**10 Simple Things You Can Do to Improve Your Data Science Skills in 2023**

**Here’s to making this next year productive, satisfying, and all-around beneficial.**

Ah, a new year.

An opportunity for growth, development, and advancement. A chance to begin anew — to put away the difficulties of 2022 and cultivate a bright, successful 12 months.

And if you’re a data scientist, a chance to continue to develop your skills in this ever-growing and impactful field in an effort to do good in the world. What better resolution could there be than that?

Allow me to help you on your way — with 10 things you simply must do to take your 2023 data science game to a whole new level.

**Sign up for an online class**

Data science is an ever-changing field. As a result, you need to stay on top of your game. I still hear people insist that there is no need to learn Pandas, because R is enough. Or that the only language data scientists will ever really need is SQL.

Sure, there are probably some data science jobs out there that thrive on SQL, or even R. But that’s not the point. If you’re unwilling to keep up with and learn the latest technologies, you’ll be quickly left behind in an industry that waits for no one.

So please, sign up for an online class, go through a detailed YouTube tutorial, learn a new technique on [Kaggle](https://www.kaggle.com/), do SOMETHING.

You’ll be better off for it.

**Commit to learning one specific, new skill**

This is related to the above point, but not quite the same. In a previous article, I discussed various [underappreciated skills to make you into a next-level data scientist](https://medium.com/towards-data-science/3-underappreciated-skills-to-make-you-a-next-level-data-scientist-6b5236770651). If nothing I suggested flies your plane, there are likely a plethora of similar articles you could find online.

There are two advantages to expanding your skill set beyond your current comfort zone:

1. You’ll become a more *diverse* data scientist.
2. You’ll become a more *unique* data scientist.

And both of these advantages will lead to an even one: companies will want you, and they’ll want you bad.

**Teach someone, anyone, that machine learning isn’t magic**

Last year, I worked on a data science project involving a predictive machine learning model. I only joined during the initial phases, before handing it off to someone in pursuit of brighter pastures (as you will ascertain throughout this article, I have a mild skepticism toward machine learning).

Some months later, the colleague who took my place expressed frustration. Why? Because various other folks on the project team — namely, those who were there for project management, domain expertise, etc., and had limited knowledge of data science — continually pushed my colleague to develop a machine learning model when they didn’t even have access to the data yet. They insisted it was possible.

This sounds like a stretch, but it’s true. And I don’t blame the other folks either. Due to the powerful capabilities of upcoming technologies like self-driving cars and virtual home assistants (I’m looking at you, Alexa), a growing number of people have started to view machine learning as some kind of magical black box that can do anything.

I don’t mean to burst your bubble, but machine learning is just a bunch of spreadsheets and complex mathematical equations. It’s powerful, but it’s only as good as the data, and it can result in some pretty terrible results (think discriminatory facial recognition algorithms or self-driving cars that hit people) if your data isn’t up to par.

More people need to know this. As a data scientist, help spread the word.

**If you don’t know Pandas, learn it**

And if you do, [learn something new about it](https://pandas.pydata.org/docs/user_guide/index.html#user-guide).

If you go through some of my data science articles, you’ll find I am a big advocate for Pandas. And with good reason.

Put simply, Pandas is one of the most powerful tools out there for data processing, manipulation, and analysis. It’s so well-established that more specialized modules are designed to work with its primary data storage method, the DataFrame:

* Scikit-learn for machine learning
* Altair for visualization
* SciPy for scientific computing

Developers and researchers are also consistently developing tools built on top of Pandas to make it even better. [Here’s a recent example](https://ponder.io/).

If you are a data scientist, learn Pandas. Please.

Don’t make me tell you again next year.

**Learn a thing or two about qualitative research**

I work closely with someone whose research is on the cutting edge of data science. He’s working to build more equitable and accurate social recommendation systems (if you don’t know what that is, an example is the algorithm that suggests things for you on your Instagram’s Discover page).

His work has been published at some top-tier conferences and he’s on track to get his PhD from a top-tier institution. He’s also a great person to go to with questions about data science.

The kicker? He almost never writes code or runs quantitative tests. Sure, some of his projects involve these more technical elements, but they aren’t his area of expertise, nor are they the primary focuses of the projects.

He’s a qualitative researcher through and through, and he uses his combined knowledge of the technical landscape (he was a programmer in a past life) and state-of-the-art qualitative research techniques to learn some really cool things from social media data.

So, if you’re overly attached to your code and numbers (like myself), consider learning a thing or two about qualitative research.

Good data science requires a bit of both.

**Learn to design a user study**

More and more folks are starting to realize that blind trust in technology can be foolish and dangerous. As a result, good data science is becoming more human-centered. A colleague of mine who now works as a data scientist graduated to that role directly from her work in user experience at the same company.

Whether you’re designing a new model, algorithm, or visualization, it is essential to test it to determine practical usability and ensure proper ethics. However, running a proper, rigorous user study is no walk in the park. It involves careful design, a working knowledge of statistics, and the ability to put into practice what you learn.

This is a skill I myself lack; I’ll be taking a course on it in early 2023 to fill the gaps. If you’re interested, [you can access the materials online for free](http://depts.washington.edu/acelab/proj/Rstats/index.html).

**Work on a personal project**

Back when I was in college, there was a lot of discussion around internship application season regarding what the best quality to highlight on a resume was.

Some (less trustworthy) students insisted GPA should be in big, bold letters. Other (smarter) ones recommended focusing on your previous experiences and coursework.

But the students who truly knew what they were doing — the older ones who’d been through this whole process before — gave the most precious advice: **assuming you’re applying for an industry-based technical position, projects should be one of the most prominent sections on your resume.**

Employers aren’t nearly as concerned about where you’ve worked or what your degree is in as they are about *what you can do*. Even when reading about your experience, all they’re really looking for are the details of your accomplishments at those respective jobs.

Not only will building a personal project contribute to your resume — it’ll also help you hone your data science skills.

It’s a win-win situation.

**Expand your view of data science**

Since I have a penchant toward the data science sub-field of visualization, I’ll use it as an example here. Consider this example as a microcosm for all of data science.

When most folks hear the words “data visualization,” they think of line charts or bar graphs, perhaps even a histogram if they’re feeling adventurous. Viewed from this perspective, the world’s earliest data visualizations were developed some time around the 15th and 16th centuries, when such charts started to appear.

This is an extremely limited view. What’s the point of data visualization? It’s meant to take data that is in some hard-to-interpret form — be it numerical, textual, or otherwise, and represent it in a visual way that’s easy to understand. It can be *anything* that takes such data and applies a visual transformation to it.

Consider the *Imago Mundi* below, a map from the Babylonian Empire estimated to have been carved over 2500 years ago [1].

This is a visualization which takes complex geospatial data and represents it in a way which is much easier for the average person to understand. It was no less valuable to the Babylonians than the maps we see on the news today, and acknowledging that provides a foundation to expand our own perspective with regard to the visualizations we might produce today.

Don’t get hung up on textbook definitions of “data science.” Successful data science requires imagination and cleverness.

Broaden your perspective.

**Write an article**

I admit to being a bit biased here, but I’m taking the opportunity to drive home this point.

I’ve taught introductory computer science and data science for many years, and I’ve been lucky enough to interact with some experts in the fields. Specifically, experts in technical *education.*

Although individual preferences and opinions varied greatly, everyone I’ve spoken to held one belief in common: **one of the best ways to learn something is to teach it**.

Every time I write an article about some technical data science topic, I come out of it with a deeper understanding than I previously had. That’s simply the nature of writing. Being forced to articulate a concept clearly requires me to understand it deeply. On top of that, motivated by the fear of misleading others, I’ll often research the topic and review it with various resources first, adding even further to my own learning.

Is it often time-consuming and intimidating? Yes. But is it well worth it? Most definitely.

Give it a shot — you never know what you might learn.

**And finally, take a break**

The tech industry is notorious for its intensity. Friends of mine who are Amazon software engineers routinely describe how overworked folks are. Elon Musk openly stated after his Twitter takeover that anyone who stayed needed to be willing to work in extremely demanding conditions.

It’s easy to fall into a pattern which focuses on working and bettering yourself 24/7, all the while forgetting what life is truly about. You want to do data science to make the world a better place, right? Then start with yourself. If you overwork yourself, you’ll become bitter — at everyone and everything. When that happens, it’ll be hard to remember why you started in the first place.

Trust me — I’ve been there. It’s a place better left unseen.

So be kind to yourself, take a breath, and maybe worry about that data set after a nice tropical vacation.

**Recap + Final Thoughts**

Here’s your 2023 data science cheat sheet:

1. Take an online class.
2. Learn one new specific skill.
3. Spread word that machine learning isn’t magic.
4. Please, for the love of all that is good in this world, learn Pandas.
5. Pick up a thing or two about qualitative research.
6. Look into the powerful potential of user studies.
7. Start working on a personal project.
8. Broaden your perspective of data science.
9. Write something and post it somewhere!
10. Breathe. Relax. And take some time off.

Cheers to a wonderful, ethical, data-informed new year.