

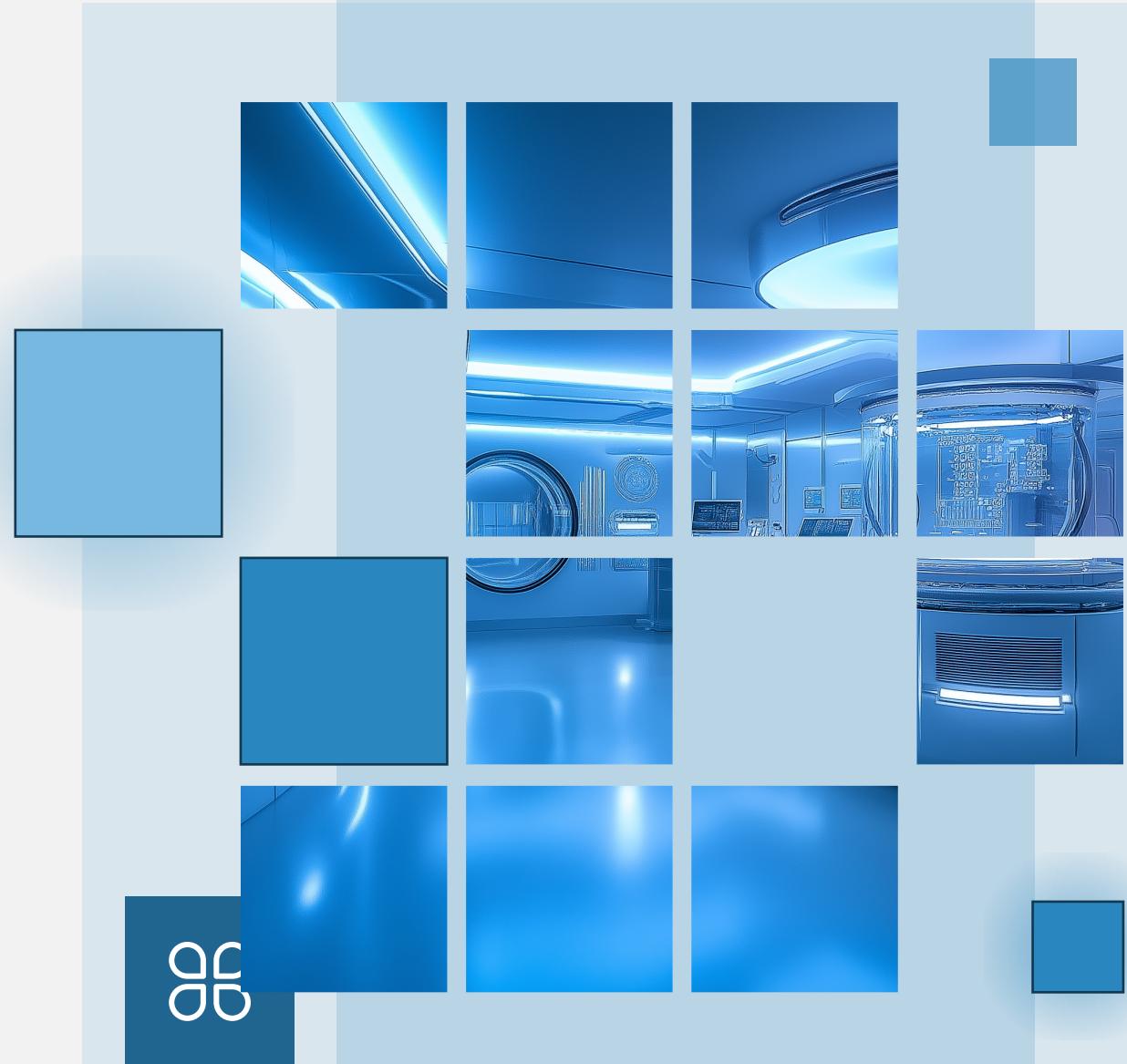
Socioeconomic Factors and Healthcare Burden during COVID-19 in the Americas



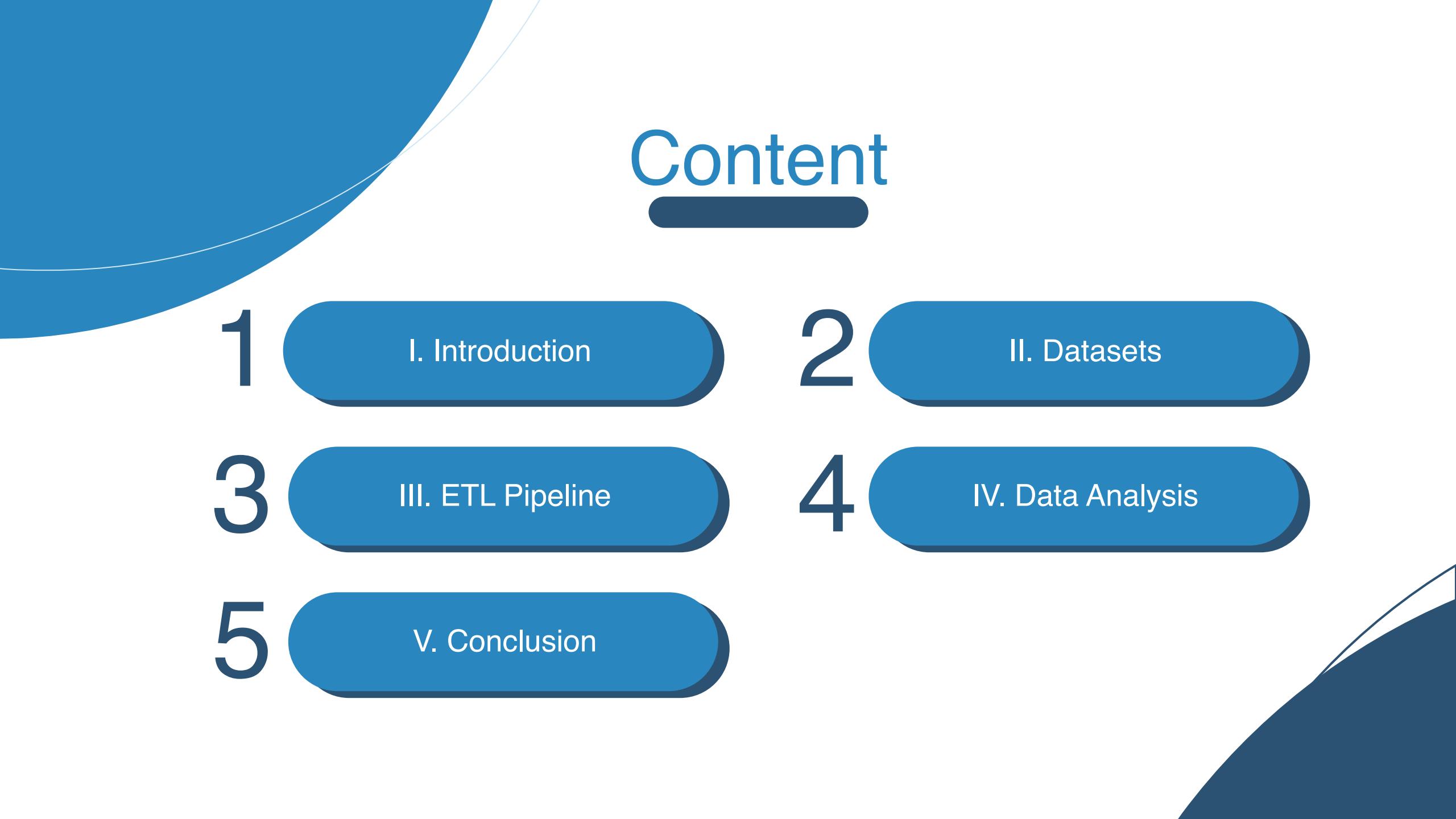
Yulin Wang



24.01.2025



Content



1

I. Introduction

3

III. ETL Pipeline

5

V. Conclusion

2

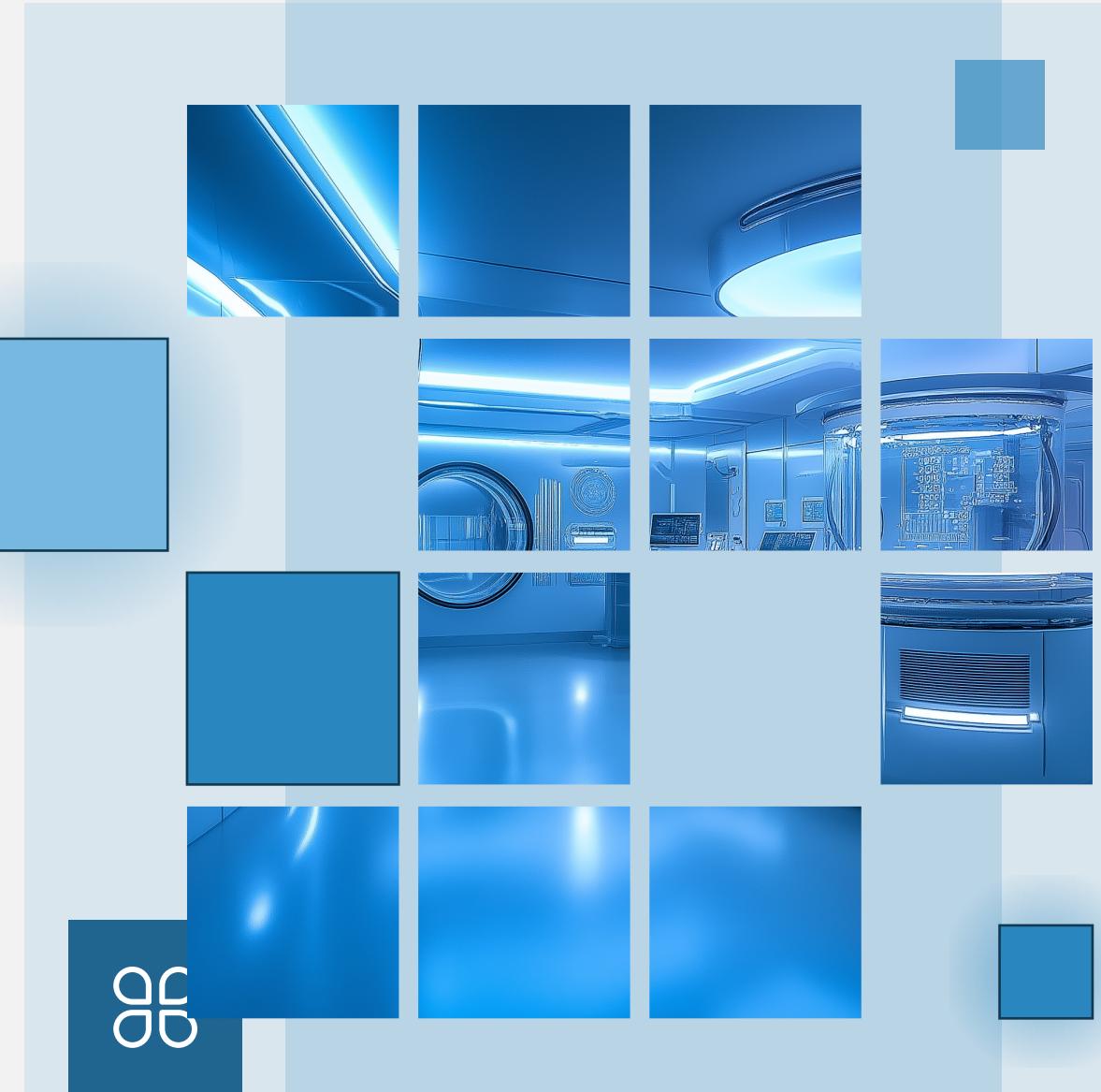
II. Datasets

4

IV. Data Analysis

01

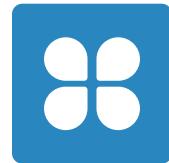
I. Introduction



Research Background



COVID-19 Impact on the Americas



COVID-19 has caused severe health and economic challenges across the Americas. Countries with diverse socioeconomic conditions have experienced varying impacts, highlighting significant inequalities in how the pandemic has been managed.

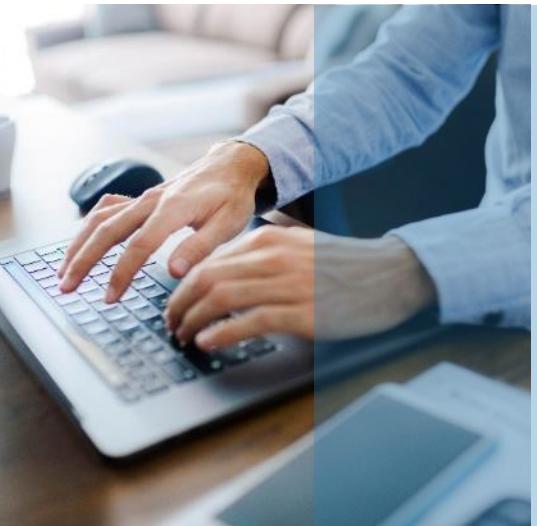


During the early stages of the pandemic, densely populated urban areas faced rapid virus spread, overwhelming healthcare systems. As the virus expanded into suburban and rural regions, limited resources further strained the capacity to respond effectively.

Research Objectives

Main Question

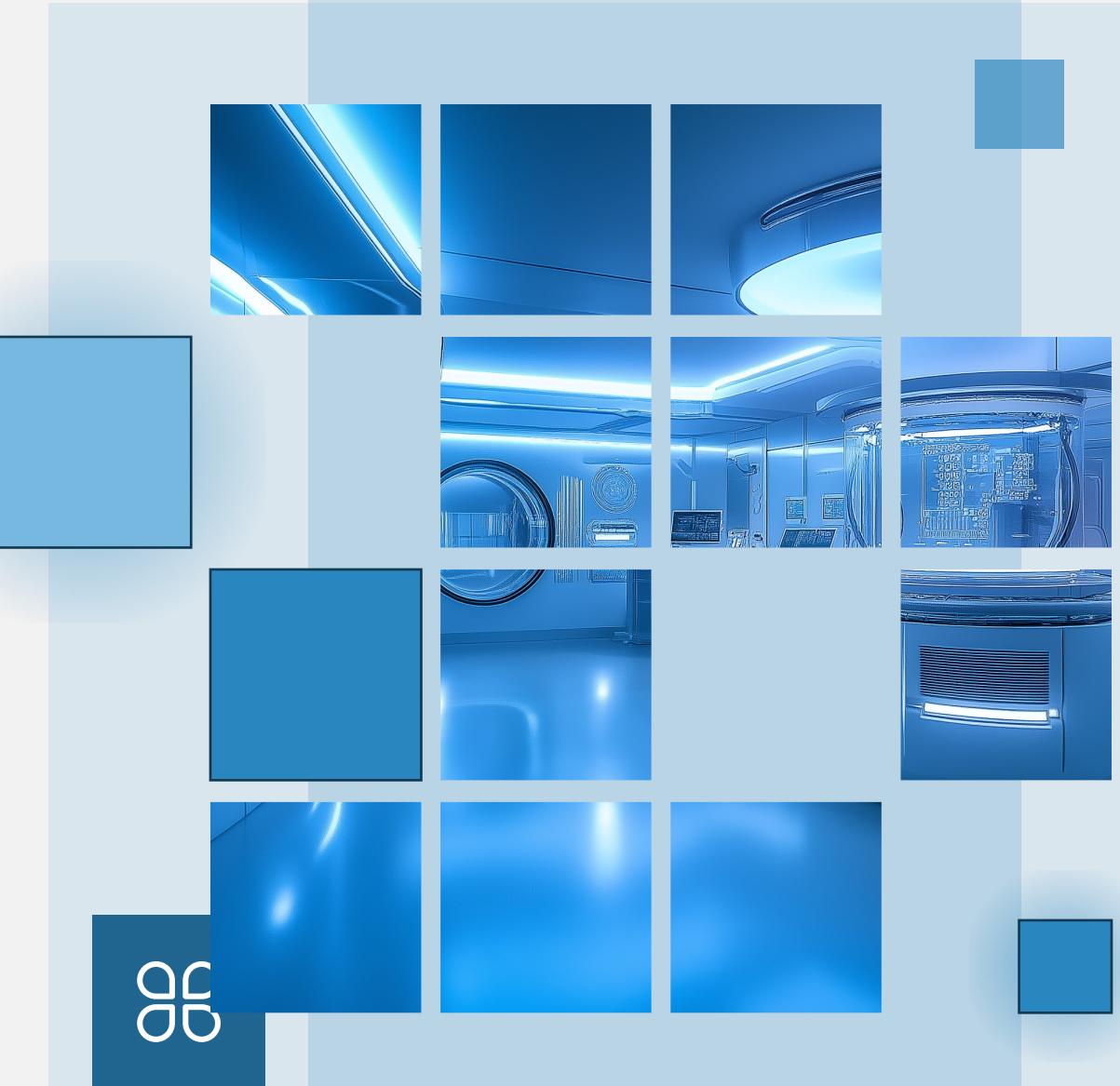
How do socioeconomic factors influence the healthcare burden caused by COVID-19 in the Americas?



This research seeks to understand how socioeconomic factors, such as GDP, health expenditure, and urbanization, influence the healthcare burden caused by COVID-19 in the Americas. This analysis aims to uncover key patterns and inform future policy improvements to mitigate disparities.

02

II. Datasets



Datasets

1. Our World in Data COVID-19 Dataset



- **Metadata URL:** <https://ourworldindata.org/coronavirus>
- **Data URL:** <https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-data.csv>
- **Data Type:** CSV
- **Characteristics after Transformation:**
Comprehensive data with time-series statistics covering 200+ countries, containing metrics such as confirmed cases, deaths, testing rates, ICU patients, and vaccination coverage.
- **License:** CC BY 4.0 License

2. World Bank Open Data



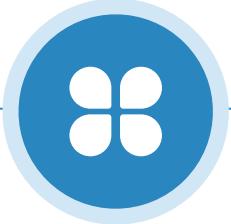
- **Metadata URL:** <https://data.worldbank.org/>
- **Data URL:** <https://databank.worldbank.org/source/world-development-indicators>
- **Data Type:** CSV
- **Characteristics after Transformation:** Includes 100+ socioeconomic indicators for over 200 countries, covering GDP, unemployment rate, Gini index, poverty rates, healthcare expenditure, and urbanization metrics.
- **License:** World Bank Dataset Terms of Use

03

III. ETL Pipeline



ETL Pipeline Design



Data Extraction

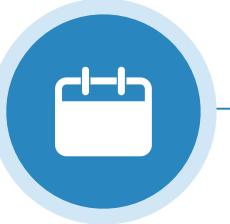
Python scripts and SQL queries were used to retrieve data from open datasets, including the **COVID-19 Open Data Repository** and **World Bank socioeconomic data**.



Data Transformation

Data cleaning and preprocessing were done using **Pandas**, with the following steps:

- Standardizing column names for consistency.
- Filtering data for the Americas region and mapping country names for alignment.
- Handling missing values by dropping or imputing data, ensuring data accuracy.
- Aggregating COVID-19 metrics by month and year for analysis-ready datasets.
- Pivoting socioeconomic indicators to reshape data for easier integration.
-



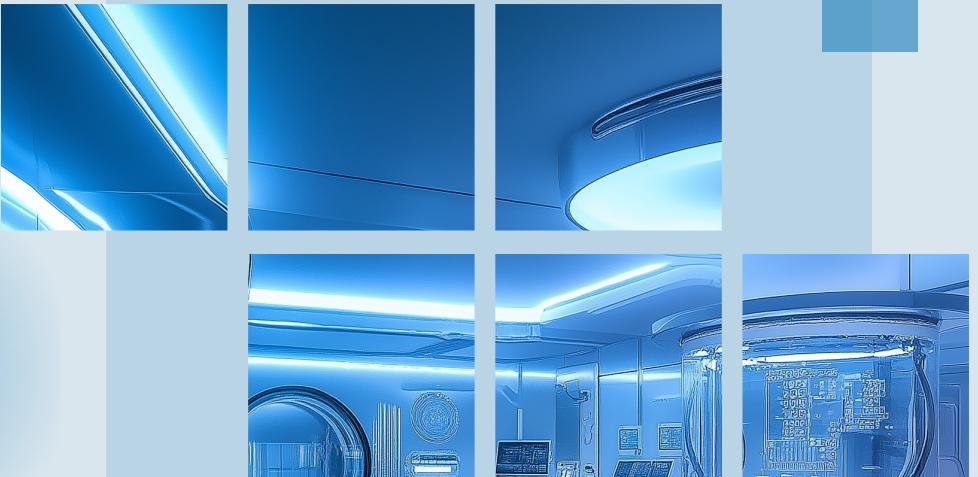
Data Loading

The transformed data was loaded into a **SQLite database** for storage and efficient querying.

Two main tables, `covid_monthly` and `merged_data`, were created to store aggregated COVID-19 data and merged socioeconomic data.

04

IV. Data Analysis



Correlation Analysis

Impact of Population Size on COVID-19 Cases and Deaths

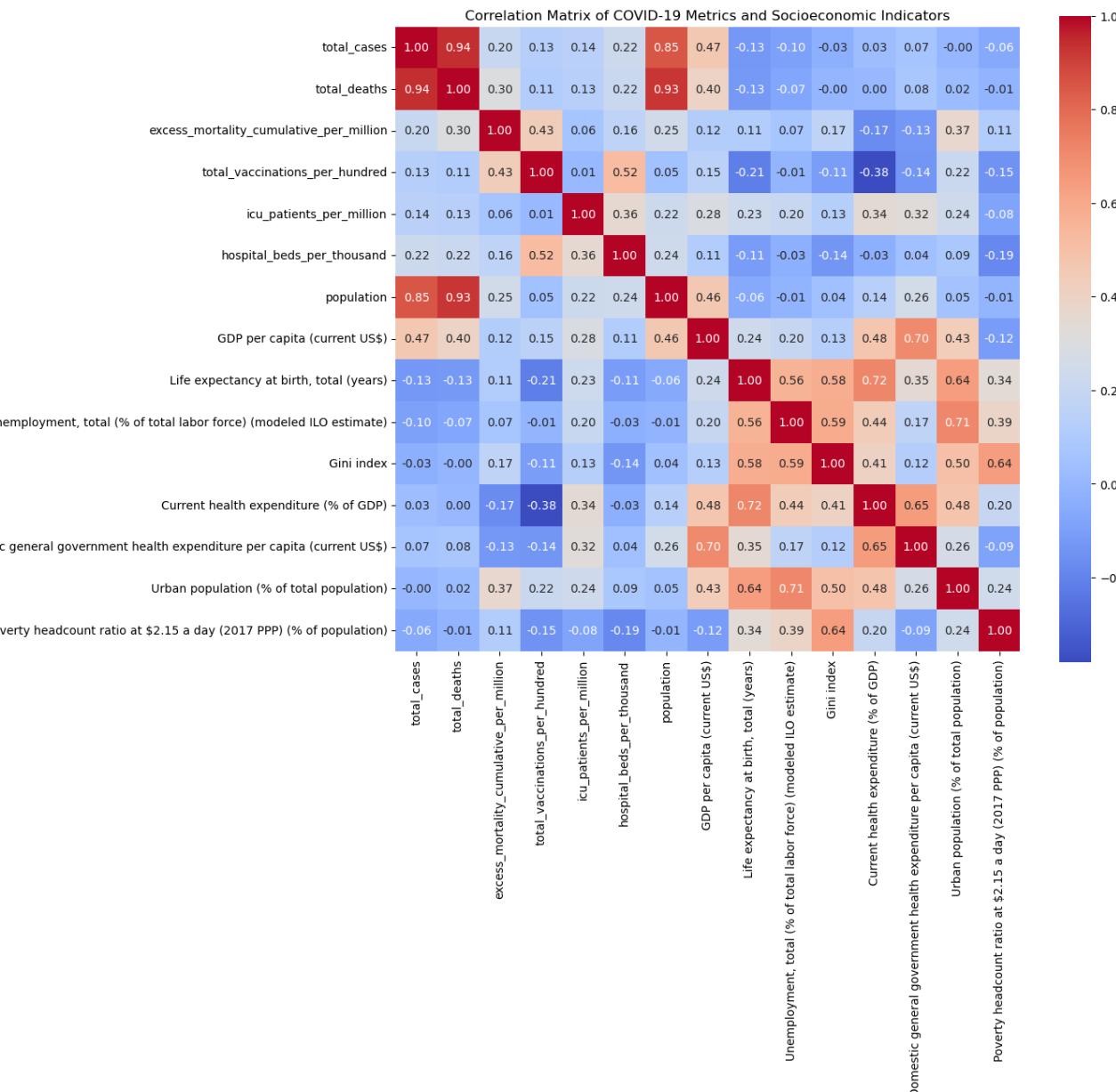
Population Size has strong positive correlations with **total cases** (0.85) and **total deaths** (0.93). Larger populations experience higher COVID-19 impacts.

GDP Per Capita and Healthcare Spending

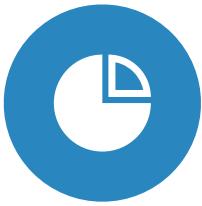
GDP Per Capita shows a moderate positive correlation with **health expenditure (% of GDP)** (0.70). Wealthier countries tend to spend more on healthcare.

Poverty and Economic Disparities

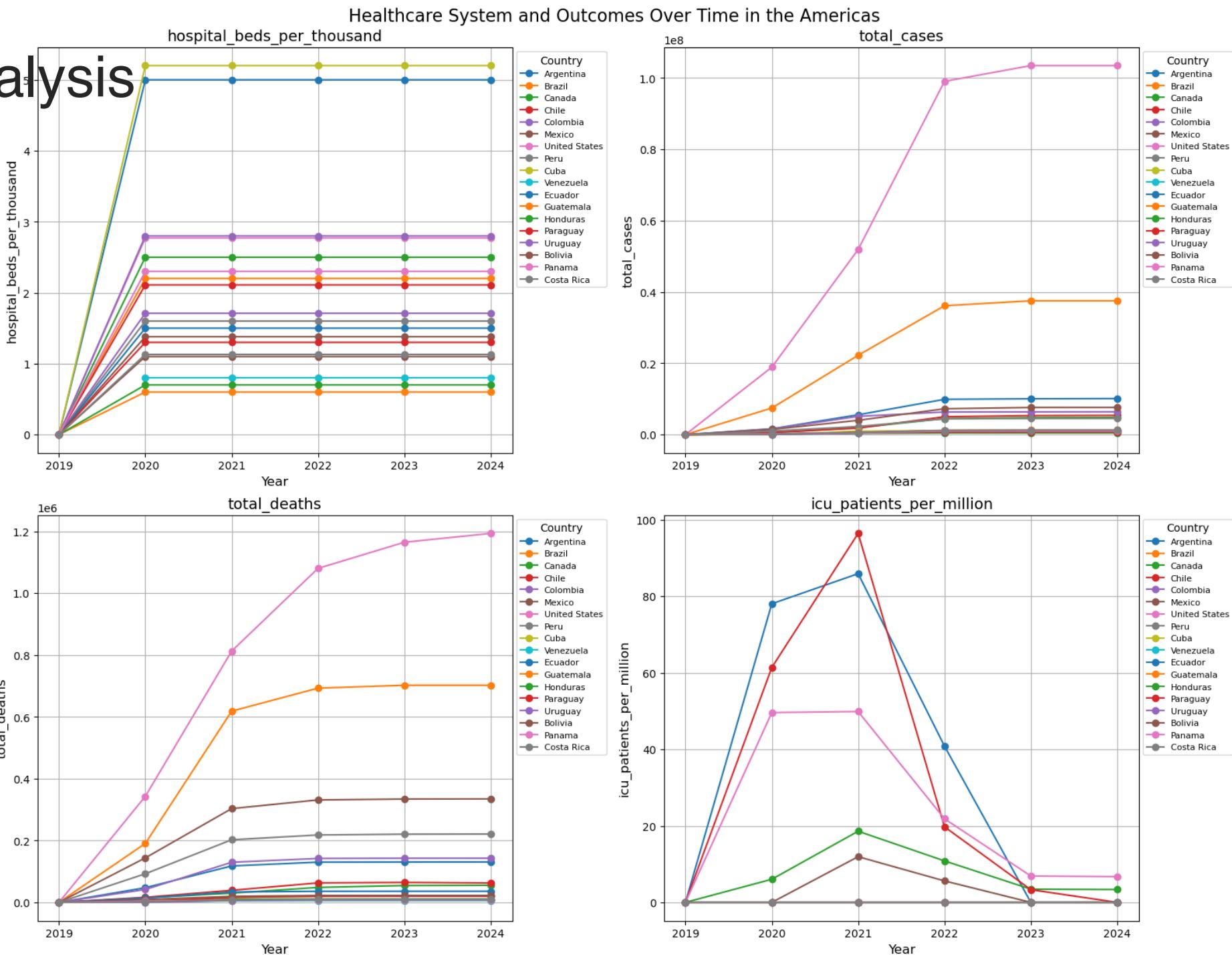
Poverty Headcount Ratio correlates negatively with **GDP per capita** (-0.71). Poorer nations have higher poverty rates.



Exploratory Analysis

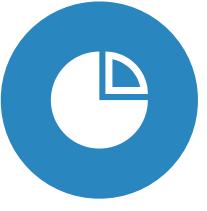


- Countries like Canada had higher values, showing better healthcare resources.
- Total deaths followed a similar pattern to total cases. The United States and Brazil had the highest death counts.

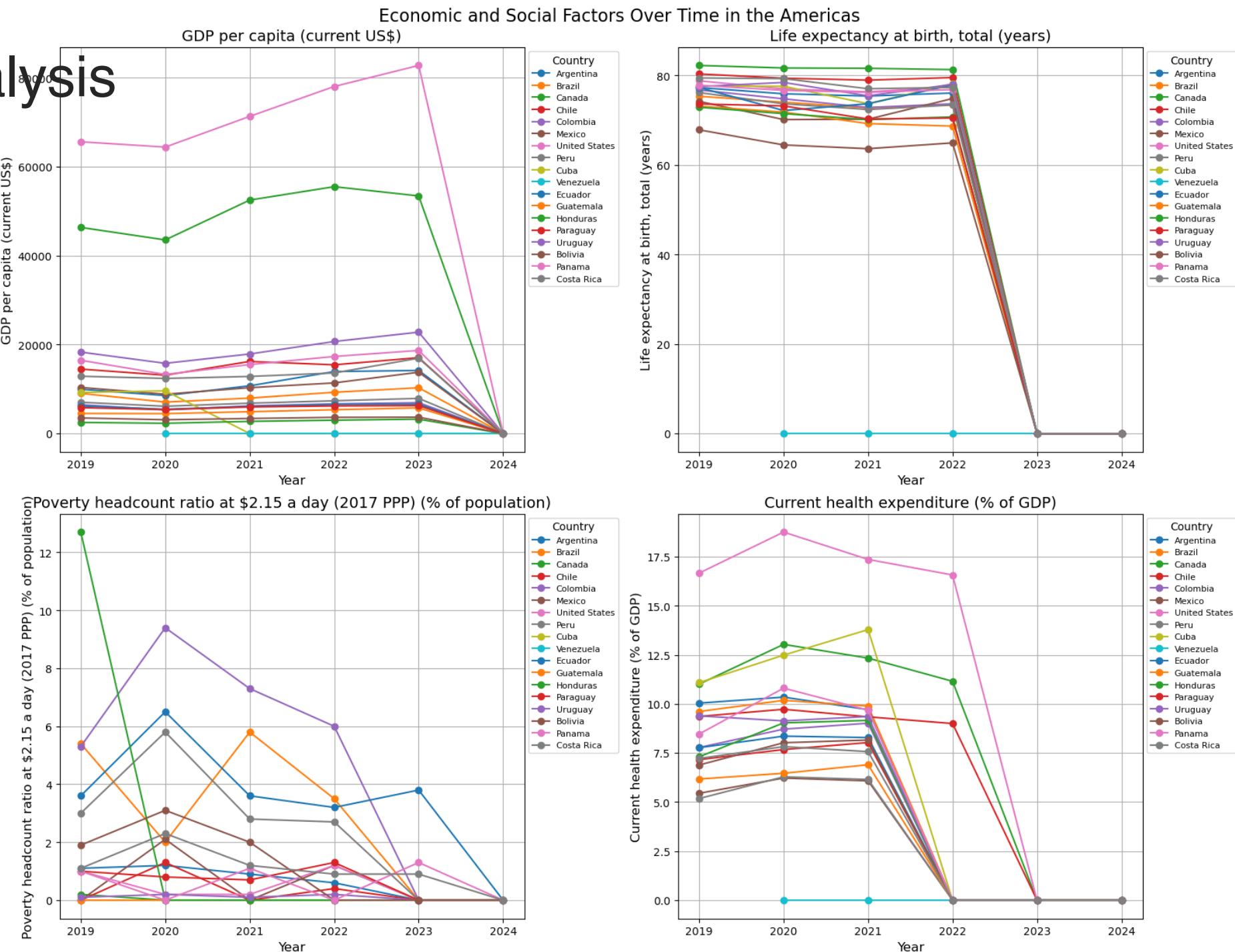


Some countries reported zero values, which means no data was available.

Exploratory Analysis

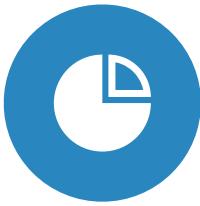


- The United States and Canada had much higher GDP levels than other countries.
- Countries like the United States spent a larger share of GDP on health. These countries handled the pandemic better.

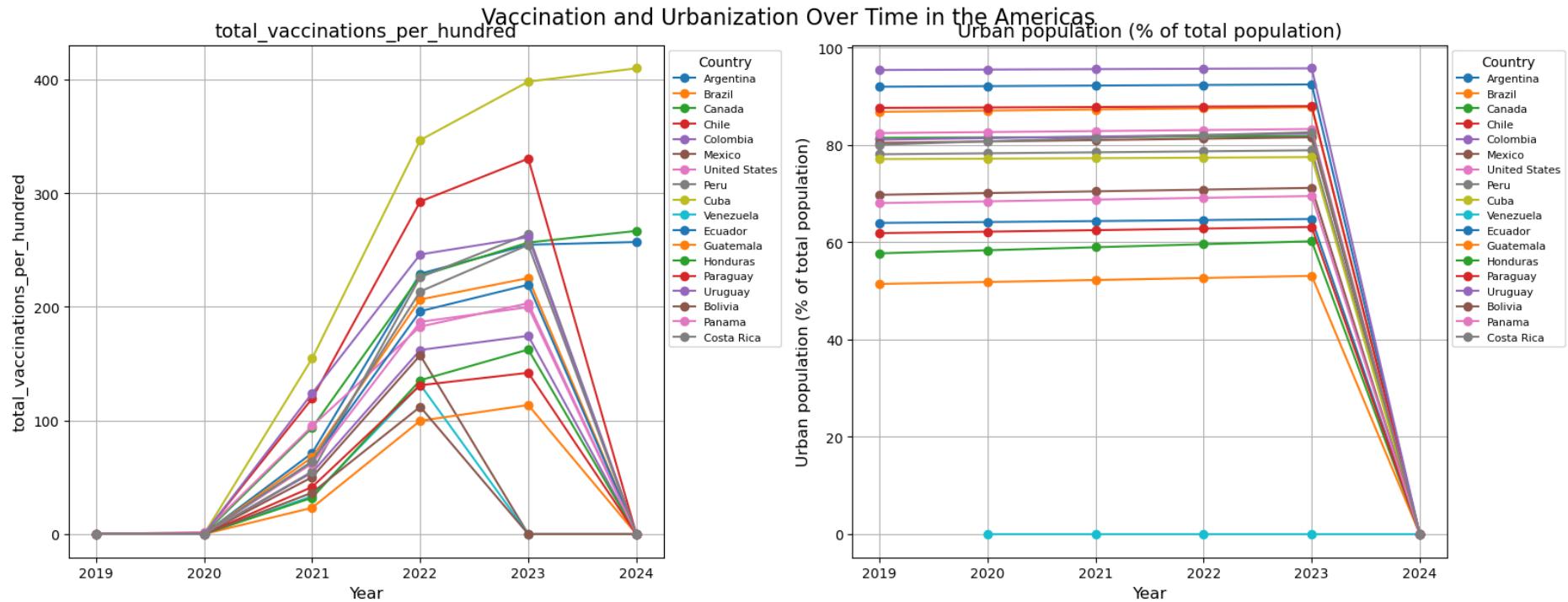


Some countries reported zero values, which means no data was available.

Exploratory Analysis



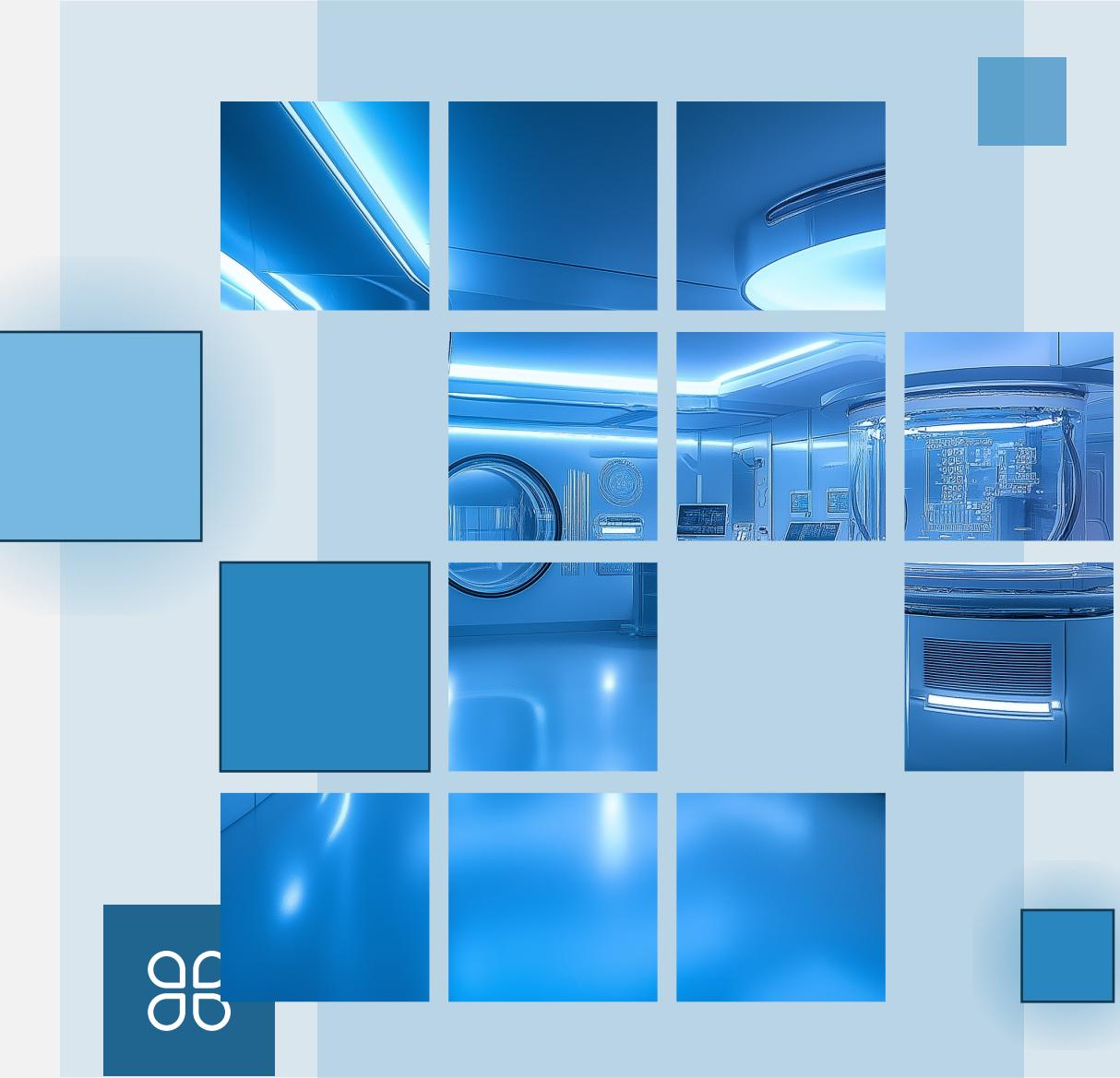
- Vaccination rates rose quickly in 2021. The United States and Canada had the highest rates.
- Urbanization levels remained steady. The United States and Canada had the highest urbanization rates.



Some countries reported zero values, which means no data was available.

05

V. Conclusion



Summary of Findings - Reducing the healthcare burden



+

01

Strengthen healthcare infrastructure

Increased health expenditure as a percentage of GDP could improve resilience to pandemics in vulnerable countries.

Investments in ICU capacity and hospital beds per capita can mitigate the impact of future health crises.

לע

+

02

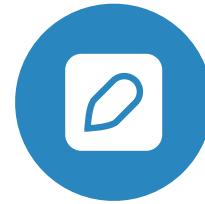
Focus on equitable vaccination programs

Expanding vaccine accessibility in low-income regions could reduce mortality rates and healthcare strain.

Addressing disparities in vaccination rates between countries can support regional pandemic recovery.

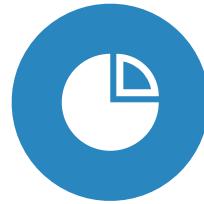
לע

Summary of Findings - Prioritizing investments in health and socioeconomic resilience



Correlation Patterns

Correlation analysis reveals strong links between socioeconomic factors like GDP per capita and key healthcare indicators such as ICU availability and vaccination rates.



Policy focus

Policymakers should prioritize improving healthcare systems in low-income countries to reduce vulnerabilities during health emergencies.

Limitations

Data completeness

Many datasets, including those on ICU capacity and vaccination rates, have missing values for certain countries or years, affecting analysis reliability.

Data timeliness

Available datasets only cover a limited time range, reducing the ability to fully capture long-term trends or outcomes.

Future Work

01



02



Further Research Directions

Investigate the long-term relationship between socioeconomic factors and the healthcare burden caused by COVID-19.

Improving Data Collection

Enhance data collection methods to reduce missing values and minimize biases in datasets.

Thank you for reading!



Yulin Wang



24.01.2025

