UNIVERSITY of GREENWICH
Alliance with FPT Education

BTEC

Higher Nationals in Computing

Business Intelligence

ASSIGNMENT 1

Learner's name: Hà Quang Thống
Assessor name: Nguyễn Xuân Sâm
Class: GCS0903A
ID: GCS200763
Assignment due: 17th October 2022 **Assignment submitted**: 16th October 2022

ASSIGNMENT 1 FRONT SHEET

Qualification	**BTEC Level 5 HND Diploma in Computing**		
Unit number and title	Unit 14: Business Intelligence		
Submission date	17th October 2022	**Date Received 1st submission**	16th October 2022
Re-submission Date		**Date Received 2nd submission**	
Student Name	Ha Quang Thong	**Student ID**	GCS200763
Class	GES0903A	**Assessor name**	Nguyen Xuan Sam

Student declaration

I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice.

Student's signature	THONG

Grading grid

P1	P2	M1	M2	D1	D2

☼ **Summative Feedback:**		☼ **Resubmission Feedback:**
Grade:	**Assessor Signature:**	**Date:**
IV Signature:		

Table of Contents

1. Introduction 5
1.1. Motivations 6
1.2. Objectives 6
1.3. Summary 7
2. Data 7
2.1. Description dataset 7
2.2. Data processing 8
2.3. Summary 8
3. Methodologies 9
3.1. Correlation 9
3.2. Regression 10
3.3. Summary 12
4. Experiments and results 13
4.1. Install Libraries 13
4.2. Results 16
5. Conclusions and future works 21
5.1. Conclusions 21
5.2. Future works 21
Reference 22

Table of Figures

Figure 1: Reality of gun violence in the US 5
Figure 2: The chaotic scene of the shooting at Highland Parks 6
Figure 3: Datasheet of Violence gun 8
Figure 4: Display anaconda 13
Figure 5: Display of anaconda app 13
Figure 6: Jupyter Notebook 14
Figure 7: Import the required libraries 14
Figure 8: Analysis Gun Violence by Year 16
Figure 9: Statistical chart of the number of casualties by year 17
Figure 10: Python code for shooting statistics 18
Figure 11: The graph shows the number of gun riots by state (1) 19
Figure 12: The graph shows the number of gun riots by state (2) 19
Figure 13: Common age of suspects (1) 20
Figure 14: The average age of the suspect (2) 21

1. Introduction

The present century we are living in is called the era of peace. As most of the nation's today have come together to form treaties and alliances to maintain order and peace all over the world. The possession and use of firearms is strictly regulated and controlled by the laws of each country. But in a country that respects human rights like the United States, owning a gun is authorized by the state for public use. If you are a licensed citizen or have completed firearms training, you will be able to own them. In the state of Texas, a citizen over the age of 21 who has not been convicted of a felony or some other prohibition on record has the right to own a firearm.

Because owning a gun is so easy, some people take advantage of guns to solve their conflicts and problems. Multiple shootings have been recorded over the years. Gun violence is considered a big problem in America. Most condemning is that most victims of mass shootings are innocent children in schools. Therefore, the report below will focus on analyzing gun violence in US states from 2013 to 2018. At the same time, it will analyze whether gun violence trends in the coming years are related. regarding the number of gun ownership in the states.

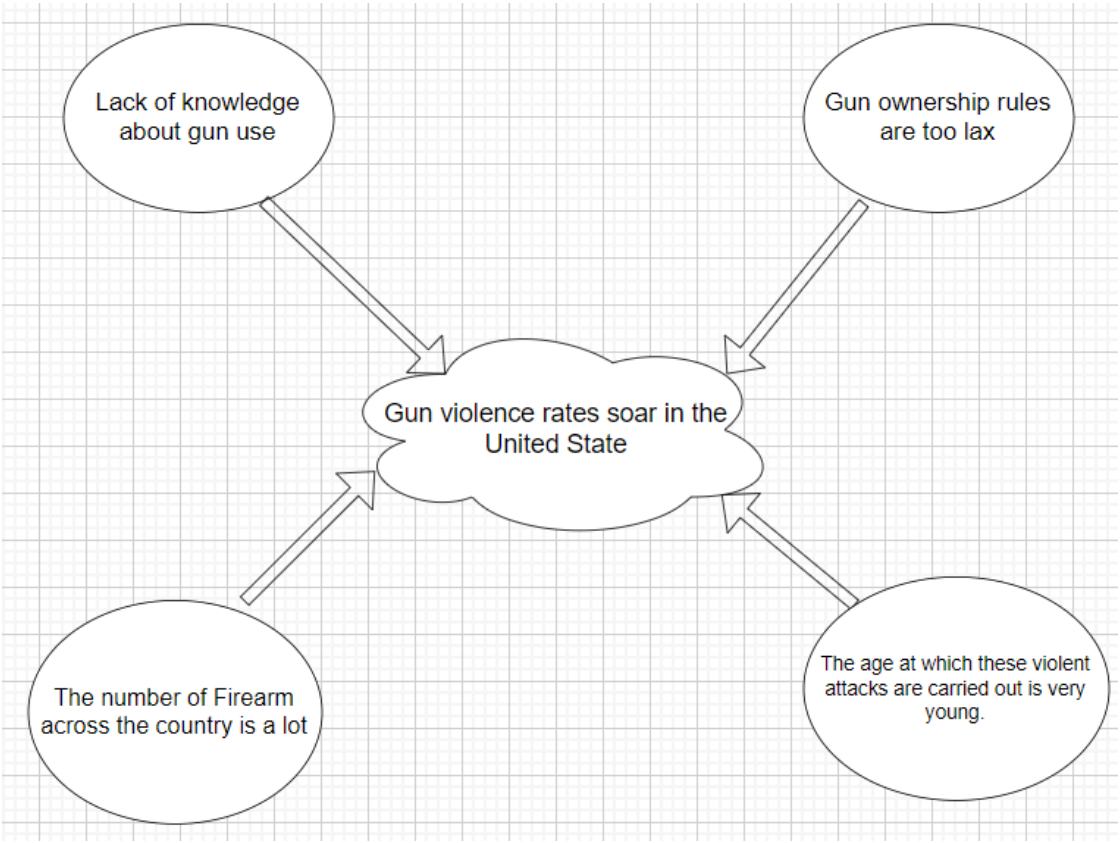

Figure 1: Reality of gun violence in the US

1.1. Motivations

Today, when reading newspapers or browsing bulletin boards. The gun question is a hot topic of debate around the world. More specifically, this topic is widely discussed in the United States, where gun riots are common. According to The Washington Post, the number of gun deaths reached more than 45,000 in the years 2020 and 2021, the highest since 1995(Mark Berman, 2022). Not only that, but bad elements also target schools, where the future of the world is sown. They opened fire again, killing the innocent children. At the same time, these tragedies show no sign of abating, creating chaos in the United States and danger to individuals around the world.

Furthermore, I find this booklet useful to assess the trends in gun violence development from 2013 to 2018. What if we could create a machine learning that recognizes signs like this so?

Figure 2: The chaotic scene of the shooting at Highland Parks

The most recent is the shooting at Highland Park on Independence Day of the United States. Causing a great commotion for the entire population when participating in the parade at that time. At the same time, there was a controversy that the use of guns in the US should be abolished. According to abc7chicago on July 6, at least 5 people died and 39 were injured in the shooting at this crowd (Evelyn Holmes, 2022). By September 29, according to the publication of this newspaper, a total of 7 people were killed and 48 people were injured(Mark Rivera, 2022). This is a number that makes the whole world concerned about gun violence in the US.

1.2. Objectives

For the current data that I have, I will focus on current data analysis with the following goals:

+ Collect and compare the number of gun violence incidents by year

+ Number of people killed and injured by year

+ Statistics of the number of gun violence incidents of each state across the US

+ What is the average age of these suspects?

1.3. Summary

In this part, I've already introduced and describe the motivations that may help me conduct research, analysis, and statistics of this data. Besides, I've placed the target of analyzing these articles. Furthermore, the work in this report includes 4-section, one of the rest is part of the data, to be more specific about the source of the data, the values of the data and how I processed the data to create a clean data.

2. Data

This dataset is taken from the Kaggle platform, it was collected by James Ko from gunviolencearchive.org. This data set is shared with the public for the purpose of studying gun violence and making informed predictions about future trends, he said. (KO, 2018)

2.1. Description dataset

In the data table that I collected, there are 9 attributes below:

- **Incident Id**: To number cases to make them easier to distinguish, or easier to find data when querying through code or by computer.
- **Date**: The date and year are clear when those accidents happened.
- **State**: Record all shootings in almost all states of the US.
- **City or County**: Counties or cities where the case of the previous state occurred.
- **Number of Killed(n_killed)**: The death toll of gun violence that took place
- **Number of Injured(n_injured)**: Number of serious or minor injuries from gun violence that took place
- **Incident URL**: URL to get data of these incidents.
- **Source URL:** Sources for statistics on this violence across the United States.
- **Congressional district:** The number of congressional districts created by the United States to divide districts during elections
- **Gun stolen:** The number of guns stolen when the incident occurred. After each incident of gun violence, police may or may not be able to recover the weapon they used. Because of that, they will rely on the testimony of the witness or the criminal himself to record the weapon they used or witnessed by everyone.
- **Gun type:** The type of gun that was used or stolen was seized at the investigative agency where it happened.
- **Incident characteristics**: The characteristics of that incident, documented. For example, how many bullets were fired, Was the injured or dead person killed or committed suicide or just accidentally shot.
- **Latitude / Longitude:** Exact location of the case

- **Location Description:** Describe the place where the crime took place, be it in the house, garden, restaurant, ...
- **Number gun involved**: Number of guns in the incident
- **Participant age/type/gender/name/status:** The age, relationship, gender, name and status of the people involved in the fight, whether the victim or the attacker.

The remaining information is intended to describe in detail the actions, location, district, relationship between the victim and the perpetrator where the fights took place. And sources of information to get this data

2.2. Data processing

This data is taken from the Gun Violence Archive site. The data provides more than 260,000 mass shootings across the US from 2013 to 2018. It shows virtually all shootings that took place during these years. Here are some pictures of that datasheet:

	incident_id	date	state	city_or_county	address	n_killed	n_injured	incident_url	source_url	incident_url_fields_mi...
1	461105	2013-01-01	Pennsylvania	Mckeesport	1506 Versailles Aven...	0	4	http://www.gunviolenc...	http://www.post-gazet...	False
2	460726	2013-01-01	California	Hawthorne	13500 block of Cerise...	1	3	http://www.gunviolenc...	http://www.dailybulleti...	False
3	478855	2013-01-01	Ohio	Lorain	1776 East 28th Street	1	3	http://www.gunviolenc...	http://chronicle.northc...	False
4	478925	2013-01-05	Colorado	Aurora	16000 block of East It...	4	0	http://www.gunviolenc...	http://www.dailydemo...	False
5	478959	2013-01-07	North Carolina	Greensboro	307 Mourning Dove T...	2	2	http://www.gunviolenc...	http://www.journalnow...	False
6	478948	2013-01-07	Oklahoma	Tulsa	6000 block of South ...	4	0	http://www.gunviolenc...	http://usnews.nbcnew...	False
7	479363	2013-01-19	New Mexico	Albuquerque	2806 Long Lane	5	0	http://www.gunviolenc...	http://hinterlandgazett...	False
8	479374	2013-01-21	Louisiana	New Orleans	LaSalle Street and M...	0	5	http://www.gunviolenc...	http://www.nola.com/c...	False
9	479389	2013-01-21	California	Brentwood	1100 block of Breton ...	0	4	http://www.gunviolenc...	http://sanfrancisco.cb...	False
10	492151	2013-01-23	Maryland	Baltimore	1500 block of W. Fay...	1	6	http://www.gunviolenc...	http://www.abc2news....	False
11	491674	2013-01-23	Tennessee	Chattanooga	1501 Dodds Ave	1	3	http://www.gunviolenc...	http://www.wrcbtv.co...	False
12	479413	2013-01-25	Missouri	Saint Louis	W Florissant Ave and ...	1	3	http://www.gunviolenc...	http://stlouis.cbslocal....	False
13	479561	2013-01-26	Louisiana	Charenton	1000 block of Flat To...	2	3	http://www.gunviolenc...	http://www.huffingtonp...	False
14	479554	2013-01-26	District of Columbia	Washington	2403 Benning Road ...	0	5	http://www.gunviolenc...	http://www.washingto...	False
15	479460	2013-01-26	Ohio	Springfield	601 West Main Street	1	3	http://www.gunviolenc...	http://www.whio.com/...	False
16	479573	2013-02-02	Tennessee	Memphis	2514 Mount Moriah	0	5	http://www.gunviolenc...	https://www.highbeam...	False
17	479580	2013-02-03	California	Yuba (county)	5800 block of Poplar ...	1	3	http://www.gunviolenc...	http://sacramento.cbsl...	False
18	479592	2013-02-07	Illinois	Chicago	2500 block of East 75...	0	4	http://www.gunviolenc...	http://chicago.cbsloca...	False
19	479603	2013-02-09	Louisiana	New Orleans	400 block of Bourbon ...	0	4	http://www.gunviolenc...	http://www.nola.com/c...	False
20	480311	2013-02-11	California	Vallejo	800 block of Humbold...	1	4	http://www.gunviolenc...	http://archive.news10....	False
21	480327	2013-02-11	Delaware	Wilmington	500 North King Street	3	2	http://www.gunviolenc...	http://www.philly.com/...	False
22	480344	2013-02-12	Utah	Midvale	8286 Adams Street a...	4	1	http://www.gunviolenc...		False
23	480358	2013-02-19	California	Orange (county)	Katella Avenue	4	3	http://www.gunviolenc...	http://www.dailymail.c...	False
24	480383	2013-02-21	Oklahoma	Tulsa	1200 block of North 8...	1	3	http://www.gunviolenc...	http://www.krmg.com/...	False
25	480401	2013-02-22	Michigan	Grand Rapids	1447 Grandville Ave. ...	0	4	http://www.gunviolenc...	http://www.mlive.com/...	False
26	480407	2013-02-23	California	Lancaster	43145 Business Cent...	0	4	http://www.gunviolenc...	http://latimesblogs.lati...	False
27	480443	2013-02-24	Georgia	Macon	2800 block of Mercer ...	0	8	http://www.gunviolenc...	http://www.macon.co...	False
28	481186	2013-03-02	Louisiana	Shreveport	7000 block of Burling...	1	3	http://www.gunviolenc...	http://www.shreveport...	False
29	481198	2013-03-03	Georgia	Moultrie	224 Second Street N...	2	2	http://www.gunviolenc...	http://www.moultrieob...	False
30	481208	2013-03-03	Michigan	Saginaw (county)	4030 Dixie Hwy	0	4	http://www.gunviolenc...	http://www.mlive.com/...	False
31	481213	2013-03-04	California	Los Banos	800 block of La Mesa ...	1	3	http://www.gunviolenc...	http://www.mercedsu...	False
32	481220	2013-03-05	Indiana	Indianapolis	1800 block of Edinbur...	1	3	http://www.gunviolenc...	http://www.wthr.com/s...	False
33	481229	2013-03-07	Mississippi	Jackson	2900 block of Greenw...	2	2	http://www.gunviolenc...	http://www.wapt.com/...	False
34	481237	2013-03-10	Missouri	Kansas City	9331 Hillcrest Rd	0	5	http://www.gunviolenc...	http://m.columbiatribu...	False
35	482771	2013-03-11	District of Columbia	Washington	1200 North Capitol St...	0	13	http://www.gunviolenc...	https://www.washingt...	False
36	482801	2013-03-13	California	Oceanside	504 Calle Montecito	2	2	http://www.gunviolenc...	http://www.sandiegou...	False
37	482856	2013-03-13	New York	Mohawk	17 W Main St	6	2	http://www.gunviolenc...	http://www.syracuse.c...	False
38	482838	2013-03-14	California	Modesto	1400 block of Wester...	0	4	http://www.gunviolenc...	https://www.youtube.c...	False

Figure 3: Datasheet of Violence gun

Because the data we obtained includes too much discrete data and data that is not relevant to the analysis. So, we removed a few extraneous columns like "incident_url" and "source_url", Furthermore, we created some new features by splitting date from values like 'date', 'year', 'month', 'monthday', 'weekday'. Besides, we also combine the values of two columns 'n_killed' and 'n_injured' into a 'loss' formula to facilitate data analysis and synthesis.

2.3. Summary

After retrieving and processing the data as mentioned above. Let's move on to part 3. This is the description and demonstration of the regression and correlation algorithms. For more details on correlation coefficients, linear regression and multiple linear regression

3. Methodologies

3.1. Correlation

The statistical link between two entities is referred to as correlation. In other terms, it's the relationship between the movements of two variables. Additionally, correlation may be used to a variety of data types. You may have guessed how certain events would link to one another in certain circumstances, while you may have been surprised by the relationship in others. It's crucial to realize that correlation does not always imply causation.

It's critical to comprehend the following concepts in order to comprehend how correlation functions:

- Positive correlation: One example of a positive correlation is This indicates that the two variables changed in the same direction, either up or down.
- A negative correlation is represented by the number -1. The two variables therefore went in opposing directions.
- A zero correlation indicates that there is no connection between the two variables. In other words, while one variable changes, another, unrelated variable changes as well.

Common formula to calculate correlation coefficient

The linear link between two variables is determined using Pearson's correlation coefficient. It has a value between -1 and 1. It is written as follows:

$$r = \frac{n\sum[(x_i-\bar{x})(y_i-\bar{y})]}{\sqrt{n\sum(x_i-\bar{x})^2 * n\sum(y_i-\bar{y})^2}} \quad (1)$$

In formulation (1) where:

n: Number of observations

$\mathbf{x_i}$: The value of x

$\mathbf{y_i}$: the value of y

$\bar{\mathbf{x}}$: The mean of value-x

$\bar{\mathbf{y}}$: The mean of value-y

where n is the sample size, $\mathbf{x_i}$ **and** $\mathbf{y_i}$ are the itch sample points and $\bar{\mathbf{x}}$ **and** $\bar{\mathbf{y}}$ are the sample means for the random variables x and y respectively.

The sign of r indicates the strength of the linear relationship between the variables.

- When r is close to 1, the two parameters have a strong linear connection.
- If r is near 0, then the two variables have no linear relation.

- When r approaches -1, the two variables exhibit a weak (negative) linear connection.(MATH, 2022)

What Exactly Is Linear Correlation?

Correlation coefficients range from -1 to +1. A correlation value +1 shows that there is a complete positive correlation. Variable y grows as variable x grows. Variable y reduces as x decreases. A correlation value of -1 implies that the relationship is completely negative. Variable z lowers as variable x rises. Variable z grows as variable x lowers.

The Bottom Line

The linear correlation coefficient could be useful in identifying the link between an investment and the broader market or other securities. It's frequently used to forecast stock market returns. This statistical measurement is valuable in many ways, notably in the banking business.

For example, it may be used to assess how well a mutual fund performs in comparison to its market index, or it can be used to establish how a mutual fund performs in comparison to another fund or asset class. Diversification advantages are acquired by adding a low-correlated, or negatively correlated, mutual fund to an existing portfolio. (NICKOLAS, 2021)

3.2. Regression

Regression is a statistical technique used in finance, trading, and other fields that aims to ascertain the strength and nature of the relationship between a single dependent variable (often represented by Y) and several other factors (known as independent variables).

Although it is a strong tool for identifying relationships between variables in data, causality cannot be clearly shown by regression analysis. It has a variety of applications in business, banking, and finance. It is employed, for example, to assist investment managers in valuing assets and comprehending the connections between variables like commodity prices and the stocks of companies that trade in those commodities. (BEERS, 2022)

Explain in a more understandable way

Regression assesses whether correlations between variables that are seen in a data collection are statistically significant.

Although there are non-linear regression techniques for more complicated data and analysis, simple linear regression and multiple linear regression are the two fundamental forms of regression. While multiple linear regression employs two or more independent factors to predict the outcome, simple linear regression just uses one independent variable or predict the result of the dependent variable Y. (while keeping all other variables constant).

Purpose of Regression

Regression is used in statistical analysis to uncover the relationship between the variables in data. It can indicate both the amplitude and statistical significance of such a correlation (i.e., whether the association is likely due to chance). Regression is a strong statistical inference method that has been used to make predictions based on past data.

Calculating Regression

3.2.1. Simple linear regression:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2)$$

In equation (2) where:

- **y**: for each given value of the independent variable, (x). is the anticipated value of the dependent variable (y).
- **x:** The independent variable(s)
- β_0: is the intercept, the predicted value of y when the x is 0.
- β_1: is the regression coefficient – how much we expect y to change as x increases.
- ε: is the error of the estimate, or how much variation there is in our estimate of the regression coefficient.

Linear regression determines the best fit line across your data by looking for the regression coefficient (β_1) that minimizes the model's total error (ε).

While it is possible to run a linear regression by hand, it is a time-consuming operation, thus most people utilize statistical software to assist them rapidly examine the data.(Bevans, 2020)

Assumptions of linear regression

If you think basic linear regression could be acceptable for your project, check sure it fits the linear regression assumptions given below. For additional details on each, see our analytical checklist: (graphpad, 2022)

- Relationship that is linear
- Scattered normally dispersed
- Homoscedasticity
- There is no uncertainty in predictions.
- Observations made independently
- For estimate, variables (rather than components) are employed.

Graphing linear regression

The Linear Regression calculator displays a general graph of your data as well as the regression line.(graphpad, 2022)

Graphing is useful not just for display but also for detecting outliers in your data. If there are a couple points that are significantly different from the others, there are a few possibilities: they might be excessively impacting your regression equation, or the outliers could be a significant finding. Use this outlier checklist to help you determine which is most likely in your specific situation.

3.2.2. Multiple linear regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ ... + \beta_t X_t + \varepsilon \quad (3)$$

In equation (2) where:

- **Y**: dependent variable
- **X:** explanatory variables
- β_0: The y-intercept
- β_t: slope coefficients for each explanatory variable
- ε: the model's error term

Statistical software is often used to compute the least-squares estimates—β_0, β_1, β_2... β_t. The regression model can include as many variables as desired, with each independent variable denoted by a number—1,2, 3, 4...t. The multiple regression model enables an analyst to forecast an outcome based on data from numerous explanatory factors.

However, the model is not always fully correct since each data point might diverge somewhat from the model's expected conclusion. To account for such minor fluctuations, the model includes the residual value, E, which is the difference between the actual and projected outcomes. (Bevans., 2022)

3.3. Summary

In this section, we have learned in-depth about linear and correlation algorithms and know the definitions and how to apply that algorithm. The next, which is also considered the last part in this project, aims to provide end-to-end results when retrieving and synthesizing data. In addition, this section also describes in detail how to download the tool and apply which libraries are needed in the analysis.

4. Experiments and results

4.1. Install Libraries

4.1.1. Install anaconda

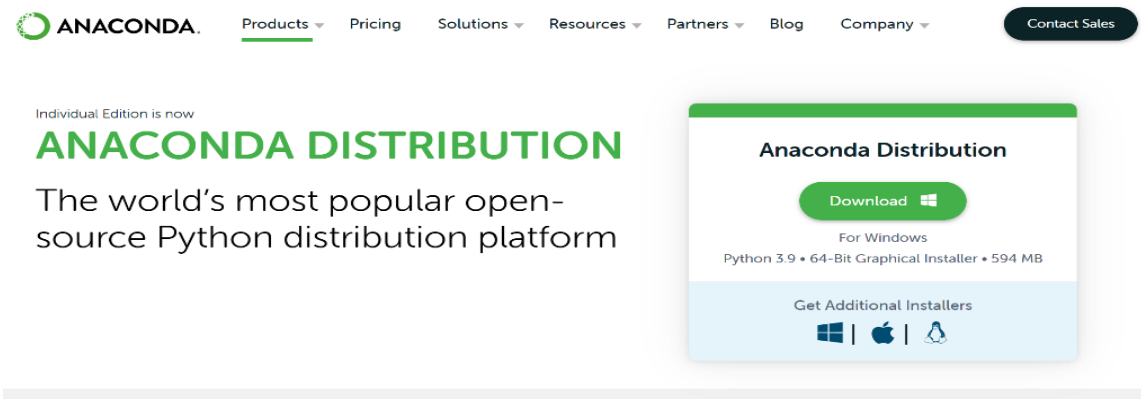

Figure 4: Display anaconda

To analyze data on Jupyter using Python language, the first thing we need to do is install Anaconda. To install anaconda, we need to do the following steps:

Step 1: Go to Anaconda Website, select Download and select the operating system compatible with your device.

Step 2: Since almost everyone uses Python 3.9 now, the default for anaconda is to install Python 3.9, select the equivalent system type for your machine and click install.

Step 3: After the installation is complete, open the application. Then install according to the instructions system steps

4.1.2. How to get into Jupyter and create New Folder?

After downloading the Anaconda application. Then we will start it up and have the interface as shown in Figure 5 below

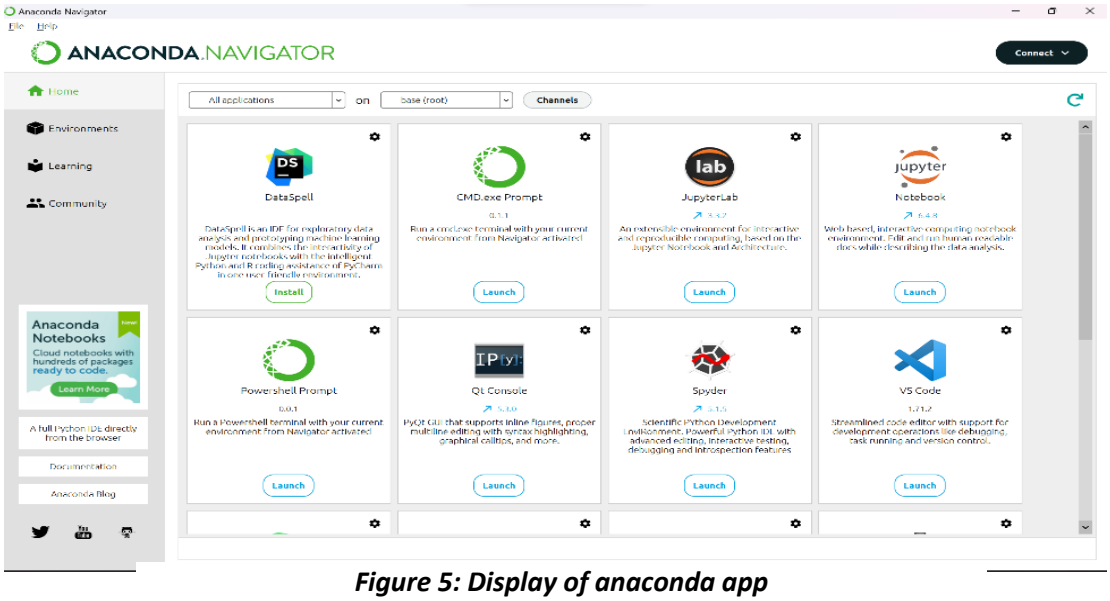

Figure 5: Display of anaconda app

To login to create a new folder we need to perform the following steps:

Step 1: Click the Launch button in the Jupyter Notebook application to enter the home page of Jupyter, it will look like Figure 6 below:

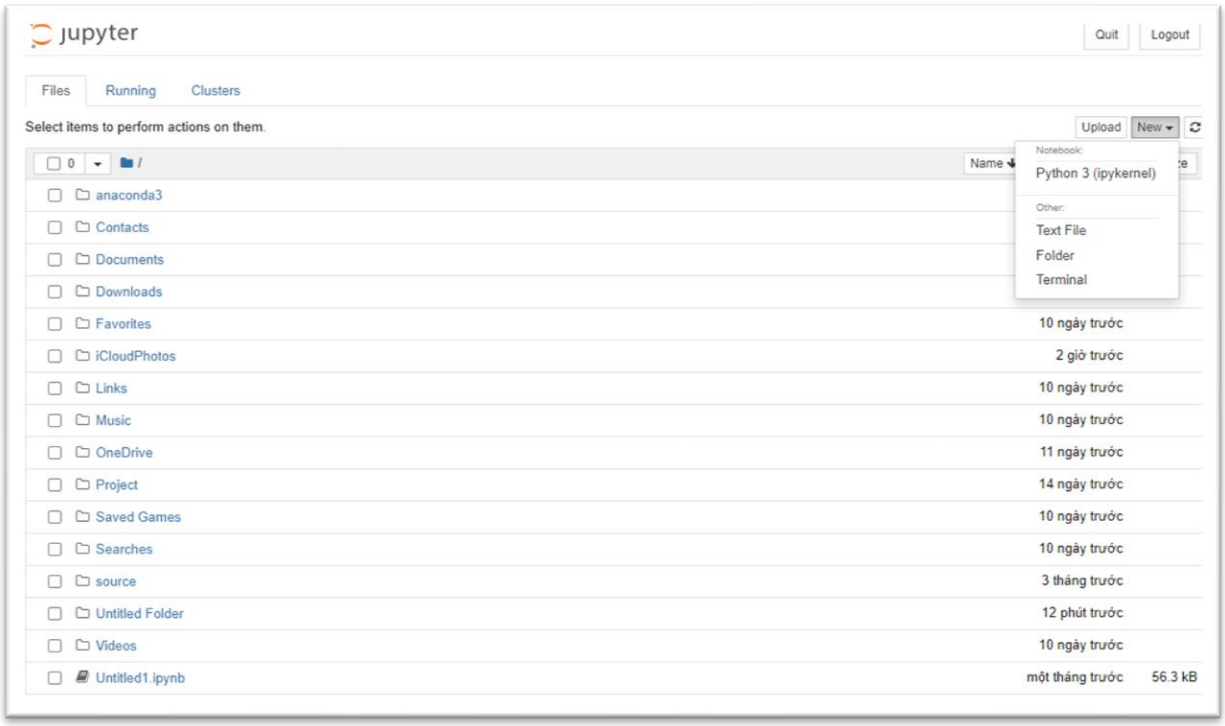

Figure 6: Jupyter Notebook

Step 2: After entering the interface of Jupyter Notebook, to create a new folder we need to click the new button and select Python 3 as shown above.

4.1.3. Enter library

To analyze and synthesize data from the Python language, it is necessary to import the necessary libraries used to analyze the images. In **Figure 7** below are some of the libraries needed for my analysis.

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

from plotly.offline import init_notebook_mode, iplot

init_notebook_mode(connected=True)
from datetime import datetime
```

Figure 7: Import the required libraries

Those libraries include:

- **NumPy**

NumPy is one of the most popular open-source Python libraries. It can handle big matrix and multidimensional material and has built-in mathematical tools for quick processing. "Numerical Python" is defined by the term "NumPy." It may be used in linear algebra, as a random number generator, and as a multi-dimensional container for generic data. (Vats, 2021)

- **Pandas**

Panda is a BSD-licensed open-source library. They are typically employed for data analysis, processing, and cleansing. Panda allows for simple modeling and data analysis activities without the need to convert to another language such as R.

- **Seaborn**

The package includes tools for visualizing statistical models. The Matplotlib-based library enables for the generation of statistical visuals via:

- Variable comparison using an API based on datasets.
- Heat chart
- Options of various color palettes to display the patterns.

- **Matplotlib**

This package is used for numerical data visualization and data analysis. This open-source toolkit is used to create high-quality figures like as graphs, pie charts, scatterplots, histograms, and so on.

- **Plotly.offline**

Plotly lets you create graphs offline and save them to your own system. The plotly.offline.plot() function generates a standalone HTML file that is saved locally and may be viewed in your web browser.

While working offline in a Jupyter Notebook, use plotly.offline.iplot() to display the plot in the notebook. (tutorialspoint, 2019)

4.2. Results

- **What is the year with the highest amount of gun violence?**

After identifying candidate models as well as metrics, we processed the data using Python. During our review, we found that the violence index of 2017 is very high, and they increase over time. According to the results we analyzed, in 2014 the number of violent cases reached 51,854,000 people and those violent incidents in 2017 reached 61,402 thousand people. This said that the number is very alarming if it keeps growing like this in the future. This is not a good indicator, but it speaks to the lack of gun control in the United States. Those numbers are clearly shown in **Figure 9** below.

```
year = pd.DatetimeIndex(dat['date']).year.value_counts()
year = year.rename_axis("Year")
year.sort_index().plot(kind='bar', title = "Gun Violence Incidents by year", figsize=(15,10))
```

```
<AxesSubplot:title={'center':'Gun Violence Incidents by year'}, xlabel='Year'>
```

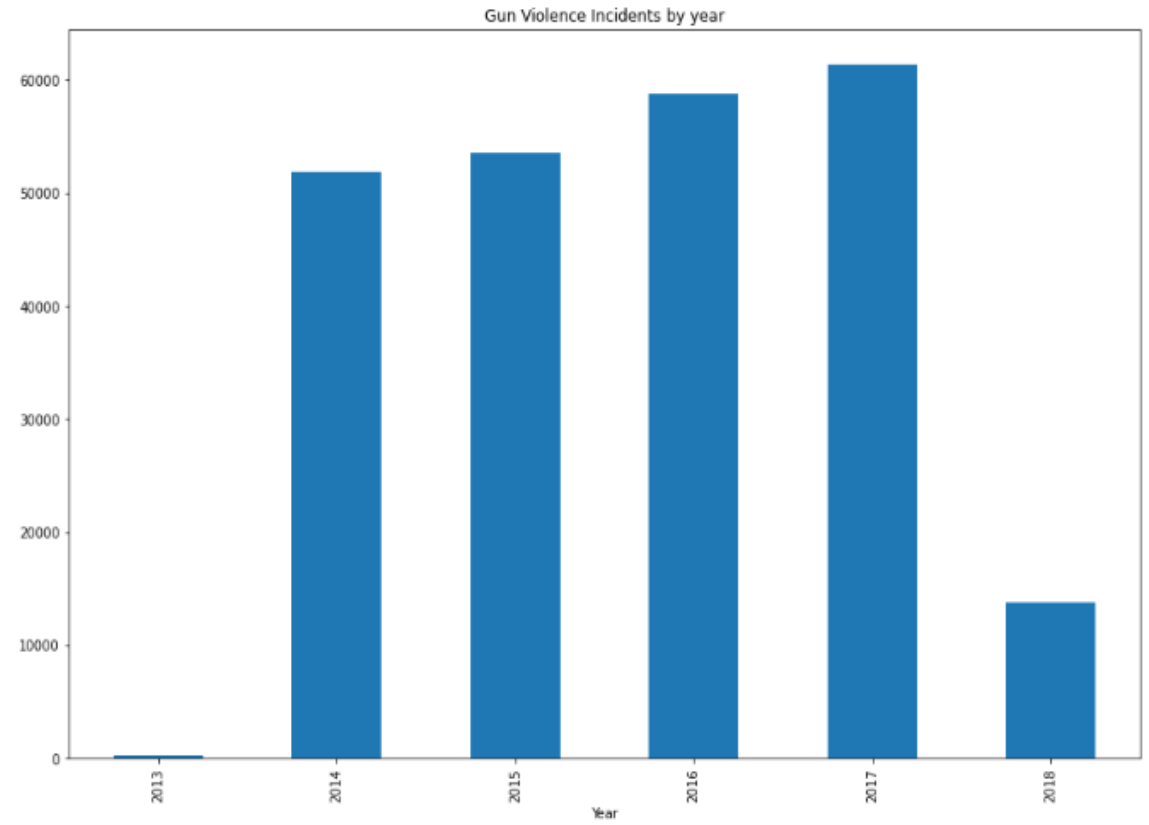

Figure 8: Analysis Gun Violence by Year

As you can see on the chart, from 2013 to the end of 2014, the rate of violence spiked suspiciously. Shows the government losing gun control over its citizens. In the years that followed, the number of gun violence cases increased dramatically. The peak of this figure was in 2017 when there were 61,402 estimated incidents of violence in the country. In addition, because the data was only aggregated until the beginning of 2018, the index in the statistics table is at the lowest level in the table.

- **How many were killed and died in each of those years?**

```
df = dat.groupby('year')[['n_killed','n_injured']].agg('sum')
df = df.rename(index=str, columns={"n_killed":"People Killed",'n_injured':'People Injured'})
# df.head()
df.plot(kind='bar', rot=0, title="Number of Casaulties by Year", figsize=(15,10))
```

```
<AxesSubplot:title={'center':'Number of Casaulties by Year'}, xlabel='year'>
```

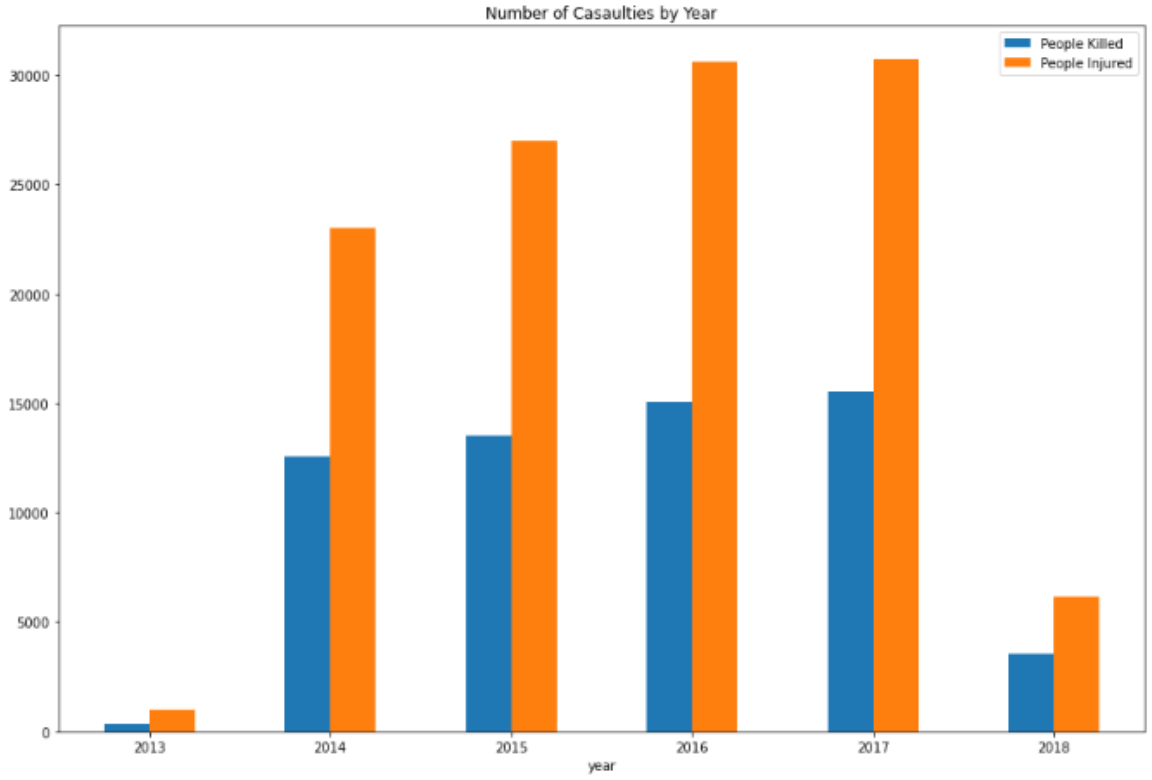

Figure 9: Statistical chart of the number of casualties by year

When looking at the chart above, as you can see. 2017 was the peak in terms of the highest number of deaths. Because as the number of gun violence increases, so does the number of casualties.

Fortunately, the victims of gun riots are mostly injured. It is worth mentioning here that the numbers increase rapidly and there is no restraint or constraint. This is to say that the legal and gun control agencies in the US in these years did not have any timely intervention in terms of the number of firearms sold and used.

In 2016 and 2017, the number of casualties was relatively equal, but the number of deaths in 2017 reached more than 16,000 deaths.

- **Which state has the most gun violence in the US?**

This statistic is surprising. I thought that the number of gun violence cases would be equal to the number of guns, but no. In 2017, the number of guns in sub-Texas was the highest in the country. According to Guns & Ammo, the state of Alaska is the easiest place to own a gun in the US, people can bring a gun and use it without a license. Thought these two places, those riots would happen more than most other states. But according to the data we have, the state of Illinois is the place where gun riots take place the most with more than 17,000 cases. In second place is the state of California with more than 16,000 cases. And Texas is only 3rd in this ranking. Illinois is the site of many guns' riots

```
states_df = dat['state'].value_counts()

statesdf = pd.DataFrame()
statesdf['state'] = states_df.index
statesdf['counts'] = states_df.values

scl = [[0.0, 'RGB(205,200,177)'],[0.2, 'RGB(255,127,80)'],[0.4, 'RGB(255,114,86)'],\
            [0.6, 'RGB(238,106,80)'],[0.8, 'RGB(205,91,69)'],[1.0, 'RGB(139,62,47)']]

state_to_code = {'District of Columbia' : 'dc','Mississippi': 'MS', 'Oklahoma': 'OK', 'Delaware': 'DE', 'Minnesota': 'MN', 'I
statesdf['state_code'] = statesdf['state'].apply(lambda x : state_to_code[x])

data = [ dict(
        type='choropleth',
        colorscale = scl,
        autocolorscale = False,
        locations = statesdf['state_code'],
        z = statesdf['counts'],
        locationmode = 'USA-states',
        text = statesdf['state'],
        marker = dict(
            line = dict (
                color = 'RGB(240,248,255)',
                width = 2
            ) ),
        colorbar = dict(
            title = "Gun Violence Incidents")
        ) ]

layout = dict(
        title = 'State wise number of Gun Violence Incidents',
        geo = dict(
            scope='usa',
            projection=dict( type='albers usa' ),
            showlakes = True,
            lakecolor = 'RGB(240,248,255)'),
             )


fig = dict( data=data, layout=layout )
iplot( fig, filename='d3-choropleth-map' )
```

Figure 10: Python code for shooting statistics

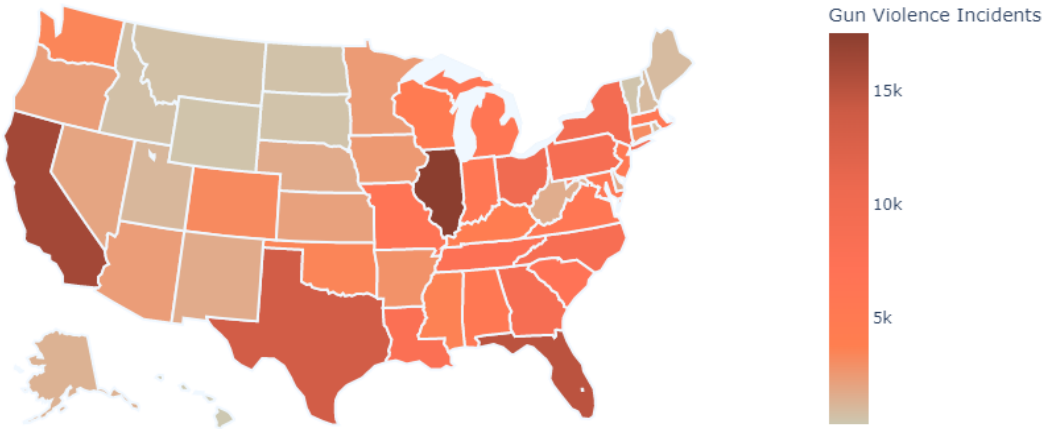

Figure 11: The graph shows the number of gun riots by state (1)

In the map showing the gun violence incident above, there are four states located in the dark red area: Illinois, California, Florida and finally Texas. The chart shows in detail every number of gun violence incidents across the United States. Most notably, states across the United States have had at least hundreds of cases of gun violence since 2013. And California has 16,306 thousand, followed by Florida with 15,029 thousand. A number as if counting the number of people in a certain country. But these are gun riots in America

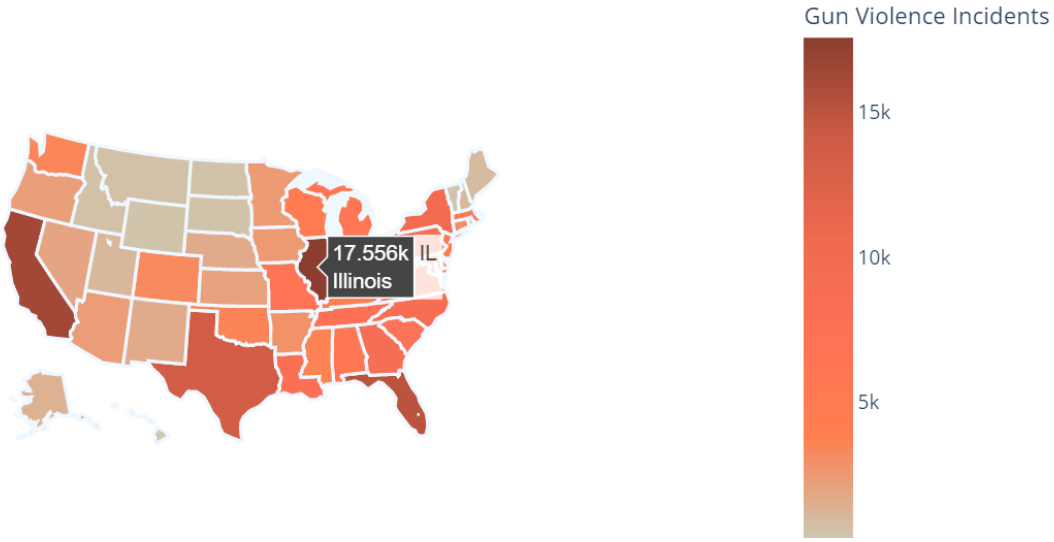

Figure 12: The graph shows the number of gun riots by state (2)

So why are this violence happening more and more, is it because of poverty that people fight and fight each other. Probably not, because the index of homeless here is low compared to the whole country, according to the United States interagency council on homelessness, the state of Illinois has only about 10,431 homeless people. So that's probably due to other evils, like the amount of gang fights that happen. This has yet to be verified, as there is no definitive answer to these questions from the government yet. (GOV, 2022)

- **What is the average age of these suspects?**

```
## Finding the Suspect Age Groups
suspect_age_groups = {}
for i, row in gun_violence.iterrows():
    suspects = []
    for k,v in row['participant_type_map'].items():
        if "suspect" in v.lower():
            suspects.append(k)
    for suspect in suspects:
        if suspect in row['participant_age_map']:
            ag = row['participant_age_map'][suspect]
            if ag not in suspect_age_groups:
                suspect_age_groups[ag] = 0
            else:
                suspect_age_groups[ag] += 1

trace1 = go.Bar(x=list(suspect_age_groups.keys()), y=list(suspect_age_groups.values()), opacity=0.75, name="month", marker=dict(color
='rgba(200, 20, 160, 0.6)'))
layout = dict(height=400, title='Suspects Age - Distribution', xaxis=dict(range=[0, 100]), legend=dict(orientation="h"));
fig = go.Figure(data=[trace1], layout=layout)
iplot(fig)
```

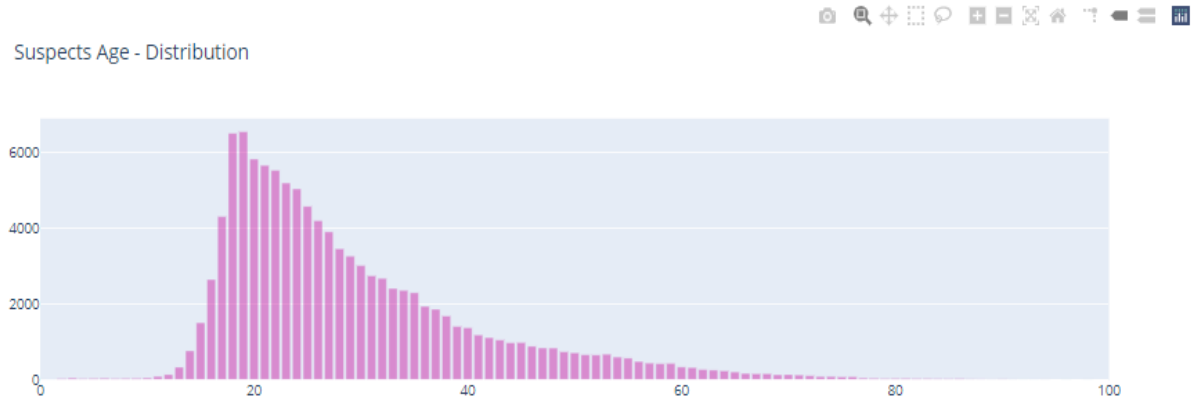

Figure 13: Common age of suspects (1)

Perhaps most people will wonder, what class do the suspects belong to and why they open fire so blatantly.

The answer is statistically through figure 13 above shows that. Most of the suspects range in age from teenagers to young men. The remainder is divided equally among different age groups.

In my opinion, they are probably quite dissatisfied with their early failures. Or maybe because of the daily conflicts they face, and they don't know how to handle them, they instead choose ever more extreme forms.

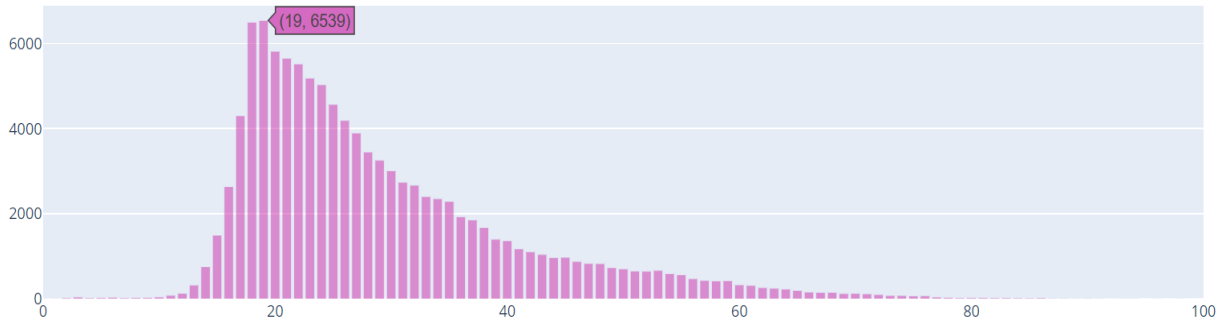

Figure 14: The average age of the suspect (2)

As you can see in the image above, the common age of these riot suspects is in their teens. First, they are not aware of the consequences they will receive after the violence they cause. The next thing, they are just entering the juvenile age, the age to be punished before the law, they have not yet controlled their own behavior, so these problems occur. This last one, in my opinion, is due to the conflicts of youth in a freedom, leading to them not being able to control themselves and using guns to do these negative things.

5. Conclusions and future works

5.1. Conclusions

To sum up, I have somewhat enumerated the reasons why gun riots take place more and more. It also answered my questions about the areas with the most riots, thereby giving me a better overview of gun violence in the US. In other words, it will help me somewhat better understand the dangerous areas with high risk of gun violence in the future. Besides, the statistics on the age of the suspects Crime will help the US government to somewhat enforce gun laws in specific states, moreover, by educating children about the harms and dangers to avoid the situation that is happening. out right now.

5.2. Future works

The data retrieved from GUN VIOLENCE ARCHIVE is somewhat limited, so we can only analyze a part of it at the time of 2018. Up to now, the US government has not improved this problem, so this is only possible. will probably be a prominent future work of the US government.

This is not just America's responsibility, because the issue of guns should be eliminated in the world, and it should not appear in human life. Moreover, America is a multi-racial country, so color discrimination happens very often, and it is impossible to use guns to solve these problems.

Reference

BEERS, B. 2022. *What is Regression? Definition, Calculation, and Example* [Online]. Available: https://www.investopedia.com/terms/r/regression.asp [Accessed].

BEVANS, R. 2020. *Simple Linear Regression | An Easy Introduction & Examples* [Online]. Available: https://www.scribbr.com/statistics/simple-linear-regression/#:~:text=What%20is%20simple%20linear%20regression,Both%20variables%20should%20be%20quantitative. [Accessed].

BEVANS., R. 2022. *Multiple Linear Regression | A Quick Guide (Examples)* [Online]. Available: https://www.scribbr.com/statistics/multiple-linear-regression/ [Accessed].

EVELYN HOLMES, K. J., JOHN GARCIA, CATE CAUGUIRAN,. 2022. *Highland Park parade shooting: Victims ID'd; grandfather, lifelong synagogue member among them* [Online]. Available: https://abc7chicago.com/highland-park-victims-mass-shooting-illinois-parade/12019612/ [Accessed].

GOV, U. S. 2022. Illinois Homelessness Statistics.

GRAPHPAD. 2022. *Linear regression calculator* [Online]. Available: https://www.graphpad.com/quickcalcs/linear1/#:~:text=The%20formula%20for%20simple%20linear,b%20is%20the%20estimated%20intercept. [Accessed].

KO, J. 2018. Gun Violence Data.

MARK BERMAN, L. B., DAN KEATING, ANDREW BA TRAN AND ARTUR GALOCHA 2022. The Washington Post.

MARK RIVERA, E. H. 2022. *Highland Park shooting survivors sue Smith & Wesson, Robert Crimo III and shooter's father* [Online]. Available: https://abc7chicago.com/highland-park-parade-shooting-lawsuit-smith-and-wesson/12276050/ [Accessed].

MATH, C. 2022. *Correlation Coefficient Formula* [Online]. Available: https://www.cuemath.com/correlation-coefficient-formula/ [Accessed].

NICKOLAS, S. 2021. *Correlation Coefficients: Positive, Negative, & Zero* [Online]. Available: https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-correlation-coefficient-positive-negative-or-zero.asp#:~:text=The%20correlation%20coefficient%20is%20determined,of%20data%20from%20its%20average. [Accessed].

TUTORIALSPOINT. 2019. *Plotly - Online and Offline Plotting* [Online]. Available: https://www.tutorialspoint.com/plotly/plotly_online_and_offline_plotting.htm [Accessed].

VATS, R. 2021. *Libraries in Python Explained: List of Important Libraries* [Online]. Available: https://www.upgrad.com/blog/libraries-in-python-explained/ [Accessed].