

Avoid Hiring Data Ninja Rockstars

How to build effective data teams

Dr. 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)

The reality is ...

Nobody should or even can do everything

Responsibilities should be distributed

Example project

Setup: Large fashion e-com
Trasos

Task: recommender system



What the business says

- Increase retention
- Reduce cost of acquisition

What does the business mean?

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

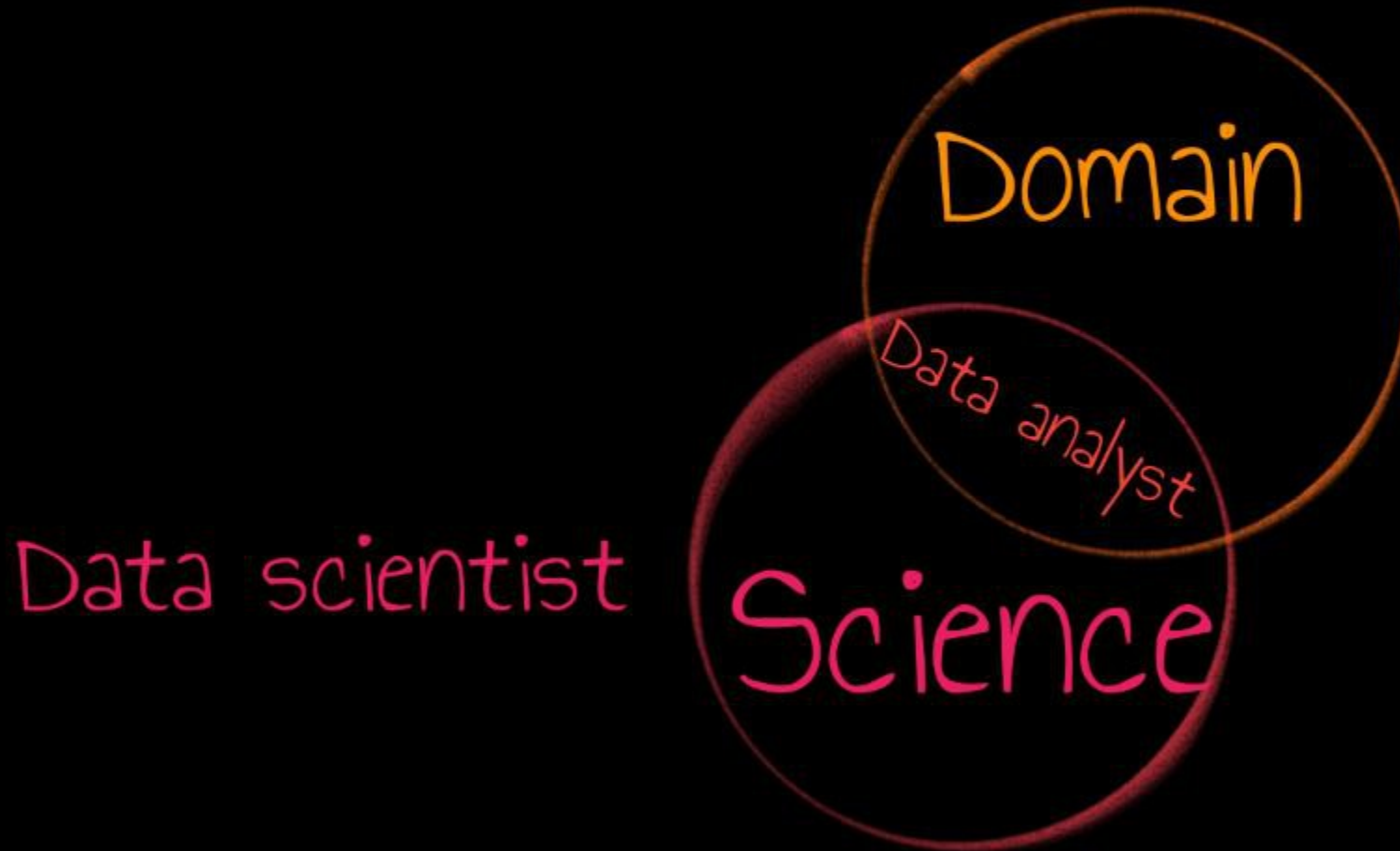
1. Define business questions
2. Define required data
3. Prepare data
4. Research models
5. Productionise model
6. Monitor

Data skills

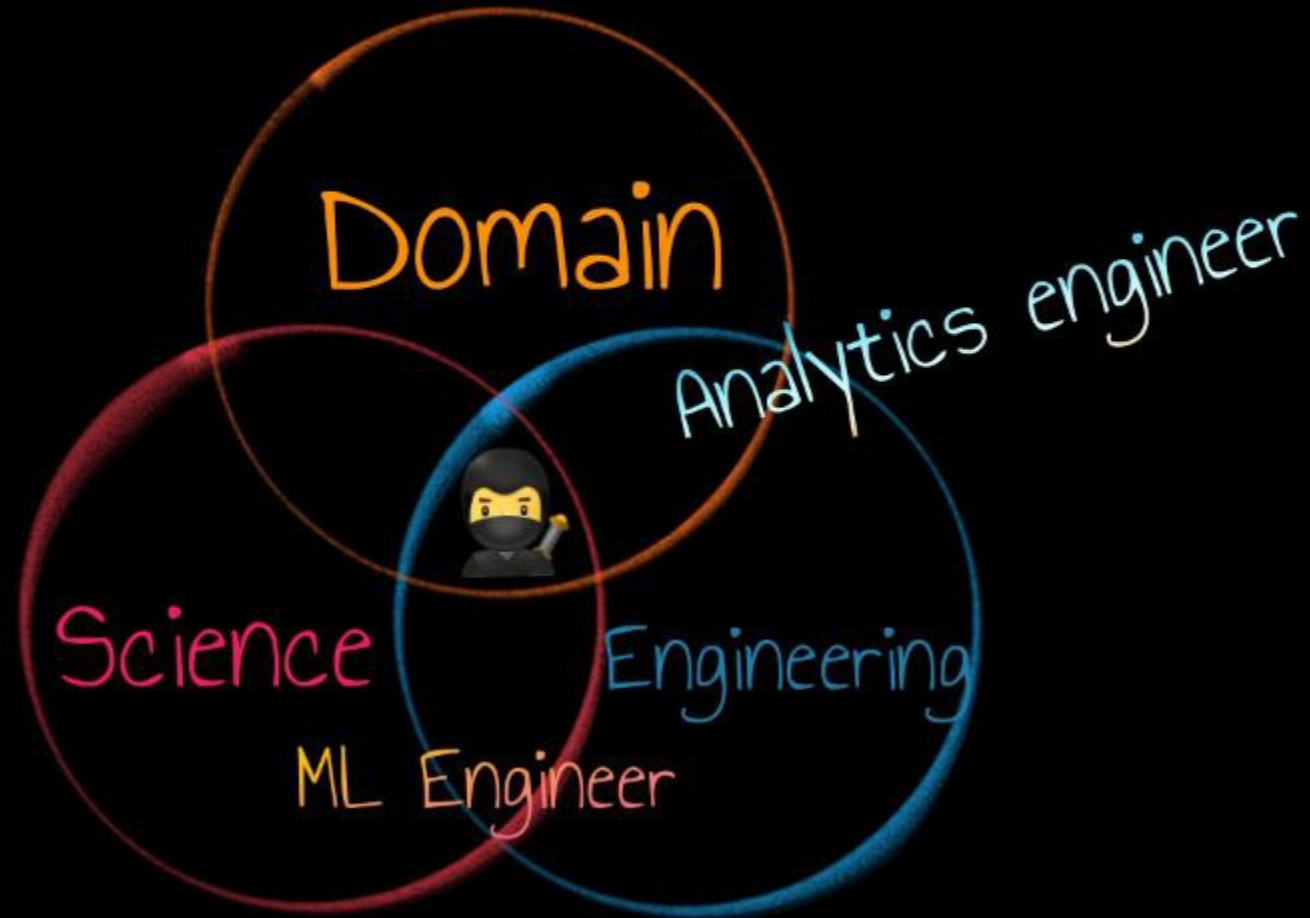


Analysts

Data skills



Data skills



The three roles

Data Scientists

The research folk

Data scientist skills

- Data + science!
- Automate manual processes
- Build data products
- Statistics
- Machine/deep learning
- Programming + data wrangling
- Data 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
- Handover of model for productionisation
- Define model metrics, request monitoring

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

Data Engineers

The big data folk

Data engineer skills

- Data lifecycle
- Strong engineering skills
- Monitoring data pipelines
- Manage data needs for multiple teams
- (Sometimes) responsible for ML models in production

Data engineer responsibilities

- Understand whether the data will support the project
- Tracking down data
- Extracting samples

Data engineer responsibilities

- Build data pipelines for project
- (Potentially) implement model
- Monitor data

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