COURSERA CASE STUDY

Inclusive Products with Data Science

coursera

Applying Data Science to Build Inclusive Products

Coursera is the world's leading platform for higher education. Our learners come from diverse backgrounds and almost every country in the world. They differ widely in baseline abilities, goals, and learning styles. The rich data collected in the online medium, when harnessed through data science and machine learning, can be applied to build effective, scalable, and inclusive products that serve these diverse individuals' needs.

BENEFITS

With high quality data, we can benchmark the status quo and uncover where we can improve.

- Set goals to increase the number of female learners on the platform, the share of those who enroll in STEM content, and the quantity of all STEM enrollees who go on to successfully complete the course
- Learn definitively what solutions work for whom through experimentation
- Develop working hypotheses using inferential techniques on ways to improve our products and services to better meet the needs of our diverse learner base

IMPLEMENTATION

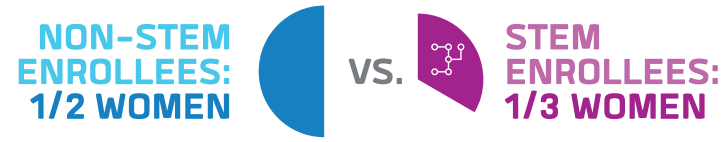
The journey to building inclusive products starts with understanding the status quo and uncovering where we can do better. Easy access to high-quality data — including socio-demographic data — is a central ingredient. At Coursera, data engineers collect, curate, standardize, and make available critical data to understand the status quo. This includes building demographic features such as gender into our core data lake.

Much of the core data lake is accessible to the full company – in anonymized form – via a third-party analytics tool. This is important because the best ideas, especially around questions of inclusiveness, will come from a diverse set of employees.

Understanding what is driving our observations is crucial for designing the best solutions. By combining observational data with causal inference techniques and experimentation methods, we obtain an initial understanding of the mechanisms at play and potential solutions. When experimental results indicate that different users benefit notably from different experiences, we use machine learning to build personalized products to meet individuals' unique needs.

IMPACT

Consider enrollments on Coursera by content domain and learner gender: While women account for almost half of enrollees in non-STEM fields, only about one-third of STEM enrollees are women.



In the raw data, we saw that when a course had a female instructor it had about a nine percentage point increase in the share of enrollees that were female. We wanted to know if having more female instructors, or marketing them better, would drive more women into STEM classes or whether the correlation was caused by something else (e.g., female instructors are more likely to teach in the domains and subdomains within STEM that already appeal to female learners). From the experiment, we learned that the effect of instructor gender on whether or not a learner chooses to enroll in any content in the field is substantial: Female learners who received an email about a machine learning specialization emphasizing a female instructor were 26 percent more likely to enroll in a STEM course on the platform than those who received an otherwise identical email about the same specialization emphasizing a male instructor. The instructor's gender had no impact on male enrollments.

ACTIONABLE RECOMMENDATIONS

Collect and curate key socio-demographic data and ensure broad access to this data, for example via self-serve tooling software.

Build a balanced data team – from data engineering to machine learning.

- Data engineers can drive the data curation, standardization, and democratization.
- Data scientists focused on decision science can develop working hypotheses from observational data and design and run experiments to validate those hypotheses.
- Data scientists focused on machine learning can, in cases where different users benefit from different experiences, build personalized products to meet users' unique needs.

Make experimentation easy, including set-up and analysis.

- Whether you go for a buy or build solution, invest in a robust AB testing platform.
- The platform will ideally allow folks around the company — from Project Managers to Engineers to Marketers — to easily launch new experiments in production.
- It will also make it easy to analyze and correctly interpret results.
- As part of that, incorporating segmentation into the reporting of results allows the end user to observe how the experiment performed along different cuts — like by geo or by gender.

Engage employees with diverse backgrounds on projects to gain the best ideas and the most critical thinking, especially around questions of inclusiveness and accessibility.

To ensure a collection of key data to measure inclusivity:

1. Identify what sociodemographic data are high value;

2. Define ownership and accountability for securing that data; and

3. Allow the owning team to figure out how to get it. For example, we received self-reported gender information when the individual created their user profile and we invested in inferring gender from a combination of first name, geo location, and language.

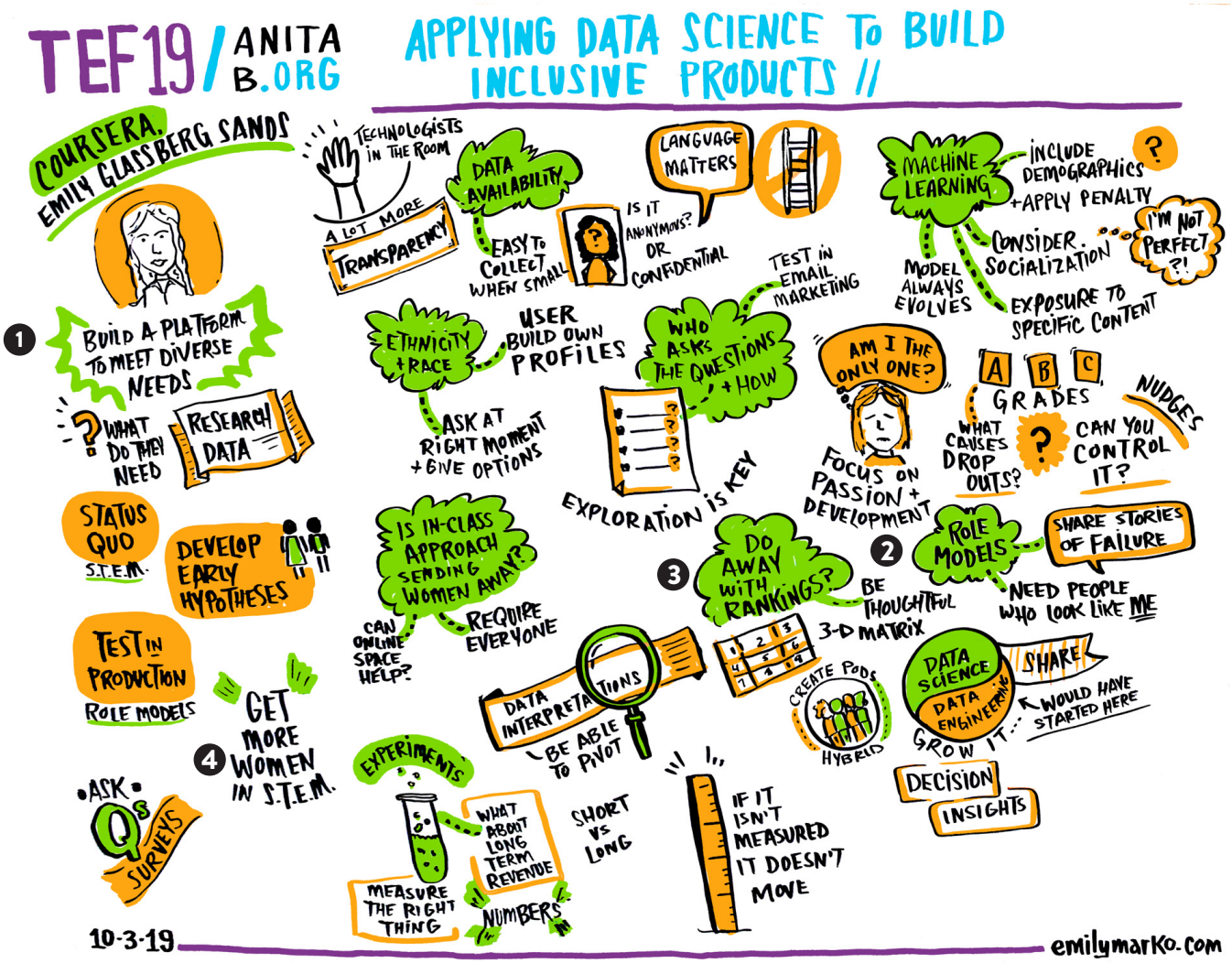
CONTRIBUTORS >

Emily Glassberg Sands, Vice President, Data Science, Coursera

Christine Chiu and **Whitney Walton,** AnitaB.org Partnership Success

*AnitaB.org is a social enterprise founded on the belief that women are essential to building technology the world needs. We envision a future where the people who imagine and build technology mirror the people and societies for whom they build it. Learn more about TEF and our other programs at **AnitaB.org/Events-Organizations***

GRAPHIC RECORDING OF EVENT



1 Build a platform to meet diverse needs

"Female learners who received an email about a machine learning specialization emphasizing a female instructor were 26 percent more likely to enroll in a STEM course on the platform than those who received an otherwise identical email about the same specialization emphasizing a male instructor."

2 Role models

"I'm an economist by training, and I think that's probably the reason I was at first rejected from Coursera. My primary advice to young women is to not get too lost in the data science, the math, and the algorithms and instead remember those are all means to an end and the end is impact. Consider the problems in the world that you care about, and remember that data science provides this vast set of tools that you can use to solve the ones you care about most."

3 Build more equitable credentials

"There's unequal access to the school you go to or the jobs you get. Part of the motivation behind skill scoring is to create more equitable, fair, and accessible signals for the labor market. Companies are hungry to find more diverse talent and skill scoring provides really valuable input. We are actually seeing it in use already from a lot of our Coursera for Business customers to identify who of their internal employees are well positioned for new opportunities or new roles."

4 Get more women in STEM

"It's not just on the individual to learn, but also on their companies to train them. Coursera for Business provides learning content through employers and governments to reach a wider swath of individuals who might not be able to afford training, re-skilling, and the opportunity for advancement."