Lessons learned from teaching Data Science

Gianluca Campanella

21st April 2018

Hello!

My name is **Gianluca** [dʒanˈluːka]

What I do nowadays

I'm a Data Scientist at

Microsoft

in **Algorithms and Data Science**

What I do nowadays

I also run my own company

that provides

Data Science training and mentoring

How I got into teaching

It all started at

Imperial College London

Today I want to talk to you about…

1. The skills gap in Data Science
2. The Data Science potential
3. How we implement all this

What I mean by ‘Data Science’

From S. Geringer (originally from D. Conway)

What I mean by ‘Data Science’

Data-driven decision-making

- Focus is on the problem-solving process
- Multidisciplinary but domain-centric
- Tools are secondary

The Data Science skills gap

The shortage of data scientists is becoming a serious constraint in some sectors.

— T. H. Davenport and D. J. Patil
Harvard Business Review (2012)

How do we close the skills gap?

Higher Ed

‘Traditional’ degrees

- Lots of theory
- Take a while to catch up
- More recognition?

Up-skilling

Bootcamps, MOOCs…

- Mostly hands-on
- Adapt faster
- ‘Show your skills’

How do we close the skills gap?

How do we ensure…

- Relevance?
- Quality?
- Consistency?

The *actual* question

How do we realise
the potential of Data Science?

The *actual* question

~~How do we realise~~

What even is
the potential of Data Science?

The Data Science potential

Data Science promises…

- Automation
- Risk minimisation
- Innovation

How do we realise this potential?

Data Science needs to be

embedded

within companies and processes

How do we realise this potential?

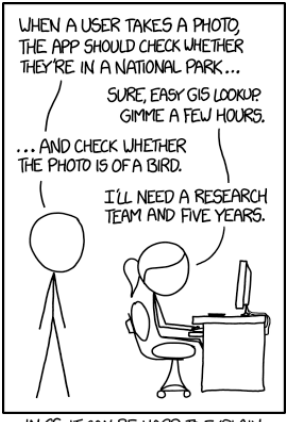
This means…

- Build, don’t buy
- Cultivate internally, don’t outsource
- Humans in the loop

How do we realise this potential?

1. Numeracy
2. Culture
3. Adoption

Numeracy

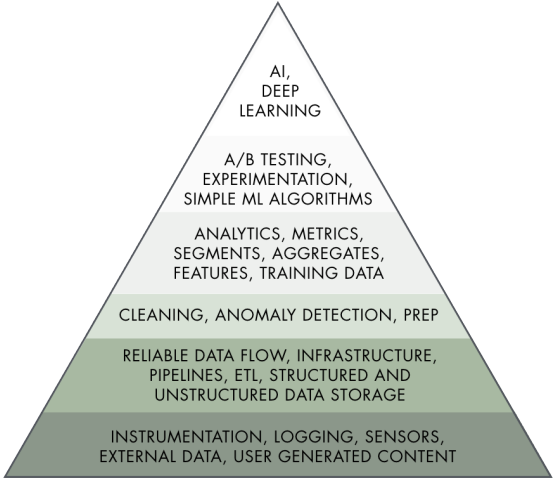


From xkcd

Numeracy

- Realise what's possible
- Determine existing capacity
- Understand the Data Science process

Don’t try to run before you can walk



From M. Rogati

Culture

The ROI of Data Science projects
is very difficult to predict!

- Power law-like distribution of returns
- Failure is *always* an option

Culture

Embrace a

high-risk, high-reward innovation culture

- Iterate quickly $\rightarrow$ fail fast
- Operationalise

Adoption

If it’s not used in production…

Adoption

If it’s not used in production…

It never happened!

How to realise the Data Science potential

- Embed Data Science starting at the top
- Build and re-build… fast
- Actually use it!

How to realise the Data Science potential

How?

How to realise the Data Science potential

You need
good people and **good teams**

Attracting and retaining good people

- Have a roadmap
- Hire for potential
- Let them choose their tools
- Give them resources
- Nurture their curiosity

Training good people

Data Analyst

- Understands the business
- Values automation

→ Teach them the pragmatism of the Software Engineer

Software Engineer

- Understands the tech
- Knows automation

→ Train them to recognise what matters to the business

Training good people

Data Analyst

- Understands the business
- Values automation

→ Teach them the pragmatism of the Software Engineer

Software Engineer

- Understands the tech
- Knows automation

→ Train them to recognise what matters to the business

Are we up-skilling them properly?

Teamwork

A successful Data Science team is…

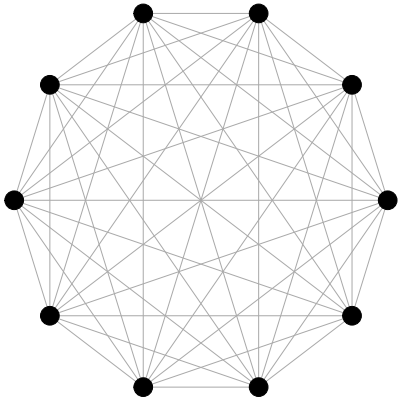
- Diverse
- Flexible
- Collaborative

Communication complexity

Communication complexity is

quadratic

in the number of team members



Communication complexity

Make sure there's
shared understanding

Particularly of…

- The Data Science workflow
- Software Engineering best practices

Take-aways

- Embed Data Science and its process
- Have a roadmap
- Don’t look for unicorns