

Gun Violence Analysis and Prediction Using Logistic Regression

By

Namra Maniar – 20MIS1054

A Project Report

submitted to

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

In partial fulfilment of the requirements for the course of

BIG DATA ANALYTICS

In

Int. MTECH SOFTWARE ENGINEERING

Vandalur – Kelambakkam Road

Chennai – 600127

November 2022

**Abstract:** Gun violence analysis and Prediction could be a systematic approach for distinguishing and analysing patterns and trends in crime. Our system will analyse regions with a high chance of gun violence and visualize gun violence prone areas and try to predict it using the machine learning algorithm Logistic Regression. With the increasing advent of computerized systems, crime knowledge analysts can facilitate enforcement officers to hurry up the method of finding and preventing crimes. rather than that specialize in causes of crime incidence just like the criminal background of the offender, political hostility, etc. we have a tendency to are focusing primarily on crime factors of every day.

**Keywords:** gun violence, machine learning, logistic regression, algorithm.

**Introduction**: Gun violence in the US results in tens of thousands of deaths and injuries annually. In 2013, there were 73,505 nonfatal firearm injuries which included 11,208 homicides, 21,175 suicides, 505 deaths due to accidental or negligent discharge of a firearm, and 281 deaths due to firearms use with "undetermined intent". In this project, I have performed a deep exploration on this Dataset of gun violence incidents reported in US from 2013 – 2018 and tries to build a machine learning model that helps in predicting whether the suspect is guilty or not. This model is build using Logistic Algorithm. However, the performance of these algorithms is affected due to various conditions such as no injured and no killed victim are there.

Is the past few years, new theories and algorithms have made deep learning increasingly dominant in the field of prediction.

**Tools Used:** For the detailed study and inferring the results about the data and the things we have done. The language which was used to infer the results was Python. We could have used any other language but we used this because this allows us to transform and gather thousands of columns and rows at a time and extract useful insights from them.

I used Jupyter Notebook for this project because it is quite powerful and allows us to write and execute arbitrary python code through the browser and it is very well suited to Machine Learning and Data Science and may libraries be preinstalled and we need to put much of the RAM into use. As for dataset, I used the real time gun violence data happens in USA from 2013 – 2018.

**Libraries Used:**

*NumPy* - This library aims to provide an array object which is up to 50 times faster than the traditional Python lists. It provides a lot of supporting functions that make the working on Python easy.

*Pandas* - This library is a fast, powerful, flexible and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.

Here, we used it for mathematical functions mainly to derive the matrix.

*Matplotlib* - This is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits.

*Seaborn* - This is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

Here, it used for plotting the analyzed data using bar graphs.

*Folium* - This library folium makes it easy to visualize data that’s been manipulated in Python on an interactive leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map.

Here, it is used to plot location related trends on map of USA.

*SK Learn* – It is a machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support-vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Here, this library has been used for making a predictive model.

**Process Involved:**

I took the dataset and imported it into jupyter notebook file. After importing it into notebook I observed many of the columns are not required. So, we have to drop all of them. Some of the data are not in appropriate order so I distributed it into different columns.

Firstly, I focused on analyzing the data after observing some trends in dataset. Trends Like Time Related Trends, Location Related Trends, etc.

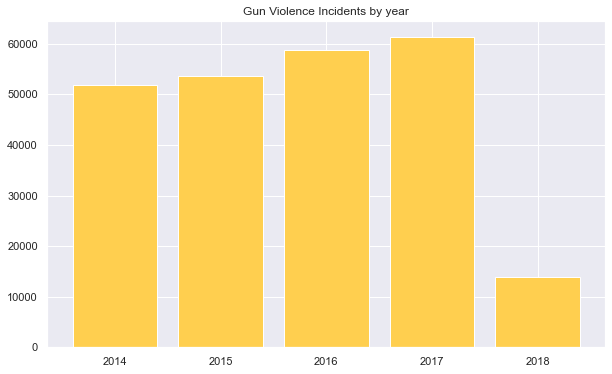
Then, I created a different data frame which has the data of no. of killed and no. of injured. Then according to that I created an algorithm to judge whether the suspect of that case is guilty or not.

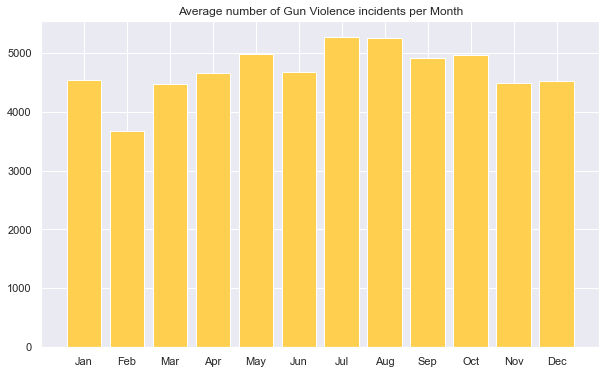
After I got the results, Logistic Regression was preferred to use to make a prediction algorithm.

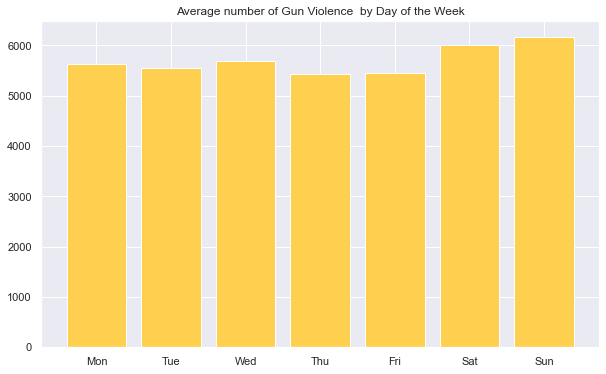
**Results:**

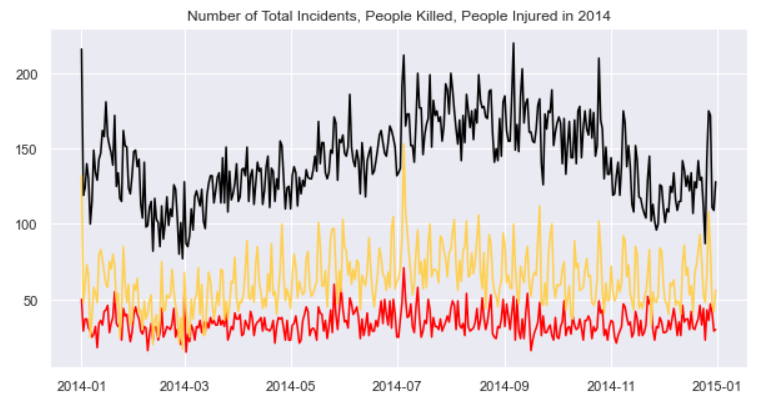
*Analysis Work:*

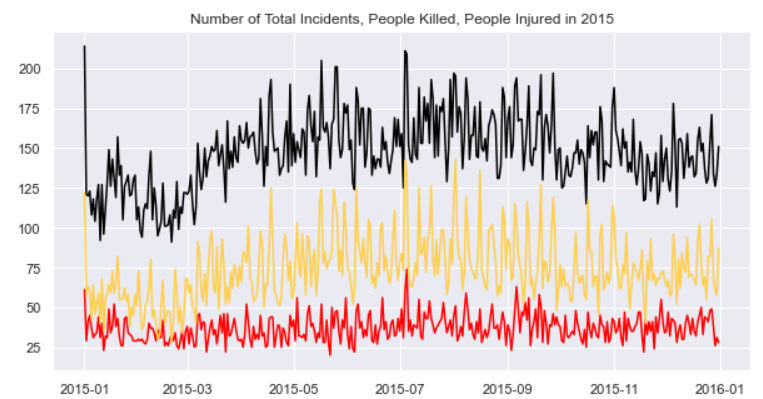
1. Exploring Time Related Trends:

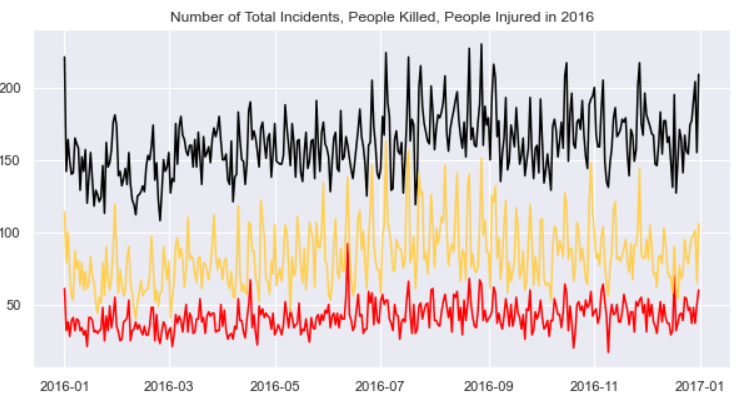


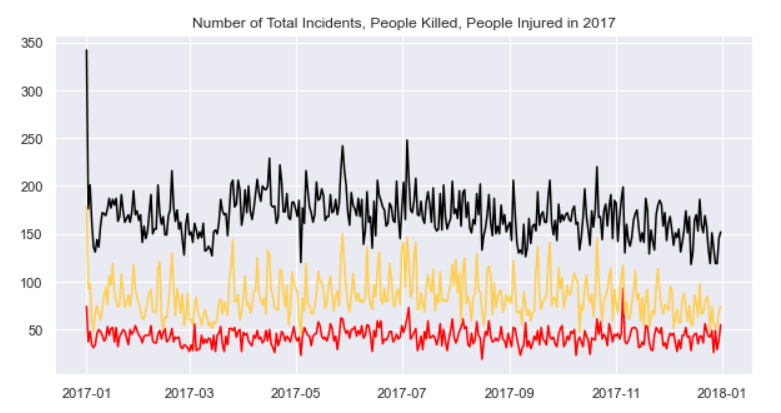




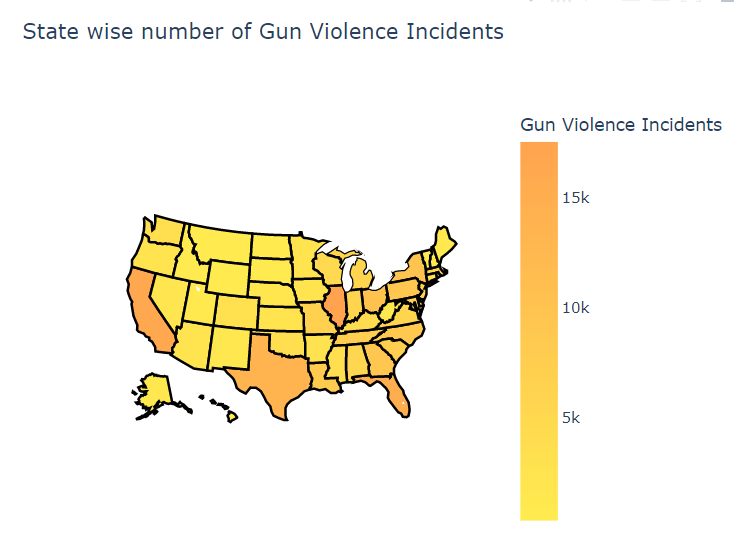


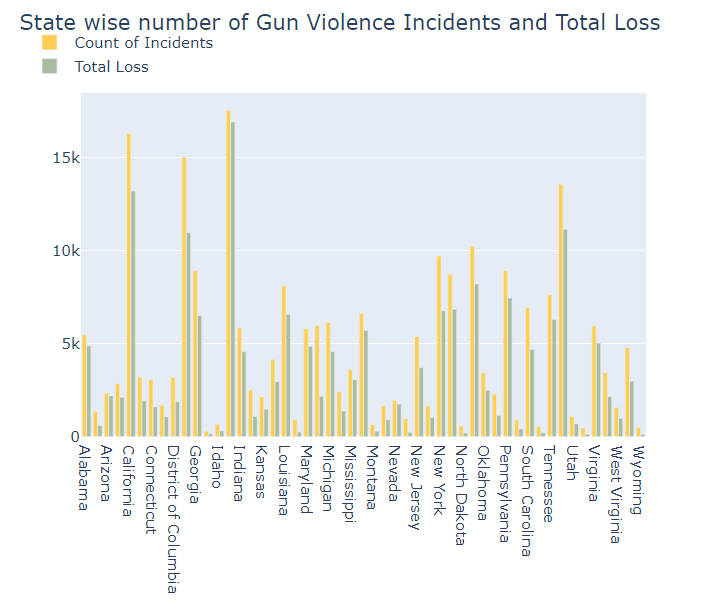


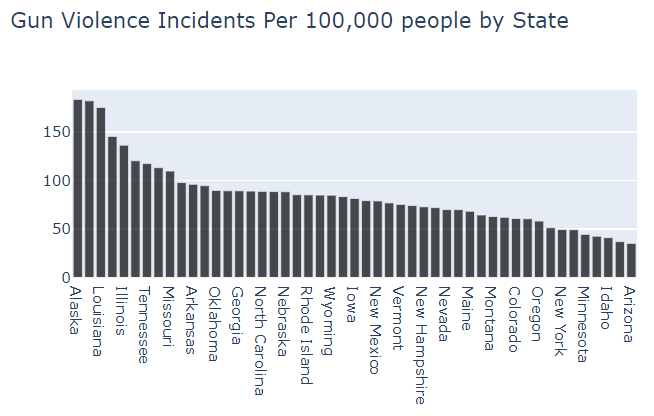


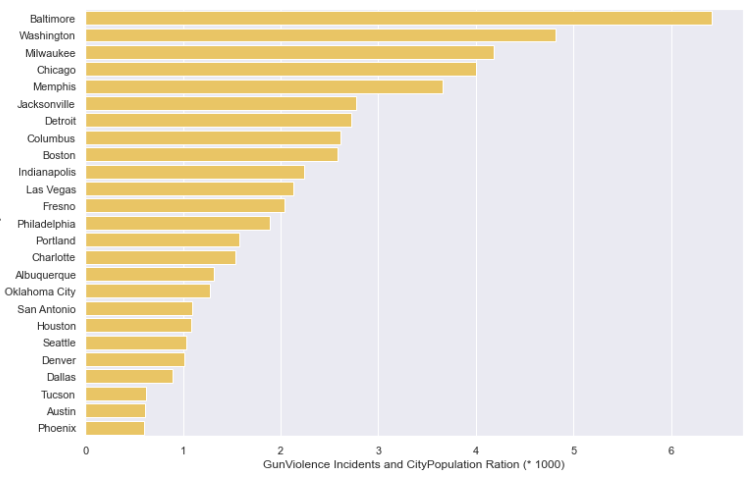


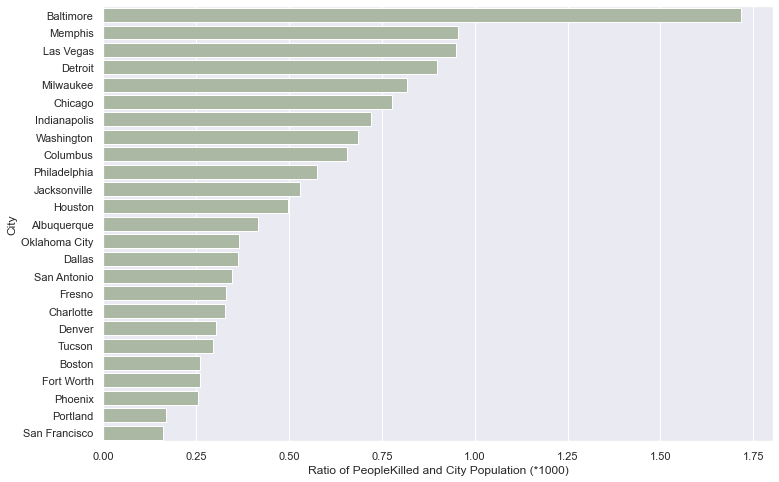
1. Exploring Location Related Trends:

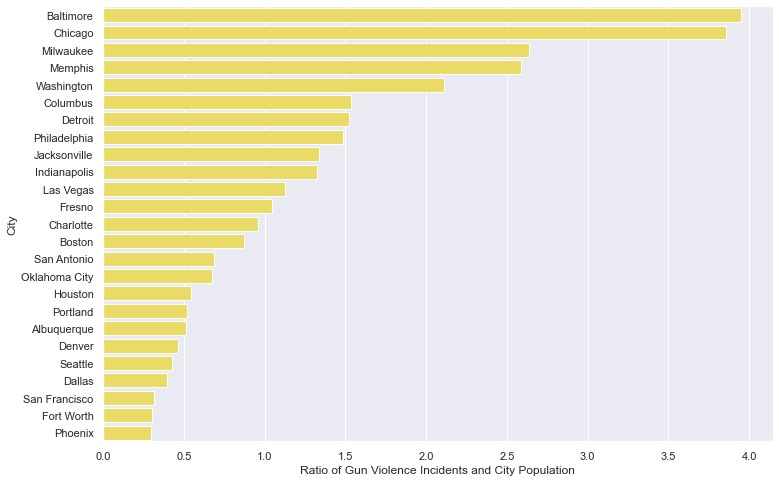




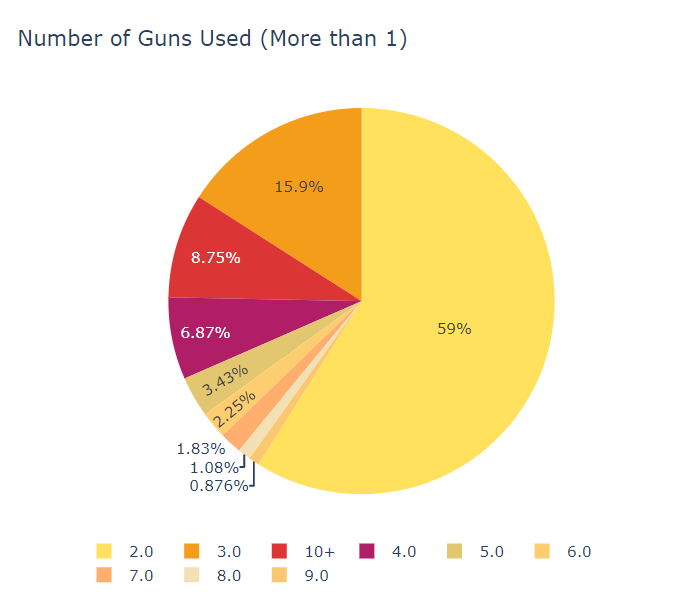








1. Other Characteristics of Gun Violence Incidents.



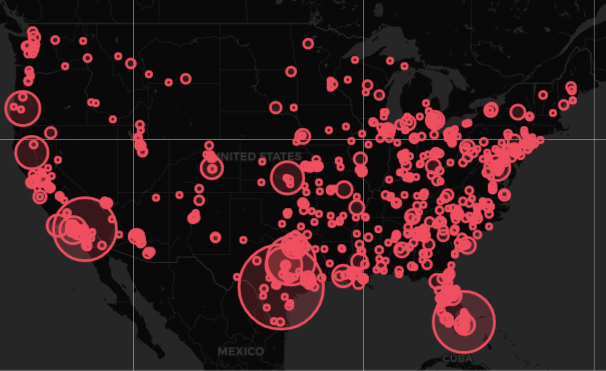
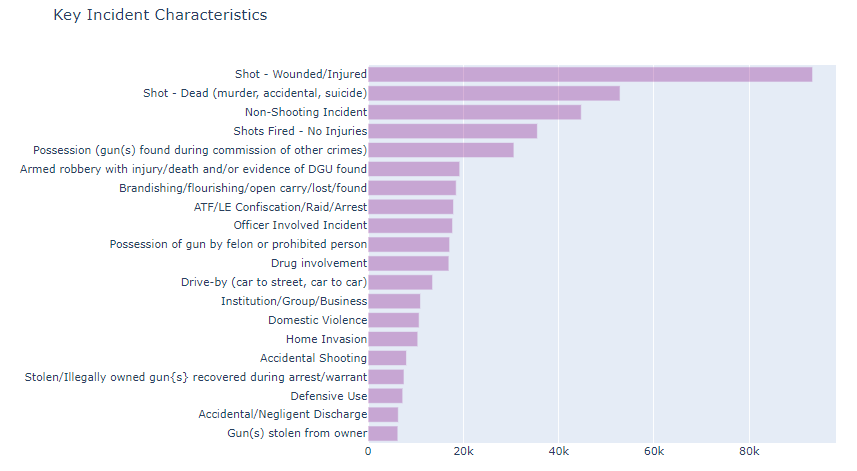
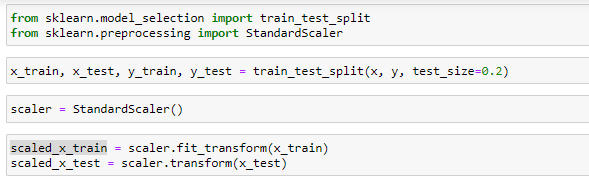


Fig: Most Serious Gun Violence Incidents in USA



*Prediction Result:*

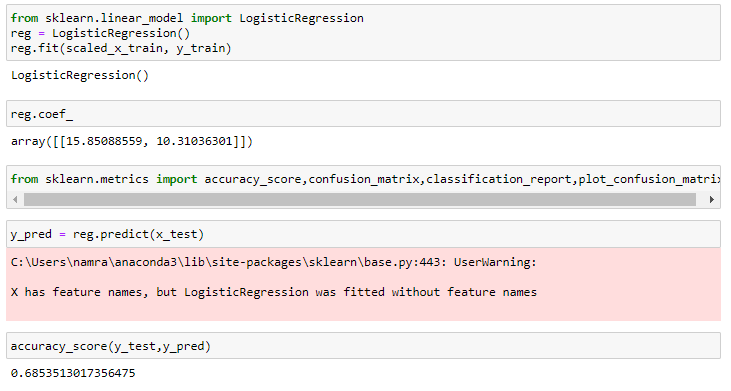
For Predicting whether suspect of that case is guilty or not I divided the prediction data frame into train and test data.



Then importing the Logistic Regression Model using Sk Learn library.

Then fit the scaled train data into Logistic Regression model.

Predict the best Y axis data using X\_Train data.



After Predicting the Accuracy Score acquired is **0.685351.**

**References:**

* <https://pandas.pydata.org/>
* <https://matplotlib.org/>
* <https://www.kaggle.com/code/shivamb/deep-exploration-of-gun-violence-in-us>
* <https://www.kaggle.com/datasets/jameslko/gun-violence-data>
* <https://scikit-learn.org/stable/>
* <https://numpy.org/>
* <https://www.ibm.com/in-en/topics/logistic-regression>
* <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html>