The Erdős Institute

2023 May Data Science Boot Camp

Introduction

Welcome!

- Welcome to the May 2023 Data Science Boot Camp!
- In this boot camp we will:
  - Learn some python
  - Learn some data science
  - Complete a data science project

Top two resources

1. Boot Camp Website, https://www.erdosinstitute.org/programs/spring-2023/data-science-boot-camp
2. Erdős Institute Slack
   a. Look for spring-2023-cohort channel

Lecturer

- Matthew Osborne, PhD
  - Head of Boot Camps at the Erdős Institute
- Graduated from OSU Math in 2020

Group Project Coordinator

- Alec Clott, PhD
  - Head of Data Science Projects
  - Sr. Principal, Quantitative Analytics and Data Science @ Gartner
- Graduated from OSU Political Science in 2021

Community Manager

- Olivia Haimerl
  - Graduate of The Ohio State University with a Bachelor of Arts in Medical Anthropology
  - About to start law school at the University of Virginia
- Your top contact for:
  - Slack channel access
  - GitHub repository access

Group Project Information

- Turning presentation over to Alec

The Erdős Institute Projects

May 2023

Goals

- An opportunity to work with real-world data and produce findings in a short time-span
- Focus on substantive areas (quant finance, wearables, health, policy, UX, etc.) and techniques (NLP, time-series, etc.)
- Building your portfolio is crucial in the data science market, provides a framework for job interviews

Projects

- Portfolio-worthy data science project/product
- Includes:
  - 5-minute overview video and presentation
  - Annotated GitHub
  - Executive Summary
- Reviewed by project judges
- Top 5 projects will present to all participants in our closing ceremony for the May 2023 Bootcamp

Team Formation

Background of Bootcamp attendees

- Hundreds of students from all over the world
- Some of you may know other attendees, others of you won't
- Many different backgrounds (subject areas, experience with coding)*
- Various types of data science career goals
- Various goals for the bootcamp
- Various goals for the projects

*And that is totally fine and expected!

Read these documents

https://www.erdosinstitute.org/programs/spring-2023/data-science-boot-camp

(Project Information at Bottom)

Project Expectations

Overall Structure

- **Team size:** 3-4 people
  - Can choose group or...
  - Can be matched with others based on your job goals or data science interests
- **Goals:** “portfolio” project
  - Can be used in job interviews (when the time comes)
  - Results have business value
  - Communicate to lay-people **and** team of data scientists
- **Structure**
  - Group meetings -- each group decides how much time they want to spend
  - Weekly office hour w Alec (open to all)
  - Check-in with project mentor on a regular basis (15-30 min)

Project Requirements

- **Instructions at the bottom of the 2023 Data Science Bootcamp Page**

- **In order to get an Erdős certificate, you must complete a data science project start to finish**
  - Project must be coded in Python
  - Have an annotated GitHub repository
  - Executive summary of your project results and implications
  - ***For presentation day:***
    - **5-min** pre-recorded PowerPoint presentation detailing project process from start to finish
    - Judges will vote on winners!
    - More info will be given closer to project day

Your To-Do List

This Week:

Project/Homework Deadlines

May 12, 2023 12:00 AM
Project Pitch Day (Live on Zoom)
Opportunity to meet with other Erdos Fellows and form teams and propose topics.

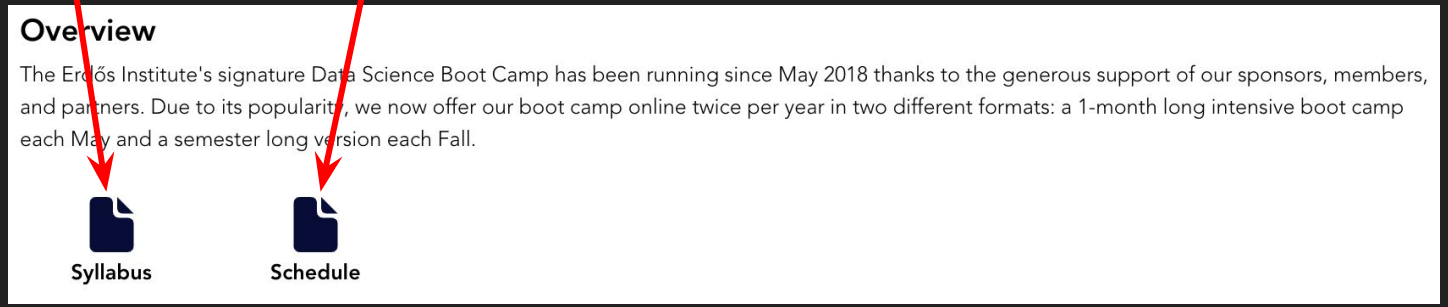
May 12, 2023 12:00 AM
Submit Team Proposal to Project Formation Page
If you want to propose a project, or have an idea for a project, submit it by this date.

May 14, 2023 12:00 AM
Finalized Teams with Preliminary Project Idea
Teams need to be finalized by this point. If you proposed or created a project, you must have others in your group. If you did not propose or create a project, you must join an open group.

Note: Times are not accurate, check the course website for the correct deadline times.

Boot Camp Format: Non-Project Portion

- 11 Live Lectures
- 11 Problem Solving Sessions
- All Zoom links can be found in your Erdős profile or on the course website
- Syllabus and Schedule can be found on the course website



Lectures

- Live lectures 5:30 - 7:00 PM ET every MTWT until May 25, 2023
  - Will be recorded and uploaded to the website
  - Will not be able to cover entirety of GitHub repository
- Every lecture jupyter notebook already has a pre-recorded lecture on the website

Problem Sessions

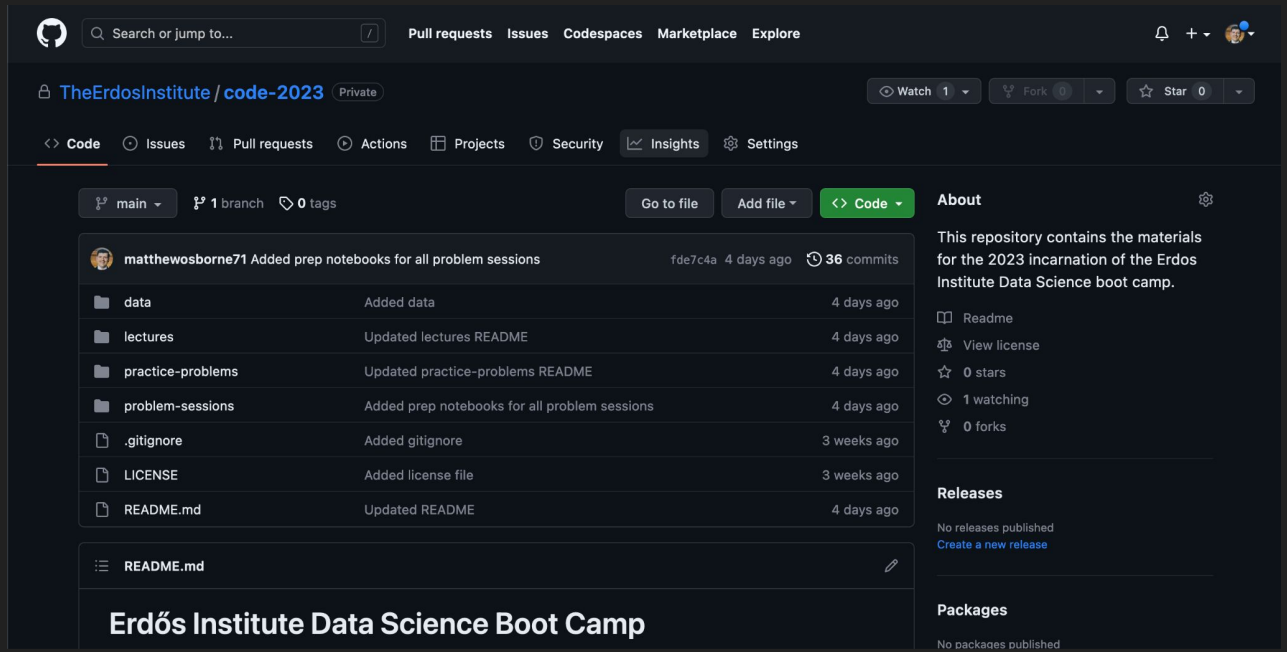
- One hour to work on problem sets in small groups
- Every MTWT from 5/9/23 to 5/25/23
- Two identical sessions:
  - 10-11 AM ET
  - 4-5 PM ET
- TAs will rotate between groups to assist and observe
- Each problem session also has a “prep notebook” for those wanting to practice
- Start tomorrow (5/9/23)

Getting Set Up

- Clone the repository
- Be able to open a jupyter notebook

The GitHub Repository

- Link can be found on the course website
- Contains all of the educational content for the boot camp



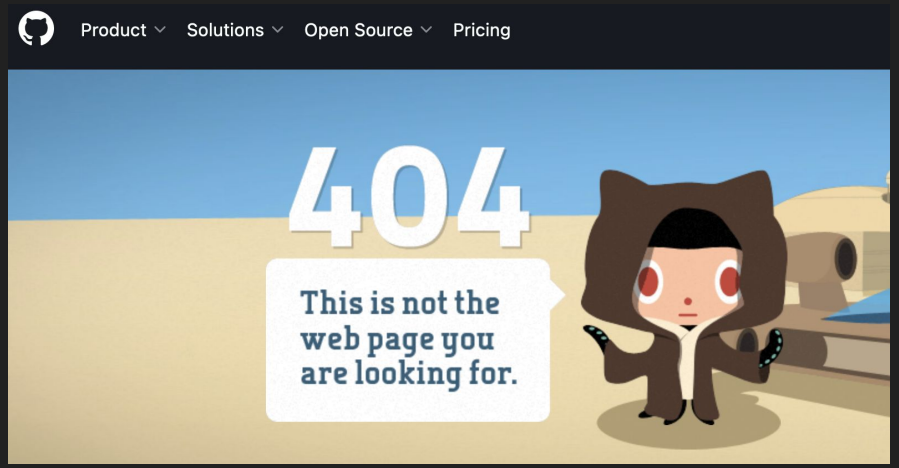
The GitHub Repository - Steps

1. Sign into your GitHub account
2. Clone the repository onto your computer
    a. Can find instructions in the “First Steps” section of the website
3. Everyday of the boot camp you will need to “pull” the updates to the repository
    a. Look for “Giting Started with GitHub” in the “First Steps” section of the website
4. Can be useful to make a copy of the folder in your computer for your edits

The GitHub Repository - 404 Issue

If you receive the 404 Error when clicking repo link:

- Check you are signed in
- Check that you have added your GitHub link to your Erdős profile
- Message Olivia about being added to the repository



Jupyter Notebooks

- All educational content contained in jupyter notebooks
- Allows combination of Word document style text and python code
- Let’s look at an example

Jupyter Notebooks - Getting Set Up

- Follow Step 3 Under “First Steps” on website
- Can use Anaconda Navigator, https://docs.anaconda.com/free/navigator/index.html
- Can also just install like any other python package, https://jupyter.org/install

Questions & Concerns?