correlation one

STUDENT
WELCOME PACKET

DS4A: AI-JOB-READINESS TRAINING FOR COLOMBIA
October-December 2019

correlation one

Content

- Welcome & Program Background
- Program Calendar
- Program Locations
- Curriculum Overview
- Classroom Overview
- To do before the first class: Setting Up Your Local Development Environment
- To do before the first class: Access Slack and the Forum
- Certification
- About Correlation One

Welcome & Program Background

Welcome to the 2019 Colombia Data Science for All (DS4A) program. We are thrilled to welcome you to be part of this unique opportunity to learn data science for the digital future.

You are part of a highly impressive group of 300 students, selected by MinTIC and Correlation One from thousands of applicants via a meritocratic and transparent process. Your class of 300 represents managers, entrepreneurs, and innovators from across Colombia, and from industries including software, medicine, consumer goods, and industrial products, among others. Your group has excelled in leading schools, companies, and government agencies both in Latin America and around the world.

We look forward to working with you over the next 10 weeks as we explore data science and the many ways it can be used to impact the world around us. Through lectures taught by renowned instructors, immersive homework, and hands-on group projects, we hope you will gain a close understanding of what is possible through skills as varied as data analysis, linear modeling, SQL and advanced machine learning.

We hope not only that you will learn technical skills, but also that you will be stewards of the data science community in Colombia. This country has the potential to be the main AI hub in Latin America – but that promise can only be fulfilled if we succeed in building a sustainable, vibrant, and local AI-ecosystem. Towards that end, we will facilitate you getting to know your classmates and other local stakeholders who are working to ensure that a robust AI community is built in Colombia.

This packet will serve as an overview of the program, what it covers, and what your role in the program is. Welcome to DS4A-Colombia.

MinTIC / Correlation One

Program Calendar

OCTOBER						
SUN	MON	TUE	WED	THU	FRI	SAT
		1	2	3	4	5
6	7	8	9 Class #1	10 Class #2	11 Class #3	12 Class #4
13	14	15	16	17	18 Class #5	19 Class #6
20	21	22	23	24	25 Class #7	26 Class #8
27	28	29	30	31		

NOVEMBER						
SUN	MON	TUE	WED	THU	FRI	SAT
					1 Class #9	2 Class #10
3	4	5	6	7	8 Class #11	9 Class #12
10	11	12	13	14	15 Class #13	16 Class #14
17	18	19	20	21	22 Class #15	23 Class #16
24	25	26	27	28	29 Class #17	30 Class #18

DECEMBER						
SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6 Class #19	7 Class #20
8	9	10	11 Class #21	12 Class #22	13 Class #23	14 Class #24

Program Locations

● Bogotá
Centro de Convenciones Gonzalo Jiménez de Quesada
Segundo Nivel Salon C
Cl. 26 #13A-10 Bogotá

● Medellín
Universidad EAFIT
Avenida Las Vegas - Carrera 49 N° 7 Sur-50
Bloque 19 piso 4 aula 414 (415)

● Cali
Universidad Santiago de Cali – Campus Pampalinda
Calle 5 carrera 62-00 Cali
Salón 1528 (Bloque 1 piso 5)

● Barranquilla – Week 1:
Universidad del Atlántico
Sede Norte: Carrera 30 Número 8- 49 Puerto Colombia - Atlántico
Bloque G --> Biblioteca "Orlando Fals Borda"
Lugar: 101G

Barranquilla – Weeks 2-10:
Universidad del Norte
Km.5 Vía Puerto Colombia
Salón: Laboratorio L-J-5

Class Requirements

- In this intensive program, each of the 24 classes counts. Participants must plan to attend each class.
- Classes will begin at 7am sharp. Participants should plan to arrive at the classroom at 6:45am, sign in, and be fully prepared to work at 7am.
- Classes end at 7pm
- Participants should bring their own laptops to class.
- Participants should bring chargers for their laptops.
- Before the first class, participants should set up their laptops with a data analysis environment. Instructions for setting up your data analysis environment are on the class site: https://ds4a-colombia.correlation1.com

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 student 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 or five people at the outset--this team assignment will persist throughout the program and will form the core of your peer "Learning Group".

Class Structure

Classes begin in the morning with a lecture. Typically, Lectures will be from 7am - 12pm.

In the afternoon, participants will work in their Learning Groups on practical projects and lab work related to the material presented in the morning. They will be assisted by Teaching Assistants, who will be available on-site and remotely for help.

As discussed above, Learning Groups will be assigned on the first day of class, and will consist of the core group of peers you will work with on projects. Learning Groups have been strategically formed to pair together individuals with complementary skills.

In addition to class work, certain blocks of time have been allocated for participants to meet employers who are actively hiring in the data science & AI space. Social time will also facilitate classmates meeting one another and building a local data science community.

Specific project and lab work will be given to you throughout the program. Participants are eligible for certification upon successful completion of the program. Certification Requirements will be given to participants on the first day of class.

Accessing Slack, Forum & 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://ds4a-colombia.correlation1.com with your email and password.

Class Materials:

Class materials will be made available on the class website (following the registration process, per above): https://ds4a-colombia.correlation1.com

On the class website, you will be able to download cases, view the class schedule, and ask questions on the class forum.

Slack Communication:

Slack will be the primary channel of communication for most program topics. Students may use Slack to ask questions during lecture, should critical questions related to curriculum arrive; TAs will monitor for critical questions and answer as appropriate. 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

Once you have downloaded Slack, click the link below to join the DS4A_Colombia slack workspace.

https://join.slack.com/t/ds4acolombia/shared_invite/enQtNzcwMDYzOTY0NzY5LTllYTQxY2EwZWU3YWQ4MWJkYmZiMTZmYWQxZWFiMmNkNWE4MzlkZjliYWRiODlhZjk4Y2ZlNTJjZDg5ODBiOTA

Finally, complete your profile by adding your full name, phone number, and a short blurb about yourself.

We have already set up a few channels for class-wide communications, however you are encouraged to create new channels to facilitate other activities, like group projects or social gatherings.

Curriculum Questions (Forum):

A forum will be made available on the class website:
https://ds4a-colombia.correlation1.com

The forum should be used primarily for questions outside of class time. Students can post questions and answer one another's questions on the forum. TAs will also be moderating forums and answering questions. Students are encouraged to use the forum to not just ask questions, but as a general learning resource – to answer their fellow students' questions and check the key themes and focus issues as they evolve during the course.

Curriculum Overview

The lectures will cover the following topics:

- Week 1: Data Science, Machine Learning, and Artificial Intelligence at a Glance; Basics of Python; Data Interpretation
- Week 2: Data Investigation & Exploratory Data Analysis
- Week 3: AWS; SQL in a Big Data World
- Week 4: AWS (cont.); Data Wrangling & Cleaning; Interactive Data Visualization
- Week 5: Data Wrangling & Cleaning (cont.); Data-Driven Decision-Making
- Week 6: Linear Modeling - Variations & Extensions
- Week 7: Classical Machine Learning Models & Cross-Validation
- Week 8: Regularization; Deep Learning; NLP & Sentiment Analysis
- Week 9: Deep Learning (cont.); Model Selection
- Week 10: Project Work & Presentations

Generally, there will be a pair of lectures each morning from 7AM – 12PM on the days when the program is in session. Each lecture will consist of a mixture of instruction, exercises, and built in Q&A time.

The program will also consist of three other crucial components:

Weekly Extended Cases: These serve as the "homework assignments" of the program. They are structured as complex, multi-part business problems, and students must leverage all of the skills they have learned from that week's lectures in order to solve them properly. Cases are given out at the end of the day on Saturdays, and participants are expected to work on them from Sunday through Thursday of the following week, to be completed and turned in by that Friday for class. Each extended case is designed to take approximately 3 - 4 hours to complete.

Datathon Project: Each trainee will be assigned to a five-person Learning Group for the duration of the program. Trainees will work with their groups to produce both a Datathon Project and a Final Project (discussed later).

The Datathon Project is based on Correlation One's Datathon competition series, and is designed for the trainees to apply and showcase the data science and analysis skills

they have accumulated throughout the program. Students will be given a particular topic of focus and several datasets related to this topic. They are asked to explore the provided data and based on this, pose a question which they believe would be interesting to answer. They will then write a report detailing why they believe this question is important and how they went about answering it. This report is expected to detail the team's execution of the entire data science process, from start to finish.

Final Project: The Final Project is meant for trainees to apply and showcase the data engineering and visualization skills they have developed throughout the program. Teams will be asked to scour available datasets from a number of pre-provided data repositories (the full list will be provided separately) and build a production-level application centered around a subset of those datasets. This application should either solve an existing problem or improve upon an existing process. It will be up to each team to identify a problem or area of improvement to focus on.

Teams will also be expected to write a report which details what their application does, how they chose to design it and why, and how they implemented each major piece. Final projects will be showcased to both government officials and private sector employers on the last week of the program, and should be comprehensible and usable by non-technical persons.

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, Master Class" graduate, students must:

1. Attend all lectures & activities barring some emergency, in which case student needs to get permission from Correlation One.

2. Complete all weekly extended cases and achieve a 'Satisfactory' rating in at least 6 out of 8 of them. To obtain a 'Satisfactory' rating, the students need to perform at least 75% of the exercises.

3. Contribute sufficiently to the completion of both the Datathon and final projects. At the end of the program, we will conduct a survey whereby we ask each student to rate how each of their teammates contributed to the Datathon and final project; this information will be used in determining certification.

Certification can help program graduates earn promotions or new jobs, and also indicates membership in Colombia's most prestigious AI community.

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 trans- formation.

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 scientist 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.