**ETL Project**

*Mass Shootings vs State Gun Laws*

Due Date: February 1st, 2020

Team Members: Michelle Fitzpatrick, Tim Cronin

**Synopsis:**

2020 is a big election year, and most of the general public can agree that there are a number of hot topics that will be heavily debated. Gun violence and corresponding gun control initiatives have come to the forefront of political chatter. Our aim in this ETL project is to shed some light on the data behind gun laws by state as they correlate to mass shootings over the last 12 months. In essence, logic would clearly say that the tighter the gun laws are, by state, the fewer shootings will occur. By extracting information from three distinct data sets (2019 mass shootings, gun ownership laws by state, and registered weapons by state) and massaging the data accordingly, we illustrate the clear and jarring correlation between states with loose gun laws and mass shootings over the last year.

**Extract:**

Data was loaded into Jupyter Notebook for analysis and transformation.

Data Source: Gun Violence Archive

<https://www.gunviolencearchive.org/reports>

Format: CSV

Data Source: World Population Review

<http://worldpopulationreview.com/states/gun-ownership-by-state/>

Format: CSV

Data Source: Statista: Crime & Law Enforcement

<https://www.statista.com/statistics/215655/number-of-registered-weapons-in-the-us-by-state/>

Format: CSV

**Transform:**

We created our dataframe from a combination of CSVs from three external sources. Using pandas in Jupyter, we loaded and read the CSVs, then we cleaned the data removing columns that were not relevant to our hypothesis.

When we merged the CSVs, there were multiple incidents in each state. To address this and make things more aesthetically appealing, we manipulated the data to read cleaner. We deleted columns such as ‘incident date’, ‘address’, and ‘city or county’. Using a groupby function we determined the total number of people killed or injured in each incident by summation. Using the groupby function to count the total number of incidents we determined how many mass shootings occurred per state in 2019. Using these metrics we created a new, cleaner dataframe with state as the new index. In this new data frame appears gun registration, purchase, and carrying laws (in booleans), the number of registered weapons, the total number of people killed and injured, and the total number of mass shootings per state. We also renamed most of the columns, and created a data frame with all the column names underscored specifically for when we import the data into SQL.

**Load:**

We loaded our final data into SQL. Given that our final output is a data frame, SQL was chosen for its speed and simplicity when dealing with varying relational datasets, and for extracting smaller CSVs from larger datasets provided. Further, given the cleaned up version of the data we discerned, SQL proves to be much more intuitive to extracting the particular information we were looking for.

**Limitations:**

Data compiled was only from the sample size of 2019

Because firearm registration is not required in every state, the column ‘number of registered weapons’ cannot account for the many unregistered weapons a state may have.

**Conclusion:**

Ultimately, as we anticipated, the looser the State’s gun laws, the more mass shootings occur. Keep in mind, however, that this is a concentrated look at just 2019’s mass shooting data and state gun regulations.

On a state-by-state basis, there’s a general correlation between stronger gun laws and lower rates of firearm deaths. One could even assert that strengthening state firearm policies would likely prevent firearm suicide and homicide, with other benefits in neighboring states as well. As evidenced from our analysis, of the 15 states that saw a double digit figure of mass shootings over the last 12 months, only two of them required fire arm registration. Further, of the top five states that accrued the most mass shooting fatalities, none of them required fire arm registration in advance of purchase.

To conclude, gun policies, though seemingly logical, are clearly not examining the data before their very own eyes. This ETL project illustrated the limitations in extracting the types of merged information for us to deliver a cogent final assertion, but demonstrated the utility in its execution.