**Data Dictionary: Strathmore University Hackathon**

In this document, you'll find short descriptions of the variables in the datasets provided. Read it carefully to understand the data you'll be working with better.

**Variables**

***Ref\_area (iso 3 country code) :*** This variable represents geographical location/region associated with the data.

***ref\_area.label:*** country full name

***source:*** This variable refers to the source of the data, indicating where the information was obtained or collected from.

***indicator:*** It represents the indicator being measured or observed in the dataset.

***indicator.label:*** This variable provides a label or name for the indicator.

***sex:*** It categorizes the data based on gender.

***sex.label:*** Broad label for the gender categories.

***classif1:*** This variable is a classification category that may help further categorize or group the data in the dataset.

***classif1.label:*** It provides a label or name for the first classification category, aiding in its interpretation.

***classif2:*** Similar to classif1, this variable is another classification category that may offer additional grouping or categorization.

***classif2.label:*** This variable provides a label or name for the second classification category, enhancing its clarity and understanding.

***time:*** It indicates the year when the observation was made.

***obs\_value:*** This variable holds the actual observed value or data point for the corresponding combination of reference area, indicator, sex, classif1, classif2, and time. It is the numerical information you'll be analyzing and interpreting in your dataset.

**Indicators**

***NB:*** *"ILO Modeled estimates" means that the data contains imputed observations for missing years.*

***Inactivity rate by sex and age -- ILO modelled estimates, Nov. 2022 (%):*** This indicator conveys the number of persons of working age outside the labour force (that is, not employed or unemployed) expressed as a percentage of the working-age population.

***Unemployment rate by sex and age -- ILO modelled estimates, Nov. 2022 (%):*** The unemployment rate conveys the number of persons who are unemployed as a percent of the labour force (i.e., the employed plus the unemployed).

***Employment by sex and economic activity -- ILO modelled estimates, Nov. 2022 (thousands):*** The employed comprise all persons of working age who, during a specified brief period, were in one of the following categories: a) paid employment (whether at work or with a job but not at work); or b) self-employment (whether at work or with an enterprise but not at work). The observations in this dataset have been imputed meaning you get values for all years

***Employment by sex, age and economic activity (thousands):*** The employed comprise all persons of working age who, during a specified brief period, were in one of the following categories: a) paid employment (whether at work or with a job but not at work); or b) self-employment (whether at work or with an enterprise but not at work). This dataset contains values for select years only (no imputations)

***Unemployment by sex, age and education (thousands) :*** The unemployed comprise all persons of working age who were: a) without work during the reference period, i.e. were not in paid employment or self-employment; b) currently available for work, i.e. were available for paid employment or self-employment during the reference period; and c) seeking work, i.e. had taken specific steps in a specified recent period to seek paid employment or self-employment. Data disaggregated by level of education are provided on the highest level of education completed, classified according to the International Standard Classification of Education (ISCED).

***Employment by sex, age and education (thousands):*** The employed comprise all persons of working age who, during a specified brief period, were in one of the following categories: a) paid employment (whether at work or with a job but not at work); or b) self-employment (whether at work or with an enterprise but not at work). Data disaggregated by level of education are provided on the highest level of education completed, classified according to the International Standard Classification of Education (ISCED).

***Inactivity rate by sex, age and education (%):*** This indicator conveys the number of persons of working age outside the labour force (that is, not employed or unemployed) expressed as a percentage of the working-age population. Data disaggregated by level of education are provided on the highest level of education completed, classified according to the International Standard Classification of Education (ISCED).

For more details please refer to the [ILO data explorer page](https://www.ilo.org/shinyapps/bulkexplorer1/?lang=en&segment=indicator&id=SDG_0831_SEX_ECO_RT_A&ref_area=ETH+GHA+KEN+NGA+RWA+SEN+UGA&timefrom=2010&timeto=2021&latestyear=TRUE).

**Steps**

**NB : Filter total population from the csv for just youth (15-35 years). Ensure shares from ILO match relevant youth age group i.e 15-24 shares should be applied to 15-24 population and 25+ shares from ILO should be applied to 25-35 population.**

1. Apply ILO inactive rate by age and gender to corresponding population for 2015 - 2024 to get total inactive population (**ILO inactive share \* population**)
2. Apply ILO unemployment rate by age and gender to corresponding population for 2015-2024 to get the total unemployed population. Please note that the unemployed rate is expressed as a percentage of the labour force and **not the total population**. **(ILO unemployed rate \*(tot pop population - inactive population))**
3. Get total employed population by subtracting inactivity and unemployed population from total population
4. Using the c indicator in the dataset, apply the shares of each ISIC REV 4 sector to the total employed population to get absolute numbers for each sector. Please note that the obs value provided by the ILO does not represent the population. It should only be used to get shares for each sector. I.e **shares for agric = agric obs value/ total for ISIC REV 4 sectors. This share should then be multiplied by the employed population**
5. Since the **employed by sex, age and economic activity not modeled indicator** only provides values for individual/select years, you will need to get the sector growth rate from the modeled data set (***Employment by sex and economic activity -- ILO modelled estimates, Nov. 2022 (thousands)*** and assume that these growth rates would also apply to the youth population **(15-35 years)**. For example, if the share of females in agriculture in the modeled dataset is 55% in 2019 and 57% in 2020 (2 percentage points increase) and you only have 2019 sectors for Kenya in the not modeled dataset (by age and sex). Assuming the share of agriculture for female youth is 49% then in 2020 you would assume the share also grows by 2 percentage points to 51%.
6. The detailed sectors are a summarized version of the ISIC REV 4 sectors. To get the growth rate for Information and communication sector(IT), you will use the growth rate from the **Detailed: Transport storage and communication** in the ***Employment by sex and economic activity -- ILO modelled estimates, Nov. 2022 (thousands)*** indicator. Please refer to the table below to see how the sectors are aggregated.
7. For employment, unemployment and inactivity by sex, age and education, use aggregate education levels i.e less than basic, basic, intermediate and advanced. Note that the inactivity by sex, age and education indicator is expressed as a percentage and hence this rate will be multiplied directly by the inactive population. The employment and unemployment by sex, age, education indicators on the other hand are expressed in thousands and hence you will need to derive the rates(shares) from the aggregate education levels and multiply these by the employment and inactive population respectively. Education is only provided for select years. Take the latest year for each country e.g. for Kenya this would be 2019 Assume these shares remain constant till 2023. While showing trends on education, do it for select years due to limited data availability
8. Start by showing general trends in the labour market (unemployed, employed and inactive), show this by education level then go into analyzing the sectors specifically the **IT sector(** what are trends, differences between genders, differences between countries etc)

NB: In the datasets provided, sectors are aggregated into broad sectors, aggregate sectors and ISIC- Rev 4 sectors in the not modeled sectors/economic activity indicator **(Employment by sex, age and economic activity (thousands))** and into broad sectors, aggregate sectors and detailed sectors in the modeled sector/economic activity indicator **(Employment by sex and economic activity -- ILO modelled estimates, Nov. 2022 (thousands)).** For this task, please focus on getting the growth rates from the detailed sectors in the modeled economic activity for all years and apply those to the ISIC Rev 4 sectors in the not modeled economic activity to fill in the missing years. Please refer to the table below

**Sectors**

|  |  |
| --- | --- |
| **Sector(ISIC Rev 4)** | **Detailed Sectors (Modelled)** |
| A. Agriculture; forestry and fishing | Detailed: Agriculture; forestry and fishing |
| B. Mining and quarrying | Detailed : Mining and quarrying |
| C. Manufacturing | Detailed : Manufacturing |
| D. Electricity; gas, steam and air conditioning supply | Detailed :Utilities |
| E. Water supply; sewerage, waste management and remediation activities |
| F. Construction | Detailed: Construction |
| G. Wholesale and retail trade; repair of motor vehicles and motorcycles | Detailed : Wholesale and retail trade; repair of motor vehicles and motorcycles |
| H. Transportation and storage | Detailed : Transport, storage and communication |
| J. Information and communication |
| I. Accommodation and food service activities | Detailed: Accommodation and food service activities |
| K. Financial and insurance activities | Detailed : Financial and insurance activities |
| L. Real estate activities | Detailed : Real estate, business and administrative services |
| M. Professional, scientific and technical activities |
| N. Administrative and support service activities |
| O. Public administration and defence; compulsory social security | Detailed : Public administration and defence; compulsory social security |
| P. Education | Detailed:Education |
| Q. Human health and social work activities | Detailed:Human health and social work activities |
| R. Arts, entertainment and recreation | Detailed: Other services |
| S. Other service activities |
| T. Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use |
| U. Activities of extraterritorial organizations and bodies |
| X. Not elsewhere classified |  |