How I Became a Data Scientist

Bolor Turmunkh

First Steps

At the beginning of my fifth year of graduate school at the University of Illinois, with thoughts of impending graduation, I started thinking for perhaps the first time in my life about who I wanted to be. I had lived happily as an information hermit for four years. I had spared little thought for anything other than academic research. It would have been handy if I had kept up with career trends, sought-after skills, or internship opportunities. But as they say, the secret of getting ahead is getting started. So, I buckled down and got started.

After a quick Google search on trending careers of the future and cross-referencing the required skills with my own past experiences, I landed naturally on data science. In this post, I will recount the path to my current position as a data scientist, and describe some differences between academic research and industry work—so that if you are considering the same options, you might be better informed about the trade-offs.

What Is Data Science?

A famous Venn diagram (Google "data science Venn diagram") defines data scientists as having skills at the intersection of coding, statistics, and domain expertise. They are the people who take a business problem, go prospecting for available and attainable data, reformulate the question in technical terms, design and implement a statistical and machine learning task, and reinterpret the results for the business client to ultimately answer the original question. That makes it sound like to be a data scientist you need to be a statistician and a computer scientist with years of industry specific experience. That's not quite true.

The reality is, data science is both vast and new, with specializations and sub-fields quickly developing. Highly sought-after data scientists are people who are broadly familiar with all aspects of data science while being experts in one or two fields. It is a highly achievable career for mathematics graduate students—with some preparation.

When she wrote the article, Bolor Turmunkh was a data scientist at Uptake in Chicago. She has since moved to Apple in Cupertino and is working as a machine learning engineering manager within an anti-fraud engineering field. Her email address is bolor.turmunkh@gmail.com.

How Did I Become a Data Scientist?

There are plenty of resources online that outline possible paths to becoming a data scientist. I will simply describe my own experience.

From the moment I realized I would enjoy being a Data Scientist to finding my first internship, I spent nine months devouring online and free courses on Machine Learning and Python, sent out dozens of applications, got two interviews, failed miserably at one of them and lucked out with the remaining employer, who was willing to give me a chance. It was a small start-up in San Francisco developing enterprise software in Natural Language Processing. They posted internship positions on their website alone, and I daringly emailed the founder directly with my best pitch about why I would be a good match. I say daringly because it is not standard practice at larger and more established companies. But anything goes at start-ups, and this one happened to go swimmingly for me.

During the three month internship, I learned intensely. My technical knowledge deficit was overwhelming at times. But here, my academic training was an asset. Living with overwhelming stress without it paralyzing you is arguably what "PhD grit" is all about. A harder adjustment was the social aspect of the workplace. Like any profession, there are jargons and topics, popular and unpopular opinions, the latest and meanest blog posts, all exchanged electronically in an open, yet entirely silent office. But I found mentors and allies who helped me feel at home.

This internship was just the beginning of my journey to becoming a data scientist. It took another two years and one failed job search cycle before I landed my current position.

How Is Being a Data Scientist Different from Being an Academic?

You are no longer alone.

Intellectual isolation was the hardest part of my academic research experience. Apart from conferences in my particular field in mathematics, and research meetings with my advisor, I had no peers with whom to engage in frequent and technical discussions of the details of my work. That is no longer the case in industry. Not only are my coworkers ready to get as nitty-gritty into my project as I wish to go, they also possess a wealth of experience dealing with similar projects and are happy to share their expertise. Learning in such an environment is exponentially faster than learning alone.

Project time frames are shorter.

Time frames differ significantly from company to company. Larger companies tend to tackle longer-term projects. Software teams typically have shorter time frames due to the nature of the work. So, the scope and the strategic importance of your day-to-day activities will vary depending on where you work. For me, deadlines for large projects are on a quarterly basis and smaller ones are weekly. The shorter deadlines are oftentimes helpful since they force

you and your managers to clearly define goals and criteria for success. On the other hand, short-term goals can sometimes feel short-sighted, if your team's priorities change drastically.

You won't always get to decide what to work on.

This one is a spectrum. Companies such as Apple prefer to set strategic directions and product vision from the top and have them permeate downward. More bottom-up companies such as Facebook prefer a more entrepreneurial feel. Most companies lie somewhere in between, which means you are somewhat in charge of what you get to work on. My team establishes quarterly priorities and project proposals together, which then go through a review process to make sure the proposals align with company goals.

You have more resources.

As a graduate student, the main resource I had was my own time. As such, I was used to solving all my problems on my own. But as a team member, your goal is to arrive at a good solution in the most efficient manner possible. Doing everything yourself is not the most efficient way. Getting help is not only highly recommended, but expected of you.

Done is better than perfect.

It is an entirely new skill for most academics to weigh the costs and benefits of doing the job perfectly versus doing it fast. In industry, one makes this trade-off every day.

Closing Thoughts

The qualifications and projects of a data scientist are quite different from those of an academic mathematician, and yet the actual work is quite similar in nature. The great majority of a data scientist's time is spent defining and redefining an ambiguous problem until it can be clearly stated, and then solved.

Once a data scientist finds interesting results, it is crucial to communicate them to the end customer. Building a story around a complex issue, supporting that story with evidence derived from data, and interpreting the results into a concrete recommendation for the customer, are the central tasks of the job. From this perspective, your graduate training in mathematics, statistics, or operations research will provide a strong foundation for moving into data science.

Good luck with your career transition and job search!

Credits

Author photo is courtesy of Bolor Turmunkh.

BolorTurmunkh

Insights into a Day in the Life of an Industrial Mathematician: Part I

Sharon Arroyo, William Christian, and Rajeev Namboothiri

Students often ask, "What is a typical day when working as a mathematician in industry?" Typical days vary across different companies, positions, years of experience, and interests. Following are insights from our careers in industry that allow us to be curious and creative while providing value to our companies. In Part I, the focus is on how a day varies as we develop a mathematical solution. Part II is focused on how typical days change over the years of one's career.

To appreciate the daily mix of activities you should understand the six general steps taken when developing a solution to a business problem:

1. Defining requirements
2. Developing a mathematical problem formulation
3. Defining an algorithm
4. Implementing an algorithm
5. Transitioning the solution
6. Analyzing results

Each of these steps can take hours to months depending on the type of project. Also, the process is rarely a linear one as we describe below and frequently bounces back and forth as more is learned about the problem.

When a project starts, there typically will be several meetings with the end user, often called the customer, to determine the solution requirements. These meetings could be in a conference room or, as in our cases, they could be on an airplane production line, at an analyst's desk, or at an overseas military installation. The number of meetings and time frame needed to develop an initial set of requirements depends on the difficulty of the problem and the experience of the customer organization in teaming with mathematicians. During these meetings, the composition of the technical team needed will be considered. Mathematicians are always working in teams with collaborators from various departments and with varied backgrounds, which

Sharon Arroyo (PhD) is a Senior Technical Fellow at Boeing and Vice President of Industry at SIAM. Her email address is sharonfarroyo@msn.com.

William Christian (PhD) is an applied research mathematician at the US Department of Defence. His email address is researcher@wchristian .com.

Rajeev Namboothiri (PhD) is a senior scientist at GE Research. His email address is rajeev.namboothiri@ge.com.