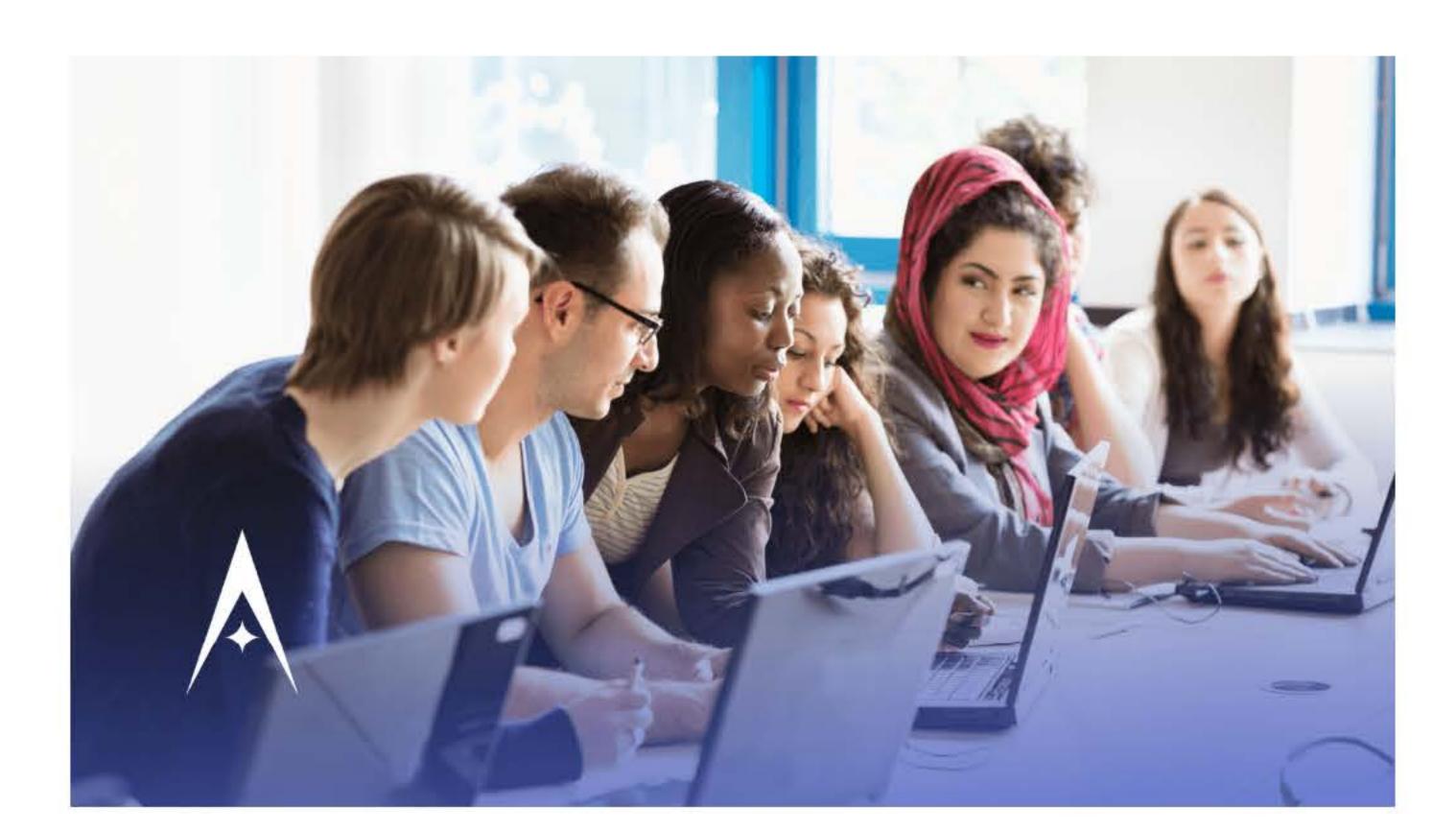# Embracing the Unexpected: The Expectations & Realities of a Student Data Scientist

Written by

Gurman Dhaliwal

Published on

June 20, 2023

For many of us, a major in data science didn't exist when we started our careers. So what's it like for the next generation of data scientists enrolled in one of these college programs? We get perspective from one of our very own Data Science Alliance interns (and Responsible Data Science ambassadors!), Gurman Dhaliwal.

Embracing the Unexpected: The Expectations & Realities of a Student Data Scientist
by Gurman Dhaliwal

Choosing a college major was difficult for me. It felt like the first step to committing to a career and I wanted a little of everything. I liked math and programming, but I also wanted a job that allowed me to be creative, gave me a platform for communication, and was versatile enough to explore different industries. After some research, the data science program at the Halıcıoğlu Data Science Institute (HDSI) at UC San Diego seemed like a good fit. Despite my decision to pursue this path, I still had doubts and the assumptions I made at the start reflected this skepticism. However, as I work through my final quarters, I am glad (and surprised!) by how the realities of my experience have diverged from those expectations.

> **Expectation #1: Data science will be a lot of repetitive math and programming classes.**
> **The Reality: While math and programming are pillars, there is actually a lot of variety in classes.**

Looking back, my classes have had much more variety than I expected. Programming and math classes are a majority but each course offers a different perspective on core topics while equipping us with a myriad of tools. There's also significantly more diversity in the field, ranging from classes on statistical fairness definitions to bioinformatics. I also found niches I especially enjoyed in healthcare, data ethics, and privacy. This helped widen my perspectives on the roles and industries I could enter as a data scientist early on.

> **Expectation #2: I'd be working alone most of the time.**
> **The Reality: I work a lot with others and I am better for it.**

I like working with people. Ideas are generated faster. I feel more creative and it's just more fun! Nevertheless, I initially gave into the stereotype and pictured myself doing my data science homework hunched over a laptop for the better part of my day, so I was surprised by how much group work there was. Nearly all my programming and math classes encourage us to work with at least one other person. Meeting and working with people I didn't know pushed me outside my comfort zone and refined my teamwork and communication skills. Even in professional settings when my work was independent, I found that working with other interns made me a better data scientist. Although we each had similar foundational skills, leaning on one another to utilize our different strengths and areas of focus allowed us to be better as a whole.

> **Expectation #3: Data science is the same as machine learning.**
> **The Reality: Machine learning is just a part of the data science project life cycle.**

To be fair, I didn't know much about data science or how machine learning (ML) was defined when I started my journey. Still, coming into the HDSI program, I thought data science was synonymous with ML. I imagined that most of my classes and work would be creating predictive models and delving into neural networks. Instead, the bulk of courses and work in data science focuses on data cleaning, data expiration, and visualization, with the ML analysis taking less time than you'd expect at the end... at least for now.

> **Expectation #4: My role could be automated.**
> **The Reality: Certain responsibilities can be automated but the creativity of data scientists as problem solvers can not.**

This concern originated during my first natural language processing class where my professor showed how quickly GPT-3 could write code. It was daunting as an entry-level data scientist — how was I supposed to compete with models that could correctly write SQL queries faster than I could read them? However, this exercise was meant to illustrate that our roles as technologists weren't just learning to use tools and understand the inherent processes that allow them to function. Large language models still can't do your homework correctly, but eventually (and inevitably) they will improve, and when they do, I'm optimistic that they'll be more of an aid rather than a detriment to data scientists. Unlike data scientists, LLMs aren't problem solvers. They can't generate original ideas, use creativity to navigate ambiguous problems, or effectively communicate with different audiences. This may change in the future but through my education and professional experiences, I am confident that I can still make a positive impact in the field.

> **Expectation #5: Working as a Data Science Alliance intern would help me become a better data scientist.**
> **The Reality: It did... and it's made me an ambassador of Responsible Data Science.**

I entered the Data Science Alliance (DSA) internship knowing this was an opportunity to further my data science skills by working with "messy" data and using my cleaning, visualization, and predictive skills. Throughout the internship, I did it all. I faced the tedious data collection methods head-on and spent many more hours cleaning raw data than I would've expected. I made interactive charts that drew awe and furthered the narrative. I also learned that out-of-the-box models aren't foolproof methods for solid results.

Beyond all the technical assignments, I also delved more into Responsible Data Science (RDS) to understand what the broader impacts of my work entail. My first assignment at DSA was to read the White Paper outlining what RDS looked like and the current prevalent issues. It was a gripping and comprehensive read about the most substantial issues facing the field of data science. Moreover, at Responsible Data Working Group meetings and mixers, I had the chance to discuss some of these issues firsthand with the paper's authors who are all respected professionals from the industry and academia.

Read RDS White Paper Here

My DSA internship experience introduced me to the conversation on how we can better understand the fallbacks of data science and how we can use it responsibly. It has helped cement my interest in the intersection of data science, policy, and ethics.

The Takeaway

As a part of my data science journey, I've learned to embrace the unexpectedness that comes with reality. I learned that the breadth and depth of data science were ideal for doing a bit of everything: to research, to program, to analyze, and to tell stories. With that, I'm confident in my decision to pursue data science and excited to see what the next phase of my career brings.