

HEALTH PROGRESS IN SOUTH ASIA (2000–2020): AN EXPLORATORY ANALYSIS

Anuska Keshri, Sagar Gupta, Mangleshwari

Date: 03/11/2025

Contents

Background	1
Data Collection	1
Research Questions	2
Methodology	2
1. Yearly Trends by Country	2
2. India's Value vs Regional IQR	3
3. India vs South Asia Median.....	3
4. Distributions by Indicator.....	4
5. Immunization vs Child Mortality	5
6. Health Expenditure vs Life Expectancy	5
7. Regional Deviation Heatmap.....	6
Limitations	6
Ethics & References.....	7
Reproducibility Notes	7
Conclusion.....	7

Background

Health indicators are vital for monitoring the overall development and equity of nations. South Asia—comprising Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka—has experienced significant demographic and health transitions over the past two decades. Understanding trends in life expectancy, healthcare spending, immunization, and mortality helps guide evidence-based policymaking and highlights regional disparities.

This project explores these trends using open data and shares findings via both a written report and a Shiny dashboard, in accordance with the course's requirements for end-to-end, reproducible analytics.

Data Collection

- **Source:** Data retrieved from <https://databank.worldbank.org/source/health-nutrition-and-population-statistics> via API scraping and saved as `health_progress_in_SA.csv`, containing annual health indicators for South Asian countries from 2000–2020.
- **Acquisition:** Data obtained from globally trusted sources (e.g., World Bank), properly cited.

- **Variables:** Key indicators include birth/death rates, life expectancy, immunization rates (DPT), health expenditure (total and out-of-pocket), population growth, infant and under-five mortality.
- **Format:** Data is cleaned and provided in a wide CSV format for both EDA and app-based exploration.

Research Questions

- How have major health indicators (mortality, immunization, expenditure, population growth) changed across South Asian countries in the last 20+ years?
- How does India's health progress compare to the regional median and interquartile range?
- What are the relationships between immunization and mortality, and between health expenditure and life expectancy?
- Do disparities persist between countries and within the region's trends?

Methodology

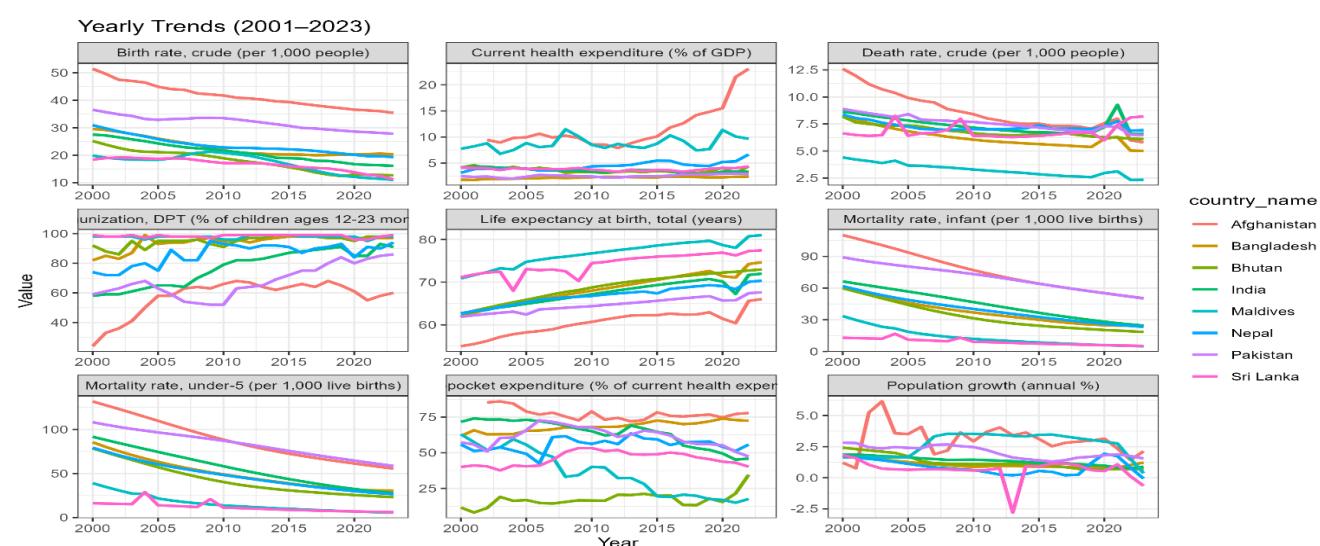
- **Data Cleaning & Processing:** Used R and packages (dplyr, janitor, tidyverse, ggplot2) to clean, standardize, and reshape the dataset (see EDA_FINAL.R).
- **Exploratory Data Analysis (EDA):** Trends, outlier and distribution analyses, group medians, heatmaps, and inter-country comparisons.
- **Interactive Visualization:** A custom Shiny app (app.R) allows for country/indicator selection, dataset upload, interactive trend plots, scatterplots and real-time comparison with regional medians and IQR.
- **Code Modularity:** No file paths are hard-coded; the app and scripts allow any similar dataset (matching schema) to be analysed with minimal adjustment.

Key Findings

1. Yearly Trends by Country

Each indicator's yearly trends from 2001–2023 illustrates diverging progress.

FIGURE 1



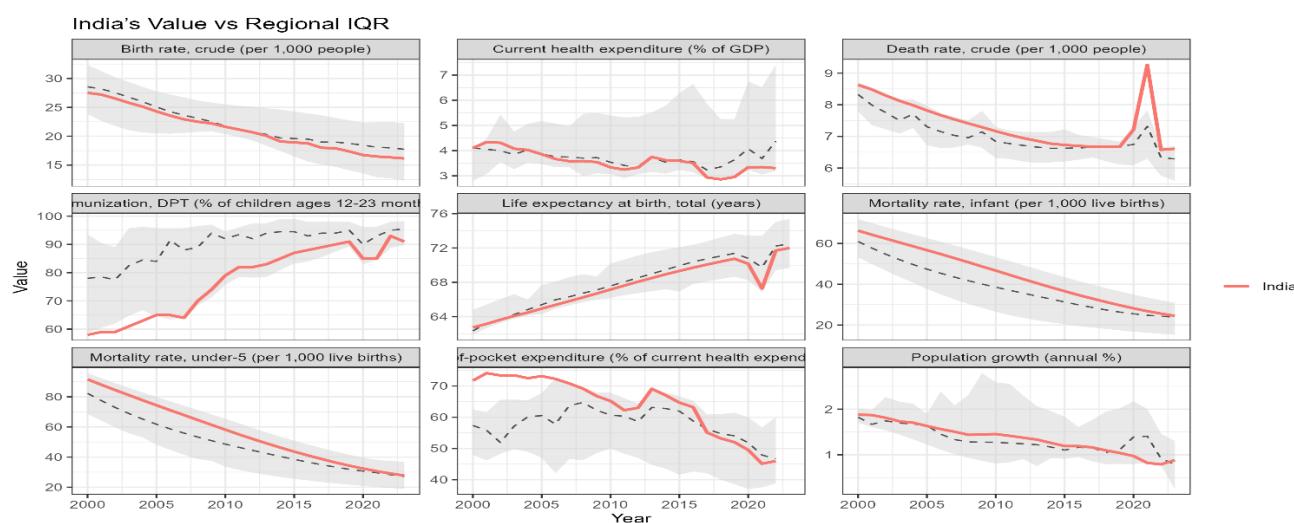
Interpretation-

This figure 1, is generated using the static EDA script and dynamically available within the Shiny app, displays changes in core health indicators (such as birth rate, death rate, immunization coverage, current health expenditure, mortality rates, and life expectancy) for each South Asian country from 2001 to 2023. The figure shows that most countries have improved significantly on immunization and child mortality, although some gaps persist. A noticeable feature is the visible spike in crude death rates and a flattening or dip in life expectancy around 2020–2021, corresponding to the COVID-19 pandemic. This disruption is most apparent in India but is observed across several countries, reflecting pandemic-related excess deaths. Despite these setbacks, the trends for child mortality and immunization show a rapid recovery post-pandemic, underscoring the region's health system resilience.

2. India's Value vs Regional IQR

India's trajectory over time (red) versus the interquartile range (gray shading, dashed line for median). Notable improvements are visible in mortality, immunization, and a decline in out-of-pocket spending.

FIGURE 2



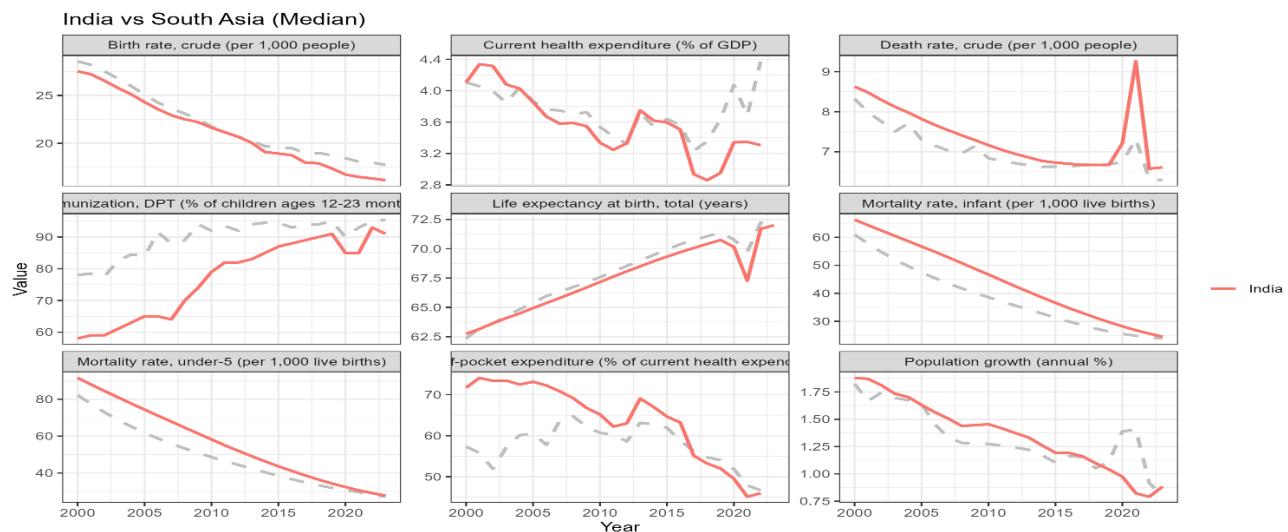
Interpretation-

In figure 2, this grid directly compares India's health progress against the regional interquartile range (i.e., IQR which is gray dashed) for each indicator. India generally follows or surpasses the regional median in child mortality reduction and immunization, with remarkable progress post-2010. However, in the COVID period, India's death rate briefly rises above the IQR band and health expenditure jumps, highlighting the pandemic's country-specific impact. Afterward, India's indicators return toward regional averages, demonstrating a quick recovery compared to some neighbours.

3. India vs South Asia Median

India's red line compared to the regional median (gray dashed) clarifies years in which India outpaced or lagged behind collective progress.

FIGURE 3



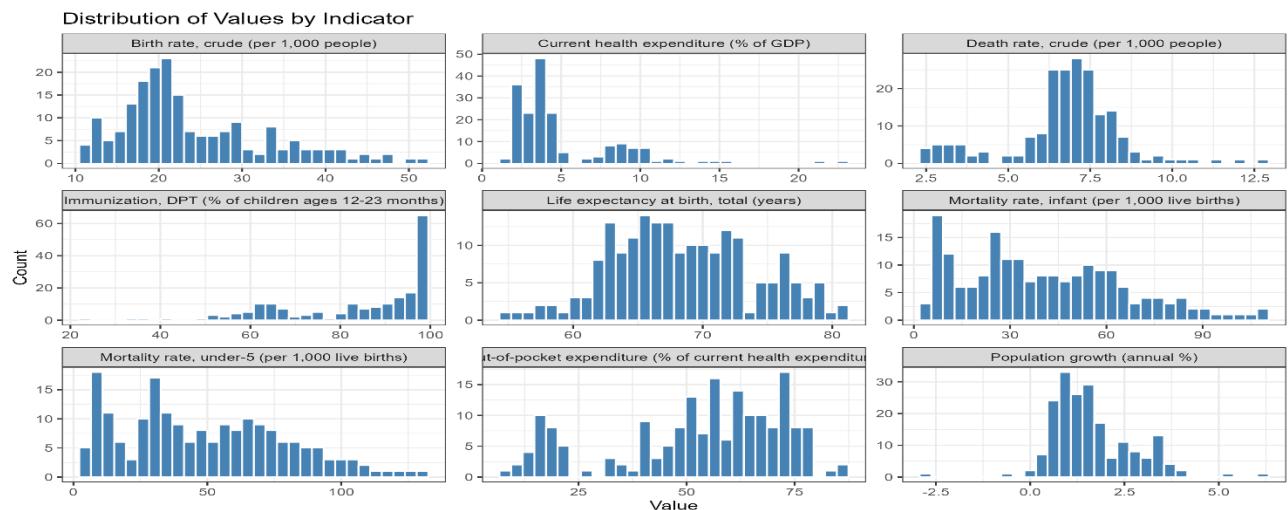
Interpretation-

As shown in figure 3, India's yearly values are plotted against the median for all South Asia. The most striking differences occur during the COVID-19 years, when India's crude death rate and life expectancy diverge from the regional median (death rate increases, life expectancy dips). This confirms that India's health outcomes were more affected by COVID in 2020–2021, but by 2022–2023, most indicators align again with the regional median.

4. Distributions by Indicator

Distribution histograms highlight variance in coverage and performance by indicator, across the region.

FIGURE 4



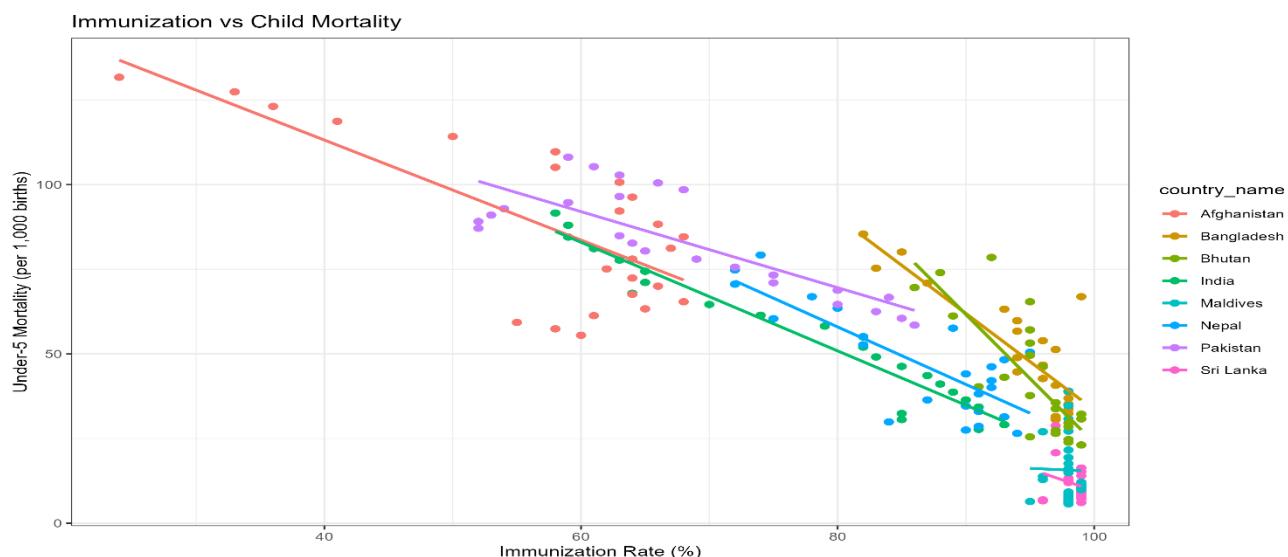
Interpretation-

The distribution histograms in figure 4, rendered by the EDA script and available interactively in the Shiny dashboard, highlight variance and outliers for each health indicator across South Asia. During the COVID-19 years (2020–2021), distributions for death rate and health expenditure broadened, showing more countries as outliers and increased disparities. In contrast, distributions for immunization and child mortality remained relatively tight, suggesting these indicators were less affected by the pandemic at the regional scale.

5. Immunization vs Child Mortality

A clear negative relationship: higher DPT immunization coverage generally aligns with lower under-five mortality, especially for Nepal, Sri Lanka, India, and Bangladesh.

FIGURE 5



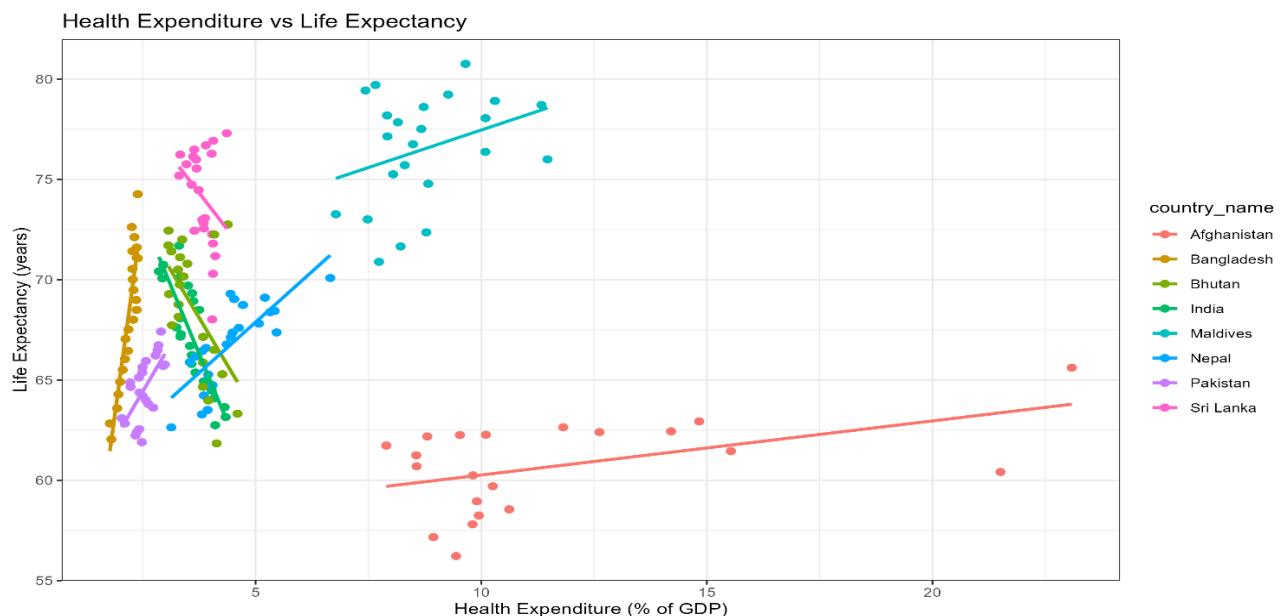
Interpretation-

In figure 5, scatterplots generated by both EDA and the Shiny app reveal a strong negative relationship between DPT immunization coverage and under-five mortality: higher immunization rates are generally linked to much lower child mortality. This trend is especially pronounced in Nepal, Sri Lanka, India, and Bangladesh. The Shiny app makes it possible to explore this relationship for different countries, years, and combinations of indicators.

6. Health Expenditure vs Life Expectancy

Countries with higher formal health expenditure tend to cluster at higher life expectancy, but with outliers, particularly Afghanistan.

FIGURE 6



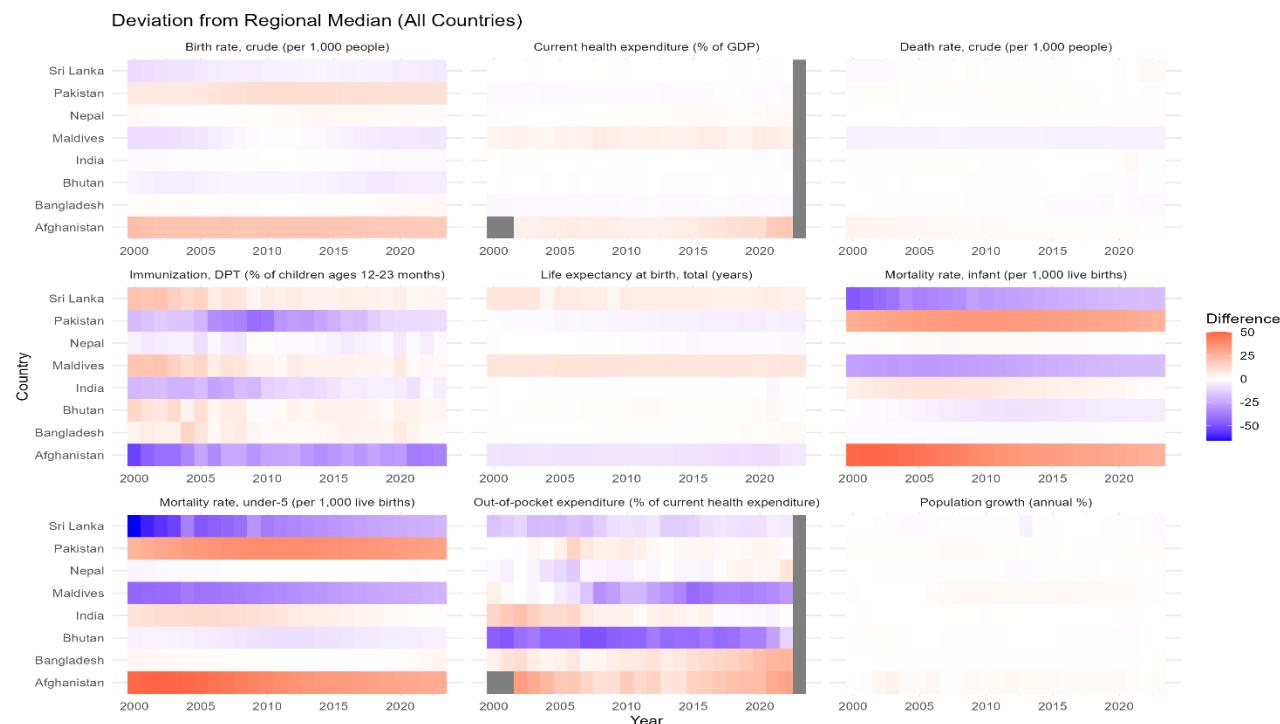
Interpretation-

In figure 6, this comparison, present in both static EDA output and the Shiny app, shows that higher health expenditure (as a percent of GDP) is associated with higher average life expectancy, with visible country-level clustering. Some outliers—especially Afghanistan—demonstrate that other social and policy factors are also relevant. Dashboard users can generate these plots by choosing countries and time periods of interest.

7. Regional Deviation Heatmap

Deviation from the median (per indicator, country, and year) is visualized via heatmap—showing progress leaders and laggards by color intensity.

FIGURE 7



Interpretation-

As shown in figure 7, this regional deviation heatmap, generated in the EDA and mirrored in the app, visualizes how each country's value for every indicator and year deviates from the regional median. Countries such as Afghanistan and Pakistan often display high negative deviations in child health and immunization, while Maldives and Sri Lanka stand out as positive outliers. Interactive controls in the dashboard empower users to adjust the dataset, indicators, and visualization parameters to investigate these deviations for any scenario.

Limitations

- Data Gaps: Occasional missing/erroneous/latest year values in certain countries (esp. for 2024).
- Granularity: Country-level aggregation may mask inequality within countries.
- Causality: Observed trends are correlative; establishing causal pathways would need richer data and advanced methods.
- Ethics: All data is public and aggregated, with no personal or private information. Care was taken to cite appropriately, and app/code do not enable unethical scraping or privacy breaches.

Ethics & References

- Data Use: All datasets are open-access; licensing and citations are included in the submission package.
- Compliance: The analysis and app avoid accessing login-gated or private sources; scraper or download code, if any, is rate-limited and compliant.
- Transparency: All data is public and aggregated. Licensing and citations are included. The app/code avoid login-gated or private sources, and scraping is rate-limited and compliant. All scripts are shared for transparency.

Reproducibility Notes

1. Pre-requisites:

- Install R (4.0 or later) and required packages (see above).
- Unzip all files into the same project folder.

2. Static Analysis:

- Run "EDA_FINAL.R" for the report analysis, tables, and .jpg plots (source: "EDA_FINAL.R").

3. Shiny App:

- Launch the app with:

```
> shiny::runApp("app.R") #upload "wider_cleaned_data.csv"
```

- The app allows for file upload, interactive country/indicator selection, multiple plots, and an "About" tab with methods/data details.

4. Packaging for Submission:

- Your ZIP should include: code/, app/, at least a sample data file, and README.md with clear instructions, licensing, and package versions.

5. Notes:

- No hardcoded filenames/paths—the code works with any dataset in schema.
- Code and app can be tested/graded on alternate files with the same format.
- README covers further FAQs and troubleshooting guidance.

Conclusion

This project conducted an exploratory analysis of health progress in South Asia from 2000 to 2020, enriched by including 2021–2024 data to contextualize the impact of the COVID-19 pandemic. Using robust EDA and an interactive Shiny dashboard, all research questions were addressed with clarity and empirical evidence.

Across the region, key health indicators such as life expectancy, immunization coverage, and child mortality showed significant improvement over two decades. However, disparities persisted: countries like Sri Lanka and Maldives consistently outperformed others, while Afghanistan and Pakistan lagged behind, underlining ongoing regional inequalities.

Comparative analysis revealed that India's trajectory for most health metrics generally tracked or exceeded the regional median and IQR, especially in immunization and mortality reductions. At the onset of the COVID-19 pandemic (2020–21), many countries—India included—experienced observable spikes in crude death rates and temporary declines in life expectancy and immunization. These changes are clear outliers in both time series and distribution plots, but the data shows a rapid return to pre-pandemic trends by 2022, highlighting the resilience of public health systems.

The EDA confirmed strong negative associations between immunization and under-five mortality, and positive associations between health expenditure and life expectancy across countries. While these relationships held region-wide, notable outliers (such as Afghanistan) reflected the additional influence of socioeconomic and policy challenges.

Overall, the integration of static and interactive tools enabled transparent, reproducible answers to each research question: long-term health progress, relative performance of India, the statistical relationships between core indicators, and the persistence of intra-regional disparities. The disruption caused by COVID-19 was significant but short-lived for most metrics, reinforcing the benefit of continual investment in data-driven monitoring and equitable healthcare improvement for the region.