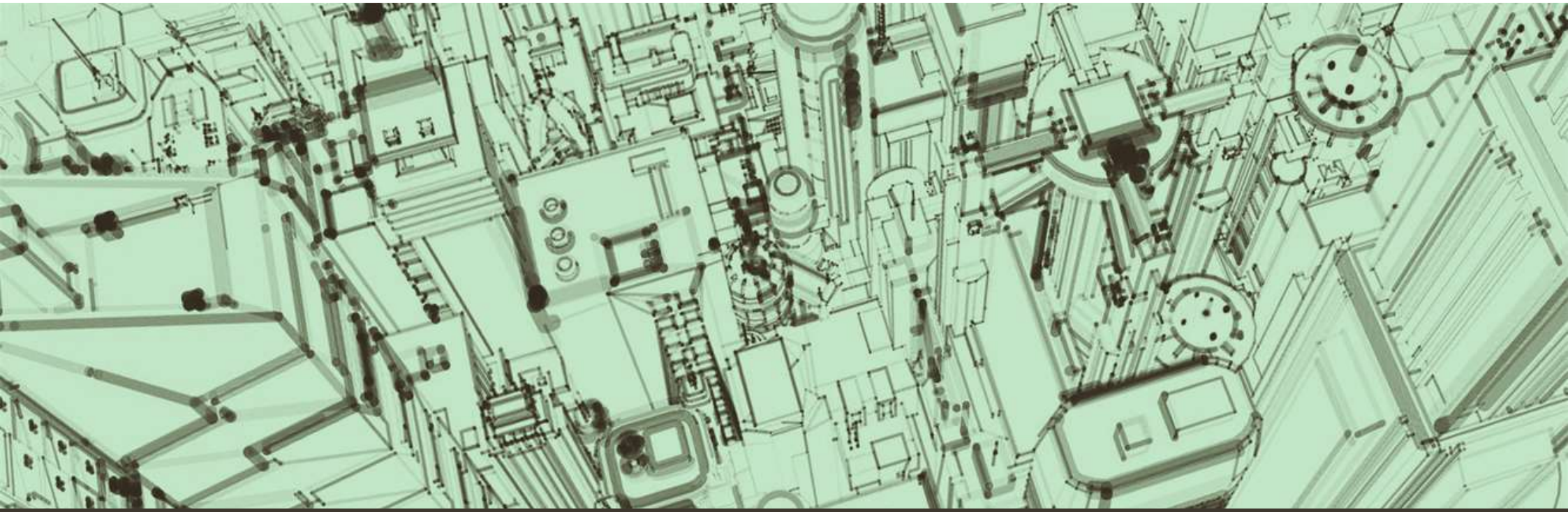


Geospatial Analysis of Wealth and Poverty Distributions in Manhattan Island

Coursera IBM Data Science Professional Certificate Capstone Project

Introduction

- New York city is often described as the financial capital of the world.
- Manhattan Island is synonymous with New York City and is described as the cultural, financial and entertainment capital of the world.
- As with all cities Manhattan is a place with high income inequality.
- Understanding the income and social inequalities and how they relate geographically would aid in better addressing the inequality concerns.
- The similarity of rich and poor neighbourhoods in terms of facilities and venues around each neighbourhood was also studied.

Data Acquisition

- The primary data sources used were as follows:
  - Geographic Data – comprising GeoJson and Shapefiles
  - Census Data – Income & Population Data
  - Foursquare API Venues Data

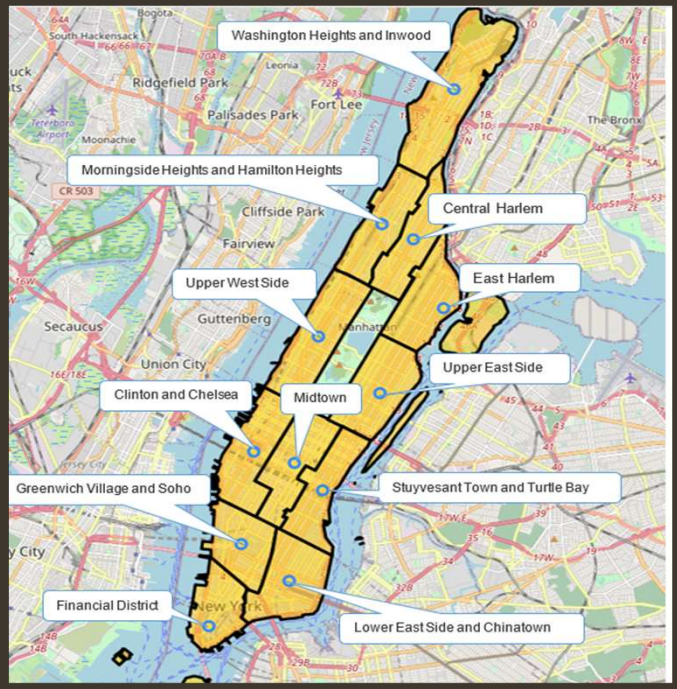
Data Acquisition - Geographic Data

- The geographic data used was obtained from the **New York University Spatial Data Repository (**https://geo.nyu.edu**)**.
- . Two sets of Geojson data was used from this repository:
  - **2010 New York City Community Districts** Geojson data (https://geo.nyu.edu/catalog/nyu-2451-34159)
  - **2010 New York City Census Tracts** Geojson data (https://geo.nyu.edu/catalog/nyu-2451-34505)
- The datasets were filtered to only contain the Borough of Manhattan.
- Census Tracts were used to map the census data directly onto an area in which the data was collected.
- Community Districts data was used in lieu of traditional neighbourhoods, since they are mandated by the city charter to review and monitor quality of life issues for New York City neighbourhoods (Insights obtained could be actioned by the Community Districts).

Data Acquisition - Geographic Data

Census Tracts of Manhattan

Community Districts of Manhattan



Data Acquisition - Census Data

- Socio-economic data was obtained from the **DATA2GO** (https://www.data2go.nyc)
- Datasets were developed by the **Measure of America (**http://measureofamerica.org/**)** initiative by the **Social Science Research Council (**https://www.ssrc.org/**)**.
- Data was downloaded in Excel format.
- Excel spreadsheets had to be cleaned prior to being imported into the Jupyter Notebook, due mainly to formatting issues.
- Datasets contained socio-economic data for Community Districts and Census Tracts.
- This dataset was filtered to limit the number of economic and social indicators used.

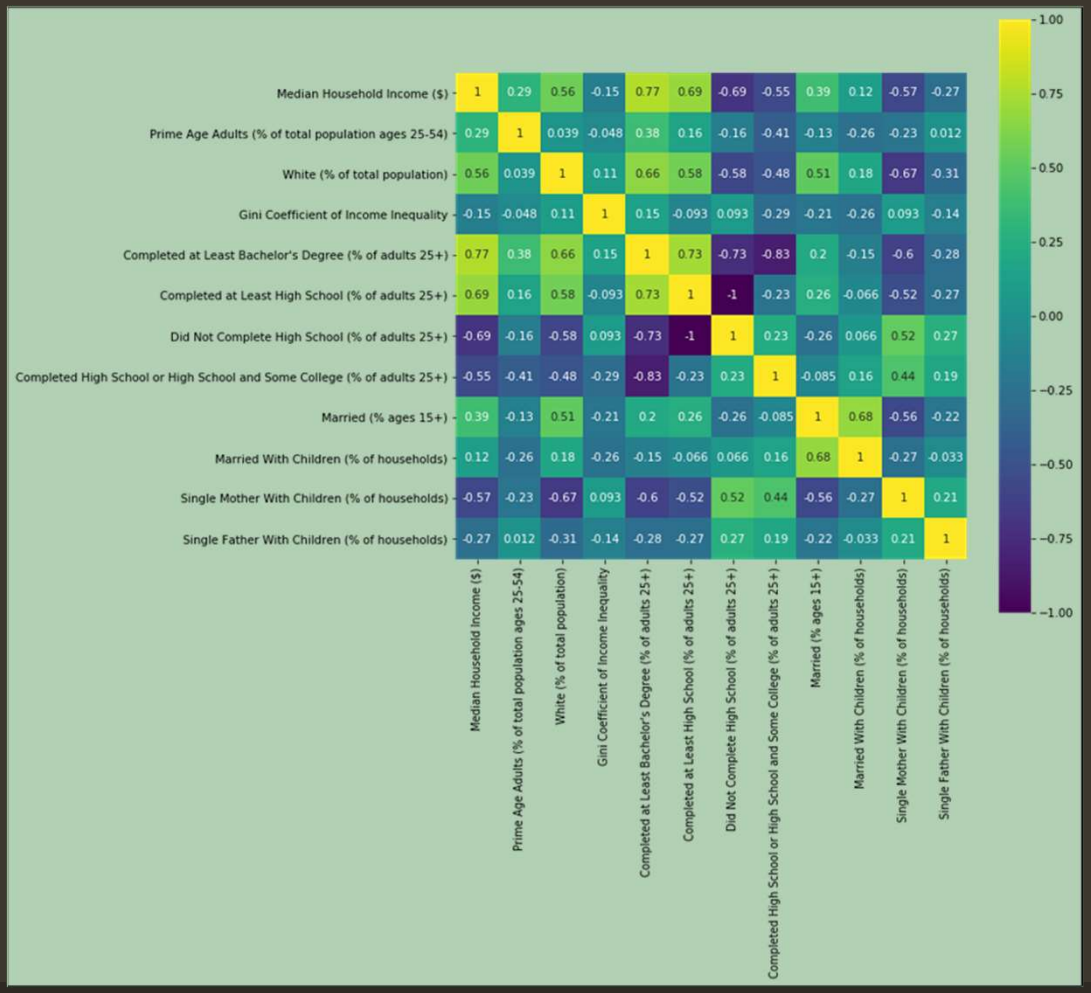
Data Acquisition – Foursquare API

- The Foursquare API was used to get the most common venues for each Community District
- This data was used in conjunction with the socio-economic data to determine the similarities between high income neighbourhoods and low-income neighbourhoods.

Exploratory Data Analysis

- Correlations between Median Household Income and various other features was calculated to determine which features correlate to Income.
- Choropleth maps were drawn to visualize the distributions of various socio-economic indicators.
- Choropleth maps also allowed visualize of the locations of high income and low income areas.
- Geopandas as well as Community Districts shapefiles were used to determine the midpoints of each Community District.
- Foursquare API data was used to determine the types of venues in each Community District based on the midpoint locations.
- Clustering of the Community Districts, based on the types of venues was done.

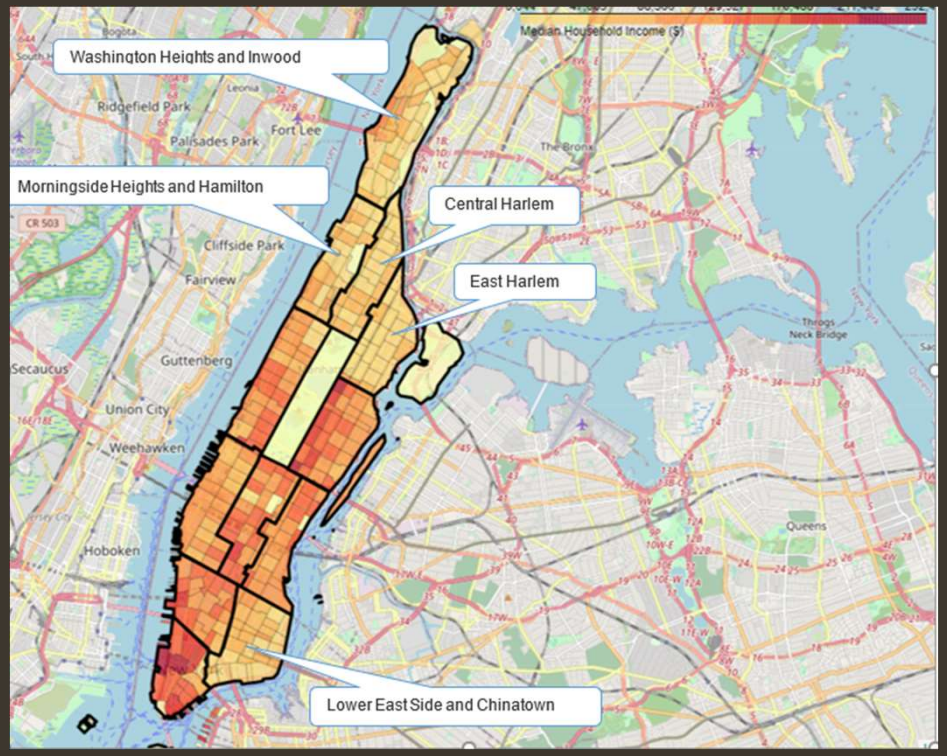
Results – Feature Correlations



- The various features were correlated against Median Household Income to produce a Correlation Matrix (data summarized in the table below)
- A high positive number (maximum 1) shows strong positive correlation, whilst a low negative number (minimum -1) shows strong negative correlation.
- From the table it is evident that one of the best predictors of success is education, with "Having completed at least a bachelors degree" and "Completed at Least High School" highly positively correlated to Median Household Income.
- "Did not complete High School" is highly negatively correlated to Median Household Income (greater number of people who did not complete high school in an area, the lower the median income earned in the area).
- The effect of strong family structures is also evident, with Single Mothers and Single Fathers with Children negatively correlated to income, whilst % married correlates positively to Income.

Features Correlated against Median Household Income	Correlation Coefficient
Did Not Complete High School (% of adults 25+)	-0.69
Single Mother With Children (% of households)	-0.57
Completed High School or High School and Some College (% of adults 25+)	-0.55
Single Father With Children (% of households)	-0.27
Gini Coefficient of Income Inequality	-0.15
Married With Children (% of households)	0.12
Prime Age Adults (% of total population ages 25-54)	0.29
Married (% ages 15+)	0.39
White (% of total population)	0.56
Completed at Least High School (% of adults 25+)	0.69
Completed at Least Bachelor's Degree (% of adults 25+)	0.77
Median Household Income ($)	1.00

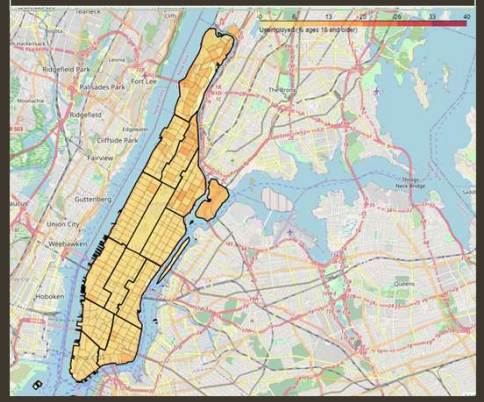
Geospatial Distributions Results

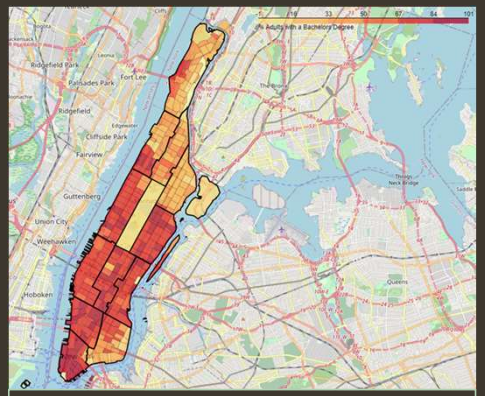


- Income distribution in Manhattan is skewed.
- Southern parts of Manhattan have a greater Median Household Income.
- Lower income areas clustered towards the North.
- The exception being “Lower East Side and Chinatown”.

Geospatial Distributions Results

- Unemployment Rates are shown in the map below.
- Distributions follow the distributions of Median Household Income.
- Areas with high unemployment are found to the north.





- The Percentage of Adults with a Bachelors Degree is shown above.
- The low income areas have a lower percentage of adults with a Bachelors Degree.
- Education is strongly correlated to success and should be a key focus area when looking at improving the socio-economic conditions within the low income areas.

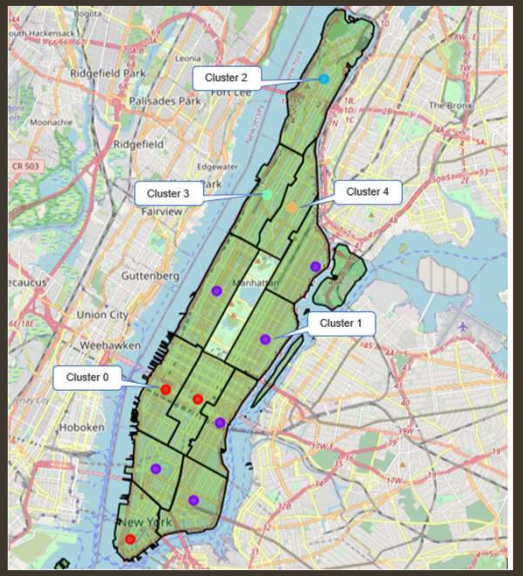
- Poverty Rates are shown below.
- Follows the same distribution as low income areas.
- It should be noted that even within the high income areas, there are blocks of poverty.

- One of the quality of life indicators is Commuting times.
- Above is shown the % of Workers with a greater than 60 minute one way commute.
- The low income areas are also the areas with the greater commute times.
- This adversely affects quality of life, since there is less free time for other pursuits (further studies, family time, etc.)

Cluster Analysis Results

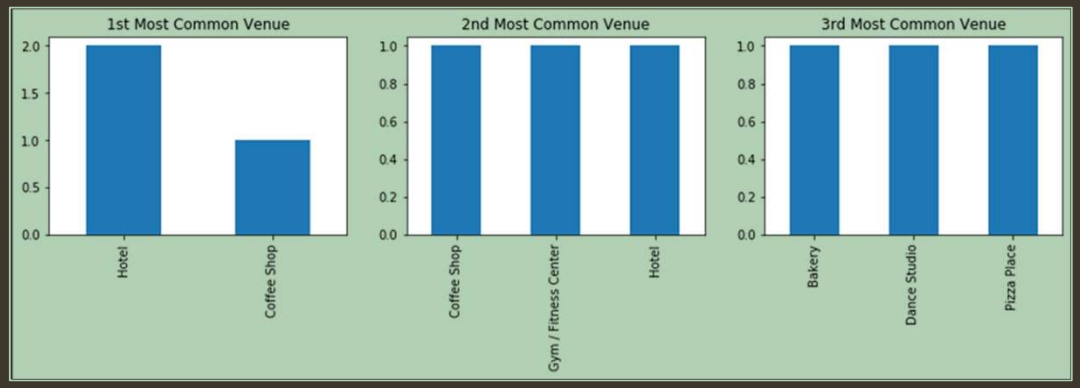
- The cluster analysis results are shown below.
- The number of venues returned by the Foursquare API calls, for the low income areas are significantly lower than the high income areas.
- This shows that there isn't sufficient economic activity in the low income areas to sustain a large number of businesses.
- Clustering indicates that there is a difference between the low income and high income areas, in terms of venues available in each area.

Community District	Cluster	Number of Venues
Clinton and Chelsea	0	84
Financial District	0	100
Midtown	0	100
East Harlem	1	65
Greenwich Village and Soho	1	100
Lower East Side and Chinatown	1	100
Stuyvesant Town and Turtle Bay	1	100
Upper East Side	1	100
Upper West Side	1	95
Washington Heights and Inwood	2	39
Morningside Heights and Hamilton Heights	3	39
Central Harlem	4	46

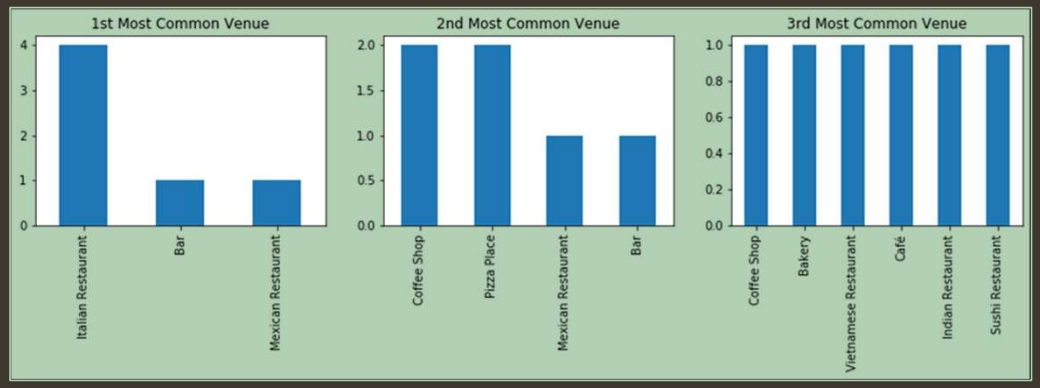


Most Common Venues in Each Cluster

Cluster 0

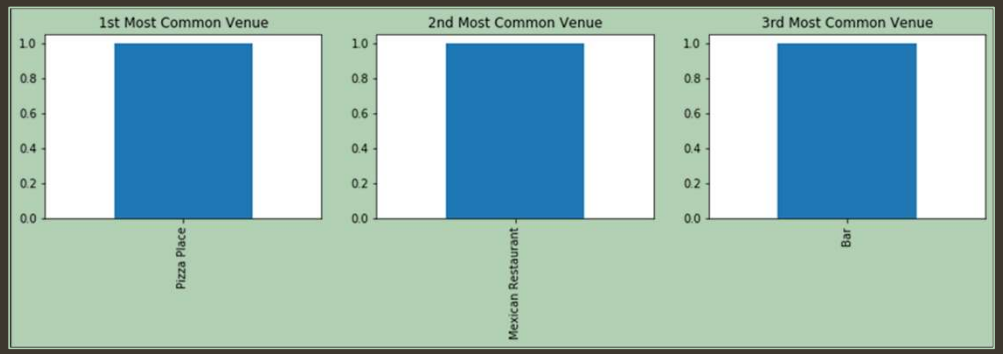


Cluster 1

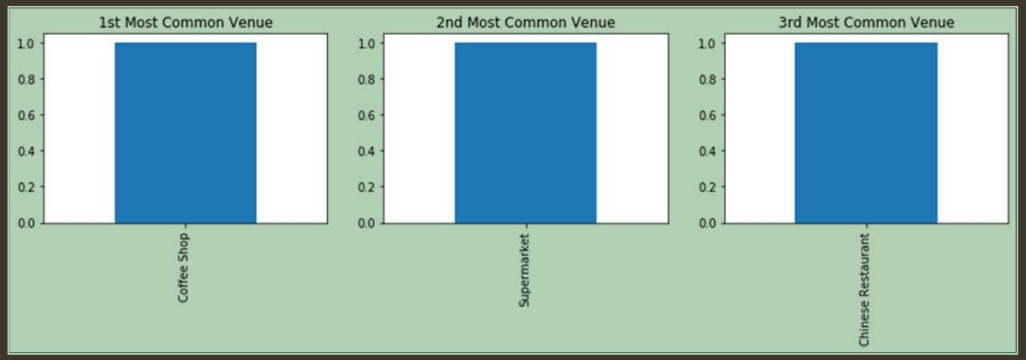


Most Common Venues in Each Cluster

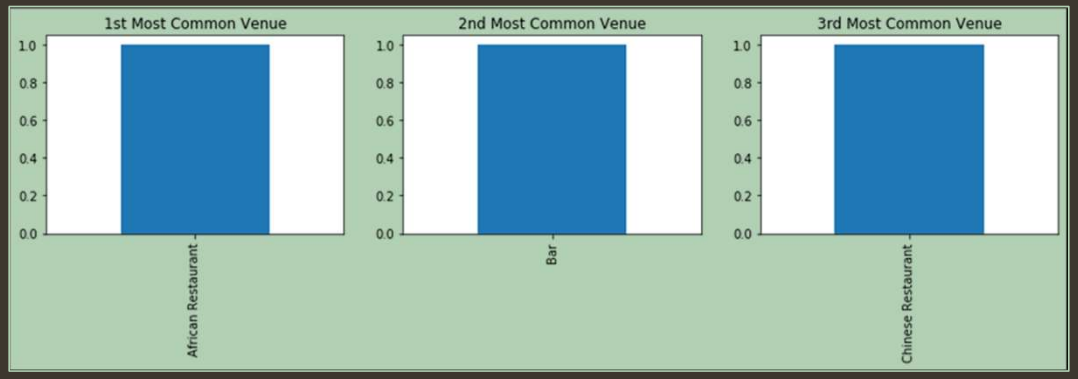
Cluster 2



Cluster 3



Cluster 4



Conclusions

- The geospatial analysis of the Community Districts in Manhattan yielded interesting insights into the wealth distribution within Manhattan. It was observed that the lower income neighbourhoods were situated in the northern parts of the island. However, even within the wealthy areas, there were Census Tracts experiencing poverty. This could be due to rent controlled apartments still being used or expensive buildings reserving a portion of the apartments for lower income groups.
- Correlation analysis showed that the best indicator of success is education.
- The commuting times were greater for people living in the low-income areas. This affects quality of life, since they have less free time for other pursuits (furthering their education, family time, etc.).
- The cluster analysis showed that the lower income areas were not similar to the higher income areas. This can be attributed to the low number of venues that are available in the lower income areas. The number of venues suggest that the lower income areas cannot sustain a large number of venues.

Recommendations

The geospatial analysis of poverty and wealth distributions in Manhattan Island showed that there are inequalities that need to be addressed. It is recommended to NGOs looking to improve the socio-economic conditions of the poorer areas, that they:

1. Focus on education as a primary method to eradicate poverty.
2. Incentivise businesspeople to establish businesses within the areas so that the unemployment rates and commute times can be reduced.
3. Look to foster entrepreneurship within those areas (thereby increasing the number of businesses).