THE BATTLE OF NEIGHBORHOODS NY

1- INTRODUCTION

- In this project we will use the Foursquare API to explore neighborhoods in New York City and group the neighborhoods into clusters. We will use the k-means clustering algorithm to complete this task. Finally, we will use the Folium library to visualize the neighborhoods in New York City and their emerging clusters.

- This project is interesting for people or companies that would like to start a business in the city and would like to know the distribution of businesses across the city.

2- DATASET

- We are going to use the dataset available with the NY data, with this data we are going to create a clustering model for the entire city.

- The data consist in the name of the neighborhood and the location values. Using the Foursquare API will create a dataset in which we will have the previous data divided in streets and the different business in those streets.

|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Allerton	Pizza Place	Chinese Restaurant	Deli / Bodega	Supermarket	Fried Chicken Joint
1	Annadale	American Restaurant	Pharmacy	Train Station	Liquor Store	Park
2	Arden Heights	Deli / Bodega	Pizza Place	Bus Stop	Coffee Shop	Pharmacy
3	Arlington	Deli / Bodega	Intersection	American Restaurant	Bus Stop	Boat or Ferry
4	Arrochar	Bus Stop	Italian Restaurant	Deli / Bodega	Pizza Place	Supermarket

3- METHODOLOGY

- The final goal of the project is to cluster the locations, firstly we have an induvial analysis of each of the neighborhood in which we are able to see the frequency of the different business, with this analysis we can observe the different division that we can have afterwards.

- Why K-Means?

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. So, for this project, this algorithm works perfectly.

4- RESULTS

- As final result we can observe the city map with the different datapoints assigned to the corresponding cluster.
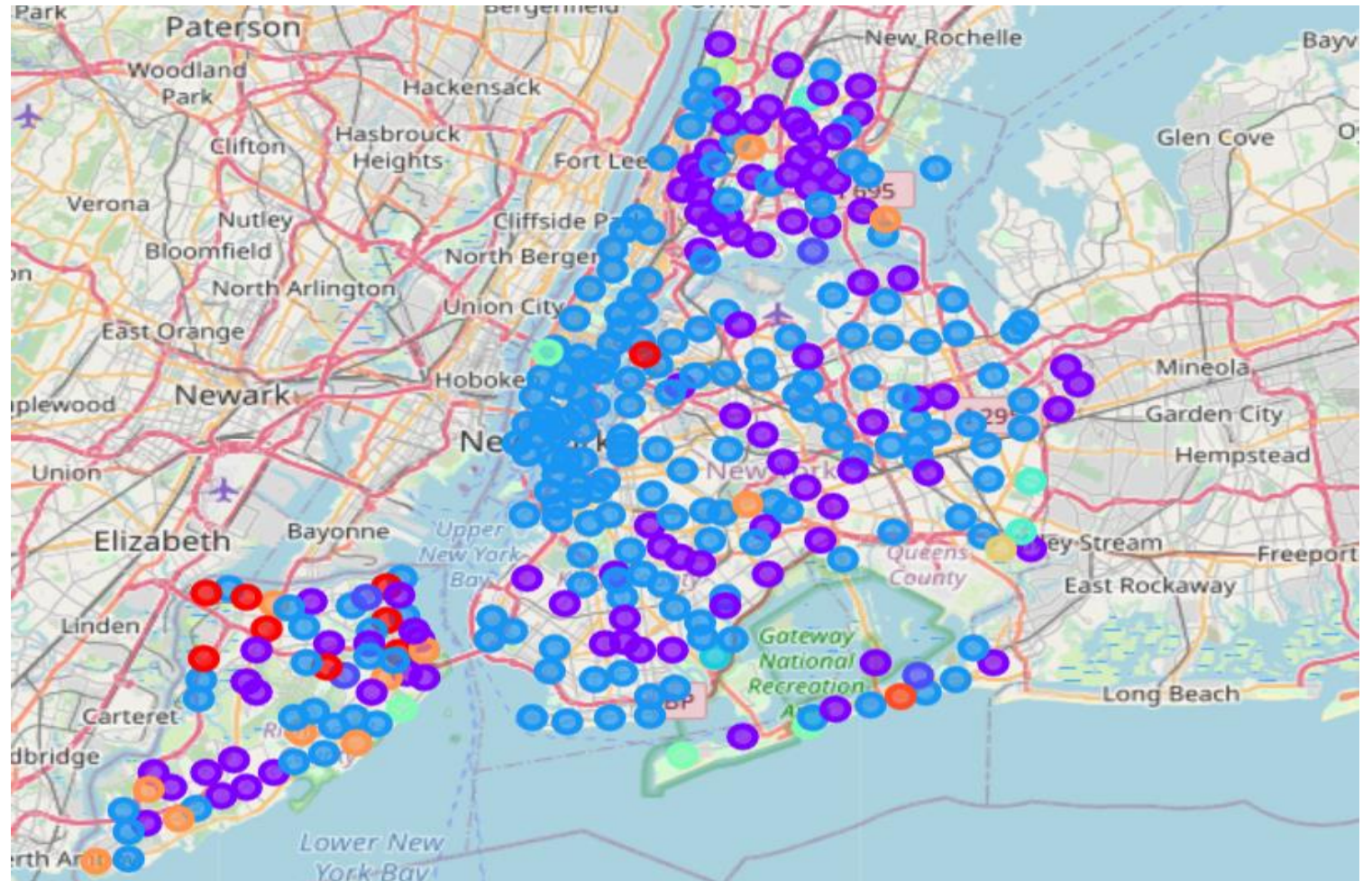
5- CONCLUSION

- In this project we used the Foursquare API to explore neighborhoods in New York City and group the neighborhoods into clusters. We used the k-means clustering algorithm to complete this task. Finally, we used the Folium library to visualize the neighborhoods in New York City and their emerging clusters.