**Hackathon for students at Strathmore University: "Unleashing the Potential of Youth Employment Data in Kenya’s IT Sector"**

**Goal:** Hosted in partnership between the World Data Lab and Advantage Austria, the goal of this hackathon is to use data science techniques and any programming language of your choice(Preferably R/Python) to uncover insights and design solutions related to the job market in Kenya’s IT sector, specifically focusing on youth employment **(15-35 years)**. While the focus should be on Kenya’s IT sector, please feel free to also make comparisons to Rwanda and Ghana. Participants will have access to labor market datasets and will be challenged to use these datasets to build data visualizations, models, or other tools that can help policymakers, educators, and other stakeholders better understand and address youth job market trends in Kenya’s IT sector.

The Hackathon presentations will be held August 29th, 2023. Participants have until 23rd August, 2023 11:59pm to complete the task. Please send your results to [linda.karani@worlddata.io](mailto:linda.karani@worlddata.io)

**The task should be worked on individually**

**Datasets**: Participants will have access to the following :

1. The dataset provided by the International Labour Organization (ILO) for Kenya, Rwanda and Ghana which includes labour market indicators . Filter to get relevant indicators: e.g. *employment by age, gender, economic activity etc.*
2. A dataset containing **IIASA** **population projections** for Kenya, Rwanda and Ghana to aid in forecasting and rescaling of the ILO data to get total populations.

***Please note that a dictionary with definition of variables will also be provided separately***

**Deliverables:** At the end of the hackathon, participants will be asked to present their projects to a panel of judges. Projects can be in the following forms:

1. **Data Visualizations:** Participants can create infographics or interactive data visualizations using the IT employment datasets to illustrate trends, patterns, and relationships in the data. This could include visualizations such as bar charts, line graphs, scatter plots, maps, and more. The visualizations should aim to answer questions such as: What are the current employment trends for the Kenyan IT sector nationally and by demographics? How does this compare to Ghana and Rwanda? How does education level impact employment outcomes? What is recommended to improve youths’ chances of obtaining a job? What are the drivers of success for securing a job in the formal IT sector?
2. **Predictive Models:** Participants can use machine learning algorithms to build predictive models that can help forecast future outcomes for employees or businesses in the IT sector. For example, participants could use data on demographic characteristics to predict the likelihood of finding employment in the IT sector. The models should aim to answer questions such as: Given a young person’s background, what is their likelihood of finding employment in the IT sector? What is the predicted number of youth employed/unemployed/inactive each year out to 2030? Probabilistic models or time series models (e.g. exponential smoothing or ARIMA) could be considered.
3. **Tools for Employment Monitoring:** Participants can build tools that allow policymakers and other stakeholders to monitor the employment status of youth in the IT sector. This could include dashboards, web applications, or other tools that provide real-time updates on employment status and allow stakeholders to track trends and patterns over time. The tools should aim to answer questions such as: How has the employment status of youth in the IT sector changed over time? Which country is providing the most employment opportunities for young people in the IT sector?

**Judging criteria:** Projects will be evaluated based on their originality, thought process, technical merit, potential impact, and presentation quality. The code should **also be well-documented, reproducible, easy to run, and ideally recorded in GitHub.** Any statistical relationships or predictive models should be accompanied by a quantification of statistical significance and accuracy.

**Presentation:** The presentation should be around 6 slides, and the structure should be as follows:

1. **Slide 1:Individual background:** Briefly introduce yourself and share your aspirations for participating in the hackathon.
2. **Slide 2: Methodology**
   1. Explain the process you followed to analyze the data and develop your project.
   2. Mention the tools, techniques, and programming language you used.
3. **Slide 3: Data selection and processing**
   1. Highlight the variables you selected from the datasets and why you chose them.
   2. Discuss any data cleaning, processing, or preparation steps you took to get the data ready for analysis.
4. **Slide 4:** **Results**
   1. Present the outputs of your project, such as visualizations, models, or tools (i.e dashboard).
   2. Show how you addressed the goal of uncovering insights and designing solutions related to job market trends for youth in the IT sector in Kenya.
5. **Slide 5:** **Findings/Insights** 
   1. Discuss the insights and trends you discovered through your analysis.
   2. Highlight the most significant findings and explain their implications for the future of the job market.
6. **Slide 6:** **Recommendation/gaps/potential impact**
   1. Provide recommendations for policymakers, businesses, and other stakeholders based on your findings.
   2. Discuss the potential impact of your project.
   3. Summarize the contributions of your project.

**Prizes:** The top project will receive an interview for a paid internship with World data lab

**Criteria**

1. **Technical competence:** Participants should have a good understanding of data science techniques and the ability to use programming language, preferably R (and python), to work with large data sets and uncover insights.
2. **Data visualizations skills:** Participants should be able to tell a compelling story with their visualizations.
3. **Predictive modeling and machine learning skills**: Participants should have the ability to build predictive models that can forecast future youth employment/unemployment outcomes for the IT sector at a national/subnational level.
4. **Excellent communication and presentation skills:** Participants should be able to present their projects effectively, highlighting their key findings and recommendations, demonstrating their potential impact, and communicating this effectively to a non-technical audience.
5. **Teamwork:** Team player with the ability to work effectively in a team.
6. **Reproducibility:** Participants should have well-documented code that is reproducible, easy to run, and recorded in GitHub.