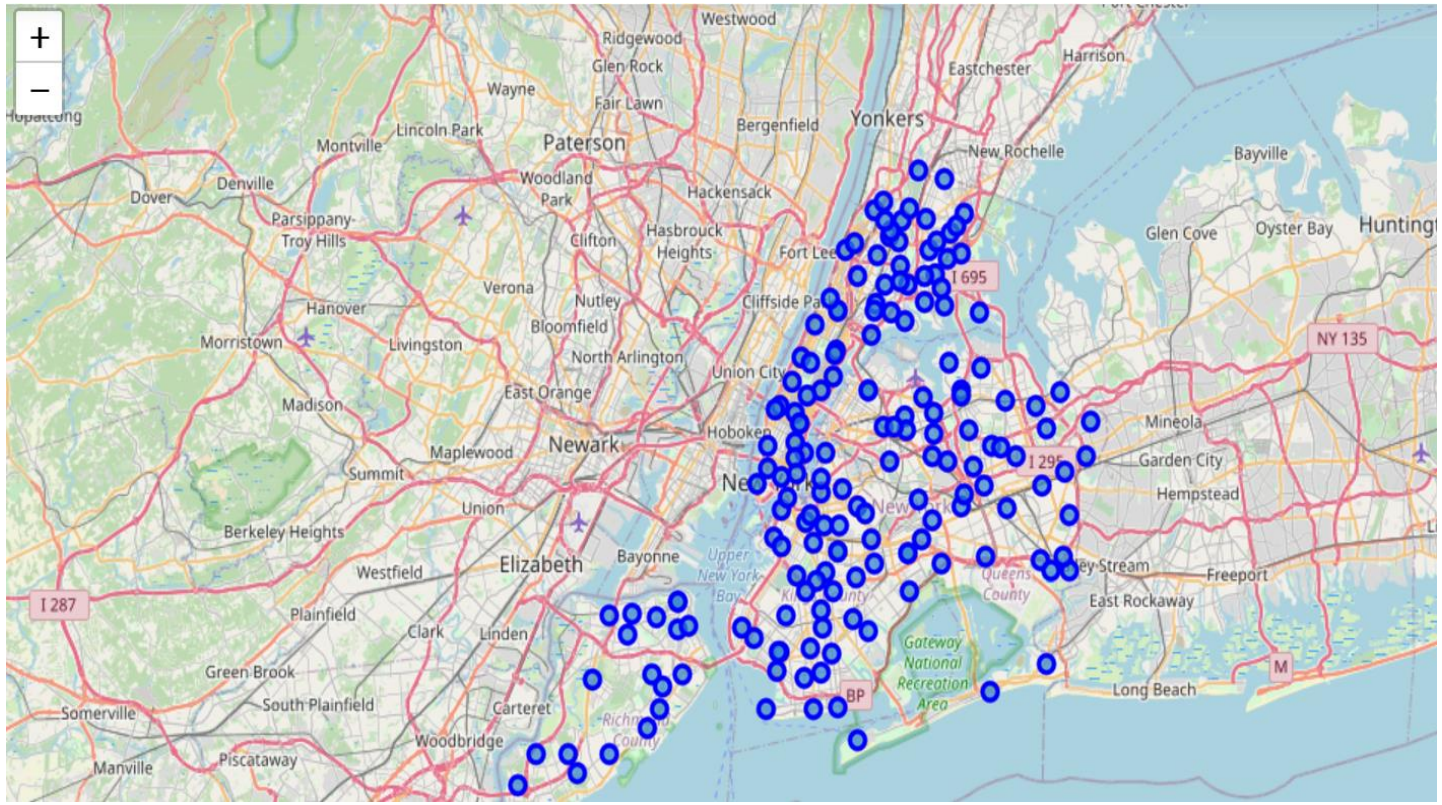
Predicting the best neighborhood for a new restaurant

Dharni Shah

April 30, 2020

Business problem



- To find the best location options to open a new restaurant in New York city
- Exploring the neighborhoods further to determine which type of cuisine would be favorable for the selected neighborhood.

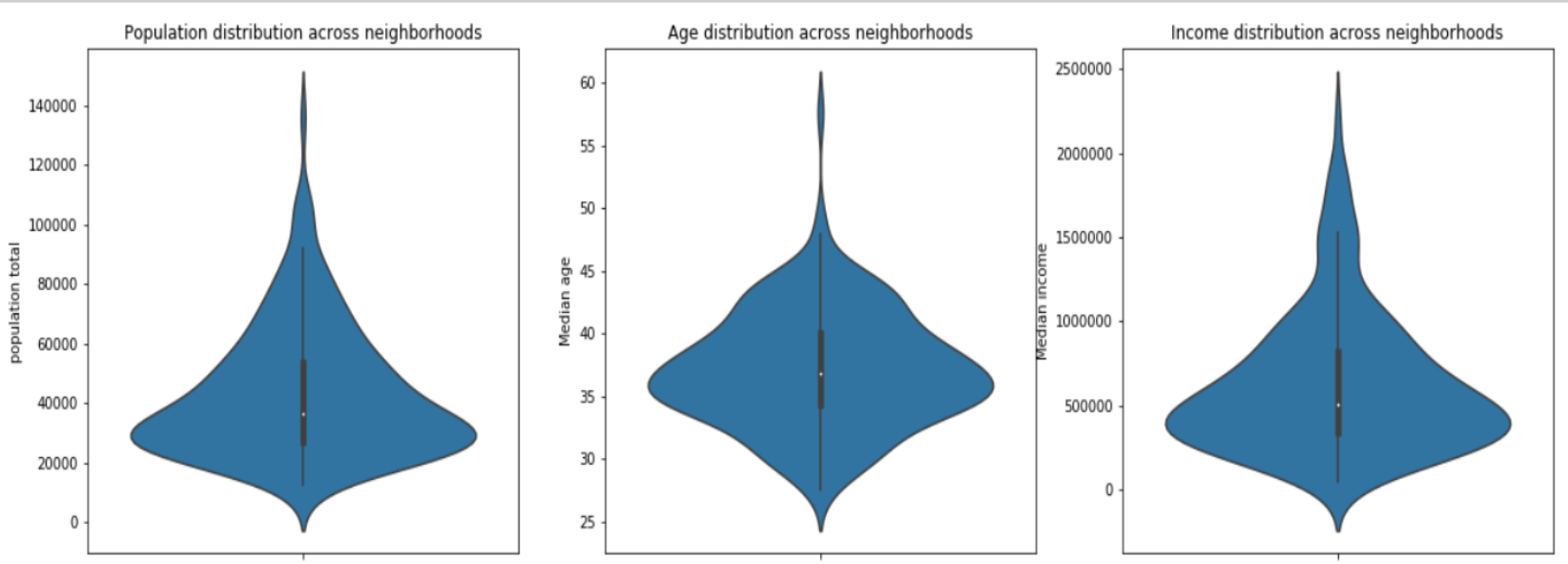
Data sources

Borough	Ncode	Neighborhood	Latitude	Longitude	population total	Median income	Median age	Venue
Brooklyn	BK45	Georgetown-Marine Park-Bergen Beach-Mill Basin	40.623845	-73.916075	48351	1520979	36.8	4.0
Brooklyn	BK61	Crown Heights North	40.670829	-73.943291	100130	980637	34.6	1.0
Brooklyn	BK90	East Williamsburg	40.708492	-73.938858	33155	519058	34.1	2.0
Queens	QN23	College Point	40.784903	-73.843045	24199	354073	38.7	8.0
Staten Island	SI11	Charleston-Richmond Valley-Tottenville	40.530531	-74.232158	24083	342708	39.5	2.0

The neighborhood demographics data was obtained from census data https://www.census.gov/ and https://popfactfinder.planning.nyc.gov/

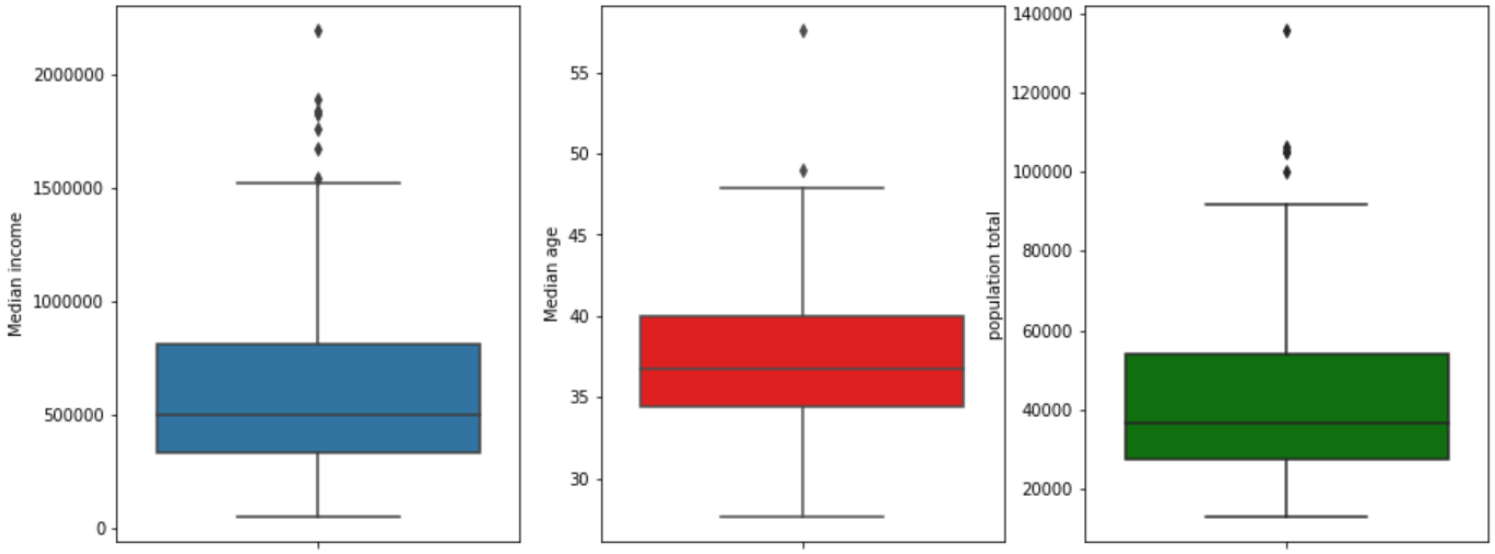
The data consisting of the venues for each neighborhood was obtained using the Foursquare API for the city of New York.

Exploratory Data Analysis



The frequency of demographics parameters is determined by visualizing the population size, median age and median income of all the neighborhoods.

Exploratory Data Analysis



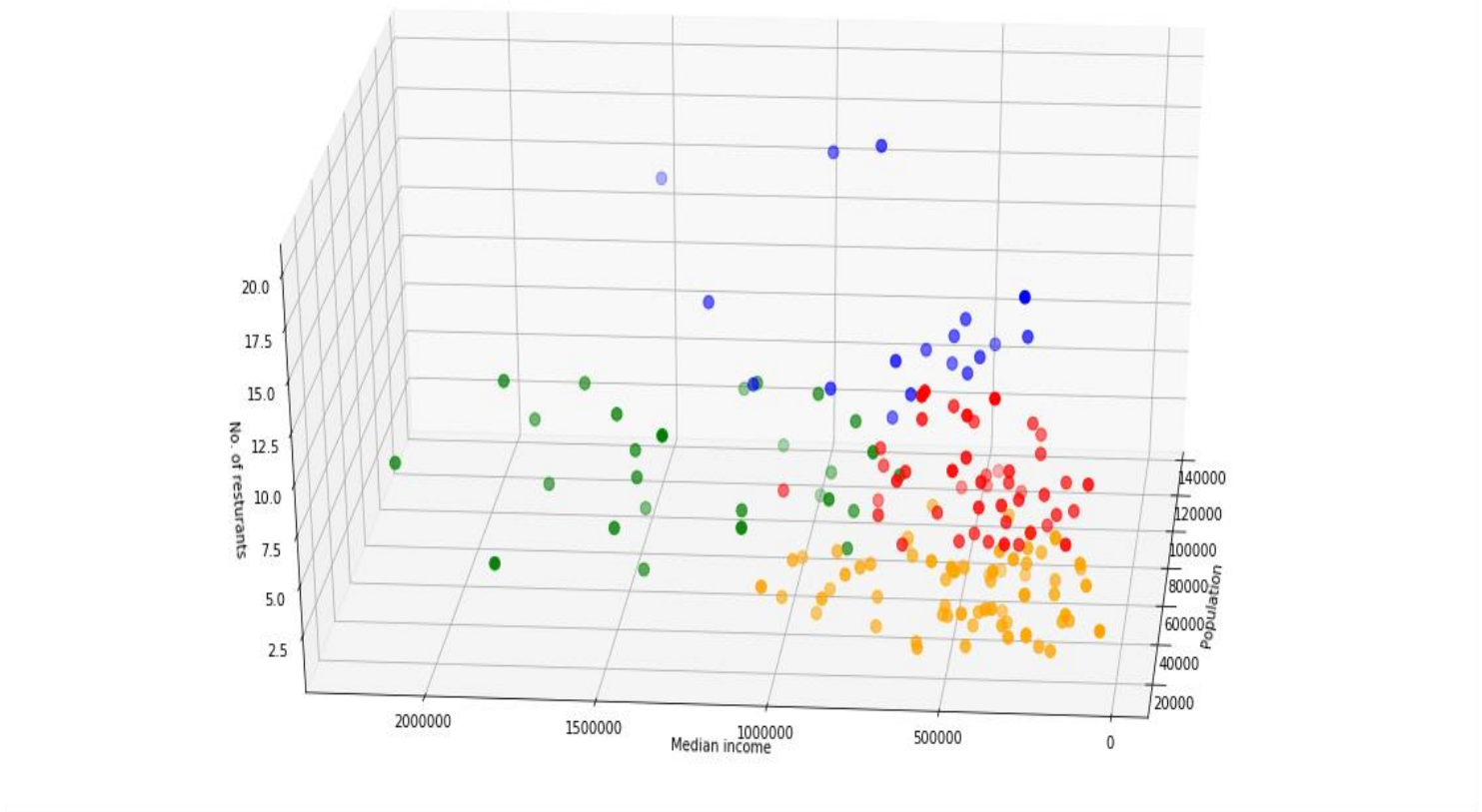
Box plots to better visualize the distribution range

Results

Cluster label	Cluster description
0	Very small population, low income, very less number of restaurants
1	Very large population, high income, very high number of restaurants
2	Small population, very low income, high number of restaurants
3	Large population, very high income, less number of restaurants

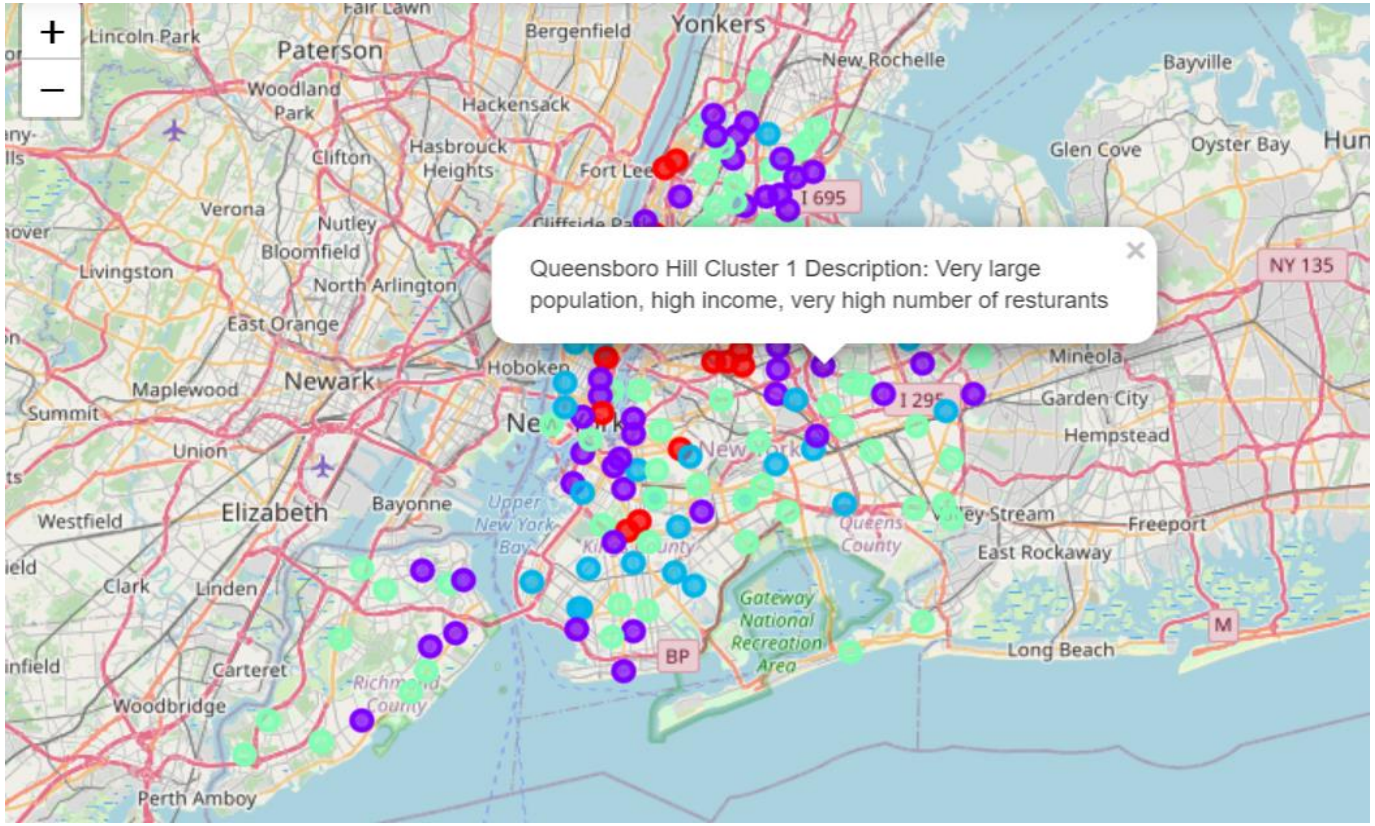
After examining the centers of 4 clusters. The following description is added across the corresponding labels.

Results



The clusters formed are visualized using 3D scatter plot. The parameters used for clustering are Median income, population and number of restaurants. There 4 clusters formed which separates the neighborhoods w.r.t these features.

Results



To visualize which neighborhoods belongs to which cluster, Folium map with neighborhoods superimposed on it was created. The label for each neighborhood shows the cluster number and its description. Further, on using K-means clustering on all the “Venue Category” to predict the best type of restaurant for the new outlet

Conclusion

In this study, I analyzed cluster 3 would be the best option from which the neighborhoods for a new restaurant should be selected. Depending on the budget and size of population the owner plans to serve, other clusters can be investigated. The clusters can be made non-overlapping with more demographics, geographically accessibility of the location and competitors' data.