



UN
Big Data
Hackathon

WEBINAR

Introduction to the 2022
UN Big Data Hackathon
(UNBDH)

The 2021 UN Youth Hackathon in numbers...

Theme: the impact on SDGs by the COVID-19 pandemic (focus on SDGs 1, 2, 3, 4, 5, 8)



3 days



33 countries



80 teams



300+ participants



In words...

INNOVATE...

to participate in the achievement of the SDGs

NETWORK...

with data-enthusiastic students and professionals

EXPOSE...

yourself to state of the art data analysis tools and methods

COMPETE...

with leading data scientists across the globe

LEARN...

from your peers and strengthen your teamwork



UNBDH Organizing Team



During this hackathon, the [United Nations Statistics Division](#), [Badan Pusat Statistik](#), [UNCTAD](#), [Major Group for Children and Youth](#), [UN Global Pulse](#), [UN Global Platform Regional Hubs for Big Data \(UAE, Brazil, Africa and China\)](#), [Office for National Statistics](#), [Statistics Canada](#), [Islamic Development Bank \(IsDB\)](#) and [Asian Development Bank \(ADB\)](#) will be on the lookout for **disruptive** and **innovative** solutions that use data, machine learning and artificial intelligence to fast track progress towards the **UN Sustainable Development Goals (SDGs)**.



OPENING STATEMENT



Ronald Jansen

Assistant Director, Chief of Data Innovation and Capacity Branch
@ United Nations Statistics Division/United Nations Department of
Economic and Social Affairs



United Nations - UNSD, UN DESA, UN Global Pulse & UNCTAD



Ronald Jansen



Jiayue Zeng



Markie Muryawan



Clarence Lio



Karoly Kovacs



Bogdan Dragovic



Sean Lovell



Daniel Hopp

BPS Indonesia



Muchammad Romzi



Setia Pramana



Efliza



Maulana Faris



Choerul Afifanto



Ana Fitriyani



M. Yunus Hendrawan



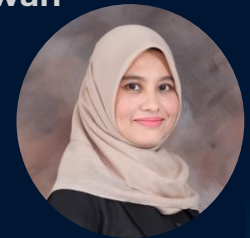
Rifqi Aulan Nisa



Silvia Arini



Kiki Rizki Amalia



Nori Wilantika

Badan Pusat Statistik Indonesia



Muchammad Romzi

Director of Analysis and Statistics Development
@ Statistics Indonesia



Big Data in BPS-Statistics Indonesia



Technology Adoption

BPS works to adapt to changes in technology and access to information by creating Big Data ecosystems

Official Statistics



Big Data

We utilize Big Data as effectively as possible as one of the potential sources of Official Statistics



Web Crawling/ Web Scrapping



Satellite Image/ Space Technology



Mobile Positioning Data and Other Index

Come and Join Us in Yogyakarta



About
Event



Indonesia,
Yogyakarta



UN Big Data
Hackathon



8-11 November
2022



Students or young
professionals



The winning teams will present their work at the UN World Data Forum in Hangzhou, China in 2023 and receive certificates and rewards.



FINAL EVENT



Seruni
Beach



Prambanan
Temple



Embung
Tambakboyo



Big Data in BPS-Statistics Indonesia



WEB-CRAWLING



Mobile Positioning Data (MPD)



- ✓ Tourism statistics
- ✓ Metropolitan Area Statistics
- ✓ Commute statistics



Satellite Image



- ✓ Poverty mapping
- ✓ Economy activity
- ✓ Agricultural statistics



Relative Index Facebook

- ✓ Economy activity
- ✓ Poverty mapping



Google and Facebook Mobility Index

- ✓ People mobility

	Resource		Result
	Marketplace	➡	E-commerce data
	Flight tracker, transportation booking websites	➡	Transportation data
	Job search websites	➡	Labour statistics
	Hotel booking websites	➡	Tourism statistics
	Air quality websites	➡	Environment and disaster statistics
	Mobil123, rumah123	➡	Housing and transportation statistics
	Online news and social media	➡	Phenomenon and population image
	Google Map	➡	Society activity and mobility
	IDx	➡	Financial report and shares



MGCY



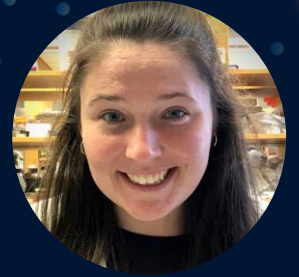
Oumaima Makhoulouk



Abhijith Mathew



Camila Ramirez



Tori Lovins



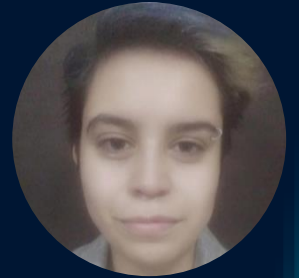
Ona Ambrozaite



Rufai Balogun



Erica Eggleton



Alinne Olvera

Major Group for Children and Youth (MGCY)



Ona Ambrozaite

Ph.D. Candidate @ **John Hopkins University**
& Science-Policy Interface Co-Coordinator @**MGCY**



UN Global Platform Regional Hubs for Big Data



Rene Wearing
(UAE)



Ruaa Alshehhi
(UAE)



Aisha Khalid
(UAE)



Bryan Gao
(China)



Patricia Tavares
(Brazil)



**Andrea Diniz
da Silva** (Brazil)



Therese Uwimana
(Rwanda)



Samuel Nahimana
(Rwanda)

UN Global Platform Regional Hub for Big Data **in the UAE**



Her Excellency Hanan Ahli

Acting Director

@ Federal Competitiveness & Statistics Authority



Watch the video again [here](#)



UN Global Platform Regional Hub for Big Data **in China**



Bryan Gao

Staff of Division

@ UN Global Platform Regional Center Management Office in China



UN Global Platform Regional Hub for Big Data **in Brazil**



Andrea Diniz da Silva

Professor

@ National School of Statistical Sciences – ENCE/IBGE



UN Global Platform Regional Hub for Big Data **in Africa**



Therese Uwimana

Director of Data Revolution and Big Data
@ National Institute of Statistics of Rwanda



ONS UK, Statistics Canada, ADB & IsDB



Eric Deeben
(ONS UK)



Alison Baily
(ONS UK)



Cherryl Chico



Eduardo Valdez



David Evans
(Stats Canada)



Stephanie Goetz
(Stats Canada)



Kathryn Stevenson
(Stats Canada)

The Office for National Statistics (ONS UK)



Eric Deeben

Technical International Programme Lead
@ ONS UK



Statistics Canada



David Evans

Innovation Leader and Strategist
@ Statistics Canada



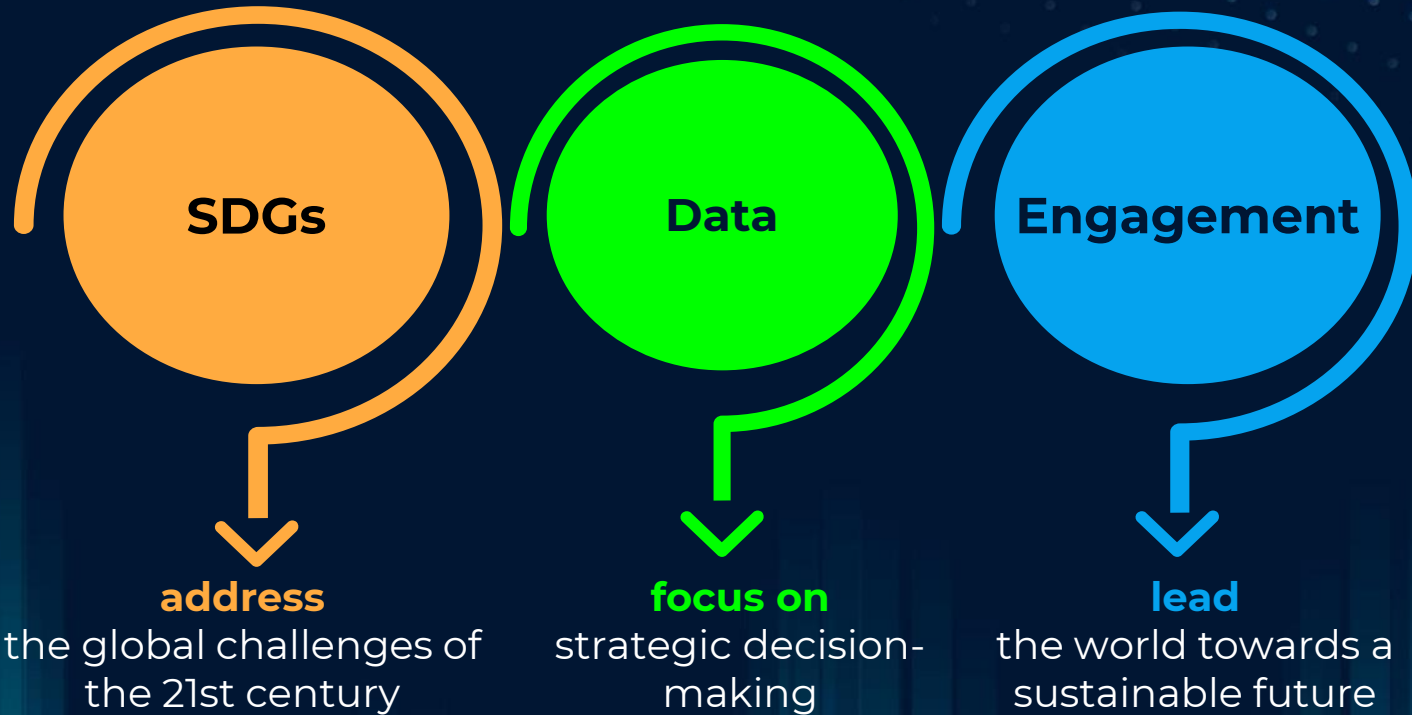
Objectives

The hackathon aims at:

- Providing a space for data-enthusiastic 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.
- 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
- Supporting youth engagement in evidence-based decision making through the analysis of data.



What brings us here?



UNBDH Tracks

Big Data Experts Track

Refer to the 2020 AIS Hackathon

- Government officials, professionals, and researchers with a background in data science, statistics, economics, or social sciences
- **Requires the usage of data coming from the Automatic Identification System (AIS) in line with the theme**
- Requires familiarity with data processing using pyspark, geospatial indexing techniques, big data analytics and modeling or statistics
- Participants will form **20 teams of 3-5 people each team**. An ideal grouping of teams would include data scientists, statisticians, economists, and social scientists

Youth Track

Refer to the 2021 UN Youth Hackathon

- Data enthusiasts, young professionals, or students that are **32 years old or younger**
- **All levels in data science are welcome!**
- Participants will form ~100 teams of 3-5 people each team. An ideal grouping of teams would include participants from different professional and academic backgrounds.



UNBDH Logistics

	Big Data Experts Track (20 teams)	Youth Track (100+ teams)
Logistics	On-site in Yogyakarta, Indonesia OR Virtually	Virtually OR Satellite locations (for now, Yogyakarta, Dubai and Rio de Janeiro) if you live close by
Technology	<ol style="list-style-type: none">1. Analytics tools: Online Jupyter Hub (Python & R), locally-installed tools (Python, R, PowerBI, Tableau, SAS...)2. Code collaboration platform + submission of outputs: GitHub3. Communication: Slack & emails	
Datasets	AIS datasets Trade datasets Datasets specifically related to the theme	Datasets specifically related to the theme



Join the [mailing list](#) to be updated
(webinars, deadlines, winners!)

UNBDH Timeline



- Access to the databases will be granted starting from **November 8th at 2:00 PM Yogyakarta Time (UTC+7)**.
- Proposals must be submitted on **November 11th at 12:00 PM Yogyakarta Time (UTC+7)**.



Expected outcomes

Teams will hack their way through **an innovative analytical solution** to address the [Global challenges](#) and help achieve the [Sustainable Development Goals](#) following a theme that will be revealed closer to the UNBDH.

For both tracks, teams must submit:

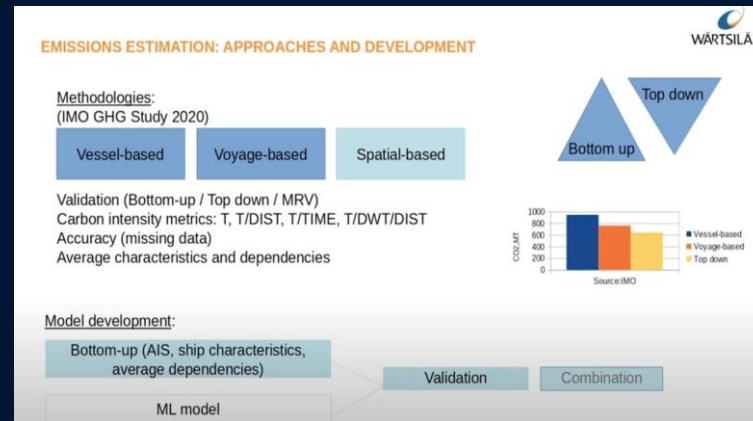
1. Presentation explaining their solution (free format)
2. Video with maximum length of 10 minutes = voice-over of the presentation
3. Coding scripts

Big Data Experts Track

Examples of outputs

Example 1: BlueCarbon, overall winning team of the 2020 AIS Hackathon

Using the AIS data provided by the UN Global Pulse platform, the team geographically distributed CO2 emissions from shipping based on individual vessel locations and activity using a Machine Learning Model. They subsequently developed an interactive dashboard to map these emission distributions for different time periods.



Reference: BlueCarbon's presentation video

<https://www.youtube.com/watch?v=qmybyzV5R8A>

Big Data Experts Track

Examples of outputs

Example 2: DogCat, student winning team of the 2020 AIS Hackathon

Using the AIS data provided by the UN Global Pulse platform, the team analyzed the impact of COVID-19 on three major sectors: commodities, bulk carriers and trade countries

Topic : The difference in the impact of COVID-19 to container shipping

Motivation

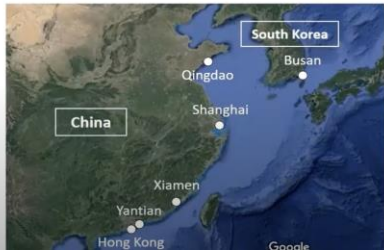
Is there any difference in the impact of COVID-19?
e.g.) Deployed vessels becomes smaller by the low demand

Objective

To measure the difference in the impact of COVID-19 to several factors

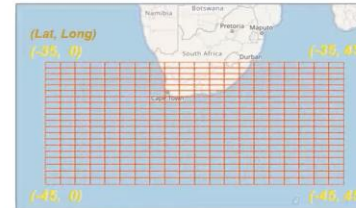
Target

Data : Arrival data on AIS data
Ship : Container ship
Port : Busan, Qingdao, Yantian, Shanghai, Xiamen, Hong Kong, Los Angeles
Term : 2019/01/08 - 04/23
2020/01/08 - 04/22
Factor : Shipping Alliance and Vessel size



Research Policy

- ① Set up a polygon area around the Cape of Good Hope and generate meshes in the area. (20 × 20 split)



- ② Extraction of speed and position data for each vessel in the area.
- ③ Using AIS data and other ship movement data, GHG emissions are calculated for each mesh and individual ship and the total values are tabulated.

Reference: DogCat's presentation video

https://www.youtube.com/watch?v=X_jmNy4qbjE



Youth Track

Examples of outputs

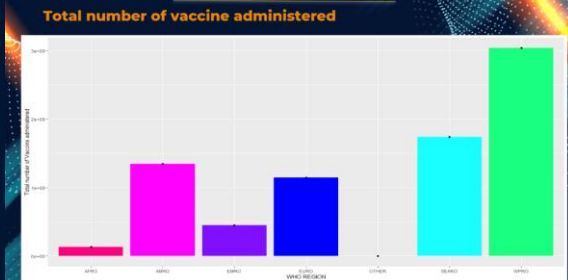
Example 1:

Reports providing detailed statistical analysis or data correlations between different datasets

Data Guardians



SOME EXPLORATION RESULTS



UN Youth Hackathon projects (sample):
<https://drive.google.com/drive/folders/1MxLU0NvDnD224fVHPcZKUatx-s3OJ-T7>

CAS-ZJSO

Section III

Most of the industries were shutdown to lower mobility. There have been negative economic implications due to restrictions and decreased travel business worldwide.

The negative economic implications may increase inequality, discrimination and medium and long-term unemployment if not properly addressed by right policies.

Most countries are currently trying to combat the virus spread by screening for COVID-19 in large numbers and maintaining strict quarantines, entry bans policies, as well as other limitations for citizens in or recent travelers to several countries in the most affected areas.

Unemployment rate yearly by global and regions

As a key performance indicator of economic health and a crucial metric for judging policy outcomes, unemployment rate typically calculated on a monthly, quarterly or annual basis.

The ILO provides unemployment rate data covering most of countries and areas until 2021. After the 2030 agenda for sustainable development was adopted by world leaders in 2015, the global yearly unemployment rate constantly decrease to 5.4% in 2019.

The yearly unemployment rate in G20 and G7 countries decreased by 0.6% and 3.3% respectively from the peak in 2011 to 5.5% and 4.2% respectively in 2019. At the SDG summit held in September 2019, SDG Member States renewed international commitment to fulfil the 2030 agenda for sustainable development.

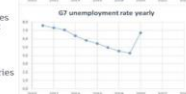
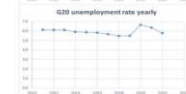
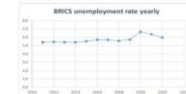
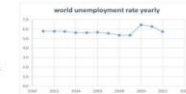
Yet, in only of brief period of time, the COVID-19 has disrupted efforts to achieve the 2030 agenda for sustainable development.

The COVID-19 pandemic has upended the global labor market, with massive job losses and spike in unemployment to its highest level.

The yearly global unemployment rate increased by 1.1% to 6.5% in 2020. The dramatic increase in BRICS, G20, G7 countries by 1.0%, 1.2% and 2.4%, respectively to 6.4%, 6.7% and 6.7%.

Thanks to strong policy support, accelerated vaccine rollout and the application of digital solutions to maintain business and consumption, growth resumed forcefully in the world in 2021.

The unemployment rate dropping to 5.5% showed a continued recovery from 2020 to 2021 after the COVID-19 pandemic. The BRICS and G20 countries also have the same economic recovery trends.



Youth Track

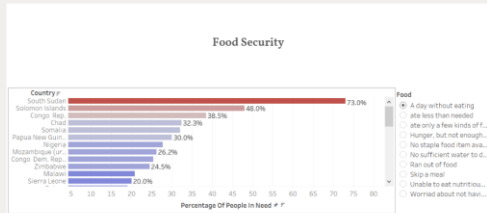
Examples of outputs

Example 2: Interactive dashboard with enhanced analytical visuals

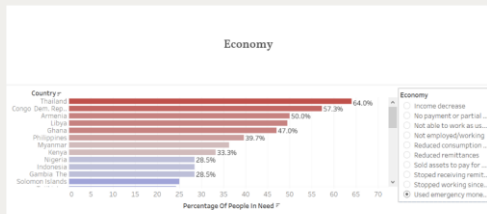
UN Youth Hackathon projects (sample):
<https://drive.google.com/drive/folders/1MxLU0NvDnD224fVHPcZKUatx-s3OJ-T7>

Team IdeaX

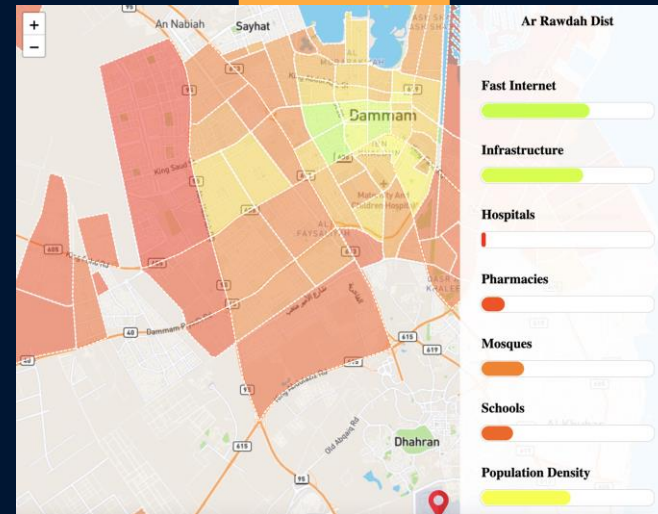
visualization



visualization



Doves



Youth Track

Examples of outputs

Example 3: Advanced Machine Learning Models for predictions

UN Youth Hackathon projects (sample):
<https://drive.google.com/drive/folders/1MxLU0NvDnD224fVHPcZKUatx-s3OJ-T7>

Team Sustainability, overall winner

Data Rockstars

Conclusion: Our solution

A statistical segmentation to better understand the impact of a household socio-economic characteristics on their vulnerability to COVID-19 and their consequences.

A integrated prediction model in order to assess the vulnerability of households to COVID-19 regarding their income, food security and education access

Pandemic Score Methodology Overview

- 1. Researching Data**
 - Look for data specifically for the pandemic period.
 - Education Survey
 - Time spent at home
- 2. Preprocessing Data**
 - Survey data, questions and answers were converted into indicators and flags.
- 3. Creating Factors**
 - Each factor is built upon some metrics derived from the available variables. The final factor is an equally weighted average over these metrics scaled to the maximum points attributed to the main category.

What does the score components measure?

Online Learning <ul style="list-style-type: none"> Adherence Effectiveness assessment Inclusion Support to teachers Engagement 	Learning <ul style="list-style-type: none"> Adjustments to curriculum and calendar Learning gaps Learning losses 	School Closures <ul style="list-style-type: none"> Total time of full closure Current state of full closures 	School Reopening <ul style="list-style-type: none"> Adherence Strategies Support to staff
Vulnerable groups <ul style="list-style-type: none"> Special support for distance learning and reopening 	Tests and Vaccines <ul style="list-style-type: none"> Rate of people fully vaccinated Rate of tests Rate of vaccination 	Mobility <ul style="list-style-type: none"> Relative change of time spent at home relative to the pre-pandemic period 	Forecasting <ul style="list-style-type: none"> Relative change in new cases and deaths (forecasted) relative to the previous period (2 weeks window)

Forecasting Methodology

To predict COVID-19 new cases and deaths, we applied an additive model with non-linear trends fit daily. This approach is robust to missing data and shifts in the trend.

Algorithm: Prophet
 Y: Smoothed new cases and deaths per million
 Training data: JAN/2020 - OCT/2021
 Validation: NOV/2021
 Out of time prediction: JAN/2022
 Mean Squared Error: 8,65

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.



Special Prizes & Certificates

(more to come, stay tuned...)

In each track:

- **Outstanding team:** This prize would be granted to the overall best solution.
- **Best team of the region:** Best project from a region (Africa, Europe, North America, Latin America, MENA, Asia, Oceania)
- **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.
- **Best presentation:** Based on the presentation skills of the team.

Youth track only:

- **Most promising team:** Outstanding solution from beginners teams
- **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.

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



Prizes

Grand Prize:

- The Outstanding teams will present their work at the **UN World Data Forum in Hangzhou, China in 2023**. This will be give them the opportunity to present their project to obtain potential feedback from international experts.

Other Prizes

- Granting full access to the UNGP AIS platform for one year.
- Further training, mentoring, and assistance from UK Data Science Campus for a specific project in 3-6 months.



If you want to participate or receive all the updates:

1. **Register** for the hackathon with your team or individually:
<https://input.un.org/EFM/se/3995D1A40525E4CE>
2. Read about **all the information** about the hackathon on the **website**:
<https://unstats.un.org/bigdata/events/2022/hackathon/>
3. Join the **mailing list** to hear about the upcoming webinars, registration deadline & logistics: <https://input.un.org/EFM/se/3995D1A46342E490>

Q&A

Do you have additional questions?

un-big-data-hackathon@unmgcy.org

Follow us on:



[@unbigdatahackathon](https://twitter.com/unbigdatahackathon)



[@unbigdatahack](https://www.instagram.com/unbigdatahack)



[@unbigdatahackathon](https://www.facebook.com/unbigdatahackathon)

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

