Project title: Gun Violence in the USA

Group members: Ikran Askar and Salieu Barrie

-------Project Proposal-----------

Areas of focus

1. The team aims to get data from at least three sources.
2. Ensure a consistent format of data are accurate values
3. Narrow down the data to the year of interest
4. Load the data to a single database for further analysis

-----

Project Description/Outline

Gun control is a major issue in the United States. There is an enormous human and economic toll due to gun violence in all 50 states. There are frequent incidence of shootings – some can be attributed to be accidental and some and deliberate mass shootings. These shootings mostly results to death.

This project is aimed at extracting, transforming and loading the data to the database for further analyses. Our work will principally focus on ensuring we get a ready-to-use data for the team. The work will mainly focus on data cleaning, joining, filtering and aggregating.

-----

Datasets to be used

Source: <https://www.gunviolencearchive.org/mass-shooting>

<https://www.gunviolencearchive.org/accidental-deaths>

<https://www.kaggle.com/datasets/ecodan/school-shootings-us-1990present?select=pah_wikp_combo.csv>

1. The field of interest mass shooting data set includes incident date, state, city/county, number killed, and number injured.
2. The field of interest accidental data set includes incident date, state, city/county, number killed, and number injured.

Table

Description automatically generated

Table

Description automatically generated

Graphical user interface, table

Description automatically generated

Some initial findings with the data are:

1. The data are not readily available in uniform format
2. Different sites have different interest and data are skewed to their interest
3. The school shootings data may have been part of the mass shooting data. There might be duplications. (A limitation of our data)

Rough Breakdown of Tasks:

1. Preprocessing data
2. Extract
3. Transform
4. Load