1. **Problem Statement:**

In this data analysis project, we explored gun shoot data with the objective of understanding the intention of gun shoot incidents and their relationship with the race of individuals in the United States. The dataset provided information on various attributes, including the year, month, intent, police involvement, sex, age, race, Hispanic origin, place, and education.

1. **Data Collection:**

The dataset used in this analysis was sourced from Kaggle, a well-known platform for open-source datasets. The dataset comprises gun shoot data collected over a span of 15 years, from 2006 A.D to 2020 A.D.

1. **Dataset Description**

The dataset used in this analysis spans from 2006 to 2020 A.D. It contains a total of 10 columns, namely: year, month, intent, police, sex, age, race, Hispanic, place, and education. These columns capture various attributes related to gun shoot incidents, providing valuable information for analysis. In total, the dataset consists of 529,233 rows, each representing a specific gun shoot incident.

1. **Perform EDA [Exploratory Data Analysis]**

The EDA (Exploratory Data Analysis) for this dataset was conducted using the Python programming language within the Jupyter Notebook environment. The analysis utilized popular open-source data manipulation and analysis library called Pandas as well as matplotlib a widely used Python-based visualization tool.

The key findings from the EDA are summarized as follows:

1. The majority of gun shoot incidents occurred in residential settings, specifically in people's homes.
2. Male individuals were found to be involved in a higher number of gun shoot incidents compared to females.
3. The primary intent behind the gun shoot incidents was identified as suicide.
4. Among the different racial categories, individuals belonging to the White race accounted for the highest number of gun shoot incidents, surpassing other racial groups.

The further analysis is shown in python file.

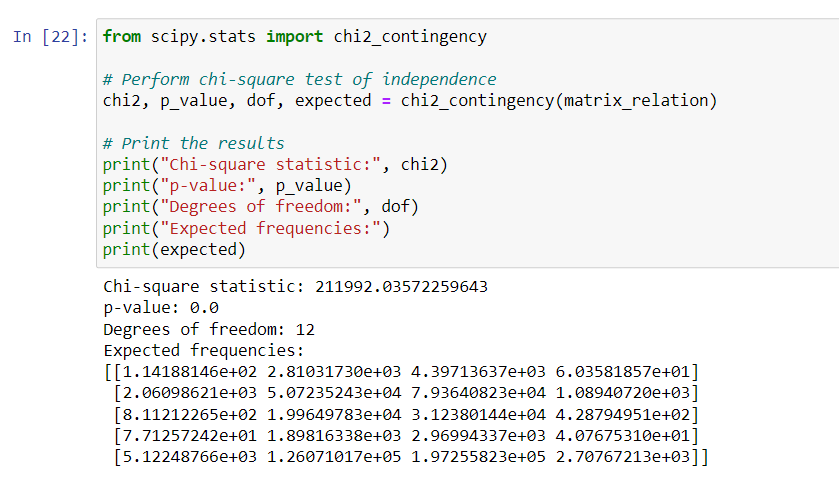
1. **Hypothesis Formulation**

Null hypothesis (H0): Race and intent of gun shoot incidents are independent.

Alternative hypothesis (H1): Race and intent of gun shoot incidents are associated.

1. **Hypothesis Checking**

The hypothesis checking have been performed using chi-square test:



The analysis revealed a significant statistical association between the variables 'race' and 'intent' in the gun shoot incident dataset. The chi-square statistic computed was 211,992.04 which indicates a substantial deviation from the expected values. With 12 degrees of freedom, the p-value obtained was 0.0 which provides compelling evidence to reject the null hypothesis of independence. This implies that there is a meaningful relationship between race and intent in gun shoot incidents indicates that these variables are not independent of each other.