BATTLE OF THE NEIGHBORHOOD

Manhattan

"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"

Prepared by : Azman Ali

What ?

What is the project about?

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 etc) Neighborhoods are clustered and categorized based on similar profiles or attributes

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

Why?

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.

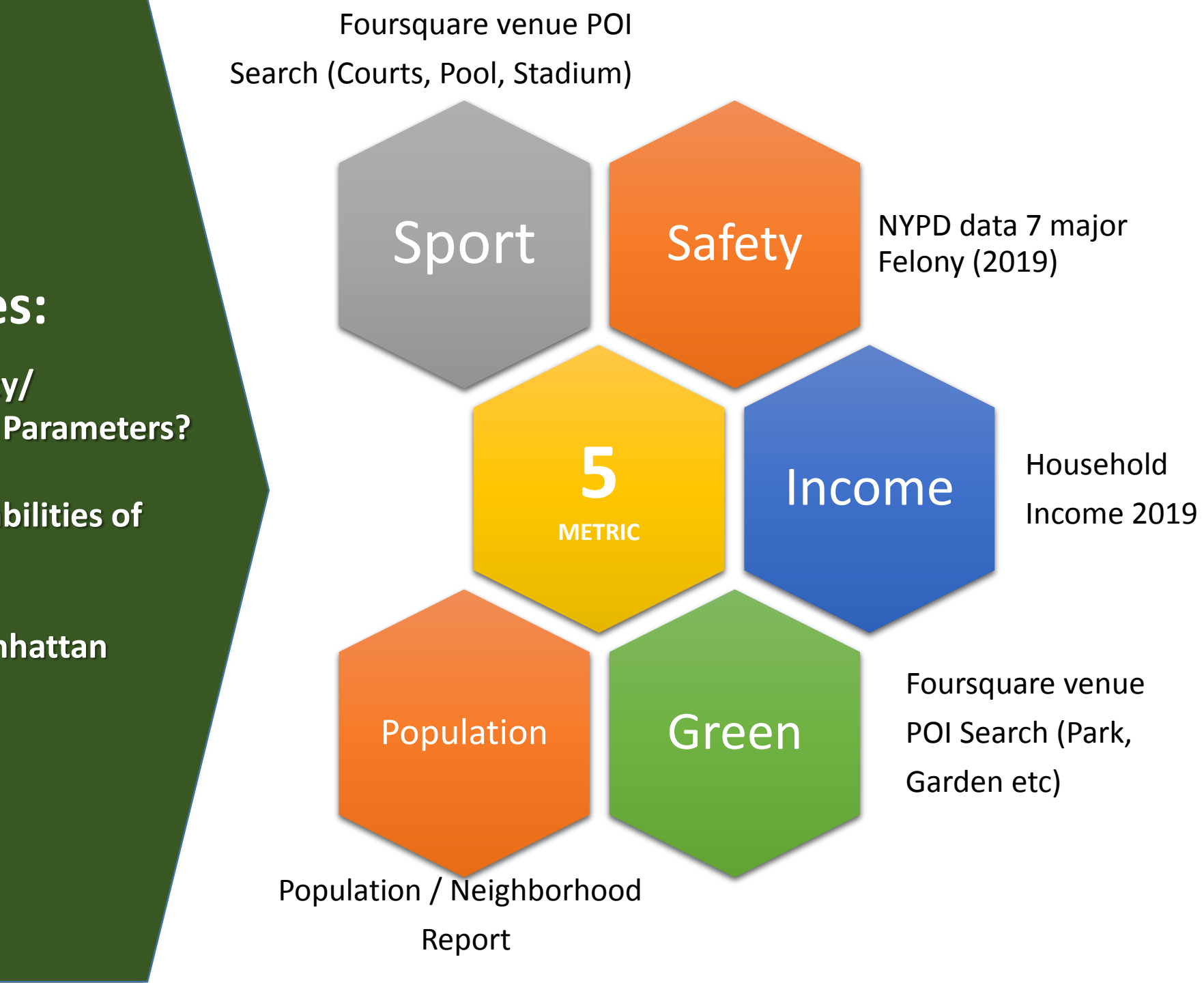
For the citizen’s benefit, the public or tax payers are expecting more transparent government and information that involve public interest. Travelers can also benefit from the city profiling to look for particular interest

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.

How ?

Setting up Objectives:

- Can we Measure/Profile a City/ Neighborhood based these 5 Parameters?
- How strong are interdependabilities of these parameters?
- Conduct a Case Study on Manhattan

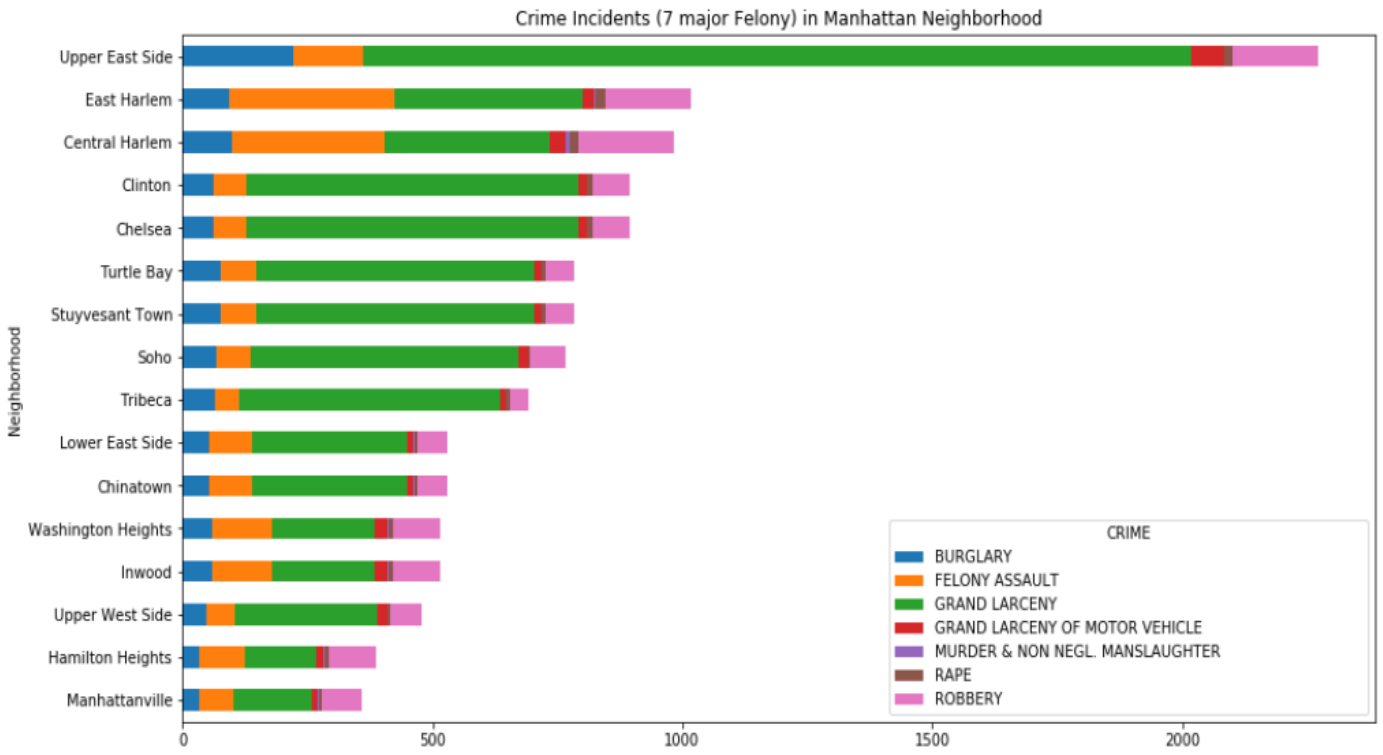


INSIGHTS
&
DISCUSSION

Highest Total Crime in Upper East side

Grand Larceny is the most frequent Crime in Manhattan

Robbery, Felony Assult and Burglary are equally distributed across all neighborhood



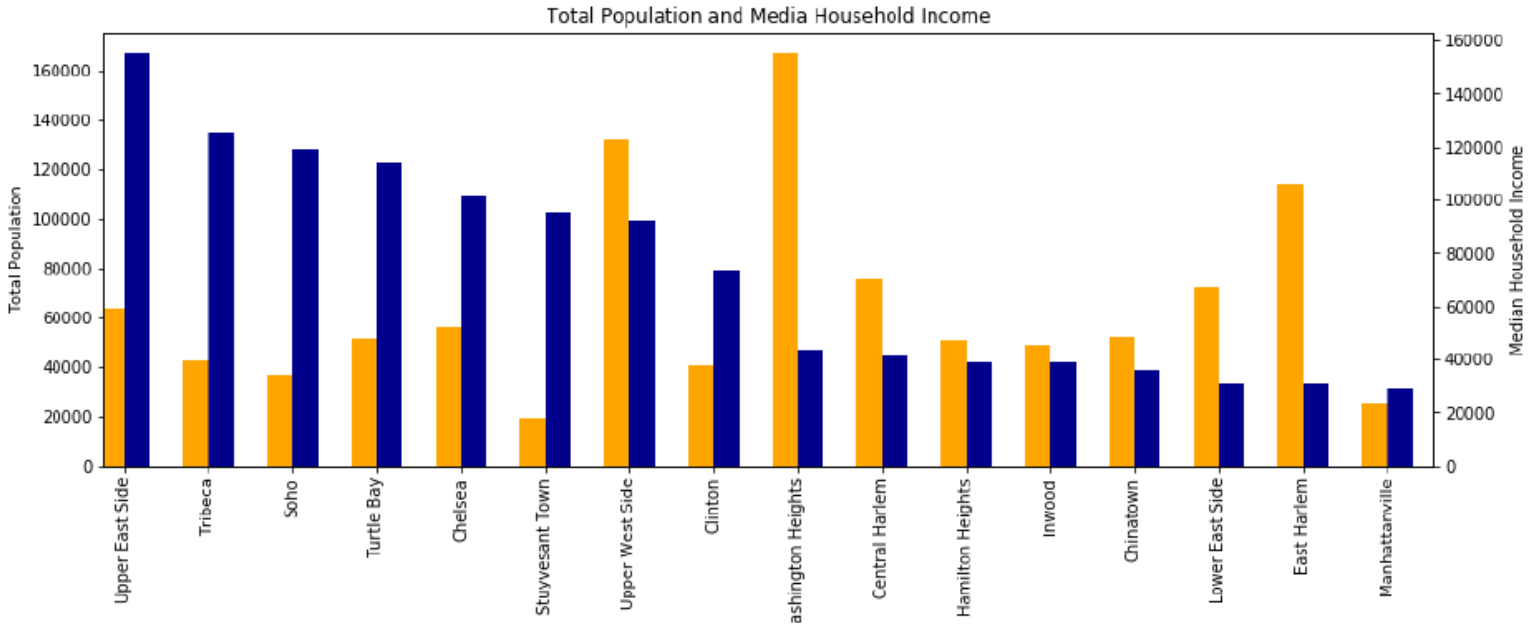
Top 6 highest median income are from population **lower** than average neighborhood **population**

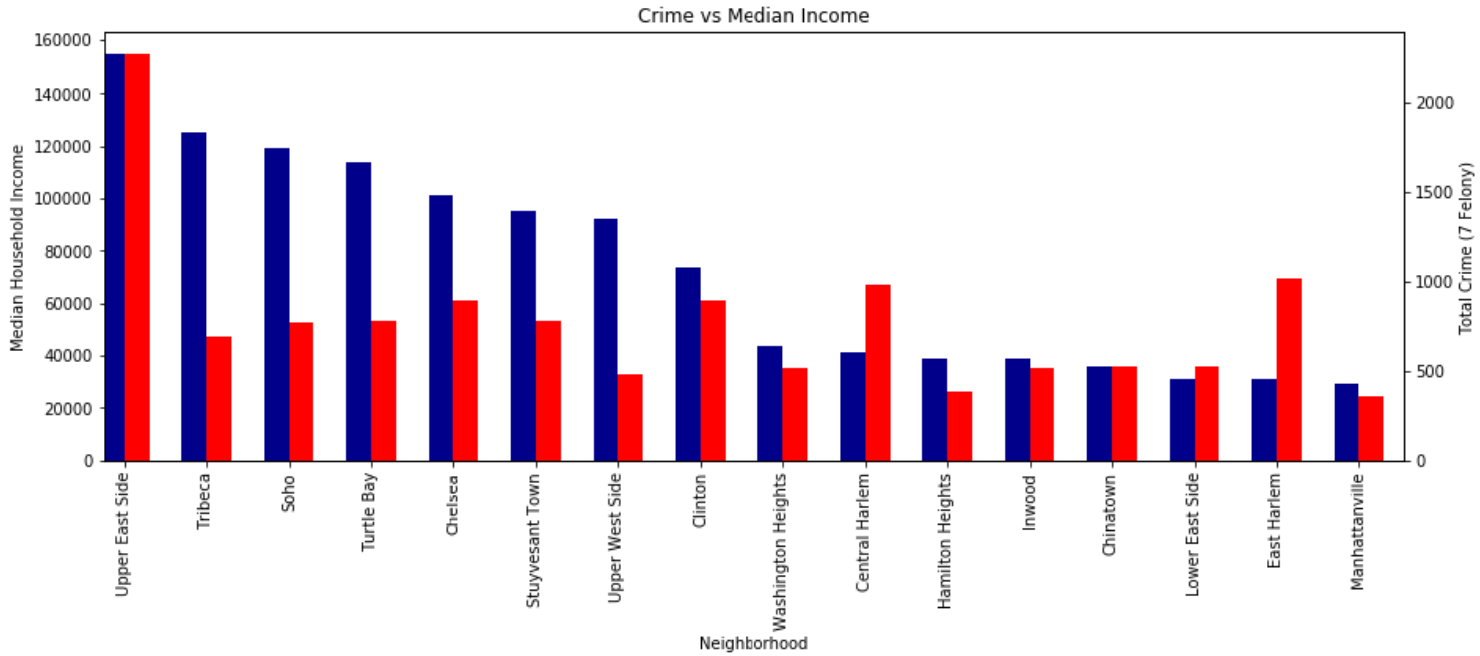
The **6 lowest Populated** neighborhood in contrary earned **less than average income**

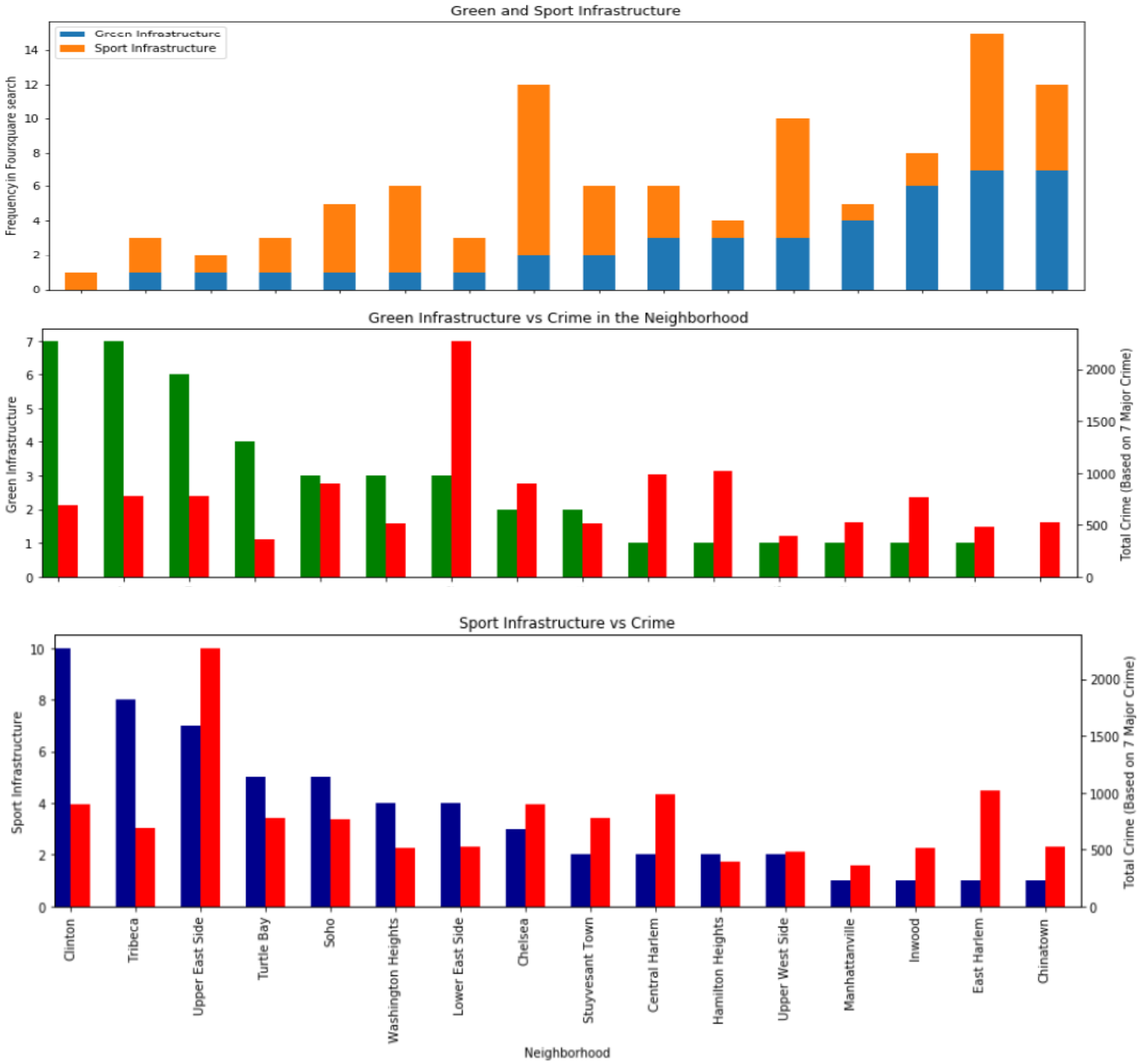
Washington Height and **East Harlem** are most populated area with **below average household income** (anomaly)

Crime distribution are generally fairly distributed despite median household income

Upper East side ,East Harlem and Central Harlem are the most area with reported **crime** in 2019







Green area are hardly visible in **Chinatown**

>50% of neighborhoods has below average Green infrastructure availability

Clinton, Turtle Bay, Tribeca and **Upper East Side** has more sports and **Green infrastructure** than others

NO significant influence of **Green infrastructure** toward crime statistic

Upper East side has **highest crime** reported in 2019

CRIME incidents are **fairly distributed** across manhattan

How Strong are Parameters Co-related?

Parameters with Positive (+ve) Correlation

Household Income and Total Crime, (coef: 0.5946)

"There are moderately high correlation between average income a nd Crime in the neighborhood. So crime are likely to increase in a higher income community"

Sport Infrastructure and Crime (coef: 0.415397)

"Crime are likely happening in the area with sporting facilities"

Sport Infrastructure and Income (coef: 0.60320)

Higher income area have more sport facilities then the lower inco me area.

Green infrastructure vs Income (0.503112)

"Similarly Green infrastructure are more when people earns more"

	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

Negative (-ve) Correlation are :

Green Infrastructure vs **Population**

(Coef : -0.402467)

The higher population means lesser green spaces in the area

Multilinear Regression for Target Variable= Crime

Target (Crime) =

Independent Variables:

(0.014168250784884876) * **(Population)**

+ 0.49353995203090123 * **(Median Income)**

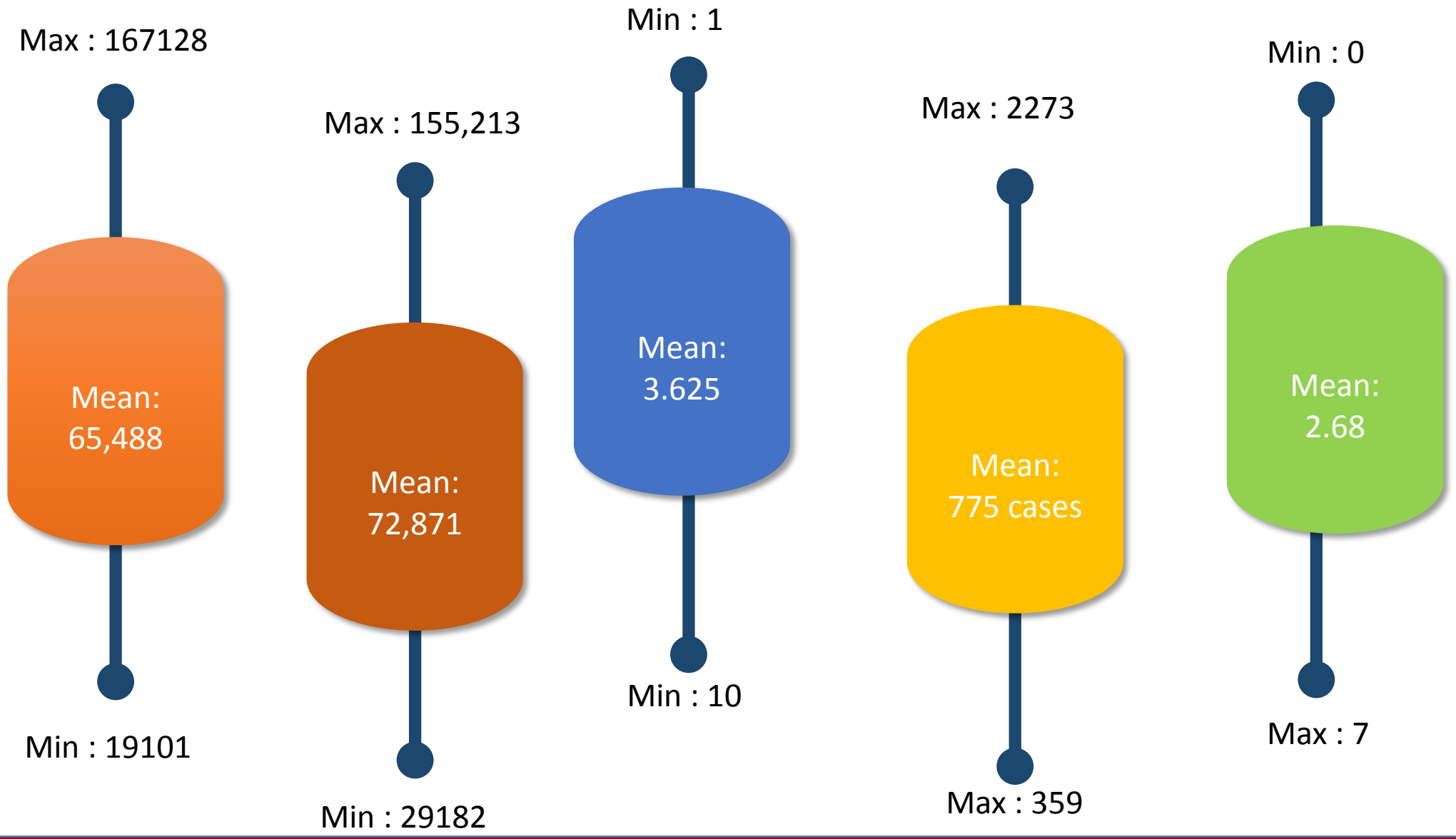
+ 0.06861183698536587 * **(Green Infrastructure)**

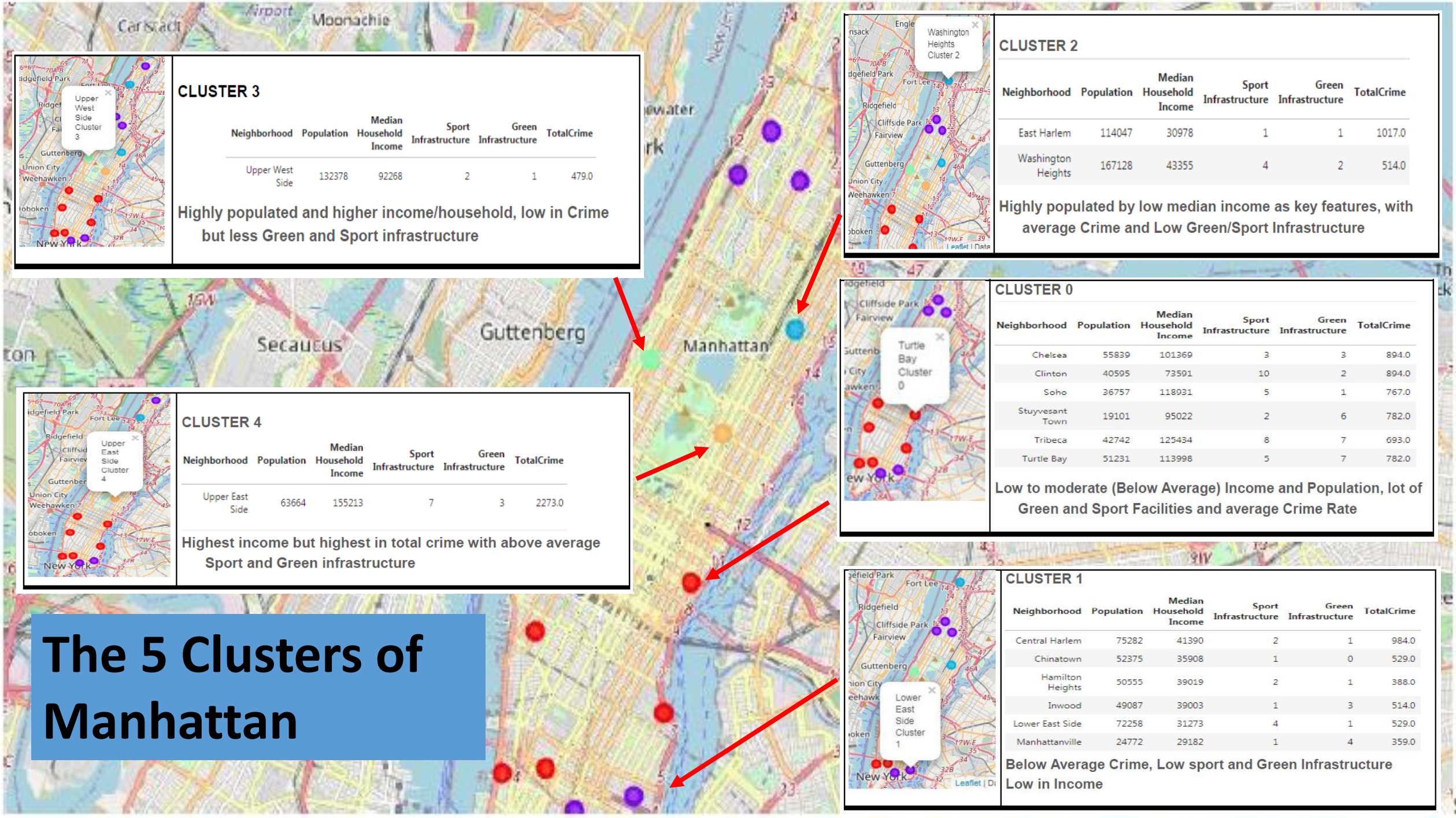
+ -0.2148155862830695 * **(Sport Infrastructure)**

Intercept: 0.10421619754242245

"Crime are likely to increase in the area with higher household income and Sport Facilities"

OVERALL STATISTICS FOR 5 PARAMETERS





CLUSTER 3

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

Highly populated and higher income/household, low in Crime but less Green and Sport infrastructure

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 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 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

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

The 5 Clusters of Manhattan

DISCUSSION

City Metric can indeed be developed but not limited to the 5 Parameters used in this case studies. The complexity of behavior we are trying to measure required more parameters .This can be other life aspect such as education, businesses, healthcare, transportation/mobility etc.

In general no single parameters are directly contributed to lower crime rate regardless of income & population or whether there are relatively high availability of sport and green infrastructures.

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

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

(Cluster 0) : 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 , 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 :
(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 within Manhattan.

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's 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 required almost similar open data to be available (in this case NYPD data) which might not be openly available in other part of the globe. This will highly depending 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 providers

Thank You

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