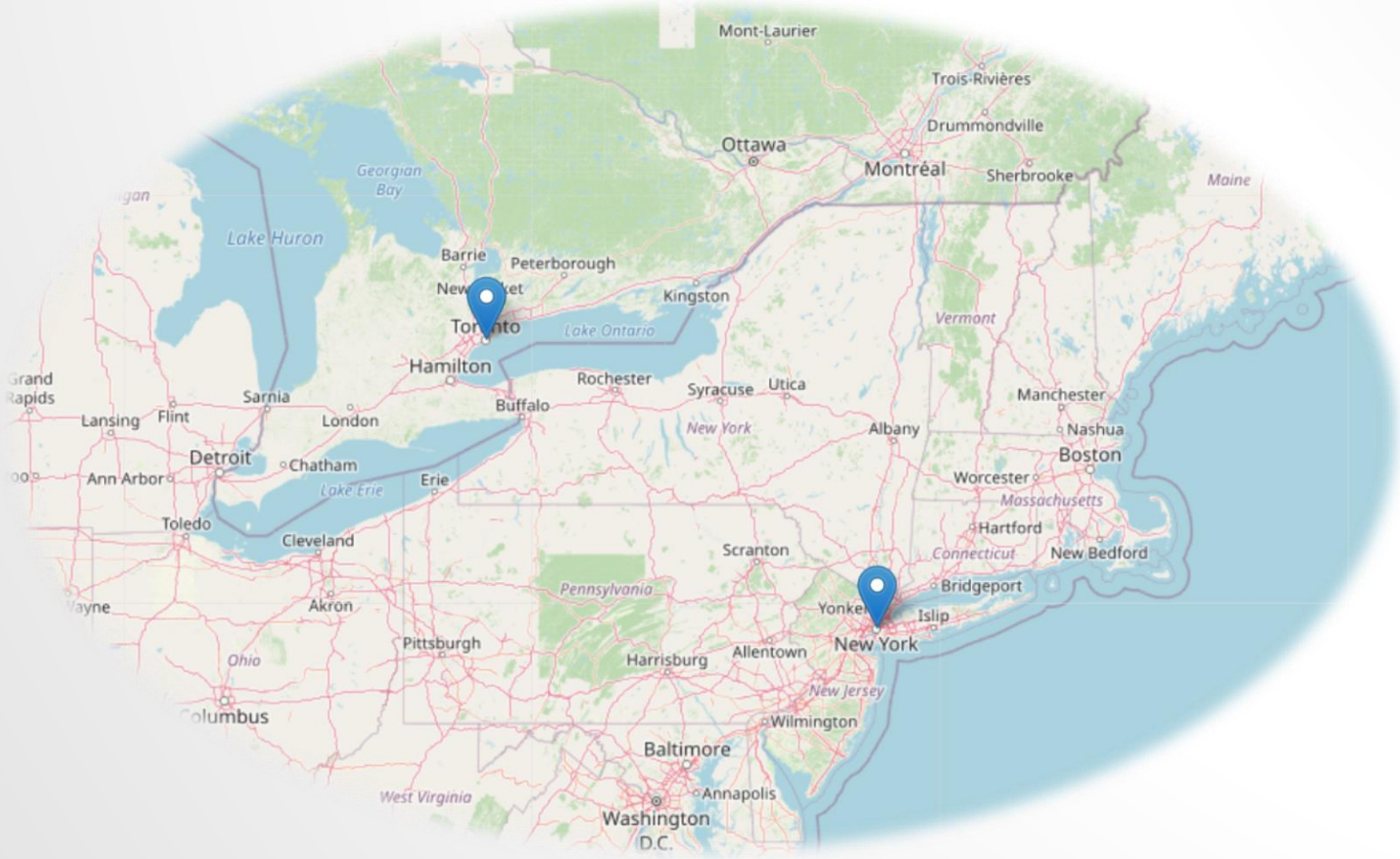
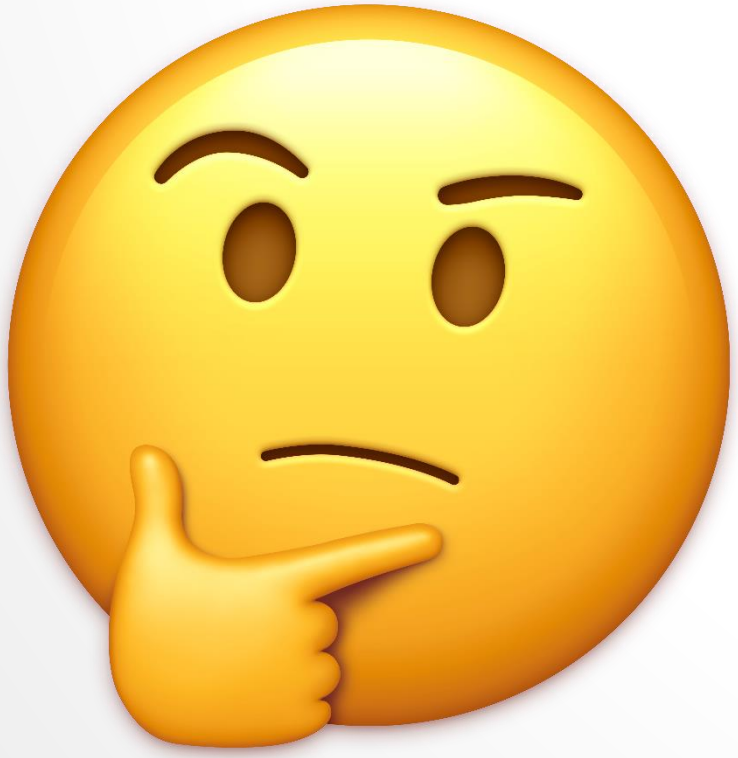
# A JOURNEY TO TORONTO AND NYC

## EXPLORING AND CLUSTERING NEIGHBORHOODS

# DOES DIFFERENT LOCATION MEAN DIFFERENT NEIGHBORHOODS?



# WHAT DO YOU FEEL IF YOU FIND SIMILAR NEIGHBORHOOD AS THE ONE  IN YOUR CITY?



# TORONTO VS QUEENS

- Toronto is the provincial capital of Ontario.
- It is the most populous city in Canada and the fourth most populous city in North America.
- The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada.

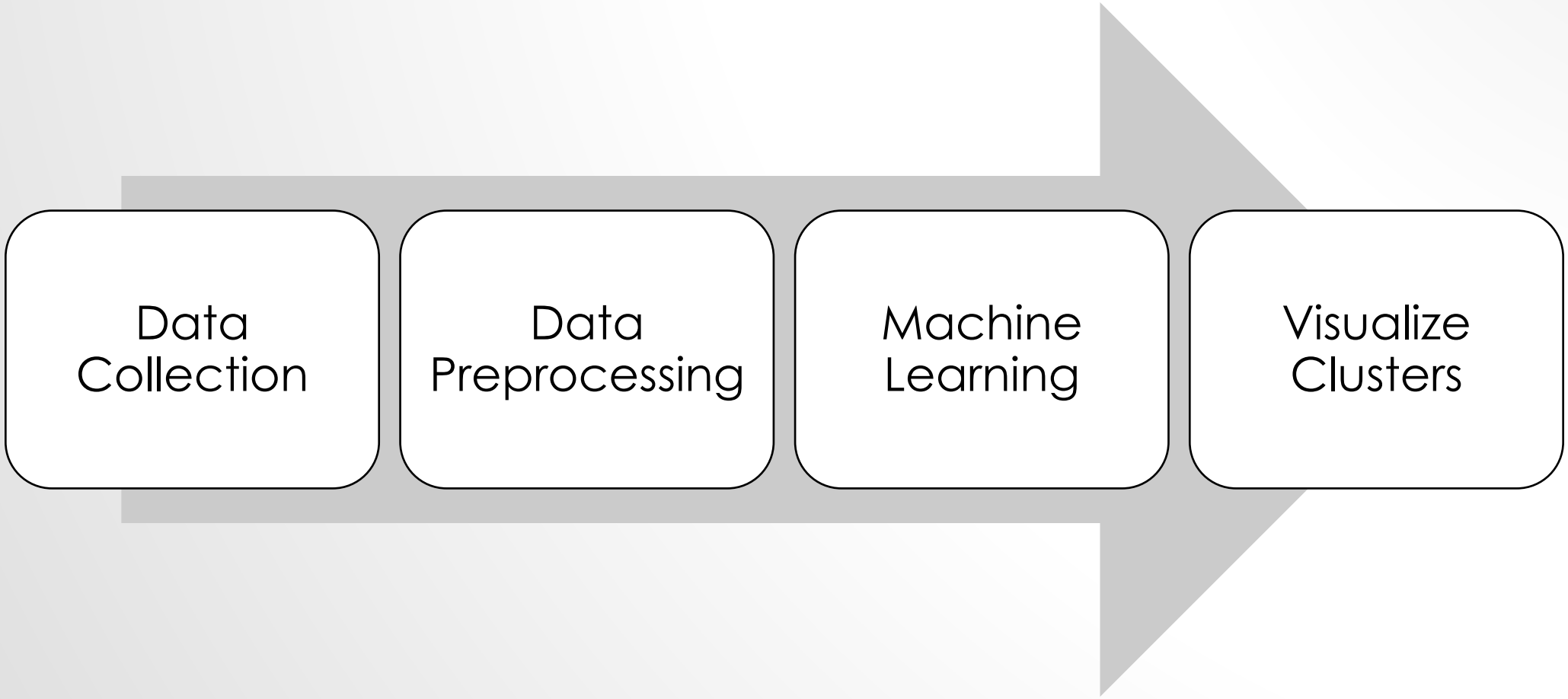
- Queens is a borough of New York City.
- It is one of the most ethnically diverse counties in the United States.
- It has the most diversified economy of the five boroughs of New York City.

# TORONTO VS QUEENS

# OBJECTIVE

- Analyze the neighborhoods of Toronto city and Queens in New York city
- Group them into similar clusters.
- Those clusters can be used to find out neighborhoods that are same as your current neighborhood or at least similar.

# THE JOURNEY HAS BEGUN



# DATA COLLECTION

**Wikipedia** Toronto City data

- which contains Postal Code, Boroughs and Neighborhoods

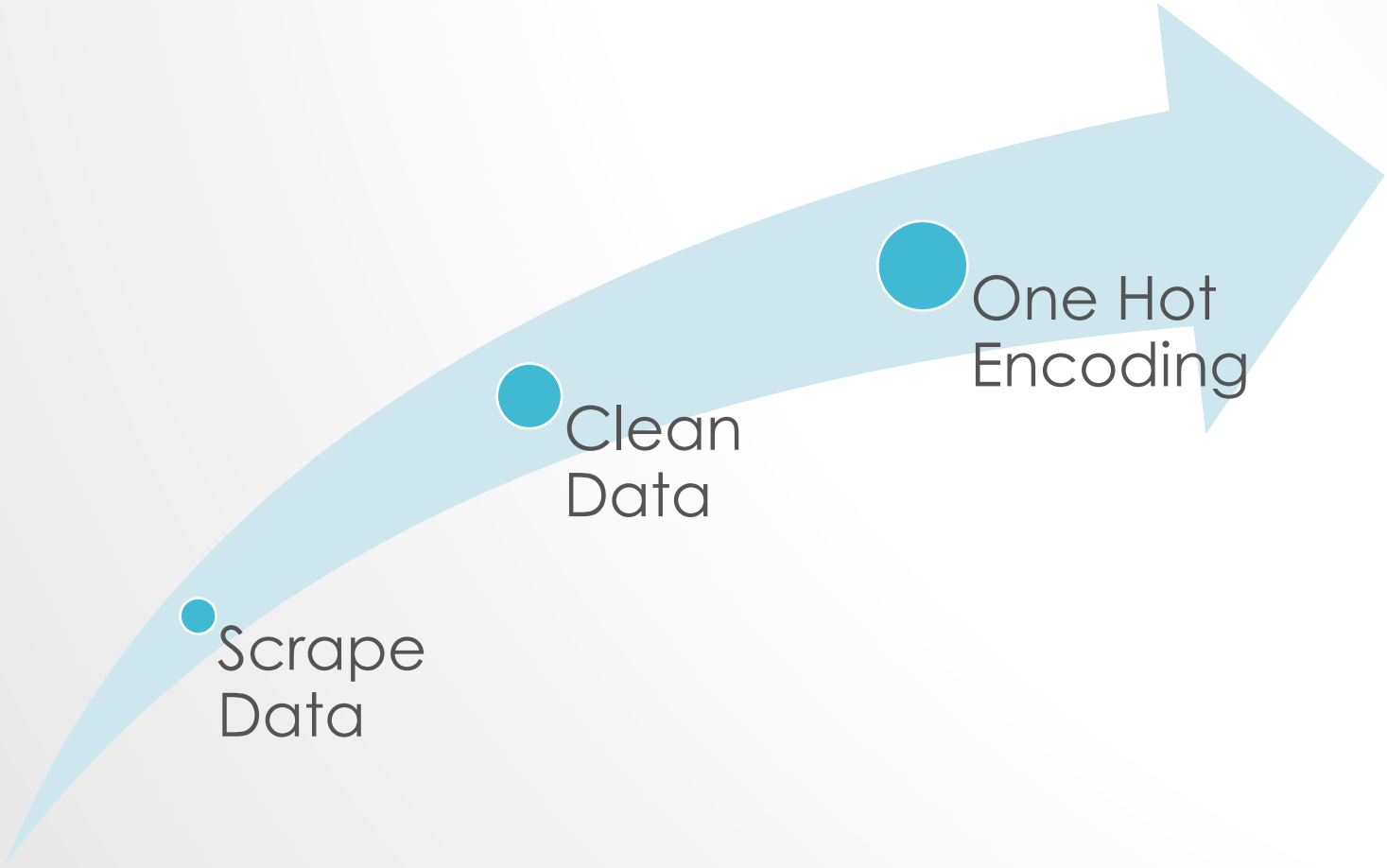
**json file** New York City data

- which contains list Boroughs, Neighborhoods along with their latitude and longitude.

**Foursquare** Venues data

- which contains Venue name, Latitude, Longitude, Category

# DATA PREPROCESSING

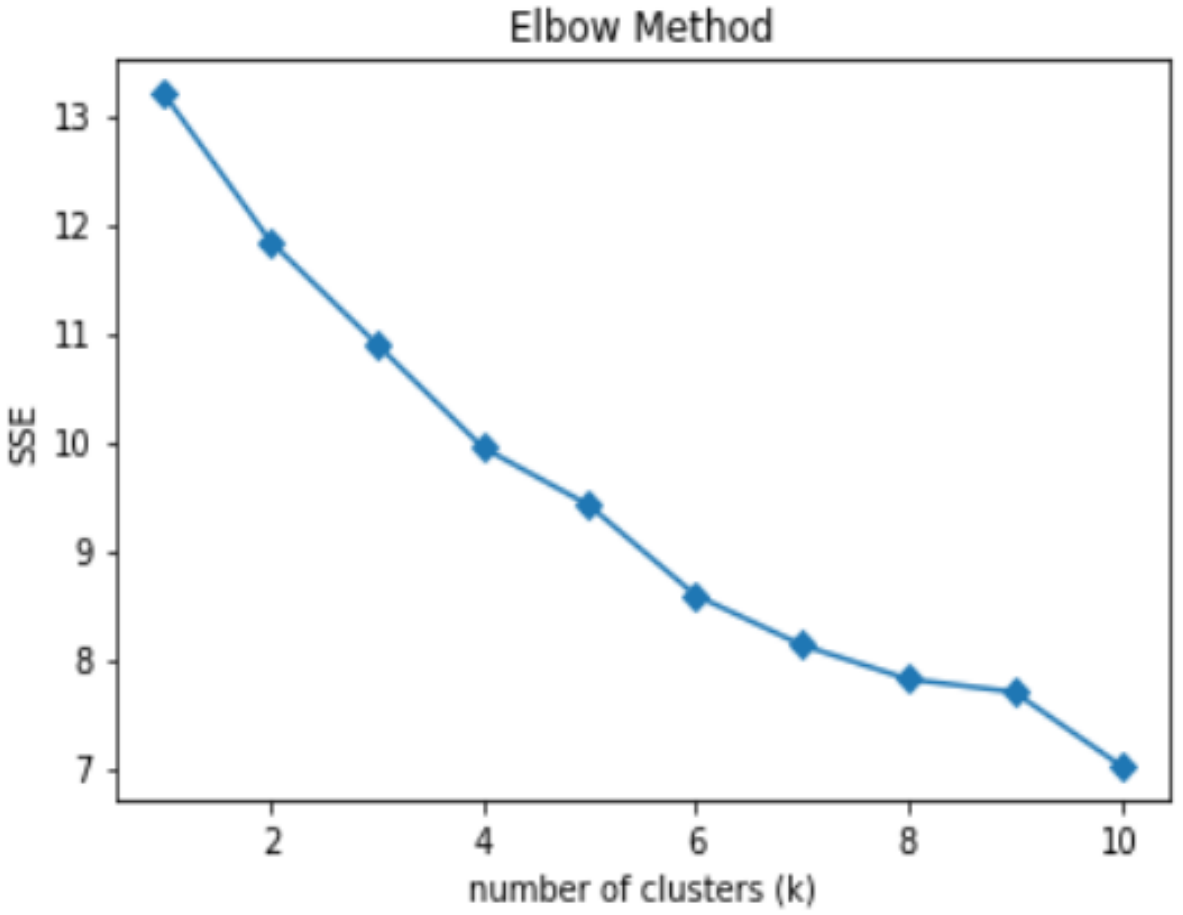


# MACHINE LEARNING

- K-means clustering is an unsupervised machine learning algorithm that creates clusters within your data, which can help you to discover categories or groups that you might not have seen on your own.
- To implement the k-means, it is very important to determine the optimal number of clusters ‘K’.
- Determine the value of optimal k with the Elbow Method and Silhouette Score.

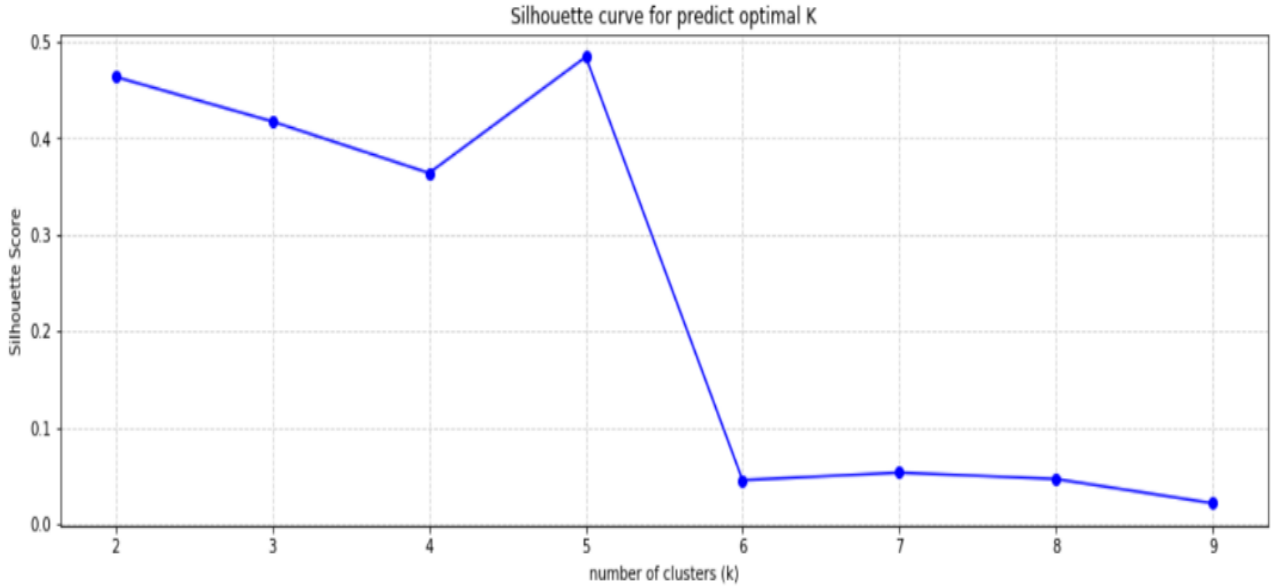
# THE ELBOW METHOD

- Calculates the sum of squared distance of samples to their closest cluster center for different values of “K”.
- The optimal number of clusters is the value after which there is no significant decrease.



# SILHOUETTE SCORE

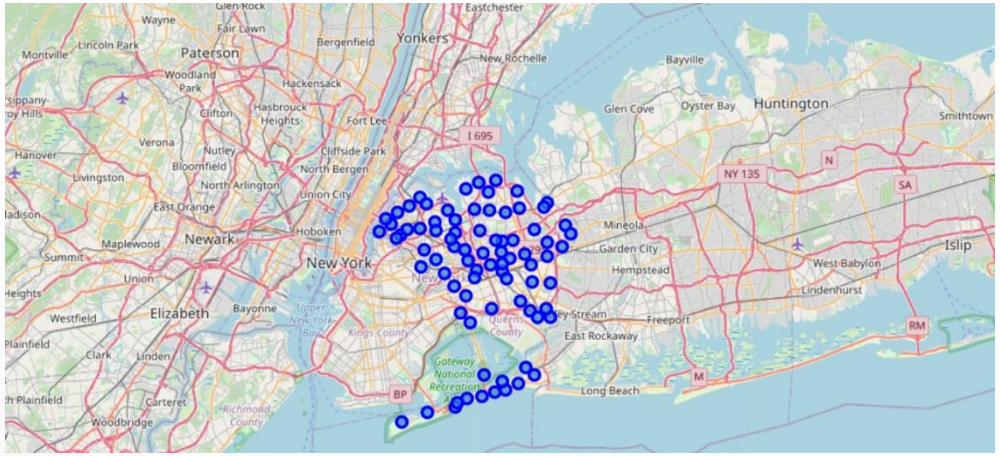
- The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation).



# BEFORE CLUSTERING

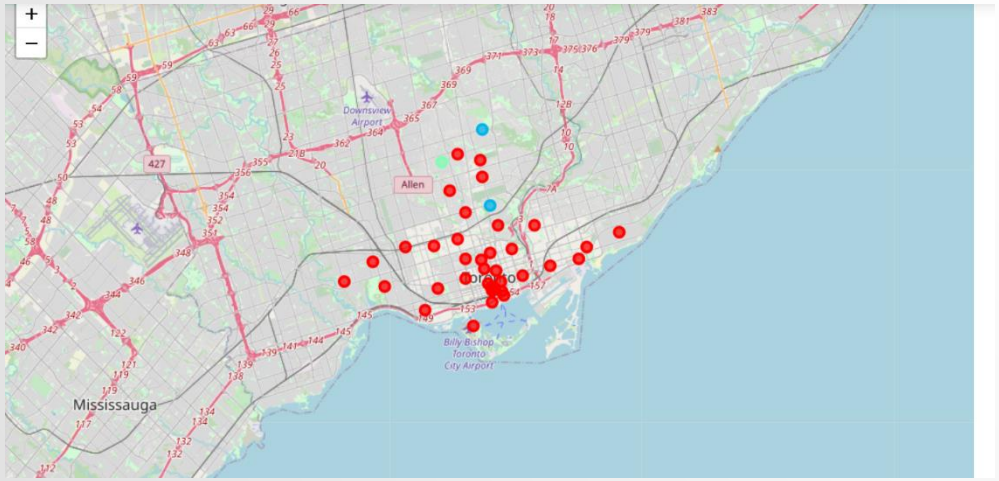
## TORONTO


## QUEENS

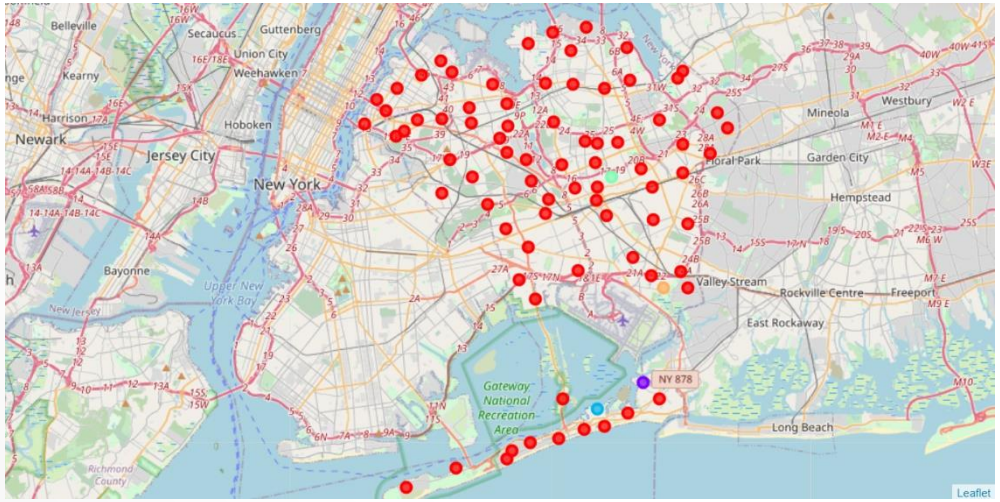


# VISUALIZE CLUSTERS

## TORONTO




## QUEENS



# LUCKY CLUSTER

**Cluster 3 (Light Blue "Indigo")**

```
qu_tor_merged.loc[qu_tor_merged['Cluster Labels'] == 2, qu_tor_merged.columns[[0, 1] + list(range(5, qu_tor_merged.shape[1]))]]
```

| | Borough | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 63 | Queens | Somerville | Park | Women's Store | Dumpling Restaurant | Distribution Center | Dive Bar | Dog Run | Doner Restaurant | Donut Shop | Dosa Place | Dry Cleaner |
| 85 | Central Toronto | Lawrence Park | Park | Bus Line | Swim School | Women's Store | Dumpling Restaurant | Dog Run | Doner Restaurant | Donut Shop | Dosa Place | Dry Cleaner |
| 89 | Central Toronto | Moore Park, Summerhill East | Park | Trail | Women's Store | Diner | Distribution Center | Dive Bar | Dog Run | Doner Restaurant | Donut Shop | Dosa Place |

# DISCUSSION

- As seen in the table above, if a person wished to move from Queens to Toronto.
- If a person's current location were in the Neighborhood of Somerville in Queens, which has venues like Parks, Dog Runs, stores and Restaurants nearby, the person, would like to relocate to a neighborhood like Lawrence Park or Moore Park, Summerhill East in Central Toronto which also has venues like Parks, Dog Runs and Restaurants.
- This is just one example of how our data analysis can help people relocate from one city to another which similar to their current localities or other neighborhood in the same city.

# CONCLUSION

- there are many real-life problems or scenarios where data can be used to find solutions to those problems.
- Like seen in the example above, data was used to cluster neighborhoods in Toronto and Queens NYC based on the most common venues in those neighborhoods.
- Similarly, data can also be used to solve other problems, which most people face in metropolitan cities.


Q
A

# THANK YOU