



EMERITUS  
INSTITUTE OF MANAGEMENT

# Postgraduate Diploma in Data Science

IN COLLABORATION WITH



COLUMBIA | ENGINEERING  
EXECUTIVE EDUCATION



# Columbia, Applied Artificial Intelligence - Agenda

## **Introductions**

Columbia Faculty

Emeritus

Course Leaders

Students

## **Learning Journey**

Course Features

Course Calendar

Webinar Schedule

## **Course Support**

Office Hours

Support Tab

## **Course Tools**

Learning Platform: Canvas

Vocareum



# Introductions - Emeritus



## Academic Delivery

Boston, US



**Tom Dougherty**  
Practice Head, Analytics



**Adriana Heredia**  
Associate Practice Head,  
Analytics



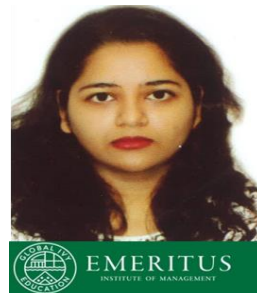
**Marie Mize**  
Teaching Support Team



**Kristen Kehr**  
Data Scientist

## Course Support

Mumbai, India



**Sabeena Merchant**  
Program Coordinator

# Introductions - Course Leaders



**Carleton Smith**

- MS Predictive Analytics, DePaul University
- Researcher, Visual Informatics & Data Analytics, DePaul University
- Data Science Instructor at General Assembly



**Eduard Valera**

- MS Electronics Eng, UPC & MS Electrical Eng Northeastern University
- Researcher, underwater communications, MIT
- Bioinformatician, Center Genomic Regulation



**James Faghmous**

- PhD CS University of Minnesota
- Researcher, techniques to predict algorithms for Big Data, University of Minnesota
- Assistant Professor, Stanford University



## Introductions – Columbia Faculty



**Hardeep Johar**

Lecturer, Industrial Engineer  
& Ops Researcher

- MA Economics, Birla Institute of Technology & Science
- PhD Information Systems, Stern School of Business, NY University
- Quantitative Trader at Morgan Stanley, Credit Suisse and Deutsche Bank
- Adjunct Professor & Lecturer at Columbia



**Vineet Goyal**

Assistant Professor, Industrial  
Engineer & Ops Researcher

- PhD in Algorithms, Combinatorics and Optimization, Carnegie Mellon
- Postdoctoral Associate at MIT's Operations Research Center
- Assistant Professor at Columbia

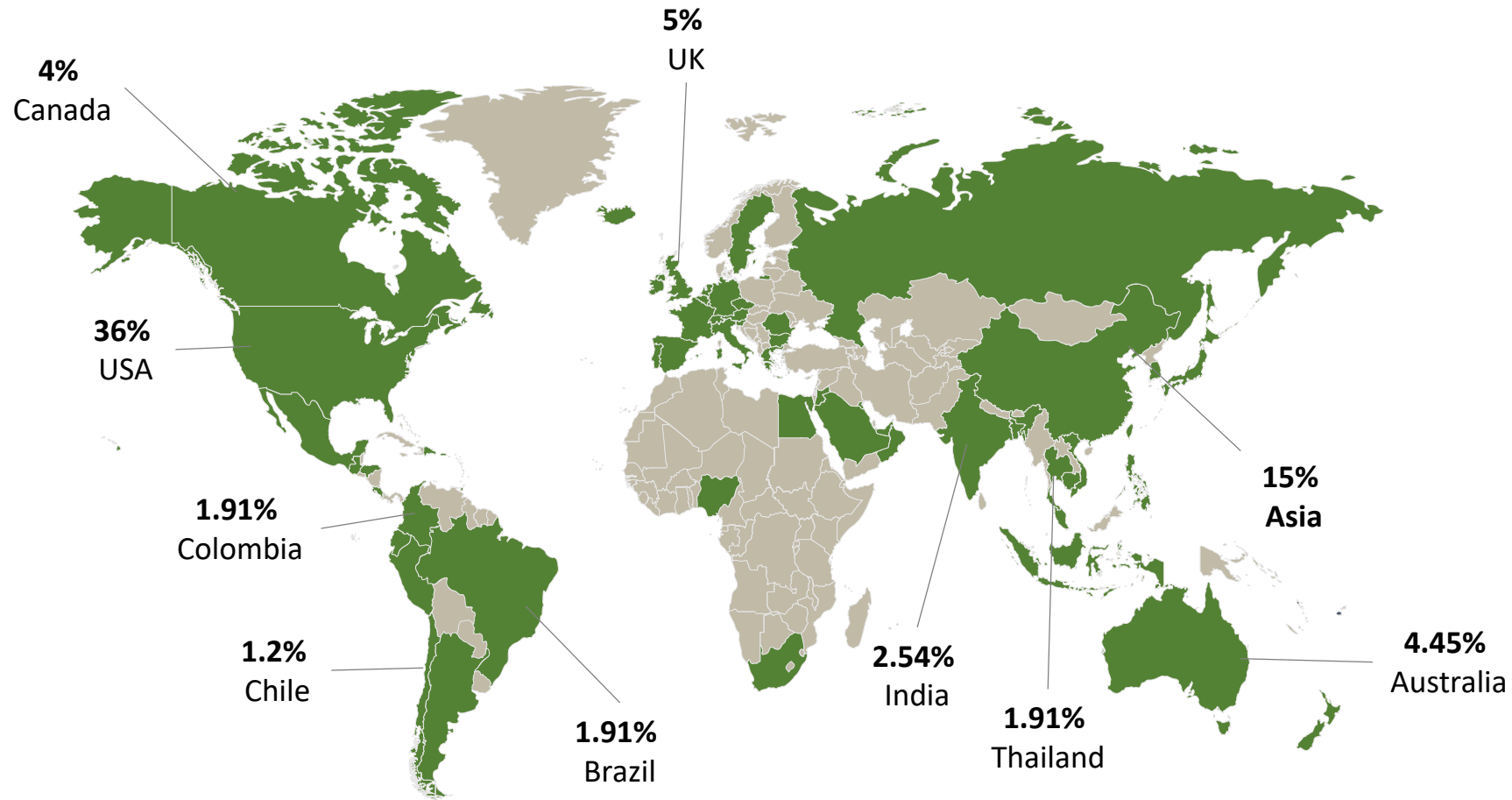


**Costis Maglaras**

Professor, Columbia Business  
School

- PhD in Electrical Engineering, Stanford
- Executive member of Columbia University's Data Science Institute
- Research on the interface of stochastic modeling and data science

# Introductions - Getting to know you



**157**  
participants

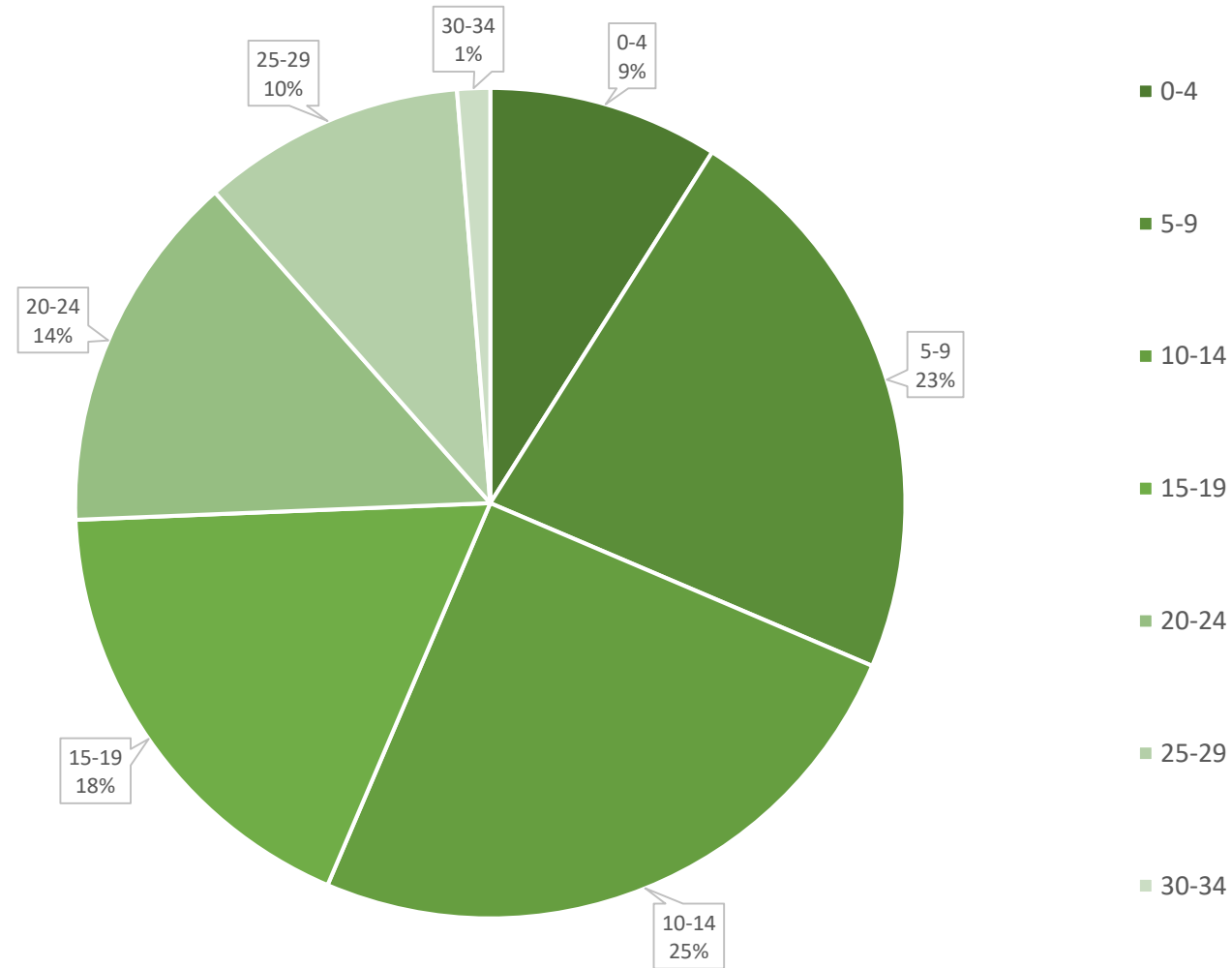


**37**  
countries

# Introductions - Getting to know you



## Work Experience



# Introductions - Getting to know you

## Industries



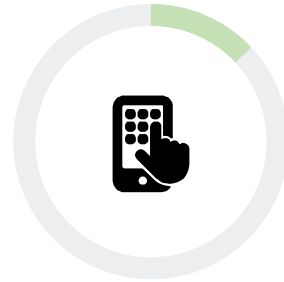
22%

Banking & Financial Services



21%

IT Services



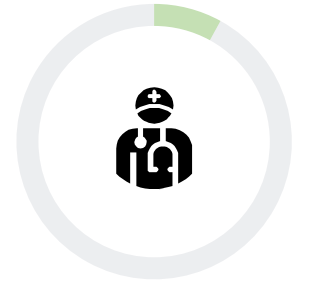
13%

IT Products



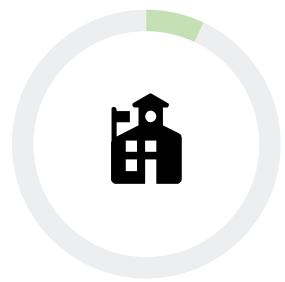
12%

Consulting



8%

Healthcare



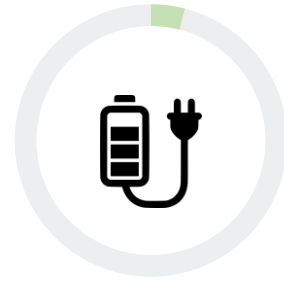
7%

Education



6%

Energy



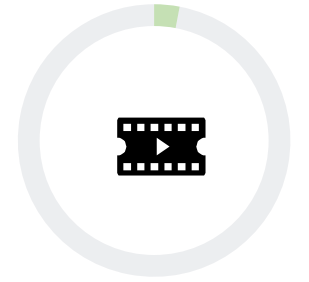
4%

Electronics / Hardware



4%

Telecommunications



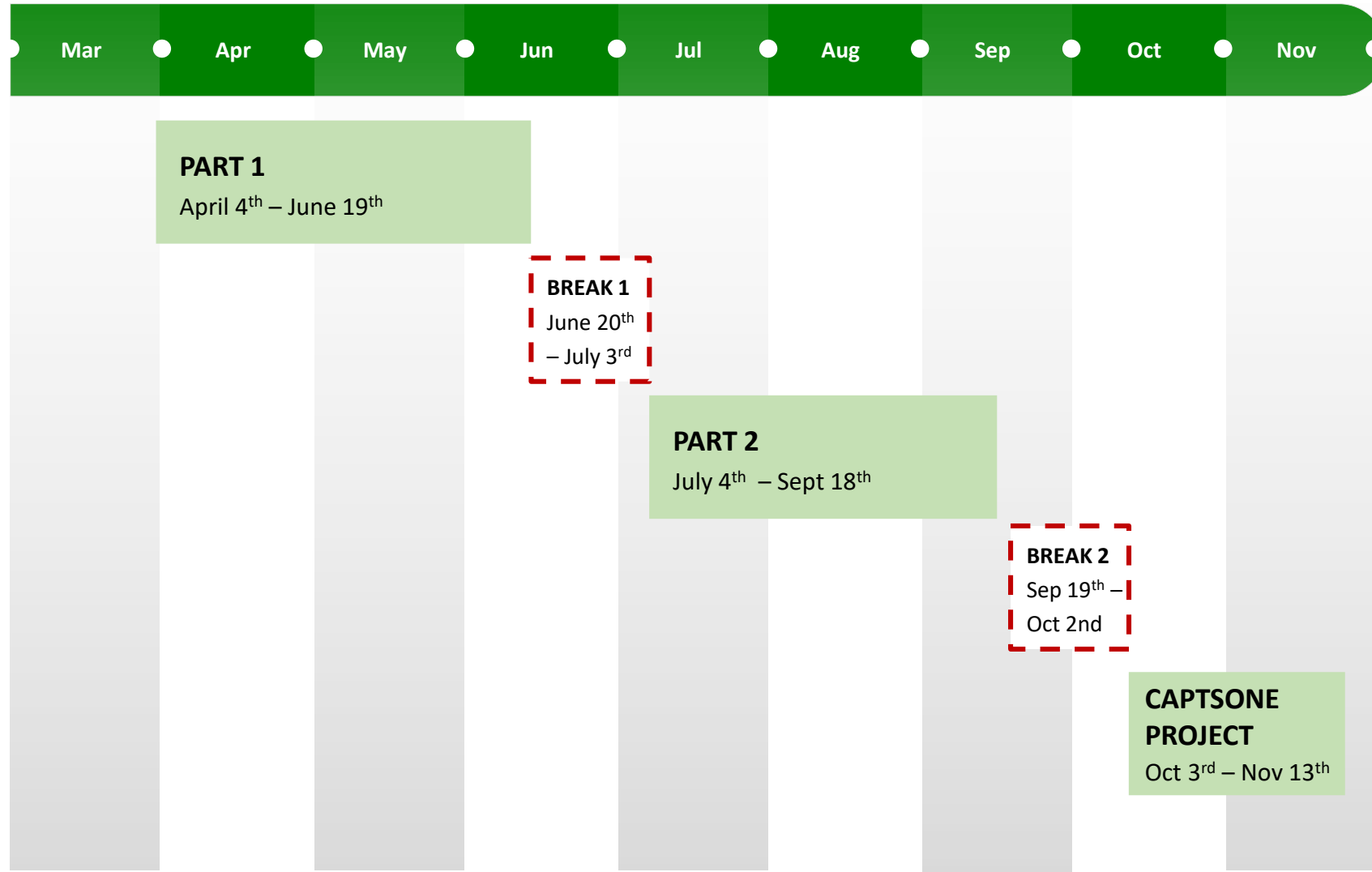
3%

Media

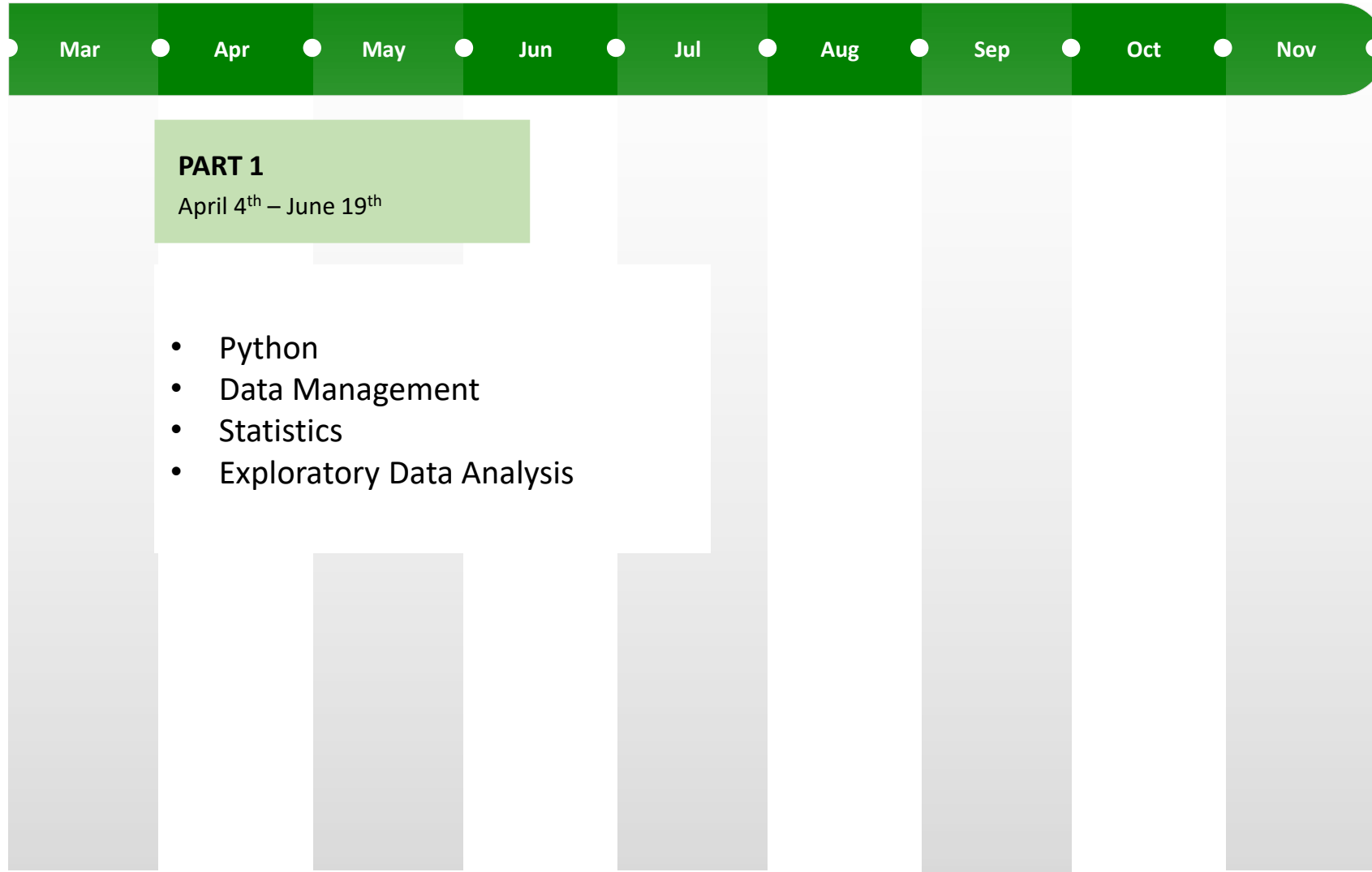




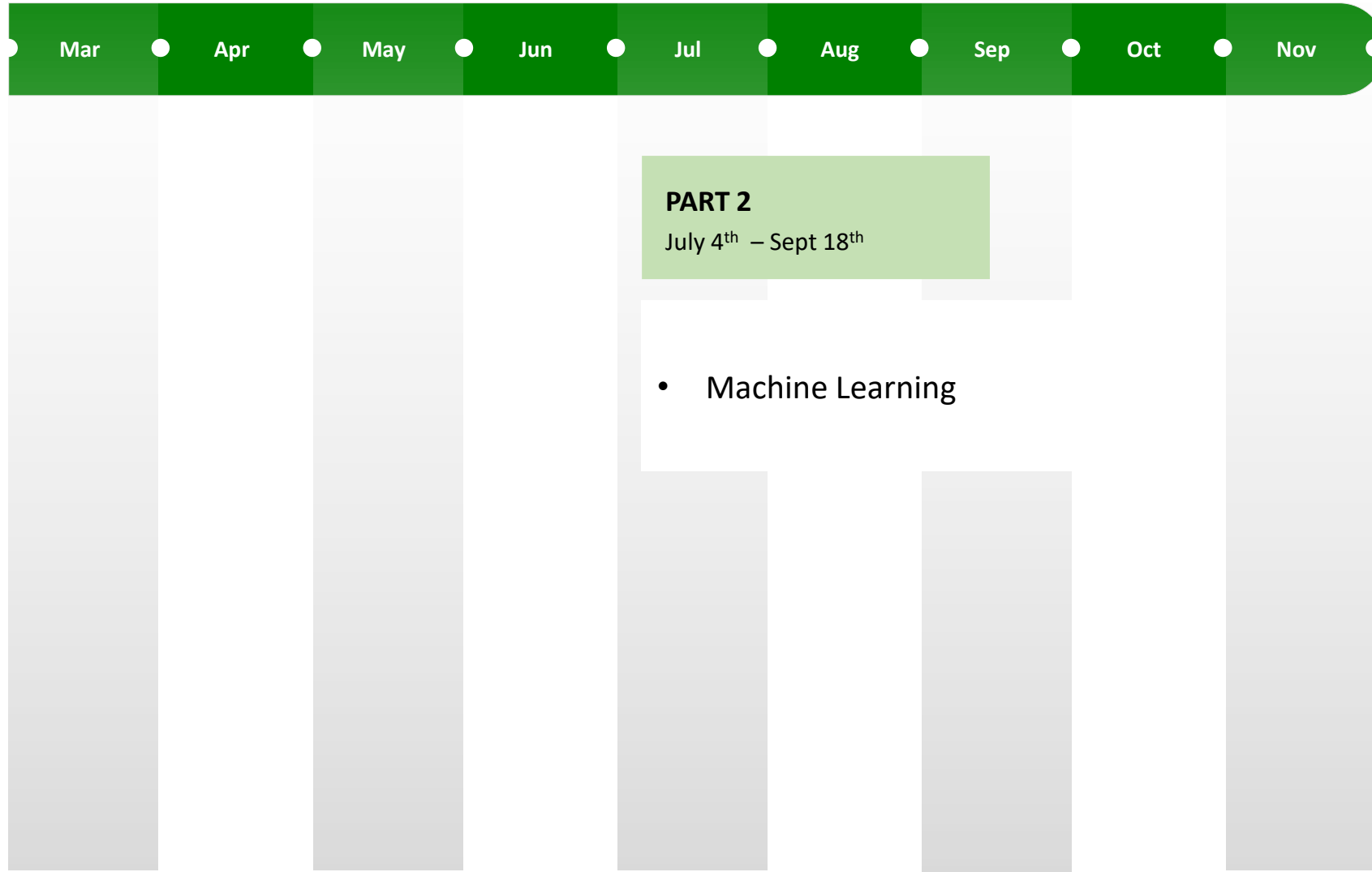
# Learning Journey - Calendar



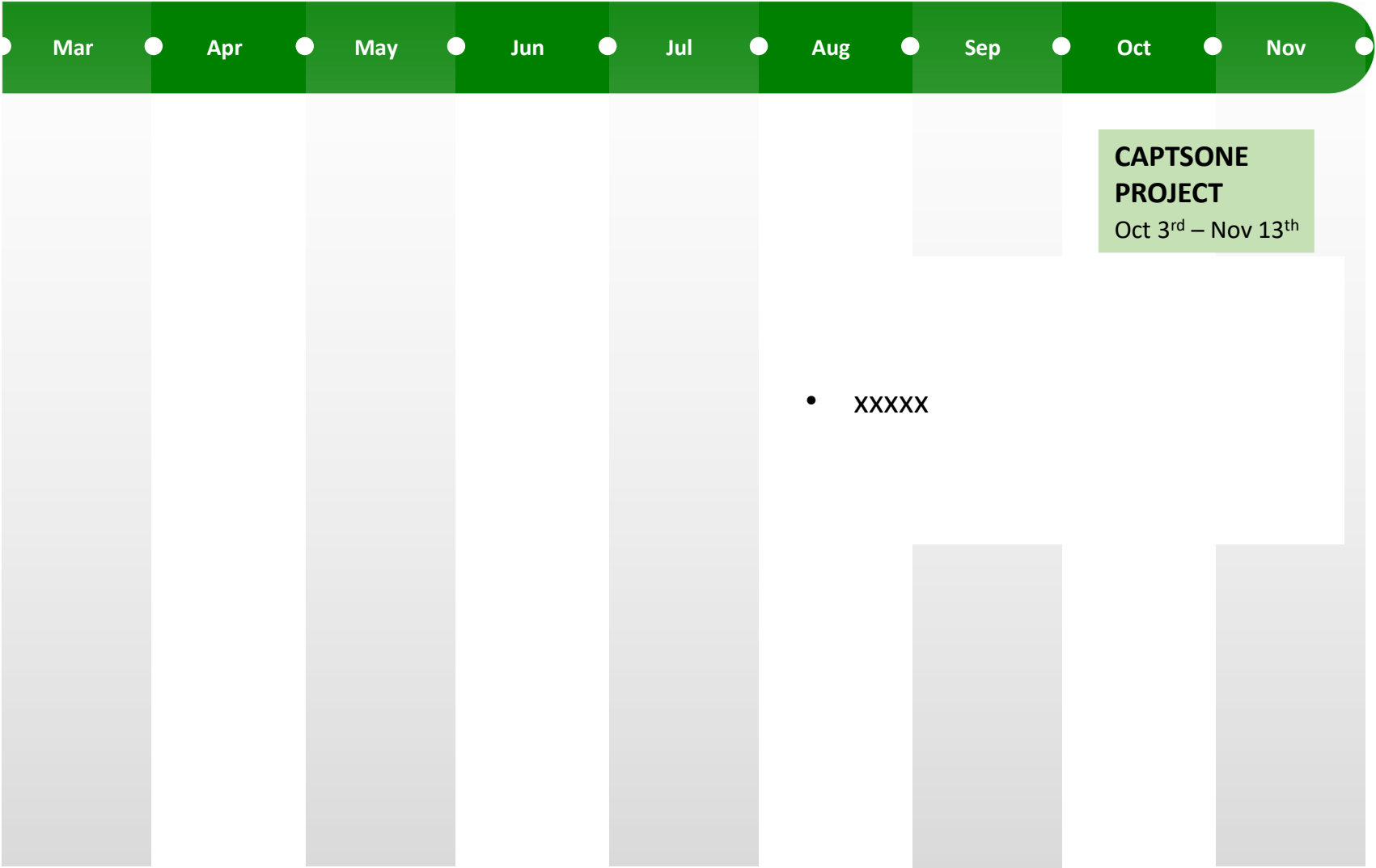
# Learning Journey - Calendar



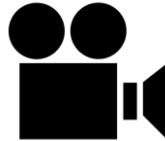
# Learning Journey - Calendar



# Learning Journey - Calendar



# Learning Journey - Course Features



## Faculty Videos

Professor from Columbia University explaining content of course.



## Assignments

Apply models and methods learnt into to real world data sets.



## Topic Webinars

Webinars on specific topics conducted by special practitioners



## Assignment Webinars

Weekly webinars to explain how to complete the assignment



## Office Hours

Extra help provided by Course Leaders on content and assignments.



## Support

Tell us your problem through Canvas and we will take care of it

# Learning Journey – Assignments

## Assignment Details

- One assignment per week.
- Total number of assignments: 22.

## Assignment Release Process

- Assignment release: every Thursday.
- Assignment deadline: Thursday, 7 days after the assignment release.
- Solution release: Thursday, 21 days after assignment release.

## Assignment Grading

- Use of Vocareum, equipped with Jupyter Notebook
- When you submit your assignment, Vocareum generates a grading report
- Unlimited opportunities to submit your code for grading
- If multiple attempt, we will select highest score for final grade

| Week #  | Assignment Name   |
|---------|---|
| Week 3  | Foundational programming concepts: variables, loops, conditions and functions             |
| Week 4  | Python data structures  |
| Week 5  | SQL introduction: operations on columns and rows  |
| Week 6  | Sorting, grouping, and joins: functions on databases                                      |
| Week 7  | Data extraction - Part 1  |
| Week 8  | Data extraction - Part 2  |
| Week 9  | Representing statistical distributions in Python: comparison of distributions             |
| Week 10 | Sampling using Python modules: advantages of sampling                                     |
| Week 11 | Obtaining quantitative estimates using hypothesis testing                                 |
| Week 12 | NumPy visualization: NumPy analysis tools   |
| Week 13 | Pandas objects, viewing data, selecting data, indexing & setting, operations using Pandas |
| Week 14 | Basic regression and classification   |
| Week 15 | Linear regression   |
| Week 16 | Logistic regression   |
| Week 17 | Decision trees and random forests   |
| Week 18 | Ensemble methods  |
| Week 19 | Naive Bayes classifiers   |
| Week 20 | Neural networks   |
| Week 21 | K-means clustering  |
| Week 22 | Dimensionality reduction with PCA and LDA   |
| Week 23 | Text mining   |
| Week 24 | Time series analysis  |





## Learning Journey – Webinar Schedule (Part 1)

| Week     | Webinars   | Date     |      |
|----------|--|----------|------|
| Week 1   | Intro Webinar  | 3/29/19  | Live |
| Week 3,4 | Intermediate & Basics Python   | TBD      |      |
| Week 5   | Relational Databases   | TBD      |      |
| Week 6   | SQL  | TBD      |      |
| Week 7   | Data Extraction – Part 1   | 05/08/19 | Live |
| Week 8   | Data Extraction – Part 2   | 05/15/19 | Live |
| Week 9   | Statistical Distributions: the shape of data                               | TBD      |      |
| Week 10  | Sampling: when you can't or won't have ALL the data                        | 05/30/19 |      |
| Week 11  | Hypothesis testing - answering questions about your data                   | 06/06/19 |      |
| Week 12  | Data analysis and visualization - using Python's NumPy for analysis        | 06/12/19 | Live |
| Week 12  | Data analysis and visualization - using Python's pandas for data wrangling | TBD      |      |



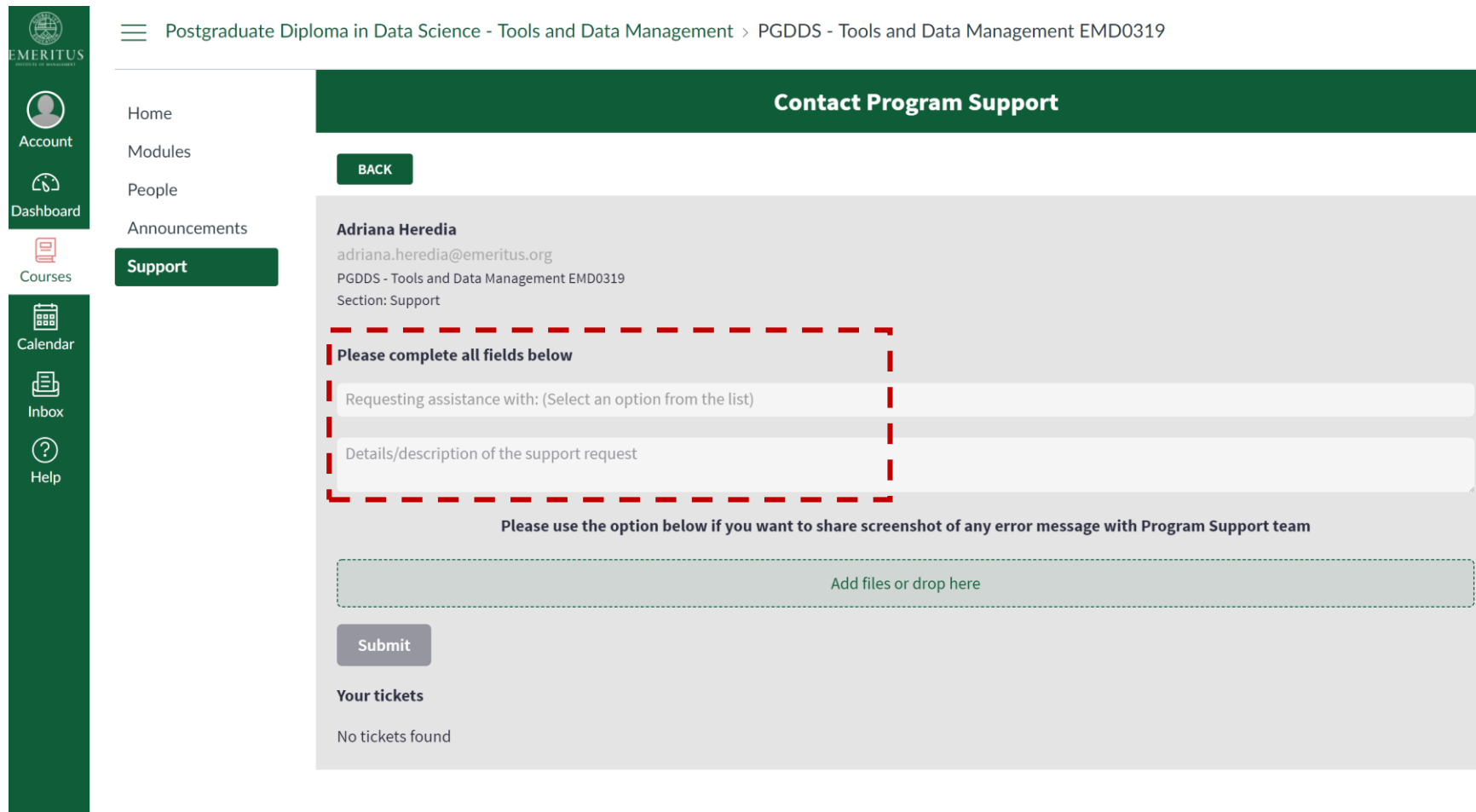
## Learning Journey – Office Hours

---

|   |                |            |              |
|---|----------------|------------|--------------|
|  | Carleton Smith | Tuesday    | 3:00 PM UTC  |
|  | Eduard Valera  | Wednesdays | 4:00 PM UTC  |
|  | James Faghmous | Wednesdays | 11:00 AM UTC |

- Attend the office hour that best works for you.
- Participation in office hours is not required. Attend with specific questions you need answered.

# Learning Journey - Support



The screenshot shows the 'Contact Program Support' page on the Emeritus platform. The left sidebar contains navigation links: Home, Modules, People, Announcements, Support (highlighted), Courses, Calendar, Inbox, and Help. The main content area has a breadcrumb trail: 'Postgraduate Diploma in Data Science - Tools and Data Management > PGDDS - Tools and Data Management EMD0319'. Below this is a 'Contact Program Support' header. A 'BACK' button is visible. The user's profile information is displayed: 'Adriana Heredia', 'adriana.heredia@emeritus.org', 'PGDDS - Tools and Data Management EMD0319', and 'Section: Support'. A red dashed box highlights the form fields: 'Please complete all fields below', 'Requesting assistance with: (Select an option from the list)', and 'Details/description of the support request'. Below the form is a note: 'Please use the option below if you want to share screenshot of any error message with Program Support team', followed by a dashed box for file upload with the text 'Add files or drop here'. A 'Submit' button is at the bottom. The 'Your tickets' section shows 'No tickets found'.

- Complete the field you need support with
- Describe what is your problem



The right team will assist you with your problem



# Program Grading & Passing Criteria

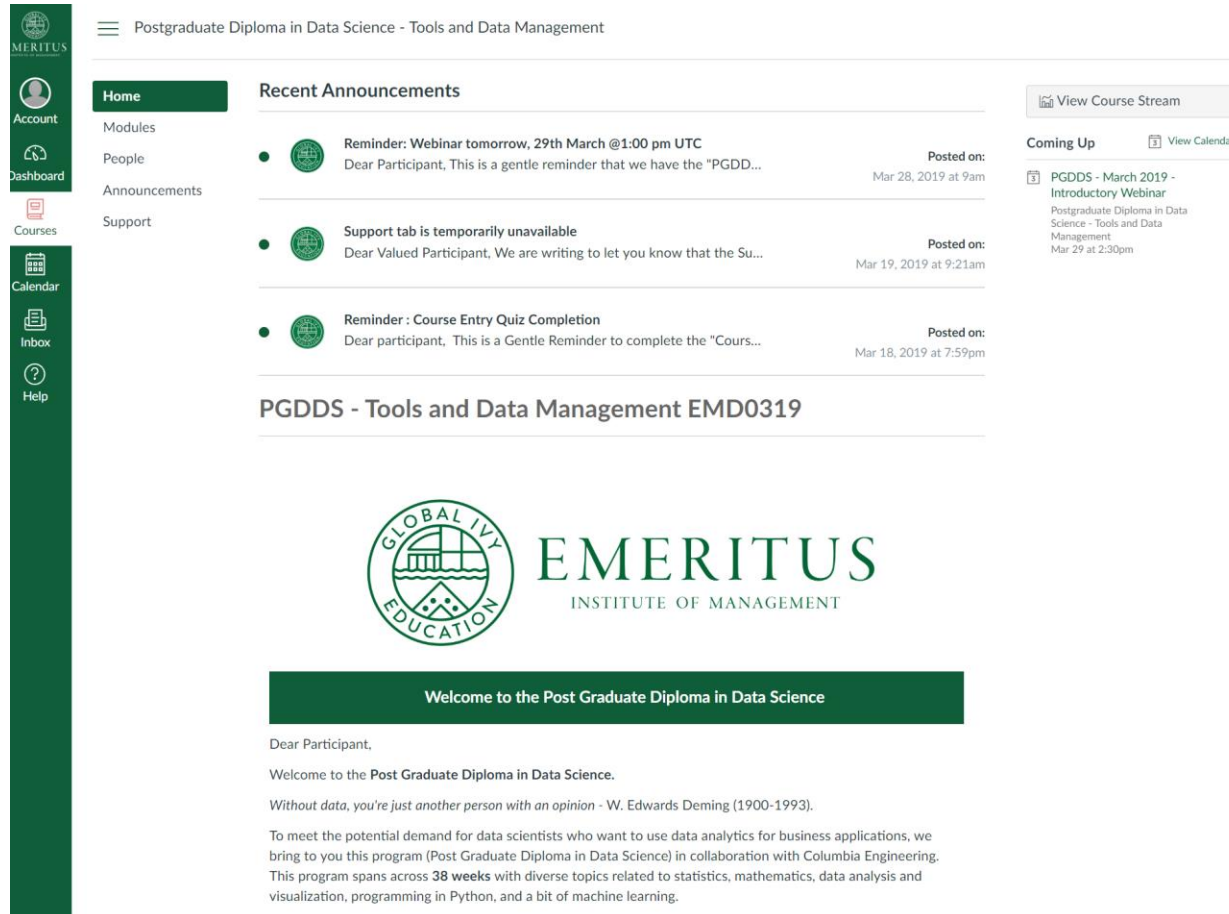
## Program Grading

- GPA scale, that runs from 0-10
- GPA is calculated individually for each course – Course 1, Course 2 and Course 3
- Capstone Project: Exceeds Expectations; Meets Expectations; Below Expectations

## Passing Criteria

- Distinctions: awarded to the top 25% of the cohort
- Pass: awarded to students who receive a GPA of 5 or higher
- Low Pass: awarded to students who receive a GPA of 4 or higher but lower than 5
- Fail: Students who receive GPA below 4 or do not complete the course.

# Learning Platform - Canvas



The screenshot displays the Canvas LMS interface for the 'Postgraduate Diploma in Data Science - Tools and Data Management' course. On the left is a dark green sidebar with navigation icons for Home, Account, Dashboard, Courses, Calendar, Inbox, and Help. The main content area has a top navigation bar with a hamburger menu and the course title. Below this is a 'Recent Announcements' section with three items: a webinar reminder for March 29th, a support tab unavailability notice, and a course entry quiz completion reminder. To the right of the announcements are buttons for 'View Course Stream' and 'View Calendar'. Below the announcements is a section for 'PGDDS - Tools and Data Management EMD0319' featuring the Emeritus logo and a welcome message. The welcome message includes a quote by W. Edwards Deming and a description of the 38-week program in data science.

Postgraduate Diploma in Data Science - Tools and Data Management

**Home**

Modules

People


Announcements

Support

**Recent Announcements**

- Reminder: Webinar tomorrow, 29th March @1:00 pm UTC**  
Dear Participant, This is a gentle reminder that we have the "PGDD...  
Posted on: Mar 28, 2019 at 9am
- Support tab is temporarily unavailable**  
Dear Valued Participant, We are writing to let you know that the Su...  
Posted on: Mar 19, 2019 at 9:21am
- Reminder : Course Entry Quiz Completion**  
Dear participant, This is a Gentle Reminder to complete the "Cours...  
Posted on: Mar 18, 2019 at 7:59pm

**PGDDS - Tools and Data Management EMD0319**

 **EMERITUS**  
INSTITUTE OF MANAGEMENT

**Welcome to the Post Graduate Diploma in Data Science**

Dear Participant,

Welcome to the **Post Graduate Diploma in Data Science**.

*Without data, you're just another person with an opinion - W. Edwards Deming (1900-1993).*

To meet the potential demand for data scientists who want to use data analytics for business applications, we bring to you this program (Post Graduate Diploma in Data Science) in collaboration with Columbia Engineering. This program spans across **38 weeks** with diverse topics related to statistics, mathematics, data analysis and visualization, programming in Python, and a bit of machine learning.

**Coming Up** [View Calendar](#)

**PGDDS - March 2019 - Introductory Webinar**  
Postgraduate Diploma in Data Science - Tools and Data Management  
Mar 29 at 2:30pm

- You will find all the course material: videos, assignments, webinars and schedules for office hours on Canvas.
- Use the learning platform to contact your Course Leader with any course related questions.
- Use the support tab for any technical issues.

Link to access PGDDS Course: <https://student.emeritus.org/courses/540>

## Grading Tool – Vocareum

The logo for Vocareum, featuring a stylized blue 'V' with an orange double-headed arrow above it, indicating a range or grading tool.

# Vocareum





# Q&A