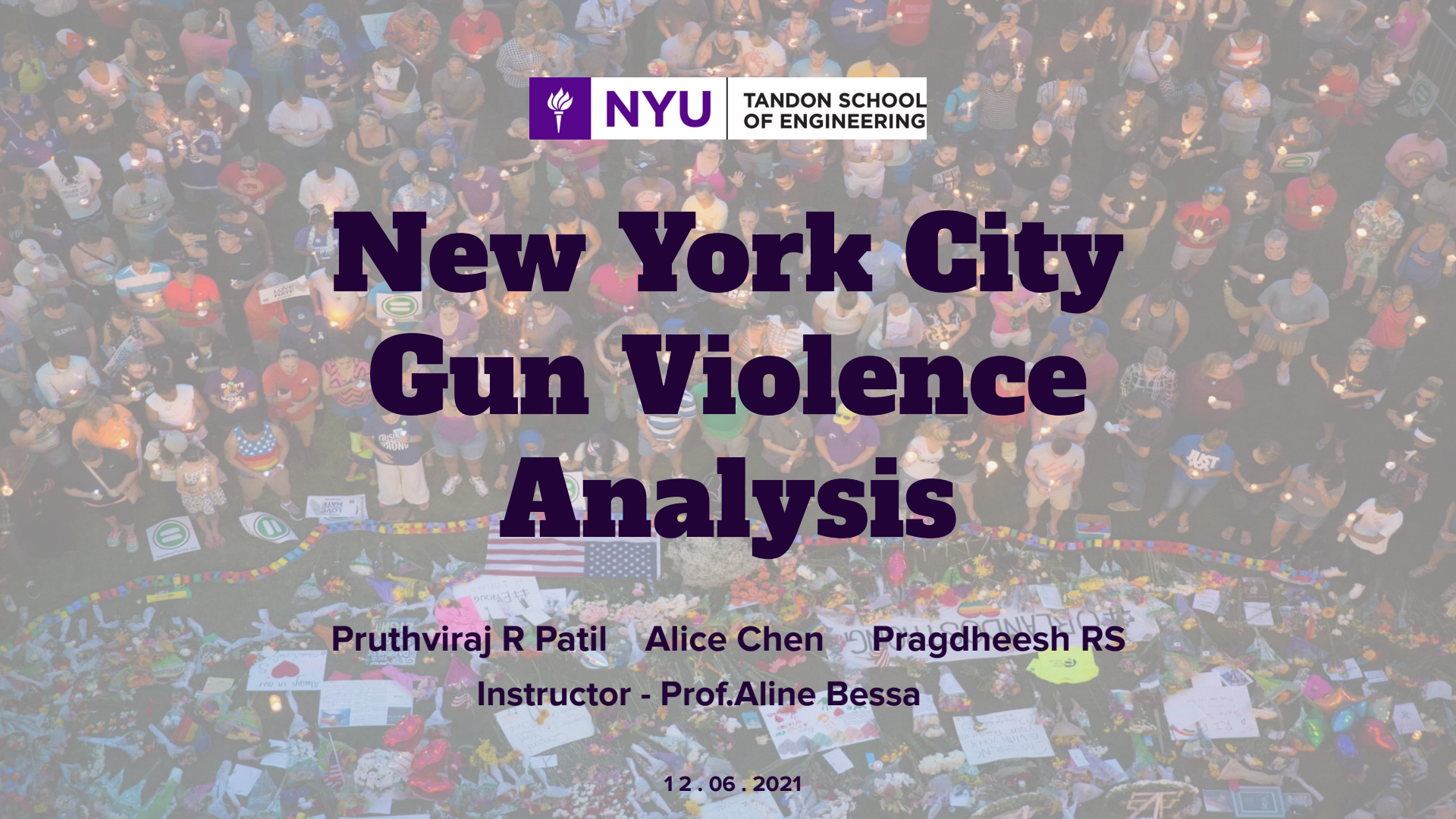


NYU TANDON SCHOOL OF ENGINEERING

New York City Gun Violence Analysis

Pruthviraj R Patil Alice Chen Pragdheesh RS

Instructor - Prof.Aline Bessa

12 . 06 . 2021

NYU

ETA PIPELINE

01 → 02 → 03 → 04

ARTICULATION OF PROBLEM STATEMENT

To Analyse the Gun Violation incidents in NYC based on various factors and bring out useful insights.

COLLECTION OF DESIRED DATA

Importing two main datasets from NYC Opendata and Kaggle. Analysing and Aggregating them.

DESIGNING MODEL BASED ON ALGORITHM

Clustering the gun violence incidents using K-Means (Top Down) and Agglomerative Clustering (Bottom Up)

FURTHER ANALYSIS AND CONCLUSION

Further Analysis using additional data and getting useful insights using NYC census tract files. Also, concluding the research.

DATA ENGINEERING

NYU

Datasets used

- NYC Census tract shapefiles. Source - https://data.cityofnewyork.us/City-Government/2010-Census-Tracts/fxpq-c8ku
- Kaggle Dataset - Gun violence in USA : Each data point with respect to perpetrator (1 tuple denote = 1 incident). https://www.kaggle.com/duttadebadri/gun-violence-in-usa-insights-forecast
- NYC Open data - NYPD Shooting Incident Data : Each data point with respect to victim (multiple tuples denote = 1 incident). https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8

Kaggle Dataset

'incident_id', 'date', 'state', 'city_or_county', 'address', 'n_killed', 'n_injured', 'incident_url', 'source_url', 'incident_url_fields_missing', 'congressional_district', 'gun_stolen', 'gun_type', 'incident_characteristics', 'latitude', 'location_description', 'longitude', 'n_guns_involved', 'notes', 'participant_age', 'victims_age_group', 'participant_gender', 'participant_name', 'participant_relationship', 'participant_status', 'participant_type', 'sources', 'state_house_district', 'State_senate_district'

(TOTAL- 29 VARIABLES)

'Date' (further splitted into 'day', 'month', 'year', 'day of week'), 'city_or_county', ('n_killed'+'n_injured'), 'latitude', 'longitude', 'victims_age_group', 'participant_gender'

(TOTAL - 10 VARIABLES)

DATA ENGINEERING

NYC Open Dataset

'INCIDENT_KEY', 'OCCUR_DATE', 'OCCUR_TIME', 'BORO', 'PRECINCT', 'JURISDICTION_CODE', 'LOCATION_DESC', 'STATISTICAL_MURDER_FLAG', 'PERP_AGE_GROUP', 'PERP_SEX', 'PERP_RACE', 'VIC_AGE_GROUP', 'VIC_SEX', 'VIC_RACE', 'X_COORD_CD', 'Y_COORD_CD', 'Latitude', 'Longitude', 'Lon_Lat'

(TOTAL- 19 VARIABLES)

'OCCUR_DATE (further splitted into 'day', 'month', 'year', 'day of week')', 'OCCUR_TIME', 'BORO', 'Latitude', 'Longitude', 'Lon_Lat', PERP_SEX, VIC_AGE_GROUP

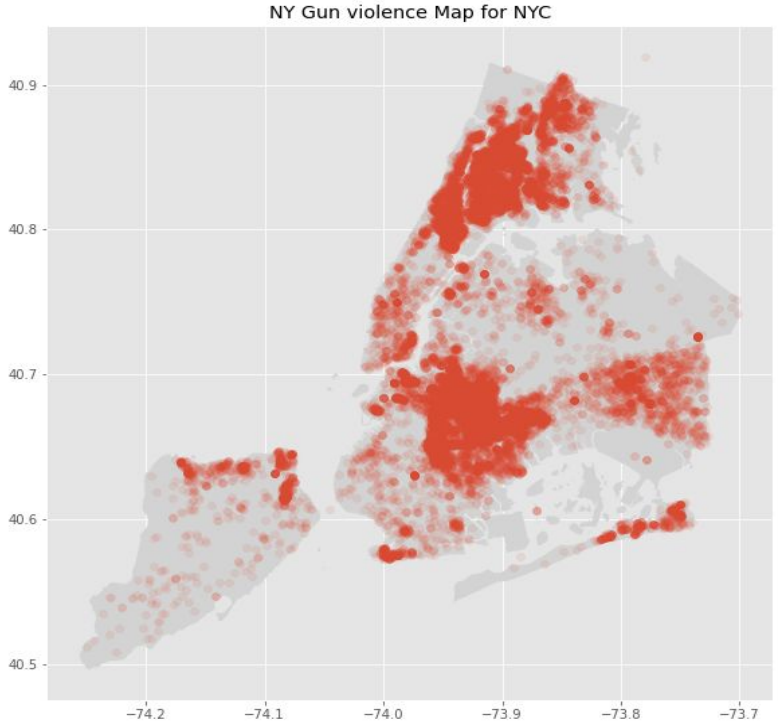
(TOTAL - 11 VARIABLES)

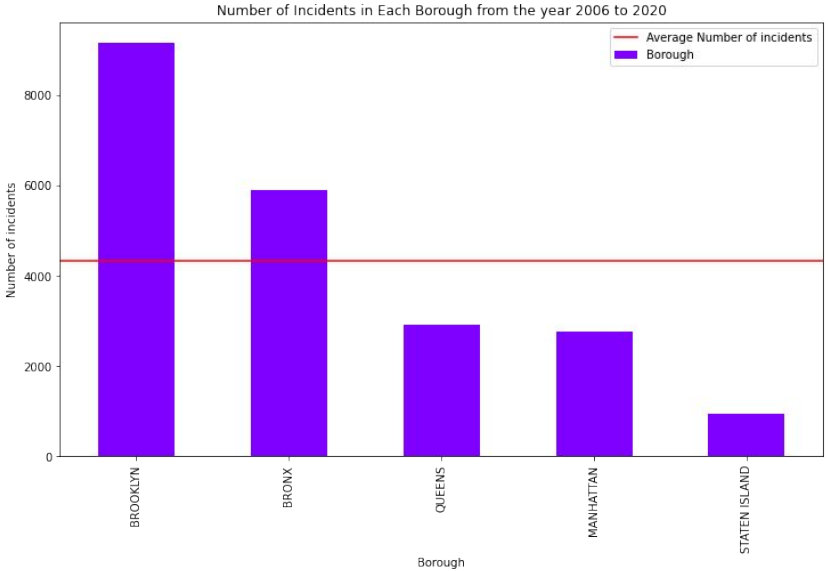
NYU

Data aggregation strategy :

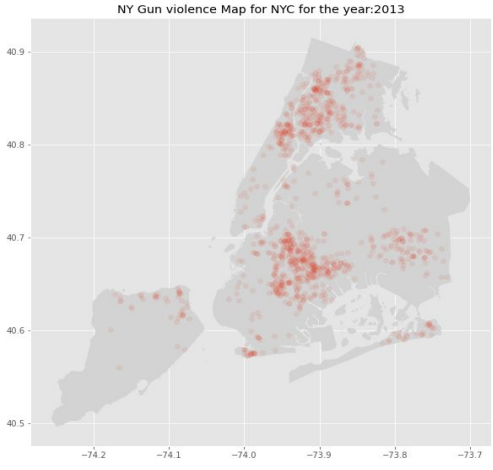
- Combined the tuples in the dataset-2 to convert it into 1 incident = 1 tuple format.
  - Latitudes and Longitudes alone were useless as there can be incidents occurring at same place at multiple times.
  - Used Date and Time along with Latitudes and Longitudes to club tuples into one incident.
  - ‘VIC_AGE_GROUP’ column contained list in every tuple denoting every individual’s age group in that incident.
- Vertically joined both the datasets, totally getting around 18466 incidents from the year 2006 to 2020.
- Removed the Null values with respect to Latitudes and Longitudes.

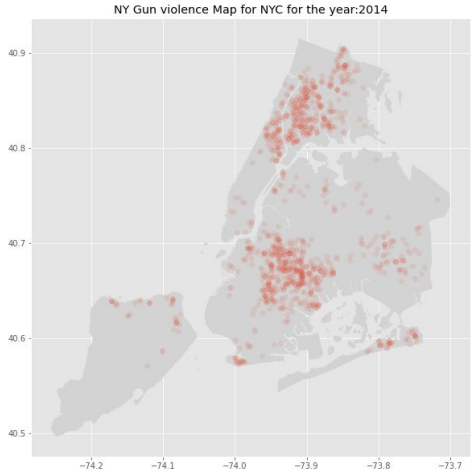
DATA-ANALYSIS

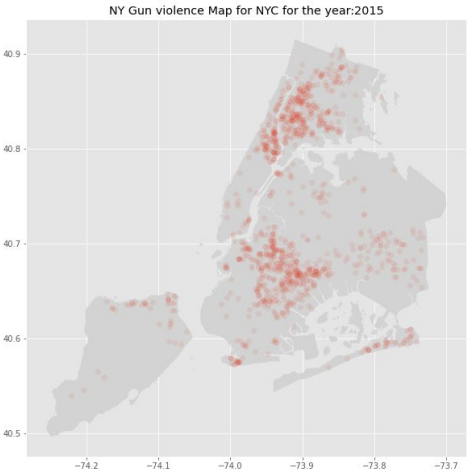


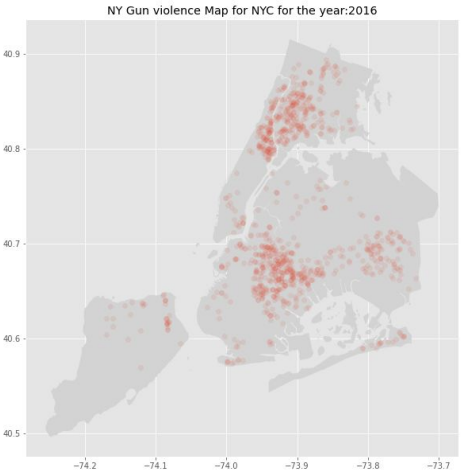


! Brooklyn having the max average crime rate !

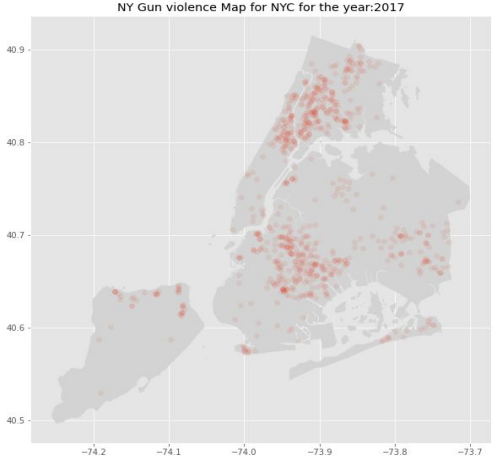


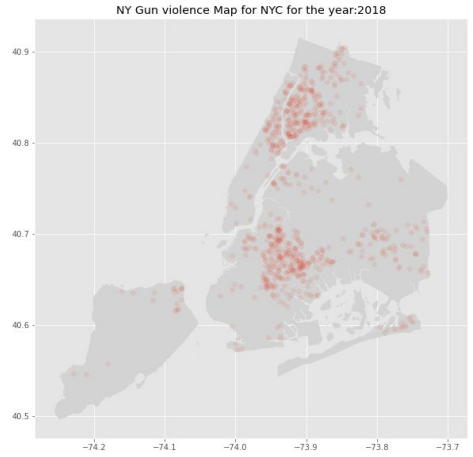


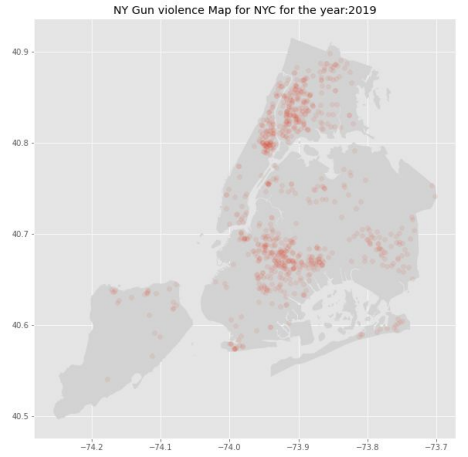


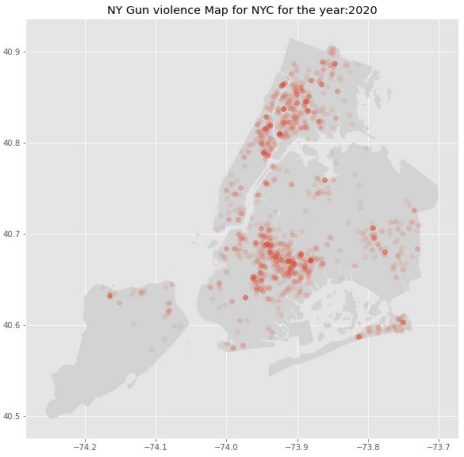


GUN VIOLENCE MAPS FROM THE YEAR 2013-2020



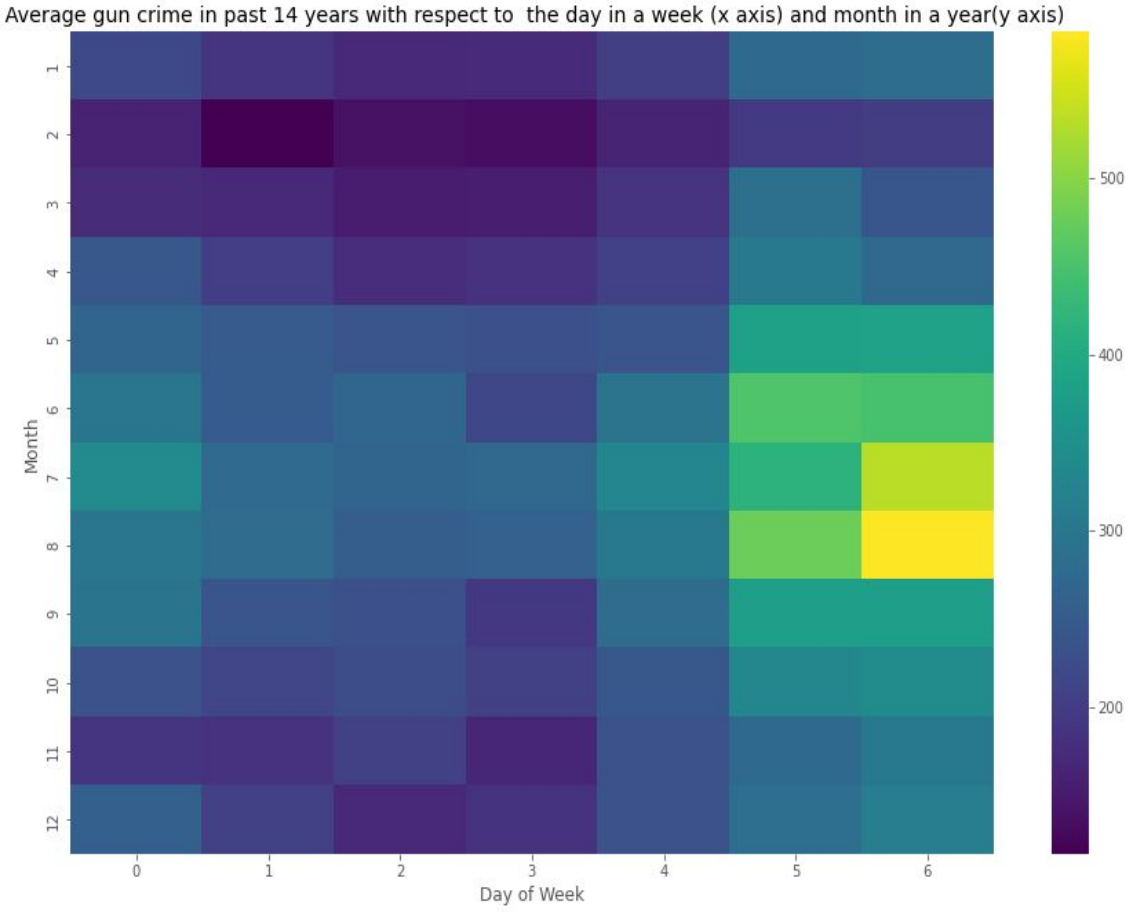




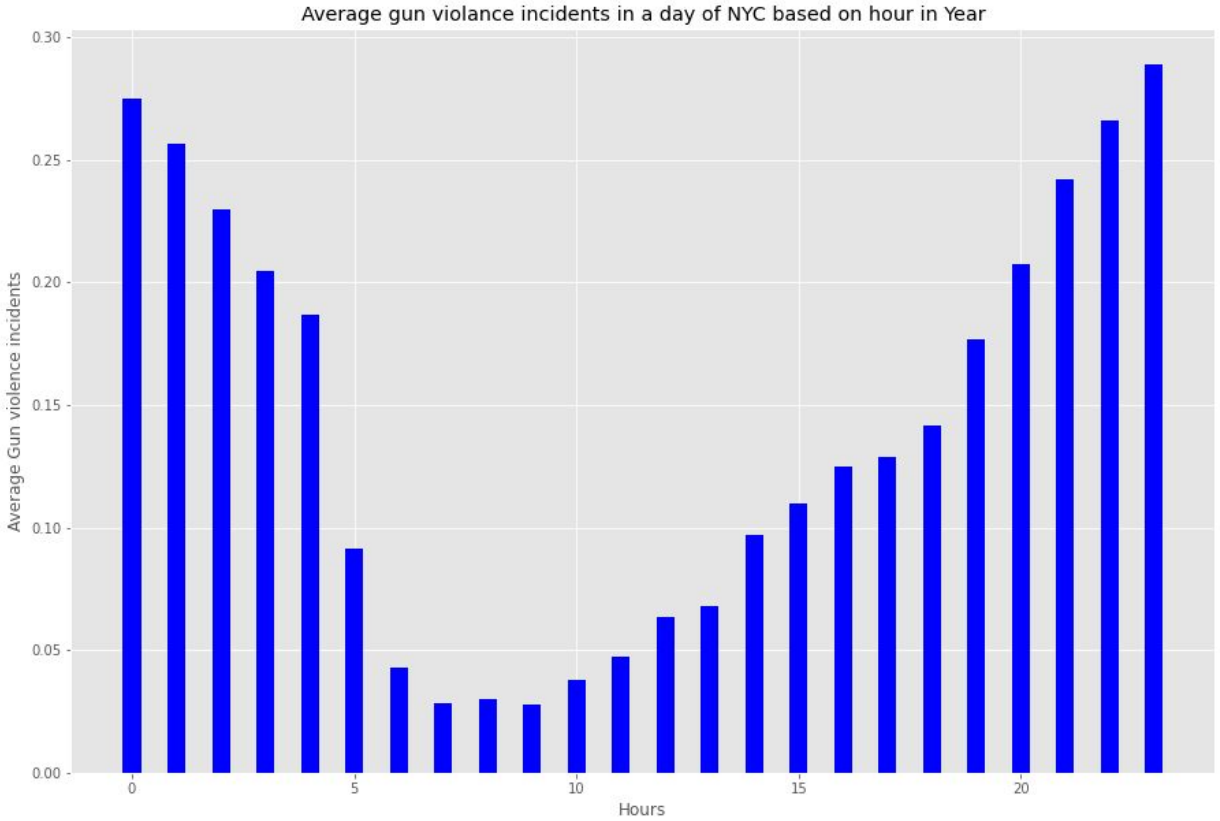


Max violence during Summer-Fall weekends!!

Can we attribute weather (in turn, population density on roads) to this aspect?



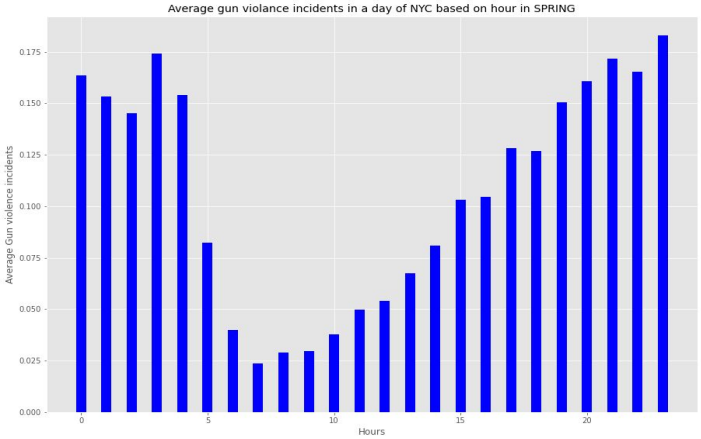
NYU



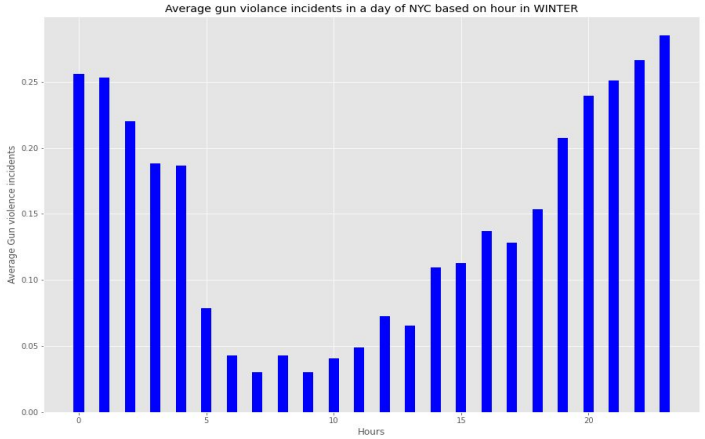
The chances of Gun violence at night in NYC is approximately 7 times more than day time!

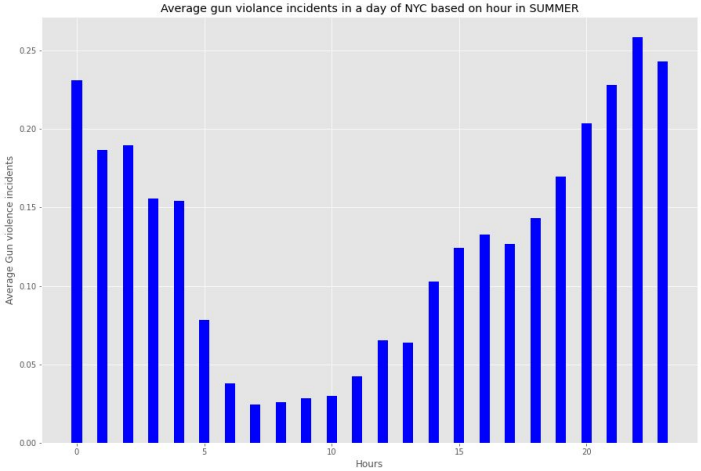
DATA CONSIDERED : NYC OPEN DATA (2013-2020)

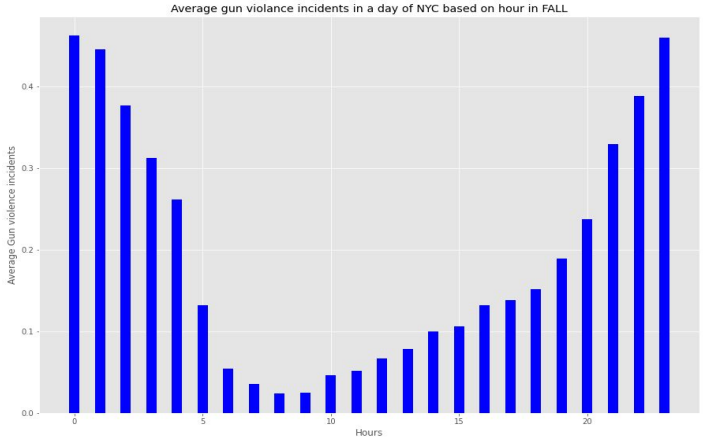
SPRING



WINTER



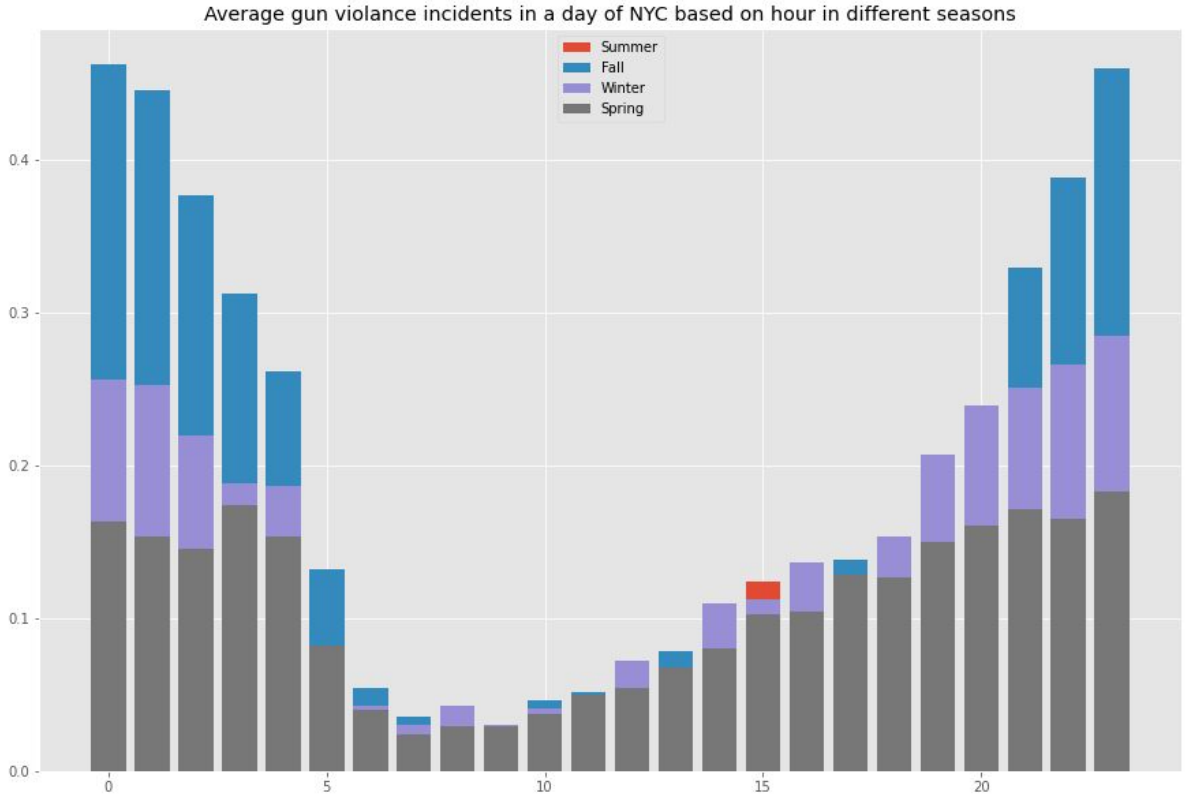




SUMMER

FALL

NYU

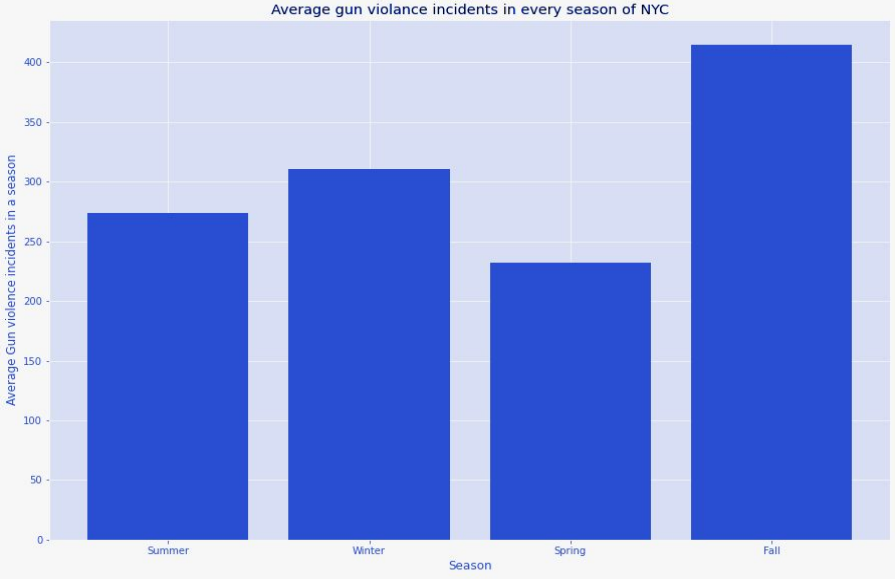


To be precise, Fall weekend nights are shady!

DATA CONSIDERED : NYC OPEN DATA (2013-2020)

55%

More gun violence in Fall than Spring season!



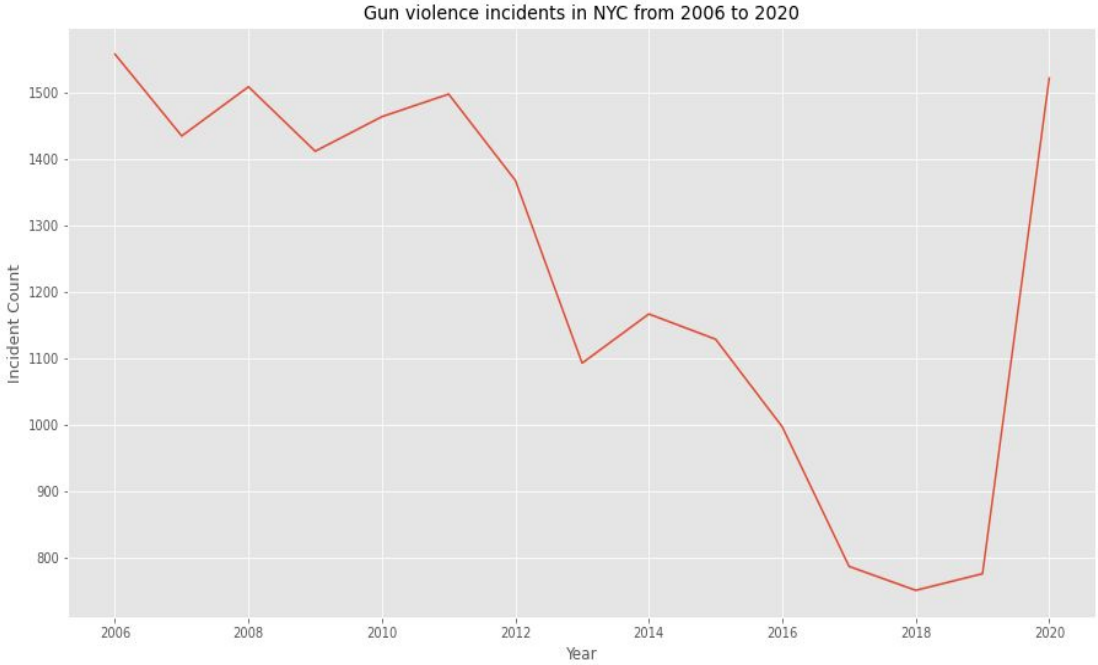
Certainly, pandemic has worsened the gun violence rate. Is unemployment the reason?

New York City + Add to myFT

New Yorkers fear return of 'bad old days' after shootings surge

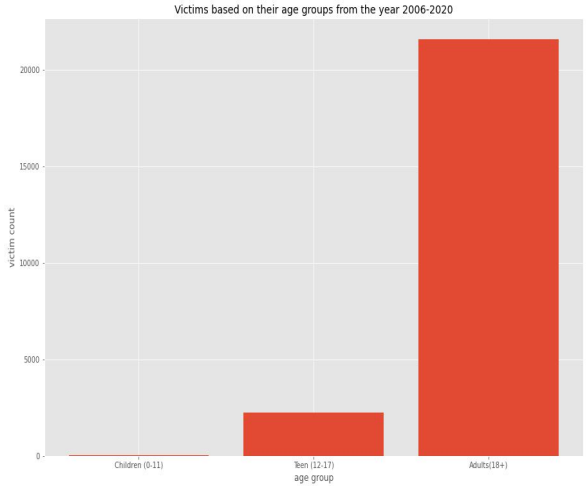
City is rattled by rising tide of violence following a long period of relative peace

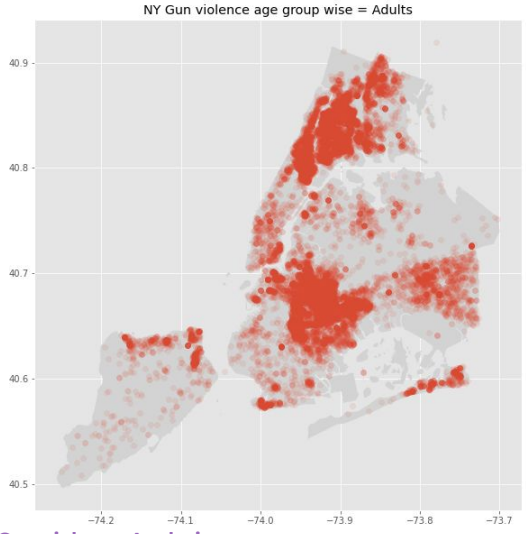
A police officer stands near the scene of an afternoon shooting that left one person dead on July 7 in Brooklyn © Spencer Platt/Getty

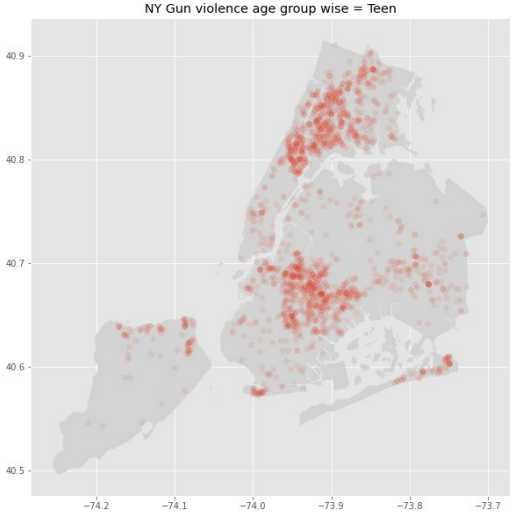


Source:
https://www.ft.com/content/a444b43e-ba40-48b5-a064-fa51c591cc32

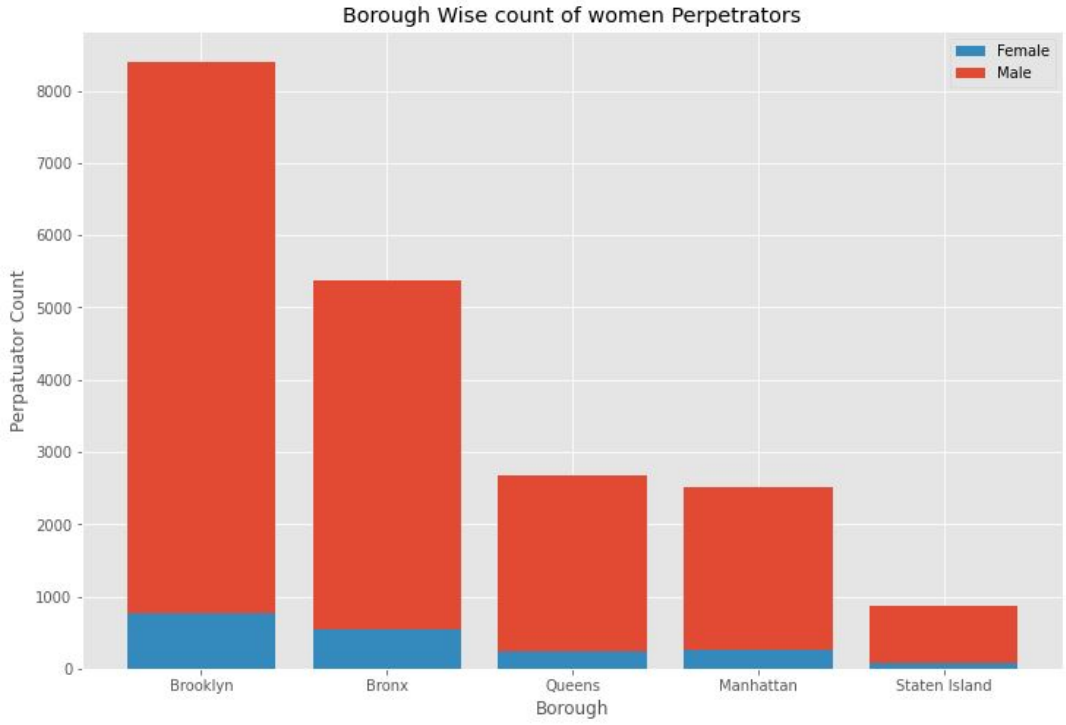
It is alarming that even Teenagers (age - 12-17) and Children (age <12) are the victims of the gun violence at NYC!!

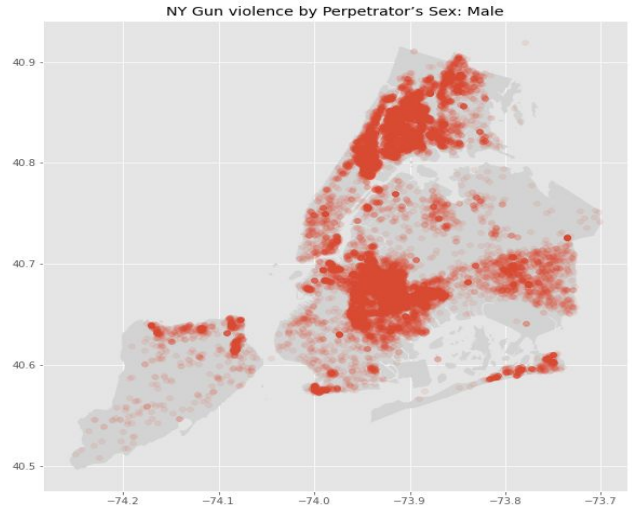


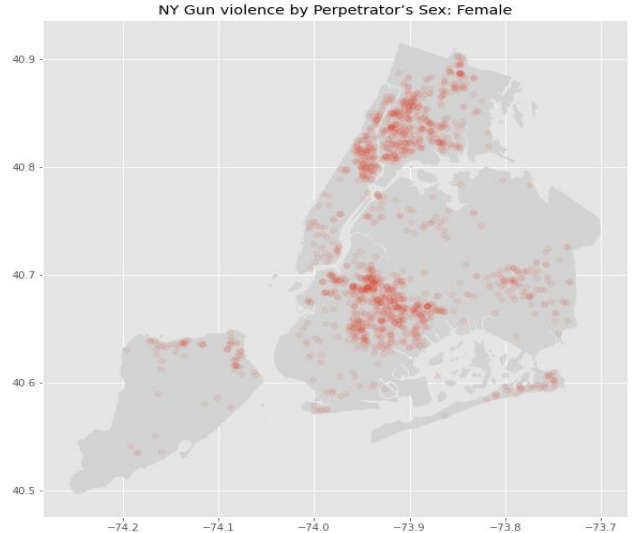




Perpetrator's Analysis (Gender Wise):







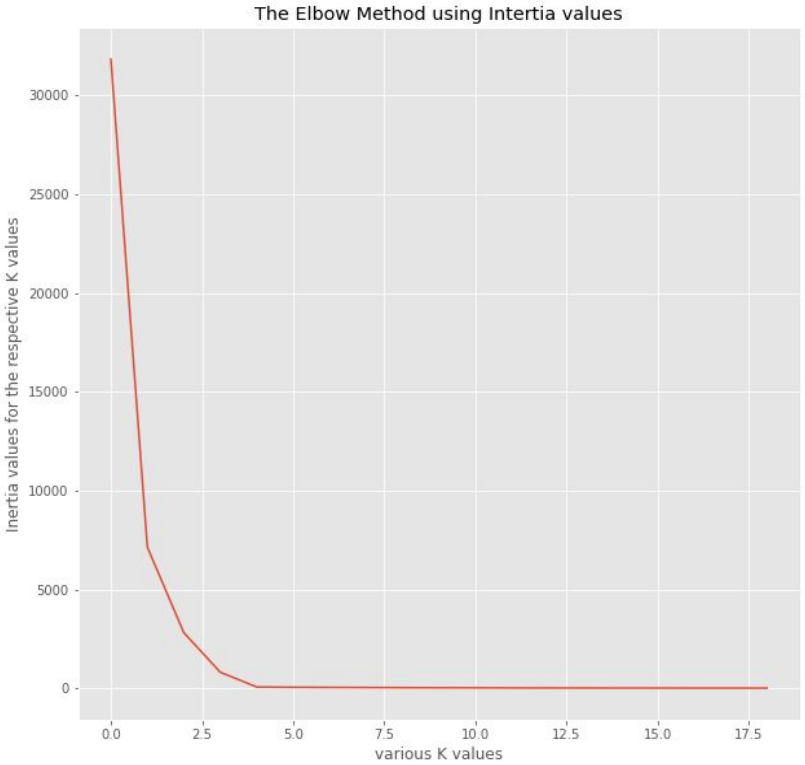
NYU

GEOSPATIAL-CLUSTERING ANALYSIS

NYU

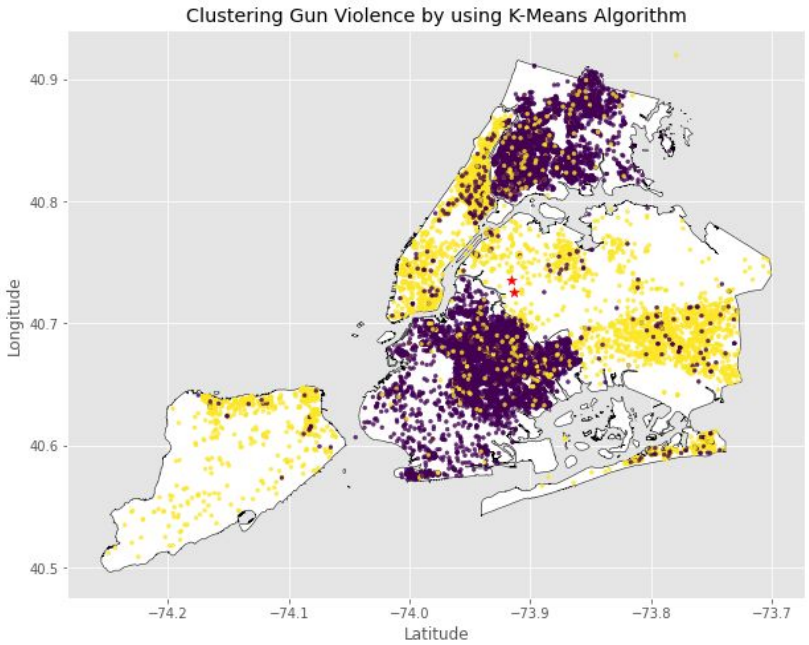
How did we negotiate number of clusters?

- Used Elbow method for deciding the number of clusters.
- As the Elbow is somewhere around 2.3. So, we will consider K=2 as well as K=3.
- Must be fun! Let's do the clustering!

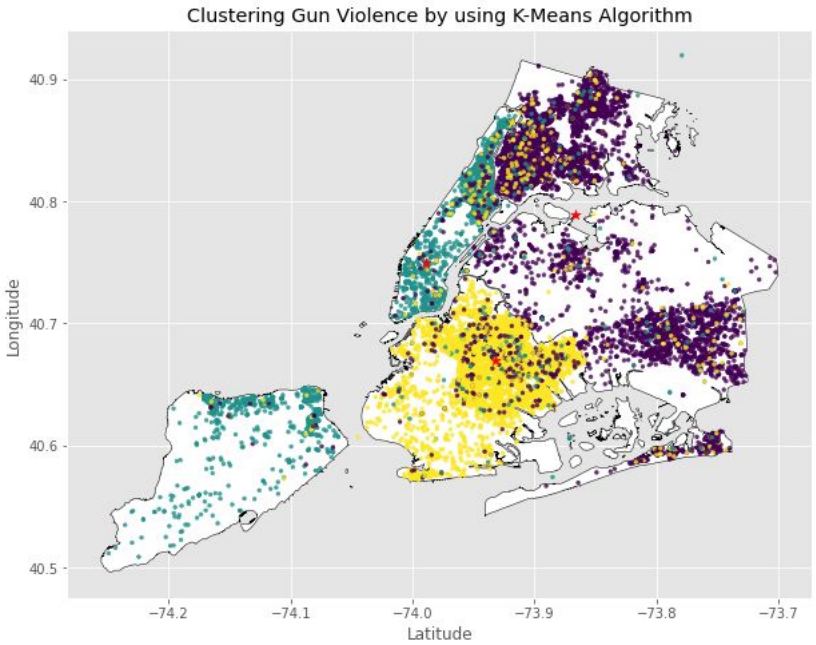


NYU

What do the K-Mean clusters have to say?..

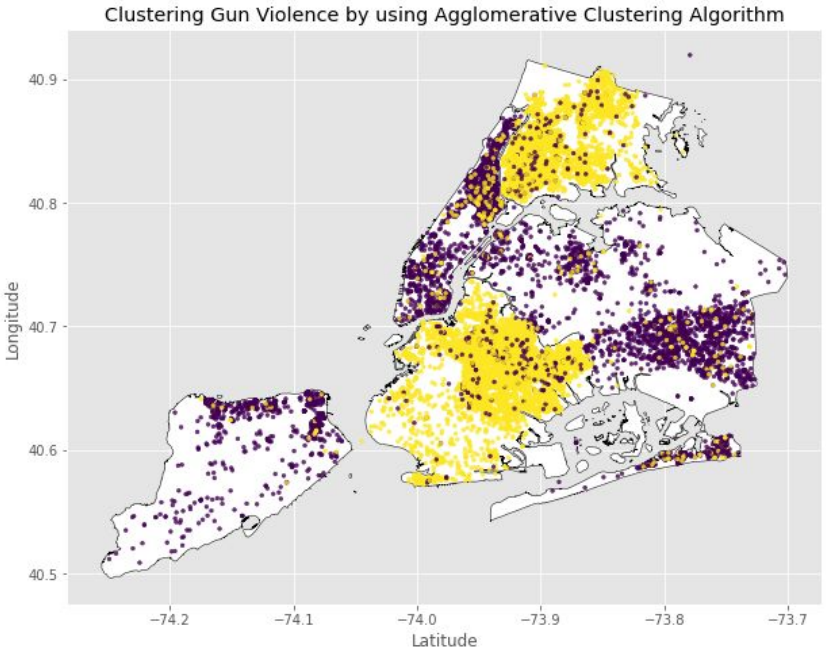


K=2

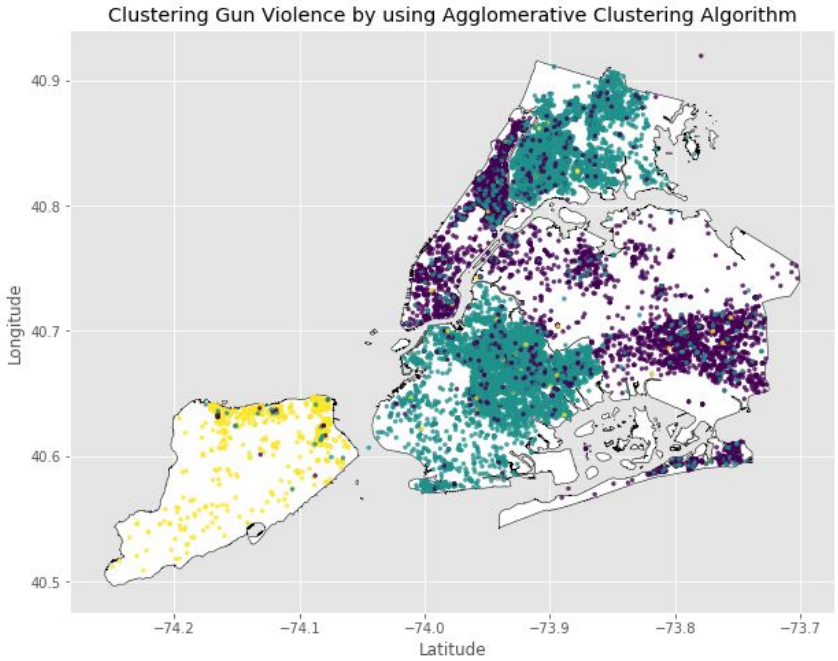


K=3

What do the Agglomerative clusters have to say?..



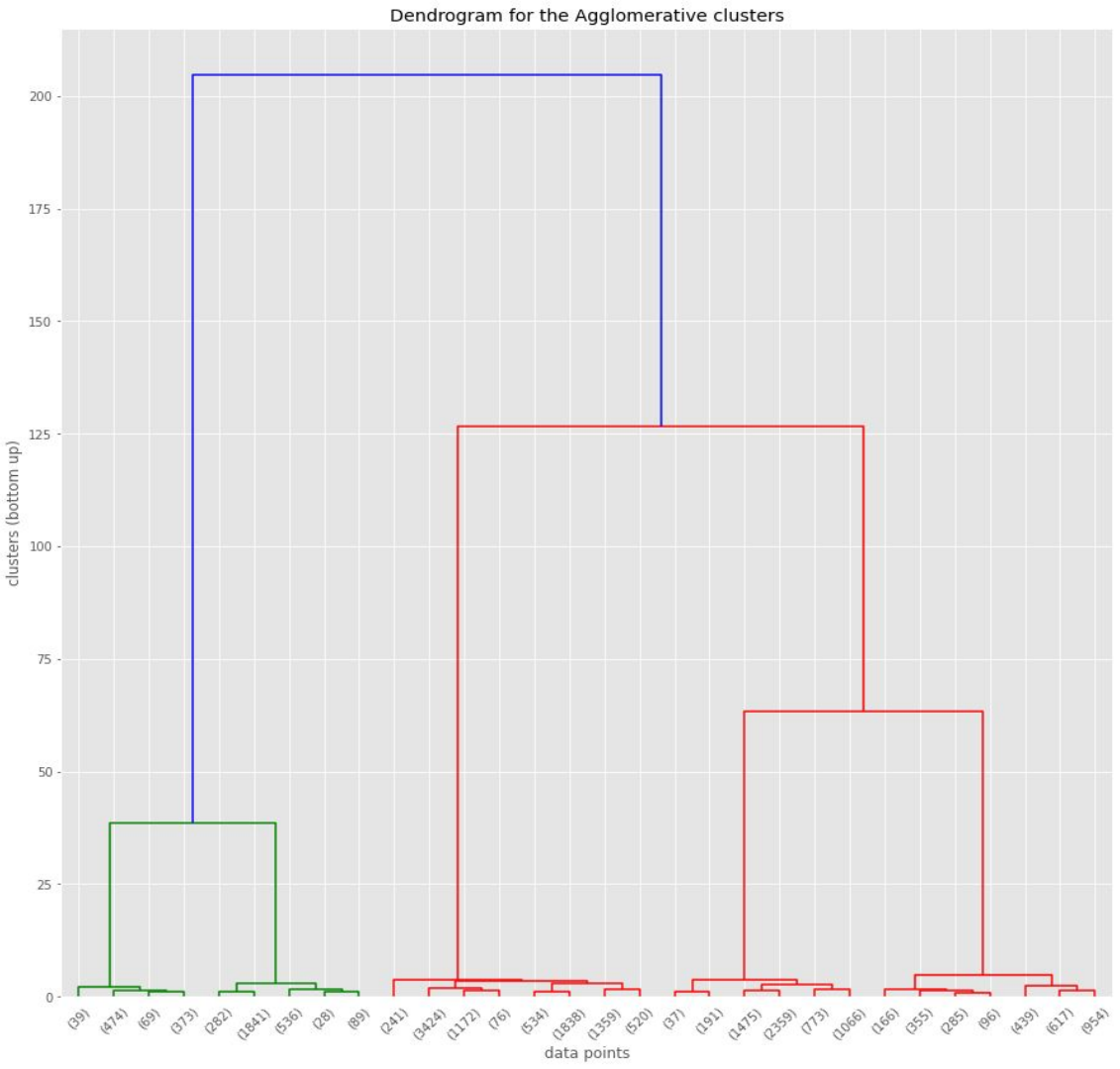
K=2



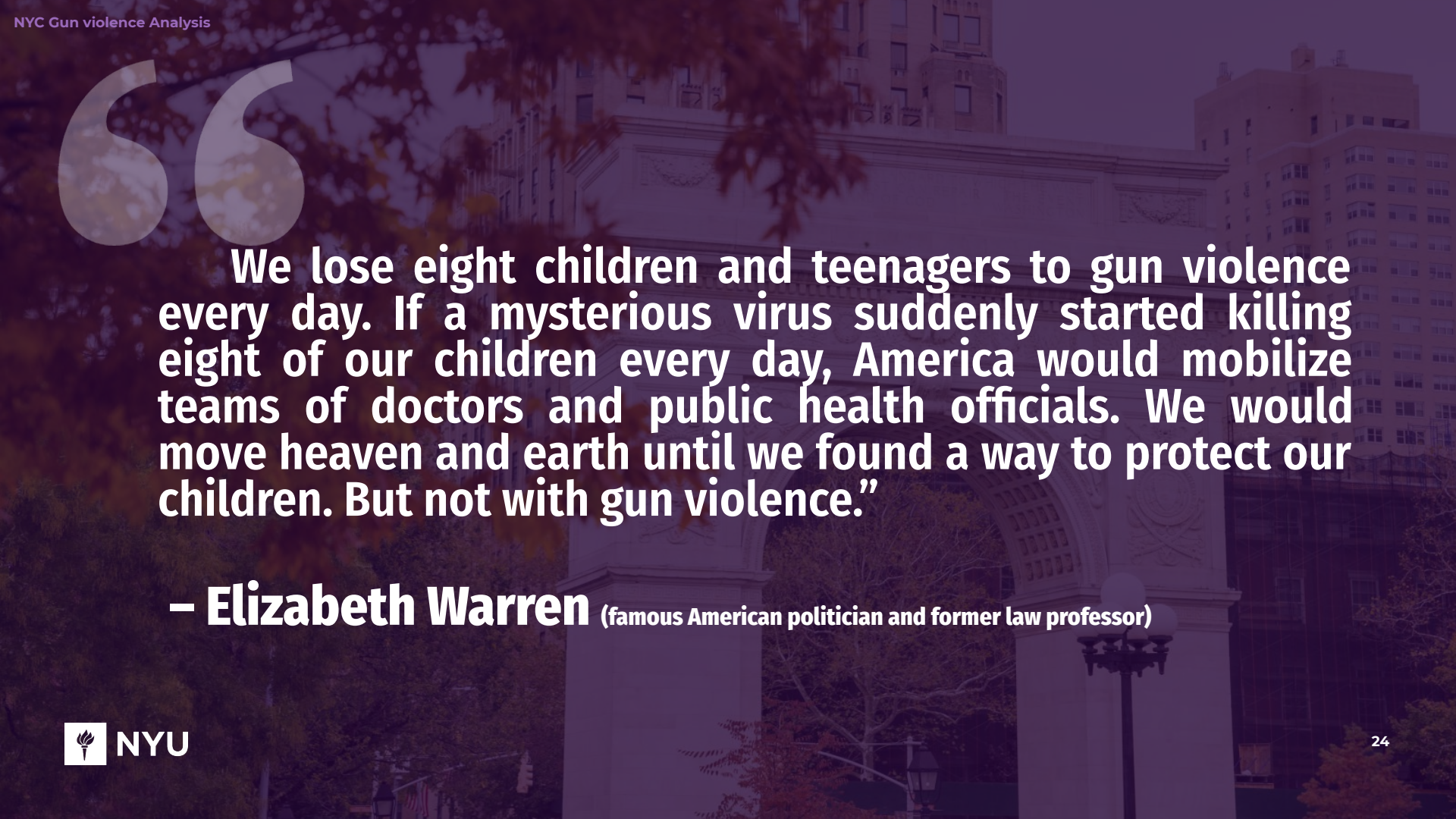
K=3

NYU

The Dendrogram that Speaks..



CONCLUSION



“We lose eight children and teenagers to gun violence every day. If a mysterious virus suddenly started killing eight of our children every day, America would mobilize teams of doctors and public health officials. We would move heaven and earth until we found a way to protect our children. But not with gun violence.”

– **Elizabeth Warren** (famous American politician and former law professor)

Thank you!