**CRIME OR SAFETY??**

**Name:Vedasamhith Alloori**

**Net ID: va278**

**Email Id:va278@scarletmail.rutgers.edu**

**What is Crime and Why I have chosen Crime Data ??**

Imagine waking up in the morning and watching the news on TV. You hear a devastating news which will ruin your mood the whole day that a celebrity got murdered or there has been multiple gun shots at mall. You keep thinking about it at work the whole day and you couldn’t perform as well as you do at work. You may not think it, but it affects us every day. While everyone desires safety, achieving 100% safety is an unreachable goal. Nevertheless, thanks to the efforts of law enforcement and dedicated individuals, our world is made safer each day... A dream of mine maybe ,FBI to see the work I did so I can work with them as a data analyst. My experience working with the Rutgers University Police Department as a CSO has sparked a passion for making small changes that contribute to a safer environment for students, faculty, and professors. This realization led me to choose crime data as the focus of my capstone project. There is a lack of awareness among people about the intricacies of crime data, and my goal is to make a modest contribution to creating a safer society. I wanted to bring that goal combined with my Data Analytics skills using SQL to validate the data , clean it and did few analytics to preprocess it & Tableau to present my findings .

**Preprocessing the Data**

* Have loaded the 3 data files into SQL Server Using Microsoft Access . After loading it , have validated the loaded data if the data got loaded properly checking any row duplicates. Got 200 row duplicates and used a SQL code to delete them. Have checked for E+ values for numerical value columns. For one data , found 100 E+ values. So, I have deleted the table in SQL. Changed the following column in the raw data excel from number format to general and loaded it again and got no E+ Values . Have done some analytics by changing the date format and concating few columns to the proper format. Did grouped by for age column doing analysis on it and saved it in the proc. Then after analysis on SQL , have loaded the SQL data into Tableau.

**Tableau**

* Have created 15 Visualizations which I have put them into 4 different dashboards of different datasets.

**Dashboard 1:**

A screenshot of a computer screen

Description automatically generated

* This Dashboard represents the Crime Across US in the year 2018 and have added 3 Visualizations in this dashboard.
* The first visualization is a map chart of the US total crimes where have showed total crimes as legends and added the total population in the tool tip. we can infer that CA to be the first crime happening state and if we can hover on any state, it will show the state name, total crimes happened in 2018 and the total population of that state.
* The 2nd Visualization is a butterfly chart where the pink color states the number of employment people, and the blue color states the unemployment people . We can state that because CA has the highest Population it’s shown as that CA is the highest in the employment and unemployment people. Whereas TX is the second highest which has highest unemployment people. We can conclude that maybe because people are unemployed, there may be more crimes happening since they are illiterate and have no job.
* The 3rd Visualization is also a butterfly chart where the orange color states the number of violent crimes per population and the blue color states number of non-violent crimes per population. CA has the highest number of violent crimes and non-violent crimes.
* So from the Dashboard, We can come to a conclusion saying that maybe the laws are less strict and has the highest population in CA around 22.2M that’s why lot of crimes happen in CA.

**Dashboard 2:**

**A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated**

* This Dashboard represents the Crime Across per state specifically in the year 2018 and have added 4 Visualizations Which is related to each other in this dashboard. Added the state filter for this dashboard and have kept NJ for now since I’m living in NJ right now.
* In the 1st Visualization you can see the total number of crimes happening in that state. We can state that in NJ ,325k Crimes happened.
* In the 2nd Visualization, which is a Lollipop Chart, have kept the crime type wise and can see how many crimes happened for each type of crime in that state. So we can see that in NJ , the highest type of crime is Larcenies of 1.6M crimes, followed by Burglaries of 65K, Robberies of 51K and so on. We can conclude that maybe people or the public are not careful to be secured with their personal properties like jewelries or Cars or etc there is lot of larcenies happened in 2018. And maybe can suggest people to keep their property safe and changing their lock system in the car and their houses as well as larcenies may reduce in the future.
* In the 3rd Visualization, which is a tree map , have kept the total crimes by community name and can see how many crimes happened for each community name by sorting the highest crime for the top community. So from the tree map , we can infer that Newark has the most crimes than Jersey city and Camden. Since the place is the downtown in NJ , we can conclude maybe there are lot of crimes happening and there may have been less police force there . I would suggest the public from inferring that to be safe and avoid to travel alone at late hours .
* In the 4th Visualization, which is a scatter plot, have considered the median family income and the median income per community to see if that is related to the crimes happening in that community and have put total crimes as the legend. So from the graph we can see that Milburn township has the highest income which has one of the least crimes . Economic hardship may lead some individuals to resort to theft or other property crimes as a means of survival where we can see Newark City one of the lowest income. It makes sense because maybe Milburn has the highest income and budget for safety measures and can conclude they have focused on safety measures to make it a less crime community.

**Dashboard 3:**

**A screenshot of a computer

Description automatically generated**

* This Dashboard represents the Police overview in US for the year 2018 and have added 2 Visualizations
* In the 1st Visualization , you can see the total amount of budget for police forces and etc for each state. We can see FL has the highest budget of $8668K and followed by CA. So we can conclude that a higher police budget can lead to increased law enforcement presence, better equipment, and improved training. In theory, this might contribute to better crime prevention and response where we could see but in the case of CA we could tell that the Inefficiencies, lack of strategic planning, or mismanagement could hinder the effectiveness of law enforcement efforts.It may be important to assess the practices and tactics employed by the police force. Excessive use of force, lack of community engagement, or issues related to accountability and transparency could contribute to challenges in crime reduction.
* In the 2nd Visualization , We can see that it is a bubble chart and it states the total number of hours police have worked overtime and the number of police which have been requested for the year 2018. Have kept the bubble size based on the sum of hours worked. We can infer that CA has the highest no of hours overtime worked of 8878 hours which is around 370 days of work 24 X7 approximate and cops requested per office is 2432. We can say since CA is the highest crime state , it makes sense for cops to overtime around 8878 hours.

**Dashboard 4:**

**A screenshot of a computer screen

Description automatically generated** **A screenshot of a video game

Description automatically generated**

* This is the final Dashboard which comprises of 6 Visualizations which have been taken from a different dataset of a NYC data of the total crimes and the complaints over the years. This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD).
* In the above 1st Visualization , which is an interactive line chart which states the offence description by crime complaints over the years from 2019-2022. From analyzing the visualization , can see that in the year 2019 Criminal mischief around 150 is in the highest followed by Miscellaneous penal law around 122 and assault 3 & related offences around 53 etc. By the year 2022, Miscellaneous penal law has reached to 218 which increased to the top and followed by assault 3 & related offences which has also increased drastically around 170 to the second most offence category.
* In the 2nd Visualization, it states the No of crimes occurring at the most occurring places and kept the size by the number of crimes which have shown as bubble chart. So, from the chart, we can see that most of the crimes happen mostly in the street around 26M crimes data over the years followed by residence-apt house around 17M crimes . So have made this Visualization to see the highest number of crimes happening at which places and to suggest reducing the crimes happening. To conclude to reduce the No of crimes happening maybe increasing the police forces and follow a bit strict timeline curfew to be safe. The integration of technology, such as surveillance cameras and data analytics, can also further enhance law enforcement capabilities. In conclusion, the coordinated implementation of these strategies, rooted in community engagement and preventative measures, is vital for effectively mitigating street-level crimes or etc. in New York City and fostering a safer urban environment for all residents.

**A map of the city

Description automatically generated**

* In the above 3rd Map Visualization , it shows the No of crimes happening per area of the borough type .Have kept the latitude column and latitude column for the map viz. Have kept the borough name as legend. As seen from the map above as we hover any area from the borough, we can see the name of the borough and the No of crimes happening.
* In the 4th Visualization which is a line chart states the number of crimes happened over the years. From analyzing the graph we can infer that there has been a drastic increase in the No of crimes for the year 2006 which was 520.34K Crimes as in the year 2005 it was 10.69K crimes. It has increased a lot . We can conclude that the surge in crime rates in New York City during the year 2006 can be attributed to a confluence of factors that collectively contributed to a challenging socio-economic landscape. Economic disparities and rising unemployment rates, exacerbated by broader national economic trends, likely fueled criminal activities, especially in economically marginalized neighborhoods. Additionally, shifts in law enforcement strategies or resource allocations during this period may have impacted crime reporting practices, leading to an apparent spike in recorded incidents.
* In the 5th Visualization, which is a bar chart which states the suspect and victim age grouping by crimes. Wanted to show which age group of suspects are targeting which age group people. As seen from the visualization above , I have grouped the ages as show below
* **>18 as Group Z**
* **18-24 as Group A**
* **25-44 as Group B**
* **45-64 as Group C**
* **65+ as Group D**

So, from analyzing the graph, we can state that most B group Suspects have been targeting the Group B age Victims. Individuals in the 25-44 age group may interact more closely with others in the same age range due to shared social circles, workplaces, or residential areas. Crimes may occur within these social networks.

* In the 6th Visualization, have kept a pie chart of the crimes completed & attempted. From the above chart we can state that 98.32% of the crimes have been completed and 1.68% of the crimes just have been attempted which have not been completed.

**Conclusion:**

* Visualizations and analyses conducted using the datasets have yielded insightful findings. From a macro perspective, the dashboards showcase the correlation between crime rates, economic indicators, and law enforcement budgets. Specifically, the visualizations reveal patterns across different states, delve into crime types within a specific state, and explore the relationship between income levels and crime. Notably, the examination of NYC crime data over the years highlights a significant spike in 2006, emphasizing the importance of understanding socio-economic factors and law enforcement strategies in interpreting crime trends.
* In conclusion, this project underscores the necessity of a comprehensive, data-driven approach to tackling crime. By combining technology, community engagement, and strategic interventions, we can pave the way for effective crime prevention and contribute to building safer, more resilient communities. The findings and visualizations presented here serve as a foundation for ongoing discussions and actions aimed at creating a society where safety is a collective endeavor.
* One effective way to inform law enforcement and contribute to crime reduction is through the establishment of community crime watch programs. These programs empower local residents to actively participate in the safety of their neighborhoods by collaborating with law enforcement agencies. By fostering open communication channels, residents can share information about suspicious activities or concerns, enabling law enforcement to respond promptly and strategically. Regular community meetings, where law enforcement officers provide updates on recent crime trends and safety measures, help build trust and strengthen the partnership between the police and the community.
* Additionally, leveraging technology, such as community-specific mobile apps or online platforms, can facilitate real-time reporting and information-sharing. By actively involving citizens in the crime prevention process, law enforcement gains valuable insights, and the community becomes an integral part of the solution, contributing to safer and more secure neighborhoods.

**Datasets / References:**

|  |
| --- |
| * <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i> |
| * [Crimes in US Communities Dataset (kaggle.com)](https://www.kaggle.com/datasets/michaelbryantds/crimedata) |
| * <https://data.cityofnewyork.us/Public-Safety/NYPD-Hate-Crimes/bqiq-cu78> |