

A data scientist's guide to saving time

As a data scientist, you stand at the crossroads of science, engineering, business intelligence, and mathematics. You have a lot to juggle — everything from coding and data visualization to cleaning up datasets and reacting to ad-hoc day-to-day responsibilities, and unless you strictly track your time, you might be surprised how you spend it. In a recent survey conducted by HP of 350 data scientists worldwide, 48% claimed they spent more time organizing their data than actually analyzing it!

Your time is valuable, so efficiencies are essential tools of the trade. Who would have guessed that excelling in your role could have just as much to do with time management as mastering the intricacies of SQL? Here are six tips for optimizing your workflow and making the best use of your time.



Exceptional performance with Intel® Xeon® or Intel® Core™ i9



1

Saving time with proactive communication

Establishing communication touchpoints to ensure you're making the right decisions throughout the project is essential to saving time. Forty percent of surveyed data scientists mention that they often start working with data before fully understanding the business objectives.¹ This lack of communication often leads to managers having unrealistic expectations about the project's outcome.

The problem can be surprisingly nuanced. After all, business stakeholders and data scientists are all fairly technical people — they simply use different vocabularies and communication styles, which leads to both groups sometimes talking past one another. A critical distinction: Business stakeholders tend to think in binary outcomes, while data science is painted in shades of uncertainty. Aligning your data-driven approach with that of your stakeholder's is essential.

“Business stakeholders tend to think in binary outcomes, while data science is painted in shades of uncertainty.”

No matter how technical the topic, it's important to communicate in a way that makes sense to the people who have to implement it. Ken Jee, Z by HP Ambassador² and the head of data science at Scouts Consulting Group, works in sports analytics. “A lot of athletes are not going to be willing to implement a solution that they cannot understand. Starting out with more simple linear models, even though results might not be as good, can be a really powerful way to show the people why we're making a specific decision in a certain way.” In other words, your project can be implemented faster if everyone in the loop understands its value.

2

Save time by getting to know your data upfront



Louise Ferbach

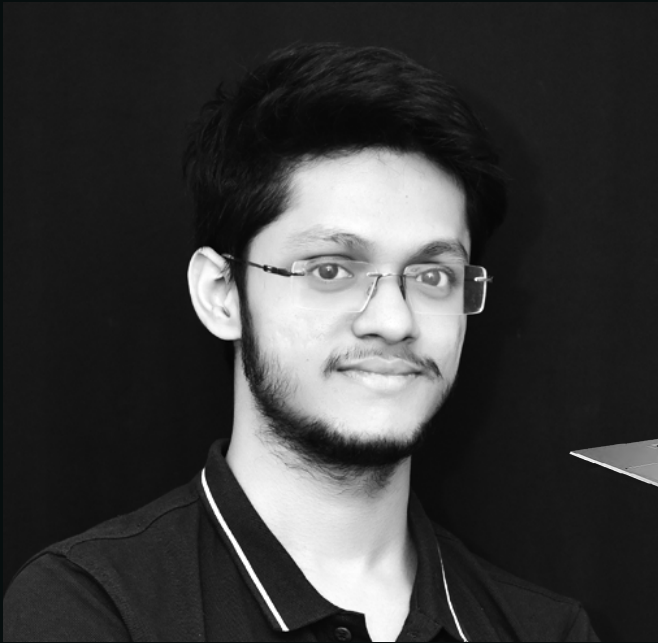
Not only is your time precious, but it's often split between projects, both long-term and short-term. That means making the best use of your time is paramount and optimizing the time at the start of any project can pay enormous dividends later. One common mistake that can cost you time downstream is to start the modeling phase too soon, before you really understand your data. When you start to work on a new project, you're no doubt eager to start modeling – after all, that's the exciting part of the job - but experienced data scientists know better.

Louise Ferbach, Z by HP Ambassador² and an actuary data scientist in France, says that time spent here can be the most fruitful in the entire project – finding correlations and, as Ferbach says, “getting to know your data.” By dedicating time, even as much as a day or two, it's possible to discover patterns that'll help inform your model. In the end, that's actually a huge time gain.

Likewise, a huge part of any data science project is documentation. Getting the appropriate documentation locked down early is a critical way to improve efficiency and save time. Don't neglect properly documenting your own code as you go. Poorly documented code is a bad habit common to both software engineers and data scientists alike, so never assume you'll remember what you intended even a week later, much less a month. Spending time on documentation means you won't have to decipher your code every time that you go back to it.

3

Save time by adding the right accessories to your workspace



Paras Varshney

When it comes to choosing hardware, CPU and GPU are likely your first priority. High-compute power right out of the box is a must-have to meet the needs of your everyday work and to save time.

Often, it's also simple things that make a big difference day to day. A lot of accessory choices are personal decisions, driven by comfort, convenience and preference — there's no one-size-fits-all accessory that will save time and increase efficiency for everyone.

For example, everyone has their own preference when it comes to a mouse accessory. Paras Varshney, Z by HP Ambassador² and data scientist at LogicAI in Poland, rarely uses the clickpad in his laptop; instead, one of his favorite gadgets is his mouse. He says he can work much faster with a mouse. Jee also lives by his trackball.

He works on planes a lot, and he loves that you don't have to move the trackball, which makes him faster while working on the tiny tray table.

Varshney focuses on the display, "I love using a split display. Since I have the HP Z38c curved display, I use it to keep multiple windows open at the same time. I don't have to change windows or switch between tabs — they all are open at the same time in front of me." The nature of your work demands multitasking in many open windows that include apps, browser tabs and dashboards.

That means any laptop screen will slow you down as you open and close or manage your workspace. So, adding the right accessories — everything from the right mouse to curved displays — can help you optimize your workflow and save time.



Exceptional performance
with Intel® Xeon® or Intel® Core™ i9



4

Save time by recognizing your productivity patterns

As a data scientist, pattern recognition is in your DNA. In the same way you spend your days optimizing models, there are opportunities to optimize your daily routines for better productivity.



Qishen Ha



Ken Jee

Because no two people are the same, the same workflow isn't optimal for everyone. In an age when lots of people are opting to work from home, Qishen Ha, Z by HP Ambassador² and a machine learning engineer at LINE, recognizes that the solitary work environment doesn't suit him; he feels too easily distracted. Instead, he prefers being in the office, surrounded by other people who all are working. "You push yourself to work more intently, I think," he says.

On the other hand, working from home full time, Jee has thought a lot about his workflow and has established blocks of time throughout the day for processing emails. He doesn't like the inefficiencies that arise from multitasking his core work and sending emails, and mentions that every time he looks at an email while doing other work, he loses precious minutes while trying to re-engage on his primary task. In an age when everyone is expected to react to messages in email or Slack virtually in real time, bucking that trend by scheduling blocks of time for emails can be critical to making the best use of time.

5

Save time with the right tools & configurations

Configuring a new computer is always a challenge and when surveyed, 42% of data scientists lament they spend too much time configuring their data environment, with an average of five hours per week lost.¹



One significant way to improve your efficiency is by adopting Windows Subsystem for Linux[®] — WSL 2³ lets you virtually run Linux tools, utilities, and applications directly within Windows without resorting to a dual-boot configuration or virtual machine. Jee, for example, says that he no longer has to remotely access his Linux workstation while working at his Windows desktop — an improvement that reduces friction and speeds up his workflow.

Likewise, preconfigured software stacks have proven to be nothing short of a revelation for data scientists who have had the pleasure to use one.

“I didn’t know such a thing actually existed before I was confronted with it, and it has absolutely changed things,” says Ferbach. “When Z by HP sends you your Data Science Software Stack.” That software stack is essentially a comprehensive suite of applications and environments — everything pre-loaded with automatic updates, avoiding the inevitable software incompatibilities and troubleshooting time that plague routine setups.



Exceptional performance with Intel® Xeon® or Intel® Core™ i9



Save time with workflow automations

Depending on where you are in your data science career, you likely already optimize your workflow in some ways. However, it's important to be mindful of additional ways to optimize your time with increasing sophistication as your skills and experience grow. "Simple skills, like screen splitting, can be done by people who are new to data science," says Varshney, referring to how to best make use of a large display.

A more sophisticated tool for your toolkit? Automation. Any data scientist with some experience in the rear-view mirror knows the value in automating their workflow. After all, it doesn't take long to see that some tasks require a lot of manual processing and may need to be done again and again, so automating those tasks can save enormous amounts of time. Just a few commands can enable you to run an entire process autonomously.



The Z by HP advantage

As evidenced by all the above tips, there are so many habits you can adopt that will make the work you do as a data scientist more efficient. Even a few small steps can make the time savings add up. That said, the most fundamental time-saver for data scientists will always be determined by the power of their workstation. That's why Z by HP is constantly innovating to bring data scientists the high-compute workstations, displays, and tools they need to manage their tasks as seamlessly as possible. Check out the [tools and libraries included in the preconfigured software stack](#) that Z by HP offers, which can help you save time or read about [WSL 2](#), offered on select Z by HP data science workstations.

¹ HP proprietary research: Understanding Data Scientists, November 2021.

² Disclaimer: Ken Jee, Louise Ferbach, Qishen Ha and Paras Varshney are part of the Z by HP Data Science Ambassador program and are provided with products.

³ WSL 2 requires Windows 10 or higher, Intel Core i5 processor or higher and is available on select Z workstations. You must be running Windows 10 version 21H2 and higher (Build 19044 and higher) or Windows 11.