The Battle of Neighborhoods - Coursera Capstone Project

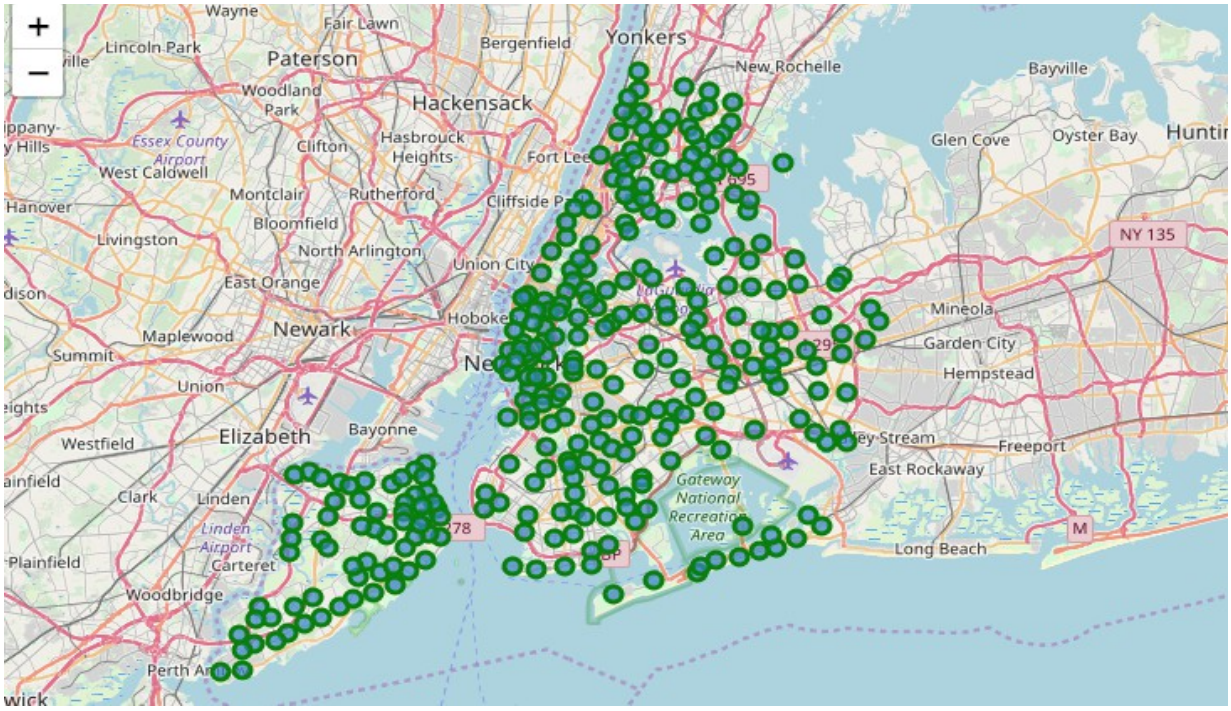
Analyzing spatial data to find the best neighborhoods for the establishment of a new Italian Restaurant in New York City.

Introduction:

The first Italian restaurants in New York, date back to the 1860s. New York has a history when it comes to Italian food and restaraunts, with a plethora of immigrants from Italy having been established there and having brought their culture as well as cousine.
So it is fair to say that anyone wishing to establish a new Italian restaurant, will have to deal with a lot of already existant competition on this specific field.
Using the Foursquare API, we will be able to eliminate many competitors, simply by finding the top rated restaurants. Finally we will be able to visually find neighbourhoods far away from these restaurants, that are potential places.

Project Description:

The aim of this project will be to decide which Neighborhood(s) of New York is more appropriate as a choose for a bussinessman to open a new Italian restaurant.
Using the Foursquare API we will find the Italian restaurants for each neighbourhood and choose the ones that have the best review score.
Finally we will use these restaurants as reference points, and find the neighbourhood(s) furthest away from them is the best to open the new restaurant.



New York Data:

The data we will be using, is derived from the NYU Spatial Data Repository.
It includes many features, but we will only use the Neighborhood names and their coordinates (Latitude, Longtitude).
The link to download the data is this : https://geo.nyu.edu/catalog/nyu_2451_34572

Foursquare API Data:

We will need to find the Italian restaurants in every neighborhood. In order to get this information we will utilize the Foursquare location information. For every neighborhood, a plethora of venues will be retrieved with information such as name, coordinated, type of venue, tips, reviews and photos. Of course because of the nature of our subscription to the Foursquare API, we will not be able to retrieve all of this information, as some require premioum calls.
Summarizing, the information for each venue we will get is the following:

1. Neighborhood
2. Neighborhood Coordinates (Lat,Lon)
3. Venue
4. Name of venue
5. Venue Coordinates (Lat,Lon)
6. Venue Category

Foursquare API:

Using the Foursquare API we will search each neighborhood for Italian restaurants, as well as their ratings. Due to the request limitations the limit parameter will be set to 100 and the radius parameter will be set up to 500.

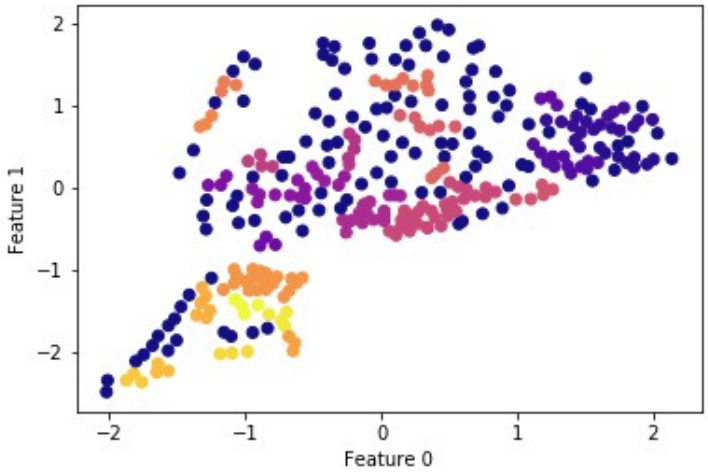
Methodology:

We first start on loading the already existent New York data, and processing this information. After that, we put the Foursquare API to use and request the venues for each neighborhood. After we have selected only Italian restaurants from the search above, we use the API once again to request for the rating of each restaurant.
Finally we process over the entirety of the data to visualize our information and get it ready to be analyzed.
The final step includes the DBSCAN analysis.
In summary the work can be described as follows:

1. Data acquisition and processing
2. API requests
3. Data preparation
4. Clustering

Clustering Analysis / Results:

In this case, since we need to find the neighborhoods that are the furthest away from the Italian restaurants, it is better to use the DBSCAN algorithm, since this algorithm will produce outliers, which in our case will be the neighborhoods we are looking for.
The result from the analysis reflects our goal, as any candidate neighborhood is shown as an outlier in our DBSCAN analysis.



Discussion and further Recommendations:

As a recommendation, one could go into further analysis, and run the same analysis for the top rated restaurants in a suggested neighborhood derived from our analysis. In this way you eliminate even more competitors, and have a clearer view of which parts of the neighborhoods are not close to Italian restaurants.

Conclusion:

In this project, through a DBSCAN analysis, we were able to identify possible locations for the establishment of a new Italian restaurant in the area of New York. Using the data derived from the Foursquare API and the analysis, we can clearly identify many neighborhoods that are far away from the top competitors in that business area.