

# The Erdős Institute

## Fall 2022 Data Science Boot Camp

### Schedule

Each week will have a single live virtual lecture from Matthew Osborne on Wednesday from 5:00 - 6:30 PM ET. All lecture materials also have pre-recorded asynchronous lectures that you can watch on your own time, <https://www.erdosinstitute.org/boot-camp>. I will attempt to cover all notebooks not listed as “Optional” below. All lecture notebooks listed as “Optional” will not be covered directly, but I wanted to put them in context with the relevant notebooks. Also, even if you plan on attending the live lecture I encourage you to watch the pre-recorded videos as well.

There will also be multiple problem solving sessions every week, with each session working through the same notebook. Exact times are TBA until we get our TAs in place. Each problem solving session will cover the material covered in the most recent lecture. For example, if there was a problem solving session on a Tuesday it would cover the material covered in lecture on the previous Wednesday.

## Prior to Boot Camp

- Set up your computer for the boot camp by following the steps found in the “First Steps” section here, <https://www.erdosinstitute.org/data-science>,
- Ensure you have access to the boot camp GitHub repository, <https://github.com/TheErdosInstitute/code-2022/> and
- Make sure you understand Python to the extent covered in the “Python Prep” materials in the GitHub repository, <https://www.erdosinstitute.org/python-prep>.

## Week 0: 8/29/2022 - 9/6/2022

- Attend the Erdős Fall orientation or watch the recorded session,
- Check your computer setup by trying to run any lecture jupyter notebook,
- Start watching the lectures in week 1.

## Week 1: 9/7/2022 - 9/13/2022

Live Zoom Lecture: Wednesday 9/7/2022 5:00 - 6:30 PM ET

### Lecture Notebooks to Cover

- Introduction/0. Welcome! (Optional)
- Introduction/1. A Broad Overview
- Data Collection/1. Data File Types (Optional)
- Data Collection/2. Data Repositories (Optional)
- Data Collection/3. Data Competition Sites (Optional)
- Data Collection/4. Data in Databases
- Data Collection/5. Web Scraping with BeautifulSoup
- Data Collection/6. Python and APIs
- Data Collection/7. Summary and Conclusion (Optional)

### Problem Session

#### Session Times

- Thursday 9/8/2022 5:00 - 6:00 PM ET
- Sunday 9/11/2022 2:00 - 3:00 PM ET
- Tuesday 9/13/2022 3:00 - 4:00 PM ET

### Notebooks Covered

- Fall Prep Notebook 1 (Optional)
- Fall Problem Session Notebook 1

## Week 2: 9/14/2022 - 9/20/2022

Live Zoom Lecture: Wednesday 9/14/2022 5:00 - 6:30 PM ET

### Lecture Notebooks to Cover

- Supervised Learning/1. Introduction (Optional)
- Supervised Learning/2. A Supervised Learning Framework
- Supervised Learning/3. Data Splits for Predictive Modeling
- Supervised Learning/Regression/1. Simple Linear Regression
- Supervised Learning/Regression/2. A First Predictive Modeling Project

### Problem Session

#### Session Times

- Thursday 9/15/2022 5:00 - 6:00 PM ET
- Sunday 9/18/2022 2:00 - 3:00 PM ET
- Tuesday 9/20/2022 3:00 - 4:00 PM ET

### Notebooks Covered

- Fall Prep Notebook 2 (Optional)
- Fall Problem Session Notebook 2

## Week 3: 9/21/2022 - 9/27/2022

Live Zoom Lecture: Wednesday 9/21/2022 5:00 - 6:30 PM ET

### Lecture Notebooks to Cover

- Supervised Learning/Regression/3. Multiple Linear Regression
- Supervised Learning/Regression/4. Categorical Variables and Interactions
- Supervised Learning/Regression/5. Polynomial Regression
- Cleaning/1. Introduction (Optional)
- Cleaning/2. Scaling Data
- Cleaning/3. Basic Pipelines

### Problem Session

#### Session Times

- Thursday 9/22/2022 5:00 - 6:00 PM ET
- Sunday 9/25/2022 2:00 - 3:00 PM ET
- Tuesday 9/27/2022 3:00 - 4:00 PM ET

## Notebooks Covered

- Fall Prep Notebook 3 (Optional)
- Fall Problem Session Notebook 3

## Week 4: 9/28/2022 - 10/4/2022

Live Zoom Lecture: Wednesday 9/28/2022 5:00 - 6:30 PM ET

### Lecture Notebooks to Cover

- Supervised Learning/4. Bias-Variance Trade-Off
- Supervised Learning/Regression/6. Regularization
- Supervised Learning/Regression/7. Interpreting Linear Regression (Optional)
- Supervised Learning/Regression/8. Linear Regression Diagnostic Plots
- Supervised Learning/Regression/9. Feature Selection Approaches
- Supervised Learning/5. Data Splits and Overfitting (Optional)

### Problem Session

#### Session Times

- Thursday 9/29/2022 5:00 - 6:00 PM ET
- Sunday 10/2/2022 2:00 - 3:00 PM ET
- Tuesday 10/4/2022 3:00 - 4:00 PM ET

## Notebooks Covered

- Fall Prep Notebook 4 (Optional)
- Fall Problem Session Notebook 4

## Week 5: 10/5/2022 - 10/11/2022

Live Zoom Lecture: Wednesday 10/5/2022 5:00 - 6:30 PM ET

### Lecture Notebooks to Cover

- Supervised Learning/Time Series Forecasting/1. What are Time Series and Forecasting
- Supervised Learning/Time Series Forecasting/2. Adjustments for Time Series Data
- Supervised Learning/Time Series Forecasting/3. Time and Dates in Python (Optional)
- Supervised Learning/Time Series Forecasting/4. Baseline Forecasts

### Problem Session

#### Session Times

- Thursday 10/6/2022 5:00 - 6:00 PM ET
- Sunday 10/9/2022 2:00 - 3:00 PM ET
- Tuesday 10/11/2022 3:00 - 4:00 PM ET

## Notebooks Covered

- Fall Prep Notebook 5 (Optional)
- Fall Problem Session Notebook 5

## Week 6: 10/12/2022 - 10/18/2022

Live Zoom Lecture: Wednesday 10/12/2022 5:00 - 6:30 PM ET

### Lecture Notebooks to Cover

- Supervised Learning/Time Series Forecasting/5. Averaging and Smoothing
- Supervised Learning/Time Series Forecasting/6. Stationarity and Autocorrelation
- Supervised Learning/Time Series Forecasting/7. ARIMA
- Supervised Learning/Time Series Forecasting/8. Next Steps for Time Series (Optional)

### Problem Session

#### Session Times

- Thursday 10/13/2022 5:00 - 6:00 PM ET
- Sunday 10/16/2022 2:00 - 3:00 PM ET
- Tuesday 10/18/2022 3:00 - 4:00 PM ET

## Notebooks Covered

- Fall Prep Notebook 6 (Optional)
- Fall Problem Session Notebook 6

## Week 7: 10/19/2022 - 10/25/2022

Live Zoom Lecture: Wednesday 10/19/2022 5:00 - 6:30 PM ET

### Lecture Notebooks to Cover

- Supervised Learning/Classification/1. Adjustments for Classification
- Supervised Learning/Classification/2. k Nearest Neighbors Classifier
- Supervised Learning/Classification/3. The Confusion Matrix
- Supervised Learning/Classification/4. Logistic Regression

### Problem Session

#### Session Times

- Thursday 10/20/2022 5:00 - 6:00 PM ET
- Sunday 10/23/2022 2:00 - 3:00 PM ET
- Tuesday 10/25/2022 3:00 - 4:00 PM ET

## Notebooks Covered

- Fall Prep Notebook 7 (Optional)
- Fall Problem Session Notebook 7

## **Week 8: 10/26/2022 - 11/1/2022**

Live Zoom Lecture: Wednesday 10/26/2022 5:00 - 6:30 PM ET

### **Lecture Notebooks to Cover**

- Supervised Learning/Classification/5. Diagnostic Curves
- Supervised Learning/Classification/6. Bayes' Based Classifiers
- Unsupervised Learning/1. Introduction (Optional)
- Unsupervised Learning/Dimension Reduction/1. Principal Components Analysis

### **Problem Session**

#### **Session Times**

- Thursday 10/27/2022 5:00 - 6:00 PM ET
- Sunday 10/30/2022 2:00 - 3:00 PM ET
- Tuesday 11/1/2022 3:00 - 4:00 PM ET

#### **Notebooks Covered**

- Fall Prep Notebook 8 (Optional)
- Fall Problem Session Notebook 8

## **Week 9: 11/2/2022 - 11/8/2022**

Live Zoom Lecture: Wednesday 11/2/2022 5:00 - 6:30 PM ET

### **Lecture Notebooks to Cover**

- Supervised Learning/Classification/9. Decision Trees
- Supervised Learning/Classification/8. Support Vector Machines
- Supervised Learning/Ensemble Learning/1. What is Ensemble Learning
- Supervised Learning/Ensemble Learning/2. Random Forests
- Supervised Learning/Regression/10. Regression Version of Classification Algorithms (Optional)

### **Problem Session**

#### **Session Times**

- Thursday 11/3/2022 5:00 - 6:00 PM ET
- Sunday 11/6/2022 2:00 - 3:00 PM ET
- Tuesday 11/8/2022 3:00 - 4:00 PM ET

#### **Notebooks Covered**

- Fall Prep Notebook 9 (Optional)
- Fall Problem Session Notebook 9

## **Week 10: 11/9/2022 - 11/15/2022**

Live Zoom Lecture: Wednesday 11/9/2022 5:00 - 6:30 PM ET

## Lecture Notebooks to Cover

- Supervised Learning/Ensemble Learning/3. Bagging and Pasting (Optional)
- Supervised Learning/Ensemble Learning/4. Boosting
- Supervised Learning/Ensemble Learning/5. AdaBoost
- Supervised Learning/Ensemble Learning/6. Gradient Boosting
- Supervised Learning/Ensemble Learning/7. XGBoost (Optional)
- Supervised Learning/Ensemble Learning/8. Voter Models
- Supervised Learning/7. GridSearchCV (Optional)

## Problem Session

### Session Times

- Thursday 11/10/2022 5:00 - 6:00 PM ET
- Sunday 11/13/2022 2:00 - 3:00 PM ET
- Tuesday 11/15/2022 3:00 - 4:00 PM ET

### Notebooks Covered

- Fall Prep Notebook 10 (Optional)
- Fall Problem Session Notebook 10

## Week 11: 11/16/2022 - 11/20/2022

Live Zoom Lecture: Wednesday 11/16/2022 5:00 - 6:30 PM ET

## Lecture Notebooks to Cover

- Neural Networks/1. Perceptrons
- Neural Networks/2. The MNIST Data Set
- Neural Networks/3. Multilayer Neural Networks
- Neural Networks/4. keras

## Problem Session

### Session Times

- Thursday 11/17/2022 5:00 - 6:00 PM ET
- Sunday 11/20/2022 2:00 - 3:00 PM ET
- *Note:* No Tuesday session this week
- Fall Prep Notebook 11 (Optional)
- Fall Problem Session Notebook 11

## Thanksgiving (US) Break: 11/21/2022 - 11/27/2022

No events this week, enjoy the break.

## Week 12: 11/28/2022 - 12/6/2022

Live Zoom Lecture: Wednesday 11/30/2022 5:00 - 6:30 PM ET

## Lecture Notebooks to Cover

- Neural Networks/5. Introduction to Convolutional Neural Networks
- Neural Networks/6. Introduction to Recurrent Neural Networks
- Neural Networks/7. Loading Pre-Trained Models
- Neural Networks/8. Future Directions (Optional)
- Supervised Learning/6. Gradient Descent (Optional)

## Problem Session

### Session Times

- *Week 11 Makeup Session:* Tuesday 11/29/2022 3:00 - 4:00 PM ET
- Thursday 12/1/2022 5:00 - 6:00 PM ET
- Sunday 12/4/2022 2:00 - 3:00 PM ET
- Tuesday 12/6/2022 3:00 - 4:00 PM ET

### Notebooks Covered

- Fall Prep Notebook 12 (Optional)
- Fall Problem Session Notebook 12

## Week 13: 12/7/2022 - 12/9/2022

No Live Zoom Lecture this week.

## Lecture Notebooks to Cover

- Presentation Tips and Tricks/1. General Presentation Tips
- Presentation Tips and Tricks/2. Plotting Tips

## Problem Session

No problem sessions this week.

## Final Project

- Project Presentation Video Due Friday 11:59 PM ET 12/9/2022

## Final Note

Any `jupyter notebook` not mentioned in this schedule will not be covered directly in the boot camp and should be considered “Optional”. However, feel free to ask me about any `jupyter notebook` on Slack or in my office hours (see syllabus) regardless of whether it was optional.