

Methods of Advanced Data Engineering
Winter Semester 2024/25

Analysis of Shootings Incidents and Weather Data

Presented by:
Abantika Bose (23082335)



1. Motivation
2. Introduction
3. Data Sources
4. Methodology
5. Findings
6. Discussion
7. Conclusion
8. Reference

I chose this project to understand the factors influencing shooting incidents in the United States. With over 30,000 incidents recorded between 2015 and 2020, this data provides an opportunity to uncover meaningful patterns. Combining information about demographics, locations, and weather offered a unique way to study these incidents.

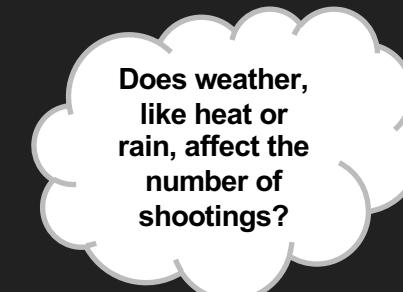
As I explored the data,
Few simple questions came to mind:



Do shootings happen more often at certain times of the year or week?



Are some places more dangerous than others ?



Does weather, like heat or rain, affect the number of shootings?



1. Motivation
- 2. Introduction**
3. Data Sources
4. Methodology
5. Findings
6. Discussion
7. Conclusion
8. Reference

Gun violence remains a pressing concern in the United States, affecting thousands annually. This analysis explores patterns using **demographic, geographic, and weather data** to uncover trends and inform solutions.

This study focuses on three primary dimensions

Temporal Patterns

Spatial Clustering

Environmental Influence

1. Motivation
2. Introduction
- 3. Data Sources**
4. Methodology
5. Findings
6. Discussion
7. Conclusion
8. Reference

Dataset	Source	Original Timeframe	Filtered Timeframe	Attributes	Purpose
Shooting Incidents	Kaggle - US Police Shootings Dataset	2015–2020	2015–2020	Date, city, state, race, gender, armed status, mental illness, body camera use.	Analyse temporal, spatial, and demographic patterns in shooting incidents.
Weather Data	Kaggle - Weather Dataset for the US [United States Daily Climate Observations (GHCN-Daily)]	1992-2021	2015–2020	TMAX, TMIN, PRCP, snowfall, snow depth.	Explore correlations between weather conditions and shooting frequencies.

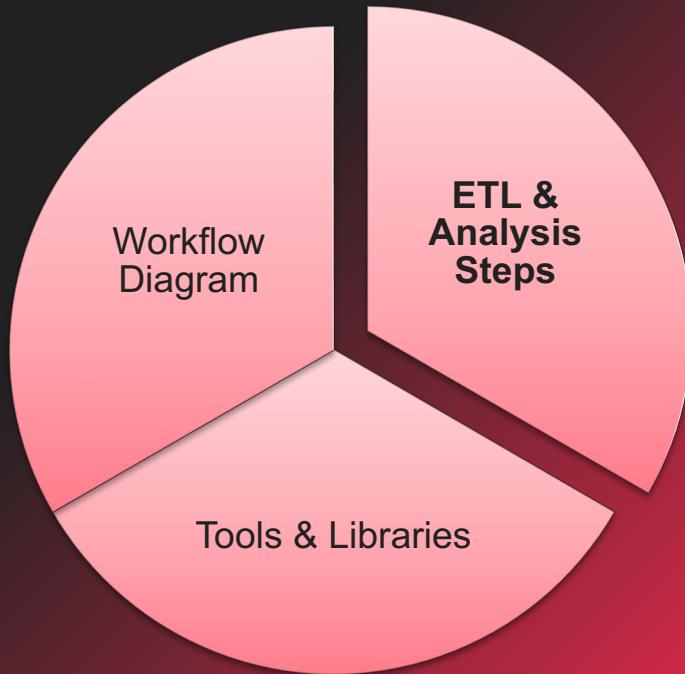
	<code>id</code>	<code>name</code>	<code>date</code>	<code>manner_of_death</code>	<code>armed</code>	\
0	3	Tim Elliot	2015-01-02	shot	gun	
1	4	Lewis Lee Lembke	2015-01-02	shot	gun	
2	5	John Paul Quintero	2015-01-03	shot and Tasered	unarmed	
3	8	Matthew Hoffman	2015-01-04	shot	toy weapon	
4	9	Michael Rodriguez	2015-01-04	shot	nail gun	
...	\
4890	5916	Rayshard Brooks	2020-06-12	shot	Taser	
4891	5925	Caine Van Pelt	2020-06-12	shot	gun	
4892	5918	Hannah Fizer	2020-06-13	shot	unarmed	
4893	5921	William Slyter	2020-06-13	shot	gun	
4894	5924	Nicholas Hirsh	2020-06-15	shot	gun	
...	\
age	gender	race	city	state	signs_of_mental_illness	\
0	53.0	M	Asian	Shelton	WA	True
1	47.0	M	White	Aloha	OR	False
2	23.0	M	Hispanic	Wichita	KS	False
3	32.0	M	White	San Francisco	CA	True
4	39.0	M	Hispanic	Evans	CO	False
...	\
4890	27.0	M	Black	Atlanta	GA	False
4891	23.0	M	Black	Crown Point	IN	False
4892	25.0	F	White	Sedalia	MO	False
4893	22.0	M	White	Kansas City	MO	False
4894	31.0	M	White	Lawrence	KS	False
...	\
threat_level	flee	body_camera	arms_category			
0	attack	Not fleeing	False		Guns	
1	attack	Not fleeing	False		Guns	
2	other	Not fleeing	False		Unarmed	
3	attack	Not fleeing	False	Other unusual objects		
4	attack	Not fleeing	False	Piercing objects		
...	\
4890	attack	Foot	True	Electrical devices		
4891	attack	Car	False		Guns	
4892	other	Not fleeing	False		Unarmed	
4893	other	Other	False		Guns	
4894	attack	Car	False		Guns	

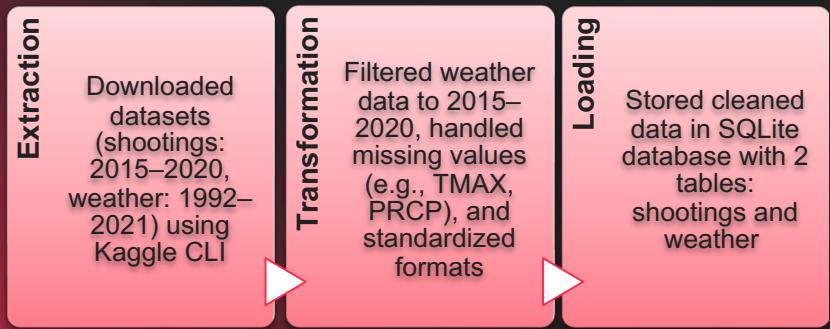
Shootings Data

Weather Data

ID	DATE	TMAX	TMIN	EVAP	PRCP	Latitude	\
0	US1MDHW0012	2015-10-14	189.0	56.0	NaN	0.0	39.3387
1	USC00206510	2018-08-02	267.0	156.0	NaN	28.0	45.3614
2	US1PALH0005	2016-07-20	189.0	56.0	NaN	0.0	40.6474
3	US1NDV0152	2020-01-29	189.0	56.0	NaN	0.0	36.1124
4	US1TXUV0032	2018-12-29	189.0	56.0	NaN	0.0	29.2242
...	\
46230812	US1WIRS0001	2017-03-28	189.0	56.0	NaN	5.0	45.4818
46230813	US1OKCC0009	2018-05-04	189.0	56.0	NaN	5.0	34.5898
46230814	USC00144608	2017-06-26	189.0	56.0	NaN	5.0	38.4161
46230815	US1IACG0011	2019-06-04	189.0	56.0	NaN	5.0	43.1329
46230816	US1NEDG0029	2019-06-15	189.0	56.0	NaN	5.0	41.2140
Longitude	Elevation						
0	-76.9468	166.1					
1	-84.9511	228.0					
2	-75.6505	143.9					
3	-86.8117	191.4					
4	-100.0654	294.7					
...					
46230812	-91.1468	347.5					
46230813	-98.4766	353.9					
46230814	-95.8511	358.1					
46230815	-93.2079	358.1					
46230816	-96.1908	366.1					

1. Motivation
2. Introduction
3. Data Sources
- 4. Methodology**
5. Findings
6. Discussion
7. Conclusion
8. Reference





Temporal

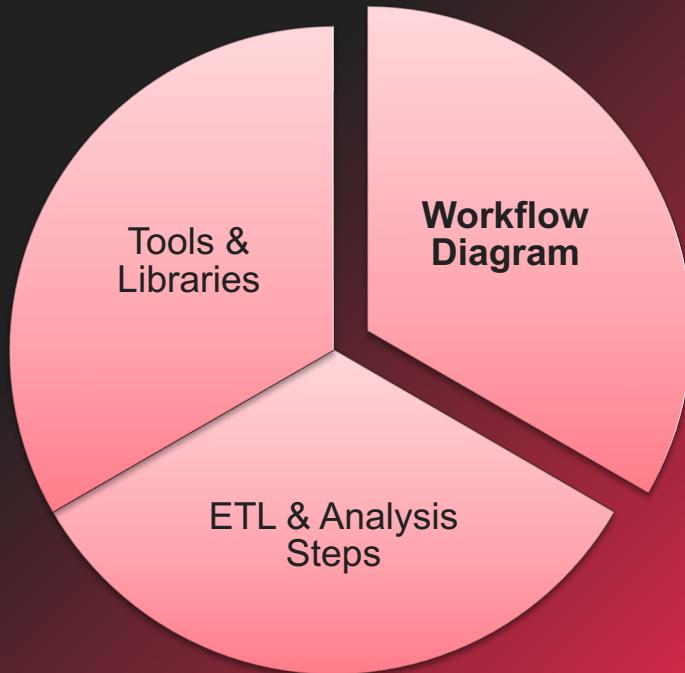
Identified trends by year, month, and day of the week

Spatial

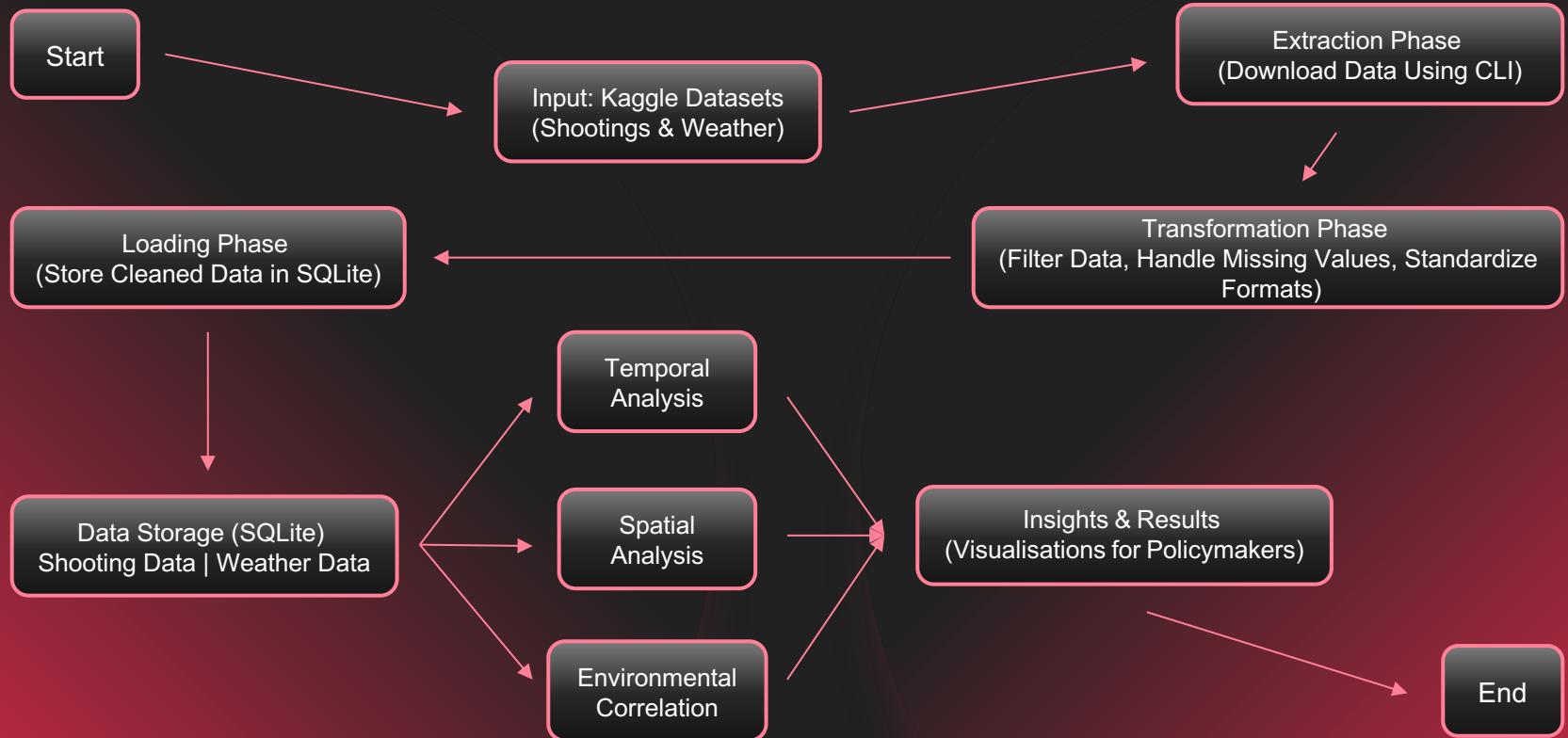
Visualized hotspots using Folium maps and heatmaps

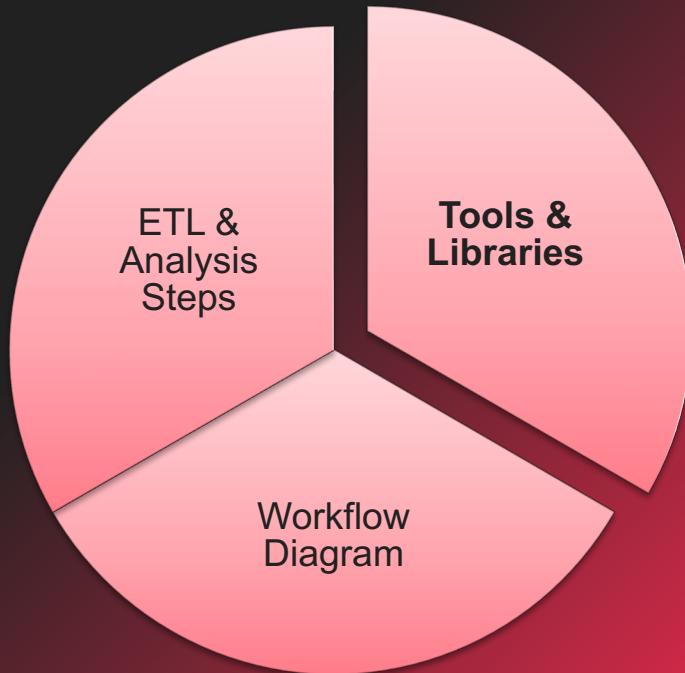
Environmental

Correlated weather variables (e.g., TMAX, PRCP) with shooting frequencies



Workflow Diagram



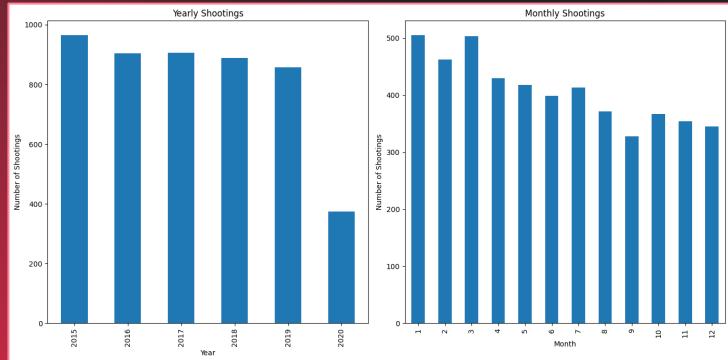
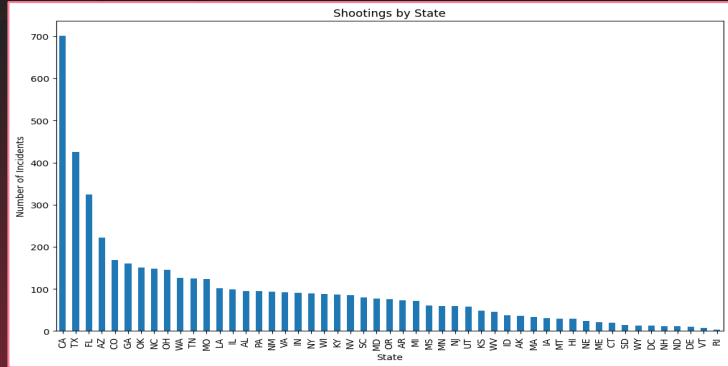


Python: pandas, matplotlib, seaborn,
Folium, sqlite3

Kaggle CLI: For dataset extraction

SQLite: For structured data storage

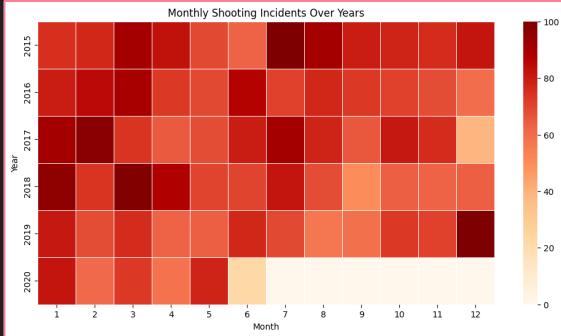
1. Motivation
2. Introduction
3. Data Sources
4. Methodology
- 5. Findings**
6. Discussion
7. Conclusion
8. Reference



QUESTION 1

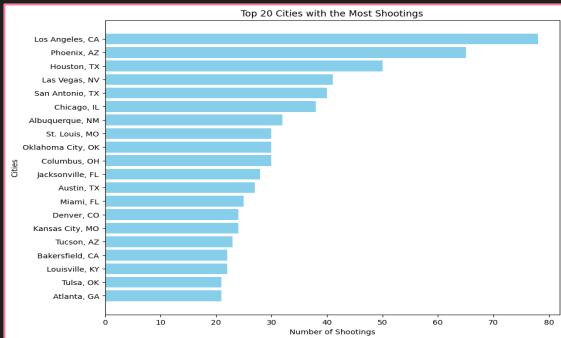
What are the spatial and temporal patterns of shooting incidents, and can these trends help identify high-risk areas and times for potential future incidents?

Shooting incidents peak during January – March, then again during August and are concentrated in high-risk states like California, Texas, and Florida, with significant urban clustering.

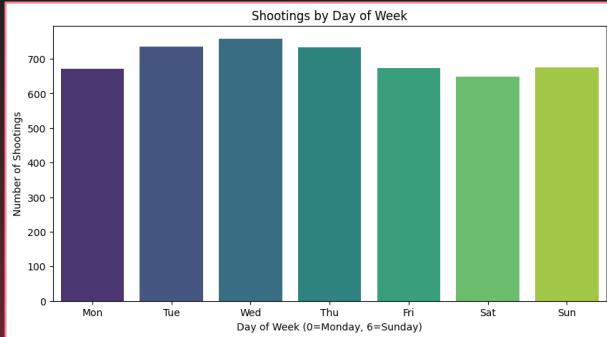


QUESTION 2

What are the trends in shooting incidents across different locations and times?



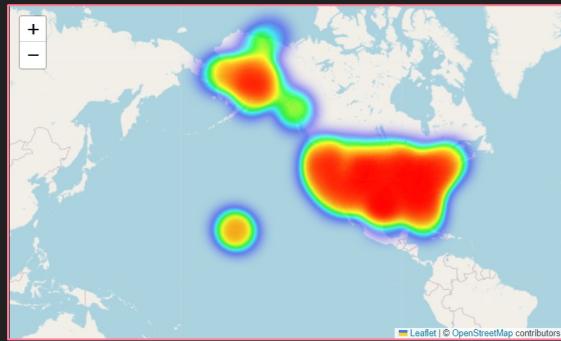
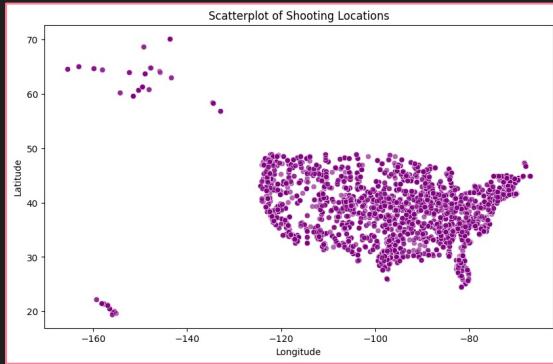
Shootings are more frequent during January – March, then again during August, and cities like Los Angeles, Houston, and Phoenix emerge as urban hotspots for incidents.



QUESTION 3

How do factors like time, day, or location influence the frequency of shootings?

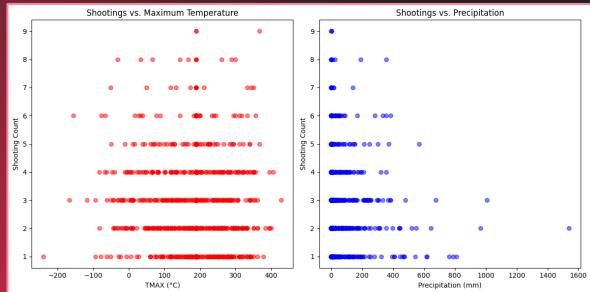
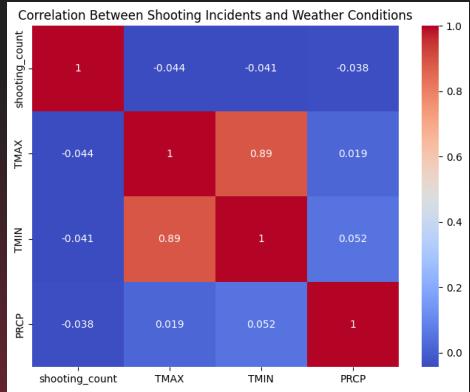
Tuesday – Thursday has slightly higher shooting frequencies, with dense regional clusters observed in urban areas.



QUESTION 4

Can we identify hotspots or predict shooting occurrences based on historical data?

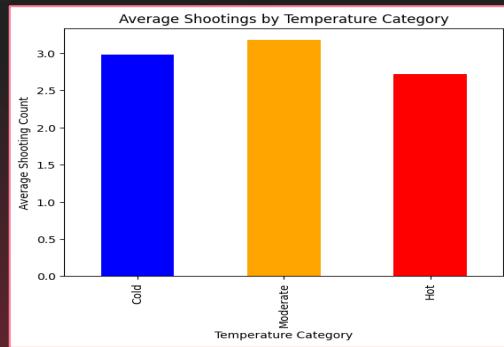
Heatmaps highlight high-risk areas, and scatterplots reveal detailed spreads, aiding hotspot identification and predictive modelling.



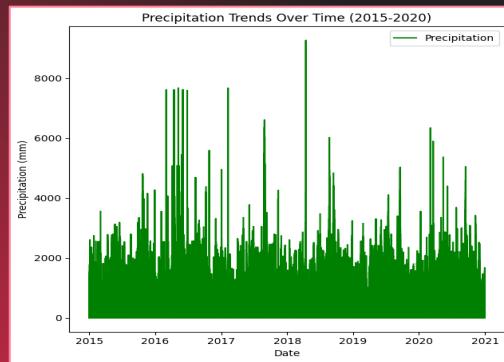
QUESTION 5

How do weather conditions (such as temperature, humidity, and precipitation) correlate with the frequency and distribution of shooting incidents across different locations and times, and can certain weather patterns be linked to higher shooting occurrences?

The correlation matrix shows that as temperature and precipitation increase, shootings decrease. This suggests that warmer or wetter days see fewer shootings, as illustrated in the scatterplots.

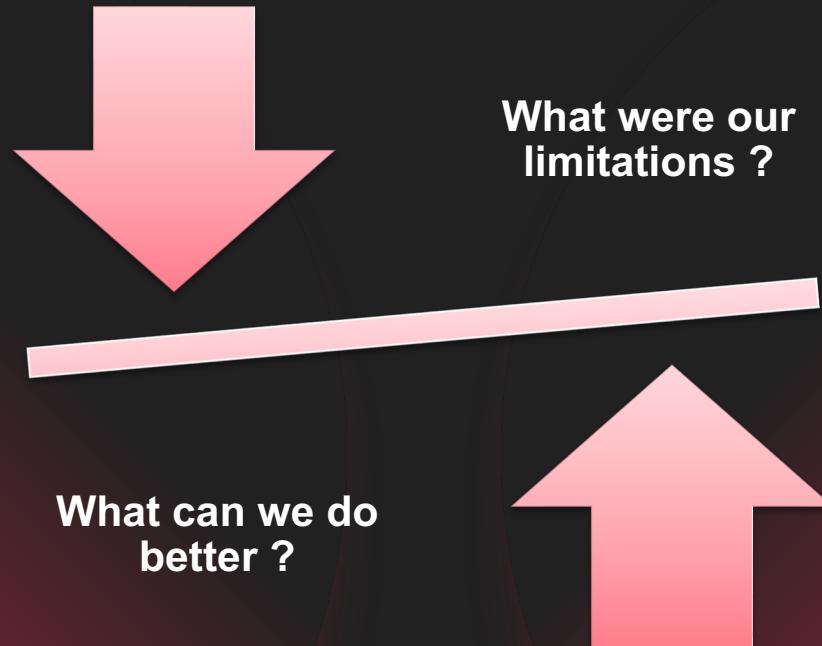


Temperature Trends: Moderate temperatures see the highest shooting rates, with incidents dropping in colder and minimal in hotter conditions.

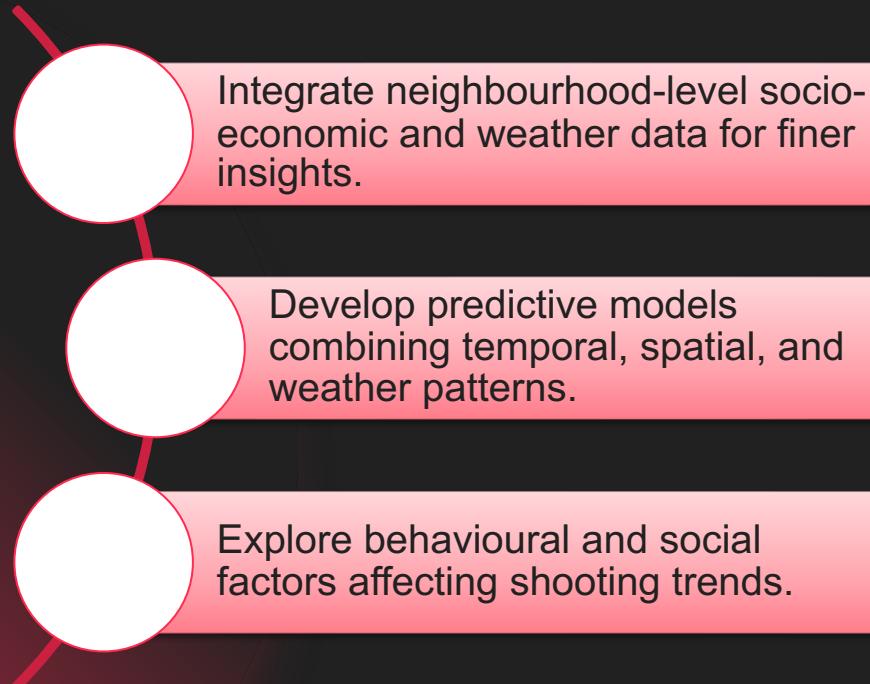


Precipitation Trends: The precipitation graph shows seasonal rainfall variations, with higher rainfall periods correlating with fewer shootings, likely due to reduced outdoor activities.

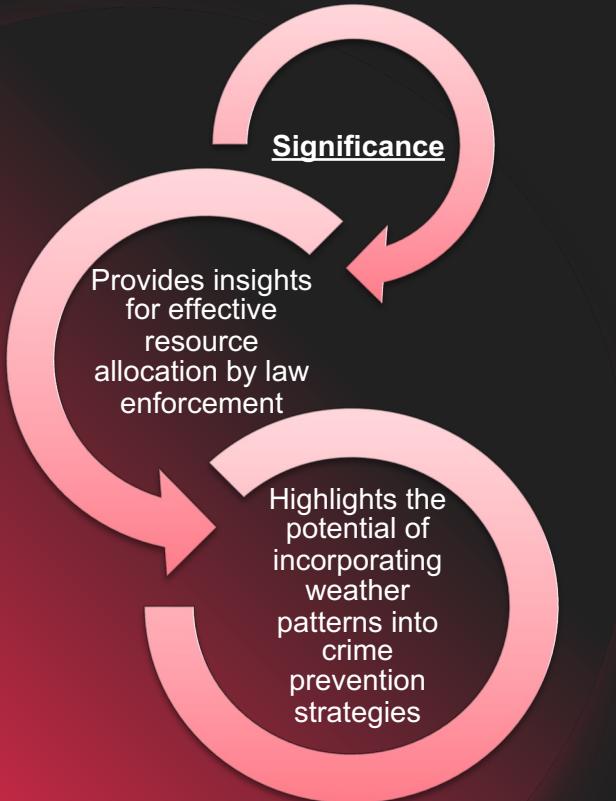
1. Motivation
2. Introduction
3. Data Sources
4. Methodology
5. Findings
- 6. Discussion**
7. Conclusion
8. Reference



Data Granularity	<ul style="list-style-type: none">• Weather data limited to city-level, missing neighbourhood-specific trends.
Data Imputation	<ul style="list-style-type: none">• Filling missing data may introduce biases.
Correlation vs. Causation	<ul style="list-style-type: none">• Findings show trends but don't confirm direct causation.



1. Motivation
 2. Introduction
 3. Data Sources
 4. Methodology
 5. Findings
 6. Discussion
 - 7. Conclusion**
 8. Reference
-



Key Takeaways:

- ❖ **Temporal Patterns:** Shootings peak January–March and mid-week.
- ❖ **Spatial Trends:** Hotspots in California, Texas, and urban areas.
- ❖ **Environmental Influence:** Lower temperatures increase incidents; rain reduces them slightly.

Implications:

- ❖ Encourage targeted interventions in high-risk areas and times.
- ❖ Address socio-economic factors in identified hotspots to reduce long-term incidents.

1. Motivation
2. Introduction
3. Data Sources
4. Methodology
5. Findings
6. Discussion
7. Conclusion
8. Reference

Datasets:

- ❖ Kaggle - [US Police Shootings Dataset](#)
- ❖ Kaggle - [Weather Data US](#)

The background of the slide features a movie clapperboard. The clapperboard is dark with white text. At the top, it says "PROD NO.", "SCENE", "TAKE", "ROLL", and "SOUND". In the center, there is a large number "4". The word "CRIME" is partially visible at the bottom left, and "TIME" is partially visible at the bottom right. A large, solid pink circle is positioned in the center-left area of the slide, overlapping the clapperboard.

Thank you for listening !!