

A decorative graphic on the left side of the slide consisting of white lines and circles on a blue gradient background, resembling a circuit board or neural network.

# CAPSTONE PROJECT – THE BATTLE OF NEIGHBORHOODS

EDER BRAZ

# 1.INTRODUCTION

- Stakeholder want to create a Brazilian steakhouse restaurant.
- which neighborhood the restaurant would gain the most attention from customers and face the least competition from other steakhouses.

## 2. DATA DESCRIPTION

- New York Dataset that contains the 5 boroughs and all the neighborhoods that exists in each borough as well as the latitude and longitude coordinates of each neighborhood.
- The Foursquare API will be used to explore neighborhoods in New York, more specifically, we will be using the explore function to get the most common venue categories in each neighborhood.

# 3. METHODOLOGY

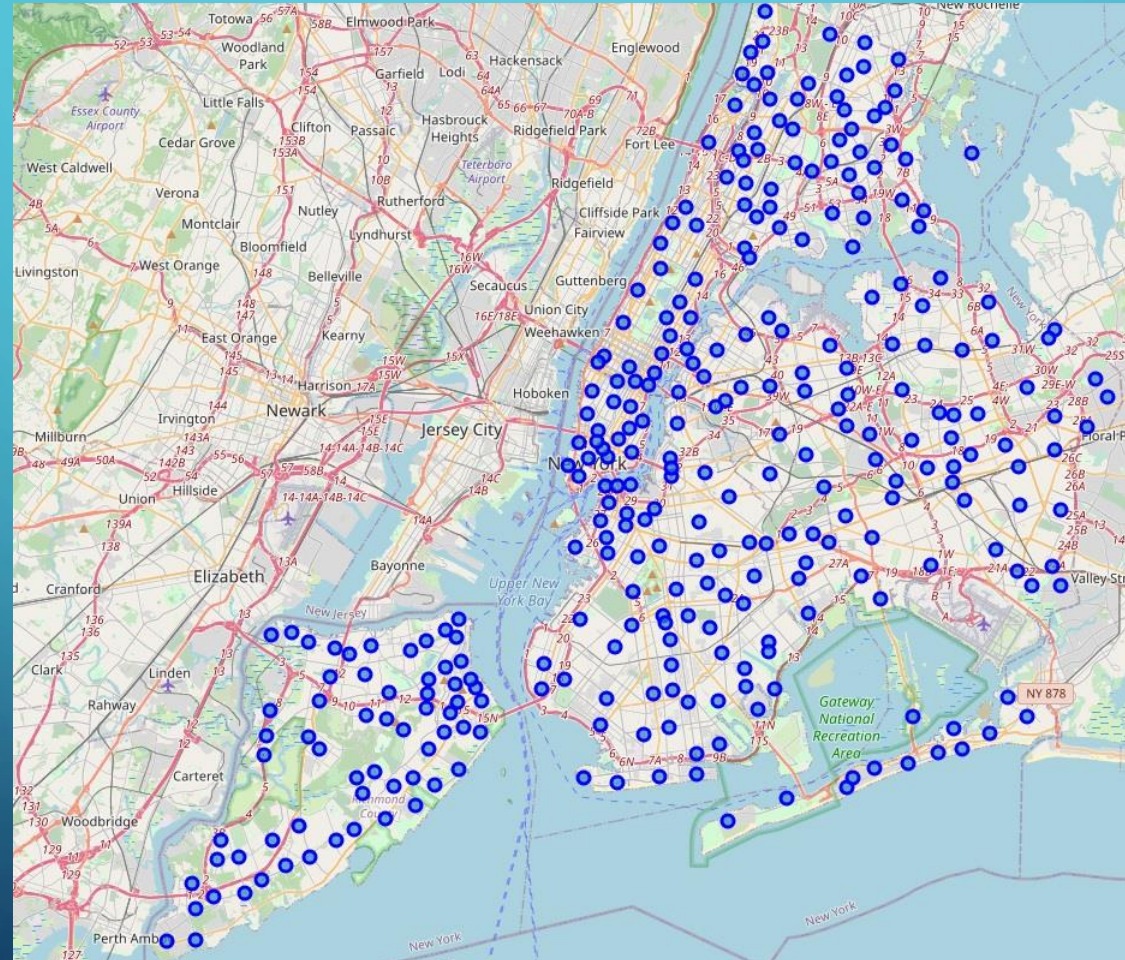
## 3.1 Extracting data and transforming

- Dataset downloaded from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) Then converted into a Pandas Dataframe.

	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



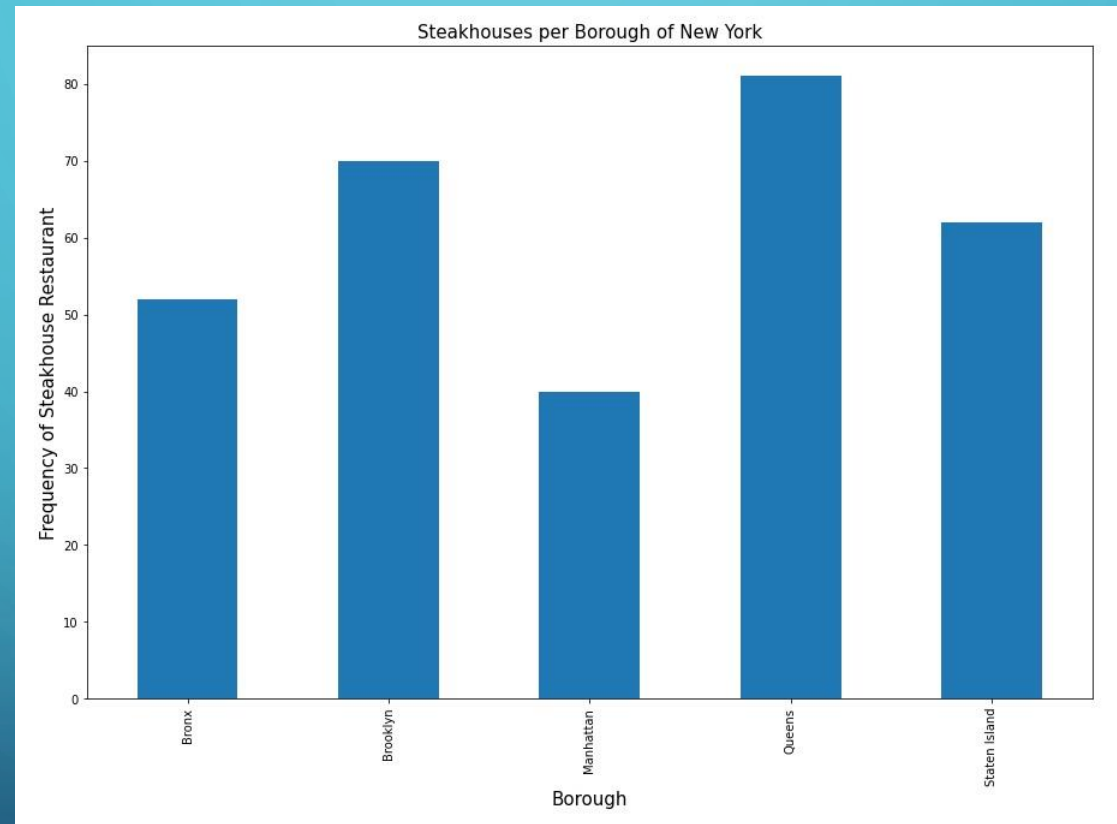
# Using Python folium library to plot all neighborhoods on map



# Result from Foursquare API search

	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	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

# How many steakhouses each borough have?



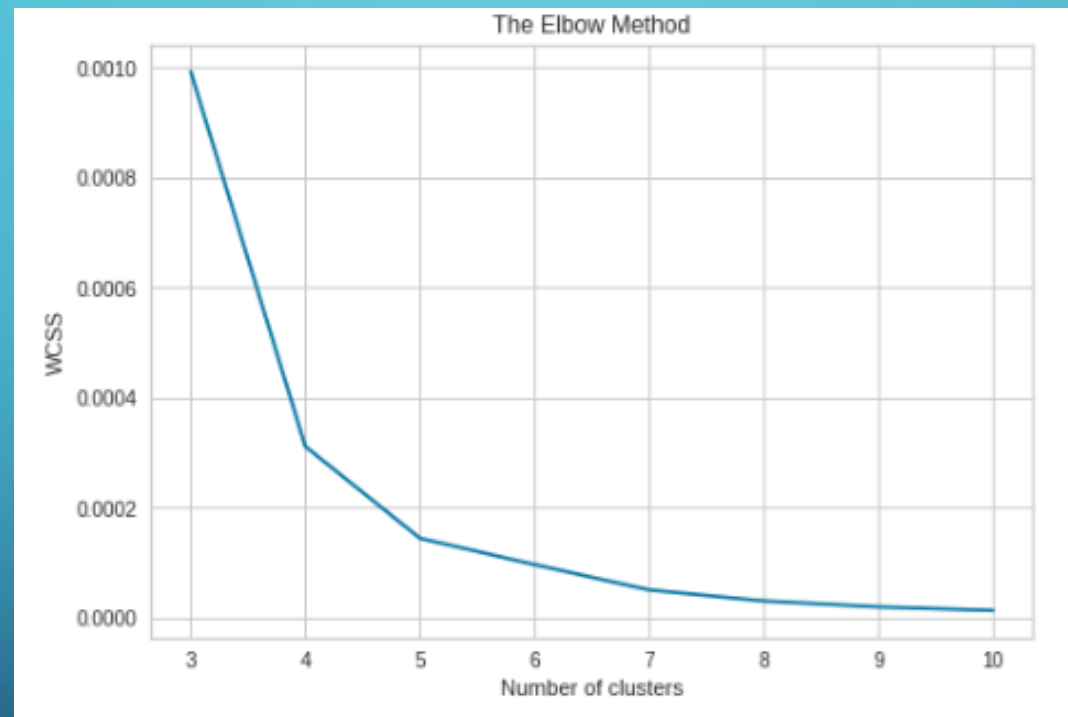
## 3.2 CLUSTERING

- K-Means algorithm to cluster the neighborhoods.
- Elbow method to determine the optimal value of K.



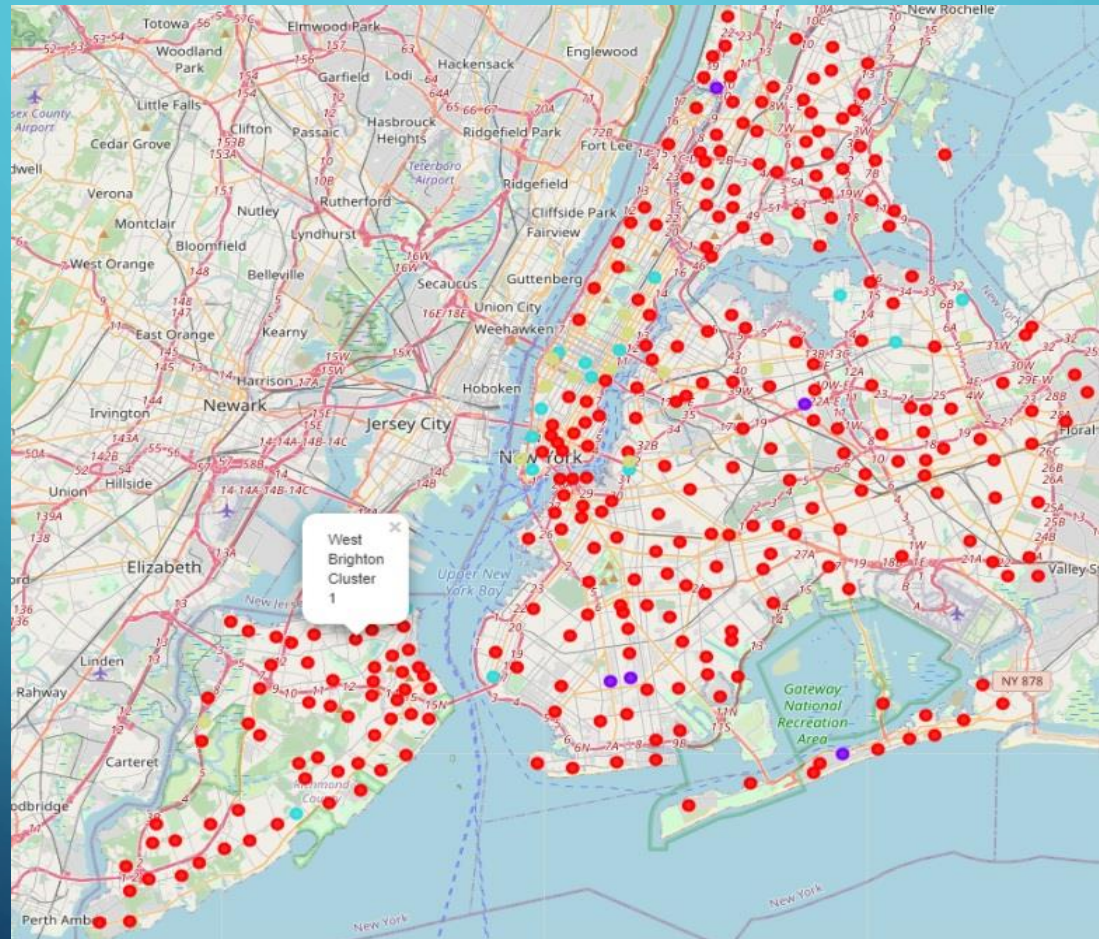
# Using the elbow method

k=4 was the value defined for K



## 4. Results

### Plotting the clustered map



Cluster 1 has all neighborhoods that don't have  
steakhouses

	Borough	Neighborhood	Latitude	Longitude	Steakhouse	Cluster Labels
0	Bronx	Wakefield	40.894705	-73.847201	0.0	0
205	Staten Island	Grymes Hill	40.624185	-74.087248	0.0	0
204	Staten Island	West Brighton	40.631879	-74.107182	0.0	0
203	Staten Island	Rosebank	40.615305	-74.069805	0.0	0
202	Staten Island	Stapleton	40.626928	-74.077902	0.0	0
201	Staten Island	New Brighton	40.640615	-74.087017	0.0	0
199	Queens	Forest Hills Gardens	40.714611	-73.841022	0.0	0
206	Staten Island	Todt Hill	40.597069	-74.111329	0.0	0
198	Queens	North Corona	40.754071	-73.857518	0.0	0
196	Queens	Brookville	40.660003	-73.751753	0.0	0



# MAP WITH BEST THE LOCATIONS



# CONCLUSION

- The stakeholders wanted to avoid unnecessary competition against existing steakhouses.
- They should avoid cluster 2 and 3 that are the neighborhood with more steakhouses.
- Cluster 4 has little competition and a very interesting neighborhood (Jackson Heights, Queens) with Latin and South American restaurants.
- Cluster 1 has 65 neighborhoods without competition.
- Recommended zones should therefore be considered only as a starting point for more detailed analysis.