

# Capstone Project - The Battle of Neighborhoods (Week 1)

## INTRODUCTION

Prepared By –  
Pranav Narayanan

# Introduction

- ▶ The City of New York is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.
- ▶ Due to this influence, many restaurants are operational in the city catering to a wide variety of audience. This also means that there is a huge demand for the supply of items needed to run these restaurants.
- ▶ In this analysis, we will discuss the optimal place to start a restaurant supply chain shop which can cater to maximum restaurants.

# Target Audience

- ▶ The target audience for this analysis would be any person or company who is interested in opening a restaurant supply store in New York.
- ▶ After the analysis they would obtain the optimum position to open the store which will minimize the cost of transport as well as generate sufficient income due to its prime location.

# Business Problem

- ▶ My client is eager to open a restaurant supply store in New York area. Excited by the prospect that there are thousands of restaurants to cater to, they want to find the optimum location within New York city to place their supply store. This will in fact be the exact center of our relevant cluster and equidistant from outer lying restaurants thus minimizing transport time and other charges.

# Data Selection

- ▶ The primary objective is to retrieve the data pertaining to the Boroughs and neighborhoods in New York city along with their geographical coordinates.
- ▶ We use the given website to obtain the GeoJSON file which contains the data we need - <https://geo.nyu.edu/catalog/nyu-2451-34572>
- ▶ Below is how the data looks like

	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

# Data Selection

- ▶ Next objective is to retrieve the list of all food related establishments in New York. This is achieved using Foursquare API. Below is a sample after the operation is completed –

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Margaritaville Restaurant and Lounge	40.898564	-73.837016	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898083	-73.850259	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Essa Deli	40.892181	-73.854427	Deli / Bodega
4	Wakefield	40.894705	-73.847201	Big Daddy's Caribbean Taste Restaurant	40.899767	-73.857135	Caribbean Restaurant

# Methodology

- We then wrangle the data using one-hot encoding and then group it by Neighborhood and calculate the mean value for each venue. This is fed as the dataframe for clustering using K-Means.

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bangladeshi Restaurant	Bar	Bath House
0	Allerton	0.0	0.0	0.020000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.060000	0.0	0.0	0.0
1	Annadale	0.0	0.0	0.107143	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.071429	0.071429	0.0	0.0	0.0
2	Arden Heights	0.0	0.0	0.096774	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.032258	0.032258	0.032258	0.0	0.0	0.0
3	Arlington	0.0	0.0	0.055556	0.0	0.0	0.0	0.0	0.027778	0.0	0.0	0.000000	0.027778	0.027778	0.0	0.0	0.0
4	Arrochar	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.000000	0.066667	0.000000	0.0	0.0	0.0

# Methodology

- ▶ Using silhouette\_score we obtain the optimum number of clusters the data can be segmented into. Below is the silhouette\_score results for different cluster sizes –
- ▶ silhouette\_score results : {2: 0.17743357998254314, 3: 0.14066439991387775, 4: 0.12914863106062802, 5: 0.1328577503914053, 6: 0.11736892508853244, 7: 0.12027984810897324, 8: 0.08326902182481445, 9: 0.07694941825109619}
- ▶ We can see that cluster of size 2 has maximum silhouette\_score



# Methodology

- We now plot the clusters in folium using different colors –



# Methodology

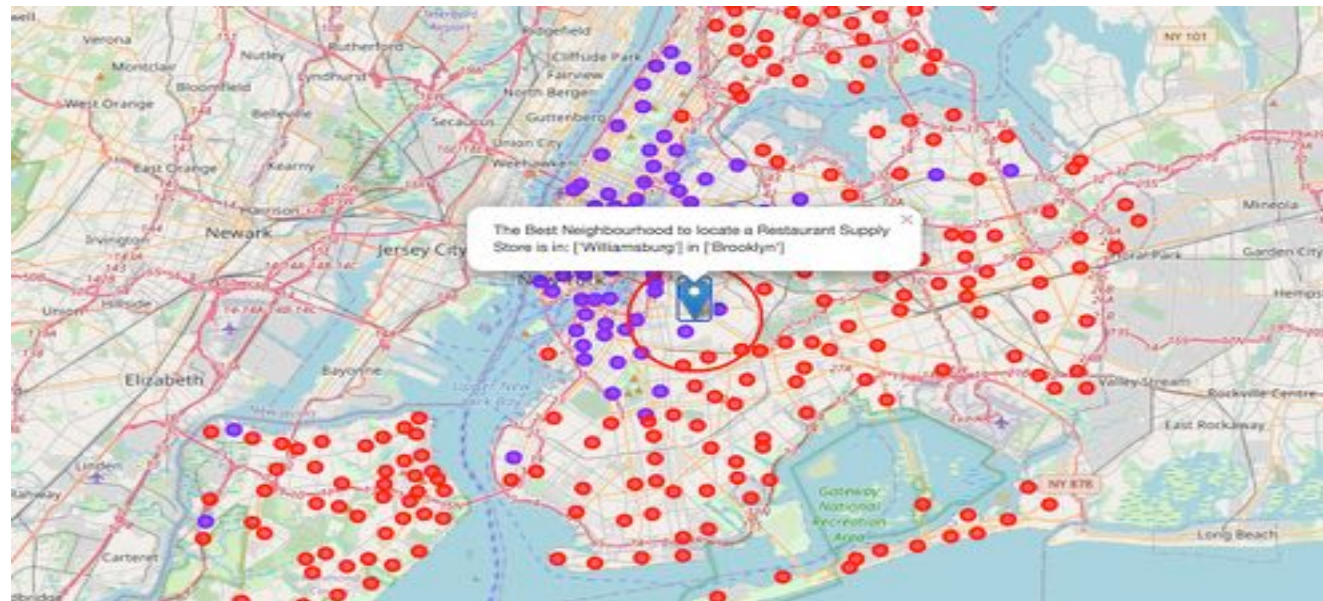
- ▶ From the previous map, we can visually observe that Cluster 0(in red) has the maximum number of values. To find the optimal place for locating a restaurant supply store, we find the centroid of the geographical coordinates of the points in Cluster 0.
- ▶ We obtain the centroid by averaging the latitude and longitude information.
- ▶ Centroid obtained is as follows –
- ▶ Latitude : 40.69394045210022, Longitude : -73.93700185136717

# Methodology

- ▶ Once the latitude and longitude of centroid has been obtained, we use OpenCageGeocode to reverse lookup the coordinates and find the address and other details.
- ▶ We obtain the following information from reverse lookup –
- ▶ `'building': 'Building 7', 'city': 'New York', 'city_district': 'Kings County', 'continent': 'North America', 'country': 'United States', 'country_code': 'us', 'house_number': '93', 'postcode': '11206', 'residential': 'BEDFORD STUYVESANT/ EXPANDED STUYVEVSANT ' HEIGHTS HISTORIC DISTRICT', 'road': 'Lewis Avenue', 'state': 'New York', 'state_code': 'NY', 'suburb': 'Brooklyn', 'confidence': 10, 'formatted': 'Building 7, 93 Lewis Avenue, New York, NY 11206, United States ' of America', 'geometry': {'lat': 40.6941276, 'lng': -73.9369924}`

# Results

- We plot the centroid on the map using folium.



# Results

- ▶ As observed in previous map, the geographical centroid offers the optimal position for locating a restaurant supply store since it is equidistant from outlying points and can save cost involved in transport and other logistical activities.
- ▶ Placing the store within the vicinity of the centroid will generate sufficient revenue for the client and help in cutting costs.

# Conclusion

- ▶ I feel confident with the recommendation I have given my client as it is backed up with demonstrated data analysis. While nothing can ever be 100% certain they will certainly be better informed than they were prior to this comprehensive analysis.
- ▶ Much more inference can be obtained with more work. A potential side business for my client might be assisting new restaurant owners where they might locate a new restaurant, who their competition is and who their clientele might be.