Tensor 2.0

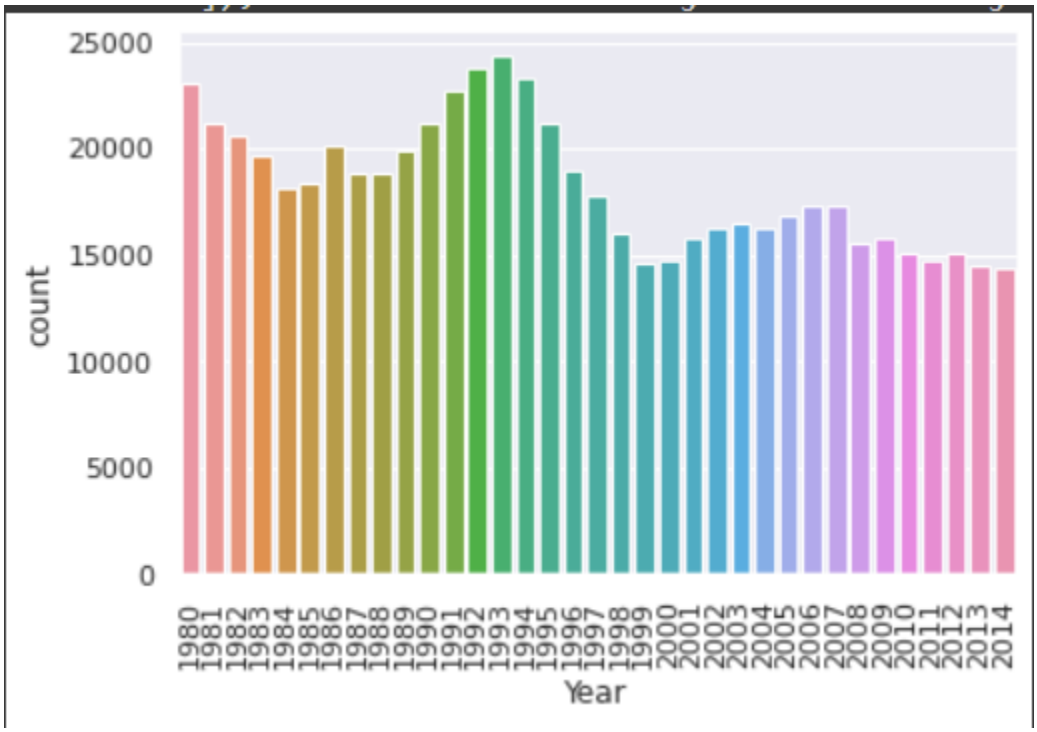
Homicide Report 1980-2014

By Srinivasa Raghavan S, Vihar Devalla, Jaaswin D Kotian

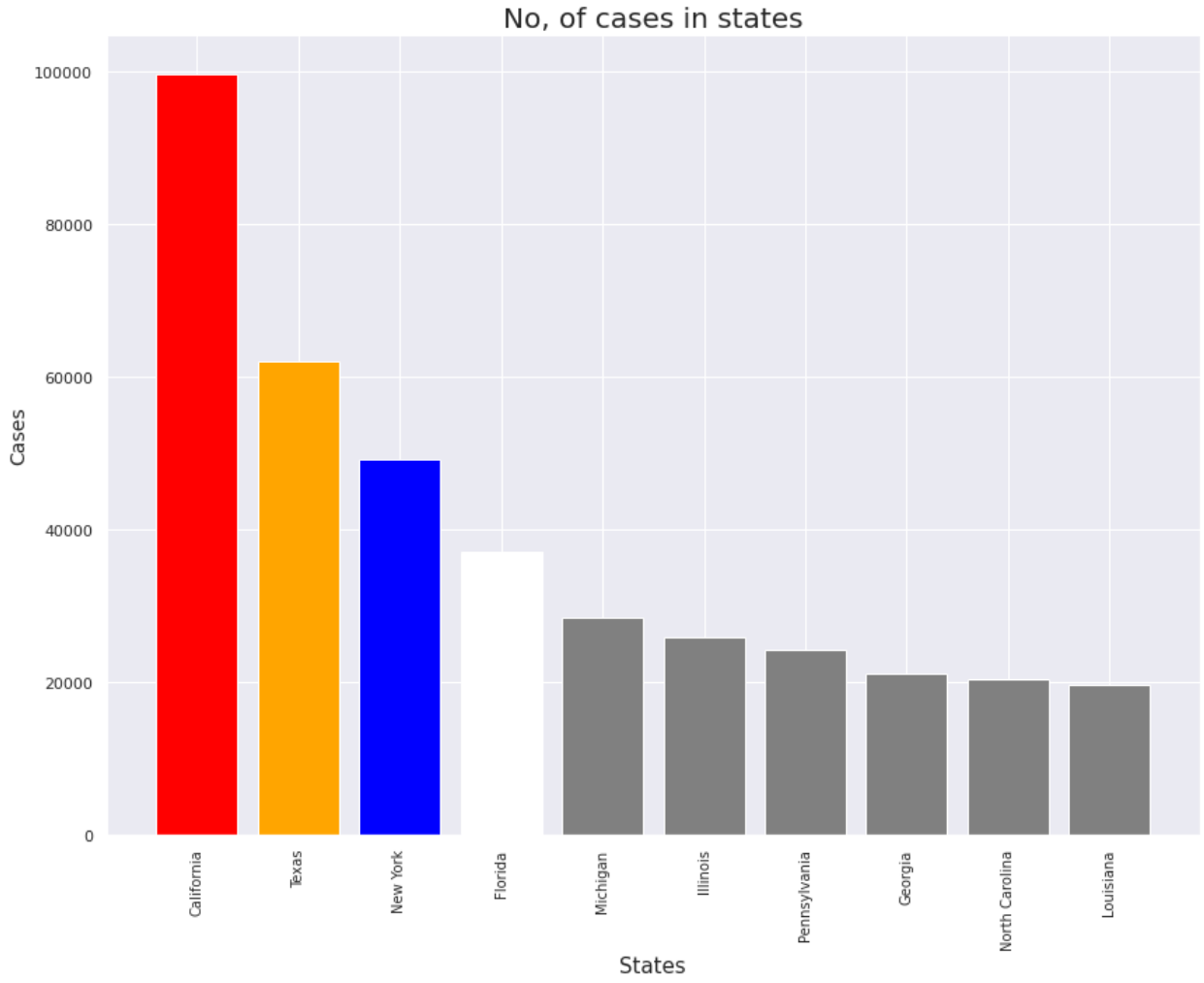
Problem Statement : "Last year, an alarming increase in homicides left communities reeling as officials searched for answers. Over the past few decades, the rate at which this horrendous crime is increasing is alarming, and Governments and Police all over have been looking for insights and solutions to help them understand the situation more."

Solution:
We used the Homicide Report 1980-2014 Dataset from Kaggle (https://www.kaggle.com/murderaccountability/homicide-reports) to understand the different cases related to homicides in the US for the years 1980-2014 and plotted necessary graphs to correlate possible features and developed an ML model to predict if a homicide case can solve or not using the given dataset.

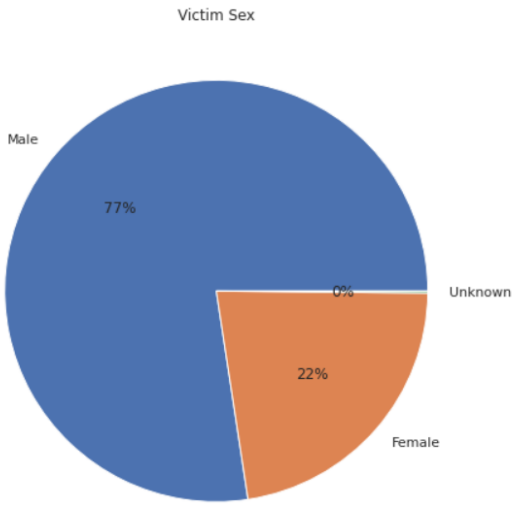
Inferences

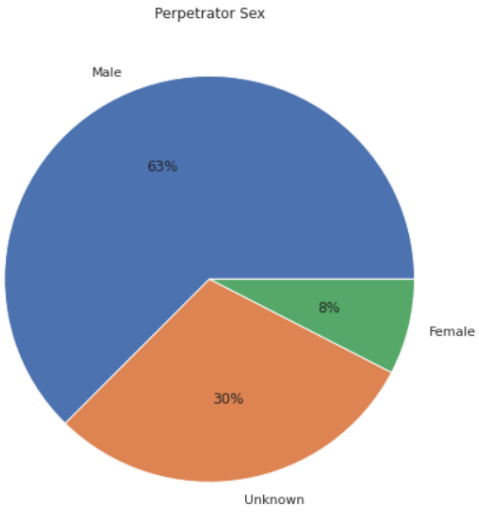


The given plot shows the number of homicide cases from the years 1980-2014. The graph hits a peak around the 1990s and since has almost stagnant to declining cases through the years. The increase in the 1990s was due to “crack epidemic” that fueled violent crimes in the late 1980-90s. New York was the worst offender during this time.

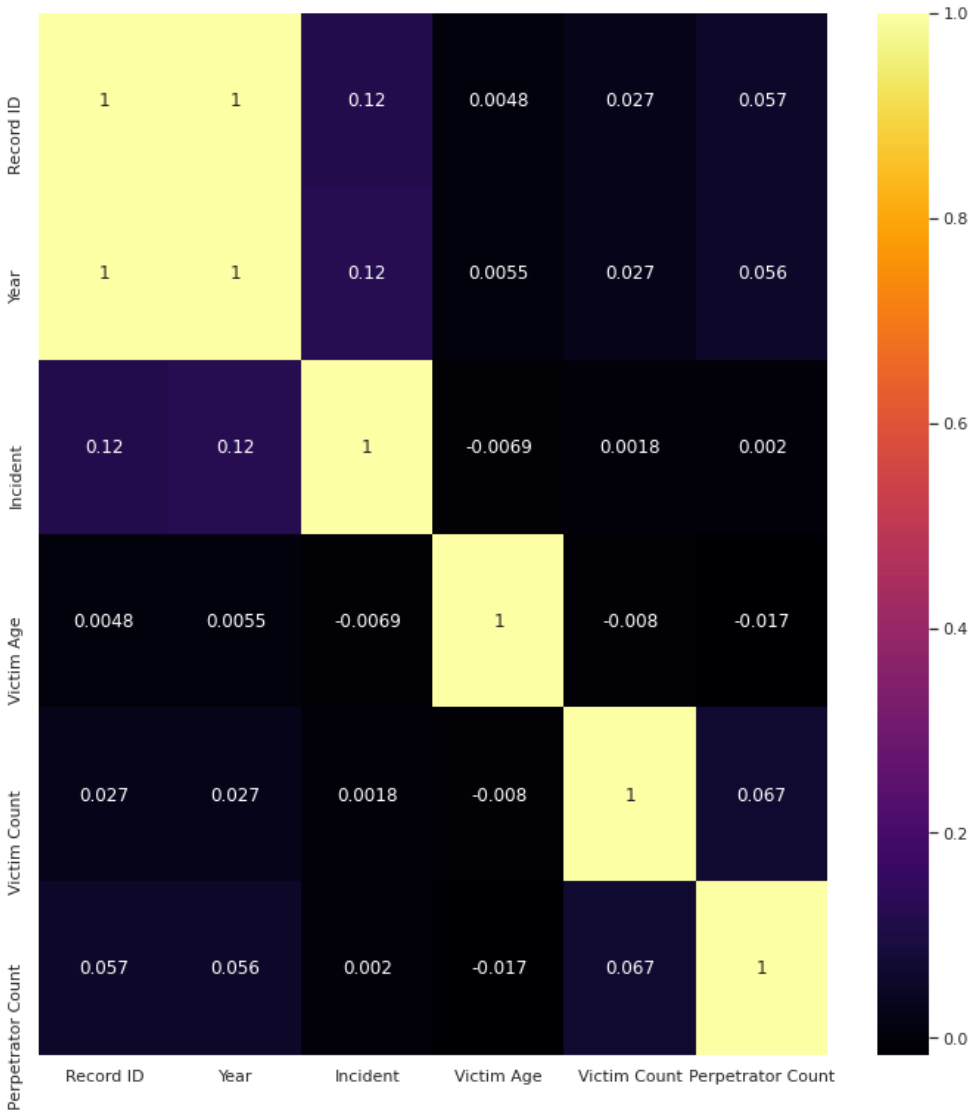


This plot is used to display the number of cases recorded in the different states of the USA. California has the maximum cases recorded during this time. This could be due to the riots, gangs first seen during the crack epidemic and continue to increase after it.

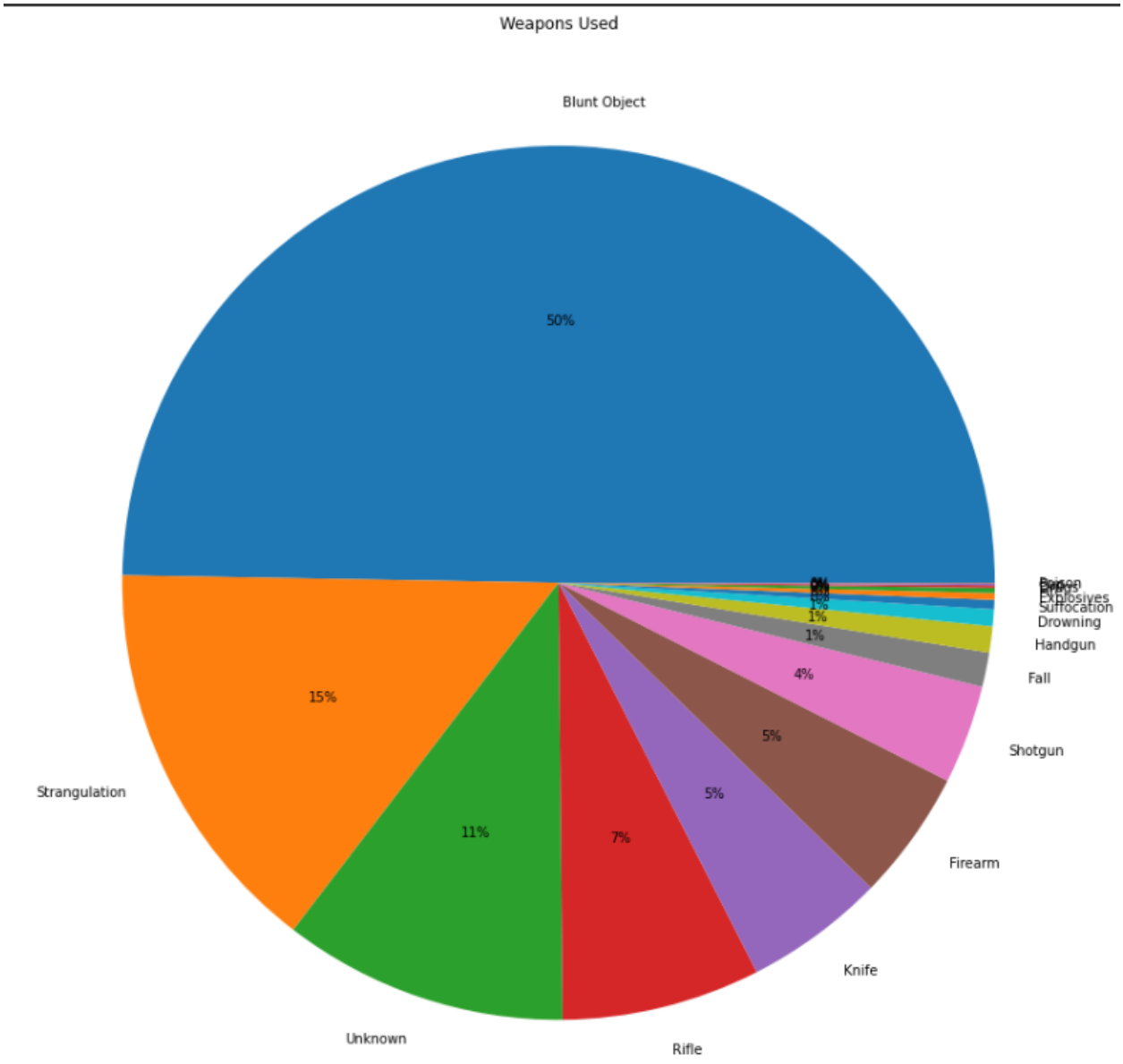




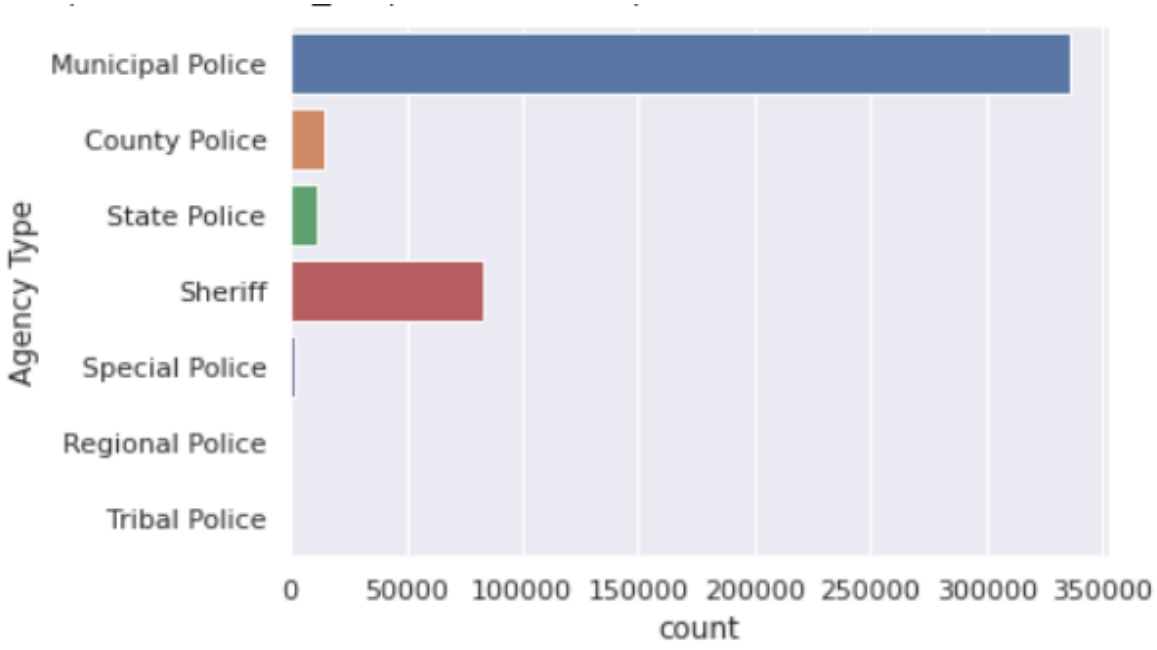
The given graphs show that men are killed the most by other men (at least in theory) due to possible gang wars and political agendas.



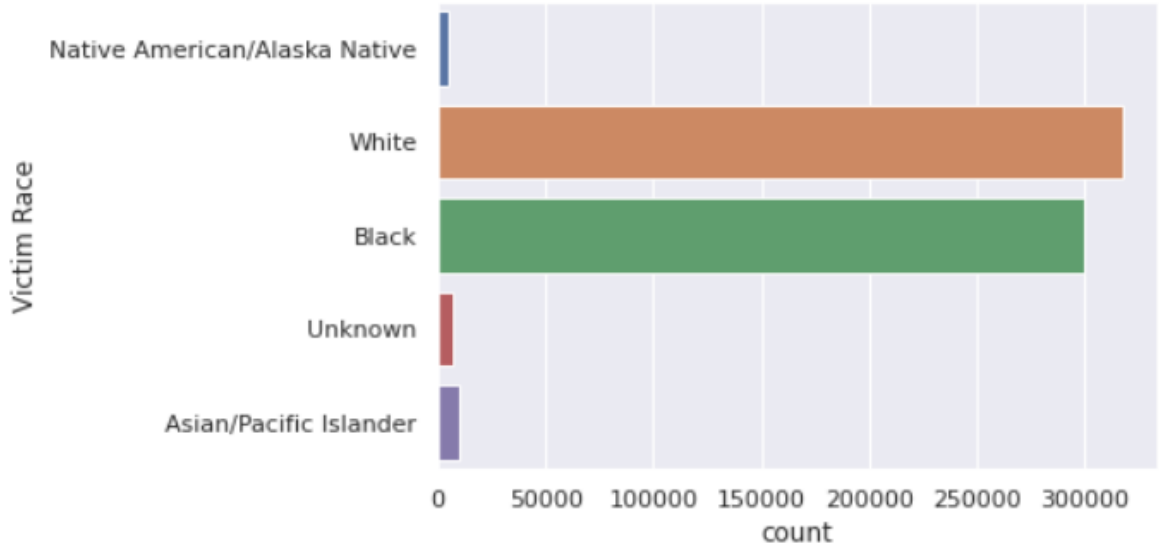
Here is a heatmap to correlate the different columns in our database.



This shows the different weapons used for the homicide. Strangulation is the second highest killer after blunt objects (could be a wide variety of objects), this could be due to domestic violence which usually involves strangulation, followed by rifles which is due to lenient gun laws in the US.



Most cases are involves municipal police as most crimes are recorded in cities, and it comes under municipal jurisdiction.

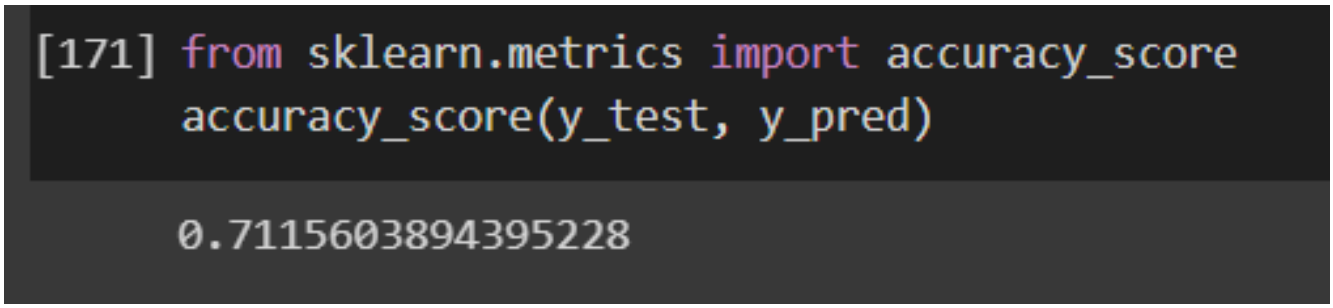


A count plot to display the Victim Race

Model

We have created an ML model to predict if a given case can be solved or not based on the features such as 'Agency Type', 'State', 'Year','Incident', 'Crime Solved', 'Victim Age', 'Weapon','Victim Count', 'Perpetrator Count'.

The model used is a RandomForest Classifer, LabelEncoder for encoding, and we got around 70% accuracy for this model.



Colab Link - https://colab.research.google.com/drive/1yDLXvVdy5LZMJy-OMi7hzRvh66LqK-Kn?usp=sharing