



Background

The United Nations' fifth Sustainable Development Goal (SDG) is to achieve gender equality and empower all women and girls. One of the targets under this goal (Target 5.1) is to end all forms of discrimination and violence against women and girls in public and private spheres. This includes ensuring equal rights to access education, healthcare, economic opportunities, political participation, ownership and control over property, inheritance, participation in decision-making processes at all levels, and equal access to financial services, including microfinance, by 2030.

The [datasets](#) provided host the latest sex-disaggregated data and gender statistics covering demography, education, health, access to economic opportunities, public life and decision-making, and agency. The datasets are at various levels of data granularity for eight African countries, including Egypt, Ethiopia, Ghana, Kenya, Morocco, Rwanda, Nigeria, and South Africa.

Demography

Demography is the statistical study of human populations, including age, gender, income, socioeconomic status, and geographical location. These datasets include various demographic statistics such as GDP (Gross Domestic Product) per capita and urban and rural population shares.

Education

The education statistics in these datasets include school enrollment for primary, secondary, and tertiary education and literacy rates. However, enriched datasets could include other education statistics such as educational attainment (the highest level of education completed by a person).

Health

The health statistics collected by the UN cover a wider range of diseases, access to medication and healthcare, and mortality rates per age group. Some examples of health statistics in this dataset include access to anti-retroviral drugs and incidence of HIV.

Access to economic opportunities

This category of statistics, often also known as *Decent work and economic growth*, is a set of statistics that relate to employment. This set of statistics is often categorised in the following way:

1. Wage and salaried workers, also known as employees.

A high proportion of wage and salaried workers in a country can signify advanced economic development.

2. Self-employed workers
 - a. Self-employed workers with employees (employers).
 - b. Self-employed workers without employees (own-account workers).
 - c. Members of producers' cooperatives and contributing family workers (also known as, unpaid family workers).

If the proportion of own-account workers (self-employed without hired employees) is sizeable, it may be an indication of a large agriculture sector and low growth in the formal economy.

3. Vulnerable employment

- a. Own-account workers
- b. Contributing family workers

A high proportion of contributing family workers - generally unpaid, although compensation might come indirectly in the form of family income - may indicate weak development, little job growth, and often a largely rural economy.

Public life, decision-making, and agency

This category of statistics often relates to the empowerment of females. It includes indicators that track whether or not a woman can apply for a passport, be head of the household, travel outside of the country, open a bank account, etc., in the same way as a man. This is a large category of indicators that are quantified as boolean values, where a value of 1 represents “yes” and 0 “no”.

Gender Parity Index (GPI)

The indicators under Target 5.1 include a ratio called the Gender Parity Index (GPI). It is a term used to describe the ratio of females to males in the specific target. The Institute for Statistics of UNESCO (United Nations Educational, Scientific and Cultural Organization) also uses this general definition of GPI: For any development indicator, one can define the GPI relative to this indicator by dividing its value for females by its value for males.

A GPI between 0.97 and 1.03 indicates parity between the genders. A GPI below 0.97 indicates a disparity in favour of males. A GPI above 1.03 indicates a disparity in favour of females.

Data models

Table 1: The data model for the Gender_parity_2022 dataset.

Column name	Data type	Feature description
Country	String	The country name.
Labour force participation rate, female	Float	The percentage of the working-age female population, ages 15 and older, that are actively taking part in the workforce.
Labour force participation rate, male	Float	The percentage of the working-age male population, ages 15 and older, that are actively taking part in the workforce.
Unemployment, female	Float	The percentage of the labour force female population that is unemployed.
Unemployment, male	Float	The percentage of the labour force male population that is unemployed.
Literacy rate, adult female	Float	The percentage of the adult female population, ages 15 and older, that can read and write.
Literacy rate, adult male	Float	The percentage of the adult male population, ages 15

		and older, that can read and write.
Rural population, female	Float	The percentage of the female population that lives in rural areas.
Rural population, male	Float	The percentage of the male population that lives in rural areas.
Urban population, female	Float	The percentage of the female population that lives in urban areas.
Urban population, male	Float	The percentage of the male population that lives in urban areas.
GDP per capita (in USD)	Float	The Gross Domestic Product (GDP) per capita, i.e., a country's GDP divided by its total population.
Population size	Int	The estimated number of people living in a country.

Table 2: The data model for the Gender_Egypt dataset.

Column name	Data type	Feature description
Country_name	String	The country name.
Country_ISO3	String	The country name abbreviation.
Year	Float	The year for which the indicator value was recorded.
Indicator_name	String	The indicator name and description according to SDG 5.
Indicator_code	String	The indicator code; includes substrings that identify the category of indicator and the gender of the specific indicator.
Value	Float / Int / Boolean	The value for a specific indicator, year, and country. Either float, int, or boolean, depending on the indicator.

The other tables included in this module, such as Country, Gender_inequality, and Indicator, have similar feature descriptions and data types as the data models in *Tables 1* and *2*.