Global Education and Socioeconomic Dynamics: Analyzing Gender, Performance, and Economic Correlations

1 Introduction

Education is a fundamental driver of social and economic progress; however, access to quality education remains highly unequal worldwide. Disparities based on gender, socioeconomic status, and geographical location continue to impede efforts toward achieving global educational equity. While education promotes economic growth, improved health outcomes, and active citizenship, it does not always succeed in reducing social inequalities, particularly in nations with low Human Development Index (HDI) scores (World Economic Forum, 2023a). This inequality is often exacerbated in developing regions where limited resources, insufficient infrastructure, and socio-cultural barriers restrict access to quality education for marginalized communities.

Furthermore, socioeconomic development is closely linked to academic outcomes, such as literacy rates and proficiency in core subjects like reading and mathematics. Countries with higher HDI scores consistently exhibit better educational performance and literacy achievements, reflecting investments in educational infrastructure, teacher quality, and policy frameworks. In contrast, regions with lower HDI scores face significant barriers to accessing quality education, including lack of funding, shortages of trained teachers, and inadequate learning environments. These challenges perpetuate cycles of underdevelopment, as children in underprivileged areas are often denied the opportunity to acquire the skills necessary for future economic participation (UNDP, 2023; OECD, 2023). Addressing these systemic issues requires a focus on inclusive policies that prioritize equitable access to education for all socioeconomic groups.

Unemployment rates are another critical indicator associated with educational outcomes. Individuals with higher levels of education generally experience lower unemployment rates, as they possess the skills and qualifications needed to compete in the labor market. This highlights the importance of aligning education systems with labor market demands to reduce unemployment and foster sustainable development (World Economic Forum, 2023b). In many regions, particularly those with developing economies, a mismatch exists between the skills taught in schools and the demands of the workforce. As a result, even educated individuals may struggle to find suitable employment, perpetuating economic stagnation. Integrating vocational training, technical education, and job-oriented curricula into national education systems can bridge this gap and ensure that graduates are equipped with relevant skills to thrive in the modern workforce.

Additionally, the interconnected nature of education and socioeconomic outcomes is evident in its broader implications for poverty alleviation, gender equality, and population dynamics. Higher levels of education empower individuals, particularly women, to make informed decisions about family planning, healthcare, and career opportunities, contributing to reduced birth rates and improved quality of life (UNESCO, 2020). This is particularly significant in developing regions, where investments in education can lead to a transformative cycle of empowerment, poverty reduction, and sustainable growth.

This report examines key global education indicators, including the relationships between literacy and math proficiency, socioeconomic influences, and global unemployment trends. Through comprehensive analysis, the study aims to provide actionable insights to guide policy development, address educational inequalities, and promote equitable and sustainable progress worldwide. By

identifying gaps in access, performance, and economic outcomes, this report highlights the importance of targeted interventions and strategic investments in education systems to create a more equitable and prosperous future.

2 Objectives

This project aims to perform comprehensive data cleaning, processing, and exploratory data analysis (EDA) to derive meaningful insights regarding global education and socioeconomic indicators. Specifically, the project aims to:

- Examine Education Trends: Analyze trends in education completion rates across different levels (primary, lower secondary, and upper secondary) and evaluate gender disparities in educational outcomes.
- Identify Relationships: Explore correlations between socioeconomic factors such as unemployment rates, birth rates, and education-related metrics (e.g. tertiary enrollment rates and proficiency levels in reading and math). This will help uncover connections between education and economic development.
- Highlight Global Disparities: Investigate differences in educational performance and socioeconomic indicators between countries, emphasizing the gap between developed and developing regions. Assess the influence of socioeconomic challenges on educational access and completion.

The findings from this analysis are intended to provide insights that can support decision-making processes at various levels:

- Policy Development: Governments can utilize the analysis to address disparities in education access, improve literacy rates, and implement policies that align education outcomes with broader economic goals.
- Strategic Planning: Education stakeholders and organizations can use the insights to allocate resources effectively, target underperforming regions, and promote equitable access to education opportunities.
- Economic Insights: Investors and international agencies can assess how educational trends influence broader socioeconomic factors such as unemployment and birth rates, aiding in strategic investments to foster long-term development.

3 Methodology

3.1 Data source

The dataset used for this project was sourced from Kaggle World Educational Data. It contains 202 records and 29 attributes covering various aspects of global education and socioeconomic indicators. These attributes include education completion rates across different levels, gender-based literacy rates, unemployment rates, and tertiary enrollment figures.

This dataset is well-suited for conducting a thorough analysis of global education trends, as it provides valuable insights into both educational outcomes and their connections with socioeconomic factors. By including data from multiple countries and diverse metrics, it enables a comprehensive evaluation of gender disparities, regional performance, and the relationship between education and broader economic indicators.

3.2 Database creation

The database consists of five interconnected tables (Figure 1):

- Countries Table: Stores country-specific data with country_id (PK), country_name, latitude, and longitude.
- Out_of_School_Rates Table: Captures out-of-school rates by gender and education level. References country_id (FK)
- Completion_Rates Table: Contains education completion rates for primary, lower secondary, and upper secondary levels, split by gender. References country_id (FK)
- Education_Proficiency Table: Tracks reading and math proficiency across grade levels. References country_id (FK).
- Social_Indicator Tables: Includes socioeconomic metrics like literacy rates, birth rate, tertiary enrollment, and unemployment rate. References country_id (FK).

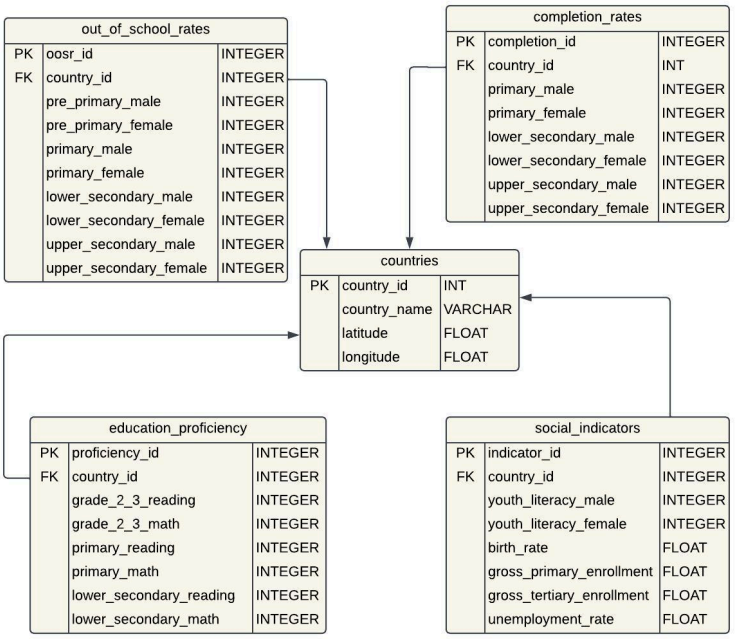

Figure 1: Relational database schema diagram

The countries data acts as the central hub, with all other tables linked through country_id as a foreign key. The current design maintains a 1:1 relationship since no year column exists. If a year column were present, the tables could support a 1:N relationship for time-based analysis.

The database was created and populated using SQLite with Python's SQLite3 library to ensure data integrity and efficient querying.

3.3 Data processing

Preparing the dataset for analysis is an essential process to ensure its accuracy, consistency, and dependability. The following steps were carried out to clean and refine the data, laying a solid for exploratory data analysis (EDA):

1) Initial Assessment of the Data

   The structure of the dataset was verified using `df.info()` and `df.shape` to understand the number of records, columns, and data types. The dataset initially had 202 rows and 29 attributes. To facilitate data processing, a separate index column (country_id) was added, increasing the total number of attributes to 30.

2) Handling Missing and Duplicate Values

   A check for missing values confirmed that the dataset had no missing entries, eliminating the need for imputation or row removal. Additionally, duplicate rows were examined using `df.duplicated()`, and no significant duplicates were found, ensuring the dataset remained intact without data loss.

3) Resolving Inconsistencies

   Inaccurate country names were corrected to ensure consistency. For instance, text encoding issues were resolved, and typos like "Guinea0Bissau" were replaced with the correct name, "Guinea Bissau".

4) Outlier Detection and Management

   Outliers in numerical attributes were visually inspected using boxplots for all key columns, including completion rates, proficiency metrics, literacy rates, and socio-economic indicators.

   - Outliers were identified using the Interquartile Range (IQR) method:

$$IQR = Q3 - Q1$$
$$Lower\ Bound = Q1 - 1.5 * IQR$$
$$Upper\ Bound = Q3 + 1.5 * IQR$$

   - Values outside these bounds were flagged as outliers.

   Outlier removal significantly reduced the dataset size from 202 to 64 rows, leading to the decision to retain outliers to preserve data integrity. Instead, rows with 15 or more zero values were removed, as they provided little meaningful information.

These steps laid a solid foundation for meaningful Exploratory Data Analysis (EDA) and ensured that the dataset was robust and ready for further insights.

3.4 Exploratory Data Analysis (EDA)

The exploratory data analysis (EDA) aimed to identify key trends, relationships, and insights within the dataset. Using statistical summaries and visualizations like bar charts, scatter plots, heatmaps, and choropleth maps, the analysis provided a clear understanding of the data.

1) Gender Comparison in Education Completion Rates

   To analyze gender disparities in education completion rates across primary, lower secondary, and upper secondary levels:

- Average male and female completion were calculated using SQL queries.
- A bar chart was created to compare male and female completion rates visually.

2) Correlation Analysis Between Reading and Math Proficiency

   Correlation analysis was performed to explore relationships between reading and math proficiency at different education levels:

   - A correlation heatmap was created to visualize relationships between reading and math proficiency scores across grade 2-3, lower secondary, and upper secondary levels.

3) Relationship Between Tertiary Enrollment Rate and Birth Rate

   To investigate the relationship between tertiary enrollment rates and birth rates:

   - The correlation coefficient was calculated, showing the strength and direction of the relationship.
   - A scatter plot was used to visualize the trend between the two variables.

4) Global Comparisons Using Choropleth Maps

   Choropleth maps were created to provide a geographical perspective on key socioeconomic and educational indicators:

   - Tertiary Enrollment Rate: Countries with higher gross tertiary enrollment rates were highlighted in shades of blue.
   - Birth Rate: Birth rates across countries were visualized using shades of red, with higher birth rates represented by darker colors.
   - Unemployment Rate: The global unemployment rate was mapped using shades of purple to identify regions with higher unemployment

4 Discussion of key findings

- Gender Disparity in Education Completion Rates Across Levels:

  The bar chart (Figure 2) illustrates the comparison of average education completion rates for males and females across primary, lower secondary, and upper secondary levels. At the primary level, both genders show high completion rates, averaging around 42%, with females slightly outperforming males. In the lower secondary level, completion rates decline compared to the primary level, though the gender gap remains small, with females maintaining a marginal edge. At the upper secondary level, the completion rates decrease further, but females continue to exhibit slightly higher rates than males.

  Overall, the analysis reveals a consistent trend where females outperform males across all education levels, but both genders experience significant drops in completion rates as education levels progress.

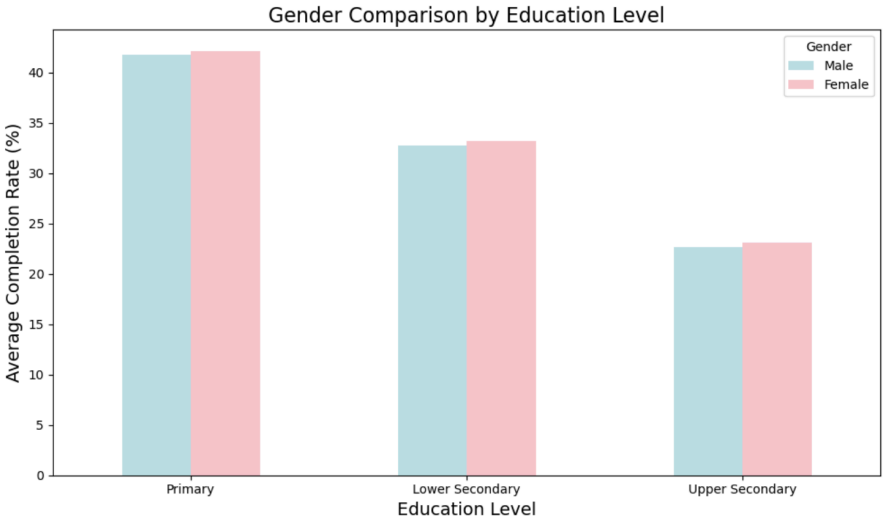

Figure 2 : Gender Comparison by Education Level

- Relationship Between Reading and Math Performance Across Education Levels:

  The correlation heatmap (Figure 3) highlights the relationship between reading and math proficiency across grade 2-3, primary, and lower secondary levels. Within the same education stage, strong correlations were observed, such as 0.82 between primary reading and math proficiency and 0.94 at the lower secondary level, indicating a close alignment of reading and math skills as students advance.

  However, the analysis also shows weaker correlations between different stages, such as grade-level performance and subsequent primary or lower secondary levels. This suggests that early education performance may not strongly predict outcomes in later stages, emphasizing the importance of early intervention to improve learning outcomes as students progress.

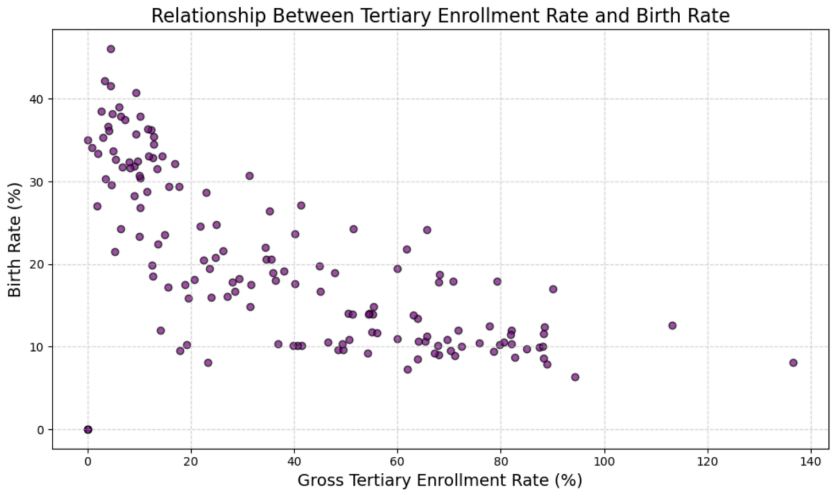

Figure 3 : Relationship Between Tertiary Enrollment Rate and Birth Rate

- Global Analysis of Tertiary Enrollment Rates and Birth Rates:

  The choropleth maps (Figure 4 and 5) Provide a geographic analysis of tertiary enrollment rates and birth rates across countries. Developed regions like North America, Western Europe, and Austrailia show high tertiary enrollment rates, often exceeding 50%, reflecting strong education systems and accessibility. In contrast, parts of Africa, South Asia, and Latin America display lower enrollment rates due to socioeconomic challenges and limited educational infrastructure.

  Birth rates follow an inverse trend, with Sub-Saharan Africa and parts of South Asia exhibiting the highest birth rates, often exceeding 30-40%. In comparison, regions with high tertiary enrollment, such as Europe, North America, and East Asia, show lower birth rates (below 10%). These results reinforce the negative relationship between tertiary education and birth rates, highlighting how improved access to education can influence population trends.

Figure 4 : Global Tertiary Enrollment Rate

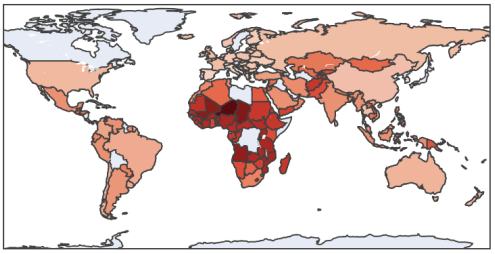

Figure 5: Global Birth Rate

- Global Unemployment Rate by Country:

  The global unemployment rate map (Figure 6) showcases significant regional disparities, with countries in Sub-Saharan Africa, particularly South Africa, experiencing the highest unemployment rates, as indicated by the darkest purple shades. Regions in South America, parts of Europe, and Asia exhibit moderate unemployment, represented by lighter purple shades, while developed economies in North America, Australia, and parts of Asia show relatively low unemployment rates.

  The map reveals that high unemployment rates are concentrated in developing regions, whereas developed economies display lower rates. This disparity highlights the influence of economic development, labor policies, and social conditions on unemployment levels. Further analysis could explore the relationship between unemployment and education or other economic indicators.

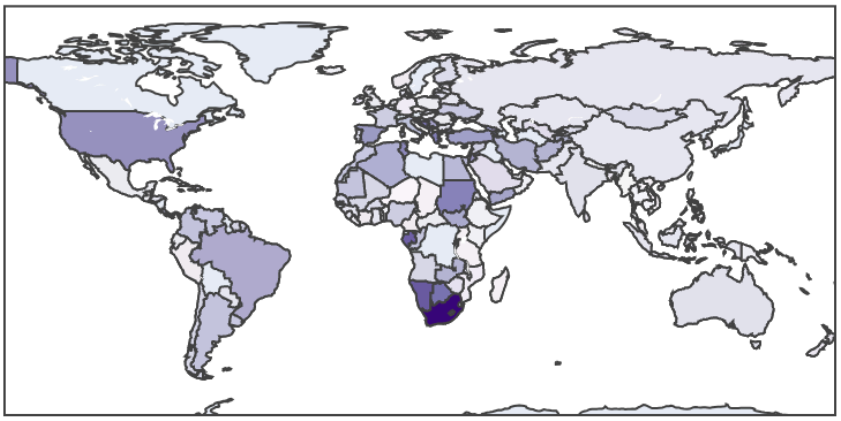

Figure 6 : Global Unemployment Rate by Country

5 Conclusions and recommendations

The analysis reveals key trends in global education and socioeconomic indicators. Gender disparities show that females consistently outperform males in education completion rates, though both genders experience a decline from primary to upper secondary levels, indicating challenges in education continuity. Reading and math proficiency demonstrate strong correlations within the same stage but weak connections across stages, highlighting gaps in foundational learning.

A negative relationship between tertiary enrollment rates and birth rates suggests that access to higher education, especially for women, contributes to lower birth rates through improved family planning. Regional disparities are evident, with developed regions showing higher tertiary enrollment and lower birth rates, while developing regions face challenges such as limited access to education and high unemployment rates, particularly in Sub-Saharan Africa. These findings underscore the need for targeted interventions to address education gaps and socioeconomic inequalities globally. The following recommendations have been derived from this study:

- Strengthen Education Continuity:

  Efforts should be made to reduce dropout rates at higher education levels by improving access to resources, offering financial support, and implementing targeted policies, especially in regions with low completion rates.
- Focus on Early Education:

  Address weak correlations between education by investing in early education programs. Strengthening foundational learning in reading and math can improve long-term educational outcomes.
- Promote Higher Education for Women:

  Policies that enhance access to tertiary education, particularly for women, should be prioritized. Such efforts can lead to broader socioeconomic benefits, including reduced birth rates and improved family planning.
- Address Regional Disparities:

Developing regions with lower tertiary enrollment rates and higher birth rates should receive targeted investments in education infrastructure, teacher training, and resource allocation to bridge existing gaps.

- Link Education to Employment Opportunities:

  To combat unemployment, education systems should align with labor market demands by promoting vocational training, technical education, and skill development programs, particularly in regions with high unemployment rates.

By adopting these recommendations, stakeholders can enhance global education outcomes, minimize socioeconomic inequalities, and foster inclusive growth while contributing to long-term sustainable development and improved quality of life across communities worldwide.

References

1. Organisation for Economic Co-operation and Development. (2023). *Education at a glance 2023: OECD indicators*. OECD Publishing. https://doi.org/10.1787/eag-2023-en
2. United Nations Development Programme. (2023). *Human development report 2023*. United Nations. https://hdr.undp.org
3. World Economic Forum. (2023a). *Global gender gap report 2023*. World Economic Forum. https://www.weforum.org/reports/global-gender-gap-report-2023
4. World Economic Forum. (2023b). *The future of jobs report 2023*. World Economic Forum. https://www.weforum.org/reports/the-future-of-jobs-report-2023
5. UNESCO. (2020). *Global Education Monitoring Report 2020: Inclusion and education*. United Nations Educational, Scientific and Cultural Organization.