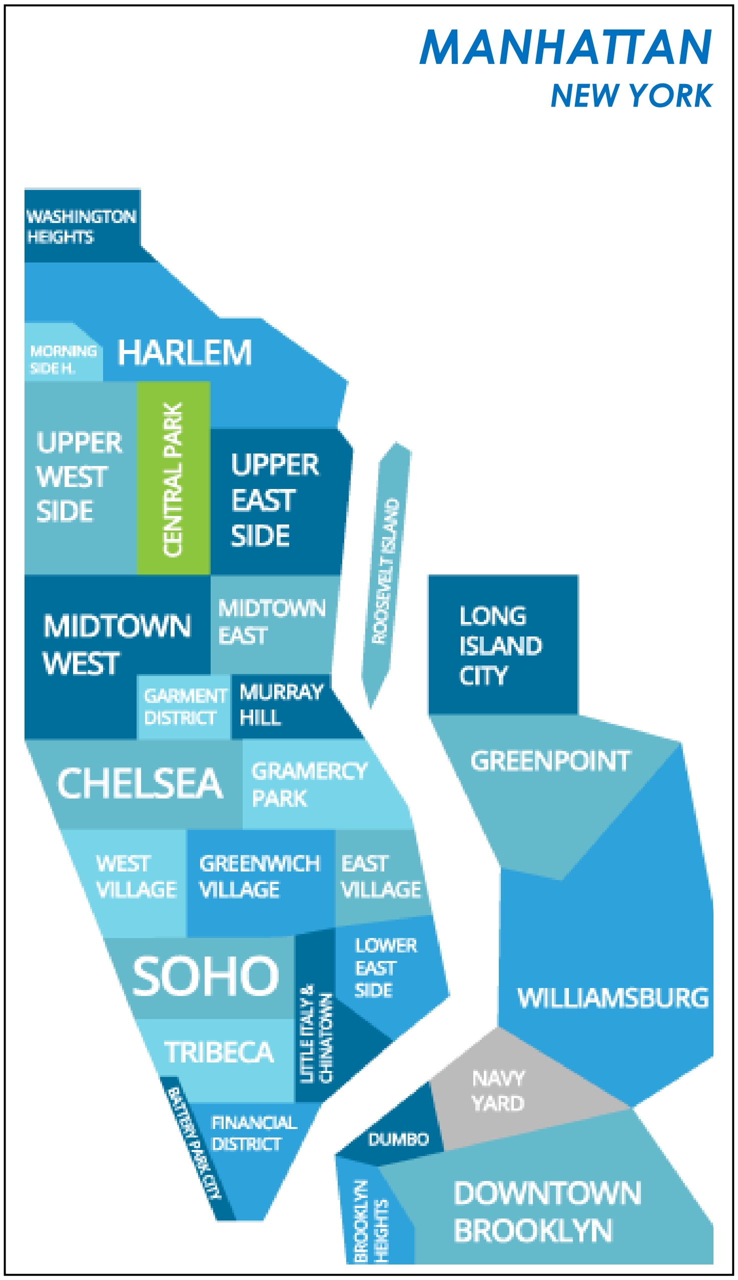
# The Battle Of The Neighborhoods



## REPORT 2020

*“Smart Cities used to be about technology and governance, but future cities pays greater attention on citizen. Thus cities need a new definition and metric to measure and improve”*

**Azman Ali**

**Tel** [Telephone]
**Fax**

Kuala Lumpur
Malaysia

https://github.com/Azman-Ali/CapstoneIBM/blob/master/BattleCode.ipynb

# Contents

*“Location and Spatial Analytic are keys in addressing many challenges in managing urban infrastructure and community”*

# Executive Summary

The notion of Smart City has evolved from being technology centric to more citizen centric. Over the years, our perception on the places has changed and evolved from the traditional notion of smart infrastructure to the like of ‘Livable City”,”Happy City” or “Green City” instead. While City infrastructure will always be one of the indicator for a healthy city posture, other parameters such as safety, economic or demographic can be useful indicator. This reports aims to provide a framework for measuring and profiling neighborhoods that can be used in decision making across many sectors such as tourism, retails or urban planning.

## The Importance of Measure-ability

One could not improve on what one could not measure. To date, city governors are adopting different metric for measuring city performance, so our hypothesis is that there should be a framework to guide how a city/neighborhood can be measured and where they stand compared others.

## Feature Highlights

This report will be based on our own assumption on 5 basic features that in our opinion representing some key parameters to measure score of city performance. These features are discussed in detail in the Data section.

## Who will benefit

City managers, governors or stakeholder may adopt the approach build a more sustainable city dashboard for their own users or publish public.

At citizen level, the public or tax payers are expecting more transparent government and information that involve public interest

Business Users can also use the framework to make assisted decision in either Business Planning or operation. Businesses can understand more about the neighborhood as the target market and optimize their resources and business offering.

***Azman Ali***

**Senior CONSULTANT**

# The Report Structure

The elements that will be covered in this report are organized as follows:

- **Introduction** where we discuss the problem statements, hypothetical solutions, objectives and who would be interested in this project.
- **Data** where we describe the data that will be used to solve the problem, their sources and particular parameters selected..
- **Methodology** section which represents the main component of the report where we discuss and describe any algorithm, flow, exploratory data analysis that we did, statistical testing performed, if any, and what machine learnings that were used and why.
- **Results** section where we discuss the results and draw any business values
- **Discussion** section where we discuss any observations made and any recommendations based on the results.
- **Conclusion** section where we conclude the report

# Introduction

## Background

Future governance of city infrastructure and it’s community/neighborhood demands for structured decision making approach. As such, the ability to measure how a city or community are progressing is crucial. Managing cities and their community are complex job and thus requires various integrated aspects to be measured. On the other hand, with the advent of digitalization, infrastructure operation and citizen activities can now generate more data in real time that can be useful, which can contribute towards new insights and knowledge. Today, Citizen’s awareness are increasing more than ever and city governance are expected to be more transparent especially in the matter involving public interest

This project aims to conduct a comparative study on neighborhoods within Manhattan, New York. Analytical framework developed for this study would be applicable to other cities or further developed into a proper data product.

## Statement of Problems

- Understanding analyzing location and or spatial data are key to urban planning and to address many different problem within community and neighborhood,
- However, the overall posture or profiling metric for city and its neighborhood is lacking and often measured in silo, prompting for the needs of structured analysis, tools or framework.
- A new parameters or Metric need to be explored to enable profiling of a city or neighborhood.

## Research Questions

- Can we profile a city or neighborhood based on several multi sector data? Such as safety, economic or infrastructure?
- Can we establish a metric or structured analytic framework to measure current state vs future state of a city (or against other neighborhood)?
- What are the candidate parameters for such measurement and how are they are influencing others?

## The Solution Hypothesis

- A metric or scheme to measure and profiling a neighborhood would be useful for stakeholders. The measurement will help understanding and establishing current state, identify gap and compare with other neighborhood.

- The metric is aimed to be comprehensive by looking into multiple integrated criteria. As a very basic requirement, it should measure readiness of relevant infrastructure that reflects the needs of citizen including economy, safety, social/recreation as well as other demographic criteria.

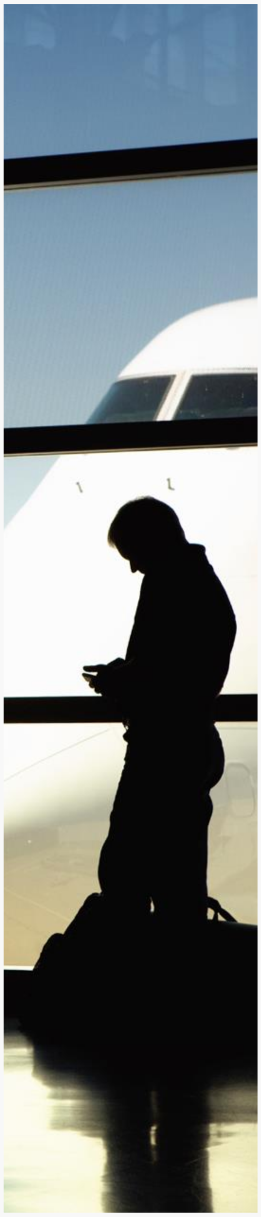
## Who Will Benefit

- City managers, governors or stakeholder may use the framework to start profiling their cities and neighborhood as part their planning exercise. They can explore to build a more sustainable city dashboard for their own use or published for public engagement (open data platform)
- At citizen level, the public or tax payers are expecting more transparent government and information that involve public interest. If used for public consumption, it can help city to increase citizen engagement and participation in city management which promote transparency in governance.
- Travelers are always curious about the places they are visiting and where to find the best location for their stay/visit
- Business Users can also use the framework to make assisted decision in either Business Planning or operation. Businesses can understand more about the neighborhood such as the target market and optimize their resources and business offering accordingly.

## Objective

- Comparative analysis between neighborhood within a Manhattan based integrated aspect of criteria including safety, economic, and social & demographic data.

*Note: The analysis and metric proposed are based on limited infrastructure and demographic data which might not be sufficient to address complexity of the community being measured. The main ideas of this project is to establish a prototype metric and framework for profiling based on multiple integrated aspect of city/neighborhood.*



Location Data and Demographic Data will be used in this project

# Data

Data is the most important aspect of this project. However, before any implementation or commencement of study, a good understanding of business case is crucial to ensure we will engage and acquire the right set of data and information. This is to eliminate 'Garbage in Garbage out' issue at the later stage. While there are many methods of measuring city performance and benchmarking existed, our approach aims to strike balance in the way we evaluate our city/neighborhood. As such we believed that the life elements such as health, social, economy and some demographic aspect would be worth experimenting or studied, especially when comparing places for living or work. Driven by these different needs we can then decide on the data that is needed, their sources as well as analytical and presentation approaches.

## Data Requirement & Hypothesis

Data Selection is made based on our Hypothesis of the following critical parameters to be measured. They are  grouped into 5 key criterias as follows:

| Criteria | Indicator | Data Required |
|---|---|---|
| **[S} Safety** | Crime/Incident report | Crime/Incidents report (NYPD open data) |
| **[E] conomy** | Household income can be a good indicator of lifestyle , property price/rent rate | Census data or any report on household income and neighborhood (open data) |
| **[P] opulation** | Population is  good indicator of popularity and affordability | Population in neighborhood data/report |
| **[S] port Infrastructure** | Our Hypothesis is that Sport infrastructure availability will shape certain behavior such as crime rate etc. | Foursquare Data on POI count/frequency within neighborhood. This include facilities such as gym, court, pool, |
| **[G] reen Infra/ Recreation** | Green infrastructure, Recreation POI.  Our Hypothesis is these infrastructure will shape certain behavior such as crime rate etc | Foursquare Data on POI count/frequency within neighborhood. This include all category of POI such as park, playground, marina, garden etc |

## Data Sources and Description

Based on the table above we have then identify the Data Sources as follows

**FOURSQUARE** is a location data providers that have served business and developers that require location information for their application**.** *Location Data that we needs and stated in the table above such as POI, Categories, Users (check in, tips) are data endpoints that are accessible as free and limited API calls. (Reference : www.foursquare.com)*

**NYPD OPEN DATA** Multiple open data on crime and incidents downloadable from the official website. We will use data based on 7 major felony happening in Manhattan in 2019

- *https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i*
- *https://www.researchgate.net/figure/An-Approximate-Mapping-of-NYPD-Precincts-to-NYC-Community-Districts_tbl1_247923137,*
- *NYPD Report Data 2019 by Precint : https://www1.nyc.gov/site/nypd/stats/crime-statistics/historical.page*

### Demographic Data

- Population by Neighborhood (google)
- Household Income Data by zip code (IRS open data) (google)

## Scope

For the data analysis, the scope of the study will cover Neighborhood within Manhattan Borough, New York.

# Methodology

As stated in our objectives, our deliverables will include comparative analysis between neighborhoods, employing various statistical and machine learning approaches which will be reported using relevant visualization approaches. Basic data preprocessing and analysis (such as neighborhood ranking) will be the dominant part of this report with some application of basic regression, prediction and recommender systems included.

In general the framework will be implemented based on the following flow:

1. Understanding Business Requirement, Problem Statements and Data Requirements/Specification (as highlighted in previous segments)
2. Identify 5 key features representing parameters to be assessed for City/Neighborhood profiling.
3. Identify relevant data end points and sources/providers. The data will be Extracted, Transferred and preprocessed accordingly.
4. Understanding data, formatting and  performing basic analysis on Data including using various statistical, machine learning and visualization tools
5. Perform comparative study and generate using appropriate techniques/approaches
6. Discuss any observation, findings , challenges and limitation of the studies.

## Data Preprocessing

*The following table outlines the indicator for respective pillar and respective data endpoint needed (for example particular point of interest data to be retrieved)*

| criteria | Indicator | Data Required |
|---|---|---|
| **[S} Safety** | Crime/Incident report | Crime/Incidents report (NYPD open data)<br>Format : Table/csv/excel |
| **[E] conomy** | Household income can be a good indicator of lifestyle , property price/rent rate | Census data or any report on household income and neighborhood (open data)<br>**Format** :Table/csv file/html |
| **[P] opulation** | Population is  good indicator of popularity and affordability | Population in neighborhood data/report<br>**Format :** Table/csv/html |

| | | |
|---|---|---|
| **[S]port Infrastructure** | Our Hypothesis is that Sport infrastructure availability will shape certain behavior such as crime rate etc. | Foursquare Data on POI count/frequency within neighborhood. This include facilities such as gym, court, pool,<br>**Format :** JSON data/geojson from Foursquare API services |
| **[G]reen Infra/ Recreation** | Green infrastructure, Recreation POI. Our Hypothesis is these infrastructure will shape certain behavior such as crime rate etc | Foursquare Data on POI count/frequency within neighborhood. This include all category of POI such as park, playground, marina, garden etc<br>**Format:** JSON data/geojson from Foursquare API services |

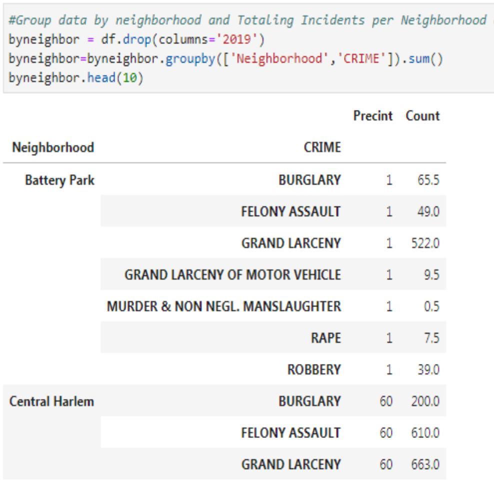
## Proposed Method For Comparative Analysis

*The following table maps the indicator and data to be analyzed, It specify the proposed data presentation or output for each pillar as well as for integrated metric (overall result)*

| criteria | Indicator | Data Presentation |
|---|---|---|
| **[S} Safety** | Crime/Incident report | • Ranking/Distribution Table<br>• Bar Chart, scatter<br>• Folium Map Crime Distribution (Marker) -Bar chart<br>• K-means Clustering marker maps<br>• Cluster Table |
| **[E] conomy** | Household income can be a good indicator of lifestyle , property price/rent rate | |
| **[P] opulation** | Population is good indicator of popularity and affordability | |
| **[S] port Infrastructure** | Our Hypothesis is that Sport infrastructure availability will shape certain behavior such as crime rate etc. | |
| **[G] reen Infra/ Recreation** | Green infrastructure, Recreation POI. Our Hypothesis is these infrastruture will shape certain behavior such as crime rate etc | |

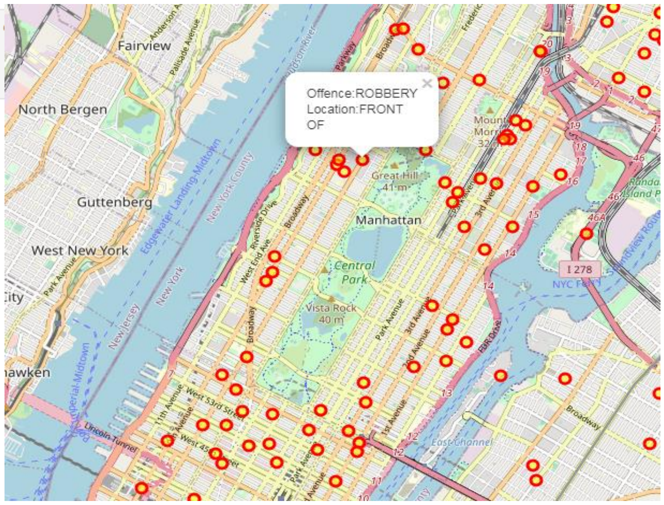
# Results

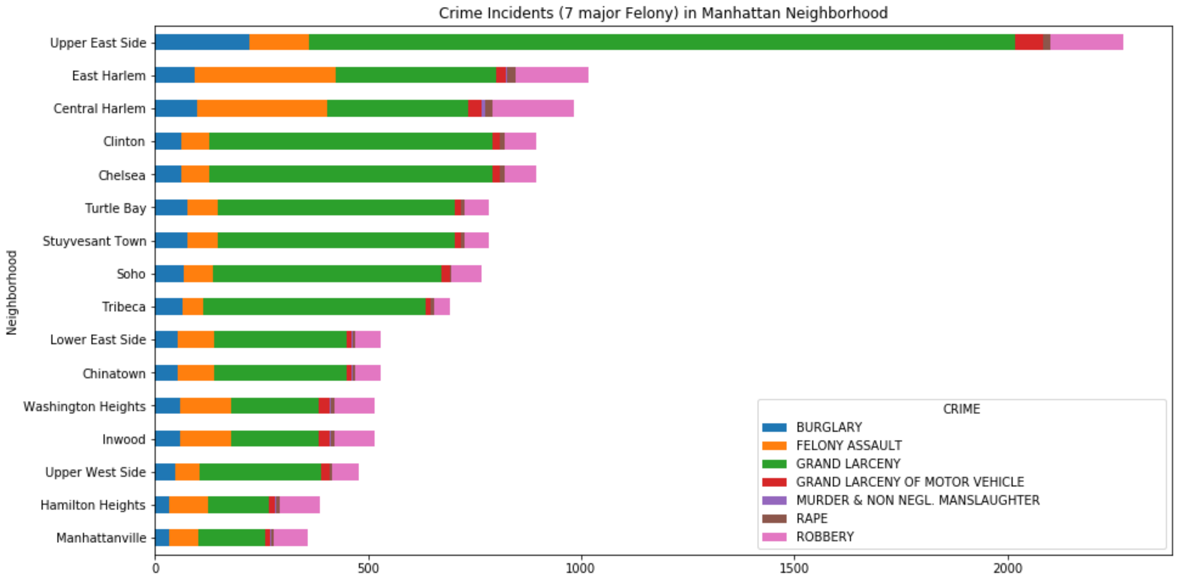
1. Historical Data from NYPD 2019 Incidents for 7 major crime are downloaded
2. Crime and incidents desciption populated on Manhattan maps
3. Interactive marker indicating type of offence and incidents location
4. Incidents are sparsely distributed across Manhattan island (not densely located)



```
#Group data by neighborhood and Totaling Incidents per Neighborhood
byneighbor = df.drop(columns='2019')
byneighbor=byneighbor.groupby(['Neighborhood','CRIME']).sum()
byneighbor.head(10)
```

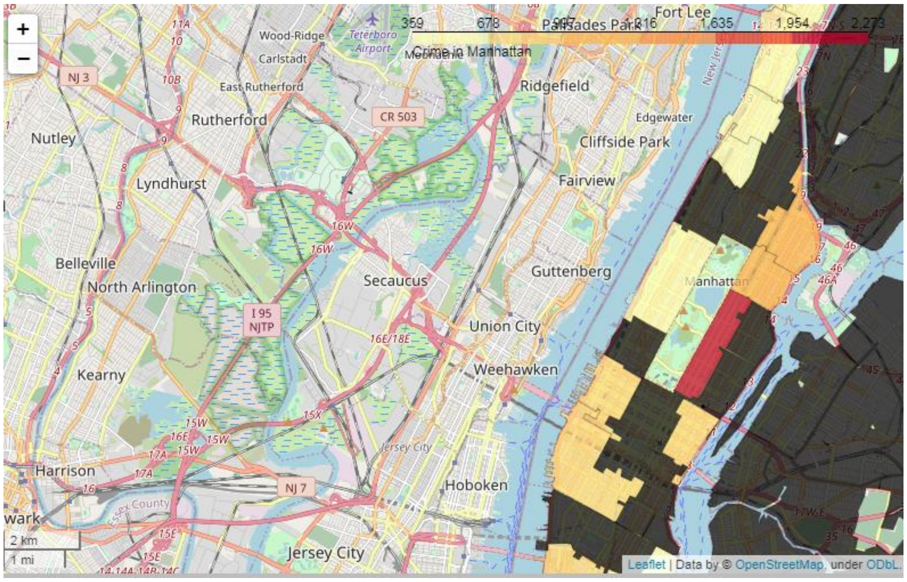
| | | Precint | Count |
|---|---|---|---|
| Neighborhood | CRIME | | |
| Battery Park | BURGLARY | 1 | 65.5 |
| | FELONY ASSAULT | 1 | 49.0 |
| | GRAND LARCENY | 1 | 522.0 |
| | GRAND LARCENY OF MOTOR VEHICLE | 1 | 9.5 |
| | MURDER & NON NEGL. MANSLAUGHTER | 1 | 0.5 |
| | RAPE | 1 | 7.5 |
| | ROBBERY | 1 | 39.0 |
| Central Harlem | BURGLARY | 60 | 200.0 |
| | FELONY ASSAULT | 60 | 610.0 |
| | GRAND LARCENY | 60 | 663.0 |



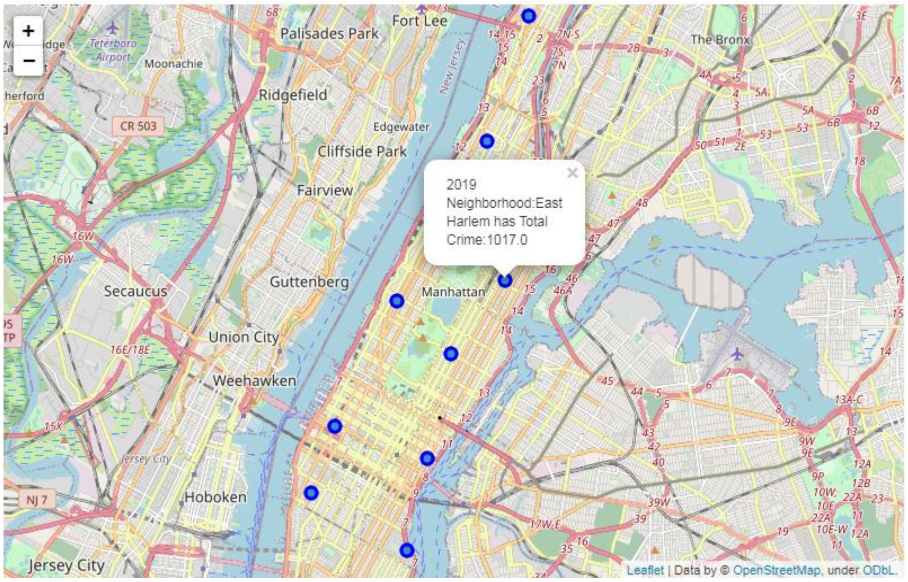


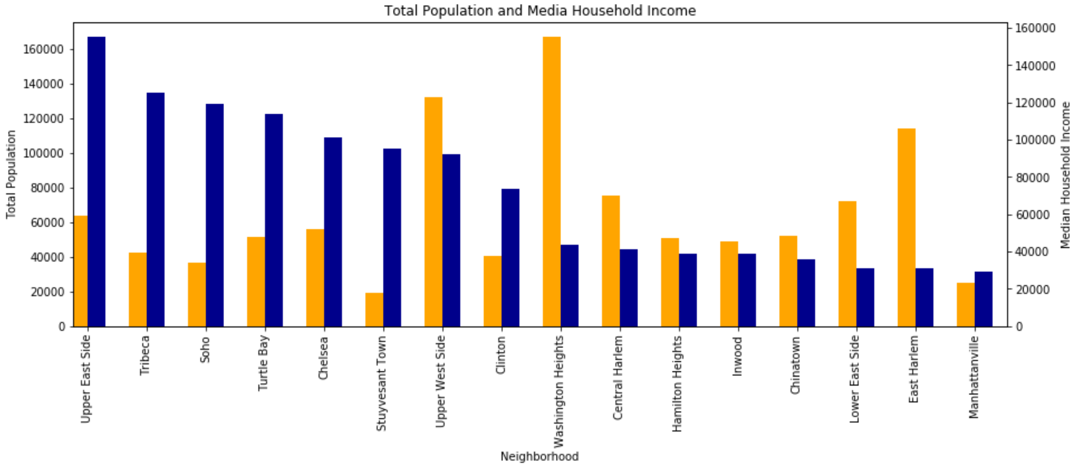
1. Highest TotalCrime in Upper East side
2. Grand Larceny is the most frequent Crime happened in Manhattan
3. Robbery , Felony Assult and Burglary are equally distributed across all neighborhood

Population density are seen is some area especially on upper east area (note that due to lack of naming consistency, only neighborhood  with standard naming are displayed.



Total Crime are aggregated and marked based on Neighborhood


2019 Neighborhood:East Harlem has Total Crime:1017.0

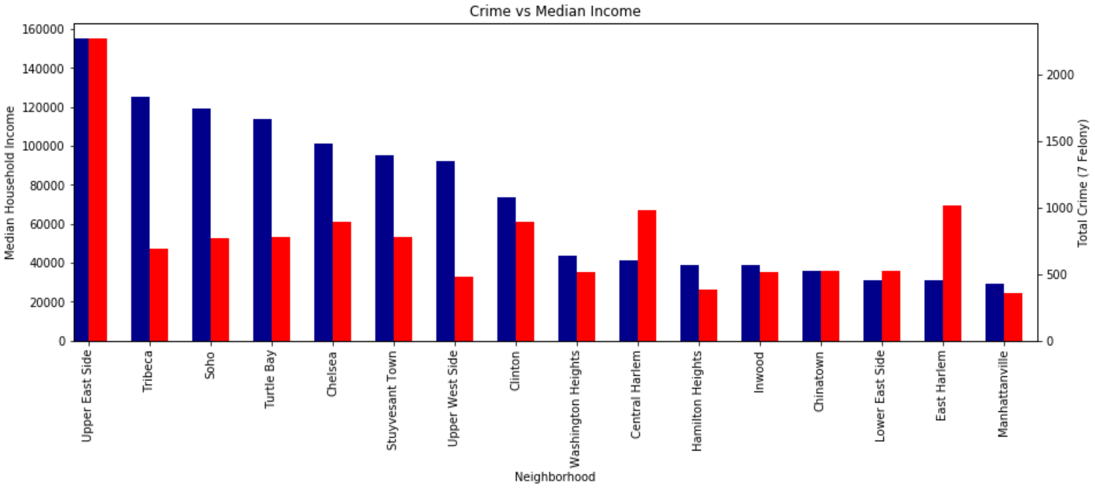


From above chart,

1 Top 6 highest median income are from lower than average neighborhood population

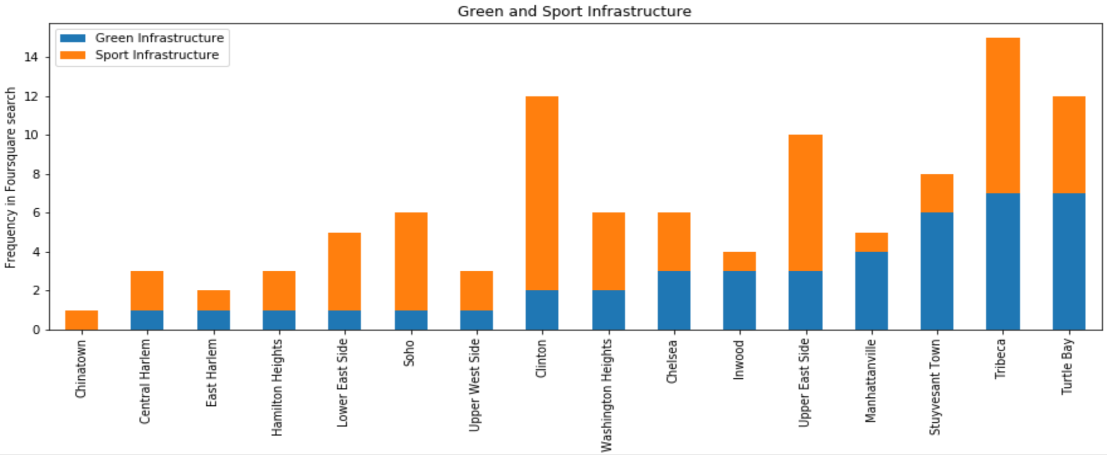
2. The lower 6 Populated neighborhood in contrary earned less than average income

3. Washington Height and East Harlem are 2 most populated area with below average household income (anomaly)

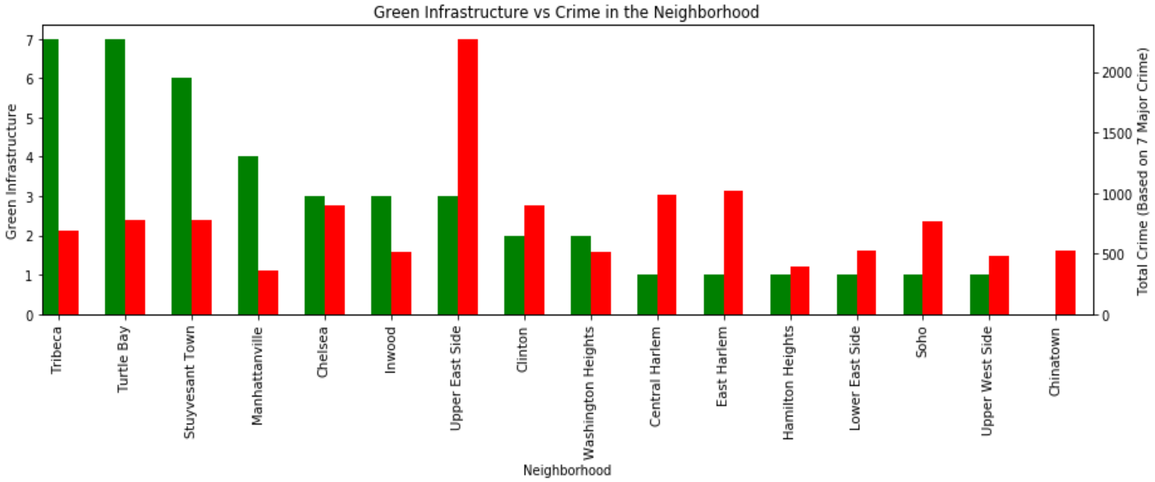


1. Crime distribution are  fairly distributed regardless of median household income
2. Upper East side ,East Harlem and Central Harlem are the most area with reported crime in 2019

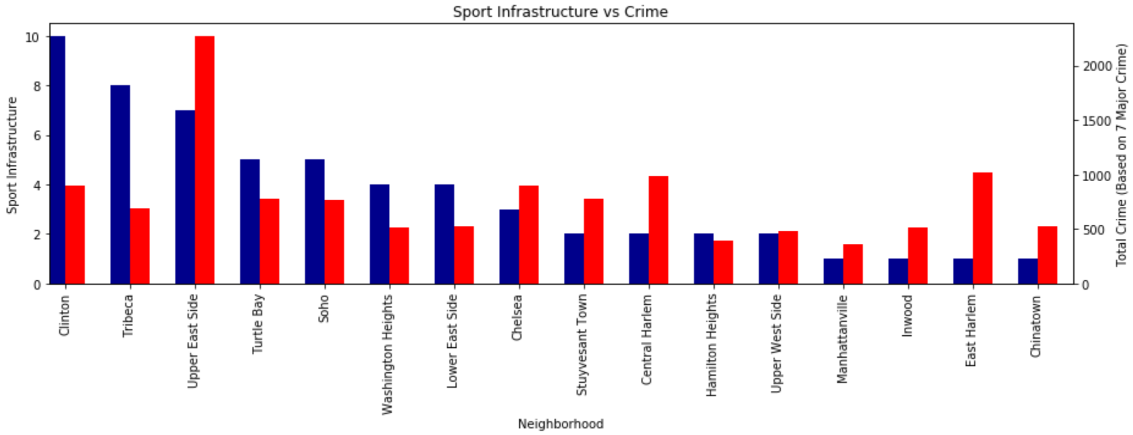
# Results



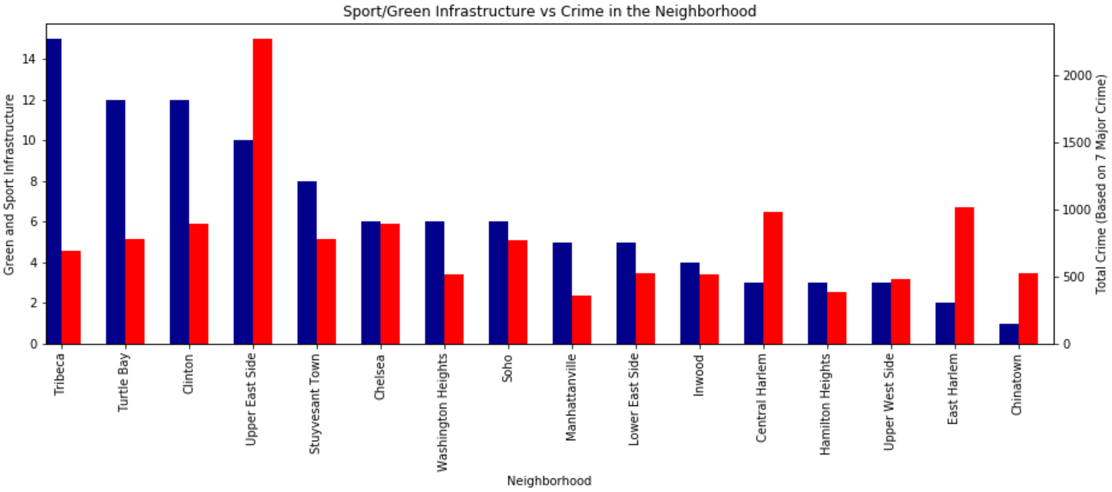
1. Green area are hardly visible in Chinatown
2. More than half of neighborhoods has below average of Green infrastructure availability
3. Clinton, Turtle Bay, Tribeca and Upper East Side has more sports and Green infrastructure (combined) than others



1. Generally we see no significant influence from Green infrastructure toward crime statistic
2. Upper East side has significant/highest crime reported in 2019

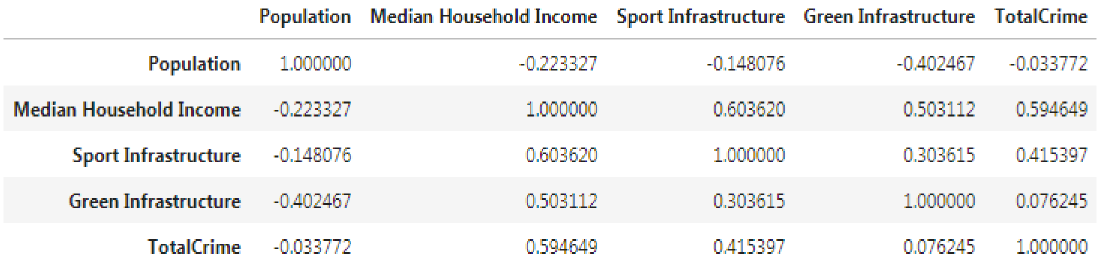


1. Crime/ incidents reports  are fairly distributed regardless of Sport Infrastructure availability



1. Combination of Sport and Green has no visible significance to crime distribution which are seen to be fairly distributed across Manhattan

**To further Validate our general assumption based on the diagram above, we can assess the data pair to get some measurement. The following correlation are applied  to our data (Based on Pearson Correlation)**



| | Population | Median Household Income | Sport Infrastructure | Green Infrastructure | TotalCrime |
|---|---|---|---|---|---|
| **Population** | 1.000000 | -0.223327 | -0.148076 | -0.402467 | -0.033772 |
| **Median Household Income** | -0.223327 | 1.000000 | 0.603620 | 0.503112 | 0.594649 |
| **Sport Infrastructure** | -0.148076 | 0.603620 | 1.000000 | 0.303615 | 0.415397 |
| **Green Infrastructure** | -0.402467 | 0.503112 | 0.303615 | 1.000000 | 0.076245 |
| **TotalCrime** | -0.033772 | 0.594649 | 0.415397 | 0.076245 | 1.000000 |

With the above Correlation assessment, we can highlight the following observation:

**Parameters with Significant Positive (+ve) Correlation are:**

1) Between Household **Income** and **Total Crime**, (coef: 0.5946)
   *(There are moderately high correlation between average income and Crime in the neighborhood. So crime are likely to increase in a higher income community*

2) **Sport** Infrastructure and **Crime** (coef: 0.415397)
   *Crime are likely happening in the area with sporting facilities*

3) **Sport** Infrastructure and **Income** (coef: 0.60320)
   *Higher income area have more sport facilities then the lower income area.*

4) **Green infrastructure** vs **Income** (0.503112)
   *Similarly Green infrastructure are more when people earns more*

**Parameters with Significant Negative (-ve) Correlation are** :

1) **Green** Infrastructure vs **Population** (Coef : -0.402467)
   *The higher population means lesser green spaces in the area*

## Based on Multiple Linear Regression (Target = Crime)

Not lets try to build a Multiple regression model that uses Crime as Target and the rest of our 5 parameters as Independent Variable. This is to show the parameters interdependency or strength of influence towads crime incidents

***Target (Crime Rate)=***

***Interception +V*Population + X*MedianIncome+ Y*GreenInfrastrcuture + Z*SportInfrastucture***

## Result of Running Multilinear Regression

**Independent Variables:**

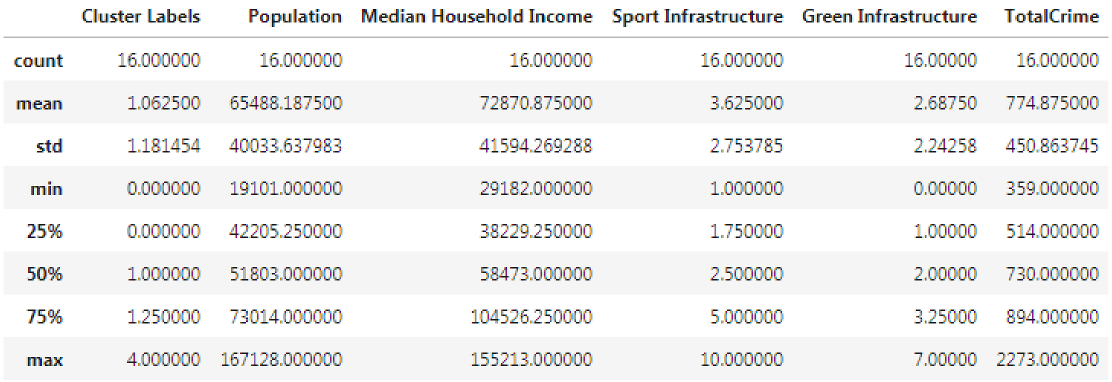
(0.014168250784884876) **(Population)** + 0.49353995203090123 **(Median Income)** + 0.06861183698536587 **(Green Infrastructure)** + -0.2148155862830695 **(Sport Infrastructure)**

**Intercept:** 0.10421619754242245

The Strongest influencer of Crime Rates is the "**Median Income**" parameters

## Basic Statistics of all parameters :

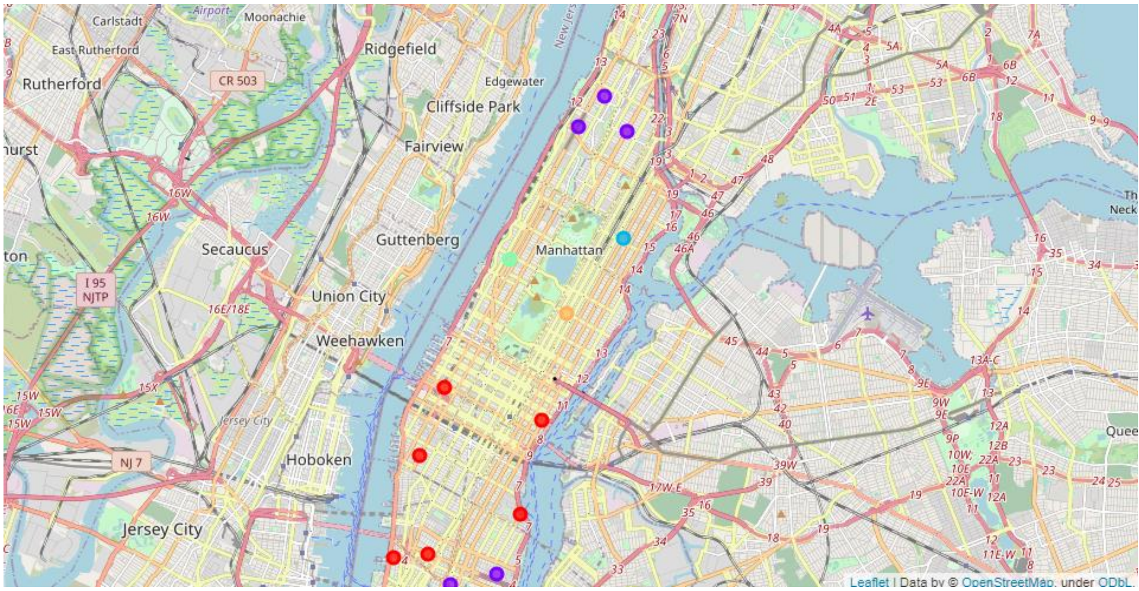
Overall Statistic are useful in analysis as we need to compare various findings with the overall statistic. For example the Mean Income for Manhattan city is around 73K per household. Average Population per neighborhood is 65K while 775 average major felony were reported in Manhattan for 2019. The Green and Sport infrastructure basically between 2-3 venues on average. Other statistic such as  min and max an percentile distribution are also handful for analysis



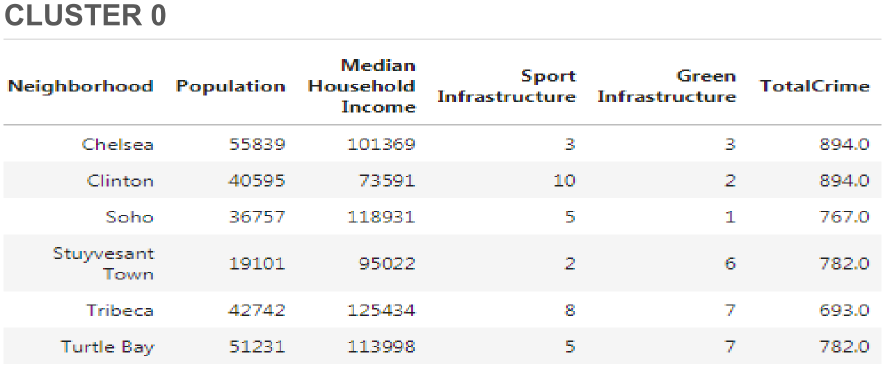
| | Cluster Labels | Population | Median Household Income | Sport Infrastructure | Green Infrastructure | TotalCrime |
|---|---|---|---|---|---|---|
| **count** | 16.000000 | 16.000000 | 16.000000 | 16.000000 | 16.00000 | 16.000000 |
| **mean** | 1.062500 | 65488.187500 | 72870.875000 | 3.625000 | 2.68750 | 774.875000 |
| **std** | 1.181454 | 40033.637983 | 41594.269288 | 2.753785 | 2.24258 | 450.863745 |
| **min** | 0.000000 | 19101.000000 | 29182.000000 | 1.000000 | 0.00000 | 359.000000 |
| **25%** | 0.000000 | 42205.250000 | 38229.250000 | 1.750000 | 1.00000 | 514.000000 |
| **50%** | 1.000000 | 51803.000000 | 58473.000000 | 2.500000 | 2.00000 | 730.000000 |
| **75%** | 1.250000 | 73014.000000 | 104526.250000 | 5.000000 | 3.25000 | 894.000000 |
| **max** | 4.000000 | 167128.000000 | 155213.000000 | 10.000000 | 7.00000 | 2273.000000 |

### Clusters of Neighborhood

Based on data we collected, clustering algorithm are used to segments the neighborhoods based on similarity. This will be useful for stakeholders wanting to profile their cities and make strategic action or plan to address gaps or improve livability of the area. For our analysis, K-mean methods with :5 clusters are applied and populated based on colors as in this map.

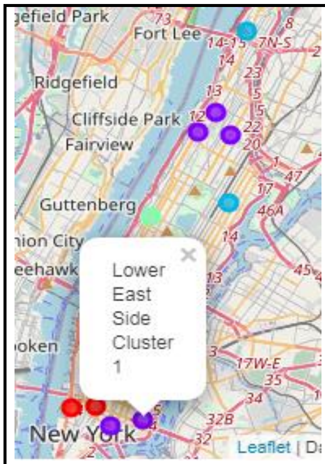


S

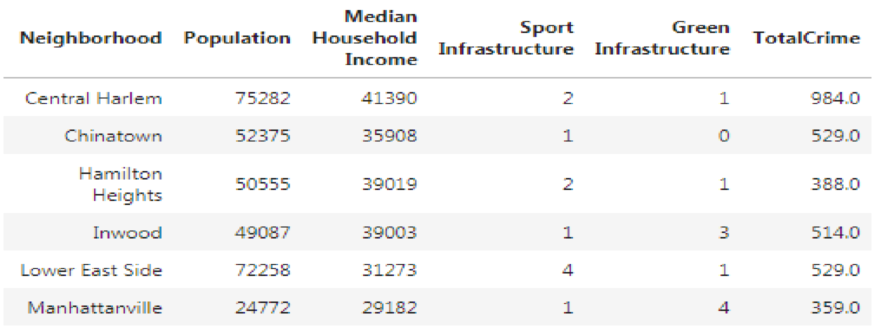

CLUSTER 0

| Neighborhood | Population | Median Household Income | Sport Infrastructure | Green Infrastructure | TotalCrime |
|---|---|---|---|---|---|
| Chelsea | 55839 | 101369 | 3 | 3 | 894.0 |
| Clinton | 40595 | 73591 | 10 | 2 | 894.0 |
| Soho | 36757 | 118931 | 5 | 1 | 767.0 |
| Stuyvesant Town | 19101 | 95022 | 2 | 6 | 782.0 |
| Tribeca | 42742 | 125434 | 8 | 7 | 693.0 |
| Turtle Bay | 51231 | 113998 | 5 | 7 | 782.0 |

**Low to moderate (Below Average) Income and Population, lot of Green and Sport Facilities and average Crime Rate**

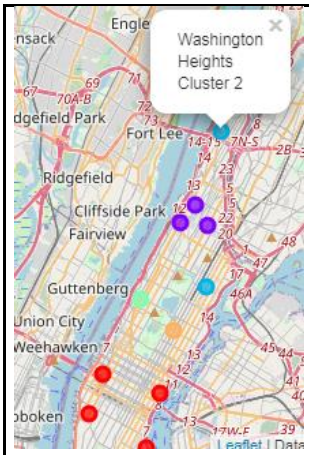


## CLUSTER 1

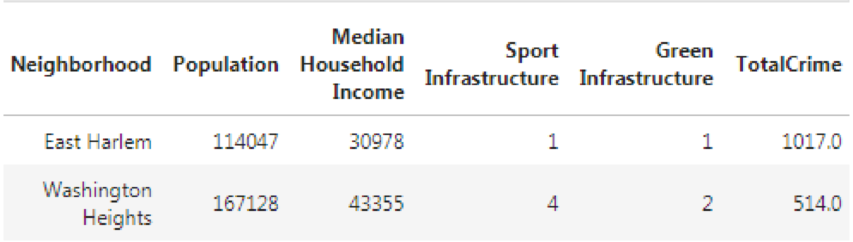


| Neighborhood | Population | Median Household Income | Sport Infrastructure | Green Infrastructure | TotalCrime |
|---|---|---|---|---|---|
| Central Harlem | 75282 | 41390 | 2 | 1 | 984.0 |
| Chinatown | 52375 | 35908 | 1 | 0 | 529.0 |
| Hamilton Heights | 50555 | 39019 | 2 | 1 | 388.0 |
| Inwood | 49087 | 39003 | 1 | 3 | 514.0 |
| Lower East Side | 72258 | 31273 | 4 | 1 | 529.0 |
| Manhattanville | 24772 | 29182 | 1 | 4 | 359.0 |

**Below Average Crime, Low sport and Green Infrastructure Low in Income**

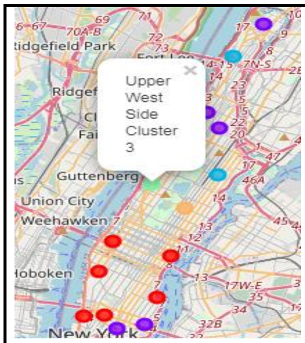


## CLUSTER 2

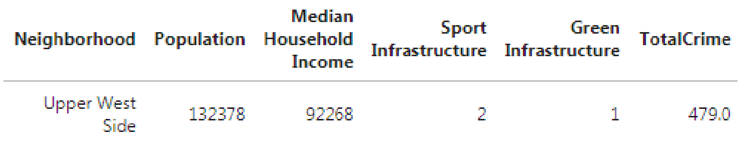


| Neighborhood | Population | Median Household Income | Sport Infrastructure | Green Infrastructure | TotalCrime |
|---|---|---|---|---|---|
| East Harlem | 114047 | 30978 | 1 | 1 | 1017.0 |
| Washington Heights | 167128 | 43355 | 4 | 2 | 514.0 |

**Highly populated by low median income as key features, with average Crime and Low Green/Sport Infrastructure**

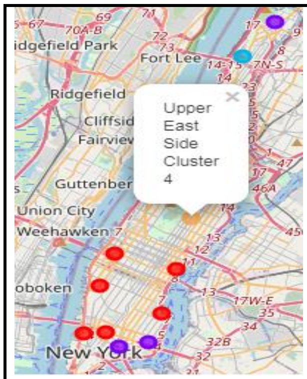


## CLUSTER 3

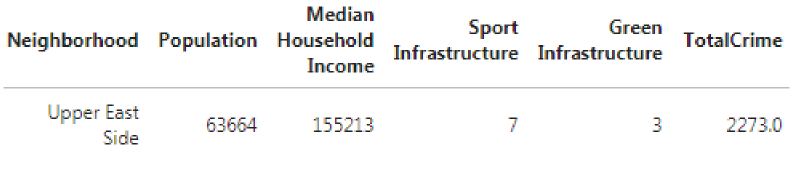


| Neighborhood | Population | Median Household Income | Sport Infrastructure | Green Infrastructure | TotalCrime |
|---|---|---|---|---|---|
| Upper West Side | 132378 | 92268 | 2 | 1 | 479.0 |

**Highly populated and high income/household, low in Crime but less Green and Sport infrastructure availability**



## CLUSTER 4



| Neighborhood | Population | Median Household Income | Sport Infrastructure | Green Infrastructure | TotalCrime |
|---|---|---|---|---|---|
| Upper East Side | 63664 | 155213 | 7 | 3 | 2273.0 |

Highest income but highest in total crime with above average Sport and Green infrastructure availability

# Discussion

**Overall Cities or Neighborhood  Measurement Metric**

- City Metric can indeed be developed but not limited to the 5 Parameters used in this case studies. In general no single parameters are directly contributed to lower crime rate, mostly because the crime or incidents are fairly distributed across all neighborhood regardless of income & population or whether they have high availability of sport and green infrastructures.

- The complexity of behavior we are trying to measure required more than what we are measuring so far. This can be other life aspect such as education, businesses, healthcare, transportation/mobility etc.

- Having said that, the Strongest correlation are found in Median Household  income (with moderately high correlation). Sport infrastructure also contributing probability of crime happening in the area.

**Profiling by clustering Manhattan's Neighborhood**

With the help of clustering algorithm we can work segment the neighborhoods and mapped them to 5 different profiles. 5 Clusters are chosen  to mimic a distribution curve where we would like to see the dominant criteria for majority of neighborhood but at the same time would like to see if there are unique criteria which applies to the lowest or highest percentile of the neighborhood.  As the result we can conclude the following :

Most of the Neighborhoods falls into either of these clusters/criteria:

- (Cluster 0) : Low to moderate (Below Average) Income and Population, lot of Green and Sport Facilities and average Crime Rate.  This can be a decent place for middle income group who wants to be around less crowded a balance lifestyle
- (Cluster 1) Below Average Crime, Low sport & Green Infrastructure and Low in Income. This is probably less crowded area for lower income group who are not into sports or recreational.

These are the clusters that can be anomaly or unique is a sense :

- (Cluster 2) Highly populated but low income as key features, with average Crime Incidents and Low Green/Sport Infrastructure. This could be the busiest area in the town preferred by lower income segment.
- (Cluster 3)Highly populated and higher income/household, low in Crime and less Green and Sport infrastructure. Could be preferred by Moderate to High income people.
- (Cluster 4)Highest income but highest in total crime with above average Sport and Green infrastructure. This could represent an Elite clusters within Manhattan.

## Discussion

**What can be improved?**

- This study can be extended to examine other potential parameters (ie health, business, education) that could be contributor to a more accurate predictor to the target variable (ie crime). Similarly other target variables can be explored in the future.

- Similarly the size and scope of data set can be improved by analyzing larger set of historical data.
- The study conducted are based on external perspective by using data available openly. Subject matter expert or local people involvement might improve the framework further.

**Challenges and will it works on other cities?**

- Data availability is a challenge, especially when relying on API and location services which would probably require subscription to a more premium services for smoother data extraction.

- Replicability of the study to other cities/countries would require almost similar open data to be available (in this case NYPD data) .This might not be openly available in other part of the globe. As it will highly dependeding on country or city policies on open data

- Naming and Border of neighborhood and are loosely defined compared to other more established parameter such as country name or zip code. Therefore, suitable geojson data with context detail and consistency need to be sought from perhaps local location data providers

# Conclusion

This work is done as part of requirement for IBM capstone Datascience certificate by Coursera.

We presented a prototype of city metric that can be used to assess city or neighborhood similarity or strength. In realizing that we introduced 5 parameters (population, income, crime, sport and green infrastructure availability).

The metric aims to provide a measuring framework that can benefit various stakeholders including the authorities, city planners, travelers, business owners or other social or economic researchers.

Case studies are conducted on Manhattan neighborhood by using data from Data from various sources (NYPD, Foursquare location data, geojson) 5 Clusters of neighborhoods are formed by using K-Means clustering which categorizes the neighborhood into similar profiles.

While framework is aimed to be replicated to other cities, we discussed some of the challenges, gaps and recommended improvement for future projects.

## References :

**Code** : https://github.com/Azman-Ali/CapstoneIBM/blob/master/BattleCode.ipynb

**Foursquare:** *www.foursquare.com*

**NYPD Open Data**
*https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i*

*https://www.researchgate.net/figure/An-Approximate-Mapping-of-NYPD-Precincts-to-NYC-Community-Districts_tbl1_247923137,*
*NYPD Report Data 2019 by Precint :https://www1.nyc.gov/site/nypd/stats/crime-statistics/historical.page*

**Coursera :** https://www.coursera.org/learn/applied-data-science-capstone/home/welcome