THE BATTLE OF NEIGHBOURHOODS

Exploring Indian restaurants in New York City

- ***Monica. J***

Introduction:

Data plays a vital role in today's world. Every industry for example medicine, healthcare, retail, banking, tourism, etc. have started to lay business strategies around the outcome of data analytics. An on-going situation where data is playing a vital role for many decisions related to medicines, travel, statistics, etc. would be the Coronavirus outbreak. The medicine and pharmaceutical industry are focussing on data to identify the increasing demand for masks, flu-related medications and hand sanitizers and alert the service providers to be able to supply to these demands .The travel and tourism industry are analysing the risk involved across the globe due to the consequences of the outbreak.

There are a number of scenarios in today's world where a lot can be inferred with data. The business scenario which I have picked for my project is as follows:

A company in the "Travel and Living" category decided to make a show to explore the most popular cuisines in New York City. They conducted a survey among people randomly and observed that the Indian cuisine is pretty popular in NY among others such as Mexican, Chinese, Thai, etc. They turned to their data analysis team to help them narrow down on the options and get the list of the best restaurants in terms of ratings, likes, etc. which they would like to consider exploring for their show.

Data:

For this project we need the following data:

1. New York City data that contains Borough, Neighbourhoods along with their latitudes and longitudes

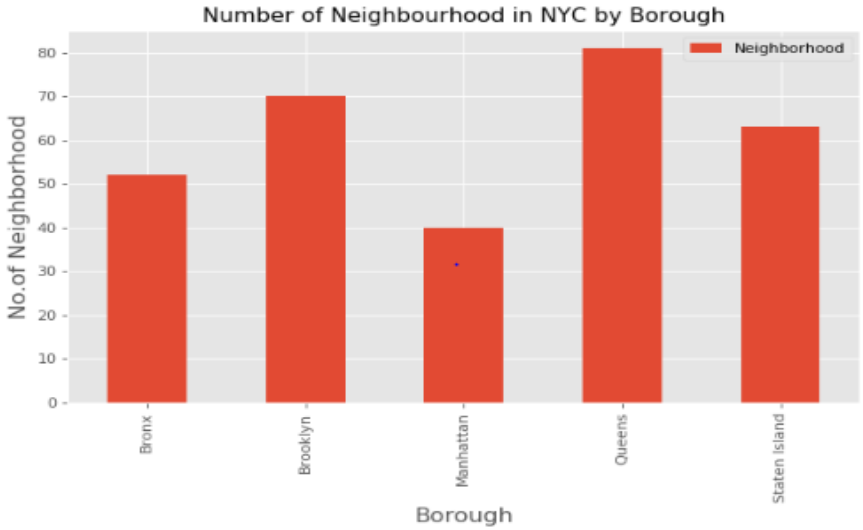
- Data Source: https://cocl.us/new_york_dataset
- Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of New York City.

```
new_york_data.head()
```

Out[3]:

	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

```
In [4]: new_york_data.shape
Out[4]: (306, 4)
```



2. Indian restaurants in New York City.

- Data Source: Foursquare API
- Description: By using this API we will filter all the venues to get only Indian restaurants and their attributes.

Sample data:

```
In [18]: indian_rest_ny.head()
```

Out[18]:

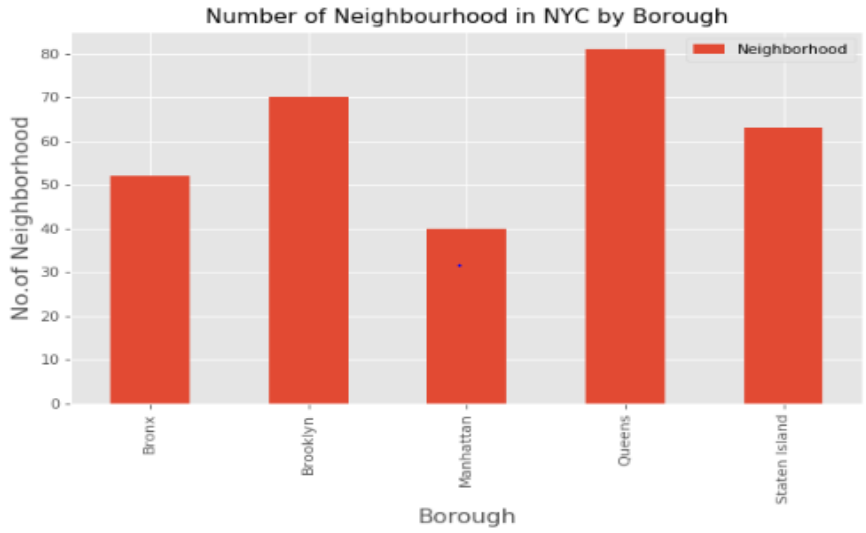
	Borough	Neighborhood	ID	Name
0	Bronx	Woodlawn	4c0448d9310fc9b6bf1dc761	Curry Spot
1	Bronx	Parkchester	4c194631838020a13e78e561	Melanies Roti Bar And Grill
2	Bronx	Parkchester	55dfa36a498e164ef19bef7b	Premium Sweets & Restaurant
3	Bronx	Spuyten Duyvil	4c04544df423a593ac83d116	Cumin Indian Cuisine
4	Bronx	Concourse	551b7f75498e86c00a0ed2e1	Hungry Bird

Methodology:

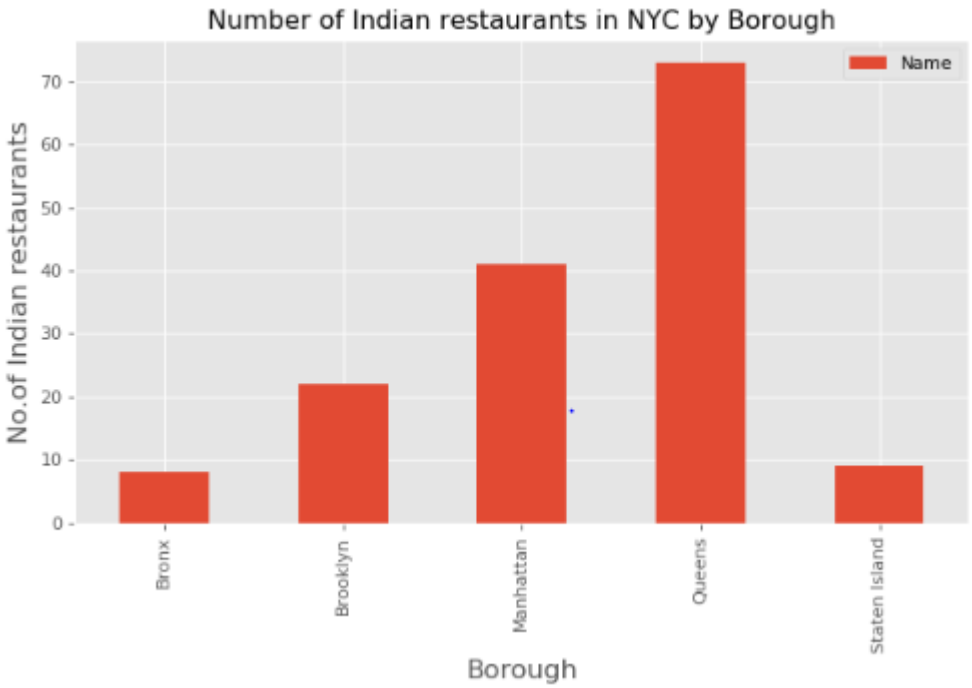
Data from the above mentioned data sources is loaded into data frames using the Panda's library. The various functions and operations under the Pandas library are used to select, filter, sort and analyse the data to identify all Indian restaurants having a rating > 8 and identifying the Top 10 restaurants based on ratings, likes and tips. Additionally, the geopy.geocoders library is used to convert the address to latitude and longitude values. Finally, the top restaurants are plotted on a map using the Folium library.

Results:

1) On analysing the data in https://cocl.us/new_york_dataset. It is observed, that NYC has 5 Boroughs and Queens has the highest number of Neighbourhoods with a count of 81.



2) Using the Foursquare API, I extracted the data for all Indian restaurants across Boroughs and could see the Queens borough has the highest number of Indian restaurants with a value of 73.

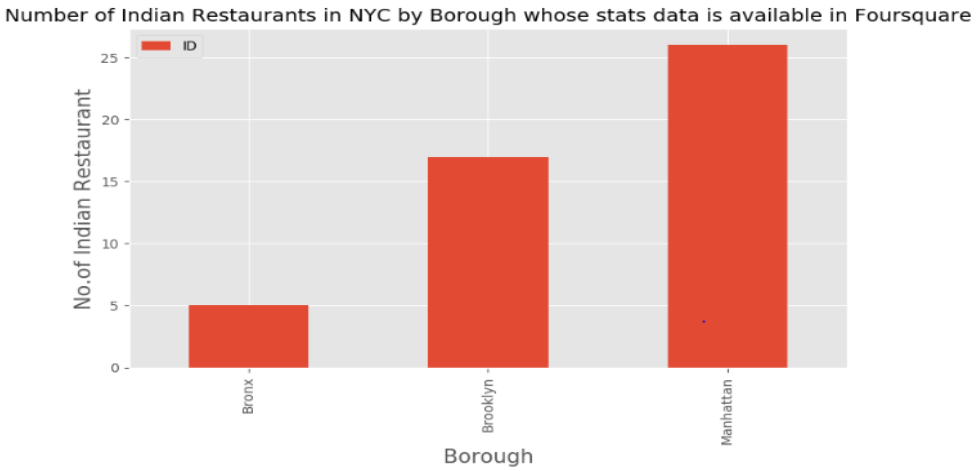


3) There are total 153 Indian restaurants in NYC. Out of which only 48 restaurants have the required data such as Tips, Rating and Tips. On analysing further it looks like restaurants in Manhattan are more frequently rated when compared to other Boroughs.

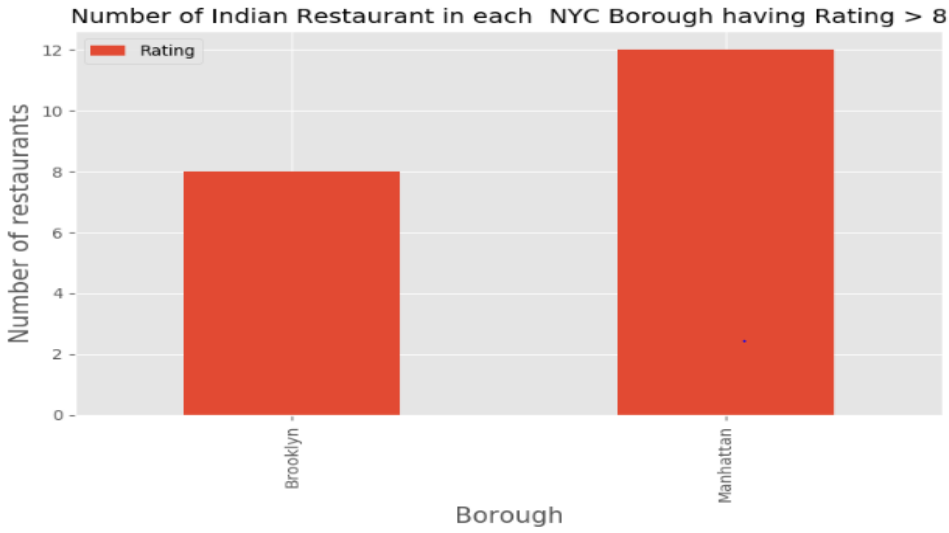
```
In [43]: indian_rest_ny.shape
Out[43]: (153, 4)
```

```
In [56]: indian_rest_stats_new =indian_rest_stats_ny[indian_rest_stats_ny.ID !=0 ]

In [58]: indian_rest_stats_new.shape
Out[58]: (48, 7)
```



4) Implying to point 3, most of the highly rated Indian restaurants are in Manhattan followed by Brooklyn.



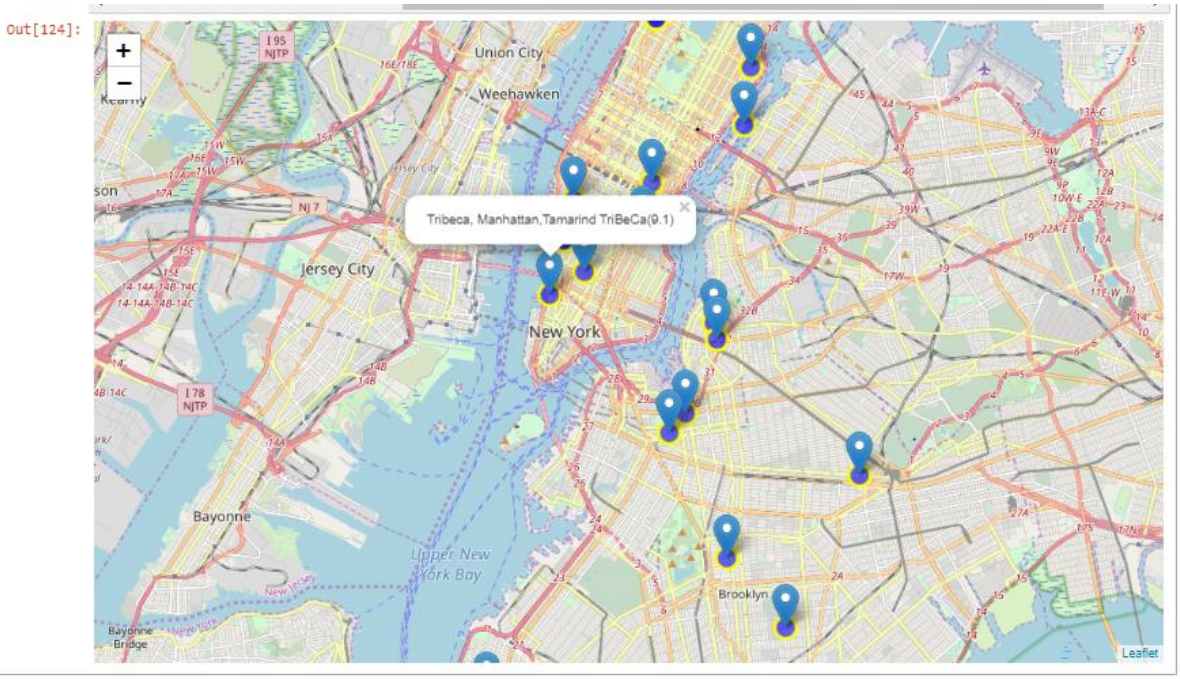
5) The top 10 restaurants based on the rating, likes and tips, is as follows:

```
In [30]: indian_res_top_10 =indian_res_rating_8.sort_values(by=['Likes','Tips'], ascending=False).head(10)

In [31]: indian_res_top_10
Out[31]:
```

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
40	Manhattan	Tribeca	4bbb9dbded7776b0e1ad3e51	Tamarind TriBeCa	590	9.1	148
49	Manhattan	Gramercy	4a70a75bf964a52016d81fe3	Bhatti Indian Grill	421	8.7	160
32	Manhattan	Yorkville	5272ca4511d22488f6895caf	Drunken Munkey	208	8.5	61
33	Manhattan	Roosevelt Island	4fe4fb50c2eee335e4fea69d	Moti Mahal Delux	186	8.4	83
42	Manhattan	West Village	591794df2be42556988e4a8e	Rahi	115	8.7	24
38	Manhattan	Chelsea	4b4aab62f964a520978c26e3	Dil-e Punjab Deli	108	8.7	44
35	Manhattan	Upper West Side	42489a80f964a5208b201fe3	Swagat	106	8.4	39
22	Brooklyn	North Side	5631511b498e3d6d7e0a4df0	Tikka Indian Grill	95	8.5	29
23	Brooklyn	South Side	5631511b498e3d6d7e0a4df0	Tikka Indian Grill	95	8.5	29
18	Brooklyn	Prospect Lefferts Gardens	4bb93b70cf2fc9b6fe64a002	Gandhi Fine Indian Cuisine	81	8.6	47

6) The top restaurants are plotted on the map of NYC



Discussion:

Though the data shows Queens has more Indian restaurants, the top restaurants in New York City are spread across Manhattan and Brooklyn. This implies multiple things:

- Is the data accurate?
- Is it a good option to go by the outcome of this data analysis or should we consider other methods to shortlisting the restaurant? For example: word of mouth or advertisements look for Michelin starred restaurants, etc.
- The number of people using the Foursquare application may be very less or the number of people rating Indian restaurants in particular maybe less or inconsistent. Should Foursquare work more on promoting the usage of the application or create awareness among people?
- Should data be considered from other similar apps such as the Google Places App or Yelp?

Conclusion:

Based on the data set taken, Queens Borough has the highest neighbourhoods as well as the highest number of Indian restaurants.

Most of the Indian restaurants having a rating of more than 8.0, highest likes and tips are in Manhattan, followed by Brooklyn.

Out of the top 10 restaurants listed as part of the analysis, 7 are in Manhattan and 3 are in Brooklyn.

Manhattan is the better option as there is more number of highly rated restaurants.

Tamarind TriBeCa looks like a go-to option for the show with a rating of 9.1, 591 likes and 148 tips.