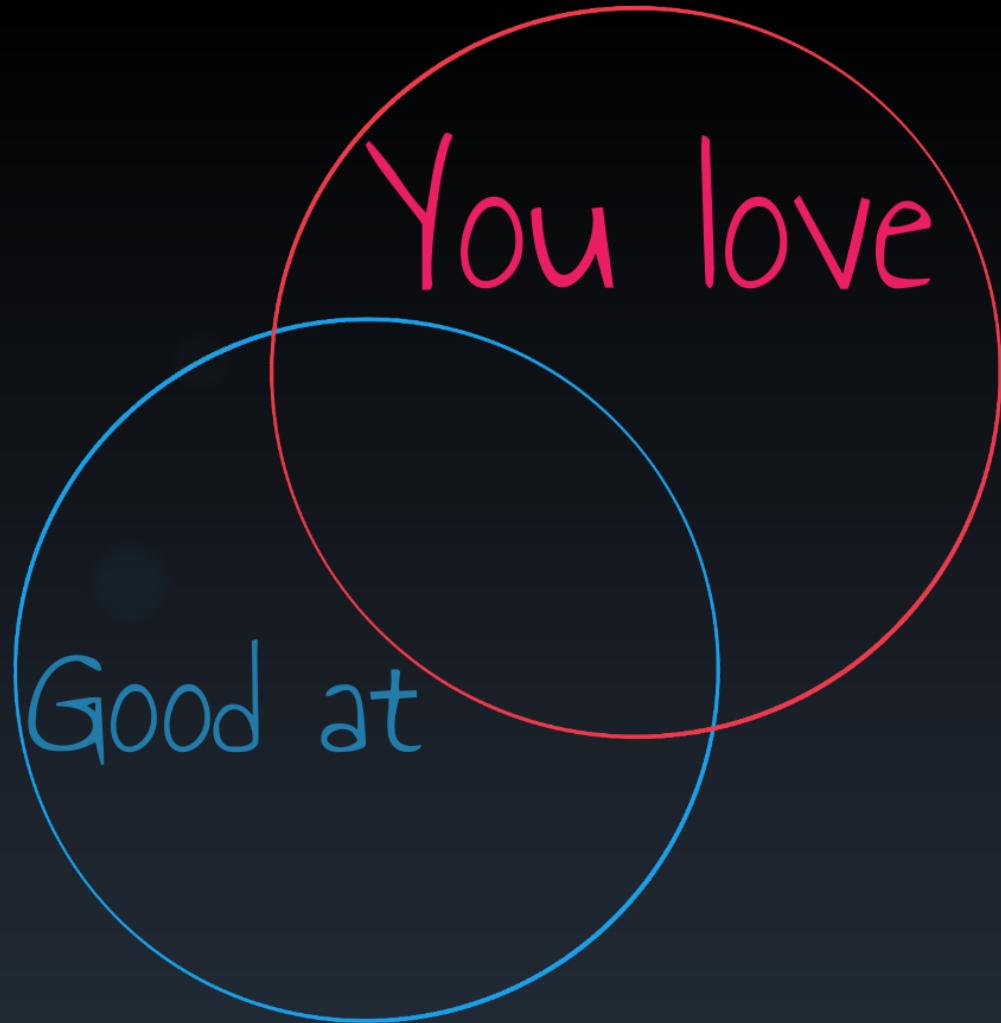
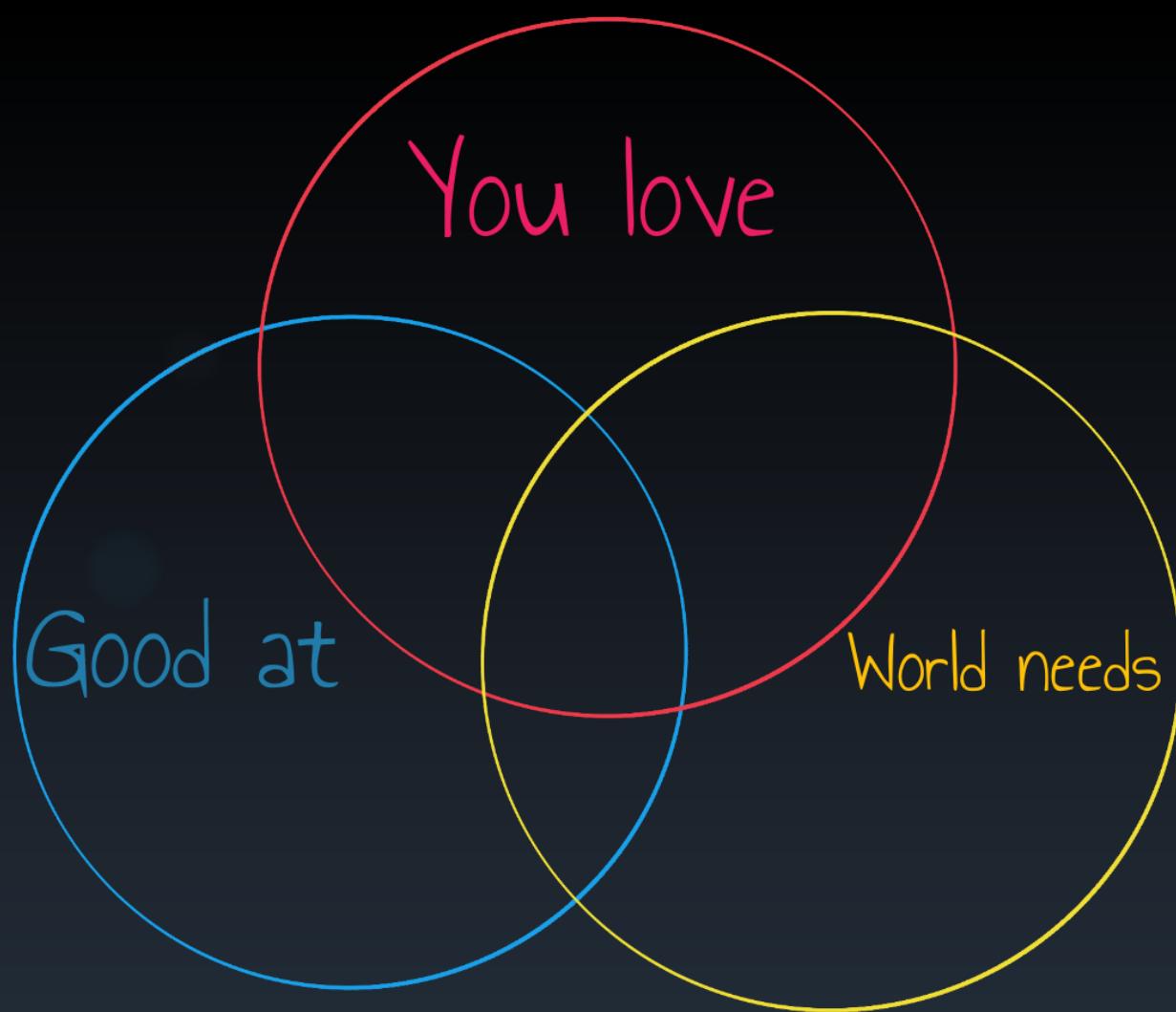


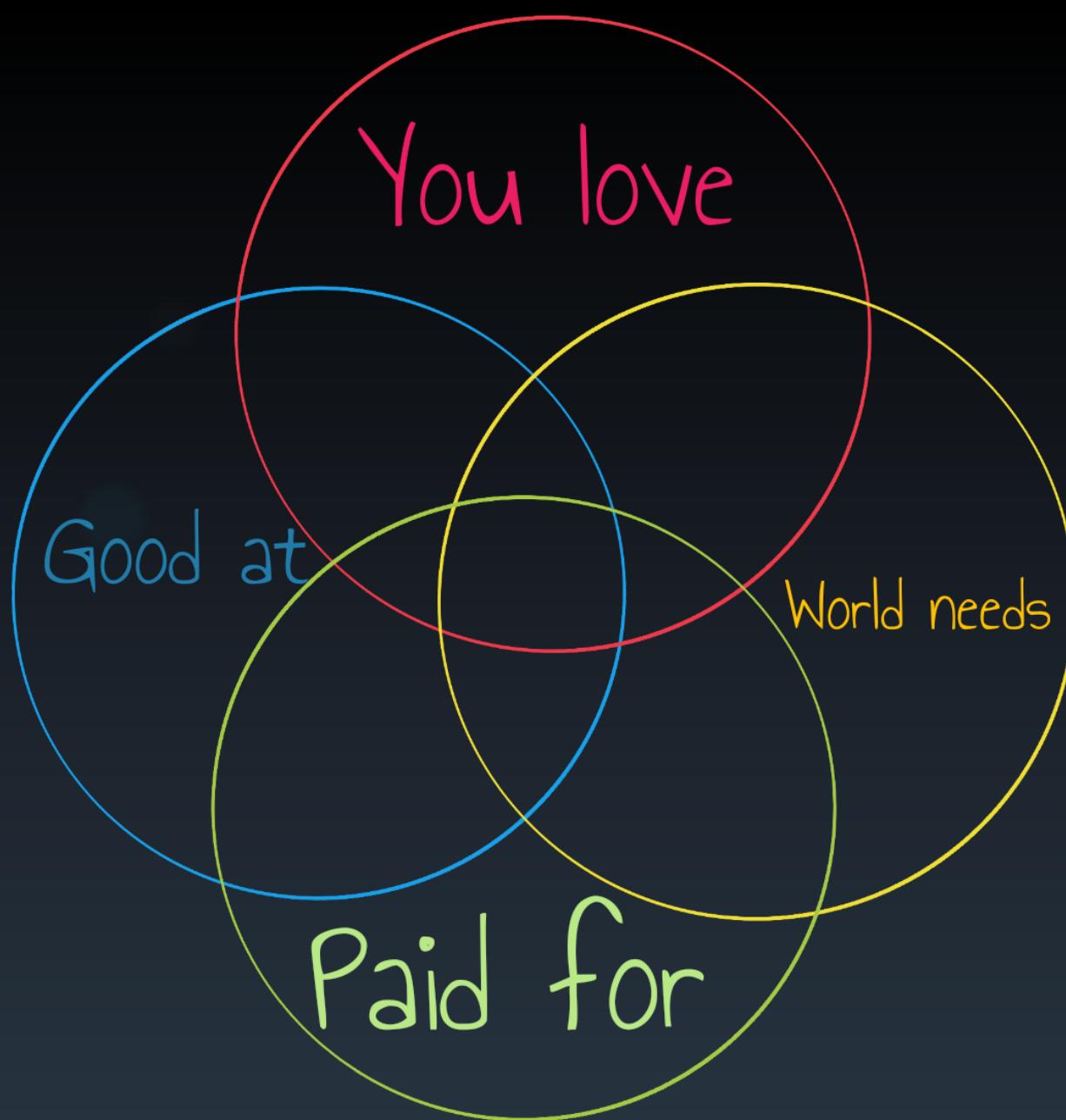
How do we know where to work?

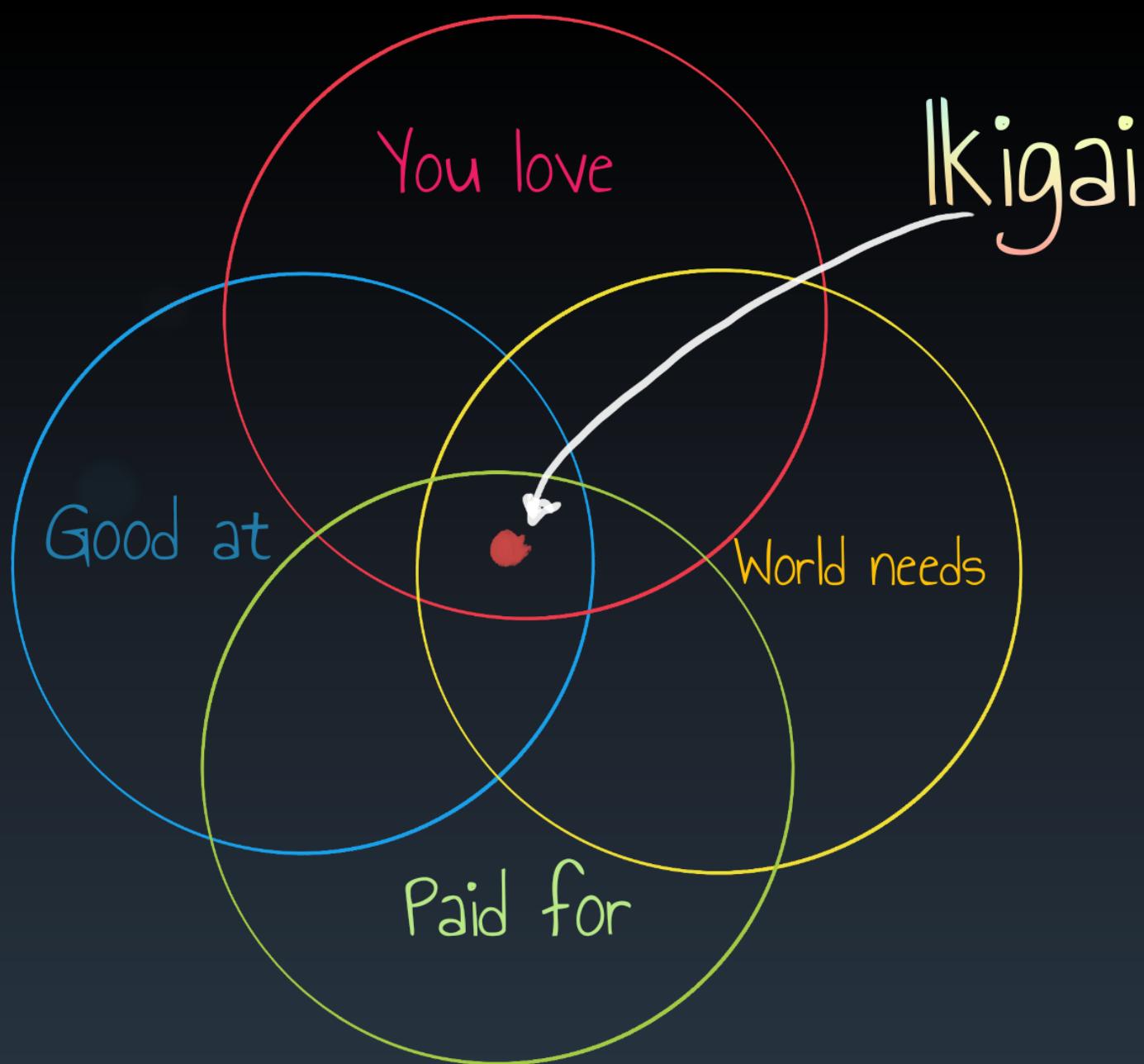


Good at









<

Ikigai

“ Ikigai (生き甲斐, lit. 'a reason for being') is a Japanese concept referring to something that gives a person a sense of purpose, a reason for living.

”

Wikipedia, the most reliable source of valid data

DE, DS, and others

Find your Ikigai

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?

Setup: Large fashion e-com (*Sapando*?)

Task: recommender system

What business says:

- Increase retention
- Reduce cost of acquisition

What business means?

Clients be like



What business means?

- Customers and even potential customers should return earlier
- It should be easier to convert non-customers to customers

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

People involved

draw diagram here:

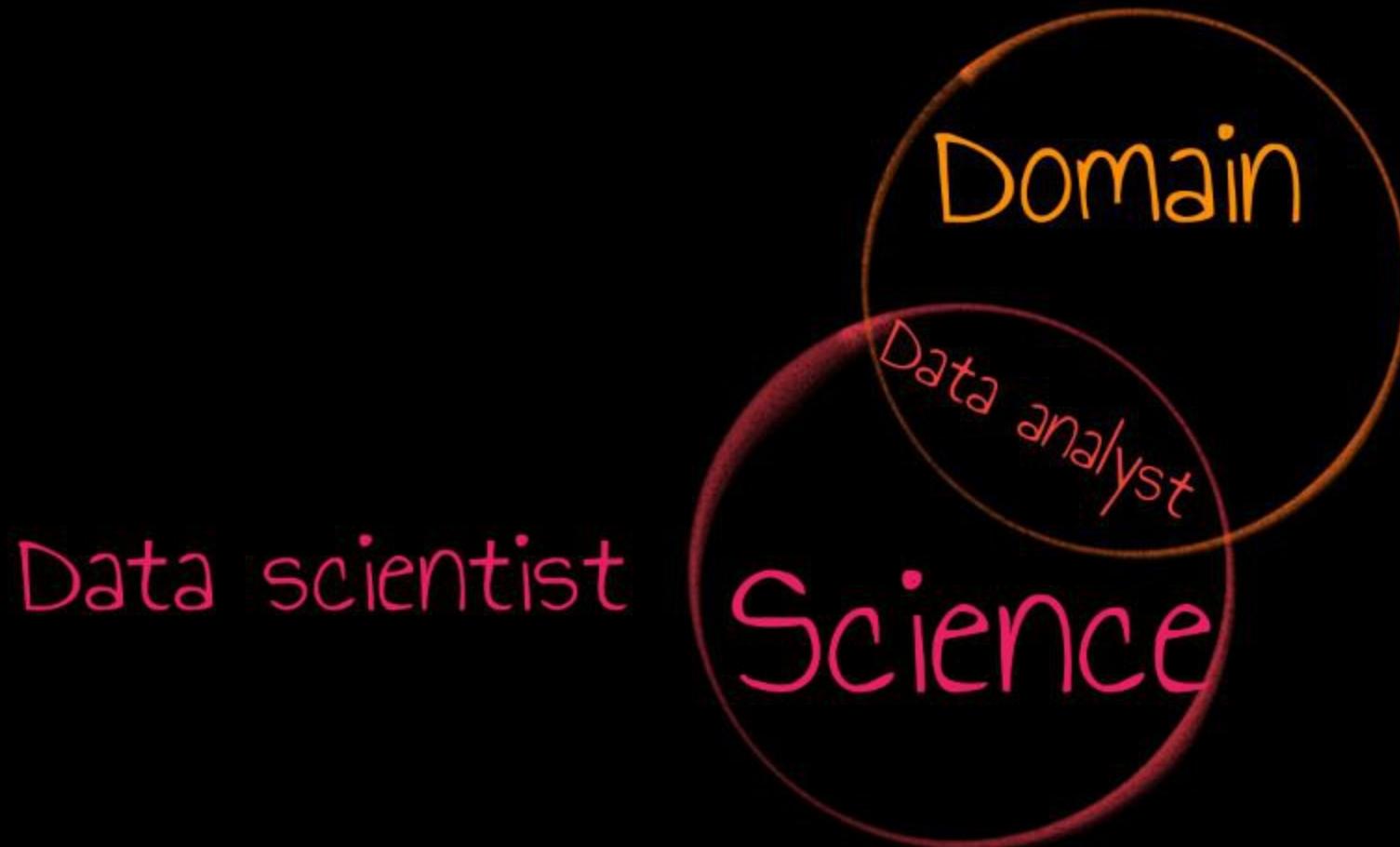
analysts+dss+des inside, other engineers, stakeholders outside

Data components

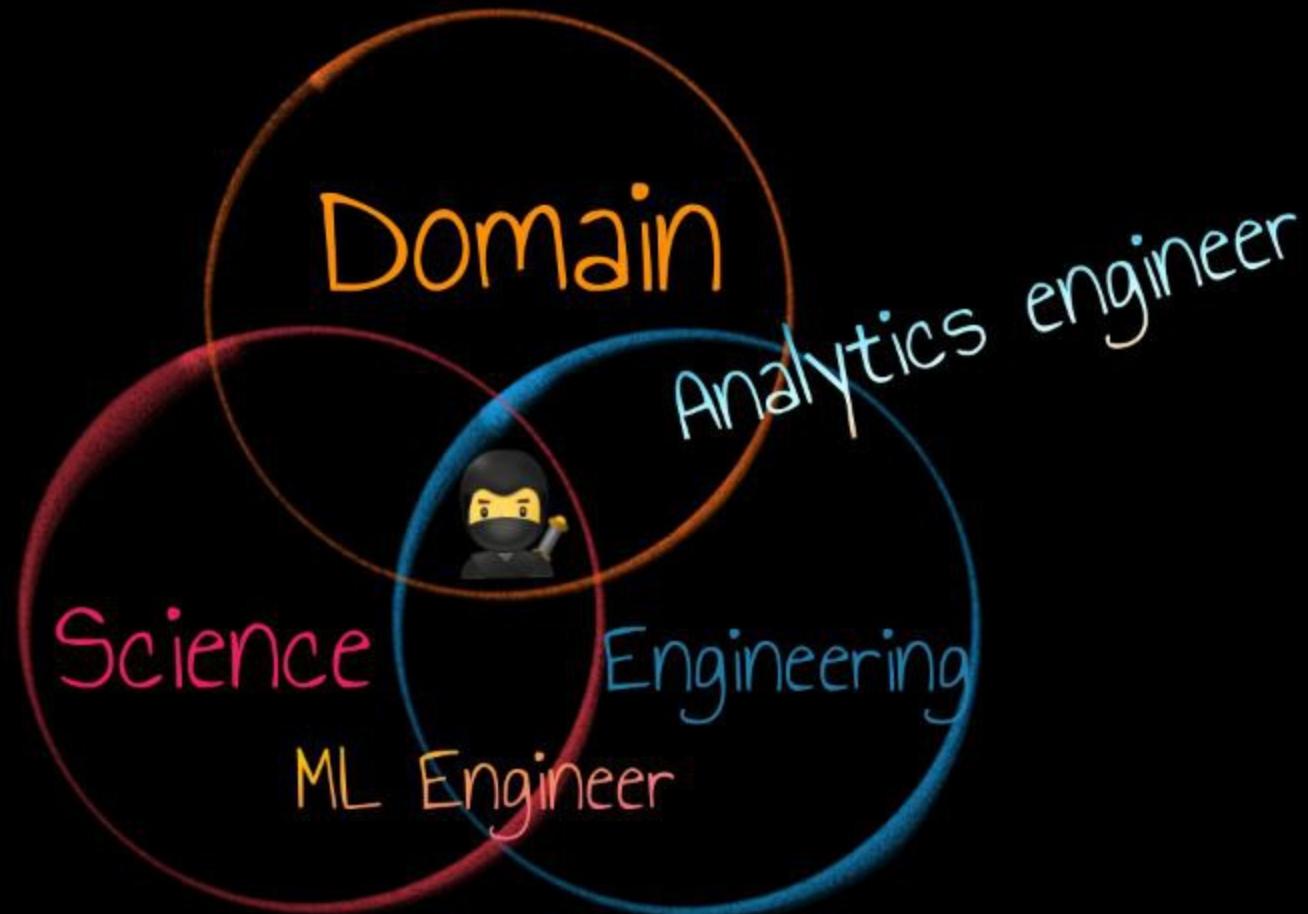


Analysts

Data components



Data components



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
- Model serving - custom solutions, ready built solutions like ML

Data engineer responsibilities

In production:

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

Data scientist responsibilities

- Find a suitable technical solution
- Discuss constraints with business and engineering
- Request suitable data from DE
- Build MVP, assess success, test
- Define DoD for “production ready” in collaboration with DE team
- Extracting features from images and item metadata
- Data science metrics monitoring

Data scientist responsibilities

Collaborate with data scientists to understand what business outcomes should be for the product

Downstream customer of data engineers, request required data for reporting (might involve asking DE to make dedicated ETLs because data will be required for long term reporting)

Create business metrics from raw clicks and scrolls - retention, acquisition and conversion behaviours

Build dashboards for reporting impact of business metrics on

Connect business metrics with financial impact of project - relationship between retention, acquisition and conversion and financial performance

Overlaps: Productionisation

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

Overlaps: Diagnose/fix

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

Other roles

TY!