

SDS PODCAST

EPISODE 821:

THE SKILLS YOU NEED TO BE AN EFFECTIVE DATA SCIENTIST, WITH MARCK VAISMAN



- Jon Krohn: 00:00:00 This is episode number 821 with Marck Vaisman, Senior Cloud Solutions Architect at Microsoft.
- 00:00:05 Today's episode is brought to you by Gurobi, the decision intelligence leader.
- 00:00:12 Welcome to the Super Data Science 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.
- 00:00:48 Welcome back to the Super Data Science Podcast. Today we've got the exceptionally experienced and insightful Marck Vaisman on the show. Marck has been at Microsoft for seven years; for more than five of those years now, he's been a Senior Cloud Solutions Architect for them, specializing in data, data science, and AI ML. For nearly a decade, he's been an adjunct professor at both Georgetown University and George Washington University, teaching graduate-level courses on math, stats, analytics, and decision sciences. He co-founded a nonprofit in Washington, D.C. that runs both the data science D.C. and statistical programming D.C. meetups. He holds a bachelor's in mechanical engineering from Boston University and an MBA from Vanderbilt.
- 00:01:25 Today's episode will be of interest to anyone who is, who manages, or who aspires to be a data professional. In today's episode, Marck details the skills, competencies, and personas, the data scientists and related professionals, such as analysts, data engineers, ML engineers, and AI engineers, can have. He also fills us in on the academic research on why the term data scientist is such a difficult job title and word to define. Most importantly, today he provides a comprehensive characterization of all the essential skills that every data

professional needs to be effective, as well as the skills that allow you to specialize as a particular subtype of data scientist. And he talks about the implications of all this for both folks hunting for a data job and the companies that are looking to hire them. All right, are you ready for this deeply practical episode? Let's go.

00:02:29 Marck, welcome to the Super Data Science Podcast. I love that you are on the show. We've been talking about doing an episode with you for a while, and now the stars have aligned and you're here.

Marck Vaisman: 00:02:38 Thank you so much for the invitation. I was really, really happy when you reached out.

Jon Krohn: 00:02:41 Yeah, where are you calling in from today, Marck?

Marck Vaisman: 00:02:43 I'm just outside of Washington, D.C., in Maryland.

Jon Krohn: 00:02:47 Nice. And so we met in person for the first time, two years ago, at the New York R Conference. I was doing an interview actually for this show, on stage with Hilary Mason, and so people can check that out. Episode number 589. Hilary Mason iconic-

Marck Vaisman: 00:03:09 Yes.

Jon Krohn: 00:03:09 -person in data science, one of the original data scientists period. And yeah, we did an episode with her live on stage at the New York R Conference.

Marck Vaisman: 00:03:19 New York R Conference.

Jon Krohn: 00:03:20 And yeah, you and I met there in person. I signed a copy of my book for you.

Marck Vaisman: 00:03:25 That's right. I have it right here. I know this is mostly for audio, but for the folks on video, you can actually see me holding the book, and it's actually signed by Jon.



- Jon Krohn: 00:03:33 Nice. Yeah, so there you go. That's how we knew each other, and I think we've run into each other at New York R conferences since as well.
- Marck Vaisman: 00:03:42 Right. I saw you also this year when you were hosting the panel on the meetups.
- Jon Krohn: 00:03:47 10-Year Retrospective Panel.
- Marck Vaisman: 00:03:49 Correct.
- Jon Krohn: 00:03:49 Yeah. Yep. And we do actually have recordings of that as well. The episode numbers are 794 and 790. So there was two panels this year where we had a number of iconic people from the data science community over the years, live on stage, and they answered a couple questions for this podcast as well, but enough about the past and how we met. Let's get into the content for today's episode. So what sparked the topic for this episode today? We have, over the years, we've had various episodes for our listeners on what is a data scientist? What's a data analyst? What's a data engineer? And how to get into these fields, what skills you need. And we haven't done one in a while. So that's part of what, when you reached out and suggested this content, based on a talk that you gave at the Data Council in Austin earlier this year, the talk was called, What Makes for an Effective Data Practitioner in 2024?
- 00:04:53 And I will include a link to that full talk. It's a half hour talk, and I'll also include a link to the slides. And what we are going to do in this episode is we're going to basically go through the content of that talk, and we're going to let our audience, our listeners today know, "What makes an effective data practitioner in 2024?" And a lot of this does center around the data scientist itself.
- Marck Vaisman: 00:05:18 Yes.



- Jon Krohn: 00:05:20 If there's kind of one thing that you're trying to do with this episode, Marck, it seems to me like what you're trying to do is to help disambiguate and clarify and create a common nomenclature around data scientists subcategories and how it relates to data analysts, data engineers, ML engineers, AI engineers, all these different titles.
- Marck Vaisman: 00:05:43 Yes, absolutely. And you and I, and I think every other member of the data science community, we've been talking about this thing called data science for a really long time, and it's like... Yet you ask 10 different people, and you got 20 different answers, right? But more so than that, it is just that I feel, especially as an educator, because one of the things that I do is I've been teaching now for about nine years. Yeah, about nine years. And obviously, some of the content is pretty, I would say static, but there's other types of content that I needed to tweak. And part of me has been thinking, especially a lot, is like, "All right, what should I be doing to make my students be successful?" But more so than that is just this constant overloading of this term, right? We say data science.
- Jon Krohn: 00:06:37 The term data science. Yeah.
- Marck Vaisman: 00:06:38 Yes. The constant overloading of this term. And you can ask many people about whether it's overloaded or not. And you're going to get a million different answers, so.
- Jon Krohn: 00:06:55 So you spoke where you're teaching there, you teach at George Washington University and Georgetown University you are adjunct faculty member at both.
- Marck Vaisman: 00:07:02 Yes.



- Jon Krohn: 00:07:02 Teaching in decision sciences at George Washington U. and analytics, as well as graduate math and stats at Georgetown.
- Marck Vaisman: 00:07:03 Yeah, and more on the analytics side today. The math stats was several years ago.
- Jon Krohn: 00:07:08 Nice. But yes, so lots of experience teaching in and around data science, including things like natural language programming, big data courses, that was probably a few years ago.
- Marck Vaisman: 00:07:19 That's been my bread and butter, pretty much big data and cloud computing, because I've been doing big data and cloud computing since probably 2010, 2011. And data vis, which I think... I love that field in particular, that specific domain on data visualization and data analysis. I mean, that's one of the core skills if you want to communicate... Because it's really about communication, you can make great charts, but you need to be able to communicate and communicate effectively. It's not just about the process of making the charts and making them nice and relevant, but it's also about telling the story.
- Jon Krohn: 00:07:57 Yeah, exactly. I agree with you. And before we get to 2024 and start talking about, What Makes for an Effective Data Practitioner Today? Let's talk about the history of this research that you've been doing. So this stretches back to 2012 when you started working on a little book for O'Reilly called Analyzing the Analyzers, where at that time data science was a really new term. And so it sounds like you focused more on this term data analysis and you broke down. I mean, you covered tons of different skills. I'm going to read these really quickly just to give a sense of, even in 2012, what it involved to be in a kind of data science field at that time. Algorithms, back-end programming, Bayesian stats, big data, business,

classical stats, data manipulation, front-end programming, graphical models, machine learning, math, optimization, product development, science like experimental design, technical writing, publishing, simulations, spatial statistics like geographic stuff, working with structured databases, doing surveys, system administration, like UNIX kind of stuff, database administration, temporal statistics like forecasting, time series analysis, unstructured data like NoSQL, text mining, and the visualization that you were just talking about.

00:09:13 So all of those kinds of skills were already a part of being a data analyst or a data scientist more than a decade ago. In your paper or in your booklet, *Analyzing the Analyzers*, you broke all of those skills down into five skill areas and then used those to define four distinct personas of data professional. Do you want to tell us about that?

Marck Vaisman: 00:09:38 Yeah, and let me give you a little bit of the backstory of that publication. So one of the things that I started getting early on involved was in the data science community, right? With the meetups and that sort of thing. So I've been in the D.C. area now for 18 years, but basically, I befriended a lot of the folks across the country in the early days of Twitter, but then I got really close to the New York community. So yeah, you mentioned Hillary and obviously Drew Conway, Jared, these folks that were also running meetups. So I basically took what they were doing, and I volunteered to start running the meetup here in the D.C. area, started with the R meetup. And then when Harlan Harris, who used to live in New York, he moved to D.C., we started the data science meetup just as a sort of sister meetup, which became the flagship.

00:10:29 But what ended up happening is that people would actually come to us in the meetups and talking about 2012, and they'd be like, "I'm looking for X, for a machine

learning engineer, or I'm looking for a data analyst, or I'm looking for whatever." I mean, we're not recruiters. And it kind of started for us really wanting to know who was coming to the meetups and where they came from, because again, everyone was talking about data science, about what they were doing. But to your point, all of those skills sort of came out of the analysis that you mentioned. So, long story short, is, we said, "Let's do something about this. Let's learn a little bit about our folks." So we created a survey, we published it online, we tweeted it, we got a bit.ly link. I think Hillary was a bit.ly back at the time, so we used bit.ly, and we got about 250 responses worldwide.

00:11:21 Most of them were US, obviously not necessarily a very small sample, and probably very biased because we were kind of talking to our own audience, but nonetheless, we quantified this. And this was the first attempt to really quantify this thing that we were calling data science. And the results that you see is the results of the data, and when you look at them, they make sense. I don't think; maybe there are some slight variations at the time, but for the most part, I think those made a lot of sense. So to your point, yeah, we gathered the data; we actually asked people; we didn't ask them about tools; we asked them about skills. So we didn't ask them about specifically which programming language they were using. We just asked them. And we asked about all those skills that you mentioned about, and then processed the data and did non-negative matrix factorization, and that got to the clusters.

00:12:17 And we got four personas in five skill areas, major skill areas. So the four personas are data business people, data creative, data developer and data researcher. And then the five skill areas are business, machine learning/big data, kind of lumped together, today probably separate, math/operations research, because

there's a lot of operations research. There's a lot of overlap there. Programming, just general programming and then statistics. And you sort of see these four personas, and the data business person. Obviously, their focus is on business. They might know a little bit of programming; they might do a little bit of this, a little bit of that. But their core focus is on skills. There was one called the data; I think it was a data... I don't remember if it was the data creative or the data research. I think the data creative, which was more of a generalist, which sort of had even spread skills across the board.

00:13:11 Then we had a data developer, which was obviously, more maybe like a software person or someone that did a lot more programming. And then there was the researcher, right? Folks that were coming out of perhaps social sciences or traditional research areas that obviously had a lot of, their strength was statistics, not necessarily programming or other things, but just, of course, core statistics. And we used this language to give some context, and it was a successful publication. I think it was the first time that anybody did that, and I wrote this with Harlan Harris and Sean Murphy, who were... Sean was one of the other folks that was part of the local meetup group, organizers. And it was a lot of fun to do, and we presented it. And then O'Reilly was at one of the conferences we presented this at, and they said, "Hey, we'd love to publish this." And they did.

00:14:04 And I've sort of always related back to this thing, and I think the names of the clusters may have changed a little bit; there's probably a couple more personas here today. But a lot of what we did back then, I think, still holds truth. It still holds today.

Jon Krohn: 00:14:22 In a recent episode of this podcast, the mathematical optimization guru, Jerry Yurchisin, joined us to detail how you can leverage mathematical optimization to drive



commercial decision-making, giving you the confidence to deliver provably optimal decisions. This is where Gurobi optimization comes into play. Trusted by most of the world's leading enterprises, Gurobi's cutting-edge optimization solver, lightweight APIs, and flexible deployment simplify the data to decision-journey. And thankfully, if you're new to mathematical optimization approaches, Gurobi offers a wealth of resources for data scientists, including hands-on training, comprehensive Jupyter Notebook examples, and extensive free online courses. Check out episode number 813 of this podcast to learn more about mathematical optimization and all of these great resources from Gurobi. That's episode number 813.

- 00:15:10 Nice, and so, I think a key thing here is that when you talk about these personas, these personas are not what we would call today as a job title. So it's not like these personas are data analyst, data scientist, data engineer, these are... I mean, you could be either a data analyst or a data scientist, and you could embody one of these four personas more or less. So you could be more. Is that right? That you could be kind of more... Yeah.
- Marck Vaisman: 00:15:36 I think so because a lot of, and this is sort of going back to the original question, right, about what is data science? I think data science is whatever the organization wants it to be. In some organizations, a lot of it's just BI, and it is, business intelligence and slicing and dicing data, creating reports, doing visualizations.
- Jon Krohn: 00:15:59 So that what you just defined there, that might be more like a data business person.
- Marck Vaisman: 00:16:04 Maybe? You could be a data business person, but you could also be a data developer working in that type of organization. So it doesn't necessarily map one-to-one.



- Jon Krohn: 00:16:14 But yeah, so you have these clusters of skills, and those then break into particular personas. And there is a chart, which is in the slides or in the original book. So I'll have links to both, that book from 2013, as well as your slides from 2024. Either way, people can see this full-color illustration. It kind of looks like modern art. It looks like-
- Marck Vaisman: 00:16:39 I think Miró or [inaudible 00:16:43]. No yeah.
- Jon Krohn: 00:16:50 But yes, this is really...
- Marck Vaisman: 00:16:52 Pop art, think pop art.
- Jon Krohn: 00:16:53 Pop art. Yeah, exactly. But it breaks down, so you can see things like the persona that's a data business person; they index very highly on the business skill set. The data researcher indexes really highly on the statistics skill set and the math skill set more than the other groups. And the data developer has more so the programming skills and the machine learning or big data skills. So it's a cool... I really like this way of understanding personas of the different kinds of subtypes of data scientists or machine learning engineer or data analysts that you could have out there.
- 00:17:35 And I think something that's actionable already, and we're going to talk about this more later in the episode, but something that's actionable then, when you think about these kinds of different personas, is that it means then, as a hiring manager or as a recruiter or as somebody looking for a role, instead of looking for just any data scientist job or any data analyst job or any data engineering job, you can be thinking about, "Oh, actually, which of these personas do I really need in this role? Or which of these personas best encapsulates me? Or which blend of them best encapsulates me?" I mean, you don't need to classify yourself just into one bucket. You could

be a data creative and a data business person, or whatever.

- Marck Vaisman: 00:18:22 And obviously, these are more obviously trends. I mean, there are clusters, but they give you probably a common language and definition to a certain extent. But by all means, I wouldn't necessarily... Because it's really hard, this is made up by so many different competencies and skill areas, and I know we'll talk about that. If we created a cluster for every possible combination, I mean, we'd have millions of permutations, and that's just not feasible.
- Jon Krohn: 00:18:54 Yeah. I mean, this is exploratory data analysis, summary statistics. The whole point of that is to create a small number of tangible groups for us to work with and discuss. So that's really cool. And then, since 2012, there's been an evolution of data science. So you've got them captured in your presentation here; I'll just kind of really quickly reel through them. But at the time that your survey came out, you had kind of these core data science skills, which actually have really stood the test of time: SQL, R, Python, scikit-learn, pandas. Maybe R is to some extent less essential in a lot of contexts than it was a decade ago. But there's other things like Python, SQL, scikit-learn, pandas, and I mean, R is still, they're... I don't mean to knock it. It's hugely valuable.
- Marck Vaisman: 00:19:42 Well, yeah, you're talking to an R lover here, so... No, I do agree with that. I think R, again, R is a tool, and at the end of the day, it's a means to an end. Personally, I have embraced it and used it for a long time; more so, the tool was the community. But I find the tool just, there are certain things that are very easy to do with R that still holds truth today, and I still use it very much so. But nonetheless...
- Jon Krohn: 00:20:14 Data visualization.



- Marck Vaisman: 00:20:15 Absolutely.
- Jon Krohn: 00:20:16 Even creating presentations?
- Marck Vaisman: 00:20:18 Even just data wrangling. Yeah. Yeah, creating presentations, like my presentation that we're looking at, was created with Porto. That's not an R thing specific, but again, it kind of comes from that ecosystem. But yeah, so yes, I think when data science first started, and one thing that I wanted to mention that I don't think we did was, we actually did the survey because I think we had all been relying on the OG Venn Diagram, which we haven't talked about, right? Drew Conway's famous 2010 Venn Diagram that said, "That data science was a combination of three things," and it was very general, and it's a combination of hacking skills, math stats and knowledge, and then substantive expertise. And when you look at the Venn Diagram, which I also have in my presentation, and it's funny, you may have never seen this. For folks that are listening today, you may not necessarily know what this is, but this was sort of the reference in 2010, 2011.
- 00:21:17 And I think what we did sort of expanded on that. It just gave it a lot more context. But as we look at the evolution, one of the things that I've observed, because... Obviously, a lot of this is from my observation and experience, but one of the things that I've seen is that, as a community, we still call this thing data science, but it's a lot more than those things that we just talked about today. And there have been sort of hot skills or competencies that have been very popular, and I don't want to make this conversation about the hiring side of the equation, but because I think, collectively, we all know that that's broken and there's just a lot of changes...
- Jon Krohn: 00:22:08 We should talk about that, and we'll get to that. I think we're going to have some takeaways at the end of the episode, kind of summarizing all of the points made in

today's episode. To have guidance either for hiring managers, for people organizing processes, for hiring data scientists, or just for people that are looking for data science jobs. I think we will have some good takeaways, but that, we'll get to.

Marck Vaisman: 00:22:29 Sure. But it's just that the term has meant different things over time. And we just keep adding more things to the mix, to the pot. So I just sort of think about this; I just call it the data science soup or the brew, where you're just standing, like in a pot, and you're just throwing things in it, and you just keep adding and adding and adding, and it's still called the same thing.

Jon Krohn: 00:22:54 Yeah, yeah. So that brew kind of captures the hot skills over time, over the past decade. So I was getting into this with a, you started with SQL, R, Python, scikit-learn, pandas. Then you had the big data era from 2010 to 2014, where things like Hadoop, MapReduce, Hive, and Spark became popular things that "all data scientists needed to know."

00:23:17 And then, around 2011, you had the Kaggle era where everyone was crushing Kaggle competitions with XGBoost. So all of a sudden, everyone needs to know those kinds of boosted trees approaches. Then you had data product era; people need to know product management. Deep learning was a huge one starting in 2014 and still now, today, with more specialized models like Transformers, generative AI, large language models. Also, neural nets, convolutional neural networks, generative adversarial networks, we don't hear about this as much anymore, CUDA programming. You had after the deep learning era, we got into the MLOps era, AutoML, models as API, MLflow. You've done a great job summarizing. I'm just reading Marck's work here, so don't credit me as you hear this.

Marck Vaisman: 00:24:03 By all means, this is not; there is variations. I just attempted to illustrate the evolution. And I think that the eras, some of the eras I think are distinct, are definitely distinct. But a lot of these are just; they started, and they just keep going.

Jon Krohn: 00:24:22 I think you did a great job. Yeah. So yeah, what you're saying there is that some of them had a beginning and an end. So the big data stuff: no one talks about Hadoop anymore; you don't need to know Hadoop. XGBoost kind of had a big phase; you don't necessarily need to know XGBoost. Well, it's useful to know. And we've had some great episodes on XGBoost on the show. But yeah, some things like deep learning, MLOps, responsible AI cloud stuff, which you know really well, you do that professionally and you teach on that a lot, that kind of stuff. That's not going to go away because that's something; cloud skills are critical because we're dealing with such large data sets now, such large machine-learning models, that you can't be doing things on your laptop. Well, there are still some things you can do on your laptop, but there's a lot of things that you can't.

00:25:10 And so you need to have these cloud skills. And so yeah, cloud skills, the cloud era, responsible AI since 2018. That's been something: explainable AI, data privacy, data governance, safety, and then, most recently, since ChatGPT, we've had the generative AI era, large language models, LangChain, retrieval, augmented generation, prompt engineering, all that kind of stuff. So I don't know, I realize I'm just kind of reeling through lots of skills, but if you are listening and you're new to data science, I think it's a helpful framing to know how we've gotten to where we are today, where we now have this, as you describe it, this overloaded data practitioner brew, where over the past 10, 15 years, these huge areas evolve that have so many applications and so much detail. And you could build a whole career focused on being an XGBoost expert,



but yet the data scientist is somehow expected in job interviews to know all of these things somehow magically.

- Marck Vaisman: 00:26:16 Job definitions. I mean, you look at the job definitions, and then there's still a laundry list of buzzwords and keywords, right? At the end of the day, many are, not all of them, but many are.
- Jon Krohn: 00:26:31 No, you're right. And you have a really comical cartoon in your slide deck that says, "Great expectations versus reality." When you're an AI engineer in these interviews, you have to do things like be able to define in detail how XGBoost works or how a Transformer architecture works, like layer by layer, what are the neural nets and the connections in that Transformer architecture. But then, when you're in the job, you never need to know that stuff. I mean, very rarely. There's some Google DeepMind roles, some OpenAI roles, some research roles where you might actually be trying to make Transformers better, but the vast majority of people who are using Transformers aren't trying to reinvent the Transformer and aren't going to need to program it on a low level. Instead, you're just going to import Transformers, and it's going to be a one-liner.
- Marck Vaisman: 00:27:32 Or scikit-learn or Tidyverse, or whatever library that does the work for you.
- Jon Krohn: 00:27:38 Exactly. So it's a funny thing. Yeah, the data science interview process is broken for sure, and you've done a great job highlighting it already. I think in this episode, how there's this insane, this impossible amount of stuff that data scientists are expected to keep up with. And at least with things like the Super Data Science podcast, with this show, it allows people to have some kind of high-level overview of all the things going on, so that you're aware... I think that that is important. I don't think that everyone should need to know how, everyone

working in data science or a related field. I absolutely do not think that you need to know how a Transformer architecture works and that you need to be able to program that at a low level. But I think it's useful to know what's possible with Transformers and LLMs, so that when you're confronted with some business problem in your job, you can say, "Oh, I heard in a podcast or in a blog post or paper or Twitter or whatever. I think I have an idea of a way that we can solve this problem."

00:28:47 And then you can look into it in more detail. So, keeping wide breadth on all of these different topic areas, everything in the data brew, I think is probably worthwhile doing because you never know when you're going to encounter a business problem, where knowing that possible solution exists is out there. And then, now we have LLMs, you can really easily, you can learn how these things work. You can apply them to whatever real business problem you have with so little effort.

Marck Vaisman: 00:29:16 Yeah. And interesting on that specifically, since that's something that I'm actually doing professionally a lot of right now, is working with customers who are using our LLMs as a service, because that's one of the many services that we're offering. That really falls under the realm, in my opinion, of application development and not necessarily data science, although there is a data science aspect to using LLMs for a business problem, because you do want to use the data science methodology in the sense of evaluation. You build your prompts; you build a business application. Obviously, the typical, I think the most common use case right now is RAG, right? Retrieval Augmentation Generation, which is there's the retrieval aspect, which is, you take whatever data that you want to send to an LLM, you put it in a search engine or a vectorized database or whatever, you do the search, you send that to the prompt, but also you want to evaluate those prompts.



00:30:14 And obviously, there's a lot of data science methodology that you use to evaluate those prompts, and then you can create... You want to quantify it, but from a business perspective, from an application perspective, it really, I think, falls more on the side of application development rather than data science. Yet, if you go back to the history, and I think you look at, I actually haven't seen a data science definition or a job posting recently, but I don't know how much Gen AI skills are there, but my point is that it's somewhat tangential, yet it's still sort of part of this big bucket.

Jon Krohn: 00:30:52 Do you ever feel isolated, surrounded by people who don't share your enthusiasm for data science and technology? Do you wish to connect with more like-minded individuals? Well, look no further, Super Data Science community is the perfect place to connect, interact, and exchange ideas with over 600 professionals in data science, machine learning, and AI. In addition to networking, you can get direct support for your career through the mentoring program, where experienced members help beginners navigate. Whether you're looking to learn, collaborate, or advance your career, our community is here to help you succeed. Join Kirill, Hadelin and myself, and hundreds of other members who connect daily. Start your free 14-day trial today at superdatascience.com and become a part of the community.

00:31:38 Yeah, these Gen AI skills, I think they're such a transformative technology.

Marck Vaisman: 00:31:44 Oh, absolutely.

Jon Krohn: 00:31:46 That as a potential tool for ways that you can be building a better product or creating data, literally simulating data to be training or machine learning models, or, obviously, as an approach to solving machine learning problems

themselves. There's so many uses to Gen AI, and I think more than anything, everyone should be using one or more of the click-and-point user interfaces to these for all manner of everyday work and real-life problem.

- Marck Vaisman: 00:32:22 Oh, I mean, it helps in so many ways. I can't tell you the number of times I ask a Gen AI a question, or I'm trying to do this and I got some code that, obviously, it sends me in the right direction. And sometimes it works; sometimes it's not perfect, but I don't have to remember everything. So things like GitHub Copilot are really amazing, or if you just ask a question, if you go to a search and you ask a question, you got some reasonably good answers today.
- Jon Krohn: 00:32:54 For sure. All right, so let's talk about how we can be more effective. Given this situation that we're in with all these different kinds of skills, so I think you did a great job in your talk and in your slide deck of defining, so you have this over several slides, and now I think we're going to break it down in an audio-only way, but you've got this a la carte menu that you describe, so I don't know if you wanted to describe that menu at a high level, and then we can dig into the parts of the menu piece by piece.
- Marck Vaisman: 00:33:33 Sure. Before we do that though, I did want to say that part of the reason I got to this work or these ideas that we're going to be talking about and presenting today is that, I think data science is taking the levity out of the equation, sort of stopping making fun of ourselves. We know we haven't done a great job in defining this, and by all means, I'm not necessarily trying to solve that problem today. That's a big problem to tackle, but just raise awareness. But interestingly, when I presented this talk at Data Council a few months ago, I wanted to, rather than work on anecdotal evidence, sort of like Marck's personal opinions, which can be pretty strong sometimes. I wanted to see if there was actually data on it. And interestingly enough, so I looked up our Analyzing the

Analyzers publication, and I went to Google Scholar, and I saw it was referenced by about 150 works over the last 10 years. And I kind of started going up that path up the graph, what's referencing it, and I guess I was humbly surprised that our work has been quoted by a lot of other folks that are also publishing some really great ideas on data science as a field and data science education, skilling, and competencies.

00:35:09 So, long story short, is, I downloaded a lot of the publications that were published, I'd say, in the last maybe three or four years that quoted our work in some way. And when I looked through the works, I found that there had been some research done that attempts to define this thing called data science, and it's hard, and it's also hard to teach because of all of the different elements. So I attempted to summarize all of these potential skills and competencies to give folks a roadmap, whether you're starting out or you're mid-level, or you're a hiring manager, a writer or a director or an executive, and you're actually looking to hire people. I think it's fair to understand that all of the different aspects that make up this thing that we call data science, and try to be reasonable with the expectations and not just think that everything... I mean the unicorn, kind of going back to the whole mythical unicorn, which I know we all talked about a long time ago, and I think we still talk about today. Unicorns, we know they're mythical creatures, and they may exist.

00:36:19 I haven't seen one yet, but I don't know about you. So in the attempt to map all these things, I came up with, I would say, a framework where you look at the skills. You can't say... So here's a beef that I have, right? You see all these blog posts out there. There's like, "You want to become a data scientist? And here's a four-week roadmap to become a data scientist. Do X, do Y, do Z, and with whatever programming language." It's like, no, no. I

mean, that may only get you maybe 20% of the way there. Like, really, there's so much more to this than the technical side, right? Yes. The technical side is important; understanding the technical, the algorithms, the programming, being able to...

00:37:13 But there's so much other. What is the application? What is the result? The communication aspect. I think data science today, it's modern business practice. I mean, it has been adopted a lot in different organizations, some more than others, because some are more mature, and large orgs typically are much more mature in that sense, and they have specialized teams. In smaller organizations, you sort of have to do a little bit of everything because there's only a little bit of you. Nonetheless, if you think about the competencies, right? Because it's not really about skills; it's about competencies, and there's a good framework in one of the...

Jon Krohn: 00:37:53 What does that mean? What's the difference between a skill and a competency?

Marck Vaisman: 00:37:56 My sense, or at least I think the best way I can define it, is I think a competency is a combination of skills to be able to perform a specific task. Because sometimes like a simple task, you might just need one skill, but a lot of the tasks that we do, especially in the modern workplace, they're complex. It's not just a simple thing. It is a combination of skills. Competencies are not. I wouldn't say they're mutually exclusive because there's obviously overlap between competencies and, I think, skills, and there's not a one-to-one mapping of skills to competencies because the competency can be made up of... Many competencies can still use similar skills as part of the framing process. But there is this chart and in one of the works that we looked at that Cuadrado-Gallego and Demchenko. So one thing I did want to comment though is, interestingly, I did find that there has been a lot of

academic research in the last couple of years, not necessarily in the US, which surprised me.

00:39:08 There has been formal academic research in trying to define data science as a profession, as a field, and also try to establish a roadmap, sort of an academic roadmap of skills and competencies. That's where I sort of got all of this stuff. And interestingly, all of these new works reference our work, so that's how I got here. And I was surprised to see this because I've been thinking a lot about this recently, and I was surprised to see this, and I was really pleasantly surprised to find that the ideas are similar. When I read this, it's like, "This makes sense. I've sort of been thinking about this too." And obviously, I never published it or did any form of research, but it makes a lot of sense.

Jon Krohn: 00:40:01 Yeah. So your summary point there is that you did this research, or you created this publication a dozen years ago, and you're pleasantly surprised that so much subsequent research since has been citing that work, and the work that has been coming out since is largely aligned. The academic research is largely aligned with the thinking that you have been doing about what defines a data science career and what makes an effective data scientist.

Marck Vaisman: 00:40:33 Yes, I think that's a great way to put it. Thank you for summarizing.

Jon Krohn: 00:40:38 That is my job. Okay. So, Marck, let's get to how we can be effective. So let's talk about this a la carte menu. So there are, I mean, you can describe it. There's...

Marck Vaisman: 00:40:53 I think a good way to think about skills and competencies today for a data practitioner because let's take data scientists out of the equation. Let's make it more about a data practitioner because data practitioners is broader,

but that's my intention because there's different functional areas. However, when you look at the models, there's the skills model; there's a competency model. There's two or three different models that have been proposed by different folks; I try to take those ideas and compile them into a framework. And what I came up with is this idea that you've got, from a skills' perspective, or let's call it skills and competencies perspective. There's sort of three major areas. There's what I call a baseline set of data skills, which are made up of probably four broad areas. The one is kind of like what I call data literacy, is being able to work with data.

- 00:41:54 You don't have to be a data scientist, but you need to be able to work with data. And I think everyone today needs to be data-literate in the world. In the modern world, we all need to be data-literate. So it's about being able to understand information presented, numerical information presented. I'm not talking about creating statistical models or anything, just understanding the data, how it got analyzed, whether it makes sense, that sort of thing.
- 00:42:21 Second is data wrangling. You can't do machine learning, you can't do data visualization, you can't do just simple data analysis without data wrangling. That's just a given. It's unsexy yet it makes... And we all know it makes up a lot of our work, but you need to know how to work with different tools to extract data, whether it's databases, cloud, structured/unstructured data, text files. Wherever the data is, obviously, it's mostly digital these days, and if it's not digital, oh, boy. That's a whole other story. But it's that.
- 00:43:01 Then there's the computational skills. I mean, look, folks, we use the computer; we need to know how to use a computer more beyond using a browser and just applications. You got to use the terminal, you got to use the command line, you got to know how to make the

computer do what you want it to do. I cannot stress that enough, and I still see this in students coming into the programs. They come in, I mean, yes, they've never programmed before, and that's fine, but you have to effectively use your computer, and your computer's going to talk to other computers, and you have to have computational skills. And then the last part is, obviously, your basic data analysis and visualization skills. And I'll talk about communication separately. So those are what I call baseline data skills, and regardless...

- Jon Krohn: 00:43:51 Yeah, so I'll recap those. So the baseline data skills that you defined in this a la carte menu for effective data science are: data literacy, data wrangling, computational skills, and data visualization.
- Marck Vaisman: 00:44:04 Yes. And I am saying here that it's much more than an a la carte menu because it's not really a la carte. There's a lot of overlap here. Obviously, this is just one framework. So it's not about just picking the boxes; it's about looking at the whole thing and focusing on the areas that make sense. But there's a lot of overlap, and it's very hard. And the other thing, kind of going back to what I said before, is you see a lot of publications, people, whether they're blog posts or media posts, like, "Hey, I became a data scientist in two weeks," and there's a little roadmap. It's like, "Okay, start with the Iris dataset." Like, no, you're not using the iris data set anymore. I'm sorry, or the Penguins dataset or the Eurostat. Okay, fine. But that's an a la carte; it's like you have to check the boxes. This goes a lot further. It's not about checking the boxes; it's about understanding how all of these elements really come together. I don't know if that makes sense.
- Jon Krohn: 00:45:08 Since April, I've been offering my machine learning foundations curriculum live online via a series of 14 training sessions within the O'Reilly platform. My curriculum provides all the foundational knowledge you

need to understand modern ML applications, including deep learning, LLMs, and AI in general. The linear algebra, calculus, probability, and statistics classes are all in the rearview mirror. But the final three classes in the series, which are all on computer science, they are still to come. Registration for the first of these computer science classes is open now. That's Intro to Data Structures and Algorithms, on September 25th. And Data Structures and Algorithms, Level 2 on Hashing, Trees, and Graphs on October 23rd. And registration will open soon for the 14th and final class, Optimization, that will be held on November 20th. If you don't already have access to O'Reilly, you can get a free 30-day trial via my special code, which is also in the show notes.

00:46:02 Totally makes sense. Yeah. You're saying that there isn't a one size fits all shortcut-

Marck Vaisman: 00:46:09 No. No.

Jon Krohn: 00:46:09 -to being the ultimate data scientist. The practicality is that there are lots of different facets. You will never in your whole life be a data scientist that can tick all the boxes of data science skills because it's a fast-moving field. As we went through, over the past decade, we've had all those different eras: the big data era, the data product era, deep learning, MLOps, cloud, responsible AI, generative AI. Each of those eras comes with a huge amount of different tools and approaches, many of which are highly technical and distinct. And so, it is impossible to keep up and be like this ultimate unicorn data scientist. However, despite that, you are providing us with a candle in the darkness and confusion, in the storm-

Marck Vaisman: 00:46:58 I'm attempting to.

Jon Krohn: 00:46:59 Of all these things where, like you just said there are baseline data skills that everyone needs.

Marck Vaisman: 00:47:06 Yes.

Jon Krohn: 00:47:06 Regardless of which kind of data scientist you want to specialize in becoming, and like we just said, those are data literacy, data wrangling, computational skills, and data visualization. So then what's next? So then the next thing is, I guess, there are different data domains that we can choose to be more expert in. And like I said earlier, I think it's probably a good idea to, through things like podcasts or blog posts or whatever people like, to be keeping abreast of what's going on in all of the kind of adjacent domains, but to actually develop experience and deep expertise, you could pick and choose the different domains, one or two, or three.

Marck Vaisman: 00:47:44 Yeah, and so back to this mental model or framework that I'm talking about. Obviously, we're seeing it here, Jon and I, but you folks are listening it. So one is the baseline data skills, and think of that as one of the basic components. On the other side of this, there's another set of skills and competencies that are also basic but they are non-technical, right? So I would call it the baseline technical skills; what we were just talking about where the baseline data technical skills. Now I'm going to talk about what some people might call soft skills, or I really like the term that Cassie Kozyrkov uses, which is non-automatable skills. So it's the human side. And those are things like your cognitive skills, the communication, interpersonal and intrapersonal, and ethics, like ethical skills. And then there's a whole decision intelligence aspect, which is being able to make decisions based on information that sort of take all of these things.

00:48:52 So we've talked a little bit about the technical baseline skills. This is the, let's call them, the non-technical, or let's call them the human skills that you need. And I also call them like 21st-century skills and not me, I've seen this being quoted elsewhere, like 21st century skills,

because any modern business practitioner, any modern worker today, needs to have these skills pretty much regardless of whatever domain you're in, not just in the tech, but regardless of any other domain. So I think you have to have those skills. And then in the middle, there's more of the specialization.

- Jon Krohn: 00:49:31 One sec, quickly before you get to the middle and the specialization, let me just recap those. So there's overall, there are three broad categories of kinds of skills that people need to be an effective data scientist today. We've gone through two of them now, and those first two are all what you would describe as baseline skills. The first bucket are the baseline technical skills that we went over first. Those are again, data literacy, data wrangling, computational skills, and data visualization. And then there's the baseline non-automatable skills per Kazi Kozarkoff, or what a lot of people call soft skills, which is, I think we all know what we mean when we say soft skills. It's just kind of a funny thing because they are learnable skills just like anything else. But I think when we say soft skills versus hard skills, people know what that means. And so, that includes things like interpersonal skills, intrapersonal skills, ethics, decision intelligence like you described. The one in there that I didn't understand at a glance. What does cognitive skills mean?
- Marck Vaisman: 00:50:36 So this is an attempt. Again, I tried to summarize some of these works that are quoted in my presentation, and there is a set of competencies called cognitive competencies, and it's just thinking about... I wish I could give you a better definition, honestly.
- Jon Krohn: 00:50:56 Is it kind of general problem-solving skills, that kind of idea?



- Marck Vaisman: 00:51:00 Probably. I think so. And then if you look right, but then if you look at this sort of model, there's that, and then there's these three arrows that are kind of coming from different directions. There's statistics, computational and domain. And the domain, so we haven't gotten there yet. But yes, cognitive, I think just that simple word is attempting to qualify or describe, yes, problem-solving skills, how you think about problems, and how to get from a problem to a solution.
- Jon Krohn: 00:51:36 Cool. Yeah, that all makes perfect sense. And so, now I'll let you continue on with what you're about to do, which is that, so having covered those two big buckets of baseline skills that everyone needs to be effective data scientists, the technical skills, the soft skills that we talked about most recently. In addition to that, there are the data domains that we can specialize in. This is the third bucket, so go rock and roll on that.
- Marck Vaisman: 00:51:59 Yeah, so the data domains have been around for some time. I think we have seen some specialization. Again, 10 years ago it was called data science, but there is now data engineering, there is now data management and data governance. There is just more research methods, like research statistical... So some of these are interestingly related to the personas that came out of our work, not necessarily, but they're somewhat related. And some of these are lumped together. And I didn't come up with these data domain areas; I'm actually quoting one of the works, and I think it was the quadrata work; I can't remember at this point, but they actually lumped; there's a couple of things that are kind of lumped together. They actually lumped together research methods and project management interestingly, but there is some specialization, and I would say that maybe in AI engineer today is probably one of those data specializations.



00:52:58 But again, to be a data AI engineer, you still have to have the baseline data skills. You still have to have the technical skills, the baseline non-technical skills, or human skills. And if you look at the bottom here, there's a box that's called domain that crosses all of these areas because a lot of these are domain-specific, and a lot of skills are transferable across domains, but some are not. So a lot of this is going to depend on the domain that you're into. So there's going to be a specific set of domain skills that you're going to need based on the industry that you're in and the types of problems that you're trying to solve. So it's very hard to define and describe. And then, at the even higher level, these are skills and competencies, but then there's a third level, which is a thinking set of skills or thinking models. And one of the works quotes, they define three thinking models. One was the statistical thinking model, the other one was the computational thinking model, and the third one was the domain thinking model. So you can sort of see, right?

Jon Krohn: 00:54:04 All right, let's pause just here one second Marck. So I can kind of try to make sure I am wrapping my head around this correctly.

Marck Vaisman: 00:54:10 There's a lot, I know.

Jon Krohn: 00:54:11 Hopefully, the audience is here as well. So you've now gone through the three buckets of data science skills. So we've got the baseline technical skills, the baseline "soft skills." Then you've got the domain-specific kind of skills. So there's many different buckets of these, so there's like a data science/data analytics bucket, there's a data engineering bucket, there's data management and governance, there's research methods and project management. There's kind of like an other bucket that's like AI engineering or generative AI. And so you have the baseline technical skills you need to know, you have the baseline soft skills you need to know, and then you can

choose based on what interests you or what maybe you think is going to make the biggest impact or be most lucrative or whatever.

00:55:01 You can choose to specialize in some of these other data domains that I just listed. And so, that covers over, at least in this framework for describing data science skills. That covers almost everything that's in this framework that you're describing. And then you're saying on top of all that, on top of all the skills that I just listed, there's these three other buckets of statistical, computational, domain-

Marck Vaisman: 00:55:34 Thinking.

Jon Krohn: 00:55:35 -Thinking. So those aren't skills. It's thinking.

Marck Vaisman: 00:55:39 Yes.

Jon Krohn: 00:55:40 I see, I see, I see.

Marck Vaisman: 00:55:40 Yes. And I think it's the ability to apply; it's kind of meta, and I know I'm having a hard time describing it because I wrote this a few months ago and I haven't gone back to it, but that's related to the cognitive question that you were asking me before.

Jon Krohn: 00:56:01 Gotcha.

Marck Vaisman: 00:56:01 So it's just how you think about these things, right? Because it's not just about knowing a tool or knowing a skill. It's not just about knowing Python and knowing scikit-learn to build a predictive model. It's on top of just knowing the basic computational skills; it's like, "How am I going to computationally solve this problem?" It just sort of goes beyond, beyond that. If that makes sense?

Jon Krohn: 00:56:25 I gotcha, I gotcha. It's a mindset, it's an approach; it's kind of like a way of seeing the world. If you spend a lot of

time thinking about statistical problems, you start to see the world in data distributions and interacting data distributions, and you think about problems you could be solving as a regression model where you're thinking about, "Okay, this is the outcome that I'm trying to maximize, and what are the different kinds of data distributions that I can be using to inform that outcome?" So I think I understand that; I'm probably more of a statistical thinker myself and a computational thinker, for example.

Marck Vaisman: 00:57:02 Well, I'm glad I'm not the only one that sees the world and looks at it, and thinks about a distribution. So I'm glad we're having this conversation.

Jon Krohn: 00:57:09 That's all there is out there. Nice. Okay. So that is super helpful. And again, so you can get more on all of this kind of framework for ways of thinking of data science roles, data science skills, data science thinking frameworks in Marck's talk and in his slides, What Makes an Effective Data Practitioner in 2024? Again, we've got a link to those in the show notes.

00:57:37 So Marck, to kind of wrap things up here, armed with all of these kinds of terms. So having now a sense of going back to the beginning of this interview, we were talking about the data personas. So things like business people, data business people, data creatives, data developers, data researchers. And then, we also had, at that time, the kind of skill areas, but that's now been maybe updated in 2024 to what we were just talking about more recently with the baseline technical skills, baseline soft skills, the data domain skills that people can specialize in, as well as these thinking frameworks. And there's also, I think there's one other thing we might've forgotten to talk about from that schematic, which is general, so not data science domain-specific knowledge, but just domain knowledge for whatever kind of application area you're in. So you

could specialize in financial applications or in digital advertising, or in agriculture.

- Marck Vaisman: 00:58:41 So the domain typically refers to either an industry or maybe a functional area within an organization. But yeah, so obviously, there's going to be some domain-specific skills and thinking that you're going to need that, and again, some of it is transferable, some of it is not; it really depends. So, as you are thinking about your career progression, I think one of the things that you always want to think about is what are those things that are transferable? Maybe you've never worked in industry X, but you've got a lot of experience in industry Y, and there's a lot of similarities that you can relate. So those are the kinds of things that are transferable because maybe the domain is different, but the application is very similar.
- Jon Krohn: 00:59:24 Excellent. So armed with this information, armed with all of these categories and definitions, what can our listeners do to make the world a more sensible place for data scientists practicing, for data scientists interviewing, for data scientists hiring?
- Marck Vaisman: 00:59:42 That's a loaded question. So for the hiring side, I'd love to see hiring managers or organizations have more reasonable expectations in the kinds of... Really try to map what are their needs to the skills and competencies that they need and not make people go through like a rigmarole of hoops and take-home assignments and trivia problems and things that are not applicable to the core skill sets that you need to do for the job. That's on the hiring side.
- 01:00:18 On the growth side, on the personal side, on the skilling side is, you have a roadmap, and you don't have to check every box; as Jon said before, you really don't. You can't because it's going to take a long time and time is finite,

but you should have general knowledge of all of these things, right? It's going to be hard; you can't be an expert in every single part here. It's impossible. I mean, I've been doing this for 15 plus years, and by all means there's a lot of holes in my knowledge, and I think it's unreasonable to expect that level of mastery. Again, coming back to the unicorn, I feel that that's what sometimes it comes across is that design expectation that you need to be a master across all of these areas, and it's just unfeasible.

- Jon Krohn: 01:01:12 Brilliant. Yeah, that is a great tip. I agree 100%, that it is insane how the standard in data science interviews today has become, this getting deep into the theory of decision trees or Transformers or whatever, when the practice of the job in most roles, and again, there's exceptions, that kind of stuff-
- Marck Vaisman: 01:01:38 Yeah, of course.
- Jon Krohn: 01:01:38 -Those types of evaluations.
- Marck Vaisman: 01:01:39 Of course.
- Jon Krohn: 01:01:40 Make sense if you're going to be a researcher at Google DeepMind, but for most data science roles, it should be about applications. The kinds of things that make somebody a great data scientist are their grit, their perseverance, their curiosity.
- Marck Vaisman: 01:01:59 Curiosity is a big one.
- Jon Krohn: 01:02:01 Communication skills.
- Marck Vaisman: 01:02:03 You can't teach curiosity. That's one thing that you cannot teach, and I think you can build it, but I think it's part of those core skill sets. And the reason I say it why is because right now I'm working on some stuff and I'm looking at data and I'm like, "This is wrong; this just

doesn't make sense." And you have to start thinking, "Okay, wait, where is this data coming from? How is it generated? Who's creating this? I want to know what the upstream data source is." Again, these are just really small examples, and I think a lot of it's just written by curiosity, just wanting... Again, I think I consider myself an everlasting learner, and, obviously, I think that for me, the teaching helps with the learning because I always want to be up-to-date, like a perpetual student, I guess, for lack of a better term. But I've always been very curious, and I think that's a benefit. You can definitely build it, but just ask questions and ask questions and ask more questions.

Jon Krohn: 01:03:00 100%. All right, well, fantastic. Marck, this has been an eye-opening episode. Thank you for your work over the past decade defining these kinds of categories and making it easier for us to understand what it means to be a data scientist, for providing frameworks for the kinds of skills that everybody should be learning, and for providing us with a kind of sense that we shouldn't be overwhelmed because these kinds of expectations heaped on us by job descriptions or by hiring managers aren't realistic. But the kind of framework that you've described today where there are some baseline skills that everyone should know, but there are also these domain skills where you can just pick ones that you like, dive deep into those, find JDs that feature those, and find a company and a hiring manager that is willing to be realistic about [inaudible 01:03:52].

Marck Vaisman: 01:03:52 And I think that's the hard problem. Unfortunately, that's a hard problem. I think some organizations are better than others in that respect, but if you come in armed and saying, "Okay, look..." Part of the reason why you want to know about all of these things is because you also want to ask specific questions about the job. "What does the job really entail?" If it's something being, and I think

probably we've all had our fiascos over the course of our careers where we go in and for an interview and they're looking for X and what they really want is Y, but they call it X because that's what everybody else is calling it. If you read *Analyzing the Analyzers*, there's a couple of flick stories there, so about that, and that's not new. But it also helps you frame yourself to be that problem solver for whatever the need is of the organization.

- Jon Krohn: 01:04:48 Nice. All right. Well done, well concluded. Marck, before I let my guests escape from the program, I always ask them if they have a book recommendation. Do you have anything for us?
- Marck Vaisman: 01:04:59 I am reading right now a book on ADHD and relationships, and I've known, for about two years, formally, that I guess I have ADHD, probably like many other people do. And it can cause problems sometimes, so I'm reading about that, and I want to read more. I really want to understand the why's, and it explains a lot about a lot of things, but I also want to... For me, it's been really eye-opening.
- Jon Krohn: 01:05:35 Nice. That's a great recommendation. And yeah, I think I certainly know people who have only discovered as mature adults that they're on that kind of spectrum.
- Marck Vaisman: 01:05:48 Yeah.
- Jon Krohn: 01:05:48 And it is provided them with actually invaluable tools and actually, again, a lot of peace and solutions to problems that they kind of thought were problems; it turns out that they not problems. They can actually be... There's advantages to that kind of mindset as well. There's things that you can do that other people can't do. So yeah, so just learning about it more, understanding these kinds of frameworks helps people find peace and make the most of their unique skill set.

- Marck Vaisman: 01:06:22 And it's actually the first book I'm reading in a long time because I haven't really read a lot of books recently. I think part of it is because of the ADHD too. I read a lot of short stuff, but it's been hard for me to do sustained reading even though I read a lot.
- Jon Krohn: 01:06:37 Nice. All right, well, it's interesting to hear that from you, Marck. And last question for you is how people should follow you. Obviously, you've had a lot of useful insights for us on the show today, and I have a question for you, which is: Several of your social media accounts, your GitHub account, your Twitter account, it's called wahalulu?
- Marck Vaisman: 01:07:01 Wahalulu, yes.
- Jon Krohn: 01:07:01 What's that?
- Marck Vaisman: 01:07:04 That was something, I don't know. I was probably a teenager when I came up with that word. It doesn't mean anything; I just thought it was funny, and when I, in the early days of online stuff, I started using it as my handles and I've kind of kept it. It's my Twitter handle, it's my GitHub handle, it's my identifier in forums, and yeah, it's unique. Actually, I've put this... There was a close word that I think is in the, I don't know what the right term is in one of the Polynesian or Hawaiian dialects that is very close, and I don't know what it means, but because of the lulu kind of that sort of thing. But yeah, it has absolutely no meaning, but it's just one of those things that I sort of came up with in one of my goofy days, and it's sort of stuck around.
- Jon Krohn: 01:08:06 Nice. All right. Well, so let us know what are your social media handles? Where should we be finding you?
- Marck Vaisman: 01:08:12 Yeah, I'm not great with social media, honestly. I mean, I definitely have the accounts. I don't really publish a lot. I

have not been on Twitter or Mastodon recently, like I haven't posted. I do post occasionally on LinkedIn. That's probably the best way, and yeah, well, there's the link to the presentations, which we'll share. I definitely, that's one of the things that I need to work on is just having a better social media presence.

- Jon Krohn: 01:08:38 No, I mean, I wouldn't worry about it. I think that you're probably getting more done by not being on social media.
- Marck Vaisman: 01:08:46 Yeah, the time that you need to sort of foster a social presence is a significant investment, which I just don't have time to do.
- Jon Krohn: 01:08:54 For sure. Makes perfect sense. Awesome Marck, thank you for coming on the show today. I learned a lot, and yeah, you provided lots of useful information to us today. Really appreciate it. And yeah, we'll catch up with you again in the future.
- Marck Vaisman: 01:09:07 Yeah, Jon, thank you so much; this was awesome. I love talking about this, and if you listeners want to meet up and talk about this, by all means, please reach out. I'm happy to talk about it. I'm happy to present, talk to your organizations, whatever it is. And so, thanks again.
- Jon Krohn: 01:09:29 Such an exquisitely detailed episode with Marck Vaisman today. In it, Marck filled us in on the four personas of data scientists he uncovered through a cluster analysis of skills, namely data business people, data creatives, data developers, and data researchers. He talked about how data science is difficult to define because it is a complex field with multiple mutually non-exclusive definitions. How to be an effective data professional? You need specific baseline technical skills, namely data literacy, data wrangling, computational skills, and competency with data visualization, how you also need specific baseline "soft skills," namely interpersonal, interpersonal,

ethical, decision intelligence, and problem-solving skills. He talks about how data professionals can optionally specialize in one or more data specializations, including data analytics or data science, data engineering, data management research methods, and other newer specialized areas like generative AI.

- 01:10:21 Finally, he talked about how hiring managers should begin to come to grips with the breadth of data professional subtypes out there so that they can make their talent acquisition more effective and their interviews more fair and relevant. As always, you can get all the show notes, including the transcript for this episode of the video recording, any materials mentioned on the show, the URLs from Marck's social media profiles, as well as my own at superdatascience.com/821.
- 01:10:44 And if you'd like to connect with me in real life, as opposed to just online, I'll be giving a keynote and hosting a half day of talks at a Web Summit coming up on November 11th to 14th in Lisbon, Portugal. With over 70,000 people in attendance, I'm pretty sure it's the biggest tech conference in the world. Certainly one of the biggest ones, and it would be cool to see you there. Other folks speaking include Cassie Kozyrkov, the CEO of Groq; his name is Jonathan Ross; and the Brazilian football legend, Roberto Carlos.
- 01:11:14 All right. Thanks to everyone on the Super Data Science podcast team: our podcast manager Ivana Zibert, media editor Mario Pombo, operations manager Natalie Ziajski, researcher Serg Masis, writers Dr. Zara Karschay and Silvia Ogweng, and founder Kirill Eremenko. Thanks to all of them for producing another deeply practical episode for us today. For enabling that super team to create this free podcast for you. We're super grateful to our sponsors. You, yes, you. You can support this show by checking out our sponsor's links, which are in the show notes. And if



you yourself are interested in sponsoring an episode, you can get the details on how by heading to jonkrohn.com/podcast.

01:11:53 Otherwise, share this episode with folks who would enjoy it. Review the episode online, subscribe if you're not a subscriber, but most importantly, I just hope you'll keep on tuning in. I'm so grateful to have you listening and hope I can continue to make episodes you love for years and years to come. Till next time, keep on rocking it out there, and I'm looking forward to enjoying another round of the Super Data Science podcast with you very soon.