6 Things About Data Science that Employers Don't Want You to Know

I want to shed light on the *darker side* of being a data scientist. This article is not meant to discourage you, but like any other job, data science as a career has its downfalls. What I believe is important is that you're made aware of these things so that when you encounter them in your life, it doesn't hit you like a truck (like it did for me!).
And depending on your personality and interests, you might not find any of these much of a downfall at all, which is a good thing! So with that being said, here are 6 things about data science that employers don't want you to know.

1. A vague term, like "data science," means vague responsibilities.

The more you read about data science, the more you'll realize how broad is data science. In fact, it's so broad that there are articles specifically on the different types of data science jobs out there (data scientist, data analyst, decision scientist, research scientist, applied scientist, data engineer, data specialist… you get the point).

Furthermore, as it's a multi-disciplinary field, the term "data science" covers a wide variety of skills, most likely more than you'll be able to perfect in your lifetime.

Therefore keep these things in mind throughout your data science journey…
1. Have an open mindset and try not to stay so fixed on the glamorous parts of data science. For example, if you find yourself querying tables or working on data architecture instead of working on machine learning models, don't be discouraged. Any data-related skill is a valuable skill to know and will most likely come in handy in the future!
2. Similar to the first point, there is no fixed path in data science. Thus, take whatever opportunities come your way and learn as much as you can from each opportunity. The more experience you gain, the more choice you'll have for yourself in the future.
3. And as a last overarching statement, try not to set such strict expectations for what you want to do until you are experienced and knowledgeable enough to be able to. Beggars can't be choosers!

TLDR: Have an open mind when navigating through your data science journey. It's not only going to be about building models.

2. You'll most likely be using SQL a lot more than you think.

When I first started my career, I always thought that SQL was a skill that only data analysts used. And because I had that mindset initially, I never appreciated my developed knowledge in SQL.

This is not the way that you should think about SQL, ever!

If you are working in a data-related role, whether it's a data science role or not, SQL is never going to leave you.

As a data scientist, you'll need data if you want to build machine learning models, which means you're either going to have to query your data or you'll have to build pipelines if the data doesn't exist yet. And it's extremely important that you know SQL well so that your data is robust and scalable.

TLDR: SQL will always be your best friend, so make sure you take the time to be proficient in it.

3. Data in the real world is messier than you can imagine.

If you've ever worked with data on Kaggle, the real world is nothing like it. On Kaggle, the data is typically clean, descriptions are provided for each table, and each column and feature names are fairly intuitive.

This is not the case in the real world. Not only will you not likely have any of the things that I listed above, but you probably won't have reliable data to start with.

I wrote an article on [10 times that I had to work with really messy data](#), but just to give a couple of examples:

- Dealing with differently spelled categories, i.e., United States, USA, US, United States of America.
- Working with data where the logic is compromised. An example of this is if there's a record that shows that a given user uninstalled the same app twice without re-installing it in between… yikes.
- Working with inconsistent data. For example, one table may have told me that our monthly revenue was $50,000, but another table with similar information might have said that our monthly revenue was $50,105.

TLDR: A majority of your time is going to be spent on cleaning your data. It's very unlikely that you'll be able to jump straight into modelling.

4. A large portion of time is spent understanding the business problem at hand.

Whether you like it or not, a data scientist is very much a business analyst. Why? Because you need to have a full understanding of the domain that you're working in and the business problem at hand. Without this, you'll miss out on key relationships, assumptions, and variables that could be the difference between a 65% accurate model and a 95% accurate model.

For example, if you're a data scientist in the marketing department, it is essential that you fully understand every type of marketing channel, what purpose it serves, where it lies in the marketing funnel, what type of users it typically attracts, and what metrics are used to evaluate that given channel.

To give an example, trade shows are generally much more expensive than affiliate marketing ([CAC's](#) are higher). However, the [LTV](#) of customers acquired from trade shows is also higher. Had you built a model only focused on CAC's, you may have given incomplete information resulting in no longer marketing through trade shows. **TLDR:** A significant portion of time should be spent understanding the business problem and the domain that you're working in before diving into any model building.

5. You're not expected to know every tool, but the more you know, the better.

I've previously said that it's better to focus on a few tools and be really good at them. I still stand by that statement, but the sad reality is that your employer(s) will most likely expect you to evolve and learn more tools as you go.

You should definitely know your basic tools well. This means Python, SQL, and Git, as well as several Python libraries including Pandas, NumPy, scipy, scikit-learn, etc…

However, don't be surprised if your employers throw new tools for you to learn asap, like Airflow, Hadoop, Spark, TensorFlow, Kubernetes, the list goes on.

Also, if you switch employers in your career, you'll likely have to learn a new set of tools because every company has its own desired tech stack, so be wary of this when choosing new employers.

TLDR: The learning never ends. If you don't like the sound of that, data science might not be for you.

6. Communication skills are your best friend.

This one is more for those who think that being a data scientist means that you can hide in a room building models all day. No matter what any employer tells you, even if they say you can work at home 24/7 or work as a team of one, you'll be required to collaborate and communicate with other stakeholders.

Even if you're a team of one, you're going to **have to** communicate with upper management about the work that you're doing and the tangible business impact that it's having. You'll also likely have to collaborate with other teams and business analysts to build that domain knowledge that we talked about a few points earlier.

TLDR: Data Science requires a lot more communication than you think, and it's instrumental in being a successful data scientist.