correlation one

NOVEMBER 2019

DATA SCIENCE FOR ALL

DEVELOPING LOCAL AI TALENT

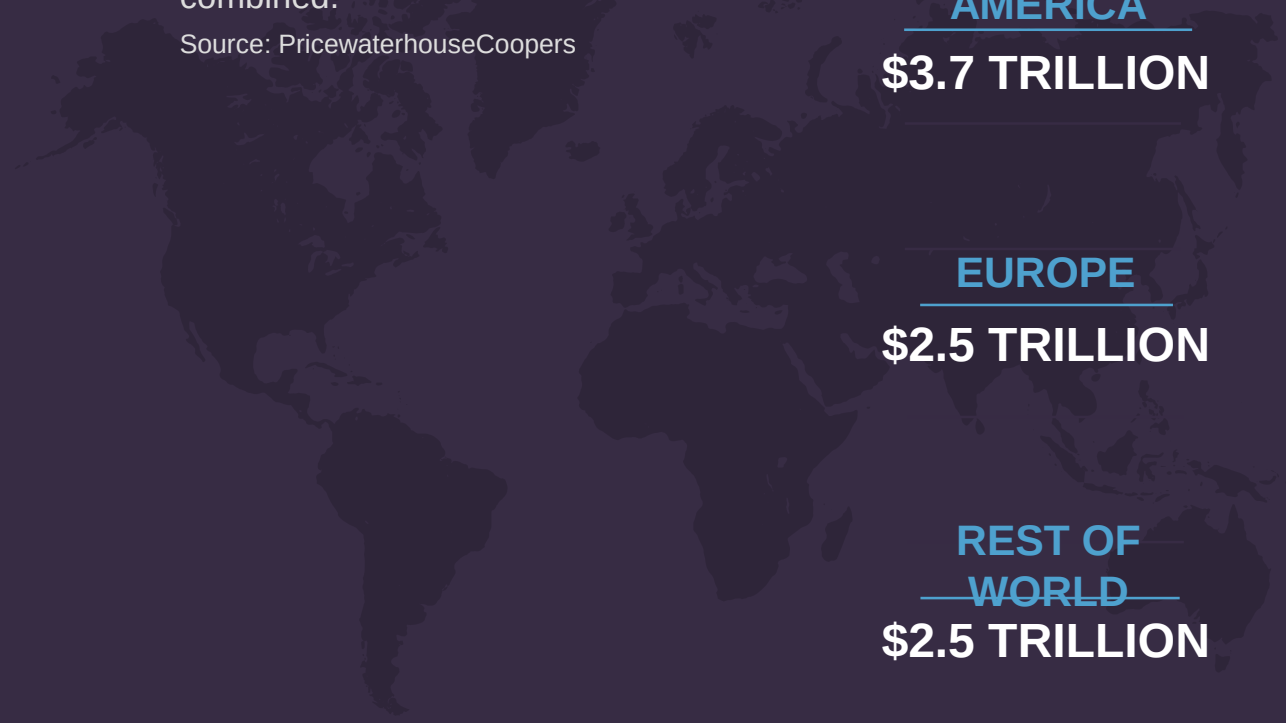
AI will prove to be the biggest economic force in history

AI IS PROJECTED TO CONTRIBUTE

$15.7 TRILLION

to global GDP by 2030, more than the current output of China and India combined.

Source: PricewaterhouseCoopers



CHINA

$7.0 TRILLION

NORTH AMERICA

$3.7 TRILLION

EUROPE

$2.5 TRILLION

REST OF WORLD

$2.5 TRILLION

To participate in the AI revolution, **countries must develop local AI talent**

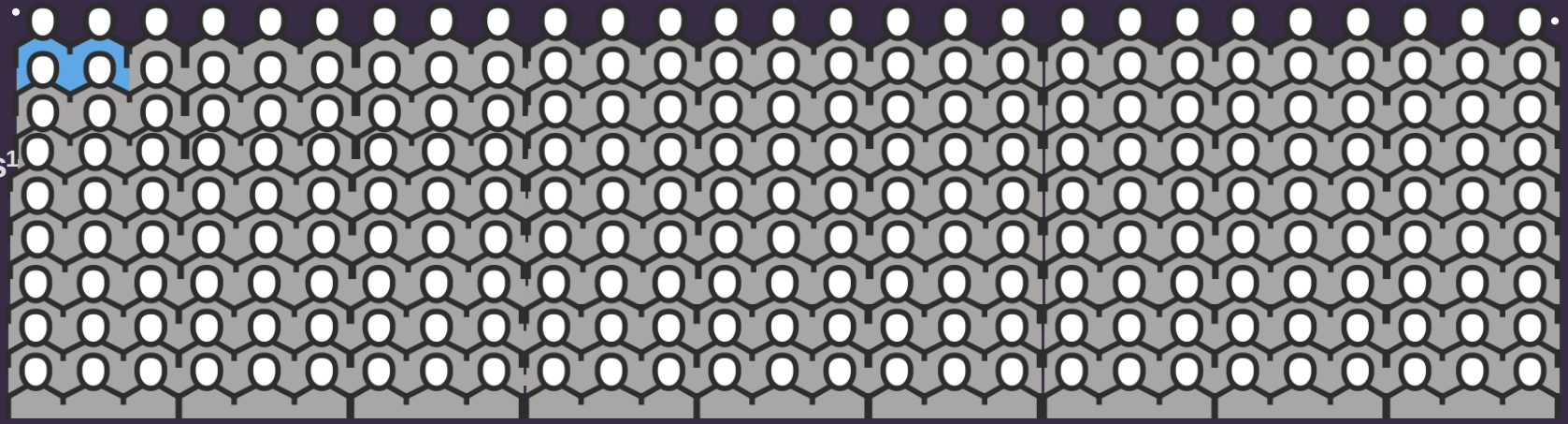
"Across the board, companies report that talent is the biggest hurdle to integrating data and analytics"
- McKinsey

"As the personal computer did 30 years ago, data science and analytics… are creating a new revolution in work."
- PricewaterhouseCoopers

"You have to have the data...But without talent, it is meaningless."
- AT&T

Data literacy is becoming necessary for an enormous number of jobs

2M
Data scientists[1]



230M
Analysts and managers across finance, operations, accounting, marketing[2]

In the data economy, all analytical workers will need to be data literate

1 KDNuggets | 2 McKinsey

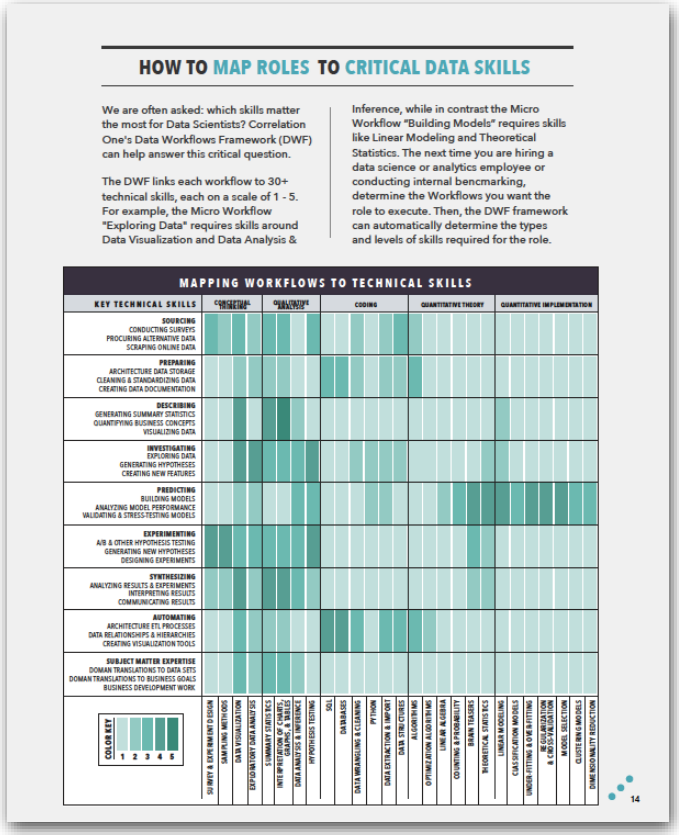
Correlation One helps organizations develop their data literacy

Correlation One's solutions help organizations build AI capabilities

CITADEL wayfair Point72 [XTX] MARKETS experian Jefferies KORN FERRY imc

Correlation One set the standards for data science and AI skills

We built a proprietary framework that taxonomizes data roles and skills



HOW TO MAP ROLES TO CRITICAL DATA SKILLS

We are often asked: which skills matter the most for Data Scientists? Correlation One's Data Workflows Framework (DWF) can help answer this critical question.

The DWF links each workflow to 30+ technical skills, each on a scale of 1 - 5. For example, the Micro Workflow "Exploring Data" requires skills around Data Visualization and Data Analysis & Inference, while in contrast the Micro Workflow "Building Models" requires skills like Linear Modeling and Theoretical Statistics. The next time you are hiring a data science or analytics employee or conducting internal benchmarking, determine the Workflows you want the role to execute. Then, the DWF framework can automatically determine the types and levels of skills required for the role.

MAPPING WORKFLOWS TO TECHNICAL SKILLS

KEY TECHNICAL SKILLS

CONCEPTUAL THINKING | QUALITATIVE ANALYSIS | CODING | QUANTITATIVE THEORY | QUANTITATIVE IMPLEMENTATION

SOURCING — CONDUCTING SURVEYS; PROCURING ALTERNATIVE DATA; SCRAPING ONLINE DATA

PREPARING — ARCHITECTURE DATA STORAGE; CLEANING & STANDARDIZING DATA; CREATING DATA DOCUMENTATION

DESCRIBING — GENERATING SUMMARY STATISTICS; QUANTIFYING BUSINESS CONCEPTS; VISUALIZING DATA

INVESTIGATING — EXPLORING DATA; GENERATING HYPOTHESES; CREATING NEW FEATURES

PREDICTING — BUILDING MODELS; ANALYZING MODEL PERFORMANCE; VALIDATING & STRESS-TESTING MODELS

EXPERIMENTING — A/B & OTHER HYPOTHESIS TESTING; GENERATING NEW HYPOTHESES; DESIGNING EXPERIMENTS

SYNTHESIZING — ANALYZING RESULTS & EXPERIMENTS; INTERPRETING RESULTS; COMMUNICATING RESULTS

AUTOMATING — ARCHITECTURE ETL PROCESSES; DATA RELATIONSHIPS & HIERARCHIES; CREATING VISUALIZATION TOOLS

SUBJECT MATTER EXPERTISE — DOMAIN TRANSLATIONS TO DATA SETS; DOMAIN TRANSLATIONS TO BUSINESS GOALS; BUSINESS DEVELOPMENT WORK

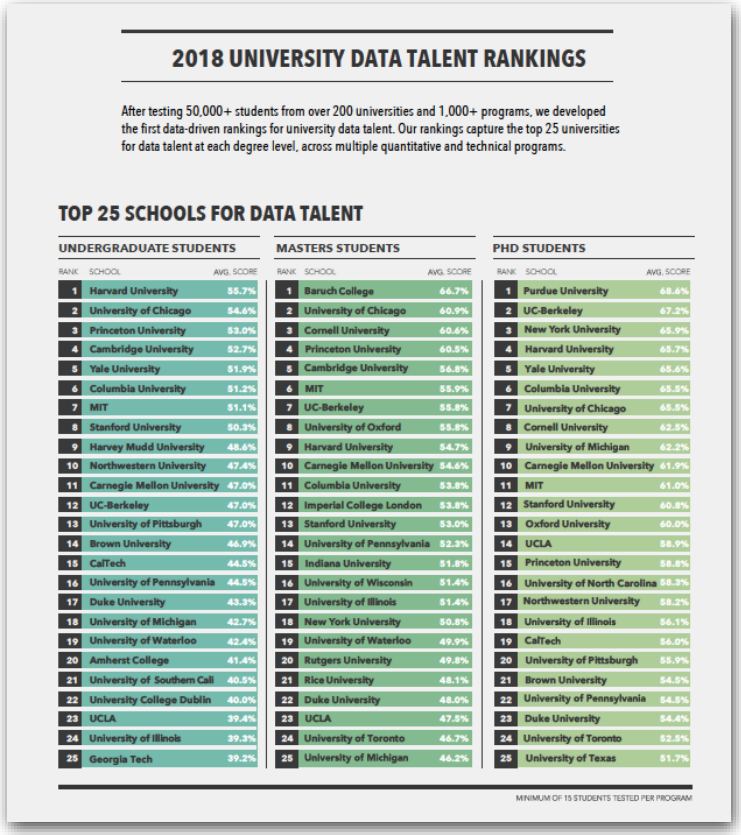
COLOR KEY 1 2 3 4 5

SURVEY & EXPERIMENT DESIGN; SAMPLING METHODS; DATA VISUALIZATION; EXPLORATORY DATA ANALYSIS; SUMMARY STATISTICS; INTERPRETING RESULTS; CODING & TOOLS; DATA ANALYSIS & INFERENCE; HYPOTHESIS TESTING; SQL; DATABASES; DATA WRANGLING & CLEANING; PYTHON; DATA EXTRACTION & IMPORT; DATA STRUCTURES; ALGORITHMS; OPTIMIZATION ALGORITHMS; LINEAR ALGEBRA; COUNTING & PROBABILITY; BAYES THEOREM; THEORETICAL STATISTICS; LINEAR MODELING; CLASSIFICATION MODELS; UNDERFITTING & OVERFITTING; REGULARIZATION & CROSS-VALIDATION; MODEL SELECTION; CLUSTERING MODELS; DIMENSIONALITY REDUCTION

14

(Future of Data Talent 2019 Annual Report, p14)

We developed the first standards and benchmarks for data skills



2018 UNIVERSITY DATA TALENT RANKINGS

After testing 50,000+ students from over 200 universities and 1,000+ programs, we developed the first data-driven rankings for university data talent. Our rankings capture the top 25 universities for data talent at each degree level, across multiple quantitative and technical programs.

TOP 25 SCHOOLS FOR DATA TALENT

UNDERGRADUATE STUDENTS			MASTERS STUDENTS			PHD STUDENTS		
RANK	SCHOOL	AVG. SCORE	RANK	SCHOOL	AVG. SCORE	RANK	SCHOOL	AVG. SCORE
1	Harvard University	55.7%	1	Baruch College	66.7%	1	Purdue University	68.6%
2	University of Chicago	54.6%	2	University of Chicago	60.9%	2	UC-Berkeley	67.2%
3	Princeton University	53.0%	3	Cornell University	60.6%	3	New York University	65.9%
4	Cambridge University	52.7%	4	Princeton University	60.5%	4	Harvard University	65.7%
5	Yale University	51.9%	5	Cambridge University	56.8%	5	Yale University	65.6%
6	Columbia University	51.2%	6	MIT	55.9%	6	Columbia University	65.5%
7	MIT	51.1%	7	UC-Berkeley	55.8%	7	University of Chicago	65.5%
8	Stanford University	50.3%	8	University of Oxford	55.8%	8	Cornell University	62.8%
9	Harvey Mudd University	48.6%	9	Harvard University	54.7%	9	University of Michigan	62.2%
10	Northwestern University	47.4%	10	Carnegie Mellon University	54.6%	10	Carnegie Mellon University	61.9%
11	Carnegie Mellon University	47.0%	11	Columbia University	53.8%	11	MIT	61.0%
12	UC-Berkeley	47.0%	12	Imperial College London	53.8%	12	Stanford University	60.8%
13	University of Pittsburgh	47.0%	13	Stanford University	53.0%	13	Oxford University	60.0%
14	Brown University	46.9%	14	University of Pennsylvania	52.3%	14	UCLA	58.9%
15	CalTech	44.5%	15	Indiana University	51.8%	15	Princeton University	58.8%
16	University of Pennsylvania	44.5%	16	University of Wisconsin	51.4%	16	University of North Carolina	58.3%
17	Duke University	43.3%	17	University of Illinois	51.4%	17	Northwestern University	58.2%
18	University of Michigan	42.7%	18	New York University	50.8%	18	University of Illinois	56.1%
19	University of Waterloo	42.4%	19	University of Waterloo	49.9%	19	CalTech	56.0%
20	Amherst College	41.4%	20	Rutgers University	49.8%	20	University of Pittsburgh	55.9%
21	University of Southern Cali	40.5%	21	Rice University	48.1%	21	Brown University	54.5%
22	University College Dublin	40.0%	22	Duke University	48.0%	22	University of Pennsylvania	54.5%
23	UCLA	39.4%	23	UCLA	47.5%	23	Duke University	54.4%
24	University of Illinois	39.3%	24	University of Toronto	46.7%	24	University of Toronto	52.5%
25	Georgia Tech	39.2%	25	University of Michigan	46.2%	25	University of Texas	51.7%

MINIMUM OF 15 STUDENTS TESTED PER PROGRAM

(Future of Data Talent 2019 Annual Report, p20)

We help organizations and countries build AI talent ecosystems

We are helping Colombian professionals lead the AI revolution

Correlation One has the largest global network of AI experts and programs

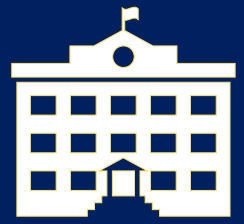
Correlation One has hosted 100+ data science programs in 6 countries

- Partnerships with 600+ academic groups around the world
- 250+ expert contributors to Correlation One's assessment and training platform
- Support for multiple languages on our assessment and training platform
- 40+ live data science competitions per year in 6+ countries

Data Science for All: Colombia

Colombia's stated goal: become the most educated Latin American country by 2025

Government support
Strategic 2025 plan published by Colombia's Ministry of Education



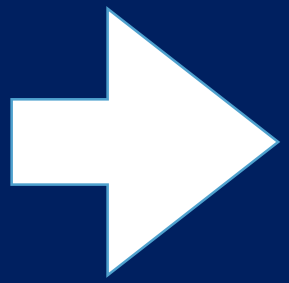
Recognized universities
Universidad Nacional de Colombia, Universidad de Los Andes, and others

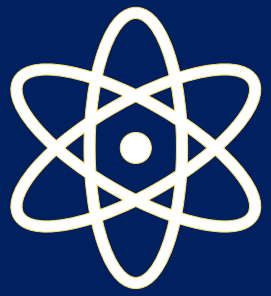
Student population
Colombia has over 2,100,000 students in higher education

Colombia has the foundational ingredients to participate in the AI economy

The Challenge: bridging the gap between STEM degrees and AI skills

STEM-Educated
Colombia has STEM students and professionals with basic quant skills





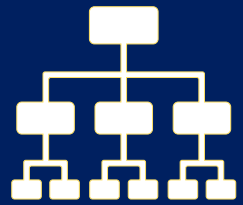
AI Professionals
We can close the last mile between STEM foundation and cutting-edge AI skills

Colombian scientists, researchers, and engineers have the ability to become qualified AI professionals, but they need top-down programs and support to close the skills gap

Solution: AI job-readiness training

10-week programs
Correlation One will host a series of live and online training programs for local professionals and students

Practical skills taught
Students will learn practical data analysis, visualization, wrangling and cleaning, predictive modeling, machine learning, etc.



Customized curriculum
Coursework and projects will be set in the context of Softbank's portfolio companies and other firms hiring in Colombia

Objective: Train professionals and students on practical data and AI skills

Introducing DS4A

correlation.∴one

DATA SCIENCE 4 ALL

COLOMBIA

Convocatoria proved talent pool and demand

Convocatoria was extremely successful; C1 developed a transparent, meritocratic methodology to help MinTIC select from 10,000+ applicants

Applicant selection funnel

- **September 14:** Convocatoria opens
  Positive coverage in major media outlets
- **September 20:** Convocatoria closes
  Applicants from wide variety of cities, economic backgrounds, and industries
- **September 22:** Technical Exam
  Applicants show high aptitude and fit for DS4A coursework

>1k applicants in 48 hours

10,424 total applicants

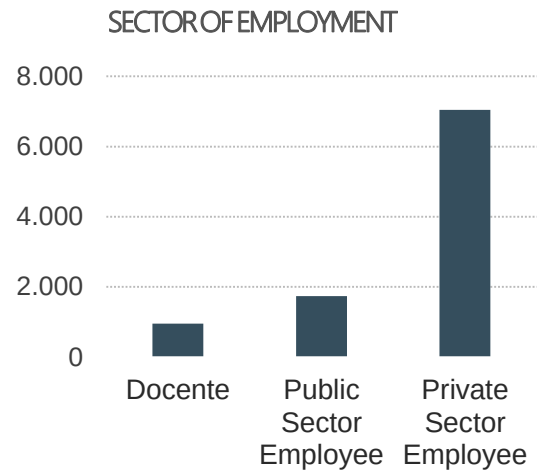
6,450 total exam takers

September 25: MinTIC selects ~300 applicants for admission (~3%)

Colombian applicants were highly diverse

With strong technical scores, applicants represented a full range of backgrounds

APPLICANTS BY CITY	
Bogotá	5066
Medellin	910
Cali	591
Barranquilla	328
Bucaramanga	185
Envigado	129
Manizales	115
Pereira	113
Cartagena	101
Sabaneta	81
Bello	70
Floridablanca	65
Soacha	60
Mosquera	59
Popayan	58
Tunja	56
Palmira	48
Ibagué	95
Armenia	47
Dosquebradas	41



APPLICANTS BY INDUSTRY OF INTEREST	# APPLICANTS APPLICANTS
Software	4607
Servicios de Informacion y Procesamiento de Datos	4441
Telecomunicaciones	2344
Colegio, Entidadesde Formacion para el Trabajo, SENA, Universidad	1676
Servicios Cientificos Tecnicos	1491
Finanzas, Banca, Seguros	1442
Gobierno Administracion Publica	1180
Venta al por Menor	480
Venta al por Mayor	464
Fabricacion de Computadoras, Electronica	460
Transporte Almacenaje	456
Otra Fabricacion	450
Asistencia Sanitaria, Asistencia Social	360
Mineria	355
Construccion	352
Arte, Entretenimiento, Recreacion	352
Publicacion	331
Utilidades	318
Servicios de Hostelería y Alimentacion	153
Servicios Juridicos	153

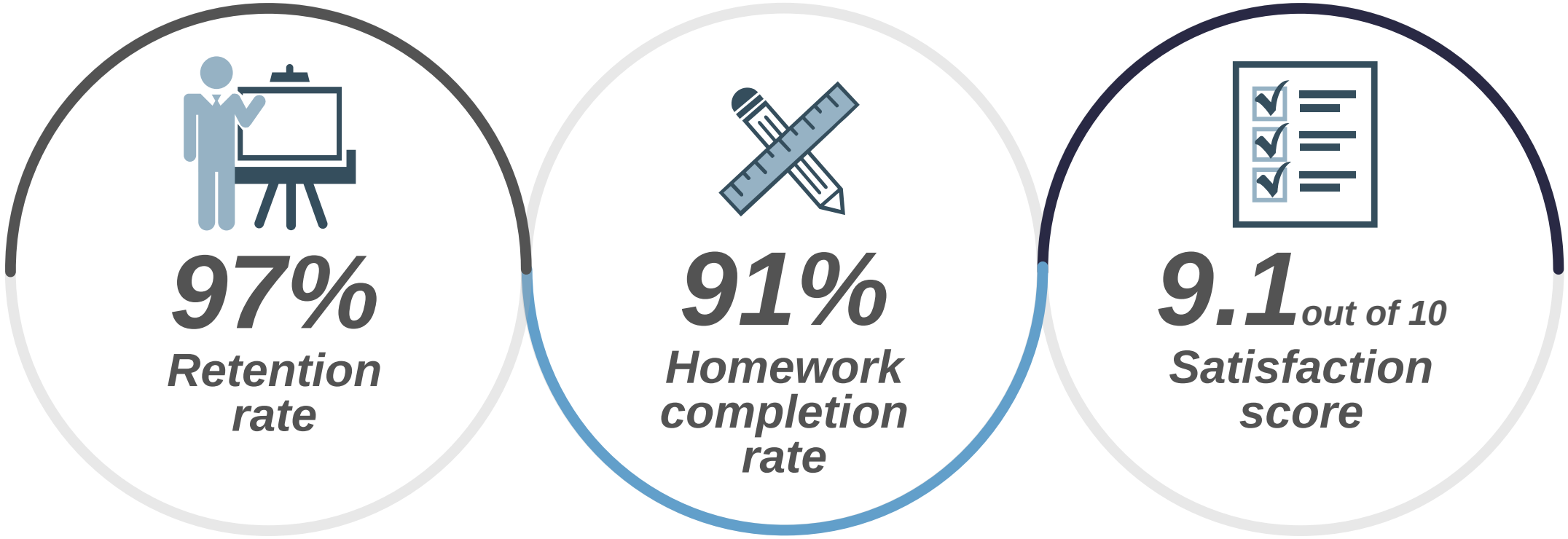
These results do not include long tail of highly diverse additional applicants

DS4A has successfully launched in 4 cities

We are in week 7 of 10, with exceptional results in Bogotá, Medellín, Cali, Barranquilla

Participant experience has been exceptional

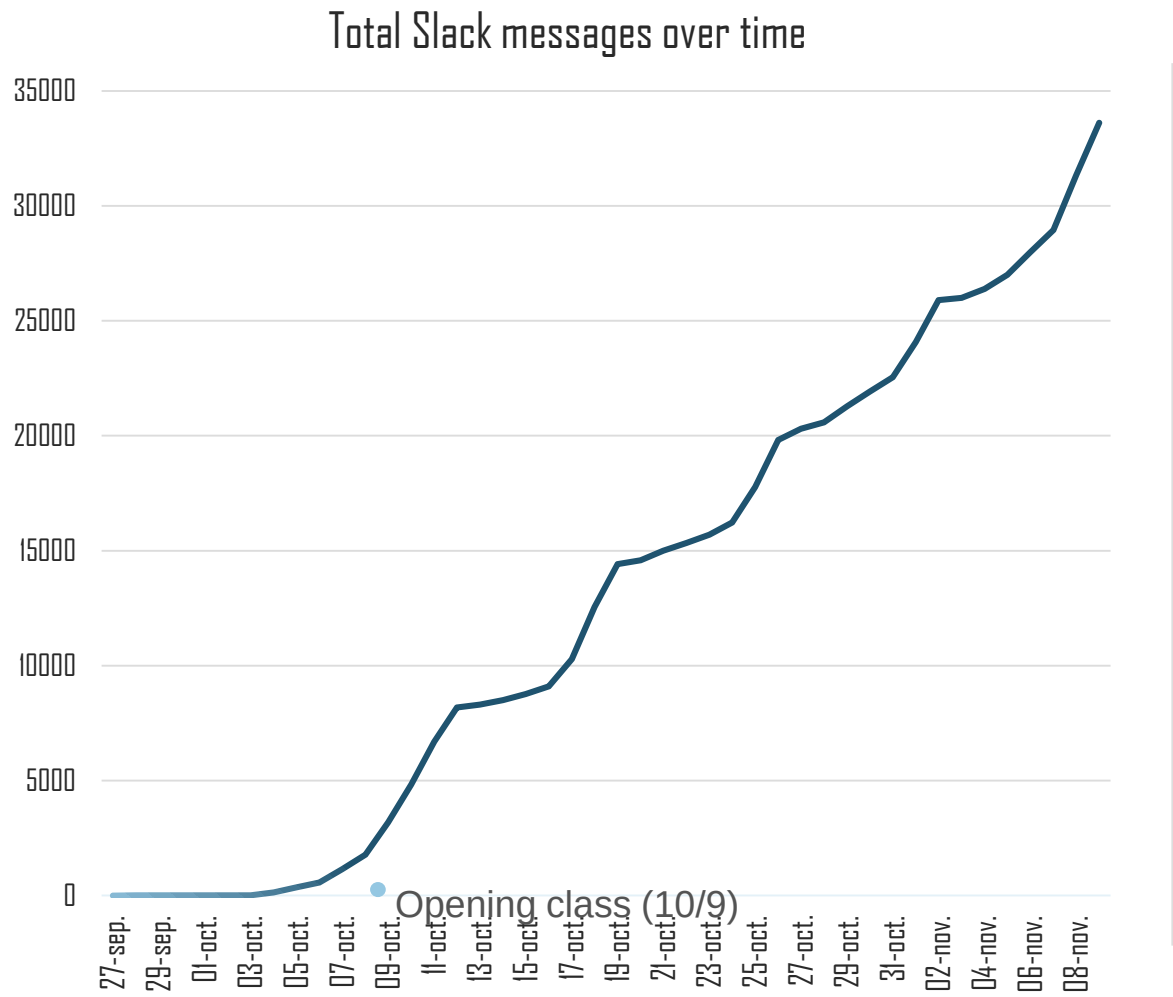
Class attendance, homework completion, and project satisfaction remain exceedingly high

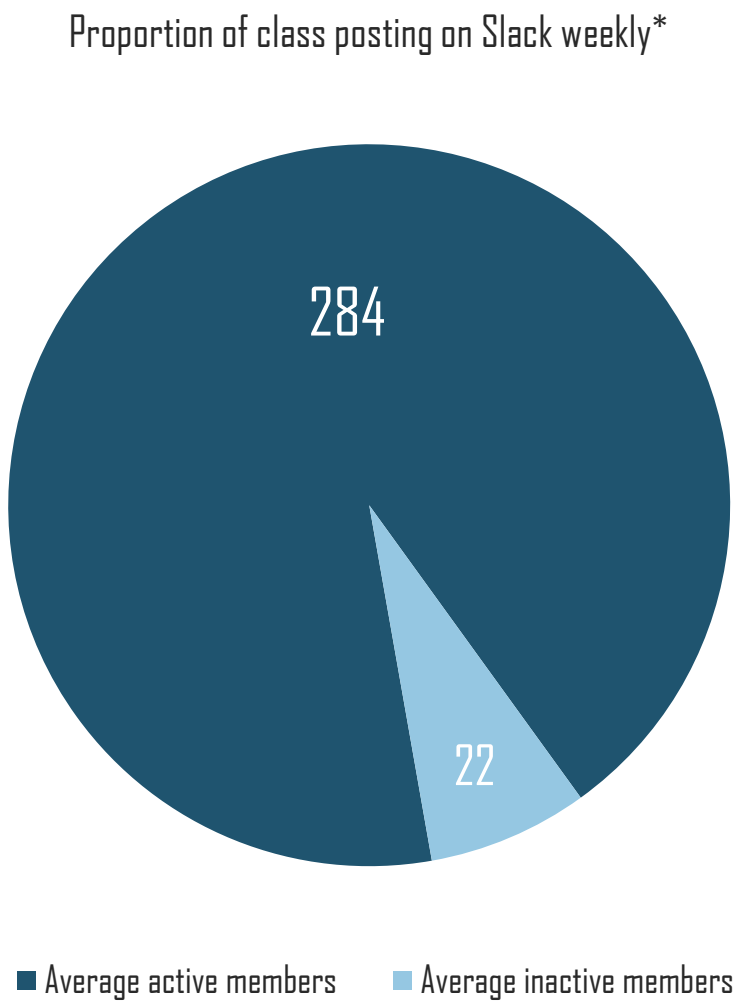


Source: * EDx website; ** Chronicle of Higher Education, expert interviews; † Expert interviews

Participants are forming an ecosystem

A community has solidified and continues to steadily grow across Colombia

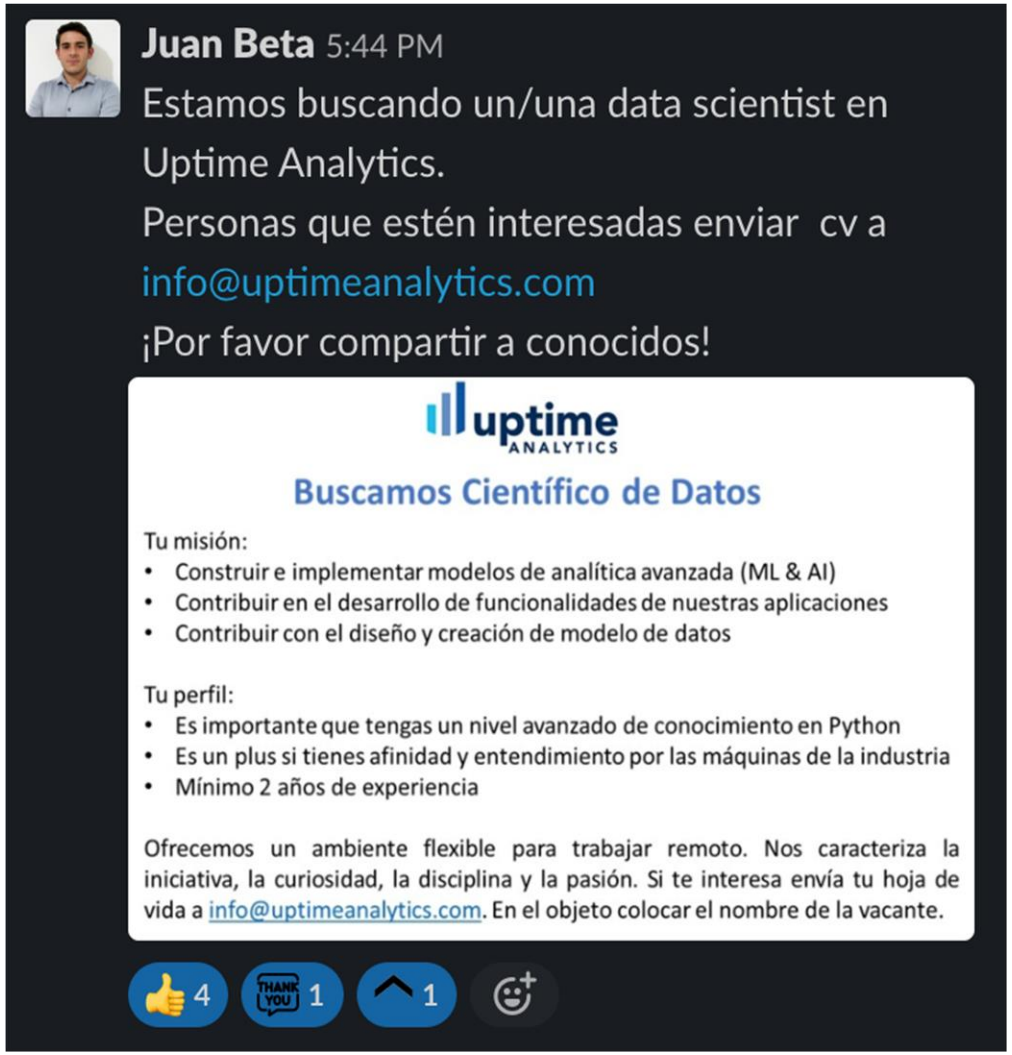




* Average across 5 weeks of class

33,000 messages and counting

From posting jobs to celebrating birthdays, and everything in between



In-depth collaboration

The Forum is providing a platform for layers of exploration and collaboration

1 **Question**

mariana 1d

How did you find out your EC2 instance's public IP address to access the app from the internet? I'm running the app and I supposedly have this IP address http://52.91.243.157/ 3 but I get a 'This site can't be reached' error in the browser.

2 **Answer**

N

normansimonr 1d

Hi mariana! You have to check that:

- Your EC2 instance has the correct inbound rules (EC2 Instance Ports Solution 8)
- Your app has host 0.0.0.0 at the bottom of the py script.
- You access the public IP of your EC2, and add :8050 to the end.

3 **Clarification**

JMiraV 1d

For people working on Windows, it'd be useful to point that they need to get rid of pypiwin32 and pywin32 in requirements.txt in order to properly run `pip3 install -r requirements.txt --user` in the EC2 instance.

4 **Continued use**

jaimeandresalvarez 15h

Hi, @normansimonr, thanks for the help. Your post was really useful!

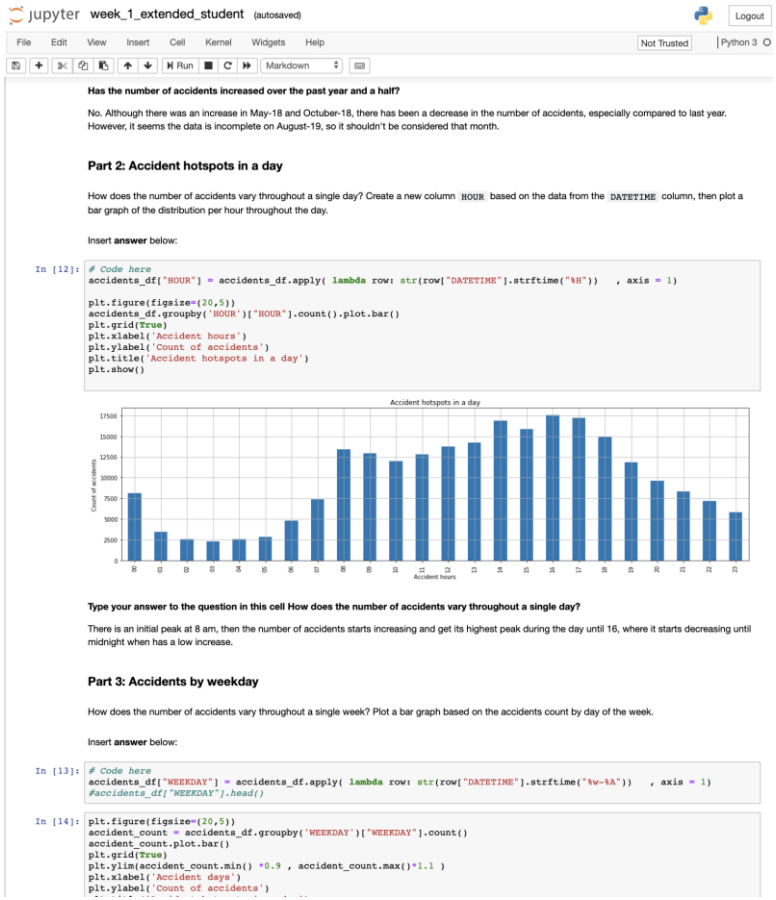
Learning By Doing

Teaching staff from Harvard, MIT, Stanford personally grades and advises students

Homework overview

- Participants independently complete Extended Cases, which are reviewed by hand, by world-class teaching staff
- Cases are done in participants' Jupyter notebooks and reflect substance of previous week's lectures
- Teaching staff offers additional assistance via Forum and Office Hours

Example participant submission



jupyter week_1_extended_student (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3

Run | Markdown

Has the number of accidents increased over the past year and a half?

No. Although there was an increase in May-18 and Octuber-18, there has been a decrease in the number of accidents, especially compared to last year. However, it seems the data is incomplete on August-19, so it shouldn't be considered that month.

Part 2: Accident hotspots in a day

How does the number of accidents vary throughout a single day? Create a new column `HOUR` based on the data from the `DATETIME` column, then plot a bar graph of the distribution per hour throughout the day.

Insert **answer** below:

In [12]:
```
# Code here
accidents_df["HOUR"] = accidents_df.apply( lambda row: str(row["DATETIME"].strftime("%H"))   , axis = 1)

plt.figure(figsize=(20,5))
accidents_df.groupby('HOUR')["HOUR"].count().plot.bar()
plt.grid(True)
plt.xlabel('Accident hours')
plt.ylabel('Count of accidents')
plt.title('Accident hotspots in a day')
plt.show()
```


Type your answer to the question in this cell How does the number of accidents vary throughout a single day?

There is an initial peak at 8 am, then the number of accidents starts increasing and get its highest peak during the day until 16, where it starts decreasing until midnight when has a low increase.

Part 3: Accidents by weekday

How does the number of accidents vary throughout a single week? Plot a bar graph based on the accidents count by day of the week.

Insert **answer** below:

In [13]:
```
# Code here
accidents_df["WEEKDAY"] = accidents_df.apply( lambda row: str(row["DATETIME"].strftime("%w-%A"))   , axis = 1)
#accidents_df["WEEKDAY"].head()
```

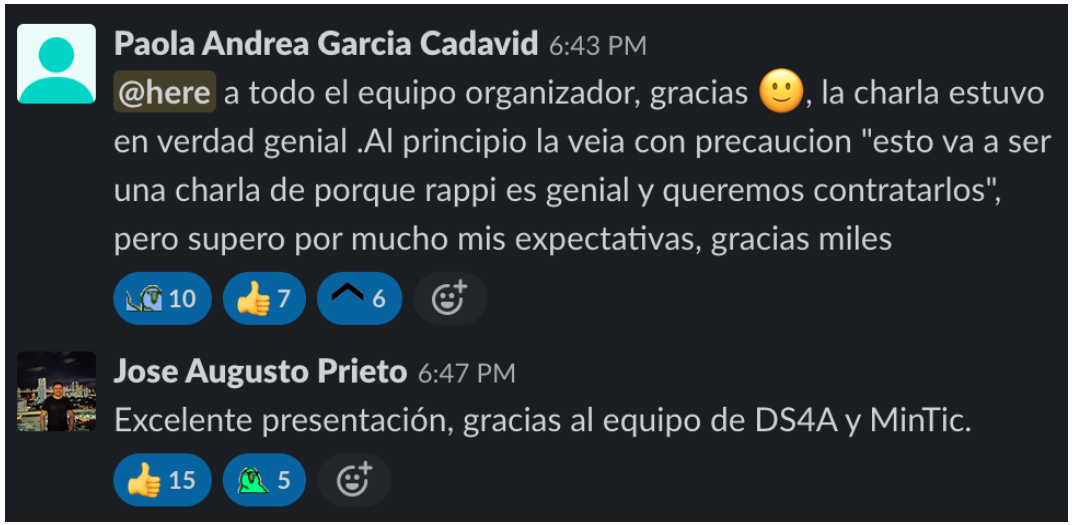
In [14]:
```
plt.figure(figsize=(20,5))
accident_count = accidents_df.groupby('WEEKDAY')["WEEKDAY"].count()
accident_count.plot.bar()
plt.grid(True)
plt.ylim(accident_count.min() *0.9 , accident_count.max()*1.1 )
plt.xlabel('Accident days')
plt.ylabel('Count of accidents')
```

Speaker Series for practical connections

Our speaker series has launched, creating value for participants and employers

IMPACT FOR PARTICIPANTS

- Participants interact directly with Colombian visionaries to learn about the power of data science



IMPACT FOR EMPLOYERS

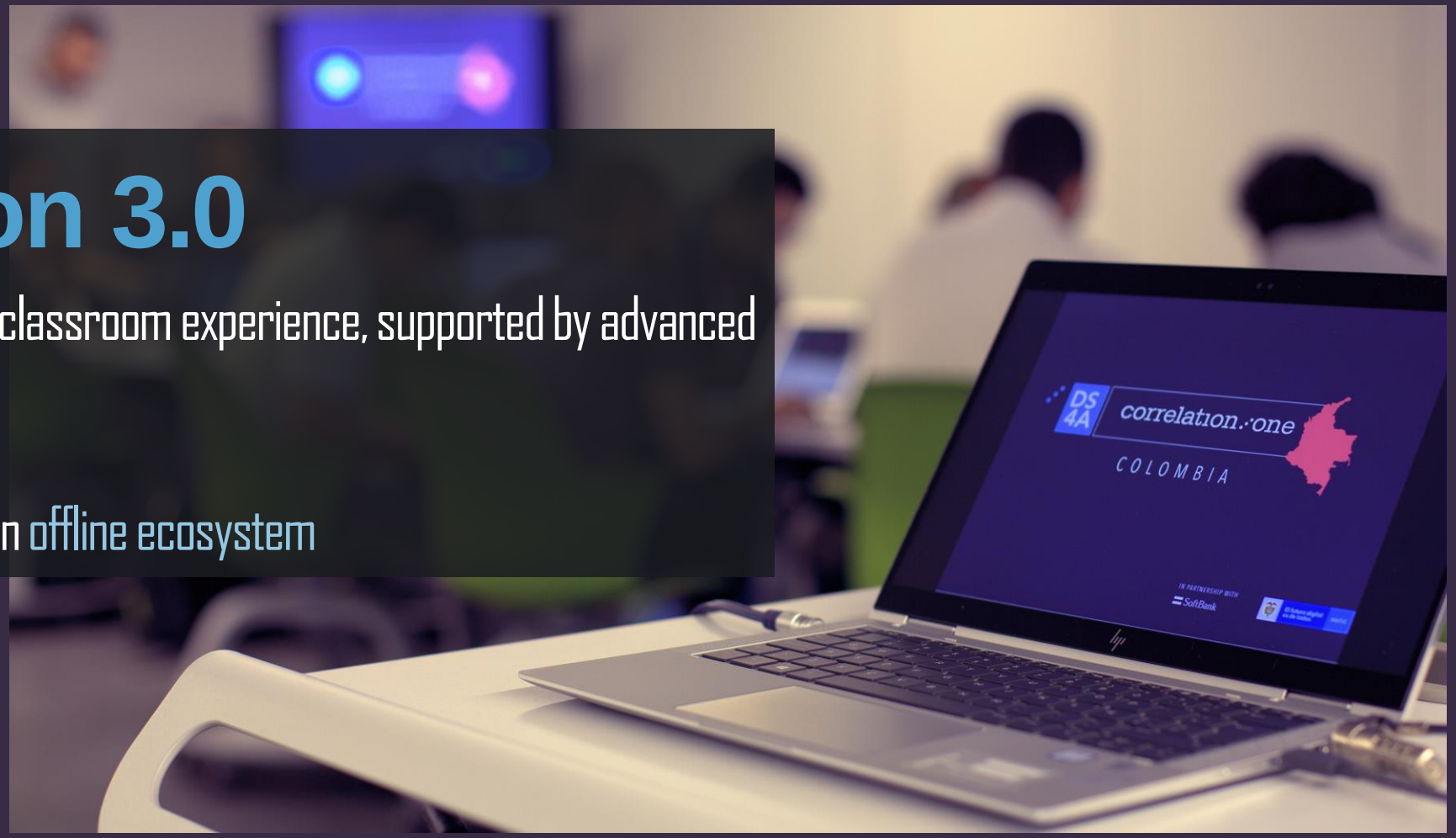
- Employers like Rappi engage Colombia's best talent and share their unique goals and needs

DS4A is a new model of education

Combining the best of in-person and online learning to create Education 3.0

Education 3.0

A distributed, interactive classroom experience, supported by advanced online functionality and engineered to seed an offline ecosystem



Class design

Classes feature in-person and virtual elements to enhance participant experience

IN-PERSON ELEMENTS

- Expert feedback
- Group work on real-world projects
- Personalized TA training
- Employer visits / networking

VIRTUAL ELEMENTS

- World-class instructors
- Customizable learning modules
- Platforms for peer networking
- Scalable

NEW ELEMENTS

- Calibrated to local region's employers
- Tied to ecosystem-building mandate

We can scale much larger in Colombia

Our Colombia program is easily scalable (5-10x+), and can include new "track" offerings

Geographic expansion

Grow in current cities and beyond, including in rural areas

"Track" expansion

Offer AI training for higher/ lower sophistication levels, managers, or engineers

We target an exponential impact

By giving resources to key stakeholders, DS4A impacts more than its participants

Teaching the teachers

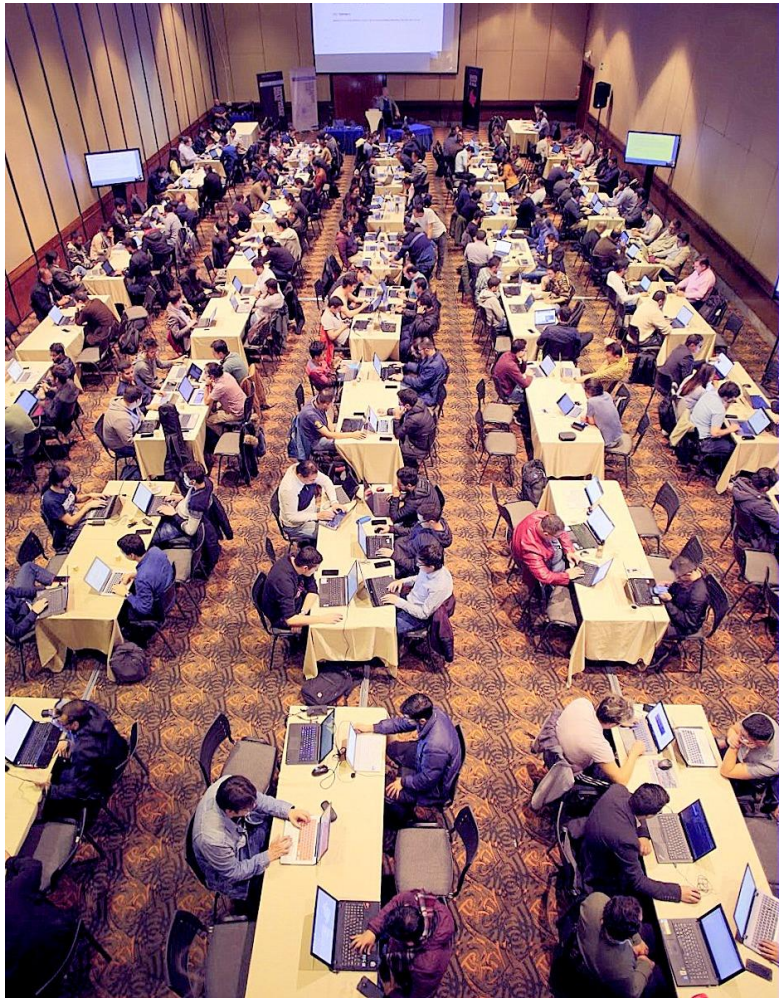
Training Colombia's data science professors, from universities like Pontificia Universidad Javeriana and Universidad Nacional

Partnering with employers

We build curriculum that is tailored to employers' needs, and engage businesses through Career Fairs and Speaker Series

Open sourcing curriculum

We are open sourcing our curriculum, so all Colombian students can access world-class training



Summary

Preparing for Colombia's digital future

AI is the future

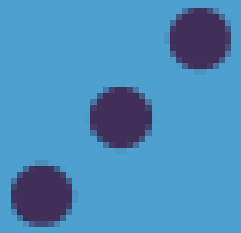
- AI is the defining force of 21st century economies – with it, new possibilities open for people across industries

Talent underpins AI

- Data science talent is the foundation of the organizations of the future, from startups to government agencies
- Countries must develop their local AI talent

DS4A is the solution

- DS4A trains participants and seeds a local AI ecosystem
- DS4A is making Colombia Latin America's leading digital economy



Thank you

correlation one