A Data Ninja Rockstar Is Not Enough

Jodie Burchell
Pasha Finkelshteyn

Who works with data?

Who works with data?

Of course, data scientists!

Who works with data?

Of course, data scientists! (not only)

Reality is

Nobody should or even can do everything

Responsibilities should be distributed

Some responsibilities can be "outsourced" to SaaS's

Example project

Setup: Large fashion e-com
Trasos

Task: recommender system

What business says:

- Increase retention
- Reduce cost of acquisition

What business means?

Clients be like

Business outcomes

- It should be easier to convert non-customers to customers
- Customers and should return sooner

Technical constraints

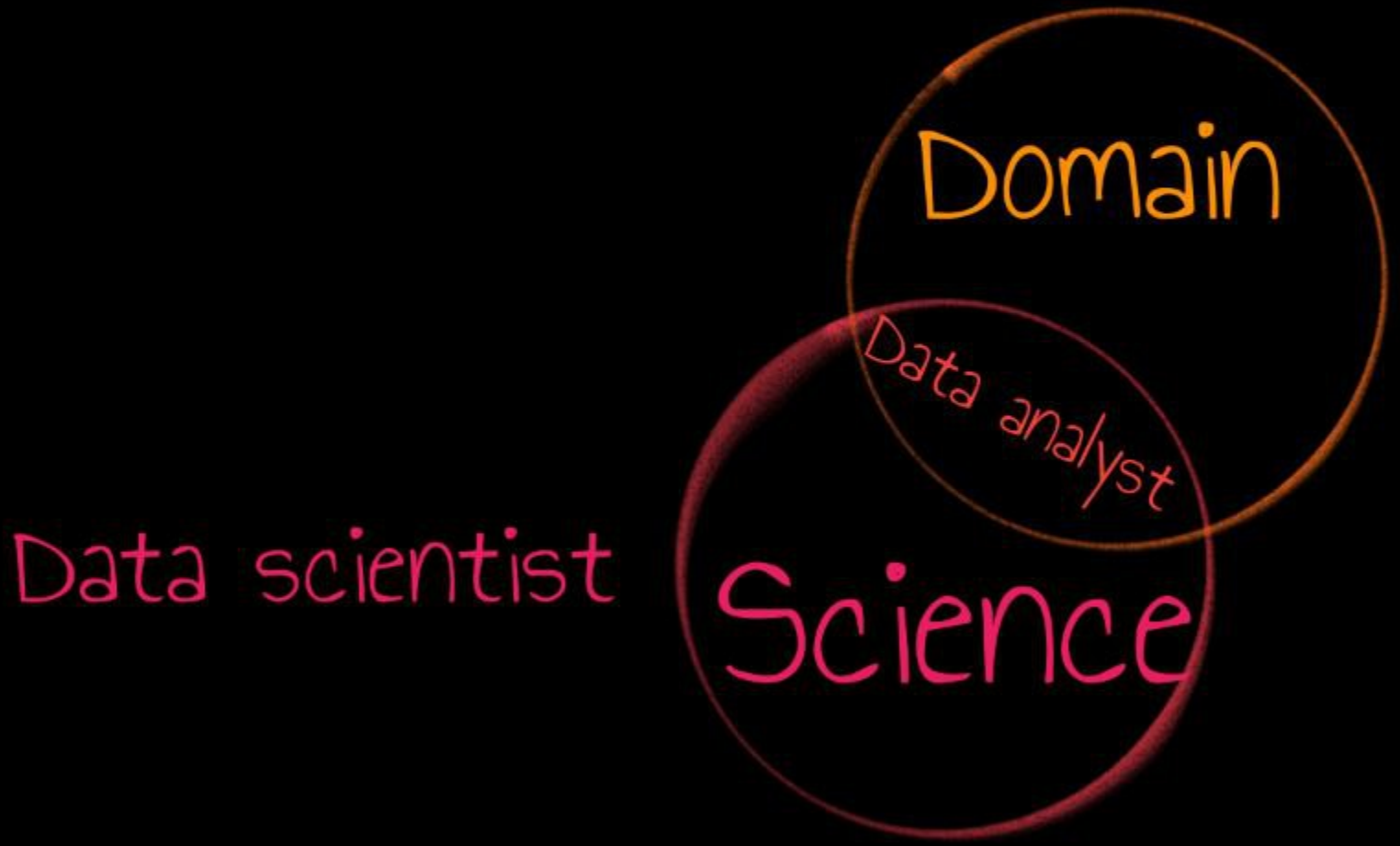
- Lots of simple events (clicks, scrolls, interactions)
- Backend is in Scala, but Data Science team works with Python

Approach

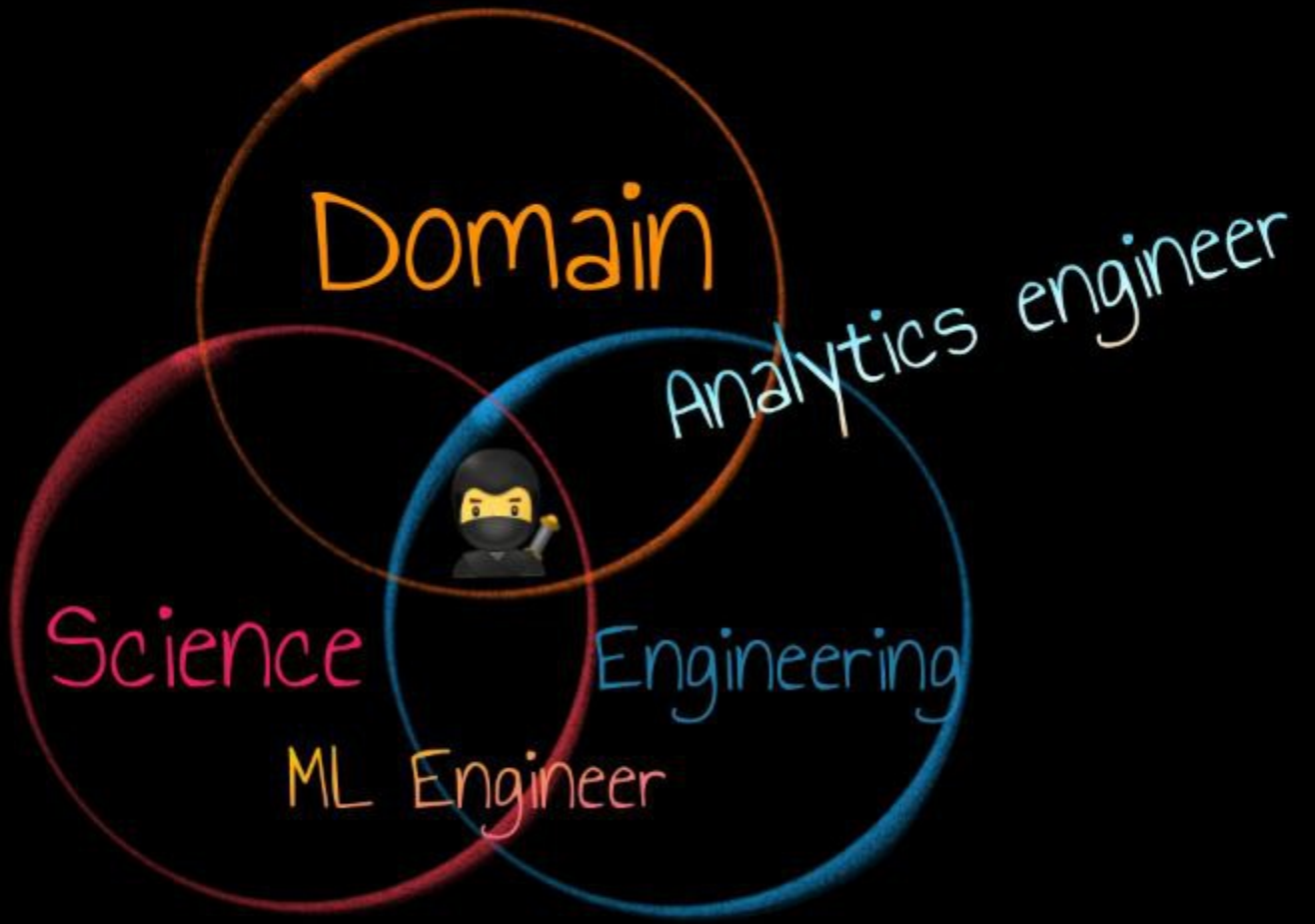
1. Define business questions
2. Define required data
3. Prepare data
4. Research models
5. Productionise model
6. Monitor

Data components

Data components



Data components



The three specializations

Data Engineers

The big data folk

Big data?

- Data that won't fit a single node
- Data that scales on 3V
  - Velocity
  - Variety
  - Volume
- Data on which we can make reliable business decisions

Data engineer responsibilities

Data:

- Events

Data engineer responsibilities

Data preparation:

- Collection and storage (managing costs)
- Verification (data engineering, upstream customers)

Data engineer responsibilities

Technical solution:

- Configuration
- Technical solution and choice

Data engineer responsibilities

In production:

- Monitoring (ops)
- Provide access to suitable data to BI and DS

Data Scientists

The research folk

Data scientist skills

- Ability to do research
- Mathematics
  - Statistics
  - Machine/deep learning
- Programming
- Data wrangling and visualisation

Data scientist responsibilities

1. Translate business question
2. Research available data and make shortlist
3. Request sample from data engineering

Data scientist responsibilities

4. Modelling:
    - Feature engineering
    - Iteratively find suitable MVP
5. Define DoD for “production ready” in collaboration with DE team
6. Define model metrics, request monitoring

Data Analysts

The data communication folk

Data analyst skills

- Deep understanding of business
- Knowledge of what data answers business questions
- Data communication and visualisation
- Data wrangling
- Data exploration and analysis

Data analyst responsibilities

1. Understand feasibility of project
2. Define metrics for measuring business success
3. Request sample from data engineering

Data analyst responsibilities

5. Request final pipelines from DE for reporting
6. Reporting on business success:
    - Business metric dashboards
    - Adhoc analyses (e.g., for marketing)

Overlaps: Productionisation

- What is the definition of done for the finished model?
- What “production ready” code means
- Who is responsible for each bit?
  - Responsible for algorithm: DS
  - Responsible for upstream data: DE
  - Responsible resourcing: ops

Overlaps: Diagnose/fix

- Monitoring (from DS, DE and ops) should pick up issue
- It’s a team effort to diagnose
- Major changes should be an agreement between all affected teams

Other roles

Summary

1. There are lots of responsibilities around data
2. These responsibilities are broad
3. You need to split your work between 3 dimensions
4. Hire team accordingly

TY!