



Member-only story

The Most Undervalued Skill for Data Scientists

Why writing is crucial for technical roles, and how to get good at it



Tessa Xie · Follow

Published in Towards Data Science · 8 min read · Jul 3, 2024



693



8



Image by author (Midjourney)

“Why is my manager nitpicking my write-up? What difference does it make changing the wording from X to Y?”

You have probably caught yourself thinking this when you see your managers’ numerous suggestions all over your document; I know I have. In fact, I used to think that writing is the most trivial part of the job of a data

scientist; because the analyses and numbers should speak for themselves right? **Wrong!**

Over the last years, I have realized that writing is an essential skill for data scientists, and that the ability to write well is one of the key things that sets high-impact data scientists apart from their peers.

In this article, I will first convince you that writing is at least as important as your technical skills, and then give you concrete tips to help you improve your writing.

Why is writing so important for data scientists?

1. It's used everywhere in the corporate world — I have highlighted the importance of communication in my previous [articles](#) and like it or not, majority of communication in the corporate world happens in a written form. From project-scoping documents to weekly updates, analysis and experiment write-ups, feedback and performance reviews, JIRA tickets and wiki pages, everything counts on effective written communication to get the message across.

2. Writing helps to bring clarity to your thinking process — Paul Graham, cofounder of the famous startup accelerator Y Combinator (who's a computer scientist AND writer) famously said in [one of his memos](#):

If writing down your ideas always makes them more precise and more complete, then no one who hasn't written about a topic has fully formed ideas about it. And someone who never writes has no fully formed ideas about anything nontrivial.

— Paul Graham

Very often, when you start writing things down, you realize how little you know about a subject and the potential gaps in your thinking/analysis.

3. Writing is the “last mile” of your data science work. None of your stakeholders will read your SQL query or look at your Jupyter Notebook (a lot of engineers and data scientists would like believe the opposite but trust me,

they likely won't). If you want your work to be understood by others and influence decisions, then you need to do the final step of packaging it in an effective write-up. If you skip this step, it's like leaving the package in the warehouse instead of delivering it to the customer.

Image by author

What does “good” writing look like in data science?

Be clear about your audience. If you are writing for everyone, you are writing for no one. Be very specific about who this particular piece of writing is for, and tailor it to that audience and their needs.

Focus on the “so what”; the sausage-making goes in the appendix. As data scientists, we love to talk about the complex analysis we did or how we designed the experiment. Because we put in all that work it feels so wasteful to NOT talk about it. But the harsh truth is, most of the time, our audience does not care; they just want to understand the takeaways.

You can describe the technical details of your work in the appendix in case someone wants to go deep, but the main part should focus on the insights and recommendations.

Have a clear storyline. Fiction or not, every piece of (long-form) writing should be a story. Because that's how humans communicate and that how

our brains process information. Usually the storyline for analysis goes like this:

➔ We found out about something interesting and this is why you should care about it / what you should do (**summary to get your readers hooked, including a recommendation if applicable**)

➔ Here's how we arrived at these insights (**analysis details for the curious explorers**)

➔ Here are the caveats and alternative paths forward (**optionality in case someone challenges the recommendation**)

➔ Here are additional resources you might find interesting (**appendix for those that really want to go deep on the topic**)

It might help to build the skeleton first before adding the details. If the story depends on how the analysis goes (which is often the case for DS analysis since the nature is more exploratory), at least figure out the structure of the doc before diving into the details.

If you are building a deck/presentation, I have a little

Have a clear summary. If you remember the pyramid principle I mentioned in my previous post about communication, it's especially important to written communication. Because the summary is your first touch point with your readers, it should be interesting enough to capture their attention so they want to read on; at the same time, it should capture all the essence so if they decide to stop reading after the summary, they got all the most crucial information they need to know.

Be succinct. When it comes to writing, less is more.

Keep it simple. We work in a technical field and use technical jargon all the time. Often, data scientists think it makes them seem more competent if they use technical language. If you look closely, though, you will notice that the more senior people become, the simpler their choice of words. VPs and

C-Level executives can explain complex topics in language that anyone can understand, regardless of their (technical) background. You can use tools like the [Hemingway app](#) to check if your writing is too complex.

Use signposting. Signposting is a technique that makes it easier for the reader to understand your document. The core idea is to use words and phrases that make it immediately clear what the sentence or section is about, so that readers can quickly skim the text and make sense of it. For example:

- Using the phrase “for example” before you give an example
- Writing “in conclusion” before you summarize
- Labeling sequences of arguments with “Firstly / secondly / finally”

Always have your readers in mind – what do your readers care about the most from this piece of analysis? What do your readers already know about the context/background and what else if necessary for them to know?

Add visualizations. It’s a cliché for a reason: “A picture says more than a thousand words.” When you are trying to communicate dense technical content, a crisp diagram, framework or flowchart can help a lot to get your point across. For example, illustrating what “pyramid principle” means like the graph below will hopefully give you a better idea about how to carry it out in your own writing.

Image by author

How can you improve your writing?

Read a lot. This includes both guides on how to write well (by reading this post, you've made the first step!) as well as strong technical writing that you can imitate (you can find some examples here).

If you want to dig deeper into the science of writing well, I recommend you take a look at ["On Writing Well"](#) by William Zinsser.

Practice, practice, practice. As with everything else, practice makes perfect. Here are a few concrete things you can do to practice your writing:

1. **Document your work in a personal wiki.** Few data scientists do this in my experience, but it's a very useful resource to have and a great way to get more writing practice.
2. **Write structured Slack messages.** Most of the Slack messages we send and receive all day feel like a stream of consciousness (or worse, like teenagers' text messages). People tend to type what comes to their mind and hit "Send" without taking the time to structure the message in a way that makes it easy for the reader to understand it. Writing succinct,

structured Slack messages using the principles discussed above is a great way to stand out.

3. **Write online.** Writing these posts on Medium is ongoing writing practice for me. Try it out; you might even enjoy it and find an audience that enjoy your insights.

Challenge yourself. “You are your own worst enemy” might not be a bad thing when it comes to writing. You need to be able to read your own writing like it’s your first time seeing it so you can be objective about what’s missing, what’s confusing and what needs to be shortened.

Ask others to be your devil’s advocate. Being your own devil’s advocate can be extremely hard sometimes, because true objectivity requires you to abandon your current knowledge about the topic and your ego. It’s sometimes just easier to find another challenger for your work. Ideally this is someone who truly knows nothing about the subject matter and is willing to be very honest with you about their opinion.

What are some good examples of strong technical writing?

I described above what good writing looks like in theory but it’s easier to understand once you see a few examples of it. Here I’m providing some concrete examples for some of the points I mentioned above so you can have a better idea about how to put those suggestions into practice.

Top highlight

Clear audience

The [Data-Driven VC](#) newsletter is targeted specifically towards Venture Capitalists and startup founders who want to take a data-driven approach to investing in and growing companies. While this creates a niche blog that might not appeal to everyone, picking this specific target audience makes it easier to provide value for them.

Strong visualizations

For a crash-course on how to visualize complex systems and technical subject matter in general, check out [ByteByteGo](#). Their diagrams make it

super easy to understand things that would take multiple paragraphs of jargon to describe accurately.

SeattleDataGuy is also using plenty of visualizations, but typically in a slightly less serious way (e.g. see his post on Apache Iceberg [here](#)).

Keeping it simple

Gergely Orosz, who writes *The Pragmatic Engineer*, does a good job summarizing complex topics in relatively simple terms. E.g. check out his [post on how AI Software Engineering Agents work](#).

Combining best practices: Simple, succinct language with clear visualizations

[Daily Dose of Data Science](#) is a prime example for how to combine multiple best practices to produce easy-to-understand but still insightful data science content.

For example, check out their recent post on [Confidence Intervals and Prediction Intervals](#). Or their super brief, but informative post on [Cross-Validation Techniques](#).

In conclusion

Being able to write (well) is crucial for your work, even (or, you could argue, especially) for technical folks. Being able to succinctly communicate your thoughts on paper takes practice. Reading a lot, writing a lot and being open to feedback are the keys to getting better at this craft.

For more hands-on tips and interesting topics about data science, definitely follow me here on [Medium](#), or on [LinkedIn](#) or on [Substack](#)

[Data Science](#)[Careers](#)[Writing](#)[Editors Pick](#)[Career Advice](#)



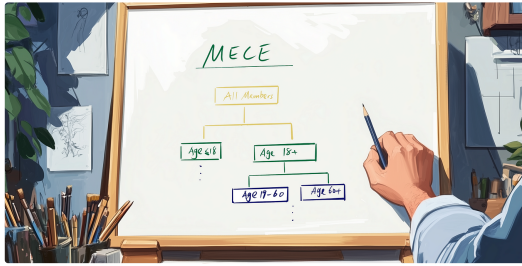
Written by Tessa Xie

3.5K Followers · Writer for Towards Data Science



Data Science Manager in Tech, Ex-McKinsey Data Scientist; Avid Traveler, Diver and Painter. 🐼🍷 Writing a DS newsletter at <https://www.divingintodata.com/>

More from Tessa Xie and Towards Data Science

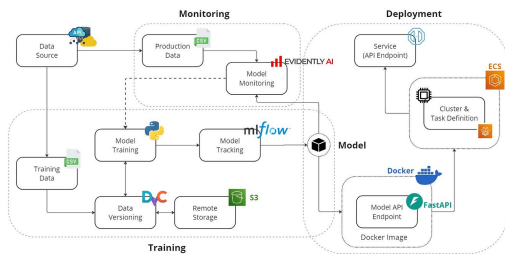


Tessa Xie in Towards Data Science

Analytics Frameworks Every Data Scientist Should Know

Why I believe my experience at McKinsey made me a better data scientist

★ Aug 28 🖱️ 1.1K 💬 19 📌 ⋮

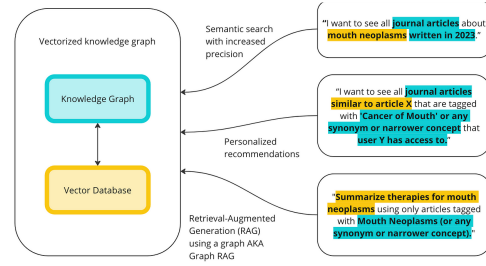


Prasad Mahamulkar in Towards Data Science

Machine Learning Operations (MLOps) For Beginners

End-to-end Project Implementation

Aug 29 🖱️ 1.3K 💬 13 📌 ⋮



Steve Hedden in Towards Data Science

How to Implement Graph RAG Using Knowledge Graphs and

A Step-by-Step Tutorial on Implementing Retrieval-Augmented Generation (RAG),

Sep 6 🖱️ 1.3K 💬 14 📌 ⋮



Tessa Xie in Towards Data Science

How to Better Communicate as a Data Scientist

Most valuable lessons from my time at McKinsey

★ May 29 🖱️ 886 💬 11 📌 ⋮

See all from Tessa Xie

See all from Towards Data Science

Recommended from Medium



Alexander Nguyen

I Wrote On LinkedIn for 100 Days. Now I Never Worry About Finding a

Everyone is hiring.

★ Sep 21 🖱️ 22K 💬 403 📌 ⋮



Egor Howell in Towards Data Science

Advice from 15 Top Data Scientists

Going over the main skills you need to be a "good" data scientist

★ Sep 27 🖱️ 459 💬 5 📌 ⋮

Lists



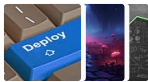
How to Career Plan When You've Already Started a

10 stories · 467 saves



My Favorite Writing Advice & Inspo

15 stories · 1214 saves



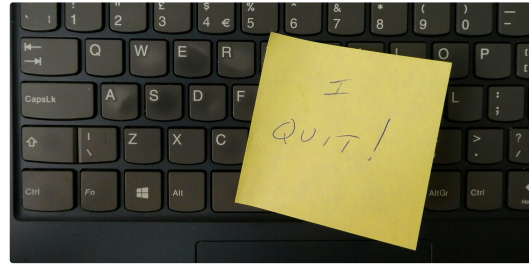
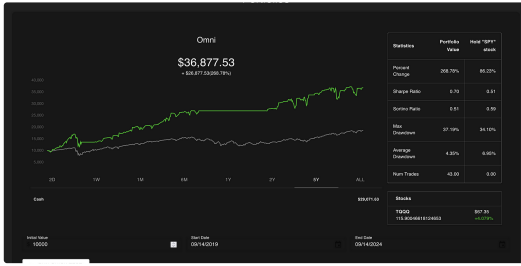
Predictive Modeling w/ Python

20 stories · 1572 saves



Our favorite writing prompts and inspiration

15 stories · 858 saves



Austin Starks in DataDrivenInvestor

I used OpenAI's o1 model to develop a trading strategy. It is

It literally took one try. I was shocked.

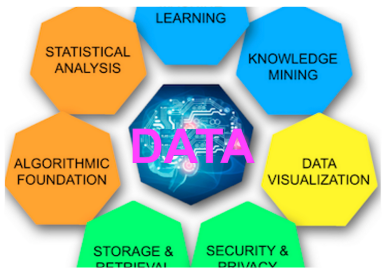
Sep 15 3.1K 88

Imad Adrees in The Pythoneers

3 Reasons Most Data Scientists Resign after 1.7 Years

The reality of the 'sexiest' job in the world.

Sep 12 188 10



Joseph Robinson, Ph.D. in Towards AI

5 Essential Machine Learning Techniques to Master Your Data

A Comprehensive Data Science Guide to Preprocessing for Success: From Missing

Sep 25 510 4

Bhavik Patel in Product Coalition

Analysing Tech Layoffs: Which Roles Were Hit Hardest?

In the aftermath of tech layoffs, it would appear that engineering was hit hardest, but

Sep 17 1K 18

See more recommendations