SDS PODCAST
EPISODE 41
WITH
NICHOLAS CEPEDA
Kirill: This is episode number 41 with Aspiring Data Scientist Nicholas Cepeda.

(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)

Welcome everybody to the SuperDataScience podcast. Today I've got a very energising guest with us. Nicholas Cepeda is calling in from Chicago. And what you need to know is that just before this podcast was recorded, Nicholas got off the phone from an interview for a data analytics internship. And unfortunately, during the podcast he couldn't say the name of the company, he just called it a large media corporation, because he didn't know if he got the job or not. But now he got in touch with me a few days later and he said that he got the job and he's super excited about it, and he said he can finally share what company it is, and he got a job with Disney. How cool is that? He got a job with Disney to do some analytics there, like an internship in analytics and data. And so we're super excited for him and super big congrats and shout out to Nicholas and throughout this podcast, you will see why it's so obvious why this person got this job.

And the funny thing is that Nicholas hasn't been doing analytics for all his life. He's only been doing it for a couple
of years, he's only starting out, he's only studying all these things. And a lot of you are probably in that same boat, that you're just learning new techniques, new skills, you're just starting out. And this will be a very, very inspirational podcast to show what you can achieve in such short time if you are driven, if you are passionate.

So we'll definitely talk a lot about his interview and how it went, and what questions he was asked, and things like that, and how he felt about it. So you'll learn a lot of things there. Plus we will talk a lot about the tools that Nicholas is studying and Nicholas has used here and there. And we'll talk about R, we'll talk about Tableau, we'll talk about SAS. We'll actually dive quite a bit into SAS, and this will be a good podcast to get a feel for what SAS is like and what it's about. We'll talk about Hadoop and how Nicholas has used it before and what exposure he's had. We'll talk about the different machine learning algorithms and what are the advantages of some, what are the disadvantages of some.

And you will actually see how well-versed Nicholas is and how he operates with these terms on a very easy basis, that he understands these things. And it's very inspiring to see how much you can learn if you really want to in such a short period of time, if you really get excited about what you're learning. And I would really love for that excitement to pass on to you, and I think if you listen to this podcast, there's no way around it because Nicholas is so energetic and so enthusiastic. So that's the other thing why I really like this episode, is because it just gets you so enthusiastic after it. I walked away with so much energy, wanting to
conquer mountains and change the world. So without further ado, I bring to you Nicholas Cepeda.

(background music plays)

Welcome everybody to the SuperDataScience podcast. Today I've got a very interesting guest on the show, Nick Cepeda calling in from Chicago. Nick, how are you going today?

Nicholas: I'm doing fantastic, Kirill, how are you?

Kirill: I'm well as well. First of all, how's the weather in Chicago? We were just talking about it. It's pretty crazy there, isn't it?

Nicholas: It's crazy. We set the record of all time with 70+ consecutive days with snow in winter. I know in your Advanced R course, you talked about -- we analyze some weather patterns and weather data, and Chicago always had the biggest fluctuation.

Kirill: Yeah, totally. And now you have so many days. New York just got snowed out just a few weeks ago, and you're like, no snow at all.

Nicholas: Yeah, it's been the weirdest thing. My snow blower actually broke down. I'm so glad I haven't had to use it!

Kirill: Your what? Snowboard?

Nicholas: My snow blower.

Kirill: Oh, ok. I didn't even know those existed! Wow, I haven't been in snow for ages! Anyway, so today was an exciting day for you, you're calling in and you just had a big event. What happened, tell us, today?
Nicholas: I just had an interview. I got off the phone with a big media company for an advanced analytics and optimization internship for this summer.

Kirill: Wow, that is so cool. So how are you feeling about that interview? How did it all go?

Nicholas: I think it went really well, they were really impressed with what I had to say, and they were really impressed with the knowledge I had on the tools that I expressed. And a lot of that knowledge came from your courses.

Kirill: Oh wow.

Nicholas: I was on spring break last week from graduate school, and I don't think I got any sleep! I was up til like 3 o'clock in the morning knocking out these courses. I did the Advanced R, I did the Tableau A-Z, and I did the Tableau Advanced as well.

Kirill: Oh yeah, that's right, that's right. Because I saw you publishing these certificates, and I was like, "Wow, you're smashing through these courses!" That is so, so great to see. The tools are predominantly R and Tableau, and there's a couple of other tools that they're using as well, right?

Nicholas: Right, they're also using SAS. Which I don't have as much familiarity with, but my university, Northern Illinois University, they just created a university alliance with SAS, so I have open access through their cloud environment to get into SAS 9.4 and SAS Enterprise Minor. So I've been doing a little bit of work on my own trying to teach myself SAS skills.

Kirill: Nice. And what are you studying at university?
Nicholas: Management Information Systems with an emphasis on Business Intelligence and Analytics.

Kirill: Ok. That's pretty cool. So have you ever worked in analytics before, or is this the first job you're applying for?

Nicholas: This is my first job. I had an internship last summer with an Aviation and Aerospace company called AAR. I was a Business Intelligence Developer, an intern, and the project they had me on was developing a world map, a dashboard of all 73 locations worldwide, and then you'd be able to click and drill into them with IT specific information for the IT executives. It was an awesome project because it was something that was the first map dashboard that had been built at the company. It was a project that had been sitting on the shelf for quite some time because of lack of resources.

And because they had an intern come in, they threw me on pretty much as the lead of developing this dashboard. We were able to save the company $15,000 because they were using Oracle Business Intelligence Enterprise Edition to build their dashboards, and when they upgraded from the old software to the new software, Oracle promised that they’d be able to build all these beautiful map dashboards, but what they weren’t told is that there was a catch. They wouldn’t have all the data they needed to propagate the longitude and latitude for each country and each city. So in order to get all that data, they would have to pay $15,000 dollars to have that extra package. Of course, right?

So me and another developer, we put our heads together for about two weeks and we were able to crack the syntax. It was like a 90-column huge metadata table that we had to
figure out and how it worked, but we were able to figure it out and recreate the points in-house, which saved them from that $15,000 purchase. And then after that we also created a spatial technical training guide so they can recreate it after I had left the company.

Kirill: That’s so cool. And that’s also a great thing to do, like very altruistic or genuine thing to do, not to take away knowledge with you because some people are like that. They do something and they hope that that’s going to keep them in the company, but no. It’s a good idea to create these documentations so that if you have to move on, you have to move on. Things happen in life, and that’s good to make it easier for the next person that comes your way.

Nicholas: Right. With lots of pictures. (Laughs)

Kirill: Yeah, totally. That’s very impressive. Take us back to the interview. I’m sure so many listeners are dying to hear about the process because we’ve got so many people who are looking for jobs, who are just preparing for their first interview, who’ve just taken a few courses. Walk us through it. So, it was a phone interview, right? This was a second phone interview. What happened in the first one and how did they react, what did they tell you after the first interview?

Nicholas: First interview was a pretty basic interview, it was an HR interview. She wasn’t technical at all. It was just basic information, it was only 15 minutes long. You know, “projects you’ve worked on, why did you apply for this position, tell me about yourself, what makes you qualified for this position.” It was very quick, it wasn’t too detailed at
all. And then, something stood out to them about that and then I got another call. It was a long process, maybe about two months, and I got a call for a second interview.

Kirill: Okay. That’s the one that was today?

Nicholas: Correct. And then as soon as I got a call, probably about a week and a half ago, I don’t think I’ve slept since. I’ve just been studying non-stop, trying to prepare. I don’t think I’ve ever prepared this much in my life. I completely lack sleep right now. (Laughs)

Kirill: (Laughs) All right. Okay. So you studied, prepared, you’ve taken the courses. Is there anything else you prepared? Did you prepare for soft skill questions? Did you prepare for questions about what you want to do in the company? What exactly were you focused on when you were preparing?

Nicholas: Mostly technical. I think I have the soft skills, I think I’m spoken enough. I’ve had enough interviews, I’ve spoken with enough people. I was a Marine, so I’m not afraid to talk to people. So it was more fine-tuning those technical skills because I don’t have the experience of being in a technical role. I’m coming from a Marine Corps background, aviation supply, and then I had an IT internship as a BI developer, but I don’t really have that analytic experience. Most of my experience comes from data science courses in college, business intelligence courses in Udemy. So I was really focusing on getting that technical skill and then being able to translate that technical skill into a story.

Kirill: Gotcha. And you mentioned before the podcast that the skills or the experience that you built by doing projects in the courses – specifically, in some of my courses I try to
focus on portraying, not just the technical knowledge, but actually putting it in a way that you do a project while you’re learning R or Tableau or something. You said that that somehow helped you in the interview. Could you go into a bit more detail? Because that’s a common question I get from people that say, “I don’t have the real world experience. What do I do about it?”

Nicholas: Sure. The department I would have been working with, that I was interviewing for, was the Consumer Insight and Management Analytics, primarily for media data. So anything that marketing strategy would send out, e-mails or TV advertisements, or paper things in the mail, they needed to somehow quantify that. So, the course that I mentioned was the WeWashYouSleep [Romey] clustering marketing analysis, where we took the locations of I think 50 companies and we separate them into two regions and we analysed their average revenue, their average marketing spend, and the return on marketing investment and then we clustered them by revenue, marketing spend and then population to normalize them. And then after we clustered them, we ran a regression to identify which cluster had the greatest return on marketing expenses. And I mentioned that company, that project that I had done, and it sounded like her eyes lit up. You know, she was really interested in what I had to say about that, especially because it applied to marketing and specifically to what they were doing.

Kirill: Hold on, I’m going to stop you there because probably a lot of people are just—you are so excited, you’re saying so many things, and I just wanted to clarify. So Nick is talking about the case study that we have in the basic Tableau course,
“Tableau A-Z,” about k-means clustering. So, if you’ve done the Tableau course and you’ve gotten to the very last section, that’s the case study he’s talking about. And it’s exactly that, so there’s these companies in the U.S. that do washing of laundry and other items. So you outsource your laundry to them. There’s some analysis that we did around that. Basically you talked them through the case study that you did in the course and they appreciated that you have that knowledge.

Nicholas: Correct. Absolutely.

Kirill: And what else did they ask you on the interview?

Nicholas: It was two people on the interview, a male and a female. The other guy, he was more interested in technical things. He said, “What are some of the biggest challenges you have in some of these projects that you’re doing?” and I told them the general 70-30 rule, that 70% of my time is spent on creating a strategy on how I’m going to prep and clean the data. I talked about some other strategy I learned in your course with median imputation, I talked about removing those records if they don’t pertain to your analysis, maybe formatting it differently, splitting it, pivoting it, and being able to just clean it up. If it’s monetary values, maybe take the logarithm to normalize it a little bit more. Just spending a lot of time understanding your fields, understanding your data that you’re working with, and what meaning it has so you could be able to analyse it and interpret it.

Kirill: Yeah. I really like how you took knowledge that we discussed in a specific course. I think that a lot of that is coming from the R programming advanced course. And you put it into a

Show Notes: [http://www.superdatascience.com/41](http://www.superdatascience.com/41)
context of just generalized analytics skills so you don’t have to do it in R, right? You can do of all those things that you mentioned in any tool. If they’re using R, fine; if they’re using SAS, you can also do that in SAS. I think that’s a great way of pitching yourself at an interview, that you do know some tools, but if there’s a tool that you don’t know that they use, you can still take all of these analytics skills and transfer them onto that tool.

Nicholas: Absolutely.

Kirill: Okay. So that’s really cool. It sounds like you’re pretty excited about this opportunity, but for the benefit of our listeners out there I always say, “Don’t put all your eggs in one basket,” and I was just wondering what your thoughts are on that. Are you rock set on this one company, or are you keeping your options open and exploring other opportunities out there?

Nicholas: Absolutely not. I’m always exploring. The field of analytics is growing so much. There’s such a need and there’s such a demand for it in the future. There’s more of a demand than there are people to fill the demand so there are so many opportunities out there and I’m constantly looking.

Kirill: Yeah, awesome. That’s great. It’s great to hear that you’re so excited about this field and that you are looking to continue growing your expertise here and taking courses and looking for new opportunities. That’s very cool. So tell us a little bit about how you found it coming back from the Marine Corps back into the civil world, into—what do you call it in Marine—

Nicholas: Civilian.
Kirill: Civilian world, not civil world. (Laughs)

Nicholas: Well, it was actually a very past-dependent thing, me getting into data science and analytics. So, I joined the Marine Corps, I was aviation supply, I ran a warehouse with hundreds of millions of dollars’ worth of aeronautical parts, so F-18s, helicopters, you name it. So when I wanted to get out – I did four years, I got out, I used my GI Bill and I didn’t want to just throw four years of experience away because I think that would have been such a waste.

So when I went to go look for a degree program, I was looking more for supply chain, you know, logistics and supply chain, that kind of thing. I moved back to my home state of Illinois and I went to Northern Illinois University and I found a program called OMES, which was Operations and Information Management. Which was perfect, because it kind of had a little bit technical, a little bit of operations. It was just a great leaping point from my past experience.

During that program, I took a lot of supply chain classes, I learned about Edwards Deming and statistical process control and logistics and linear models. But then I took a database course and I fell in love with SQL. It was like a puzzle to me. It was like, “How can I word it and get the questions that I need out of these couple lines of code?” and I fell in love with it.

And so I switched my emphasis from operations to more information systems and then I went and I did an internship at AAR which highlighted both my technical skills that I had learned in college, and my Marine Corps skills because they
were an aerospace and defence aviation company, so it felt like the right fit for me.

While I was there I worked a little more IT, it was Business Intelligence. I was building dashboards, I was doing reports, I was doing ETL – Extract, Transform and Load. And while I was there, there was another developer there and he said “big data is the future, big data is the future.” And I said, “What is big data?” I didn’t even know about it at the time and I said, “What is big data?” He said, “Look at Hadoop. Look at R. Look at these things. There’s so much volume of data growing and it’s being able to take that volume and transform it into actual insight for businesses.”

And I said, “Wow! That sounds interesting. That’s something I want to be involved in.” So after I finished my internship, I immediately wanted to continue on with my GI Bill and to continue on with my education, and my college, Northern Illinois University, also offered an MIS program. So I thought that was the perfect next step. And the electives focused on big data analytics, machine learning, business intelligence and that kind of fit with what I was being told. So everything kind of led up to this point. Now I’m about to finish my degree. I’ll be done in December and then we’ll see where it goes from there. I’m building up my toolkit.

Kirill: Yeah, and once you’ve done that, the world is your oyster, right? You can go anywhere and do anything.

Nicholas: Absolutely.

Kirill: That’s so cool. Yeah, it sounds like you’ve been set out for success. Pretty much wherever you go you’ve been given all these hints about what you should do next in life.
Nicholas: Yeah, I’ve been very fortunate. Very, very fortunate.
Kirill: That’s awesome. Also, MIS – is that Management of Information Systems?
Nicholas: Correct. Management Information Systems, yeah.
Kirill: And when you were hearing about big data from that developer friend at AAR and other places, how did you feel about the overwhelm of tools that you have to learn? Were you not scared by the fact that you have to pick up R, you have to learn what this big data thing is, you have to understand all these new techniques and tools and add them to your toolkit? How did you feel about that? Were you just excited and powering through, or did you have that bit of scepticism about this whole new field?
Nicholas: I was completely overwhelmed. I had no idea where to start. I took a course on Hadoop and I started getting into Hadoop and we set up a semi-distributed cluster on a machine, we configured it. And as I was getting into this Unix shell scripting and setting up these XML files – I had never done that before. I’m not a computer science person, I’m not IT. I came from Marine Corps, I was in a warehouse, running a warehouse, and I never touched Linux before. I was like, “What is this? This is like the backend of the computer? What is this?” It made me so nervous. My wife was like, “If you don’t know it, this is not your degree.” She said, “You’re doing the wrong thing.” I was like, “But this is the future. I keep being told this is the future. I have to learn this.”

So I took it little by little. In undergraduate they taught a little bit of R and I was like, “What is this? This is like a programming. I’m not a computer science guy. This has
nothing to do with the business.” But it completely does. You know, there’s so much out there and as long as you can understand one, I think you can understand them all. Because they all do pretty much the same thing, it’s just a different language. It’s the same action, different language.

Kirill: Yeah, I totally agree with you. And it’s kind of good that you started into this world from SQL because it’s a very easy language to learn. That’s how I kind of got into it. I knew a few other programming languages, but my link to data science was through SQL. And I’m really thankful for that because it’s a very easy language to learn, hence Structured Query Language. Yeah, therefore you pick it up very quickly and then from there, it’s very easy to open up to others.

That’s the next thing I wanted to ask you. You said you’re very passionate about SQL. Can you tell us a bit about your journey? How did you get into SQL in the first place? And then maybe tell us some of the tips and tricks that you know about SQL that you’d like to share with our audience.

Nicholas: Sure. So, when I was in the Marine Corps they ran ad hoc queries, which is pretty much like SQL. I ran them but I never understood how they were built, and how they worked in the backend. When I took my first database course, that’s when I really started learning SQL. It was asking the database any questions I wanted. It was solving puzzles, it was manipulating the data, it was looking at different things. It was incredible. Some tips on SQL—I don’t know...

Kirill: For instance, why would you say SQL is so much better than or would you say SQL is better than Excel? There’s a lot of people out there who think, “I’m using Excel. I’m pretty good
at it and I probably will never need to know SQL.” So what do you think are the advantages of learning SQL over just knowing Excel?

Nicholas: I think Excel has a cap on it. You know, it’s only going to handle so much, it’s only going to be able to give you so much of an output. If you’re going to deal with a bigger dataset, more structured dataset, you need something like a database to be able to understand and ask the right questions to this database. You can always create a star schema in SQL and then bring it into Excel and use Power Pivot to kind of make pivot tables and analyse your data that way. But I think to have the structure, to have the ability to ask any question to the database in SQL is the advantage.

Kirill: That’s awesome. Just listening to you talk how you operate with these terms like, “Oh, I just do a star schema in SQL, put it into Excel, use a Power Pivot...” I mean, it takes years and years and years of practicing experience to develop the wealth of knowledge and you have that across the board, across all the tools. Whatever company gets you, they are going to be super lucky because you seem like a person who is super passionate about data science. You’ll go a long way. It’s really exciting to hear it.

Nicholas: Thank you. I have an insatiable hunger to learn. I’m very passionate. I’m very driven. I’m very motivated to get the work done and I think a lot of that comes from the Marine Corps. You know, just having that discipline to want to do better, to challenge myself and to be hungry to learn and never stop learning, being coachable and being able to mould and adapt and not just saying, “Oh, I know R,” and if
the company switches to something else like SAS, “Oh, I can’t do that. I don’t know that.”

Being able to adapt and learn—in the Marine Corps we had a saying called “Semper Gumby”. Instead of “Semper Fidelis,” which means “always faithful” in Latin, “Semper Gumby” means “I’m always flexible, I’m always willing to change, I’m always adaptable and I adapt and overcome”.

Kirill: That’s awesome. Great advice! Everybody listening to this podcast, if you want to become awesome data scientists, go to the army first, develop those skills, and then come back and learn. I’m joking. You don’t have to do that, of course. You can develop those skills another way. But I like what you said about adapting and being flexible. I think that’s very important for people, anybody in the modern world that is changing so rapidly, especially in the space of data science with all these tools coming out, with all of them taking over. Even the Gartner BI Magic Quadrant report came out in February, and it’s a completely different landscape, you know.

Now Power BI is overtaking Tableau, and where is that going to go and all the other ones are kind of falling off behind. I think the landscape of analytics and business intelligence is changing very rapidly. So it’s great to hear that you’ve managed to develop those. What would your advice be to people listening to this podcast who haven’t been to the army, who don’t have that self-control, self-discipline to sit down and learn and satisfy their insatiable hunger for knowledge? What would your advice be to them on how they
could develop these types of skills that are really benefitting you right now?

Nicholas: I think you have to do what you love. If you find something you’re passionate about and you really put the motivation in because you’re passionate about it, it’s not work to you. It’s a puzzle, it’s a game, it’s combining art and science, it’s painting a beautiful picture. To me, that’s something that I really enjoy, so it’s not hard to put in all these hours of work, to be up until three o’clock in the morning doing your courses. It just comes naturally when I have that type of passion and that type of interest in doing this kind of thing and just knowing that this is what the future is going to be. It’s the combination of data analytics and machine learning and artificial intelligence and deep learning and all these things to come. It’s going to completely change the way the world operates in 10 years.

You know, self-driving cars are not very far away. Within 5 years, we are going to be seeing these self-driving cars on the road and it’s going to completely change the landscape of not only how business operates, but how the world operates. And it’s just being able to adapt and see that insight of what the world will be. My father, he’s a printer, he’s been a printer for all his life and it’s an industry that’s just completely phased out. They constantly talk about cutting the cost by closing down the printing shop. That’s something he’s done all his life but it’s completely being phased out by technology. And being able to see and look ahead at industries that are growing and are growing tremendously is trying to get a step ahead of my competition, of the world.
Kirill: Yeah, gotcha. That’s very true. Definitely, that’s interesting and exciting, to see where the world is going. On that note, I wanted to also get your opinion on—there are so many different things that you’ve tried in data science, from Hadoop to SQL to R to visualization to business intelligence to geocoding to logistics, all these different areas. What is the area that excites you the most? Where do you think your career will take you? I know this can change with time, you don’t know what will happen tomorrow. But right now, how do you feel, where is your career taking you in terms of what you’ll be doing in the next 2-3 years?

Nicholas: What I’m really interested in is machine learning, is taking this data and running it through models, whether it’s linear regression to multivariate linear regression to see what factors influence on the profits, different factors of consumer behaviour to influence on their purchase decisions. Or taking logistic regression as a binary output and a yes or no output and saying, “What are the chances of this customer churning? What are the chances of this person defaulting on this loan?” And not only being able to say yes or no, but giving a probability to that. Being able to run these models is what really interests me, I think, and it’s what I really like to get into.

Kirill: Okay. That’s pretty cool. And what steps are you taking in that direction?

Nicholas: I’m currently in a big data and analytics course through EMC. My college is kind of running through a course in that sense, through EMC, and then I’m also enrolled in your Udemy “Machine Learning A-Z” course as well.

Show Notes: [http://www.superdatascience.com/41](http://www.superdatascience.com/41)
Kirill: Okay, that’s a big one. (Laughs) That’s one you probably won’t get done in one evening.

Nicholas: No, not even close.

Kirill: Okay. That’s pretty cool. And I’ve just thought of a really challenging question for you. What would you say to somebody who is listening to this podcast now who has gotten into the field of data science because, like you, they’ve heard that it’s super popular, it’s going to change the world, it’s like the new big thing, and they’ve gotten into it but they really don’t like what they’re doing? Let’s say they’re learning Python and they’re struggling, they hate it, but they’re kind of pushing themselves to actually do it because they know that’s the future. But at the same time, they don’t like it. What would you say to them and how would you say it to change their mind-set or change their framework or change what they’re doing in order to still be successful in the space of data science?

Nicholas: There’s multiple things. You don’t have to just be running Python. You can do visualization in Tableau, you can do visualization in R. You can do it in different languages. Maybe if Python is not your thing, maybe try R, maybe try SAS, maybe try doing something else. It doesn’t just have to be coding or hard coding. That’s a very computer science type of skill, but there’s other things. There’s business analytics, there’s being able to be the voice in-between of IT and management, you know, being able to have that communication of knowing a little bit IT, not too technical, but being able to tell a story, to be able to take complicated data and visualize it in a way that makes sense and be able to...
to tell a story. I keep saying “the combination of science and art,” being able to have that balance and show managers how they can have actionable insight with the data that they’re being presented.

Kirill: Yeah, totally. That’s exactly the kind of answer I was hoping for, that data science is such a broad field. If you’re hating or not liking that one thing you’re doing, you really should get in line with your passion and keep exploring. You can still be in the space of data science and still be successful, but find something that is more closer to your heart, whether it’s machine learning, whether it’s visualization, or like Nick said, being that connector between data science and management and the decisions makers. There’s lots of areas and you can definitely find your passion. Yeah, that’s really cool.

And the next thing I really want to ask you is, what would you say has been your biggest challenge you’ve ever had in the space of data science, whether it’s learning, whether it’s using data science? What’s the first thing that comes to mind that’s been your biggest struggle ever?

Nicholas: You know what? I think it was just being an undergraduate and—I was doing SQL and I got pretty familiar with SQL, but being introduced to R for the first time and never having had any programming experience at all, what is a vector, what is this. You know, he was doing an example, something like “I like apples” and it was like logical vectors and printing an answer based off a logical vector and I was like, “This makes absolutely no sense. When am I ever going to use this?” You know, I just didn’t see the value of it, so I couldn’t wrap my
head around being passionate about it because I didn’t see the value in it. It was just too early of an introduction to it for me.

Kirill: Yeah, totally. And then you got over that. Like, what did it take for you to really appreciate R programming after that?

Nicholas: I had to take a step back and see what it’s doing, not the nitty-gritty of creating a vector, creating a data frame. I had to take a step back and say, “What is our overall goal? What is our objective? I’m coming from a college of business. How is this enhancing the bottom line? How is this giving us a return on investment?” And once I understand that there is a huge gain in this and the companies who aren’t doing this are going to be non-existent in the future. That’s when I really said, “Okay, now that I understand what we’re trying to achieve, now I can trust R on how to achieve that.” I had to take a bigger picture first before I can get into the lines of code that I needed to get into.

Kirill: Yeah, okay. That’s a very useful tip, to look at the bigger picture of whatever it is that you’re doing, like, find the purpose. I think that not only helps understand the purpose behind your actions and therefore make them more meaningful, but it also helps understand where you need to limit yourself. Because a lot of the time people get carried away with analytics and they do this and they do that and they come up with these insights and those insights, whereas they only needed to do 10% or 20% of that work to really get the results that matter. Yeah, I think that’s a very valuable tip. Is that something that you apply generally, not
just in that one specific example? Do you try to apply that in every single project that you do?

Nicholas: Yeah, because you have to understand what you’re trying to do. If you’re tasked with something and you don’t understand what the person who tasked you with it wants, you can do all these analytics and all these models and do all this work and you’re just multiplying the expenses. If you go into the discovery and data prep and model building and model selection and doing all of this and then you come back with a final product that doesn’t even meet the needs of your customer, then you did all that work for nothing. You need to understand what we’re trying to achieve. It needs to be a teamwork, it needs to be a collaboration, it needs to be a bigger picture understanding of the overall goal.

Kirill: Okay, gotcha. And you’ve already shared a couple of wins with us. Like this interview that you had sounded pretty awesome. Then also you’ve talked about a project that you did for geocoding applications and so on. Can you share another one? What is another big win that comes to mind in the space of analytics?

Nicholas: I actually just finished a project in SAS where I was analysing a marketing campaign for a retail organization and what we did was we took prior marketing solicitations in order to better target for the next solicitations. We took traits that were similar to the most profitable customers and I ran it through a model in a project at school. So we took the raw data in, we explored it, we looked at the standard deviations, the missing values, we partitioned it through training and validation, we replaced it. Some of the missing values, we
took the logarithm of income and monetary values and then we ran five different models.

We ran an interactive decision tree, a decision tree at gradient boosting, decision tree, a regression, and a neural network. And then I compared all the models and their performance. Being able to do all that simultaneously with just a click of a button through SAS, I was able to run five different models and see which one was the best performing one overall.

Kirill: Which one was it?

Nicholas: It turned out multivariate regression was the most profitable for that situation.

Kirill: That’s so interesting.

Nicholas: Absolutely. It used factors of frequency, status of purchase, the last purchase amount, and then the lifetime purchase amount. Those were the main factors that were driving this model.

Kirill: That’s so cool. So out of all of them, it sounds like the simplest one was the most effective one.

Nicholas: Absolutely. My ROC curve looked great.

Kirill: (Laughs) Tell us why do you think that was the case. When you have multivariate regressions versus a neural network, why do you think that one wins?

Nicholas: I’m not sure. It was just kind of trial and error of what I was doing and that was the one that had done a better job. Maybe it was the variable selection, but it turned out the regression was the best one.
Kirill: Okay. That’s pretty cool. All right. So, SAS, talk us through this. You’re saying, one click of a button and you can run all of these different things. For somebody who doesn’t know SAS at all, what is SAS and how does it work?

Nicholas: Well, what I’m speaking about is SAS Enterprise Miner, and what it is it’s a very user friendly graphical user interface where you have different tabs like ‘explore,’ ‘modify,’ ‘model,’ ‘utility.’ There’s different elements within it and you’ll drag it onto your data flow chart. One thing is your data source. You’ll drag your data source in and then you’ll drag another thing called StatExplore in and then you’ll connect one box to the other box, kind of like a data flow diagram. So you’ll click run, you’ll import your data, and then you’ll automatically run your StatExplore and get the results because they’re connected. Anything you connect, it’s like a chain, a domino effect, so it’ll just run through.

Kirill: Gotcha. So it’s kind of like a drag and drop tool. You don’t have to do any coding.

Nicholas: There is coding involved, a little bit, but very minimal with this interface.

Kirill: Okay, that’s pretty cool. So you basically have a dataset and then you run all of these models just to find which works better?

Nicholas: Correct. And you can compare them across and then your final output will be all of the models plotted with their ROC curves as well as their regression [indecipherable 36:54].

Kirill: Okay. So, for those who don’t know, can you tell us a bit about the ROC curve? You’ve got so many terms that you’re
operating with, it’s hard to keep up. So tell us what ROC curve is.

Nicholas: Sure. The ROC curve is a measure of your model accuracy. It’s your true positive rate over your false positive rate. It’s how well you can split your binary classifier. It’s how well your model can do, how well it can perform. And the best curve would hug the upper left corner with the line. So, if you think of a plot, you’re plotting your true positive rate on one axis and your false positive rate on another axis. So perfect would be a complete up the left and across the top, hugging the left upper corner of the plot. That would be perfect classification. Every single time you ran your model, it would 100% hit every time if it was completely accurate.

And then on the opposite end, if it just went diagonally from across the axis on a diagonal angle, that’s no better than random guessing. So depending on where that falls and then you plot the area under the curve, your AUC, it’s a numeric value anywhere from 0 to 1 and it is how well your model performed. You want to optimize that number. You want the highest number.

Kirill: Okay. And you used the ROC curve to judge the performance of your models basically?

Nicholas: Correct.

Kirill: Okay, gotcha. And the other thing I wanted to ask you, when you’re running so many models, are you not fearful that you’re not focusing on one of them enough and that you’re not really optimizing the parameters that well? How much time did you spend on optimizing the parameters of each of those models?
Nicholas: I spent days on each one, just playing around with things and seeing which worked better, taking variables out, adding variables in, doing different things, changing my leaves on my decision tree. I did a lot of work just to kind of figure out and see what the best thing was.

Kirill: All right, cool. Thank you very much for walking us through that case study. I think it was very cool to hear about all of the different algorithms and how you’re comparing them. I can see why you’re excited about machine learning now when you’ve put all of these things together. It’s very cool to get your hands dirty and understand those. The next thing I wanted to ask you was what would you say is your one most favourite thing about being a data scientist?

Nicholas: It’s solving puzzles. I’ve always loved solving Rubik’s Cube and when I was a kid I loved doing jigsaw puzzles. To me it’s solving a problem, you know, anything huge that seems too overwhelming can always be solved by taking little baby steps. You know, solving one little problem, solving another little problem and solving another little problem, and if you solve enough of those little problems, you can complete your overall goal. That’s what data science is to me. It’s breaking up a huge problem into little manageable chunks and being able to do those manageable chunks. It’s very interesting to me to be able to solve that puzzle and tell a story.

Kirill: Yeah. Okay, that’s really cool. And if you hypothetically had all the knowledge and skill about data science and machine learning in the world, what puzzle would you be solving? How would you change the world with data science and machine learning?
Nicholas: I’d be an entrepreneur, that’s for sure. (Laughs) I’d be doing my own thing. I’d love to get into self-driving cars, it seems like something very interesting. Or wearable technology, to be able to analyse and interpret some wearable technology and GPS technology, because it’s constantly taking in data. You know, to be able to take that much data in at the time and be able to understand it in real time, I think it’s fascinating. It’s a little beyond my scope right now, but being able to do that in real time is incredible, I think.

Kirill: That’s really cool. And building on top of that, you’ve already mentioned what you think is going to happen, that data is going to be playing a big role. In terms of career opportunities, what do you think people should look out for in the next 3 to 5 years? What do you think they should start preparing for?

Nicholas: Automation. You know, more and more companies are out for bottom line. They’re going to want to reduce the cost and the easiest way to do that is automated processes. Tesla just built a factory and it’s almost 100% automated. Robots are completely building these cars and it’s very self-sufficient. And not only are they building these cars but they are learning as they go, they are becoming more efficient, more optimal every day. The more you apply deep learning, machine learning, artificial intelligence, it keeps iterating and it gets even smarter. I think that’s the future right now.

Kirill: Yeah. Fantastic, love it. I can totally agree with you on that. It’s just going to be a huge shift in terms of what humans do and what robots do. And no wonder we have that—what is it called? Universal income or something like global universal
income that is slowly being tested out. What are your thoughts on that?

Nicholas: I’m not familiar with that.

Kirill: That’s like—because automation is removing so many people’s jobs, lots of people are not going to have jobs so some governments are piloting this test where they’re going to be just paying people for being alive, basically. You just get a salary for doing nothing. It’s called a universal income. In Canada in some places they are paying that and I think in Denmark, they already implemented that in some areas. To your point that automation is becoming so huge, it’s a good thing in terms of progress, but it has these negative effects on people’s jobs and people’s employment so this is kind of a solution.

Yeah, we’ll see where that goes. It sounds very interesting, the tests they’re doing. It’s slowly starting to ramp up and more and more places are starting to pilot it out.

Nicholas: Interesting.

Kirill: That’s pretty exciting. Anyway, thank you so much for coming on the show and sharing this wealth of knowledge. It was great to hear about your interview and I really hope that goes well. In any case, I’m sure you’re going to find lots and lots of cool opportunities with your great aspirations and wealth of knowledge. Where can you say our listeners can contact you and follow you and get in touch with you if they’d like to learn more about how your career moves on from here?
Nicholas: Please contact me on LinkedIn. I’m always open to sharing and connecting with people, especially if I have a mutual connection with Kirill. Send me a message, add me on LinkedIn, follow me. I’d love to respond to you.

Kirill: Okay, beautiful. And one more question I have for you today is, what is your one favourite book that you can recommend to our listeners to help them become better in what they do?

Nicholas: Right now my Bible is EMC’s “Data Science and Big Data Analytics”. It’s just walking through, discovering, analysing, visualizing and presenting data through different statistical techniques, through R and a little bit of Hadoop at the end as well. This has been my Bible. I’ve carried this around with me everywhere.

Kirill: Why do you like the book so much?

Nicholas: I think it does a really good job at breaking down these topics simply. You know, a lot of the times you’ll get into looking at statistics but you need a PhD in Statistics to understand this stuff. They make something that isn’t so complicated way more complicated, and I think it does a really good job at explaining the topics and breaking it down bit by bit with practical exercises as well.

Kirill: Okay. Wow! Fantastic. So have you gone through all of it or do you still have a little bit left?

Nicholas: No, I still have a bit left. I’ve done k-means clustering, I’ve done linear regression, logistic regression, and I think I’m going to do a little bit of classification and text analysis as well. I think that’s the last chapter. And a little bit of Hadoop as well.
Kirill: Okay, wonderful. Again, thank you so much for coming on the show. I really appreciate your time and all of the things that you’ve shared. I’m sure so many people are going to find it so valuable.

Nicholas: Thank you so much. I really appreciate that you had me on the show.

Kirill: So there we go, guys. I hope you enjoyed today’s show. I hope you’re pumped and full of energy after this conversation. This is a testament. Nicholas’s example is a testament to where there’s a will, there’s a way. From not knowing anything about analytics, from coming from a completely different background, this is a person who’s managed to master R, Tableau, SQL and lots of different algorithms. You just heard him talk with such ease about multivariate regression, decision trees, regression boosting, neural networks, the ROC curve.

All of that knowledge, it’s coming from his drive, from his passion to learn, from his ability to sit down and just get things done, right? This is a person I have no doubt will have no problem whatsoever finding jobs. In fact, it’s just a matter of time before people are going to be asking him to join their company, before recruiters are going to be coming to him. Just that wealth of knowledge that he has and that passion for this field that he conveys, he’s not afraid to show his passion and get other people enthusiastic about analytics. He’s definitely somebody to look up to and I’m very excited that I had this opportunity to speak with Nicholas. It really got me very excited about the different things that he’s up to and the different things that he was
able to bring to the table at the interview. I’m very excited for Nicholas and I hope that you got a lot of takeaway value from here.

You can get the show notes for this episode at www.superdatascience.com/41. There you’ll find the show notes, all of the links, the link to Nicholas’s profile, so definitely hit him up, follow his career, connect with him and get in touch. Maybe you will be able to learn a few things from how his career progresses in the future. And on that note, thank you so much for listening to this podcast. If you enjoyed this episode, don’t hesitate to leave us a review at iTunes. We’d really, really appreciate that. That would help us spread the word about data science and get more people listening to this show. And that brings us to the end of the show for today, but don’t worry, we’ll be back in a couple of days and I’ll see you there. Until next time, happy analyzing.