**Coursera Capstone Project — the Battle of Neighbourhoods**

Indian Techie relocating to America

1. **INTRODUCTION**

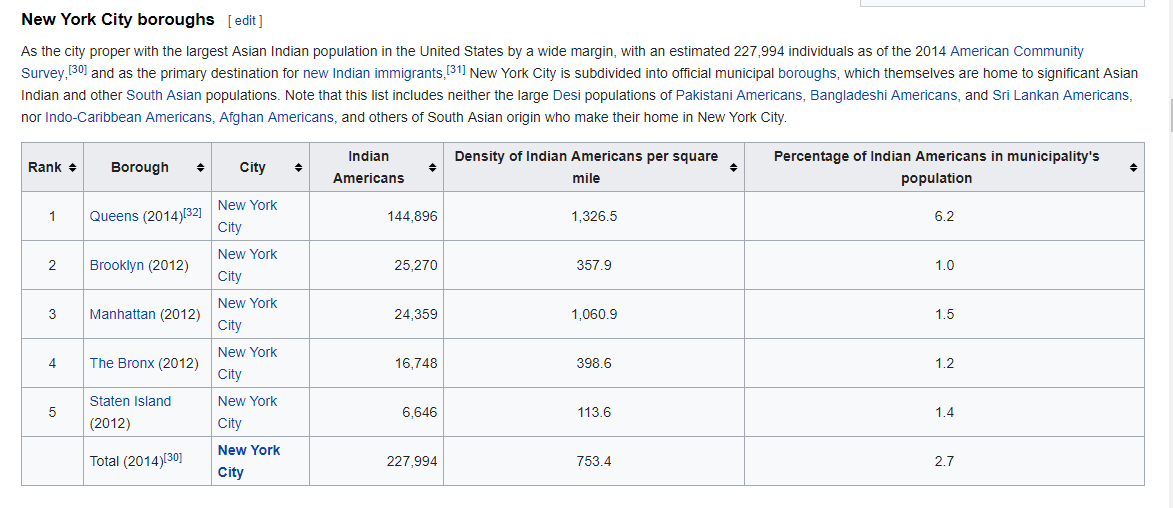
***Problem Statement:*** An Indian Techie has been offered a job in America. He needs to find out an accommodation and explore the neighbourhoods for his recreation, food etc. As he’s new to the city, this project is aimed to help him explore the neighbourhood and give insights into the same.

***Target Audience:*** Thisproject 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.

1. **DATA**
2. 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



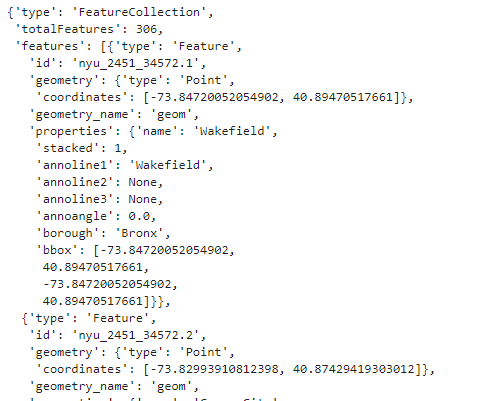
1. New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude.

**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.

1. Using FourSquare API we will find all venues for each neighbourhood.
2. geopy library to get the latitude and longitude values of New York City.
3. **METHODOLOGY**
4. ***Download and Explore New York City Dataset -***   
   <https://geo.nyu.edu/catalog/nyu_2451_34572>
5. ***Load and explore the data -***<https://cocl.us/new_york_dataset>

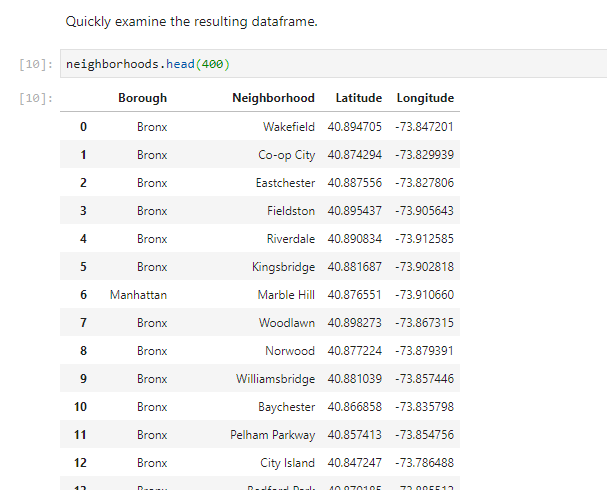
We get a JSON File with output as below:



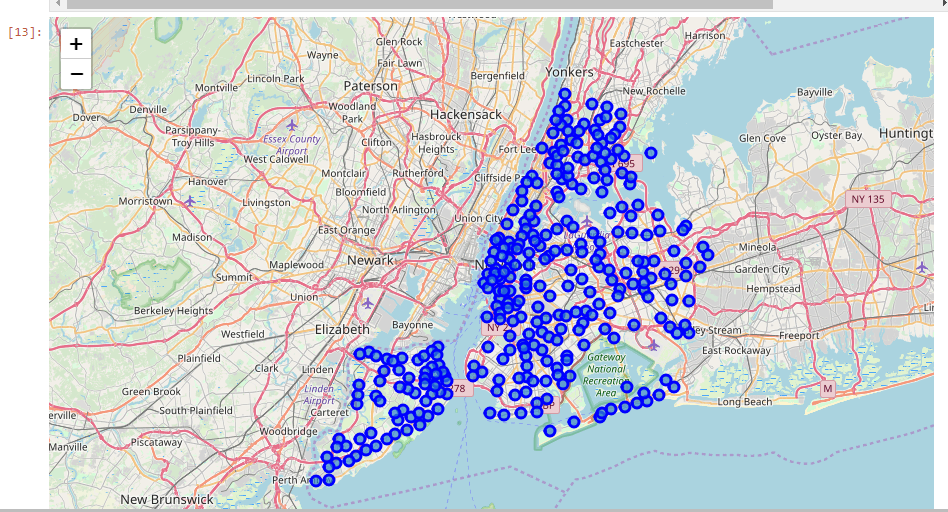
We get all the details of the neighbourhoods of New York City from this data

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

We get the following Output after loading the data into dataframe

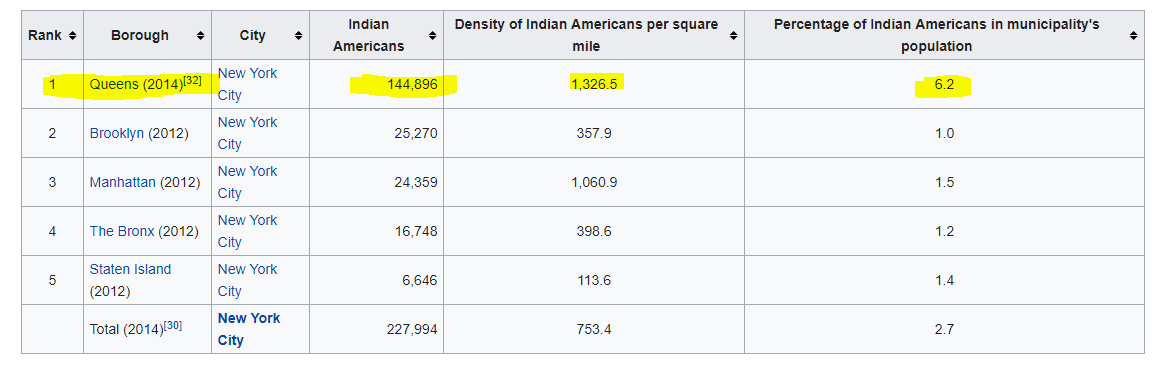


1. ***Create a map of New York with neighbourhoods superimposed on top***



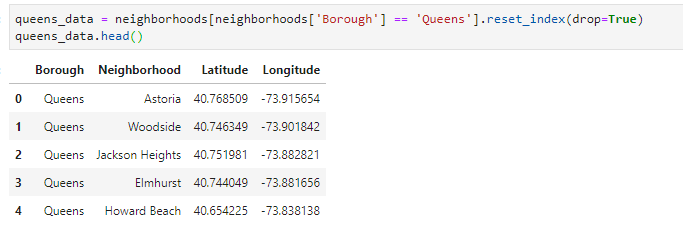
As seen from the Wikipedia report: <https://en.wikipedia.org/wiki/Indians_in_the_New_York_City_metropolitan_region>

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

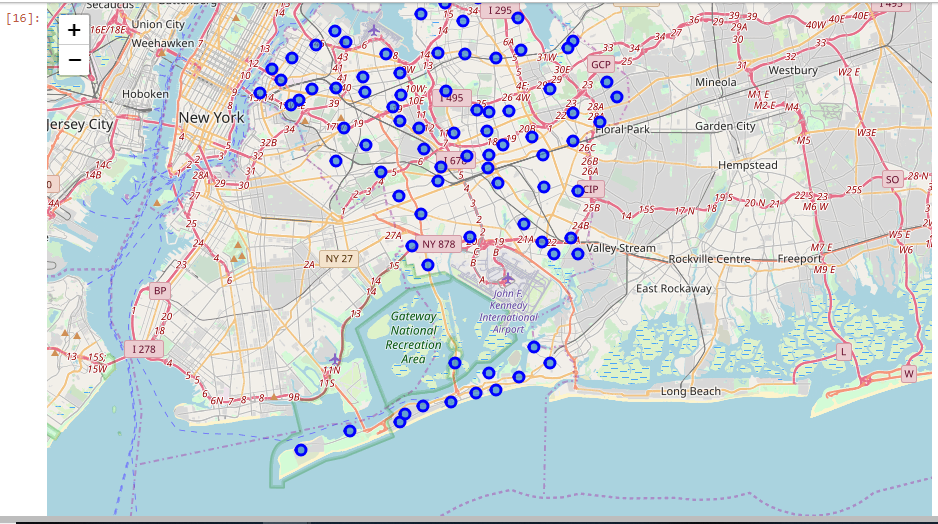


1. ***Analysing Queens Neighbourhood:***

We find Neighbourhood details of Queens with Latitude Longitude details



***Now we create map of Queens using latitude and longitude values***

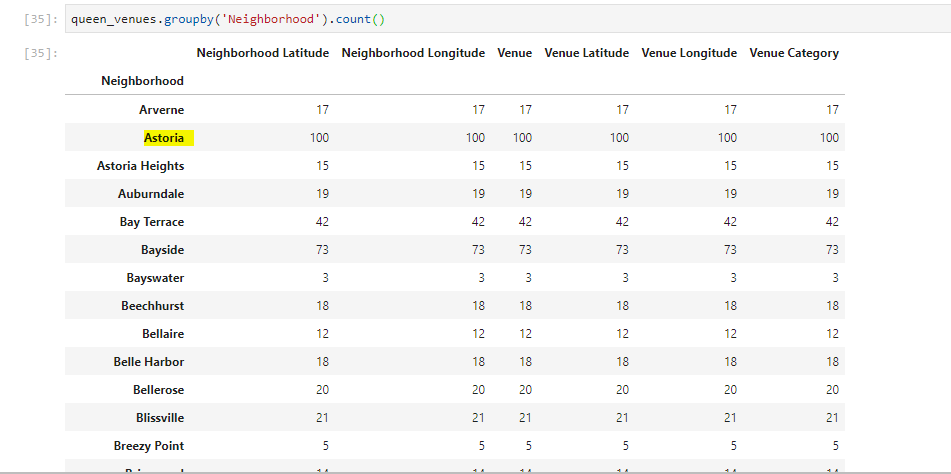


1. ***Next, we are going to start utilizing the Foursquare API to explore the neighbourhoods and segment them.***

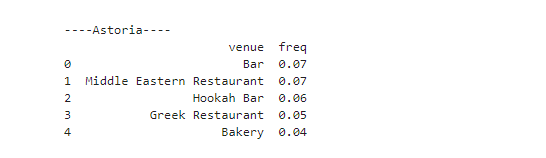


1. ***Explore Neighbourhoods in Queens***





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

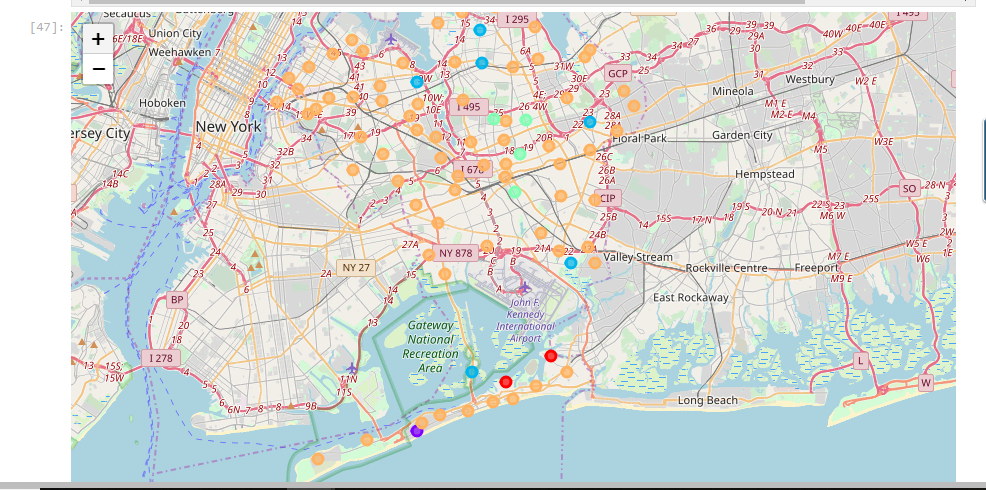




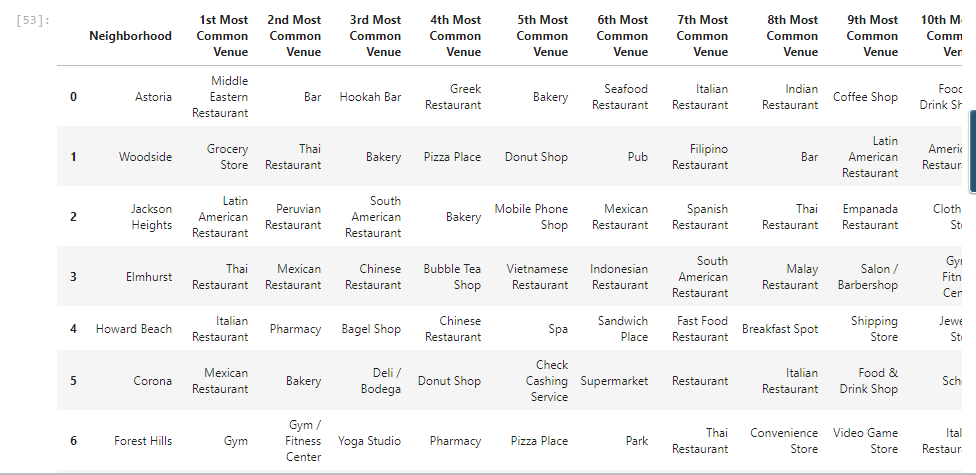
1. ***Now we cluster the Neighbourhoods***



1. ***Visualizing Clusters using Folium Map***



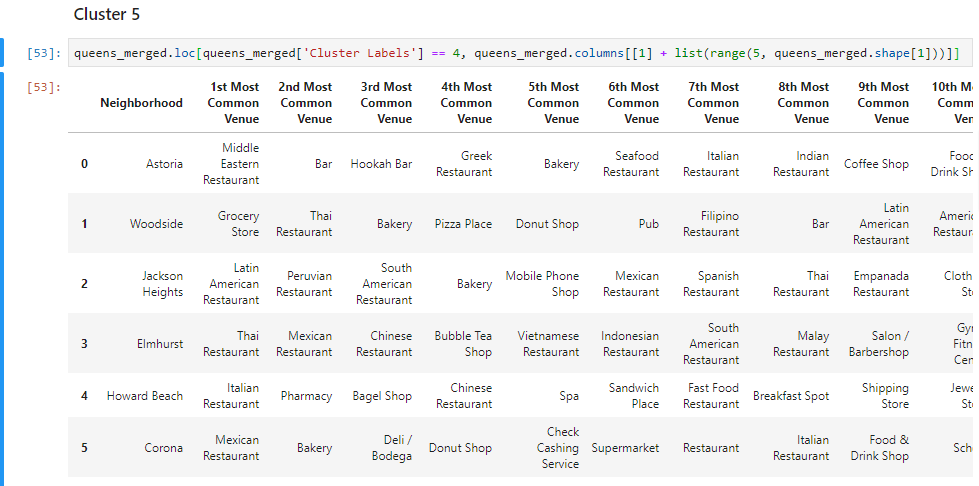
After k-means clustering, we found that Cluster – 5 has maximum traffic and top Venues are listed as below



1. **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 is the most happening cluster with many options for exploring.

1. **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.