

A Data Ninja Rockstar Is Not Enough

Jodie Burchell
Pasha Finkelshteyn

Who works with data?



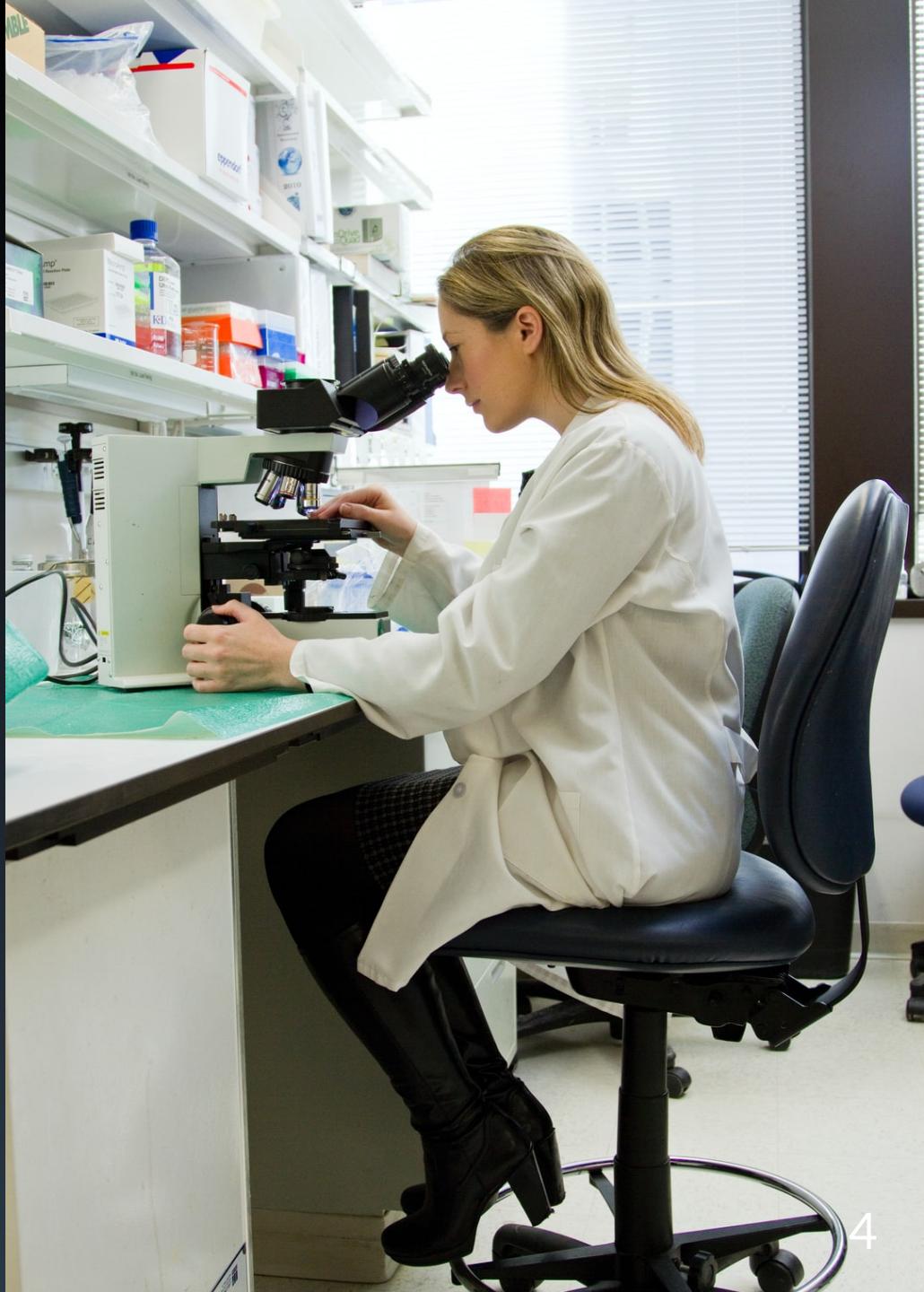
Who works with data?

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

Anthony Goldbloom ✅
@antgoldbloom

Just finished v1 of the new recommender system I'm building. Results so far are incredibly promising

CUSTOMERS WHO BOUGHT THIS ITEM



ALSO BOUGHT THIS



10:58 PM · Aug 11, 2022 · Twitter Web App

221 Retweets 22 Quote Tweets 2,189 Likes

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 should return sooner

Technical constraints

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



Approach

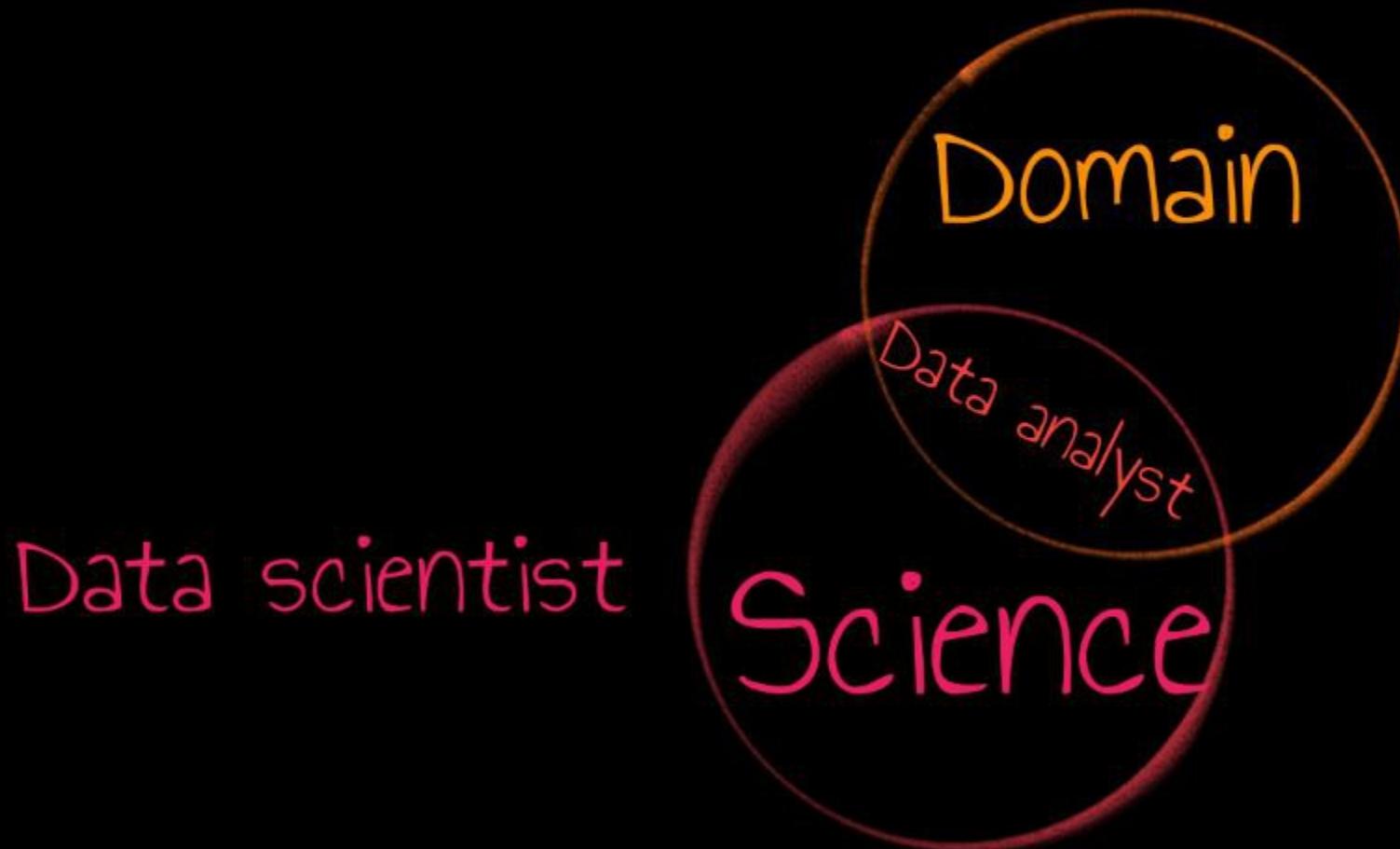
1. Define business questions
2. Define required data
3. Prepare data
4. Research models
5. Productionise model
6. Monitor

Data components

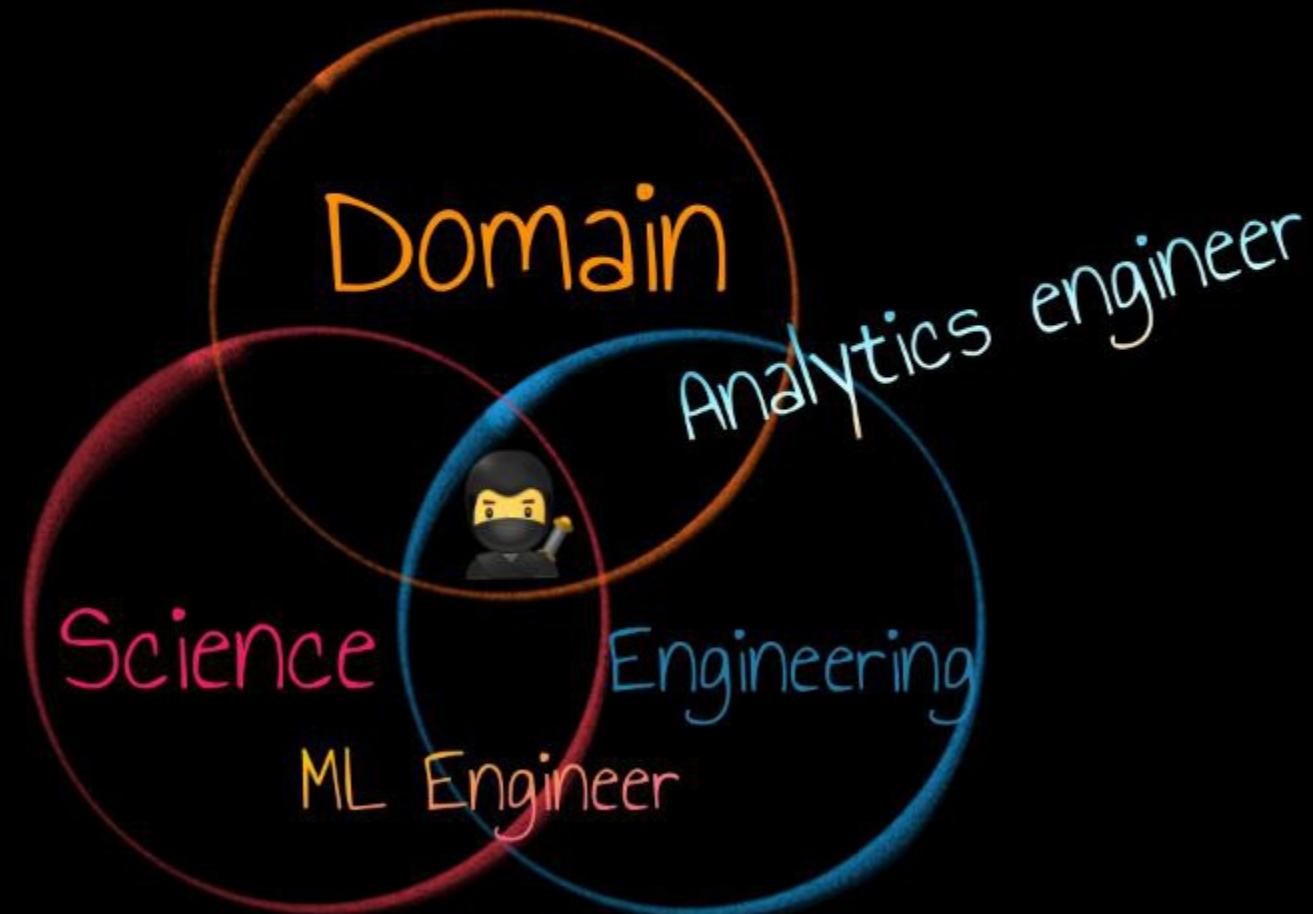


Analysts

Data components



Data components



The three specialisations

Data Scientists

The research folk

Data scientist skills

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

Data scientist responsibilities

- Translate business question
- Research available data and make shortlist
- Request sample from data engineering

Data scientist responsibilities

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

Data Engineers

The big data folk

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

BI Analysts

The data communication folk

BI analyst skills

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

BI analyst responsibilities

- Understand feasibility of project
- Define metrics for measuring business success
- Request sample from data engineering

BI analyst responsibilities

- Request final pipelines from DE for reporting
- Creating dashboards for business metrics
- Adhoc reporting

Overlaps: Productionisation

- What is the definition of done for the finished model?
- Who is responsible when things go wrong?

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



Thank you!

Twitter: @asm0di0

Blog: t-redactyl.io

Twitter: @t-redactyl