The Battle of Neighborhoods Report

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.

1. Data acquisition and cleaning

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.

The table format is the following:



1. 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.

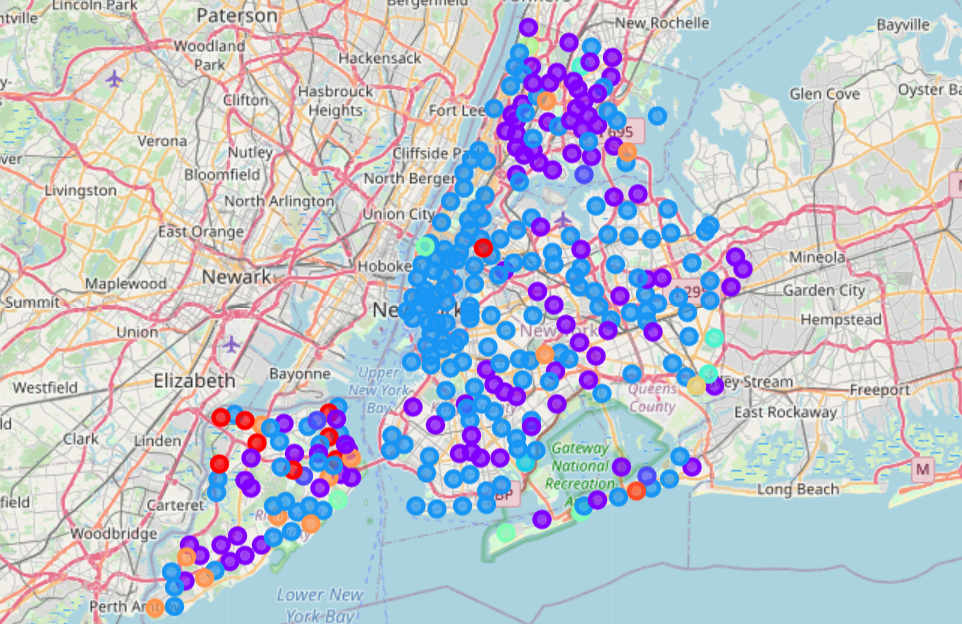
Secondly, we apply the K-Means algorithm, using k=5.

k-means clustering is a method of [vector quantization](https://en.wikipedia.org/wiki/Vector_quantization), originally from [signal processing](https://en.wikipedia.org/wiki/Signal_processing), that aims to [partition](https://en.wikipedia.org/wiki/Partition_of_a_set) n observations into k clusters in which each observation belongs to the [cluster](https://en.wikipedia.org/wiki/Cluster_(statistics)) with the nearest [mean](https://en.wikipedia.org/wiki/Mean) (cluster centers or cluster [centroid](https://en.wikipedia.org/wiki/Centroid)), serving as a prototype of the cluster. So, for this project, this algorithm works perfectly.

1. Results

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

I will not enter in detail as the map is very easy to interpret.



1. 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.