This could probably use a reorg / update

Contents

1. Today's Goals

- Understand DS orgs, how they work
- Know the different trade-offs between different organization styles, how they work and the effect that they have on the day-to-day job
- Try to plan your own career, where you want to go and why
  Understand your role
  Understand that role's effect on your career

2. Models of Data Science within an Organization

- What is an organization?
- It's a structure (people and tasks) within a company which has the following properties:
  (1) A set of roles (formal and informal)
  (2) A set of tasks to be completed
  (3) A hierarchy of authority.
- We are interested in understanding these orgs at different companies.
- Organizations can change depending on a number of different factors and the things that we talk today are very much in-flux.

- Major factors I look at when evaluating data science orgs:
  1. Value of data science to the company (what is it used for? All of the "strategy stuff". What is the value of DS to the firm).
  2. Age of company
  3. Size of DS team
  4. Personality of leadership
  5. Background of leadership
  6. Reporting Structure
- Lets go through each of these and the impact it has on the data science org.

2.1. Value of DS to the business.

- We have spoken about this a lot, no need to hem and haw about it.
- Moral: The closer DS is to the company's "Moat" the more power the DS org has, the more stable it will be, etc.
- Companies where DS is of secondary importance to their core treat DS as a second class citizen.

2.2. Age of company.

- The younger the company, the more "ad-hoc" / informal and disorganized the DS team will be. The DS team will also be much more reflective of the people who work on the team, rather than more "process" based.
- Younger companies tend to also have greater opportunities. It's more of a blue ocean of opportunity.

2.3. Size of DS team.

- The smaller the team the more "informal" / "ad-hoc" the organization will be.
- Small teams tend to strain against the cost of process.

2.4. Personality of Leadership.

- The more informal the organization the more the personality of leadership matters.
- If the head of the DS org has a strong personality (positive or negative) that will also reflect in the organization.
- Strengths and weaknesses of leadership of the head of an org also tend to be reflected within the organization itself (examples: good at hiring?)

2.5. Background of Leadership.

- There is no common path to being a DS leader
- DS leaders will often rely on their previous experience when defining their DS team, including how they hire, how they reward / promote and how they define success.

- Coming from an engineering vs. consulting background. How is the team run (sprints? Jira?)? How is hiring done (how much emphasis placed on technical skill?)?
- DJ Patil: Former Chief Data Officer of the United States under Obama and 1st head of Data Science / Chief Data Scientist at Linked In. When he left LI they tried to find someone to replace him, but couldn't find someone with his skills, so they remade the entire data science org.

2.6. **Reporting Structure.**

- Data Science tends to report to a different vertical within an organization (CTO? CMO? COO?).
- Depending on that vertical there are different expectations around the day to day of the team.
- Where DS is in the org has an outsized impact on what you do.

What do you do?	Org
Building Product	CTO / Engineering
Forecasting Revenue	CFO / COO/ OPS / Finance
Optimizing Advertising	CMO / CTO / Marketing
Predicting Employees Quitting	HR / Head of HR / CPO

- All of the above tasks are DS tasks, but reporting to different orgs has an effect on the tasks.
- They also have an outsized impact on how you are evaluated.
  - Bad Engineer reporting to the CTO? Penalized
  - Great communicator reporting to the CMO? Rewarded

3. Disorg / Informal

- What do we mean by "Informal" and "disorganized"?
  1. Roles tend to lack definition and overlap with other roles
  2. Tasks tend to more ad-hoc and ill-defined.
  3. Breadth of tasks tends to be larger

     On my DS team it's not uncommon to have to help with budgeting
  4. Hierarchy tends to be more fluid

     Draw traditional org-chart vs. random graph.
- Is this bad? *Not necessarily*
  - Reorganize around problems more efficiently
  - Lower costs (organizationally) "cost of process"
  - Higher exposure (which is especially helpful when starting your career)

  - If leadership has a good personality, then that tends to positively impact the entire org and imprint it in a positive way. People will pitch in, work hard, etc.
  - Downside:
    - Inconsistent performance appraisal
    - Opaque measures of success

Disorg/Informal		Attribute		Org / Formal
		Larger	→	
	←	Personality Drive		
		Better Role Def	→	
	←	Opportunity		
		Better Consistent Communications	→	
	←	Task Variety		
	←	Learning		
	←	Turnover		

4. Centralized vs. Decentralized

- Most data science teams exist on a spectrum of decentralization:

Centralized

- All DS in a central Org
- Single Reporting Structure
- Usually a single Jira Board / Job queue
- Positives:
  - Report to a person who is in DS
  - Work with other DS people
  - Consistent management assessment
- Negatives:
  - Less effective b/c silo'd from the team
  - Fire first
  - "Fun" data tasks often given outside of the data org (to people who are on the ground)

Decentralized/Embedded

- Each team has their own DS team (hire own / Rental)
- Working with a team on a functional or project basis
- Positives:
  - Highly effective, work directly with a team
  - Continuity of project (rather than job queue)
  - Exposure to other teams and skillsets
- Negatives:
  - Manager tends to not be DS
  - Very easy to be pigeonholed as an SQL or Excel Person
  - Less ability to learn DS skills

- Most companies operate in a hybrid mode:
  (1) At TinyCo there were two data science teams:
    (a) Central Data Science Team (LTV, CAC, Cross Promotion, User Acquisition, Marketing and Biz Development)
    (b) Game Data Science Teams: DA/DS on a single game working with the product managers/engineers.
  (2) AT FB/Instgram:
    - "Pods" for functional tasks (UI/Eng/BE/PM/DS)
    - Two reporting structures: your functional team (e.g. the PM in charge of the pod) and a shadow manager who is a DS manager.
  (3) Riot Games (League of Legends)
    - Central Team and Embedded teams.
    - DS rotate on 3-6 month assignments on an embedded team and then are rotated either to central or onto another embedded team.

5. Model #1: DA/DS/DE

- "Classic Model" of data organization in a firm
- 3 Readings: "What is a Data Scientist", "DA, DS and DE" and "DA vs. DE"
  (1) **Data Analyst:** Provide insights, job is EDA, data prep. Story-telling.
  Tools used: SQL, Notebooks, R, Excel, Python, Looker, tableau, etc.
  (2) **Data Scientist:** "The Chamption" (← blah), building models
  Tools used: SQL, Notebooks, Python, Stats, ML, etc.
  (3) **Data Engineer:** Infrastructure, database, deployment, automation and ETL Processes
  Tools: Python, SQL other "real" languages like Java / Scala / Spark, command line, docker
- DA vs. DS: Use Data to talk to people vs. Using data to talk to computers.
- We have all seen the Venn diagram of "Stats, Programming and Domain Knowledge" with the center being "DS".
- DRAW VENN DIAGRAM.
- This isn't a useful diagram. Draw one with "Knowing Recipes, How to Boil water and using a knife" calling the middle a "chef".
- You can do this with anything, but it doesn't help explain what a chef does.
- Describing something this way puts too little focus on the "outcome" of the job.
  For a chef we would want a job description as "Make Sushi"
  For a lawyer we would want a job description as "Write Contracts"
- Pay structure: DE > DS > DA (I'd guess)

6. Nick's Take On the Model

- IMO: The job title "Data Scientist" is a historical relic that does a lot to misvalue the power that working with data can have. A lot of companies have the need for a specific of skills that fall somewhere under the umbrella of data science. "Data Science" became a placeholder because we didn't have a better phrase.
- Goldilocks Problem
- Because the job title doesn't describe what a person does specifically we end up in this situation where:
  - "Easy" tasks become data analyst tasks (we don't want to be SQL Monkeys)
  - "Hard" tasks become ML Engineers / Data Engineers (it's too hard for us)
- The job of "Data Scientist" is the job of "just right" – not too hard and not too easy.
- This is very bad for career prospects because you have essentially boxed yourself in.
- Pay structure: Research > Scientist

7. DS @Lyft/FB

- Story: Faced with previous definition (the DA/DE/DS model), hiring DA was essentially impossible.
- There was an implicit ranking (DE > DS > DA) and thus no one wanted to be at the bottom.
- Basically every DA would leave as soon as they got a DS offer.
- Solution? Call everyone a Data Scientist! Specifically also given them some other functional responsibility (Metrics, Insight, etc.)
- Previous Data Scientists would be called a "Research Scientist" or "Applied Researcher" or "Applied Scientist".
- This is pretty common now. Lots of organizations have just reclassified their data analysts as data scientists. It doesn't change the underlying problem (poor definition of job, nebulous requirements, etc.) but at least it makes people happy.
- This model is (IMO) part of the reason that we see so many ML Engineer titles flourishing. If everyone is a Data Scientist then traditional data scientists are putting themselves into the role of Data Engineer (usually not great coders), so the actual good coders needed a new title.
- Who puts models into production?

8. DS @ Twitter: A vs. B

- Basic Idea is that "A" is for analyst and "B" is for building.

- "A" is for static analysis – closer to a statistician
- "B" is for stronger coders, probably building models.
- I'd guess that "B" is paid a bit more.
- Who puts models into production?
- I don't think that this model is used at Twitter anymore, but I do see similar designations in other companies, usually on the smaller side.

9. DS @ Airbnb

- Airbnb was experiencing explosive growth and were having difficulties managing the proliferation of job titles and functions related to data.
- A lot of ad-hoc job definitions, etc.
- Having trouble doing consistent evaluation of candidates, being fair with promotions and management. Job expectations all over the place

Year	DS Job Applications
2015	5K
2016	9K
2017	20K
2018	35K

- There solution was to break up the Data Science Job into three distinct functional roles:
  1. **DS, Analytics** Traditional Data Analyst
  2. **DS, Inference** More of a Data Scientist. For Airbnb this means statistician, design of experiments.
  3. **DS, Algos** Similar to a DE, but less coding. More model creation and evaluation.
- Who do you think gets paid the most?
- Who puts models into production?

10. Overall thoughts

- No one has a good model for data science in organization
- For all companies DS is relatively new – we have good models for other firm verticals (HR, engineering, etc.), but not as many for data science.
- DS tends to be more personality driven than other verticals with more established cultural norms.
- Organization structure tends to have lots of different dimensions.

- As you interview and think about what you want to do with your career, be aware of "the way things are here" isn't necessarily the way things are everywhere.