The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

# Capstone Project - The Battle of Neighborhoods

# Introduction

## Business Problem

- ▶ For the people who loves to explore Indian cuisines in New York area, It is helpful to have below information readily available so that they can experience more variety of Indian dishes.
  - What are the various Indian cuisine based restaurant options one has, in New York neighborhoods
  - Recommend with a neighborhood which has more Indian restaurants

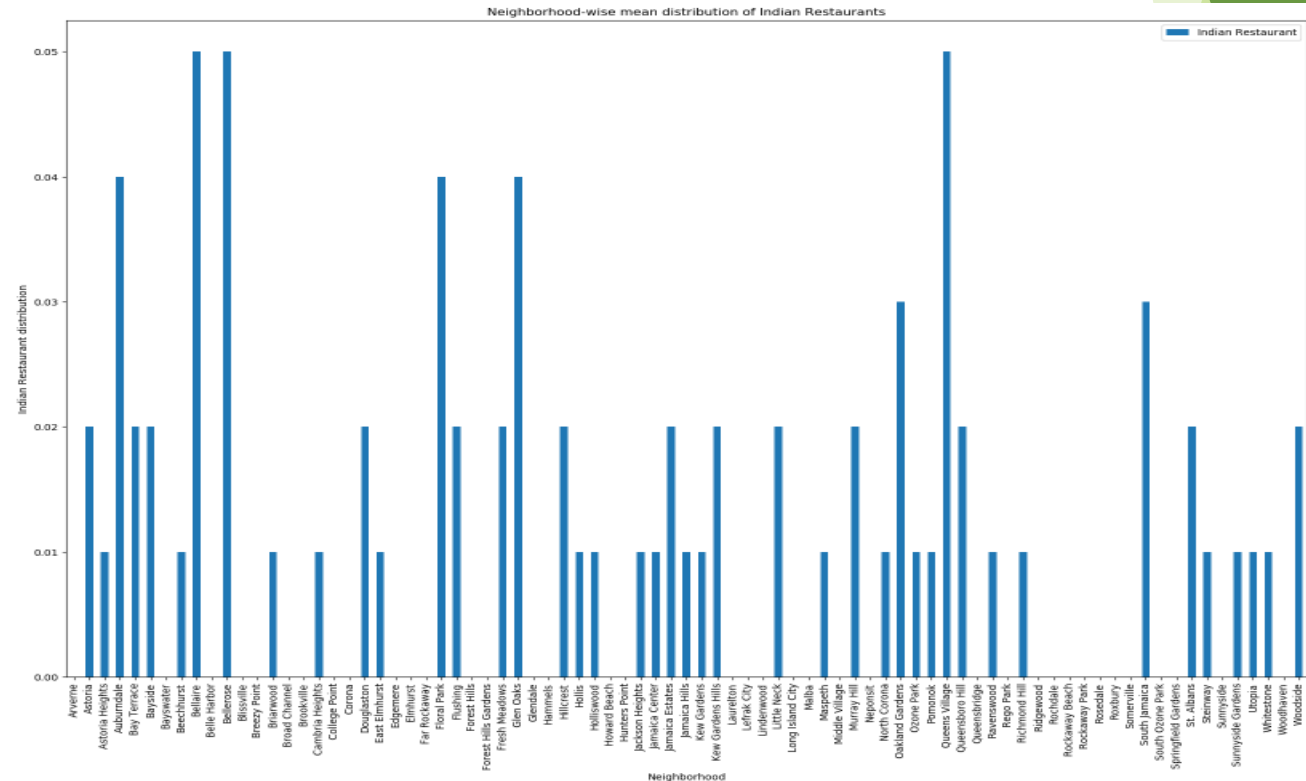
## ▶ Goal:

*This project aims at above aspects and recommend the people with neighborhoods having more Indian cuisine restaurants across Queens borough of New York city.*

# Data acquisition and cleaning

- ▶ To realize our goal, we need to have access to data with list of boroughs, neighborhoods, location co-ordinate information like latitude and longitude values.
  - Data source: [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
- ▶ Location data of Indian restaurants in New York city neighborhoods.
  - Data source: Foursquare database
  - Link: [api.foursquare.com](https://api.foursquare.com). We used 'explore' API end point in our project.
- ▶ We converted the JSON file of New York data into a data frame with features Borough, Neiborhood, latitude and longitude.
- ▶ We have data related to 5 Boroughs and 306 neighborhoods in New York. Out of this, we worked mainly on Queens borough data
- ▶ Used various Data preprocessing & analysis techniques as mentioned below
  - ▶ Creating new data frame from existing one, merging of data frames
  - ▶ Sorting of data points
  - ▶ One hot encoding to handle categorical variables
  - ▶ Plotting for visualization,
  - ▶ Creation of maps

- ▶ Queens borough has highest number of neighborhoods (80) which is important input for our further study
- ▶ There are a total of 264 unique venues categories in Queen borough across all neighborhoods including Pizza place, Caribbean and Indian restaurants, Chinese etc....

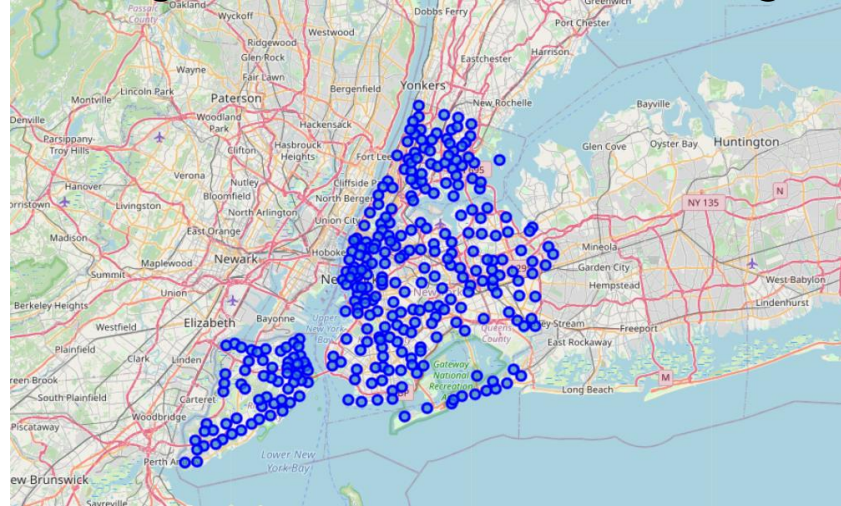


- ▶ The neighborhoods Bellaire, Bellerose and Queen Village has high number of Indian restaurants with mean of 0.05.
- ▶ These are followed by Auburndale, Flora park and Glen oaks with mean of 0.04

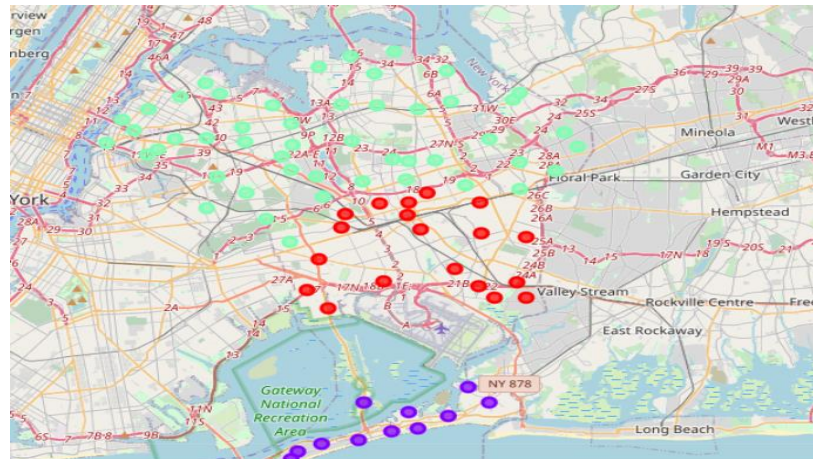
# Modelling

- We have chosen to split our neighborhood data set into 3 clusters.  
After clustering analysis below are the outcomes.
- **Cluster-3** has relatively more number of Indian restaurants as common venues, compared to other clusters. we can recommend this cluster of neighborhoods to people to explore Indian restaurants.
- **Cluster-2** has no Indian restaurants as any of top-10 common places across all neighborhoods. Hence it is not cluster of our interest.
- **Cluster-1** has very few Indian cuisine based restaurants and one can explore it if cluster-3 based restaurants are already experienced.

## Neighborhoods before Clustering



## Neighborhoods After Clustering



# Conclusion & Next actions

- ▶ As per analysis and results, we could recommend Bellaire, Bellerose and Queen Village neighborhoods & cluster area to users to explore Indian restaurants, as they have highest number of Indian restaurants presence.
- ▶ **Next actions:**
  - ▶ it is important to understand that we have tried to solve the problem by considering the biggest Borough of New York i.e. Queens. The results may vary if we consider all the other Borough's neighborhoods as well.

We may need to consider neighborhoods of other boroughs to get more accurate results.
  - ▶ We can extend this problem by recommending the users with best restaurants based on the rating of previous users

Thank you