Introduction to Data Science

Gianluca Campanella

Contents

What is Data Science?

What is Data Science?

Mathematics and
Statistics / Operational Research

Computing and
Software Engineering

Visualisation and
Communication Skills

Domain expertise

What is Data Science?

A problem-solving approach
based on the scientific method

What does Data Science deal with?

Problems!

Can we improve…

- The quality of offers we send to our customers?
- Road safety?
- How we identify people at high risk of cancer?

What does Data Science deal with?

Predictions?

How likely…

- Is a customer to respond to some offer?
- Are traffic accidents to occur in a certain area?
- Is a person to develop cancer in the next 10 years?

What does Data Science deal with?

Mechanisms?

Why…

- Does a customer decide to respond to some offer?
- Do traffic accidents occur regularly in certain areas?
- Do people develop cancer?

What is Data Science?

Statistics

- Predates computers
- Understand why something happens in the face of uncertainty

Machine Learning

- ‘Algorithmic modelling’ (L. Breiman)
- Computers can learn rules without explicit programming

Deep Learning

- Less structured inputs
- Computers can learn structure without explicit programming

What is Data Science?

	Predictions	Mechanisms
Analysis	**Descriptive** What's happening?	**Diagnostic** Why is it happening?
Building	**Predictive** What's likely to happen?	**Prescriptive** What do I need to do?

Recap

Data Science is…

- Evidence-based problem solving and decision-making
- Multidisciplinary but domain-driven
- Analysis-focused or building-focused

Who is a Data Scientist?

Who is a Data Scientist?

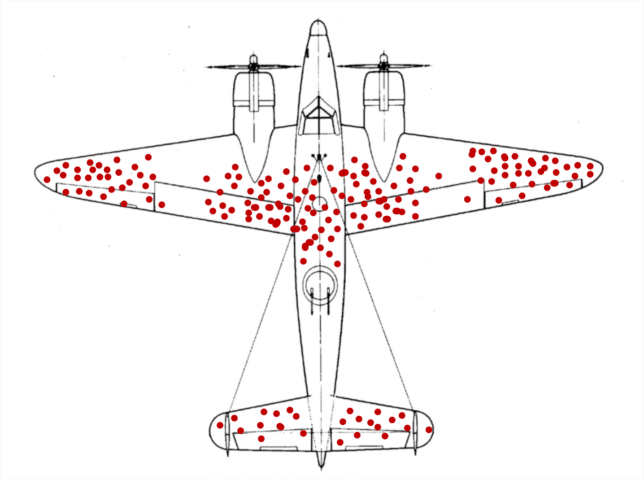
Someone who can…

- Get a ‘feel’ for the data
- Communicate effectively
- Work well in a team

What's this 'feel' for the data?

- Passion for the domain
- Curiosity about the data
- Intuition and creativity
- Common sense
- Rigour and accuracy
- Relevance

What's this 'feel' for the data?

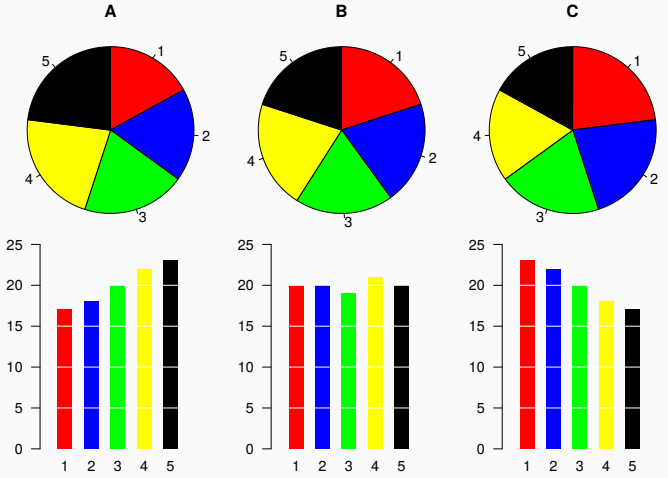


Via *Wikimedia Commons*

How do I communicate effectively?

- Condense findings into recommendations
- Use storytelling techniques and visual aids
- Understand limitations and don’t overstate results

How do I communicate effectively?



Via *Wikimedia Commons*

Recap

A successful Data Scientist…

- Is insatiably curious — and a bit stubborn!
- Never stops learning
- Is a practical, impact-driven, dependable person
- Can tell a story
- Knows the limitations of Data Science