**Repo Link:** [**jennyxu12/MA346-summer-final (github.com)**](https://github.com/jennyxu12/MA346-summer-final)

**Dashboard Link:** [**https://police-shooting.herokuapp.com/**](https://police-shooting.herokuapp.com/)

**Deepnote Link:** [**https://deepnote.com/@yajie-xu/Final-Project-v733C8atR-SabacCrB02Xg**](https://deepnote.com/@yajie-xu/Final-Project-v733C8atR-SabacCrB02Xg)

The goal of our project is to perform data analysis to investigate the factors that could potentially raise the possibility of police shooting by drawing maps and charts in plotly. From our data analysis, we found out the state and city with the highest number of police shootings, and at the same time, the influence of gender, age, armed, mental illness, and race towards the tendency of police shootings to occur from 2015 to 2020 in the US.

We obtained the US police shooting file from Kaggle. This dataset contains the victims biological information, including name, killed date, manner of death, armed, age, gender, race, city, state, signs of mental illness, threat level, flee, body camera, and arms category. Our analysis focuses on seven of those factors, including state, city, gender, age, armed, signs of mental illness, and race by order. So first of all, we drew a map of police shooting counts for all states in the US to compare the numbers by creating a dataframe indexed by unique states and counting the times they appeared in the dataset. By looking at the level of darkness, we found out that California had the highest number of police shootings among all states, followed by Texas and Florida. An interactive bar chart on the dashboard was also created for the users to choose the particular year they want to look at, from 2015 to 2020, for the number of police shootings in all states in the US. We found out that California is again the state with the highest number of police shootings for every year. Then by using another dataset that contains latitudes and longitudes of all cities in California, we plotted a density map specifically for California, so that we could see the area that encountered the police shooting most often, which is Los Angeles downtown, according to the color of that region being more reddish. Meanwhile, Oakland, San Jose, Bakersfield, and San Diego also had relatively high numbers of police shootings.

After figuring out the state and region with the highest number of police shootings in the US, we were interested in finding the safest state. We first checked the length of the dataframe to make sure all states had some numbers of police shootings. Then, by sorting the number of police shootings, we found the five states are Rhode Island, Vermont, Delaware, North Dakota, and New Hampshire. In order to make it easier to read, we merged the US state abbreviation table to the dataframe to attach the full name to each state. Then we would like to look into the biological factors after we evaluated the geographical ones. In terms of gender, male are 21 times more likely to be shot by police compared to females in the US. What’s more, the age group from 23 to 40 years old is dominant among all victims’ age groups. It is reasonable because male and youngsters are generally considered to be stronger. The police are more likely to choose to shoot them in order to protect themselves as well as avoid more damages from the youngsters.

Furthermore, we were curious about the situation of the victim when the shooting happened, so we looked at if the suspects are armed or not and their signs of mental illness. It shows that being armed with guns and not showing signs of mental illness are the two categories for each column that causes the highest number of police shootings. It makes sense that more shootings happened when the victims were armed, especially with guns because they are more dangerous and harmful. However, it might be counterintuitive that the police shot more people who were not showing signs of mental illness. One factor could be that mental ill people are only a very small portion of society. But the high counts still suggested that whether the suspect shows the sign of mental illness might not be the police’s main consideration in a situation where a police shooting might take place.

Besides, race is always a hot topic in the US, and some discriminatory shootings happened recently. A lot of voices are saying that the police tended to kill African Americans because of their skin colors. In order to further analyze the relationship between race and police shooting and see whether this statement is correct, we counted the numbers of police shootings for each race group. We figured that the majority is mostly white people. Then we found a flaw of only conducting this measure because white people indeed consist the largest percentage in the US population. So we calculated the percentage of the counts for each race category in each state by dividing its corresponding counts with the total number of people killed in its corresponding state, and then comparing it with the US population by race. We assumed that all states in the US have the same percentage of population by race to make the process easier. A sunburst pie chart is created, showing each state by percentage of people killed for different races as values and population percentage by race as colors. Theoretically, we should see the race of white, which is colored as dark red, being the largest area for each state on the outer sections because as the percentage of white population is the largest, there is supposed to be a greater percentage of white people being killed in each state. However, we can see the race of black, which is colored as blue, occasionally takes the largest area for some states, meaning that although they only consist a small portion of the population, they take a greater percentage in the population of people being killed by police, leading to a potential discriminatory police shooting. Therefore, according to the pie chart, Maryland, Illinois, Washington DC and Delaware are some states where the black take the greatest percentage of people shot. Also, we can occasionally see that other races, like Hispanic and Asian, are more likely to be treated unfairly in some states. Admittedly, the results might not be that precise since we assumed all states have the same population percentage by race when it is not in reality, but overall, it could still reflect the fact of discriminatory police shootings in the US to some degree.

In conclusion, through our data analysis, we found California is the state that has the highest number of police shootings in the US from 2015 to 2020 and the highest for any particular year during the five years. Across the state of CA, Los Angeles is the area that has the most number of shootings. In the US, Rhode island, Vermont, Delaware, North Dakota, and New Hampshire were the five states that had the lowest number of police shootings during the five years from 2015 to 2020. In addition, people who were male, in the age group of 23 to 40 years old, and armed with guns could potentially raise the possibility of police shooting since they had the highest number of police shootings. Mental illness does not seem to be a very important factor that could lead to the likeness of police shooting. More importantly, discriminatory police shooting against minorities, especially black, does exist in the US.