

2021 DataHacks Attendee Resource Guide

Table of Contents:

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

I. Introduction

Welcome to the DataHacks 2021! This document will tell you everything you need to know about how things will run. We hope that even though this event will be virtual, you will all take away with you valuable skills and friends. Should there be any questions, reach out to us on Discord!

DataHacks is an annual data science competition organized by DS3, a data science student community at the University of California, San Diego. DS3 organized DataHacks every year in hope of bringing more students into the data science student community. 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](#)!

II. Communications - Discord and Hopin

- **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 on the HACKERS subchannel. 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

- **Hopin**

- F. This year we will be using Hopin as our main platform for the **opening ceremony, award ceremony, and workshops!** If you're not familiar with Hopin, please check out these quick guides to learn how to use the platform!
- G. Hopin Guide Links
1. [The Hopin Attendee Experience](#)
 2. [Hopin Attendee Troubleshooting/Questions](#)
 3. [Hopin Session FAQs](#)
- H. Please register through this [link](#) if you haven't already. You will need to create an Hopin account in order to join the event. (**password: datahacks2021**)
- I. At the time of the event, all you need to do is login in to your Hopin account (or create an account) and join the event page.

III. Schedule

DataHacks 2021 Schedule

(all times in PST)

<u>Date/Time</u>	<u>Workshop/Event</u>
Friday, April 9	
6:00 pm - 9:00 pm	Team Building Channel active
9:00 pm - 10:00 pm	Opening Ceremony via Hopin
10:00 pm	Hacking begins
10:00 pm - 10:30pm	Workshop - Beginner (Science) Track (all workshops on Hopin)
	Workshop - Economics Track
	Workshop - Bitcoin Track
	Workshop - Advanced (Vulcan) Track
10:30 pm - 11:30 pm	Workshop - Intro to Python/Pandas
Saturday, April 10	
10:30 am - 11:15 am	Workshop - Using GitHub
11:15 am - 12:00 pm	Workshop - Hypothesis/Experimental Testing
12:00 pm - 1:00 pm	Workshop - ML Modeling & Feature Selection

1:00 pm - 2:00 pm	Workshop - Data Visualization
2:00 pm - 3:00 pm	Workshop - Time Series (including data viz)
3:30 pm - 4:30 pm	Workshop - Report Writing
4:30 pm - 5:30 pm	Workshop - Deep Learning
5:30 pm - 6:30 pm	Workshop - Transfer Learning
8:00 pm - 9:00 pm	Destress Activity
Sunday, April 11	
11:45 am	Submission Reminder
12: 00 pm	Code & submissions freeze
12:00 pm - 3:00 pm	Judging
3:00 pm - 3:30 pm	Closing Ceremony

IV. Competition Tracks

Beginner Track: Religious Text

- Dataset description: In this track, you will answer questions that will aim to provide context regarding religious texts.
- Goals: Use Natural Language Processing tools to find patterns, trends and changes between different religious books and their wordings. You will present a report which mainly includes data visualization and experimental testing to find relationships between different texts.
- Submission format: pdf report

Intermediate Track I (Bitcoin/Cryptocurrency)

- Dataset description: This track addresses and motivates analysis of ransomware payments in the Bitcoin network, between the years 2009 and 2018.
- Goals: Create models to answer questions that may help identify the root Bitcoin addresses of malicious actors. Also, create a report with at least 3 data visualizations and a hypothesis test to help identify addresses involved in such payments
- Submission format: pdf report

Intermediate Track II (Prosperity)

- Dataset description: This challenge focuses on prosperity and growth of a given set of countries between the years 2007 and 2016 with regard to various factors like education,

governance, etc. We aim to identify the most prosperous countries, as well as factors behind their growth.

- Goals: Use machine learning and data visualization to provide a growth report for listed countries. Your final product should include at least 3 data visualizations and a model that ranks countries by their overall prosperity scores.
- Submission format: pdf report

Advance Track

- Dataset description: In this project, you will have the privilege to use real-world stock market data given to us directly by our industry partners, Vulcan Value Partners.
- Goals: Your task is to create several different time series models over varying time frames. After creating your model, you will create a report which answers several questions about your model, and includes visualizations for the same. You are free to use any method that you deem fit, as long as you are not leveraging any data which we have not already provided to you.
- Submission format: pdf report

V. Prizes

- **Free Swag for All Participants!**
 - All participants will get free swag including a T-shirt after participating in DataHacks!
 - How to participate (only need to fulfill one condition from below):
 - Submission
 - Attend workshop twice
- **Beginner Track Prizes: Religious Text**
 - Prize: Sony WI-XB400 Wireless Headset
 - Number of winner: 1 team (4 students max)
 - Criteria:
 - Compete with a team with 4 members or less
 - Meet the goal(s)
- **Intermediate Track I Prizes: Economics**
 - Prize: Amazon Fire Stick Lite
 - Number of winner: 1 team (4 students max)
 - Criteria:
 - Compete with a team with 4 members or less
 - Meet the goal(s)
- **Intermediate Track II Prizes: Bitcoin**
 - Prize: Amazon Fire Stick Lite
 - Number of winner: 1 team (4 students max)
 - Criteria:
 - Compete with a team with 4 members or less

- Meet the goal(s)
- **Advance Track Prizes: Vulcan**
 - Prize: Amazon Echo Dot (3rd gen)
 - Number of winner: 1 team (4 students max)
 - Criteria:
 - Compete with a team with 4 members or less
 - Meet the goal(s)

VI. Workshops

Workshops Descriptions:

- **Beginner Track Intro: Religious Text Analysis**
 - Join this workshop for detailed explanation of the dataset, goals, and expected submission for the Religious Text Analysis Track
- **Intermediate Track I Intro: Economics**
 - Join this workshop for detailed explanation of the dataset, goals, and expected submission for the Economic Track
- **Intermediate Track II Intro: Cryptocurrency**
 - Join this workshop for detailed explanation of the dataset, goals, and expected submission for the Cryptocurrency Track
- **Advance Track: Vulcan**
 - Join this workshop for detailed explanation of the dataset, goals, and expected submission for the Vulcan Track
- **Introduction to Python/Pandas**
 - Python and the Python Pandas library are commonly used for data cleaning, data analysis, and machine learning. Highly recommended for beginners!
- **Using GitHub**
 - Learn how to use GitHub for version control and team collaboration (combine codes).
- **Hypothesis/Experimental Testing**
 - Learn how to use hypothesis/experimental testing in generating meaningful analysis.
- **ML Modeling & Feature Selection**
 - Introduction to machine learning and feature selection. Learn how to use machine learning in your project and how to select the right feature.
- **Data Visualization**
 - Introduction to data visualization and data viz tools. Learn how to create powerful visualizations that would help your project stand out.
- **Time Series (with Data Viz)**

- Learn how data points can be represented over a time interval and how to visualize these points.
- **Report Writing**
 - Learn how to write an efficient report for your final submission.
- **Deep Learning**
 - Introduction to deep learning. Deep learning is a subset of machine learning. It could be a really powerful tool in certain situations.
- **Transfer Learning**
 - Introduction to transfer learning. Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task.

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 12:00 pm
- **How to submit?**
 - Please submit your project through this [link](#).
 - All submissions need to be a pdf file.
 - Please name your submission following this format:
 - ReligiousText_TeamName
 - Economics_TeamName
 - Cryptocurrency_TeamName
 - Vulcan_TeamName
- **How to convert a Jupyter Notebook to pdf?**
 - Right click and print
 - Set Destination or Printer to Save As PDF

VIII. Sponsors (Link attached)

- [Data Science Alliance](#)
Data Science Alliance is a community of data enthusiasts committed to advancing responsible data practices. As champions of data science, we believe in the democratization of data and pursue efforts to build public trust. We work to promote San Diego's data science ecosystem and invite you to join us.
- [Vulcan Value Partners](#)
Vulcan Value Partners are value investors—business analysts with a long-term time horizon focused on purchasing publicly traded companies that are competitively entrenched at significant discounts to intrinsic worth.

As an independently owned and operated firm, we believe that our capital should be invested alongside our client partners. Our goal is to compound capital at real rates of return significantly in excess of inflation over our five year time horizon.

IX. Acknowledgement

1. We would like to advertise for this event as an acknowledgement to the Rady Analytics Club!

- a. **The Petco Challenge!**

- The Petco Challenge, a case competition of the Rady School of Management, empowers undergraduate and graduate students by taking them beyond the classroom and giving them exposure to real-world business situations. The Challenge offers students the opportunity to apply the concepts they have learned and showcase their knowledge, critical thinking skills, leadership, and presentation abilities.

- b. Entry deadline: April 23rd, 2021
 - c. When: April 29th 2021 - May 2nd 2021
 - d. Where: Virtual

RADY DATA ANALYTICS CLUB PRESENTS:



CASE STUDY COMPETITION



Questions? Email dataanalyticsclub@rady.ucsd.edu

BEGINS: APRIL 29, 2021 | 9 AM
ENDS: MAY 2, 2021 | 11:59 PM



petco.
THE HEALTH +
WELLNESS CO.