

The first national exploratory data analysis (EDA)  
competition returns



# 2022 PJDS

Philippine Junior Data Science Challenge

EVENT PRIMER

Brought to you by



UP DATA SCIENCE  
SOCIETY  
Powered by Data-Driven People



GCash

Corporate partners

Special thanks to



In cooperation with



Organization partners

Special thanks to



# GREETINGS!

The UP Data Science Society invites you to the second installment of its **Exploratory Data Analysis (EDA) competition**, the Philippine Junior Data Science Challenge. As the pioneering organization dedicated to building a community of Data Science enthusiasts and aspiring professionals in UP Diliman, **we offer this event to undergraduates all over the Philippines** in order to provide an exciting way for students to apply the discipline in a competitive setting. This competition allows aspiring Data Scientists of the Philippines a chance to work with and against themselves to cultivate their holistic statistical thinking. With **cash prizes** and potential **internship opportunities** awaiting participants!

We highly encourage your registration and we look forward to all the useful insights and analyses that you have to offer.

Best regards,

**Joaquin Eala**

PJDSC 2022 Head director



UP Data Science Society



UP Data Science Society

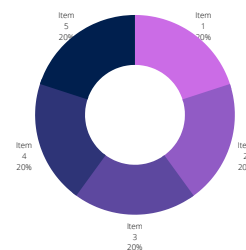
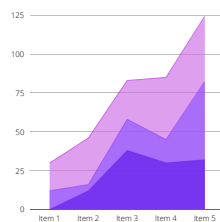
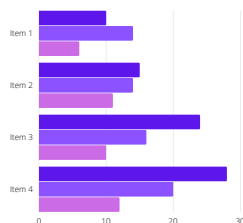


updatasciencesociety@gmail.com

# What is an EDA?

**Exploratory Data Analysis or EDA** is a method of analyzing a given data set in order to extract its primary attributes as well as to provide deeper insight on the factors which may have caused certain trends in the data.

**An EDA is conducted using data visualization techniques such as statistical graphs.** This may include box plots, bar graphs, histograms, and scatter plots. These graphs are created in order to summarize the qualities of the data set and to **present these qualities in a way in which people can easily understand the information embedded in the data.**



Aside from data visualization, **descriptive statistics is also used to help make conclusions regarding the given data.** Techniques such as taking the measures of central tendency as well as the measures of variability are crucial in any good EDA.



# MECHANICS

## Registration

- A. The competition is **open to all undergraduate college students** whether enrolled or not for the duration of the competition
- ✗ Individuals who possess a **4 to 6 year baccalaureate degree may NOT** join the competition
  - ✗ Individuals who have **not started a baccalaureate degree may NOT** join the competition
- B. Teams will be composed of a **minimum of two (2)** to a **maximum of four (4)** individuals
- C. A registered **team with less than 4 individuals may request for the organizers to match them to teams who also have less than 4 individuals**. These teams may be merged as long as they satisfy the team composition of 2-4 individuals
- D. Teams must **submit their registration** and other required documents **before the deadline** set by the organizers
- E. Organizers will select **thirty (30) teams** based on merits found on the teams' answers to questions in the registration form. Additionally, the composition of the final list of teams will be as follows, **based on the a majority of a team's campus location**:
- NCR: 4 teams | Luzon (non-NCR): 4 teams | Visayas: 4 teams**  
**Mindanao: 4 teams | Any location: 14 teams**
- F. Teams that will proceed to the next round of the competition will be **notified via email**

# MECHANICS

## Preliminary Round

- A. Teams will be **notified via email of the dataset to be used** throughout the competition
- B. **A single dataset** will be given to the participants
- C. The dataset will contain one or more tables
- D. Teams **must strictly adhere to the agreements** stipulated in the contract signed during the competition. If the organizers discover any breach of contract, the team will be **immediately disqualified** from the competition and **legal action** may be taken against them
- E. Teams **may use any analysis tool and resource available to them**. Additional data not part of the provided dataset can be used to supplement the analysis, as long as this data is publicly available (i.e. any third party has independent access to the data)
- F. Teams **must include** the following in their submission:
  - ✓ **3-slide infographic** detailing the insights and analysis of the chosen dataset. There must be exactly 3 slides of size A4 in PDF format
  - ✓ **1-2 page Technical Journal** outlining the thought process and methodology of the EDA

# MECHANICS

## Preliminary Round

- G. Judges will select the **top five (5) submissions** based on the following criteria:
- Quality of Insights (35%)**
  - Storyline Cohesion (30%)**
  - Communication of Insights (20%)**
  - Methodology (15%)**
- H. **Selected teams will be notified via email** if they will proceed further into the competition

## Workshops

- A. To guide teams in the Exploratory Data Analysis presentations, the organizers will provide **online seminar-workshops**
- B. These workshops are **only open to the top 30 teams** qualified to enter the preliminary round and will be administered via Zoom
- C. Attendance to these **workshops are optional** and are not a prerequisite to qualifying for the final round
- D. A **30-minute Q&A session** will be open to the participants where they can consult with the workshop speakers

# MECHANICS

## Final Round

- A. The final round will be **livestreamed to the public**
- B. Each team will give a **20-minute live oral presentation** about their submitted Exploratory Data Analysis
- C. Teams are **free to provide additional analysis** to supplement their initial submission from the preliminary round, provided that they do not stray too far from their previous insights, and follow the limitations stated in the preliminary round
- D. Teams will be given **15 minutes to answer questions** from the judges after their presentation

## Awarding & Post-event

- A. Judges will decide the winners of the first, second, and third prize and shall be **announced on the same day of the Final Round**
- B. The first, second, and third awards are as follows:

**1st Prize:** Php 15,000 cash + Internship\*

**2nd Prize:** Php 10,000 cash + Internship\*

**3rd Prize:** Php 5,000 cash + Internship\*

\*To be finalized and specified in a separate primer



# COMPETITION TIMELINE

**Release of EDA Dataset**

**May 12, 2022**

**May 13, 2021**

**Deadline of  
Registration**

**May 21, 2022**

**Seminar workshop**

**May 28, 2022**

**End of Preliminary  
Round**  
(Submission of infographics)

**June 3, 2022**

**Release of  
Top 5 Teams**

**June 11, 2022**

**FINAL ROUND**  
(Presentation of Top 5 Teams)



## FAQs

### **Is there a registration fee?**

There is no registration fee.

### **I am not from Computer Science, Engineering, or any course directly related to Data Science. Am I still allowed to register?**

The competition is open to all undergraduates from any university with no restrictions on degree program.

### **Can I register as an individual?**

Yes, you may register as an individual. The event team will be adding these individuals either into teams composed of other individuals or teams that are looking for more teammates. They will be placed into these teams according to their compatibility with the rest of the members

### **We are a team of 2 or 3 members, can we request for more members?**

Yes. There is an option in the registration to request for more members if your team signs up with only 2 or 3 members so far. However, this does not guarantee that your team will receive additional members should the organizers be unable to find them.

### **I'm an undergraduate but currently on LOA, can I still participate?**

Yes, you may still participate if you are an undergraduate that is currently on Leave of Absence.

## FAQs

### **Is this open to graduate students?**

No. This competition is open only for undergraduate students who have not yet completed a previous undergraduate degree.

### **Do I need to have any programming experience to join the competition?**

It is not a requirement, but it would help. We will hold a workshop with learning sessions to be able to widen your knowledge on data visualization but most of the programming aspect is left for you to learn.

### **What will the dataset be about?**

The specific details on the dataset will be disclosed to the participants on the scheduled date of release on May 12 (Thursday) as stated in the Competition Timeline.

### **What programming language are we allowed to use?**

You are allowed to use anything as long as you are able to successfully analyze and create an insightful exploratory data analysis.

### **Can our team have members from different universities and regions?**

Yes. You are free to team up with members from all over the Philippines as long as they meet the criteria for joining.

### **What if I just graduated college, can I still join this competition?**

No. Only students currently taking up their undergraduate program can join the competition.

# ABOUT THE ORGANIZATION

Founded in 2020, UP Data Science Society (UP DSSoc) is the **first data science organization in UP**. The organization aims to provide an environment for aspiring data scientists to engage in insightful data science discussions and to hone their skills, regardless of skill level, through workshops and other opportunities. UP DSS also envisions itself as a multidisciplinary organization that equips and **empowers members to use data science to produce data-driven solutions for social good**, aligning both internal and external events to contribute to the advancement of communities at large. The work and activities held in UP DSS revolve around the organization core values of **Purposeful Learning, Community for Innovation, and Utak at Pusò**.

## OUR VISION

A multidisciplinary organization whose members contribute to society through **innovative solutions** to community problems using Data Science, and help **responsibly develop the field of Data Science in the country**.

## OUR MISSION

**To provide an avenue for Data Science enthusiasts of UP Diliman to safely share ideas and experiences, hone their skills, and collaborate to build socially relevant and impactful projects while promoting data science and its real life applications.**



### Data Science Challenge

Team Registration  
(3-4 members per team)  
\*All applicants are required to join

November 12 to 18, 2021  
Link: <https://bit.ly/31ZneAd>



### General Mechanisms

For the Recognition of the  
Biweekly Top DataCamp Users



UP DATA SCIENCE  
SOCIETY  
Powered by Data Driven People



UP Data Science Society



UP Data Science Society



[updatasciencesociety@gmail.com](mailto:updatasciencesociety@gmail.com)