**Project Dataset Options Team: AbracaData**

**1. Gun Violence in the United States**

<https://www.kaggle.com/jameslko/gun-violence-data>

The use of Guns for personal defense has been in debate for some time now with genuine frustration among people about their impact on society at large. Gun violence in the United States results in thousands of deaths and injuries. Under such, having gun-violence data publicly available will help to reach to the collective conscience of society.

Gun Violence Archive is a non-profit organization that provides online public access information about gun violence in the United States. The dataset has limitations in terms of the number of incidents that are returned from a single query and the website’s ‘Export to CSV’ option has some missing fields, hence we decided to consider the gun violence dataset on Kaggle. One of the Kaggle users obtained dataset using web scraping techniques from [gunviolencearchive.org](http://www.gunviolencearchive.org/). Since the data is publicly available in GVA and Kaggle, there are no restrictions on terms of use, or licensing up till due credit is given to GVA for using their data.

The chosen dataset contains all the information about gun violence that has happened in the United States from January 2013 to March 2018. It is intriguing to overview statistics where mortality from gunshots is as high as 100 people per day. Needless to say, the USA’s gun death rate is 11 times higher than that of other high-income countries. Under such, it is very important to understand the impact of gun violence in the US. We are particularly interested in figuring out what is causing gun violence. Nearly two-thirds of gun violence are suicides. Having access to the dataset may reveal an association between gun ownership and the violence caused.

The data can be used by researchers, government bodies, advocates, writers in the analysis of current regulations related to gun safety usage.

This dataset can help answers questions like

● What areas and states are most prevalent in gun violence in the United States?

● Are there any particular age group or gender are targeted in gun violence?

● What is the gender ratio of shooters?

● What guns are most frequently used in gun violence?

Additional Sources

Everytown for gun safety

<https://everytownresearch.org/gun-violence-america/>

Wikipedia

<https://en.wikipedia.org/wiki/Gun_violence>

**2. Emergency 911**

The data set is downloaded from Kaggle, and it is under Open Database Contents License (DbCL)

<https://www.kaggle.com/mchirico/montcoalert>

The dataset is about calls to emergency service. 911 being the “to go to point” for any emergency situation, it would be interesting to understand the nature of these emergencies in a temporal and geographical context. The chosen dataset contains information about the location of a phone call, the nature of emergency along with the timestamp. It would be interesting to understand the type of incidences served by emergency service, for instance, what percentage of reported cases included violence vs help call for a sick pet.

Potential users of this dataset could be people who wish to determine the crime rate in a region. The information can help people to de preemptive preparation before visiting the given region. For decision-makers, it can mean allocating appropriate resources to emergency service. For example, the people assigning dispatch can use this data to determine how many officers should be allocated to a given region.

This dataset can help answer questions like:

● What are the most common emergency situations reported to 911?

● Is there a trend in geographical locations for reporting emergencies to 911?

● Is there a trend in terms of timestamp for emergency calls to 911?

**3. Global Terrorism Dataset**

<https://www.kaggle.com/START-UMD/gtd#globalterrorismdb_0718dist.csv>

<https://www.start.umd.edu/pubs/START_GTD_Overview2017_July2018.pdf>

The Intellectual Property of this database is owned by the University of Maryland and is maintained by the researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START). START is headquartered at UMD and provides revocable, non-exclusive, non-transferable right and license to use this dataset for non-commercial purposes only.

(<https://www.kaggle.com/START-UMD/gtd>)

The dataset contains over 181k incidents of terrorist attacks from all over the world. Apart from the city and the country, the dataset also specifies the latitude and longitude of the attacks thereby giving an accurate location for data visualization. There is a 20-30 words long summary for most of the attacks that will help with a better assessment of results. Additionally, the dataset includes notes and proper citations with a focus on the socio-political repercussions of given attacks.

The potential users of this dataset would include news media, university researchers, governmental and non-governmental organizations working around public safety.

This dataset can help answer questions like:

● What is the intensity of attacks by region for a particular year?

● Is there a particular temporal or geographical trend in these attacks?

● What can be the motives behind attacks?

# 