**REPORT ON EDA CONDUCTED ON US SHOOTING**

From the exploratory data analysis performed by group 7 on the US police shooting, below were our observation.

Based on the race of the survey, we can infer that all the races in US experienced the killing, although some races recorded fewer killings.

The shooting consists 50.58% of whites and 49.42% of other races (Black, Hispanic, Asian, Native, Other), which indicate the that the number of killing among the other races are equivalent to the number of killing among the whites.

The whites recorded the highest number of killing but most of the incidence where not recorded. This means that the whites are not the focus of the killing so they are not interested in keeping their records.

The census conducted by the United state census Berea shows that, interestingly enough, the north-eastern-most part of the United States has the highest percentage of white people compared to the other white populations across the nation.

The states with the highest overall white populations are California (28,409,288), Texas (22,819,758), Florida (16,602,290), and New York (13,539,678). This is likely because these are the four most populous states in the U.S., and over three-quarters of the country's overall population is white. Despite their large white populations, these four states are considered to be among the most diverse in the US. The states with the lowest percentage of white people are Hawaii (25.5%), the District of Columbia (46.0%), and Maryland (58.5%). link <https://worldpopulationreview.com/states/states-by-race>

From the above census it is clear that the whites are the most dominate people living in the states that recorded lower number of killing and they are the least dominate at the states that recorded higher number of killing

From this we can infer that the whites are not the target of the killing

We also observed that the top five cities with the highest number of killings are the cities are found in the states where the other races live. This means that the shooting was aimed at all the other races but not the whites.

The number of whites that uses arms are twice the number of all the other races and the whites that were unarmed are almost the same as the others, which means that the whites are more likely to attack than all the non-whites.

**INSIGHT FROM VISUALIZATION**

We realized that the non-whites are the target of the killing, therefore there is racism

**PROCESSES**

This talks about the processes we went through to present our findings. It involves the step-by-step approach to have finally arrived at our current results. Below are the stages;

Firstly, the pandas and seaborn will help us in the data visualization of the work. The panda method helped us in converting string date into python date time object. The NumPy will help us in the numerical aspect of the work. The next step will help examine the data by looking at the first 5 rows of the data. This is due to the fact that the rows are many, hence in order to get the exact number of rows to work with, we initiate this process.

The ‘In [9]’ helps us to know the number of rows and columns in our data. To find the type of data in dataframe, we initiate ‘In [15]’ which is ‘data.dtypes’. this is important because there are different range of data with their distinct characteristics and hence knowing the type of data helps in making accurate assumptions. To check whether there is any missing values, we initiated ‘In [18] which is ‘data.isnull() .sum()’. To check whether there is any duplication of values we initiate ‘In [20]’ which is ‘data.duplicated() .sum()’.

**Examining the data-basic analysis**

The ‘data.gender.values\_counts()’ helps to know the number of males and females who were killed. The ‘data.race.value\_counts()’ counts the number of people in a particular race.

Looking at the case before us, it is prudent to know the number of people who were armed with a particular weapon therefore, the ‘data.armed.value\_counts()’ helps to achieve this result.

The ‘b\_camera.head()’ shows the number of killings which were recorded. The ‘In [20]’ shows the number of arms in terms of category used.

All the ‘out […]’ gives us the outcomes of all the ‘Ins’ functions. For instance the ‘In [9]: data.shape’ which is initiated to know the number of rows and columns in our data gives ‘Out [9]’ gives (4895, 15) which means that there are 4895 rows and 15 columns in the dataset.