



Final Competition Guidelines

Competition Overview

The University of Calgary is hosting a competition to give its students an opportunity to showcase their skills in programming and data science. This competition is open to students enrolled in the university, and it provides a platform for them to demonstrate their abilities to the wider community. The competition organizers have provided a comprehensive dataset that the participants will work with.

The objective of the competition is for the students to perform a variety of data science tasks, including visualization, manipulation, and analysis. This will require the students to use their programming and data analysis skills to uncover insights and trends within the data. The participants will have the opportunity to apply the concepts and techniques they have learned in the classroom to real-world data, and to demonstrate their ability to work with large datasets.

In addition to demonstrating their skills, the competition provides an opportunity for students to gain recognition for their accomplishments. This can be valuable for students who are looking to build their portfolios and further their careers in programming and data science.

Overall, the competition aims to foster a sense of community among University of Calgary students and to provide an avenue for them to demonstrate their skills to the world. By participating in this competition, students can gain valuable experience and make connections that can help them achieve their career goals.

In-Person Symposium

The culmination of the competition is an in-person symposium event on March 27th, 2024 (location and time TBD). Attending this event is highly encouraged, as it allows for you to showcase your submission with other competition attendees, as well to network with other

individuals within the data science community at the university, and professionals in the industry! In the past, competitors and attendees of the in-person symposium gained the opportunity to network with industry professionals, and received full time offers, internships, as well as research opportunities!

***Winners for each category will be announced at the symposium. We hope to see you all there!**

Competition Categories

The competition mentioned is structured into three distinct categories, allowing participants to compete in the area that best showcases their skills and interests. These categories are:

- **Data Visualization:** This category focuses on the visual representation of data and aims to test the participants' ability to create compelling and informative visualizations that communicate insights and patterns in the data. Participants will be expected to use their knowledge of data visualization techniques and tools to produce high-quality visualizations that effectively communicate their findings.
- **Machine Learning Approach:** This category focuses on the application of machine learning algorithms to the data. Participants will be expected to demonstrate their ability to clean and pre-process data, select appropriate algorithms, and evaluate their results. They will also be expected to use their skills in programming and data analysis to implement and apply machine learning techniques to the dataset provided.
- **Research Proposal Approach:** This category focuses on the formulation and presentation of a research proposal based on the data provided. Participants will be expected to use their skills in data analysis and critical thinking to identify important questions and trends in the data, and to propose a research question(s). They will also be expected to present their findings in a clear and concise manner, demonstrating their ability to communicate complex ideas effectively.

Each category offers its own unique challenges and opportunities for participants to showcase their skills and knowledge. By competing in one of these categories, participants can demonstrate their expertise in a particular area and gain recognition for their

achievements. The competition provides a valuable learning opportunity for students who are interested in advancing their careers in data science and related fields.

Submission Criteria

Document Submission

1. Submission deadline will be March 22nd at 11:59 PM MST (Midnight). Any submissions past this deadline will not be considered for the competition.
2. Participants are required to submit the following documents
 - a. Source code as .py or other appropriate format based on the chosen programming language. If participants choose to use external softwares such as Excel or Tableau for the data visualization category, then they must submit all relevant files.
 - i. If an external software is used, please submit a one-page document indicating the work process undertaken to design the visualizations. Refer to the section specific for 'Data Visualization Guidelines'.
 - b. Document containing a 16:9 poster presentation with all the figures and relevant notes

Relevant Participant Information

1. Participants are allowed to work alongside another partner. This must be declared when signing up for the competition. Failure to declare this may result in disqualification.
2. The dataset used for this competition has been retrieved from 'Our World In Data'. You cannot recreate the figures or charts present in the 'Our World In Data' website. All figures generated must be unique, and any references to the 'Our World In Data' website must be cited.
3. Participants are allowed to use up to two (2) additional open source datasets to supplement the provided data. These datasets must be open access, and should reflect the chosen submission category. Please reference the chosen datasets.
4. The supplementary data must aid the provided data, and must not be the main aim of study. Judges may decide that use of supplementary data was in excess, which can result in disqualification. Please refer to the example section in this document for input on how to use additional data.