Explore Bias in the NYC Shooting Incident Dataset

Toan Luong

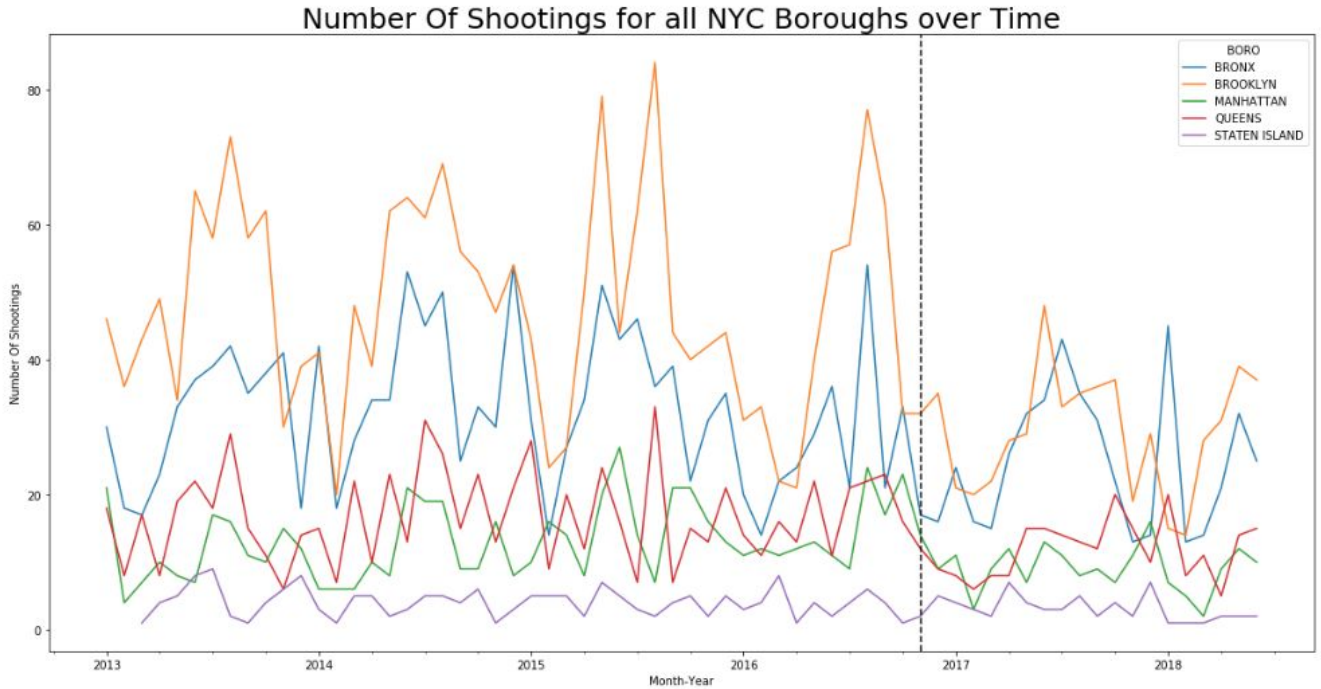
NYC OpenData

Data & Questions

Breakdown of every shooting incident that occurred in NYC between 2013 and 2018. “This data can be used by the public to explore the nature of shooting/criminal activity”

- Bias in statistics
  - Time-series plots of reports counts, age group, race, sex of perpetrators and victims
- Bias in missing data
  - Bar plots of victims’ demographics whose perpetrators’ information are missing
- Bias in geographical distribution
  - Folium interactive map based on the incidents’ GPS lat-long data

Time Trends - Report Counts



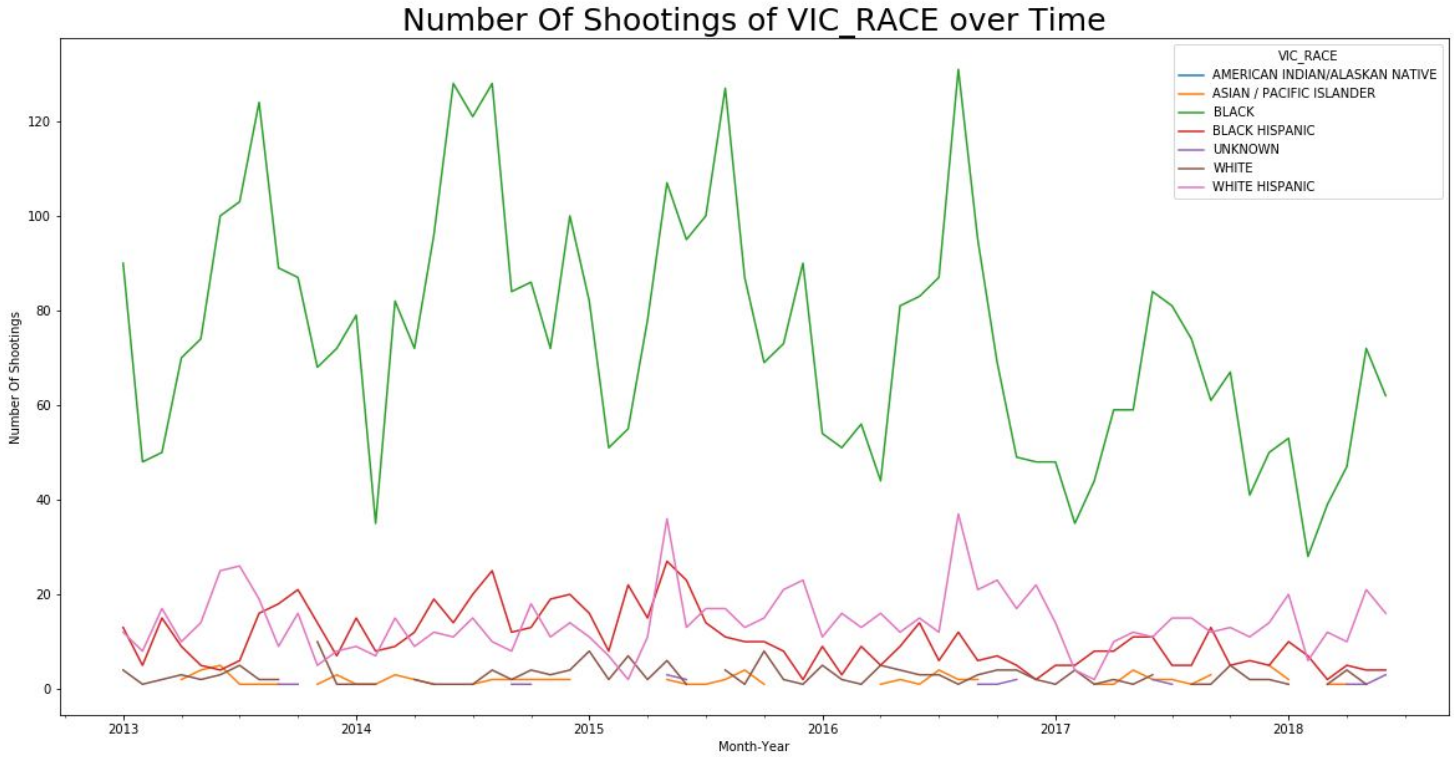
The New York Times

Crime in New York City Plunges to a Level Not Seen Since the 1950s

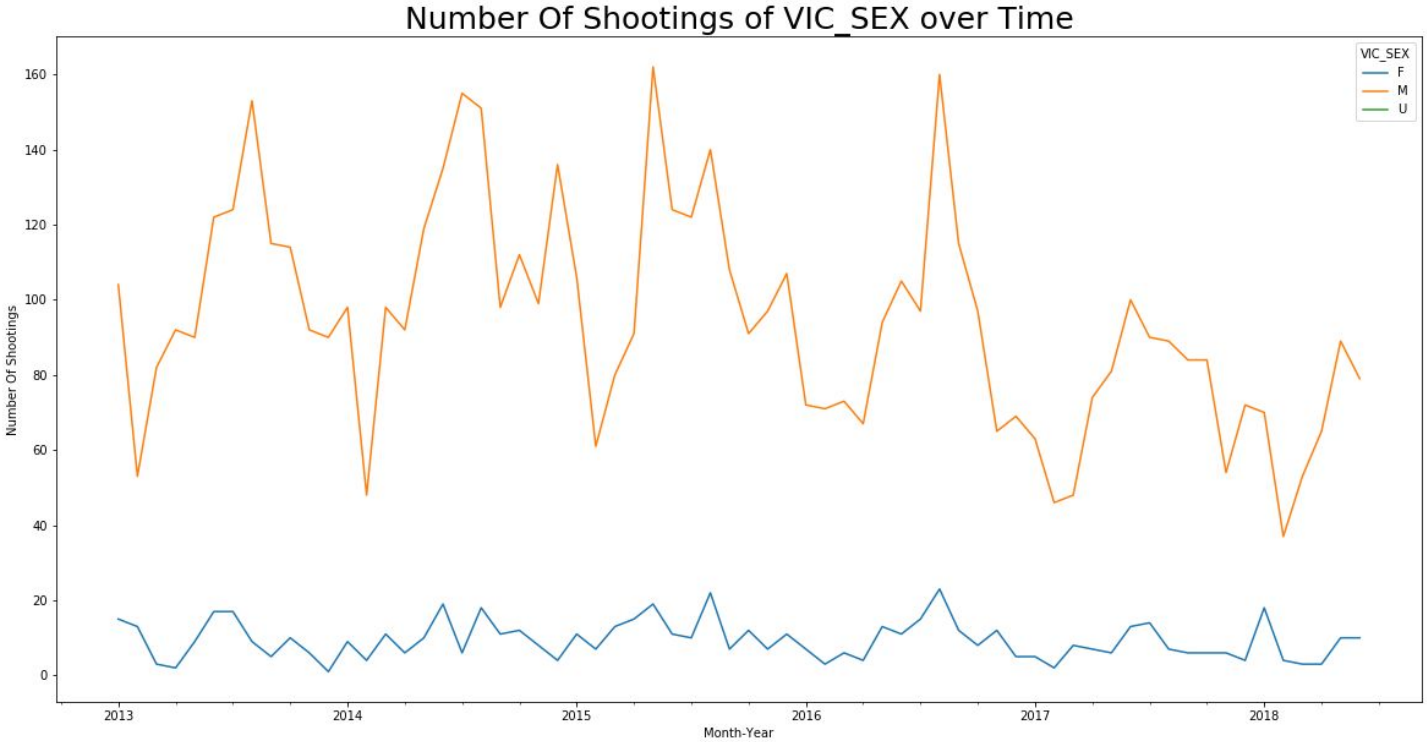
Area	Median House-hold Income	Mean House-hold Income	Percent-age in Poverty
The Bronx	$34,156	$46,298	27.1%
Brooklyn	$41,406	$60,020	21.9%
Manhattan	$64,217	$121,549	17.6%
Queens	$53,171	$67,027	12.0%
Staten Island	$66,985	$81,498	9.8%
New York City	$48,631	$75,809	18.5%
New York State	$53,514	$77,865	13.7%
United States	$50,140	$69,193	13.0%

Source: Wikipedia

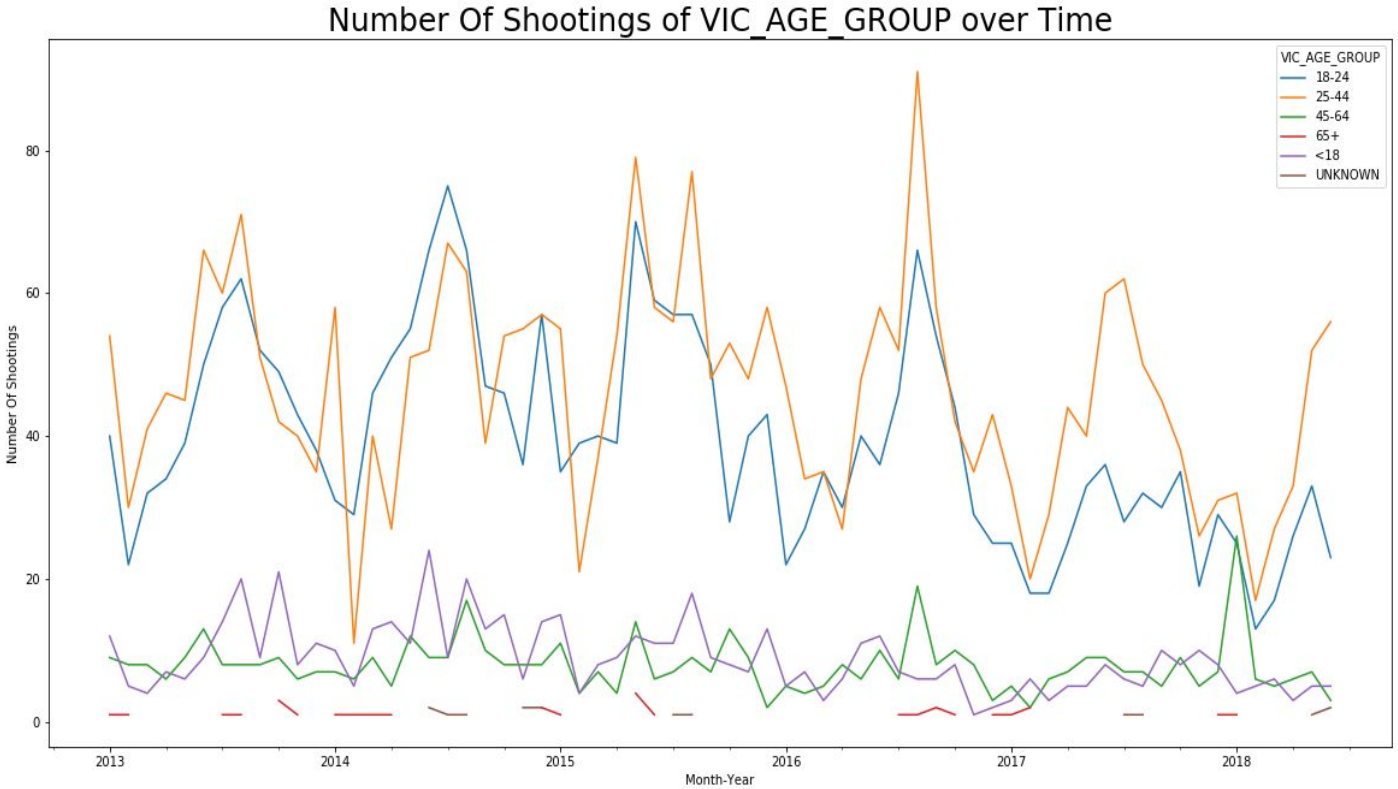
Time Trends - Race



Time Trends - Sex



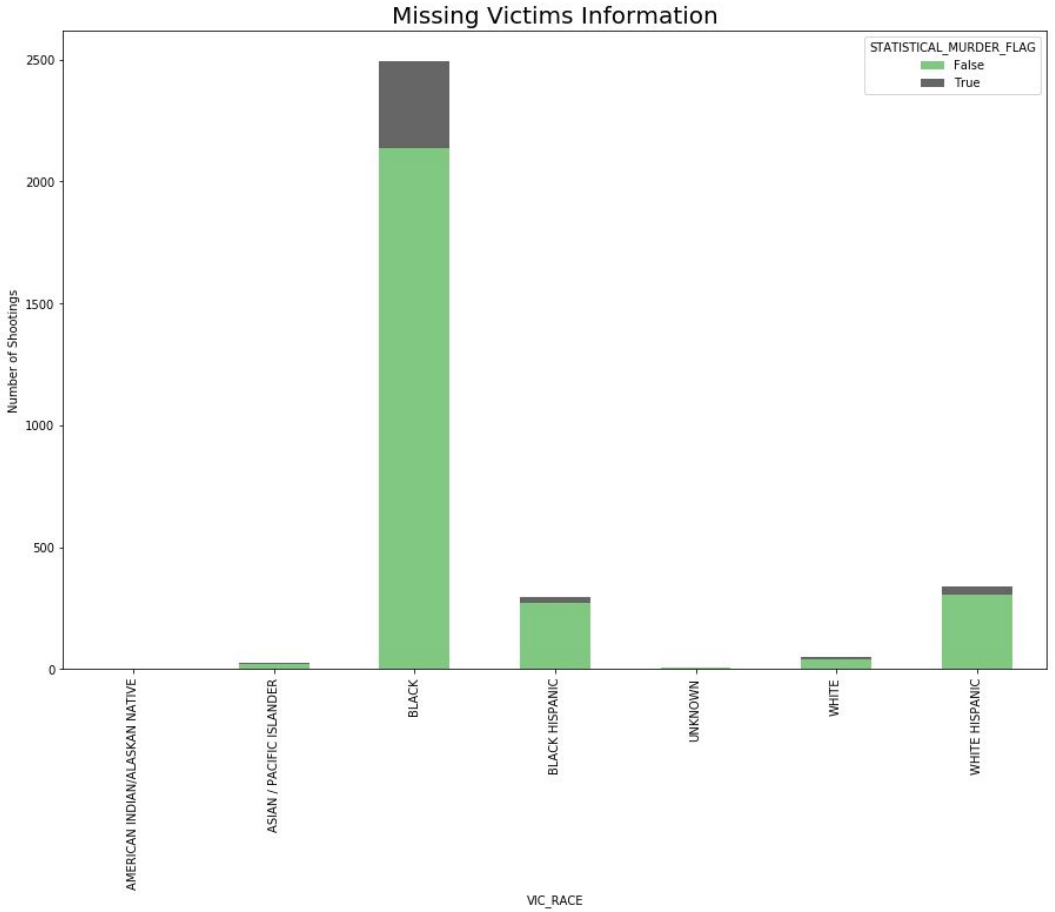
Time Trends - Age Group



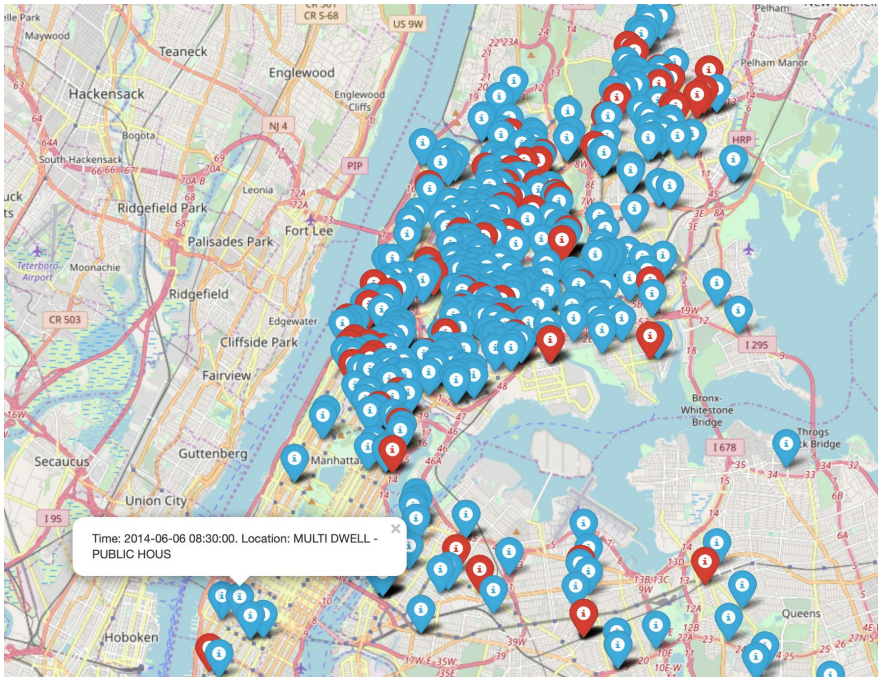
Missing Data Analysis

3,123 out of 6,848 victims' reports have missing perpetrators' information

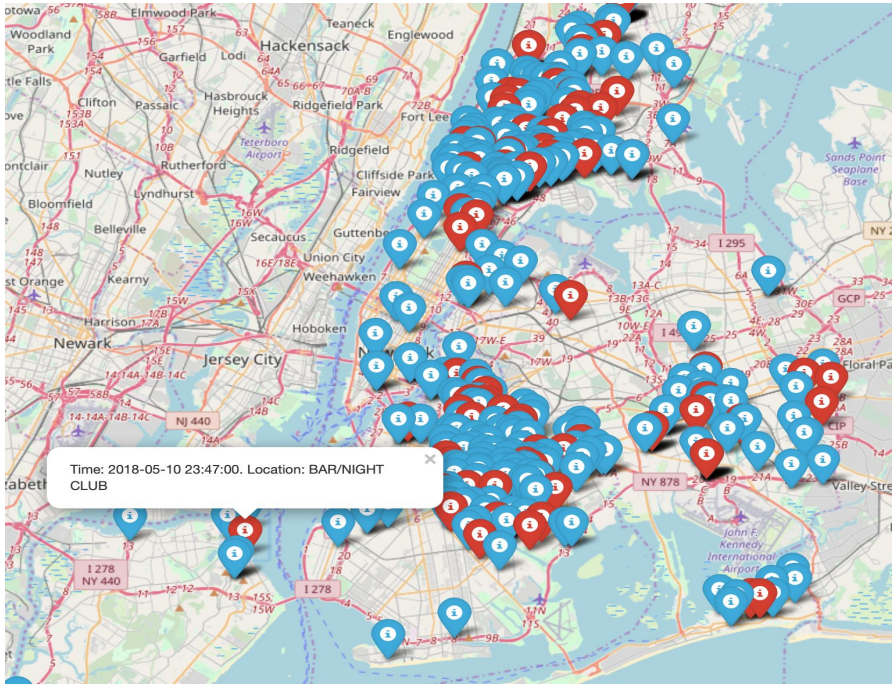
A majority of Blacks & Hispanic victims



Geographical Scatterplot



2014 Reports



2018 Reports

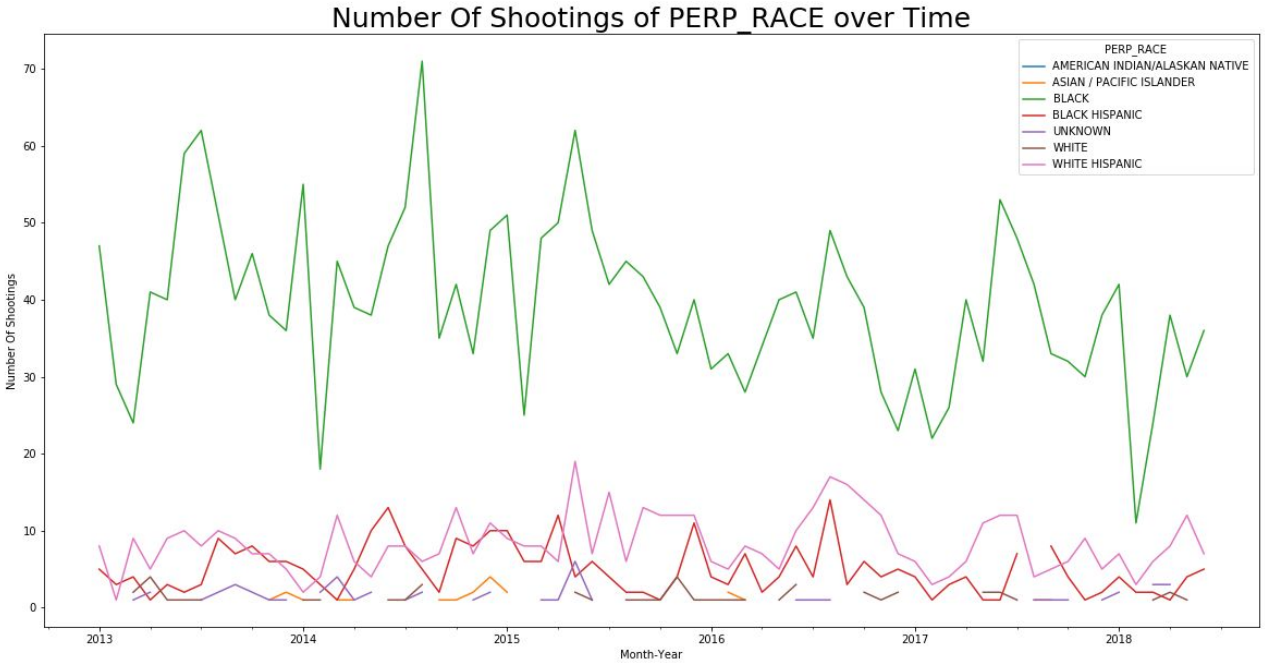
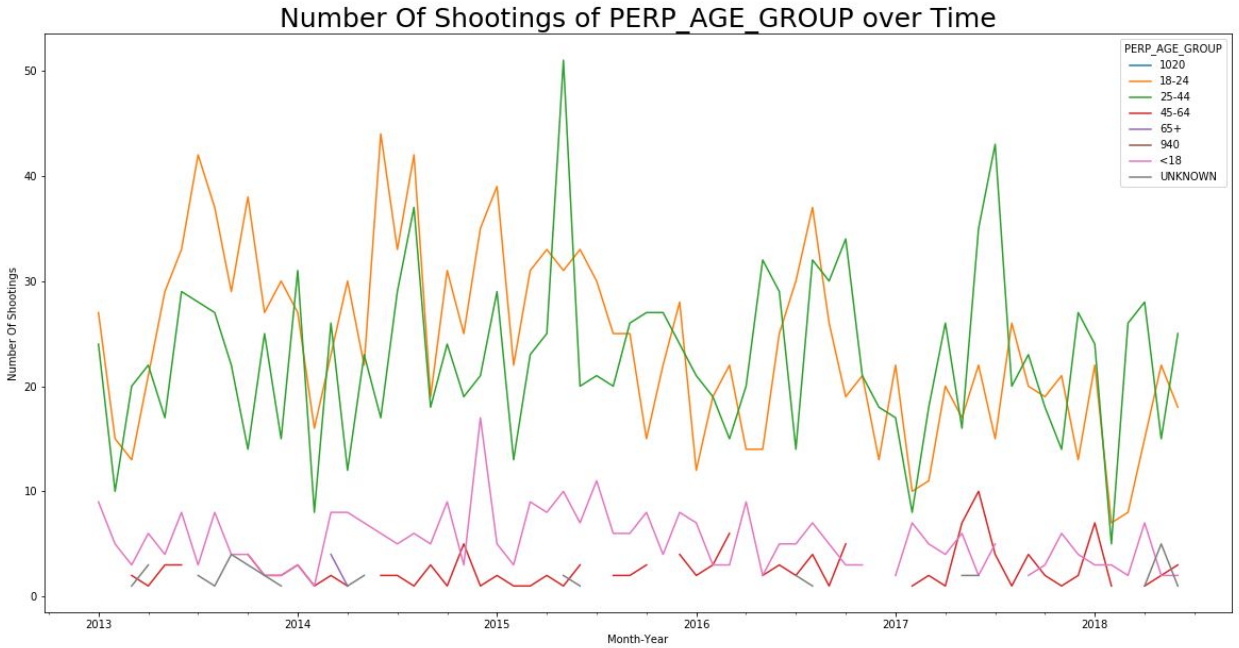
Learning Points

- Heavy bias toward minorities especially Blacks and Hispanics
- Data do not express the full extent of shootings threats in NYC and could skew the numbers and perception of the city's safety.
  - "Only valid shooting incidents resulting in an injured victim are included in this release"
- Potential analysis pitfalls that can heighten racial tensions or create excuses for neighborhoods gentrification

NYC Open Data Disclaimer

By accessing datasets and feeds available through NYC Open Data, the user agrees to all of the Terms of Use of NYC.gov as well as the Privacy Policy for NYC.gov. The user also agrees to any additional terms of use defined by the agencies, bureaus, and offices providing data. Public data sets made available on NYC Open Data are provided for informational purposes. The City does not warranty the completeness, accuracy, content, or fitness for any particular purpose or use of any public data set made available on NYC Open Data, nor are any such warranties to be implied or inferred with respect to the public data sets furnished therein.

Time Trends - Race



Time Trends - Age Group