World Education Data Analysis

Atul Satam
Department of Computer Science
St. Francis Xavier University
Antigonish, NS
x2022ekc@stfx.ca

Sanket Vagal
Department of Computer Science
St. Francis Xavier University
Antigonish, NS
x2022fjg@stfx.ca

***Abstract*—In the 21st century, understanding the need and impact of education is very vital. This project helps us take a deep dive into the world of statistics related to Education. We study various indicators, and trends and try to understand the challenges faced on multiple levels across the world. The idea is to explore the data obtained through sources of UNICEF, UNESCO and the World Bank to understand the underlying correlation between various indicators & implore if the obvious factors are the mainstay reasons or if there could be other indicators that significantly affect the data as well. We visualize various indicators based on their correlation to each other to our best knowledge for a better impact study and understanding for the general readers.**

I. Introduction

Education is a vital force shaping individual lives and the overall progress of societies. In this project, we'll be looking at key indicators like enrollment rates and literacy levels, aiming to understand the broader trends and challenges that influence education on a global scale. By tapping into widely available data, we hope to shed light on the patterns that impact the educational experiences of children worldwide. The visualizations and analysis would help an end-user understand how Education is currently being focused or neglected by regions based on their literacy rates, GDP per capita, etc.

II. Data Exploration

We started by collecting the data from the UNESCO, UNICEF & World Bank Dataset websites. They have their own set of datasets which follow various indicators to select & download the data from. As an end user, you can select what region, year or indicator data you want to download. The UNICEF data ranges on various topics such as Climate change, Health, Literacy, Covid-19, Poverty, Urbanization, etc. The UNESCO website on the other hand has its own designated portal to select. The various choices to select from the above are

1) Indicators
2) Regions & Countries
3) Target i.e. certain Frameworks.

This allowed us to go through almost 110-120 indicators & choose based on our preferences whether they would make sense against our base data which was selected from the UNICEF website.

The data we collected from multiple sources was then combined from the following datasets:

1) Education Statistics
2) Early childhood development Statistics
3) Economic indicators
4) Demographic Statistics
5) Global Literacy rates

Data cleaning & pre-processing was then done on the data. The header values were full of acronyms & numbers that had to be looked up in the appendix to make logic & would not be code-ready. There was a lack of benchmark indicators, there were only 4 primary benchmark indicators available on the UNESCO website. So, we changed the header values to more legible values & decided to drop any acronyms unless extremely necessary. This would help

any new end-user understand the data without going through a hassle. The incomplete data was a major issue since the data wasn't widely available for the developed countries. We also fixed NaN values as much as possible without filling in any Mean/Median data since it would skew the data & change it to an extent where it wouldn't match the real data. So, even if we did implement a method to complete the data, there would be no proper methodology to confirm the data. We also performed feature engineering by grouping countries by region, so it would be easier to understand if the trend we are analyzing only belongs to certain countries or a specific region. Thus, we now begin with the visualization of the data.

III. Result & Analysis

A. *Literacy Rate*

The first goal was to understand how the literacy rate grew over the years and how it is currently across various regions. We created an interactive map for the ones across the years from 1860 to 2010. Here we can see how the literacy rate has increased over the years. Some developed nations had higher literacy rates for a longer time as compared to some less developed countries. For example, you can see developed countries like China, and the USA having a 95+ percentage of literacy for a long time now. Whereas developing countries like India can show where it has been rising gradually and increasing a bit more after the independence of the country. African countries have a slow growth in terms of literacy since they are on the lower end of development due to various factors such as GDP, percentage of the budget that is allocated to Education, etc. We were unable to show the interactive map in the report due to restrictions on the document. Figure-1 below shows the literacy rate for various regions in the year 2018. As compared to previous years, the subcontinent and other regions have shown an increase in literacy while some regions are still lacking behind.

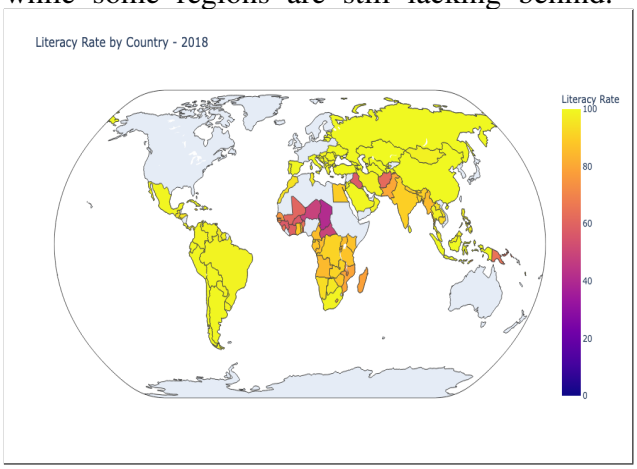

Fig 1- Literacy Rate in 2018

B. *Youth Literacy Rate*

Here we filtered the youth literacy rate by region (Fig-2) and by bottom 25 countries(Fig-3) to understand which regions were at the top and which were struggling due to underlying factors. We could drastically see that the Middle East & African regions were lacking behind which could be due to factors such as Infrastructure Challenges, Conflict and Instability, Health Challenges and access to Technology. The bottom 25 countries also consisted of the majority from the same regions.

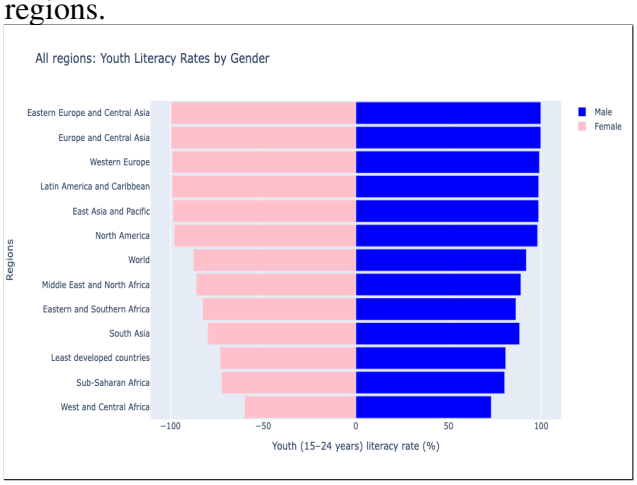

Fig 2- Youth Literacy Rate by Region

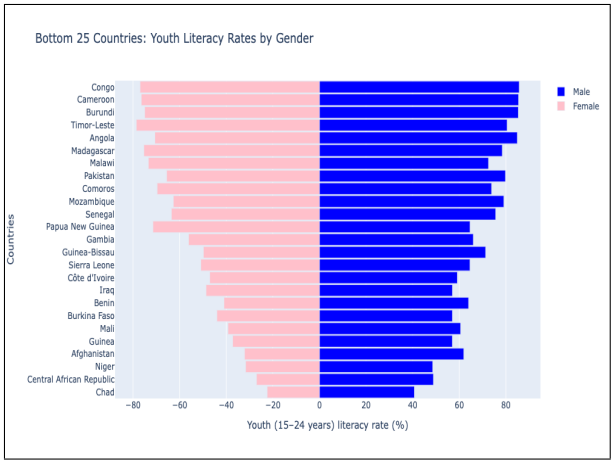

Fig 3- Youth Literacy Rate by Bottom 25 countries

C. Urban & Rural Population Proportion

Here we see how the regions with more urban populations have an extremely high literacy rate and the regions with more rural populations have a decreasing rate of literacy. The general average for the world lies in the region of 60 percent of the urban population with the literacy rate being in the higher 80s.

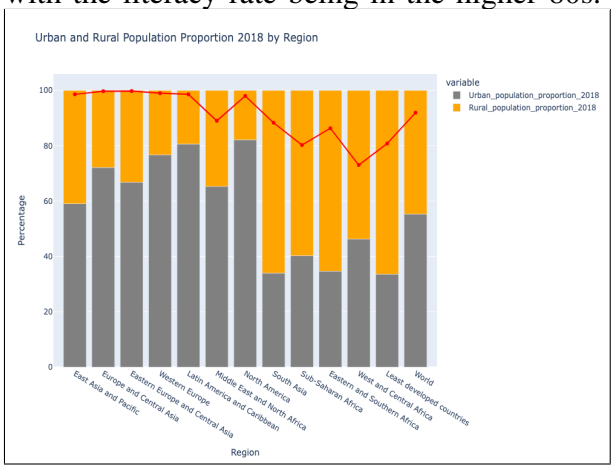

Fig 4- Urban & Rural Population Proportion by Region in 2018

D. GDP & Literacy Rates

This graph shows the relation between the GDP per capita and the literacy rates of that country, sorted by the GDP. We can observe that although a high GDP corresponds to a high literacy rate, as on the left side of the graph, the same is not true for the other way around. A high literacy rate may not always mean that the GDP is high. This is because, many developing nations contribute a higher percentage of their annual budget to education which although translates to higher literacy rates at the time of the survey, will only increase the GDP once they are in the workforce.

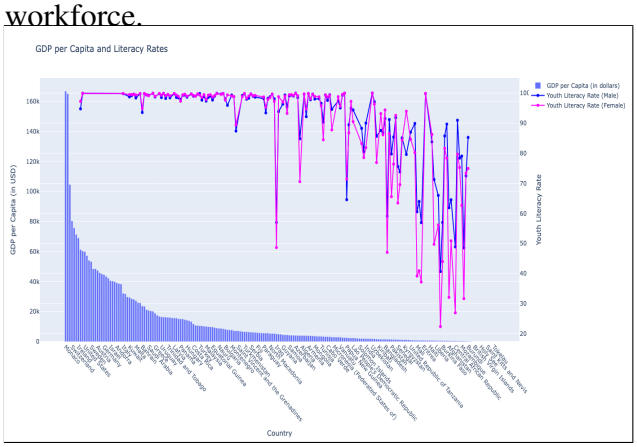

Fig 5- GDP per Capita (USD) & Literacy Rates

E. Dependency Ratio & Literacy Rate

The total dependency ratio is the ratio of the population aged 0-14 and that aged 65+ to the population aged 15-64. Ratio of number of dependants per 100 persons of working age (15-64). We can observe that for countries, where there are more dependants per working adults, the lit rate is lower,i.e. inversely proportional. This is because the people need to find work to support their families, as they have too many dependents and are the only ones who can earn a living.

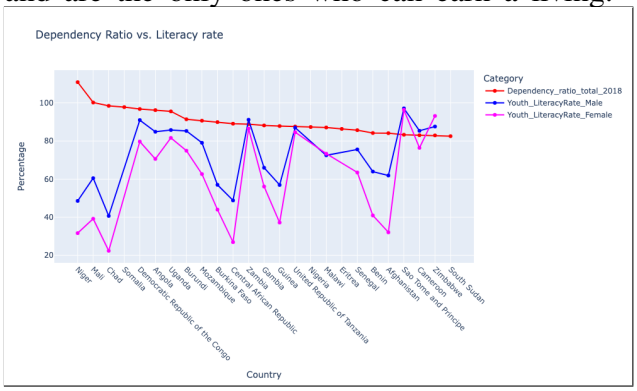

Fig 6- Dependency Ratio vs Literacy Rate

F. Male Out of School Rate - Region Based

Out of School rate is the number of students of an education level age who are not enrolled in that education level. This gives a closer look at OOS rates for males for each region. We can observe that most OOS rates are in the Pre-primary and upper-secondary levels, Which correlates to our previous findings with dependency. The Pre-primary OOS rates are high due to: Families might delay enrollment, thinking that formal education is less critical at such an early age, contributing to higher out-of-school rates. Government policies and priorities may focus more on primary and lower secondary education.

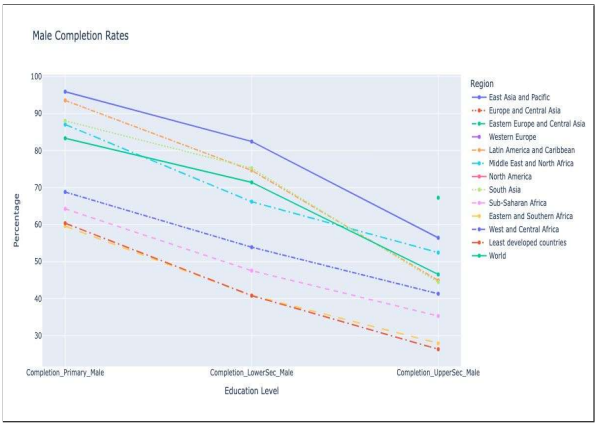

Fig 7- Male Out of School Rate Region Based

G. GDP & Out of School Upper Secondary Rate

As we saw before in the GDP vs Literate graph, a similar trend is found, where the lower the number of students who are not enrolled in education, the higher the GDP. One anomaly which can be found at around 43 percent is Qatar, which, although has a higher than average OOS rate, has a much higher GDP as compared to the similar OOS countries. This is because the majority of the population in Qatar consists of immigrants from either Western countries or those looking for labour jobs and coming with their families. This results in higher-than-normal GDP, whereas the OOS rate is higher as well.

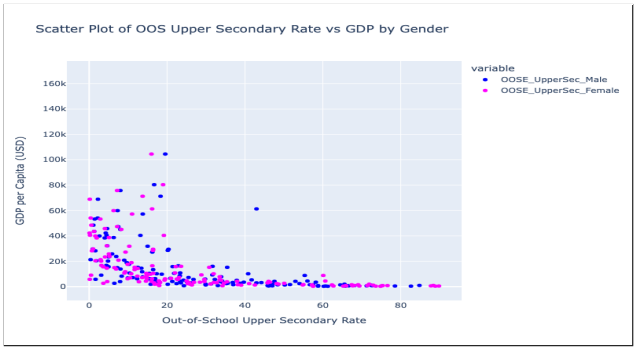

Fig 8- GDP vs OOS Upper Secondary Rate by Gender

H. Completion Rates - Overview

The completion rate stands for the number of students of an education level age who completed that education level. Higher Completion rates are better. The reasons stated before which affect the OOS rate also affect completion rates. It means that students who are enrolled in an education are more likely to complete it. The overall bar indicates the total completion rate for all education levels for that region.

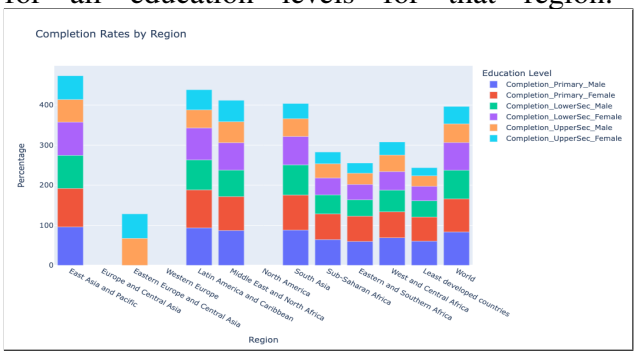

Fig 9- Completion Rates: Overview

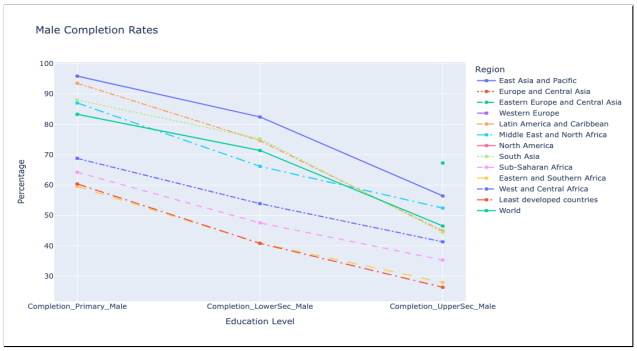

Fig 10- Male Completion Rate Region Based

IV. Conclusion

Thus, we discussed the trends in terms of Education & Literacy, concerning different Countries and Regions. Different education and social indicators were related to the overall economic and development factors of the countries. The visualizations gave us an insight into how the indicators are related to each other.

References

[1] https://data.unicef.org/resources/resource-type/datasets/

[2] http://sdg4-data.uis.unesco.org/

[3] https://databank.worldbank.org/source/education-statistics-%5e-all-indicators

[4] Hannah Ritchie, Veronika Samborska, Natasha Ahuja, Esteban Ortiz-Ospina and Max Roser (2023) - "Global Education" Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/global-education' [Online Resource]