Data Report on the Impact of COVID-19 in Correctional Facilities

1. Introduction This report examines the effects of COVID-19 in correctional facilities. Due to their crowded conditions, these facilities face unique challenges. This study compares infection and death rates in these facilities with the general population, to identify key differences and propose focused health responses.

2. Research Questions

- How do infection rates in correctional facilities compare with those in the general population in similar areas?
- Are death rates in correctional facilities higher than in the general population, and if so, how much higher?

3. Data Sources : Source Description

- **Facilities Data:** Provides details on COVID-19 cases and deaths among inmates and staff at each facility. Link
  data_facilities.head()

Facilities Table:

	nyt_id	facility_name	facility_type	facility_city	facility_county	facility_county_fips	facility_state	facility_lng	facility_lat	latest_inmate_population	max_inmate_population_2020	total_inmate_cases	total_inmate_deaths	total_officer_cases	total_officer_deaths	note
0	F3EFE858	Alex City Work Release prison	Low-security work release	Alex City	Coosa	1037	Alabama	-86.009015	32.904507	188.0	NaN	77	0	17	0.0	None
1	5B910220	Alabama Therapeutic Education Facility prison	State rehabilitation center	Columbiana	Shelby	1117	Alabama	-86.624067	33.180755	272.0	NaN	11	1	2	0.0	None
2	02FB1675	Bibb Correctional Facility	State prison	Brent	Bibb	1007	Alabama	-87.162781	32.920754	1725.0	1825.0	164	3	61	0.0	None
3	6378F6C4	Birmingham Women's Community Based Facility an...	State prison	Birmingham	Jefferson	1073	Alabama	-86.808344	33.531101	192.0	NaN	17	0	28	0.0	None
4	EAABF900	Bullock Correctional Facility	State prison	Bessemer	Bullock	1011	Alabama	-85.673927	32.147144	1477.0	1577.0	162	5	80	1.0	None

- **Systems Data:** Provides aggregated data on tests, cases, and outcomes at the state system level. Link
  data_systems.head()

Systems Table:

	system	inmate_tests	total_inmate_cases	total_inmate_deaths	latest_inmate_population	max_inmate_population_2020	total_officer_cases	total_officer_deaths
0	Alabama	15505.0	1601	64	19144.0	21900.0	1019.0	3.0
1	Alaska	27756.0	2428	5	5011.0	5143.0	85.0	0.0
2	Arizona	43652.0	12229	44	38248.0	41674.0	2739.0	0.0
3	Arkansas	0.0	11378	52	14729.0	15886.0	424.0	5.0
4	California	121207.0	49511	217	110471.0	117767.0	16090.0	26.0

- **US General Data:** Provides data on nationwide COVID-19 cases and deaths for comparison. Link
  data_us_data.head()

US Data Table:

	date	cases	deaths
0	2020-01-21	1	0
1	2020-01-22	1	0
2	2020-01-23	1	0
3	2020-01-24	2	0
4	2020-01-25	3	0

3.2 Structure and Quality Our data is structured in organized relational tables featuring numeric and categorical types. Verified for completeness and accuracy. Missing data points are addressed by techniques to maintain data consistency.

3.3 Licensing and Compliance The data was sourced from public government records, compliant with open data licenses allowing free academic use. We maintain full transparency and adherence to these licenses, ensuring that all usage meets the necessary legal requirements. This dataset is governed by terms that permit copying, distribution, display, and derivative works, provided that such use is non-commercial and includes appropriate credit to The New York Times. Detailed terms of this licensing can be seen here Link.

4. Data Pipeline : Python and SQLite are employed for data management, with Pandas for data manipulation and Matplotlib for visualizations, enhancing our analytical capabilities.

4.2 Pipeline Overview

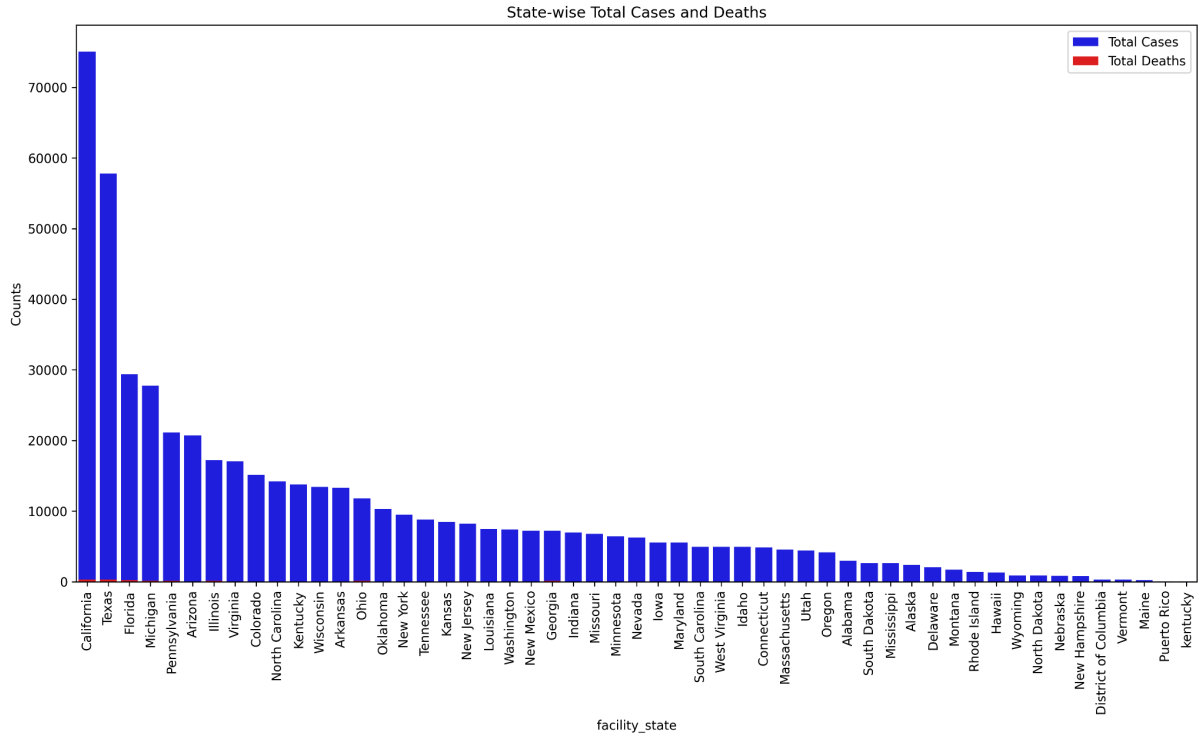
- **Extraction**: Automated scripts fetch data directly from public URLs using the `requests` library, ensuring a reliable data flow from source to our servers.
- **Transformation**: Data is cleansed and preprocessed with Pandas in Python. Key operations include standardizing date formats, normalizing geographic data, and removing duplicates to ensure data quality.
- **Loading**: Cleaned data is then stored back into structured SQLite databases, ready for comprehensive analysis.

4.2 Challenges and Solutions Handling missing data was a significant challenge for me. I addressed this by employing median imputation to ensure data accuracy without bias.

4.3 Meta-Quality Measures Error logging is implemented to address data inconsistencies, with continuous updates to our data handling rules ensuring reliability as data collection methods evolve.

5. Results and Analysis

"State-wise Total Cases and Deaths" -



Data Structure and Quality

Data is well-organized in a SQLite database, making it suitable for deep analysis. This structure ensures reliable performance and maintains data integrity, which is necessary for handling detailed queries effectively.

Data Format and Rationale

I chose SQLite because it is easy to use, carries well across different platforms, and works seamlessly with Python. This compatibility is beneficial for performing extensive data analyses without complications.

Currently, there are no anticipated issues with the data. The system is running smoothly, and the data management processes are stable.