



SDS PODCAST

EPISODE 165

WITH

MATT DANCHO 2



Kirill Eremenko: This is episode number 165 with founder of Business Science, Matt Dancho.

Welcome to the Super Data Science 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.

Welcome back to the Super Data Science Podcast ladies and gentlemen. Today I've got a returning guest, Matt Dancho for the second time on the show. Last time Matt was on this podcast was episode 109, which went live in November 2017. So it's been over six months since then and it has been great ... It was very great to catch up with Matt and see what he's been up to. So most importantly Matt is now a father. Congratulations to Matt, he's got a baby girl and she's four months old and of course that's been the biggest event in his life in the past six months. And also, we talked about very interesting things that have been going on his life. Matt ... If you don't know Matt, Matt is a consultant at Business Science. He works with companies to help them implement data science in their operations or in their HR, finance and so on in different aspects of their businesses, and get insights from there and understand how to make better data driven decisions. And so his consulting business has been growing since the last time we talked, but also Matt started into education. He's been actively giving back to the community, have been actively giving back to data scientists.



He's launched a course on data science, on how to use data science in business and we talked quite a bit about that. So you'll get insights into how he structured the approach to creating this course. So you will know in which order to learn things and what are the most important things when looking into problems, and specifically we're talking about an attrition problem. Employee attrition, which is an HR problem, very [inaudible 00:02:37] problem that can be solved with data science. So that was a very powerful concept that we covered off in the podcast.

Another thing that we talked about is deep learning with R. So Matt is an R fanatic. He's actually developed some very powerful packages that have been very well received by the world and used quite ... their use is quite widespread. And so we talked about deep learning and R, that's an up and coming topic for 2018 and we all know that Python supports Deep Learning and artificial intelligence quite well. And R even a year ago, it wasn't a case, but now there have been some upgrades and we'll be talking about very interesting things such as Keras in R. And of course at the end, Matt will give some very interesting tips for those of you who want to give back to the data science community and want to take that step further and move from being a good data scientist to a great data scientist, and to a data scientist that cares about others and that helps others in the world of data. All in all, a very interesting podcast. I was very happy to have Matt back on the show. We had a great chat, can't wait for you to check it out. And without further



ado, I bring to you Matt Dancho, founder of Business Science.

Welcome back to the Super Data Science Podcast ladies and gentlemen. And today for the second time, we've got a returning guest, Matt Dancho. Matt, welcome to the show. How are you doing?

Matt Dancho: I'm doing awesome Kirill. Thanks again for having me back. It's been a pleasure.

Kirill Eremenko: It's been ... Like [Yazzy 00:04:21] messaged me just before the show. It's been over six months. I cannot imagine. It feels like ... It feels like it was that ... not that recent. Not that long ago I mean, like a month ago that we chatted and since then so much has happened. Like first of all, congratulations, you have a baby girl. [Shea 00:04:38] ... [Shea 00:04:38]? Is that ... Am I pronouncing it right? Four months old.

Matt Dancho: Yes, four months old. I'm a new dad and my baby girl, her name is [Shea 00:04:48] and she's just awesome. It's been a different experience for me, but it's been amazing. It's been really cool.

Kirill Eremenko: Fantastic. Fantastic. Very excited to hear that. It's always ... Does it ... Like 'cause I hear people saying that your life is completely different, like you enter into a brand new realm. How do you feel like after four months? Is that true or do you feel it's just like ... it's a bit of a different ... a little change in your life? Or is it like completely different?

Matt Dancho: Yeah. It's one of those things where ... So sleep is at a premium, but other than that, it's been really awesome. And you know what? For me, I found that I

still actually get the same amount of work done, 'cause that's one of the things that you always wonder about, is like how is this gonna change ... Your schedule really changes. So I think we were talking about ... I wake up now at like 4:30 in the morning and every morning I'm ... 'cause that's kinda my quiet hours and I can do my stuff. But then ... And then during the day, I'm able to spend time with my little one and my wife, and help her out. So it's ... Yeah, it's one of those things. You adjust, it's like anything.

Kirill Eremenko: Yeah. Yeah. That's critical. I'm very impressed that you managed to almost the same amount of work. Did you have to like give up on other parts of your life in order to achieve that or just more efficient now?

Matt Dancho: Yeah. So fitness, I definitely have cut back 'cause I used to do like about an hour a day of working out. And now it's probably like 15 minutes to 30 minutes a day and there might be a few days that I just kinda give it up and say, "Hey, certain things are more important." So.

Kirill Eremenko: Yeah. Yeah. Got you. All right. Well that's awesome. What else has been happening? Six months, for some people sounds like a lot of time and for some people not that much. How about you? What have you been up to since then?

Matt Dancho: So the past six months has been a blur. So besides having a child, it's been ... So I think we met ... What was it? It was in November, just like right after ...

Kirill Eremenko: Yeah.



- Matt Dancho: Or no, no, December right after H2O World. So since then, January is ... kinda kicked the year off and Davis my software manager, he went to RStudio because that was right around the time that I was getting ready to have a child, and I was just a little bit concerned with traveling. He gave an awesome presentation on tibbltime and then the next several months, what we've been doing is-
- Kirill Eremenko: Sorry, just to remind our listeners. Sorry to interrupt, tibbltime is your package. Is that correct?
- Matt Dancho: Yes. Yeah. So tibbltime is a time series package, which traditionally in the tidyverse has been very difficult to do ... to use. So now what tibbltime does is make that super easy. It's a newer package for us, but it's one of probably three or so that are very popular with the data science community.
- Kirill Eremenko: Got you. Got you. Okay. So ... Sorry, let's get back to your story. So you're business partner Davis, or like the person that you work with went to RStudio to present on that package.
- Matt Dancho: Yeah. So that kicked off January of last year and then since then, what we've doing is ramping up for Business Science University, which was just released last month and opened ... We opened up the doors with our first course and now we have 91 students and then-
- Kirill Eremenko: Congratulations. That's so cool.
- Matt Dancho: Yeah, I know. It's been really exciting 'cause you're always curious if people are gonna adopt it and wanna



be part of it and it certainly has been the case. So we've been just super excited. And then on top of that-

Kirill Eremenko: Tell us a bit more. I'm sorry. Tell us a bit more. So Business Science University ... So it's Business Science University ... business-Science.io. Is that correct?

Matt Dancho: That's ... Yeah, that's our website. So you can just google Business Science and we'll be the first hit and then I think actually Business Science University is the second hit. But it's basically ... Yeah, our way of kind of ... So we've seen this challenge that happens out there in our dealings with various clients. So over the past probably year and a half now, we've really interfaced with a lot of clients, some of which have helped us develop open source software, others are very like heavyweights in the financial industry. But what we've seen is this challenge of really being able to help those data scientists connect with the organizational needs and more importantly, the business needs. So what Business Science University aims to do and what I feel it does, is really bridges that gap. It gives them the tools, the systems, the frameworks and really just helps them see how I personally and the other instructors that work with me, really implement data science in an actual organization to show financial benefit. And to also align stakeholders up front in the process, while also showing them cool technologies like H2O, Lime and all sorts of ... just very like cutting edge data science. So that's what it's all about.



- Kirill Eremenko: Fantastic. That's really cool. Okay. So sorry, I keep interrupting. So Business Science, that launched off, 91 students signed up. Awesome. What next?
- Matt Dancho: So ... So, yeah. And then we've also been working very closely with a few key clients. One of them is S&P Global, which is a relationship that's just beginning and really what we did was ... Actually last month, it was really cool, RStudio another company that I wanna shout out, is ... been an amazing resource and partner for us. So they brought us into S&P Global, I guess it was in April now to do a training on time series deep learning, which we also have a nice blog article out about. But yeah, it was really cool. It was just kinda spur of the moment. We came out there, we gave them this training. It was me and and Max Kuhn at RStudio. He did one on ... He did a nice educational piece on his time series packages ... or not time series, his data science packages. But it was really cool, we got to meet the guys at S&P Global. We gave a presentation for about 65 or so people, just at ... both at S&P and other various financial institutions, and really ... It was cool because now it looks like we're gonna take the next step with them and actually do some corporate trainings, and it was just a really cool thing. So things are going well for us. It's been busy, just got back from R/Finance, that conference last week. So it's a blur to be honest-
- Kirill Eremenko: That's so cool.
- Matt Dancho: But it's been awesome.
- Kirill Eremenko: That's awesome and it makes me so excited. I gotta do a little plug here for our event for the listeners



listening to this. So I'm really excited 'cause Matt presenting to S&P Global about time series, deep learning and all these other cool stuff. Matt is actually coming to Data Science Go as one of the speakers and he'll be presenting there, and Matt you're also doing a workshop. Is that correct, at Data Science Go?

Matt Dancho: Yeah. I am basically open to whatever I can do.

Kirill Eremenko: Yeah.

Matt Dancho: So if I can ... Because at the end of the day, it's about helping the data scientists and giving them the tools they need to succeed. And I really see that a lot in Super Data Science and Data Science Go specifically, I think that that's where you guys are doing an excellent job.

Kirill Eremenko: Fantastic.

Matt Dancho: So I wanna be part of it, I wanna be out there. I heard ... I did not attend last year, but I heard it was amazing. Bo Walker, one of my long time friends, I actually took his ... I actually got to know him by talking his marketing class that he offers related to data science and he is ... he's the man. So he actually introduced me to you.

Kirill Eremenko: Yeah.

Matt Dancho: And told me I need to do it.

Kirill Eremenko: Yeah, definitely.

Matt Dancho: So I'm doing it.

Kirill Eremenko: Thank you. Thank you very much for the kind comments. And so yeah, for those listening, if you



wanna meet Matt in person and get to attend some of his speeches and workshops, Data Science Go is the place to come, October 12, 13, 14. All right, back to you. Back to all of your amazing accomplishments. That's been a heck of a journey, I can see that already, and so tell us a bit more about this course. So like you emailed me when you launched it, I think or after you ... or you had some initial traction and I've had a look at the [inaudible 00:14:11] and how you described the course. Very interesting idea, very interesting application.

So what's ... Just for the purposes of our listeners if you're not looking at your screens right now. The course is about data science for business and as Matt described, it entails the best practices and methodologies, techniques that he uses in consulting, but in this case, you specifically chose HR. So you're applying data science to a very ... like a selected problem of employee attrition, of how often ... or why are people leaving, what are the characteristics, what are the class ... how can you class them into different groups and deal with this attrition better. Can you tell us a bit more? Like why did you choose specifically this problem and what techniques? Maybe like give us a quick ... a few quick insights into your course. Why exactly ... Like what kind of techniques do you use to combat attrition.

Matt Dancho:

Sure. Sure. So the reason that we chose the attrition problem is because it's actually a part of a broader problem called churn, which is just binary classifications. So you can still take this course and



even if you're not an HR professional, it's still amazingly valuable because really you can swap this problem out for say a customer churn, or even like if you're in finance like fraud detection, that's a binary classification problem. So ... But the reason we chose HR and the employee attrition, it's actually quite simple. The top rated blog article on business science is our HR attrition blog post. So we kinda ... We used that as the feedback that we needed and we said, "Hey. This is the ..." Because we said ... When we set out to create this course, me and my team, we wanted to create real ... using a real world case study as an example. And it walks the student through the entire process and actually it's me coding and then giving them challenges and things to do. So they see my process, how I analyze it from beginning to end with this HR attrition problem and it's ... We felt that that was necessary in order to give it the context that it needed to be able to solve a real life problem, and so that's why we chose it.

Kirill Eremenko: Got you. Got you. Okay. So it kind of like ... It stops being just machine learning concepts and actually you see it in action. It like adds the flesh to the problem or to the theoretical components of machine learning and makes it more I guess engaging for students to watch.

Matt Dancho: Yeah. It just ... It gives them ... So my wife's in real estate and she often tells me that a lot of times when she takes a person into an empty home, that it's very difficult to get a concept. So I equate that data science since if you're trying to like figure out if that home is the right for them, you need to fill it with some



furniture. You have to put some context around it and I that HR problem, while it may not appeal to everyone, I think the vast majority benefits because they see the data science being applied to a specific problem, rather than trying to talk about it in the abstract and it helps with the learning.

Kirill Eremenko: Fantastic. That's a very cool analogy. Your wife must ... she should be quoted in the data science community for that one, like the empty home/furniture. Really puts it in perspective, indeed. When you go into an empty home versus a furnished home, you can like ... you're much more likely to ... You see yourself in there and maybe you'll buy it or rent it out. That's really cool. And so tell us a bit about the techniques, like obviously you probably can't disclose everything. So you wanna keep some value in the course, but tell us some of the main things that you look at when you are doing some HR analytics or when you're dealing with this attrition problem or any kind of binary classification problem.

Matt Dancho: Sure. Sure. So there's a few things that we do a little bit differently and I think that these are really some of the reasons that we're so successful with our clients, which is where we've tested these tools. And then that's kind of how I use ... I use that feedback from the clients to say, "Hey. Whether or not over time, that these are gonna be valuable to the broader audience." So the first thing that we do differently, is we start it off with a framework and it's actually ... it's called the Business Science Problem Framework, BSPF for short.



And it's actually something that we can give away to your listeners if you want-

Kirill Eremenko: Of course. But-

Matt Dancho: As part of the podcast.

Kirill Eremenko: That would be so cool.

Matt Dancho: Yeah. It's a really cool ... So what we did was we put a cheat sheet together. I mean just a PDF of what we're calling the Business Science Problem Framework and it's been something that I've implemented based on ... Actually, it's kind of ... It's got a funny story. So I read this book called 'Principles' by Ray Dalio and he's a ... he's very famous in the financial community. He's the founder of Bridgewater Capital, an amazing firm and what he was he put his principles down on paper and it turned into this book, and it's a really thick book. But what I did was when I was reading through those principles and understanding his management philosophies, a lot ... I found that a lot of it tied to data science. So I ended up integrating some of his theories, some of my own theories and also another project management framework for data science called CRISP-DM. And I integrated those three into this one framework and then I had to give it a name, so I wanted to put ... basically it became the Business Science Problem Framework. And so that's what we start the course off with and basically every chapter kind of walks them through another stage of that progression through the framework.

It starts them off with evaluating whether or not you even have a problem that's worth your time and effort



to try and solve. Because what we find is that a lot of times data scientists, they'll think they have a churn problem and then they'll wanna present that to management, but they don't have ... they haven't sized the problem. And when I say sized, they haven't put dollars to that problem to show, "Hey ...", to their executive team or stakeholders, "Is this a one million dollar problem? Is it a one thousand dollar problem or is it a 10 million dollar problem?" If it's a one million dollar problem, it's probably worth their time. If it's a million, it's definitely-

Kirill Eremenko: Yeah.

Matt Dancho: But if it's only a thousand ... if it's only gonna save that company a few thousand dollars, it's definitely not worth that data scientist's time. So that's where the framework really helps. Starting them off, aligning them with what the organization really needs to see in order to be able to kick the data science project off. And then throughout the process, it really just gives them kind of the steps ... step by step ... how to accomplish that data science project and when to get in touch with the stakeholders. So that's one thing. H2O, that's another big key piece. So we at Business Science have been implementing a lot of H2O for kind of like the ... or we'll say like structured data or like the tabular data-

Kirill Eremenko: So that ... Can you tell us a bit more about H2O? I know you explained it in the previous podcast, but for my personal benefit and for others who are listening who might not be too familiar with this concept. Can you explain again what is H2O?

Matt Dancho: So H2O is actually a company and they do a free and open-source platform also called H2O, where ... And it's available in R and in Python, and what it does is ... and one of the main reasons that I use it, is for this thing called automated machine learning. And that's what we teach in the course, is how to use automated machine learning because there's a huge benefit. It has this algorithm that under the hood, it applies a bunch of different models and a bunch of different techniques, including grid search to figure out which parameters to use for the different models that it tries, which ... It regularizes the data for you. It does like all of the stuff that normally takes a lot of time and effort for the data scientist part. So that way from my personal opinion, it actually helps me do what I need to do a lot faster.

So that's the big benefit, is you get a really high accuracy solution, and you don't just get one solution, you get a bunch of different models. So like GLMs, deep learning models, it actually stacks a bunch of the models and returns what's called a stacked ensemble. And so you get all these different models in what they call a leaderboard, which ranks the models based on different accuracy measurements. So for this classification problem, which is for churn, they rank them on AUC and log loss and you get this leaderboard that has all of these different models that you get to then explore further, that it's kind of done a really good start for you.

Kirill Eremenko: Nice.

Matt Dancho: Yeah. So that's-



- Kirill Eremenko: So it's kind of like Data ... what DataRobot does.
- Matt Dancho: Yeah. It's like what DataRobot does, but it's free and open-source and you can actually ... And it integrates great with Python and R, so it's really like the best of both worlds. So you don't have to have a DataRobot account or anything like that to be able to use this.
- Kirill Eremenko: Nice. Nice. Okay. Thank you. So that's H2O. All right [inaudible 00:24:47]. So that's also part of your course, how to build these automated machine learning models.
- Matt Dancho: Yep. Yeah, that's a big part of the course. We do two chapters on that, one chapter is just automated machine learning, and then the second chapter is more related to the broader task of taking that H2O and measuring the performance through like AUC, through ROC curves, through precision versus recall. And then also what the executives care about, which is gain and lift. We have a very detailed section on those types of graphs.
- Kirill Eremenko: Fantastic. I just wanted to mention here for our listeners that the purpose of this discussion is we're not trying to promote Matt's course in anyway. We just ... I'm really grateful Matt is sharing these things because ultimately ... Like if somebody doesn't wanna take the course, they can find all these things online separately. I find the value of course is that all these things are already aggregated for you and you jump in, and learn along, like learn in a row and in a structured format. So just for our listeners, if you're taking notes you can implement these things or any of these things, especially H2O, which is like an open-source free



software on your own. I'm sure Matt will share the link to his course further down the track, but yeah. So thanks a lot Matt for going into these, really appreciate you disclosing the content and the structure of the course. So let's continue.

Matt Dancho: Yeah. Okay. And Kirill, what I would say to that too is I agree with you a 100%. For those that wanna learn more, we actually have on our blog, which is free. You can check out the HR article if you just google I think employee attrition and Business Science, it'll pop right up and that's free. So if you ... That gives you just kind of ... very quickly, just the H2O algorithm and those things that you can implement for yourself.

Kirill Eremenko: Yep.

Matt Dancho: The course like you said is really ... it's much broader and it takes you through the entire program that we offer. But for those that just wanna learn H2O, that's probably a good start.

Kirill Eremenko: Got you. Okay. All right. So in the course so far we have the BSPF, the cheat sheet, which you've kindly said that you can share with our listeners and we'll include that in the show notes. Then you mentioned that sizing the problem is one of the key steps and data scientists forget about that quite often. Then we talked about H2O, the automated machine learning approach. What else? Is there any other things that you look at?

Matt Dancho: Yeah. So at the end of the course, basically what we're doing is then bringing it all home. So what we need to do is then ... We've got a really good model, we've got



... we can actually explain it using another technique called LIME, which stands for ... Well it's basically for a feature explanation. So throughout the course, you create this really high performing model and then you also figure out what's causing the attrition using this approach called LIME, like the fruit lime.

And then once you have that, then what we do is we bring it all home by showing ... We have a couple of chapters. One is dedicated to what we call a sensitivity ... trying to understand if you ... You have these levers now, so two of the levers specifically are overtime. So if you reduce overtime, how does that affect ... They predict like what effect will that have on the model, more people should stay. And then if you provide stock options, which is a cost to the organization, how would that affect the model. And then taking that effect and then doing what's called a sensitivity analysis to show, "Okay. I'm gonna adjust these levers and we feel that if we adjust them in these types of scenarios, then how does that affect the total cost of attrition." And we actually show that there's a benefit potentially of reducing your attrition cost by about 20%, which ... In dollars and cents, if you have 200 high performing people turning over, that could save the company a half million per year, maybe even more. So it's a ... Yeah. What we do is we really try and have it be business focused. So it's not just the tools and technique. Yeah, those are really cool and really neat, but if you can't provide business value, then they're worthless. So we have to be able to show that business value.



Kirill Eremenko: Yeah. Totally agree and feature explanation is definitely an important component, otherwise people are looking at it and they're like, "Yes, these are numbers, these are parameters. But what does it mean in a physical world? What does it mean for my business?" And once you ... I think it's important to focus on that because once you have the description of what it actually means, they can relate to it better and they can see the consequences or they can see the links between things. And also that helps ... In my experience, that helps find potential flaws. Like sometimes for instance, you might build a model and you might not take some domain knowledge into account. And then once you explain it somebody, if you go through the effort of actually explaining those features and explaining what they mean in the physical world, then they can give you feedback and they can say, "Hey, that doesn't sound right, like these two shouldn't be linked." Or, "Hey, maybe you can add this other thing that you didn't know about before.", and that might help and that can also increase your model maintenance or model performance in the long run.

Matt Dancho: Yeah. One of the things that really separates I think a good data scientist from a great data scientist is really that communication with the people that actually do the job, that do the work, that are familiar with the problem that you're trying to solve. Because at the end of the day like you said, that helps with your features, they're gonna know way more than you know about that problem and they're gonna be able to help you identify features like you said, if there's flaws in your

model or if there's things that you're just not considering. That's really where it helps out and then to your point too.

Kirill Eremenko: Yeah.

Matt Dancho: So the other thing is ... And to your point about feature explanation, that is critical to really being able to make ... really that's what the business cares about at the end of the day, that they need to be able to make decisions. You can have the best high performing model with H2O or any of these other great technologies that are out there, but if you can't explain it to the business in terms of levers that they can pool and what they can do to make better decisions, that's really what they care about. So you have to be able to do that. LIME is a tool that we use to be able to do that and we show you how to do that, but it's very important.

Kirill Eremenko: Totally agree. Totally agree. Okay. Thank you very much. Great overview and we're not going to go into discussion of R versus Python, we did that last time and you gave some great comments there.

Matt Dancho: I'm an R guy-

Kirill Eremenko: Yeah.

Matt Dancho: For everyone that doesn't know that.

Kirill Eremenko: Yeah. And the course is in R as well, right?

Matt Dancho: Yes. Yeah, the course is in R and down the road, I think we probably will do a little bit more in Python.

Kirill Eremenko: Yeah.



- Matt Dancho: Just because we know that there has been several people requesting it.
- Kirill Eremenko: Yeah.
- Matt Dancho: But really what we wanna do is show people how you can use R to be able to implement ... And you actually get to see me live code in R.
- Kirill Eremenko: Yeah.
- Matt Dancho: Like as I'm ... All of the courses are me basically kind of talking through my own coding and it can get a little raw sometimes, but ... And I say raw because it's me actually just kinda like actually ...
- Kirill Eremenko: Making mistakes on screen.
- Matt Dancho: Yeah, making mistakes. Be like, "Oh, I shouldn't have done that."
- Kirill Eremenko: But that's what people love, right? Like we live in a world full of ... Everything is so artificial and fake and plastic and you sometimes just want this sincere approach where you can connect with people. We all make mistakes. Nobody codes and gets all the code a hundred percent right all the time. People make mistakes and that's totally fine. Like I don't even edit those out of my videos. I think it's ... it helps people see that it's fine to make mistakes, we're all human.
- Matt Dancho: Yeah. And I think one of the funny things is like I'll run into an error and like onscreen, I'll have to troubleshoot it.
- Kirill Eremenko: Yeah.



Matt Dancho: And so they actually see the process of troubleshooting-

Kirill Eremenko: Yeah.

Matt Dancho: And it's totally like spur of the moment and ... But I think it actually helps them because then it sees, "All right. When I run into an issue, this is what I do."

Kirill Eremenko: Yeah.

Matt Dancho: "This is ... I'll have to google it or something like that."

Kirill Eremenko: Yeah.

Matt Dancho: And then, yeah.

Kirill Eremenko: How about support? Like for us for instance, Super Data Science, it's ... Like support is a very critical component of what we do because people ... some people have installation problems and people will have problems ... like run into errors or maybe there's compatibility issues with their operating system and so on. Do you find that that takes up ... like requires a lot of time to help students through this journey?

Matt Dancho: It hasn't been too bad just because I think right now we have 91 students, so we're not really dealing with millions. But I think as the program scales and as it just ... as we get more students, I think we will probably run into that a little bit more.

Kirill Eremenko: Yeah.

Matt Dancho: The problems that we've had, it might be like, "Hey ..." Because H2O depends on Java, "H2O, I couldn't get it installed when I updated to the most recent version of Java." And we just had to say, "All right. Well just

reinstall Java eight. If you go back to eight, eight's the supported version and they're working on getting the next version supported." So it's those types of things.

Kirill Eremenko: Yep.

Matt Dancho: And then also I think down the road, we do have a solution to that problem, which is this thing called RStudio Cloud, which I've just recently begun using. It's a free service that RStudio offers where you can actually set up ... It's kind of like a sandbox and you can set up ... like have all the software preloaded. So people can just hop on and they'll be able to code. So we're exploring that as an option as well, which just ... people don't have to install software and you can just jump right in things a lot quicker.

Kirill Eremenko: That's really cool. Another thing that we use in some of our courses ... I think in the computer vision course that we have in Python, what we use is a virtual machine. So we've actually prepared a virtual machine, we've already loaded all of the ... We've installed Python inside it, we've installed all the deep learning tools, all the AI things that we need, computer vision libraries and so on. And then we just wrap it up in a virtual machine and students just download that virtual machine, install it on their computer and therefore they kind of like ... as if they're logging into a computer which is on their computer. And everything is already set up and then they just have to code there. That could be one of the other things.

Matt Dancho: Yeah. Yeah. That's the same type of thing that we're looking into just to make that time from starting the course to actually coding faster.



Kirill Eremenko: Yeah. Yeah.

Matt Dancho: We wanna get that as short as possible.

Kirill Eremenko: Got you. And speaking of scaling, do you have any ideas for the next course in mind?

Matt Dancho: So right now we're working on another course that's actually an extension of this first course. So in this course, what the student really learns is the tools and techniques to kinda go from ... all the way through the process of understanding the business problem to developing a model to actually developing a recommender algorithm, that recommends decisions to the managers to help with HR. But we don't get into deployment and that's kinda the next step. So right now I'm working with Kelly O'Brien. She actually just started working at RStudio and she's been a long time friend of mine and she's working on a course that's the extension which turns that recommender algorithm into a web application. So this is using a tool called Shiny and it's developed by RStudio, but it's an amazing tool, it's free to use. People build lots of web applications with it because it's just so easy and it's the perfect way to get your organization ... actually using your data science. Like making decisions with data science, but without having to know how to run a model. All they do is they use a web based application that you build and it's just like drop downs and gauges and all sorts of different ways to interact with the user interface. And be able to allow the executives or the managers or whoever you wanna be able to make better decisions, make those decisions essentially using a web based GUI.



Kirill Eremenko: Got you. And just for our listeners, Shiny's like ... It's like an interactive dashboard tool similar to Tableau. That's-

Matt Dancho: Yeah.

Kirill Eremenko: Basically to summarize it.

Matt Dancho: Yeah. The benefit of Shiny is that if you're an R programmer, you're able to include R right into like everything that you do.

Kirill Eremenko: Yeah.

Matt Dancho: So it's the most flexible option that we've seen out there to be able to build these web based applications.

Kirill Eremenko: Got you. Got you. Okay. Cool. Well that's exciting news, that you're moving on to that. And I wanted to like shift gears a little bit and ask you ... pick your brain a little bit about something you mentioned before the show, that you're actually working on TensorFlow for R. Tell us a bit more about that, because TensorFlow for those who are not familiar, is the deep learning ... one of the deep learning approaches or packages for Python. There's TensorFlow, which was developed by Google and there is PyTorch, which was developed by Facebook. But they're all on Python and one of the biggest concerns that people have about Python versus R, is that Python is able ... in Python you're able to create artificial intelligence, deep learning and so on. Whereas in R, the packages are not so widespread and ... Actually it was you first Matt who told me that ... like you are not skeptical about it. You're very optimistic because there are things that ... going on in

the space, and now I hear that you're working on this whole TensorFlow for R. What's that all about? Tell us a bit more please.

Matt Dancho:

So right now it's been a ... probably over the past six months, it's really gained a lot of steam. And TensorFlow for those ... the guys ... don't know ... Probably the most popular package that uses TensorFlow is the Keras library and as of recent, probably within the past six months or so, Keras has been now mainstream in R using ... It's the library Keras, K-E-R-A-S. And it's actually built by RStudio and Company, and they worked with the guy who actually developed Keras, his name is Francois Chollet. So they actually worked with the gentleman who built Keras in Python and developed it into R. So what we've been doing is really using that a lot. We're actually working with a relatively new addition to RStudio. Her name is Sigrid Keydana and she's a TensorFlow developer for RStudio. And we're working on a combined blog post right now, where we're doing a lot of stuff with time series, with LSTMs and really just doing it all within R, which I think is pretty amazing, because a year ago, you couldn't do this. So this is really a new technology and that's what ... one of the things that excites me personally about using R and also the amazing technology and how things just are rapidly changing.

Kirill Eremenko:

Got you. Got you. That's really cool, very exciting. And so like already now, you can work ... you can start building deep learning models with Keras and R. Is that correct?



- Matt Dancho: Yeah. Yeah. We're doing a lot of deep learning right now. It started with a blog actually that we did a few months back on churn ... on customer churn. We used the Keras library and then since then last month when I presented at S&P Global, we presented an LSTM model that was developed using the RStudio and the Keras libraries in R. And since then, now we're taking the next step. So I actually just at the R/Finance conference that I was at last week, was talking with JJ Allaire, the founder of RStudio and he hooked me up with Sigrid. And since then, we've hit the ground running this week and we're moving towards probably within the next couple of weeks, getting a nice blog tutorial out that'll be available to everyone for free, both on our blog and their TensorFlow ... the RStudio TensorFlow blog as well. We'll both ... Both the Business Science blog and their blog will have it. So I think it'll be a nice edition.
- Kirill Eremenko: Fantastic. Wow. That's so cool. How do you find time for all these things and you have a child?
- Matt Dancho: Oh, dude I don't know. It's ... Yeah. It's like ... It's a lot of moving things. So it's kinda like juggling. You just don't want one of the balls to hit the ground because everything's so cool-
- Kirill Eremenko: Yeah.
- Matt Dancho: And everything deserves your time.
- Kirill Eremenko: Yeah.
- Matt Dancho: And really the thing that keeps me motivated is honestly helping the data scientists-



Kirill Eremenko: Yeah.

Matt Dancho: Because as you know, we're a big open-source company. We're also ... Like we're very committed to giving out a lot of free tools, a lot of ... We just want these data scientists to grow and to be effective. That's what it's all about.

Kirill Eremenko: Fantastic.

Matt Dancho: That's what keeps us moving.

Kirill Eremenko: That's so cool. So probably that and coffee. I loved your answer at the start. I asked Matt before the podcast, I was like, "So what's your morning routine?" 'Cause he wakes ... You wake up at 4:30 AM and so I asked him, "Do you have like a morning routine or something?" He's like, "Yeah. I drink lots of coffee."

Matt Dancho: Yeah. Yeah. So for those of you that are considering trying to get into a similar type of program or a routine, the toughest start ... the toughest part is actually taking that first step out of bed. Once I do that-

Kirill Eremenko: Yeah.

Matt Dancho: Then the next like three steps are headed towards the coffee machine and once I get that coffee into my mouth-

Kirill Eremenko: Yeah.

Matt Dancho: Things start to ... the ideas start flowing and everything's all good. But, yeah. There's a lot of caffeine involved.



- Kirill Eremenko: That's good. That's good. Cool. Well what I wanted to ask you as well is ... how about some tips for our listeners. So obviously, you've achieved quite a lot of success and you're helping ... you're giving back to the community. You're working on really interesting projects and creating these courses and blog posts. What does somebody ... What would you say to somebody who's just starting out and they're a data scientist, they are enjoying learning, enjoying their role and so on. What would you say they should look into to accelerate their growth? What are some of the most common things that you've maybe seen in your own journey and also that you've seen in other people's journeys in growing into not ... as you say, not just a good data scientists, but a great data scientist, somebody who can make a difference in the world, somebody who can start giving back to the community and helping others? What would you say to people who are striving for that?
- Matt Dancho: So for those that are striving and it depends where you're at in the progression. So if you're just starting out, definitely ... Like for example the Super Data Science courses, I've actually taken some of your machine learning course Kirill-
- Kirill Eremenko: Thanks.
- Matt Dancho: They're awesome.
- Kirill Eremenko: Thanks, I didn't know that.
- Matt Dancho: So-
- Kirill Eremenko: Thanks a lot.



Matt Dancho:

So, yeah. So that's a good spot. There's also some free resources like 'R for Data Science', the book. I highly recommend that for anyone that's just starting out. Once you get your chops and kind of you go from that where I'll say the beginner level to the intermediate, at that point I think you're ready to start making and impact. And when I say an impact, I don't mean just with furthering your own learning, but also helping others and being able to really empower other people to succeed.

And the way that I do that and the way that I've seen it most effectively done, is just sharing your analyses or sharing the things that you're working on, the things that you're discovering, but doing that in a public setting either through LinkedIn. I see a lot of people doing like ... I believe LinkedIn now has like the blogs or even just do a short post about things that you're working on or what's even more effective is if you develop your own blog. And I highly recommend anyone that's just starting out or even as an intermediate or whatever level, to develop your own blog, and really do that as a way to showcase your work because that's actually how we started Business Science. It was just me as a blogger blogging about the weird things that enjoy, like finance and stuff that ... Like I'd show my family and they'd be like, "What are you doing?" And they're like, "This is silly."

Kirill Eremenko:

Yeah.

Matt Dancho:

And ... But then it got a following, people just took to it. And it was not necessarily ... The target audience wasn't my family.

Kirill Eremenko: Yeah.

Matt Dancho: It was other data scientists like me who just had a passion for this type of ... this really cool software and technology and kind of like the intersection of everything.

Kirill Eremenko: Yeah. Yeah. That's some great advice and definitely I agree with starting your blog and putting things up there. But even if somebody just wants to take the first step, I recently discovered for myself Medium. It's a great place to publish blogs, it's so easy. So I ... One of my recent ones was about blockchain, about how blockchain mining works ... blockchain and bitcoin mining works, and it took ... Like I'm a very slow writer. It's much easier for me to record an audio or a video. But their ... It took me like a few weeks to write it, but their system is so good and their fonts are so pleasant. You just write it and then you like put it off, it automatically ... it auto saves, then you can go back to it, keep writing later on. I think it's a really cool system, plus as far I know, you can integrate Medium into your blog afterwards. So that's another way to look into it.

Matt Dancho: Yeah. Kirill, Medium's great. I have actually blogged on it a little bit. But actually I don't know if know Favio Vázquez.

Kirill Eremenko: No. Who is that?

Matt Dancho: He's a ... Okay. So he's a data scientist at OXXO and I think he's like a senior scientist or a lead data scientist there, so he's pretty high up the chain. But he's actually a guy that I've been working with a little



bit more closely. He blogs on Medium and he does a lot of actually Spanish blogging too-

Kirill Eremenko: Oh, wow.

Matt Dancho: But all around data science. So he's been able to do both in English and in Spanish and hit both of those audiences using Medium as his platform too. And he also integrates code into it, which I believe you can do. Right?

Kirill Eremenko: Yeah. Yeah.

Matt Dancho: Yeah. So it ends up being a really good platform of him.

Kirill Eremenko: Yeah. That's really cool. Okay. Well that kind of brings us to the end of the podcast. We've mentioned quite a lot of resources along the way, like the cheat sheet so that will be on the show notes. And what else? We've mentioned your course, we'll definitely put a link. Oh, by the way, I wanted to ask you. Are you ... Will you be able to give like our listeners some sort of coupon or special offer for your course?

Matt Dancho: Oh, yeah. Absolutely. I almost forgot about that. Yeah, if you do DSGO20 for the Data Science Go conference, we'll give you 20% off of the course. And so again that's DSGO20 and yeah, just sign up for the course. We definitely are encouraging new signups, there's a 30 day money back guarantee. So you can also test it out and then if you don't like it, which we haven't had any cancellations yet, it's definitely a good place to start if you're in that intermediate and looking to take it to the advanced level.



- Kirill Eremenko: Fantastic. Fantastic. Thank you so much. That's very generous of you. Yeah, so guys if you're interested in the course, then you've got the coupon. If you're not interested in the course or you're on the fence, then there's that blog that Matt mentioned. We'll also include that in the show notes about HR, which is like one of the most popular blogs like really ... How many views did you get on that blog?
- Matt Dancho: I don't know how many LinkedIn shares. I used to do it off of the LinkedIn shares-
- Kirill Eremenko: Yeah.
- Matt Dancho: Because that was ... that was kind of a nice barometer, but it was like four or five thousand LinkedIn shares-
- Kirill Eremenko: Wow.
- Matt Dancho: I mean it just went viral there for quite a bit of time.
- Kirill Eremenko: That's crazy. Yeah. So-
- Matt Dancho: And it still gets a lot of traffic too.
- Kirill Eremenko: Yeah, definitely worth checking out as well. Yeah, so those are some things that we mentioned on the show. Definitely ... All that's gonna be in the show notes. Anything else that you'd like to share with us before we wrap up?
- Matt Dancho: Did you wanna talk about maybe the book? I do have a good idea for-
- Kirill Eremenko: Oh, yeah.
- Matt Dancho: Based on our conversation.
- Kirill Eremenko: Yeah, yeah, yeah. Okay.



Matt Dancho: So-

Kirill Eremenko: Is there ... Let me officially ask the question.

Matt Dancho: Okay.

Kirill Eremenko: So Matt, this brings us to the end of the show. Is there ... Oh, before I ask about the book, I always ask where can our readers and listeners find you and get in touch with you. What are the best places to connect with you?

Matt Dancho: So I'm on LinkedIn, just Matt Dancho. You guys feel free to reach out. I love interacting. Send me a message, I love talking to anybody, especially from the Super Data Science Podcast. We got a really-

Kirill Eremenko: You got a lot of contacts the last time. Right?

Matt Dancho: We got a ton of contacts and it was great because it was ... just fantastic.

Kirill Eremenko: That's awesome.

Matt Dancho: So LinkedIn's the big one. Twitter, I'm also on there and actually you can find our company @bizScienc, which is on Twitter and also we're on LinkedIn as well.

Kirill Eremenko: Yeah.

Matt Dancho: So definitely check out our company.

Kirill Eremenko: That's so cool. And ... Yeah, and just wanted to say like Matt's one of those people that you can get in touch with and like if you have any mentoring questions or you have questions about like setting up your own consultancy. I'm sure Matt's gonna be very open to helping you out and answering those. So yeah,

definitely get in touch and finally Matt, this brings us to the end of the show. I just have one more question for you, what's a book that you would like to recommend to our listeners that can help them in their careers?

- Matt Dancho: So the book that I will recommend 'cause I already talked about 'R for Data Science'. So that one's out there for the beginners and we'll kinda set that one aside since I've already mentioned it. But the other book that I would recommend is called 'Deep Learning in R'. It's book that I just ... I got actually about a month ago and it's been amazing. So if you're interested in doing deep learning in R ... in the R programming language, it's a must read I mean-
- Kirill Eremenko: Who-
- Matt Dancho: And I think it's available. It's on ... What's that?
- Kirill Eremenko: Who's the author?
- Matt Dancho: So it's actually Francois Chollet and JJ Allaire.
- Kirill Eremenko: Oh, wow.
- Matt Dancho: So Francois is the ... He's the-
- Kirill Eremenko: Keras.
- Matt Dancho: He's the creator of Keras, yeah. And JJ Allaire is the founder of RStudio, so yeah. Actually JJ gave me an autographed copy of it at the R/Finance conference-
- Kirill Eremenko: Oh, wow.
- Matt Dancho: So that was cool.
- Kirill Eremenko: That's so cool.



- Matt Dancho: Yeah, and we're actually giving it away. We're gonna be raffling it off or giving it away as part of a promotion we're running.
- Kirill Eremenko: Awesome.
- Matt Dancho: So very cool.
- Kirill Eremenko: Okay. Well thank you so much for sharing that and all the other great insights from today's show. I'm definitely excited for people to check it out. And on that note, I'll see you at Data Science Go 2018 in October.
- Matt Dancho: Yeah. I'm really looking forward to it. Again, I've heard really good things about it, both from people that I really have a lot of respect for. So I hear it's pretty exciting too, pretty energy intensive, which is exactly where I wanna be.
- Kirill Eremenko: Yeah. For sure. For sure. There'll be lots of energy. All right. Thanks a lot Matt. Talk to you soon.
- Matt Dancho: All right. Thanks Kirill.
- Kirill Eremenko: So there you have it. That was Matt Dancho, founder of Business Science LLC, an up and coming consulting firm and also an online educator and online instructor in the space of data science, and somebody who loves giving back to the data science community. I'm sure you felt Matt's passion for data science and for giving back to the community and definitely ... I'm confident you've picked up some very valuable insights from this show. What I would say my biggest take away from today was that very cool quote, which Matt and his wife coined, that together ... through that story he



described with the furniture, that data science or machine learning algorithm on it's own is like an empty house. But once you apply it to a specific business problem, whatever it is, operations, HR, finance, anything else that's to do with the business, once you apply that problem to something specific, it becomes like a house full of furniture. And it's much easier to relate to, much easier to see yourself living in there, much easier to understand and actually make use of, and that's a great analogy. I really liked that and that's why for instance in our Super Data Science courses, we have a very keen focus on case studies. We introduce lots of case studies in our courses and that's why Matt's course is also structured around that HR problem of employee attrition.

So there we go. That was the podcast with Matt. Make sure to connect with him on LinkedIn, hit him up, get in touch, connect. And if you have any questions, ideas, suggestions you wanna run by him, I'm sure he'll be open to hearing those. Also, we'll include the links to Matt's business if you're interested in some consulting done for your business, some data science consulting. We'll include a link to of course his LinkedIn. We'll include a link to his academy or university where he does courses and where this current course is up and running. We'll include the coupon that Matt shared, DSGO20. What else? We'll include the book that Matt mentioned and the transcript for the episode. All of these things will be included in the show notes, which you can find at www.superdatascience.com/165. That's www.superdatascience.com/165.



And of course, last but not least, Matt will be at Data Science Go 2018, which is happening in October, 12, 13, 14. October, in San Diego. So if you haven't picked up your tickets yet, now is the time. Head on over to datasciencego.com and you can get your ticket for Data Science there. By the way, we've added quite a lot of speakers. Right now there's 12 speakers on the website as I'm looking at it. Have a browse through, see what topics we're covering, see what you might be interested in. It's gonna be a fun time and I can't wait to see you there. And on that note, thanks for being here today and sharing this hour with us. And until next time, happy analyzing.