

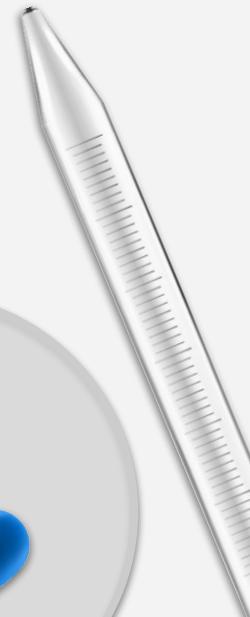
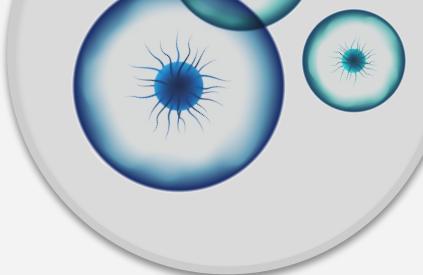
# **Analysis of COVID-19 Case Rates in Correctional Facilities vs. States in US**

Methods of Advanced Data Engineering (MADE WS2024/25)

**Harshvardhan Joshi  
og52yvov; 23162173**

# Introduction

**Aim:** Examine COVID-19 case rates in correctional facilities versus state-level data; and to identify patterns to manage outbreaks in high-risk settings/environments.



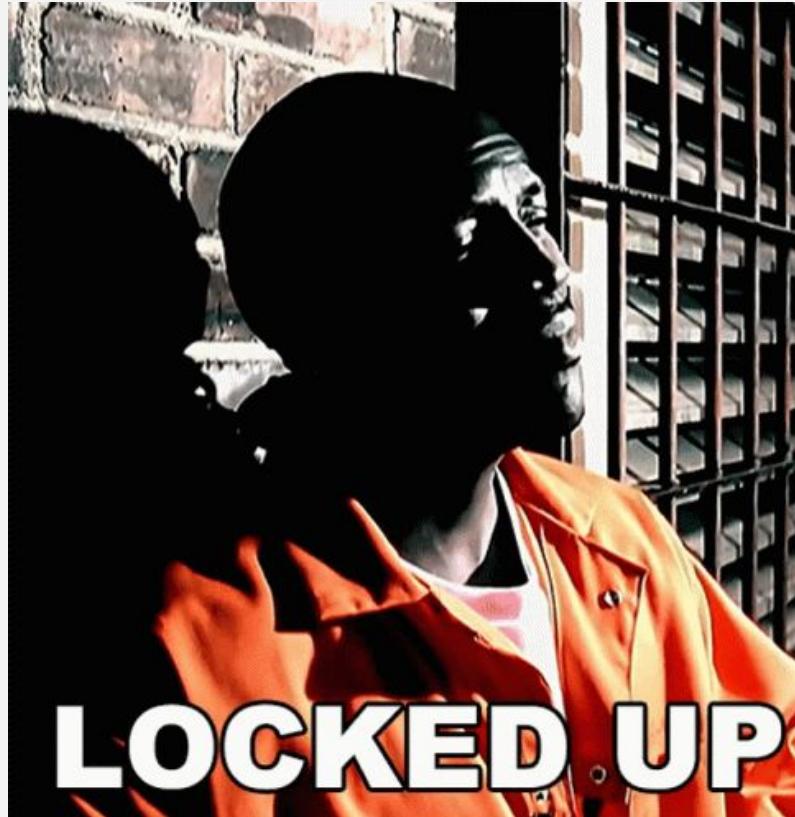
# Data Description

**Data Sources:** Utilized state public health databases

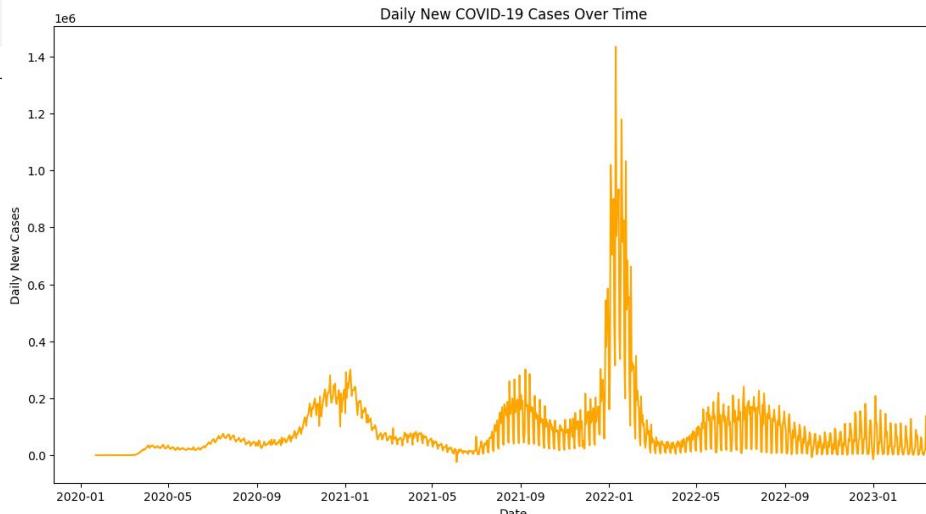
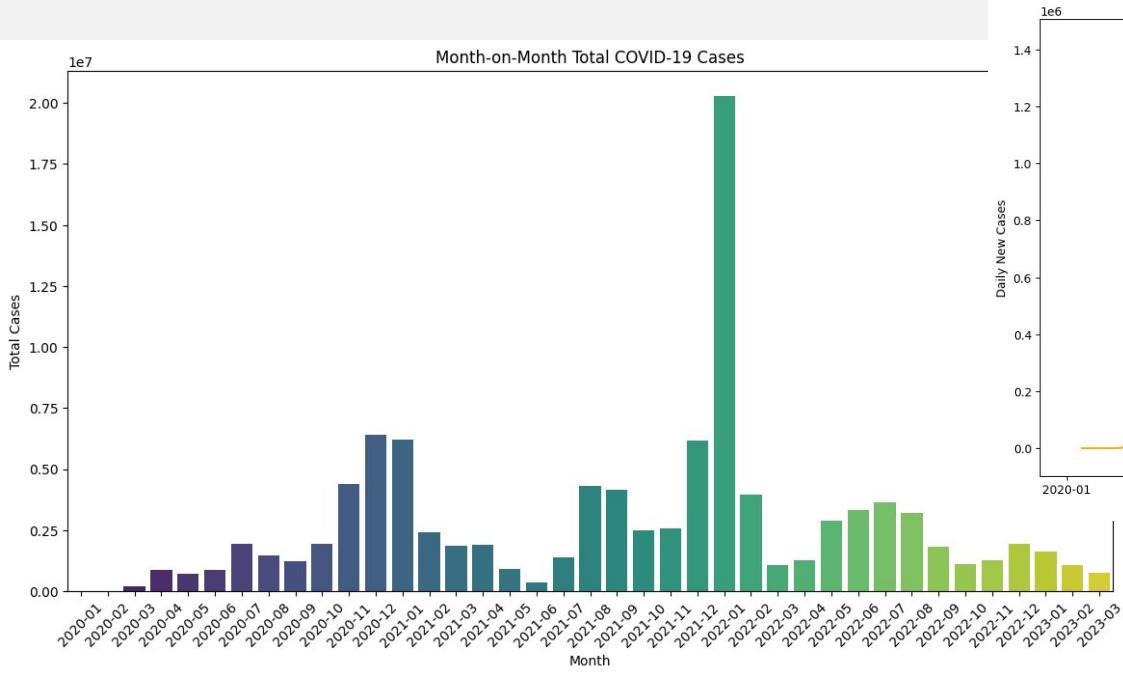
Using SQLite as a database as databases are stored in a single file. It makes it easy to manage and transport the database file across systems or env's without a server.

## Datasets Used:

- **Facilities Table:** Case statistics for individual facilities.
- **Systems Table:** Health infrastructure and measures.
- **US Data Table:** data for the entire United States, tracking the daily cumulative count of COVID-19 cases and deaths
- **US States Table:** Specific state-level case trends.



# How data in US is distributed:



Daily New Cases Over Time: This graph tracks the daily new COVID-19 cases. Also shows fluctuations and peaks which may correspond to sudden outbreaks.

Month-on-Month Total Cases: Displays the month-to-month changes in COVID-19 case totals. It shows the last months of 2021 with the highest surges or declines, useful for assessing the impact of interventions over time.

# Methodology



## Data Aggregation

Aggregate COVID-19 case data by state and facility.



## Standardization

Calculating case rates per 100,000 population to ensure comparability.

As directly comparing these raw counts of cases across populations of different sizes can be misleading

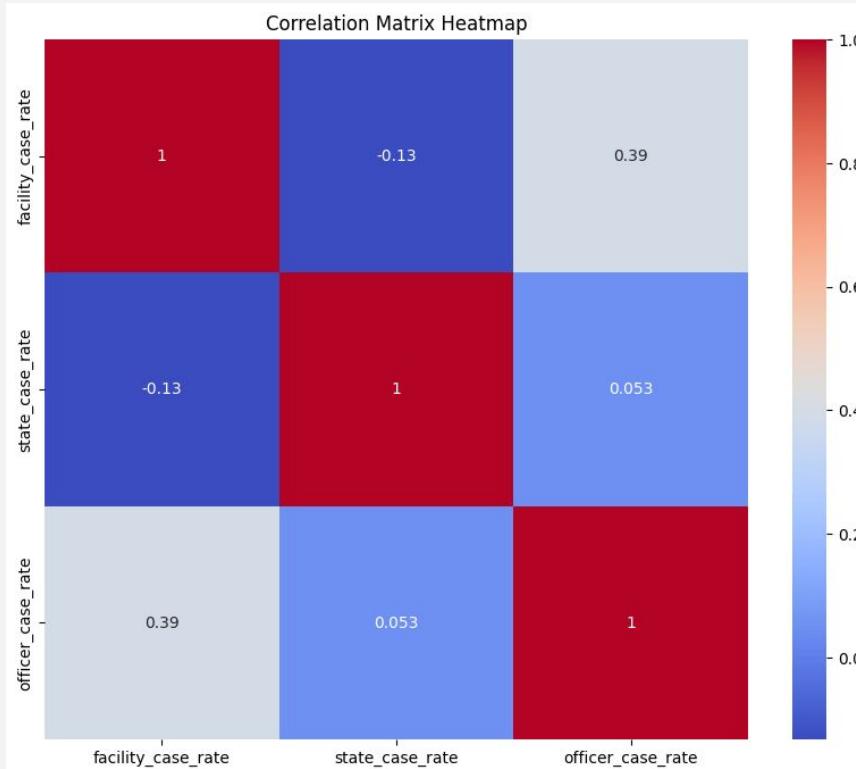


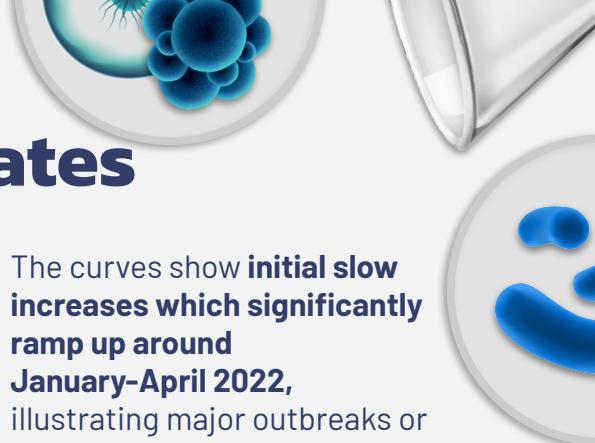
## Analysis Tools:

Use SQLite for data querying and processing.  
Using Python libraries for data analysis and visualization.

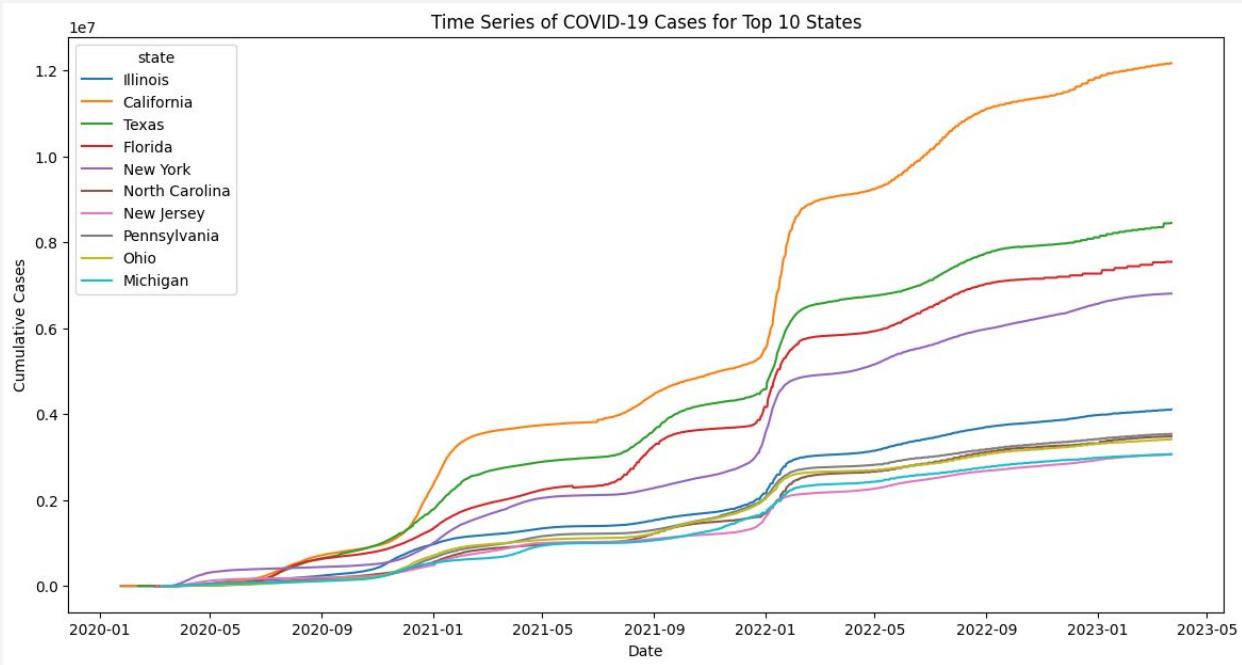
# Correlation Matrix and Heatmap

- **Positive correlation between facility case rates and officer case rates :** There is a moderate positive correlation of 0.39, showing shared risk factors between facilities.
- **Negative correlation between facility and state :** Negative correlation of -0.13 indicates that outbreaks in facilities may not mirror broader state infection patterns.
- **Trends between officer case rates and state case rates:** Spread among correctional officers is relatively independent of state trends.





# Time Series Analysis of Top 10 States

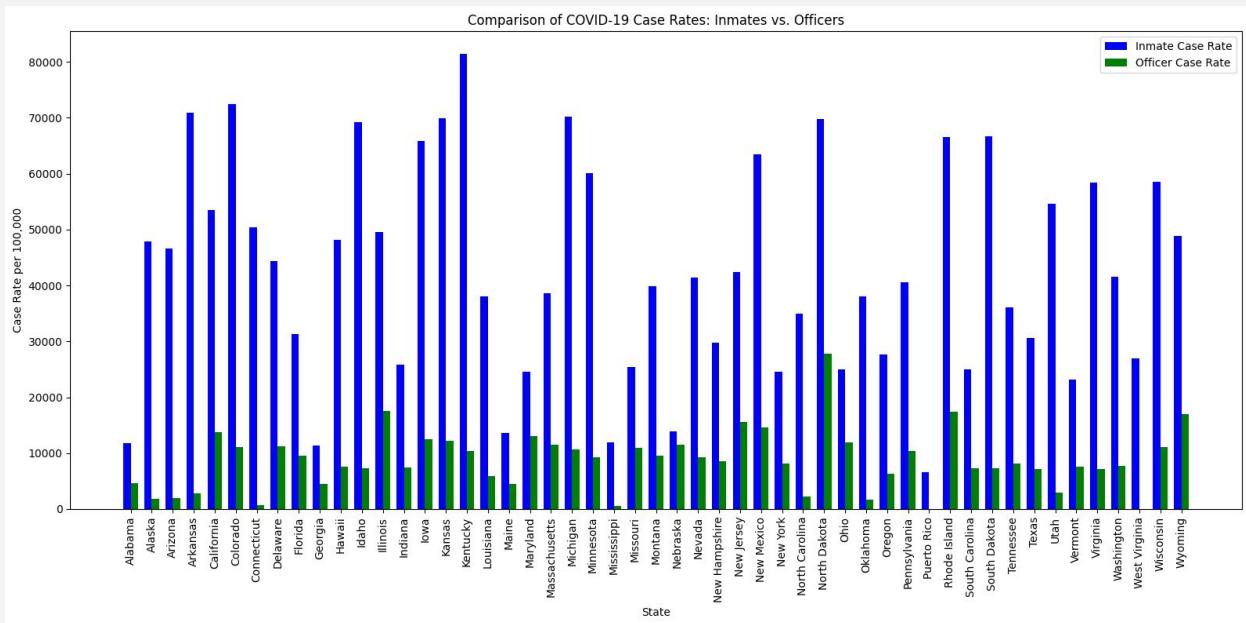


The curves show **initial slow increases which significantly ramp up around January-April 2022**, illustrating major outbreaks or changes in testing and reporting.

States with steeper curves may indicate delayed or less effective containment measures, and flatter curves might suggest earlier intervention and control. (but population is a major factor to play role in this)

California, Texas and Florida shows very high peaks, representing large outbreaks (likely due to their population sizes and density).

# Comparative Case Rates Inmates vs Officers



It shows that more severe conditions are faced by inmates.

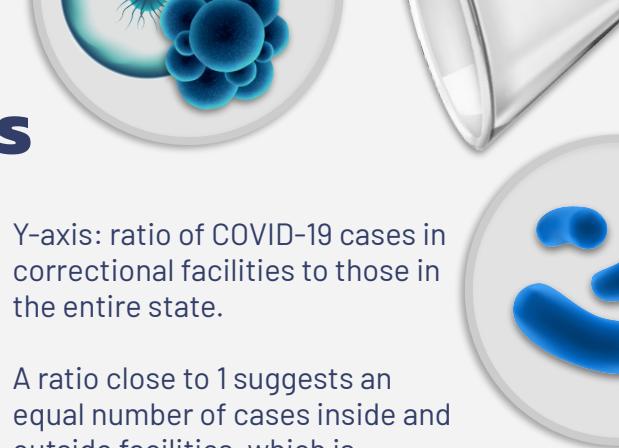
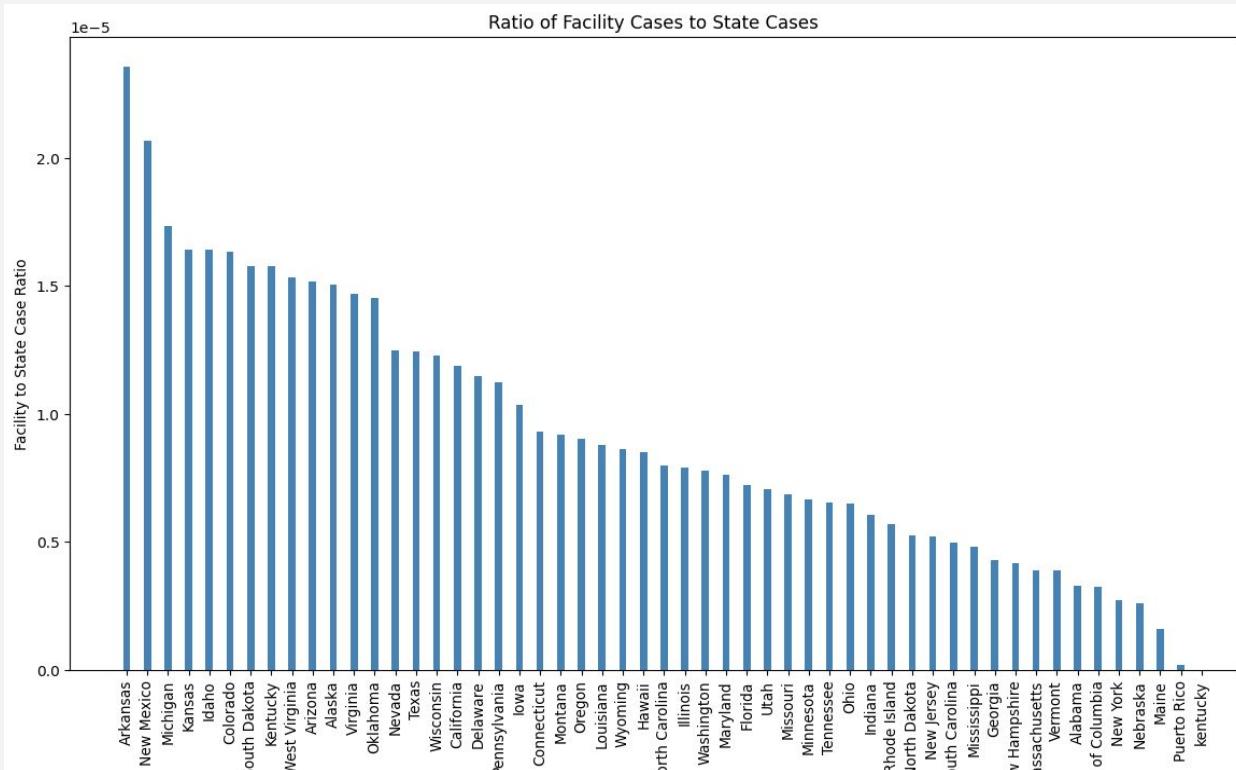
Prisons may **experience overcrowding and limited space for social distancing**.

Added with a lack of personal protective equipment, leads to poorer health conditions.

These factors drive higher case rates among inmates. It shows the need for targeted health interventions within prisons to manage and prevent the spread of COVID-19.

The graph compares COVID-19 case rates per 100,000 for inmates versus officers across various states.

# Facility to State Case Ratio Analysis



Y-axis: ratio of COVID-19 cases in correctional facilities to those in the entire state.

A ratio close to 1 suggests an equal number of cases inside and outside facilities, which is uncommon.

Higher ratios indicate many more cases within facilities than in the broader state population. Lower ratios suggest fewer cases within facilities.

Identifying states with higher facility-to-state case ratios, public health officials can prioritize focused testing, containment, and treatment efforts in these areas.

This graph shows the concentration of COVID-19 cases within correctional facilities compared to the entire state population

# Conclusion :

This analysis explores COVID-19 case rates in correctional facilities compared to state data, showing various impacts.

Factors like overcrowding and inadequate health measures may contribute to higher case rates within these facilities. It highlights the challenges of managing outbreaks in such environments.

The findings the need for precise health interventions for these settings. Strategic policies and decisions can help reduce the virus's spread in such hotspots, protecting inmates and staffs.

