The Accidental Data Scientist

Source: Unsplash
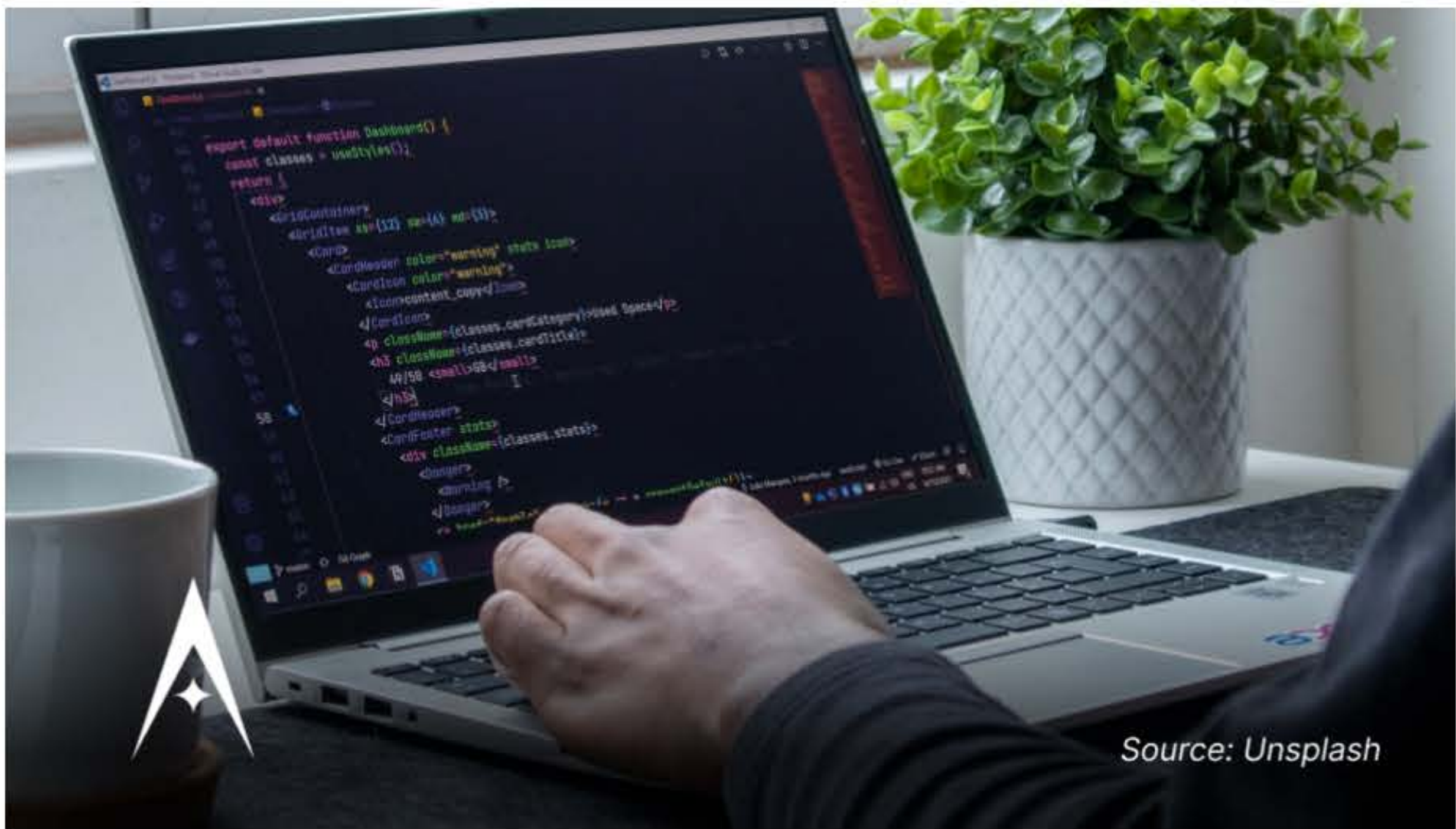
Written by

Taner Halıcıoğlu

Published on

January 17, 2022

As we all gear up for the challenges of 2023, it's imperative that we reflect on where we've been so we can appreciate where we want to go. So it's the perfect time to share a story you may have missed when it was first published in 2020—the "data scientist" origin story of our very own Data Science Alliance Founder, Taner Halıcıoğlu.

It's a tale many in our community can relate to: Taner didn't plan to become a data scientist. He simply followed his passions, put in the work, and became one along the way.
Wherever you are in your Data Science journey, a newcomer or a legend, we hope Taner's story inspires you to pursue what you love and forge your own path. And may all your roads lead to Responsible Data Science!

Told with little fanfare, but tremendous sincerity, we proudly present...

ICYMI: Stories from an Accidental Data Scientist
by Taner Halıcıoğlu

I never considered myself a "data scientist." I'm certainly not one by training, and I didn't go to school to learn how to understand data and use it for business. That being said, when I reflect back on what I've been doing the last nearly 25 years of my life, it turns out data science has been there the entire time.

Since I was a kid, I've always been interested in how things work. I was driven to understand the mechanics behind things and see how all the pieces fit together to make one cohesive unit. Perhaps that's why I've also been interested in numbers and math, which led me to earn a degree in Computer Science from the University of California, San Diego. Looking back — data science and numbers — they were there all along. I just didn't know how they'd all come together until years later.

After graduating, I started a career in technical operations, where one of my key responsibilities was designing, creating, and maintaining "statistics and monitoring" systems for various companies. This meant figuring out how to gather and store technical data about the systems and services, as well as determining what constituted a "problem" with something.

After college, I joined the startup Internet Systems, Inc. (later Globalcenter), where one of my main roles was writing software to do this statistics and monitoring work. I gathered data about the computer and network systems we ran in our data centers, and provided ways to track, monitor, and alert key stakeholders about that data. A lot of that data was used to bill customers, so naturally, the business relied on that data and its accuracy. Then, in 1999, I joined another start-up — Loudcloud (later Opsware) — where I wrote similar tools, as well as did some software development around operationally-related systems.

In 2002, I entered the world of online commerce when I went to eBay and helped create and maintain their monitoring and alerting systems for machines and services. In the beginning at eBay, some of the monitoring systems were very rudimentary. "Is the server up and running?" was not an uncommon question asked around the office. But simply having a server "up and running" is not the heart of a successful online business (although it's certainly a critical component of it). It's the data and information being collected and processed that is the heart of eBay. I enhanced some of the monitoring tools so that eBay would extract useful information (data) from the servers ... data such as, "How much load is on the server?" After that, we started gathering even more service-specific information and all this data was displayed in dashboards that eBay operations center folks could use to tell what's going on at a glance.

Two years later, I was hired by Facebook to build the initial system and network architecture and determine how to solve some technical limitations in order to keep the site running under the increasingly heavy load it found itself facing. This meant looking at tons of data and trying to figure out how best to make adjustments to keep the site going strong.

Facebook was challenging, because — due to how quickly the (now) social media icon was growing as a website and service — we were facing issues that had never really been experienced before. For example, we had to figure out how to solve issues with storing and serving up the millions of photos that people uploaded. This made us work on interesting solutions that needed their own tracking, and we used data to do that.

In 2009, I was hired by Blizzard Entertainment to build a new group called "Reliability Engineering." One of the goals I gave myself was to ensure any new service would have a specific way to monitor it, so we could determine if it was working correctly or not. I did this by extracting service-specific data. For example, for a chat service, we wanted to see how many messages were being sent, so we could quickly tell if something was going wrong. If there was a sudden drop or spike in those numbers, we could receive an alert and look into what might be going on.

In 2013, I came full circle, if you will, when I returned to UC San Diego to teach in Computer Science. I started teaching a seminar class about all the things I'd learned over the last couple of decades of my career. It was then that I really began to realize I'd been using data science to help companies make decisions from a technical-operations standpoint for my entire career. This prompted me to create the Halıcıoğlu Data Science Institute (HDSI) at UC San Diego, where students would be able to really learn about how data is all around us and helps us make smart choices and interesting discoveries.

Whether you're working with a social media giant, or as an entrepreneur at a small start-up, data is one of the most powerful tools we have to successfully make and drive decisions. We don't need huge technical systems (although those are certainly helpful). But everyone can use data to evaluate something, or at least to think critically about something. Data is all around us, and sometimes you don't even realize what you're doing is (technically) "data science."

About the author:
Taner Halıcıoğlu is the founder of the Data Science Alliance as well as the Halıcıoğlu Data Science Institute at UC San Diego. He's always been fluent in the practical and the visionary. To solve intractable problems, he embraces both. Taner practiced the art of data science before it had a common name — leveraging data in concert with his skills as a software developer and systems engineer to help tech giants Cisco, Globalcenter, Loudcloud, and eBay scale their capacity to match their goals. He continued to turn founders' visions into realities when he joined Facebook as its first full time employee.

Originally published at https://bahija.medium.com on December 17, 2020.