SDS PODCAST
EPISODE 233:
HIGH OCTANE
DATA SCIENCE
LEADERSHIP
AT RED BULL
Kirill Eremenko: This is episode number 233 with Director of Data Science at Red Bull, Josh Muncke.

Kirill Eremenko: 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.

Kirill Eremenko: Welcome back to the SuperDataScience Podcast, ladies and gentlemen, super excited to have you on the show. And today we've got a very interesting guest from a very exciting company, Josh Muncke, Director of Data Science at Red Bull. I literally just got off the phone with Josh and we had an amazing conversation. This podcast is going to be full of valuable insights. For instance, we covered off a couple of case studies of how Red Bull uses data science, so if you're a fan of Red Bull, then this is going to be very cool for you to learn. Also, we talked about topics such as data science leadership and how that is such an important area for businesses to consider when they're starting out into the world of data science and for data managers to think about how data science leadership is different to leadership in other areas of the business.

Kirill Eremenko: We talked about asking good data questions, the importance of data science, and the decision making process in any kind of business. And of course we went through Josh's background, how he went from consulting into industry and what he learned along the way. So all in all, a very exciting podcast is coming up.

Show Notes: http://www.superdatascience.com/233
Can't wait for you to dive straight into it and without further ado, I bring to you Josh Muncke, Director of Data Science at Red Bull.

Kirill Eremenko: Welcome back to the SuperDataScience Podcast ladies and gentlemen. Today I've got a very exciting guest, Josh Muncke, Director of Data Science at Red Bull. Josh, how are you going today?

Josh Muncke: I'm really good. Thank you Kirill. Excited to be on the show. Thanks for inviting me.

Kirill Eremenko: That's awesome. So pumped. We're going to have an adrenaline filled podcast. So it was really cool to meet in person the first time we met was a couple of months ago in October at DataScienceGO, in fact, DataScienceGOx for those of our listeners who are not aware of what this is GOx is, it's a conference that we have for executives, data leaders and business owners and yeah. What was your experience like at DataScienceGOx, tell us how you felt at the event and if you've got any value out of it.

Josh Muncke: Yeah, I thought it was an amazing event. It was really the first time I'd been to an event like that which was really centered around the leadership aspects of data science. So it was great to have kind of a smaller, more focused session that was really dedicated to those folks that are leading and managing data science teams. So it was hugely valuable made a lot of great connections and contacts. Folks that I'm still in touch with now and have been speaking to and lots of interesting conversations and debates about what is
still a pretty new discipline, leadership within the data science world.

Kirill Eremenko: Thanks man. Thanks. Really appreciate the feedback. And indeed, one of the parts that I liked the most was having those conversations about data leadership. I remember we were at dinner and you mentioned that right now there is simply just is no platform for leaders to understand how to better set up data science teams, how to manage data science talent, how to retain data science talent and how to set up these projects and move forward. And yeah, that's a quite an important question in the world right now. I think it's popping up quite recently given how data science has been developing and it hasn't been an issue that before, but do you think this is like a indication of how data science is slowly maturing? What would you say?

Josh Muncke: Yeah, I think that's probably correct as data science as a discipline has become more mature and more and more companies are kind of creating and setting up data science teams and departments. They're realizing that actually you need good talented leaders to run those departments. And so in the early days of data science I think a lot of companies previously just hired one or two data scientists, gave them the keys to the data warehouse and said, "Hey, go and play and come back with something interesting or valuable." And now companies are trying to actually embed data science into the way that they work and the way that they make decisions. I think they're figuring out that actually keeping those teams happy and engaged and tied to the objectives of the company is not just a case
of putting them in a room with the database, you actually need people who can create the vision and the strategy and and the career paths for those people too. And that is what data science leadership is. And it's not easy.

Kirill Eremenko: Yeah. Yeah. And in your personal journey, so you've moved from consulting in IBM to Deloitte and now you're a Director of Data Science at Red Bull. How have you gone about getting this knowledge of data science leadership? Obviously there's a lot of trial and error, but how would you recommend somebody in a similar position to you to develop these leadership skills specifically in data science and lead their teams correctly?

Josh Muncke: Yeah, I mean, it's hard now, and I'll be the first person to say that it's been a learning experience for me too. I think like myself, a lot of people come from kind of a more technical data background. I studied physics and I was in data consulting, as you said, IBM and then at Deloitte. And so my kind of early part of my career if you like was as a data scientist and so that kind of training and that kind of experience doesn't necessarily prepare you well for being a leader in data science. So I think a lot of people who are now kind of leaders and managers and data science teams kinda ended up there by chance, like they got promoted into that role not necessarily because they were naturally great leaders and natural managers of people and talent.

Josh Muncke: So I think first thing to realize is like everyone finds it difficult in that space and it is a new set of skills,
right? It's a new ladder to learn how to climb and you shouldn't feel bad if you find it difficult or if you find it to be something that you need to take time to learn. The big change that I made from consulting to joining Red Bull was one that came with the need to go from managing projects and groups of people to deliver a single goal to the managing my own team or creating and setting up and then managing my own team.

Josh Muncke: And one of the things that I found to be really, really helpful and just how I did that effectively and how I did that well was to find coaches and mentors within my company and outside of my company. So there were no other data science mentors and coaches. So what I had to find is people who I felt could see were good leaders at Red Bull and outside of Red Bull and then speak to them about leadership challenges and problems and questions that I had and even just getting that outside perspective I've found to be really, really helpful.

Kirill Eremenko: Very interesting. How would you say the leadership in data science differs to leadership in other areas of the business for instance in Red Bull? Of course there's a lot of things that you can copy and take away, but what are the main differences that people need to look out for?

Josh Muncke: Yeah, I think there's a few things that kind of make data science a little bit unique. And one thing that I think makes data science harder to manage than maybe some other aspects, it's just a fact of it can be very exploratory and open ended. So Angela Bassa who is the director of data science at iRobot actually has a
great article on Harvard Business Review and she's done a few podcasts as well, talking about this. The fact is you're not managing a process that has a really clearly defined start, middle and end where the objective is always super clear and as long as you kind of point in the right direction, you know you're going to get there eventually.

Josh Muncke: You're managing something which is usually very exploratory, which has many different paths and routes. It can go down and on occasion might actually not return something of value. And so within data science, you need to figure out a way to keep the people who are doing that work motivated and pointed in the right direction even if there might not be a right direction, very obvious and also provide air cover for those people in the wider business if things don't pan out as people would have liked or hoped or expected.

Kirill Eremenko: Gotcha. So you need to prepare your team for that as well, build your team appropriately and prepare them morally, mentally, or for these differences and these uncertainties that are facing them.

Josh Muncke: Yeah, definitely. I think you need to help the team understand that not every data science projects is going to have a really clear, nice deployed product or output. You need to help the business understand that as well. And then you need to throughout the course and the duration of those projects, making sure you're making the best decisions you can do and helping the team make better decisions they can do to kind of keep pointed towards something that's going to be valuable.
Kirill Eremenko: Yeah. Gotcha. Well, before we dive further deeper into your work at Red Bull, I would like our listeners to get to know you a bit better and I'm very curious about your background because it's very similar to mine actually. You have a bachelor of physics, I also studied physics in my bachelor's. You worked at Deloitte and I also went through that at Deloitte. So it's going to be fun going through this. Give us a big bit of background, maybe just for my benefit. What kind of physics did you study?

Josh Muncke: So my bachelors of physics was a really broad degree. I ended up actually kind of specializing more in kind of nuclear and plasma physics. So the thing that I kind of wrote my bachelor's thesis on was on the confinement mechanisms of plasma, super-heated plasma in nuclear fusion reactors. If you ask me to remember anything more about it, it's broad and I think I would forget it. But that is what I studied.

Kirill Eremenko: Yeah, gotcha, I'm in the same boat as you. Though I remember the name of my thesis or probably not even the name, but yeah, wouldn't be able to dive deep into that stuff. But what I like about physics is it structures your brain in a certain way that then you can like, once you've learned something like nuclear physics is much easier to learn anything else, that you kind of have this confidence that you can master anything that's come about.

Josh Muncke: Right, I think physics for me, I never, I don't think I quite realized it at the time, but it's one of these things that everything we're doing is about the application of master to some kind of applied problem with the real
world. And so that is so true, the job I ended up doing, I don't know if that was intentional or accidental or just kind of a good luck, but yeah, I really think my education prepared well for that because it's really the idea of applying those techniques to get to some answer or uncover some insights about the real world. True in physics and true in data science.

Kirill Eremenko: Gotcha. And so how did you go from physics to being a consultant at IBM?

Josh Muncke: Right. Well, that's a funny story in itself. I basically was planning to continue my education and I was going to continue to do a masters and then maybe even beyond-

Kirill Eremenko: A PhD.

Josh Muncke: ... A PhD yeah maybe. And I was dating a girl at the time who had an expensive taste in handbags and I remember thinking, I've got to get a job if I'm going to be able to afford those handbags. But it was very late in the year and so I was kind of out of options for a lot of the most populated graduate programs that some of the big employers in the UK, IBM was one that had year round application. And so I went down to the IBM headquarters in Portsmouth in the UK, met the graduate recruitment team there, really, really excited by the role and the kind of foundation program that they had. I applied and was lucky to get the job.

Josh Muncke: So it was a little bit serendipity that they were still accepting applications that late in the year and ended up like I said, yeah, doing a three and a half years IBM in a team called Business Analytics and Optimization.
So that was kind of data science before it was called data science. All consulting working with a lot of different companies, really understanding how data is used at companies. And then that’s when I first started to look at data visualization and modeling as a way to solve problems in business as opposed to in academia.

Kirill Eremenko: Interesting. Do you remember your very first project?

Josh Muncke: Yeah. My very, very first project actually was a big clothing store retailer in the UK and they were doing a project called single view of customer. So they were trying to pull together all of their different data sources about their customers from credit card data, online eCommerce and customer service call centers to kind of stick together this profile, which they were then going to use for marketing purposes. And I remember my first day on the project, I had just come out of my graduate training at IBM, really felt good about myself and I was told to go and write this data test plan or something, completely bombed, have no idea what I was doing, ended up, sat down with the project manager and he said, "I don't think I was supposed to see this yet, was I?"

Josh Muncke: I remember feeling pretty bad about my choice of consulting career. But I think everyone feels like that their first day at work. So. Yeah. I mean that in itself was a great project. It was a great learning experience and had some fantastic mentors and managers that really kind of helped shape those early parts of my career. And ultimately where I am now.
Kirill Eremenko: And speaking of data science leadership, it's so ... Like especially in those early phases, so up to the manager or director to encourage, reassure the new graduate or analyst that's it's okay to fail, it's okay to learn because it can be so discouraging at the start.

Josh Muncke: I think that is something that is key and I mean that is key in any kind of leadership role, but especially in data science where you do have this iterative, exploratory, kind of work environment where things don't always go right. It's really important that the more experienced folks, less experienced folks know that sometimes things just don’t work out. And that's just the price of doing something which is ultimately kind of an innovation role that is exploratory in nature.

Kirill Eremenko: Gotcha. All right, so you did three years at IBM and then you move to Deloitte. What made you make the move?

Josh Muncke: I think that was just kind of a time you get to after about three years in your first job where I think you start to think about what could you do now and is there something else that could be interesting?

Josh Muncke: I really loved consulting and I loved the variety of different problems and projects that I've got to work on in consulting. I wanted to work for somewhere where there was going to be kind of less focus on the software and the tools specifically that IBM had a bit more focused on the business problem and the commercial side. And Deloitte offered that. So I joined a great team again with a great group of people and
fantastic managers at Deloitte in London, in the consumer business teams. So that was kind of like retail and consumer products. And yeah, that was after about three years and I was at Deloitte for ... I think you and I had this conversation about the same time, about two years, two and a bit years was my [inaudible 00:16:33].

Kirill Eremenko: Yeah, same for me, it was two years and yeah, it's kind of like these consulting firms, they usually have this unspoken rule two years up or out. And not to say that I cut out because I couldn't go up, but you just kind of like that two years or two or three years mark is when you kind of like sit down, reassess, like do you want to continue or is it time to move on of then you do another two or three years and again you reassess. I guess that how it works. Yeah, for me I realized okay, I've learned a lot, I love variety, I had a lot of things. Now I know what I want, now I know where I want to go and how was it for you like after two and a bit of years at Deloitte? Why Red Bull? How did that happen?

Josh Muncke: Well, yeah, I mean it was slightly different for me that my last project at Deloitte was actually at Red Bull so I originally came to Red Bull.


Josh Muncke: I was poached. I was kind of in a situation where as with consulting folks who are consulting companies will know you're incentivized to go out and do a project and then move onto the next big thing. And so I had done a couple of projects with Red Bull. Actually in
Austria, which is where Red Bull is globally headquartered.

Kirill Eremenko: Oh really, I didn’t know that.

Josh Muncke: Yeah. Lesser known fact.

Kirill Eremenko: Oh well.

Josh Muncke: Red Bull's global headquarters are actually in Salzburg just in Austria. And so I had done a couple of projects there and I was really, really passionate about the company and what we were building and wasn’t really ready to leave, just felt so strongly about the team there and what was being created. So decided yeah, after the offer came that it was kinda the right time to make a move, was really, really lucky that, that move was to Santa Monica, California, which is also pretty hard to say no to. So packed up my flat in London and I moved out here.

Kirill Eremenko: Fantastic. Well, and being at Red Bull ever since.

Josh Muncke: I’ve been at Red Bull ever since. So yeah, I’m nearly coming to three years now.

Kirill Eremenko: Wonderful. And so what was the position that you moved to? Were you joining a data science team or were you starting a data science team? Describe the environment, the circumstances at the time?

Josh Muncke: Yeah, so I joined as the director of data science and I was the only person in the data science department at that team, there was no existing team or department. There was no real strategy about what the data science should be at Red Bull. So that was kind of my first job.
It was to say what should data science be at Red Bull, what should we do, what kind of projects should we work on, who should be higher and what should we deliver? So yeah, it was an interesting few months, especially kind of going around just introducing myself to people as the new director of this department that they've never heard of. But that's what I always find exciting is having the opportunity and the sponsorship to be able to create and set something new that is really, really exciting, really motivating. And ultimately one of the reasons I came here was to be able to do that. And I'm lucky Red Bull gave me that opportunity.

Kirill Eremenko: I love it, I totally love their approach in like, oh, we don't have a data science department, we're not going to start by hiring an analyst, let's hire a director right away. Let's go all in. That's so like Red Bull like from what we see that adrenaline sports and stuff like very courageous, very straight to the point. We don't have a data science department, let's hire Josh as the director of data science. Wow, that's so cool. And what is your team like right now to almost three years later?

Josh Muncke: Yeah. So right now we're a team of four people. So I've got three data scientists that work with me. Three really talented folks at the ... I'm really excited to have hired and are still here, none of them left. And so we are working on projects at Red Bull from the openness in the beverage side of our business. So presumably everyone knows that we make and sell energy drinks. So we do projects with the sales team and the distribution team on the beverage side and we also do projects with the media side of our business. So with
Red Bull TV and RedBull.com. We also have those lots of events and marketing that we run to. So we do projects on that side. So we are still a pretty small team I think and especially considered the variety and the scope of projects that we're working on. But never let anyone tell you that a small group of people, if they're committed can't change the world is my motto.

Kirill Eremenko: That's very, very wise words. Okay. And so very interesting. Let's move on a bit into the work that you guys do. So you mentioned you're in two sides of the business, the beverage side of things and the media side of things. Could you give us an overview more, what I'm interested in is for our listeners, it'll be very cool to hear and there's plenty of fans. I'm sure there's plenty of fans of Red Bull listening to this. It would be really cool for them to hear kind of like an industry case study like maybe if you're going to share a project that you recently did or are the type of work that you do, the approaches that you have. Some specific case study if you will, to go into [inaudible 00:22:14].

Josh Muncke: Sure. Yeah. I as I said I think one of the really interesting things about Red Bull is just kind of very broad and diverse business that we have. And so as a data person, the ability to go and play in other people's back yards is really great at a company like that because it means there's a great variety of projects to do. And so maybe I'll give you two examples to kind of illustrate the scope of different kinds of things that we work on. So one project is kind of very core sales analytics. So as you probably know, we sell Red Bull at many different bars, clubs, restaurants across the
country. And so one natural question we might ask is, are there additional bars and clubs out there that are not selling Red Bull that maybe should be.

Josh Muncke: And so to answer a question like that it's actually a great machine learning question because we want to get to something really, really tactical which is the list of prioritized places that we're not selling Red Bull that we should be. And the inputs that are going to be things like what type of bar and club is this place, what are the demographics around that location? Maybe we can pull some data from external data sets like Google, like I said, demographic data is also helpful there and we're trying to build a model that basically is predicting the volume opportunity based upon our current set of bars and clubs, for bars, clubs, restaurants that we're not selling Red Bull at. So the output there it's not really a dashboard, it's not particularly sexy, it's something that we can hand over straight to the sales team, really a list of locations that we think would be a high priority places for them to go and see if they're interested in selling our product.

Kirill Eremenko: That's really cool. So you're using experience with your current data sets and like your bars that you have already and the geodemographics around them, the drive times, the profiles of those bars and anything else that you can find on those bars and then you're looking at the bars that you're not servicing and finding kind of like for like matches or it kind of like even a recommender type of system where you're looking at your existing data and trying to learn from
that to make predictions for the other bars out there that you have never ever dealt with.

Josh Muncke: Exactly. That's right. And if you know anything about the US, it's a huge country and the number of bars and restaurants is changing. And there's lots of turnover, right? So there's lots of new bars and restaurants opening all the time. So what we want to do is make sure that we're rerunning this model fairly frequently so that new bars and restaurants are brought in and we can prioritize them for our sales guys as quickly as possible.

Kirill Eremenko: Gotcha. And if you are able to share, could you let us know a bit about the model? What kind of a machine learning algorithm did you use for that?

Josh Muncke: Yeah, so it's actually an interesting project because one of the things that we wanted to do with this project is give the team kind of an opportunity to compete on model selection. So for this project, we actually ran a mini internal Kaggle competition. So we didn't load on Kaggle and open up to the public. A lot of the data we were using was proprietary but we actually set up a little test hold out set and we said, "Okay guys, over the next two weeks we will compete to see who can build the best model, the best supervised model to predict volume for these accounts." And so the model that ended up winning is quite often seems to happen at the moment was actually an XGBoost model. And, but really the beauty is in the features, right? So the winning model is actually the model that... where the data scientist that built it had taken some time to create some new

Show Notes: http://www.superdatascience.com/233
powerful features that were really productive and helpful in getting to that optimum easy.

Kirill Eremenko: Very interesting. I've seen that before as well, where you use XGBoost. It sometimes can even outperform deep learning algorithms. It's surprising, maybe because deep learning requires so much more data and so much more training.

Josh Muncke: I think XGBoost is still generally considered to be better for most structured supervised learning problems than deep learning. I think certainly for me, I would always go to like some kind of boosted or tree based model on a structured dataset before starting on something like deep learning. That's much easier to get up and running more quickly and you're probably going to catch up most of the value and not modeling problem with something like that without having to go to a deep learning approach.

Kirill Eremenko: Gotcha. As you mentioned, feature engineering, super important, right? The way you select your columns or parameters of this model, it's like how do you create new ones? How do you combine existing ones? Do you look at just the number of customers that go into the bar or do you look at number of customers divided by the drive time distance or the revenue that the bar is making multiplied by the average spending or divided by the average spending of the customer. Like kind of those types of things. And what I wanted to ask you is I find that when you use XGBoost or like recently I had an example when you use XGBoost and then you do feature engineering you end up with like, I don't know, maybe six or eight features which are very highly

Show Notes: http://www.superdatascience.com/233
predictive, but I find that it's very sensitive. As soon as you remove one of those features or you add a new one in, results can go completely change. Did you have that experience?

Josh Muncke: Yeah. I mean definitely with XGBoost, that is one of the things you'd expect. It's a tree based model, so it's considering a lot of interactions between variables and so making even small changes so the input data you put in are going to have pretty big outcomes in terms of the final predictions. I think a lot of people attempted to think of that feature engineering step as kind of just like a data cleaning process where you just kind of line up your training data set and you push it into your model and then what you get out is, or how you improve that is on a further tuning hyper parameters. And I think that's a shame when people do that because there's a lot of opportunity to be obtained by thinking cleverly and more like a human with your business knowledge about how to frame that training data set.

Josh Muncke: So for example, one of the features that ended up being pretty predictive here in this model was actually looking at the, for each bar club and restaurant, looking at the volume of other bars and clubs and restaurants around it, requires a little bit of little bit of like geospatial feature engineering, right? You have to kind of calculate those trade areas and you have to look at other places that are nearby and then calculate the average amount of volume that they're selling. And so to do that, it's not something that the model itself is going to automatically calculate for you. So you can
actually think and be clever about the way you set that modeling problem up and the data you feed into it and you're going to get probably better performance of your model by doing that.

Kirill Eremenko: Gotcha. I love that example because it speaks to the creativity that data science requires. I hear quite a bit of a concern that data science is going to be automated that companies like DataRobot that are going to edge out the data science and not to say that there's no room for services like DataRobot and automated data science. But still there is so much creativity involved unless you think about in advance and think of it as you said as a business problem, use your business knowledge and then go out there and put some effort to derive those additional features like the volume of the other bars around. The automated algorithm for data science will never actually even know that there is such a possible feature. It's not going to just go out there and understand how bars work and suggest that feature. It's just going to use what you're given and unless you think about it creatively and come up with this feature, you're gonna miss out.

Josh Muncke: I totally agree. I totally agree. I think the automated data science engines and things like DataRobot or even auto ML that definitely going to have a role in the toolkit of the data scientist. I really see the outputs of some of those things and you've got a very, very clearly structured and well frame problem with a nice clean data set and your output is all about predictive performance. I definitely see that those tools are going to play a role. Do I think they're going to do away with
the need for a data scientist you can creatively think about a business problem and the strategy of the company and then translate that into the data right by creating sensible features that make sense? I don't think so. I think that there will still be a need for that. Absolutely.

Kirill Eremenko: Totally. And then on other end as well you've got to have a data scientist who can communicate the result.

Josh Muncke: Yeah.

Kirill Eremenko: Right? That's the big part for you guys as well.

Josh Muncke: Yeah. Last I checked, DataRobot wasn't that good at standing up in front of the board and presenting their results in front of a skeptical sales people.

Kirill Eremenko: Yeah. Yeah. All right, cool. So that was a wonderful example. Thank you so much. And you mentioned you have two case studies. What was the second one?

Josh Muncke: Yeah, so the other example is kind of right in the other side of our business and is something that you will almost certainly be aware of this type of problem which has recommendation models. So we have Red Bull TV, which is a fantastic repository of content. You can watch it on your phone, on your laptop, on your apple TV or other device and we make a lot of great content and we put it out there for people to watch and enjoy and consume and it's free.

Kirill Eremenko: Wow, it's free. Everybody listening, it's free.

Josh Muncke: It's free.
Kirill Eremenko: Download it now. I was expecting that it's going to be like Netflix.

Josh Muncke: No.

Kirill Eremenko: How come I don't have that? I'm getting it right now.

Josh Muncke: Yeah. Everyone listening to do me a favor and go and sign up for Red Bull TV, get an account and let us know what you think. So one of the problems that we actually never implemented on Red Bull TV previously was recommendations, right? And so that's a very, very well told story is how can you use algorithms to better present what kind of content you put in front of someone and specifically what the problem we were interested in solving was content to content recommendations. So how do we find content that is similar to other content? So that when is looking at one piece of maybe downhill mountain biking videos, what else should we show them to potentially watch next? That was previously a problem that was always solved by humans at Red Bull, always done by kind of editors manually creating lists and we we're able to show the power of kind of algorithms to help find additional similarities in our content and put those recommendations in Red Bull TV.

Kirill Eremenko: Interesting. So tell us how do you actually go through this process? Because I imagine it's like video content. Do you like use the metadata? Do you use some NLP to get the text out of the images or do you use some computer vision? How do you get into what's in that video?
Josh Muncke: Yeah, I don't think I can go too much into the nitty gritty of it, but I will say that you're on the right track.

Kirill Eremenko: Okay. Gotcha. Gotcha. Well, yeah, as we move forward into the world, it becomes more and more advanced and yeah, I heard like a couple of years ago, I actually heard that Google had plans to ... You know how like when you search for something, you are recommended pages on the web, but videos only if the title of the video has it. But Google had plans to actually go into the spoken text inside the video and pull out information from their [inaudible 00:34:01] wouldn't be surprised.

Josh Muncke: So one of the areas I think has been really, really productive for deep learning and AI models has been how do you get data out of places that were previously not considered data, so all that unstructured data like raw, transcriptions or video content pitches were previously kind of taking up space on people's disk drives and cloud server, but not really able to be analyzed in a way that could actually be then used to drive a decision or an action.

Josh Muncke: And so one of the things that Google for sure many of the companies and Red Bull is finding is that actually starting to apply some of these text, image, audio, video analytics techniques on that data, you're able to extract a huge amount of really, really actionable data from them that can then be used to drive things like recommendation or search products. So there's been an amazing transformation in the industry just in the last, call it 5 to 10 years. And it's proven really, really
valuable for companies that are now getting stuff out of that previously unavailable data.

Kirill Eremenko: Gotcha. I actually read an article recently about recommender engines and wanting to get your thoughts on this. So I heard that there's two types of recommender engines and often they're combined. So one is where it looks, as you described, it looks at the content and looks at similarities between the content to recommend to the user. So if somebody liked I don't know, Stephen King movie, they might like Stranger Things like the TV show because they're both like kind of scary horror and stuff like that related. So there's a relationship between the content is like a network between the content that the algorithm taps into.

Kirill Eremenko: Whereas the other one is, it looks at similarities between the users. So if, for instance, I liked I don't know, a movie like Lion King about the cartoon but then I have somebody that's, maybe I don't know, but they're similar to me in terms of the geography, the kind of like transactions that they perform on the website or any other data that's available on the person. And they have never even watched the cartoon, they've never watched like Pixar movie or anything like that, but because of the similarities, they might be recommended the content that I've seen. So and that pops up completely different recommendations. What are your thoughts on that? I don't expect you to go into detail whether Red Bull uses either of those or the second one, but just what are your thoughts on the differences in the power of the two types of recommender system?
Josh Muncke: Yeah, I mean, I think it’s a really interesting space and there’s loads of great research that’s been done on this. One of the way I typically see the split is you’ve got kind of like content to content where I’m looking at which content is similar to other content. You’ve got kind of like a user to item, like user to content models and those are gonna be kind of like your more standard collaborative filtering type models where you’re kind of saying like, other people who watched or voted this tend to like this other piece of content that you haven’t seen yet. The tradeoffs there are kind of interesting because those collaborative filtering models are great and kind of really unpick. Not just good recommendations, but also these really interesting vectors of users and tastes where you can kind of look at the results of the Matrix factorization and kind of say, hey, these are the kind of types of users or types of contents that we have.

Josh Muncke: But after you do that Matrix factorization. So those give that really nice understanding of the interaction between your user and your content. But they’re not very good if you get a brand new piece of content, right, because no one's watched this, so how do you recommend it? So there you need something that's going to be content based where you can actually say, hey, this content for whatever reason, based on whatever characteristics is similar to the other piece of content, therefore this is how we're going to place it. What I think is really interesting is now the application of deep learning techniques to recommendation where the really advanced approaches are actually combining
kind of content based with behavioral based with kind of like personal features or personalized features and information about the users to produce really, really like granular recommendations that are really high performing. So that is a really interesting area of research. And I'm pretty sure that you can guess that folks like YouTube are using stuff that is state of the art in deep learning for recommendation.

Kirill Eremenko: I recently checked how many research papers Google published this year in 2018 on this stuff like it's 434 research papers on just AI, machine learning, computer vision.

Josh Muncke: That's wild.

Kirill Eremenko: Yeah. It's like more than one per day if you think about it, ridiculous it's like a printing press for research papers. Crazy.

Josh Muncke: Yeah, it's crazy.

Kirill Eremenko: Okay. Okay. That's very cool, fascinating topic and thank you very much for those case studies. I'm sure a lot of people will get some great ideas, guidance out there. I wanted to switch gears a little bit and talk about, we mentioned data science leadership. I want to talk about mentoring. When you were in DataScienceGOx, we had this exercise where during one of the lunches, the lunch on Sunday, I think it was, no the lunch on Saturday, the DataScienceGOx at [inaudible 00:39:39] where we had, I think over a dozen of leaders and directors and business owners would go to the DataScienceGO conference the main event with 300 attendees.
Kirill Eremenko: And you guys were placed into different tables to mentor the audience or mentor the attendees who were at your lunch table. How did you find that exercise? Because like I've had so much interesting feedback from many, from both sides. Tell us a bit about that and in general, because I know like I've read a bit about mentoring and there's been some exercises where companies have sent their teams to Red Bull to get mentored. So I'm assuming you have some experience. What are your thoughts on mentoring in [inaudible 00:40:23]?

Josh Muncke: I think it's incredibly important and it's not just limited to data science. I think mentoring is one of the most ... Or finding a good mentor is one of the most important things that you can do for your career. And I think that applies whether you're at the beginning of your career, halfway through or towards the end. The exercise at DataScienceGOx was excellent. It was really good. I had some great conversations with some folks that were kind of pretty new to data science and we're trying to figure out specific problems that they were working on at their companies or become more generally just how they get started and what they were supposed to do to find their first job. So I thought it was great. I really enjoy that kind of exercise.

Josh Muncke: I think it's important for us folks who are a little bit more experienced in the data science world to make sure that we are out there and making ourselves available and giving back to the community for those junior people that are just getting started. So and it's something I feel really passionately about. I think it
can be incredibly valuable. You're ultimately helping kind of the next wave of talent come up and one day those people might be applying for jobs at your company and say you want to make sure that you really give back and mentor where you can because I think it's a good thing to do.

Kirill Eremenko: Yeah, and that's the feedback I've heard around that people who have some experience in data science are so passionate about giving back to the rest of the community and helping others grow. I honestly don't really know why it's so ... I haven't seen this in other fields. It's very pronounced in data science, maybe it's due to the steep learning curve, once you get up the learning curve, you're like, oh wow it's actually, it all makes sense. Let me explain it to somebody.

Josh Muncke: Yeah. I listened to the podcast that you did with Kristen Kehrer and Kate Strachnyi a little while ago, and those guys are just inspirational in terms of the amount of mentoring that they do and the amount of give back they do. The blog postings they write, the training courses they create, the books they're doing so much like inspirational stuff and tend to give back. I'm not that good at that stuff, the really public platform stuff. But I do think that it's important to give back. And so one of the things that I've done a couple of times that I've really enjoyed is going to judge at hackathons, there's one at UCLA called DataFest.

Josh Muncke: It's pretty popular and I was a judge out early this year and I think those kind of events are great as well because those are also people that are new in their career. Given a data set and 48 hours to go and find

Show Notes: [http://www.superdatascience.com/233](http://www.superdatascience.com/233)
something interesting in it. And being there to kind of mentor and judge those kinds of events are really, really good experience and maybe doesn't involve for the people like myself who aren't great at writing for public, doesn't evolve the scariness of putting yourself out on the platform.

Kirill Eremenko: Gotcha. And what would you say is your most common advice that you give to people who are starting out into the space of data science?

Josh Muncke: That's a hard question. I think the one that I find myself saying most frequently is you've got to go and find real world projects. I think a lot of people who do, they decide bootcamps and online courses. Those are great and those are a great start to your career as a data scientist. But for a hiring manager or a leader, you're pretty aware that most of the problems and the projects that you work on those types of course are pretty artificial. Their structure, the data is usually set up pretty nicely, you've got a fairly concrete metric to train to. And so I think one piece of advice that I always find myself giving to junior folks is go out and find projects that you're passionate about and being passionate is important because it means that you're going to see it through, but also that are real world projects, right where you actually maybe need to go and be creative about how you obtain the data.

Josh Muncke: You need to think carefully about the features you haven't got kind of like a cheat sheet on what features to create and where there are real tradeoffs between the different types of model you use. That is one piece of advice I find myself giving a lot because I think it's
much more impactful as a hiring manager to see projects where someone's actually gone out and solved a real world project where things aren't pretty, than it is to see kind of a project that was solved as part of a bootcamp or an online course.

Kirill Eremenko: Gotcha. And similarly, when people go out there and find something of interest to them, like at the DataScienceGo we had, Nadieh Bremer presenting how she ... One of the projects she's done is she took the Lord of the Rings books or movies and then just analyzed like in which movie, who got to speak and how many words they said and build a visualization around that. And it's not going to change an industry. Is not really like, it's not a business problem but somebody who has that passion about a certain topic and then they apply data science to it, it really shows that not only can they wield the tools and make those insights happen, but also there’s believers in data science that apply to things that they just consider their hobby.

Josh Muncke: Right. Yeah. I think it's just important to see that someone cannot just write the commands to build a regression model, but that they can actually think creatively about the ways to apply those in the real world. That's really what doing those projects are all about. And so yeah, I mean, at first I want to say I've a huge data science crush on Nadieh. I think she's amazing and the work that she does in data visualization is just unbelievable. Hers like many other example is people who are passionate about the field and the domain of data science and are able to kind of
translate that passion into something which maybe it doesn't change the world, but actually really shows these techniques that we have, this field that we work in can give really powerful answers to sometimes pretty difficult questions.

Kirill Eremenko: Yeah. By the way, did you get to catch up with her? Because I remember you mentioned.

Josh Muncke: No, I didn’t.

Kirill Eremenko: So bad. Sorry about that, I should have introduced you guys. I'll make sure to make the intro somewhere else. Yeah, that’s really cool. It’s good to catch up with people who inspire you, right? Meet them in person or even over email.

Josh Muncke: Surely yeah.

Kirill Eremenko: So that’s really cool. Thanks for the tips on mentoring. And there’s some other topics that I want to cover from like... you know of choice paralysis but before we get to the end of our podcast, I guess one thing I would like to get your opinion on or thoughts is something that you mentioned that you’re quite interested in is data science and the decision making process. Could you tell us a bit about that? What are your thoughts on how data science impacts the whole decision making process within a business?

Josh Muncke: Yeah, I think this is so interesting because a lot of data scientists, when they're first brought to a company kind of make the mistake of thinking that the whole data science process is really focused around the data. So I've got to get to another data, I've got to build
models and that's kind of like the output of my work and I think that the disillusionment then comes when you see the outputs of those models is not then used by the business or ends up being kind of like either ignored or discarded. And so for me, what I always talk to my team about and really anyone I mentor is this idea that you need to think less about the data and the model, but more about the decision that needs to be made. So there's actually some teams in some companies that are resigned to reframe data science into decision science.

Josh Muncke: And one of the people here who is really, I think leading the pack in terms of just best practice and what is good really look like is Cassie Kozyrkov who's at Google. She's the chief decision scientist.

Kirill Eremenko: Oh yeah. I watched her talk. I don't remember, I think it was a Ted Talk.

Josh Muncke: Amazing.

Kirill Eremenko: Yeah. So good.

Josh Muncke: She's done a Ted talk and she's got some fantastic articles and podcasts that she's done. And what she says about this whole decision science thing is that the problem is that when you see data, you can't help but be influenced by it. So you need to think at the beginning of a project with your business stakeholders and asking them what would your default decision be if you didn't have the results of this analysis? What would you do? What would be the targets for either accuracy that you need to set or model predictive
performance or outputs before you can make a decision one way or the other?

Josh Muncke: And so by doing that, what you do is you set kind of like a framework by which as the data scientist when you do your analysis, you then know what kind of success looks like, right? So that you can then kind of say, when I’m building this model, or doing this analysis what am I working towards? And then you've got those kind of fixed set of goalposts as opposed to having something where I think a lot of people in data science will have seen this idea like, okay, build the model and I’ll tell you what the decision is, yes or no. Once I see the results and it’s like very, very hard as a data scientist then because like how do you know if the results of the outputs of what you’re doing is ever really going to drive any kind of decision in the business.

Kirill Eremenko: And adding on to that I would say also a lot of data scientists don’t consider this whole process of integration of their findings, of their models into the business. Data science projects used to be more kind of one off, all right, let's find the insights, what's going on, let's do this thing and okay, let's inform a decision. But more and more they're becoming ongoing thing. So where you deliver a model but then it has to be deployed into the business and it has to be developed and it has to be integrated and then it has to be maintained and so on. And that sets a whole new part so like supporting these ongoing decisions constantly. And I'm sure like you mentioned this with your model, the first case study that you've carried that you have to
retrain it with time, right? Otherwise new stores come into the world, new bars and also the model might deteriorate over time. So that's another thing that people need to keep in mind as well.

**Josh Muncke:** Yeah, definitely. I think the difficult thing is really making sure you're clear on what kind of decision that needs to be made, right? Is this a decision that is kind of like it's a one off decision and we just need to know the answer and that could be a prediction, like a predictive decision or it could be an inference, right? I actually need to look at the coefficients in this model to understand the strength of some effect. And that's one type of analysis. Another type of analysis is going to be more like what you said, where actually, I need to make this decision many, many times in an automated way ongoing and that will probably require a different kind of approach, potentially a different kind of model. And certainly that model management and maintenance once the model is deployed for the first time to make sure that that decision that is being made by the model continues to be the best decision that can be made. And those are things that you want to know before you start the project and not find out at the end.

**Kirill Eremenko:** Yeah. True. And all of that ties into something else since you're quite passionate about is asking good data questions. Well, how do people ask better data questions? Because that's such a common issue that I've seen hundreds of times where people just hand you the data, like find me some insights or ask you a question and then halfway through the project they...
realize they were asking the wrong question. What advice would you give to business leaders and data scientists to agree on the questions that started for the first party to ask better questions and for the data scientists who guide the business leaders into asking the good data question, what are your tips there?

Josh Muncke: Right. Yeah, I think there's a few things, like you said, I'm really passionate about asking good questions. I think it's kind of the trick up the sleeve of the data scientist is as you said to themselves, ask your questions and to coach the business into asking good questions and I think there's a few things you can do to really make sure that you're doing your best to achieve that. One of those things is kind of my secret weapon, which is to ask who is going to do what with the answer to this? Right?

Kirill Eremenko: That's so good, that's so good.

Josh Muncke: Right? Because it really forces whoever that business stakeholder is to kind of say like, okay, who is the stakeholder that's going to be making the decision, what are they going to do with the answer to it? Because too often what you find is that the question that you're framing up is actually being framed by someone who isn't going to be using the answer. Right? So if you're building a model that's going to go to sales people that are out in the field selling cases of Red Bull, and that question is being posed by the head of sales. Well, the likelihood is that he may have misinterpreted the needs of those people, right? And the needs of their answer.
Josh Muncke: So you want to find people that are out representative of the answer, the people that are going to consume the answer to that to be in that project with you. So I think the first part of a good question is that just figuring out who's going to do what, what with the answer or the output? The second thing, which I think is I kind of stole it and I'm sure you've heard it, Smart Targets, right?

Kirill Eremenko: Yeah of course.

Josh Muncke: I think you can translate that to smart questions, right? So you can think about the questions that you're asking your frame of your data science project and in this kind of smart framework. So are they specific, right? Do they relate to something that you can really put your finger on or are they kind of more general? Of course it's data science, right? So they need to be measurable. And Mr. Measurable, if you can't measure the thing that you're trying to ask a question about, really difficult to do any data science on them, everything needs to be actionable. Everything needs to be actionable. That's why we're doing it. We're by and large and mostly applied data scientists not research people. So we're looking for something where if we get the answer, we can actually do something with it. You want to have something that's realistic and realistic care can take a number of different dimensions.

Josh Muncke: But realistic is for me, means can we actually make this decision if we actually get this answer, can we actually make the decision? Do we have the organizational mandate, do we have the sponsorship,
do we have the ability with our consumers to make this kind of decision if we get this answer? And then T, you want to have some kind of timeframe, right? So when do we need a decision by an what timeframe are we doing this analysis on to make sure that we're clear that is this a previous 30 days analysis or is this a previous 5 years analysis? And that's really important to note before you actually start doing the work.

Kirill Eremenko: Love it. I love the adaptation over the Smart Targets to data science and I never thought of it that way.

Josh Muncke: Yes. Smart targets, smart questions.

Kirill Eremenko: Smart targets, smart questions. Awesome. Well Josh, we'll leave it at that. Thank you so much for all the wisdom and the insights. Before I let you go, where are the best places for our listeners to get in touch or follow your career so that they can learn more things from?

Josh Muncke: Yeah. Like I said, I'm not great with public promotion so there's no blog I have unfortunately, but I would be more than happy for anyone is interested in getting in touch to please reach out to me on LinkedIn. Send me a message whether you just want to chat, whether you want to meet up and go for coffee or you're looking for a job, just get in touch. And I'd be more than happy to have a conversation with anyone that's interested.

Kirill Eremenko: Fantastic. Fantastic, thanks Josh, and one final question for you. Is there a book that you can recommend to our listeners that has perhaps changed a career or life that you think would be useful for them to read as well?
Josh Muncke: There have been loads of books. One of my real favorites was a Thinking, Fast and Slow by Daniel Kahneman. So that is a book all about how humans make decisions and some of the fallacies that we maybe make or that we don't realize we're making as we make decisions. So I would really, really encourage data scientists to read it because it opens up the world of understanding about how people make decisions and potentially some of the incorrect things that people do when they do make those decisions. And as we talked about, decision making is one of the most critical things for a data scientist to be able to understand and influence.

Kirill Eremenko: Gotcha. Okay. There we go, so it's Thinking, Fast and Slow by Daniel Kahneman.

Josh Muncke: Daniel Kahneman.

Kirill Eremenko: Daniel Kahneman. Thanks so much Josh for coming on the show, being amazing, really enjoyed our chat and I'm sure lots of people will get very valuable insights.

Josh Muncke: Thank you Kirill.

Kirill Eremenko: So there you have it, ladies and gentlemen, that was Josh Muncke, Director of Data Science at Red Bull. I hope you enjoyed this conversation as much as I did. It was so cool, of Josh to share two case studies of how data science is applied at Red Bull and hopefully you are able to extract some examples of industry applications of data science from that. And another important topic that we covered off in this podcast was data science leadership, an extremely important area.
to focus on for businesses as we go more and more into the world where data science matures and it becomes a function. A separate function within business.

Kirill Eremenko: On that note, make sure to connect with Josh. You can get the URL to his LinkedIn and all the show notes at www.superdatascience.com/233. That's superdatascience.com/233. And there you'll also find the transcript for this episode, any materials we mentioned as well. And if you know anybody who's in data science leadership, who is a leader in the space of data science, a manager, a business owner, a director in the space of data science and is interested or might benefit from knowing and learning more about data science leadership, then send them this episode, forward this episode and help them get these insights and maybe after this podcast, connect with Josh and brainstorm some ideas about data science leadership. On that note, thanks so much for being here today. I look forward to seeing you back here next time. And until then, happy analyzing.