**Data Set Title**

**Exploratory Analysis**

Jameson Buie [jbuie@bellarmine.edu](mailto:jbuie@bellarmine.edu)

Name, [student@bellarmine.edu](mailto:student@bellarmine.edu)

1. **INTRODUCTION**

This data set contains data from nearly 400 shootings that occurred at schools from the primary to secondary level. These shootings occurred before, during, or right after class times. There are many other instances that were not included in this data set because they did not occur while school was in session. Additionally, this data set does not include shootings that happened on college campuses because shootings on large public campuses happen more often and do not necessarily involve students as shooters or victims, which is what we are exploring. We chose this data set because we want to explore the relationships between the shooters, the victims, time, and motives that were involved in these tragedies. Also, school shootings seem to be occurring more often so we will check to see if there is an uptick in number since 1999 and if there is a reason for this.

1. **DATA SET DESCRIPTION**

This data set contains 387 samples with 50 columns with various data types. There are 22 columns with string values and 18 columns with integer values. The rest of the columns have different data types.

**Table 1: Data Types and Missing Data**

|  |  |  |
| --- | --- | --- |
| *Variable Name* | *Data Type* | *Missing Data (%)* |
| uid | Ordinal/Int64 | 0% |
| nces\_school\_id | Nominal/object | 1.55% |
| school\_name | Nominal/object | 0% |
| nces\_district\_id | Nominal/float64 | 6.2% |
| district\_name | Nominal/object | 6.46% |
| date | Interval/object | 0% |
| school\_year | nominal/object | 0% |
| year | Interval/int64 | 0% |
| time | Nominal/object | 6.2% |
| day\_of\_week | Nominal/object | 0% |
| city | Nominal/object | 0% |
| state | Nominal/object | 0% |
| school\_type | Nominal/object | 0% |
| enrollment | Ratio/object | 0% |
| killed | Ratio/int64 | 0% |
| injured | Ratio/int64 | 0% |
| casualties | Ratio/int64 | 0% |
| shooting\_type | Nominal/object | 0% |
| age\_shooter1 | Ratio/float64 | 29.2% |
| gender\_shooter1 | Nominal/object | 22.7% |
| race\_ethnicity\_shooter1 | Nominal/object | 61.75% |
| shooter\_relationship1 | Nominal/object | 31.3% |
| shooter\_deceased1 | Ratio/float64 | 32.3% |
| deceased\_notes1 | Nominal/object | 90.1% |
| age\_shooter2 | Ratio/float64 | 96.9% |
| gender\_shooter2 | Nominal/object | 96.9% |
| race\_ethnicity\_shooter2 | Nominal/object | 98.2% |
| shooter\_relationship2 | Nominal/object | 98.4% |
| shooter\_deceased2 | Ratio/float64 | 98.4% |
| deceased\_notes2 | Nominal/object | 99.7% |
| white | Ratio/float64 | 2.6% |
| black | Ratio/float64 | 2.6% |
| hispanic | Ratio/object | 2.6% |
| asian | Ratio/float64 | 2.6% |
| american\_indian\_alaska\_native | Ratio/float64 | 2.8% |
| Hawaiian\_native\_pacific\_islander | Ratio/float64 | 32.3% |
| two\_or\_more | Ratio/float64 | 32.3% |
| resource\_officer | Nominal/int64 | 0% |
| weapon | Nominal/object | 35.7% |
| weapon\_source | Nominal/object | 72.9% |
| lat | Nominal/float64 | 0.25% |
| long | Nominal/float64 | 0.25% |
| staffing | Ratio/float64 | 6.2% |
| low\_grade | Nominal/object | 1.3% |
| high\_grade | Nominal/object | 1.3% |
| lunch | Nominal/object | 12.9% |
| county | Nominal/object | 0.5% |
| state\_fips | Nominal/int64 | 0% |
| county\_fips | Nominal/int64 | 0% |
| ulocale | Nominal/float64 | 0.77% |

1. **Data Set Summary Statistics**

Narrative introduction to the section.

**Table 2: Summary Statistics for School Shootings Data**

**A table with numbers and text

Description automatically generated**

A screenshot of a graph

Description automatically generatedA table with numbers and a few black text

Description automatically generated

There should be a table for **EACH** categorical variable.

Table 3: Proportions for XXX (n=yyy)

A screenshot of a computer

Description automatically generated

Table 4: Correlation Table/TablesA table of numbers with black text

Description automatically generated

A table of numbers with black numbers

Description automatically generatedA screenshot of a phone

Description automatically generatedA table of numbers and digits

Description automatically generatedA screenshot of a phone

Description automatically generated

A screen shot of a diagram

Description automatically generated

1. **DATA SET GRAPHICAL EXPLORATION**

Narrative introduction to the section. In each section below, indicate any interesting distributions, anomalies, imbalance, etc. that you notice.

* 1. *Distributions*
  2. *ScatterPlots / Pairwise Plots (continuous variables)*
  3. *Barcharts (categorical variables)*
  4. *Other Plots - don’t skimp – there are likely other plots that would be useful that I haven’t already specified. Include those in this section.*

A graph of blue rectangular bars

Description automatically generated

A graph of blue squares

Description automatically generated

A graph of a number of people

Description automatically generated with medium confidence

A graph of a graph with blue dots

Description automatically generated with medium confidence

A pie chart with different colored sections

Description automatically generated

A graph of blue bars

Description automatically generated with medium confidence

A graph of different colored lines

Description automatically generated with medium confidence

1. **SUMMARY OF FINDINGS**

Finish up with a paragraph or two of summarizing your findings about this data set.

Our school shootings data set consisted of 50 columns and 387 rows and the earliest recorded shooting in this set was in 1999 with the most recent happening this year in 2023. The columns included things like the school name, date it happened, the state, enrollment of the school, killed, injured, casualties, shooter age, gender, and race, along with many more categories.

As we dove into the data, we were able to find some interesting facts and correlations. Just starting with a basic but useful finding, there averaged one person injured and .5 people killed per shooting. As we moved into the correlations, most were fairly low so when analyzing the data we took that into account when looking at “stronger” correlations. This being said, there was a very weak correlation between the age of the shooter and if someone was killed or injured, with correlations of -.07 and .01. There was one of the stronger correlations of .2 when looking at if the shooter was deceased and if they killed anyone. There was also the strongest correlation within the races of the shooter with white people and casualties. Based on our regplot, we also found that there is almost no correlation between teh number of casualties and the enrollment of the school.

A couple other interesting findings based on some graphs that we had were that the most shootings have happened in recent years. 2022, 2021, 2018, 2023, 2019 are the top five years for the number of shootings, showing an upward trend. Most casualties happened on Tuesdays with the least amount on Thursdays, and around 120 of the shootings were random while around 70 were targeted. And finally we found that most shootings took place in California with 10%.