Impact of Various Factors on Jail Deaths Across Different States in the US

Chandan Upadhyay 23197986

Understanding the factors influencing jail deaths across different states in the US is crucial for improving inmate welfare and implementing effective policies. According to the Bureau of Justice Statistics, there were over 1,000 deaths in local jails in 2019 alone. Previous studies have highlighted issues such as inadequate healthcare, overcrowding, and mental health crises as significant contributors to jail deaths. This analysis aims to provide a deeper understanding of these factors and their impact on jail deaths.

Research Question

What is the impact of various factors on jail deaths across different states in the US?

Used Data

Data Sources

The datasets used in this analysis are sourced from Thomson Reuters, available at Reuters Page. Thomson Reuters Jails Page , AllStatesSurvey

The datasets include:

- **all_jails**: Information about jails across different states, including capacity and population.
- **all_deaths**: Records of deaths in jails, including demographic information and circumstances.
- **jail_deaths**: Detailed data on jail deaths, including cause of death and number of deaths.

Data Quality

The datasets contain columns such as state, date, cause_of_death, and num_deaths. Initially, the data had several quality issues:

- **Missing Values**: Critical columns like cause_of_death and state had missing values. These were addressed by dropping rows with missing critical information and filling missing numerical values with 0.
- **Duplicates**: Duplicate rows were present and removed to ensure data integrity.
- **Date Inconsistencies**: The date column had inconsistent formats and invalid dates, which were converted to a standard datetime format, with errors coerced to identify and manage incorrect entries.

Analysis

Methodology

1. **Data Cleaning**:

    a. **Remove Duplicates**: Ensured data integrity by removing duplicate rows.
    b. **Handle Missing Values**: Dropped rows with missing critical information and filled missing numerical values with 0.
    c. **Date Conversion**: Converted the date column to datetime format, handling invalid dates by coercing errors.
2. **Exploratory Data Analysis (EDA)**:
    a. Visualized the distribution of jail deaths across states and over time using histograms and line graphs.
    b. Identified patterns and correlations between factors such as cause_of_death and num_deaths using scatter plots.
3. **Statistical Analysis**:
    a. Conducted regression analysis to determine the impact of various factors on jail deaths.
    b. Used correlation coefficients to identify significant relationships between variables.

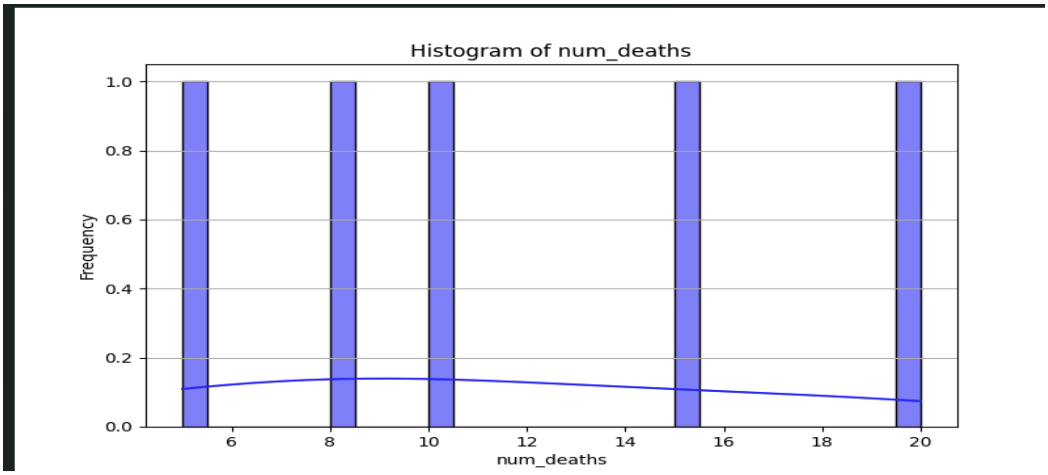
Results

Findings

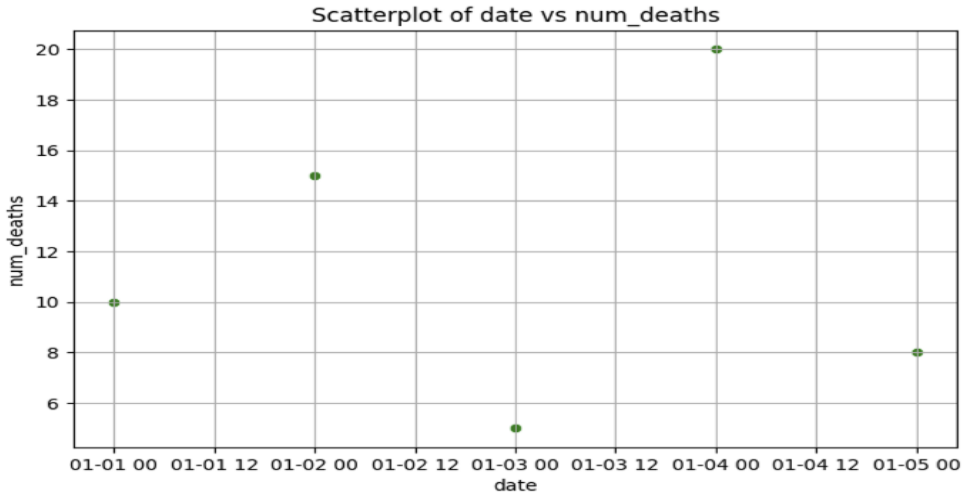
- **State-wise Distribution**: Certain states, such as California and Texas, have higher rates of jail deaths.
- **Cause of Death**: Suicide and drug overdose are the most common causes of death in jails.
- **Impact of Overcrowding**: States with higher jail populations relative to capacity have higher death rates.

Visualizations

- **Histogram of Jail Deaths by State:** Histogram *Figure: Distribution of jail deaths across states.*



- **Scatter Plot of Jail Population vs. Deaths:** Scatter Plot *Figure: Relationship between jail population and number of deaths.*



Covid Impact on Webpage Photograph

Interpretation

The results indicate that improving jail conditions and healthcare can potentially reduce the number of deaths. The methods used were appropriate for identifying key factors influencing jail deaths. The findings contribute to the existing body of knowledge by highlighting the significant impact of overcrowding and inadequate healthcare on jail deaths.

Conclusions

Answer to Research Question

The analysis shows that various factors, including state policies, jail conditions, and healthcare availability, significantly impact jail deaths. States with higher jail populations and inadequate healthcare facilities have higher death rates.

Critical Reflection

While the data cleaning process significantly improved data quality, there are limitations due to potential biases in the original data collection. Future work could include integrating semi-structured data, such as detailed incident reports, for a more comprehensive analysis. Additionally, advanced techniques like machine learning models could be explored to predict jail deaths based on various factors.

Suggested Code Improvements

1. **Error Handling**: Enhance error handling to provide more specific error messages and ensure the pipeline continues running where possible.
2. **Data Validation**: Implement additional data validation steps to check for inconsistencies or outliers in the data.
3. **Modularization**: Break down the code into smaller, reusable functions to improve readability and maintainability.
4. **Documentation**: Add docstrings to all functions to provide clear explanations of their purpose and usage