



MAY - AUGUST 2020

PARTICIPANT WELCOME PACKET



WELCOME & PROGRAM BACKGROUND

DATA SCIENCE FOR ALL / COLOMBIA

Welcome to the DS4A / Colombia 2020 program, sponsored by MinTic. We are thrilled for you to attend this world class program for data science learning and career development.

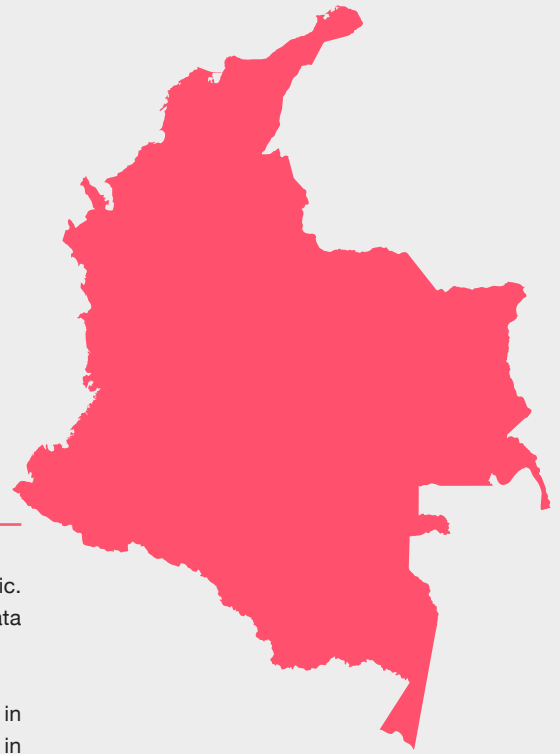
The program features the premier data science and AI training in the world. Lead Instructor Natesh Pillai of Harvard University, in combination with a Teaching Assistant team from top institutions in the US and top graduates of previous DS4A programs, will work with you to explore data science in ways unlike any other program. Our training is hyper-practical, based on business cases and real-world commercial contexts.

In addition to training, DS4A supercharges your career by matching you with actual companies and job opportunities. Throughout the program, you will work side by side with data scientists and managers from Colombia's leading companies.

Because program entry is merit-based, you will be surrounded by top talent. You will work with fellow participants who are engineers, scientists, managers, entrepreneurs, and innovators. Participants are from Colombia's leading institutions and come from a variety of industries.

We look forward to working with you over the 12 weeks of the program, as we master data science and the many ways it can impact our world. Through lectures, interactive cases, and immersive group projects, you will learn what is possible from skills like querying big data, robust linear modeling, and advanced machine learning.

In addition, you will be plugged into the region's most prestigious data science community. [For a sneak peak, see the video from our DS4A 2019 program here.](#)



/ CONTENTS

PROGRAM CALENDAR	04
CLASS REQUIREMENTS	05
OUR LEARNING PHILOSOPHY	06
CLASS STRUCTURE	07
CURRICULUM OVERVIEW	08
CERTIFICATION	09
TEACHING ASSISTANTS	09
ACCESSING CLASS MATERIAL	10
ZOOM	10
SLACK COMMUNICATION	11
THE FORUM	11
PRACTICUM PERIOD	12
PAST PARTICIPANT TESTIMONIALS	13
ABOUT CORRELATION ONE	14

12 WEEK PROGRAM

PROGRAM CALENDAR

MAY						
S	M	T	W	T	F	S
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

JUNE						
S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

CLASS 1-16

PRACTICUM PERIOD

FINALE

JULY						
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

AUGUST						
S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					



CLASS REQUIREMENTS

This class is virtual, in our “Live Online” mode. However, we simulate a live classroom throughout the program.

In this intensive program, each class counts. Participants must plan to attend each class.

Classes will begin at 8am sharp. Participants should plan to log in to their computers and be fully prepared to work at 8am. Classes end at 6pm.

Participants need their own computers for class.

Participants need to have a strong internet connection during class days.

Before the first class, participants should set up their laptops with a data analysis environment and with the proper communication tools used during the course:

Instructions for setting up your data analysis environment are on the [class website](#).

Download Slack’s desktop and mobile applications and set up your profile by following [these instructions](#).

Download Zoom to your desktop and create a profile by following [these instructions](#).

You will be given pre-work before the program starts. This is a short case that will help prepare you for the course. You are expected to complete the pre-work before the first day of class.

OUR LEARNING PHILOSOPHY

We have a learning philosophy with three core tenets:

- 1** Data Science is best learned via practical examples
 - 2** Data Science is best learned by hands-on doing, not by reading or watching
 - 3** Data Science is best learned by working collaboratively in groups, not individually
-

We operationalize this learning philosophy in our Data Science For All program. Specifically, each topic is taught via a practical case study, i.e. a practical problem that one would encounter in the real-world that needs to be solved. Furthermore, while lectures are an important part of the program, the majority of participant time will be spent learning by doing, via homework assignments, cases, problem statements, and capstone projects. Finally, participants will be assigned to a team of four to six people at the outset--this team assignment will persist throughout the program and will form the core of your peer "Project Group".



CLASS STRUCTURE

The program takes place in Correlation One's Live Online format. This format is conducted fully online, using programs like Zoom for interactive lecture broadcasts and Slack for dynamic class communication. The purpose of Live Online is to combine the world-class lectures and resources of online-only education with the personalized advising of in-person education.

Throughout the course, you will have a pair of Teaching Assistants to support you. These TAs are here to answer your questions and help you through the course, including with your Final Project. Always feel free to reach out to them if you need anything. Further details on your TAs can be found in the "Teaching Assistants" section of this document.

Here is an example of a typical class day:

7:45AM-8:00AM	8:00AM-12:30PM	12:30PM-1:30PM	1:30PM-6:00PM
Participants log on	Lecture stream	Lunch	Group work/Office Hours with Teachers/Extended Case Review/Extra Programming

LECTURES

Taught by Lead Instructor Natesh Pillai, these are interactive lectures which you will follow with a Jupyter notebook. Questions via Slack are encouraged, and will be answered by your "TA"s (Teaching Assistants) and Professor Pillai throughout the lecture.

LUNCH

Participants are free to have lunch as they choose for 1 hour between lecture and the afternoon activities.

GROUP WORK

A large part of the course is a Final Project (described in greater detail in the "Curriculum Overview" section). For your Final Project, you will work in a team of 4-6 people to solve a real-world data-science problem. There will be allotted time every class day for your Project Group to meet and work on your Final Project. Use this time to advance your project and ask your TA critical questions.

SPEAKER SERIES

Office Hours: TAs will conduct team office hours with each project team. TAs are also available for individual office hours, where you can speak with them one-on-one to ask any questions you have.

EXTENDED CASE REVIEW

TAs will review Extended Cases each Saturday, allowing time for questions. Extended Cases are discussed further below, and are the "homework" of the class.

EXTRA PROGRAMMING

Throughout the program, Correlation One will offer additional enriching opportunities to complement the curriculum. Examples include guest speakers and networking sessions.

CURRICULUM

The lectures will cover the following topics:

PRE-WORK / BASICS OF PROGRAMMING

WEEK 01 /	WEEK 02 /	WEEK 03 /
DATA SCIENCE, MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE AT A GLANCE; BASICS OF PYTHON	BASICS OF PYTHON (CONT.); DATA INTERPRETATION	CODING DATA VISUALIZATIONS; DATA INTERPRETATION (CONT.); EXPLORATORY DATA ANALYSIS
WEEK 04 /	WEEK 05 /	WEEK 06 /
EXPLORATORY DATA ANALYSIS (CONT.); DATA WRANGLING & CLEANING; AWS	SQL IN A BIG DATA WORLD; DASH	TEXT PROCESSING; DATA-DRIVEN DECISION-MAKING
WEEK 07 /	WEEK 08 /	WEEK 09 /
LINEAR MODELING - VARIATIONS & EXTENSIONS	CLASSICAL MACHINE LEARNING MODELS & CROSS-VALIDATION	REGULARIZATION; SENTIMENT ANALYSIS; DEEP LEARNING
WEEK 10 /	WEEK 11 /	WEEK 12 /
SURVEYS & EXPERIMENTS; WEB SCRAPING; CONVOLUTIONAL NEURAL NETWORKS; REINFORCEMENT LEARNING	PROJECT WORK	PROJECT WORK

Each lecture will consist of a mixture of instruction, exercises, and built in Q&A time. The program will also consist of two other crucial components:

EXTENDED CASES

These serve as the “homework assignments” of the program. They are structured as complex, multi-part business problems, and participants must leverage all of the skills they have learned from that week’s lectures in order to solve them properly. Cases are given out in the middle of the day on Fridays, and participants are expected to work on them throughout the following week, to be completed and turned in by the beginning of class on the following Friday. Each extended case is designed to take approximately 5 - 9 hours to complete.

FINAL PROJECT

Each participant will be assigned to a 4-6 person Project Group for the duration of the program. Participants will work with their groups to produce the Final Project. The Final Project is meant for trainees to apply and showcase the data analysis, modeling, engineering, and visualization skills they have developed throughout the program. Each project group can either (i) choose a real-world data science challenge submitted by a Colombian business or public entity; or (ii) create a relevant project based on their own interests.

Final projects will be showcased at the Finale at the end of the program. Winning project teams will be awarded several prizes based on research quality, presentation clarity, and impact.

CERTIFICATION

Participants enrolled in the program are eligible for certification upon successful completion of the program. To receive certification as a “Data Science For All / Colombia” graduate, participants must:

Attend at least 90% of lectures & activities

Complete at least 75% of Extended Cases with a ‘Satisfactory’ rating

Contribute sufficiently to the completion of their Practicums. At the end of the program, we will conduct a survey whereby we ask each participant to rate how each of their teammates contributed to the Practicum; this information, in combination with TA judgment, will be used in determining certification.

Program graduates can highlight their certification to their existing employers or in interviews with potential employers, and also indicates membership in Latin America’s most prestigious AI community.

TEACHING ASSISTANTS

Throughout the course, you will have two TAs overseeing your learning experience and project-related work. All of our TAs are very experienced and eager to help you learn during the program. You should contact these TAs regarding questions related to the course material or your project, and generally use them as a resource throughout the course. You will interact with your TAs in three main ways:

Slack Your TAs will have a specific Slack channel that you and other participants will use to interact with them directly. Please see the “Slack” section later on in this document for additional details.

Team Office Hours Once per each lecture day, you will meet with your TAs at a scheduled time with your project team. These meetings will be used to help keep your team on track with your project and answer any questions you may have.

Individual Office Hours During lecture days when your TAs are not meeting with teams during Team Office Hours, they are available to meet with you one-on-one. You should message your TAs on Slack through the TA-specific channel. If both of your TAs are busy, you can post in the #extra-help Slack channel -- another TA who is available will see your post and meet with you via Zoom.

ACCESSING CLASS MATERIALS

REGISTERING FOR CLASS WEBSITE

Class participants will receive a unique link by email which will allow them to register for the class site. Once you receive the email, you will be directed to a page in order to provide a password. After that point, you will be able to log in at any time into the class site at <https://colombia2.ds4a.io/> with your email and password.

CLASS MATERIALS

You will be able to access all class materials on the class website. Lectures and other programming will be recorded and will be added to the class website shortly after they occur. Class materials will be made available on the class website on the Monday or Tuesday before the corresponding lectures. On the class website, you will be able to download cases and make submissions as required.

ZOOM

Lectures, meetings with project teams, office hours, and most of the live parts of the program will occur via Zoom.

You should have the Zoom app installed and have your own Zoom account before the first day of lecture. You will need to login into Zoom with your email address to make an account (you should use the same email address you used to sign up for DS4A). [You can download Zoom here.](#)

Zoom links to lectures and other programming will be posted on the #official-announcements channel on Slack. Please make sure you have accepted the invite to join Slack before the first day of class.

SLACK COMMUNICATION

DOWNLOADING SLACK

Slack will be the primary channel of communication for program topics. Because all important program announcements will be made through Slack, all participants should download Slack's desktop and mobile applications from the links below (or app store on your phone), and enable notifications:

Mac: <https://slack.com/downloads/mac>

Windows: <https://slack.com/downloads/windows>

Android: <https://slack.com/downloads/android>

You will receive an invite to join the private slack for the training program via email. Finally, complete your profile by adding your full name, phone number, and a short blurb about yourself.

GUIDE FOR USING SLACK

We have already set up a few channels for class-wide communications, which are described below. Please review the guide before the first day of class.

#<ta names>_ta-questions (e.g. jane-george_ta-questions) This is your TA-specific channel you will use to communicate with your TAs. You will be added to this channel prior to the first day of lecture. You will have two TAs that are responsible for assisting you and a subset of other participants. If you have questions during the lecture days (both during lecture and outside of lecture) you should ask them here (or message them directly). This will also serve as a channel for you to ask project-related questions. For questions on extended cases, you should ask them on the Forum.

#official-announcements Correlation One staff and TAs will post official announcements to this channel.

#extra-help If you want to speak with a TA one-on-one but your TAs are busy meeting with project teams, you can post in this channel to request individual office hours. An available TA will see your post and set up a Zoom call to talk about your questions.

#tech-support Post any tech issues with the training materials in this channel and a TA will assist you.

#live-chat Public channel for participants to post during lecture. This channel does not have a set topic, it's purpose is for you and other participants to interact with each other and the lecturer during lecture.

#resources Public channel for all participants and TAs to post data science-related resources for each other, such as interesting articles, videos, and data sets.

#social-<city> (e.g., #social-bogota) City-specific channels for you to chat with other participants in your local area, plan social events (post-COVID), etc. Use these channels to build relationships with other participants. Feel free to create a channel for your city if one does not currently exist.

#networking Public channel for you to share your LinkedIn profile, network professionally with other participants, post about job opportunities, and discuss career-related topics.

Making New Slack Channels You are encouraged to create additional Slack channels to facilitate communication and other activities, like group projects or social gatherings. Slack channels can be related to class-material, or something casual like sharing favorite recipes or home-office perks during quarantine.

THE FORUM

The Forum, which can be found on the [class website](#), will be used to ask questions related to [extended cases](#). In addition to asking questions, you should also answer questions other participants ask on the Forum if you know the answer. This will increase not only their learning but also your understanding of the material as well. TAs will also be monitoring the Forum and responding to questions.

Please do not use Slack to ask extended case questions - you should post them on the Forum instead. The purpose of this is to make sure all participants benefit from the answer. If you post an extended case question on Slack, TAs will tell you to post it on the Forum and answer it there.

PRACTICUM PERIOD

PURPOSE

The purpose of the Practicum Period is to give participants dedicated time to advance their Practicums after learning the full slate of DS4A curriculum. The period will involve consistent team work on your Practicum, advised by your TAs, in the lead up to your Practicum presentation at the Finale.

TIMELINE

From Sunday, July 19th - Thursday, August 6th, you will work full-time on your Practicum. This will largely consist of teamwork, with your TA serving to answer critical questions and guide your thinking.

Your Group will present your Practicum on Saturday, August 8th, at your Finale.

PARTICIPANT ACTIVITY

During the Practicum Period, your team will be free to work on your Practicum whenever you see fit. We encourage teams to schedule formal times to meet remotely in a consistent fashion. You should dedicate at least the normal class hours on the Fridays and Saturdays during the Practicum Period to working on your Practicum.

Your TAs will be fully available during the normal class hours on the Fridays and Saturdays to assist with Practicum-related questions. TAs will also be able to assist Sunday - Thursday according to your group needs.

OUTCOME

The outcome of the Practicum Period is a recorded presentation that will be presented during the Finale on August 8th. Your presentation should be done with all sensitive information redacted, and should show your problem solving approach in a compelling way.

Accompanying your presentation will be a summary of your project to easily demonstrate the value of your project to others, along with a detailed report on your project. The detailed report will give a comprehensive description of the data sets, analysis performed, and outcomes of your project. Correlation One will provide additional guidance on your project deliverables further along in the course.

Top teams will receive a number of prizes including media interviews, awards presented by top officials in the Government of Colombia, and recognition among the entire DS4A program. All teams will receive multimedia of their presentation to be used in career promotion.

PAST PARTICIPANT TESTIMONIALS

DS4A propels graduates' careers and changes lives. DS4A graduates have experienced career advancement, found new job opportunities, started companies, built professional networks and made friends for life. We pride ourselves on your success, and testimonials from some of our graduates are below:



KAREN

"Learning from Natesh, a Harvard professor and one of the best data science teachers in the world, has been so inspiring! You can't find this level of teaching anywhere else."

ADRIÁN

"DS4A changed my life! By the final week, I had already received three job offers, including one from my dream company, Ecopetrol."

JESSICA

"DS4A gave me the inspiration to make a major career change. I got an incredible job in healthcare through DS4A!"

CRISTIAN

"The program instructs you in completely practical skills. Everything is very useful. DS4A has allowed me to take on new projects at my company, like building a dashboard visualization for business performance...I received a big promotion in less than one month!"

JAIRO

"I love the DS4A community. To be part of a group of passionate and talented people is a tremendous resource. I started a company with two fellow participants and our entire class still keeps in touch, sharing job opportunities & ideas!"



ABOUT CORRELATION ONE

The AI & data talent market is broken. The shift towards a data economy is rapidly transforming the roles of 200M+ analytical workers, and enterprises and governments lack the tools to manage this transformation.

Correlation One is the market leader in the data talent space. We enable companies to attract, assess and train analytical talent to build their data advantage. We have built an expert community of 150,000+ data scientists and 600+ relationships with the world's leading universities in the US, China, UK, Canada, and Ireland.

Our global data science programs, including our signature Datathons, have helped foster the AI talent ecosystem in multiple regions around the world. Our programs have also earned worldwide recognition, including a Harvard Business School case study.

We have developed custom talent solutions for some of the most sophisticated employers in the world, including Citadel LLC, JP Morgan, Point72, Deloitte, and the National Science Foundation.

For more information on Correlation One, please [visit our website](#).