**PROJECT 3: CRIME ANALYTICS**

Reflective Diary on US CRIME DATA SET by:

**Oladimeji Muiz Fadeyi -21045690**

Module:

**BUSINESS INTELLIGENCE AND DATA VISUALISATION**



**INTRODUCTION**

According to the International Association of Crime Analysts (IACA), crime analysis is a field of study that covers a range of techniques. Experts conducting crime analysis and the tools they employ strive to help police agencies function more effectively by providing better data. Analyst data can be useful in the following ways:

* Investigate and resolve crimes
* Develop effective methods and tactics to avoid future crimes.
* Find and catch criminals, the accused must be prosecuted and convicted, and find a way to improve personal safety and quality of life
* Internal processes need to be optimized, patrols and surveys should be a top priority, also identifying and resolving community issues
* Be prepared for future resource requirements, and establish a solid policy, and public education is essential. (Iaca.net, 2022)

Crime Analysis is one of the most important activities of intelligence and law enforcement agencies, where crime is on the rise around the world. This project analyzes data collected by the FBI and FOIA regarding crime in different parts of the United States from 1980 to 2014. This record consists of all crime-related data, the date the crime occurred, and its relationship to other events.

This record shows the state in which the crime occurred, the gender of the crime, the age of the crime, the race of the criminal, the race of the victim, the age of the victim, the relationship between the victim and the criminal, and the type of weapon used, the crime has been resolved. Most law enforcement agencies can face the challenge of efficiently and accurately analysing the increasing amount of crime.

The main purpose of this project is to find criminal patterns and trends in US criminal records. It also aims to provide useful information on how to minimise crime and stay safe.

**PROJECT SELECTION**

**February 16**

I and 3 other coursemates decided to form a group called GROUP 10, the group members are Oladimeji Fadeyi, Teslim Kazeem, Ifenyinwa Dauda, and Chiazokam Orji. We chose to work on the crime dataset because we found it to be more interesting than the remaining project topics. This puts us in a position to identify the pattern of criminal behaviour. Thereafter, we tasked ourselves with watching a video on Tableau by Edureka via Youtube. This Reflective Diary will visually acquire, filter, enhance, tune, and examine crime patterns in several states across the United States of America. It will also analyse the importance of crime analysis in the industry and its use by potential stakeholders. Finally, I aim to help law enforcement professionals improve the outcome of crime analysis.

**LITERATURE REVIEW**

We evaluated various earlier research efforts and papers that had been done on crime analysis. Every day around the globe data is generated in all forms and manner. This data creates digital trails, which are mostly harmless but can help the DOJ (Department of Justice) in a number of ways. Forensic experts make use of fingerprints and DNA that is maintained and stored in databases which is unique to each individual to identify criminals. Investigators can discover when accounting records were manipulated to hide a fraud plan by looking at the metadata of a document. The data also helps law enforcement agencies identify and respond to crime patterns. Traditional procedures to detect and prosecute white-collar crimes, on the other hand, may be thrown off by exponential data expansion.

By focusing on how to identify the concentration of a particular crime and predict crime patterns in a timely and accurate manner. To improve decision making and crime mapping. It contributes significantly to the literature on crime visualisation. Researchers have created a system model to use to identify and predict a range of crime patterns, this can help law enforcement agencies improve their crime-fighting tactics. (2010, Zhou et al.)

A study by a group of people in 2017 evaluated and investigated the strengths and shortcomings of predictive analytics in spatial criminology. They examined the existing crime data to see how well they could forecast the crime. We used spatial visualisation to investigate three types of crime: burglary, robbery, and assault. Environment variables, demographic variables, environment variables, indicator variables, and criminal history variables are all examples of predictors. From 2011 to 2014, statistics on burglary, robbery, and battery were collected every two weeks.. Researchers wanted to investigate whether there was a difference in predictive ability between day and night (Rummens, et al., 2017)

The Department of Justice is aware of this and is working together to revise and study better ways to screen large amounts of data (O'Connell & Gaunt, 2020).

Since at least the 1900s, crime maps have been used in crime analysis.. The NYPD uses them on a daily basis. The visualisation consisted of a standard crime map with pins showing the location of the incident and some of the criminal attributes. Presently, forensics and analysis has progressed from using pins to using computer tools like GIS technology, which has had a significant impact on crime analysis (National Institute of Justice, 2017)

**Team Meeting**

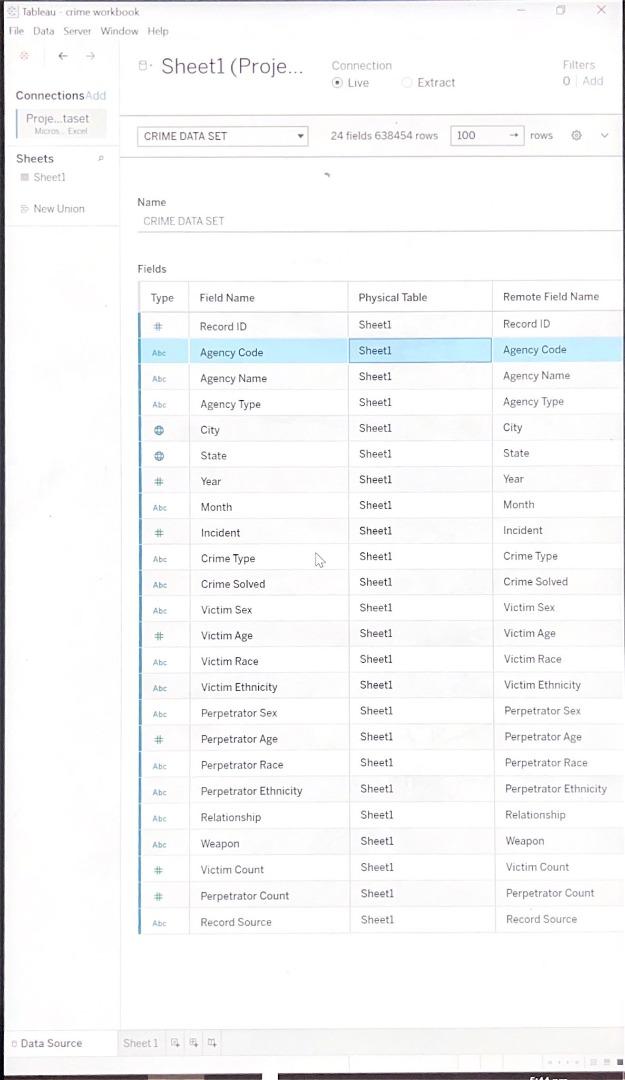
**February 18**

We had a meeting with my fellow group members (Group 10) on collaborate (accessed via blackboard), we discussed the dataset, deliverables, milestones, and examples of useful insights to show aside from the five listed in the assignment outline. In addition, we discussed the marking criteria and reference styles, as well as what makes a competent, outstanding, or outstanding piece of evidence.

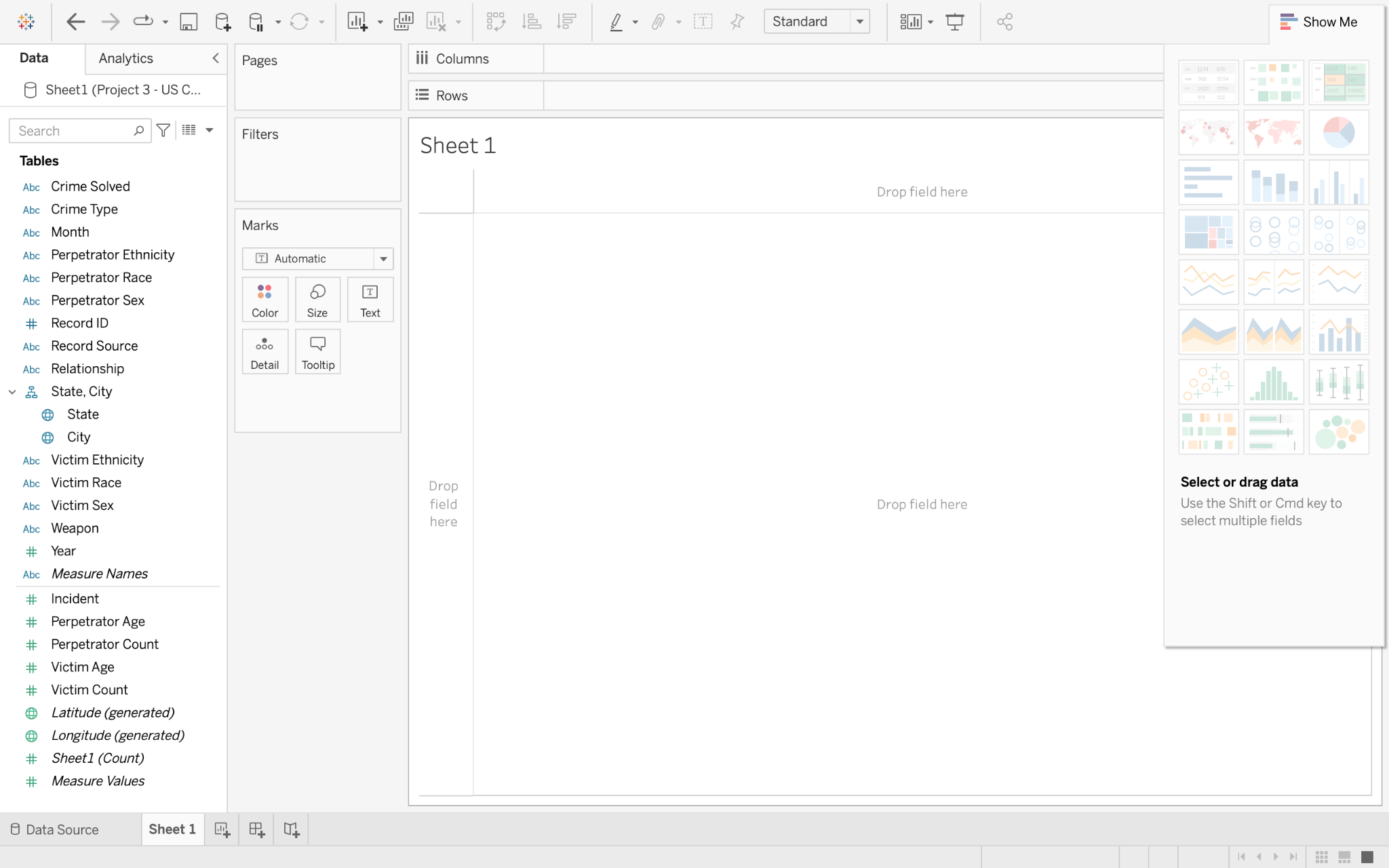
As seen in the images below, the US crime data set consists of 23 fields and 638454 rows.

Graphical user interface, application, table

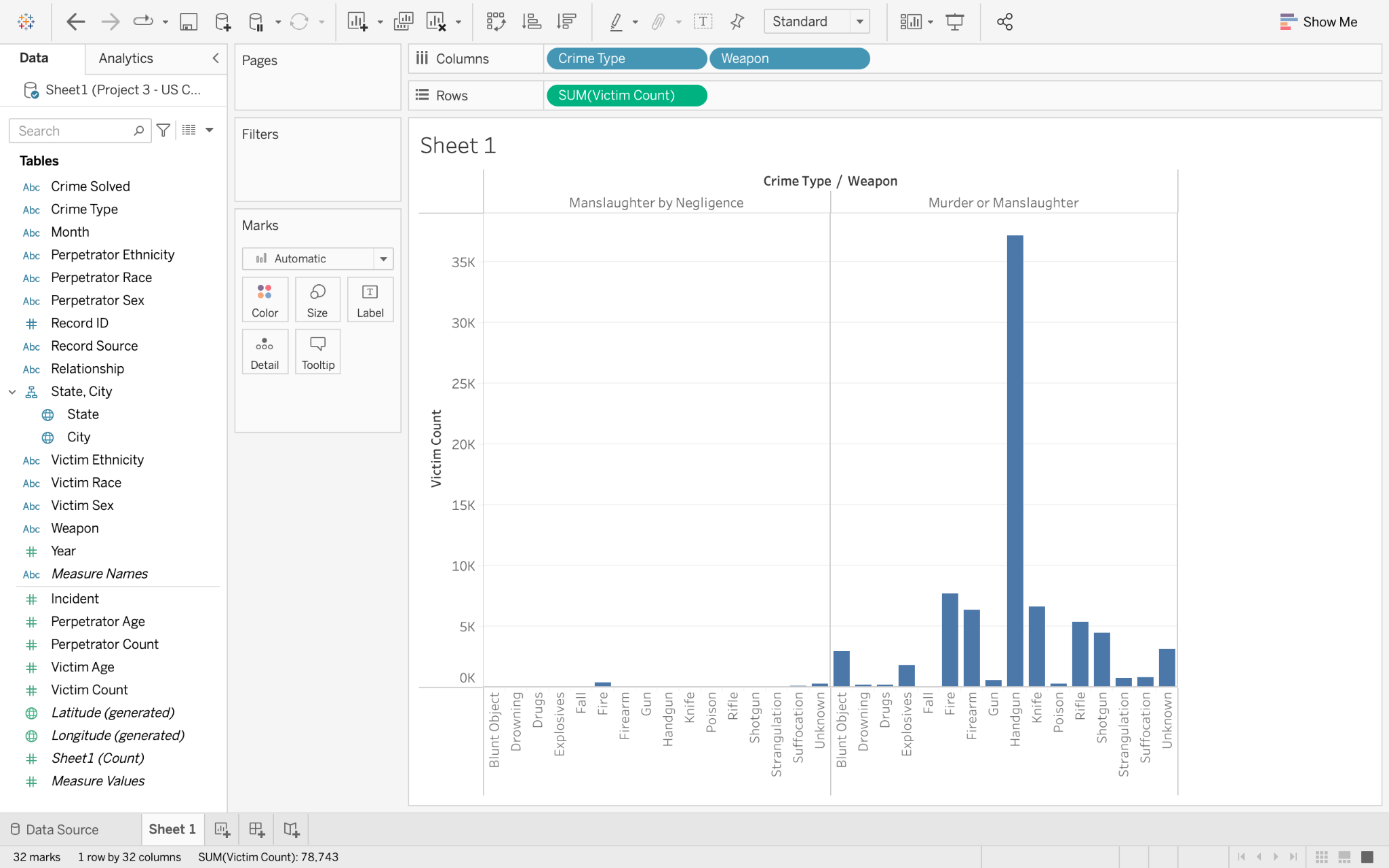
Description automatically generated



From the sheet we were able to first visualise our data from the generated fields i.e we could see the aggregation of all measures present in the dataset.



We proceeded to compare measure names and values e.g comparing crime type/weapon vs the victims. **(as shown in the image below)**



According to the dataset (as shown above), two types of crime are recorded:

* Manslaughter by Negligence
* Murder or Manslaughter

To gain a better understanding of these offences, I conducted some research during my self-study.

Manslaughter is that the act of killing another person during a way that's less culpable than murder. In contrast, a murder occurs when one human being unlawfully kills another human being, and homicide is when one human being causes the death of another. Some homicides are manslaughter, and a few are legal, such as when justified by an affirmative defence, such as insanity or self-defence.. Manslaughter by Negligence is the Killing of another person through gross negligence (the voluntary and conscious conduct (including a failure to act) by a person who, at the time of the conduct, knew that the conduct was likely to be harmful to the health or well-being of another person.) (Cornell Law School, 2022)

**Team Meetings**

**March 4 & March 18**

**MILESTONE 1:** Data exploration

We talked about the first milestone and tried to answer each question using information from a YouTube video (Tableau Training for Beginners | Edureka, 2019).

**1. Is it necessary to clean the dataset?**

Data cleaning is the process of eliminating or changing data that is inaccurate, incomplete, irrelevant, redundant, or incorrectly formatted in order to prepare it for analysis. (Sisense, 2022)

Yes, the United State crime dataset needs cleaning. We started a worksheet to generate a geographical visual representation of the data by loading the victim count against the location. We were familiarising ourselves with the dataset, but it came up with a blank graph and a 3k unknown value at the bottom right. This made us realise that we needed to clean up our dataset more to get rid of the unknowns. (See image below)

**Graphical user interface, application, website

Description automatically generated**

To fix any structural issues, we expected to have to write Python code or utilise Microsoft Excel, however we realised the error wasn't shared by other members. This indicated that the issue was not with the data format, but rather with the tableau workbook's settings.

We discovered that this was due to the fact that the dataset originated in the United States, and our system settings had attributed the dataset's location to our respective home countries. As a result, tableau did not recognize the locations of the data. The majority of the mistakes were resolved when we moved the location of our tableau from Nigeria to the United States.

A screenshot of a computer

Description automatically generated

This process **(as seen above)** cleared the bulk of the unknown states, leaving 1 unknown state. To do this, Manually we changed the spelling of the state "Rhodes Island" to "Rhode Island" to remove the last unknown state.

**A screenshot of a computer

Description automatically generated with medium confidence**

To remove errors from the city, we modified the city's geographical role in the dimension field to county, which removed even more unknowns. **(see image below)**

**Graphical user interface, application, website, map

Description automatically generated**

**2. Is there any missing data in the dataset that needs to be addressed?**

The loaded dataset on Tableau appears to have no missing fields, but we still need to do a lot more to guarantee that our data is clean and ready for analysis. As seen in the image below we used Microsoft Excel to search for empty fields in the dataset, we were unable to find any. ( see image below)

**A screenshot of a computer

Description automatically generated**

**3. Is there any other form of pre-processing required for the data?**

The goal of data pre-processing is to ensure that the data is of high quality (Data Preprocessing in Data Mining -A Hands On Guide - Analytics Vidhya, 2022). Absolutely, there is no need to pre-process the data in any way.

**4. What methods were used to compile and calculate the data?**

The dataset was compiled by FBI and FOIA, it consists of information gathered on crimes that occurred in various parts of the United States of America (USA). From 1980 to 2014, there were over 638,454 criminal events reported throughout all 50 states.

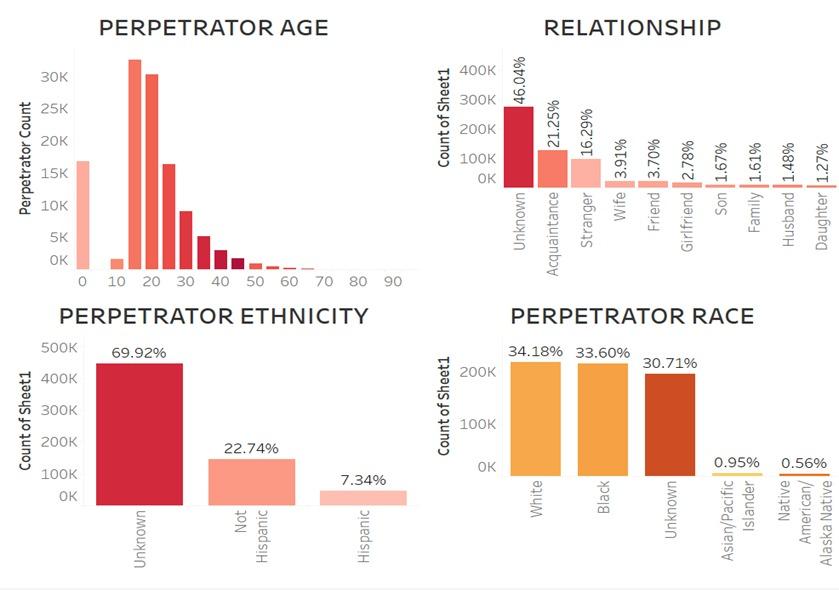
**5. Are there any other fields or calculations that could be included to make the data easier to understand?**

We did not see the need or reason to add extra fields or calculations to the dataset during our analysis.

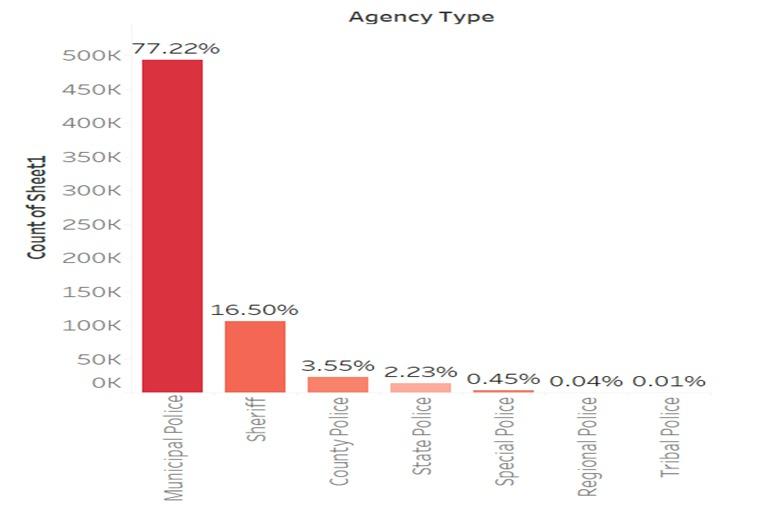
**6. What are the data's characteristics? It may be beneficial to explain the facts in a variety of ways.**

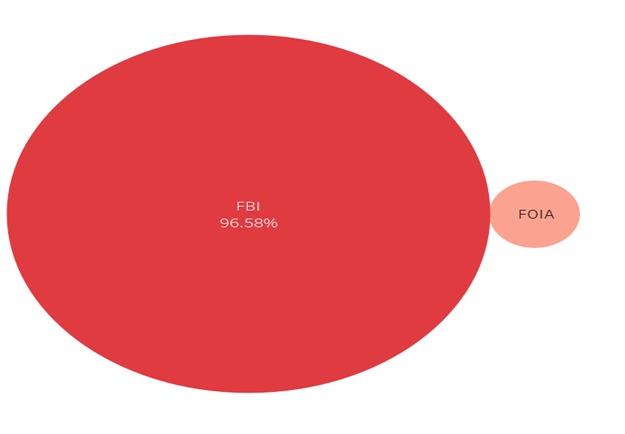
The data consist of the nature of the crime (Manslaughter by Negligence and Murder or Manslaughter), number of incident, victims (ethnicity, sex, race, age and count), perpetrators(ethnicity, sex, race, age and count), relationship between victim and perpetrator, weapon used to commit the said crime, year of crime, location of crime, agency type and record source.

Many of the value columns in the data had "unknown values". So, because the dataset contains so many unknown values, it is possible that some of the values that should be contained in this information wasn't captured.. (**see image below**)

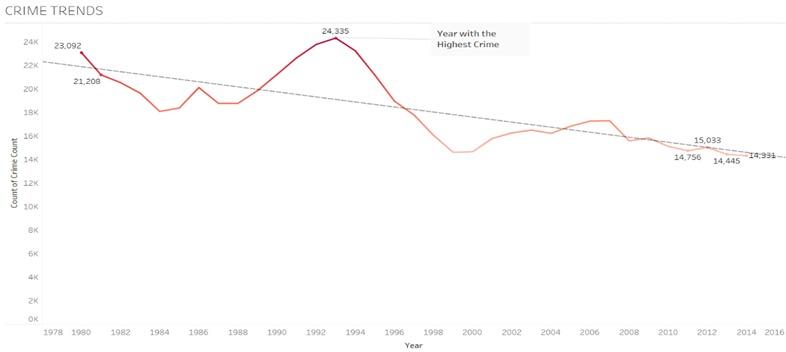


The **images below** show the agency type that took charge to solve the crimes committed in the data and the source of the data. The FBI and the Freedom of Information Act (FOIA) are the source of the data, with the FBI responsible for compiling 96.58% of the dataset and 4.42% from FOIA.





The last image (**as seen below**) shows the crime trend from 1980 to 2014 indicating progression and regression trends. The peak of our line graph occurred in 1993 with a crime of 24,335 recorded which is indicated with a point on our graph.



**February 17 - Project Meeting with industry advisor**

**MILESTONE 2 :** Understand the business requirements and identify stakeholders

There was a meeting with Azhar Awan, an expert in crime analysis. He gave some insights on the requirements of the project and how to tackle it. From the datasets collected, we can predict that police, insurance companies, lawyers, health authorities, district attorneys, government agencies, and other law enforcement agencies will be interested in this analysis. Instead of guessing from my assumptions, I did more research on the people involved in the criminal justice system to get a better understanding.

The people involved in the criminal justice system are the affected people. In other words, someone who has something to lose. In this case there will be participation with both Internal and external stakeholders. The criminal justice system can be considered an internal stakeholder, but an external stakeholder is an entity that the criminal justice system serves and influences in some way. Identifying these parties can provide insight into the criminal justice system.

**Internal Stakeholders**

Police officers, judges, protection observers, protection observers, correctors, lawyers, court officials, etc. are known as internal stakeholders. Internally, the criminal justice system includes those responsible for the enforcement and interpretation of the law. Insiders of the criminal justice system may include individuals suspected of being criminal and convicted. All of these parties are interrelated and influence each other. These stakeholders need to play their respective roles in effectively preventing crime, arresting criminals, and treating them.

**External stakeholders**

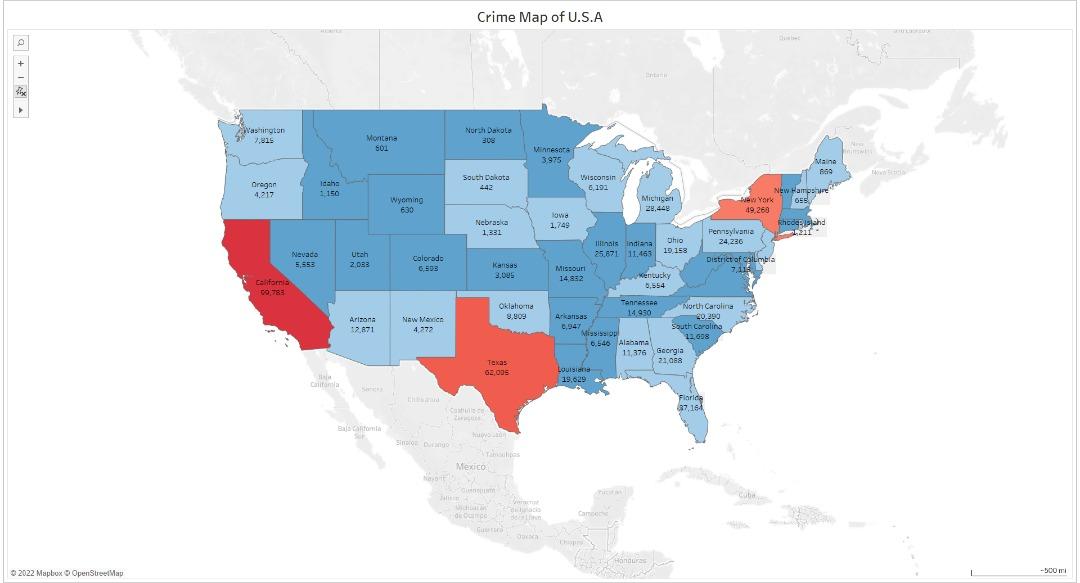
Elected authorities, media, victim support organisations, businesses, families, schools, and societies are all external stakeholders. When people become victims of crime, they are affected. It is important to address the individual's impact on the criminal justice system as a stakeholder. For example, by defending a defendant, a criminal defence counsel affects everyone else involved. How criminal accusations are defended is whether they are deprived of their liberty, whether their release endangers society, and how the general public sees lawyers. And how the court oversees the case (Study.com, 2022).

Some of the questions that law enforcement authorities will be seeking answers to in the dataset are as follows:

* Which city in the United States has the highest and lowest rates of crime?
* What percentage of crimes are solved in the United States?
* Is there a pattern to how crime has progressed through time?
* What is the relationship between victim vs perpetrator?
* What were the common weapons used to commit crime in the U.S?
* Which year saw the greatest increase in crime in the U.S?
* At what age is a person more likely to commit a crime?
* What are the perpetrators' demographic characteristics?

**February 25, 2022, April 1, April 8 & April 15 - Team Meetings**

MILESTONE 3: Visualise the dataset to try and answer the business questions



The image above shows in details parts of the United States were most of the crime gathered took place

**Is there a pattern to how crime has progressed through time?**

Chart, line chart

Description automatically generated

The chart above shows the progression and regression of crime rate in the entire dataset. We used a line chart because a line chart is used to compare data over different periods. The graph is ideal for providing visual context to my viewers.

**What is the most typical crime technique in terms of weapons used?**

Chart, bubble chart

Description automatically generated

The weapon was displayed in bubbles or circles on the bubble chart. It displays the percentage, the sizes depict the percentage difference between a weapon and other weapons, and the colours easily identify each weapon. We used a bubble chart because it clearly shows the weapon used to commit the crimes.

**Which city has the highest and lowest rates of crime (Solved & Unsolved)?**

The graph below (see image below) shows the states with high rate solved and unsolved crimes. We used the chart to show the states with the highest number of crimes across the period of data collection. This graph takes into account both solved and unsolved crimes.

Chart, waterfall chart

Description automatically generated

**At what age is a person more likely to commit a crime?**

Graphical user interface, application, Word

Description automatically generated

From the chart displayed above we can deduce that people are more likely to commit crimes in their 20’s compared to other age groups

**What is the relationship between victim vs perpetrator?**

Graphical user interface, text, application

Description automatically generated

We used a vertical bar chart to show the relationship between perpetrator and victim so as to display the variables appropriately.

**What Percentage of These Cases Are Solved?**

This data has been gathered over time, and we decided to use this graph to show the percentage of crime cases reported that have been solved. This is an essential graphic for our stakeholders because it is one of the most important indicators of law enforcement performance in the US. (**see image below**).

Graphical user interface, application

Description automatically generated

**April 22 - Team Meeting**

**MILESTONE 4: Developing a narration for telling a story around the data**

While working with this dataset, I observed certain distinct elements that I needed to use to develop a narrative to convey a story. Firstly, popularly and populated states like New York, California, Texas and Florida, had the greatest levels of crime. This made me ask the question of if the population is a factor in crime, I researched the factors that affect crime and I came across an article that mentioned 4 factors that lead to crime.

Gun rights, rate of unemployment, population density per square mile, and the percentage of the population living in real estate are four statistically significant explanatory variables that lead to high crime in a city (Abhijeet Bhattacharya, 2020).

My investigation verified that municipal police receive the most reported crimes, and also the municipal agency has the highest rate of crimes solved. In overall performance, law enforcement agencies did a good job, as we can see by the high percentage of crimes solved (70%).

Furthermore, I was able to make a conclusion from the chats below that people in their 20s are the most likely to be victims of crime, as well as the most likely perpetrators of crime.

Chart, treemap chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart

Description automatically generated

Chart

Description automatically generated

Chart, waterfall chart

Description automatically generated

**April 25, 2022**

MILESTONE 5: Identify recommendations for future improvements and enhancements

What measure can help law enforcement agencies reduce crime?

**Recommendations**

* In order to avert tragic fatalities such as manslaughter, I would suggest a more tactical approach to crime prevention. In high-crime areas like California, New York, Texas, and Florida, law enforcement agencies should outlaw the illegal use of handguns, which account for 49.73 percent of all weapons used in these crimes. In addition, the use of firearms has increased in recent years.
* I strongly believe that law enforcement authorities may efficiently reduce crime in these communities by concentrating resources on the gender with the high rate of perpetrators and establishing rehabilitation programmes for crime victims.
* Also, since the age group with the highest number of crimes are individuals in their 20’s, I’ll recommend that the community leaders/governing bodies should do a sensitization/outreach programme in a way to learn soft skills and start up. I strongly believe this can actually reduce crime and violence among youths or A poverty alleviation scheme would reduce crime rate across all age groups.
* In conclusion, I'd like to advise that a crime should be treated as a crime regardless of background, religion, gender, age or . As a result, bias would be minimised, and a larger proportion of crimes would be solved.

**Reflective Report**

Working in a four-person group to begin with, we are all different, with different ideas and backgrounds, so we decided that dividing the work amongst ourselves would be the smarter way to handle and complete the task, which was a good idea as it proved effective, because as a group of four, we worked well together and would regularly check in on group meetings, showing what they had accomplished and relating ideas on what could be done better.

The dataset was tedious to work with as there are too many unknowns in the data: there's a whole lot of value columns with "unknown values". Also, deciding as a group on which charts and maps to utilise was difficult. Colour mapping and shape were also limitations for us at the start of the project. This stalled a little since there was some back and forth about which colour and chart would be best for our target audience.

I believe that the fact that we are all unique in our own way, with varied backgrounds, strengths, and shortcomings, is why things worked or went well. As a result, dividing the project amongst us was a wise and strategic decision. Coming together weekly was also a wise decision because it allowed us to plan, structure, correct, and add to what we were all doing on a weekly basis rather than waiting until submission week to try to piece it all together.

I was able to properly comprehend and present a data-driven narrative after reflecting on my analysis thus far and thinking about things I've done by observing and experimentation. This platform, as well as the techniques I employed to create this data visualisation solution, will benefit me in efficiently communicating with any relevant business stakeholder dealing with crime in the future.

Finally, with the support of the module tutors Especially Haxia, I was able to put my good people skills to good use, learned how to communicate more, and improve my management skillsthrough this project. From the perspective of business intelligence and data visualisation, I will continue to practise data processing, data analysis, pattern discovery, visual storytelling, and effective communication.

**References**

Iaca.net. 2022. [online] Available at: <https://iaca.net/what-is-crime-analysis/> [Accessed 20 February 2022].

Zhou, W., J, W. & B, Y (2010). Analyzing Spatio-temporal Distribution of Crime Hot-spots and Their Related Factors in Shanghai, China. Shenzhen, 11 - 14 May 2010. pp. 1-4.

Rummens, A., Hardyns, W. & Pauwels, L (2017) The use of predictive analysis in spatiotemporal crime forecasting : building and testing a model in an urban context. Applied Geography [Online]. 86, pp. 255-261. [Accessed 20 February 2022].

O’Connell , S. & Gaunt, K. E. (2020) United States: Use of Data to Detect Crime and Evaluate Corporate Compliance. Americas Investigations Review 2021 [Online] [Accessed 18 March 2022].

National Institute of Justice (2017) Mapping and Analysis for Public Safety. Available:https://www.nij.gov/topics/technology/maps/pages/welcome.aspx. [Accessed 18 March 2022].

Cornell Law School. 2022. [online] Available at: <https://www.law.cornell.edu/definitions> [Accessed 29 March 2022].

Sisense. 2022. What is Data Cleaning? | Sisense [online] Available at: <https://www.sisense.com/glossary/data-cleaning/> [Accessed 30 March 2022].

Analytics Vidhya. 2022. *Data Preprocessing in Data Mining -A Hands On Guide - Analytics Vidhya*. [online] Available at: <https://www.analyticsvidhya.com/blog/2021/08/data-preprocessing-in-data-mining-a-hands-on-guide/> [Accessed 2 April 2022].

Study.com. 2022. Identifying criminal injustice stakeholders [online] Available at: <https://study.com/academy/lesson/identifying-criminal-justice-stakeholders.html> [Accessed 2 April 2022].

Bhattacharya, A., 2020. *Analysis of the Factors Affecting Violent Crime Rates in the US*. [online] Papers.ssrn.com. Available at: <https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3732028> [Accessed 4 April 2022].