



# DRUGGED OUTLIERS

This project aims to detect and visualize outliers or unusual spikes in the number of drug overdose deaths.

**DATA BATCH 9**



# RAW DATASET LINK

<https://drive.google.com/file/d/1HZnECLKLEOFIYqJ8dGZlodKVwzKwGqTZ/view?usp=sharing>





# PROJECT OVERVIEW

This project, titled "DruggedOutlier", focuses on detecting and analyzing anomalous spikes (outliers) in weekly drug overdose deaths across U.S. states using early model-based provisional death estimates from CDC. The analysis aims to identify unusual surges in overdose mortality, which could signal critical public health events or emerging drug crises.

Using the official dataset of weekly predicted and actual deaths, the project flags weeks where actual overdose deaths exceed the predicted upper bound as outlier events. The analysis incorporates both temporal trends and state-wise distribution of these anomalies to support early intervention and decision-making.





# ANALYSIS PROCESS

01

DATA  
PREPARATION

03

OUTLIER  
DETECTION

05

INTERPRETATION

02

DATA  
SUMMARIZATION

04

VISUALIZATION



## INSIGHT & FINDINGS

An analysis of drug overdosed death data across all US states identified 147 significant drug outliers, based on predictions from a rolling 4-week upper bound model. These spikes were detected spatially and temporally, suggesting a pattern of exceptional events in some regions.

States with the highest number of spikes:

- United States (national level): 63 outliers
- Florida: 14 outliers
- Ohio: 10 outliers
- California: 8 outliers
- Pennsylvania: 7 outliers
- New York: 6 outliers



## INSIGHT & FINDINGS

Meanwhile, more than 20 states showed no spike in deaths at all during the observation period.

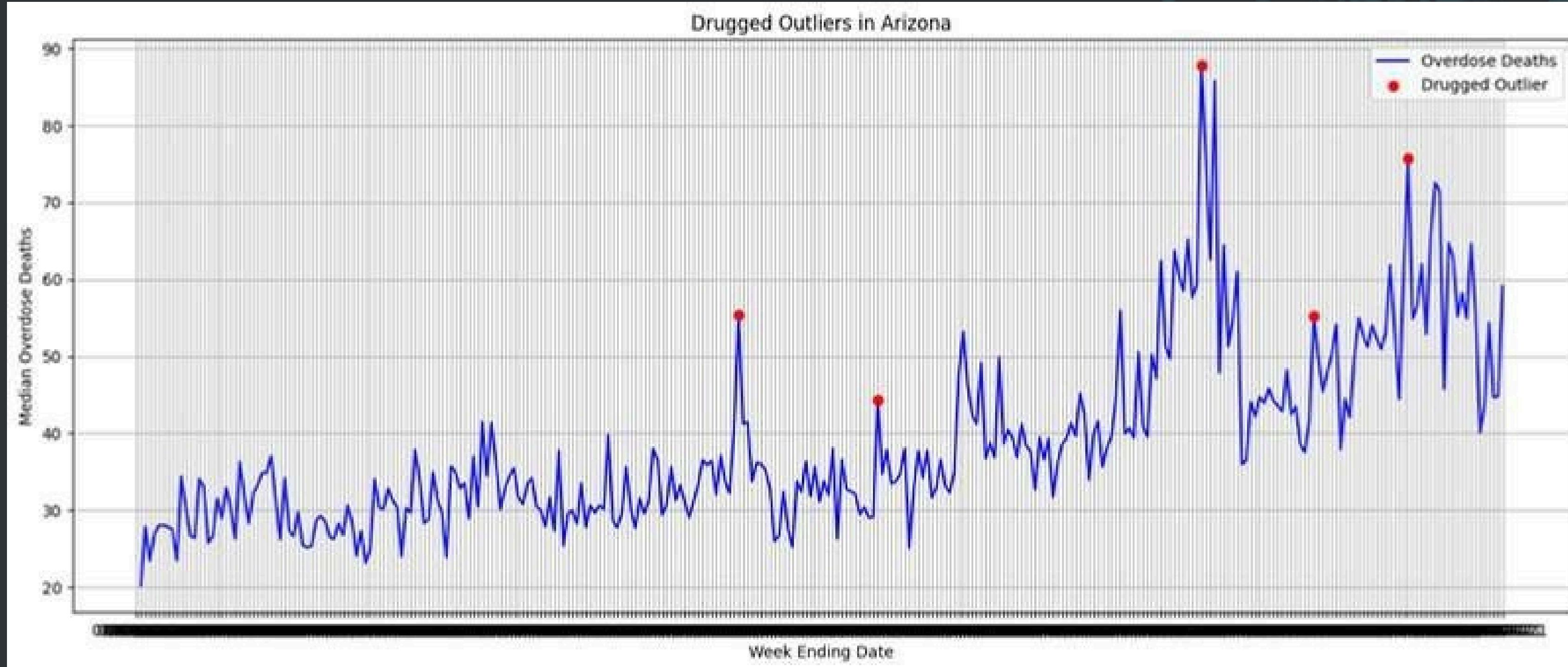
Spikes tend to occur in states with large populations and drug-prone areas.

Spatial analysis shows a concentration of spikes in the eastern and southern United States.

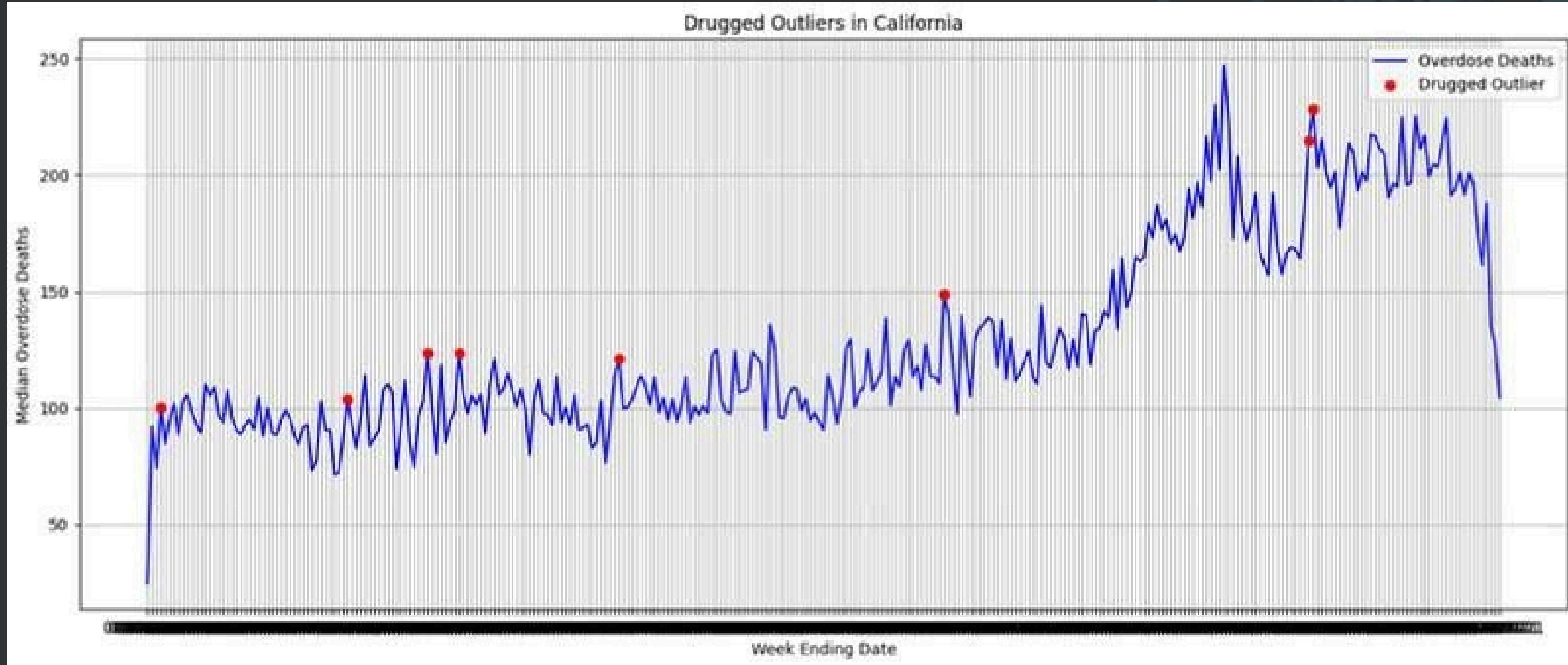
Rolling average-based predictive models are effective in identifying crisis periods by detecting actual deaths that exceed the upper limit of predictions.



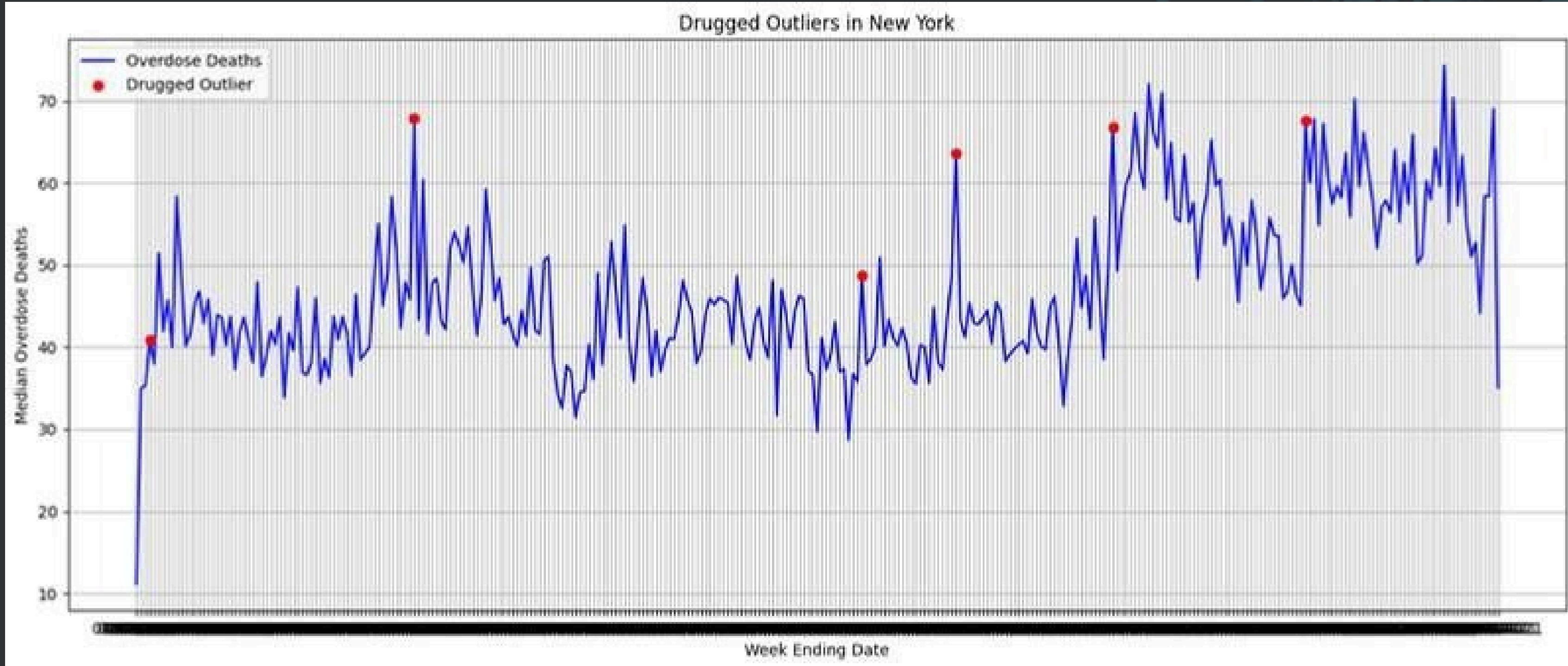
# VISUALIZATION



# VISUALIZATION



# VISUALIZATION





## CONCLUSION & RECOMMENDATIONS

Analysis of the drug overdose death dataset revealed 147 outliers (unpredicted spikes in deaths) spread across various states. These findings highlight a potential hidden health crisis in several regions, particularly Florida, Ohio, and California. The 4-week rolling upper bound method was effective in detecting these extraordinary events.

- Prioritize interventions in states with the highest number of outliers, such as Florida and Ohio.
- Implement regular predictive monitoring using similar models for early detection.
- Integrate the results of this outlier analysis into the national drug control strategy.



# AI SUPPORT

Initially, I used the IBM Granite 3.3-8B-Instruct model to assist in the analysis of this dataset. However, during the implementation process, the model encountered technical difficulties, particularly in output parsing and basic command execution such as data summarization or visualization (e.g., when displaying data or creating plots). As a result, I was unable to optimally implement the new LLM-related knowledge directly in Python.





**"OUTLIERS  
HIGHLIGHT WHAT  
THE AVERAGE  
CONCEALS."**

Thank you for following this project journey!

