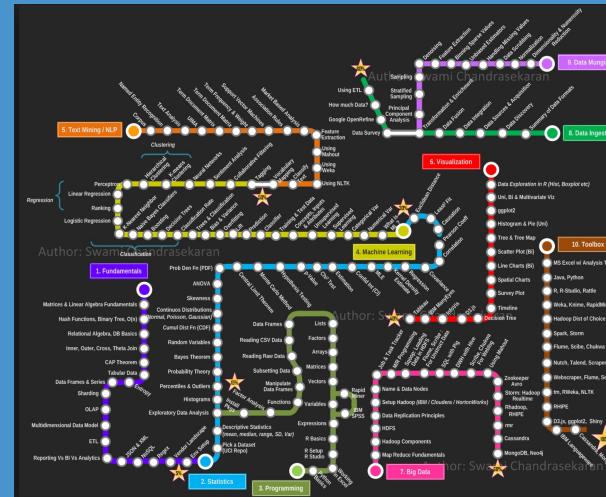


An Introduction to your Data Science career





Pinned Tweet



Josh Wills @josh_wills · 3 May 2012

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.



49



1.5K



1.1K

...



what is a data scientist



All

Videos

News

Images

Shopping

More

Settings

Tools

da·ta sci·en·tist

noun

a person employed to analyze and interpret complex digital data, such as the usage statistics of a website, especially in order to assist a business in its decision-making.

"Silicon Valley technology companies are hiring data scientists to help them glean insights from the terabytes of data that they collect everyday"

Data Science @ Coursera

- Decision Scientist
- Machine Learning Engineer
- Data Infrastructure Engineer
- Business Intelligence Engineer



Decision Scientist

*Develop & test hypotheses key
to product / business direction*

Methods:

- Experimental design
- Statistical modeling

Machine Learning Engineer

*Build data products powered
by ML models*

Methods:

- ML
- Software Engineering

Data Infrastructure Engineer

Build the data infrastructure pipeline and develop analytic tooling

Methods:

- Software Engineering
- Tool Development

Business Intelligence Engineer

Build and develop the data model and tools to access it

Methods:

- Data Modeling
- Software Engineering

DI Engineer

BI Engineer

Data Warehouse

Decision Scientist

ML Engineer

Discovery of Data Insight

Quantitative data analysis to help steer
strategic business decisions

Development of Data Product

Algorithm solutions in production, operating at scale
(e.g. recommendation engines)

Business Value

What Skills Do You Need?

Table stakes - Have a broad base. Prereqs and core curriculum will provide most of these.

Mathematics:	Statistics:	Programming:	Machine Learning:	Business/Product
Calculus	Hypothesis Testing	SQL	Supervised and Unsupervised Learning	Intuition:
Linear Algebra	Regression	R/Python	Model Fitting	Interpret and communicate results to non-technical audience

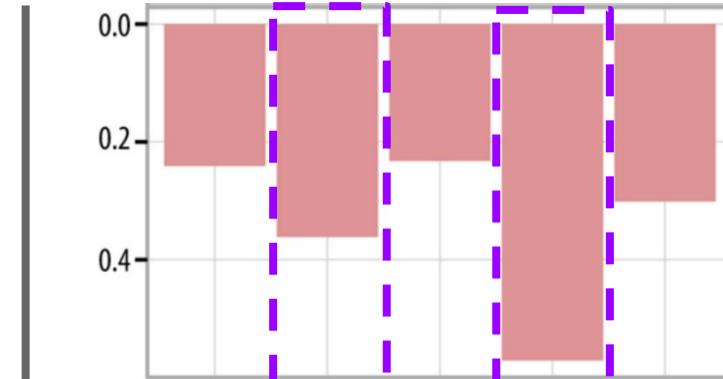
Then depending on area of interest, specialize in one or more of these buckets.
Choose electives to give an in depth knowledge in chosen area(s).

What's your Flavor?

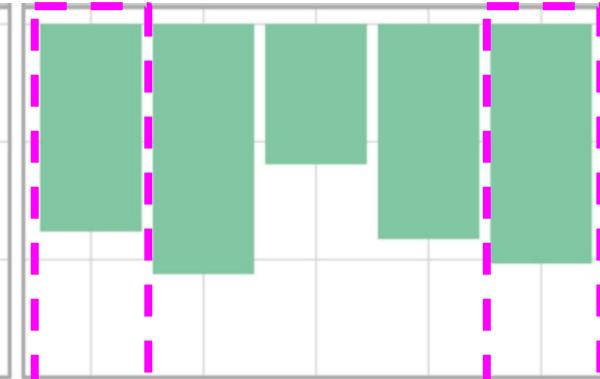
	Decision Scientist	ML Engineer
One of my favorite subjects is...	Statistics	Machine Learning
I also enjoy...	Experimentation	Coding
At night I dream about...	Analyzing data	Building things
I love thinking like & working with...	Product managers, Business people	Back-end engineers

Increasing
skill

Decision Scientist



ML Engineer



Programming
Statistics
Math/OR
Business
ML/Big Data

Skill Group

From
Analyzing
the
Analyzers

coursera |

Which Classes?

Decision Science	ML Engineer
CS 498 Applied Machine Learning	CS 498 Applied Machine Learning
STAT 420 Methods of Applied Statistics	STAT 420 Methods of Applied Statistics
STAT 571 Multivariate Analysis	CS 410 Text Information Systems
<u>Experimental Design / AB Testing</u>	STAT 542 Practical Statistical Learning
<u>Causal Inference / Econometrics</u>	CS 498 Cloud Computing Applications

Tips

Tip #1: Identify the area of data science you want to work in and tailor your curriculum to give you expertise in the necessary skills.

Tip #2: Build a portfolio of practical and applied projects to get relevant experience in your area of interest. This should involve communication.

Tip #3: Don't worry about picking an area and being stuck there. You can always pivot later. Hence try to blend your interest with ability.

Becoming a Data Scientist

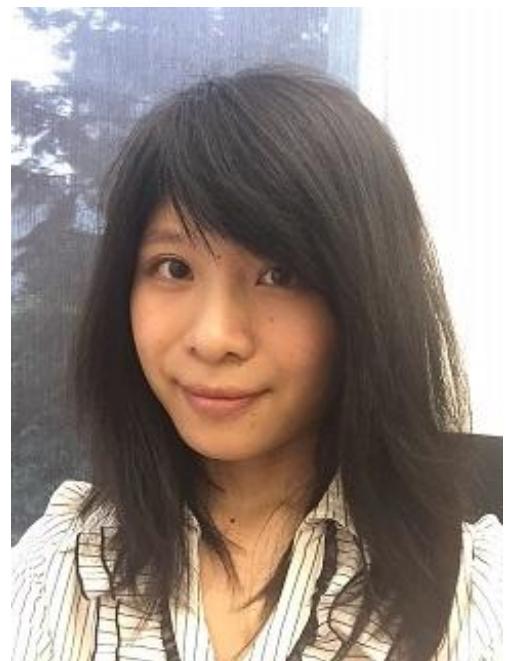
Miaomiao Wen

Data Scientist @ Coursera



About me

- PhD@Carnegie Mellon
Computer Science
- DS@Learning Analytics



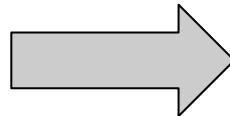
Online education
Natural language processing

School → Industry

In this talk,

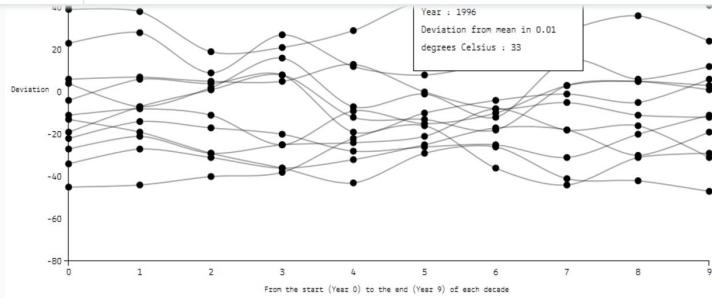
1. My experience transitioning from school to a data scientist role at Coursera
2. Tips on how to prepare for data scientist interviews

Transition



Data science projects

coursera



Explanation:

This is a sample submission of a visualization of the data from the GISTEMP site. Every line cor



Assignments and course projects

Make a measurable difference

coursera

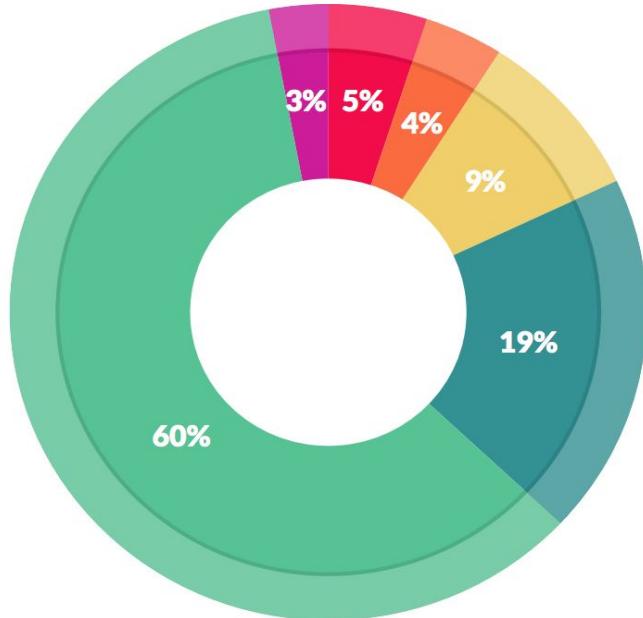
Compare & Contrast

Course project

Clean
datasets

Industry

Too much
data,
where to
start?



What do data scientists spend their time doing?

Collecting Data

Cleaning Data

Compare & Contrast

Course project

Clean datasets



Well-defined questions

Industry

Too much data,
where to start?



Which project is
more impactful?

Compare & Contrast

Course project

Clean datasets



Well-defined questions



Which methods were taught?

Industry

Too much data, where to start?



Which project is more impactful?



Which methods are most effective?

Compare & Contrast

Course project

Clean datasets

Well-defined questions

There's usually a solution.

Which methods were taught?

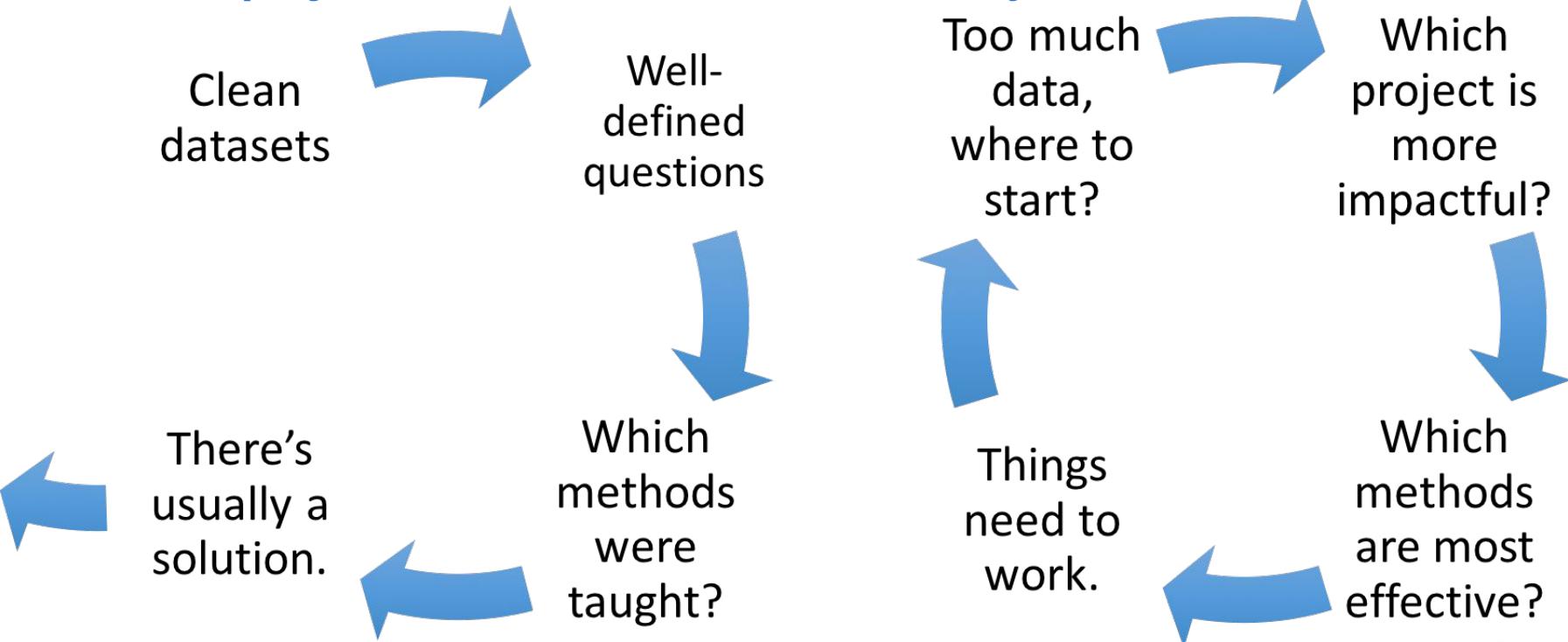
Industry

Too much data, where to start?

Things need to work.

Which project is more impactful?

Which methods are most effective?



Programming



Java™



JavaScript



I Py:
IPython



Learning on the job



coursera

< Back to Week 2 X Lessons

Two Sample Binomial Tests

Relative Risks & Odds Ratios

- Relative Risks & Odds Ratios - Relative Measures 4 min
- Relative Risks & Odds Ratios - The Relative Risk 8 min
- Relative Risks & Odds Ratios - The Odds Ratio 7 min

Delta Method

Review

MATHEMATICAL
BIOSTATISTICS
PART 2
Lecture 5:
Relative
Risks and
Odds
Ratios

Eric Cai's
BioStat
classmate
for winter
2014

Example

For the relative risk, $\hat{p}_A = 11/20 = .55$, $\hat{p}_B = 5/20 = .25$

$\hat{RR}_{A/B} = .55/.25 = 2.2$

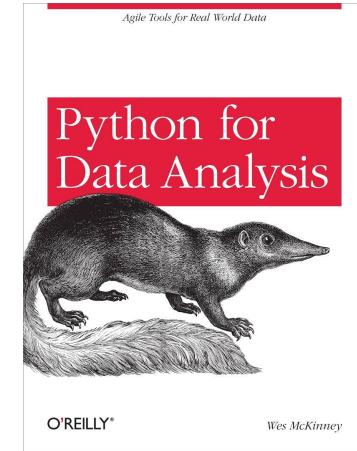
$\hat{SE}_{\log \hat{RR}_{A/B}} = \sqrt{\frac{1-.55}{.55 \times 20} + \frac{1-.25}{.25 \times 20}} = .44$

Interval for the log RR: $\log(2.2) \pm 1.96 \times .44 = [-.07, 1.65]$

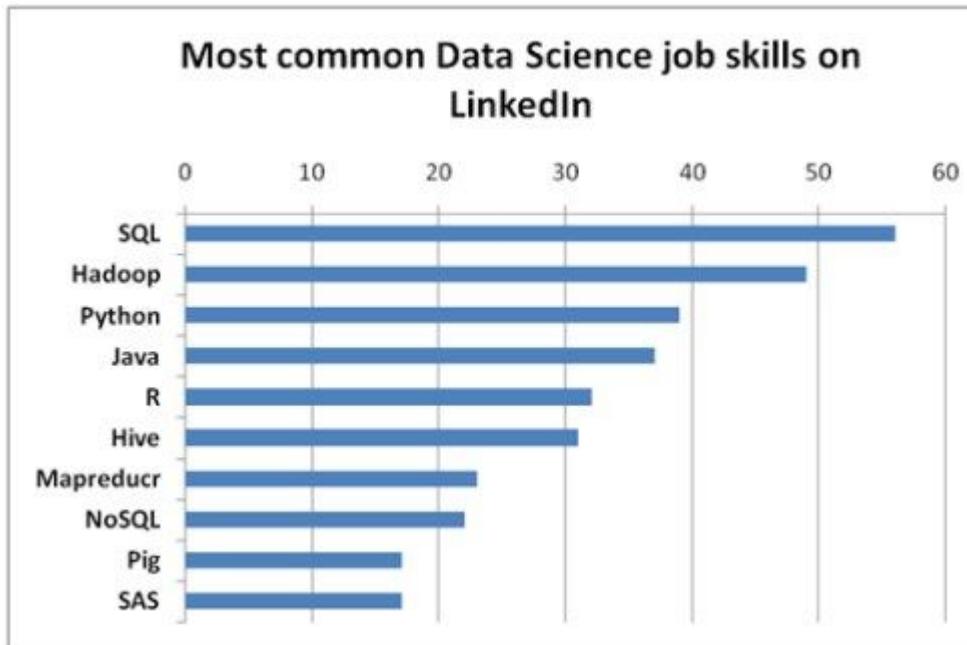
Interval for the RR: [.93, 5.21]

0:01 / 7:19

Relative Risks & Odds Ratios - The Odds Ratio



How I prepared for DS interviews



Two companies' DS interviews

Company X

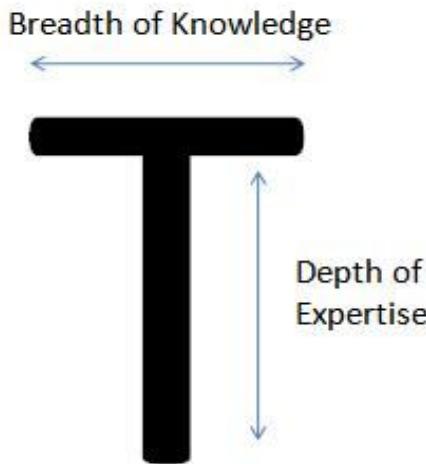
- SQL data wrangling question
- SQL + visualization
- Probability (dice rolling)
- Metric design
- Website feature analysis
- Hypothesis generation

Company Y

- Data analysis exercise
- Regression analysis
- Probability (dice rolling)
- Whiteboard coding
- Machine learning
- A/B testing, experiment design

How I prepared for DS interviews

<https://leetcode.com/>
(whiteboard coding)



About This Specialization

Courses

Pricing

Creators

FAQs

Data Science at Scale Specialization

\$49.00 USD per month

Enroll

Financial Aid is available for learners who cannot afford the fee. [Learn more and apply.](#)

Tackle Real Data Challenges

Master computational, statistical, and informational data science in three courses.

About This Specialization

Learn scalable data management, evaluate big data technologies, and design effective visualizations.

This Specialization covers intermediate topics in data science. You will gain hands-on experience with scalable SQL and NoSQL data management solutions, data mining algorithms, and practical statistical and machine learning concepts. You will also learn to visualize data and communicate results, and you'll explore legal and ethical issues that arise in working with big data. In the final Capstone Project, developed in partnership with the digital internship platform Coursolve, you'll apply your new skills to a real-world data

What are your strengths and weaknesses?

Coming from an engineering background?



Coming from maths, statistics, science background?



Tips

Tip #1: Interviews look different across companies, and across roles within company

Focus on preparing yourself based on your priorities

Tip #2: Get your hands dirty

Practice your skills using real, messy data. Find a mentor and get expert feedback on your work.

Tip #3: Develop your business savvy

Domain knowledge, Metrics, Hypothesis, Intuition

Q&A

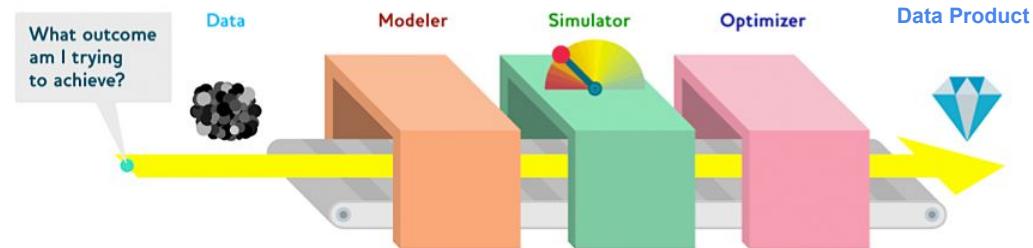
Keep it simple
Capstone, Kaggle

Typical DS projects

Decision science



Data Product



A photograph of three people standing side-by-side against a light blue background. On the left is a young man with short dark hair, wearing a dark t-shirt. In the center is a man with glasses and short dark hair, wearing a dark hoodie over a red shirt. On the right is a woman with long dark hair and glasses, wearing a denim jacket over a dark top.

AB Testing our way to Better Products

Katherine Wong

Data Scientist @ Coursera

The Road Not Taken



Business Case

Validate Ideas

Test hypotheses
informing product /
business direction

Measure Impact

Understand what users
will do by observing
their actual behavior

Improve Ideas

“Smart iteration” on
features



Example: Choosing Shirt Colors



Group A



Group B

Group B



Group A



Example: Onboarding

Onboarding Survey

Welcome to Coursera
No matter what you want to do, we have the courses to get you there.

What's your main goal?

- Advance my career
- Start a new career—I'm starting from scratch
- Start a new career—I have some of the skills
- Advance my education
- Learn for fun

Enrollments

Updates 1

Accomplishments

Recommendations

Your Recommendations

Start a new career

Personal Development Probability and Statistics Machine Learning

-  Photography Basics and Beyond: From Smartphone to DSLR
Michigan State University
4-courses [Enroll](#)
-  Academic English: Writing
University of California, Irvine
4-courses [Enroll](#)
-  Career Success
University of California, Irvine
9-courses [Enroll](#)
-  Coaching Skills for Managers
University of California, Davis
4-courses [Enroll](#)
-  Inspirational Leadership: Leading with Sense

Example: Onboarding

Onboarding Survey

Welcome to Coursera
No matter what you want to do, we have the courses to get you there.

What's your main goal?

- Advance my career
- Start a new career—I'm starting from scratch
- Start a new career—I have some of the skills
- Advance my education
- Learn for fun

Loading Page?

Personalizing for Meng Tran.
We're selecting the best courses for you on

Marketing specialist
Data analyst



Recommendations

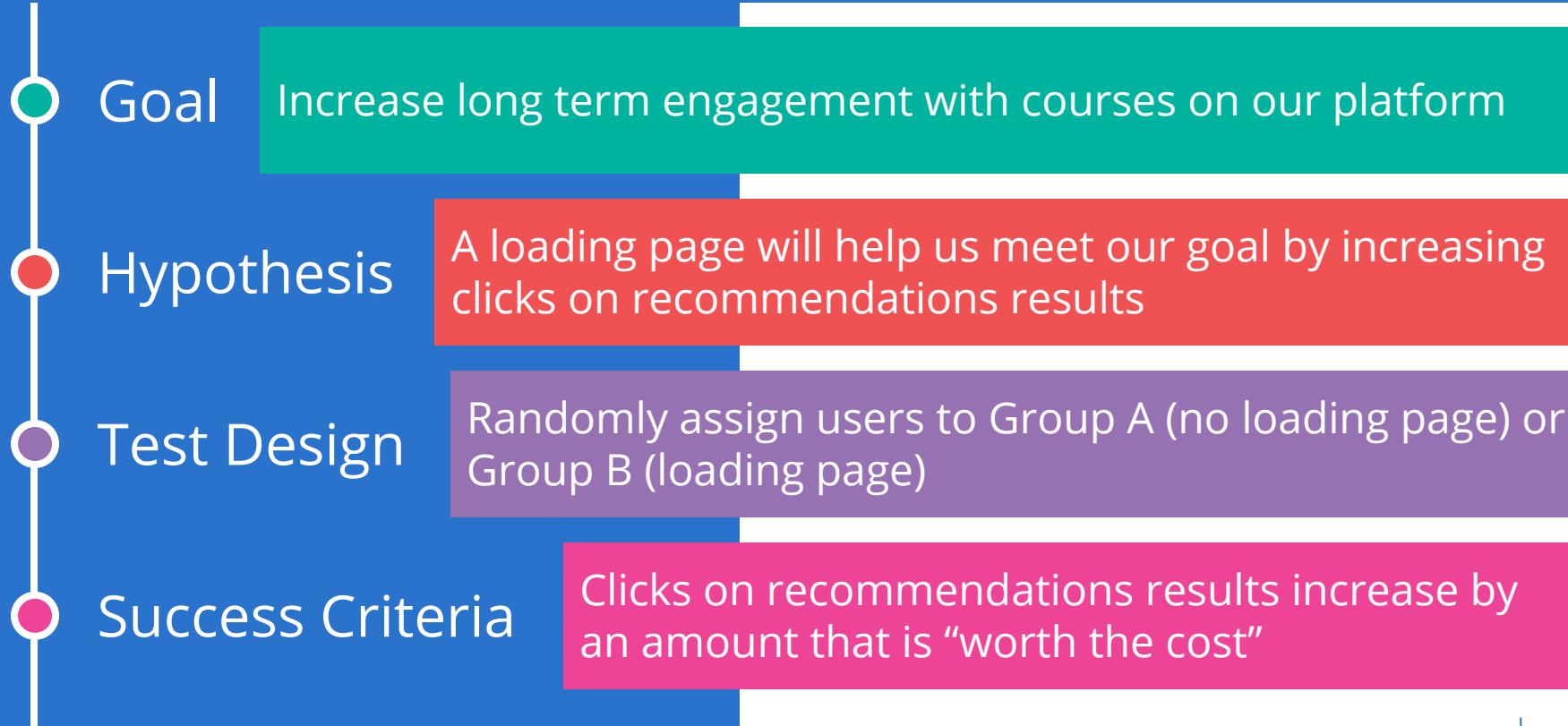
Your Recommendations

Start a new career

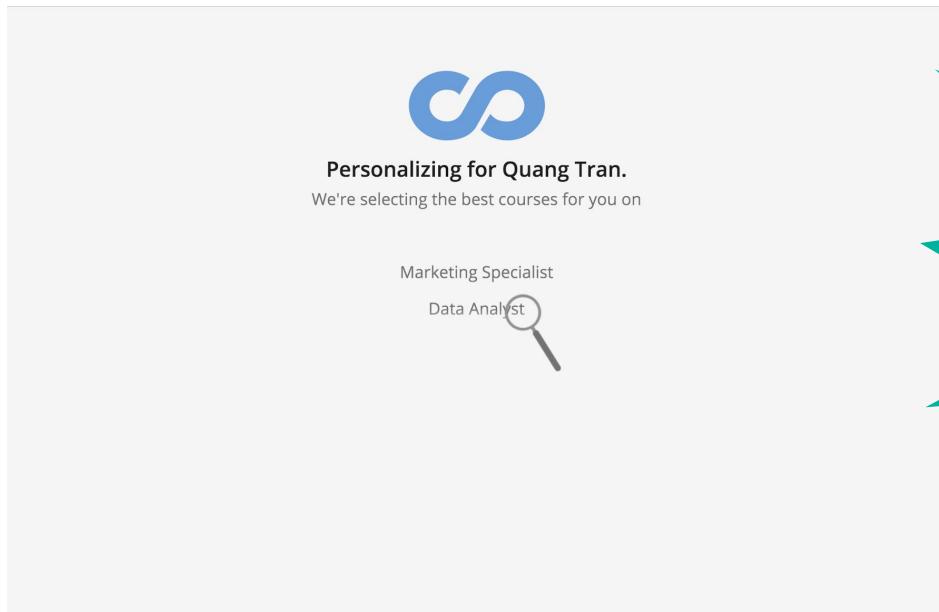
Personal Development Probability and Statistics Machine Learning

Course Title	University	Courses	Action
Photography Basics and Beyond: From Smartphone to DSLR	Michigan State University	4-courses	Enroll
Academic English: Writing	University of California, Irvine	4-courses	Enroll
Career Success	University of California, Irvine	9-courses	Enroll
Coaching Skills for Managers	University of California, Davis	4-courses	Enroll
Inspirational Leadership: Leading with Sense	University of California, San Diego	4-courses	Enroll

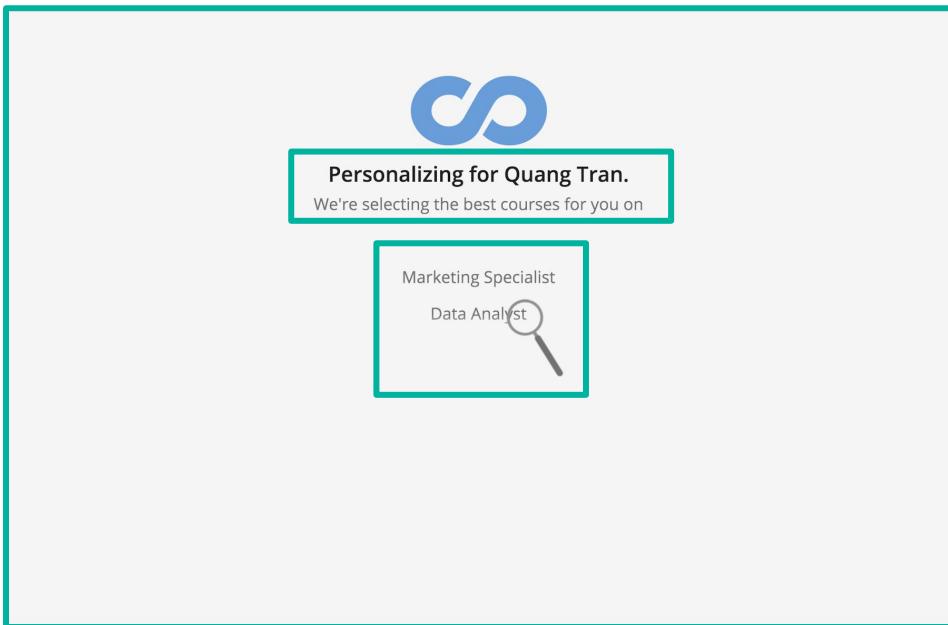
Example: Onboarding



Example: Onboarding



Example: Onboarding



Mechanisms

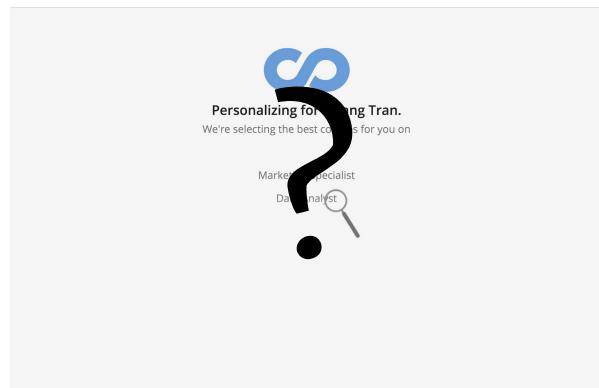
- New loading page
- Personalized copy
- Reminder of which recommendations are being generated

Example: Onboarding

Search Query

The screenshot shows the Coursera homepage. At the top, there is a navigation bar with links for 'Catalog', 'Search catalog' (which has a magnifying glass icon), 'Institutions Log In', and 'Sign Up'. Below the navigation bar, there is a large teal banner with the text 'Take the world's best courses, online.' and a 'Join for free' button. The Coursera logo is in the top left corner. At the bottom, there are logos for various universities: Penn, Johns Hopkins University, University of Michigan, and Stanford.

Loading Page?



Search Results

The screenshot shows the Coursera search results page for 'data science'. At the top, there is a search bar with 'Catalog' and 'data science' entered, along with a magnifying glass icon. Below the search bar, it says 'You searched for data science. 718 matches' and 'Active filters: English'. There are sections for 'Course Languages', 'Degrees', 'Subtitle Languages', and 'Courses and Specializations'. The 'Courses and Specializations' section highlights the 'Master of Computer Science in Data Science (MCS-DS)' from Johns Hopkins University, which is described as an 'Accelerate your career with a flexible, fully-accredited professional Master of Computer Science in Data Science from one of the world's top programs.' The Coursera logo is at the bottom right.

Tips

Tip #1: Leverage AB testing to inform direction and strategy

Tip #2: Align on the problem you're trying to solve before running the test

Tip #3: Dig into mechanisms to better understand the bounds of what the test measures