Introduction to Data Science Sustainable Development Goal 8 Report

Group C13

Ishaan (ish248), Dominic (dtcolligan), ZiYanWu (wuziyan312), Inaya (inaya-j), Serena (serenayerusalmi)

GitHub: https://github.com/ish248/C13-Data-Science.git

Introduction

In 2015, all UN Member States adopted 17 Sustainable Development Goals to give a framework to address global challenges. This report looks at Goal 8: Decent Work and Economic Growth, specifically, Target 8.1: *"Sustain per capita economic growth in accordance with national circumstances and, in particular, at least 7 per cent gross domestic product growth per annum in the least developed countries"* and *8.6 "By 2020, substantially reduce the proportion of youth not in employment, education or training".* Using the continent, GDP per capita, NEET rate and the supplemental population CSV files, this report analyses trends over time across six continents to assess whether countries are moving towards the set goals.

Methodology

2.1 Selection of the Additional CSV file

A limitation of the *"gdp-per-capita-worldbank.csv"* is that it includes GDP per capita, not total GDP. Since Target 8.1 refers to annual GDP growth, using GDP per capita alone would be misleading. For example, a falling population with constant GDP would artificially raise GDP per capita.

To solve this, we selected *"devstatus_population_gdp.csv"* as our additional CSV. This lets us reconstruct GDP for any given country and year by multiplying the population by GDP per capita. With total GDP, we could calculate annual GDP growth, which helped us assess the proportion of LDCs achieving the target annual growth rate of above 7%. Tot additional CSV was essential as it let us convert the GDP per capita into a form that aligns with the UN's measurement of progress.

2.2 Cleaning the Datasets

Before merging the datasets, each file required cleaning to ensure consistency and reliability. For the GDP per capita dataset, we removed the unused "Code" column and standardised the column names by changing "Entity" and "Year" to "country" and "year".

The continent dataset was cleaned by filtering entities that didn't appear in the World Bank GDP file, e.g. regions like the EU, leaving us with 193 countries from 6 continents.

The population dataset also had redundant columns, e.g. population between ages 15-24. We removed all columns and kept only the country, year and total population.

The NEET dataset contained values every 3 years on average. We removed rows containing N/A values and converted the NEET/year columns into numeric formats. Finally, we grouped countries by year and continent to obtain a continental average.

2.3 Processing the Data

After cleaning, the datasets were joined based on country and year to form a master dataset.

Every three years, the list of LDCs is reviewed to determine whether any countries should be added or graduated from the list. The UN defines countries as "Least Developed" based on three major criteria: Income criterion, Human Asset Index (HAI) and Economic and Environmental Vulnerability Index. The income criterion is the three-year average estimate of the gross national income (GNI) per capita, while the remaining indices are constructed from multiple indicators that are combined into certain scores.

We extracted the current list of Least Developed Countries from the United Nations' database and used this data to add an extra column to our combined data set. We used this list to explore GDP growth in LDCs.

We then calculated a trimmed mean of our NEET data using the summarise() and trim = … functions, removing the highest and lowest 10% of NEET values before calculating the continental average. This reduced the influence of extreme values in the continental NEET. Trimming outliers smooths the data and gives a clearer view of each continent's underlying trend towards achieving the target.

The GDP per capita dataset did not contain the total GDP values, which are required to calculate annual GDP growth. We therefore reconstructed GDP for each country-year by multiplying GDP per capita by the total population in that given country-year, which allows us to measure changes in economic output directly. We then calculated annual GDP growth by comparing consecutive years for each country, ignoring those without a valid previous year. This provided consistent time series data of country-level GDP growth for all countries with sufficient data.

2.4 Chosen Graphs

To evaluate a continent's performance, we constructed 6 graphs in 3 categories:

1. The share of countries achieving ≥ 7% growth
2. The distribution of annual GDP growth
3. 3-year moving average annual GDP growth.

Each category had a graph for the general continental average and the average of LDCs in that continent.

The share achieving ≥7% GDP growth lets us assess the proportion of countries in each continent that meet the threshold, translating the target into a measurable indicator. The aggregation helps to reveal whether high growth is widespread or driven by a few economies.

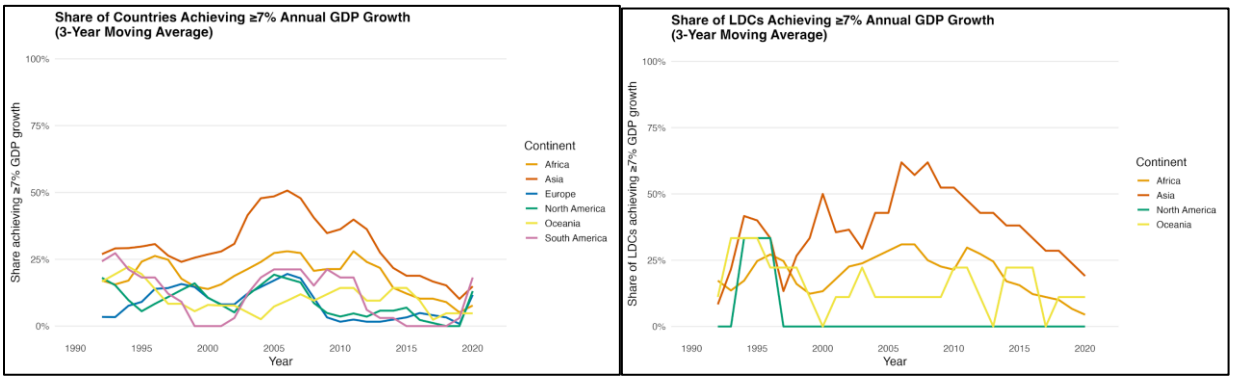
The boxplot was used to show GDP distribution as it clearly shows medians, quartiles and outliers, making it easy to identify typical growth from extreme fluctuations and whether continents are achieving stable growth or if performance is driven by outliers.

The 3-year moving average smooths out volatility and aligns with Target 8.1's emphasis on consistent patterns. The dotted line allows for visual comparison with the threshold growth rate.

Separate LDC graphs were created as Target 8.1 focuses specifically on LDC performance. Averaging the results stops non-LDCs from distorting results, letting us draw valid conclusions

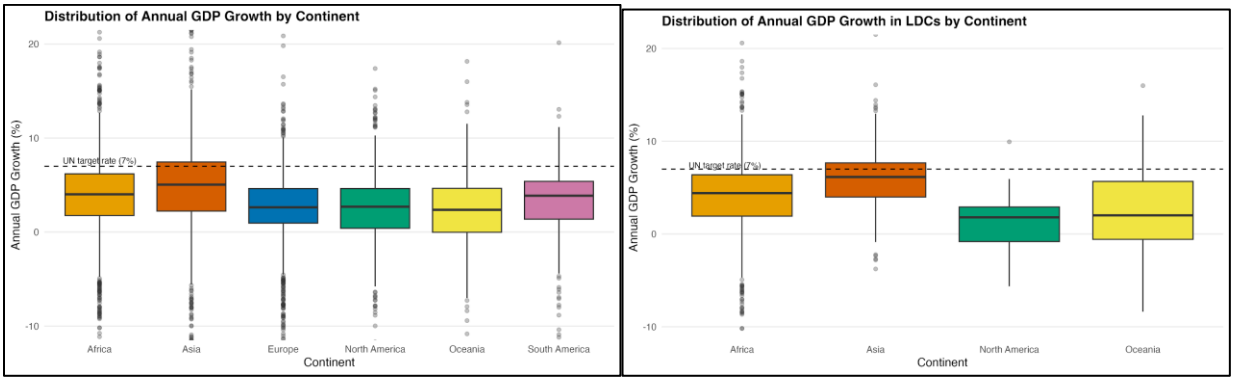
Finally, for the NEET rate, we chose a line graph as it illustrates how it changes over time. Since target 8.6 is about reducing NEET, the trend itself is important as it lets us track overall movements alongside being able to identify turning points or volatile periods.

Results



The first graph shows the proportion of countries achieving the target of 7% annual GDP growth, smoothed using a three-year average. It highlights how ambitious this target is: almost no continent (barring Asia in the mid-2000s) exceeds 50%. Africa and Asia have occasional spikes, but these drop quickly, showing that growth is brief and concentrated. Sustaining growth is, therefore, extremely rare.

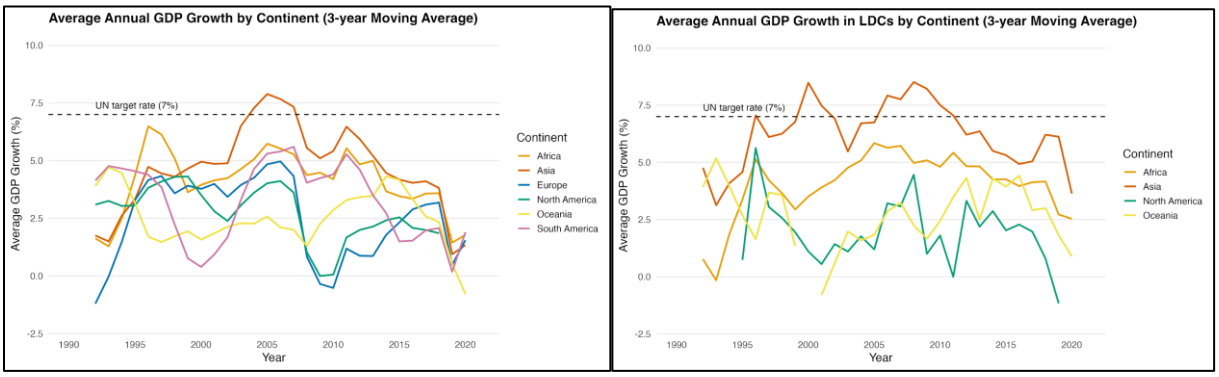
The second graph applies the same metric to LDCs. This shows that LDCs are more likely to exceed 7% growth, but this does not indicate consistent progress toward the target. By the late 2010s, the share of LDCs reaching the target falls, signalling regression rather than progress. Much of the higher incidence of ≥7% growth is due to volatility: countries often have high growth after downturns, and so these peaks do not represent structural, long-term improvement in LDC growth rates.



The set of boxplots shows the annual GDP growth by continent. In every region, the median growth rate is below the 7% target, confirming that rapid growth is not typical. Asia has the highest median and upper tail,

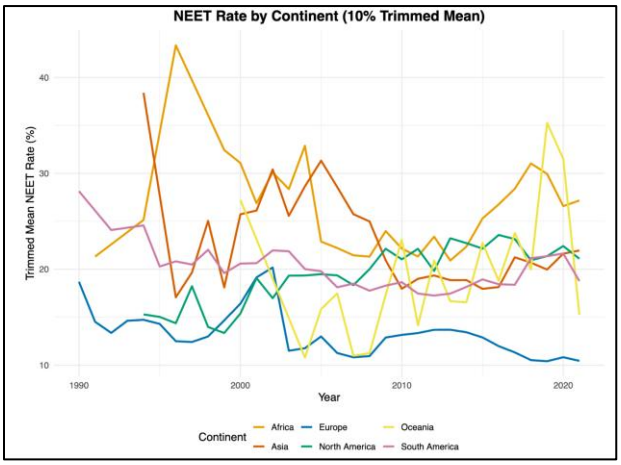
showing that while some countries achieve rapid growth, it is not typical. Africa is more volatile with frequent spikes, while Europe and North America cluster around 1-3%, reflecting slow, stable growth.

The LDCs-only box plots show that the target may be somewhat attainable, with a median growth rate just below 7% for Asian LDCs. This may reflect integration into global value chains. However, the median still does not reach the benchmark, with no other region coming close either, indicating that sustaining 7% growth remains structurally difficult even among LDCs.



This graph shows the general trend of GDP growth for all countries in each continent and shows that many continents are experiencing modest or declining growth.

The moving average lines for LDCs show that their growth rates often approach but rarely exceed 7%. Throughout the early 2010s, all regions faced a downward trend, placing LDCs further from the target. However, relative to all countries, LDCs do seem somewhat more capable of achieving target GDP growth than the rest when compared to the previous graph.



We observe that, despite using the trimmed mean, there is high fluctuation in the NEET values. Nevertheless, Europe and Asia are successful in reducing the NEET level. South America has also achieved this as its NEET rate has decreased by around 10%. However, in Oceania, there's been high volatility alongside a general upwards trend and an increase in the late 2010s. Africa has seen an increase, along with a huge spike in the late 1990s. Finally, North America has faced a slow and steady rise over the years.

Analysis

LDCs often struggle to meet the 7% target largely because their economies face high volatility, often relying on the production of primary commodities whose prices fluctuate. Structural challenges like low human capital, weak infrastructure and rapid population growth restrict long-term productivity. LDCs also face political instability and are vulnerable to external shocks, which can erase periods of high growth. As a result, even though some LDCs occasionally experience expansion, sustaining it over many years remains extremely difficult.

Analysing our trends gives us valuable insights into the differences between the continents' NEET rates. Europe's decrease could be explained by sustained investment in vocational education/training programmes alongside the expansion of higher education. Moreover, Asia's economic rise from the 1990s can be attributed to the fall in NEET, as younger people have more chances for employment and education. However, NEET levels are still volatile, possibly due to regional disparities. In South America, a similar trend can be seen. With increased government spending in programmes alongside higher accessibility to education, the continent-wide NEET fell.

Contrastingly, many African countries have high population growth, resulting in a lack of employment opportunities, forcing people out of work. Moreover, political instability contributes to a lack of investment in other programmes, leading to a higher NEET rate. Oceania's trend could be caused by a small sample size, as the trimmed mean still includes outliers, resulting in volatility. Also, differing socioeconomic conditions between countries lead to a regional rise, as the region can be sensitive to supply shocks, meaning employment opportunities keep changing. In North America, the decline in manufacturing industries from the 1990s reduced opportunities and recessions such as the 2008 Financial Crisis also put a burden on the economy, meaning it became challenging for workers to find employment, especially younger ones who lack experience.

Limitations

A key limitation is that the datasets are missing entries, meaning only countries with available data contributed to the continental calculations in a given year. Thus, some continents and LDC groups were represented by a changing set of countries, introducing inconsistency. Another limitation is that we don't weigh countries by population, so small and large countries have the same influence on the results, despite different population structures.

Another limitation of our analysis is that we used GDP growth, rather than GDP per capita growth, when evaluating Target 8.1. The UN indicator is defined as the sustaining per capita economic growth; the target text explicitly emphasises sustaining "at least 7 per cent gross domestic product growth per annum in LDCs". We chose to follow the language of the target itself, which places the threshold on economic output rather than per-person output. However, using GDP growth means that our findings reflect changes in total economic output rather than changes adjusted for population size, meaning that countries with a rapidly increasing population may appear to perform better than they would under the per-capita indicator, creating an important distinction when interpreting results.

Data quality also constrained our analysis. For Target 8.1, COVID-19 produced contractions followed by rebounds, reflecting short-term disruption rather than overall trends. This volatility made it harder to assess sustained growth. For the NEET rate, the data excludes the informal economy and doesn't distinguish between voluntary/involuntary inactivity. Therefore, it doesn't capture gap years, young parents temporarily out of the labour force or people in informal training contributing towards NEET.

Furthermore, Target 8.1 (minimum 7% GDP growth per annum in LDCs) depends on which LDCs had valid data that year, meaning changes may reflect missing values rather than performance shifts. For NEET, we can say if it has increased or decreased, but the % reduction to count as "substantial" is not defined, meaning judgment about this goal is subjective.

In addition, for data on GDP, the three-year smoothing reduces volatility, but hides information on short-term shocks. Analysis on the NEET rate uses the trimmed mean, which smooths out large fluctuations and removes some extreme values. However, policy failures (such as spikes caused by conflict) are not shown, understating severe youth unemployment crises in specific countries.

Finally, the graphs show correlations, not causation. Other influences, such as labour market structure, economic cycles, social norms (e.g., attitudes toward female employment), and government policies, influence GDP growth and youth NEET rates but are not captured in our visualisations.

Conclusion

In conclusion, the graph for Target 8.1 shows that while some regions reach the 7% GDP growth target, sustained progress is rare. LDCs sometimes achieve high growth, but this is driven by volatility rather than long-term improvement. Median growth rates across all continents remain far below the benchmark. Developed regions experience low, stable growth, while growth rates for all countries have declined since the 2010s. Overall, despite periods of rapid expansion, the evidence suggests that significant structural change would be needed for LDCs to achieve at least 7% GDP growth by 2030.

For Target 8.6, our analysis shows that although some continents have made progress, it hasn't been the general worldwide trend. Europe, Asia and South America show good progress in reducing NEET, but this is contrasted by the increase in Africa, Oceania and North America, taking them further away from achieving the target.

Bibliography

1. NCTAD (2024). UN List of Least Developed Countries. [online] unctad.org. https://unctad.org/topic/least-developed-countries/list.

2. United Nations (2025). Goal 8 | Department of Economic and Social Affairs. [online] United Nations. https://sdgs.un.org/goals/goal8#targets_and_indicators.

3. United Nations (2023). Economic Growth - United Nations Sustainable Development. [online] United Nations Sustainable Development. https://www.un.org/sustainabledevelopment/economic-growth/.

4. United Nations (2023). Goal 8 | Department of Economic and Social Affairs. [online] sdgs.un.org. Available at: https://sdgs.un.org/goals/goal8#progress_and_info.

5. United Nations (2025). THE 17 GOALS. [online] United Nations. https://sdgs.un.org/goals#history.

6. C13 (2025), C13-Data-Science – Scripts Folder [online] GitHub. https://github.com/ish248/C13-Data-Science

All graphs used in the report were generated using the code stored in the scripts folder in the C13 Data Science GitHub repository.