

Summer 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/summer-2025/data-science-boot-camp>
- Erdős Institute Slack
 - Gain access through the course homepage.
 - #summer-2025-launch-cohort
 - #summer-2025-job-help
 - #summer-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



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!

Phase I: Data Science instruction and Group Project

- 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

Overview

The Erdős Institute's signature Data Science Boot Camp has been running since May 2018 thanks to the generous support of our sponsors, members, and partners. Due to its popularity, we now offer our boot camp online twice per year in two different formats: a 1-month long intensive boot camp each May and a semester long version each Fall.



Syllabus



Schedule

Lectures

- Live lectures 1:30 - 3:00 PM ET every Tu/Th until June 19th
 - Will be recorded and uploaded to the website
- Most lectures already have a pre-recorded video on the website broken down by notebook.

Problem Sessions

- One hour to work on problem sets in small groups
- Every W/F 1:00 PM - 2:00 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 W/F 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 by appointment.

The GitHub Repository

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

The screenshot shows the GitHub repository page for 'data-science-fall-2024'. The repository is private and has 3 stars and 0 forks. The main branch is selected. The repository description is 'Materials for the Fall 2024 instantiation of the Erdős DS bootcamp'. The repository contains a table of files and folders, including '00_math_crash_course', '01_introduction', '02_regression_I', '03_regression_II', '04_regression_III', and '05_inference_I'. The repository was created by Steven Gubkin and Steven Gubkin, and it has 157 commits.

data-science-fall-2024 Private

Edit Pins Unwatch 35 Fork 0 Star 3

main 1 Branch 0 Tags

Go to file Add file Code

About

Materials for the Fall 2024 instantiation of the Erdős DS bootcamp

Readme View license Activity Custom properties 3 stars 35 watching 0 forks

Steven Gubkin and Steven Gubkin	Improving problem sessions	92ce921 · 4 days ago	157 Commits
00_math_crash_course	Cleaning up for May	5 months ago	
01_introduction	Improving problem sessions	4 days ago	
02_regression_I	Improving problem sessions	4 days ago	
03_regression_II	Improving problem sessions	4 days ago	
04_regression_III	Using more pythonic code	last week	
05_inference_I	Using more pythonic code	last week	

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 “Getting Started with GitHub” in the “First Steps” section of the website
- Make a folder where you copy over files you want to work on (leaving the git repo folder “clean”).

The GitHub Repository - 404 Issue

If you receive the 404 Error when clicking repo link:

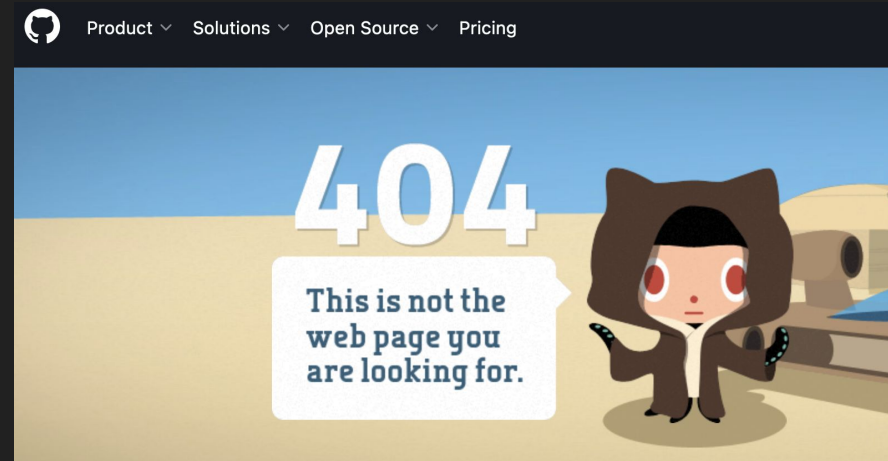
- Check you are signed in
- Check that you have added your GitHub link to your Erdős profile
- Request access on the course homepage

Course materials are available on github through the following link:



<https://github.com/TheErdosInstitute/data-science-spring-2025>

Request Access to GitHub



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`
- Enter the secret code here:
<https://www.erdosinstitute.org/ds-boot-camp-prep>

Data Science Group Projects

- 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

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 Submission

- Deliverables:
 - 5-minute overview video and slide show presentation
 - Annotated GitHub
 - Executive Summary
- Reviewed by project judges
 - You will get individualized feedback on your project.
 - You will earn a certificate as long as the project is deemed acceptable for a resume.
 - If the project would actually detract from your resume we will delay issuing a certificate until you have addressed the issues raised by the judges.
- Top 5 projects will present to all participants in our closing ceremony.

Team Formation - Live Demo

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

Phase II: Technical Interview Preparation

- If you completed a portfolio worthy Data Science project in Phase I you will continue on to Phase II: Technical Interview Preparation.
 - This will consist of self-study and mock interview questions on probability, statistics, machine learning and SQL.
 - You will also complete mock “take home” mini-projects.
- If you did not complete a portfolio worthy group project in Phase I we encourage you to work on completing such a project in Phase II.

Questions & Concerns?