



# **SDS PODCAST**

## **EPISODE 21**

### **WITH**

## **SINAN OZDEMIR**



Kirill: This is episode number 21, with Data Science Author Sinan Ozdemir.

(background music plays)

Welcome to the SuperDataScience podcast. My name is Kirill Eremenko, data science coach and lifestyle entrepreneur. And each week we bring you inspiring people and ideas to help you build your successful career in data science. Thanks for being here today and now let's make the complex simple.

(background music plays)

Hello and welcome to the SuperDataScience podcast. I'm super excited to have you on board, and today we've got a very special guest. Today we've got Sinan Ozdemir joining us for the show, and you cannot imagine how many titles this person holds. So first and foremost, Sinan is the Head Data Science Instructor at General Assembly in San Francisco. So General Assembly is a continuing education organization where after you've done your Bachelors and Masters and even PhD, you can go and get additional skills that you may be requiring for your profession. And so Sinan instructs data science at General Assembly.

But on top of that, Sinan is also the founder and Chief Technology Officer of Legion Analytics. So it's a technology startup that intelligently sources leads for businesses based on previous customer retention. So quite an involved topic, and he applies data science methods there. And also, Sinan is an author. He's just published his first book just at the start of this year, and we'll talk about that and the contents and you'll get some insights from this book. It's also about data science. You'll get all that in this show.



So we go into a lot of very interesting topics. We discuss the topic of AI, so Sinan's got another side project in his business Legion Analytics, which is called Kylie. Kylie is an Artificial Intelligence that helps people answer emails. So it's a very, very interesting topic. And also we go into different discussions around R versus Python, and why Sinan thinks that Python is a better option, and what libraries he recommends. And also we talk about the workflow of data science projects and the crucial steps in a data science project.

So all in all, I think you're going to really enjoy this episode. This is a person with vast experience in the field of data science and education. So we tried to really deliver some of value here so that you can walk away with some additional insights into data science and into how you might want to structure a career. And on that note, without further ado, I bring to you Sinan Ozdemir.

(background music plays)

Hello everybody, and welcome to the SuperDataScience podcast. Today I've got a very exciting guest with me, Sinan Ozdemir, who is an author, a head instructor of data science at the General Assembly, and the founder of Legion Analytics. Sinan, welcome to the show. Such a pleasure to have you on.

Sinan: Thank you so much for having me. I'm very excited to be here.

Kirill: That's awesome. So how're you going today? Where are you calling from?



- Sinan: I am calling from San Francisco, California. Our day has just ended and I feel great!
- Kirill: Fantastic. And you probably would, because just recently, your book has been published and is now available for sale. So you've just completed your first book, how are you feeling about that?
- Sinan: It went very well. I spent the better part of 2016 writing this book, and I'm very happy to say that it is out for purchase from Packt Publishing, and I hope you all pick up a copy!
- Kirill: Fantastic. I actually ordered my own copy on Amazon just recently, still waiting to get it in the mail. And I'm just going to read out a quick sentence from the book description. So here it says "'Principles of Data Science' is created to help you join the dots between mathematics, programming, and business analysis. With this book, you'll feel confident about asking and answering complex and sophisticated questions."
- So, as I can imagine, you not only have a lot of experience in the world of data science, but you also have a lot of experience in the world of education, and you've put all of that into your book. So can you tell us a bit more about what does your book focus on, and what exactly do you teach in it?
- Sinan: So the book focuses a lot on the culmination of mathematics, computer science, and also the application of data science. I actually come from a background of teaching oral presentation skills, speech giving, calculus, statistics, programming, Java and Python. So I come from a very diverse background in teaching and I wanted this book to not really focus on anything very specifically, like only math or only programming, but I wanted to give a more holistic



view of where data science is, where it's going and how we can use data science in our day to day life.

Kirill: That's so fantastic. So you'd say that this book is not just for people who are pursuing a career in data science, it's also for people who in any way, shape or form encounter data in their day to day life. Is that about correct?

Sinan: That's exactly right. I wrote this book not only for the aspiring data scientist, someone who wants to learn the math and code behind it, but I also wrote it for the managers of the world. I wrote it for the recent MBAs who just want to learn more about what it's going to be like working with a data team, or working with a data scientist. So it's really for anyone who wants to learn more about the field.

Kirill: That sounds awesome. I can't wait to get my hands on it. And definitely, to our listeners out there, definitely check out this book and order your copy today. So let's talk a little bit about you. How did you get into the field of data science?

Sinan: That's actually a great question because I think it's not like most people's story. I actually began my professional career as a graduate student in theoretical mathematics at the Johns Hopkins University in Baltimore, MD. I was actually working on algebraic geometry as well as cryptography. I had actually heard about a start-up that a friend of mine was starting and he asked me if I'd like to be their first intern. I was actually in school at the time, so I joined the start-up not really knowing how to code but knowing the math. So they had spent the entire summer teaching me the code behind everything. For you listeners out there, actually I was coding in R, for what it's worth.



And by the end of the summer I actually had fallen in love with this concept of not only can I apply math theoretically, but I can apply it with programming languages. And that really sparked my transition from theoretical math into data science. From there I taught myself more and more programming, more and more applications, and about two or three years later I was finally able to comfortably call myself a data scientist and the founder of a data science company.

Kirill: Fantastic! And actually, speaking of your company, it's Legion Analytics, the one you're mentioning, right?

Sinan: Yes. Yes, it is, Legion Analytics.

Kirill: What does Legion Analytics do?

Sinan: Legion Analytics creates a suite of optimization tools for businesses. So we build an automated sales platform, an automated communications e-mailing platform and basically we build products that help make businesses more efficient.

Kirill: Okay, fantastic. I checked out the website. It's definitely something very interesting. Is it like all R-based, or are you using some other tools inside it?

Sinan: Actually, I have switched from R to Python, and the site is built on a Python backend.

Kirill: Oh, very interesting. Okay. So that raises the million dollar question, as I like to call it. R versus Python – what are your thoughts on that?

Sinan: That is such a good question. And I actually still get asked, "Why do you teach Python and not R? What is the real difference?" And I like to answer that question, "Honestly, in the end it comes down to a preference. And it's not only your preference, but the company you work for preference." The

reason people like Python so much is because more and more people are trying to get their feet wet in the field of data science. And they feel that R is a bit too hard to approach, and they think Python is much easier to look at, it looks like normal English. And I think this data alone has made this large influx in the users of Python because when they were first starting out they chose the—quote, unquote—easier language. Now that’s not to say that R is better or worse or Python is better or worse; it honestly just came down to “Python seemed easier. I’m going to learn Python.” That reads into the whole “more people using Python.”

Now, I started out using R because that’s what my colleagues were using when I was a mathematician. We used R. And the more and more I needed more general applications and web development or machine learning or even visualizations, I found that Python just had a larger community and a larger third-party base for all the things that I needed to do.

Kirill: Yeah, gotcha. So the Python story that it’s so popular is kind of a self-fulfilling prophecy, right? That more people tend to start with Python because it seems easier and then more companies start adopting Python because more people know Python. Like, it self-propels into the world of data science. Very interesting thoughts on that, yeah.

Sinan: Exactly. Actually, at General Assembly we used to teach both R and Python in our data science course. So we would teach them both concurrently. And the only reason we had to pick one was because we wanted to teach more material and less back and forth between R and Python. So the only reason we chose Python, again, was because most people who are coming in are beginners. Python is easier to pick up



than R on average. Now General Assembly teaches only Python.

Kirill: Yeah. Okay, very interesting. And moving on to your General Assembly experience—it feels almost like you’re doing everything. You’ve got a start-up, you’ve got a book, you’re teaching in General Assembly. So, moving on to General Assembly, can you tell us a bit more about them? Because I’ve heard of General Assembly and I think there’s a branch in Sydney and they also teach data science there and I think they also teach other topics. But just for our listeners and also for me in the sense that I don’t know enough about General Assembly, can you give us an overview of what General Assembly is and what is your role in General Assembly San Francisco?

Sinan: Absolutely. So General Assembly is an international continuing education service that provides both in-person and online curriculum. The curriculum can range anywhere from data science, data analytics to product management, web development, frontend/backend design. So it’s actually a lot of very diverse skills that people need to get into the workforce. We find that most of our students are actually looking for a change in career or they’re trying to level up their career by learning more all the time. My role in General Assembly, going on for several years now, is not only limited to the in-person classes that I teach but I’ve also developed curriculum for their online data science courses. I do a lot of work with headquarters to develop data science material for students and companies alike.

Kirill: Very interesting. So is that limited to just Python programming or is that like the whole suite of data science





tools and methodologies and just the philosophy of data science?

Sinan: I focus primarily on the Python tools and the philosophy and the overall workflow of data science. So I talk about statistical modelling, I talk about metrics and how to measure metrics and when to measure metrics and things like that. So it's not only limited to coding and math, but I teach a lot of when to use certain models or when not to use certain models, more importantly.

Kirill: Okay. Yeah, gotcha. So a bit of machine learning—I can sense a bit of machine learning coming out there. Is that true?

Sinan: Of course, yeah. So I do specialize in teaching, at least on the Python side, I specialize in teaching natural language processing, machine learning and AI tools. I teach everything from basic data manipulation using Pandas all the way to TensorFlow and neural networks and how to practically apply them and actually create production-ready websites using TensorFlow.

Kirill: Okay. Yeah, that's interesting. Because you keep mentioning there's a lot of web development involved as well, like websites and so on, are these tools that you're teaching, are they also applicable to people wanting to start their own business, to people who want to disrupt an industry? Is that kind of the idea behind it, or is this just for people working in like Netflix creating a recommender system, or Amazon or some other place?

Sinan: It's definitely a mix. We have students who are coming in from big companies—you know, Facebook, Google, Tesla—and they're looking to bring back to their company a little bit



more knowledge. But then again you have another subset of students who are looking to make a completely radical change in their lives. And just like my book, just like my company, I believe in this philosophy that data science is useless unless someone else thinks it's useful.

Kirill: Very nice. I like that quote.

Sinan: So, in my book and in my class, I focus very heavily on the combination of math, computer science and application. You can make the most predictive model in the world predicting cancer or the weather or whatever, but if no one knows how to use it, no one wants to use it, no one gets to use it it's effectively useless. So the web development here is really a part of the application of data science. You've created a TensorFlow model, you've created a scikit-learn model, whatever you've created, now it's time for you to get it out there. Let other people use what you've built. And a blog post is so static nowadays that—a blog post doesn't really allow a user to utilize or play with your machine learning model. So we build websites that actually deploy the machine learning models, that actually let other people use them for their benefit.

Kirill: Wow, that's really interesting. Could you give us an example of a recent machine learning model that you developed in that way, that is available for people to play around with?

Sinan: Sure. So one of my students actually created a great application where they would take open data about crime statistics and what they would do is they would actually scan to see where the user is in San Francisco and what time it is in San Francisco and effectively lay out a map of



where the most highest probability chance that a crime will occur within a half mile radius.

Kirill: Oh, wow. That's so cool!

Sinan: Yeah! And it was great because not only is it an application of predictive analytics, but it's a very useful application of predictive analytics. People actually want to know if "Where I am is safe at this time of night," or "Where should I go that is less likely for a crime to occur?"

Kirill: Fantastic! Is that available online for us to check out?

Sinan: You know what, I'd have to check on that because that was a year ago, and I really hope that this person kept their site open, and if it is, I'll give it to you.

Kirill: Alright, sure. If it's still available we'll definitely include it in the show notes. Yeah, that's a fantastic application. You know, it reminds me of that example where some kids in the UK created this machine learning system that would help people dispute their fines, their parking fines. And I don't remember in which period of time—maybe six months or a year—he helped people save like £200,000 just through this online tool that he created using machine learning. Kind of a similar concept to what you've discussed here.

Sinan: A lot of innovation happens and it takes a very particular mind-set to create that kind of work.

Kirill: Yeah. And I like that notion of innovation. I'm sure you probably talk more about this in your book. What do you think is the future for innovation in data science? Where do you think people who want to apply data science, who feel that there is an opportunity to apply data science somehow, but they don't know where or how to get started, what would



your recommendation for them be to get into this field of innovation and data science?

Sinan: Step one: read my book. (Laughs) But seriously, step one: educate yourself on where data science is at the moment and what kinds of problems are we tackling. And then once you've done that, there's actually a fork in the road. The fork on the left leads to a very innovative side of data science, the most powerful image recognition software, the most powerful audio recognition software and satellite imagery and counterterrorism and all these very, very difficult and very, very useful applications of data science attempting to be solved in laboratories by TensorFlow and other deep learning technologies. That's one fork.

The other fork actually leads to developing data science tools for those people who don't even know they exist. And that's actually a part that I take very heavily in my class is, how can we apply this technology to an industry that doesn't even know it exists. How do you take image recognition at its current state and apply it to a logistics chain or a factory line or a post office? So how do we take what's already out there and apply it to new industries? That's kind of a two different ways, I think, that a data scientist can go. Really pushing the envelope in innovation or taking what's already out there and applying it to a field that doesn't have it yet.

Kirill: Yeah, gotcha. And like prime example of that – Uber, right? Prime example of using data to change an industry that existed for decades, the taxi industry. “Let's just add a button to our phones to call a taxi.”

Sinan: Exactly. So Uber is taking kind of both forks in the road and combining them. They're taking very envelope-pushing



technology. You know, self-driving cars, advanced image recognition, and they're applying it to a very old industry. But I would argue that you can do a very similar thing with technology that has existed for almost a decade. You know, taking a basic linear regression and applying it to mom and pop cookie shops, bake shops. And I think that's where a lot of the money in the data science field is going to be pouring in in the next decade or two. It's not really in the very, very envelope-pushing innovative software, but "How do we take what's already out there and apply it to people who don't have it?" And you're actually already seeing that. For example, IBM Watson, Facebook and Amazon are just now releasing their assistants, their AI assistants, and that's a very commercial product. It's not necessarily meant for CEOs, but it's meant for students to use them, students in high school and college to use them. So this whole idea of democratizing AI and giving it, into the hands of people who never had it before, that's where I think a lot of the work will be done in the coming decades.

Kirill: Okay, yeah. That's a very interesting notion that the actual difference that people who have data science skills—the difference they can make is not only in the innovation that's completely brand new to the world, you know, creating supercomputers or like artificial intelligence and so on, but it's also about taking existing methodologies and just applying them somewhere else, somewhere to companies that don't have it. It kind of reminds me of the situation where the Internet just came out and first a few companies had their websites and many companies didn't. But then more and more companies started having their websites.



And it wasn't about creating, you know, pushing the limits of the Internet. It was about taking companies such as—like you mentioned, your mom and dad cookie shop—and getting them up and running with a website, and that made a huge difference. Those cookie shops that got up and running earlier, they outsmarted their competitors. And right now the same thing is happening with machine learning. So I totally agree with that sentiment.

Sinan: Yeah. That's exactly right. I think the ability to give AI into the hands of people who could not code it themselves, that's really going to make a big difference because people who have ideas but can't necessarily create or build programs or software or websites—that doesn't mean that they don't get to have ideas. So giving AI capabilities to these people is really going to be what's new in the coming decade.

Our new product at Legion Analytics is called Kylie.AI, and what Kylie is, is an automatic e-mailer. She'll actually draft e-mails for you in your inbox, whether you are a sales person or not, or a marketer, a customer support person. And the whole idea is we've built a very complex AI system but we want to give it to you. It's actually free to sign up and use Kylie. We want to give it to you so that you have the ability to see the power of AI and use it for yourself.

Kirill: That's a really cool idea. That would help out a lot of people, especially when you open up your inbox and you have like hundreds of e-mails. If somebody could draft up your e-mails for you, how much more efficient would people get and save so much time. That's a super cool idea! What else do you guys have at Legion Analytics?



- Sinan: So our main products actually are automated sales platforms where we will actually take in your sales information and automatically recommend and e-mail new potential customers for your product. So we'll take in your existing customers, learn from them and actually recommend, but not only recommend, but automatically reach out to them, track their engagements, follow up with them—again, all automated—and use that information in a feedback loop that will enhance your sales and make it better every single day. That's one of our products, and Kylie, as I just mentioned, is actually only about two weeks old. We released Kylie about two weeks ago, and we've already seen a couple hundred signups. And again, it is free so anyone who wants to sign up on Kylie.AI is more than welcome.
- Kirill: How do you spell that?
- Sinan: Kylie is K-Y-L-I-E-A-I. Yeah, go ahead and sign up. I'd love to hear your thoughts on it.
- Kirill: Yeah. So she's only two weeks old?
- Sinan: She's only two weeks old, but she is almost a year into development, so only about two weeks in production, but about a year's worth of research.
- Kirill: (Laughs) Well, two weeks old and she can already talk. That's pretty fast.
- Sinan: (Laughs) Yeah.
- Kirill: That's awesome. There you go. That's AI for you, so guys, check it out – Kylie.AI. Yeah, that's a pretty cool concept. So how did you come up with this idea of Kylie or of all the





other tools that you have inside Legion Analytics? What prompted you to create this company?

Sinan:

So my co-founder, Jamasen Rodriguez, he was actually a former salesperson/sales manager of a few companies and actually a former founder himself, and he had figured out the problem that sales is very inefficient because you're only as good as the person behind the computer. And he thought that it must be possible to automate some parts of the sales process. And when he met me—we had both gone to Johns Hopkins together—we had figured out that yes, we could automate it. And so what we did was we had actually gone into the Y Combinator fellowship batch one in California and we had worked on our automated sales platform – automated e-mail campaigns, prospecting, CRM tools.

What we had realized was our customers were having a new problem. We were recommending people, we were reaching out to them, but then their inbox started filling up with responses and they wanted a way to better respond to people. And that actually was the impetus to create Kylie, the ability to auto-draft and auto-send e-mails for you. So if you're getting e-mails every day, about something like "How much does it cost?" or "When is it released?" or all these questions about your product, Kylie will look back at your previous e-mail history, figure out your style and the best way to write that e-mail and actually respond. So the whole ability for Kylie to adapt to your personality for every single user, that's what we realized was the big problem here and that's what we created – an individualistic AI, an AI for everyone.





- Kirill: That's really cool. So it's not just that Kylie uses a preconfigured set of templates to answer to e-mails. It actually creates its own text. Is that right?
- Sinan: That's exactly right. So unlike other platforms that use templates or a single text corpus to learn from, every single user actually gets their own Kylie. It trains on their previous e-mail history, it learns how they speak, when to speak, what words to use – “Hi” versus “Hello” versus “Hey” – in what situations. And it actually adapts to your style and the person you're talking to for every single conversation. So it's not just a bland monotonic AI. It's really trying to be you.
- Kirill: Wow, that's fantastic! So if you can disclose—or to the level you can disclose—what kind of algorithms went into Kylie that facilitate this type of AI?
- Sinan: Sure. There's actually really two parts of it. The first part is taking in an unstructured text data from the users, whether that's their Twitter, their e-mail, their Slack conversations, and applying a structure to that corpus. So we're talking about unsupervised models here, some of which are open source, some of which are built by myself and I cannot disclose, unfortunately. (Laughs) But the ability to add structure to text is the first part. Second part was creating a natural language generation model. So, our generation models are actually part generative, part retrieval because we really want to give the ability for users to create their own text but also maintain grammatical structure and make sure that the actual conversation is flowing. So structuring text data and then creating, actually generating natural language – those are the two main steps that we actually have to focus on for Kylie.



- Kirill: Okay. Very interesting. And so by combining them, she can come up with a template text that somebody—do users need to adjust it and then send it off? Or do users need to—like, they can just click ‘send’ right away?
- Sinan: They can do either. They can either set up threshold and say, “If you’re 99% confident, please send this e-mail out without my consideration.” Or they could always say, “Please only send a draft.” So they’ll put a draft directly in their e-mail inbox. You know, “No need to come to Kylie. We’ll put it directly in your inbox and then you can revise, change, whatever you’d like and then send it on your own time.”
- Kirill: That is so cool. And is there a limit to how large these e-mails get? Like, can she only do like 20 lines or can she do it unlimited?
- Sinan: No. There is unlimited. She can do it short or as long as you’d like. You know, if the response is “Sounds good,” or a four paragraph description of your company, that is what she will draft.
- Kirill: Awesome. And what is the success rate that you’ve been seeing? Have you had any feedback from—I know she’s only two weeks in production—
- Sinan: We have.
- Kirill: So what are people saying?
- Sinan: So we’ve had a lot of feedback, and the main thing that we’re hearing is “This is saving so much time.” And here’s the interesting thing. We knew there were going to be slipups; we knew there were going to be mess-ups. She’ll draft something that’s not exactly correct. We anticipated this. What we did not anticipate was even when this happened,



even when Kylie would create a draft where you can kind of see where she was going with it but it wasn't exactly correct, our users still see this as a great success because it saves them time from really thinking about this e-mail from the ground up. So even if parts of the e-mail aren't exactly correct and they have to adjust it, it still saves overall time from writing it from scratch.

Kirill: That is very interesting. So is this only for business owners or can anybody with an inbox e-mailing their own friends—can they use Kylie as well?

Sinan: Anyone can use it right now. We are currently offering the services for sales and marketing people for free. Everyone else is put on a waitlist but that waitlist will start to become actual users in a couple of months. So I would get on that waitlist now. Or if you're already in sales and marketing or working in that kind of field, go ahead and sign up for free. But to your point, anyone can really use the software, anyone can use Kylie whether or not you're a CEO, an assistant or a barista. You know, it doesn't matter what you do or who you are. The ability to learn from your communications and actually communicate for you – that is open to everyone. It comes back to what I was saying before, really democratizing AI. Not only selectively giving it to a few people, but offering it to a wide variety of people.

Kirill: Okay. Yeah, that's really very interesting so, yeah, something to check out. I think that was a very valuable conversation about an example of application of AI which can actually change lives of so many people and increase efficiency and save time. So best of luck with that project. I think it's going to go places.

Sinan: Thank you. Thank you very much.

Kirill: Fantastic! And I wanted to ask you about something from your book description. I think this will be an interesting thing for us to talk about as well. In your book, you talk about the five most important steps of data science. And I know you probably don't want to – or can't even, maybe – give away too much from your book. But at the same time, if you could somehow outline these five steps and give us some insight into them, I think a lot of our users could benefit from that.

Sinan: Absolutely. So the basic steps that I outline in the book and I actually go over one by one, step by step with many, many examples in both high level and code is basically first you have to come up with an idea. And that sounds like a very simple thing when you say it, you know, very quickly and in passing by, just "I have an idea." But that idea really has to come from somewhere.

So like I mentioned, the idea for Kylie came when we got a lot of requests from our users to help us automate their inbox. So asking an interesting question and asking the right question can be very, very difficult. And as I'll say later, the ability to go back on yourself and actually go backwards in time and try again, humility almost is a big step. So even if you think it's the right question, you have to be able to prove it's the right question. So have an interesting question, something like "Can I predict heart attacks? Can I predict cancer? Can I add structure to tweets?"

Once you've come up with an idea or a question to ask yourself, what you have to do then is obtain data. Now, obtaining data in and of itself can be difficult or easy,

whether it's open dataset or you have to collect it yourself. That in and of itself is a very large question. Once you've obtained the data, you have to start actually not only cleaning and modelling, but evaluating. So I'm kind of mixing, I'm blending some of the steps together, but once you have the data, being able to clean the data, understand the data, manipulate and visualize the data. And you'll notice that I'm spending a lot of time not on the modelling or the evaluation, but asking the right questions, getting the data, cleaning the data. Because those steps themselves are almost as important, if not more important, than the actual models we'll apply to them. Asking the right question, obtaining the data, cleaning the data is almost as important, if not more important, than which machine learning model should I throw at it. Because if you're asking the wrong question, if you're applying a model to dirty data, your outcome is going to be either nonsensical or no one will care about it. So you have to be really careful.

So I think going throughout the entire book, asking the right question, obtaining the data, cleaning the data, modelling the data, evaluating your models. And the final step is application. It's communication and visualization. It's taking what you've built and giving it to the world in a digestible format whether it's a graph, a website, whatever it is, because if you're not sharing it with the world or the company or your dog or your cat or whoever needs it – again, it's useless, effectively.

Kirill: Yeah, I totally agree. It reminds me of—yesterday I was doing a case study, like presenting a case study on our website, on our platform. And at the very start, we were talking about U.N. analytics, analytics of U.N. votes. And if you hadn't

cleaned up the data properly, then the whole of the rest of the case study, of the presentation of analytics was totally useless. It really reminds me, or it kind of aligns with what you said here, that the very first couple of steps are about getting the idea, obtaining the data and cleaning it, preparing it, are so crucial for the rest of the analytics. Because so many times people do analytics and then they realize that the data initially was wrong or the assumptions were wrong, or that idea, nobody actually needs that idea. And all that effort has gone to waste.

Sinan: Right. And I think it's because a lot of people rush into the last steps, which is the modelling, the evaluating and the application, is that they forget the first parts: "Did I ask the right question? Did I get the right data? And did I clean it enough?" and they wasted time, they wasted resources. And it's unfortunate. So I think that people need to slow down and really start from the beginning and question themselves every step of the way and talk to people, ask someone, "Hey, if I built this, would you use it?" or "Who would use this?" or "Do you think this is worthwhile?" Data science is a lot more a team effort than a solo hobby. You know, talk to your colleagues. Talk to people who don't know data science because hopefully they'll use it too.

Kirill: Yeah, totally. And they can use it without knowing how it works in the background.

Sinan: Absolutely.

Kirill: And an interesting thing I wanted to ask you is, a couple of times we've brought up on this podcast TensorFlow. So you've mentioned this methodology and I'm assuming it's not something that you cannot talk about, it's not something



completely sensitive. If you could give our listeners a bit of an overview of what TensorFlow is and how you use it or how people could generally use it in business or in other applications.

- Sinan: Of course. TensorFlow is an open source machine learning library that's actually out of Google, so you know it's good. (Laughs) And they actually do—interestingly enough, people always think that “Oh, they're a neural network deep learning package.” And yes, they do that, but they also do regression. They do linear regressions, for example. You can do actually fairly basic things in TensorFlow, not just neural networks and things like this. We at Legion Analytics, and at Kylie mostly—we use TensorFlow primarily for the deep learning neural net technology. We apply their neural nets – specifically LSTMs and sequence-to-sequence models – to our text data. So in our generative and retrievable natural language generation, a big part of that is actually powered by TensorFlow and their deep learning technology.
- Kirill: Okay, gotcha. So it's an open source package that pretty much anybody can use?
- Sinan: Absolutely. And I will say this, I'll give them a big plug. Their documentation is some of the cleanest and easiest to read I have ever seen. You know, they actually take examples and they break them up into beginner, medium and advanced level. And not a lot of documentation does that. They break up this thing for beginners and advanced users. So I think Google and TensorFlow team did a very good job at creating materials to get started with it.
- Kirill: Oh, that's really cool. And this is a Python-based package?





- Sinan: It is Python-based, but they have API and other languages. I use it primarily in Python. I do examples with it in my book with Python.
- Kirill: Wait, hold on. So you actually have coding examples in your book?
- Sinan: Absolutely! The book is littered with coding examples, starting from basic syntax of Python all the way to our final case study. An example is using TensorFlow to do image analysis and text analysis.
- Kirill: Wow, that's really cool! I'm really looking forward to that now. (Laughs)
- Sinan: (Laughs) I'm glad.
- Kirill: Yeah, I'll check it out. Okay, so that was a great overview. So we've talked about quite a lot of stuff: about your General Assembly experience, about your company Legion Analytics, about Kylie – your little two-month-old baby, we've talked about TensorFlow. So quite a lot of topics covered and this has been a great discussion. Now let's move on to some more kind of rapid fire questions that I would like to get your thoughts on. First of all, what has been your biggest challenge as a data scientist that you've ever faced?
- Sinan: The biggest challenge I've ever faced I would definitely have to say is coming in with a very theoretical math background, having to learn not just how to code but how to code properly and how to code in a team without formal education in computer science. That was probably the hardest thing I had to do in the beginning.
- Kirill: Gotcha, gotcha. And I'll attest to that. Coding in a team is a completely different story to coding by yourself. Right?





Sinan: Absolutely. Absolutely! 100%!

Kirill: Yeah. Okay, next one is, what is a recent win that you can share with us that you've had in your career? Like, you've obviously had multiple: the book, Kylie, Legion Analytics and so on. But what would you say is the biggest one for you and why?

Sinan: I would say actually the biggest win is not my book, it's not actually the release of Kylie, but it's actually been that recently we brought on a lot of new team members in the development data science space and our new employees, our new team members saw the release of Kylie and were so excited that they wanted to join the team. So I think the biggest win I've had recently was the actual growth of our team and the growth of the people who wanted to work on this technology because they realized how important it was going to be.

Kirill: Wow! It sounds like you've got a—or you're creating a team of people who are very passionate about taking the world to being a better place through data science and machine learning.

Sinan: And that's really important because I can find someone who is good at coding or good at math or good at data science. I can find that. What is harder to find is someone who is not only good at everything, but actually wants to work on the project that we're doing and actually sees the future benefit of what we're doing and why we're doing it.

Kirill: Yeah, totally. And what would you say is the mission statement of your company?



- Sinan: I think our mission statement would be something along the lines of using machine learning and artificial intelligence to enhance the efficiency of the workplace.
- Kirill: Fantastic! I love that mission statement. I'm sure a lot of people can relate and align to that. Yeah, Kylie is just a good boost of confidence for those people to really put in their best efforts into the future of the company. And next question is, what is your one most favourite thing about being a data scientist?
- Sinan: Oh, wow! I think the number one thing I love about being a data scientist—I think actually it comes from the education side, is the ability to work on different domains and different projects and still have it be very interesting and useful. I think the ability to switch from communication to sales to marketing to scheduling to coffee shops and I think this ability to switch between gears while still maintaining the data science workflow, machine learning workflow – I think that's the most exciting part about being a data scientist.
- Kirill: Yeah, I totally agree. And on top of that, I would like to add that also what I really enjoy about data science is the transferability of your skills. Like, domain knowledge is definitely very important. But the core skills of whether it's machine learning or presentation or thinking about the problem, whether it's using programming languages – you can just take them, pick them up like a suitcase and go to the next project that you're working on and just apply them there and that's very exciting in data science.
- Sinan: Exactly.
- Kirill: Totally. And next one is a big one. So this is like a very philosophical question. Like, being in the space of data

science, you cannot ignore the things that are happening around us, whether it's Tesla, SpaceX, self-driving cars and the different types of AI that's popping up, even your own Kylie system. So the question is, where do you think the field of data science is going? And what should our listeners prepare for to be ready for the future that's coming?

Sinan:

That is a big one. I think that the future of data science, the future of machine learning and AI—I'll piggyback off a previous answer I had and say that the future of data science won't necessarily be in the most innovative software technologies. But what it's going to be is the ability to apply these technologies to fields that have not been updated, in some cases, for centuries. And I think that to prepare yourselves—the listeners—as you are learning data science, as you are listening to a lecture on chi-square test or listening about pandas or R or data frames or data manipulation. As you're learning this, never ever forget why or how you're going to use this information. Never get lost in the theory or even in the practice. Always keep yourself grounded by saying, "You know what? My wife's shop uses data and I think they're using it incorrectly. How can I bring this to them?" or "My buddy is starting a vineyard and I want to be able to use image recognition to help them know when their yields are coming in or when their yields are good or bad." So always remember, you can apply this not necessarily to the most—Uber, Tesla, SpaceX, SolarCity—not to the most innovative companies, but never forget that there is a plethora of SMBs, smaller to medium-sized companies who could also use this technology.

Kirill:

Wonderful. I love that answer. Yeah, definitely that's something we've already discussed a bit and it's good to

reiterate that. Always look out for where you can apply the knowledge that you're learning. And next one is—I've got like a bit of an interesting question, which might come off as a surprise question for you because we haven't talked about this before. But I've just thought of this—you, of course, achieved a lot and you're applying data science in many different forms and ways, and you're even teaching data science in General Assembly. My question to you is, how do you learn? How do you keep learning and refreshing your skills? Are there books you read? Are there people you follow? Are there courses you take? In a position in which you are, what are the next steps for you to keep learning and improving your own data science skills?

Sinan:

As a teacher, one of the most effective skills a teacher can have is knowing your students' learning styles, how do they learn – are they auditory, visual, kinaesthetic, do they practice, do they have to see it to believe it. And I believe that my learning style is very combination of kinaesthetic and visual. So the way I learn is actually by watching other people do it and then actually trying to mimic them and then seeing I can also get it that way as well. So what I usually do end up doing is I follow a lot of machine learning specialists, and not even like on Twitter, but I follow blogs, I follow even Coursera courses if I can find a good one or a Udemy course even or a course like that. And what I try to do is I watch the person perform the skill, whether that's machine learning, TensorFlow or whatever it is, and then I actually pause the video or what have you and actually try it myself. And I actually hope I fail. I truly do.

Kirill:

Why?



Sinan: Because if I get it right, that's boring. (Laughs) Because if I fail, that means "Great, this is a chance for me to not only figure out what I've done wrong but make sure that no one else fails here again so we can move forward as a machine learning community." So I actually hope that I mess it up the first couple of times, because I want to feel what it feels like to mess up this neural net so that I can notice it when someone else might do it as well.

Kirill: Wow! That's a very profound thought. And yeah, I agree with you, it's important to—kind of like this messing up will also point out to you how students who are taking this course, how they feel when they're doing it and that can help you understand how you can better serve your students to avoid creating pitfalls like that for them, right?

Sinan: Right. I actually call that almost an academic empathy. So your empathy is you feel what someone else feels, but academic empathy is almost learning how they learn. So I have to put myself in the mind of my students or my employees or anyone as I'm trying to tell them something and I'm trying to get in their heads. I'm like, "How are you perceiving what I'm saying?" And if I've had the experience of messing up or screwing up or having to do something over and over and over again, I'm much more likely to be patient with them and to work through their problems or even seeing an error that they're having because, "Oh, I've also had that error. I know how to fix it."

And that's really important because I feel like a lot of people put data science/machine learning on this pedestal like, "I'll never be able to know enough math," or "I'll never be able to code as well as they do." But that's okay because you don't have to code like the best coder in the world. You don't have

to be the most brilliant mathematician to apply even the most simple models in a useful way. So having that kind of academic empathy, the ability to feel what they feel as you're telling someone something, that's really going to inspire them to keep going even when you're gone. Even when you're no longer their teacher, they'll feel like it's okay to mess up because, you know, "Sinan told me that when I mess up, I have to find what I did wrong, fix it and then help the next person." And that starts a chain reaction of all these people who are better learning data science and not quitting and actually applying what they've learned eventually.

Kirill: I totally agree. That's a very inspiring thought. It's interesting how what you just mentioned about getting into other people's heads and understanding how they see the problem and they think about it. Just recently I learned this concept from one of my friends of a threshold concept. Well, concept of a concept. (Laughs) Basically this term "threshold concept" where as you live your life, as you learn things, sometimes you learn something that completely changes your view forever. You can never unlearn it back. For me, one time it was when I learned that we're all created of atoms and when you, for example, touch the table it's not you actually touching the table, it's electrons pushing away from each other and you're never physically touching the table. And for me that changed the world. Like, I can never think about touching the same way as I used to. And these threshold concepts for educators, they're very dangerous because once it gets into your head you can never unlearn it and it's very hard to look back at how somebody would think about the same problem without having learned this

threshold concept. So what you mentioned there is a very important skill for any educator to have, I think.

Sinan: It's a very difficult skill to have, I think. You're absolutely right because I might have a concept because I know how something works, but I also know that my student doesn't know it yet. So I have to somehow not just tell them what it is, but get them to believe it. They have to somehow find that answer themselves quickly and efficiently but, you know, effectively. (Laughs)

Kirill: Yeah, gotcha. Yeah, totally. That was a very, very interesting discussion. I think this whole podcast was full of value, so thank you so much for coming on the show. And finally, how would you say our listeners can contact you, follow you, follow your career? Because I'm sure there's going to be lots and lots of people who will want to learn more about what you're doing and get in touch and just keep following and see where your career takes you.

Sinan: Absolutely. You can always follow me on Twitter. My Twitter handle is @Prof\_OZ. You can always follow me there. I'm tweeting about new stuff and new technologies. You can actually find my contact information on my book. I actually put it in there because I actually very much welcome people reaching out to me with questions or suggestions or comments. You know, one of my mottos I like to say is, "How can I help?" and I ask that to everyone listening right now. How can I help? I'd like to be able to offer you guidance and answers to questions that you may have. So feel free to tweet at me or find my e-mail address, which is my first name, [sinan@legionanalytics.com](mailto:sinan@legionanalytics.com). And honestly, I look forward to hearing from everyone listening.





- Kirill: Fantastic! Thank you very much. We'll definitely include those in the comments. And I've got a very interesting question for you at the very end. One final question I normally ask. What is your one favourite book that you can recommend to our listeners? (Laughs)
- Sinan: (Laughs) Well, there is this great book called "Principles of Data Science." No. (Laughs) Well, first I would be remiss to not mention again my new book "Principles of Data Science" on Packt Publishing, also available on Amazon. But if I were to offer a book that I personally learned from to help me write this book, I think would probably have to talk about the "Introduction to Statistical Learning," the ISL, and I think it's a very, very popular book. It's a very common book. I think a lot of people use it in statistical classes and it takes a very theoretical approach to data science and machine learning. But I think it's necessary because I think that if people get too gung-ho on the coding, because they feel like that's the easiest thing to learn, then they're not really going to get the full extent of machine learning and statistical modelling.
- Kirill: Fantastic! That's a very good concept. So there you go, guys and girls listening to this podcast. Check out "Principles of Data Science" by yours truly, Sinan Ozdemir and also "Intro to Statistical Learning." And on that note, thank you so much, Sinan, for coming on the show and sharing all this depth of knowledge with us.
- Sinan: Thank you so much for having me. I hope I was helpful and I look forward to hearing from each and every one of you.
- Kirill: Fantastic. Okay, take care. Bye.
- Sinan: Have a good one. Bye.





(background music plays)

Kirill: So there you have it. And I really hope that you enjoyed today's episode. Personally, my favourite part was the discussion around applications of machine learning and artificial intelligence. We do live in a world where innovation is key and where businesses can be started in days and then grow to million and even billion dollar businesses in months. And oftentimes people find themselves looking for the next big idea that's going to really impact the world, that's going to radically change things.

Indeed, there is a space for that, but that is not always the thing that you should be looking for. Sometimes if you look around, you might find quick wins. You might find ways where you can apply these new technologies or these new algorithms, machine learning and artificial intelligence in the things around you to make the world a better place. Like, an example we talked about, where you've got a mom and dad bakery. They're not using machine learning, and maybe there's a way to help them to start using machine learning and artificial intelligence to better serve their customers. And by filling in a niche like that, you can still come up with great products that will impact the world and make it a better place.

So that was an eye-opening discussion and oftentimes we do try to think way beyond what is around us but sometimes it's just worth slowing down and looking around and seeing how you can change things right here, right now. And of course, if you enjoyed the discussions in this episode, I highly recommend picking up the book that Sinan just published. It's called "Principles of Data Science" and it is available on Amazon right now.



And as always, to get the show notes and transcript for this episode, come over to [www.superdatascience.com/21](http://www.superdatascience.com/21) and there you'll find all the resources that we mentioned and also you'll find all the links to Sinan's social media so you can follow him and his career. Thank you so much for your attention today. I really appreciate you and I'll see you next time. Until then, happy analyzing.