Data Science Remote 02/11 - Class Resources Document

[**Class Meetings**](#_pt7romic61a1) **1**

[**Instructor Team**](#_hizzxn2gukpr) **1**

[**Communication**](#_b962s4xp73) **1**

[**Course Overview**](#_5mp921kk7qv1) **2**

[**Project Submission Procedure**](#_ll3eirurpmt5) **2**

[**What We Expect From You**](#_ign7nuni7vni) **3**

[Graduation Requirements](#_e3udkszg67op) 3

[Additional Expectations](#_xfgtkatb5e11) 3

[**Technology Requirements**](#_y4h40bz3tjqk) **3**

[**Group Work Formats**](#_fqqm0whw76xr) **4**

[Small-Group Discussion Roles (3-4 people)](#_po4zd9c5cc5b) 4

[Pair Programming Roles (2 people)](#_fiogwk19kn7v) 4

[**Lesson Schedule - 10 Week**](#_nxihx7p3sng5) **5**

# Class Meetings

**Time:** Tuesdays and Thursday, 8:00 to 11:00 PM (EST/EDT)

**Format:**

* **Before class:**
  + Use the appropriate links below to download lesson materials.
  + Join the [Zoom Room](https://generalassembly.zoom.us/j/3381283256).
  + Mark your attendance in Slack where prompted.
* **In class:**
  + Review where we are in the course.
  + Work through lesson modules.
  + Preview upcoming projects/homework.
  + Fill out the [exit ticket](https://docs.google.com/forms/d/e/1FAIpQLSd6xttQyifw4it6r8QuJp14LZs2oJw3rVsKIN7xiTzRiKYBRA/viewform).
* **After class** (optional)**:**
  + Ask additional questions on Zoom and/or Slack.

# Instructor Team

**Lead Instructor:** Adi Bronshtein ([LinkedIn](https://www.linkedin.com/in/adibronshtein))

**Instructional Associate:** Kihoon Sohn ([LinkedIn](https://www.linkedin.com/in/kihoonsohn))

**Teaching Assistant:** Andrew Riddle

# Communication

You can come off mute and ask questions during class, or ask questions in the [#datr-211 channel](https://q1datr2020.slack.com/archives/CT5SYUG04). My user on Slack is my full name, **Adi Bronshtein**. My email is [adi.bronshtein@generalassemb.ly](mailto:adi.bronshtein@generalassemb.ly). You can also book time for office hours with me by appointment. I will respond to Slack posts within 48 hours (**usually much less!**), and I am always happy to talk via Zoom immediately after a class session.

Kihoon’s username on Slack is his full name, **Kihoon Sohn**, and his email is [kihoon.sohn@ga.co](mailto:kihoon.sohn@ga.co) Kihoon will hold office hours ~~twice a week on Monday and Wednesday 9:00-10:00pm EST~~ once a week on Sunday 8:00-10:00pm EST starting from March 8th, by appointment via [the link](http://kihoonsohn.youcanbook.me). (His office hours can subject to be changed, you will get notified via Slack for any schedule changes.) Also, he is available anytime on Slack (will respond as quickly as possible) Don’t hesitate to reach out to him.

# Course Overview

We will use Python to explore datasets, build predictive models, and communicate data-driven insights. Specifically, you will learn how to:

* Define many of the approaches and considerations that data scientists use to solve real world problems.
* Perform exploratory data analysis with powerful programmatic tools in Python.
* Build and refine basic machine learning models to predict patterns from data sets.
* Communicate data-driven insights to peers and stakeholders in order to inform business decisions.

# Project Submission Procedure

For each project:

* Post your materials to a project-specific GitHub repository. (You should have four project repositories in all, corresponding to the three unit projects and the final project.)
* Use [this Google Form](https://forms.gle/yQYhyhh1erdjkgB2A) to submit a link to that repository.

# What We Expect From You

## Graduation Requirements

* Attend at least 17 of 20 class sessions.
* Meet expectations on all projects on time.

You may have up to a three-day grace period period on a project for extenuating circumstances *if you get approval before the due date*.

Unexcused absences are **not allowed**. If you will have to miss a class, **notify the Instructor Associate (IA)/Lead Instructor (LI) via Slack ahead of time** so that we can mark your absence as an excused absence. You may have a 48-hour grace period for notifying the IA/LI in case of emergency. Acceptable excuses include illness, death or critical illness to a family member of significant other, critical life emergency, and religious observances.

## Additional Expectations

* Take initiative.
* Ask for help.
* Practice, practice, practice.

# Technology Requirements

* 8GB RAM
* 10GB free hard drive space (after installing Anaconda)

macOS works best for this course. It is similar to Linux, so Linux will also work well in general but may raise occasional issues for which we will not be able to provide support. Windows is also an option, but it involves jumping through a few hoops with limited support.

# Group Work Formats

We use two main formats for group work in class. In both formats you will start by quickly assigning roles to group members so that you can spend less time negotiating over how you will interact and more time focusing on the task.

## Small-Group Discussion Roles (3-4 people)

* A **scribe** to record the group’s answers and share them with the class.
* A **manager** to keep the discussion moving at an appropriate pace.
* A **QA** (Quality Assurance) **person** to make sure that everyone understands and agrees with the group’s answers before the group moves on.

Any additional group members are simply **individual contributors**.

## Pair Programming Roles (2 people)

* A **driver** who writes the code or otherwise performs the relevant tasks while sharing his or her screen.
* A **navigator** who continually makes suggestions and reviews the work.

The driver should talk about what he or she is doing, ask for input, and generally keep the navigator engaged.

# 

# 

# Lesson Schedule - 10 Week

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Class** | **Date** | **Recording** | **Repository** | **Assignments Due** | **Resources (Slidedeck, reference materials, etc.)** |
| **Unit 1: Your Development Environment**  Get familiar with Python, Jupyter, Git, the command line, and fundamental data science concepts. | | | | | |
| 1 - Welcome to Data Science | 02/11/20 | [Link to the recording](https://generalassembly.zoom.us/rec/play/tMAoJOCorG43H9eXswSDA_UrW9TpJv-s1CQdr_AMnx22VXRSYFTwYLMRYuvW6cxKlGDQ-LS5sIjU-d-V?continueMode=true) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/welcome-to-data-science) |  |  |
| 2 - Your Development Environment | 02/13/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/ztxKMLfa-EVIHZHA0XCACvE-BdzlT6a82yJP-vdbzxzmsx8QKnYsXqP7yt2ayd8r) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/your-development-environment) |  | **Command line:** <https://www.youtube.com/playlist?list=PLeL7qmseLZybA_LvPtI2CfDxt3ynabEz2>  **GitHub:** <https://www.youtube.com/playlist?list=PLeL7qmseLZyZKedANxxvyWDoQQ2tDk-fq> |
| 3 - Your Development Environment, Python Foundations | 02/18/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/xMVuK4rv7HxIHKfvzl_4XIkdL9u_aaa81yMWq_oOnxwkV_n8wiRilbGGzK_Bfy4) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/python-foundations) |  | **Python curated list:** <https://www.youtube.com/playlist?list=PLeL7qmseLZyazhKMtpg5oG7ogP1vK7zVo> |
| **Unit 2: Statistics in Python**  Get introduced to fundamental concepts in statistics, and learn to perform basic data data analysis operations in Python. | | | | | |
| 4 - Python Foundations,  Exploratory Data Analysis in Pandas - Part 1 | 02/20/20 | [Link to the recording](https://generalassembly.zoom.us/rec/play/tcZ7duispjk3HtzGtgSDA_5wW465fPmshCMfr6EKyhyxVSJSMVenMLBGY-NtHKYcJa5gz3Ix8AS-2CIv?continueMode=true) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/eda-with-pandas/) |  | **Pandas:** <https://www.youtube.com/playlist?list=PLeL7qmseLZyY50TrntAiAOHEyEZ9O8xhT> |
| 5 - Exploratory Data Analysis in Pandas - Part 2 | 02/25/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/yuxPApGprnhIZ4Ho-XvGWIgQHJbXaaa803QYqfNYnkp2RuFa-4qD7eXns70P1wqk) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/eda-with-pandas/) |  |  |
| 6 - Statistics in Python | 02/27/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/vP18cbqp0WpIY6PnxkbBBr4oPYj6eaa8hiQZrPteyxpoQGy7AuNrDtWn_ZpMXQqF) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/statistics-in-python) | [**Unit 1 Project:** Python](https://git.generalassemb.ly/bronshtein-part-time-data-science/unit-1_project) | **Stats:** <https://www.youtube.com/playlist?list=PLeL7qmseLZyYWMDmqK-5Byt7_O0hF3Otj>  **Vectors/Matrices:**  <https://www.youtube.com/playlist?list=PLeL7qmseLZyZZJUb6U8O3Wjf1toqXBAvU>  **NumPy:**  [https://www.youtube.com/playlist?list=PLeL7qmseLZyYo2UrbDvaRzgF--HISGXN](https://www.youtube.com/playlist?list=PLeL7qmseLZyYo2UrbDvaRzgF--HISGXNr) |
| 7 - Experiments & Hypothesis Testing | 03/03/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/2osyKKqgrjhLGNLO4x-GS6caLorLaaa81ylMrvALxBlKo8_Q28PhXTbbMe3MU9yH) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/experiments-hypothesis-tests) |  | **Experiments & Hypothesis Testing**  <https://www.youtube.com/playlist?list=PLeL7qmseLZya2DHqjWGNLEkU6o8PDf-ri> |
| 8 - Data Visualization in Python | 03/05/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/6dBaLaP_-mdIcpWR6XDYAKgzGZvHX6a8gylK-PMNmYyy2cwS3H3j5e_-tUxrNpc) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/data-visualization-in-python/) |  | **Matplotlib and Seaborn**  <https://www.youtube.com/playlist?list=PLeL7qmseLZyZ8pJ58cLSKq0taxwgcBkAU> |
| **Unit 3: Predictive Modeling**  Learn to build models that can predict a target variable from a given set of features in tabular data. | | | | | |
| 9 - Linear Regression | 03/10/20 | [Link to the recording](https://generalassembly.zoom.us/rec/play/u5Z5I-z5qDg3HtCWuQSDC6cqW9W8fK6s0nMZ_fYMyBq3AiUBOwGhM7JAZOZ5cFn3ow5w7YH3iTB2W9fx?continueMode=true) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/linear-regression) |  | **Linear Regression**  <https://www.youtube.com/playlist?list=PLeL7qmseLZybxNe7GVXY2b4wPNq1gXJac> |
| 10 - Train-Test Split & Bias-Variance | 03/12/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/7p17dZau9kFIGafpzm2EVo8kMo73T6a82ikZrvQNyElm5c6rrde4WQCtHfpeNwe-) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/train-test-split-and-bias-variance) | [**Unit 2 Project:** EDA](https://git.generalassemb.ly/bronshtein-part-time-data-science/unit-2_project) | **Train/Test/Split:** <https://www.youtube.com/playlist?list=PLeL7qmseLZyZvqGzp2cYtR0YLvJDRxY5A>  **Bias/Variance:** [https://www.youtube.com/playlist?list=PLeL7qmseLZyae](https://www.youtube.com/playlist?list=PLeL7qmseLZyaeCdLQc7zJvMBCxJ14hQKx) |
| 11 - KNN / Classification | 03/17/20 | [Link to the recording](https://generalassembly.zoom.us/rec/play/7MUvcuyo-js3HIDG4QSDU6AtW43sL6OshncbrqdZxBnmByYBZlqkMOMaN-EYhyol31AhRXCxiJrSdgXO?autoplay=true) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/knn-classification/) |  | **KNN and Classification**  <https://www.youtube.com/playlist?list=PLeL7qmseLZyZva8cN9WmTAiFpwDz-fAIC> [CdLQc7zJvMBCxJ14hQKx](https://www.youtube.com/playlist?list=PLeL7qmseLZyaeCdLQc7zJvMBCxJ14hQKx) |
| 12 - Logistic Regression | 03/19/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/xco2LI7hp0hLTK_w8HzZRZYwPqPOeaa80ydM8_UJnkiwx6Y_Zl5Z7sVtdn6ttCFf) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/logistic-regression/) |  | **Logistic Regression**  <https://www.youtube.com/playlist?list=PLeL7qmseLZyYbxQp-d4alOuCwVphsRg8B> |
| 13 - Decision Trees | 03/24/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/5s4oM7r752JOY5Hyq2bVR7cZMIPFT6a81yhM-6UJyE8ZoPtM38XwgA5prtPkGcNv) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/decision-trees) | [**Final Project Pt 1:** Proposal](https://git.generalassemb.ly/bronshtein-part-time-data-science/unit-4_project/blob/master/requirements.md#final-project-part-1-proposal) | **Decision Trees**  <https://www.youtube.com/playlist?list=PLeL7qmseLZybAaQzKmg3FxRjqiJNaim0y> |
| 14 - Ensembling and Random Forests | 03/26/20 | [Link to the recording](https://generalassembly.zoom.us/rec/share/weVzArbq9URLQLPM6G-FVo8HPKW9X6a80yAW8qYFzU4e_Nj_byOtsEfM2Gr5TFVJ) | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/decision-trees) |  | **Ensembling and Random Forests**  <https://www.youtube.com/playlist?list=PLeL7qmseLZyZif3V1sYwYYRpau4Tmeilw> |
| 15 - Unsupervised Learning - Clustering | 03/31/20 | Link to the recording | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/clustering) |  | **KMeans:** <https://www.youtube.com/playlist?list=PLeL7qmseLZybhFnFO6jjGukO1CuqSZ2SF>  **Hierarchical**: <https://www.youtube.com/playlist?list=PLeL7qmseLZybYt93EBA6HCglbwoYbiCrQ>  **DBSCAN**:  <https://www.youtube.com/playlist?list=PLeL7qmseLZyaX6pavpDy1g9VE7i-4DXbl> |
| **Unit 4:** **Special Topics**  Get introduced to more specialized application areas in data science. | | | | | |
| 16 - Natural Language Processing | 04/02/20 | Link to the recording | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/natural-language-processing/) | [**Final Project Pt 2:** EDA](https://git.generalassemb.ly/bronshtein-part-time-data-science/unit-4_project/blob/master/requirements.md#final-project-part-2-brief) | **NLP**  <https://www.youtube.com/playlist?list=PLeL7qmseLZyYUOazaNkAzKxIZLH-xwtg4> |
| 17 - Working With APIs | 04/07/20 | Link to the recording | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/getting-data-APIs) | [**Unit 3 Project:** Modeling](https://git.generalassemb.ly/bronshtein-part-time-data-science/unit-3_project) | **APIs**  <https://www.youtube.com/playlist?list=PLeL7qmseLZyb-WkGBfR68ilRTV1ogdlWo>  **Scraping:**  <https://www.youtube.com/playlist?list=PLeL7qmseLZyZ566w57V8UTxm31KHebJX9> |
| 18 - Time Series | 04/09/20 | Link to the recording | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/time-series/) |  |  |
| 19 - Flex, catch up lesson; What’s next? | 04/14/20 | Link to the recording | [Link to the repository](https://git.generalassemb.ly/bronshtein-part-time-data-science/whats-next) |  |  |
| 20 - Present final projects | 04/16/20 | Link | Link | [**Final Project Pt 3:** Technical Report](https://git.generalassemb.ly/bronshtein-part-time-data-science/unit-4_project/blob/master/requirements.md#final-project-part-3-technical-notebook)  [**Final Project Pt 4:** Presentation](https://git.generalassemb.ly/bronshtein-part-time-data-science/unit-4_project/blob/master/requirements.md#final-project-part-4-presentation) |  |

# 