17.368661	78.475572
2	Hotel Shah Ghousie	Snack Place	17.347899	78.470986
3	Taj Falaknuma Palace	Resort	17.330118	78.467460
4	Charminar	Monument / Landmark	17.361555	78.474677
5	Pragati	South Indian Restaurant	17.388088	78.481134
6	Govind Ki Idly	South Indian Restaurant	17.364132	78.473771
7	Laxman Ki Bandi	South Indian Restaurant	17.378895	78.463973
8	Mayur Pan Shop	Juice Bar	17.388894	78.480578
9	Santosh Dhaba	Indian Restaurant	17.388485	78.479509
10	Cafe Niloufer & Bakers	Café	17.399715	78.462881
11	King & Cardinal	Burger Joint	17.400678	78.488575
12	Cream Stone Concepts	Ice Cream Shop	17.404284	78.481458

13	Karachi Bakery	Bakery	17.383454	78.475075
14	Minerva Coffee Shop	Indian Restaurant	17.401410	78.485672
15	Hotel Nayaab	Food Court	17.368504	78.477480
16	Nizam club	Lounge	17.403221	78.468729
17	Swagath Tiffins	Vegetarian / Vegan Restaurant	17.344639	78.508231
18	Siraj's International Juice Center	Juice Bar	17.395579	78.442701
19	Cafe Bahar	Indian Restaurant	17.399595	78.478566
20	Hotel Sohall	Indian Restaurant	17.372829	78.499298
21	Chicha's	Hyderabadi Restaurant	17.403255	78.460152
22	Subhan Bakery	Bakery	17.392412	78.464712
23	Spencers	Shopping Mall	17.369601	78.429146
24	Ram ki Bandi	Food Truck	17.383416	78.475502
25	Necklace Road	Scenic Lookout	17.419384	78.465307
26	Gokul Chat Center	Chaat Place	17.384716	78.483152
27	Salar Jung Museum	History Museum	17.371773	78.480636
28	Gokul Chat	Snack Place	17.397860	78.492700

[ ]

## Creating Hyderabad map with Folium

```
[7] MAP_HYD = folium.Map(location=[latitude, longitude], zoom_start=12)
# add markers to map
for lat, lng, label in zip(HYD_venues['lat'], HYD_venues['lng'], HYD_venues['name']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=10,
        popup=label,
        color='blue',
        fill_color='#3186cc',
        fill_opacity=0.7,
    ).add_to(MAP_HYD)
MAP_HYD
```



```
[8] from folium import plugins
import seaborn as sns
import matplotlib.cm as cm
import matplotlib.colors as colors
from sklearn.cluster import KMeans
print('Additional libraries imported.')
```

```
Additional libraries imported.
```

## Extracting the data of Mumbai

In this case also the data extracted may include:

1. List of Cafes and foodstalls of Mumbai with their latitude and longitude
2. List of Hotels, Book stores, Restaurant, Bakeries, and all relevant locations
3. All data is limited with in 10km radius

```
✓ [9] address = 'Mumbai'
geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geographical coordinates of' ,address,'are {}, {}'.format(latitude, longitude))
neighborhood_latitude=latitude
neighborhood_longitude=longitude

/usr/local/lib/python3.7/dist-packages/geopy/geocoders/osm.py:143: UserWarning: Using Nominatim with the default "geopy/1.17.0" `user_agent` is strongly discouraged
  UserWarning
The geographical coordinates of Mumbai are 19.0759899, 72.8773928.

✓ [10] radius = 20000
limit = 200
# importing the data with URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    limit)
results = requests.get(url).json()
```

## Determining the categories of each location in data & Showing the list of venues

```
✓ [11] venues = results['response'][‘groups’][0][‘items’]

MB_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = [‘venue.name’, ‘venue.categories’, ‘venue.location.lat’, ‘venue.location.lng’]
MB_venues = MB_venues.loc[:, filtered_columns]
# filter the category for each row
MB_venues[‘venue.categories’] = MB_venues.apply(get_category_type, axis=1)
# clean columns
MB_venues.columns = [col.split(“.”)[-1] for col in MB_venues.columns]
MB_venues.head(100)

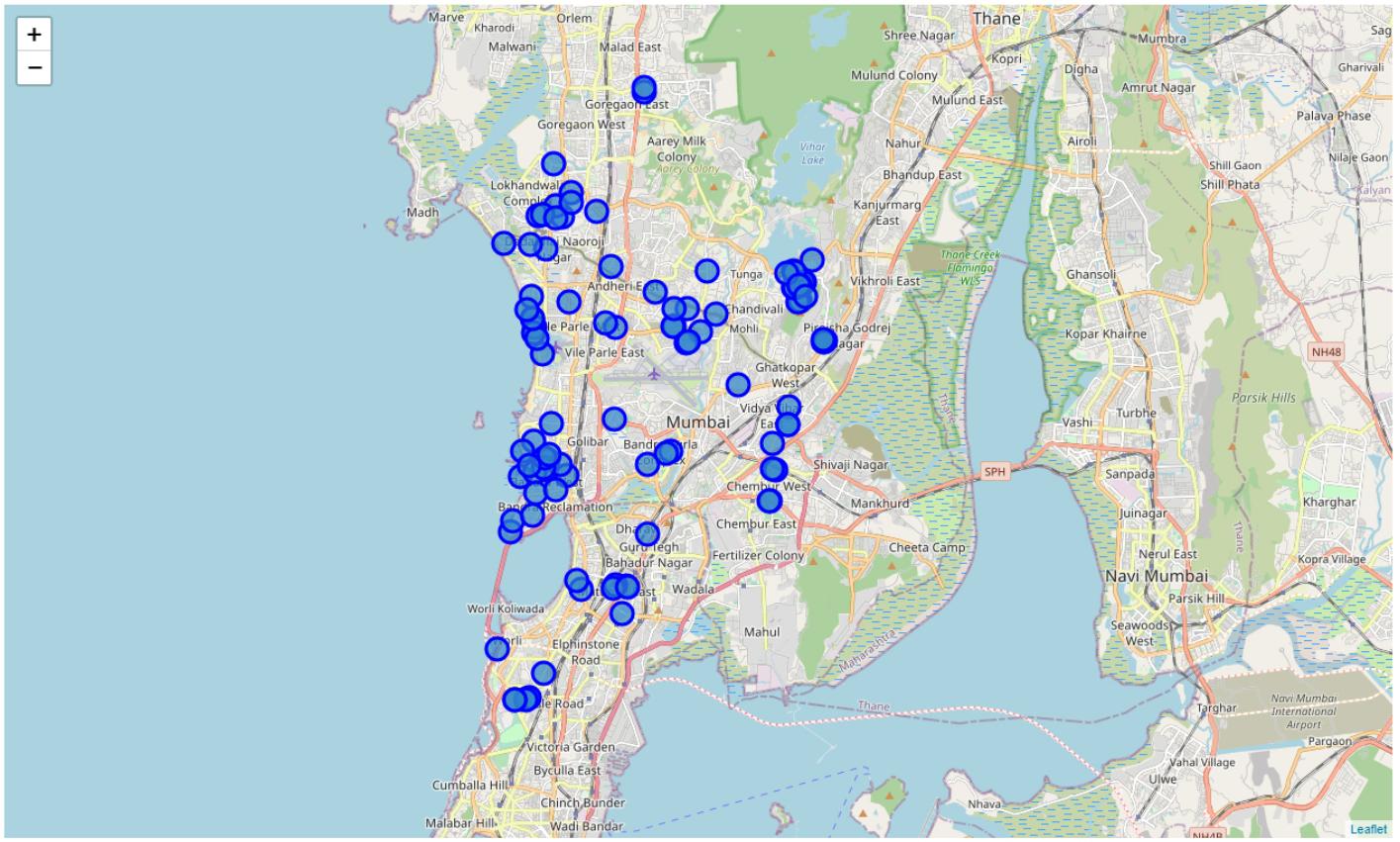
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: FutureWarning: pandas.io.json.json_normalize is deprecated, use pandas.json_normalize instead
This is separate from the ipykernel package so we can avoid doing imports until
```

		name	categories	lat	lng
0		Sofitel Mumbai BKC	Hotel	19.067448	72.869006
1		Starbucks Coffee Capital	Coffee Shop	19.063457	72.861576
2		Hamleys	Toy / Game Store	19.086655	72.889783
3		Trident	Hotel	19.066808	72.867468
4		Grand Hyatt	Hotel	19.076832	72.851270
5		Peshawari	Indian Restaurant	19.103954	72.869879
6		JW Marriott Mumbai Sahar	Hotel	19.102502	72.878236
7		Guru Kripa	Indian Restaurant	19.042955	72.861796
8		ITC Maratha	Hotel	19.104023	72.869638
9		Natural Ice Cream	Dessert Shop	19.052643	72.899760
10		Khau Galli	Food Truck	19.080160	72.905489
11		Bombay Salad Co.	Salad Place	19.064715	72.830920
12		The Leela	Hotel	19.109095	72.874114
13		Escobar	Lounge	19.060177	72.836446
14		Indigo Deli	Deli / Bodega	19.061710	72.829972
15		Starbucks	Coffee Shop	19.075523	72.831745
16		Emirates First and Business Class Checkin	Airport Service	19.099148	72.873673
17		Jogger’s Park	Park	19.059728	72.822055
18		Candies	Fast Food Restaurant	19.070288	72.826245

19	Candies	Bakery	19.061176	72.826634
20	IVY Restaurant & Banquets	Restaurant	19.069663	72.900535
21	Bombay Coffee House	Coffee Shop	19.063425	72.834671
22	Soho House	Club House	19.095805	72.829070
23	Punjab Sweet House And Restaurant	Indian Restaurant	19.062553	72.829423
24	Peshawri Restaurant, Itc, Mumbai	Indian Restaurant	19.109283	72.869891
25	Smoke House Deli	Deli / Bodega	19.065482	72.829861
26	Café Madras	Café	19.027721	72.855196
27	Ram Ashraya	Vegetarian / Vegan Restaurant	19.028092	72.851729
28	Prithvi Theatre	Theater	19.106157	72.825810
29	Mahesh Lunch Home	Seafood Restaurant	19.103680	72.826794

## Creating Mumbai map with Folium

```
✓ [12] MAP_MB = folium.Map(location=[latitude, longitude], zoom_start=12)
# add markers to map
for lat, lng, label in zip(MB_venues['lat'], MB_venues['lng'], MB_venues['name']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=10,
        popup=label,
        color='blue',
        fill_color="#3186cc",
        fill_opacity=0.7,
    ).add_to(MAP_MB)
MAP_MB
```



## Value count of HYDERABAD

```
✓ [13] HYD_count = HYD_venues['categories'].value_counts()
MB_count = MB_venues['categories'].value_counts()
print(">>>>>>>>HYDERABAD<<<<<<")
print(HYD_count)

>>>>HYDERABAD<<<<
Indian Restaurant      11
Bakery                 8
Hotel                  7
South Indian Restaurant 4
Multiplex               4
Diner                  3
History Museum          3
```

```
Café          3
Chinese Restaurant 3
Ice Cream Shop   3
Snack Place      2
Chaat Place      2
Restaurant        2
Historic Site    2
Asian Restaurant  2
Nightclub         2
Fast Food Restaurant 2
Juice Bar         2
Shopping Mall    2
Coffee Shop       2
Bistro           1
Pizza Place       1
Golf Course       1
Shoe Store         1
Breakfast Spot    1
Lake              1
Neighborhood       1
Food Court         1
Andhra Restaurant 1
Dessert Shop       1
BBQ Joint          1
Italian Restaurant 1
American Restaurant 1
Park              1
Department Store  1
Food Truck          1
Hyderabadi Restaurant 1
Resort             1
Lounge             1
Hookah Bar         1
Vegetarian / Vegan Restaurant 1
Hotel Bar          1
Bookstore          1
Steakhouse         1
Monument / Landmark 1
Scenic Lookout     1
Pub                1
Deli / Bodega      1
Performing Arts Venue 1
Smoke Shop          1
Burger Joint        1
Name: categories, dtype: int64
```

#### ▼ Value count of MUMBAI

```
✓ [14] print("")  
      print(">>>>MUMBAI<<<<")  
      print(MB_count)
```

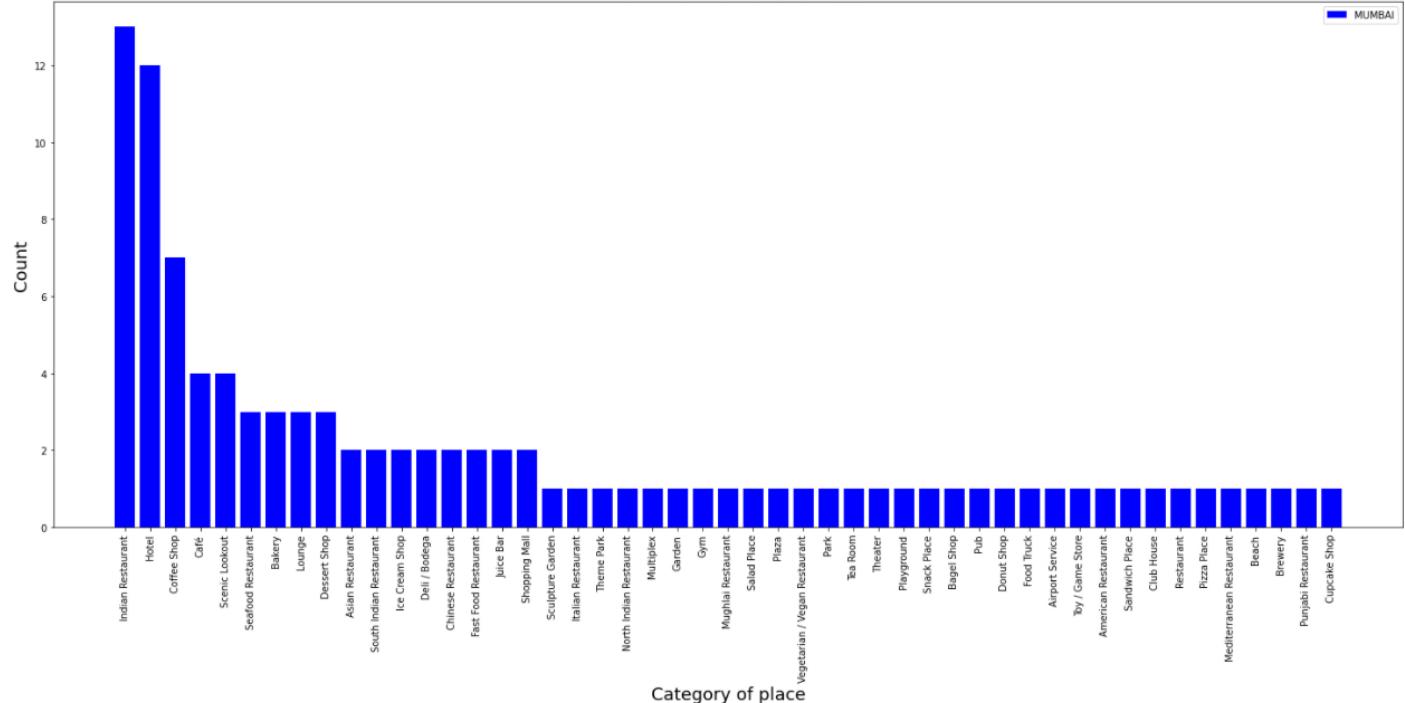
```
>>>>MUMBAI<<<<
Indian Restaurant      13
Hotel                  12
Coffee Shop            7
Café                   4
Scenic Lookout         4
Seafood Restaurant     3
Bakery                 3
Lounge                 3
Dessert Shop           3
Asian Restaurant        2
South Indian Restaurant 2
Ice Cream Shop          2
Deli / Bodega          2
Chinese Restaurant      2
Fast Food Restaurant    2
Juice Bar               2
Shopping Mall          2
Sculpture Garden        1
Italian Restaurant      1
Theme Park              1
North Indian Restaurant 1
Multiplex               1
Garden                 1
Gym                    1
Mughlai Restaurant     1
Salad Place             1
Plaza                  1
Vegetarian / Vegan Restaurant 1
Park                   1
Tea Room                1
Theater                 1
Playground              1
Snack Place             1
Bagel Shop              1
Pub                     1
Donut Shop              1
Food Truck              1
Airport Service          1
Toy / Game Store         1
American Restaurant     1
Sandwich Place          1
Club House              1
Restaurant              1
Pizza Place              1
Mediterranean Restaurant 1
Beach                   1
Brewery                 1
Puniahi Restaurant      1
```

```
Cupcake Shop  
Name: categories, dtype: int64
```

```
[15] MB_count = MB_venues['categories'].value_counts()  
plt.figure(figsize = (25, 10))  
plt.xticks(rotation = 90)  
plt.xlabel("Category of place", fontsize = 18)  
plt.ylabel("Count", fontsize = 18)  
plt.title("Venues counts of each category in MUMBAI", fontsize = 16)  
plt.bar(MB_count.index, MB_count.values,color='blue')  
plt.legend(["MUMBAI"])
```

```
<matplotlib.legend.Legend at 0x7f551076f8d0>
```

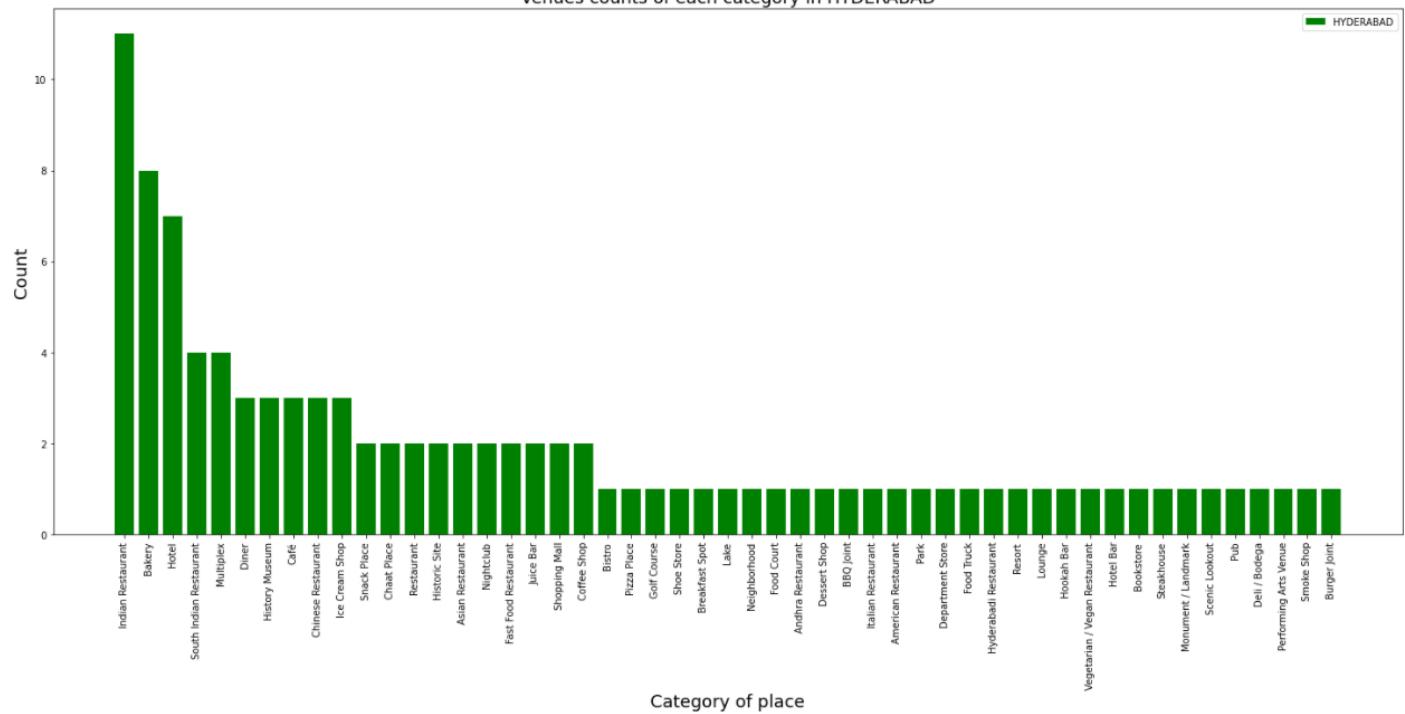
Venues counts of each category in MUMBAI



```
[16] plt.figure(figsize = (25, 10))  
plt.xticks(rotation = 90)  
plt.xlabel("Category of place", fontsize = 18)  
plt.ylabel("Count", fontsize = 18)  
plt.title("Venues counts of each category in HYDERABAD", fontsize = 18)  
plt.bar(HYD_count.index, HYD_count.values, color='green')  
plt.legend(["HYDERABAD"])
```

```
<matplotlib.legend.Legend at 0x7f55105c46d0>
```

Venues counts of each category in HYDERABAD



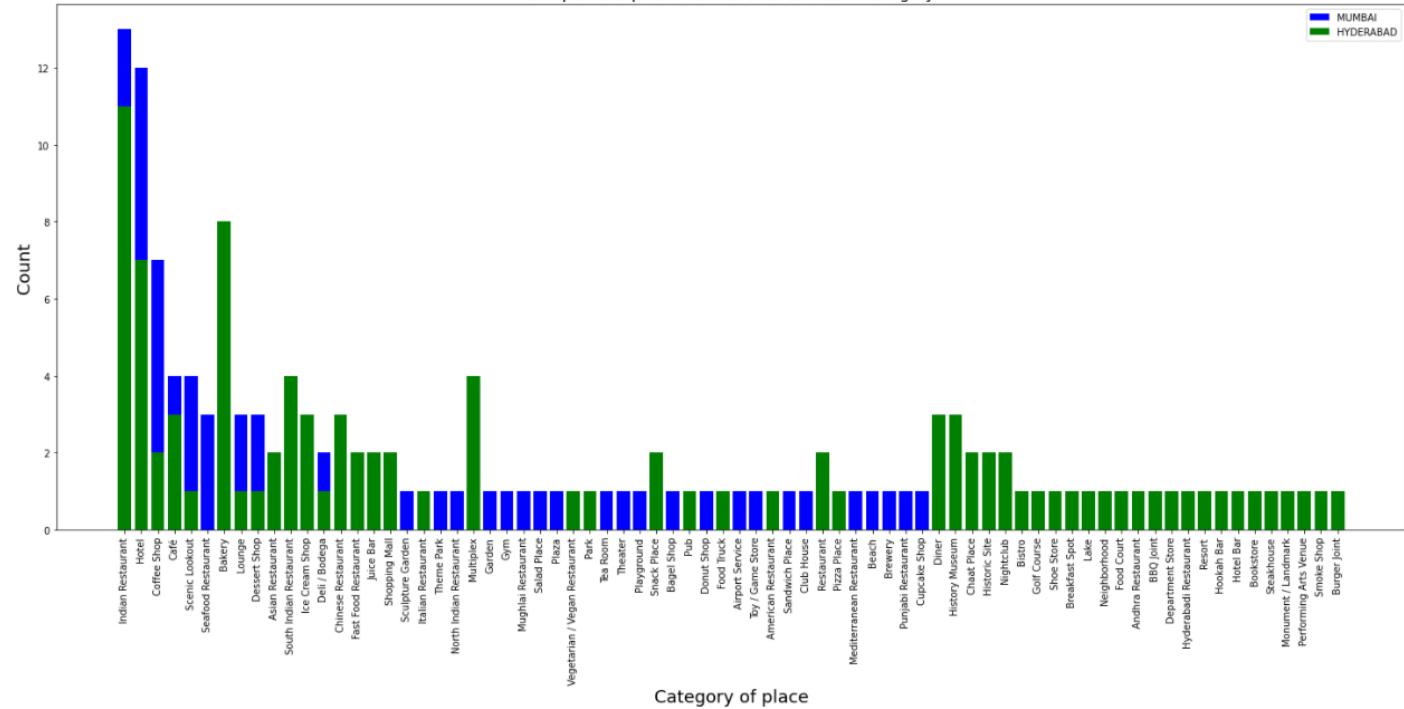
```

17] plt.figure(figsize = (25, 10))
    plt.xticks(rotation = 90)
    plt.xlabel("Category of place", fontsize = 18)
    plt.ylabel("Count", fontsize = 18)
    plt.title("Comparative plot of venues counts of each category", fontsize = 16)
    plt.bar(MB_count.index, MB_count.values,color='blue')
    plt.bar(HYD_count.index, HYD_count.values, color='green')
    plt.legend(["MUMBAI","HYDERABAD"])

```

<matplotlib.legend.Legend at 0x7f550ffe42d0>

Comparative plot of venues counts of each category



## 4. Observations & Discussion

- From above bar graph, we can observe that both HYDERABAD and MUMBAI are similar in terms of Shops, Restaurants , Cafes etc.
- However, the Mumbai is more versatile in terms of venues compare to Hyderabad.
- In Mumbai, when we observe the Folium map created with the help of Foursquare credentials, more venues are clustered in "Bandra West & Juhu" area and less venues can be spotted in "Vidya Vihar" area.
- In Hyderabad, up on similar observation, more venues are clustered in "Abids, Ameerpet & Panjagutta" area and less venues can be spotted in "Golkonda" area.
- Mumbai has significantly more number of Indian restaurants, Hotels, Coffe shops, lounges.
- On other hand, Hyderabad has more number of Bakeries than Mumbai
- In terms of Shopping malls, Chinese restaurants, Asian restaurants, Ice cream shops, Juice bar both Mumbai and Hyderabad are more similar.
- Rest miscellaneous venues are found to be single or two for both Mumbai and Hyderabad.

## 5. Conclusion

- From above observations we can conclude the following things:
- In Mumbai, if some one wants to open a restaurant, one can prefer "Bandra West or Juhu" area similarly in Hyderabad one can prefer "Abids or Ameerpet" area.
- Opening Indian restaurant or Hotel in Hyderabad will be comparatively better than Mumbai because Mumbai already has a significantly more number of this category venues than Hyderabad.
- One can always prefer starting up a miscellaneous type of venues as they were low and easy entry to the market in both Mumbai and Hyderabad
- On overall, if a person wants to shift from Hyderabad to Mumbai or vice versa , there wouldn't be any problems as both cities are more similar in terms of venues according to Foursquare API data.

## Ending Remarks

Coursera has given a very good opportunity to learn Data science. Thanks to Coursera, IBM and Peers.

Happy learning!

