Makerere University, College of Computing and Information Sciences


SEMESTER ONE 2024/2025 ACADEMIC YEAR

SCHOOL COMPUTING AND INFORMATICS TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

MASTER OF SCIENCE IN COMPUTER SCIENCE

MCS-7103 - MACHINE LEARNING

# EXPLORATORY DATA ANALYSIS ASSIGNMENT 1

ILAKUT EMMANUEL

2024/HD05/21924U

2400721924

## INTRODUCTION

Over the past few years, gun violence has risen to the forefront of public consciousness in the United States and around the world. Much of the debate has focused on gun regulation and keeping deadly weapons out of the hands of potential killers, particularly those with mental illnesses. Unfortunately, far less attention has been dedicated to the impact of gun violence on victims, its causes and trends. While individuals killed and injured in atrocities such as the Sandy Hook and Aurora Theater shootings are publicly remembered and mourned, victims of these tragedies are not limited to those men, women, and children killed, injured, or present during these horrific events. The consequences of gun violence are more pervasive and affect entire communities, families, and children.

In this study my objectives included:

identifying the most high risk areas in the United states, prevalence of gun violence among different age groups , relationships between victims and perpetrators of these crimes

## DATA COLLECTION

The data was downloaded from a  Gun Violence Archive's website. Which is a not for profit corporation formed in 2013 to provide free online public access to accurate information about gun-related violence in the United States. The project provides 260k gun violence incidents in csv formatThe file "datatoanalyse.xlsx" contains data on gun violence incidents in the United States. Each row represents a separate incident, including details like location, date, number of casualties, gun type, and participant information. This data can be used to analyse trends in gun violence, identify high-risk areas, and understand the characteristics of incidents and perpetrators.

## DATA WRANGLING

Described in this section are the data pre-processing approach that I used to clean and organise the

data. Python was chosen as the data analysis tool.

Important Python libraries: The first step in the data wrangling involved importing the python libraries

Pandas, Matplotlib, Seaborn. Pandas is an open-source python data analysis and manipulation tool,

Matplotlib is used to perform data visualisation in python and lastly seaborn is also a data

visualisation tool (based on Matplotlib) and is closed integrated with the Pandas data structures

making it easier to use, with beautiful statistical graphs.

Reading the Dataset: After extracting the raw customer support dataset, which was extracted in the

CSV file format, I performed discovery to identify the fields relevant to my objectives, I then decided to use 13,000 records to investigate incidents between January 2018 to March 2018 and converted the csv to excel format

I loaded the dataset into a Pandas Dataframe and exploratory analysis.
Assessing the Dataset: the purpose of this phase was to mainly dig into the dataset and understand
the structure, content and check if there are problems in the dataset. The detailed process involved
looking at the features attributes in the dataset for and identifying relationships and correlation between the different fields such as incident participants and participant status

EXPLORATORY DATA ANALYSIS
After performing the data assessment part, I was able to identify certain features that were imported
for my purpose.
Some of the important features I identified from the dataset were, gun violence crimes are almost always increasing in each state per month,

CONCLUSION
Below are my findings from the analysis of the customer support dataset
1. More deaths and injuries occur when a crime is done and a gun is involved

2. The trend is that crime is higher in states that are closer to the coast