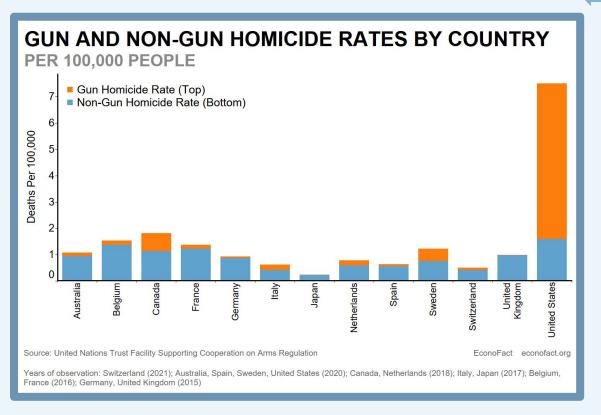# Gun Violence in the United States:

Predicting Fatality of Shooting Incidents

By: Arshia Agrawal, Beauty Kabore, Connor Morris, Insiya Raja, Joe Bernal

Background

➤ **Serious public health problem**

  ○ >48,000 firearm-related deaths in the US in 2022 alone

  ○ ~132 people die from a firearm-related injury every day

➤ **Leading cause of death**

  ○ Among top 5 causes of death for ages 1-44 in the US

  ○ #1 cause of death among children & teens

➤ **Other losses**

  ○ Costs the US tens of billions of dollars every year

  ○ Threatens our collective sense of safety and security

GUN AND NON-GUN HOMICIDE RATES BY COUNTRY
PER 100,000 PEOPLE

Gun Homicide Rate (Top)
Non-Gun Homicide Rate (Bottom)

Deaths Per 100,000

Australia, Belgium, Canada, France, Germany, Italy, Japan, Netherlands, Spain, Sweden, Switzerland, United Kingdom, United States

Source: United Nations Trust Facility Supporting Cooperation on Arms Regulation

EconoFact   econofact.org

Years of observation: Switzerland (2021); Australia, Spain, Sweden, United States (2020); Canada, Netherlands (2018); Italy, Japan (2017); Belgium, France (2016); Germany, United Kingdom (2015)

America has the most cases of Gun related incidence per 100,000 people

Our Research Goal

To build a classifier that can predict the fatality of a gun-related crime incident

Our Analysis is Based on Several Factors:

agency_name	gang_activity_ind	victim_race	victim_sex	arrest_ind	clearance_status	fatality_status	victim_age_groupings	season	year	...	shotgun
1	Y	BLACK	MALE	0	OPEN	1	Younger Adult	Winter	2015	...	0
1	Y	WHITE	MALE	0	CLEARED-EX	1	Older Adult	Winter	2015	...	0
1	Y	BLACK	MALE	1	CLEARED-ARREST	1	Younger Adult	Winter	2015	...	0
1	Y	WHITE	MALE	0	CLEARED-EX	1	Older Adult	Spring	2015	...	0
1	Y	BLACK	MALE	0	OPEN	1	Younger Adult	Spring	2015	...	0

➢ Police Department
➢ Gang Activity
➢ Victim Race, Sex, & Age

➢ Arrest Status
➢ Case Status (open/closed)
➢ Time & Date of Occurence
➢ Weapon Used

Research Importance

- Save lives.

- Aid in the allocation of appropriate emergency response and resources.

- Gain better understanding of the relationship and trends between these incidents.

- Guide preventative measures.

The model will prioritize ethical considerations to prevent misuse of technology.

A Closer Look at the Data

Narrowing Down Our Dataset

22 Police Departments	5 Police Departments	2 Police Departments

Overall Dataset:

From 22 municipal police departments in the United States

Internal data on homicides, aggravated assaults, and non-fatal shootings

Initial Data:

Newark
Los-Angeles
Baltimore
San-Francisco
Boston

Selected based on availability of fatality classification data

Final Analysis:

Los-Angeles
Boston

Selected to reduce null values, making it easier to feed the data into our models- data categorization, One-Hot Encoding
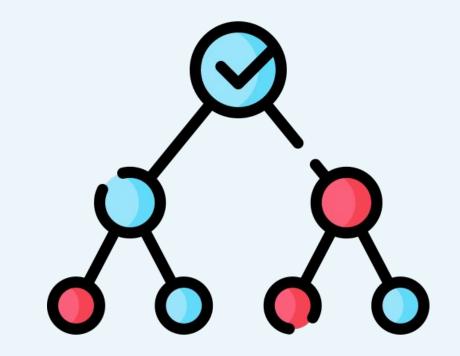
Definitions & Assumptions

➤ A firearm injury is a gunshot wound or penetrating injury from a weapon that uses a powder charge to fire a projectile.

➤ Fatal means 'Homicide-Gun' (as recorded by Baltimore PD, Boston PD, Newark PD, San Francisco PD) and 'Homicide-Criminal-Fatal-Shooting' (as recorded Los Angeles PD)

➤ Non-Fatal means 'Non-Fatal Shooting' (as recorded by Baltimore PD, Boston PD, Newark PD, San Francisco PD) and 'Agg Assault - Shooting - Non-Fatal' (as recorded by Los Angeles PD)

➤ Transgender gender had one (1) value count so was added to the 'Other' category under victim sex.
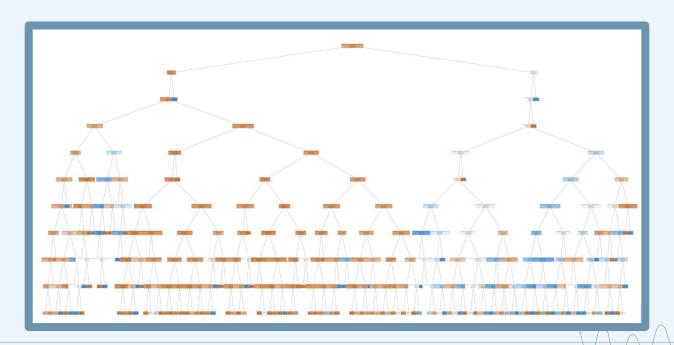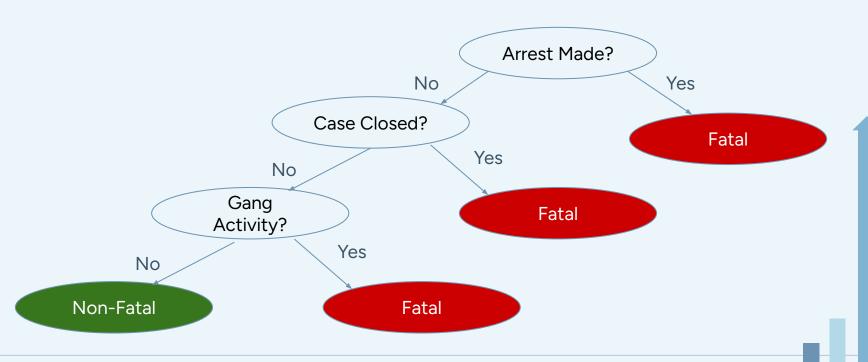
Model #1: Decision Tree

0.84% Accuracy

Top Predictive Factors of Fatality:
➢ Arrest & Case Status: arrest made and case closed -> greater likelihood of fatality
➢ Gang Activity: crime related to gang activity -> lower likelihood of fatality

Model #2: Logistic Regression

Accuracy - 0.81%

Precision, Recall, & F1-Scores

All three metrics are much better for non-fatal incidents than fatal incidents.

The precision level for non-fatal shootings was higher than fatal shootings (83% for non- fatal shootings and 55% for fatal shootings).

45% of fatal shootings were not correctly predicted.

	precision	recall	f1-score
0	0.83	0.95	0.89
1	0.55	0.23	0.32

Model #2: Logistic Regression

Confusion Matrix:

Non-Fatal (Correctly Predicted): **1410**	Fatal, Labeled as Non-Fatal: **68**
Non-Fatal, Labeled as Fatal: **287**	Fatal (Correctly Predicted): **84**

➢ False negatives are more dangerous than false positives.

➢ If a fatal incident is wrongly predicted to be non-fatal, resources sent to the incident scene may not be sufficient, resulting in tragedy.
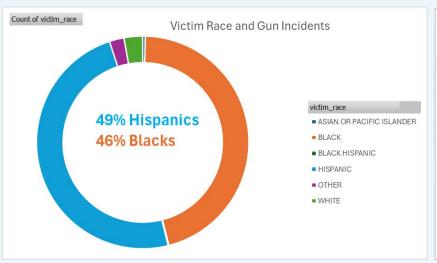
THE MAJORITY OF VICTIMS ARE MALE

Count of victim_sex

Victim Sex and Fatality Status

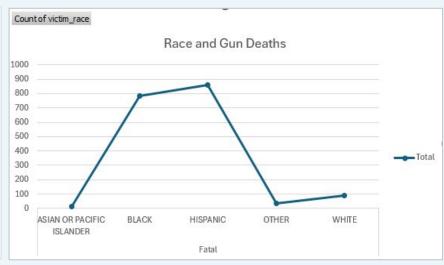fatality_status
- Fatal
- Non-Fatal

victim_sex

8,234 American men in Boston and Los-Angeles were involved in gun related violence. Indicating 89% of recorded cases. Of the 19% of gun related deaths recorded by these Police departments, 91% were young or older male adults.
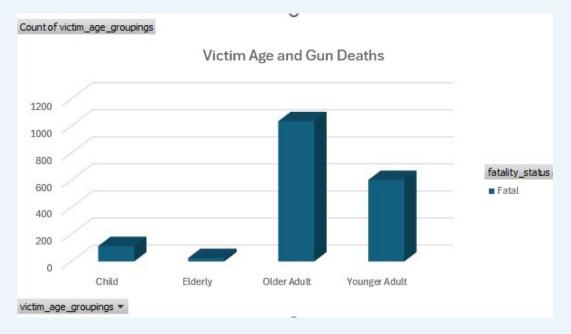
HISPANICS AND BLACKS ARE KILLED DISPROPORTIONATELY

Our data ranges from 2005 - 2018. In 2010, Blacks and Hispanics made up 58% of the Los-Angeles population.In Boston, they made up 0.7%. Yet, 95% of reported gun cases were hispanic or black people.
Contrary white people made up 94% in these areas, with 3% of recorded gun incidents. Of the 1,778 reported gun deaths, 48 and 44% were Hispanics and Blacks respectively. These minorities groups are taking the toll of gun violence.
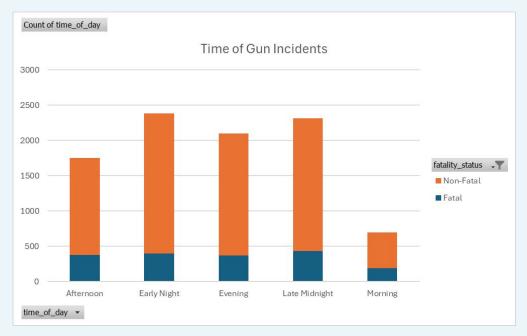
WORKING AGE ADULTS ARE LARGELY AT RISK

Count of victim_age_groupings

Victim Age and Gun Deaths

fatality_status
■ Fatal

victim_age_groupings ▾

92% of the 1,778 victims killed by gun were between ages 17 - 64. In addition, 6% (114) were children under 18.

MOST INCIDENTS OCCUR IN THE EVENING

74% of recorded incidents (6,796) occurred between 5pm and 12 am.

Sources of Bias

- Inconsistent definition of certain variables

    - Some police departments definitions of non-fatal shootings do not meet the FBI's criteria

- Data processing issues

    - Margin of error

    - Missing data fields

- Inaccuracies in dataset due to human error during data entry;

    - Typos, misinterpretations, or miscategorizations of data

- Past biases in data collection (eg, transgendered not represented in dataset), reporting, or analysis can carry forward and perpetuate inaccuracies or skewed interpretations

Impact

Our hope is that this model is used to direct life-saving resources to incidents that are more likely to result in a fatality.

Powerful tool with potential to transform how we tackle gun-related crime incidents in the US.

Use Cases

- Emergency line operators
- Law enforcements
- Decision Makers
- Health Care

Ethical Implications

➢ If the model is less effective for one group as opposed to another, it could result in an increase in proportion of fatalities in that group.

➢ What if our model is used by bad actors?
   ○ By figuring out which factors increase the fatality of a shooting, bad actors could use the model to inflict more harm.

➢ Real People
   ○ Our dataset is a collection of real people's stories: real deaths and real tragedies. It's important to keep in mind that these are not just numbers.
   ○ We don't know how this data was collected - especially in terms of consent.

Next Steps

- Deepen Our Analysis
  - Separate factors before the arrest (victim identity, time/day of occurence, etc.) from factors after the arrest (arrest, clearance status, etc.)
  - Pruning our model to increase accuracy by preventing over/underfitting

- Testing for Social Impact
  - Conduct tests of our model's accuracy on various populations - looking for significant accuracy differences between races, genders, etc.

- Looking for Other Contributing Factors
  - Socioeconomic status
  - Quality of emergency services and healthcare
  - Response time
  - Shooter relationship to victim

References

- [Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence](#)
- [Data Is Plural](#)
- [Towards a National Definition and Database for Nonfatal Shooting Incidents - PMC](#)
- [UCR Publications — FBI](#)
- [Classification in Machine Learning: A Guide for Beginners | DataCamp](#)
- [Logistic Regression in Machine Learning - GeeksforGeeks](#)
- [Firearm Violence and Injury Prevention - Fast Facts](#)
- [Logistic Regression Image](#)
- [Two FBI Agents Image](#)
- [Gun on Top of Target Image](#)
- [Paramedics Image](#)
- [Lightbulb Image](#)
- [Error Image](#)
- [Game Pieces Image](#)

THANK YOU!