

# DATA SCIENCE FOR ALL COLOMBIA

## INTRODUCTION

OCT - DEC 2019

correlation.one  
**C.ºO**



# Outline

- The Correlation One approach
- TA Introductions
- Who you are
- Why data science matters
- Class schedule
- Typical class structure and expectations
- Communication
- Certification
- Key takeaways

# The Correlation One approach

## Practical

---

*We are collaborating with long-term partners in Colombia, and aim to create a talent pipeline for AI in Colombia that is the most advanced in Latin America.*

---

## Collaborative

---

*We want to encourage you to make connections and learn from each other. This program can teach you to eventually teach others.*

---

## Adaptive

---

*Our curriculum is dynamic and tailored to your specific interests and skills. Even advanced students will find something useful here.*

---

# Teaching Assistants: Bogotá



Daniel  
Martin-Alarcon

- New Mexico State University (BS), MIT Media Lab (PhD)
- Data Science Fellow at Lambda School
- Several peer-reviewed publications in Data Science and Synthetic Biology



Jay  
Gopalan

- Harvard University (BA)
- Mentor, Harvard Education Portal; Teaching Assistant for four Harvard STEM courses
- Interests include the interaction of data analysis/modeling and large-scale business and social problems



Zach  
Hervieux-Moore

- Queens University (BA), Princeton University (PhD)
- Former software engineer with Intel and Siemens
- Educational experience in several courses, including Computer Architecture and Digital Systems

# Teaching Assistants: Barranquilla, Cali, Medellín



Jimmy  
Jing

Barranquilla

- New York University (BA)
- Data Analyst at big data firm Earnest Research, conducting credit and debit card transaction data for private sector application
- Hobbies include poker; former professional player



Julian  
Romero

Cali

- Universidad Nacional de Colombia (BA), Universidad de los Andes (MS), University of Waterloo (PhD)
- Winner, "Outstanding Teacher Assistant", Univ. of Waterloo
- Interest areas include extended formulations and stochastic optimization



Spencer  
Stebbins

Medellín

- New York University (BS), Harvard University Extension
- Experience with NYSE Top 50 companies and major government agencies
- Work as an educator includes >2,500 hours of Data Science material

# Who you are

You have been selected from a large, highly competitive pool

Convocatoria launched for all of Colombia



Over 10,000 top applicants from across Colombia



Meritocratic, transparent selection process led by  
MinTIC & Correlation One



~300 selected participants with exceptional technical skills,  
academic achievements, and professional experience

**<3%** acceptance rate for the DS4A Colombia program

# Who you are

You possess a range of impressive backgrounds and skillsets



Strong Professional Experience



Solid technical foundations



Impressive educational backgrounds



Working English-language abilities

- |  |  |  |   |
|--|--|--|---|
| <ul style="list-style-type: none"><li>• You have professional backgrounds as managers, founders, and other roles</li><li>• Your industry experience ranges from software to robotics to government</li></ul> | <ul style="list-style-type: none"><li>• You possess varied technical skillsets, shown in demanding Convocatoria exam</li><li>• Your tech stacks include R, Python, SQL, MATLAB, and Java</li></ul> | <ul style="list-style-type: none"><li>• You have been successful at world class schools in Colombia and abroad</li><li>• Many of you possess advanced degrees or professional training</li></ul> | <ul style="list-style-type: none"><li>• You possess functional English-language skills</li><li>• Many of you have school or work experience in English-speaking countries</li></ul> |
|--|--|--|---|

# Who you are

# DS4A today, the center of Colombia's AI ecosystem tomorrow



# Why Data Science?



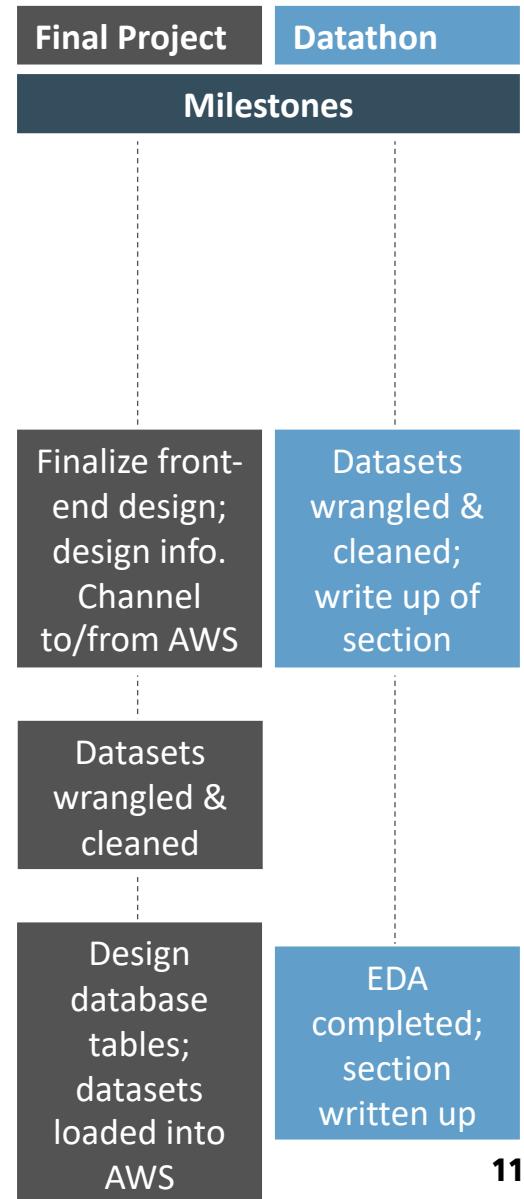
# Class schedule (1 of 3)

Week	Themes	Case	Session	Session Date	Final Project	Datathon		
					Milestones*			
1	<ul style="list-style-type: none"> <li>Data Science, Machine Learning, and Artificial Intelligence at a Glance</li> <li>Basics of Python</li> <li>Data Interpretation</li> </ul>	1	1	October 9				
		2						
		3	2	October 10				
		4						
		5	3	October 11				
		6						
		7	4	October 12				
		8						
2	<ul style="list-style-type: none"> <li>Data Investigation &amp; Exploratory Data Analysis</li> </ul>	9	5	October 18				
		10						
		11	6	October 19				
		12						
3	<ul style="list-style-type: none"> <li>AWS</li> <li>SQL in a Big Data World</li> </ul>	13	7	October 25	Select problem versions; source data; scope project	Chose appropriate topic; scope project; write proposal		
		14						
		15	8	October 26				
		16						

\* – See Student Welcome Packet for full Milestones descriptions

# Class schedule (2 of 3)

Week	Themes	Case	Session	Session Date
4	<ul style="list-style-type: none"> <li>AWS (continued)</li> <li>Data Wrangling &amp; Cleaning</li> <li>Interactive Data Visualization</li> </ul>	17	9	November 1
		18		
		19		
		20	10	November 2
5	<ul style="list-style-type: none"> <li>Data Wrangling &amp; Cleaning (continued)</li> <li>Data-Driven Decision-Making</li> </ul>	21	11	November 8
		22		
		23		
		24	12	November 9
6	<ul style="list-style-type: none"> <li>Linear Modeling - Variations &amp; Extensions</li> </ul>	25	13	November 15
		26		
		27		
		28	14	November 16
7	<ul style="list-style-type: none"> <li>Classical Machine Learning Models &amp; Cross-Validation</li> </ul>	29	15	November 22
		30		
		31		
		32	16	November 23

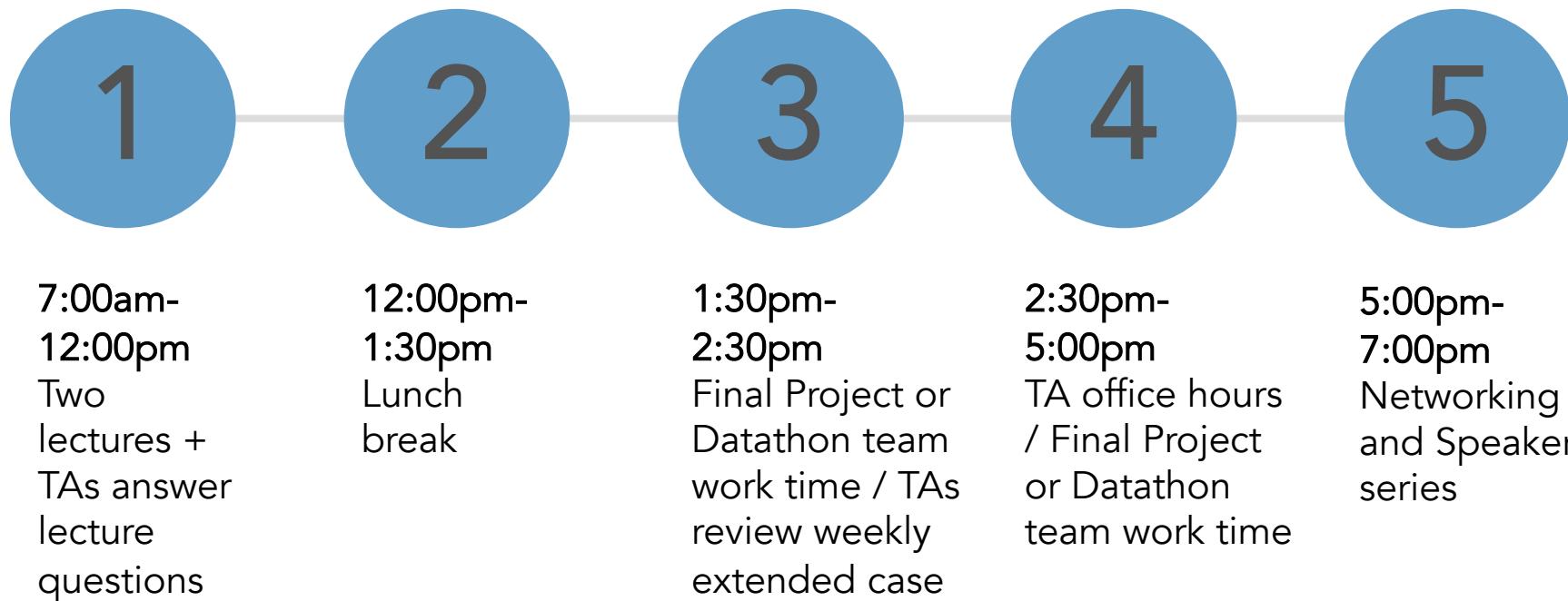


# Class schedule (3 of 3)

Week	Themes	Case	Session	Session Date	Final Project	Datathon		
					Milestones			
8	<ul style="list-style-type: none"> <li>• Regularization</li> <li>• Deep Learning</li> <li>• NLP &amp; Sentiment Analysis</li> </ul>	33	17	November 29	Finalize front-end build; link from front-end to AWS	Complete analysis, modeling; write up section		
		34						
		35	18	November 30				
		36						
9	<ul style="list-style-type: none"> <li>• Deep Learning (continued)</li> <li>• Model Selection</li> </ul>	37	19	December 6	Finish analysis of modeling of datasets	Complete conclusions; report touch ups		
		38						
			20	December 7				
10	<ul style="list-style-type: none"> <li>• Project Work</li> <li>• Presentations</li> </ul>		21	December 11	App final touches; quality checks; finish report; prepare presentation	Complete conclusions; report touch ups		
			22	December 12				
			23	December 13				
			24	December 14				

# Our class structure

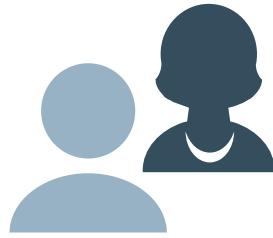
A typical class features several modules for a dynamic experience



# Our roles



## Your share



- Arrive/return on time; be prepared
- Work as “Learning Group”
- Complete projects and assignments
- Engage networking and speaking series resources

---

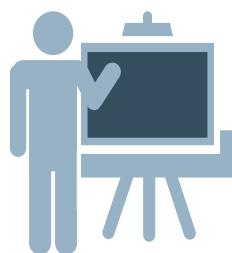
## My share



- Deliver concepts
- Problem-solve in cases
- Manage lecture discussion

---

## TA's share



- Advise project work
- Facilitate deeper insight and answer questions
- Provide real world context for class concepts

# Expectations for lectures



## Be prepared for class

- Attend all 24 classes
- Arrive 15 minutes before 7am start
- Be sure to bring your equipment
- Learning is cumulative; complete the weekly homework
- Use the Forum



## Engage during class

- Don't just watch, ask questions (during lectures, use #livequestions)
- Embrace hands-on learning (e.g., Jupyter)
- Collaborate with Learning Group
- Reflect on real world implications of material

# Asking in-class questions



#livequestions

- You are encouraged to ask substantive questions during lecture
- Use the #livequestions Slack channel to ask Lead Professor any important questions about the current lecture material
- The TAs will monitor for good questions to send to Lead Professor
- The Lead Professor will answer the most crucial questions in a dedicated question-answering time

# Lunch



- You are free to take lunch as you wish
- Please return promptly at 1:30pm

# Course materials



## Description

### Weekly Extended Cases

- “Homework assignments” of the course structured as complex, multi-part business problems

### Datathon Project

- Based on C1’s popular Datathon competition series, teams receive several datasets and use them to pose an interesting question, then explore and answer it

### Final Project

- Teams will build a production-level application (with visualization) centered around a selected dataset

# Course materials



## Weekly Extended Cases

- Apply skills from class
- Measure individual progress

Aim

Schedule

## Datathon Project

- Develop skills from class
- Practice applying data science structure to ambiguous, real-world problems

## Final Project

- Showcase skills from class to government officials and employers in final week

- 3-4 hours per week, done individually
- Distributed end of Saturday class, due the next Friday

- Done as a Learning Group throughout course
- TAs will provide guidance during class

- Done as a Learning Group throughout course
- TAs will provide guidance during class

# **Speaker series and networking**



Engage with your peers and industry and government leaders

- Discover practical uses for our material
- Create connections for business and life
- Meet the leaders of the top businesses and institutions of today – and tomorrow

# Different uses of Slack and the Forum

It is important that you use the right channel for the right issues

## Slack

- Tech support (#techsupport)
- Learning Team discussions
- Social chat (#social-...)
- In-class lecture questions (#livequestions)
- Quick TA questions or issues (private channel)

## Forum

- Substantive questions and ideas about course material
  - TAs will not reply to curriculum questions on Slack
  - You can reply to your fellow students' questions on the Forum

Do not paste code fragments from Extended Cases on Slack or Forum

# Certification requires three steps

Gain membership in Colombia's most prestigious AI community



# Conclusion

- This is a mega opportunity for all of us
- Share your ideas with us – there are many things we don't know
- We're building for the future; in future programs, we want you to be the teachers