

UN Youth Data Science Hackathon

Handbook



IMPORTANT REMINDERS

- Official kick-off ceremony: **December 6th at 7:30 AM China Standard Time (UCT+8)**.
- Start of the hackathon: **December 6th at 8:00 AM China Standard Time (UCT+8)**
 - Datasets will be available
 - Theme = the impact on SDGs by the COVID-19 pandemic
- Final submissions: **December 8th at 6:00 PM Pacific Standard Time (UCT+8)**
Send to un-youth-hackathon@unmgcy.org
 - Presentation explaining the solution
 - Video with maximum length of 10 minutes = voice-over of the presentation
 - Coding scripts
- Jury selection: **end of December**

Please check your local times!!! We will send reminders but we expect that you have your own as well!

In numbers...



3 days



33 countries



70 teams



+250 participants

1 week to go!



In words...

1. **I**NNOVATE...
to participate in the achievement of the SDGs
2. **N**ETWORK...
with data-enthusiastic students and professionals
3. **L**EARN...
from your peers
4. **S**tengthen...
your teamwork skills
5. **C**OMPETE...
with leading data scientists across the globe
6. **E**XPOSE...
yourself to state of the art data analysis tools and methods



Objectives

The hackathon aims at:

- Providing a space for young people to develop their data science skills by having direct access to relevant data sets and working collaboratively with their teams and partners of the event.
- Supporting youth engagement in evidence-based decision making through the analysis of data.
- Giving the participants an opportunity to scale up their solutions for action towards the SDGs.
- Collecting innovative solutions that may help the implementation of the 2030 agenda for sustainable development



THE GLOBAL GOALS

Agenda

Time	Activity	Presented by
9:05~9:20	UNSD, Regional Hubs, MGCY messages (1 representative per group, 1-2 minutes intervention about the role of youth in data or the importance of data science in general). Order: 1. UNSD 2. MGCY 3. Brazil 4. Rwanda 5. China 6. UAE	All
9:20~9:25	Hackathon theme reveal	MGCY
9:25~9:30	Explanation of the hackathon timeline and logistics	MGCY
9:30~9:35	Explanation of expected outcomes (deliverables)	MGCY
9:35~9:40	Review of ideas of projects	MGCY
9:40~9:55	Walkthrough of the UN Platform	UNSD (Sean Lovell)
9:55~10:05	Jury selection, prizes, platforms, action items	MGCY
10:05~10:20	Q&A Session	All



UNSD



Ronald Jansen



Bogdan Dragovic



Sean Lovell



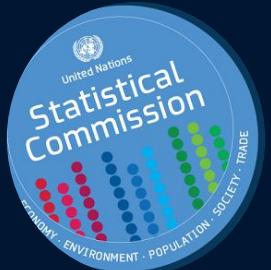
Jonathan Gessendorf



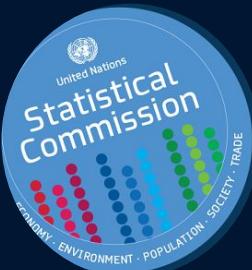
Karoly Kovacs



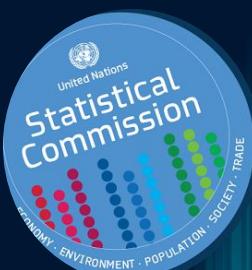
Can Ekingen



Catheryn Tajon



Clarence Lio



Markie Muryawan

MGCY



Oumaima Makhlouk



Camila Ramirez



Tori Lovins



Alinne Olvera



Rufai Balogun



Erica Eggleton



Abhijith Mathew

What brings us here?



address

the global challenges
of the 21st century

engagement

to lead the world
towards a sustainable
future

strategic

decision-making



Regional Hubs for Big Data



Latifa Alshehhi
(UAE)



Ruaa Alshehhi
(UAE)



Maryam Beyat
(UAE)



Noura Al Malek
(UAE)



**Augusto Cesar
Fadel (Brazil)**



**Andrea Diniz
da Silva (Brazil)**



Therese Uwimana
(Rwanda)



Bryan Gao
(China)

UN Global Platform Regional Hub United Arab Emirates

KNOWLEDGE
FOR PROSPERITY

UN Global Platform Regional Hub - UAE

Vision: A global collaboration to harness the power of data for better lives

Strategic Working Areas

1. Creating the largest data collaborative in the MENA region

2. Knowledge hub for the mena region

3. Data Science accelerator

4. Data tech for people and planet and for better social impact



Action Areas

Access to Data

Access to technology services

Access to Expertise

Data Science & AI practices in small offices

UAE Working Streams

Private sector engagement



4

Community & youth engagements



3

Exchange & divert knowledge from around the world

Strategic Collaborations with private companies

Academia and talent



1

Knowledge Center for Big Data

2

Private sector engagement



Strategic Collaborations with private companies

2021 Theme



THE IMPACT ON SDGS
(SUSTAINABLE DEVELOPMENT
GOALS) BY THE COVID-19
PANDEMIC.

Due to the on-going COVID-19 pandemic, people have been highly impacted financially, mentally and physically. The theme of this year's hackathon will focus in the implementation of solutions that would improve people's situations & well-being, especially at work, with their health and at school.

The theme is in line with the priority topics presented in the 2021 ECOSOC Youth Forum and in the 2021 HLPF, where the discussion centered around the **sustainable and resilient recovery from the COVID-19 pandemic**.



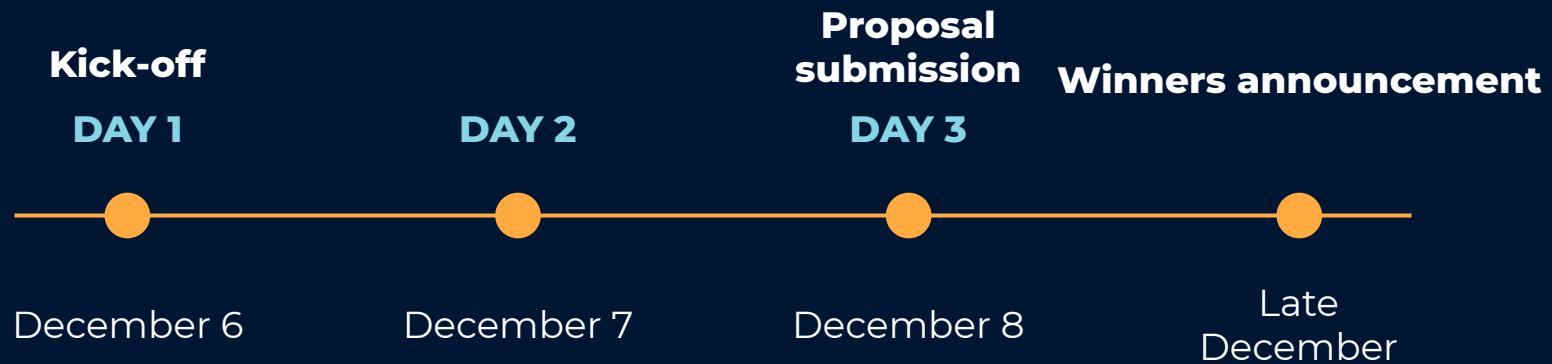
Guiding questions

- How were people impacted by the COVID-19 pandemic at work?
- Did the pandemic change the way schools operate?
- By how much did the pandemic increase socio-economic gaps in developing countries?
- Will the achievement of SDGs be impacted by the COVID-19?
- How were communities affected by the crisis?
- What data best shows the overall impact of COVID-19 on the world economy?
- How can we predict the future impacts of the crisis on the neediest communities?



THE GLOBAL GOALS

Timeline



- Access to the databases will be granted starting from **December 6th at 8:00 AM China Standard Time (UTC+8)**.
 - Proposals must be submitted on **December 8th at 6:00 PM Standard Pacific Time (UTC-8)**.



Expected outcomes

Hackathon = hack your way to an analytical solution that addresses the impact on SDGs of the COVID-19 pandemic

A **reference database** with 15+ datasets related to the theme will be provided.

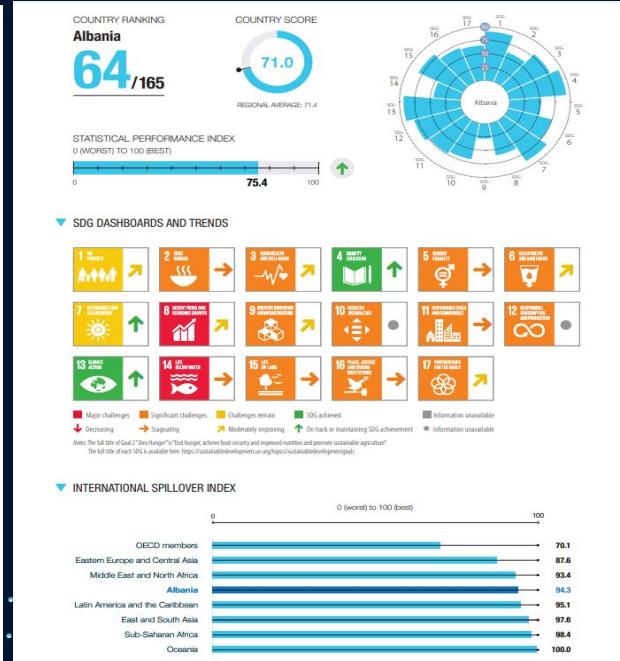
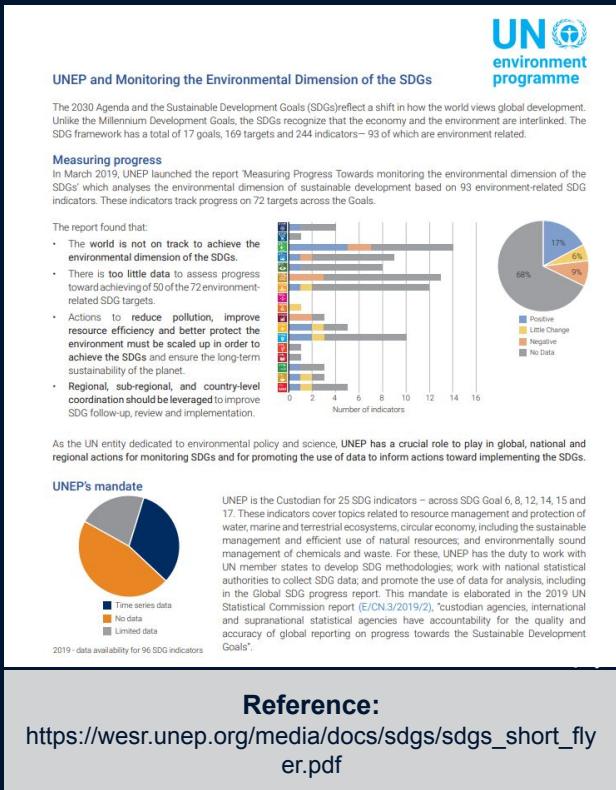
Teams must submit:

1. Presentation explaining their solution (free format).
2. Video with maximum length of 10 minutes
= voice-over of the presentation
3. Coding scripts.



Examples of outputs

Example 1:
Report providing detailed statistical analysis or data correlations between different datasets



Reference:
<https://s3.amazonaws.com/sustainabledevelopment.report/2021/2021-sustainable-development-report.pdf>

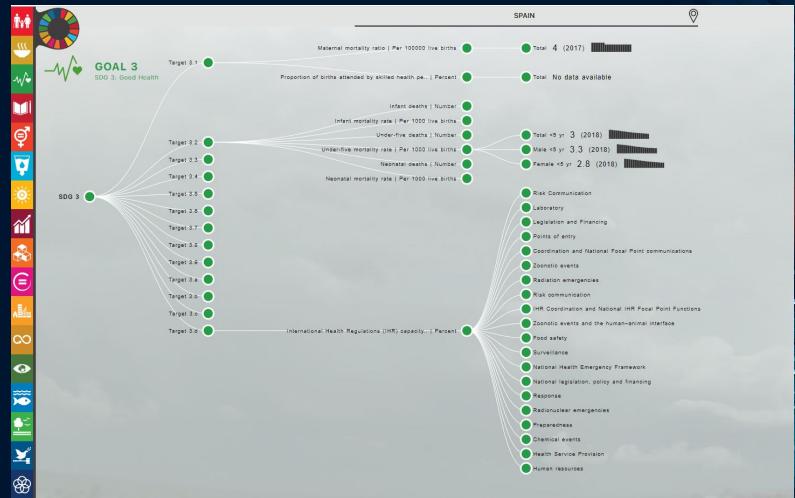
Examples of outputs

Example 2:

Interactive dashboard with enhanced analytical visuals



Reference: UNSD



Reference: <https://www.sdgssashboard.org/>

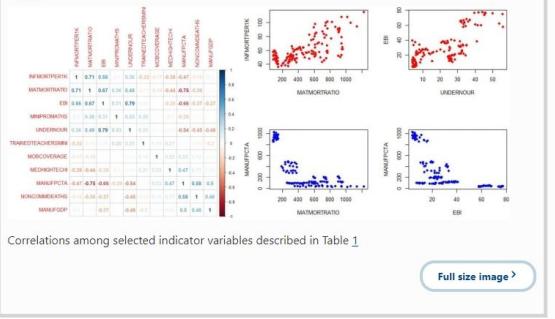
Examples of outputs

Example 3:

Machine Learning Models to predict output values

Univariate analysis has many limitations which arise from factors outside that variable. With each SDG associated with hundreds of indicators, the need to explore interactions and variations among data attributes is apparent. One way to explore such relationships is through correlation analysis which measures the strength of a linear association between two variables. The left hand side panel in Fig. 3 shows all paired correlations, with the strongest positively being between UNDERNOUR and EBI (7%) and between MATMORTRATIO and INFMORTPER1K (7%). The indicators MANUFPCTA and MATMORTRATIO exhibit strong negative correlation (~75%) while MANUFPCTA and EBI stand at -66%. The correlation plots for these four indicators are given on the right hand side panel. Effectively, the correlation function attempts to draw a line of best fit through the paired indicators without regard to which influences the other—hence the old dictum, correlation does not mean causation.

Fig. 3

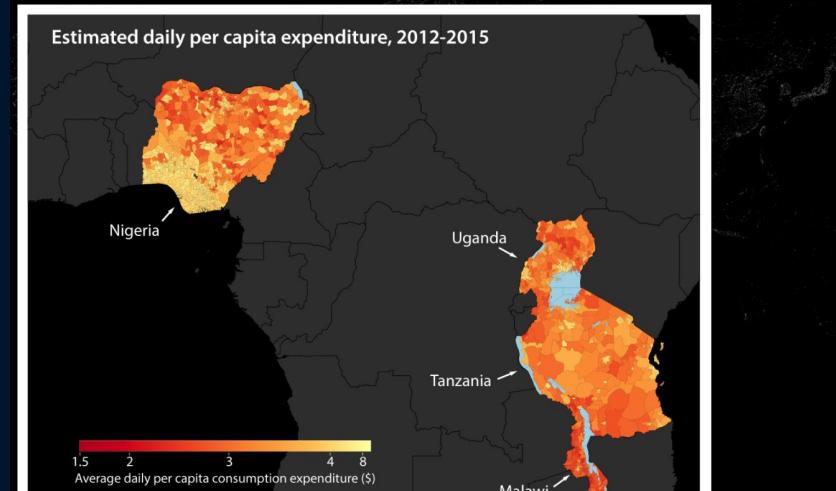


Reference:

<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00373-y>

FINE-GRAINED MAPS OF POVERTY

Using the final model that has been trained on survey data, we can estimate per capita consumption expenditure for any location where we have daytime satellite imagery. This allows us to create fine-grained maps of poverty—here are some that we put together for Nigeria, Uganda, Tanzania, and Malawi.



Reference:

<http://sustain.stanford.edu/predicting-poverty>



UN Hackathon Platform



BigDataUN Regional Hub
In China



BigDataUN Regional Hub
In Rwanda

BigDataUN Regional Hub
In Brazil

Rules of participation

The use of private and/or copyrighted datasets is not allowed for any team.

By submitting their proposal, contestants declare that the content submitted is their original work and creation and has not been presented previously in other events.



Certificate of participation

All teams that submit the expected outcomes will receive a certificate of participation.

Additional mentions will be provided for outstanding solutions of each region (Americas, Europe, MENA, Africa, and Asia).



Category of prizes

Most promising team: Outstanding solution from beginners teams that don't have access to Jupyter notebook.

Key contributor: Most active team in the communication channels that provides help to other teams' questions and shares additional data sets or insights.

Best visualization: Outstanding presentation of data in a visually compelling format.

All-student team: Awarded to outstanding team consisting of all students.

Teen award: Outstanding solution developed by teams consisting in majority of participants under 20.

Best presentation: Based on the presentation skills of the team.



Grand prize

The overall winning team will receive funding to attend the Dubai Expo 2020 and have the opportunity to present their project to obtain potential feedback from international experts.

In addition, team members will be invited to participate in a Youth Circle organized by the UAE Minister of State for Youth to boost youth engagement in data science topics.



Jury selection

Team of 5 jury members

1. Select top 30% teams
2. Judges grade all of the selected teams and standardize grades

Criterion	Description	Points
Theme	Does the team develop a solution that is in line with the event's theme?	2 pts - Completely in line with theme 1 pt - Somewhat in line with theme 0 pt - Theme not respected
Innovation	Is the team's solution groundbreaking? Is the technology behind the idea impressive?	2 pts - Brand new idea 1 pt - Implementing differently an existing solution 0 pt - Copying an existing solution (plagiarism)
Usefulness	Can the team's idea help address one of the SDGs? Does it fulfill a real need people have?	2 pts - Huge potential to contribute to SDGs 1 pt - Interesting idea but difficult to implement 0 pt - Idea already implemented, impossible to implement or not of high interest
Presentation & teamwork	How well was the project presented? Does it make the idea more appealing? Does it show true teamwork and collaboration?	2 pts - Thorough presentation showing thought process of the team and pitching the idea 1 pt - Good presentation but lacking idea's thought process or pitch 0 pt - Presentation lacking to show thought process or interest of the idea
Creativity	Did the team try to learn something new? What methods or products did the team use for the first time?	2 pts - It shows that the team worked really hard and learnt about new methods or products (even through roadblocks) 1 pt - It shows that the team worked and learnt moderately 0 pt - The team didn't work/learn at all

Platforms & tools

SLACK

- Find the channel “**questions**” to seek technical support* related to platform access or IT issues.
- Find the channel “**data**” to share useful information with the rest of the participants.
- Teams can create a private channel to have a collaborative work.

UN Youth Hackathon Platform - CKAN

- Platform provided to facilitate the access to the databases.
- Teams are encouraged to share other relevant data sets (must be open-source and publicly available) through the **Slack channel “data”**.

**Teams can support each other through Slack at any time. The UNSD and Regional Hub support teams will be available between 8:00 AM and 6:00 PM (18:00) for each respective time zone.*



Proposal submission

Teams should submit the expected outcomes through a file sharing link (such as Google Drive, Dropbox, or other alternative) via email to:

un-youth-hackathon@unmgcy.org

Deadline*: 6 PM (18:00) Standard Pacific Time (PST) on December 8th.

Please check your respective timezone in the following page.

**Proposals submitted after the deadline will not be considered.*



Submission time

NEW YORK, USA - EST, Eastern Standard Time (US)
09:00 pm Wednesday, Dec 8 2021

MEXICO CITY, Mexico
08:00 pm Wednesday, Dec 8 2021

RIO DE JANEIRO, Brazil
11:00 pm Wednesday, Dec 8 2021

PST/PDT, Pacific Standard Time (US)
06:00 pm Wednesday, Dec 8 2021

PARIS, France
3:00 am Thursday, Dec 9 2021

COTONOU, Benin
3:00 am Thursday, Dec 9 2021

CAIRO, Egypt
4:00 am Thursday, Dec 9 2021

KIGALI, Rwanda
4:00 am Thursday, Dec 9 2021

ABU DHABI, United Arab Emirates
6:00 am Thursday, Dec 9 2021

IST, India Standard Time
07:30 am Thursday, Dec 9 2021

JAKARTA, Indonesia
09:00 am Thursday, Dec 9 2021

SHANGHAI, China
10:00 am Thursday, Dec 9 2021

*Time zones expressed in 12-hour format with am (morning) and pm (afternoon)



Action items

1. Read this handbook carefully.
2. Join the Slack workspace [here](#).
3. If selected, join the Jupyter workspace [here](#).
4. In case you missed it, review the webinar recording [here](#).
5. Get in touch with your team to define your working strategy.
 - a. Create a communication channel (ex. Slack, Whatsapp)
 - b. Choose the right workspace (ex. Google collab, drive, UN platform...)
 - c. Plan the 3 days ahead of time (group meetings...)
 - d. Brainstorm on an idea based on the theme
 - e. Set goals

Remember that the **official kick-off meeting** will take place on **December 6th at 7:30 AM China Standard Time (UCT+8)**.

The datasets will be available for all teams starting from **December 6th at 8:00 AM China Standard Time (UCT+8)**.



Last advice

1. Have a good communication strategy with your team!
2. Start populating the slides as you go; don't leave it for the last day!
3. Manage your time carefully: 3 days is a short amount of time
4. Divide and conquer: each team member should work on different tasks and sync regularly so you can get the best results
5. If you're stuck, don't stop: reach out to us on SLACK or by email! Finishing is better than abandoning :)



Q&A

Do you have additional questions?

un-youth-hackathon@unmgcy.org



UNMGCY

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.