

# Police Settlements 2011 - 2019

Salvador R Sanchez Castro

University of San Diego

Master of Science, Applied Data Science

MADS 505: Applied Data Science for Business

October 24, 2022

## Data Source

The police settlements 2011 - 2019 tableau dashboard was developed utilizing data set produced by FiveThirtyEight at <https://github.com/fivethirtyeight/police-settlements/>. The data set was split into 29 CSV files for each city. The collection process from each city was independent causing slight variation in variables. In order to have a richer dashboard we assumed all city information to be comparable and a representation of the state in which the city resided.

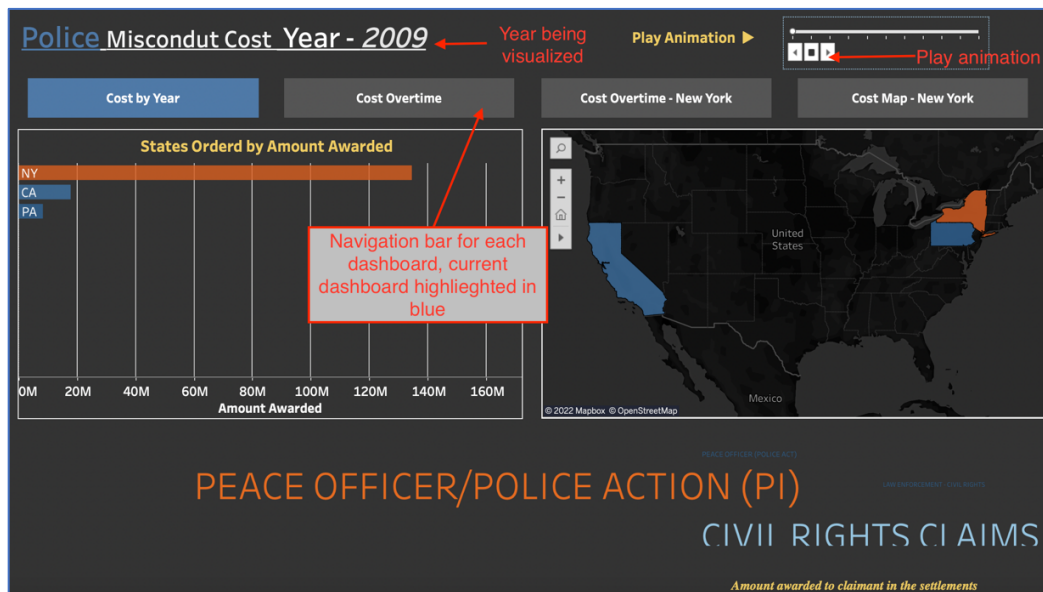
## Data Prep

In order to aggregate each CSV file a Python script was produced to loop through all the directories and subdirectories of dataset and concatenate into a single dataframe this was a two-process loop first one to identify a unique list of variables and the second one to append each file. Tableau prep software was utilized to clean up the data merging police officer and year fields. After initial visualization and data exploration, New York City was identified as a top contender with respect to settlement amounts, the data set also contained the address for each incident. Utilizing Python and the *geopy.geocoders* library longitude and latitude were calculated for each incident in New York City.

## Design

For the color scheme blue and gold were selected being representative of the police. Four dashboards were developed (not including the cover) to illustrate the story, 1) the cost of police misconduct for each year, 2) cost overtime (where New York is highlighted), 3) deeper dive into cost overtime for New York and finishing with 4) cost map of New York to demonstrate the regions with highest police activity and changes over time.

## User Guide



## Dashboard Link

<https://public.tableau.com/app/profile/salvador.sanchez/viz/PoliceSettlements2011-2019/Cover>