

# **2022 DataHacks Attendee Resource Guide**

Table of Contents:

- I. Introduction
- II. Communications
- III. Mentoring
- IV. Schedule
- V. Competition Tracks
- VI. Prizes
- VII. Workshops
- VIII. Submission

## I. Introduction

Welcome to DataHacks 2022! This document will tell you everything you need to know about how things will run. The original hope for this event was to have it in person; however due to concerns regarding the Omicron variant of COVID-19, it was decided that we would continue virtually. Despite this it is our hope that the connections you make and experiences you undergo during this competition will be just as tangible and meaningful as those you would have experienced in person.

DataHacks is an annual data science competition organized by DS3 (Data Science Student Society) at the University of California, San Diego. DS3 has organized DataHacks every year in hope of introducing more students to the exciting and innovative world of data science. Besides DataHacks, DS3 also organizes workshops and events regularly throughout the year. If you are interested, please check out and like the [DS3 page on Facebook](#) as well as the [DS3 discord](#).

## II. Communications - Discord and Zoom

- **Discord**

- A. Join the Discord using [this invite link](#) if you haven't already!
- B. Be sure to name your role! Ex: Name | Hacker
- C. If you do not have a team:
  1. Join the #team-building channel. You should be able to access this automatically but DM an organizer if you cannot! (Organizers have a role Organizer at the end)
  2. Post an introduction about yourself (Join the #self-introduction channel on the HACKERS subchannel), what projects you might be interested in and what skills you might have.

Ex: Hi everyone! My name is Wilson and I am a Data Science major at UC San Diego. I am really passionate about environmental conservation and AWS technologies, so I am interested in doing a project in the Sustainability category.. I have experience working with Python and Java, and would love to join a team. Feel free to DM me if you would like to work together!

3. Feel free to reach out to folks that post in that channel if your interests align!
- D. Please note that participants must follow the MLH code of conduct, which can be found [here](#).
- E. Discord is for:
  1. Basic communication
  2. Self introduction
  3. Team building
  4. Event update/reminder
  5. Logistic information

- **Zoom**

- F. For the duration of this competition we will be using Zoom to host our opening ceremony, awards ceremony, and workshops. For assistance using Zoom please refer to the [university's resource page](#).

### III. Mentoring

- Mentors will be available via Zoom in an office hours like format beginning at 3:30pm on Saturday and ending at 2:30am on Sunday, and then resuming again at 8:30am on Sunday through the end of the competition.
- If assistance is needed outside of these hours, check the #faq channel to see if your question has been asked previously, or use the #mentor-questions text channel to post questions; mentors will check this channel frequently to assist participants.
- Due to the larger number of participants and time constraints of the competition, only one member of a team may join an office hours at a time.

### IV. Schedule (PST)

<u>Date/Time</u>	<u>Workshop/Event</u>
<b>Saturday, April 2</b>	
3:00pm - 3:30pm	Event Kickoff
3:30pm	Hacking begins
7:00pm - 8:00pm	Workshop: Exploratory Data Analysis with Pandas (Beginner)
8:00pm - 9:00pm	Workshop: Intro to ML & NLP with Sklearn (Intermediate)
9:00pm - 10:00pm	Workshop: Deep Learning for Natural Language Processing (Advanced)
<b>Sunday, April 3</b>	
10:00am - 11:00am	Workshop: Final Submission
3:30 p.m.	Final Submissions and Coding Freeze
8:30 p.m. - 9:30 p.m.	Closing Ceremony and Awards

### V. Competition Tracks

#### Beginner Track: Workforce Salaries

- **Description:**
  - In economic research, one of the most highly studied topics is understanding the factors that determine an individual's salary. Given a dataset with many different features; how well can you predict an individual's salary?

### Intermediate Track: Shark Tank Pitches

- **Description:**
  - Shark Tank is a television show that allows prospective companies to pitch to wealthy investors. Given a dataset with the description of each pitch, can you predict whether a deal was made or not?

### Advanced Track: Analyzing Sentiment In Markets

- **Description:**
  - In financial markets, information is key. However, in the complex markets of the present, information is consumed by algorithms instantly. Can you gain an edge on the market through sentiment analysis?

#### Submission Guidelines:

- PDF Report: All visual + written content (including but not limited to infographics, visualizations, objective answers, subjective answers, justifications)
- Code: py file OR jupyter notebook OR GitHub Repo that includes all your code
- Prediction in csv file (**not required for beginner track**)
- Webpage/product (**optional, include link at beginning of the report**)

## VI. Prizes

- **Beginner Track**
  - Prize: Echo Dot (4th Gen, 2020 release)
  - Number of winners: 1 team (4 students max)
  - Criteria:
    - Compete with a team with 4 members or less
    - Meet the goal(s)
- **Intermediate Track**
  - Prize: Soundcore Headphones by Anker Life
  - Number of winners: 1 team (4 students max)
  - Criteria:
    - Compete with a team with 4 members or less
    - Meet the goal(s)
- **Advance Track**
  - Prize: Nintendo Switch
  - Number of winners: 1 team (4 students max)
  - Criteria:
    - Compete with a team with 4 members or less
    - Meet the goal(s)

In addition to these prizes all participants who submit will receive free swag!

## VII. Workshops

## Workshops Descriptions:

- **Exploratory Data Analysis with Pandas**
  - This workshop will cover the basics of how to conduct a good initial exploration of a dataset using Pandas.
- **Intro to ML & NLP with Sklearn**
  - Join this workshop for an introduction to machine learning using sci-kit learn and what NLP facilities the ubiquitous package has to offer.
- **Deep Learning for Natural Language Processing**
  - Targeted at advanced data scientists, this workshop details the importance of deep learning in NLP.
- **Final Submission**
  - Join this workshop to gain a fundamental understanding of what a good report looks like, and what is expected of a model and its subsequent analysis.

## VII. Submission Instructions

- **Submission Timeline:**
  - Submission link activated: Friday April 9th at 9:00 pm
  - Submission reminder: Sunday April 11th at 11:45 am
  - Submission freeze: Sunday April 11th at 1:00 pm
- **How to submit?**
  - Please submit your project through this [link](#).
  - Please name your submission files following this format:
    - ReligiousText\_TeamName\_report
    - ReligiousText\_TeamName\_code
    - Economics\_TeamName\_report
    - Economics\_TeamName\_code
    - Economics\_TeamName\_predictions
    - Cryptocurrency\_TeamName\_report
    - Cryptocurrency\_TeamName\_code
    - Cryptocurrency\_TeamName\_predictions
    - Vulcan\_TeamName\_report
    - Vulcan\_TeamName\_code
  - Code alternative submission format: github repository link
- **How to convert a Jupyter Notebook to pdf?**
  - Right click and print
  - Set Destination or Printer to **Save As PDF**