EDSP Challenge Mentoring Program – Ring 2 Kickoff –

August 2022

Microsoft

Agenda

- Enterprise Data Science Program Introduction

- Data Science Challenge

- Challenge Mentoring Program
  - Introduce Mentors
  - Program Logistics
  - Calendar / Topics Schedule

- Questions & Discussion

Enterprise Data Science Program

What is the Enterprise Data Science Program?

Microsoft Enterprise Data Science Program

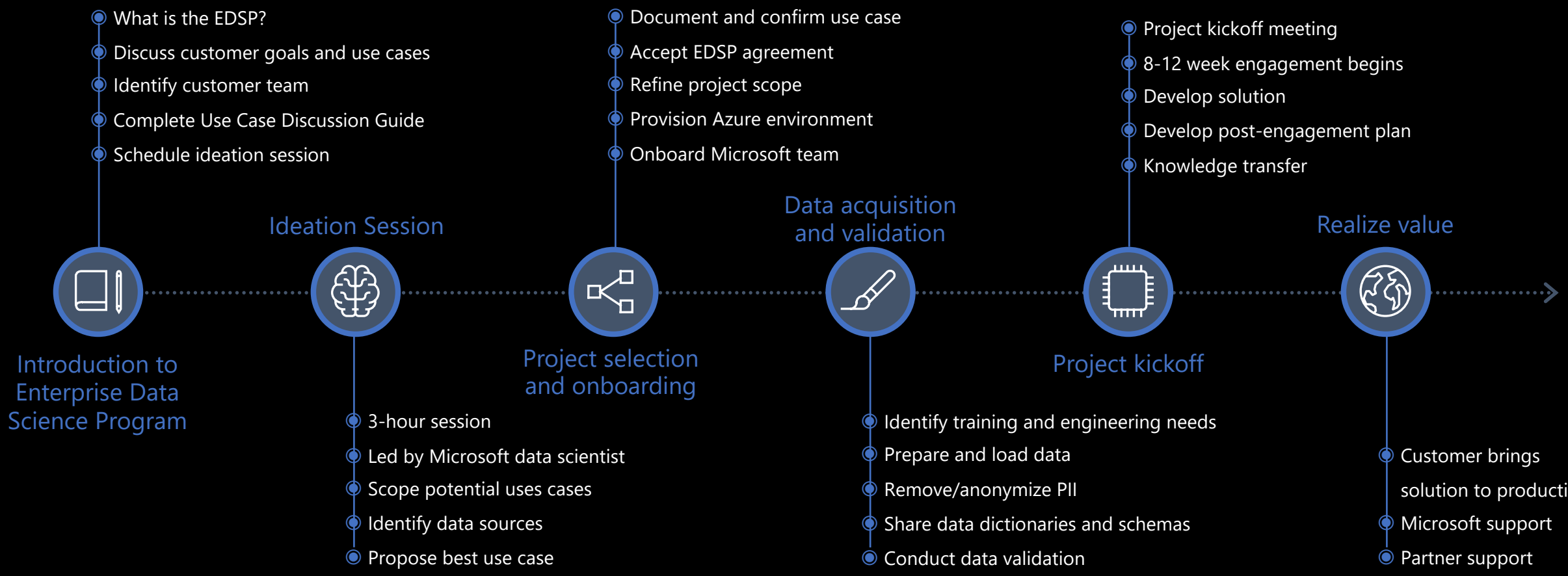
Exclusive to Our Most Valued Customers

What is the Microsoft Enterprise Data Science Program?

Microsoft is seeking to enhance its customers' Data Science footprint in the Microsoft AI Platform by leveraging a team of Microsoft Data Scientists to build Machine Learning and Artificial Intelligence models. We see this effort as a way to enhance the partnership between Microsoft and our most important customers. Because we value this partnership with our customers, we are willing to invest in this relationship by providing Data Science capability at no cost. Furthermore, at the conclusion of a Data Science engagement between this team and the customer, we will give to our customers complete ownership of all model objects created during the project. We are not seeking to build IP for ourselves with this effort; instead, we see ML and AI as such a critical workload that we are willing to dedicate staff to pushing its adoption.

With our experience in Data Science across a wide range of use cases, our team is always able to add value to each customer, whether they are just starting their Data Science practice or are already well established in the space with a full Data Science organization.

Enterprise Data Science Program Timeline



The Data Science Challenge

Microsoft Data Science Challenge

What is it?

- A Challenge that must be solved, using Python or PySpark in 24h by training a Machine Learning Model.
- It is done in a role play scenario, as if it was a real customer problem – based on a real situation.

Challenge Phases:

Phase 1	• A candidate receives the issue dataset and a business prompt. They then have 24 hours exactly to build a model using python to achieve the objectives of that prompt and to build a PowerPoint presentation detailing their work on that model.	**24 hrs.**
Phase 2a	• The day following the initial 24 hours, the candidate presents their findings to their challenge proctor in a simulated role-play scenario, including a quick walkthrough of the code.	**30 mins**
Phase 2b	• The candidate has 15 minutes to score a holdout dataset against the model that they have already built. The candidate must be ready for the holdout dataset to have differences compared with the initially provided dataset.	**15 mins**
Phase 2c	• The candidate will then have 15 minutes to receive feedback on their performance and ask any questions they may have.	**15 mins**

**Randomly generated dataset*

Microsoft Data Science Challenge

Challenge Composition

- You will receive a dataset alongside with a set of mandatory requirements. They include:
  - Bias mitigation and FairLearn.
  - Local and/or Global Explanations.
  - Deployment.

- Both the dataset and the challenges will be randomly generated.
- Independently to the final score, you will need to complete those challenges to be certificated.

Evaluation

- A Certified proctor will run the process and score the challenge using a well-defined set of criteria. (Attached).

- You will be evaluated in 4 main areas:
  - Presentation.
  - Business Understanding.
  - Computer Science.
  - Statistics.

- As you can see from the evaluation criteria, the goal is not to evaluate how good the model is, but your entire approach to solve the problem and your presentation.

Mentoring Program

Mentoring Program Objectives

- Complete (and Pass!) the Data Science Challenge
- Increase confidence in speaking with customers about data science and machine learning
- Scale the EDSP program to enable more customer projects
- Grow the community (future Proctors come from this program)

Requirements

- Attend topic kickoff sessions (or listen to recordings) every 3 weeks
- Perform self-paced study of the modules
- Attend office hours or proactively reach out to mentors for Q&A
- Sign up for the Data Science Challenge by Dec 1
- Complete the Data Science Challenge by Mar 31

Mentors

Paolo Colecchia

I'm a Data & AI Cloud Solution Architect focusing on our Azure AI portfolio. I have experience across different industries in designing AI solutions, implementing ML models and helping our customers productionise them.

Nicole Serafino

I'm a Data & AI Cloud Solution Architect in the SDP OU and co-lead the EDSP franchise for the OU. Since joining Microsoft 2 years ago, I have been working on EDSP engagements where I develop and operationalize end to end advanced analytics and AI solutions in Azure for customers across all industries. Prior to joining Microsoft, I was a Software Engineer and Data Scientist at IBM where I worked on several different Data & AI products such as SPSS and CPLEX. My educational background is in Math and Computer Science.

Mentors

Natasha Savic

AI Ranger, Cloud Solution Architect (Sydney, Australia)
Natasha is part of the Corp CSU AI Rangers Team. Prior to joining Microsoft, Natasha was a Senior Data Scientist. She was responsible for the implementation and project management of multiple Machine Learning engagements – most of them in the supervised learning area. Her technology background is in open source with Python, R & PySpark and AzureML.

Louis Li

Louis Joined Microsoft in 2014 and holds a Master of Information and Data Science (UC Berkeley) and AI Specialization Graduate Certificate (Stanford). In 2020, he published a NLP paper based on sequence to sequence T5 model translating English to SQL statements (SeqGenSQL -- A Robust Sequence Generation Model for Structured Query Language). He has also been a faculty member of W261 Machine Learning at Scale at UC Berkeley School of Information.

Mentors

Jon Tupitza

Jon is a Senior Cloud Solution Architect for Data and Analytics in the Retail and Consumer Goods OU at Microsoft. Additionally, Jon is the RCG OU franchise lead for the Enterprise Data Science Program where he is evangelizing data science and machine learning with large retailers. Jon has earned a Master of Science degree in Data Analytics from Western Governors University, master's-level certificates in data science from Harvard and Columbia, and is also a graduate of the Maryland Institute College of Art where he earned a bachelor's degree in fine art. Jon has also earned numerous industry certifications in data science, artificial intelligence, big data, business intelligence, database development & administration, software engineering, and infrastructure engineering. What's more, Jon is a Microsoft Certified Trainer (MCT), and a CompTIA CTT+ Certified Technical Trainer.

Caroline Matthews

Caroline recently joined the North America AI Ranger team in the CSU. Prior to that she was part of the Content Data Science & Engineering team at Netflix building a Knowledge Graph of entertainment information. She holds a Master of Science in Information Management and Post-Grad Certification in Data Mining from University of California San Diego. Key areas of interest include graph/network analytics, recommender systems, and expanding awareness and making AI approachable and understandable for all people.

Mentors

Sanjeev Devarapalli

Sanjeev is a Principal Data Scientist and Cloud Solutions Architect at Microsoft focused on implementing AI-driven solutions and derive insights from data using Machine Learning. Prior to joining Microsoft, Sanjeev was a lead data scientist at AT&T driving customer care insights using innovative machine learning algorithms leading to multiple patent filings. Sanjeev has a Master's degree from Virginia Tech, specializing in Software Engineering, and a Capstone Degree in Biostatistics and Medical Informatics from University of Wisconsin-Madison.

Marck Vaisman

Marck is a Sr. Cloud Solutions Architect and Data Scientist at Microsoft specializing in Data Science, Advanced Analytics, Machine Learning and Artificial Intelligence workloads on Azure. He works with customers across the U.S. Federal Government. His expertise lies in making data work for the problem at hand, drawing from experience spanning government, the commercial sector, and academia. Marck is an experienced R programmer and advocate and has been involved with the Data Science community for over 10 years. He co-founded Data Community DC, an organization that promotes Data Science and Analytics practitioners and reaches over 20,000 professionals in the D.C. Metro area. Marck is also an Adjunct Professor at Georgetown University's Masters of Analytics program and the George Washington University's Masters in Business Analytics program. Marck grew up in Caracas, Venezuela and speaks fluent Spanish.

Mentors

James Xu

James is a Cloud Solutions Architect in the Manufacturing OU at Microsoft. He holds a Masters of Science in Petroleum Engineering and a Ph.D. in Chemical Engineering from University of Texas Austin and University of California Irvine. He is a seasoned data engineer who specializes in building enterprise data pipelines and found passion in making responsible ML/AI solutions affordable and approachable to all people.

Pete Fuenfhausen

Mentors

Bala

Provide leadership on Digital Transformation with Artificial Intelligence and Cloud based Digital Solutions. Provide leadership in Machine learning, IoT and Big data and Advanced Analytical Solutions. Vast knowledge in first party Microsoft and open source technologies and Other cloud providers. Provide leadership, analysis and design tasks related to the development of a solution. Architect Cloud Based Solutions from end to end. Develop solutions viewpoint, in which the business, information and technology viewpoints are synthesized into solutions that deliver capabilities.

more

Logistics

- 64 participants

- Groups of 5-6 participants assigned to a Mentor (similar time zones, spread of skills)

- Each mentor will create a chat with his/her mentees

- Expect weekly or bi-weekly meetings with your mentor group

- Topic Leads
  - Kickoff meeting for each Topic to introduce the content
    (first one is Python on Thursday the 25th!)
  - Host Office Hours during the weeks of that topic to answer questions or discuss

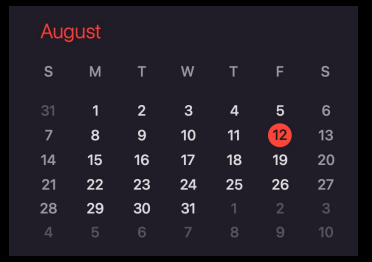
Logistics

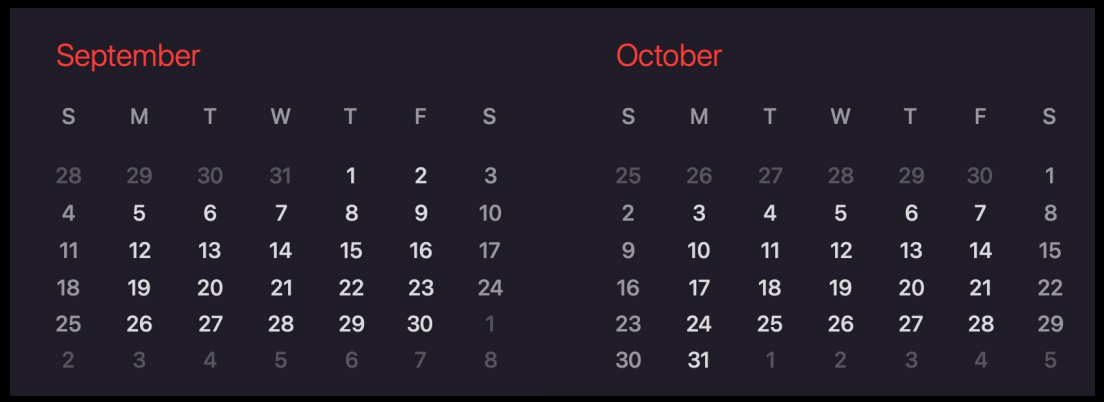
- Content is Self Guided & Self Study
  - Primary content is freely available
  - Curated by mentors to be best and most efficient coverage
  - 2-4 hours / week
  - Speak with your Mentor if you want to Accelerate

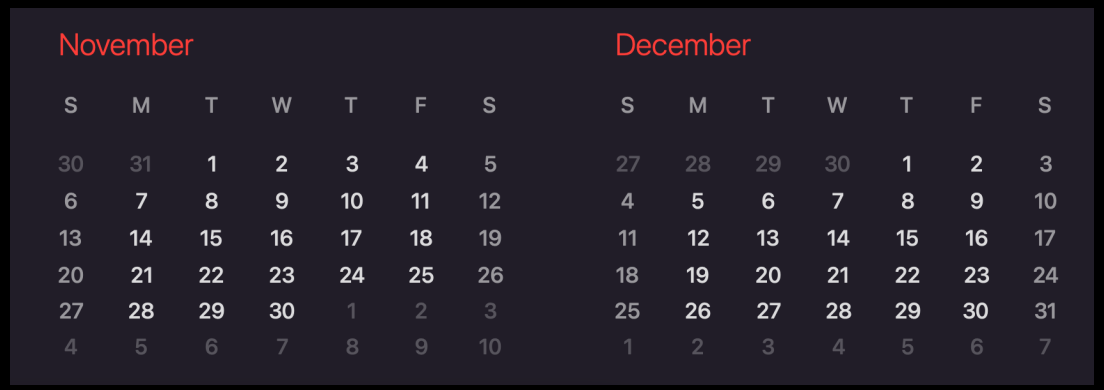
- Repo for the Program (*link to be sent shortly*)
  - Links to Self Study Materials
  - Recordings / Presentations of Topic Kickoffs

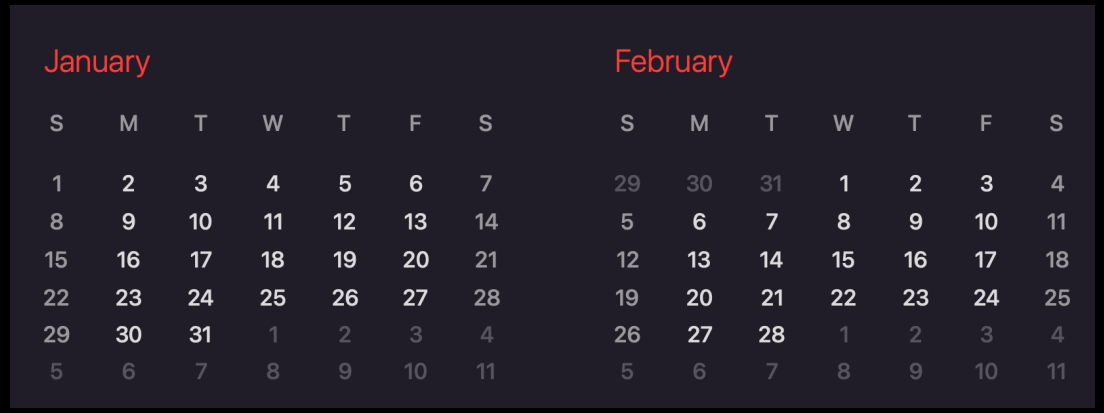
- Wrap up the program with a Sample Challenge

Mentor Programming Calendar – Ring 2 – 3 Week









Kickoff: Aug 22, 2022

Python: Aug 22 – Sep 9 (kickoff Aug 25)

Math: Sep 12 – Sep 30

Business Understanding: Oct 3 – Oct 21

Data Understanding & Preparation: Oct 24 – Nov 11

Modeling & Evaluation & Deployment: Nov 14 – Dec 2

Presentation of Analytical Results: Dec 5 – Dec 23

Responsible AI: Dec 26 – Jan 6

Sample Challenge: Jan 9 – Jan 27

MLOps: Jan 30 – Feb 17

Challenge: Must sign up by Dec 1; complete by March 31

Each module runs for 3 weeks!

Questions & Discussion