Etl project

Group: archana rajagopalan, Felix Alvin, iappa, nicole nanton, chrison mills

Date: 3/11/2020

# Overview

## Project Topic: Gun Violence in the U.S. over the last 5 years (2014-2019)

|  |  |
| --- | --- |
|  | It is election year again and gun violence is always up for debate. Our team decided to analyze data to find whether gun control or gun education/gun safety initiatives were a more viable solution for gun violence in the U.S. |

## Transformation Scope

|  |  |
| --- | --- |
|  | We used Pandas functions in Jupyter notebook to load all three CSV files. Our team reviewed the files and transformed into data frames. The operator and address columns were removed due to missing information and were not relevant to the focus of this study. The team identified duplicates by doing an inner merge on the incident id column across all three data sets. Queries were created to address our questions by grouping the data by state and getting the sum of the number of people injured and killed. We sorted the data in descending order so we could visually see which state had the highest numbers. |

## Data Tools: Postgres SQL (PG admin), Pandas (Jupyter Notebook)

|  |  |
| --- | --- |
|  | After we pulled in the CSV files and loaded them into the data frames, we did an initial connection to the Postgres database using PG admin to store our original clean data sets. We used the quick database website to create the initial table schema that got loaded into the Postgres database that generated the first set of tables. After running the queries and created the new tables with only the relevant information, we reconnected to the database and generated additional tables for the data frames. |

Datasets Used:

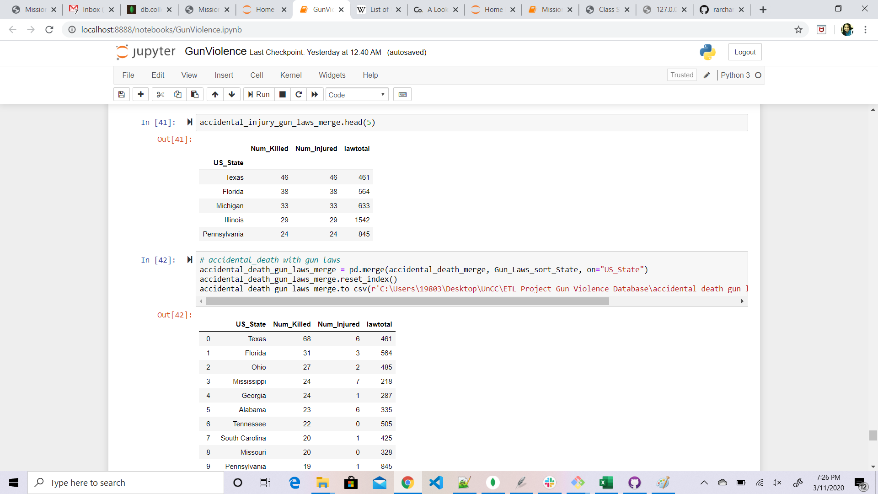
* https://www.gunviolencearchive.org/reports
* https://www.gunviolencearchive.org/mass-shooting
* https://www.gunviolencearchive.org/accidental-deaths

## Data Conclusion

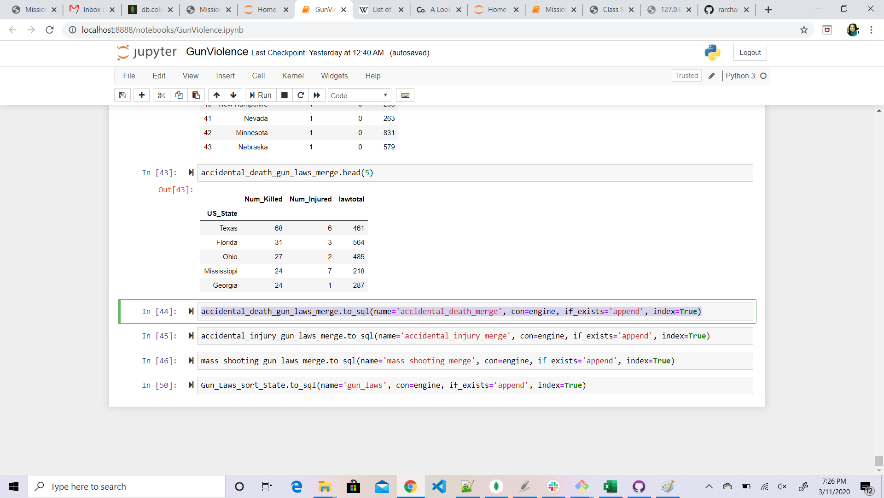
|  |  |
| --- | --- |
|  | There were some limitations to our findings due to the data available. However, we were able to address our hypothesis question in our initial project proposal below:  More gun control in the following states due to having the highest number killed by mass shootings.   * Texas * California * Missouri * Illinois * Ohio   We recommend more gun education/training in the following states due to having highest number of deaths from accidental shootings.   * Texas * Florida * Ohio * Mississippi * Georgia   The gun laws are comparatively high in the states where there are mass shootings, so the team felt there should be stricter laws and also procurement of the guns should be made stringent or gun sales overall should be banned. |

## Screenshots

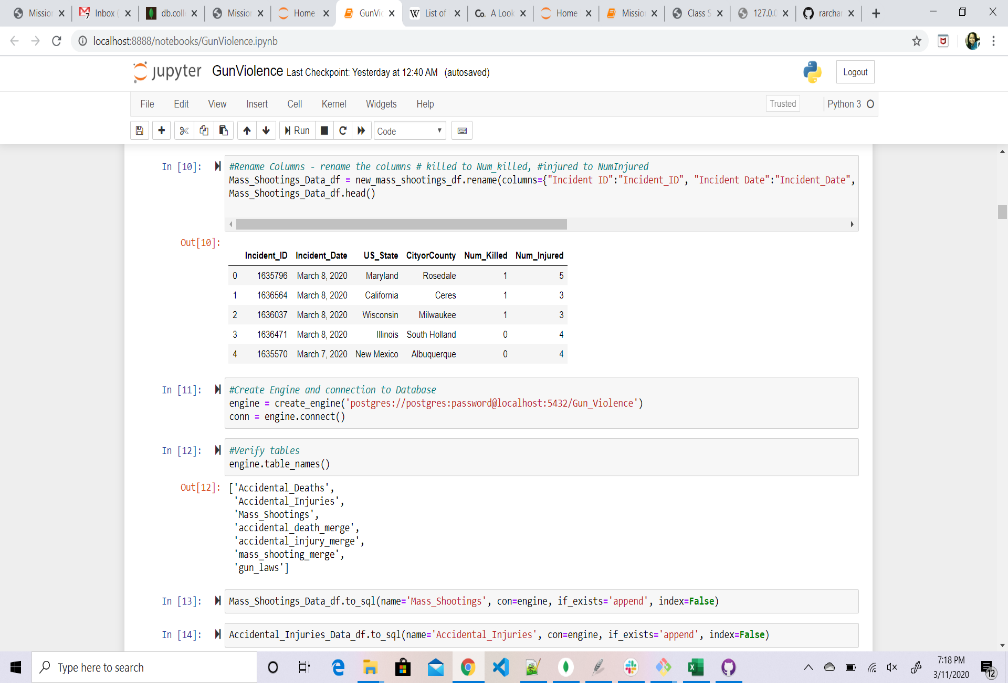
Accidental Death Screenshot



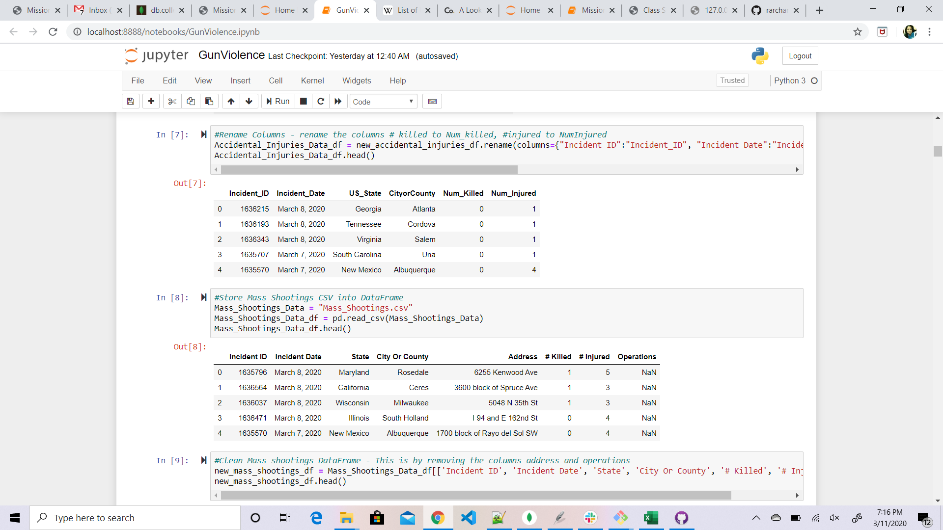
Accidental Injury Screenshot



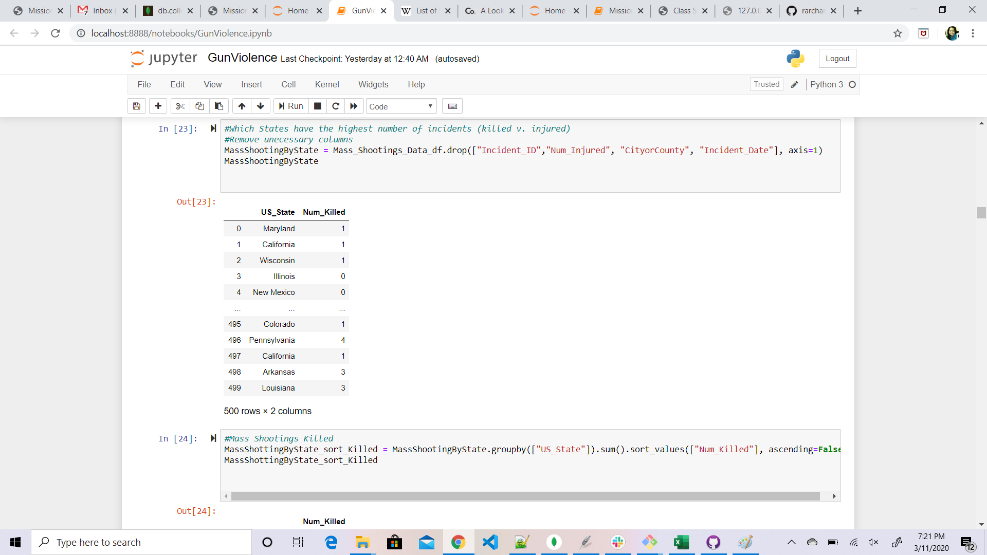
Data Frame Screenshot



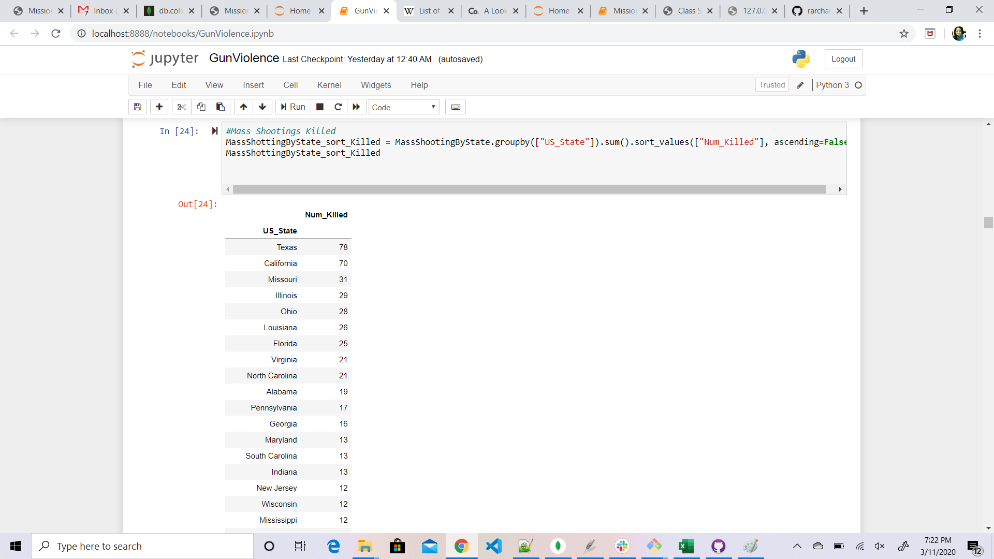
Mass Shooting Data Screenshot



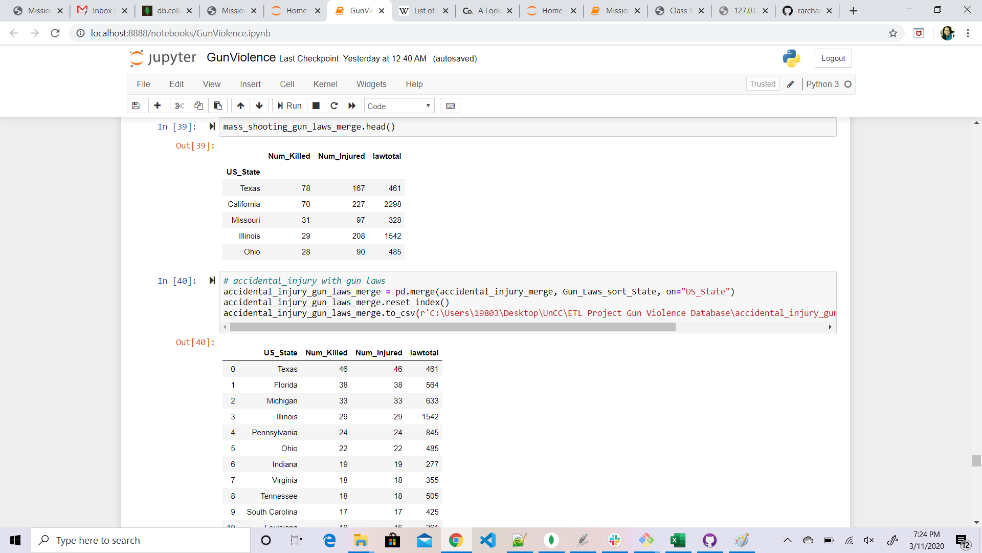
Mass Shooting by State/Injuries Screenshot



Mass Shooting by State/Killed Screenshot



Mass Shooting Gun Laws/# Killed and # Injured Screenshot



Query to get list of tables Screenshot

