



SDS PODCAST EPISODE 563: HOW TO ROCK AT DATA SCIENCE — WITH TINA HUANG



- Jon Krohn: 00:00:00 This is episode number 563 with Tina Huang, the superstar data science YouTuber. This episode is brought to you by Neptune Labs, the metadata store for MLOps, and by Einblick.ai, the collaborative way to explore data.
- Jon Krohn: 00:00:18 Welcome to the SuperDataScience podcast, the most listened to podcast in the data science industry. Each week, we bring you inspiring people and ideas to help you build a successful career in data science. I'm your host, Jon Krohn. Thanks for joining me today. And now let's make the complex simple.
- Jon Krohn: 00:00:50 Welcome back to the SuperDataScience podcast. We've got an extra fun one for you today with the beloved data scientist and content creator, Tina Huang. Tina is a YouTube superstar with over 270,000 subscribers and individual videos with over a million views. Her content focuses on getting started in data science, learning to code, SQL, productivity and study techniques. In addition to being a prolific content creator, Tina is a data scientist at one of the world's largest tech companies. She keeps the firm anonymous so she can publish more freely. She previously worked at Goldman Sachs and at the Ontario Institute for Cancer Research. She holds a Masters in Computer and Information Technology from the University of Pennsylvania and a Bachelors in Pharmacology from the University of Toronto. Today's episode should be appealing to a broad audience. Whether you're thinking you of getting started in data science, are already an experienced data scientist, or you are more generally keen to pick up career and productivity tips from a lighthearted conversation.
- Jon Krohn: 00:01:54 In this episode, Tina details her guidance for preparing for a career in data science from scratch her five steps for consistently doing anything, her strategies for learning effectively and efficiently, what the day to day is like for a data scientist at one of the world's largest tech

companies, the software languages that she uses regularly, her SQL course, and how her science and computer science backgrounds help her as a data scientist today. All right, you ready for this amusing episode? Let's go.

- Jon Krohn: 00:02:28 Tina, welcome to this SuperDataScience podcast. I'm so excited to have you here. How you doing?
- Tina Huang: 00:02:36 I'm doing good. Thanks so much for having me.
- Jon Krohn: 00:02:39 Yeah. I've been wanting to do this for a long time and now it's happening. Goodness gracious. So Tina, where in the world do you live? Where are you calling in from?
- Tina Huang: 00:02:50 I am calling from the Bay Area, California.
- Jon Krohn: 00:02:54 Nice. And then I don't know if you remember how we initially interacted, how we initially met. So about a year ago, Harpreet Sahota and Kate Strachnyi, they ran a Data Community Content Creator Award show. And in the run up to that, Harpreet, who was our guest on episode number 457, by the way, really great episode, really fun guy. Harpreet posted on LinkedIn, "Here are some amazing content creators that I recommend you check out and consider using them as nominees for the Data Community Content Creator Awards." And you were one of the YouTubers that he very highly recommended. And so I clicked on your LinkedIn profile. I loved your background. I instantly thought that you would make such a perfect guest for the show. And so I made a LinkedIn connection request. And in that connection request, I asked you if you'd like to be a guest on the program and you just never responded.
- Tina Huang: 00:03:58 I was going to say, did I enthusiastically say yes? Did I accept your request though?



- Jon Krohn: 00:04:04 You accepted the request. So I wasn't heartbroken. But yeah. You didn't respond. And I was like, "Oh, I guess she's just really busy. She's just too cool for SuperDataScience." But then more recently, I had Ken Jee on the show in episode number 555, and you and Ken are very close. And so I was able to leverage that into a connection to you. And then here we are. That's how I tricked you to being on the show.
- Tina Huang: 00:04:33 That's actually really funny though, I think it's because at that time I was doing connections, as opposed to turning on creator mode. And I actually had a script that just accepted everybody. So it is possible that it could have just accepted it, because it was a lot of effort. It's clicking each thing.
- Jon Krohn: 00:04:53 Tina, that is the most elaborate excuse to try to say something kind to me, but I really appreciate it.
- Tina Huang: 00:05:01 I can't even make something up like that. I don't have the ability of even making something up like that. So I'm quite truthful most of the time.
- Jon Krohn: 00:05:09 I believe you. All right. So you're not just famous for your LinkedIn acceptance scripts. You're also famous for your YouTube channel, which has over 250,000 followers. By the time that this episode goes live, it'll probably be much, much larger, even, growing so quickly. And your channel features technical videos on how to start in data science and things like your SQL Sundays series, your SQL Sundays series. Say that three times fast, which is super helpful for interview prep, for big tech interviews. But in addition to those technical videos, you also have more general videos on habits, study tips, motivation, and productivity. We're going to touch on lots of these topics in today's episode, but let's start off with one of your most popular video topics, how should somebody start in data science today?

- Tina Huang: 00:06:10 Excellent question. First of all, just want to put a caveat out there, a YouTuber thing, hashtag, "Not financial advice." I think first of all, it really depends on what you consider to be data science, because data science is such a vast field these days. People do so many different things and it's not a career track that's well defined at this point, compared to something like software engineering. So just putting that caveat out there. So in my opinion, my humble opinion, there are three different aspects that are extremely important for data science. The first one is coding and programming, the second one is math and stats. So my friends out there, I'm sorry, can't go away with math and stats here. And the third one is what people usually call a business where a product sense. This one's pretty interesting. And it's one that people don't really think about that much.
- Tina Huang: 00:07:08 But going through each of these quickly, programming is really important and that's actually where I would recommend people starting, mostly because it's more motivating when you can get something to work. At least for me, I have a very short attention span. So I'm just like, "Wow, at least it works," and that motivates me. And then math and stats. I also would recommend really just starting off with some of that basics, just enough for you to get started in a data science project. I think too many people freak out about math. Like, "Oh, I forgot everything. So I better start from. I don't know, calculus and addition and such." And of course calculus is extremely important, but it doesn't need to come in now. I'm preaching to the choir here.
- Jon Krohn: 00:07:47 I like that. Calculus and addition. We're going to want to brush up on the field of calculus and also putting two numbers together.
- Tina Huang: 00:07:57 Quite important. Don't you think, Jon?

- Jon Krohn: 00:07:58 I do. I do think addition is important. That is one that probably you should brush up on. That is if addition is a topic that you're struggling with, you're going to want to get that down before proceeding too far in data science. The calculus can wait a little.
- Tina Huang: 00:08:14 Yes. So get your addition down, maybe subtraction if you're really feeling it. So I would say, I want to say intro statistics, whatever you take in first year, the first couple courses enough to get you started that you don't have to go and try to deep dive into everything. If you want to be interested in machine learning, just get a general overview of how they work. It's not really necessary for you to get all the exact details, at least for now. And then the third component in terms of business sense, this is one that's pretty interesting. I see a lot of data scientists who do a lot of projects that are very cool, but they don't actually do anything. A very good example is someone goes like, "Oh, I'm going to script Spotify or something like that. And then do an analysis." Perhaps that could be useful.
- Tina Huang: 00:09:02 So maybe that's not the best example, but it's often, you don't think about the business context like what's the problem you're trying to solve? What are you actually going to do? Because that's really important, right? Because whatever fancy thing that you do, if nobody cares about it, it doesn't really matter how cool it is. And then I also recommend an iterative approach. So get the basics down for all of these things and then just start working on projects. And as you go, you're going to start having a lot of gaps in your knowledge. And then you can deep dive into these gaps and just work on more projects afterwards. And I call this a breath first approach for any of my CS nerds out there.
- Jon Krohn: 00:09:37 Nice breath first approach and a project-based learning. I definitely highly recommend that. Just get your hands

dirty. I mean, you could teach yourself addition, for example, through programming.

- Tina Huang: 00:09:50 Precisely. You're like, "I forgot how to do addition," and then you're like, "Do addition. Oh yeah."
- Jon Krohn: 00:09:56 Yeah. Experiment with that plus sign in the programming language and see what happens. Well, so even for more complex math topics than addition. So for example, calculus, in 2020, I got really into relearning partial derivative calculus as it applies to gradients in machine learning. So critical concept, a lot of machine learning algorithms learned with gradients and we need partial derivative calculus to make them happen. So I would simultaneously, I'd get out paper and pencil and I would derive the partial derivatives of different machine learning cost functions. And then I would calculate, I'd put in some dummy values and then I would compute with my hand computed partial derivatives on paper. And then I would go to a Jupyter Notebook and I would use PyTorch to differentiate the same equation automatically and make sure that I got the same numbers. So it was a really quick way to check that what I was doing by hand was right. And at that time, I didn't know PyTorch very well, so it was also an entry point for me to learn about PyTorch.
- Jon Krohn: 00:11:08 So, I don't know, there's kind of an example of how I think this project-based learning that you're describing for learning any topic, including math and stats is a really great idea. And then we can't forget about your point about commercial and product acumen. This is so, so important to real world data scientists. The only place you can get away without that third point that you mentioned is in academia. And even then, you need to be usually coming up with grants that are relevant to something, which is kind of a product acumen in a way. So cool. I love that. All right. Another really popular video of yours,

Tina, is on your study system. So even beyond data science topics, for people who want to be able to learn anything effectively, what is your study system?

- Tina Huang: 00:12:04 Yeah. So let me preface this again by saying kind of my personality. I'm very easily distractable. I give up very easily. I'm being completely serious. I procrastinate like nobody's business. I think even in university, I would be that person who has done nothing the entire semester and then before an assignment or midterm, freak out three days before and try to get something done. So that was Tina prior, right? And that's my natural state. So in order to control my issues, I come up with a system that I have heard has helped other people as well. So the way it works is first of all, have a goal. If you don't know what you're doing, when people go, "I don't want to learn programming." That's not a very good reason. It's like, what do you actually want to make from a very tangible perspective? And then break that down into specific sub goals, so you're able to work through each of these. And of course these will change over time because you don't have that much knowledge, but make an attempt.
- Tina Huang: 00:13:05 So if you break that down and you work towards that goal, so way more motivating than just randomly learning things and hoping that at some point you master the Python, which by way is never going to happen. And then the second one is more related to how to actually get yourself to do it. I have this habit of getting really excited, making a plan and then doing nothing afterwards and then feeling really bad about myself with my 50 Udemy courses that I bought. So then I never go on Udemy again. But what you do is find what pushes your buttons. For me, I'm a huge people pleaser. So if I tell someone I'm going to do something, I'm so scared that I'm going to disappoint them, that I'll make myself do it. And that way that I do this is I actually have a study live stream, which I have four times a week. And I totally hate myself every

time because I don't want to do it, but then I think about my people pleasing and I have to please these people by showing up. So I show up. And then they're going to scold me. They always scold me if I don't do this. I'm just very afraid. So that's how I get myself to actually do this.

- Tina Huang: 00:14:10 The other one is mindset. Consistency is a key. Don't really go yourself in terms of how good you are at something or how good you're becoming, go yourself on actually just showing up and doing it. That should be your lowest bar. And then after that, it's having a support system, which again, for me, livestream works out really well. Support system, they're kind of there. They're here, very supportive as well. And then finally, it's about progress. You need a way of tracking progress because how are you supposed to know how far along you've come? And also, it stops you from pretending that you've made progress, which again, I love doing that. But then if the hard numbers are there, it's hard for you to pretend that you did it. And that's how you can see your progress towards your goal. It's also very motivating because sometimes, you just feel like you're stuck on something for a very long time, but then looking back at your progress, you actually realize you made a lot of progress. You were just, as they say... What was it called? Mistaking the forest or... No, mistaking... How does that go?
- Jon Krohn: 00:15:10 Yes. That you don't see the forest for the trees.
- Tina Huang: 00:15:17 Is that how it goes? Yes.
- Jon Krohn: 00:15:17 Yes. So you get too focused on the immediate and you lose sight of the big picture.
- Tina Huang: 00:15:21 Exactly. Yes. I was trying to be poet there. So thank you, Jon.

- Jon Krohn: 00:15:28 99% of machine learning teams are doing awesome things at a reasonable scale with say, about four people and two production machine learning models. But, most of the industry best practices that we hear about are from a small handful of companies, operating models at hyperscale. The folks over at neptune.ai care about the 99%. And so they are changing the status quo by sharing insights, tool stacks, and real life stories from practitioners doing ML and MLOps at a reasonable scale. Neptune have even built a flexible tool for experiment tracking and model registry that will fit your workflow at no scale, reasonable scale and beyond. To learn more, check them out at neptune.ai. That's neptune.ai.
- Jon Krohn: 00:16:16 I love that. So you've got to define the goal tangibly and then break it up into surmountable sub goals. Then you commit. You find a way to commit yourself into it. So commit to somebody else or a group in order to kind of force your hand. If you're a people pleaser like you are, and like I am, I actually... I don't know if I've ever talked about this on air, but any project that I ever take on, I have collaborators. I never work on my own. And it's the same thing. It's just, it's so lonely and I can't get motivated if there's not somebody waiting for a result or ideally that can actually be doing it alongside me simultaneously. It makes it just so much more enjoyable. So I think it's brilliant how you do your online study sessions, where people are going to show up, you've agreed to a time, and even though it's virtual, there is a sense of community and people can ask questions and you probably catch up with people regularly. So I love that. And then your point that consistency is key is of course important for maintaining any habit. Consistency is indeed key. Nice.
- Jon Krohn: 00:17:31 So speaking of consistency, those points that you just made are wrapped up nicely in a very popular YouTube video of yours on the five steps to consistently do

anything. And one of the things that I love about that video is all of the anime that you managed to work in. That's very funny and silly. I don't know a lot about anime, but it seems that you do. And then the other thing that I really liked about it is you gave lots of specific examples of how you can have a leaderboard. So in that final piece that you talked about, about being able to visualize progress and avoid losing sight of the forest for the trees, by setting up your own custom way, your own custom metric of tracking your progress, you can then see that and have your own personal leaderboard for tracking your progress. So I definitely recommend checking that video out, both for the humor of all the anime and the consistent examples about that leaderboard. I don't know if you have anything else to add on that leaderboard or anything else about that video, Tina?

- | | | |
|-------------|----------|--|
| Tina Huang: | 00:18:36 | Yeah. I mean, the question is, did I make the video so I can talk about anime specifically? Maybe. |
| Jon Krohn: | 00:18:41 | That wasn't a question. Yeah. |
| Tina Huang: | 00:18:43 | Maybe. Maybe I just wanted to talk about anime all along. I'm like, "Just got to put in some data science here." I'm kidding. Yeah. I actually have a video dedicated specifically to that scoreboard. But just briefly, we actually review each person's scoreboard in that system during the end of the live streams. So you're keeping- |
| Jon Krohn: | 00:19:04 | Oh. |
| Tina Huang: | 00:19:04 | Yeah. |
| Jon Krohn: | 00:19:07 | It is just against yourself. |
| Tina Huang: | 00:19:08 | It is, but other people see it as well, so that helps. Again, like I said, I build in all these checklists and things |

because I have issues with getting myself to do things, but it also happened to work really well for a lot of other people. So we review the scoreboards after every single live stream and we kind of look at progress. It helps a lot because sometimes, you just don't notice certain aspects, maybe over time, you're like, "Oh, this is not important anymore." Or over time you realize that you are way too ambitious about something and you need to scope down. So kind of reminding ourselves, okay, it's about consistency and it has to be realistic.

Jon Krohn: 00:19:46 Nice. Great tips. All right. So going beyond your YouTube videos to what you do for a living, because believe it or not, despite your massive following and your consistency, producing videos that are very popular, that actually isn't your day job. So you work as a data scientist at a very well known big tech company, but that specific company name remains anonymous so that you can publish content more freely. But despite the name of said big tech company remaining anonymous, I would love to hear about your experience working at a company like this. So I imagine there are lots of data scientists, lots of software developers. When you're a data scientist at one of these big companies, what is the day to day like? What does a data scientist do?

Tina Huang: 00:20:34 So good question. So again, there's a lot of different types of data scientists. I think the larger the company gets, the more specialized people get. And also, it's very dependent on the team that you work on as well, because there's different teams that work on different problems and different products. For me specifically, I work from... I actually started working on more product and then it became more change teams, became more systems-oriented. And then from a day-to-day perspective, my day is very variable. I like to describe my job, as I'm pretty much a jack-of-all-trades, because data science is built upon-



- Jon Krohn: 00:21:12 Nice.
- Tina Huang: 00:21:12 Yeah. It's like, if we need data pipelines that are being built, sometimes I would jump into that. I think it's just because data science in itself is dependent on a lot of different factors. And if you don't have the base that's been laid there, then you can't really do data science. So sometimes I do jump and do that kind of work when there's people missing from a team. But a lot of my job is about first of all, understanding what that problem is, which is why I talk a lot about understanding what the business or what it is that you're actually supposed to be doing. And then doing a lot of iterative analyses. I also work mostly on exploratory machine learning models as opposed to deploying them into production. And a lot of this is also cross-functional talk. It's a lot of different product managers, software engineers, and then you are in a way leading the direction of the team by using data as a way of doing so.
- Tina Huang: 00:22:13 Another thing I wanted to point out is the fact that a data scientist job ultimately, if you have to think about it from a purpose, as opposed to so much exactly what it is that you're doing, your whole idea is that you're trying to use data to solve issues and to automate certain things. And if you keep that in mind, everything else makes a lot more sense. It doesn't sound as... I guess it is diverse, but it doesn't sound as random as you may think it is
- Jon Krohn: 00:22:42 Nice. So in that sense, we're kind of going back to your point about how critical it is as a data scientist to have a commercial or product acumen. So the starting point, I guess, for a lot of the work that you do as a data scientist at a big tech company is thinking about what could be better here? What could be more efficient? What could be more profitable? Where could we have a better experience for our users? So any of those kinds of questions. And so then I guess, in a company that large, you probably end

up specializing in the particular kinds of data that you're working with, or do you end up being able to kind of grab different kinds of data, solve different kinds of problems from all over the organization?

- Tina Huang: 00:23:34 It's actually more of the latter. I think we're-
- Jon Krohn: 00:23:37 Wow. No kidding.
- Tina Huang: 00:23:38 Yeah. Working in a bigger tech company, we're really lucky to have a lot of data being centralized. It's in a specific format. Obviously the nature of the data may be different, but you can have access to a lot of different data provided. Of course, we do have integrity checkpoint so you can't just randomly pull data that's not relevant. But you do get a lot of access to this.
- Jon Krohn: 00:24:01 All of your ex-partners private messages.
- Tina Huang: 00:24:06 Exactly. I think that would probably get you fired. Yeah. And then it's actually much more problem-based. So I think that's a really cool thing about working in a large company like this is that you get to think a lot about problems and you get to think a lot about things that may be problems that the company doesn't even know. And then you can take that as a project and use the data that is available in order for you to solve that.
- Jon Krohn: 00:24:33 Cool. I love that. So I would've thought that you'd end up being hyper specialized, that you'd be working on the same kinds of problems day in, day out, and maybe on the same part of the problem. So using some specific kinds of techniques or approaches over and over again, but it sounds like, not only do you get to solve all kinds of different problems, but on top of that, you described yourself as a jack-of-all-trades. And with the examples that you gave, you're not just working on some specific part of data science, you're going end to end in a lot of

situations, it sounds like, at least for the exploratory data part. So engineering data pipelines in, and then doing the exploratory data analysis, but it does sound like you often then end up passing the model off to somebody else for the machine learning engineering.

- Tina Huang: 00:25:30 Yes. That's another really good point. Again, the advantage of working in a really large company is that you get to specialize in the things that you're most interested in. I'll again, caveat to this, is that you can be specialized in a certain type of data or [inaudible 00:25:47] kind of problem. That's totally fine for me. Again, short attention span. So I like focusing on different aspects, but using different techniques in order to do that. But going back to that point. Yes. So oftentimes, I would do exploratory things, like, "Hey, this is something that we can consider." And then I would pass that along to software engineers, specifically machine learning engineers, and they're generally the ones that actually deploy and make sure that it works and just that entire system is functional.
- Jon Krohn: 00:26:18 Nice. So, now you actually have a Master's Degree in Computer and Information Technology from the prestigious University of Pennsylvania. And so you do have some computer science background. How is that helpful for you as a data scientist, even though right now you aren't involved as often with the machine learning engineering yourself?
- Tina Huang: 00:26:42 Yeah. So it was a computer science master's degree at that time. I just pretty much didn't know what to do with myself. So I figured, if I get a computer science master's degree, then probably do whatever, which was actually a good choice. I'm like, "Good job. Passed it."
- Jon Krohn: 00:26:56 It is a really good choice. I say this on episodes all the time, but if you want to be so super employable and have

your pick of whatever kinds of problems and industries you want to be working in, even what kind of work schedule you want to have, specialize in computer science. And even for being a data scientist, which I'm sure Tina, you can elaborate on more, so many things that you do as a data scientist can be made easier by that computer science background.

- Tina Huang: 00:27:26 Yep. Exactly. So computer science actually helps me out a lot. Don't tell anyone, I didn't actually get any official training in data science. Most of it is self-taught. But yeah. Having the CS background is super helpful because as you were saying earlier, Jon, you would sometimes be working through a math problem and then you would kind of just test it out. Right? So for me, CS perspective, a lot of machine learning nowadays is quite automated. So you can actually just call it and play around with it and build stuff with it, even though you don't really know what's happening under the hood. So that's been really helpful in terms me learning specific things. So I would start implementing things and then I would start building out smaller versions of it and learning very quickly. Also, the CS degree, it's very problem solving-based because I think what a lot of computer science is, if you want to specifically say software engineering, for example, which is technically what most of my peers ended up doing. It's like, you have a problem and you have no idea how to do it and then you just kind of figure it out and then you get cryptic things and then you fix it.
- Tina Huang: 00:28:36 So what you learn is a lot of patience and that has been very helpful instead of just like rage quitting. So I would attribute that to my computer science degree because I was definitely a lot less patient before, but I kind of... What is that called? How should I say it? Got my ego beaten up from doing a CS masters. So now I'm just like, "Oh, I have no idea what's happening, but that's fine."

- Jon Krohn: 00:29:01 Nice. It's great to hear that. That's such a candid way of describing what it's like to be a software developer or a data scientist most of the time, that we don't really understand what the problems are and you're constantly, every day as a data scientist or a software developer, you're getting problems thrown at you where you don't understand, maybe even the question that's being asked of you by a person or the question that's being presented to you by some data or some error. And then through Stack Overflow and through time and patience and by figuring out how to make your way through the stack, to where the problem is. There's a lot of working forwards, working backwards. You start with this big nebulous problem and then you just try to chip away at it, chip away at it until you figure out, "Okay, it's got to be this line of code. This is where the mistake must be." And then often, that can be the most frustrating part.
- Jon Krohn: 00:30:02 So I have dozens of times in my career, I've gotten to a point where I'm like, "Okay, I'm 100% sure the problem must be this line of code, but this line of code is perfect." And so somehow the computer is just not doing the command. It's doing the opposite of what it should be doing somehow, like there's a flip bit that just happened randomly somewhere, but then somehow like an hour later, you're like, "Oh. Oh, how did I miss that? Yeah, that is definitely a problem in this line of code."
- Jon Krohn: 00:30:45 So-
- Tina Huang: 00:30:45 Another-
- Jon Krohn: 00:30:46 Yeah.
- Tina Huang: 00:30:46 Sorry. Another thing I learned from computer science is that if you don't know what's happen, just turn it on and off. If you just turn it on and off, maybe it'll work. And that actually helps a lot of times.



- Jon Krohn: 00:30:59 Yep. The old off and on. Definitely tried and true computer science trick. That must be maybe a first-year grad student course at University of Pennsylvania, the old off and on 501.
- Tina Huang: 00:31:15 Yeah, pretty much. That's like every single course, especially when you're dealing with hardware and you don't know what's happening, you just literally just turn it on and off repeatedly until it works.
- Jon Krohn: 00:31:26 Great.
- Jon Krohn: 00:31:27 Einblick Is a faster and more collaborative way to explore your data and build models. It was developed at MIT and showed to reduce time to insight by 30 to 50%. Einblick is based on a novel progressive engine. So no matter the data size, your analysis won't slow down. And, Einblick's novel interface supports the seamless combination of no code operations with Python code. This makes Einblick the go-to data science platform for the entire organization. Sign up for free today at einblick.ai. That's E-I-N-B-L-I-C-K.ai. E-I-N-B-L-I-C-K.ai.
- Jon Krohn: 00:32:11 All right. So in addition to your background, your formal education in computer science, Tina, you also worked as an intern at Goldman Sachs before getting your big tech job. So how does the culture differ from a kind of traditional banking culture, like Goldman Sachs, relative to what you've experienced at big tech?
- Tina Huang: 00:32:35 Yeah, that's a really great question. So I think at places like Goldman Sachs, to give them credit, they are trying to innovate. And Goldman Sachs is probably one of the finance companies that are the most innovative and they believe in engineering and they believe in data. However, it's also an institution that's been around so long and it's very difficult to fully change the culture as well as just the infrastructure inside. Right?



- Jon Krohn: 00:33:04 Right.
- Tina Huang: 00:33:05 So one of the big things that still exists, even though they're trying to get rid of, is this difference between front office and back office. Engineering, data science, we're considered back office. People who are investment banking, traders, they're considered front office, roughly, I believe. Don't quote me on that one.
- Jon Krohn: 00:33:27 They are. I can confirm.
- Tina Huang: 00:33:28 Oh, okay. I was going to say, "Hashtag, not financial advice." So there is that difference between just the way that you're being treated at that point and the amount of resources that you do get. And not going to lie, there's also prestige differences. So people view you differently, especially within the organization, not so much outside. It's more like, "Oh, you're in engineering," or, "You're doing data stuff," then they don't take you as seriously. So that's one of the components. And so another one is just the... How do I say? The technology that's there. It's obviously not cutting edge. There's a lot of investment that goes into it, but it's really not cutting edge the way that tech is. They're finance, you don't just go, "Oh, let me innovate on this thing." So there's a lot of regulations that are there.
- Jon Krohn: 00:34:24 Yeah. Compliance issues around any change.
- Tina Huang: 00:34:28 Exactly. So there's also that. And in terms of, I would also say the caliber of people, specifically in engineering and, I guess, I kind of group them together. Us folks in the back office, I suppose. I hope that doesn't offend anyone. Basically engineering people, people in data science and such, data engineering. I'm not saying that they're not high caliber people, but generally, they are less motivated because a lot of the times, people are telling you what to do. As opposed to the tech companies, I think they make

a really big effort in giving you the power of doing what you think is right and what you think is important. And engineering of course, and data science, even if you want to go off a totem pole, which I think there's a lot less of, there's a lot more... People will listen to you if you think certain things. Leadership, business, business leaders, they will actually listen to you, if you can show them the data, as opposed to somewhere like Goldman Sachs, where a lot of it is still decided based upon seniority and the position that you're in.

- Jon Krohn: 00:35:43 Yeah. So I think my understanding from those kinds of big banks is that in order to have a big impact, you often have to develop a personal relationship with somebody higher up, and then you could potentially be feeding them these great ideas and great data. And then hopefully, that person is giving you some credit, but sometimes they don't, or so I hear. So, super cool. All right. So that gives us a little bit more on your background. Another cool thing about your background is that your undergrad from the University of Toronto is in pharmacology. So the study of pharmaceutical substances and how they act. So super cool. A biology specialization. And right after that, you worked as a research assistant at the Ontario Institute for Cancer Research. So does any of that, your background in pharmacology or cancer research, does that still make an impact on what you're doing today as a data scientist at a big tech company?
- Tina Huang: 00:36:51 Yeah. I feel like that it'd been kind of fleeting around for a while, huh? It's first biology and then I'm like, "Pre-med, I didn't work out. So let's do some bioinformatics. Let's go into finance for a bit. And now we're in tech."
- Jon Krohn: 00:37:04 Well, I believe, I did the same thing.
- Tina Huang: 00:37:08 Really? I actually didn't know that.

- Jon Krohn: 00:37:09 Yeah. So during my undergrad, my intention was to go to medical school and then I got sidetracked because I got invited to do a PhD. So I was like, "Ah, that'd be a good idea. I could still do an MD afterward, be an MD-PhD." And that was all in bioinformatics. So I have a PhD in neuroscience, but so I specialized in creating algorithms that could be applied to very large data sets, like brain imaging data sets, or genomic data, and identifying patterns, causal patterns in some cases. And then I worked in finance as a trader.
- Tina Huang: 00:37:52 Oh wow. So dude, the parallels are actually pretty crazy.
- Jon Krohn: 00:37:56 Yeah.
- Tina Huang: 00:37:59 Would you consider yourself working in tech now?
- Jon Krohn: 00:38:01 I'd definitely. Yeah. So I've been working in tech for eight years, not big tech, small tech, at a small startup. But yeah. Definitely in tech.
- Tina Huang: 00:38:13 Wow. I guess there must be some sort of parallel and personality then.
- Jon Krohn: 00:38:19 Yeah. Well, actually, I could tell you my thinking along the whole way. So I wanted to make a big impact and really help a lot of people and I enjoy working with people, and so I thought medicine would be a really good fit. But then by the time I finished my PhD, I talked to enough people who were friends of mine in my undergrad who had actually become doctors. And it started to sound like it wasn't as great on the ground, as it seemed like as an idea, that people are very difficult to change. So a lot of the patients that you see have a lot of comorbidities, smoking, alcoholism, obesity. So a lot of patients, they have all of these comorbidities and it's just saying to them, "Well, you got to exercise a bit more. You've got to cut out 100 calories a day for a few months." It's very

difficult, it's rare to find patients apparently, that actually listen to that advice and change. And so you end up having the same people come back, who aren't taking care of themselves. Apparently that can be pretty frustrating. And also for me, it doesn't scale up very well.

Tina Huang: 00:39:34 Yes.

Jon Krohn: 00:39:35 So you can only work on one patient at a time. Unlike when you develop a platform or a product where it does scale really nicely, especially with software. So that was something. So for me, and then, so by the time I finished the PhD, I was like, "Well..." You see all these academics, postdocs struggling. I was like, "I don't want to be one of them. I'll work in finance. They make lots of money." And then after doing that for a couple years, I was like, "Ugh." Making lots of money for its own sake, that, I don't know. I couldn't stay motivated about it for very long. And then so data science was this perfect for me, and maybe you feel the same way. This is really your episode. So I'm talking way too much. But for me, data science is this beautiful career path because you get to make a huge impact, lots of data that we can then use to improve products and services, automate things. You can make a massive impact. The problems that you solve are constantly changing. They're intellectually stimulating. And so, yeah. It's kind of the best of everything. You get to make a commercial impact in the end because you're solving interesting problems on a big scale. And so I absolutely love data science and I've really felt like I've been in the right place now since making that transition. How about you?

Tina Huang: 00:41:06 I mean, you pretty much summarize a lot of my thinking as well. In addition to the pre-med part, one of the reasons is I figured my personality wouldn't work so well in medicine unless I was a surgeon because I'm very much to like "Yay. Let's..." I'm a very excitable person. I

would imagine that you don't want your doctor to be very excitable. You would like them to be solemn. And that's definitely not what I'm very good at. [inaudible 00:41:33] people just saying, "Yay, you're here again." That's probably not the best. I guess in surgery you can do whatever you want because they're under anesthesia. So that was one of the reasons. In addition to that, what you're saying about the impact, I really do see that because technically there is a lot of impact in medicine, but it's really hard to do that as a doctor because you're working 72-hour shifts. I don't think you have much time to think about anything else.

Jon Krohn: 00:42:03 So there's a great organization called 80,000 Hours that tries to help you figure out how to make a big impact in your career, but also, to find something that you're good at and that you're passionate about. So try to kind of help you figure out what the optimal career choice for you is. And so we actually had Ben, Ben Todd, who is one of the founders of 80,000 Hours. And by the way, that name 80,000 Hours comes from the average number of hours that you work in your career. And so Ben Todd was on the program in episode 497. It was one of the most popular episodes of 2021. And so he is not a specialist in data science in particular, but he did a lot of research for his episode appearance. And so in addition to general career guidance, he also had a lot of specific career guidance.

Jon Krohn: 00:42:55 And so through that episode or through the 80,000 Hours blog, you can read lots about lots of different kinds of careers, making career decisions. And one of the interesting things that they have said in one of their blog posts on medicine specifically, is that in terms of your net impact, if you decide to become a doctor, your net impact on the world is on average zero. Because if you didn't get into med school and become a doctor, somebody else would have. So it doesn't actually, on average move the

needle very much. Yes, some doctors could be more impactful than others and maybe you could really make some big difference, but on average, you're not going to. It's going to net out to zero. So anyways.

- Tina Huang: 00:43:41 Kind of going back to your original question, I'm sorry. I totally sidetracked here. It's like how much-
- Jon Krohn: 00:43:43 No, I sidetracked.
- Tina Huang: 00:43:48 ... How much does that degree in pharmacology actually help out now? I think it may be surprising to say that it helps out just as much as computer science does. I think it's very interesting. Because I certainly wasn't like, "I think I'm going to be pre-med," and that I'm going to go do a computer science degree so I can become a good data scientist. That didn't happen, but it somehow really worked out. And that's because I always say, data science is oftentimes really about exploring things, understanding things through a data perspective. And in a way, that's what science is as well. It's about, you have a problem, you're trying to figure it out using a scientific method and you have the data and you try to draw conclusions about it. It just so happens that for data science, you have more data and you have different techniques of doing so. So maybe instead of doing surveys, maybe instead of just analyzing small data sets, you just happen to have bigger data sets. And instead of analyzing things, using certain software, you just code your software and you use different modules to do that. But I think the essence of data science is actually very similar. I would say it's actually the same as the scientific method. Also the stats kind of help a lot.
- Jon Krohn: 00:45:06 Yeah, definitely. The stats that you learned in any quantitative science degree, along with that kind of critical thinking you're describing, the scientific method in a lot of ways, that is what data science is, add in some

more software development usually. But yeah. It's a big component. Super cool. I love that answer. So speaking of software and using software to solve problems, what are the software languages that you use regularly now that you're in big tech?

- Tina Huang: 00:45:44 I think I'm going to spark some anger here. I actually use both Python and R, wow.
- Jon Krohn: 00:45:52 Very good. Just sitting on the fence. I also use both.
- Tina Huang: 00:45:59 Hey, because there are actually very specialized for different things. Python's a great general language to do stuff in, much better for machine learning as well, and then you can integrate that lot. But I would say in terms of stats and visualizations, R is better than that. And at the base of all of this, I use a lot of SQL. Because oftentimes, if you can solve a problem using SQL, it's the easiest and the fastest way of doing so. So that's always my first approach.
- Jon Krohn: 00:46:27 Nice. Yeah. That makes a lot of sense. I love that. Yeah. I mean Python and R are different tools and the way that you describe them, Python being a general purpose programming language, totally agree. Super helpful for any machine learning model that's going to end up being engineered in production. If you know that that's going to be the case in advance, maybe you should be starting with Python right from the scratch. However, for working with, maybe not the biggest data sets out there, but if you're going to be working with data that you can fit on a single machine, then doing statistical analysis with R and data visualization. It's so intuitive to me. I mean it was designed by statistics people, as opposed to software developers. And so, a lot of the way that it looks and feels, including being one indexed instead of zero indexed reflects that. So I love that perspective. And then I also love your perspective on how SQL, if you can just use that

is going to be the quickest way to pull your data out and do some analyses with it. So speaking of SQL, you have a SQL course on 365 Data Science. So you have your SQL Sundays series of YouTube videos on YouTube, which are freely available and then you also have a SQL course. So how's that different from what you have on YouTube? Walk us through what this curriculum is in 365 Data Science.

- Tina Huang: 00:47:55 Yeah, for sure. So it really all started off with the SQL Sundays and that's when I take a interview question and I walk through how to do it. So I have a specific system I always use when doing interviews, especially SQL interviews. So I always walk through that process, or else I get really nervous and I forget everything. So I think that's really helpful for people who are looking into doing interviews and how to approach different questions. In terms of my course, I actually made it so that it's 10 complete interviews with an interviewer, start to finish, and it's-
- Jon Krohn: 00:48:32 Oh. Cool.
- Tina Huang: 00:48:34 Yeah. So it's more expanded upon that. Because one of the challenges I faced when I was trying to study for interviews is that there really wasn't a resource that shows you what the entire process looks like. And in my opinion, grinding a lot of questions is not the best approach of doing well in these interviews. It's really understanding what the person, what the interviewer is looking for. And oftentimes after you do a problem, even before you do a problem, right? They will deliberately ask you things that are rather vague. And the expectation is for you to clarify these things. Because on the job, that's exactly what people are going to do, ask you very vague things and you need to be able to clarify it. And then after you do the problem, there's going to be follow-up questions that are going to be related to, okay, how do we

use this and what about different perspectives? And that's not something that you will get experiencing if you're just grinding problems when you're just practicing on certain platforms with interview questions.

- Jon Krohn: 00:49:38 Nice. Really cool. Sounds like a great course if I'm ever interviewing for a big tech company or maybe even just to brush up on my SQL, that sounds like a great resource. I particularly like how you highlight there, how interview questions, the interview process with a human, as opposed to when you're just getting very structured questions that are written down, that human is testing in a lot of ways, your ability to distill what the question really is and to refine down from some originally vague parameters. So very different that human experience relative to getting... If you're just doing SQL questions, because they're not interactive, if you're doing ones that are written down doing that kind of road practice, it's going to be relatively straightforward. So the answer might not be easy to figure out, but the way that the question is presented is typically going to be easy to understand. Cool. All right. So before we move on to audience questions, I've noticed from conversations that I've had with you in the past, as well as a lot of your videos, that you seem to be standing a lot of the time. So is that a productivity hack or a life hack of some kind? What other tricks do you have other than the study tips, the tricks for consistently doing anything, tricks for being a great data scientist that you covered so far? Do you have any other general productivity life hacks, like standing?
- Tina Huang: 00:51:09 I mean, the standing one just comes from the fact that my back hurts a lot.
- Jon Krohn: 00:51:12 Okay.

Tina Huang: 00:51:15 I would be honest with you. That's totally what happened. I'm sitting here because I figured, it's probably better if I not just stand there. It's a bit awkward. But yeah. That's just because my back hurts. I got a standing desk and then I hide my chair usually so I don't feel like sitting on my chair. Also, just my back hurts so much. I have a variety of different massage [inaudible 00:51:35] and different things to make my back feel better. It's probably during my undergrad days, when you just have terrible posture and sit there memorizing drugs in the middle of the night, at least that was painless for me.

Jon Krohn: 00:51:50 Because of your pharmacology degree or because of extracurricular studies?

Tina Huang: 00:51:52 Pharmacology, of course. Why would you even ask?

Jon Krohn: 00:51:57 Memorizing a bunch of drugs.

Tina Huang: 00:52:00 Yeah. There's a-

Jon Krohn: 00:52:00 Well, okay.

Tina Huang: 00:52:01 Oh, sorry. There was another one that I do have. I mean, audio folks, I apologize. You wouldn't be able to see, but I'd say physical timer that I use. So in terms of Pomodoros, I generally use a physical timer because I get very distracted by everything. Again, I just have issues with being distracted by everything. So I just have this physical timer I put here and every time I look at it, I'm like, "Oh wow, I should be studying or doing whatever." So I do that. I generally just try to make everything very, very physical and try to not use my phone. And another thing that I do, unfortunately also got some tea stains on it, but this is what I have over here. It's a pack of cards. Haven't come up with a better name. What it has on the front is memento mori. And this is very motivating to me. It means, "Remember you die." But it's not as pessimistic.

It means that you should probably live, remembering that at some point, we are mortal people. We are mortal creatures. So you should be doing the thing that-

- Jon Krohn: 00:53:01 Speak for yourself.
- Tina Huang: 00:53:03 Oh yeah. True.
- Jon Krohn: 00:53:04 I don't have to worry about that.
- Tina Huang: 00:53:07 Yeah. You can just upload your consciousness or something like that, right? So in this pack of cards, I have the things that are very important to me. So the first one is learning. The second one is health. The third one is business. And the fourth one is finance. I have this general goal of financial freedom, location freedom, time freedom and what I pretentiously call intellectual freedom, which is my ultimate purpose, where I just want to sit there and do things that I think are the most important without having to worry about things like money. And so that's one of my goals and the family, career, romance here, I have, "Not die alone." So that's kind of my goal, which I remind myself of every day. So I go through these pack of cards and then whenever I feel like I just want to sit there and watch anime, which is every day, I look at this, I'm like, "Oh right. This is very important."
- Jon Krohn: 00:54:00 "If I keep watching this anime, I'm going to die alone."
- Tina Huang: 00:54:05 [crosstalk 00:54:05].
- Jon Krohn: 00:54:06 So that memento mori, I believe you also have a YouTube video on that. I think I've seen that on your channel.
- Tina Huang: 00:54:12 Yes. I do. It's how I motivate myself to do things.
- Jon Krohn: 00:54:16 Nice. All right. So we'll include a link to that in the show notes as well. Really great tip there. And I do love the idea

of having a physical Pomodoro timer. So for those of you who aren't aware of the Pomodoro focus and productivity technique, I did an episode on that, episode 456 of the program. I just use a timer on my phone, but I can definitely see the advantage in using a physical timer because then I could force myself to have the phone out of the room. It would definitely in some situations be helpful on focusing me. Okay. So I posted on LinkedIn and Twitter asking before the episode if people had questions for you and it is one of the most popular posts I've ever made. So thank you for engaging with it. I think your giant LinkedIn audience engaged with my post. So definitely appreciate that. And we had tons of people, there are dozens of people who have written comments saying what a huge impact that you have made on their life and on their data science career. So clearly, you have a really devoted following and you're making a big impact with them.

- | | | |
|-------------|----------|--|
| Tina Huang: | 00:55:36 | That's why I keep doing this. Honestly, it's very difficult starting off, trying to go off YouTube, especially for me, I've never really used social media, but really it's what motivates me because I see that the things I do are having impact. And from a selfish perspective as well, because it forces me to become better. So I can keep creating that impact in helping people. |
| Jon Krohn: | 00:56:05 | Nice. Well, so there were a lot more comments and praise for you than there were specifically questions. So Ken- |
| Tina Huang: | 00:56:15 | It's fake flattery. |
| Jon Krohn: | 00:56:17 | ... Ken Jee here was in episode number 555. He would like to know how many cats and plants you think you will own in 10 years? |
| Tina Huang: | 00:56:26 | I believe I said at least three. |

- Jon Krohn: 00:56:32 You did. And then he wrote, more like 300.
- Tina Huang: 00:56:32 Yeah. Actually, it kind of depends on how many I can fit into this room. And the plants thing, I kind of just... I don't know. One day I was like, "I think I'm going to get plants." And then I got a lot of plants and I definitely went over my spending goals on that month. So I guess I'm not really sure I have a good reason for that. However, I'm not sure if this is interesting at all to anybody. So you see my two cats right now called Beep Beep and Boop. I think my next cat, I'm going to call Poop. I don't know, it just occurs to me. I think that'll be quite good.
- Jon Krohn: 00:57:07 That makes a lot of sense, because that's the third noise that computers make. Beep, beep, boop, and poop.
- Tina Huang: 00:57:16 Pretty much.
- Jon Krohn: 00:57:17 Okay. So we do also have a serious question here from Brenda. So Brenda works in data science and analytics and she would like to know Tina, about a time when you didn't have enough data to make a decision. What did you do? How did you handle that situation?
- Tina Huang: 00:57:33 That's a great question. I would actually say most of the time, you don't have enough data because you don't have a full grasp of everything that's happening. And what it really is you take advantage of the things that you have. You also take advantage of domain knowledge and intuition. Data science is an art as much as it is a science, as well as you try to understand things, you try to gather knowledge from the people around you as well. And that really helps drive a lot of what you can do. And actually interesting enough, one of the teams I was working on previously, there was no data because we were starting from scratch. And what's fascinating is that's really, when you start realizing that it's the scientific method. Even when you don't have any data, it's

about thinking through your problem, having a hypothesis, testing that out, as with whatever it is that you have and then just kind of reiterating throughout this process.

- Tina Huang: 00:58:33 So to answer your question, maybe it's not as specific as it would like to be, but it's just think through the problem and then use whatever resources it is that you have. And understand that even when you have a lot of data, the decision is never absolute. So you just make your best guess. And since you have the hypothesis, you can start testing things out, and maybe you were wrong. But at that point, you would know that you're wrong. You would have the data that you were wrong and then you can keep refining from there.
- Jon Krohn: 00:59:02 Nice. That is a great answer, Tina. Yeah, it wasn't very specific, but in its generalness it was very helpful indeed. All right. So yeah. So that's it. Those are all the questions that came up for you from the audience. So then we are reaching the end of the episode. Do you have a book recommendation for us?
- Tina Huang: 00:59:19 Yeah. So the book that last made me cry was A Million Miles in a Thousand Years. It's a really great book. I don't want to spoil it for you guys, but it'll make you cry a lot. I think, unless you have a heart of steel.
- Jon Krohn: 00:59:37 Oh no. I cry reading almost every book and watching almost every movie. So I'm sure it'll work for me.
- Tina Huang: 00:59:45 Oh, actually you should definitely check it out. It's about this guy who pretty much gets successful, but then he's like, "Well, okay, now what should I do with my life?" And then he kind of degenerates and such. And then he just, goes on a journey, does a bunch of things, learns a lot of life lessons, I suppose. And I don't think I'm going to spoil anything by saying this. A really big thing that he

encompasses just throughout the entire thing is, what makes a good life is also what makes a good story. And a story's pretty boring if you just sit there and do the same thing every day, right? You don't challenge yourself. So that's kind of what it is. If you want a good life is the same as trying to make a good movie. You should definitely read it because he's very good at writing. And yes, I was just crying in the middle of the night.

Jon Krohn:	01:00:35	A Million Years and a Thousand... Oh. A Million Miles-
Tina Huang:	01:00:38	A Million Miles in a Thousand Years.
Jon Krohn:	01:00:41	... and a Thousand Years.
Tina Huang:	01:00:41	Yeah.
Jon Krohn:	01:00:42	Cool. All right. Thank you for that great recommendation. And then obviously, for people to follow you, keep up with your latest, your YouTube channel is your number one recommendation. I'm sure. But in addition to that, where else should people follow you? LinkedIn? Seems like you have a lot of followers there.
Tina Huang:	01:01:01	Yeah. I'm trying to get better at LinkedIn. Like I said, I actually don't personally use any social media. So you can't really find that much to embarrass me with, as people have tried personally. But LinkedIn, I am going to try my best to be more active there, but I'm also going to start a newsletter pretty soon. So a lot of that, start posting more, just trying to understand what people are interested in on LinkedIn. Also, I have a Discord, which you can find a link. I think in the... I'm not sure if the podcast, is it called the description?
Jon Krohn:	01:01:31	Yeah. In the show notes.

- Tina Huang: 01:01:33 In the show notes, yes. And maybe we can put that there, but it's also link to my YouTube. It's a very active Discord community. People study a lot. Yeah. I mean, I troll around once in a while, but it's actually a very community run. Pretty much, the community runs itself. It's like, I'm not the one dictating what's happening or anything like that. But if you comment and talk to me there, I am very likely to respond.
- Jon Krohn: 01:01:58 Nice. Thank you. Yeah. So YouTube, LinkedIn, Discord. Wonderful. Lots of ways to keep up with Tina. Thank you so much for being on the show. It's been so much fun. You are such an engaging communicator, so bubbly and have tons of knowledge to share with us. Thank you so much for joining and hopefully it won't be too long before you're on the show again.
- Tina Huang: 01:02:24 Thank you.
- Jon Krohn: 01:02:25 Who says you can't learn and have fun at the same time? We certainly did today. In today's episode, Tina filled us in on how coding, math and commercial acumen are the three key areas to focus on if you're getting started in data science. Her five steps to consistently doing anything, namely defining your goals tangibly, breaking those into surmountable sub goals, committing to the goal to someone else, having a support system, and being consistent. She also talked about how data scientists at big tech firms focus on solving issues and creating value more so than engineering their models into production systems. Shocked about how she regularly uses Python for general purpose programming, R for statistics and visualization, and SQL for database queries. She also talked about how her SQL course practically prepares you for data science interviews and how her scientific background helps her draw conclusions as a data scientist, while her computer science background has provided her with much patience for problem solving.



- Jon Krohn: 01:03:26 As always, you can get all the show notes, including the transcript for this episode, the video recording, any materials mentioned on the show, the URLs for Tina's YouTube channel and social media profiles, as well as my own social media profiles at superdatascience.com/563. That's superdatascience.com/563. If you'd like to ask questions of future guests of the show like several audience members did of Tina during today's episode, then consider following me on LinkedIn or Twitter, as that's where I post who upcoming guests are and ask you to provide your inquiries.
- Jon Krohn: 01:04:00 Thanks to my company Nebula, for supporting me while I create content for you. And thanks of course, to Ivana Zibert, Mario Pombo, Serg Masis, Sylvia Ogweng and Kirill Eremenko on the SuperDataScience team for managing, editing, researching, summarizing, and producing another super fun episode for us today. Keep on rocking it out there, folks. And I'm looking forward to enjoying another round of the SuperDataScience podcast with you very soon.