**Coursera IBM Data Science Professional Certificate, Capstone Project:**

**Geospatial Analysis of Poverty and Wealth Distributions in Manhattan Island, New York City**

**Sunil Pakiri**

**21 June 2020**

# Introduction / Business Problem

New York city is often described as the financial capital of the world. It is also one of the cities in the United States with the highest median household income. Within New York city itself, Manhattan Island is not only the richest borough, also the most famous. Manhattan is synonymous with New York City and is described as the cultural, financial and entertainment capital of the world. Manhattan has been the setting for numerous movies, books and TV shows and hosts the world’s largest stock exchanges.

However, 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.

Non-Governmental Organisations (NGOs) are often on the frontlines of the battle for addressing social ills. However, the majority of the NGOs and aid organisations are not well funded. This means that the funding available must be put to the most effective use possible.

Income inequality and poverty are often fought on many fronts, such as minimum wage laws, better access to facilities, access to job opportunities, etc. It is also found that the poverty and income inequality is geographically segregated (“the wrong side of the tracks” problem). With this in mind, the neighbourhoods of Manhattan Island were studied to determine the geospatial distributions of wealth, quality of life and poverty.

The similarity of rich and poor neighbourhoods in terms of facilities and venues around each neighbourhood was also studied. The Foursquare API data was used to determine which venues/facilities are present in each neighbourhood and how they correlate to income levels.

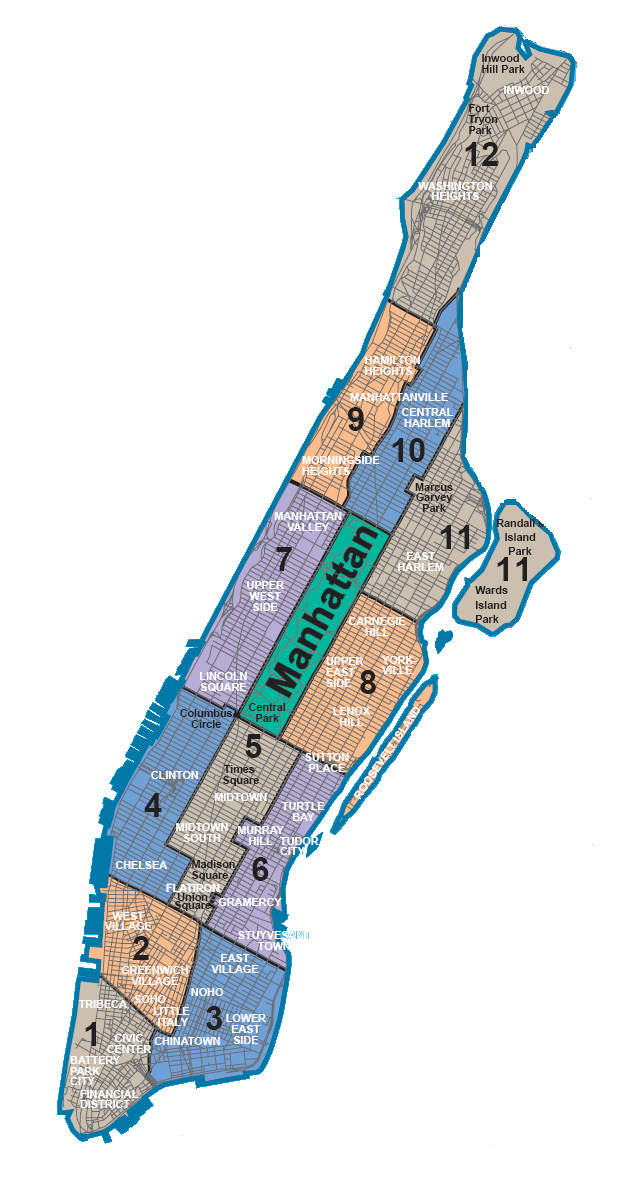
This study aims to understand the differentiating factors within these neighbourhoods and provide guidance to NGOs to more effectively address the ills of inequality and poverty, with the limited funds available.

# Data

Due to the nature of the study, various data sources were required. The primary data sources used were as follows:

* Geographic Data
* Census Data – Income & Population Data
* Foursquare Venues Data

## Geographic Data

The neighbourhood boundaries for Manhattan Island are not particularly well defined, especially with the small size of certain neighbourhoods (some consisting of only a few blocks). During the 2010 Census, the neighbourhoods were divided into Census Tracts. Census Tracts are statistical areas built from census block-groups that are designed to have an ideal size of 4,000 residents, with range of 1,200 to 8,000. Due to the small sample sizes of the Census Tracts and the large number of tracts overall (2168 for all of New York City), it was decided to use an aggregated dataset.

The aggregation chosen was the New York City Community Districts. Community Districts were mandated by the city charter to review and monitor quality of life issues for New York City neighbourhoods. Each Community District comprises of several neighbourhoods grouped together. Also, since they are directly mandated to review and monitor quality of life issues in New York City neighbourhoods, Community Districts are the ideal aggregation of the neighbourhoods for this study, since results obtained herein can be actioned by the Community Districts.

**Figure 1: Manhattan Community Districts**

The geographic data used was obtained from the New York University Spatial Data Repository. Two sets of Geojson data was used from this repository, viz. **2010 New York City Community Districts** Geojson data and the **2010 New York City Census Tracts** Geojson data.[[1]](#footnote-1)

## Census Data

Population, demographics and income levels data were obtained from the **DATA2GO[[2]](#footnote-2)** datasets developed by the **Measure of America[[3]](#footnote-3)** initiative by the **Social Science Research Council[[4]](#footnote-4)**. These were cleaned and filtered for the Manhattan data. Using these, I was able to determine the richest and poorest areas in Manhattan. I developed the choropleth maps showing income distribution as well as population distribution.

## Foursquare API Data

The Foursquare API was used to get the most common venues of each Community District in Manhattan.

# Methodology

# Results

# Discussion

# Conclusion

1. <https://geo.nyu.edu/> [↑](#footnote-ref-1)
2. [https://www.data2go.nyc](https://www.data2go.nyc/map/?id=107*36047015900*ahdi_puma!undefined!ns*!other_pop_cd_506~ahdi_puma_1~sch_enrol_cd_112~age_pyramid_male_85_plus_cd_20~median_household_income_puma_397~median_personal_earnings_puma_400~dis_y_perc_puma_102~poverty_ceo_cd_417~unemployment_cd_408~pre_k_cd_107!*air_qual_cd~ahdi_puma*family_homeless_cd_245#10/40.8278/-73.9586) [↑](#footnote-ref-2)
3. <http://measureofamerica.org/> [↑](#footnote-ref-3)
4. <https://www.ssrc.org/> [↑](#footnote-ref-4)