The Erdős Institute

Spring 2025 Data Science Boot Camp

Introduction

Welcome!

- Welcome to the Erdős Institute 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

- Boot Camp Website, https://www.erdosinstitute.org/programs/spring-2025/data-science-boot-camp
- Erdős Institute Slack
  - Gain access through the course homepage.
  - spring-2025-launch-cohort
  - spring-2025-data-science

Lecturer

- Steven Gubkin, PhD
  - Head of Training and Assessment at Erdős since 01/01/24
  - PhD in Mathematics from Ohio State University
  - Taught math at Cleveland State from 2016 - 2023

Group Project Coordinator

- Alec Clott, PhD
  - Head of Data Science Projects
  - Sr. Principal, Quantitative Analytics and Data Science at Gartner
- Graduated from OSU Political Science in 2021
- Your top contact for:
  - Project admin/requirement questions
  - Team formation questions

The Erdős Institute Projects

Goals

- An opportunity to work with real-world data and produce findings in a short time-span
- Focus on substantive areas (environment, health, finance, etc.) using techniques from the bootcamp.
  - The focus should be on using what we learn.
  - Okay to use more advanced methods. Just make sure to compare their performance to the best model you could make using methods covered in the bootcamp.
- 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 slide show presentation
  - Annotated GitHub
  - Executive Summary
- Reviewed by project judges
- Top 5 projects will present to all participants in our closing ceremony.

Team Formation

Background of boot camp 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-2025/data-science-boot-camp

(Project Information at Bottom)

Team Formation - Live Demo

https://www.erdosinstitute.org/programs/spring-2025/data-science-boot-camp/project-formation

Project Expectations

Overall Structure

- **Team size:** 3-5 people

- **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
  - Check-in with project mentor on a regular basis (15-30 min)

Project Requirements

- **Instructions at the bottom of the 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

First Important Dates:

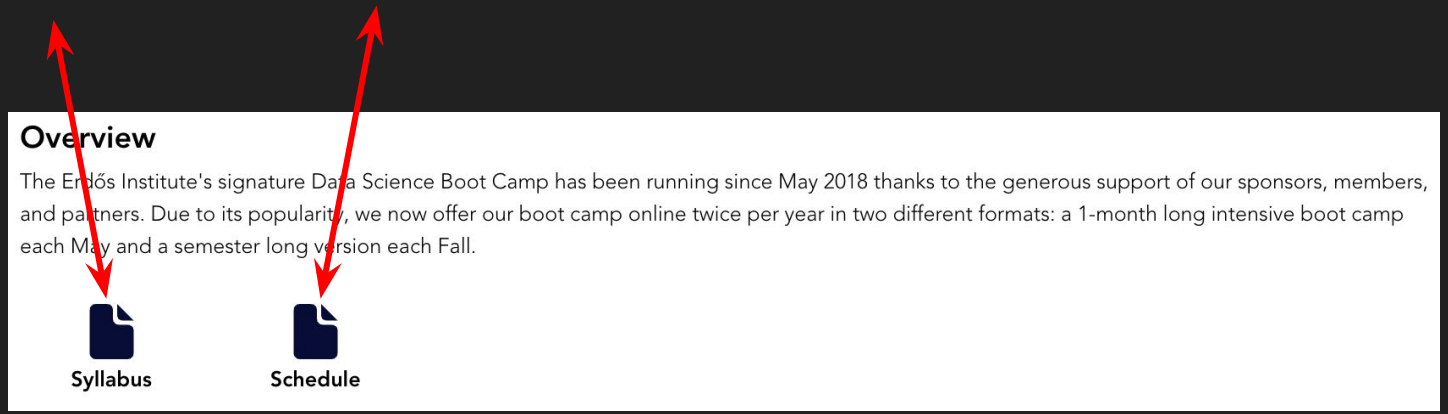
DS Bootcamp computer setup day Jan 23, 2025 at 02:00 PM EST EVENT	**Lecture 01: Introduction, Computer Setup, Q/A** Jan 28, 2025 at 12:00 PM EST EVENT	**Math Hour 01** Jan 29, 2025 at 10:00 AM EST EVENT
Office Hour 01 Jan 29, 2025 at 11:00 AM EST EVENT	**Problem Session 01** Jan 30, 2025 at 02:00 PM EST EVENT	**Lecture 02: Regression I** Feb 4, 2025 at 12:00 PM EST EVENT
Math Hour 02 Feb 5, 2025 at 10:00 AM EST EVENT	**Office Hour 02** Feb 5, 2025 at 11:00 AM EST EVENT	**Problem Session 02** Feb 6, 2025 at 02:00 PM EST EVENT

Feb 7, 2025 11:59 AM EST	**Watch video about Project Formation** This should help answer any Q's you may have going into project formation
Feb 7, 2025 11:59 AM EST	**Watch 3 Previous Top Projects** Consult the project database, and watch at least 3 previous top projects from Erdos Alumni.
Feb 19, 2025 11:59 AM EST	**Project Pitch Hour** Opportunity to meet with other Erdos Fellows and form teams and propose topics.
Feb 21, 2025 11:59 AM EST	**Finalized Teams with Preliminary Project Ideas** 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: You can find these dates at the bottom of the course website

Boot Camp Format: Non-Project Portion

- 12 Live Lectures
- 12 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 12:00 - 1:30 PM ET every Tuesday until April 15th
  - Will be recorded and uploaded to the website
- Every lecture jupyter notebook already has a pre-recorded lectures on the website.

Problem Sessions

- One hour to work on problem sets in small groups
- Every Thursday 2:00 PM - 3:00 PM ET
  - Will not be recorded
- TAs will rotate between groups to assist and observe
- Many problem sessions also have a “prep notebook” with prerequisite practice.

Math Hour and Office Hour

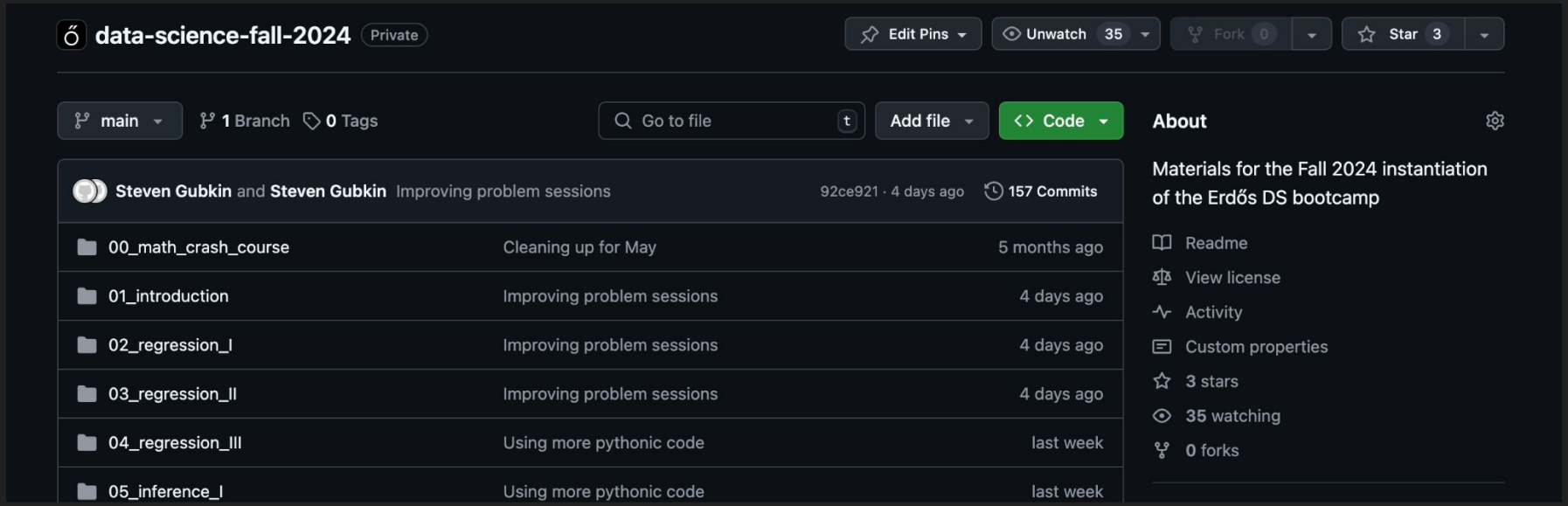
- Math Hour are every Wednesday 10:00AM to 11:00AM ET.
  - We go a little deeper into the math behind the techniques covered in lecture.
  - These are **optional.**
- Office Hour are every Wednesday 11:00AM to 12:00PM ET and by appointment.
  - Ask anything about course content, projects, debugging, etc.
  - These are **optional.**

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



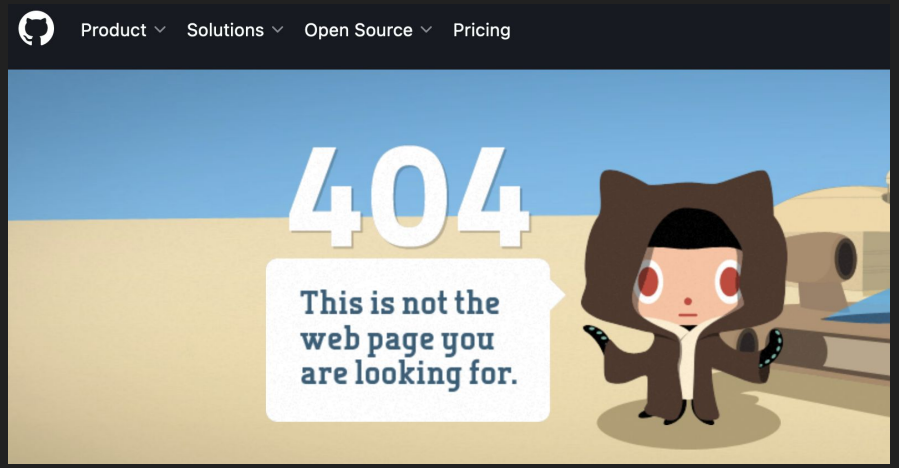
The GitHub Repository - Steps

- Sign into your GitHub account
- Clone the repository onto your computer
  - Can find instructions in the “First Steps” section of the website
- Everyday of the boot camp you will need to “pull” the updates to the repository
  - Look for “Giting Started with GitHub” in the “First Steps” section of the website
- Either make a folder where you copy over files you want to work on (leaving the git repo folder “clean”) or make a local branch where you do your work.

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
- Request access on the course homepage



Jupyter Notebooks

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

Jupyter Notebooks - Getting Set Up

- Lots of options:
  - Visual Studio Code ← this is what we officially support.
  - Jupyter Notebook
  - Anaconda Navigator
  - Many other options

Conda Environment

- If you want the most streamlined experience possible this semester, you should set up the conda environment specified in the repository and run all of the notebooks with this environment.
  - Instructions in the repo README document
- Make sure you can run the following notebooks with this environment to confirm everything is working correctly:
  - `computer_setup_day/secret_code.ipynb`

Questions & Concerns?