

# **A Global Analysis of Income Inequality (2010-2021)**

**UN SDG Goal: 10:** Reduced Inequality

**Prepared By: Terungwa Emmanuel, Flavian Kazira, Babatunde Oyedele (Cloud Data Analytics Team)**

## **Problem Statement and Objectives**

### **1.1 Introduction**

Income inequality remains one of the most pressing socio-economic challenges of our time, acting as a significant barrier to sustainable development, social cohesion, and economic prosperity. The United Nations Sustainable Development Goal (SDG) Goal 10 explicitly targets reducing inequality within and among countries. This project undertakes a comprehensive, data-driven analysis of global income inequality trends from 2010 to 2021. The aim is to move beyond anecdotal evidence and provide a quantifiable, granular understanding of the disparities that exist across different geographic and developmental contexts, thereby contributing to the global mission of achieving reduced inequality.

### **1.2 Problem Identification**

The core challenge addressed in this project is the complexity of transforming vast, multidimensional global data into actionable intelligence. While income inequality is widely acknowledged, a precise understanding of its spatial and temporal patterns, its correlation with development metrics, and the disparities that exist within world regions is often lacking. This gap hinders the ability of global organizations, governments, and policymakers to design targeted, evidence-based interventions. Without a clear data-driven foundation, efforts to reduce economic disparities can be misdirected, inefficient, or based on incomplete information.

The specific problem is the need to conduct a comprehensive analysis of income inequality across 194 countries over a 12-year period (2010-2021), to identify the key socio-economic and geographic factors that correlate with inequality levels, and to model the relationship between development and inequality across different world regions.

### **1.3 Project Aim**

To conduct a comprehensive, data-driven analysis of global income inequality trends from 2010 to 2021, identifying key socio-economic and geographic factors that correlate with inequality levels, and to model the relationship between development and inequality across different world regions.

## 1.4 Project Objectives

To achieve the stated aim, the project was guided by the following specific, measurable objectives:

1. Quantify and Compare Continental Trends: To analyze key descriptive statistics (mean, median, max, min and range) for income inequality for each continent from 2010 to 2021, identifying regions with the most pronounced and persistent disparities.
2. Conduct Time-Series Analysis: To uncover significant trends, breakpoints, and patterns of change (upward, downward, or stable) across different continents over the 12-year period.
3. Investigate Intra-Continental Disparities: To analyze the variation in inequality levels between countries within the same continent and explore correlations with key socioeconomic factors such as Human Development Index (HDI) classification.
4. Develop an Interactive Dashboard: To create an interactive data visualization dashboard that effectively communicates the spatial and temporal trends in global income inequality, enabling stakeholders to explore the data intuitively.
5. Synthesize Actionable Recommendations: To translate data-driven insights into actionable recommendations for policymakers, identifying potential intervention points and successful case studies.

## Research Findings and Analysis: Methodology & Tools

### 2.1 Methodology

**The project followed a structured data analytics lifecycle:**

1. Data Acquisition & Understanding: The project utilized the "Inequality in Income" dataset, containing records for 194 countries across 12 years (2010-2021), alongside attributes like Continent, HDI Rank, and Human Development Groups. Data gotten from Kaggle: <https://www.kaggle.com/datasets/psterk/income-inequality/data>
2. Data Preparation & Cleaning: The raw dataset was profiled and cleaned using Advanced Excel to handle missing values, validate data types, and identify potential outliers. This ensured the integrity of the data before deeper analysis.
3. Data Storage & Transformation: The cleaned dataset was imported and stored in a Microsoft SQL. SQL was employed to structure the data and perform complex transformations, most critically unpivoting the dataset from a wide format (one column per year) to a long format (a

`Year` column and an `Inequality` value column). This was essential for descriptive analysis and effective time-series analysis.

4. Analysis & Querying: Complex SQL queries were written to address each objective:

Objective I: Queries calculated descriptive statistics (AVG, MIN, MAX, Range) grouped by Continent and Year.

Objective II: Queries utilized window functions like `LAG()` to calculate yearly changes and classify trends as 'Upward', 'Downward', or 'Stable'.

Objective III: Queries analyzed inequality by grouping data by `Human\_Development\_Groups` and `UNDP\_Developing\_Regions` to understand intra-continental and developmental disparities.

5. Data Visualization & Reporting: The stored inequality in income dataset was imported to Power BI for visualizations. Interactive visualizations were crafted to tell the data story clearly. Finally, insights were synthesized into this comprehensive report.

## 2.2 Tools Used:

Advanced Excel: Used for the initial data cleaning, profiling (identifying missing values, outliers), and quick checks.

Microsoft SQL Server: Used as the relational database management system to store the cleaned data and perform advanced, efficient querying and data transformation (e.g., unpivoting, calculating aggregates with window functions).

Microsoft Power BI: The primary tool for building interactive visualizations and dashboards. It was used to create trend lines, bar charts, bubble charts and to implement interactive filters for drill-down capabilities.

## 2.3 Key Findings and Analysis

### Objective I: Quantifying Continental Trends

The descriptive statistics revealed stark contrasts between continents:

Africa exhibited the highest mean income inequality (28.04) and the largest range (59.28), indicating not only high average disparity but also extreme variation between its most and least equal countries.

America showed a similarly high mean inequality (29.14) but a smaller range (33.61), suggesting high disparity is a more consistent feature across the continent.

Europe consistently demonstrated the lowest mean inequality (15.83) and a relatively contained range, affirming its status as the most equitable region.

Asia and Oceania presented moderate levels of mean inequality (20.08 and 21.45, respectively), with Oceania showing the smallest range, suggesting more homogeneity among its countries.

### **Objective II: Time-Series Analysis**

The time-series analysis showed that trends are largely persistent:

Africa and America maintained consistently high levels of inequality throughout the entire period, with minor fluctuations but no significant downward trend.

Europe and Oceania maintained stable, low-to-moderate levels of inequality.

Asia showed a moderate and relatively stable trend. The analysis concluded that income inequality trends are deeply entrenched and require sustained, long-term policy interventions to alter.

### **Objective III: Intra-Continental Disparities**

Analysis within continents provided crucial nuance:

The Top 10 most unequal countries were dominated by nations from Africa (e.g., Central African Republic, Namibia, South Africa) and America (e.g., Haiti, Brazil).

The Bottom 10 most equal countries were overwhelmingly European (e.g., Slovakia, Slovenia, Finland), though the presence of Ethiopia (Africa) and Azerbaijan (Asia) shows that low inequality is possible even in regions with high overall averages.

The Bubble Chart analyzing HDI Rank vs. Inequality confirmed a strong correlation:

Europe: Low Inequality, High Development (Low HDI Rank).

Africa: High Inequality, Low Development (High HDI Rank).

America: High Inequality but relatively Higher Development, suggesting that development alone does not automatically guarantee equality and other factors are at play.

## **Solution Development Process**

### **3.1 From Objectives to Technical Solutions**

The development process was meticulously aligned with the project objectives.

For Objective I & II (Descriptive & Time-Series Stats): The solution involved writing a sophisticated SQL query that first unpivoted the data. A second query using Common Table Expressions (CTEs) and the `'PERCENTILE_CONT()'` function calculated the median—a more robust measure than

mean for skewed distributions. For time-series, the `LAG()` window function was pivotal in comparing each year's average to the previous year, enabling the classification of trends into 'Upward', 'Downward', or 'Stable'.

For Objective III (Intra-Continental Analysis): The solution required segmenting the data by additional attributes. SQL queries were designed to group data not just by continent, but by `Human_Development_Groups` and `UNDP_Developing_Regions`. This allowed for analysis of variance (e.g., calculating standard deviation) within these groups, directly addressing the objective of understanding internal disparities.

For Objective IV (Dashboard Development): The stored data in SQL was imported into Power BI. The results from the key SQL queries were imported as data models. The dashboard was designed with a clear hierarchy:

1. Global Overview: Key metrics (Global Average Inequality, Average HDI Rank) placed at the top.
2. Continental Analysis: Descriptive statistics table and time-series line chart placed centrally.
3. Country-Level Drill-Down: Top 10 and Bottom 10 bar charts for granular insight.
4. Correlation Analysis: Bubble chart to explore the HDI-Inequality relationship.
5. Interactivity: Slicers for Continent, Country, and Year were implemented, allowing users to filter the entire dashboard dynamically. This transformed static visualizations into an exploratory tool.

### **3.2 Iterative Refinement**

The development process was iterative. Initial query results were visualized in Power BI, which often revealed the need for additional data transformation or a different analytical approach. For example, the initial continent-level analysis (Objective I) prompted the more granular country-level analysis (part of Objective III). The dashboard layout was refined multiple times based on clarity and the logical flow of information, ensuring it was intuitive for end-users like policymakers.

## **Implementation Plan and Outcomes**

### **4.1 Implementation Plan**

The project was implemented by a three-member team with responsibilities divided based on expertise:

1. Terungwa Emmanuel: Led the database creation in SQL Server, performed the complex querying for data transformation and analysis, and architected the primary structure of the Power BI dashboard.
2. Flavian Kazira: Focused on data cleaning and preparation in Excel, contributed to writing the comprehensive project report, and assisted in generalizing insights from the SQL analysis.
3. Babatunde Oyedele: Took the lead on developing advanced visualizations in Power BI, authored sections of the report, and focused on extracting and articulating the actionable insights and recommendations from the analyzed data.

### **4.2 Outcomes and Deliverables**

The project successfully delivered on all five objectives, resulting in:

1. A completely cleaned and transformed dataset stored in a SQL database, ready for analysis.
2. A suite of advanced SQL queries that can be reused and adapted for future longitudinal studies on inequality.
3. A comprehensive Power BI interactive dashboard that effectively communicates global income inequality trends spatially and temporally. This dashboard serves as a powerful tool for evidence-based decision-making.
4. A detailed report (this document) synthesizing the methodology, analysis, and findings.
5. A set of data-driven, actionable recommendations for policymakers.

### 4.3 Actionable Recommendations & Conclusion

Based on the findings, the project culminates in the following recommendations for policymakers:

1. **Target High-Inequality Regions:** Concentrate policy efforts and resources on Africa and America, focusing on improving access to education, healthcare, and economic opportunities.
2. **Learn from Success Stories:** Conduct deep-dive studies into the policies of European nations and unexpected performers like Ethiopia and Azerbaijan to identify transferable strategies (e.g., progressive taxation, social safety nets).
3. **Invest in Human Development:** Address the development-inequality nexus by implementing policies that enhance education, healthcare, and living standards, particularly in less developed regions.
4. **Leverage Data for Decisions:** Utilize interactive dashboards like the one developed in this project to monitor trends, identify problem areas, and evaluate the impact of policies over time.
5. **Avoid One-Size-Fits-All:** Design policies tailored to specific country contexts, acknowledging the vast intra-continental disparities that continental averages can mask.

In conclusion, this project provides a robust, data-driven framework for understanding and addressing global income inequality. By quantifying disparities, highlighting trends, and providing an interactive tool for exploration, it offers a solid foundation for crafting effective policies in pursuit of UN SDG Goal 10.

### References

- Dataset: Global Inequality Data (2010–2021)  
(<https://www.kaggle.com/datasets/psterk/income-inequality/data>)
- United Nations Sustainable Development Goals (SDG 10: Reduced Inequality) (2022) by Sinan Kufeoglu