**Motivation and Introduction**

Some people believe that the rate of gun violence in the United States has dropped, however over the past half-century, both the frequency of mass shootings that were planned in advance and the severity of each shooting have increased. For instance, the notorious massacre that took place at the Columbine high school in 1999 was a well-organized assault that resulted in the deaths of 13 people and injured 25 more. Stephen Paddock, the suspect in the mass shooting that took place in Las Vegas in 2017, opened fire at a concert on October 1 of that year. There were around 58 people who passed away, in addition to roughly 500 others who were injured. The shooting rampage that took place in Las Vegas is currently considered to be the deadliest incident in recent memory. Therefore, given that there is a trend toward a larger size for each mass shooting as well as other statistics that indicate the scale of everyday gun violence, the assertion that gun violence is decreasing looks doubtful. This is because there is an upward trend in magnitude for each mass shooting. As a result, my business partner and I are interested in conducting additional research into the problem of gun violence in the United States in the hopes of unearthing some potential facts.

The importance of conducting research on gun violence lies in the fact that it has the potential to enhance our existing understanding of the subject of gun violence in the United States. Aside from that, it can assist in the determination of a variety of other enlightening information, such as the place with the highest rate of gun violence, the collateral damage that is created by gun violence, and a great deal more. As of right now, our working hypothesis is that the rate of gun violence is higher than it has ever been; hence, our primary objective is to conduct an in-depth investigation into this concept. During the course of our exploratory data research, we intend to gain a more comprehensive understanding of the factors that contribute to gun violence in the United States and arrive at a number of shocking findings regarding the current status of gun violence.

**Data Source**

**About Dataset**

**Context**

There's currently a lack of large and easily-accessible amounts of detailed data on gun violence.

**Content**

This project aims to change that; we make a record of more than 260k gun violence incidents, with detailed information about each incident, available in CSV format. We hope that this will make it easier for data scientists and statisticians to study gun violence and make informed predictions about future trends.

The CSV file contains data for all recorded gun violence incidents in the US between January 2013 and March 2018, inclusive.

**Inspiration**

I believe there are plenty of ways this dataset can be put to good use. If you have an interesting idea or feel like messing around with the data, then go for it.

I was originally inspired to compile it in the wake of the Parkland shooting and the mass media coverage that followed. Reports like this and this showed that Nikolas Cruz had exhibited plenty of warning signs on social media before the shooting; what if we could build a machine learning system that preemptively detected such signs?

**Data profile**

**Where did you get the data?**

The data was downloaded from gunviolencearchive.org. From the organization's description:

Gun Violence Archive (GVA) is a non-profit corporation formed in 2013 to provide free online public access to accurate information about gun-related violence in the United States. GVA will collect and check for accuracy, comprehensive information about gun-related violence in the U.S. and then post and disseminate it online.

**How did you get the data?**

Because GVA limits the number of incidents that are returned from a single query, and because the website's "Export to CSV" functionality was missing crucial fields, it was necessary to obtain this dataset using web scraping techniques.

**Stage 1:** For each date between 1/1/2013 and 3/31/2018, a Python script queried all incidents that happened at that particular date, then scraped the data and wrote it to a CSV file. Each month got its own CSV file, with the exception of 2013, since not many incidents were recorded from then.

**Stage 2:** Each entry was augmented with additional data not directly viewable from the query results page, such as participant information, geolocation data, etc.

**Stage 3:** The entries were sorted in order of increasing date, then merged into a single CSV file.

**Glimpse of Data**

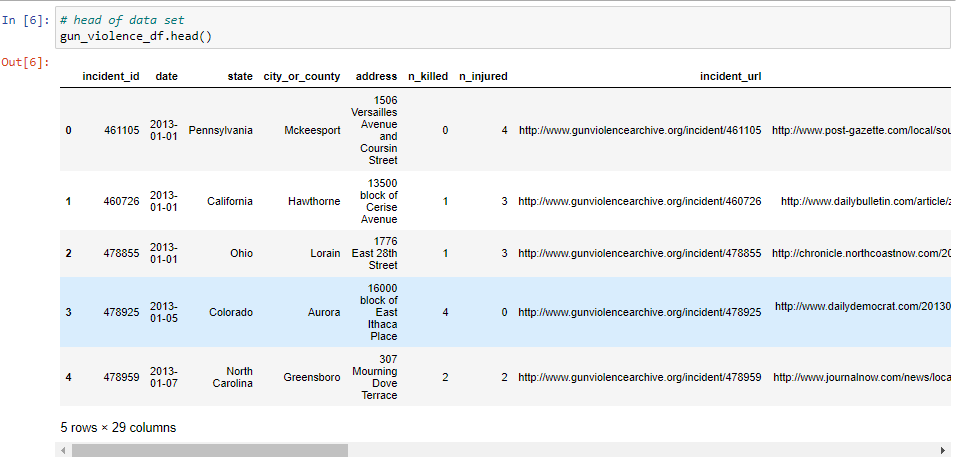


Figure 1 Head of data set

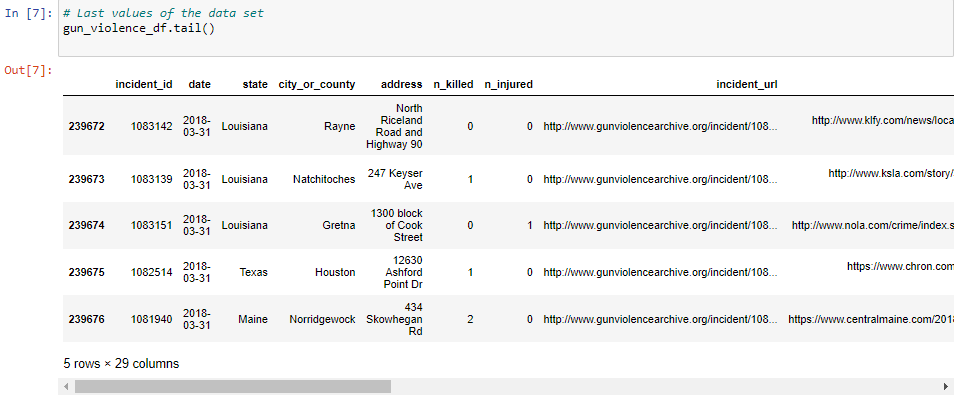


Figure 2 Tail of the data set

**Statistical Overview of the Data**

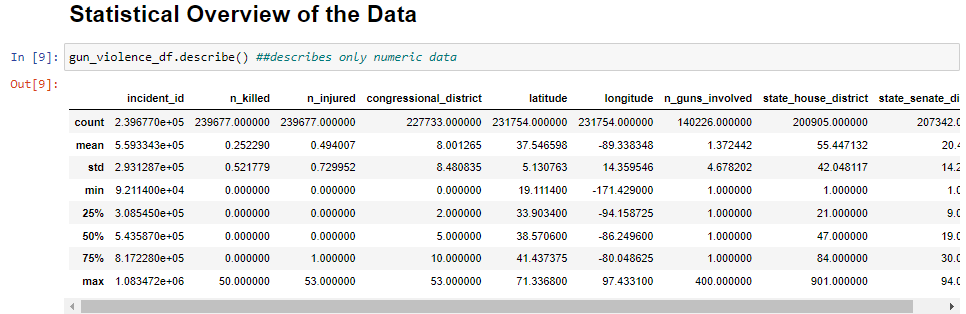


Figure 3 Describe only numerical data

The information regarding the numerical columns of the statistics on gun violence is described in the table that was presented earlier. Because the information is only provided for the numeric columns, and there is no information provided about the data that is missing, we have developed a more in-depth tool that will describe the information for all of the attributes below.

**Check for Missing Data**

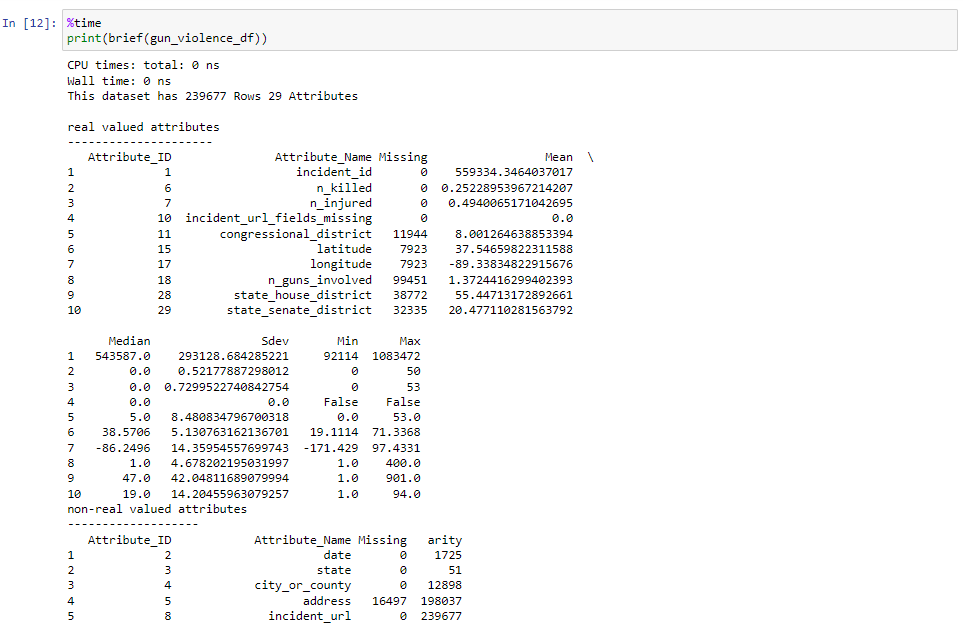


Figure 4 Missing data

Based on the analysis presented above, you can deduce that certain properties, such as participant\_name and participant\_relationship, are missing almost as many values as the total number of records contained in the dataset.

**Data Cleaning**

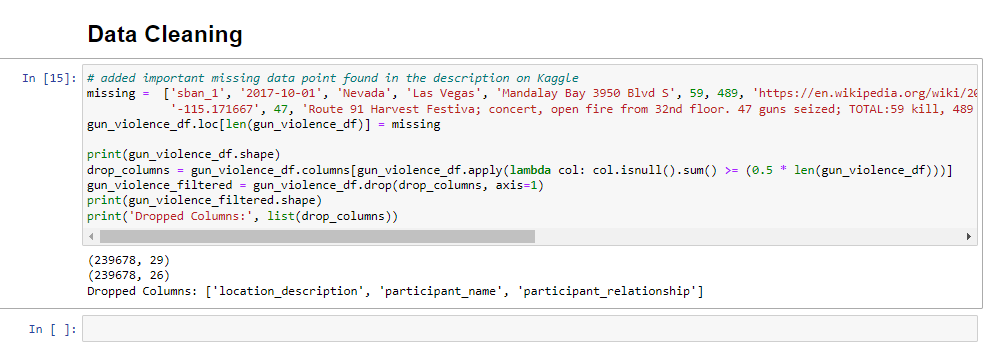


Figure 5 Data Cleaning

**Define Exploratory Questions**

1.What are some of the various trends that have been related with gun violence throughout the course of time?

2. Is it possible for us to investigate whether or not there is a geographical correlation between gun violence in the United States?

3. What differentiates the various manifestations of gun violence from one another?

4. Which demographic segment has a greater propensity to engage in acts of gun violence?