Gun Violence Analysis

submitted by

Paridhi Narayan (01501032017)
Isha Kumar (02101032017)
Jaskiran Sekhon (03101032017)
Jyoti Rani (04701032017)
Isha Singh (04801032017)

Under the kind guidance of
Mr. Rishabh Kaushal
Assistant Professor

Department OF Information Technology
Indira Gandhi Delhi Technical University for Women
Kashmere Gate, Delhi - 110006

March 2020

Contents

Chapter 1

Introduction

Gun violence is a violence committed with the use of firearms like pistols , shotguns , assault rifles or machine guns. Gun violence is a contemporary global human rights issue. Gun related violence threatens our most fundamental human right, the right to life.It is a daily tragedy affecting lives of millions of individuals across the world. According to the statics provided by amnesty.org worldwide,

- More than 500 people die every day from gun violence
- 44 percent of all homicides globally involve gun violence
- There were 1.4 million firearm related deaths globally between 2012 to 2016

The leading cause of gun violence is easy access to firearms -whether legal or illegal. There are 8 million new small arms and up to 15 billion rounds of ammunition produced every year globally. Gun violence highly hampers the physical as well as mental health of individual due to fear of life but also due to criticism for having personal relations with the participant in crime. States have an obligation to fight actual or foreseeable threats to life and should therefore take measures to protect from gun violence which includes adequately controlling how private individuals own and use firearms, controlling access to and proliferation of firearms, etc. Types of gun violence are – mass shooting, school shooting, suicide , intentional homicide shooting , domestic violence, robbery and assault , accidental shooting, pleasure shooting , psychological ,etc.

1.1 Problem Statement

Gun violence is leads to violation of fundamental human rights like – right to life and right to health. It is increasing aggressively among youngsters because of many psychological , mental, emotional reasons which not only hampers their well being but also of others in relation with them or not . It can be prevented or some how reduced by taking measures by the state and individuals. Gun violence depends on various factors , some of them can be concluded as personal relation with victim, mental state of convict , access to firearms, location , occasion , etc . But the question lies , can we model age groups in which these violence take place along with reason and predict future incidents based on previous years data in particular areas all over the world?

1.1.1 Objective

- To analyse the data and determine the age groups in which gun violence takes place maximal and minimal.
- To analyse the reason for violence and to predict future incidents.
- To inform the concerned authorities from the result predicted to bring down the rate of violence .

Chapter 2

Literature Survey

2.1 Research Questions-

1. What are the number of incidents occurring according to time series?
2. Does the number of incidents depends on whether it is a special day like Christmas, New Year etc or not?
3. Which are states where maximum violence occurs?
4. What is number of incidents where the victim is related to the suspect?
5. What are the type of guns involved in each type of incident?
6. What is the percentage of homicides, mass shootings and suicides among the incidents?
7. What are the major motives that drives different genders into gun violence?
8. What are the age groups in which most of the gun violence takes place?

Chapter 3

Proposed Methodology

3.1 Dataset

Number of Instances : 260000
Number of Attributes : 29
Label Information : Label present

Attribute Name	Attribute Type
INCIDENT_ID	Numerical
DATE	Categorical
STATE	Categorical
CITY_OR_COUNTY	Categorical
ADDRESS	Categorical
N_KILLED	Numerical
N_INJURED	Numerical
INCIDENT_URL	Categorical
SOURCE_URL	Categorical
INCIDENT_URL_FIELDS_MISSING	Categorical
CONGRESSIONAL_DISTRICT	Numerical
GUN_STOLEN	Categorical
GUN_TYPE	Categorical
INCIDENT_CHARACTERISTICS	Categorical
LATITUDE	Numerical
LOCATION_DESCRIPTION	Categorical
LONGITUDE	Numerical
N_GUNS_INVOLVED	Numerical
NOTES	Categorical
PARTICIPANT_AGE	Numerical
PARTICIPANT_AGE_GROUP	Categorical
PARTICIPANT_GENDER	Categorical
PARTICIPANT_NAME	Categorical
PARTICIPANT_RELATIONSHIP	Categorical
PARTICIPANT_STATUS	Categorical
PARTICIPANT_TYPE	Categorical
SOURCES	Categorical
STATE_HOUSE_DISTRICT	Numerical
STATE_SENATE_DISTRICT	Numerical

3.2 Attribute Description

- INCIDENT_ID
- DATE-Date of crime
- STATE-State of crime
- CITY_OR_COUNTY-City/ County of crime
- ADDRESS-Address of the location of the crime
- N_KILLED-Number of people killed
- N_INJURED-Number of people injured
- INCIDENT_URL- URL regarding the incident
- SOURCE_URL-Reference to the reporting source
- INCIDENT_URL_FIELDS_MISSING-TRUE if the incident_url is present, FALSE otherwise
- CONGRESSIONAL_DISTRICT-Congressional district id
- GUN_STOLEN- Status of guns involved in the crime (i.e. Unknown, Stolen, etc...)
- GUN_TYPE-Typification of guns used in the crime
- INCIDENT_CHARACTERISTICS- Characteristics of the incidence
- LATITUDE-Location of the incident
- LOCATION_DESCRIPTION
- LONGITUDE-Location of the incident
- N_GUNS_INVOLVED-Number of guns involved in incident
- NOTES-Additional information of the crime
- PARTICIPANT_AGE-Age of participant(s) at the time of crime
- PARTICIPANT_AGE_GROUP-Age group of participant(s) at the time crime
- PARTICIPANT_GENDER-Gender of participant(s)
- PARTICIPANT_NAME-Name of participant(s) involved in crime
- PARTICIPANT_RELATIONSHIP-Relationship of participant to other participant(s)
- PARTICIPANT_STATUS-Extent of harm done to the participant
- PARTICIPANT_TYPE-Type of participant
- SOURCES-Participants source
- STATE_HOUSE_DISTRICT-Voting house district
- STATE_SENATE_DISTRICT-Territorial district from which a senator to a state legislature is elected.

3.3 Details about the Organisation

The data was downloaded from gunviolencearchive.org. From the organization's description:
Gun Violence Archive (GVA) 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. GVA will collect and check for accuracy, comprehensive information about gun-related violence in the U.S. and then post and disseminate it online.

How did you get the data?

Because GVA limits the number of incidents that are returned from a single query, and because the website's "Export to CSV" functionality was missing crucial fields, it was necessary to obtain this dataset using web scraping techniques.
Stage 1: For each date between 1/1/2013 and 3/31/2018, a Python script queried all incidents that happened at that particular date, then scraped the data and wrote it to a CSV file. Each month got its own CSV file, with the exception of 2013, since not many incidents were recorded from then.
Stage 2: Each entry was augmented with additional data not directly viewable from the query results page, such as participant information, geolocation data, etc.
Stage 3: The entries were sorted in order of increasing date, then merged into a single CSV file.

3.4 Data Pre-processing

3.4.1 Data Cleaning

The rows "incident_url", "source_url", "incident_url_fields_missing", "sources", "state_house_district" were removed because it contained a lot of missing values and data which was not accessible.We have added 2 attributes month and year so that we can easily use the values reducing the computation.

3.5 Feature Computation

Direct Features:

These features are already present in the data set as attributes, so no computation is required as such.
Indirect Features:

We have computed only one indirect feature:
loss : This attribute is the sum of ['n_killed'] and ['n_injured'].It tells the total loss that we incurred due to the violence.

Chapter 4

Data Exploration

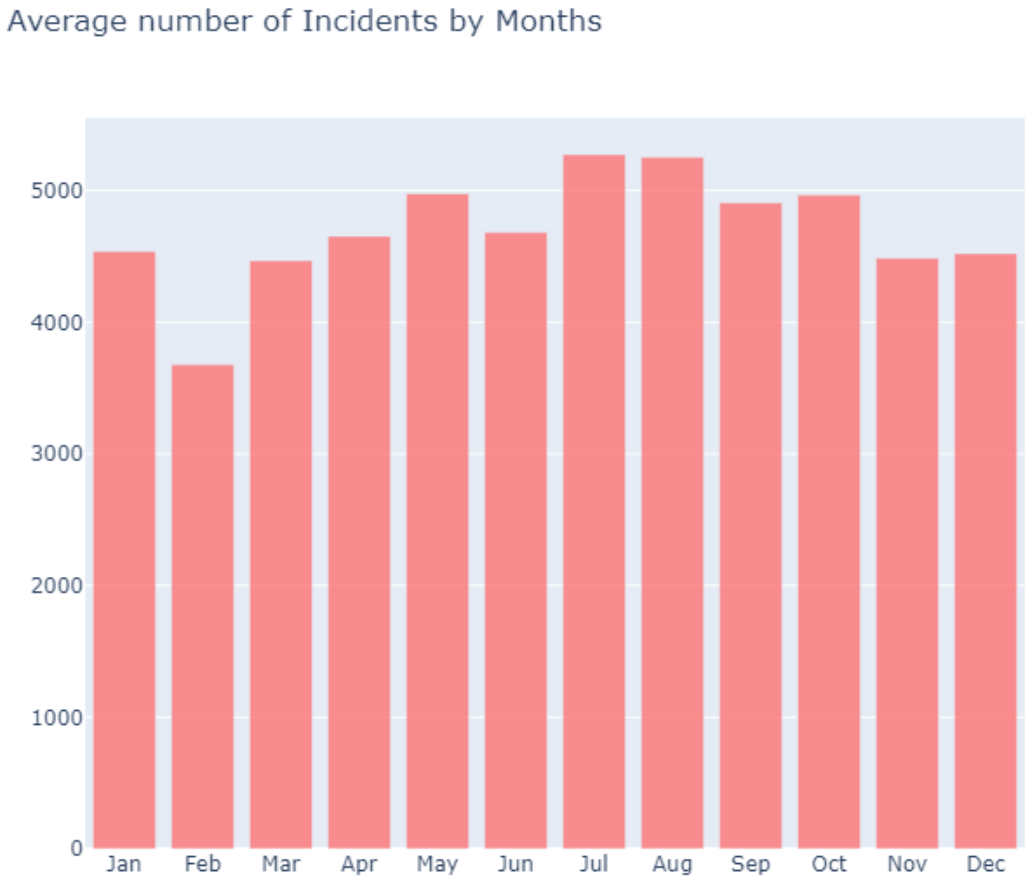
4.1 Visualizations

4.1.1 Trends by time

4.1.1.1 Incidents by year

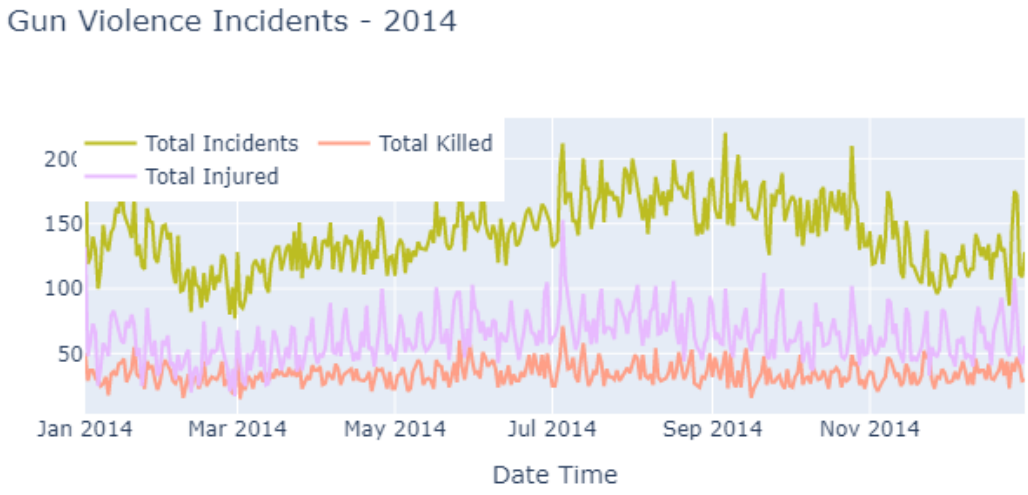
From the above graph, we can observe that the number of gun violence incidents are increasing every year.In the year 2014, there were about 51 thousand incidents reported, The number increased to 53 thousand in the next year, 2016 saw a bigger jump with close to 58 thousand incidents reported, in 2017, the number of gun violence incidents further increased to 61 thousand. A big increase of 10,000 incidents have been observed from 2014 to 2017.Year 2018's data is not complete.

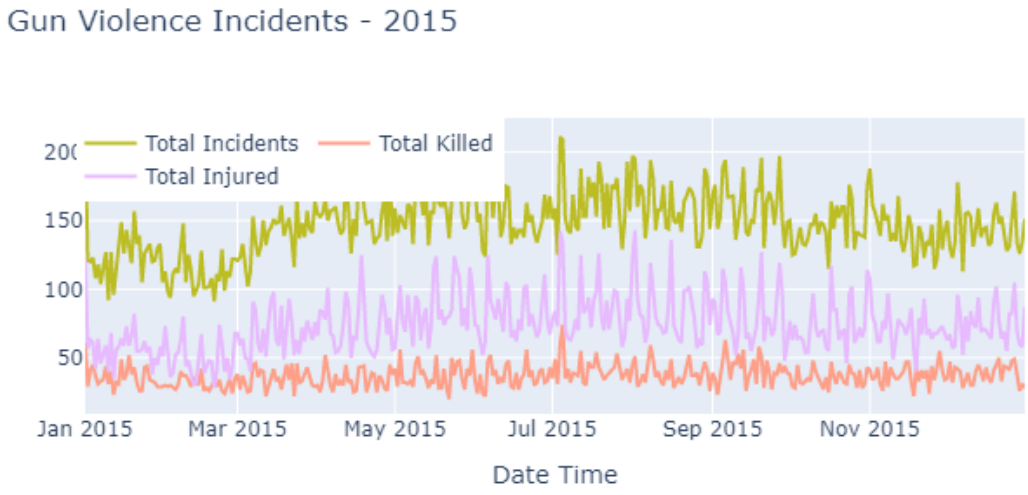
4.1.1.2 Incidents by months

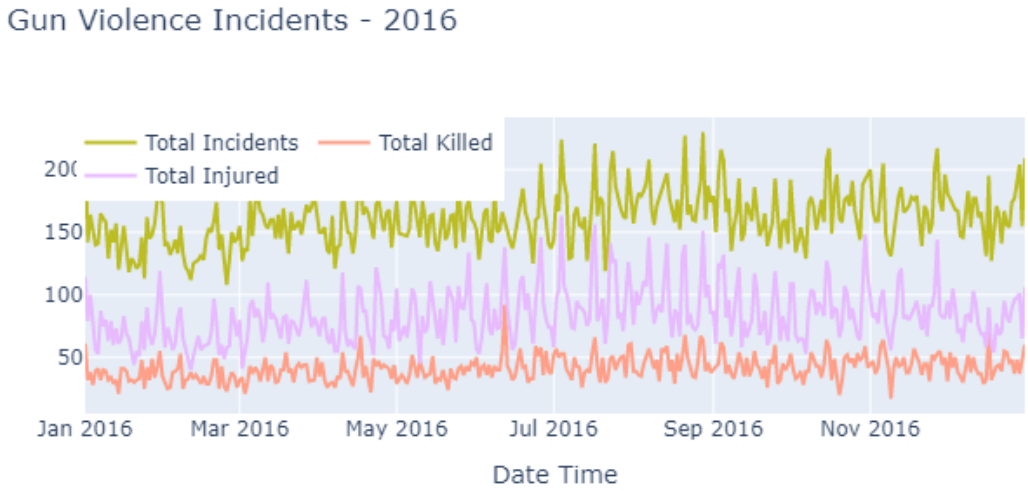


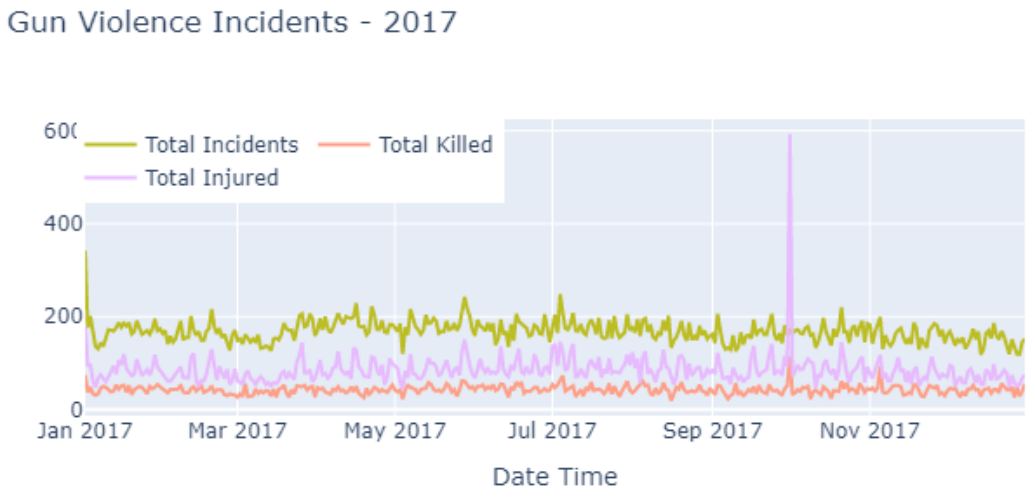
The analysis of quarters shows that more incidents occur in the warmer spring and summer periods. This seems worth diving into a little deeper. In order to compare months I will exclude 2018, as only the first three months have been recorded.The most visible 'seasonality' effect seems to me that the colder months seem to have less incidents. November, December, and February are the 3 months with the lowest number of incidents (February also only has 28 days of course).The exception seems to be January, which is worth investigating later on. My first idea is that possibly incidents on new years eve contribute to January having a high number of incidents.The other peak is the July/August period. I think that the fact that many people go on holidays in this period is the most likely explanation.

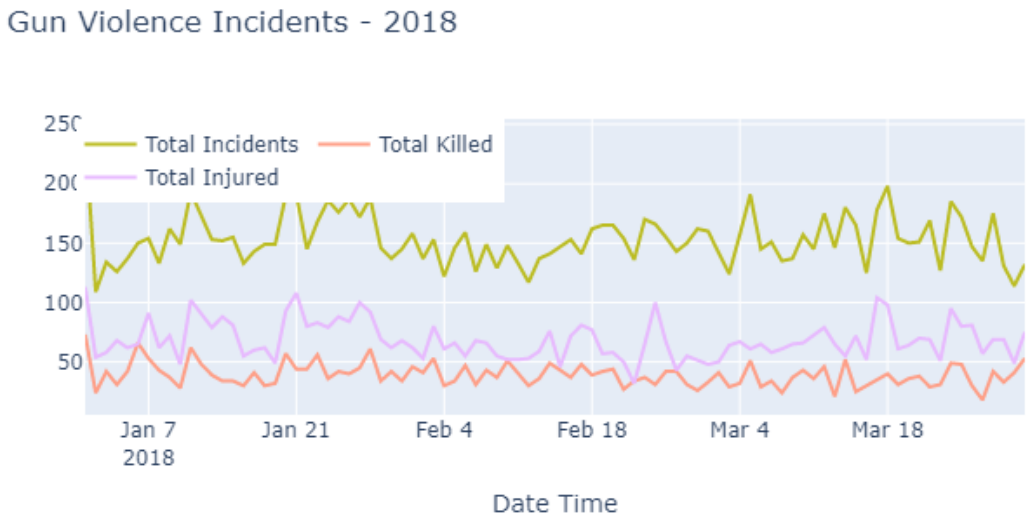
4.1.1.3 Time-Series of incidents by year





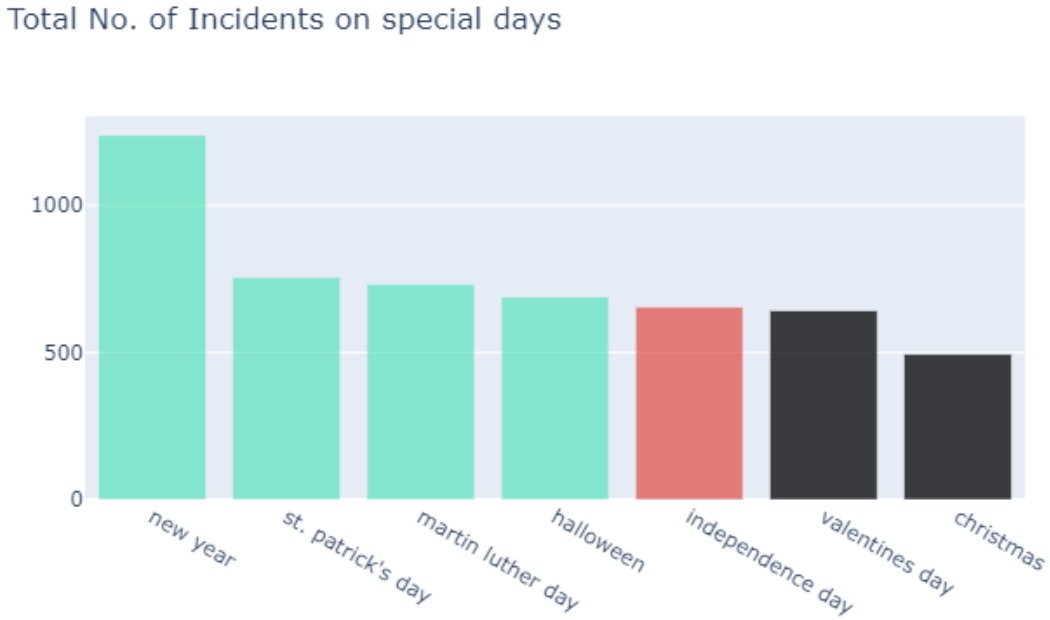






The analysis of the 4 time-series graphs(excluding 2018) shows that more incidents occur in the summer period.As per the observation made by the graphs above the summer period has the most incidents.The highest number of incidents took place in July which is the month of American independence, October (Halloween Time) and January which marks the New Year and some more holidays.

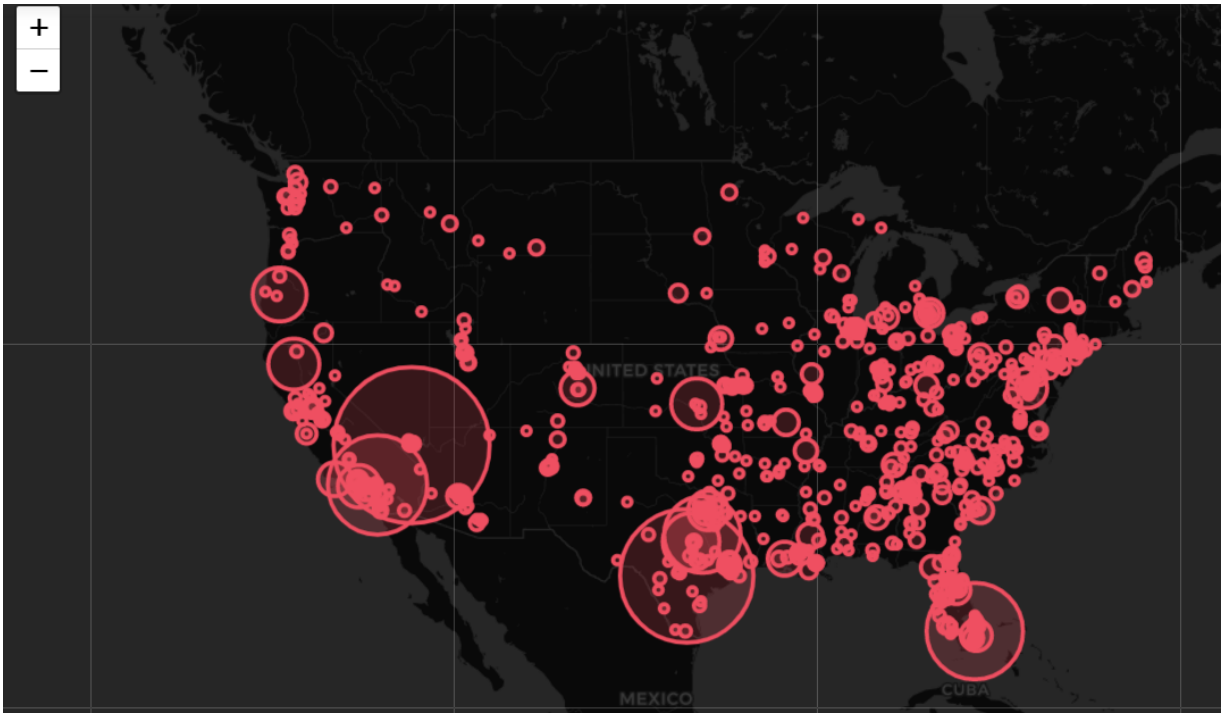
4.1.1.4 Incidents on special dates



From the above graph, we can observe that the number of gun violence incidents on New Year's are highest i.e, 1238, as compared to other special occasions reason behind this could be due to increased theft, the pressure to give gifts for the holidays often causes people to resort to theft. Lowest number of gun violence incidents occurs on Christmas Day i.e, 494, because Christmas is holy event for people and mostly people don't indulge in crimes on that day as per their culture.

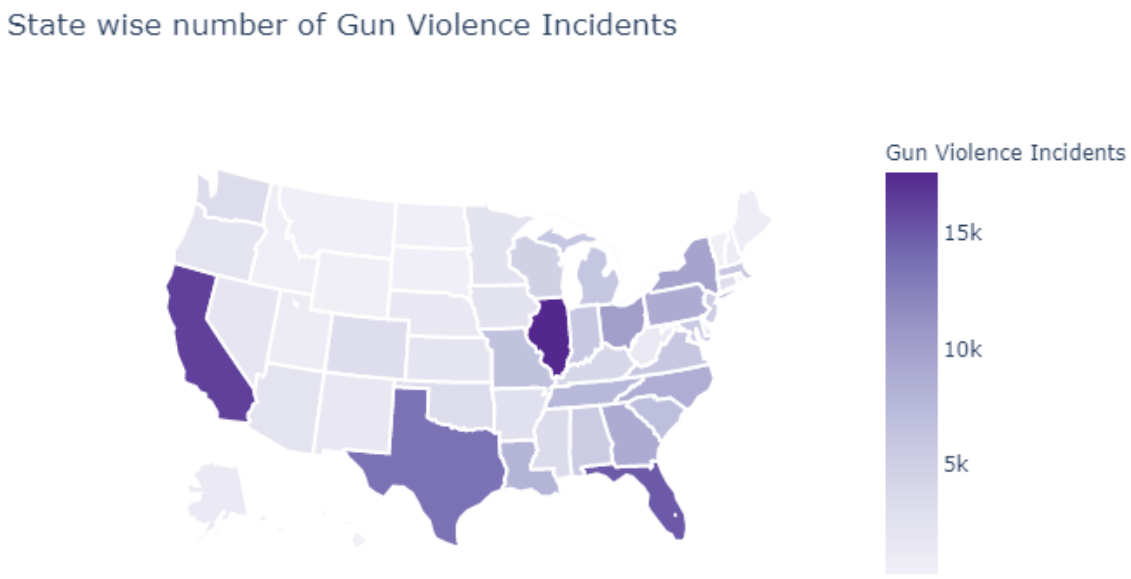
4.1.2 Trends by location

4.1.2.1 Incidents in US



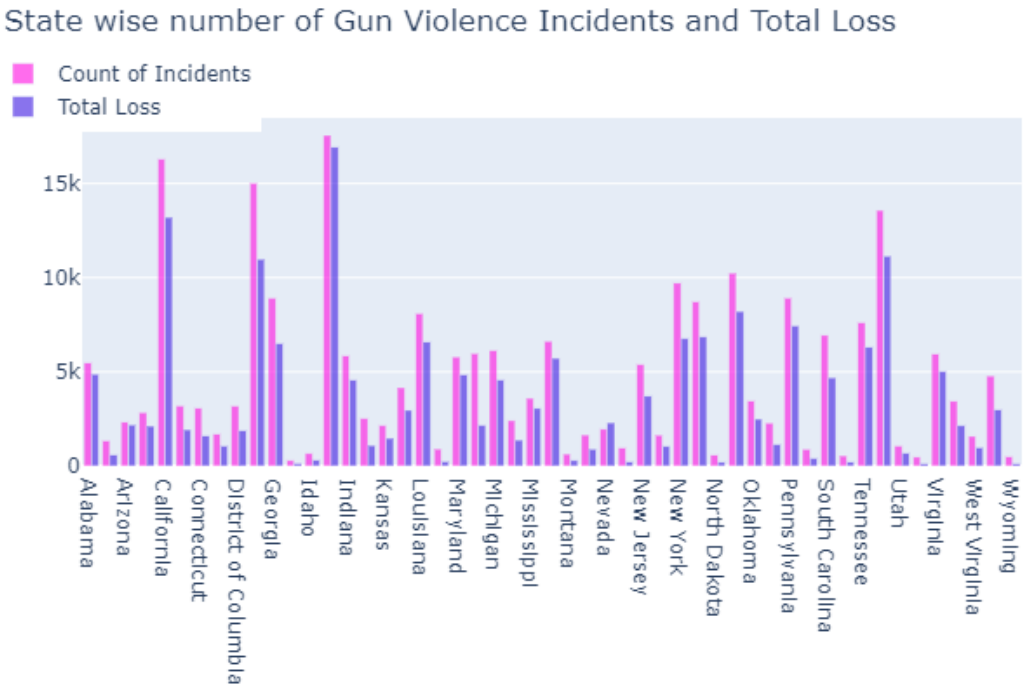
The image describes the number of incidents occurring in US represented on the map of US with circles which portray the casualty(refers to number of people killed and injured) of incident. The bigger the circle , higher is the casualty and vice versa. As it can be observed that region of California , Florida , North-west of Texas have larger circles with maximum casualty ,whereas other regions have comparatively less casualty.

4.1.2.2 Incidents by state



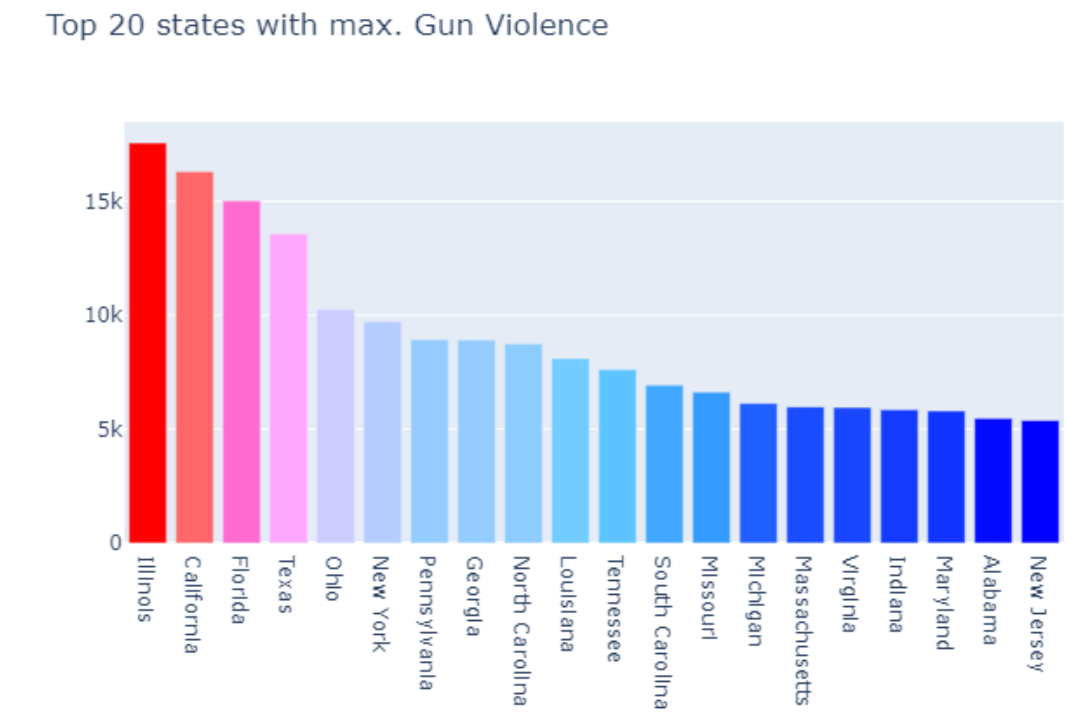
The above map depicts the number of Gun Violence incidents in different states of US. The colour bar on the right gives a better understanding of incident count, in decreasing order from top to bottom . From the graph ,it can be observed that the state of California, Indiana, Illinois have maximum crime count above 15,000.The state of Florida and Texas are no less with crime count of nearly 15,000. Whereas, the state of New Hampshire, Vermont , Connecticut , Alaska , Montana , Wyoming , North Dakota have low crime count below 5,000 (which according to Interpol is not a good record but relatively better compared to other states of America.) Rest of the states of US have crime count ranging from 5,000 to 15,000. The reason behind maximum incident count in California, Indiana, Illinois could be because of cities like Los Angeles , San Francisco , San Diego , Oakland , Chicago , Peoria, Rockford , etc which are some popular tourist destinations along with high population density therefore inviting more crime , whereas states with less crime count could be because of less population and decent lifestyle.

4.1.2.3 Incidents with loss by state

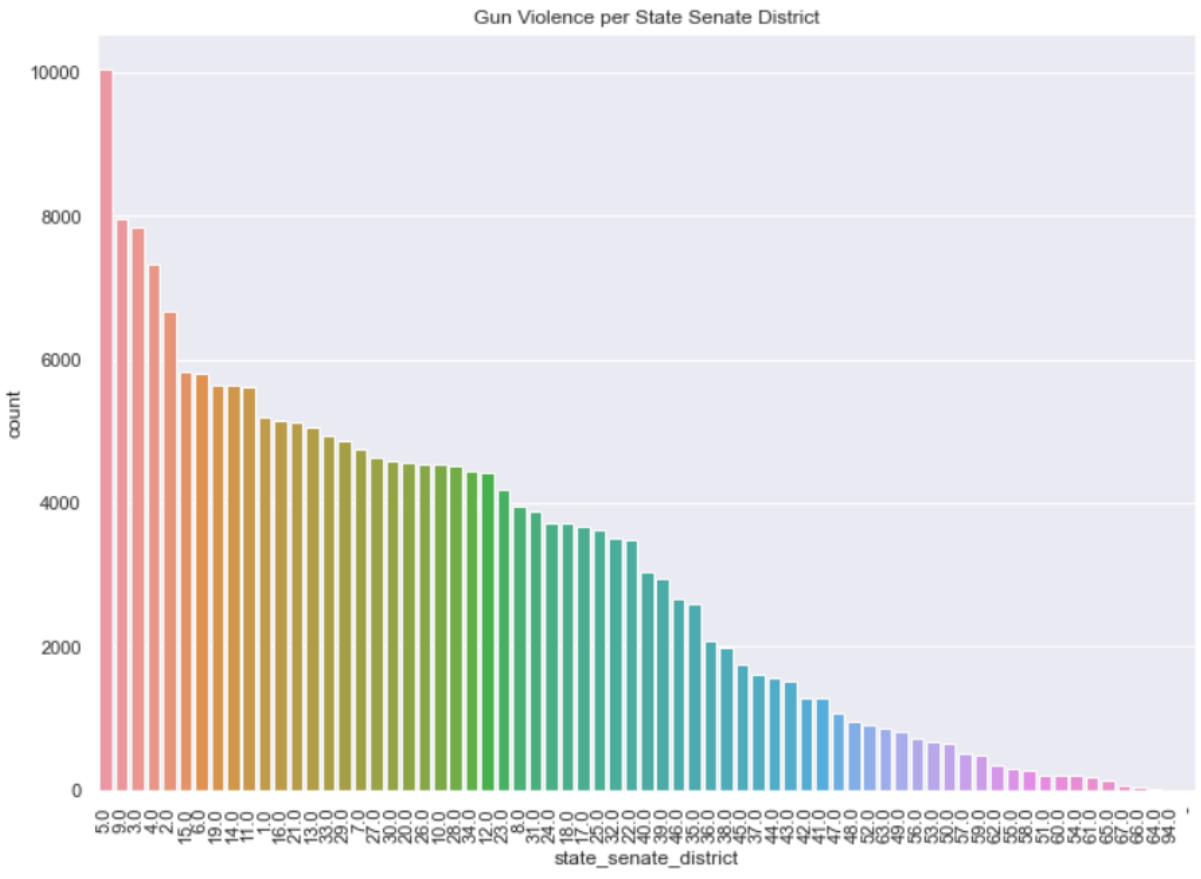


The graph depicts the number of incidents and loss count (loss count is number of people killed and injured) per state of US . From the graph it can be observed that the state of Indiana has maximum number of incidents and loss count ranging above 15,000. Then is state of California with incident count above 15,000 and loss count between 15,000 and 10,000 . The state with minimum number of incidents and loss count not more than 500 is Idaho. This graph helps us in understanding the casualty of incident.

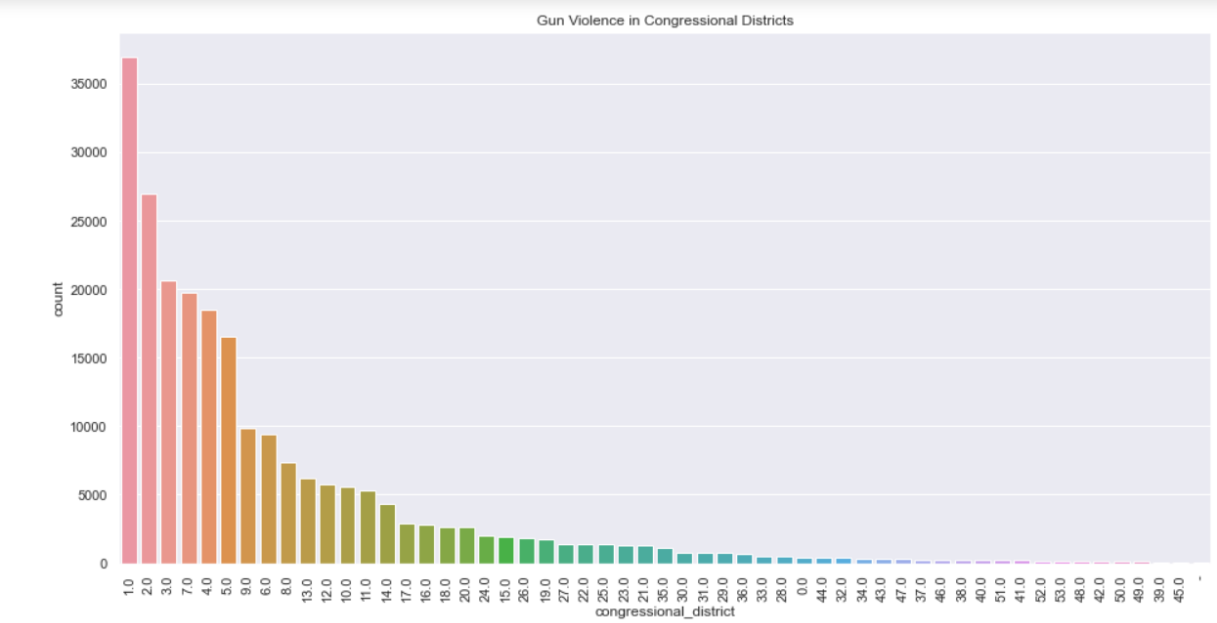
4.1.2.4 Top 20 states with highest gun violence



4.1.2.5 Incidents by state senate district

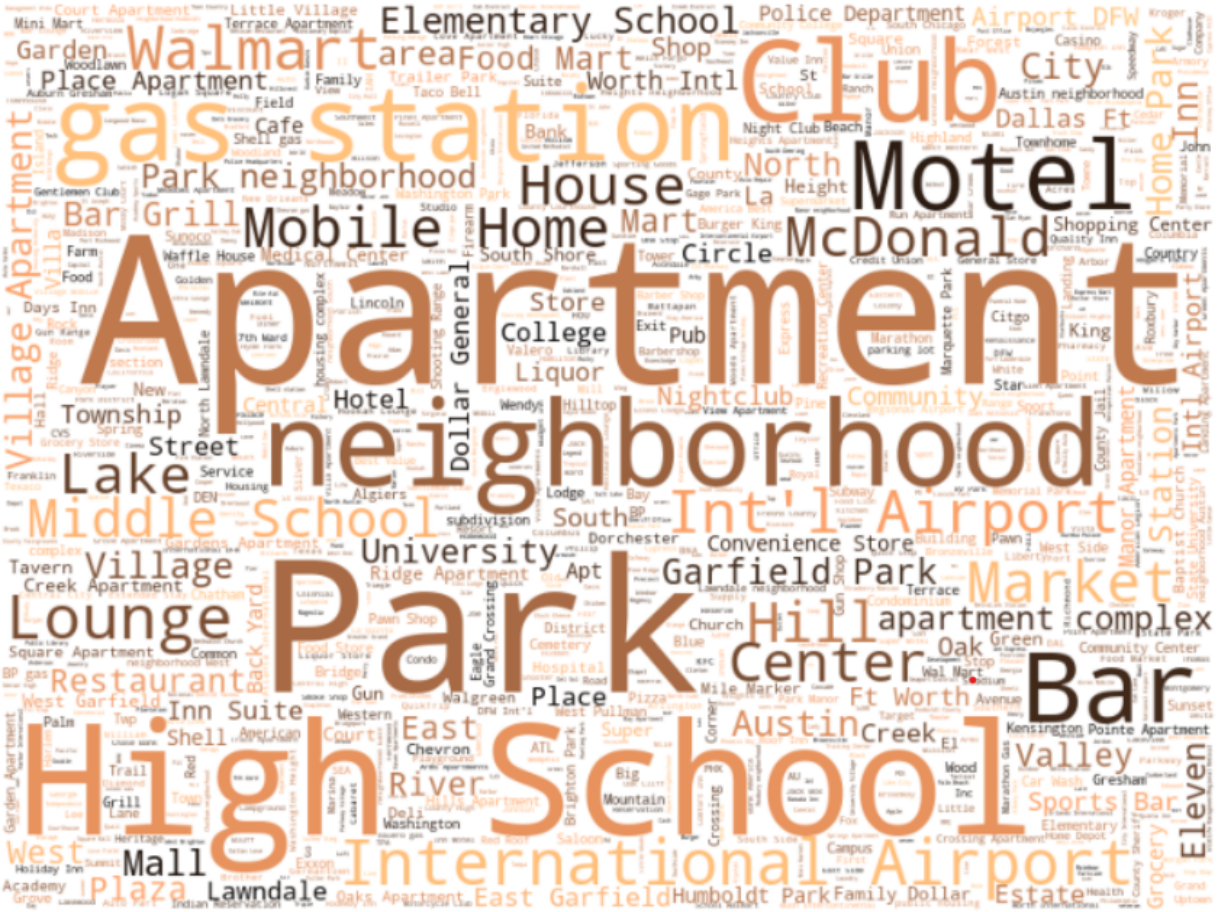


4.1.2.6 Incidents by Congressional district



The figure shows the incidents in the state senate districts of USA.Exploring data in smaller areas like districts can help solve the problem of gun violence in a better way. We can go to that places and keep a check on areas with higher number of cases. We can analyse the causes of such violence by doing some surveys, analysing CCTV footage etc.

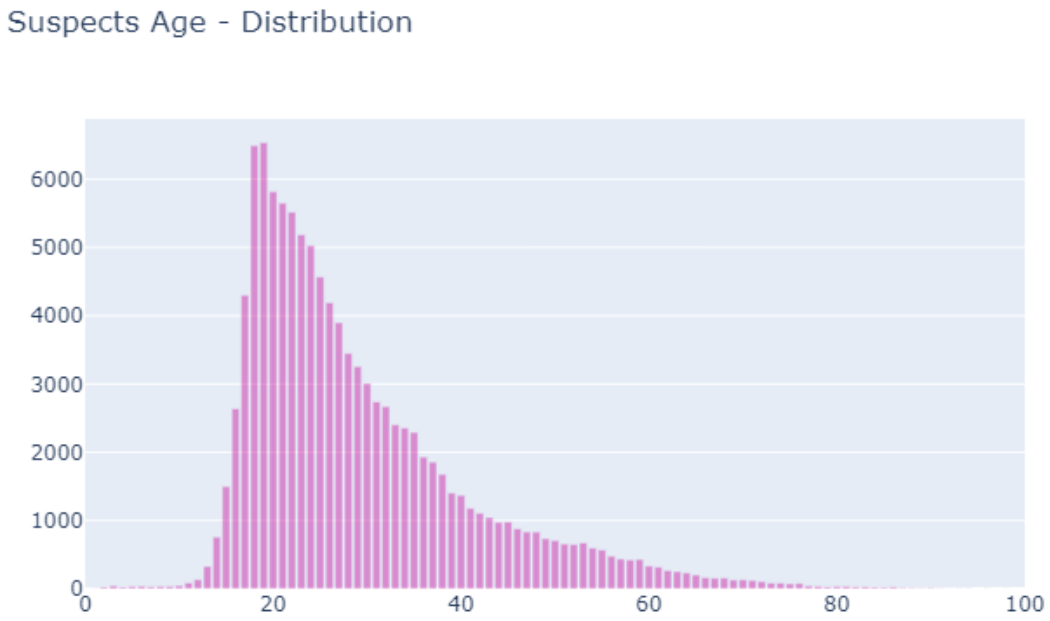
4.1.2.7 Where did most Gun Incidents took place?



As the picture above shows, mostly gun incidents occurs in Apartments, Parks and High Schools. It can be because in these places there is the most probability where young people gather and incidents mostly occur in this age group. Then at Clubs, Motels, Bars people are mostly drunk and thus gun violence incidents can occur.

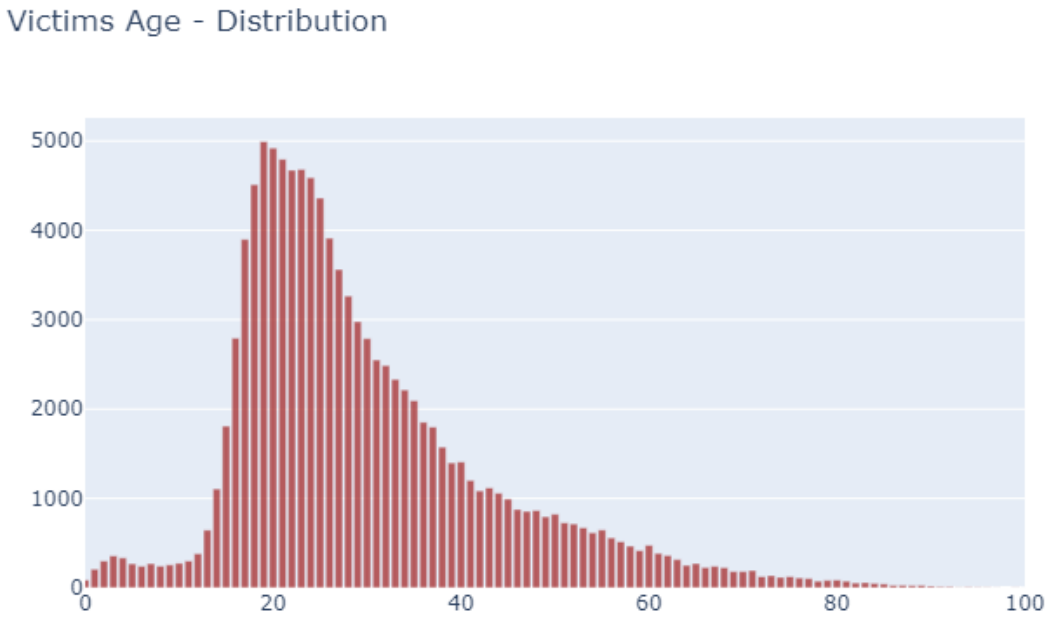
4.1.3 Trends by age

4.1.3.1 Prominent age of gun violence suspects



The above image shows the common age of the suspects of the violence. Its a sad insight that major suspects are in the group 17- 27. The major reason for this might be the urge for the youth to earn easy money, anger issues, lack of elder guidance, bad influence or drugs. It will be interesting to do another report on the reasons why young people are more involved in crimes.

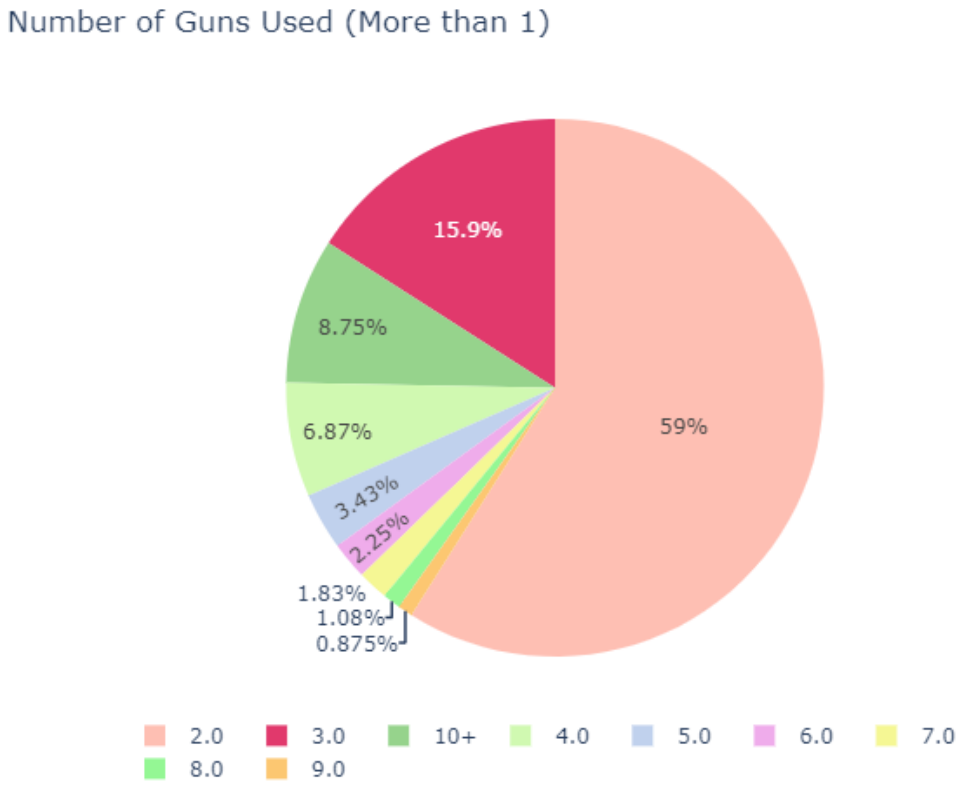
4.1.3.2 Prominent age of gun violence victims



The victims being in the same age groups as suspects is an interesting insight. Younger and older groups are easy targets still the victims lie in the younger age group. The reason might be personal grudges.

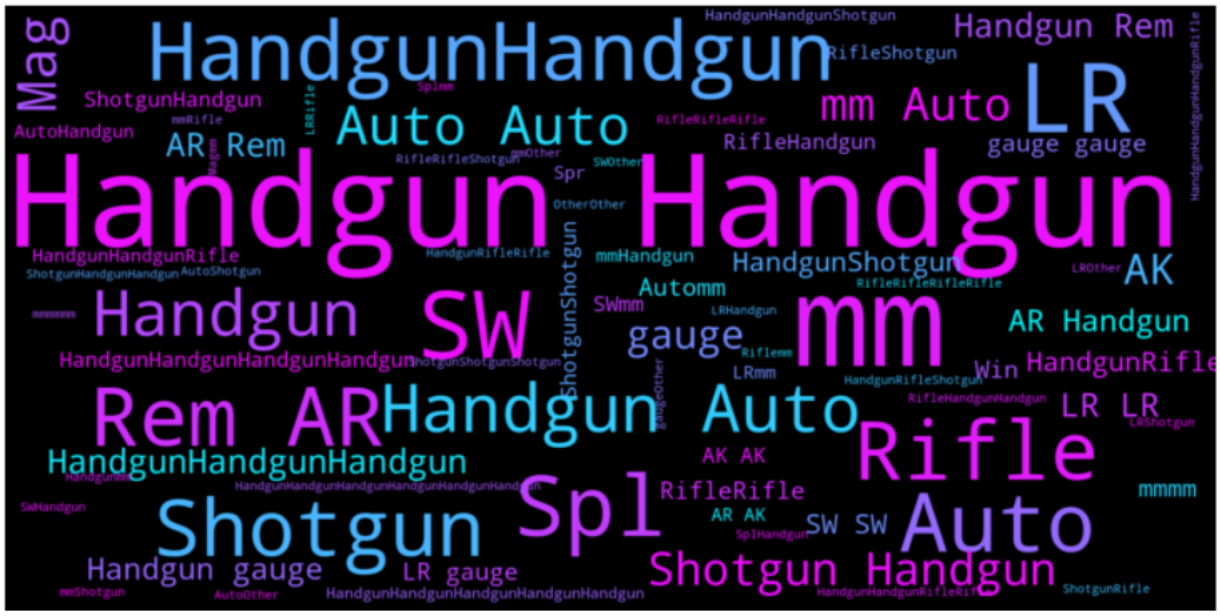
4.1.4 Other Characteristics

4.1.4.1 Number of guns used in violence



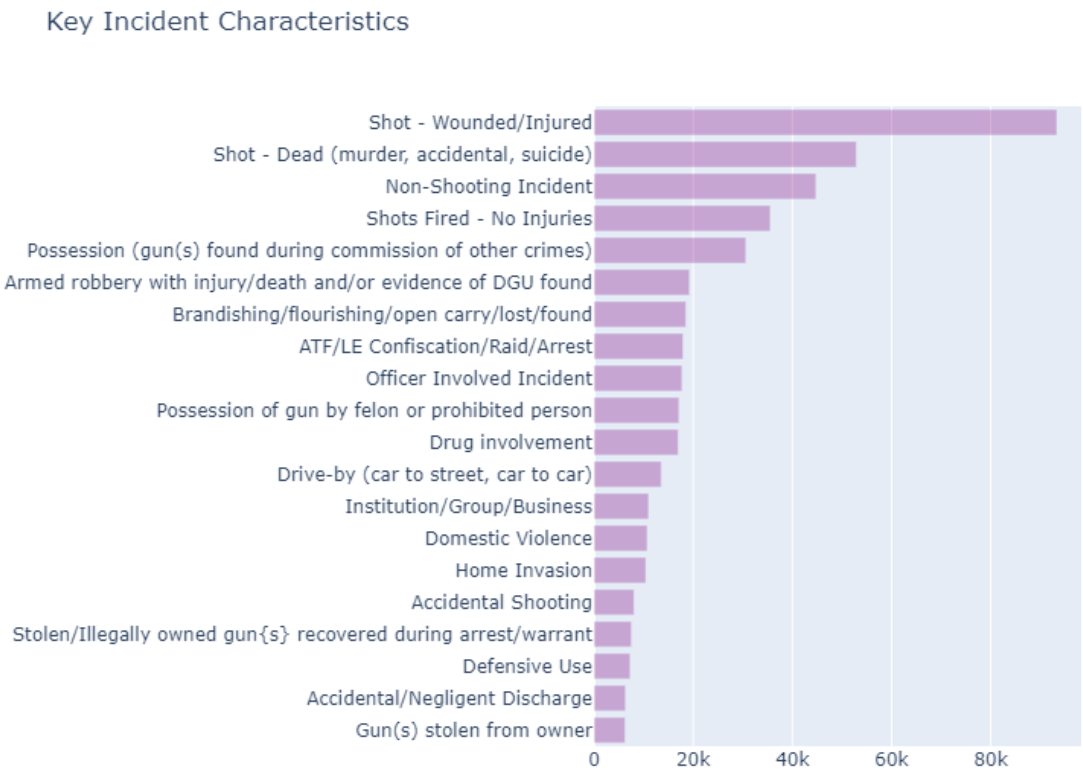
According to the pie chart shown above, it is observed that mostly gun violence incidents occurs with carrying of two guns by the participants. There are also some incidents that occur via nine guns. So, it describes that mostly incidents occur via two or three guns. This might mean that the incidents were mostly due to personal reasons and not gangs involved with it. On reading reports the suspects of most of the gun incidents are gangs but the statistics show something else.

4.1.4.2 Types of gun used

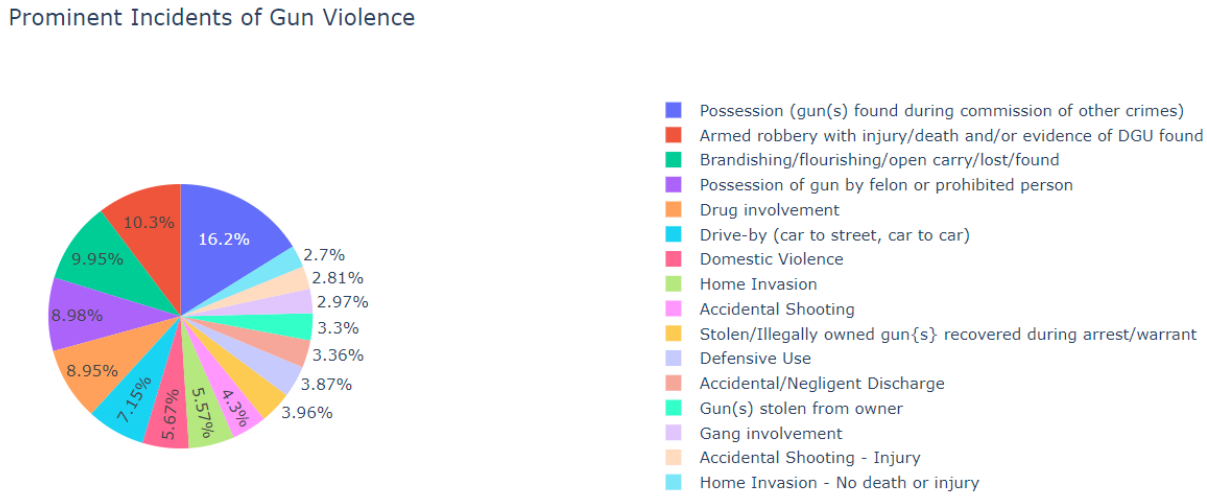


4.1.4.3 Characteristics of the incidents

The above image shows the common characteristics of the incidents that took place. In the incidents most of the people were shot wounded, murdered, shot dead, injured shot. Most of the incidents took place because of the motive of possession, suicide, murder. Some incidents were accidental as well where the motive might be to threaten or injure the victim.



4.1.4.4 Prominent incidents



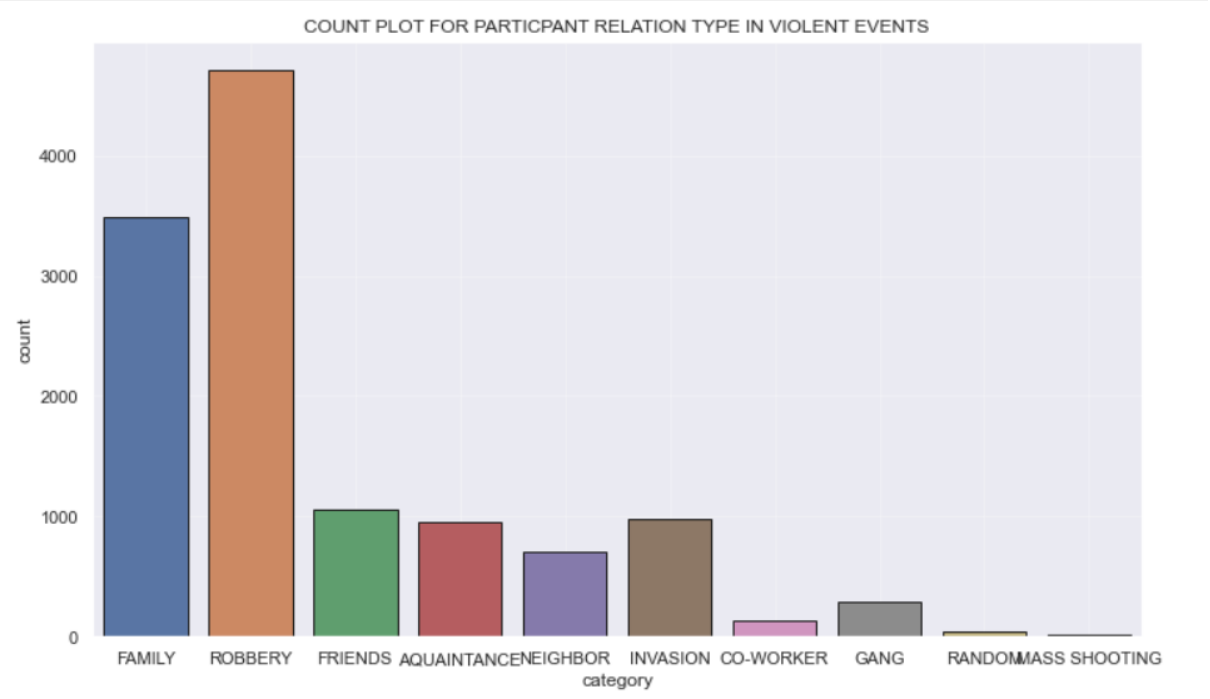
The above image shows the most prominent incidents that took place.

4.1.4.5 10 Serious Incidents

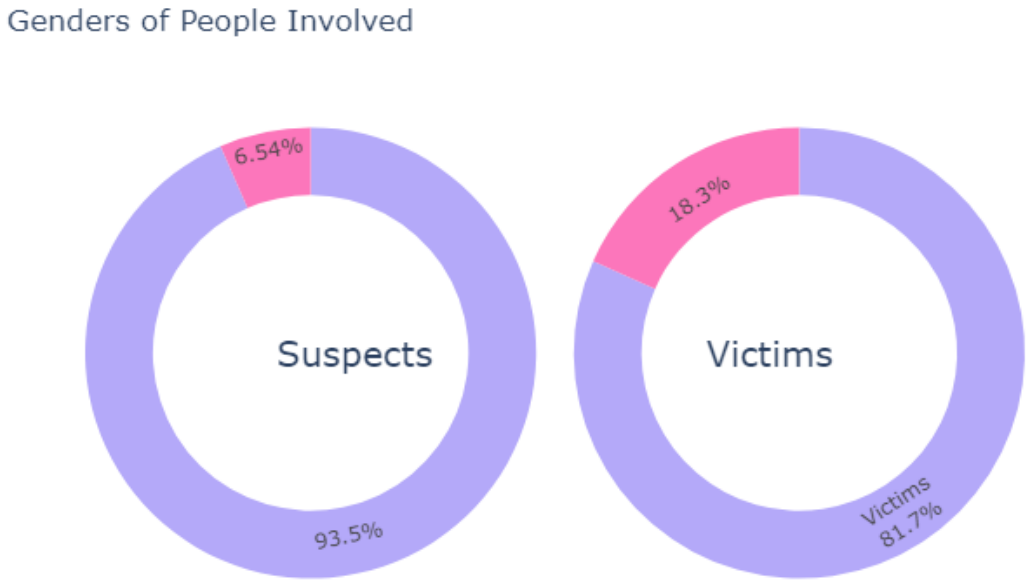
	date	state	city_or_county	address	n_killed	n_injured
239677	2017-10-01	Nevada	Las Vegas	Mandalay Bay 3950 Blvd S	59	489
130448	2016-06-12	Florida	Orlando	1912 S Orange Avenue	50	53
217151	2017-11-05	Texas	Sutherland Springs	216 4th St	27	20
101531	2015-12-02	California	San Bernardino	1365 South Waterman Avenue	16	19
232745	2018-02-14	Florida	Pompano Beach (Parkland)	5901 Pine Island Rd	17	17
70511	2015-05-17	Texas	Waco	4671 S Jack Kultgen Fwy	9	18
195845	2017-07-01	Arkansas	Little Rock	220 W 6th St	0	25
137328	2016-07-25	Florida	Fort Myers	3580 Evans Ave	2	19
11566	2014-04-02	Texas	Fort Hood	Motor Pool Road and Tank Destroyer Boulevard	4	16
92624	2015-10-01	Oregon	Roseburg	1140 Umpqua College Rd	10	9

According to the table shown, the most dangerous and lethal incident was held on 01-10-2017 in Las Vegas where 59 people were killed and 489 people were injured. Mostly harmful incidents occur in 2017. The second most deadly incident occurred on 12-06-2016 in Florida, Orlando where 53 people were killed and 53 people were injured.

4.1.4.6 Relationship of participants



4.1.4.7 Gender groups



It is interesting to get to know that most of the victims are males only. The reasons might be many. A further survey on the cases might help us know the motives behind the crimes and what led the criminals to do such a thing.

Chapter 5

Forecasting Methods

5.1 FB Prophet

5.1.1 Description

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

It is open source software released by Facebook's Core Data Science team. It follows the sklearn model API. We create an instance of the Prophet class and then call its fit and predict methods.

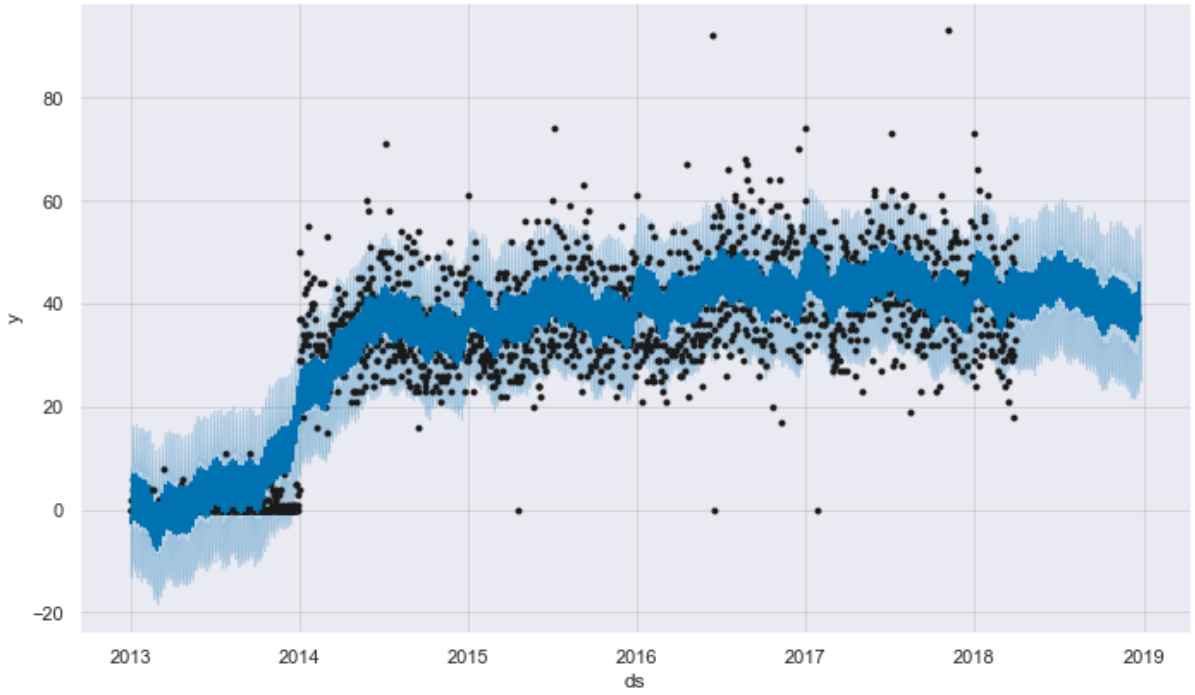
With Prophet, you are not stuck with the results of a completely automatic procedure if the forecast is not satisfactory- an analyst with no training in time series methods can improve or tweak forecasts using a variety of easily-interpretative parameters.

5.1.2 Graphical Representation Of The Results

The input to Prophet is always a dataframe with two columns: ds and y. The ds (datestamp) column should be of a format expected by Pandas, ideally YYYY-MM-DD for a date or YYYY-MM-DD HH:MM:SS for a timestamp. The y column must be numeric, and represents the measurement we wish to forecast.

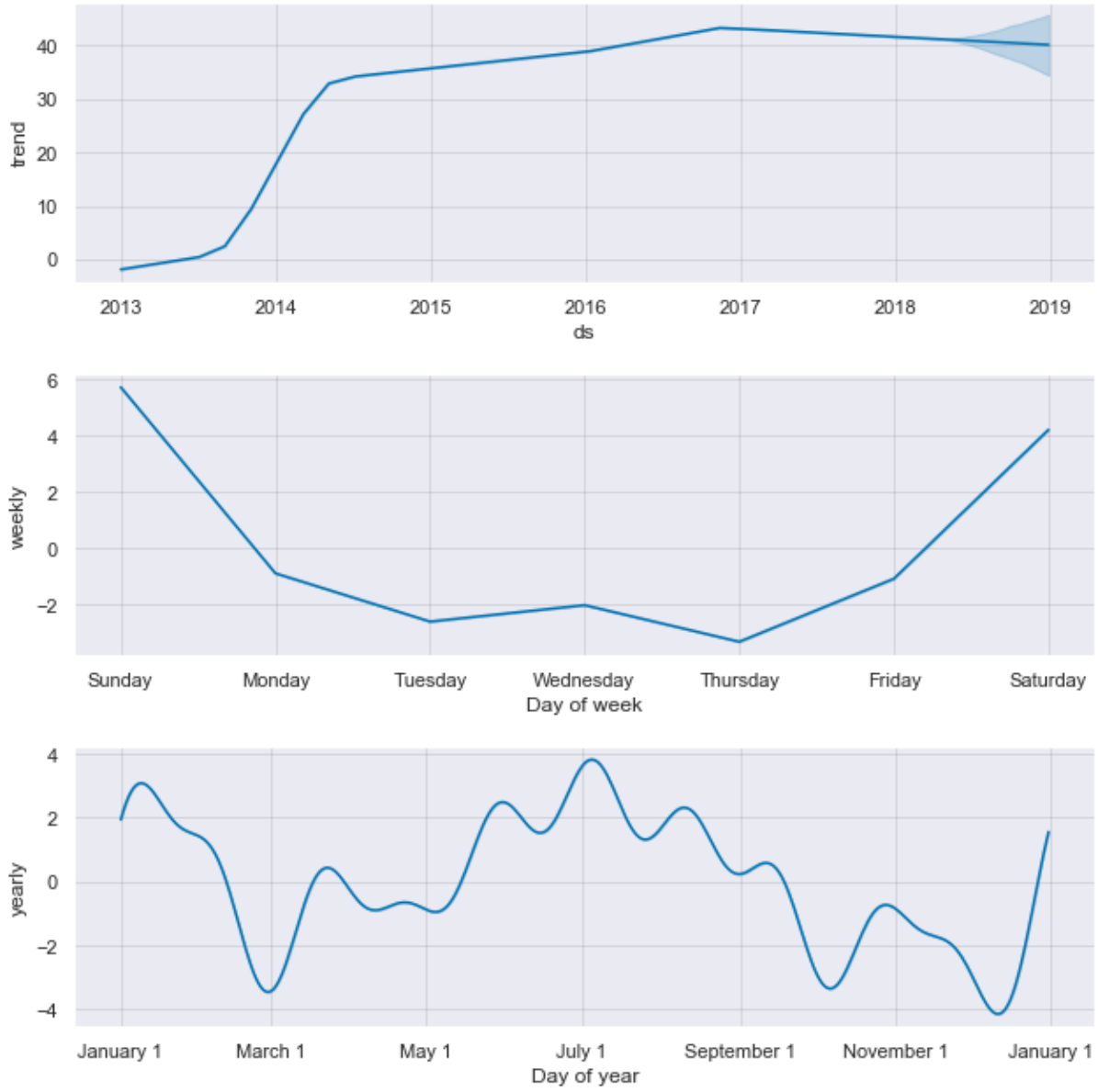
We fit the model by instantiating a new Prophet object. Any settings to the forecasting procedure are passed into the constructor. Then you call its fit method and pass in the dataframe.
Predictions are then made on a dataframe with a column ds containing the 'dates' for which a prediction is to be made. We have made a suitable dataframe that extends into the future a specified number of days(i.e. 270 periods) using the helper method Prophet.make_future_dataframe.
The predict method will assign each row in future a predicted value which it names 'n_killings'. This will forecast the number of killings that occured in different gun violence cases on a particular day.

Figure 5.1: Generic Plot of killings by years



You can plot the forecast by calling the Prophet.plot method and passing in your forecast dataframe.

Figure 5.2: Forecast Trend



To see the forecast components, we have used the Prophet.plot_components method. By default you'll see the trend, yearly seasonality, and weekly seasonality of the time series.

5.1.3 Conclusion

The graphs show the predicted trend and it also tells us about the predicted killing weekly and yearly. The generic plot of killings gave us a graph for the whole dataset and predicted the killings till 2019. The trend of 2013 is downwards as the dataset has only some values for 2013.