DSC 90

Winter 2020

Choosing a Company

Company Specific Research

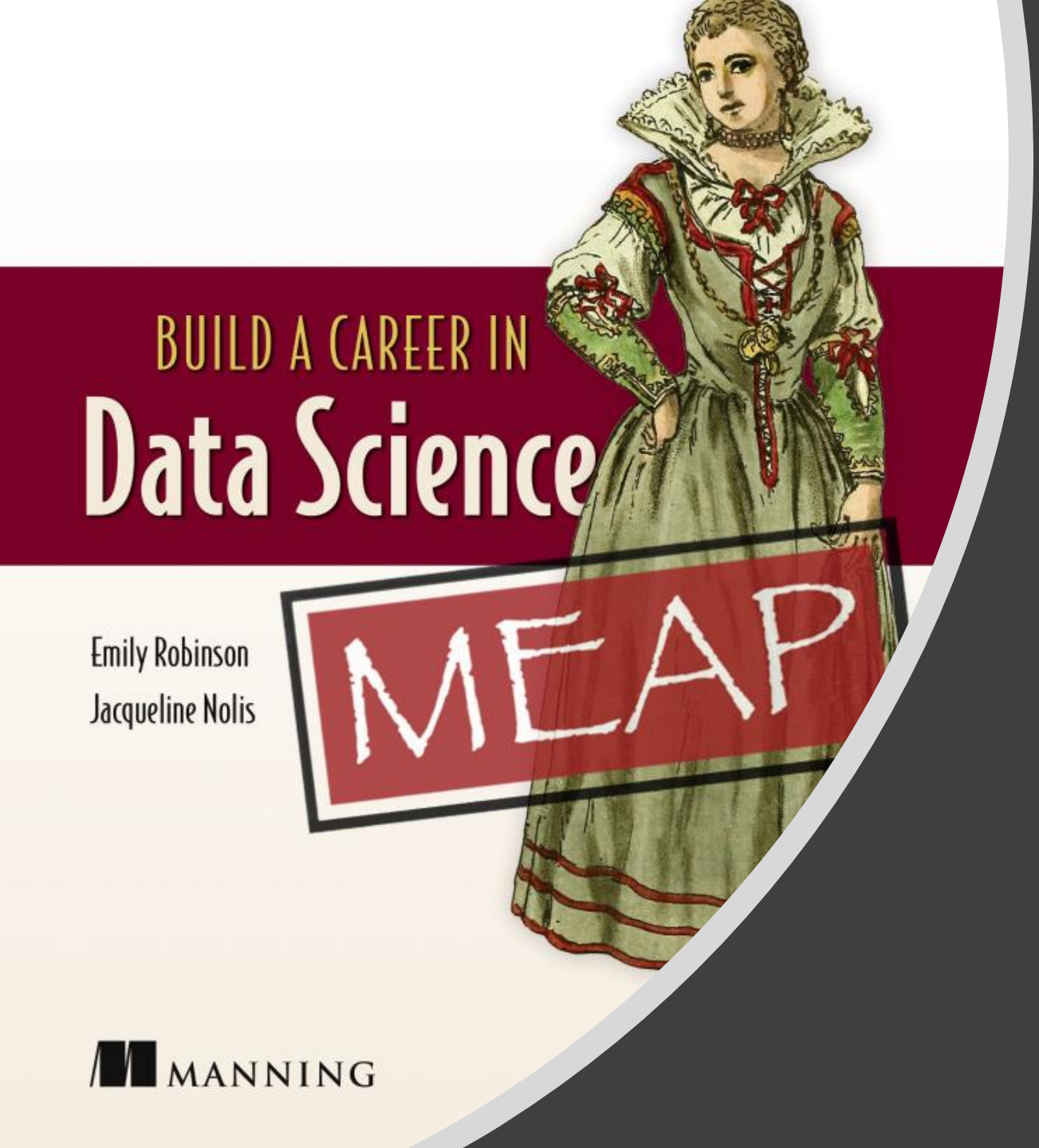
Learn about the company. How do they make money?

Understand what the data science team in each organization is working on.

Research how the company interviews

You'll be asked a lot of situational and product questions that have to do with current work the company is undertaking

Company Specific Research



Company Architypes

Massive Tech Company

Similar to: Google, Facebook, Microsoft
Company history: 20 years
Employees: 80,000

- MTC is a tech company with a massive footprint, selling cloud services, consumer productivity software like a text editor, server hardware, and countless one-off business solutions.
- The company has amassed a large fortune and uses it to fund unusual R&D projects like self-driving scooters VR technology.
- While their R&D makes the news, most of the technical workforce are engineers making incremental improvements to their existing products, adding more features, improving the user interface, and launching new versions.

Your Team

- MTC has nearly a thousand data scientists spread across the company. These data scientists are largely grouped into teams each supporting a different product or division, or individually placed within a non-data science team to fully support it.
- Large organizations hire new people every day, so the company should have standard processes for getting you a laptop, access to data, and training you on how to use any special tools.
- On the team, you'd be tasked with doing data science for your particular area of focus. That could include creating reports and charts that executives could use to justify funding the project. It could also be building machine learning models that would be handed off to software developers to put into production.
- The size of the team is a blessing and a curse: you have a large body of expert data scientists to discuss ideas with, but most of them probably don't have familiarity with the particular tasks you are working on.
- The work your team does is likely a healthy balance of keeping the company running, such as making monthly reports and providing quarterly machine learning model updates, and doing new projects, like creating a forecast that has never been done before.

The Tech

- MTC is a massive organization, and with organizations of this size it's impossible to avoid using different types of technology throughout the company.
- Worse, not only is the technology to store data disjointed, the data itself may be as well.
- Most MTC sized companies have their own homemade technology stacks.
- As a data scientist, you will likely have a number of possible tools you can use.
- The machine learning stack varies dramatically by what part of the company you are in. Some teams use microservices and containers to efficiently deploy models, while others have antiquated production systems.

Pros and Cons

- Being a data scientist at MTC means having an impressive job at an impressive company.
- The high number of data scientists in the company means you have a large support network you can rely on if you are struggling, and smooth processes for joining the company and getting access to required resources.
- The tech stack is complex and difficult to navigate because so many people have built it up in so many ways.
- It'll be harder to stand out and be noticed, because there are so many other data scientists around you.
- With MTC being an established company, working there gives you more job security.
- Something that is both a pro and con of MTC is that there are people in many specialized roles within the company.
- Another con of MTC is the bureaucracy.
- MTC is a great company to work at for data scientists who are looking to help solve big problems using cutting-edge techniques—both for decision scientists wanting to do analyses and machine learning engineers wanting to build and deploy models.
- MTC is a poor choice for a data scientist who wants to be the decision maker and call the shots—the large company has established methods, protocols, and structures that you have to fall into.

handbagLOVE

Handbag Love

Similar to: Payless, Bed Bath & Beyond, Best Buy
Company history: 45 years
Size: 15,000 employees (10,000 in retail stores, 5,000 in corporate)

- The company has been generally slow to adopt new technology, taking plenty of time before getting its first website and its first app.
- HandbagLOVE has had financial analysts employed for many years calculating high-level aggregate statistics on their orders and customers; however, only recently have they considered hiring data scientists to better understand customer behavior.
- The newly formed data science team was built starting with a base of financial analysts who had previously been making Excel reports on performance metrics for the company. As HandbagLOVE supplemented these people with trained data scientists, the team started to provide more sophisticated products.
- HandbagLOVE is far from deploying machine learning models into continuously running production. Any product recommendations on their website and app are powered by 3rd party machine learning products, rather than having been built within the company.

handbagLOVE

Your Team

- The team leans heavily towards data scientists who can do reporting rather than being trained in machine learning.
- HandbagLOVE has laid out general paths to progress into senior roles. Unfortunately, none of them are specific to data science—they are just high-level goals around being independent or things taken from the software development career management tools which don't quite apply. To progress in your career, you mostly have to convince your manager that you're ready, and hopefully your manager can get approval to promote you. On the plus side, if the team ends up growing you'll quickly become a senior person on the team.
- Since the data science team provides reports and models for departments throughout the company, like marketing, supply chain, and customer care, the members of the data science team are well known. This has given the team a great deal of respect within the company, and in turn the data science team has a lot of comradery within it. The combination of the size of the team and the level of influence within the company allows the data scientist to have far more influence than they would in other companies. It's not unusual for someone on the data science team to meet with top level executives and contribute to the conversation.

handbagLOVE

The Tech

- The common phrase you hear when talking about technology at HandbagLOVE is “well that’s how it’s always been.”
- The system has been pushed well past its limits and has had many modifications bolted on. All that being said, it still works. Other data is collected and stored in the central database as well: data collected from the website, data from the customer care calls, and data from promotions and marketing emails.
- All of these servers live on-prem (not on the cloud), and an IT team keeps them maintained.
- By having all of the data stored in one large server, you have the freedom to connect and join the data however you want. And while sometimes your queries may take forever or overload the system, usually you can find a workaround to get you something usable.
- The vast majority of analyses are done on your laptop. If you need a more powerful computer to train a model it’s a large hassle to get it.

handbagLOVE

Pros and Cons

- By being at HandbagLOVE, you have a lot of influence and ability to do what you think is wise. This is due to a combination of the size of the company and the newness of data science. This freedom is very rewarding; you are incredibly empowered to do what you think is best.

- The downside of this power is that you don't have many people to call on for help. You're responsible for finding a way to make things work or dealing with the fallout when things don't work.

- The tech stack is antiquated, and you'll have to spend a lot of time making workarounds for that—which is a hassle. You may want to use a newer technology for storing data or running models, but you won't have the technical support to do it. If you're not able to entirely set up any new technology yourself you'll just have to get by without using it.

- A data scientist's salary won't be as high as it would be at bigger companies, especially tech ones.

Seg-Metra

Similar to: a thousand failed startups you haven't heard of
Company history: 3 years
Size: 50 employees

- Seg-Metra is a young company that sells a product that helps client companies optimize their website by customizing for unique "segments" of customers. Seg-Metra sells its product to businesses, not consumers. Early on in its brief history, Seg-Metra got a few big-name clients to start using the tool, which helped the company get more funding from venture capitalists. Now with millions of dollars at hand, the company is looking to quickly scale in size and improve the product.

- The biggest improvement the founders have been pitching to investors is to start adding basic machine learning methods to the product—it was pitched to investors as "cutting-edge AI." With this new funding in hand, the founders are looking for machine learning engineers to build what they have pitched. There is also a need for decision scientists to start reporting on the usage of the tool, allowing the company to better understand what improvements to make in the product.

Your Team

- Depending on when a data scientist gets hired, they may very well be the first data scientist in the company. If not the first, they will be in the first few data science hires and likely report to the one who was first hired.
- Due to the newness of the team there will be few to no protocols—no established programming languages, best practices, ways of storing code, or formal meetings.
- Regardless of exactly how Seg-Metra's growth shakes out, the data scientists at this early stage company are in for a bumpy and wild ride.
- The work of the team can be fascinating or frustrating depending on the day.
    - These first-time analyses and engineering tasks are fascinating because they are uncharted territory within the company, and the data scientists get to be the pioneers.
    - On other days, the work can be grueling—like if a demo has to be ready for an investor and the model is still not converging the day before. A
- While the work is chaotic, all of these tasks mean that the data scientists learn lots of skills very quickly while working at Seg-Metra.

The Tech: cutting edge technology that is taped-together

- By being a young company, Seg-Metra is not constrained by having to maintain old legacy technology. Seg-Metra also wants to impress its investors, which is a lot easier to do when your technology stack is impressive. Thus, Seg-Metra is powered by the most recent and greatest methods of software development, data storage and collection, and analysis and reporting.
- Data is stored in an assortment of modern cloud technologies, and nothing is done on-premise. The data scientists connect directly to these databases and build machine learning neural network models on large AWS virtual machines instances with GPU processing. These models are deployed using modern software engineering methods.
- At first glance, the tech stack is certainly impressive. But the company is so young and growing so fast that there are constantly issues with the different technologies working together. When the data scientists suddenly notice missing data in the cloud storage, they have to wait for the overworked data engineer to fix it (and that's if they're lucky enough to have a data engineer). It would be great if Seg-Metra had a dedicated DevOps team to help keep everything running, but so far the budget has been spent elsewhere. Further, the technology has all been installed so quickly that even though the company is young it would be difficult to monitor it all.

Pros and Cons

- The growth of the company is providing all sorts of interesting data science work and an environment where you are forced to learn quickly.
- These sorts of positions can teach skills that jump start a career in data science: skills like working under deadlines with limited constraints, communicating effectively with non-data scientists, and knowing when to pursue a project or decide it's not worth it. Especially early in a career, developing these skills can make you much more attractive as an employee than people who have only worked at larger companies.
- Another pro of working at Seg-Metra is that you get to work with the latest technologies.
- While the pay is not as competitive as larger companies, especially tech companies, the job does provide stock options which have the potential to be enormously valuable.
- One con of working at Seg-Metra is that you have to work very hard. Having 50- to 60-hour work weeks is not uncommon, and the company expects everyone to contribute everything they can.
- The company is volatile, relying on finding new clients and help from investors to stay afloat, giving Seg-Metra the con of low job security. It's possible that in any year the company could decide to lay off people or go under entirely.
- Job insecurity is especially difficult for people with families so this causes the demographics of the company to skew younger. A young workforce can also be a con if you want to work with a more diverse, experienced team.

Videory

Similar to: Lyft, Twitter, and Airbnb
Company history: 8 years
Size: 2,000 people

- Videory is a late-stage, successful tech start-up that runs a video-based social network.
- They have just gone public and everyone is ecstatic about it. They're not close to the size of MTC, but they're doing well as a social network and growing the customer base each year.
- They're data-savvy and have probably had data analysts or scientists for a few years now or even since the start.
- The data scientists on the team are very busy doing analyses and reporting to support the business, as well as creating machine-learning models to help pair people with artists to commission work.

Videory

Your Team

- Videory is still at the point where you can gather all the data scientists in an extra-large conference room, or at least a small auditorium.
- Every data science person reports to a data science manager, and they all are in a single large department of the organization. They help other teams throughout the company, but ultimately they set their own priorities, and some data scientists are working on long-term projects that have no immediate benefits.
- There's specialization among the data science team at Videory given the size of the company.
  - The company has some delineation between people who do the heavy machine learning, statistics, or analytics. Videory is small enough that it's possible to switch between these groups over time.
  - There will usually be some interaction between the all the data scientists – training sessions, monthly meetings, a shared slack channel – that you wouldn't find in MTC where there are hundreds or thousands of data scientists.
  - The different sub-teams likely use different tools, and there is a group of people with PhDs who publish academic papers and do more theoretical work.

Videory

The Tech

- Videory has a lot of legacy code and technology, and probably at least a few tools they developed internally. Videory is likely trying to keep up with tech developments and has plans to either switch over to a new system or supplement their existing ones with new technologies.

- As a data scientist at Videory you'll definitely get to learn something new. All these companies have big data and systems to deal with it. SQL won't be enough – the company has billions of events they need to process a month. You may be able to try out Hadoop or Spark when you need to pull out some custom data that's not stored in the SQL database.

- The data science is typically done in R or Python, with plenty of experts available to provide assistance if things are proving difficult. The machine learning is deployed through modern software development practices like using microservices. Because of how well-known the company is for being a successful startup, lots of talented people are working at the company and bringing their cutting-edge approaches.

Videory

Pros and Cons

- Videory can be a good size for data scientists – there are enough other data scientists around for mentorship and support, but the team is still small enough that you can get to know everyone.
- Data science is recognized as important on the company-level, which means your work can get recognition from vice presidents and maybe even the C-suite (CEO, CTO, etc.). You'll have data engineers to support your work. The data pipelines may get slow sometimes or even rarely break, but you won't be responsible for fixing it.
- In a 1000+ person organization, you will need to deal with inevitable political issues. You may be pressured to generate numbers that match what people want to hear (and can tell their bosses to get a bonus) or face unrealistic expectations of how fast something can be developed.
- You can also end up working on things that aren't really needed by the business because your manager asked you to. You'll sometimes end up feeling like you have had no direction or your time was wasted. While not as much as at an early-stage startup, the organization will still be changing a lot – what's a priority one quarter can be totally ignored the next.
- While other data scientists at Videory will be more knowledgeable than you on most data science topics, you might quickly become the expert on a specific one, like time series analysis.

Ethics

- Do work that supports your values.
  - Or, at least, doesn't run counter to them.

- Culture matters
- Product matters

What to look for
- Win-win-win product
- Culture statements
- Public commitment to ethics

The reverse interview

- Any good interview will give you some time to ask questions. You should have good ones prepared!
- Figure out what you care about.

Some that matter the most to me personally:

  - Will I have full admin access to my computer?
  - What is the culture in the office/company like?
  - Can I work odd hours?
  - How do you evaluate success for this role?
  - Do you have any hesitations about my qualifications?

Why do a reverse interview?

- **Companies often use interviews as a time to figure out what they're really looking for.**
  - I suspect this is rarely intentional. But actually interviewing candidates forces a team to talk through what they're actually looking for, and they often realize they had differing perspectives prior to the interview.
- **Companies where "data science" is a new addition need your help in understanding what data science can do for them.**
  - As much as possible, use the interview to sell your vision for what data science can offer at the company, how you'll get it off the ground, and what the ROI might be.
- **Data infrastructure is important and many companies are lacking it.**
  - Many data scientists can attest to being hired at a company only to discover the data they needed wasn't available, and they spent months or years building the tools required for them to start their analysis. Many companies are naive about how much engineering effort is required for effective data science. Don't assume that a company with a grand vision for data science necessarily knows what it will take to accomplish that vision.

To whom would I report?

Why ask this question?

- Your manager is your most important work relationship.

What you should look for:

- Can you work with this person?
- Will they trumpet your success, support you through struggles, and help you do your best work?

How does your team/department/company communicate with each other?

Why ask this question?

- Do they focus on the tools (e.g., Slack, email), on content (e.g., "we discuss design decisions"), or on intent (e.g., empathy, transparency, clarity)?

What you should look for:

- Does their communication style appeal to you?

Can you tell me about current weaknesses of your team?

Why ask this question?

- Unless they're new, they know. They may not share a frank perspective, but when they do, that's usually a very positive signal

What you should look for:

- Flaws that are not superficial. Genuine weaknesses that they acknowledge and have a strategy to address.

What does a typical day in this role look like?

Why ask this question?

- Get a picture of how your time will be divided amongst various responsibiliites.

What you should look for:

- Does this match what you want?

Is there mentorship? If so, how does mentorship work at this company?

Why ask this question?

- Companies that don't invest in the growth of ALL of their employees are trash.

What you should look for:

- Not necessarily a formal mentorship, but is this common, is it encouraged, and how will you develop that relationship?

If you have concerns

You need to find a good fit. You need to **want** to work there.

If you see red flags, dig deeper. Offer to sign an NDA if they say they can't give you details on these answers.

Examples of more probing questions:

- What does success look like for this position? How will I know if I am accomplishing what is expected of me?
- What is the last project you shipped? What was the goal, how long did it take, what were the stumbling blocks, what tools did you use, etc.
- What will my first 90 days in this role look like? First 180 days?
- Why did the last person who quit this team leave? The company?
- If a startup, how long is your runway? How are financial decisions made?
- What would be my first project here? Has someone already been working on this or is this in the aspirational stage?
- What is the current state of the data infrastructure? How much work needs to be done on getting the infrastructure and pipeline into shape before we start analyzing that data?

Additional questions: https://github.com/viraptor/reverse-interview

The “Joel Test” for data science

- Are results reproducible?
- Do you use source control?
- Do you create a data pipeline that you can rebuild with one command?
- Do you manage delivery to a schedule?
- Do you capture your objectives (scientific hypotheses)?
- Do you rebuild pipelines frequently?
- Do you track bugs in your models and your pipeline code?
- Do you analyze the robustness of your models?
- Do you translate model performance to commercial KPIs?
- Do new candidates write code at interview?
- Do you have access to scalable compute and storage?
- Can Data Scientists install libraries and packages without intervention by IT?
- Can Data Scientists deploy their models with minimal dependencies on engineering and infrastructure?

https://guerrilla-analytics.net/2018/01/13/joel-test-of-data-science-maturity/

Homework

For your assigned company, try to figure out:

- Basics about the company (stage, employees, location, etc.)
- What the product is & how it makes money
- As much detail about their data science as you can find:
  - How big is the data science team?
  - What is their stack like?
  - What projects do they work on?
- What can you learn about the interview process?
- What else can you learn? Would you like working there?

Companies

- Stripe
- WalMart
- Airbnb
- IBM
- Palantir
- SalesForce
- Grand Rounds
- Geico
- SpotHero
- Harnham
- Lyft
- Highspot
- Airtable
- StitchFix
- Argo AI