UNITED STATES POLICE SHOOTINGS

# **Business Overview**

## **Overview**

In the world, law enforcement officers are given the provision to hold a gun. Due to their nature of profession, it's seen to be a necessity to protect them from harm. However, recent trends show police officers using excessive force in situations where it was not necessary. In the United States, the number of police shootings and killings of people every year is far more compared to other countries.

The trend of fatal police shootings has been increasing since 2015. The rate of fatal police shootings in the United States shows a large difference based on ethnicity. In 2019 the number of police shootings was at 1099, from which 24% of them were Black Americans. Between 2015 and 2021, the rate of fatal shootings for the black Amaericans stood at 36 per million, while the for white Americans was 15 per million.

The police shooting is a leading cause of death for young men in the United states. Young black Americans under the age of 25 to 30 years carry the highest risk of dying in the hands of a police officer. This has sparked movements and protest against discrimination due to color. The most worldly known was the black live matters movement. This was enlightened by the killing of George Floyd. The movement has spearheaded demonstrations worldwide protesting police brutality and system racism among people of color.

## **Objectives**

Our main objective for this study will be to investigate the main factors that influence police shootings in the United States in order to understand police decision making while conducting suspect arrest.

### Specific objectives

1. To determine if police are biased towards certain individuals.
2. To determine shootings distribution across states and cities.
3. To determine if suspect behaviour influences police shooting.
4. To analyze the trends in police shootings during the period under investigation.

To achieve the objective of this study, the analysis will seek to answer the following questions:

1. What are the main factors influencing police shootings?
2. Are the police biased towards certain individuals?
3. How are police shootings distributed across different states and cities?
4. What were the trends in US police shootings?

## Impact

This research seeks to influence state governments in making police reforms by analyzing police shooting data to determine the main characteristics/ features influencing police shooting.

## Hypotheses

This analysis will focus on testing the following hypotheses.

1. Police Bias

To determine if the US police are biased in cases of police shooting, these hypotheses will seek to analyze whether police shooting occurred more when the suspect was of a certain race or the suspect was within a certain age group.

* **Null Hypothesis**: Police are biased against race.
* **Alternative Hypothesis**: Police are not biased against race.
* **Null Hypothesis**: Police are biased against people aged between 19-25.
* **Alternative Hypothesis**: Police are not biased against people aged between 19-25.

1. Location of shootings

Under this section we will seek to determine whether there are states and cities where police shootings happen more than others. To do this we will group states together based on their proximity to each other.

* **Null Hypothesis**: Most police shootings occur within similar states and cities
* **Alternative Hypothesis**:Most police shootings do not occur within similar states and cities.

1. Behavior of the victim

Here we will try to determine whether having a mental illness and fleeing prompts the police to shoot.

* **Null Hypothesis**: Police shooting happens when the suspect attempt to flee
* **Alternative Hypothesis**: Police shooting happens when the suspect does not attempt to flee.
* **Null Hypothesis**:There is no relationship between mental illness and manner of death..
* **Alternative Hypothesis**:There is a relationship between mental illness and manner of death
* **Hypothesis:** There is no relationship between race and manner of death
* **Alternative Hypothesis:** There is a relationship between race and manner

## **Data Mining Goals**

Our data mining goal for this project is to investigate the main factors that influence police shootings in the United States to help understand police decision making while conducting suspect arrest.

Potential questions for consideration include:

1. What is the correlation between shooting and race?
2. What is the correlation between shooting and age?
3. What is the correlation between shooting and state?
4. What is the correlation between shooting and fleeing?
5. What is the correlation between shooting and body cameras being on or off?

## **Success Criteria**

A successful analysis will provide insights on the factors that highly influence police shootings in the United States and understand police decision making while conducting suspect arrest.

## Assessing the situation

**Assumptions**

The data provided is accurate.

**Resource inventory**

1. Datasets :
2. Softwares: Github, Google collab, tableau

**Implementation plan**

| Phase | Time-Frame |
| --- | --- |
| Formulation of Research Question | 30 minutes |
| Business Understanding | 1 hour |
| Data Understanding | 1 hour |
| Data Preparation and Cleaning | 5 hour |
| Data Analysis | 1 hours |
| Summary and Conclusion | 30 mins |

# **Data Description**

For this analysis, the data available has 14 columns and 4,895 rows collected from 2015 to 2020. The information available in the dataset is described in the table below:

| **COLUMN** | **DESCRIPTION** |
| --- | --- |
| ID | Unique identifier of the incident. |
| DATE | Date of incident |
| RACE | Race of the victim |
| AGE | Age of victim |
| GENDER | Gender of the victim |
| MANNER OF DEATH | Was the victim shot or shot and tasered? |
| ARMED | If the victim was armed and what they were armed with. |
| CITY | City of the incident |
| STATE | State of the incident |
| SIGN OF MENTAL ILLNESS | Did the victim have any indication of mental illness? |
| THREAT LEVEL | What was the threat level of the victim |
| FLEE | Did the victim try to flee from the police? |
| BODY CAMERA | Was the police body camera on or off during the incident? |
| NAME | Name of the victim |

# Data Preparation

* **Loading and reading the datasets**

We first imported the required libraries to help in the analysis. After, we loaded our dataset and created a dataframe. We then previewed both head( first 5 rows) and tail( last 5 rows) to get an understanding of the information we shall be analysing.

* **Exploring the data**

Afterwards, we checked the shape of our dataset. Our data contains 15 columns and 4895 rows. We checked the for the information of each column, to understand the datatypes of each, as well as a statistical summary of the data.

* **Data cleaning**

1. **Uniformity**: We checked for the uniformity in naming the columns names and removed any white spaces in the our columns as well as renaming our column names to lowercase.
2. **Missing Values**: checked for any null value in our data. Fortunately for us we found that our data had no missing values.
3. **Duplicates**: We checked if there were any duplicates in our dataset. Fortunate enough, our data contains no duplicates
4. **Irrelevant Columns**: Column names are irrelevant for our analysis, hence they were dropped.
5. **Outliers**: We decided to check for outliers in our dataset. Our dataset had a lot of outliers. We decided not to drop them as dropping them would affect the accuracy of the data analysis leading to results being inconclusive and incorrect.

# Analysis

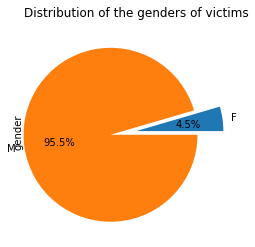
After cleaning our dataset, it was time to do analysis to answer our research objective. Detailed Analysis were conducted in google collab notebook and can be found in our github repository

[[https://github.com/MachukaJoy/US\_Police\_Shootings\_Analysis]](https://github.com/MachukaJoy/US_Police_Shootings_Analysis%5D) and tableau on <https://public.tableau.com/app/profile/asha.ahmed.deen/viz/USpoliceshootinganalysis/Story2>

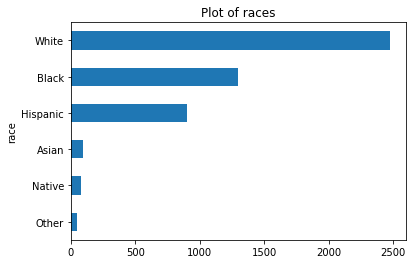
Our analysis was divided into:

## Univariate Analysis

Our first visualization was the gender comparison of our victims. We discovered that male were the most affected gender.



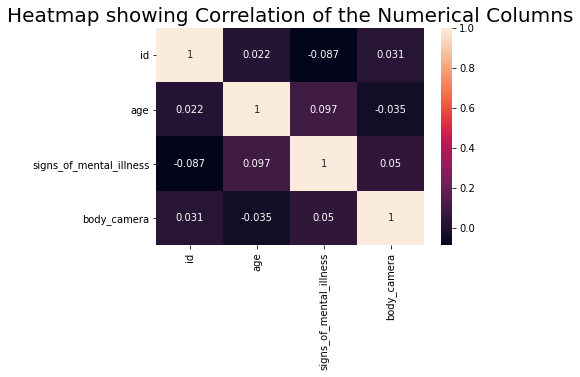
We plotted a visualization of the races in the United States. From our data,it was clearly observed that the number of white people was highest. The other races represented were blacks, hispanic, native and Asians.



After that we wanted to know how the victims died. It was clear that about 95% of the population were killed without being tasered. To understand why a great percentage were killed, we tried to understand the behaviour of our victim in that situation. From the data, most of our victims were armed with a gun. It was also clear that a great percentage of the victims attempted to attack. These could explain why the police officer used excessive force. However, it was surprising to note that most of the victims did not try to flee yet they were shot. This was strange, so we wanted to understand the mental status of the victims. What caught our attention was that the greatest proportion of the victims were of sober mind. Next we checked if the police body cams were on or off. Interestingly, more than 3000 cases occurred when the body cams were.

## Bivariate Analysis

After understanding each individual variable, we compared them against each other. We bagan by checking the correlation between each other. Most of the variables either had a weak positive correlation or a weak negative correlation.



With one of our claims being, police shootings are biased against race, we decided to gain insights on race. We compared the race against threat level and found out that the Whites posed a great threat to the police officer compared to other races. Across all races, a great number never tried to flee. Most of the whites who tried to flee were in cars while black people ran.

The analysis showed that most black and white people were killed in California State with Los Angeles and Oakland city being the top states across all races except native people, whose leading state was Arizona. The top city in death toll of blacks were los angeles and Oakland city leading in death toll of white.s angeles and San Bernardino city leading in death toll of white.

# HYPOTHESIS TESTING ANALYSIS

In this section of the documentation we intend to take you through a step by step story of how we came to the decision to either accept or reject the hypothesis. We had a total of 6 hypotheses that we broke down into different sections depending on the causing or the correlating factors.

The sections are

1. Police Bias
2. Location of the shooting
3. Behaviour of the victims

## POLICE BIAS

Under police bias we had 2 sets of hypotheses as listed in the hypothesis statements at the beginning of this document. [Go to the top](#_4ds0q8a6zovc)

1. ARE POLICE BIASED TO SHOOT PEOPLE OF MINORITY RACES?

The first hypothesis seeked to find out whether the police are more likely to shoot individuals who belong to minority races as compared to those who are white.

We went about our analysis by creating a subset to include all the incidents that white people were shot and a second subset that included all the instances that minorities were shot.

We then went ahead to compare the number of occurrences in each subset.

We will then perform our test by the standard procedure:

* **Specifying the Null Hypothesis**

People of minority races are shot by police more than white individuals

* **Specifying the Alternative Hypothesis**

People of minority races are shot by police less than the white individuals.

* **Setting the Significance Level (a)**

We set our significance level to 0.05 the standard level for most tests.

* **Calculating the Test Statistic and Corresponding P-Value**

We used a Levene’s test and got a p value of 0.14323490752467 and a test statistic of 1.46385010942280.

* **Drawing a Conclusion**

From our results, we see that our p value is above our set alpha. We therefore fail to reject the null hypothesis.

Some of the findings we got are that Police are equally likely to shoot a white person as they are to shoot a minority

However, due to our knowledge on the population distribution in the US, we were forced to sway our opinion on whether the police were biased. Some of the facts that made us sway our opinion is that

* The white population contributes to 76% of the total population in the US
* The sum of all minority races contributes to about 24% of the total population in the US
* If the police were not biased, for every 1 person belonging to the minority race shot, this would ideally count for 3 white people being shot. Which is not the case

This leads us to conclude that Police are more likely to shoot a person of colour as compared to a white person.

1. ARE POLICE BIASED TO SHOOT PEOPLE OF AGES 19 - 25?

After reading an article that stated that police are more likely to shoot people of ages between 19 and 25 as compared to people not within this age group, we set out to find out whether this statement is true.

We started by creating 2 subsets. The first group contained all instances where people who were shot were within the age 19 - 25. The second group contains people who are not within the age group 19 - 25.

We then did a sum of all occurrences in the different groups.

We then performed our test by the standard procedure:

* **Specifying the Null Hypothesis**

More people between the ages of 19 - 25 are killed than other ages

* **Specifying the Alternative Hypothesis**

More people not within the 19 - 25 age groups are killed than those within the age group

* **Setting the Significance Level (a)**

We set our significance level to 0.05 the standard level for most tests.

* **Calculating the Test Statistic and Corresponding P-Value**

We used a Levene’s test and got a p value of 0.0000 and a test statistic of 782.9762.

* **Drawing a Conclusion**

From our results, we see that our p value is below our set alpha. We therefore reject the null hypothesis.

Some of the facts we found are

* The majority of the people who are shot are not within the 19 - 25 age group.
* We also got to find out the youngest person who was shot was just 6 years old.
* The oldest person who was shot was 83 years old.

After conducting a wilcoxon t test on the data, we came to a conclusion that police are not biased to shoot people within the age group 19 to 25 as compared to other age groups.

## LOCATION OF THE SHOOTING

1. DO SOME LOCATIONS HAVE SIGNIFICANTLY HIGHER POLICE SHOOTINGS?

Under this section we will seek to determine whether there are states and cities where police shootings are more rampant than others.

* **Specifying the Null Hypothesis**

Most police shootings do not occur within nearby cities states and cities

* **Specifying the Alternative Hypothesis**

Most police shootings occur within nearby cities states and cities

* **Setting the Significance Level (a)**

We set our significance level to 0.05 the standard level for most tests.

* **Calculating the Test Statistic and Corresponding P-Value**

We used a Levene’s test and got a p value of 0.3712 and a test statistic of 0.8161.

* **Drawing a Conclusion**

From our results, we see that our p value is above our set alpha. We therefore fail to reject the null hypothesis.

We plotted a map on tableau to better visualize the spread of police shootings according to state. The map can be viewed [here](https://public.tableau.com/app/profile/asha.ahmed.deen/viz/USpoliceshootinganalysis/Story2).

Some of the interesting findings were:

* Most of the shootings occurred in the West and East Coasts of the country.
* On the East Coast, Police were more rampant in the SouthEast Region
* The Central region of the country had barely any police shooting occurrence.
* The State with most Police shootings was Carlifornia at 701 occurrences
* Followed by Texas with 426 incidents (almost half)
* California has more police shootings than any other state, but it also has more people. Even when accounting for population, California’s rate of police shootings is higher than the national average.

We arrived at a conclusion that Police shootings are more prominent within nearby states.

## BEHAVIOUR OF THE VICTIM

1. DOES A VICTIM ATTEMPTING TO FLEE RESULT IN THEM BEING SHOT?

In this section of the documentation we seek to find out whether there is a cause - effect relationship between victims attempting to flee and them being shot.

We begin by dividing our subsets of 2 groups. Group 1 will contain all the instances of victims being shot while they are not fleeing. Group 2 will contain all the instances where victims attempted to flee and they got arrested. We then go ahead and do our test.

* **Specifying the Null Hypothesis**

Police shooting happens when the suspect attempt to flee

* **Specifying the Alternative Hypothesis**

Police shooting happens when the suspect does not attempt to flee.

* **Setting the Significance Level (a)**

We set our significance level to 0.05 the standard level for most tests.

* **Calculating the Test Statistic and Corresponding P-Value**

We used Levene’s test and got a p value of 0.179712494 and a test statistic of -1.34164078649987.

* **Drawing a Conclusion**

From our results, we see that our p value is above our set alpha. We therefore fail to reject the null hypothesis.

Our results rejected the null hypothesis which was highly expected.This shows that there is a likelihood that in fact not fleeing increased your chances of being shot. From our plots initially it was showing most victims were those that did not flee. This could be recommended as an interesting factor to look at by the police.

1. IS THERE A RELATION BETWEEN MENTALLY ILL PATIENTS AND BEING SHOT

We label encoded the data and then did sampling using stratified sampling grouping by state as the strata.We then did a contingency table and then used chi square test to do our test.

* **Specifying the Null Hypothesis**

There is no relationship between the manner of death and mental illness.

* **Specifying the Alternative Hypothesis**

There is a relationship between the manner of death and mental illness.

* **Setting the Significance Level (a)**

We set our significance level to 0.05 the standard level for most tests.

* **Calculating the Test Statistic and Corresponding P-Value**

We did a Chi square test and got a p value of 0.03923581 and a test statistic of 4.2506284431704735

* **Drawing a Conclusion**

From our results, we see that our p value is below our set alpha. We therefore reject the null hypothesis. The chi square test shows that signs of mental illness have no influence on the manner of death of individuals.

1. IS THERE A RELATION BETWEEN RACE AND MANNER OF DEATH

We are going to first group the races by whites and other minority races which will then be sampled by use of proportionate stratified sampling in comparison to the US population. We decided to use this sampling method so that each strata will be a proportionate representative of the US population. We then did a contingency table and then used the chi square test to do our test.

* **Specifying the Null Hypothesis**

There is no relationship between race and manner of death

* **Specifying the Alternative Hypothesis**

There is a relationship between race and manner of death

* **Setting the Significance Level (a)**

We set our significance level to 0.05 the standard level for most tests.

* **Calculating the Test Statistic and Corresponding P-Value**

We did a Chi square test and got a p value of 0.96911597 and a test statistic of 0.0014990109805356718

* **Drawing a Conclusion**

From our results, we see that our p value is above our set alpha. We therefore fail to reject the null hypothesis. There is no relationship between race and manner of death.

# Conclusion

We had a number of conclusions from our hypotheses. Majority of the victims died through shooting and also the majority of the victims who were armed were armed with guns.

Against one of our null hypotheses, the majority of the victims were not of the age group 19-25. However most people shot were aged 25. Another distinct finding is that most males were shot and were white which was against our expectation. However the population of the United States has more whites at around 76% compared to the minority races. Most of the shootings occurred in the West and East Coasts of the country. On the East Coast, Police were more rampant in the SouthEast Region. The Central region of the country had barely any police shooting occurrence. The State with most Police shootings was Carlifornia at 701 occurrences, followed by Texas with 426 incidents (almost half). California has more police shootings than any other state, but it also has more people. Even when accounting for population, California’s rate of police shootings is higher than the national average.

Majority of victims didn't show signs of mental illnesses but were still shot. Majority of the victims threatened the police by attacking. Majority of the victims did not attempt to flee but were also still shot.

Most of the time, police did not have the body cameras on. This seems to raise a further investigative aspect.

**The strongest correlations are:**

Type Arm category and Armed

Arm category and age

Arm category and an attempt to flee

Arm category and Manner of death

Arms category and threat level

Race and State location

Armed and manner of death

Armed and mental illness

Armed and threat level

Mental illness and arms category

Mental illness and whether suspect was armed

Mental illness and suspect attempting to flee

Gender and their attempt to flee

Body camera and state

**The weakest correlations are:**

Arms category and state

Arms category and body camera

Arms category and race

Race and an attempt to flee

Race and gender of victims

Race and manner of death of victims

Manner of death and Gender of victim

Manner of death and Police body camera being on

Whether victim was armed and state

Whether victim was armed and gender

Victim's mental health and manner of death

Victim's mental health and gender

Victim's mental health and Police body camera being on

Gender and Manner of death

Gender and Police body camera being on

Gender and state where shooting occured

Body camera being on and level of threat

Attempt to flee and body camera being on

Attempt to flee and victim's attempt to flee

Level of threat and the victim's gender

# Recommendations

We recommend that there should be some strict rules to be followed. The state government should work hand in hand and create some police reforms policy. A few of these would include:

* Body cameras should always be on despite the situation.
* Handle suspects who are not fleeing and are unarmed more cautiously.
* Be careful when handling potential mentally ill suspects.
* Be careful handling unarmed suspects.
* Maybe try other handling means for suspects who are carrying a minor weapon instead of shooting.

The citizens are also to be educated on how to behave when confronted with a police. For instance fleeing even when you are innocent might get you killed. When police work hand in hand with the state government and citizens, we are likely to have a reduced rate of these shootings.