Coursera Capstone Project—the Battle of Neighbourhoods

Indian Techie relocating to America

I. INTRODUCTION

Problem Statement: An Indian Techie has been offered a job in America. He needs to find out an accommodation and explore the neighbourhoods for his recreation, food etc. As he's new to the city, this project is aimed to help him explore the neighbourhood and give insights into the same.

Target Audience: This project aims at target audience for those who are new to the city and helping him out with the neighbourhood with popular venues, places across the city.

Indians in the New York City metropolitan region constitute one of the largest and fastest growing ethnicities in the New York City metropolitan area of the United States. The New York City region is home to the largest Indian American population among metropolitan areas by a significant margin, enumerating 711,174 uniracial individuals by the 2013-2017 U.S. Census American Community Survey estimates.

II. DATA

a) For this project, I have taken this Wikipedia report as reference: https://en.wikipedia.org/wiki/Indians_in_the_New_York_City_metropolitan_region

As cue taken from the Wikipedia report, following are the Top 5 boroughs in New York City

New York City boroughs [edit]

As the city proper with the largest Asian Indian population in the United States by a wide margin, with an estimated 227,994 individuals as of the 2014 American Community Survey,[30] and as the primary destination for new Indian immigrants,[31] New York City is subdivided into official municipal boroughs, which themselves are home to significant Asian Indian and other South Asian populations. Note that this list includes neither the large Desi populations of Pakistani Americans, Bangladeshi Americans, and Sri Lankan Americans, nor Indo-Caribbean Americans, Afghan Americans, and others of South Asian origin who make their home in New York City.

Rank	Borough	City	Indian Americans	Density of Indian Americans per square mile	Percentage of Indian Americans in municipality's population
1	Queens (2014)[32]	New York City	144,896	1,326.5	6.2
2	Brooklyn (2012)	New York City	25,270	357.9	1.0
3	Manhattan (2012)	New York City	24,359	1,060.9	1.5
4	The Bronx (2012)	New York City	16,748	398.6	1.2
5	Staten Island (2012)	New York City	6,646	113.6	1.4
	Total (2014)[30]	**New York City**	227,994	753.4	2.7

b) New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude.

Data source: https://geo.nyu.edu/catalog/nyu_2451_34572

Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of New York City.

c) Using FourSquare API we will find all venues for each neighbourhood.

d) geopy library to get the latitude and longitude values of New York City.

III. METHODOLOGY

1. ***Download and Explore New York City Dataset -***
   https://geo.nyu.edu/catalog/nyu_2451_34572

2. ***Load and explore the data -***
   https://cocl.us/new_york_dataset

   We get a JSON File with output as below:


```
{'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',
   'geometry': {'type': 'Point',
    'coordinates': [-73.82993910812398, 40.87429419303012]},
   'geometry_name': 'geom',
```


   We get all the details of the neighbourhoods of New York City from this data

3. ***Transform the data into a pandas dataframe***

We get the following Output after loading the data into dataframe

Quickly examine the resulting dataframe.

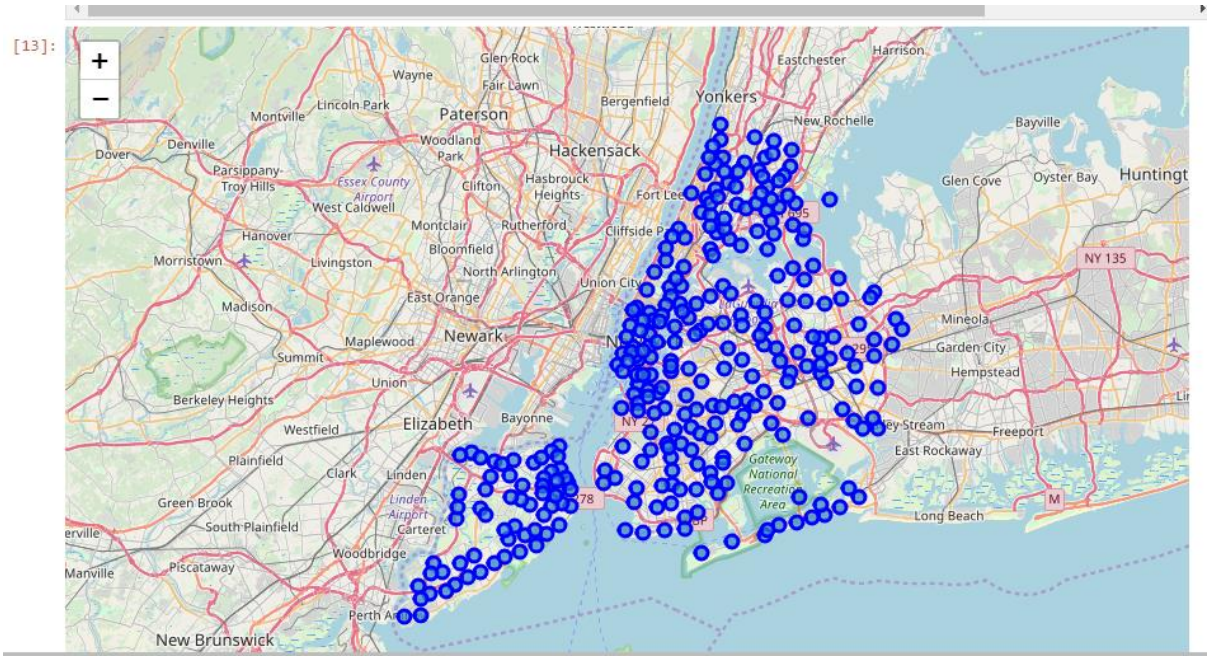
[10]:
```
neighborhoods.head(400)
```

[10]:

	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
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446
10	Bronx	Baychester	40.866858	-73.835798
11	Bronx	Pelham Parkway	40.857413	-73.854756
12	Bronx	City Island	40.847247	-73.786488

4. ***Create a map of New York with neighbourhoods superimposed on top***

[13]:



As seen from the Wikipedia report:
https://en.wikipedia.org/wiki/Indians_in_the_New_York_City_metropolitan_region

It is found that Queens in New York has the highest Indian Population, so we will further analyse 'Queens' neighbourhood and form clusters and choose the best cluster

Rank	Borough	City	Indian Americans	Density of Indian Americans per square mile	Percentage of Indian Americans in municipality's population
1	Queens (2014)[32]	New York City	144,896	1,326.5	6.2
2	Brooklyn (2012)	New York City	25,270	357.9	1.0
3	Manhattan (2012)	New York City	24,359	1,060.9	1.5
4	The Bronx (2012)	New York City	16,748	398.6	1.2
5	Staten Island (2012)	New York City	6,646	113.6	1.4
	Total (2014)[30]	**New York City**	227,994	753.4	2.7

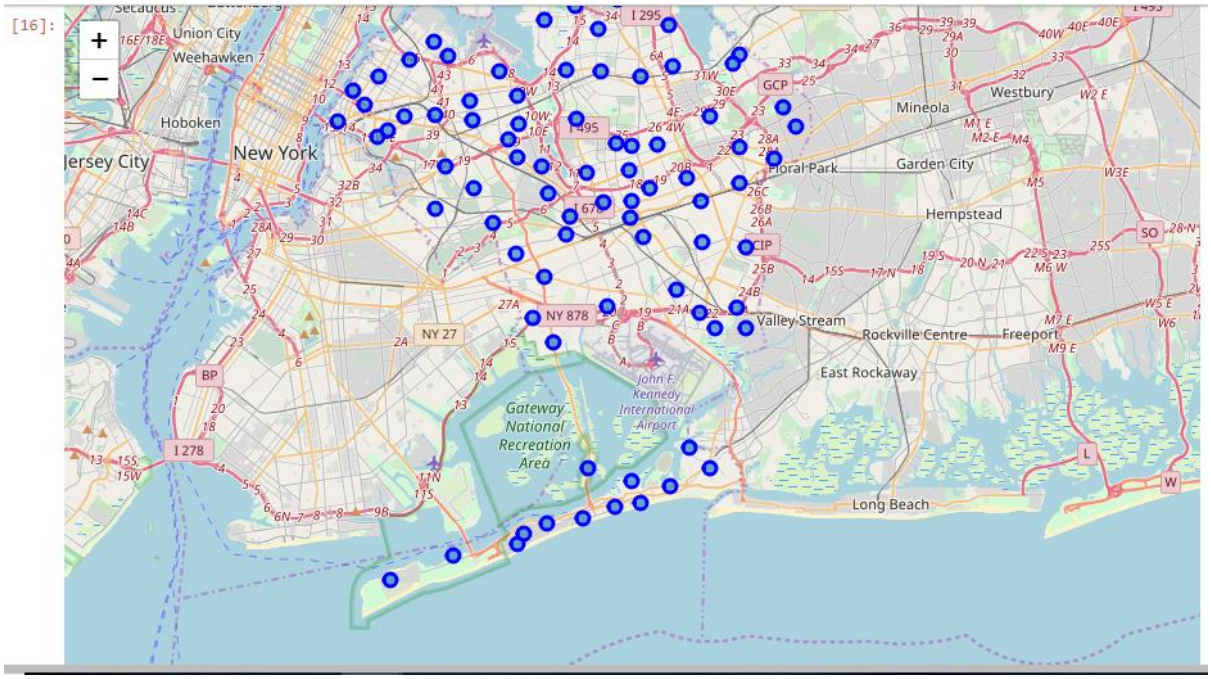
5. ***Analysing Queens Neighbourhood:***
   We find Neighbourhood details of Queens with Latitude Longitude details

```
queens_data = neighborhoods[neighborhoods['Borough'] == 'Queens'].reset_index(drop=True)
queens_data.head()
```

	Borough	Neighborhood	Latitude	Longitude
0	Queens	Astoria	40.768509	-73.915654
1	Queens	Woodside	40.746349	-73.901842
2	Queens	Jackson Heights	40.751981	-73.882821
3	Queens	Elmhurst	40.744049	-73.881656
4	Queens	Howard Beach	40.654225	-73.838138

Now we create map of Queens using latitude and longitude values

[16]:



6. ***Next, we are going to start utilizing the Foursquare API to explore the neighbourhoods and segment them.***

[23]:

	name	categories	lat	lng
0	Favela Grill	Brazilian Restaurant	40.767348	-73.917897
1	Orange Blossom	Gourmet Shop	40.769856	-73.917012
2	Titan Foods Inc.	Gourmet Shop	40.769198	-73.919253
3	CrossFit Queens	Gym	40.769404	-73.918977
4	Simply Fit Astoria	Gym	40.769114	-73.912403

7. ***Explore Neighbourhoods in Queens***

```
Astoria
Woodside
Jackson Heights
Elmhurst
Howard Beach
Corona
Forest Hills
Kew Gardens
Richmond Hill
Flushing
Long Island City
Sunnyside
East Elmhurst
Maspeth
Ridgewood
Glendale
Rego Park
Woodhaven
Ozone Park
South Ozone Park
College Point
Whitestone
Bayside
Auburndale
Little Neck
```

[35]:
```
queen_venues.groupby('Neighborhood').count()
```

[35]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Arverne	17	17	17	17	17	17
Astoria	100	100	100	100	100	100
Astoria Heights	15	15	15	15	15	15
Auburndale	19	19	19	19	19	19
Bay Terrace	42	42	42	42	42	42
Bayside	73	73	73	73	73	73
Bayswater	3	3	3	3	3	3
Beechhurst	18	18	18	18	18	18
Bellaire	12	12	12	12	12	12
Belle Harbor	18	18	18	18	18	18
Bellerose	20	20	20	20	20	20
Blissville	21	21	21	21	21	21
Breezy Point	5	5	5	5	5	5

We can see from the data of Queens neighbourhood that Astoria returned maximum venues, As a part of this project we will explore Astoria as it has maximum Venues obtained from Foursquare API Data

```
----Astoria----
                          venue  freq
0                           Bar  0.07
1  Middle Eastern Restaurant  0.07
2                  Hookah Bar  0.06
3            Greek Restaurant  0.05
4                      Bakery  0.04
```

[44]:

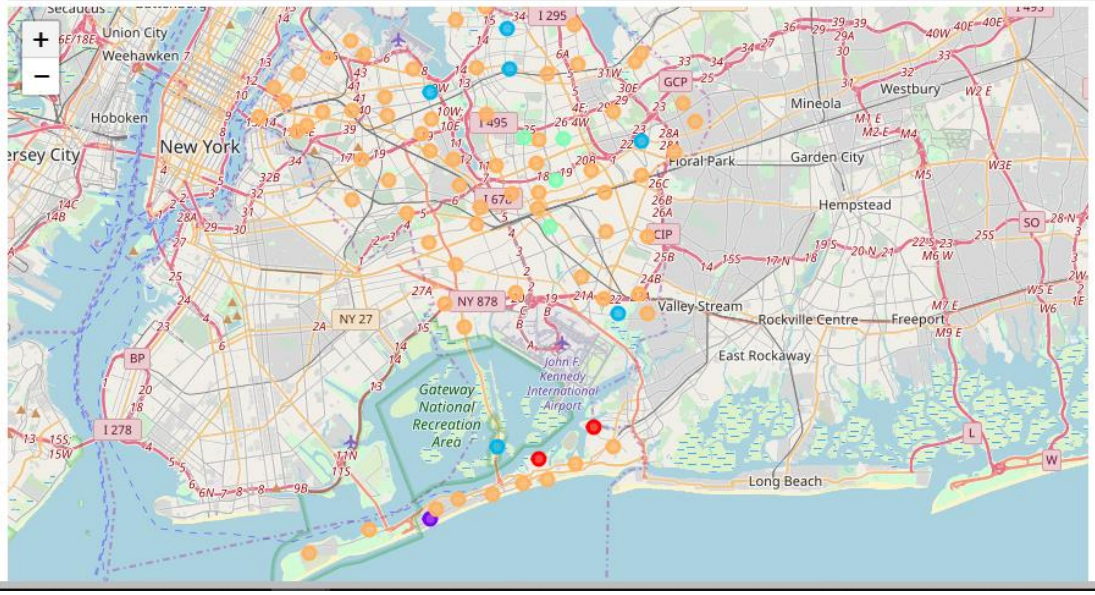
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arverne	Surf Spot	Metro Station	Bed & Breakfast	Wine Shop	Playground	Beach	Pizza Place	Donut Shop	Thai Restaurant	Bus Stop
1	Astoria	Middle Eastern Restaurant	Bar	Hookah Bar	Greek Restaurant	Bakery	Seafood Restaurant	Italian Restaurant	Indian Restaurant	Coffee Shop	Food & Drink Shop
2	Astoria Heights	Hostel	Italian Restaurant	Playground	Plaza	Music Venue	Chinese Restaurant	Moving Target	Business Service	Bus Station	Shopping Mall
3	Auburndale	Italian Restaurant	Hookah Bar	Athletics & Sports	Fast Food Restaurant	Miscellaneous Shop	Mobile Phone Shop	Supermarket	Furniture / Home Store	Korean Restaurant	Noodle House
4	Bay Terrace	Clothing Store	Women's Store	Kids Store	Donut Shop	Mobile Phone Shop	Lingerie Store	Cosmetics Shop	American Restaurant	Bank	Men's Store

8. *Now we cluster the Neighbourhoods*

eighborhood	Latitude	Longitude	Cluster Labels	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
Astoria	40.768509	-73.915654	4	Middle Eastern Restaurant	Bar	Hookah Bar	Greek Restaurant	Bakery	Seafood Restaurant	Italian Restaurant	Indian Restaurant	Coffee Shop
Woodside	40.746349	-73.901842	4	Grocery Store	Thai Restaurant	Bakery	Pizza Place	Donut Shop	Pub	Filipino Restaurant	Bar	Latin American Restaurant
Jackson Heights	40.751981	-73.882821	4	Latin American Restaurant	Peruvian Restaurant	South American Restaurant	Bakery	Mobile Phone Shop	Mexican Restaurant	Spanish Restaurant	Thai Restaurant	Empanada Restaurant
Elmhurst	40.744049	-73.881656	4	Thai Restaurant	Mexican Restaurant	Chinese Restaurant	Bubble Tea Shop	Vietnamese Restaurant	Indonesian Restaurant	South American Restaurant	Malay Restaurant	Salon / Barbershop
Howard Beach	40.654225	-73.838138	4	Italian Restaurant	Pharmacy	Bagel Shop	Chinese Restaurant	Spa	Sandwich Place	Fast Food Restaurant	Breakfast Spot	Shipping Store

9. *Visualizing Clusters using Folium Map*

[47]:



After k-means clustering, we found that Cluster – 5 has maximum traffic and top Venues are listed as below

[53]:

	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 M Comm Ven
0	Astoria	Middle Eastern Restaurant	Bar	Hookah Bar	Greek Restaurant	Bakery	Seafood Restaurant	Italian Restaurant	Indian Restaurant	Coffee Shop	Fooc Drink Sh
1	Woodside	Grocery Store	Thai Restaurant	Bakery	Pizza Place	Donut Shop	Pub	Filipino Restaurant	Bar	Latin American Restaurant	Ameri Restaur
2	Jackson Heights	Latin American Restaurant	Peruvian Restaurant	South American Restaurant	Bakery	Mobile Phone Shop	Mexican Restaurant	Spanish Restaurant	Thai Restaurant	Empanada Restaurant	Cloth St
3	Elmhurst	Thai Restaurant	Mexican Restaurant	Chinese Restaurant	Bubble Tea Shop	Vietnamese Restaurant	Indonesian Restaurant	South American Restaurant	Malay Restaurant	Salon / Barbershop	Gy Fitn Cer
4	Howard Beach	Italian Restaurant	Pharmacy	Bagel Shop	Chinese Restaurant	Spa	Sandwich Place	Fast Food Restaurant	Breakfast Spot	Shipping Store	Jewe St
5	Corona	Mexican Restaurant	Bakery	Deli / Bodega	Donut Shop	Check Cashing Service	Supermarket	Restaurant	Italian Restaurant	Food & Drink Shop	Sch
6	Forest Hills	Gym	Gym / Fitness Center	Yoga Studio	Pharmacy	Pizza Place	Park	Thai Restaurant	Convenience Store	Video Game Store	Ital Restaur

IV. RESULTS & DISCUSSION

we reached at the end of the analysis, where we got a sneak peak of the 5 major boroughs of New York City. The data exploration was mostly concentrated on the neighbourhoods and Indian localities. I have used data from web resources like Wikipedia, python libraries like Geopy, and Foursquare API, to set up a very realistic data-analysis scenario. We have found out that 'Queens' borough is the best for an Indian Immigrant settling in New York city.

We went ahead and analysed the neighbourhoods of 'Queens'. After applying K-means clustering, we found that Cluster – 5 was the best with many Venues and footfalls for recreation, food etc. which was our Objective of the Project.

The Top 5 locations from the cluster were - Astoria, Woodside, Jackson Heights, Elmhurst, Howard Beach had many Venue results obtained from the data of Foursquare API, so these are the best place for an Indian Immigrant for exploring new Venues.

Cluster 5

```
[53]: queens_merged.loc[queens_merged['Cluster Labels'] == 4, queens_merged.columns[[1] + list(range(5, queens_merged.shape[1]))]]
```

[53]:

	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 M Comm Ver
0	Astoria	Middle Eastern Restaurant	Bar	Hookah Bar	Greek Restaurant	Bakery	Seafood Restaurant	Italian Restaurant	Indian Restaurant	Coffee Shop	Fooc Drink Sh
1	Woodside	Grocery Store	Thai Restaurant	Bakery	Pizza Place	Donut Shop	Pub	Filipino Restaurant	Bar	Latin American Restaurant	Ameri Restaur
2	Jackson Heights	Latin American Restaurant	Peruvian Restaurant	South American Restaurant	Bakery	Mobile Phone Shop	Mexican Restaurant	Spanish Restaurant	Thai Restaurant	Empanada Restaurant	Cloth St
3	Elmhurst	Thai Restaurant	Mexican Restaurant	Chinese Restaurant	Bubble Tea Shop	Vietnamese Restaurant	Indonesian Restaurant	South American Restaurant	Malay Restaurant	Salon / Barbershop	Gy Fitn Cer
4	Howard Beach	Italian Restaurant	Pharmacy	Bagel Shop	Chinese Restaurant	Spa	Sandwich Place	Fast Food Restaurant	Breakfast Spot	Shipping Store	Jewe St
5	Corona	Mexican Restaurant	Bakery	Deli / Bodega	Donut Shop	Check Cashing Service	Supermarket	Restaurant	Italian Restaurant	Food & Drink Shop	Sch

Cluster – 5 is the most happening cluster with many options for exploring.

V. CONCLUSION

Finally to conclude this project, we have got a small glimpse of how real life data-science projects look like. I've made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major districts of Tokyo and saw the results of segmentation of districts using Folium leaflet map.