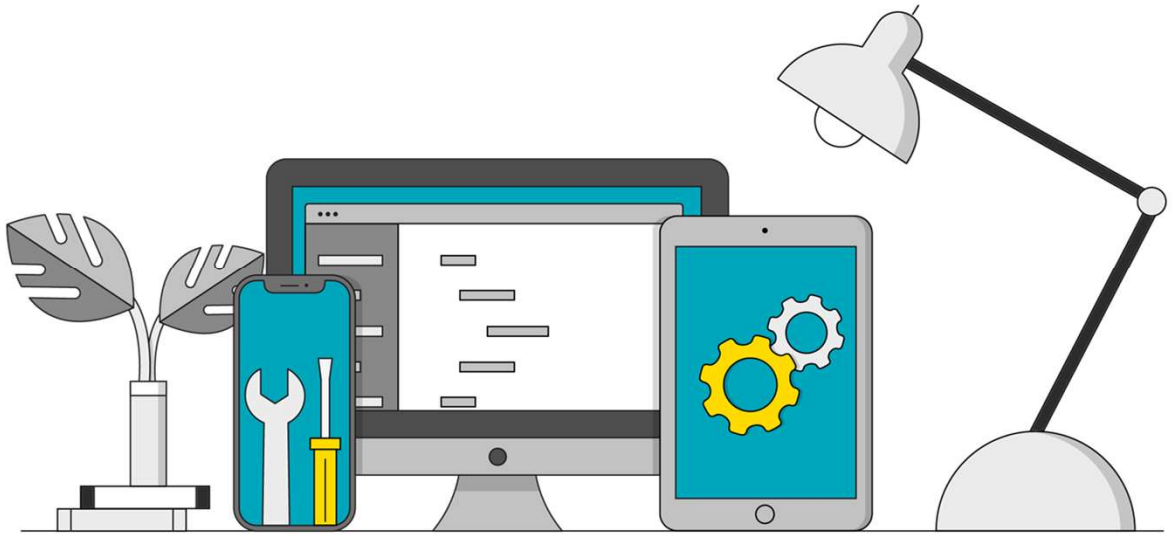
Welcome, Booz Allen Hamilton!

GENERAL ASSEMBLY

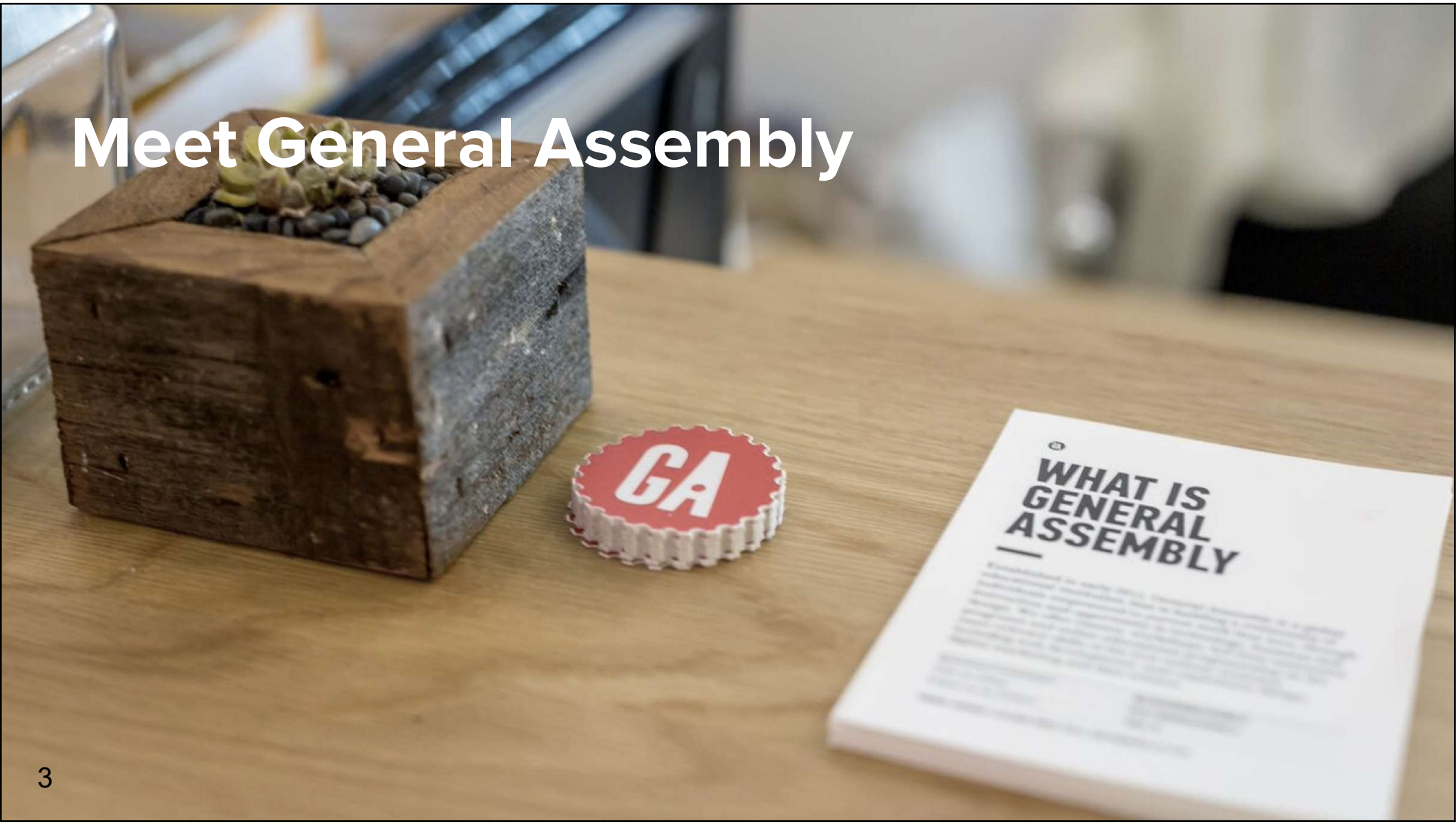
Data Science Fundamentals
Wave 11 Course Kickoff | July 2020

Agenda

- GA & Booz Allen
- Meet Your Instructors
- Program Tools
- Road to Success



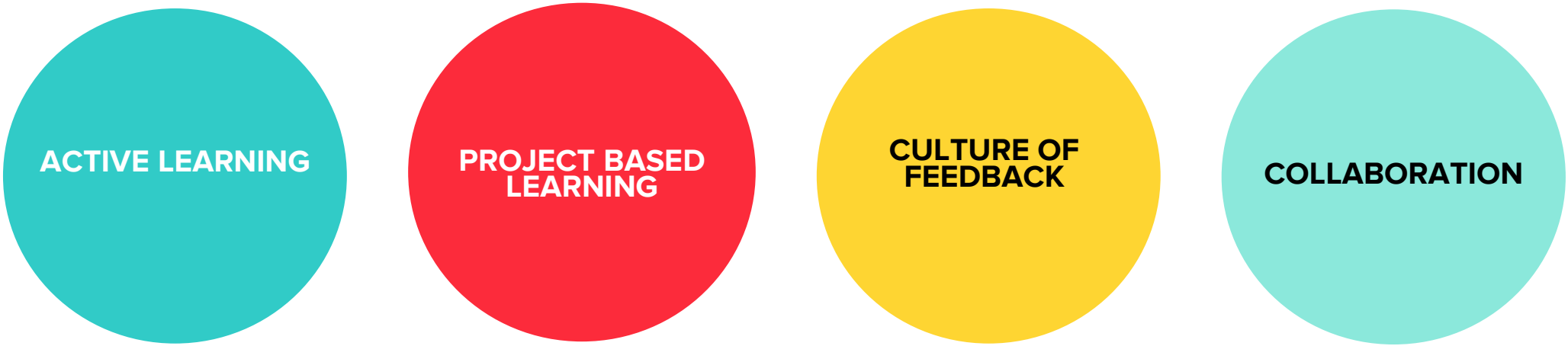
Meet General Assembly



Our Mission

We empower companies to **transform, grow and compete** through education in product, data, design, and technology.

Our Learning Approach



The Engagement Team at GA

KariAnna Eide-Lindsay
Engagement Manager

Main point-of-contact
karianna.eidelindsay@ga.co
bahsupport@ga.co

Ava Alberti
Engagement Manager

ava.alberti@ga.co
bahsupport@ga.co

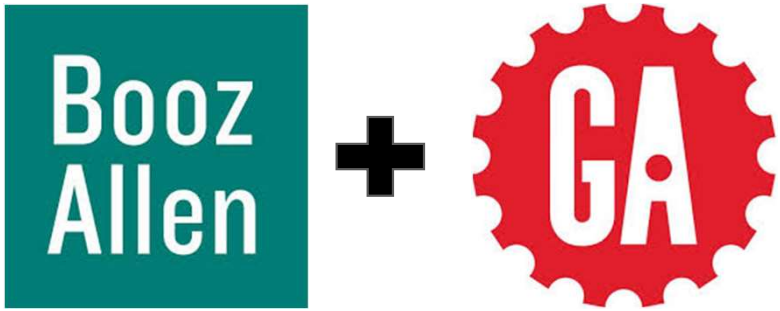
Nick Sowards
Senior Engagement Manager

Strategic direction and support

General Assembly and Booz Allen

General Assembly and Booz Allen Hamilton have partnered to deliver data courses since 2017- both Data Science Fundamentals and Data Analytics & Visualization Fundamentals.

In this time we've had over fifty-five successful cohorts and over **1,400 participants.**



Your Booz Allen Sponsors

James Hemgen
Senior Associate,
Learning & Development

Elise Picken
Lead Associate,
Learning & Development

Kadie Groh
Strategist,
Learning & Development

GA Instructional Team

GENERAL ASSEMBLY

Your Instructional Team

Preriit Souda
Lead Instructor

Ed Salinas
Instructional Associate

Alex Chicote
Instructional Associate

Program Tools

GENERAL ASSEMBLY

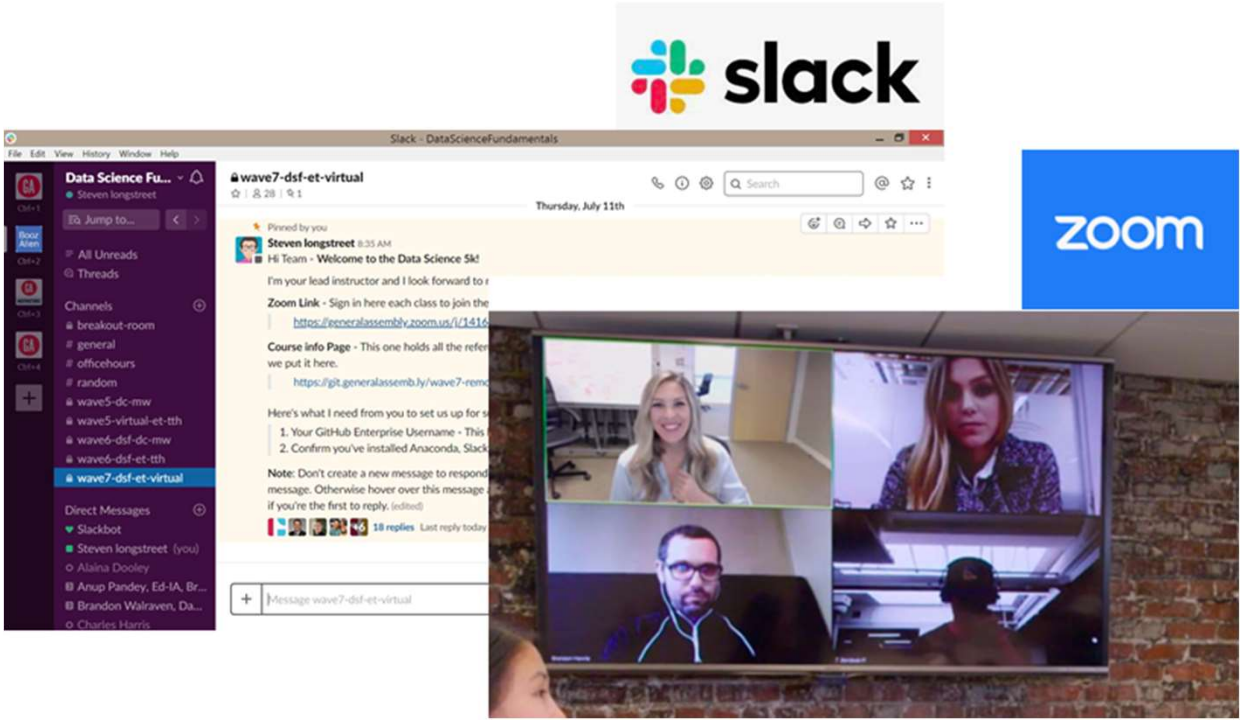
Remote Tools

Slack:

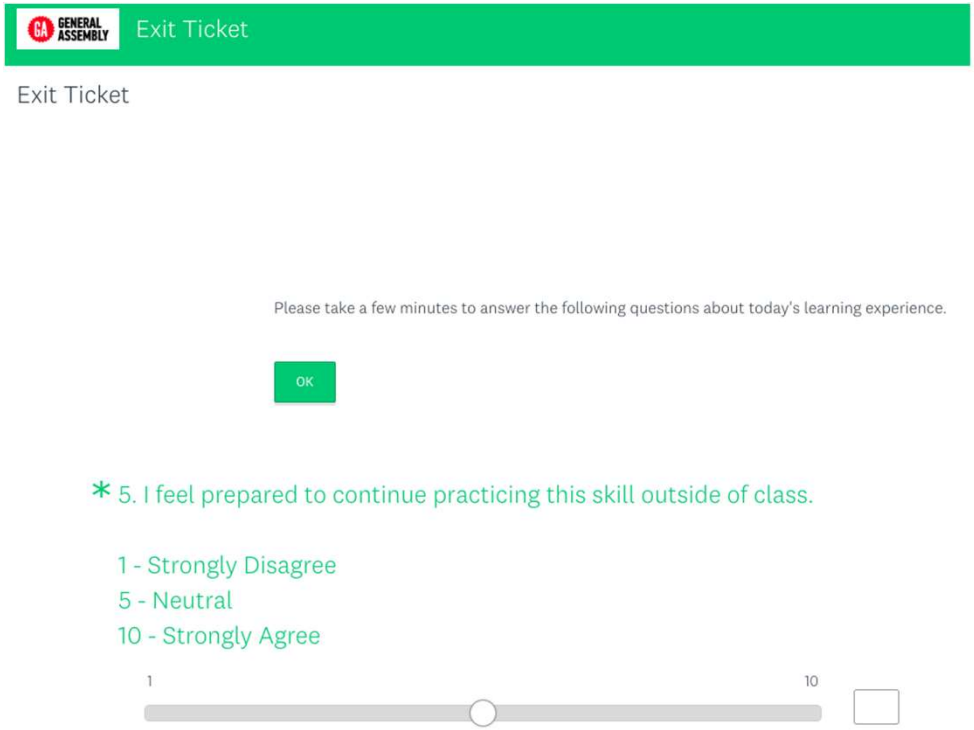
- Go to dsfundamentals.slack.com
- Log in with your Booz Allen email credentials
- If you don't see your cohort channel that begins with "wave 11," DM Kadie Groh

Zoom:

- Used for every remote session



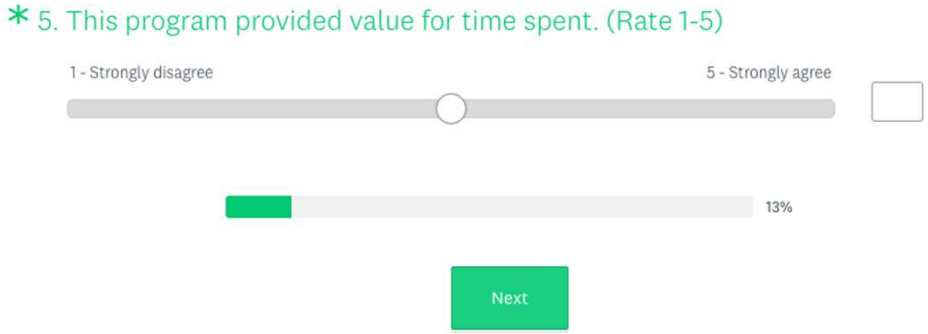
Exit Tickets



Overview:

- WHAT: “pulse check” after each class, gauging participant reaction to content and instruction
- WHY: helps GA understand impact of instruction and how/where to adapt
- TIMING: issued at the end of each class
- PLEASE NOTE: completion provides instructors with powerful visibility on where/how to support!

Program Surveys



Overview:

- WHAT: program-wide survey, covering content, instruction, and overall impact
- WHY: helps GA understand overall program effectiveness, and where to improve
- TIMING: issued at mid- and end-points of class
- PLEASE NOTE: completion is required to graduate!

Road to Success

GENERAL ASSEMBLY

Housekeeping

Be here now

Connect and collaborate!
No email/phone.

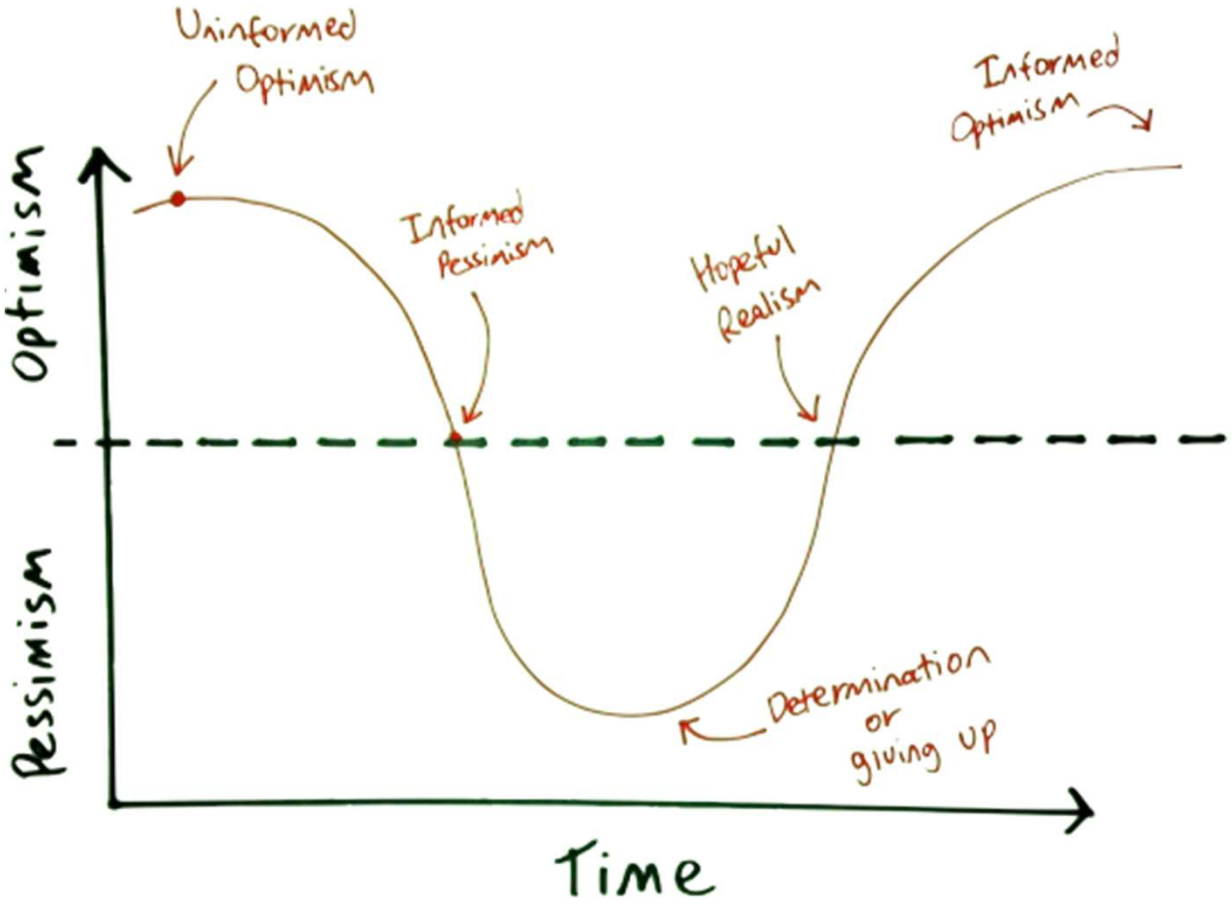
Take Notes

Take notes and ask questions!

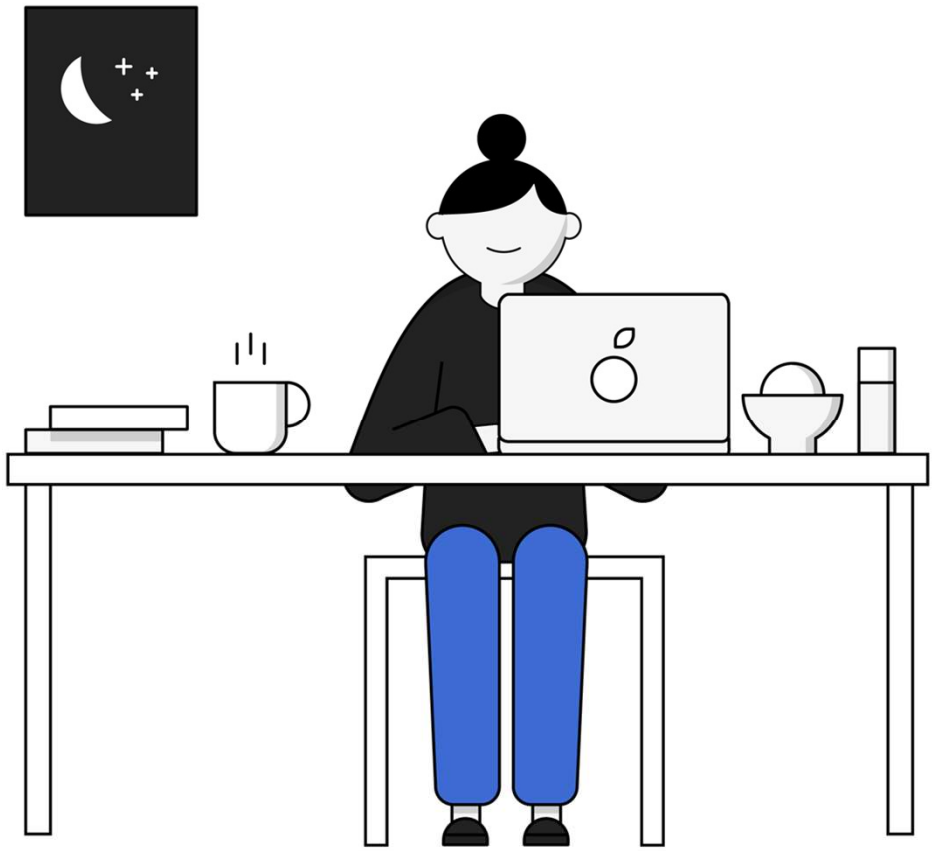
EOD Survey

Your feedback makes us better.

Planning Ahead--The “Emotional Cycle of Change”



Feedback & Support



There are multiple ways you can get support from your instructional team:

- Instructional Team: Asking questions in class (come off mute, raise your hand, or post in the Slack channel)
- Instructional Team: DM your IA to set up 1:1 office hours
- Slack: info, help, Q&A, dialogue, etc.
- Exit Ticket (after each class)

Housekeeping

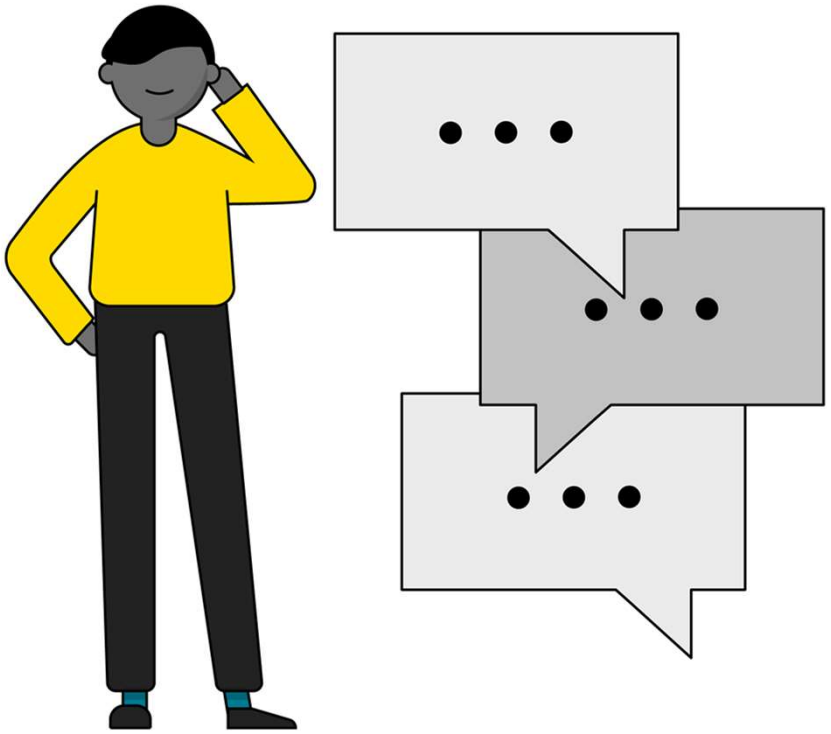
ATTENDANCE & ENGAGEMENT
(miss no more than 2 classes)

SATISFACTORY PROJECT WORK
(pass 80% or better on all course projects)

PROGRAM SURVEYS & REASSESSMENT
(must complete all to "graduate")

CERTIFICATE OF COMPLETION
(sent by GA after course concludes)

Ask Away!



Introduce Yourself!

1. Your background- Education, Work experience, Location, any/type of analytics done in the past.
2. What brought you to this course? Your expectations?
3. Something you like about the place you are presently at
4. If you know any type of magic, this is the time to tell about it

Data Science

Course Overview

Course Overview

In this program, 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.

What you will learn

- **Statistical Analysis with Python**:
Perform visual and statistical analysis on data using Python and its associated libraries and tools.

- **Data-Driven Decision-Making**:
Define and determine the trade-offs involving feature selection, model accuracy, and data quality.

- **Machine Learning & Modeling Techniques**:
Explore supervised learning techniques, including classification, regression, and decision trees.

- **Visualizations & Presentations**:
Create visualizations and interactive notebooks to present to industry stakeholders.

The curriculum materials for this course are written in Python 3.6.

Curriculum Stucture

Unit	Title	Topics Covered	Length
Unit 1	Data Foundations	Python Syntax, Development Environment	Lessons 1-4
Unit 2	Working with Data	Stats Review, Visualization, & EDA	Lessons 5-9
Unit 3	Data Science Modeling	Regression, Classification, & KNN	Lessons 10-14
Unit 4	Data Science Applications	Decision Trees, NLP, & Flex Topics	Lessons 15-19

Lesson Schedule

Date	Lesson	Unit Number	Session/Video	PaS$w0rD
Monday, July 13, 2020	Welcome to Data Science	Unit 1	Session 1	---
Wednesday, July 15, 2020	Your Development Environment	Unit 1	Session 2	---
Monday, July 20, 2020	Python Foundations	Unit 1	Session 3	---
Wednesday, July 22, 2020	Project Workshop 1	Unit 1	Session 4	
---	---	---	---	

Lesson Schedule (Unit 2)

Date	Lesson	Unit Number	Session/Video	PaS$w0rD
Monday, July 27, 2020	Exploratory Data Analysis in Pandas	Unit 2	Session 5	---
Wednesday, July 29, 2020	Data Visualization in Python	Unit 2	Session 6	---
Monday, August 3, 2020	Experiments & Hypothesis Testing	Unit 2	Session 7	---
Wednesday, August 5, 2020	Statistics in Python	Unit 2	Session 9	---
Monday, August 10, 2020	Project Workshop 2	Unit 2	Session 8	---
---	---	---	---	

Lesson Schedule (Unit 3)

Date	Lesson	Unit Number	Session/Video	PaS$w0rD
Wednesday, August 12, 2020	Linear Regression	Unit 3	Session 10	---
Monday, August 17, 2020	Train-Test Split & Bias-Variance	Unit 3	Session 11	---
Wednesday, August 19, 2020	KNN / Classification	Unit 3	Session 12	---
Monday, August 24, 2020	Logistic Regression	Unit 3	Session 13	---
Wednesday, August 26, 2020	Clustering	Unit 3	Session 14	---
Monday, August 31, 2020	Project Workshop 3	Unit 3	Session 15	
---	---	---	---	

Lesson Schedule (Unit 4)

Date	Lesson	Unit Number	Session/Video	PaS$w0rD
Wednesday, September 2, 2020	Intro to Time Series	Unit 4	Session 16	---
Monday, September 7, 2020	Holiday (Labor Day)			
Wednesday, September 9, 2020	Intro to Natural Language Processing	Unit 4	Session 17	---
	FLEX: TBD	Unit 4	Session 18	---
Wednesday, September 16, 2020	Final Project Presentations	Unit 4	Session 19	---
Monday, September 21, 2020	Final Project Presentations	Unit 4	Session 20	---

Data Science

Projects

Project Structure & Breakdown

Project Breakdown

1. Unit Project 1 (July 22): Python Technical Code Challenges
2. Unit Project 2 (August 10): Exploratory Data Analysis
3. Unit Project 3 (August 31): Modeling Practice
4. Final Project (September 16 & 21)

Unit Projects:	At the end of each Unit, you'll work on short structured projects. These activities will test your understanding of that unit's most important concepts with in-class practice and instructor support.
Final Project:	Apply your skills to a real-world or business problem of your choice. The capstone is an opportunity for you to demonstrate your new skills and tackle a pressing issue relevant to your life, industry, or organization. You'll create a hypothesis, analyze internal data, and generate a working model, prototype, solution, or recommendation.

Details on Final Project

You will get structured guidance and designated time to work throughout the course. Final project deliverables include:

- **Proposal (**End of Unit 2**)**: Describe your chosen problem and identify relevant data sets (confirming access, as needed).

- **Brief (**End of Unit 3**)**: Share a summary of your initial analysis and your next steps with your instructional team.

- **Report (**End of Unit 4**)**: Submit a cleanly formatted Jupyter notebook (or other files) documenting your code and process for technical/peer stakeholders.

- **Presentation (**Final two sessions**)**: Present a summary of your business problem, approach, and recommendation to an audience of non-technical executive stakeholders.

Data Science

Technology Requirements

Technology Requirements

Hardware:

1.8GB Ram (at least)
2.10GB Free Hard Drive Space (after installing Anaconda)

Software:

Download and install Anaconda with Python 3.6

PC only:

Install Git Bash

Others:

1. Install Google Chrome or Firefox
2.Text Editor like Notepad++ or Atom
3. Slack & Zoom!!!


GA