Data Science Bootcamp

Make Data Work
Oct 15–17, 2014 • New York, NY

@laurieskelly  @thisismetis  @dsatweet
Metis Data Science Bootcamp

In progress:

Cohort 1
Autumn 2014, Manhattan
Week 7 of 12
Data Science Bootcamp: the project specification
Data Science Bootcamp: the project specification
Bootcamp: the project specification
Bootcamp: the project specification
Bootcamp: the project specification

bootcamp

Ready-to-hire entry-level data scientists
Bootcamp: the project specification

Ready-to-hire entry-level data scientists
Bootcamp: the project specification
Bootcamp: the project specification

...and launch in September
Refining the goal

Maximize hiring rate

Ready-to-hire entry-level data scientists
Refining the goal

Train people we would hire

Ready-to-hire entry-level data scientists
Refining the goal

Maximize hiring rate

Train people we would hire

Ready-to-hire entry-level data scientists
So... what is a data scientist?
So... what is a data scientist?
So... what is a data scientist?
What is a data scientist?

Skill domains

- Statistics / Machine Learning
- Programming
- Communication
- Design
Who can be a data scientist?
Can anyone be a data scientist?
How did I get here?
Who can be a data scientist?
Who can be a data scientist?

Skill domains

- Statistics / Machine Learning
- Programming
- Communication
- Design
Who is ready to be a data scientist?

Skill domains

Statistics / Machine Learning
Programming
Communication
Design
Fighting the Dunning-Kruger effect
Ultimate inspiration
Ultimate inspiration

EUPA Player Rank = (Catching + Throwing + Athletic + Experience) / 4

A) Catching Ability

1 I have caught a Frisbee before.
2 I can catch a Frisbee thrown to me in practice most of the time.
3 I may have trouble catching throws that aren’t right to me.

*I CATCH MOST THROWS THAT ARE TO ME AT CHEST HEIGHT*
4 I can catch most of the time standing and often while running, below my knees and above my head.
5 I can catch running most of the time, and with my weak hand; I still get the dropsies.
6 I can catch most things thrown to me, some things thrown near me, and jump and catch at the peak of my jump.

*I LAYOUT ON OFFENSE*
7 I expect to catch throws to me, although I might not every time. I run through the disc and often catch layout bids.
8 I read, box, and sky. I expect to catch difficult throws, even if they are fast or not right to me.
9 I expect to catch all throws, and make it look easy.
Ultimate inspiration

Rate your overall programming skills and experience (choose the highest number which accurately describes you) *

- 0 - I have never written one line of code in any language.
- 1 - I have taken one college or online programming course, but I have not done any coding beyond the homework for the course.
- 2 - I have taken more than one programming course, or taught myself to code, and I have done “independent” coding projects that were related to this training.
- 3 - I have sometimes written code to solve problems in my work, in my research, or for personal passion projects (not directly related to coursework/training).
- 4 - Writing code is a regular aspect of my job, my research, or my personal fun time.
- 5 - I use code regularly and skillfully, and I have taught or trained others to write code.
Who is ready to be a data scientist?

Personality Traits
Who is ready to be a data scientist?

Personality Traits

- Curiosity
- Creativity
- Grit
Who is ready to be a data scientist?

Personality Traits

Compulsive, pathological curiosity
Creativity
Grit
Who is ready to be a data scientist?

Personality Traits

Compulsive, pathological curiosity
Drive to create
Grit
Who is ready to be a data scientist?

Personality Traits

- Compulsive, pathological curiosity
- Propulsive laziness
- Drive to create
- Grit
Who is ready to be a data scientist?

Personality Traits

- Compulsive, pathological curiosity
- Propulsive laziness
- Drive to create
- Irritable determination
Who is ready to be a data scientist?

**Personality Traits**

- Compulsive, pathological curiosity
- Propulsive laziness
- Drive to create
- Irritable determination
- Insensitivity to pain
Who is ready to be a data scientist?

Personality Traits

- Compulsive, pathological curiosity
- Propulsive laziness
- Drive to create
- Irritable determination
- Insensitivity to pain
- Integrity
Who is ready to be a data scientist?

Personality Traits

- Compulsive, pathological curiosity
- Propulsive laziness
- Drive to create
- Irritable determination
- Insensitivity to pain
- Integrity
- Humility
Who are bootcamps suited for?

- Focused time and cost
- Peer and community support
- Placement manager, hiring partners
- Expert practitioner-instructors

How do I become a data scientist? An evaluation of 3 alternatives

Master’s degree vs bootcamp vs self-taught (MOOCs). Bootcamps are best, but perhaps we’re biased.
Who enrolled?

Autumn 2014 cohort
Bootcamp: designing the program

???

Ready-to-hire entry-level data scientists
Program development

Bootcamp model

Training
Repetition
Culture/community
Training

Skills

Obtain
Scrub
Explore
Model
iNterpret
Training

Skills

- Obtain
- Scrub
- Explore
- Model
- Interpret
- Communicate
Training

Skills → tools

Data acquisition
Data exploration
Machine learning & statistics
Computer science
Web
Visualization
Domain awareness
Training

Tool selection

Comfort & facility with a core toolbox
Awareness of options
Some curveballs
Training

Tools

Python when possible
Training

Tools

Python, ipython notebook, git/github
Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
Training

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(flatfiles), MySQL, mongoDB
Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(Flatfiles), MySQL, MongoDB
numpy-scipy-matplotlib, pandas
Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron
(MySQL, mongoDB
numpy-scipy-matplotlib, pandas
statsmodels, scikit-learn
Training

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron (flatfiles), MySQL, mongoDB
numpy-scipy-matplotlib, pandas
statsmodels, scikit-learn
HTML, CSS, js, d3.js

Tools
Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron (flatfiles), MySQL, mongoDB
numpy-scipy-matplotlib, pandas
statsmodels, scikit-learn
HTML, CSS, js, d3.js
DigitalOcean, Flask
Training

Tools

Python, ipython notebook, git/github
BeautifulSoup, selenium, cron (flatfiles), MySQL, mongoDB
numpy-scipy-matplotlib, pandas
statsmodels, scikit-learn
HTML, CSS, js, d3.js
DigitalOcean, Flask
google
Repetition

Project-based model

“DATA SCIENCE PROJECT”
Repetition

Project-based model

Unit of repetition should be *the practice of doing data science*
Repetition

Project-based model

Unit of repetition should be *the practice of doing data science*
Repetition

Partitioning 12 weeks
Repetition

Partitioning 12 weeks
Culture & Community

Must overcome: Impostor syndrome

Perfectionism
Culture & Community

Must overcome:

Impostor syndrome
Group work
Communication

???
Culture & Community

Must overcome:

Perfectionism
- unfairly short deadlines
- jumping into the unfamiliar
- presenting projects
Culture & Community

Becoming data scientists

- Mixture of group and individual work
- Guest speakers and meetups
- Hiring partners
“I wouldn’t hire me...”
Data Science Bootcamp

Make Data Work
Oct 15–17, 2014 • New York, NY

@laurieskelly  @thisismetis  @dsatweet