**Neighborhood Comparison**

This is a ML model which finds similar neighborhoods based on venues available in the neighborhoods.

**Table of contents**

|  |
| --- |
| **Topic** |
| Introduction section |
| Data section |
| Methodology section |
| Result section |
| Discussion section |
| Conclusion section |

**Introduction section**

In this, we will convert addresses into their equivalent latitude and longitude values and will use the Foursquare API to explore neighborhoods in New York City. We will use the \*\*explore\*\* function to get the most common restaurant in each neighborhood, and then use this feature to group the neighborhoods into clusters. We will use the \*k\*-means clustering algorithm to complete this. Finally, We will use the Folium library to visualize the neighborhoods in New York City and their emerging clusters and figure out the clusters with similar restaurants.

**Data Section**

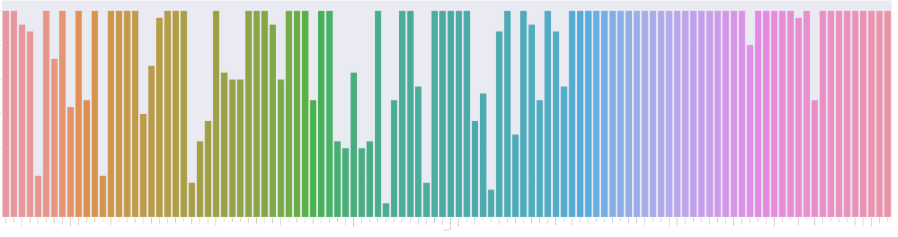
Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. Feel free to try to find this dataset on your own, but here is the link to the dataset: https://geo.nyu.edu/catalog/nyu\_2451\_34572

**Methodology :**

**Exploratory analysis:**

Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched.

Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data. 

Also while producing graph for number of cluster, I produced a graph to explore all the values for n\_clusters and then finding the best by exploring the elbow graph.

**Inferential analysis:**

Most important factors while building the recommender system were population and income. They are the most import factor because they have a nonlinear relationship according to our dataset.

It needed to make some inferential analysis to understand this nonlinear relationship. As the amount of population increases, it does not necessarily mean that average income of a neighborhood will also increase. It is true to most of the case but also many cases differ to follow this trend. Similarly, a neighborhood with less number of people may not necessarily have less average income. It is possible to have less number of people and more income and vice versa. This can be inferred from the following graph:

**Result :**

The result of the recommender system is that it produces a list of top restaurants and the most common venue item that the user can enjoy. During the runtime of the model, a simulation was done by taking ‘Washington Heights’ as the neighborhood and then processed through our model so that it could recommend neighborhoods with similar characters as that of ‘Washington Heights’.

The following image shows the result:

