Data Science Report

By margauxbonnaff, romkoval9, shyannetay17, queenaZ07, yw4111

1. Introduction

In 2015, the United Nations introduced the Sustainable Development Goals (SDGs), which are 17 goals designed to address pressing global issues by 2030.[i] This report focuses on assessing the progress of the six continents (excluding Antarctica) toward 2 key targets under Goal 8: Decent Work and Economic Growth

- Goal 8.1: Sustain per capita economic growth in accordance with national circumstances and, in particular, at least 7% gross domestic product (GDP) per annum in the least developed countries.
- Goal 8.6: By 2020, substantially reduce the proportion of youth not in employment, education or training.

Global datasets covering GDP per capita and youth NEET rates were cleaned, integrated and analysed to evaluate the performance of each country, identify disparities and track overall continental progress.

2. Methodology

For Goal 8.1, we followed UN's list of Least Developed Countries (LDCs), where LDCs are identified based on low income, weak human-assets and high vulnerability to economic shocks.[ii] When assessing "sustained economic growth", it is important to consider each country's national circumstance and historical growth trajectories, as economies experience diminishing returns to capital and productivity over time. Countries, mainly LDCs, can achieve rapid GDP growth and catch up with high income countries by adopting existing technologies from advanced economies.[iii] However, as this initial potential is exhausted, maintaining high growth requires structural transformation, which is more difficult to sustain and thus, growth tends to slow down. Applying a uniform growth target to all economies can therefore be misleading. Instead, growth expectations should reflect each country's stage of development. Based on this rationale, we defined the thresholds for "sustained growth" as shown in Figure 1 below.

Type of economy	Required GDP per capita growth rate
LDC	7%
Non-LDC	4%

Figure 1. Conditions necessary to qualify as "sustained growth"

Initial NEET (2015)	Required Reduction (%)
Below 10%	2%
10% - 20%	5%
20% - 30%	10%
30% - 50%	15%
Above 50%	15%

Figure 2. Conditions necessary to qualify as "substantially reduced"

Likewise, for Goal 8.6, the baseline youth not in unemployment, education or training (NEET) level in each country significantly influences the difficulty of achieving further reductions. To ensure a fair and context sensitive evaluation, this report applies a tiered assessment approach that adjusts expected reductions based on a country's initial NEET rate in 2015. Countries with already low NEET rates (below 10%) are near the practical minimum, since a small proportion of youth will always be in temporary transition phases or face personal circumstances preventing access to work or study. Hence, only a small reduction is expected in these cases. In contrast, countries with high NEET rates (above 30%) typically face systemic barriers such as skill mismatches.[iv] While these countries should aim for larger

reductions, progress is often constrained by these structural challenges. As a result, the tiered approach as shown below in Figure 2 allows for a fairer comparison across diverse national contexts.

For the additional CSV file, we extracted male and female youth NEET levels from World Bank's database, which is also where the other 2 given data sets were taken from, making this additional dataset highly credible and appropriate for comparison.

All datasets were cleaned to include only country-level data. Aggregate regional groupings such as "East Asia and Pacific (WB)" and income classifications like "High-Income Countries" were excluded. For comparability across indicators, the analysis periods were standardized: 2000 – 2021 for GDP per capita (Goal 8.1) and 2015 and 2020 for youth NEET (Goal 8.6). These timeframes were chosen to accommodate differences in data availability as some countries had incomplete or inconsistent time series. Also, to enable fair comparison across continents and prevent overrepresentation of countries with smaller populations, we scaled all metrics by population. Data on the population size was taken from World Bank. For GDP per capita, values were weighted by total population to calculate continent-level averages that better reflect the share of the population experiencing sustained growth. Similarly, NEET values were weighted by the youth population (ages 15-24), consistent with the official SDG definition of the target group. This adjustment reduces distortion from small-country outliers and yields a more representative regional picture.

Data reliability was tested by calculating the percentage of missing data points per continent, highlighting areas with limited or inconsistent coverage.

Continent	Missing Percentage (%)
Asia	0.49
Europe	0.97
Africa	2.36
North America	4.55
South America	0.00
Oceania	1.30

Figure 3. % of missing data for GDP (2000 – 2020)

Continent	Missing Percentage (%)
Asia	39.3
Europe	4.07
Africa	69.6
North America	27.8
South America	38.6
Oceania	72.2

Figure 4. % of missing data for youth NEET (2015 – 2000)

The results in Figure 3 indicate strong data coverage across continents as the percentages of missing data points for GDP are quite low. This suggests that the GDP per capita dataset is reliable for conducting analysis. On the other hand, the youth NEET dataset shows much higher percentages for missing data points, particularly in Africa and Oceania. Therefore, we decided to restrict the analysis to 2015 and 2020, where there was more consistent reporting across continents. Focusing on these 2 years provides the best possible estimate with the available data.

3. Results

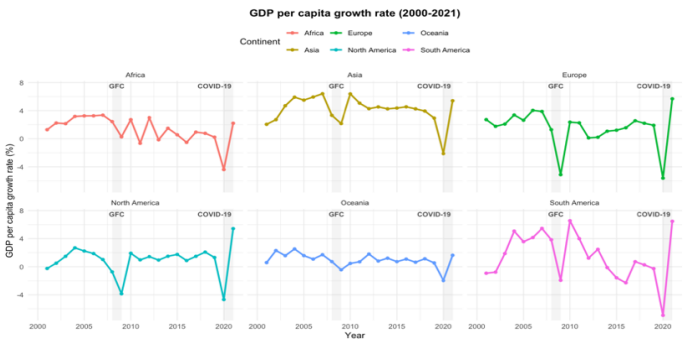

Figure 5. GDP per capita growth rate

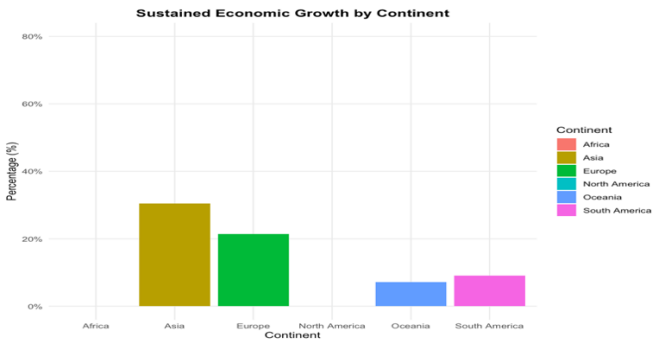

Figure 6. Sustained growth by proportion of population

Figure 5 shows that across all continents, the most significant disruptions to GDP per capita growth occurred during the Global Financial Crisis (2008-2009) and the Covid-19 pandemic (2020-2021). Beyond these shared downturns, Asia demonstrates the strongest growth, characterised by higher average GDP per capita and a relatively stable upward trajectory with fewer dips. Europe, North America and Oceania exhibit lower but steady growth, with fluctuations confined to a narrow range. Meanwhile, South America and Africa display greater volatility, with sharp and erratic swings in their growth rates.

Figure 6 further reinforces Asia's strong economic performance, with the largest proportion of countries achieving sustained growth. Europe comes next, followed by South America and then Oceania. Africa and North America both record 0%, though likely for different reasons. North America's results reflect its composition of advanced, high-income developed countries which experience diminishing returns to growth. Thus, growth is stable but below the 4% benchmark. Conversely, Africa's 0% is driven by its concentration of LDCs, where limited infrastructure hinder sustained economic growth. The findings reveal that continents with a balanced mix of LDCs and non-LDCs are better positioned to achieve sustained growth.

Looking at Figure 7 on the right, across all regions, all but 1 LDC fall short of achieving 7% average annual GDP per capita growth over the 20-year period, indicating that the global benchmark remains largely unmet. Asia stands out as the strongest performing continent, with most of its LDCs maintaining comparatively high growth rates. Among them, Myanmar is the only country to have reached the 7% average annual growth target, contributing to Asia's solid performance relative to other continents. There are no LDCs in Africa, North America and Oceania that have met the goal.

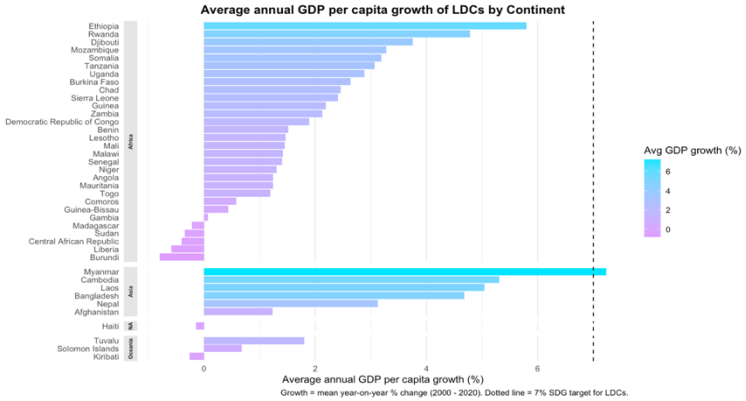

Figure 7. GDP per capita growth of LDCs by continent

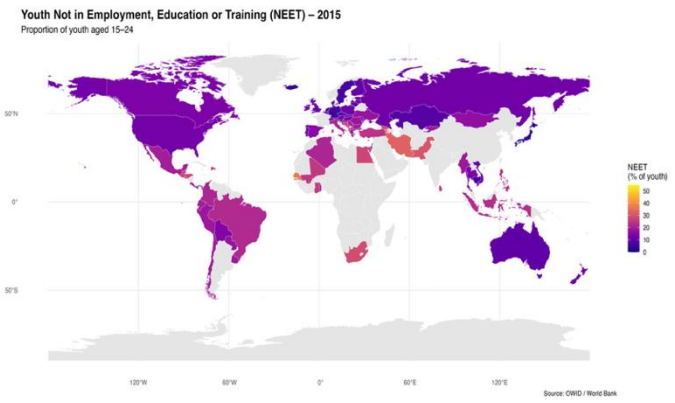

Figure 8. Global distribution of youth NEET (2015)

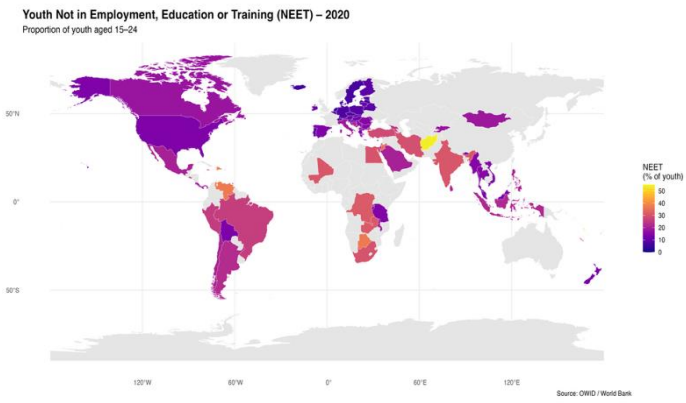

Figure 9. Global distribution of youth NEET (2020)

From Figure 8 and 9, youth NEET levels have generally worsened across countries, reflecting the combined effects of economic slowdown and early pandemic related disruptions by 2020, where many youths were laid off from work or displaced from education. The rise in youth NEET levels can be most clearly observed in North and South America, where majority of the countries in both continents shaded turn lighter in shade from 2015 to 2020. However, we can also see that large portions of Africa and Asia are grey, reflecting significant data gaps that make conclusions for those continents hard to draw. Additionally, Russia and Australia, which account for a large share of their respective continental

youth population, have available NEET data for 2015 but not for 2020. Since both countries typically report lower NEET levels as they are non-LDCs, the lack of their data in 2020 causes the averages for Asia and Oceania to appear higher than they actually are.

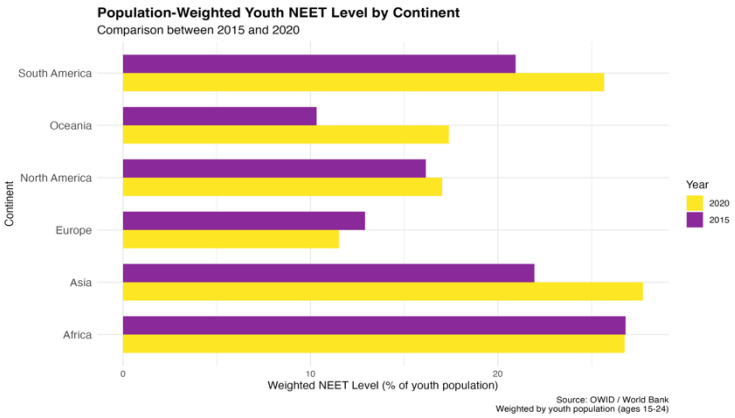

Figure 10. Change in Youth NEET level by continent

Analysing Figure 10, clear differences emerge in how youth NEET levels have changed. Europe is the only region that recorded a reduction, indicating progress in integrating youth into education or employment. Africa's NEET level remained constant and the rest of the continents experienced increases in NEET levels over the same period. The largest rise occurred in Oceania, followed by Asia, South America and North America.

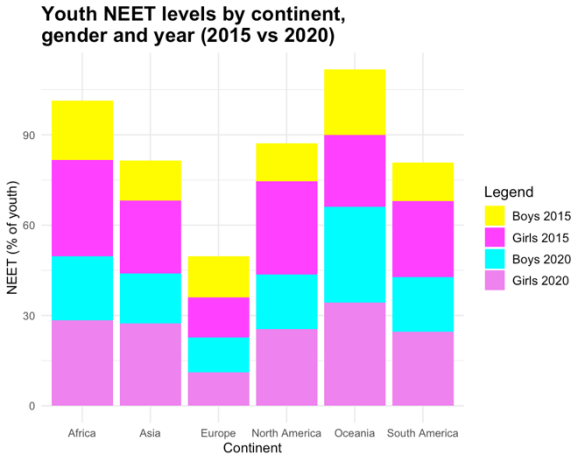

Figure 11. Youth NEET levels by continent, gender and year

Figure 11 shows increase in NEET levels for both boys and girls from 2015 to 2020 in all continents except Europe. Importantly, in all continents, the increase in NEET levels for girls is greater than that for boys. This underscores the persistent gender inequality in access to education and employment and suggests that girls were disproportionately affected by the economic downturn and social disruptions during this period. Since NEET levels for girls are consistently higher and more volatile, we can conclude that female youth across all regions are the primary contributors to the NEET outcomes for their respective regions and fluctuations in overall NEET level trends are largely driven by changes in the female population. Therefore, addressing gender-specific barriers for female youth must be a priority if substantial reductions are to be achieved towards meeting Target 8.6.

4. Discussion

The discussion seeks to address limitations of the data analysis, both in treatment and methodology.

1) The most significant challenge is the high level of missing data for youth NEET levels. This introduces potential selection bias as countries with complete datasets are typically more developed and have stronger statistical systems. As a result, the analysis may still overrepresent higher income countries and understate the challenges faced by developing regions such as Africa and parts of Asia despite scaling the data to population.
2) Furthermore, the aggregation of statistics, firstly by country and then by continent, oversimplifies conclusions and economic realities. Aggregation masks differences within continents and countries itself, and does not take into account other factors, such as the size of informal sector, that can contribute to a continent's progress in achieving the 2 goals set out.
3) The targets set create a binary classification for sustained growth, imposing rigid cut-offs. For example, the average of 7% annual GDP per capita growth rate was only attained by 1 country

out of all LDCs. LDCs that continue to achieve moderate but sustained improvement may still appear to underperform against this cut-off. This hints that goals set were unrealistic and the conclusion that most LDCs in each continent failed to attain sustainable growth may be unjustified.

4) In conjunction with Discussion Point 3, global shocks impacted both growth rates and NEET values. The 2008 Global Financial Crisis and the COVID-19 pandemic were disrupted economic growth and youth labour markets, introducing volatility into the short analysis windows. Following the 2008 crisis, many economies experienced temporary spikes in growth as they recovered to pre-crisis levels, but this increase is reflective of short-term recovery effects rather than sustainable, long-term improvement. Likewise, NEET levels between 2015 and 2020 would also have been severely impacted by Covid-19, or even the phases of an economic cycle. However, we should not exclude data from these 2 time periods even though it will make global progress towards SDG Goal 8 appear stronger. This would give a distorted picture since the aftereffects of these global shocks will continue to shape development patterns for the rest of the decade, offering a more accurate and realistic understanding of global progress in achieving sustainable growth and reduced NEET levels by 2030.
5) Classification of countries as LDCs or non-LDCs is difficult to ascertain. Although a fixed list was used to maintain consistency across years, this approach does not reflect changes in national development status over time, if any. Countries that graduate from LDC status experience structural economic shifts that affect growth potential and capability in reducing youth NEET levels and hence should be ideally reclassified in later periods. Keeping the same sample simplifies comparisons for transitioning economies.
6) Differences in sample size across continents also influence the reliability of aggregated results. Regions with fewer countries, such as Oceania, have smaller data samples, making their averages more sensitive to individual country performance. Particularly in the case of Oceania, Australia's large population size means that its data dominates the continent's data and trends, overshadowing smaller island economies within the region. This uneven distribution can lead to higher variance and volatility in smaller regions, meaning that a single country's fluctuations can disproportionately shape trends within the continent.

5. Conclusion

In conclusion, this report assessed progress towards UN SDG Goal 8, focusing on Goal 8.1 and Goal 8.6 through exploratory data analysis. Data gaps were discovered and we were able to recognise how data cleaning helped correct for these inconsistencies. For Goal 8.1, while Asia demonstrates strong growth, most LDCs and non-LDCs were unable to meet the benchmark set out for annual GDP per capita growth. This highlights that although global growth persisted, few countries maintained the pace required for sustained performance under UN's criteria. For Goal 8.6, results indicate a concerning reversal. Between 2015 and 2020, youth NEET levels increased in majority of the continents. The additional dataset of gender enriched the analysis by revealing that female youth had greater impact on NEET levels, emphasizing gender inequality as a key structural barrier to achieving this target. Overall, most countries were not fully on track to meet Goal 8 by 2020. While progress toward Goal 8.1 was observable in earlier years, the pandemic disrupted momentum, amplifying existing inequalities and slowing recovery. Achieving Goal 8 by 2030 will require stronger, more targeted policies that address gender-based barriers to inclusive economic growth.

Bibliography

---

[i] United Nations (2015). *The 17 sustainable development goals*. [online] United Nations. Available at: https://sdgs.un.org/goals.

[ii] UNCTAD (2024). *UN List of Least Developed Countries*. [online] unctad.org. Available at: https://unctad.org/topic/least-developed-countries/list.

[iii] Navarro-Pabsdorf, M., Rosa María Martínez-Vázquez and Jaime de Pablo-Valenciano (2024). Least Developed Countries: A Review of Worldwide Research. *SAGE open*, 14(2). doi:https://doi.org/10.1177/21582440241253952.

[iv] Haider, H. (2017). *Barriers to youth work opportunities*. [online] GSDRC. Available at: https://gsdrc.org/publications/barriers-to-youth-work-opportunities/.