

+ Homicides by US Police: A Data Driven Analysis +

Presented by:

Josh Hankel, Roopkanwal, Matthew Leech, Andy Armstrong, Lola Oye Balogun

Matters of Discussion



THE BIG QUESTIONS



CRISP
METHODOLOGY



VISUALIZATIONS



CONCLUSION &
RECOMMENDATIONS

The Big Questions



Predict likelihood if victims are armed, fleeing from police, where they live, and their reported threat level

The role of mental health on victims of US police shootings

Forecast the shooting trend

Cities and States that had the highest number of shooting instances

Public perspective on the current police shootings through text analysis

+

•

○

Impacts of the big questions on the society

- To understand what things are common amongst police shooting victims

CRISP Methodology

Business Understanding

- What are we trying to achieve?

Data Understanding

- What are we working with?

Data Preparation

- How are we going to work with it?

Modelling

- What insights have we found?

Evaluation

- Do our insights answer the Questions?

Deployment

- Conclusions



Business Understanding

Desired outputs and objectives

Goal: to understand what factors are most prevalent in the US police shootings

Methodology: use machine learning techniques, data analysis, statistical analysis and data visualization to find patterns and outliers in the data

Success: finding significant information, patterns, and outliers from the dataset



Assessing the current situation



Inventory of Resources: People involved, software used, virtual environments

Assumptions and Constraints

Data Dictionary

Term	Definition	Term	Definition
CSV	Comma-separated values file type	Record	A record is another word for a row in a table
ETL Process	Extract, Transform, Load process part of the data preparation stage	Integer	A datatype that represents whole numbers
Dataset	A collection of data and information	Float	A datatype that represents numbers that contain decimals
Python	A programming language used for computing	Mean	A statistical method meaning average
Datatype	A kind of data item that is defined by the values it can take	Benchmark	A statement or value that other things are compared to
String	A datatype that is meant for word and text values	Correlation	A relationship or connection between two or more things
Null Values	A term to describe missing values in data	Machine Learning	An application of Artificial Intelligence that allows the system to automatically learn and improve
Coerce	Takes values that are not allowed for the datatype and sets them as 'NaN'	Association Rule Mining	A machine learning technique that finds correlation between two or more items
Boolean	A datatype that allows only 1/0 values	<u>Mlxtend</u>	A Python library that is used for day to day science tasks
Lift	A statistical term that means the ratio of target response divided by the average response	<u>Apriori</u>	An algorithm used in finding frequent item sets for the purpose of association rule mining
Clustering	A machine learning method that divides and groups like data	<u>SciKit/Sk-learn</u>	A Python library that is used in Machine Learning



Data Understanding

Describing the Data

1

CSV and XLSX files used

2

~5000 records and 15 columns in CSV

3

Data gathered from public databases and from 100 largest police departments in US

Data Exploration

1

For the CSV file, contains Boolean values, integer (float and integer), and object (strings)

2

Each column contains ~5000 values

RangeIndex: 4895 entries, 0 to 4894

Data columns (total 16 columns):

#	Column	Non-Null Count	Dtype
0	id	4895 non-null	int64
1	name	4895 non-null	object
2	date	4895 non-null	object
3	manner_of_death	4895 non-null	object
4	armed	4895 non-null	object
5	age	4895 non-null	float64
6	gender	4895 non-null	object
7	race	4895 non-null	object
8	city	4895 non-null	object
9	state	4895 non-null	object
10	Country	4895 non-null	object
11	signs_of_mental_illness	4895 non-null	bool
12	threat_level	4895 non-null	object
13	flee	4895 non-null	object
14	body_camera	4895 non-null	bool
15	arms_category	4895 non-null	object

dtypes: bool(2), float64(1), int64(1), object(12)

memory usage: 545.1+ KB

Verifying the Quality

1

Data is clean

2

Data types needed to be changed for future modelling

3

The below picture shows any null values in our data

```
# %% checking number of rows with null values
```

```
0
```

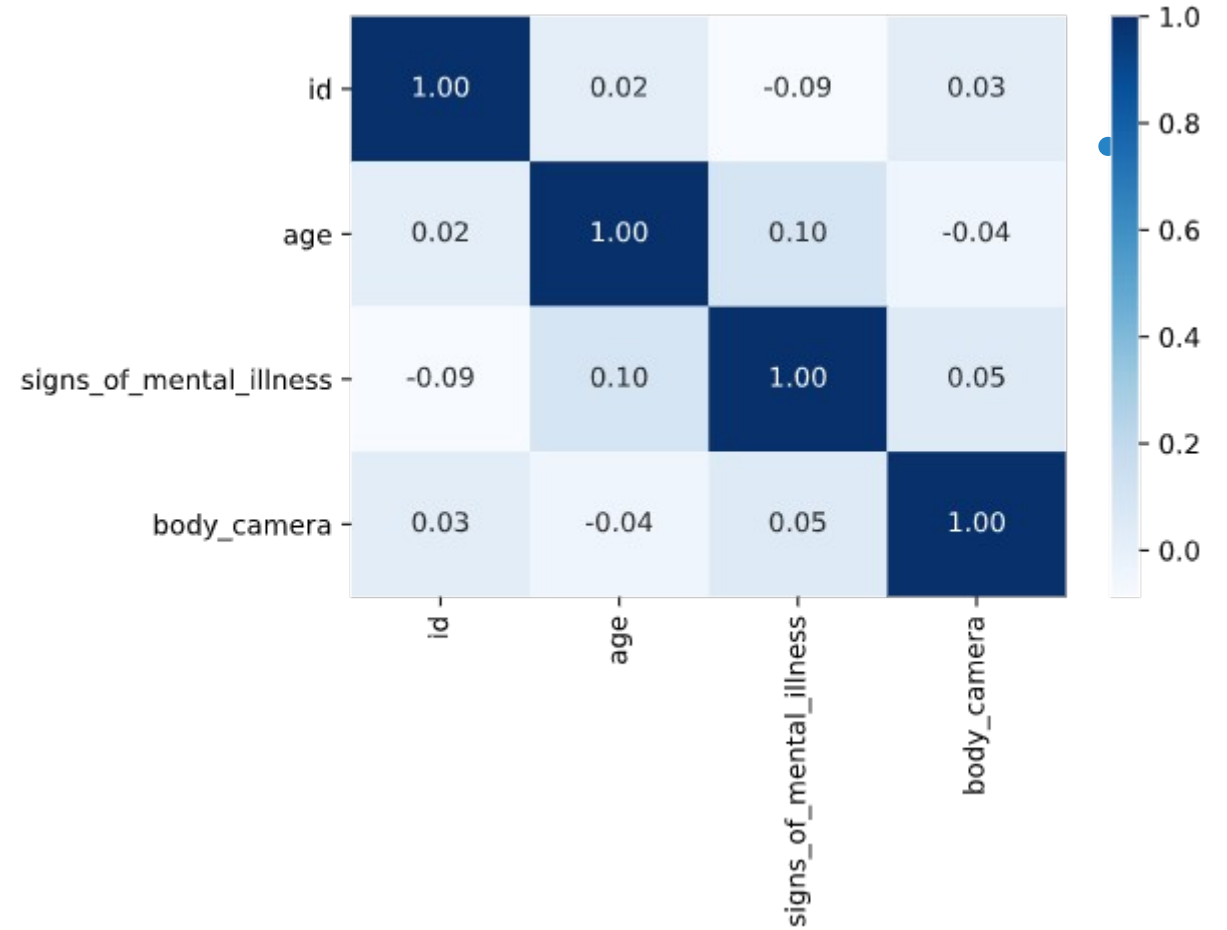
+



Explore Relationships

1

Correlation seen between age and signs of mental illness



Data Preparation

Selecting the Data

- Main dataset retrieved from Kaggle; an opensource
- All data served a purpose
- Some information used as is
- Some data needed to be changed and prepared for *data modelling*
- Data types also needed to be changed and formatted for *visuals* and *text analysis*


Cleaning and Formatting the Data


- 
- Created new columns for association mining rules and cluster analysis

- 
- Transformed categorical columns to Boolean array using one hot encoding

- 
- Converted date to datetime format

Integrating the Data

- 
- Our original datasets were in .CSV and .XLSX files, so integration between platforms was easy

- 
- Creation of a new dataset allows us to keep original data



Modelling



Association Mining Rules

Machine Learning Technique



```
graph TD; A[Machine Learning Technique] --> B[Uses frequent itemsets to find the most common attributes in the dataset]; B --> C[Apriori algorithm is used to understand correlations]; C --> D[Correlation between the attributes "Guns, Threat Level: Attack, White, Black, and Not Fleeing"];
```

Uses frequent itemsets to find the most common attributes in the dataset

Apriori algorithm is used to understand correlations

Correlation between the attributes "Guns, Threat Level: Attack, White, Black, and Not Fleeing"

Association Mining Rules

antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(arms_category_Guns)	(threat_level_attack)	0.564658	0.645557	0.450868	0.798480	1.236887
(threat_level_attack)	(arms_category_Guns)	0.645557	0.564658	0.450868	0.698418	1.236887
(race_White)	(flee_Not fleeing)	0.505822	0.627783	0.338509	0.669225	1.066012
(race_White)	(threat_level_attack)	0.505822	0.645557	0.335036	0.662359	1.026027
(flee_Not fleeing)	(threat_level_attack)	0.627783	0.645557	0.410010	0.653108	1.011697
(threat_level_attack)	(flee_Not fleeing)	0.645557	0.627783	0.410010	0.635127	1.011697
(arms_category_Guns)	(flee_Not fleeing)	0.564658	0.627783	0.352400	0.624096	0.994125

Scenario 1

antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(Black)	(Guns)	0.265169	0.564658	0.156486	0.590139	1.045126
(White)	(Guns)	0.505822	0.564658	0.297242	0.587641	1.040703

Scenario 2

Clustering

Machine Learning Technique



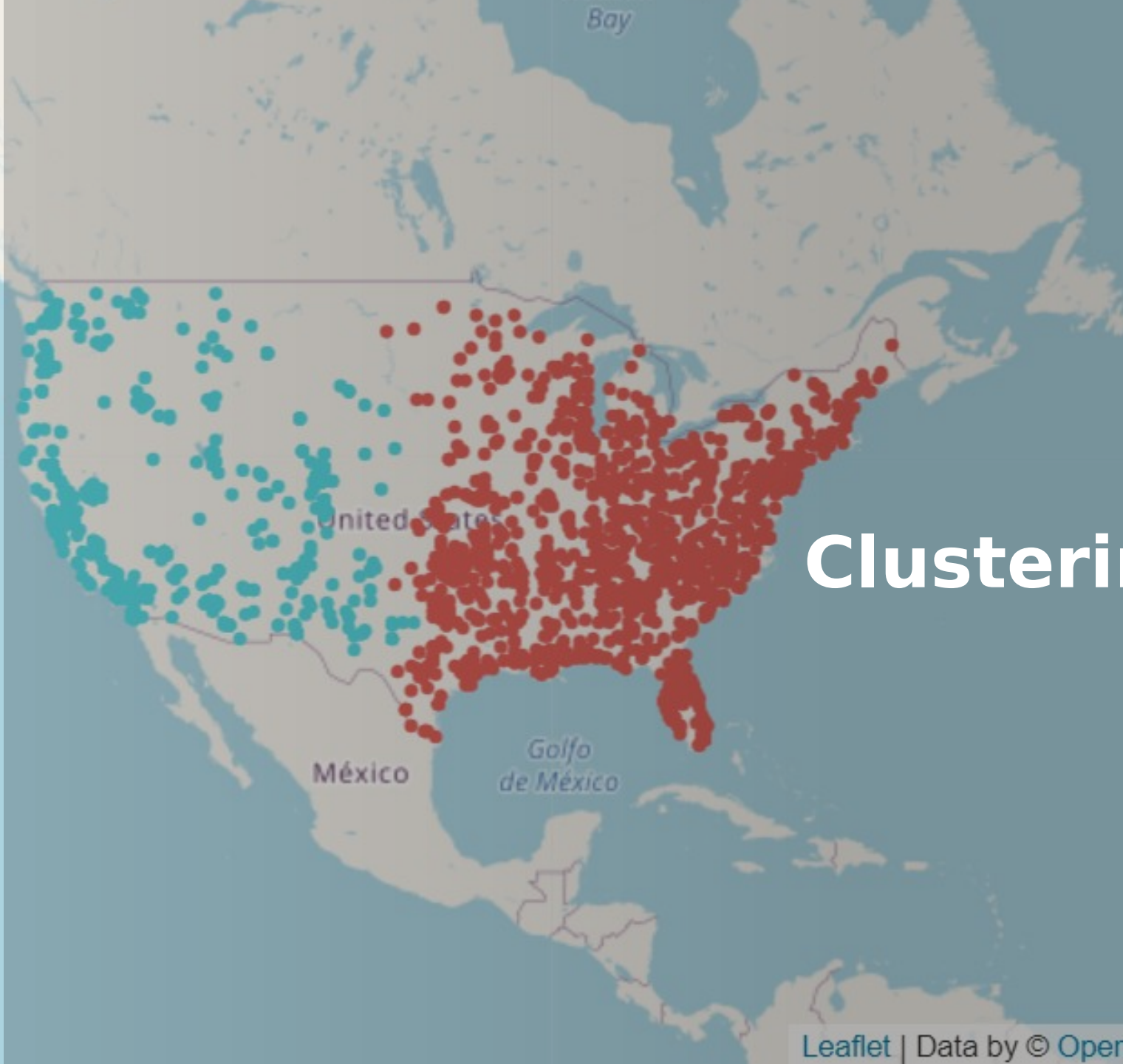
```
graph TD; A[Machine Learning Technique] --> B[Uses the SciKit/SKlearn Python library]; B --> C[Hierarchical clustering method to group like patterns in our data]; C --> D[Divide in East and West Coast];
```

Uses the SciKit/SKlearn Python library

Hierarchical clustering method to group like patterns in our data

Divide in East and West Coast

Clustering



Unstructured Text Analysis

Using machine learning to determine sentiment (Positive, Neutral, Negative)

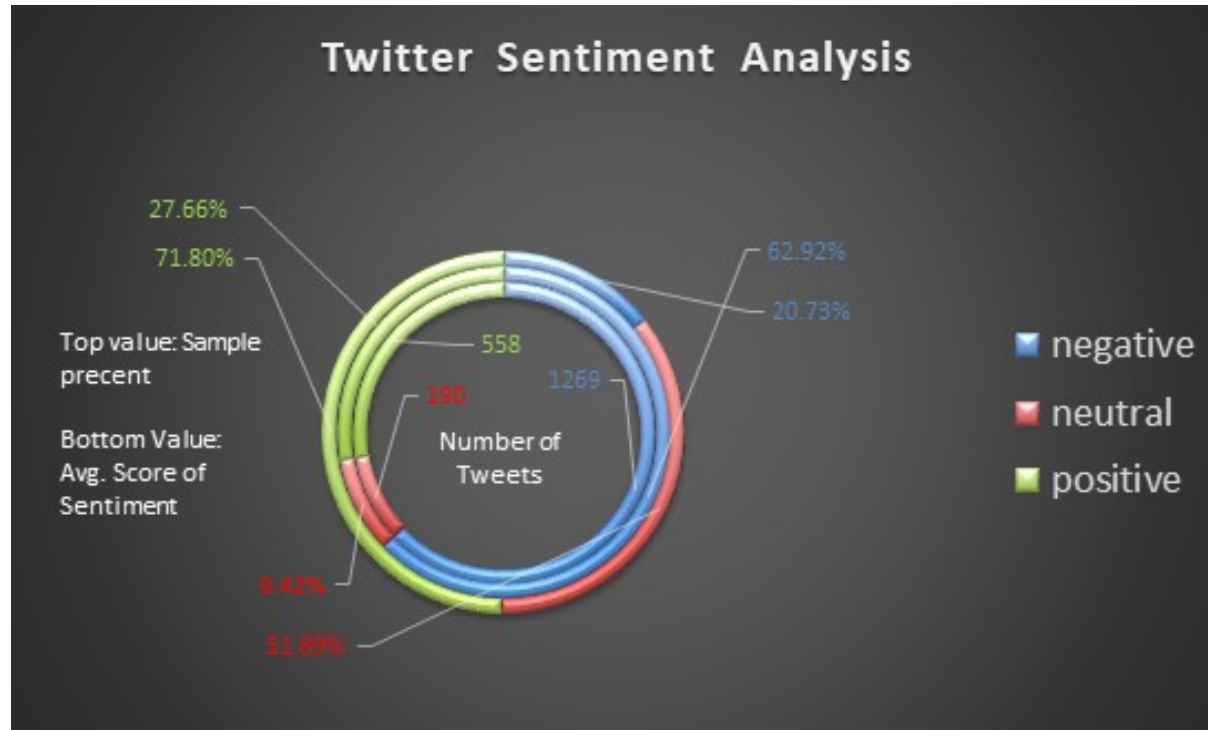


```
graph TD; A[Using machine learning to determine sentiment (Positive, Neutral, Negative)] --> B[Uses Azure Machine Learning add in for Excel]; B --> C[Able to generate sentiment from 2000 tweets in seconds];
```

Uses Azure Machine Learning add in for Excel

Able to generate sentiment from 2000 tweets in seconds

Unstructured Text Analysis



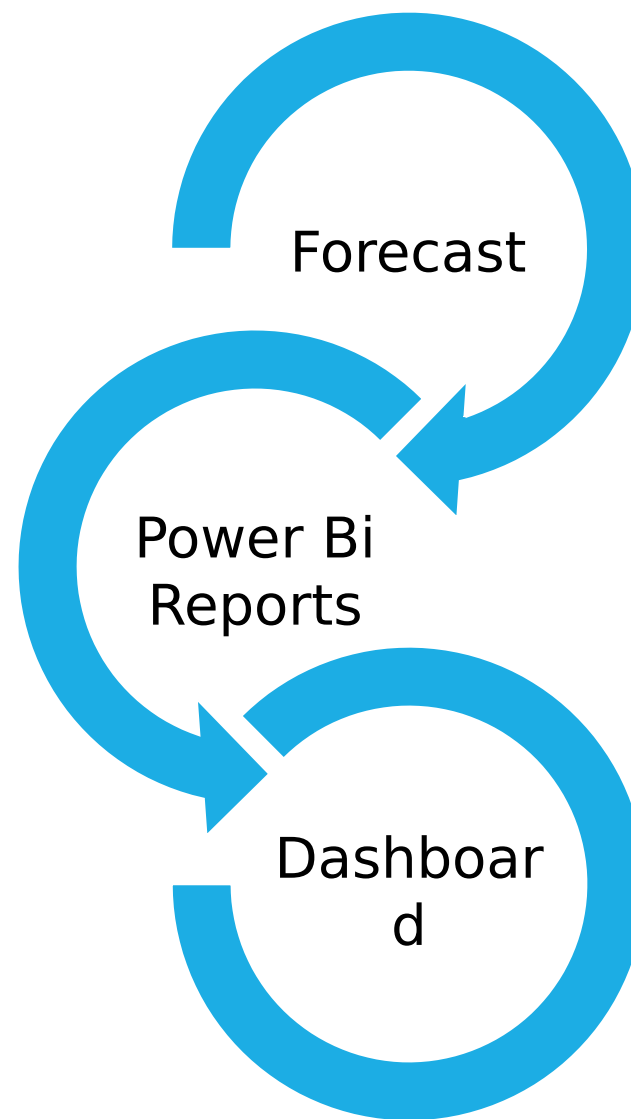
Row Labels	Number of Tweets	Percentage of Sentiment	Average Score of Sentiment
Negative	1269	62.92%	20.73%
Neutral	190	9.42%	51.89%
Positive	558	27.66%	71.80%
Grand Total	2017	100.00%	37.79%

+

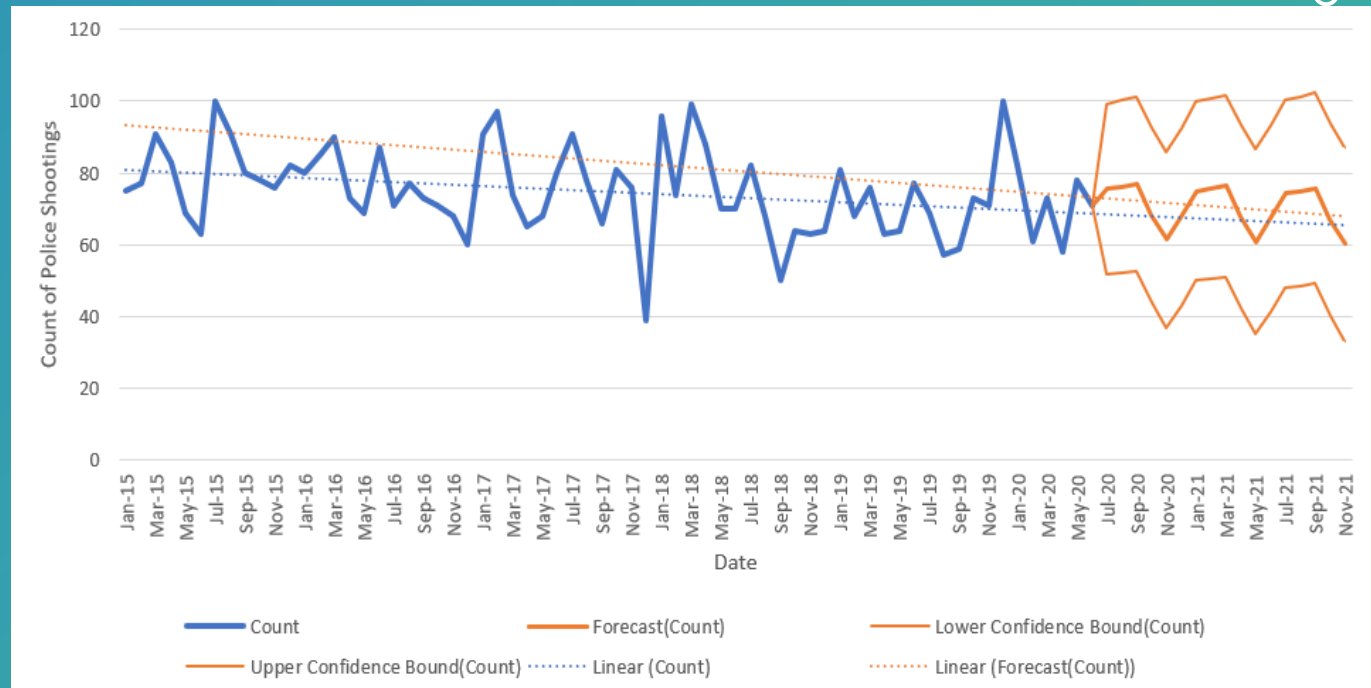
•

○

Evaluation



Shootings Forecast



Police Shootings

Total shootings
4.90K

Total Fleeing
1.82K

City-State

All

20152020

Count of instances by Race

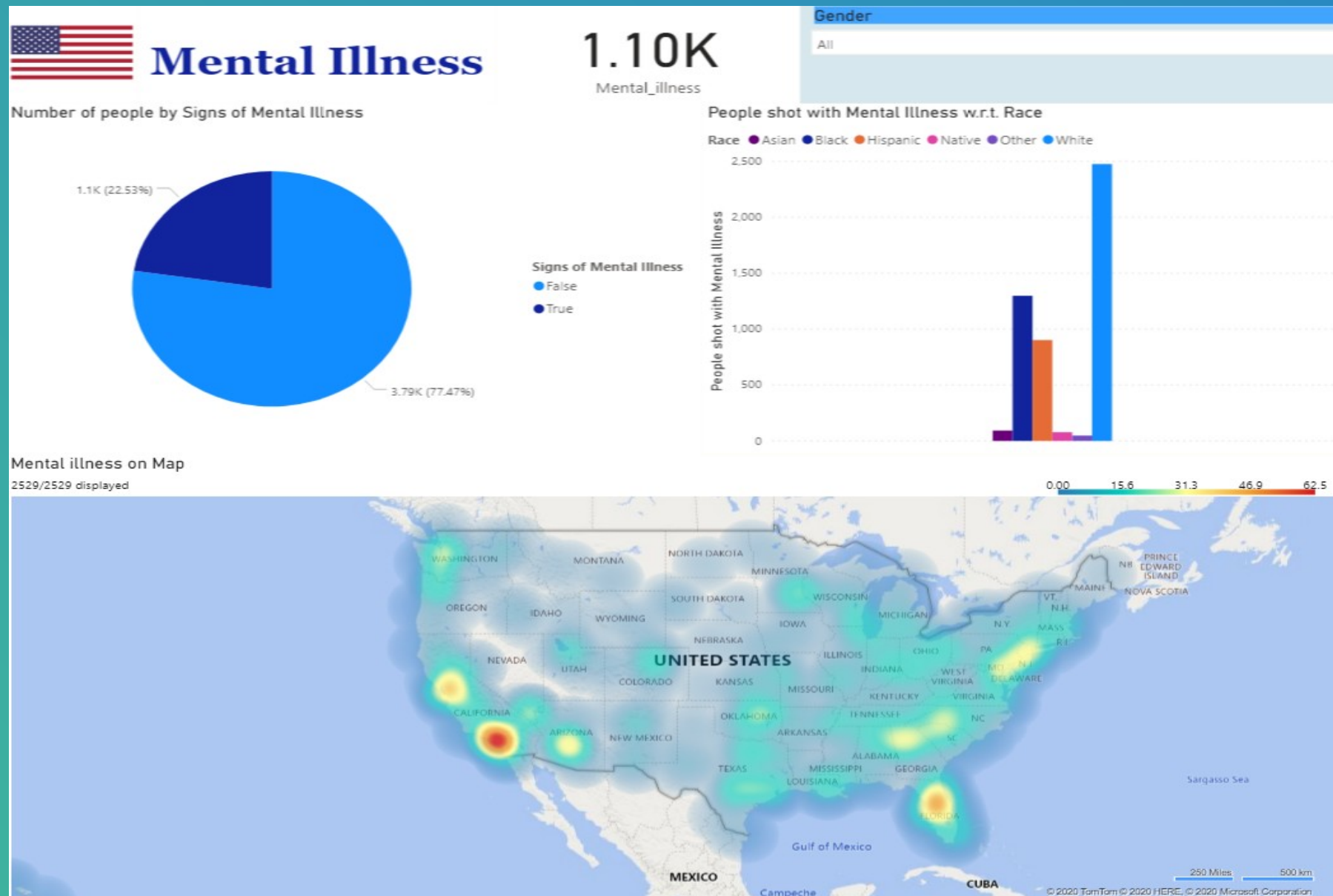
Race	Count	Percentage
White	~2650	50.58%
Black	~1700	26.52%
Hispanic	~1100	18.43%
Other	~50	1.9%

Arms Category w.r.t. Average Age

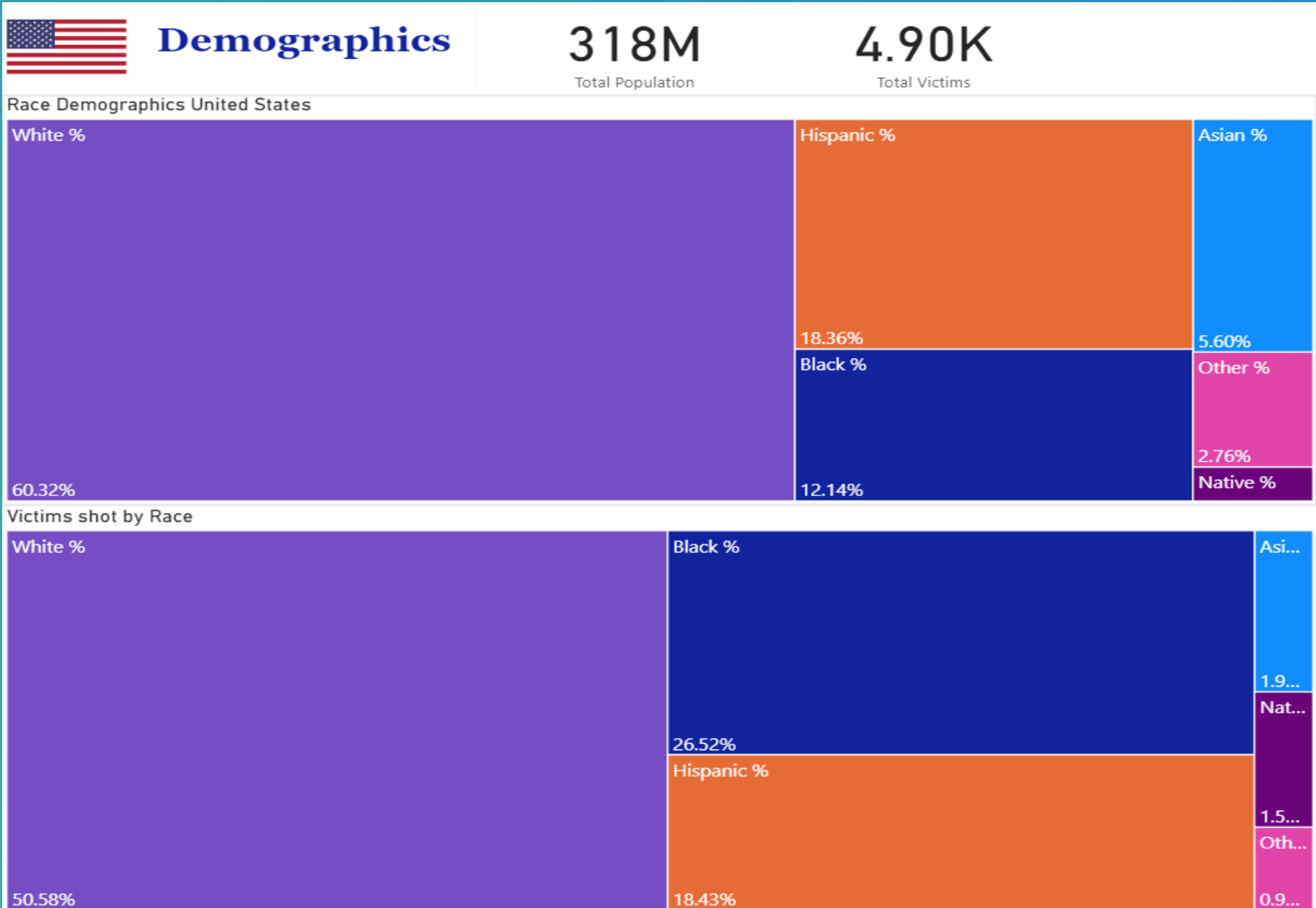
Arms Category	Average Age
Guns	~42
Piercing objects	~40
Blunt instruments	~38
Sharp objects	~37
Electrical devices	~35
Hand tools	~25

Count by Race and Arms Category

Race	Blunt instruments	Electrical devices	Explosives	Guns	Hand tools	Multiple	Other unusual objects	Piercing objects	Sharp objects	Unarmed	Unknown	Vehicles
White	~50	~5	~5	~1450	~5	~5	~100	~10	~400	~150	~200	~50
Black	~10	~5	~5	~750	~5	~5	~50	~10	~150	~100	~100	~50
Hispanic	~20	~5	~5	~450	~5	~5	~50	~10	~180	~50	~100	~20
Asian	~5	~5	~5	~50	~5	~5	~10	~5	~20	~10	~10	~5
Native	~5	~5	~5	~50	~5	~5	~10	~5	~10	~5	~5	~5
Other	~5	~5	~5	~50	~5	~5	~10	~5	~10	~5	~5	~5



Demographics Report



Homicides and Killings Report



Homicides and Killings

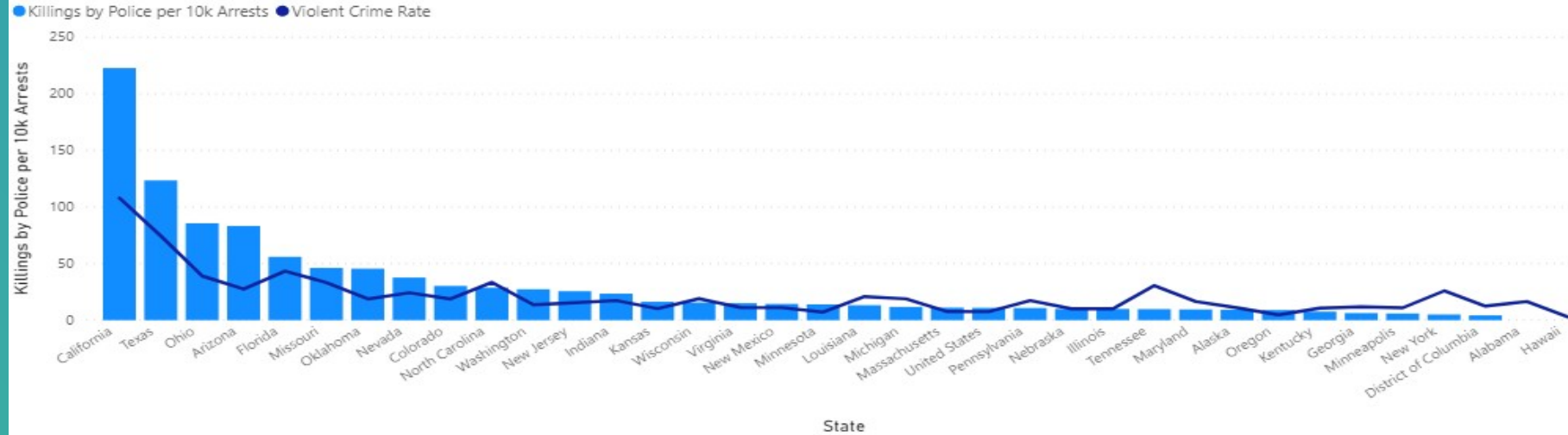
1.05K

Killings by Police per 10k Arrests

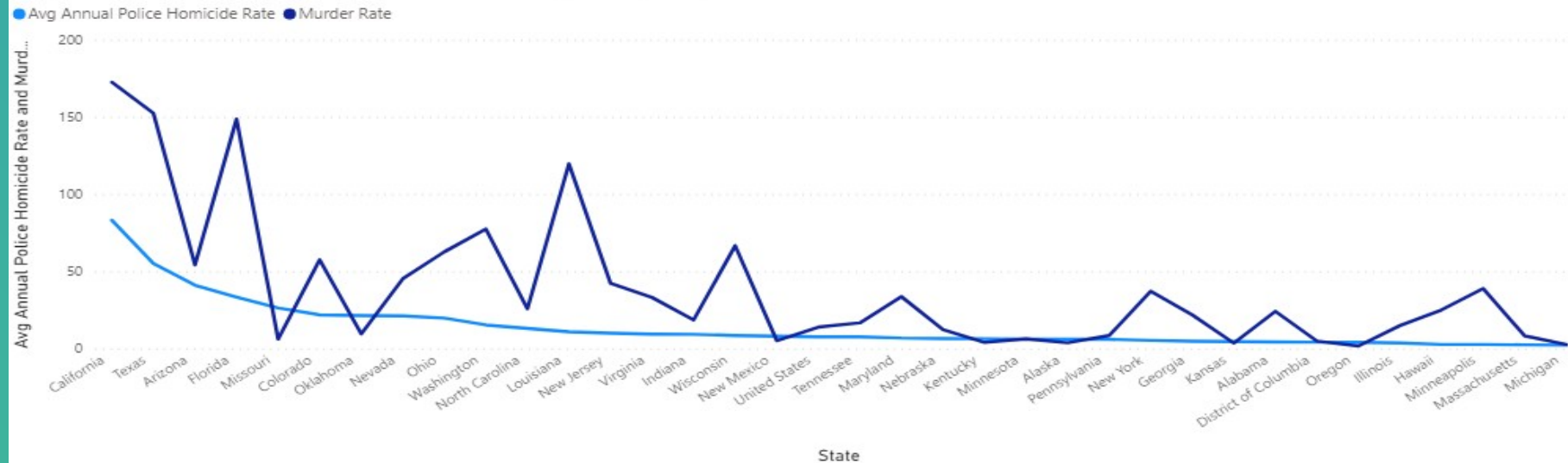
7.56

Average of Violent Crime Rate

Killings by Police per 10k Arrests and Violent Crime Rate per 1k People by State



Avg Annual Police Homicide Rate and Murder Rate per 100,000 by State

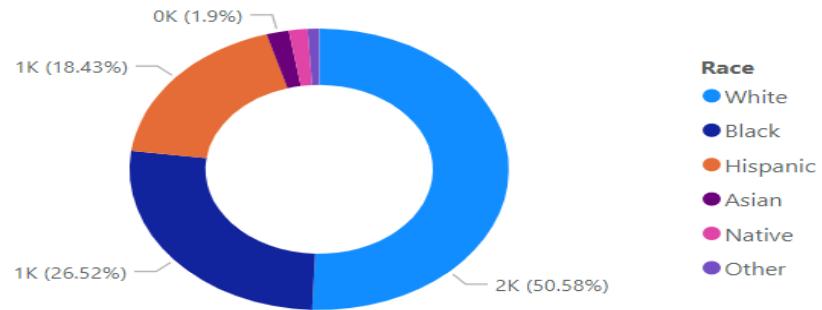




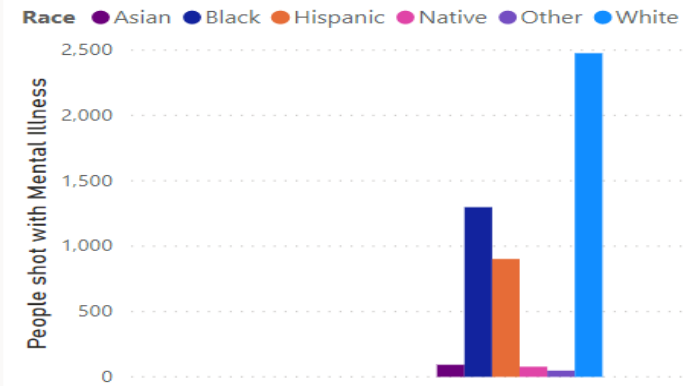
Police Shootings

4.90K

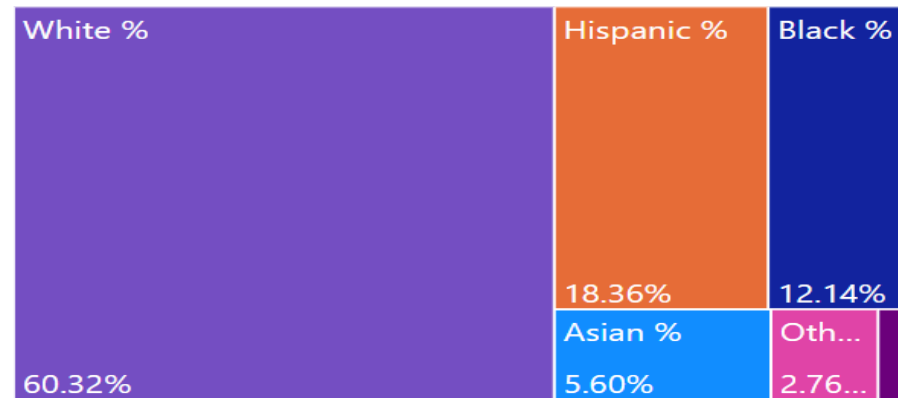
Count of instances
BY RACE



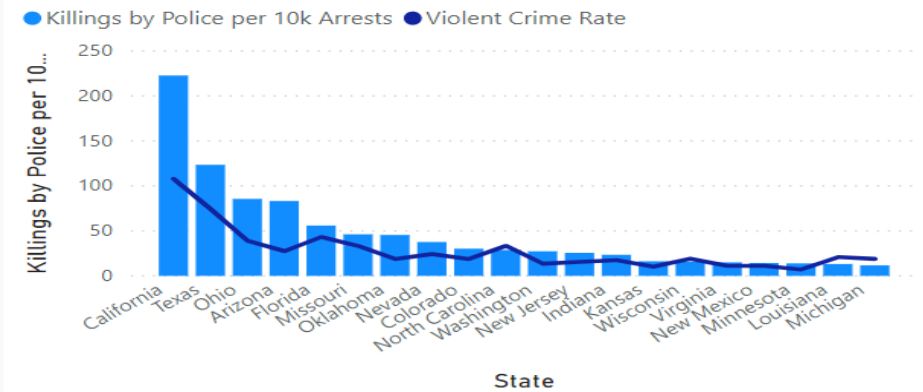
People shot with Mental Illness w.r.t. Race



Race Demographics United States



Killings by Police per 10k Arrests and Violent Crime Rate per 1k Peo...



Deployment



Conclusion & Insights



Association Rule Mining shows high likelihood of being white, being black, having a gun, being aggressive, and not fleeing were reported together in victims.

Clustering allowed helped to see distinct groups of east and west coast where shooting occurred, specifically in the east coast.

The creation of powerful visuals made in Power BI show important trends in race, gender, age, location, and many others.

Black population has the highest rate of killings per capita, victims who are armed are much more likely to be engaged by a police officer and when armed they were likely to have guns.

The number of shootings seems to be decreasing, while the population is increasing.

Recommendations



A model of policing that focuses on respect, neutrality and transparency should be implemented to address biases



Training and retraining of the police force in the area of de-escalation and mental health training



Measures to get guns out of the hands of criminals



More community Policing to change public sentiment

Questions

