GENERAL ASSEMBLY

# HELLO!


Yeo Wee Kiang

DSI Instructor

Adriel Ho

Instructional Assistant

Iana Pyrogova

Instructional Assistant

# INTRODUCTION

# UNIT STRUCTURE https://git.generalassemb.ly/dsi-sg-01/classes

| Unit | Title | Topics Covered | Duration |
|---|---|---|---|
| Unit 1 | Data Science Foundations | Fundamental Statistics, Databases, Python, & other Data Science Tools. | Week 1 |
| Unit 2 | Exploratory Data Analysis | SQL, Data cleaning, Transformation, & Visualization | Weeks 2-3 |
| Unit 3 | Classical Statistical Models | Regression, Classification, & KNN | Weeks 4-6 |
| Unit 4 | Machine Learning Models | Unsupervised Learning, CARTS, Naive Bayes, ARIMA, & NLP | Weeks 7-9 |
| Unit 5 | Advanced Topics | Bayesian Modeling, Split-Testing, & Deep Learning | Weeks 10-12 |

# INTRODUCTION

# PROJECTS

https://git.generalassemb.ly/dsi-sg-01/projects

| Projects | Deadline | Description | Requirements |
|---|---|---|---|
| Unit Project 1 | 4th Sept 2017, Monday, 6am | Python + Pokemon Data | Students will demonstrate command of basic Python data structures, implement Python syntax for control log functions, and calculate descriptive statistics. |
| Unit Project 2 | 15th Sept 2017, Friday, 6am | EDA + Double Datasets | Students will analyze SAT and drug use datasets & explicitly state their assumptions. They will form hypotheses and justify them with solid statistical testing in NumPy, while visualizing and interpreting their plots using Matplotlib and Seaborn. |

# INTRODUCTION

# PROJECTS

https://git.generalassemb.ly/dsi-sg-01/projects

| Projects | Deadline | Description | Requirements |
|---|---|---|---|
| Unit Project 3 | 6th Oct 2017, Friday, 6am | Regression + Housing Data | Students will review the Ames housing dataset & explicitly state their assumptions. Then they will clean, impute, and explore the dataset, justifying their modeling choices and any feature engineering. They will need to plot, visualize, and interpret their data while clearly outlining their modeling strategy in response to the prompted questions. They will also need to explain their results to a technical audience. |
| Unit Project 4 | 27th Oct 2017, Friday, 6am | Web Scraping + Career Data | Students will scrape and prepare their own data, then create and compare at least two models for each section: an ensemble model or decision tree, and a classifier or regression model. They will document their data and models in a notebook for technical peers, while also creating a written summary for non-technical stakeholders. |

## INTRODUCTION

# CAPSTONE PROJECT

https://git.generalassemb.ly/dsi-sg-01/projects/tree/master/project-capstone

| Stage | Deadline | Description | Requirements |
|---|---|---|---|
| 1 | *Pitch*: 22nd Sept 2017 | Student's Choice of Model + Data | •The capstone project is a culmination of the student's work and serves as their Portfolio project.<br>•Last 4 weeks of the course to work on capstone project. |
| 2 | *Dataset*: 13th October 2017 | | |
| 3 | *EDA*: 20th Oct 2017 | | |
| 4 | *Findings*:10th Nov 2017 | | |
| 5 | *Presentation*:17th Nov 2017 | | |

# SUMMARY OF DEADLINES

https://git.generalassemb.ly/dsi-sg-01/projects

| Projects | Deadlines |
|---|---|
| Unit Project 1 | 4th Sept 2017, Monday, 6am |
| Unit Project 2 | 15th Sept 2017, Friday, 6am |
| CAPSTONE Stage 1 (*Pitch*) | 22nd Sept 2017, Friday |
| Unit Project 3 | 6th Oct 2017, Friday, 6am |
| CAPSTONE Stage 2 (*Dataset*) | 13th October 2017, Friday |
| CAPSTONE Stage 3 (*EDA*) | 20th Oct 2017, Friday |
| Unit Project 4 | 27th Oct 2017, Friday, 6am |
| CAPSTONE Stage 4 (*Findings*) | 10th Nov 2017, Friday |
| CAPSTONE Stage 5 (*Presentation*) | 17th Nov 2017, Friday |

# COMMUNICATION

- *Primary* means of communication:
  - **Slack**: https://ga-students.slack.com/messages/sg-dsi-1
  - **One-on-one**
    - Please approach any one of the Instructional team directly.
- Other means of communication:
  - Email
    - Yeo Wee Kiang, wee.kiang@generalassemb.ly
    - Adriel Ho, adriel.ho@generalassemb.ly
    - Iana Pyrogova, iana.pyrogova@generalassemb.ly

# INSTALLFEST

# WHAT TO INSTALL TODAY

https://git.generalassemb.ly/dsi-sg-01/materials-installfest

| Mac Users | Windows Users |
|---|---|
| Set up **GitHub.com** & **GitHub Enterprise** accounts | Set up **GitHub.com** & **GitHub Enterprise** accounts |
| Install **Anaconda (Python 2.7)** | Install **Anaconda (Python 2.7)** |
| Install **Git** | Install **CygWin**<br>(openssh, curl, git) |
| Install **GitHub Desktop** | Install **GitHub Desktop** |
| Install **Postgres.app** | Install **PostgreSQL** |
| Install **Sublime** or **Atom** | Install **Sublime** or **Atom** |