

Global Development Insights

World Bank Indicators Analysis — Project Summary Report

200+ Countries | 26 Indicators | 2016–2024

1. Project Overview

This project involved building an end-to-end data analytics solution using real-world data from the World Bank public API. The goal was to analyse global development trends across economic, health, social, environmental, and technology dimensions, and present findings through an interactive Power BI dashboard.

Component	Details
Data Source	World Bank Open API (api.worldbank.org)
Countries Covered	200+ countries and territories
Indicators Analysed	26 key development indicators
Time Period	2016 to 2024
Total Data Points	~57,000+ rows across 7 categories
Tech Stack	Python, Pandas, Power BI, DAX, Seaborn, Matplotlib

2. Technical Stack & Skills Demonstrated

2.1 Python — Data Collection & Processing

- Fetched live data from World Bank REST API using requests library
- Handled pagination across 500+ API pages to collect 26,223 indicator codes
- Used `pd.json_normalize()` to flatten nested JSON into structured DataFrames
- Performed data cleaning: removed nulls, filtered years (2016+), renamed columns
- Used `pd.merge()` to join indicator data with country metadata
- Exported 7 clean category-level CSV files for Power BI consumption

2.2 Power BI — Visualisation & Analysis

- Built a star schema data model with countries as a dimension table
- Created DAX measures, including Poverty Reduction %.
- Designed an interactive dashboard with region slicers
- Used **Python** visuals inside Power BI for advanced analytics

- Applied a consistent dark theme with a professional colour palette

2.3 Python Visuals (Advanced)

- **Correlation heatmap** using Seaborn to show relationships between 13 health indicators
- **Regression scatter** plot using Seaborn regplot to analyse spending vs life expectancy
- Custom dark styling matching dashboard theme (facecolor, tick colours, spine styling)

3. Data Architecture

3.1 Data Collection Pipeline

The data pipeline followed a structured 5-stage process:

Stage	Action	Output
Stage 1	Fetch all countries from the API	countries DataFrame (296 rows)
Stage 2	Fetch all indicator codes from the API	final_df.csv (26,223 indicators)
Stage 3	Fetch actual values for 26 chosen indicators	7 category DataFrames
Stage 4	Merge indicator data with country metadata	7 enriched DataFrames
Stage 5	Export to CSV for Power BI	7 clean CSV files

3.2 Power BI Data Model (Star Schema)

The data model follows a Star Schema pattern — the industry standard for analytical dashboards:

- Centre (Dimension): countries table — each country appears exactly once
- Surrounding (Fact tables): health, economic, trade, labour, poverty, environment, technology
- All fact tables connect to countries via country_id (1-to-Many relationship)
- Filter direction set to Single to avoid M: M ambiguity and incorrect aggregations

Why Star Schema? It prevents many-to-many relationship errors, ensures correct aggregations, and enables clean filtering across all visuals simultaneously.

4. Key Findings & Insights

4.1 Economic Baseline

- Average GDP per capita: \$17,970 — skewed upward by high-income nations
- Average GDP growth: 2.68% — below historical average, reflecting post-COVID recovery
- Average health spending: 6.69% of GDP — just above WHO recommended minimum of 5%
- Forest area: 31.69% of total land — slowly declining year on year

4.2 Regional Health Spending Disparity

- North America spends significantly more on health than Sub-Saharan Africa
- Sub-Saharan Africa has the highest disease burden but the lowest health investment
- This spending gap directly correlates with life expectancy differences across regions

4.3 Digital Progress & Technology Trends

- Internet usage and mobile subscriptions show the steepest growth of all indicators
- Digital access is expanding faster than any other development metric globally
- Forest area declining steadily — environmental concern across all regions

4.4 Internet Access & Development Outcomes

Relationship	Finding	Strength
Internet vs. Immunisation	Positive — more internet = higher vaccination rates	Strong
Internet vs Unemployment	Weak — the internet alone does not reduce unemployment	Moderate

Key insight: Digital connectivity improves health awareness faster than it improves economic outcomes. The Internet is necessary but not sufficient for employment improvement.

4.5 Poverty Reduction Analysis

Using the custom DAX measure Poverty_Reduction%, countries were ranked by their progress in reducing poverty from their earliest to latest available data year.

Category	Countries	Key Insight
Top Performers	Bhutan (+51%), Uruguay (+49%), Nigeria (+40%)	Strong social programs and stable governance
Bottom Performers	Kiribati (-75%), Ukraine (-55%)	Raised poverty lines, climate impact, and conflict
Regional Pattern	Latin America dominates the top 10	Conditional cash transfer programs work

4.6 Health Indicator Correlations (Heatmap Findings)

- Water access is the strongest predictor of life expectancy ($r \sim 0.8$) — stronger than health spending
- Skilled birth attendance reduces maternal mortality dramatically ($r \sim -0.8$)
- TB incidence strongly negatively correlated with life expectancy ($r \sim -0.7$)
- Immunisation rates linked to overall health system strength, not vaccines alone

Policy implication: Investing in water infrastructure and skilled healthcare workers delivers better health returns per dollar than increasing general health spending alone.

4.7 Health Expenditure vs Life Expectancy (Regression Analysis)

- Positive relationship confirmed — more spending generally leads to longer lives
- Relationship is not perfectly linear — other factors matter significantly
- USA anomaly: spends ~17% of GDP on health, but life expectancy trails many European nations
- Sweet spot identified around 5-8% GDP — strongest returns on life expectancy
- Sub-Saharan cluster: low spending AND low life expectancy — confirms urgent need for investment