Philippine Datathon 2023 Contest Mechanics:

1. **Registered Participants.** The contest is open to all registered participants of the Philippine Datathon 2023 Conference who are above 18 years old, except participants who are conference mentors/speakers/lecturers or dataset lead mentors/domain expert.

2. **Goals.** The overall goal of joining the contest is to have fun. The second goal is to learn how to work as a research team. Another goal is to appreciate what others with different roles in an interdisciplinary data science team contribute to the overall output of the team.

3. **Choosing a Dataset Team.** Upon registration, participants who indicate in their registration form that they want to join the contest will be asked what Dataset Team they want to join. Participants from Scoliosis Philippines, Neural Mechanics, and CARE PH cannot join a team that is interrogating their own dataset. The following are the Dataset Teams of this year's Datathon Contest:

   - Scoliosis Philippines Dataset on Scoliosis Registry

     - The Scoliosis Philippines Dataset covers patient demographics, types of scoliosis, treatment and psychosocial needs and other conditions associated with the condition. The data were generated from their mobile app designed as the first centralized database in the country.

   - Neural Mechanics Dataset on Dengue Cases

     - PH Dataset on the disease, diagnosis, and test results.

   - CARE PH Dataset on COVID Pneumonia

     - Covid-19 pneumonia contains clinical data from over 3000 COVID-19 RT PCR +/- persons admitted at the ER or wards of The Medical City during the 1st and 2nd years of the pandemic.

   - CARE PH Dataset on Hepatitis and Hepatocellular Carcinoma

     - Hepatitis contains clinical and liver ultrasound data from over 750 persons with and without hepatitis B, with and without liver cancer, that are associated with the development of liver cancer.

- CARE PH Dataset on Hospital-Based Cancer Registry

  - Cancer registry contains demographic and baseline clinical data and pathology reports of over 60,000 individuals newly diagnosed with cancer from 28 hospitals in majority of the regions in the country.

4. **Number of Team Members and Mix.** Each team will have a maximum of six (6) members with varying expertise or chosen role. The organizers will do their best to maintain this number and the mix as well.

5. **Choosing a Role in the Dataset Team.** Upon registration, participants who indicate in their registration form that they want to join the contest will be asked what role they want to have in the team to which they belong. Examples of such roles are:

   - Patient representative
   - Clinical or domain expert
   - Data scientist
   - Statistician
   - AI or machine learning scientist
   - Epidemiologist
   - Public Health specialist
   - Social Scientist

6. **Opening of the Datasets to each Team.** The corresponding datasets will be opened to each team on the afternoon of Day 1 and will close as soon as the team presentations start on Day 3.

7. **Lead Mentor**. Each dataset team will be assigned a Lead Mentor who is familiar with the dataset they are working on – why, how, when it was created, etc. The Lead mentor is there to introduce the background of the dataset and answer questions. **The Lead Mentor is not a participating member of the team**. The Lead Mentor must be present and available to their teams throughout the Datathon conference and contest, but not to referee.

8. **Team Leader**. The teams, when they first meet, are expected to name **a Team Leader – who may also be the team's presenter – but not the Lead Mentor**. There will be time allotted for teams to meet up on the afternoon of Day 1, and on Day 2. Teams can continue to work and discuss before or after the Datathon conference hours.

9. **Dataset Questions**. The Lead Mentor assigned per dataset shall formulate questions, pertinent to the dataset, that will be used in the Datathon to be answered by members of his/her group. The following are general questions that serve as guidelines to come up with a common set of questions. The judges may expound on the questions as they see fit when they meet the participants.

- Given an initial analysis / look at the data, what set of hypotheses can we derive initially from the data sets?

- Are the datasets complete to generate such a hypothesis, and eventual conclusion? If they are not complete what are the gaps? If there are gaps, what does the team plan to do to address such gaps?

- What algorithms / techniques does the team plan to do to

  a) clean the data,

  b) validate the completeness and timeliness of the data,

  c) do their analysis.

- What key conclusions / findings were made from the data sets? Please support these with graphs, visualization, necessary analysis that were derived from the data set.

10. **Order of Team Presentations.** No later than the end of Day 2, the order of team presentations will have been determined by random picking of numbers.

11. **Team Presentations.** This will be on Day 3 during the Datathon Contest Presentations, and within a strictly monitored twelve (12) minute time allotment, the minimum requirements for each team are as follows:

- Research question
- Methods used
- Analytical tool outputs
- Key discussion points
- Conclusions

To add even more value to the presentation, each team is encouraged to focus on the gaps and issues found with their respective datasets and their proposed solutions. These datasets are real life data, so the proposed solutions must likewise be realistic and implementable, leading to real-world benefits.

12. **Panel of Judges and questioning.** There will be five (5) judges who will be announced at the start of the contest. At the end of each team presentation, each judge is allotted a maximum of five (5) minutes to query the team.

13. **Judging.** At the end of all presentations, the panel of judges are to decide based on the rubric below, will deliberate among themselves, and then arrive at a consensus for the top three team presentations. The lead judge has the role of a tie breaker.

   - Clarity and simplicity: 30%
   - Appropriateness of methods used: 30%
   - Innovation or creativity: 30%
   - Time management: 10%

   Judges' decisions are final and cannot be appealed.

14. **Prizes**.
   - 1st prize: 50,000.00
   - 2nd prize: 30,000.00
   - 3rd prize: 20,000.00
   - Consolation prize: non-sponsored registrants who joined the contest but did not win any of the top 3 prizes will be given refunds of their registration fees by CARE PH, preferably through GCash within 24 hours after the Datathon ends. Sponsored participants will not be refunded the registration fees paid by their sponsors.

15. **Publication of output.** Teams that wish to publish their output may do so if given permission by their Lead Mentor and the Organizing Committee members. The Lead mentor and dataset owners may be acknowledged in the final manuscript or be added as a last author. Publication fees will be at the expense of the team's own funding; will not be funded by the Datathon organizers.