

SDS PODCAST EPISODE 343: CAREER JUMPSTARTS THROUGH DATA SCIENCE RETREAT



- Kirill Eremenko: This is episode number 343 with Founder and CEO at Data Science Retreat, Jose Quesada.
- 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 everybody super excited to have you back here on the show. Today's guest is super interesting, super exciting, I had a wonderful chat with Jose Quesada. So there's this retreat called the Data Science Retreat. Really it isn't a holiday retreat, it's an intense training program in Data Science. It was actually founded by Jose Quesada.
- Kirill Eremenko: So Jose is a serial entrepreneur, this is one of his most recent ventures. So what it's all about is for 3 months you go to Berlin and you get intense data science training in person. Something that is tailored for you to get the data science job of your dreams after you graduate. So for two months you get lectures and tutorials run by experts in the fields of natural language processing, computer vision, deep learning, classic machine learning and many more fields related to data science.
- Kirill Eremenko: And then for the final month you do a project. But not just some sort of random project, a very cool project that you are personally passionate about. And once you're done with that you take that to job interviews



and you get hired. So the Data Science Retreat have astonishing rates of their graduates being hired. About 86% in the first three months and about 96% in the first six months after graduating.

Kirill Eremenko: Very cool program, you can check it out at datascienceretreat.com/sds, so if you go to that URL you will get a special discount that Jose has set up for our podcast listeners. Note that SuperDataScience does benefit as well if you follow that URL. When you're checking it out make sure to also specify in the field that 'How did you find out about Data Science Retreat?' specify SDS or SuperDataScience to get your special discount.

Kirill Eremenko: So in the podcast you'll learn all about Data Science Retreat, how Data Science Retreat is leading the space of data science education in Europe by offering something that is called Income Share Agreement, so with this option of an Income Shared Agreement, you don't actually have to pay for your tuition upfront, you only pay after you have gotten a job. How cool is that? So find out more on the podcast, that's what we'll be talking about in the first twenty minutes.

Kirill Eremenko: And then we'll move on to very cool projects so even if you are not going to be signing up for the Data Science Retreat, you will learn a ton from this podcast. So for instance, Jose will be sharing his views and advice for picking a portfolio of projects. He will share five steps that are very important in picking and building a portfolio of projects and you'll learn how to do that on your own. Plus, he will give us some very cool practical examples of projects that came out of the Data Science



Retreat, such as the wheelchair project, in the streets of Germany, the malaria microscope and a robot that picks up cigarette butts.

Kirill Eremenko: So that is what this podcast is all about, you're going to have a lot of fun, you'll get some really cool examples of data science applications in action. So without further ado, let's welcome Jose Quesada, Founder and CEO at Data Science Retreat.

Kirill Eremenko: Welcome back everybody to the SuperDataScience Podcast. Super excited to have you back here on board. Today's guest is a very special guest calling in from Toronto, Jose Quesada. Jose, how are you going today?

Jose Quesada: Everything is great on my side, how are you Kirill?

Kirill Eremenko: Very good as well. Thank you so much. And how's the weather in Toronto these days?

Jose Quesada: Oh, it's been snowing. February is pretty tough the winter is tough here, but probably so is... I mean in Berlin it's pretty much the same about now so not a big difference. And just to tell your audience I fly around between Berlin and Toronto right now so that's why I mentioned Berlin.

Kirill Eremenko: Got you. Interesting because I was talking to somebody from New York a couple of days ago and oh yeah a couple, maybe a week ago. And they said that the first time it snowed in New York this winter was in February but in Toronto you're not seeing that Toronto is-



- Jose Quesada: No in Toronto there was snow before yeah. January and December. I mean this is Canada after all so it snows.
- Kirill Eremenko: Nice. Is there skiing in Canada near Toronto?
- Jose Quesada: I'm sure there is. I haven't been skiing but I'm not a skier anyway so it's probably not a very common so there are not that many mountains I think around the area.
- Kirill Eremenko: Okay. Around that area. Okay, got you. Well Jose, super pumped to have you on the show, super excited. We got introduced through a common connection somebody I met at I think a business mastermind or something like that and yeah, the work that you're doing is fantastic. You're in the space of AI, data science I don't even know where to get started. You have this AI Deep Dive, Data Science Retreat, a deep learning retreat something that you had running before, so many things. So maybe give us a bit of an overview, how would you describe all the different projects that you are working on?
- Jose Quesada: So first of all, thank you for inviting me to the podcast. I know this podcast that you've been running it for such a long time. It's very high quality, so I love to be here. So let me give you some background about the things that I've been doing. So all these are bootstrap companies and my main idea is that it's really a shame that not more people are doing machine learning nowadays because machine learning gives you an incredible amount of leverage to solve real problems. And this is something that people have not seen. My

gut feeling is that businesses think that they need to be Google to get value out of machine learning and this is not what I see. So in the schools that I've been running, you see people building amazing stuff in three months and they didn't know that much machine learning when they came in, they learn on the fly and then they found an idea and implemented the solution in the time that they had.

Jose Quesada:

So if people can do it in three months why are companies not doing it? So this is the biggest motivation for me to teach as many people as possible to be effective with machine learning because then they can solve real problems. So technology is always giving you leverage and every technology makes people more powerful, but if you compare machine learning to any previous technology in this century it makes them pale in comparison. For example before Ruby on Rails you could not really do very dynamic websites and then Ruby on Rails came and everybody felt the power of being able to create websites. So many start ups were basically Ruby on Rails projects that solve real problems and this made people money. That's a consequence of solving real people's problems. Then apps came out and of course many companies were built on top of apps.

Jose Quesada:

It's a totally different environment. You have more sensors, you have more data but I think the next wave, which is machine learning makes websites and apps look ridiculously underpowered in comparison. So we can build machines that can see the world, that can identify objects, that can understand your language as

you can see when you interact with things like Alexa. This amount of power is actually in the hands of anybody with a computer. So you can literally be in your kitchen table with a laptop running a deep learning model that is detecting something that is needed for you to solve a problem because of open source libraries and pre-trained models. You can literally be in your underpants coding in your kitchen table and solving really important problems that were impossible to solve just five years ago. So I think this is where things are going and this is why I'm so passionate about teaching machine learning and deep learning.

- Kirill Eremenko: Yeah, and I like your analogy. In one of your videos you gave a great example that you can think of artificial intelligence as somebody who knows how to use the power of artificial intelligence is like a cave man who has access to fire. The fire you can draw on the wall, you can cook meat, you can scare away animals that are hunting you. It's a very strong advantage compared to what everybody else is using and on top of that, it's not that hard. Artificial intelligence and deep learning, machine learning creation of those models is becoming easier. What would you say to somebody who might be listening to this and thinking that, "Oh, well I don't have a technical background and I don't know programming."
- Jose Quesada: Yeah, so the good news is that the language that people use for machine learning nowadays, Python is very simple to learn. And of course you have to be pretty dependable with Python to be able to create

products with machine learning but it's not impossible and that's a big deal. So there are so many minority groups that will be otherwise disempowered to do anything of this relevance but just learning python and learning machine learning it's really at their fingertips. It's nothing impossible you don't need to be coming out of a prestigious university. You don't need to have powerful computers anymore, you can do things online.

- Jose Quesada: I think that we may be at the start of a time where society in general benefits big time from machine learning and this is thanks to open source libraries and pre-trained models. Of course to train a model that does something very sophisticated you need a lot of compute time but people who train those models, they polish their models and you can download them and just cut the last layer and retrain them to solve a similar but not exactly the same problem. This is very valuable. In the analogy with fire, this is more not only fire but somebody will build a steam engine and give you all you need to build more steam engines if you just bother to download the recipe.
- Kirill Eremenko: Yeah. Got you, very interesting. So how exactly do you help people? What are your businesses? Tell us about Data Science Retreat and AI Deep Dive.
- Jose Quesada: Sure. So we take people who already know a bit of programming, a bit of machine learning and we help them get to the next level so they can join the AI industry. And we do this in a way that it's very concentrated in three months and we ask them to go through 270 hours of tuition.



- Kirill Eremenko: Wow. That's a lot of tuition.
- Jose Quesada: Yeah and the people teaching are practitioners from the field that teach only the one thing that they know very well. So they come for two or three days, they teach the one thing and then they leave. The next person comes and does the same and if you do that over a few days like a month and a half or so there is no way around, you're going to get much better. But we don't let you go just with that, with tuition that's not how we operate. So the core part of the project is to create a portfolio piece that is your master work. So you can go to an interview, drop this on the table and point at it and say, I made this and that should be the end of the interview.
- Kirill Eremenko: I love it. That's awesome. So six years that's a very long time and 200 people. You've probably seen a lot of projects come out of the Data Science Retreat. Do you have a number how many projects actually have gone through this program?
- Jose Quesada: Yes I do have a number, an exact number. It must be pretty close to the number of people and 2018 where we started doing team projects. So it must be about 200 as well maybe in the last year less so because they were doing projects in teams. One thing that I've learned is that team projects go further and we highly recommend people to do team projects because the final product is more finished and it looks better in demo day. So we have a demo day at the end where they present to the companies and to the public.



- Kirill Eremenko: So it's kind of like Y Combinator, right? They do team projects as well.
- Jose Quesada: Yeah. So the Y Combinator people come with an idea that is often already a company and they may even be making money at DSR, Data Science Retreat they want the job at the end so they don't care so much about the product. So they go crazy and do a passion project that doesn't necessarily aim at being a company. But yeah, the demo day part is kind of the same, we invite companies the companies want to hire them they don't want to fund them. That's the difference but we do want them to start companies and in the history of the DSR three projects became startups and one of them is still around, maybe two of them are still around.
- Kirill Eremenko: Very cool. And how many people got hired?
- Jose Quesada: Oh, everybody gets hired eventually. So I think unless you go to a city where there is no data science whatsoever then everybody gets hired. So I think the numbers are something like 86% within three months and 90 something high in the [inaudible 00:17:45] that these are all numbers after six months.
- Kirill Eremenko: That's amazing. And before the podcast you mentioned about the ISA or income share agreement, I want to understand this a bit further. First of all how does that work and does that mean there's no upfront cost for people to actually join this program which sounds quite crazy?
- Jose Quesada: Right. So this is very interesting. So in Europe, everybody pays full price up front, everybody up to 2018 or so where we partnered with a third party, a

FinTech company that will offer a contract where they're so confident that our people are going to get good jobs, that they will pay their tuition in advance and then get them to repay them once they get the job. So you pay more, they charge I know maybe 13% or something like that but it's not a loan, it's very different from a loan. It has a lot of downside protection. For example, you only pay proportional to your salary, I think it's 10% of your salary. So if you are making zero then you pay zero. So you can be looking for a job for say six months, eight months, whatever where you want to hold it so to get a better job, they don't send you a request to pay until you get the job. And then if you don't get a job paying more than I think it's 30K or 40K I don't remember then you pay nothing so you have to get a good job not [inaudible 00:19:25] one.

Kirill Eremenko: So if your job doesn't pay you more than 30K then you don't need to repay anything.

Jose Quesada: Exactly. And also after five years, the contract disappears. So I'm telling you this from memory, so I'm maybe getting confused. So you have to definitely go to chancen.eu that's the name of the platform that does this and figure out the terms because I actually don't know the terms too closely but this is the way it works. It's really designed so that there is no burden on you so we, the company training you invest in your future. So the more successful you are the easier it is for us. So you start paying faster the better the job you get is.

Kirill Eremenko: Interesting. So ISA as you had mentioned before they are getting more popular in the U.S and we talked

about this a bit. There's a school or actually a proper university in San Francisco, which I heard about on the what was this? Actually on the Y Combinator Podcast founded by Jeremy Rossman and they have this school called Make School. It's a proper university with a campus physical location in San Francisco where that's how they teach. You go to university, you don't pay anything, which is crazy for the U.S where we've all heard the stories of the hundreds of thousands of student loans when people finish university. There you pay nothing for your whole degree, once you graduate, once you have a job, you pay some percentage whether I don't know 8%, 10% of your salary on going until you pay off your loans. So I love this approach and this is why.

Kirill Eremenko: Because the teacher's teaching you when the setup arrangement is like that, when you are not taking, as a student you're not taking on all of the risk, the risk is on the university, the teachers teaching you are incentivized to design much better curriculums. They're incentivized to make sure you're doing your tutorials and homework, make sure you're keeping up with class to give you the best knowledge and equip you the best way to get a job. Not just give you that theoretical fundamental knowledge that is good, is great but it's not applicable in real life. They prepare you for the real world, otherwise they don't get paid at the end of the day. So the incentives are aligned.

Jose Quesada: Exactly. So for me, this is revolutionary. It can make things happen that are unprecedented so social mobility for example is really a topic for me. So

imagine that you have no access to loans, you are in a part of Europe that is not doing very well economically let's say the South of Spain where I'm from. So there are people there who have a PhD in physics or in math or in engineering and they are doing not much and there are people in Madrid that are getting paid 1000 Euro per month and they have the skills to be doing much better. So they could just teleport to Berlin, go through our course in three months and pay nothing upfront and then once they get the job they start paying. So the job that they're going to get so much better than the one they had back home or maybe they have none because unemployment rate in the youth percentage of the population in Spain is ridiculous. It's like 30% overall and 50% in young people.

Jose Quesada:

So I think it's a fantastic opportunity and it's really tragic that most people don't know about this. So I say it's habitually unknown in Europe and the fact that we can provide access to everybody in Europe so that's like what 550 million people in Europe? Anybody in Europe can apply for an ISA and because of the reputation of the school, they are going to get it. So our partner is very happy to give an ISA to pretty much anybody we send. So if we interview you and you pass, we do two technical interviews. If you pass, you basically have changed your life already, assuming that you can get through the course and graduate which is so far very likely. So this is my initiative right now so I want to go all over Europe and tell people that this is happening and this is why I thank you Kirill for

having me on your podcast so I can communicate this message to more people.

Kirill Eremenko: That's awesome. Thank you very much for coming and sharing this message. I wish more people were doing this, this is really cool. And yeah, what I want to talk about next is let's talk a bit about how this program is actually structured so that people who are maybe far away from Europe will never be able even to have the chance to join this can we get some value. Maybe there's some things you can share already that people can apply in their own learning or in their own approach to projects. So how does it all get started? Somebody joins and they have to pass some interviews, assuming they pass interviews how big are the groups of people that go through this program?

Jose Quesada: Oh, it's pretty small. From maybe eight to 14, I think 14 is the maximum we've had. So it's very boutique, so we don't have big rooms or anything because it's so intense to work with every single one of them until they produce a portfolio project that is of the quality that we want. So people come from literally all over the world. We have people from Russia, from Japan, around the U.S from all parts of the world that are not Europe, but let's say like 80% are from Europe.

Kirill Eremenko: Okay, got you. And then you said 270 hours of content. So somebody comes, an instructor comes in and trains it. What are some of the topics that you include in your curriculum?

Jose Quesada: Right. So it's mostly nowadays is a lot of deep learning. So there is a lot of computer vision, a lot of NLP,

classical machine learning it's still important but not as much as before. So we renew the curriculum very faster so we throw away about 20% of the curriculum every three months.

Kirill Eremenko: Wow.

Jose Quesada: And yeah, we burned through material like crazy but this has to end because we are now teaching things that are way more advanced than what people get in interviews. For example, we are teaching reinforcement learning, Gans, things that are never going to show up in an interview in most companies.

Kirill Eremenko: Why do you teach them then?

Jose Quesada: I think there's this tension between what people want to learn and why they want to come to our course and what companies want. So if we only teach what the companies want today, then we're not preparing people for the future. So we're trying to prepare people to have a career that is successful the next two, three years without having to reinvent themselves. So the good news is that they will for sure know the answer to any interview question about classical machine learning or spark these are more vanilla let's say. But they also have the extra points let's say of knowing how to do the coolest stuff, let's say.

Kirill Eremenko: Got you. And so three months so they learn for I'm assuming a month or two and then they start on their project. So how does that work? Is that a correct assumption?

- Jose Quesada: Yeah. Now it's more like two months of classes and one month portfolio project before it was 50/50. So the number of classes is reduced progressively over time and the amount portfolio project time is increased.
- Kirill Eremenko: Sorry. So if it's two months of classes and one month of portfolio then-
- Jose Quesada: More or less.
- Kirill Eremenko: So, but then that means portfolio project time has decreased.
- Jose Quesada: In the last two or three batches yes because we added a lot of classes. We are teaching now statistics, math, all kinds of other things that may come up in interviews and we didn't teach them before. So the portfolio project it's very difficult to come up with an original topic. It is of course a passion project for you because you're going to work like crazy here so you have to really love what you're doing but also something that solves a real problem. So people spend literally weeks on one on one meetings with either myself or one of the directors in Berlin to come up with just the idea. The idea is super important.
- Kirill Eremenko: And here I want to talk a bit about I really loved your presentation, we will link to it in the show notes as well if anybody wants to watch the video version. So you had a presentation in December last year where was that? Do you remember?
- Jose Quesada: Yeah, this was in Toronto. So I have also a series of blog posts that I can send you, so you can link them in the show notes which is how to find a portfolio project

that is soon to be successful. I spend weeks on those posts, I wrote everything that I know after 200 plus portfolio projects mentoring and then I think if you read those three blog posts or watch the presentation, you get a lot of value already of the things that we do through the three months to create a new portfolio project. It's becoming increasingly difficult to find an idea. So many things are done already.

Kirill Eremenko: Yeah exactly. And that's what I wanted to talk about here so that of course anybody who comes through the program will experience this in the real world with your guidance and with your coaches guidance. But even people who won't go through the program I think this was very valuable what I learned about how to pick a project. So let's go through these points. So there was a couple of points that you mentioned in your presentation. If you don't mind let's get started on how people find their portfolio projects. Okay. So the first one was you don't have to be Google, you don't need too much data to make an idea happen. Tell us a bit more about that.

Jose Quesada: Right. So there is this a myth that you need a lot of data and lots of compute power to do anything with deep learning. That is not true in my experience. Many of the projects are done on data that the team generated themselves, they label themselves, and the compute time is not that long. So the longest compute times we've had were with Gans and those guys had a Titan running for two weeks, something like that but in general they don't need that much compute power. They just use AWS or more recently Google Colab and

they are pretty okay with that so that tells me that you definitely don't need that much compute power. Also, if you pick the right projects, you usually have a pre-trained model that has done most of the heavy lifting for you and then you just have to retrain the network for your specific circumstances.

Kirill Eremenko: Tell us a little bit more about that how does a pre-trained model work? That's quite hard to imagine. If my project is new, nobody's done it before how come there's already deep learning model I can take and it's already going to be working, I just have to change the last layer. What does that mean?

Jose Quesada: Right. Okay, so let's imagine that in your portfolio project you need to detect trees or cars to segment them out of the picture for whatever reason. You're looking at sidewalks for example and you want to just have sidewalk data, not trees or bicycles or cars. So then you can take a model that is pre-trained to detect any object and to do object segmentation and then just use that to segment the trees and the bicycles and the cars and that's it so you are done, you don't need to train the model for that.

Jose Quesada: If you need to do things for different types of sidewalks, that's very specific your model that you've downloaded somewhere, may not know anything about sidewalks then you may have to train the model to say, okay, this is a bike path, this is a stone type of sidewalk, this is a flat stone sidewalk and you may need the extra training. But this is saving you a lot of time already just by not having to train a model to detect cars, trees and so on that's fantastic and this is

a super power that we got very recently when people started sharing pre-trained models which was not the case 10 years ago.

Kirill Eremenko: But what was that you mentioned earlier about just replacing the last layer of a model?

Jose Quesada: So in transfer learning, this is how this technique is called, you reuse most of an architecture and the last layer that is giving you a mapping between the weights and categories you cut out. So imagine that you have a model that is detecting cats or dogs and you don't want to detect cats or dogs, you want to detect hamsters. So you cut the last layer that has two nodes, one for cat and one for dog and you add a new layer that is untrained, that is hamster or not hamster. And then you train the model again, but you don't update all, you just update the last layer. Sometimes people use a normal classifier from classical machine learning like an SBM or random forest on top and that's it. So you only throw away the last bit of the model.

Kirill Eremenko: So that's transfer learning in a nutshell?

Jose Quesada: Transfer learning, yeah.

Kirill Eremenko: And why does it work? Why does a model that is trained the way it's trained to detect cats versus dogs why does that then work on hamsters?

Jose Quesada: Yeah, a good question. So this is something that applied people really obsess with but in academia, transfer learning is not a big topic so if you go to NeurIPS or any of the big conferences, there is not so

much on transfer learning, but it has immediate applied value. So it's mostly the applied people who have been playing around with this. So the reason it works is because neural networks detect features for you and create nodes with neurons that are sensitive to particular features. For example, in a neural network detecting dogs, there might be a feature somewhere inside that detects ears, another one that detects noses, another one that detects legs, another one that detects tails and well if hamsters have legs and noses and ears then you are reusing that. So of course it's not the same but this is what you're going to train on top.

Kirill Eremenko: But it's not the same but it's good enough.

Jose Quesada: Yeah, exactly. And that good enough is saving you weeks of training time.

Kirill Eremenko: For instance, the way I can imagine it is detecting a dog versus a tree is a dog is more similar to a hamster than it is to a tree. So if a neural network can pick out between a dog and a tree and can say this is a dog, then it's going to be able, especially if you retrain the last layer it's going to be able to do, not perfectly, but good enough it's going to be able to pick out a hamster in the same way. So it just saves you like as you said it saves you a ton of time for maybe 80 or 90% of the same of optimal of the ultimate accuracy that you get.

Jose Quesada: Right. And remember that we have neural networks now that is really good at object categorization. So the state of the art used to be 10,000 classes I think you can do a hundred classes right now. So you have

neural networks that can classify the world in 100,000 classes, let's say and that's pretty much saying machines can see, machines can understand the world at least to the point that they can segment which objects are in the picture. So this is sincerely powerful. So if you want to build things, if you want to build technology, how cool is it that you can use machines that can see? So this is not a website where you have to click and input your data and click next this is a machine that sees the world.

Kirill Eremenko: Well, for sure and you have that wonderful example of the wheelchair project. Can you outline that? I really thought it was a very cool project that helps people, it's for social good I'd love for others to hear about it. So if you don't mind just outlining what was the project all about?

Jose Quesada: Yeah, I would love to. So at DSR, Data Science Retreat we try to do projects with social impact. It's not on the website, I don't tell people that we do this but if you look at the projects and think about what they have in common, is that. So how do you use machine learning to help people to solve problems that real people have. And this also tells you something about the diversity in the student set that we get. So this was Masanori a Japanese person who came for Japan to Berlin to do Data Science Retreat. And when he landed he looked at the sidewalks and thought "Wow, these sidewalks are very bumpy. This would be horrible for a wheelchair user." He knew that because he was a volunteer pushing wheelchair users in Japan.

Jose Quesada: So if you live in Berlin for 11 years it never occurred to me that the sidewalks would be a problem for anybody but then immediately I look down and see wow yes, they are very bumpy. So now Masanori he had a team member so the two of them try to find a solution with machine learning that will help people navigate Berlin in the smoothest way possible. And the way they did that was by analyzing sidewalks. So they went around with a wheel chair with two mobile phones attached to the sides pointing down so recording the floor as they were moving and with the accelerometer registering the bumpiness of the sidewalk. So that was a ground truth. Of course they could not go to the entire city, so they did a few samples and then they came back and tried to generalize that to all sidewalks in Berlin.

Jose Quesada: And how did they do that? Well they use the API for Google Street View and it turns out that you can just ask for the sidewalks in Google Street View. So you have to remove all the obstacles, the trees, the cars and so on and then they had a model that will tell you what the composition of the sidewalk is with their ground truth. And this was also not that big of a model, but that comes back to the point that I was trying to make that you don't need to have that much data. So they had a few samples of data and then Google Street maps to generalize. So they did that and basically for every picture in Google Street View they had a segmentation algorithm that colored differently the different parts of the sidewalk. So then for every meter of sidewalk they could assign a coefficient of how bumpy it is. Now with this information, you can



overlay another map. There's an open source version of Google Maps I think-

Kirill Eremenko: OpenStreetMaps?

Jose Quesada: Yeah. OpenStreetMaps. Yeah, exactly. So on that they overlaid their data and then they could do recommendations. That part they didn't get to do they had run out of time. So if you are in a wheelchair then you can say, "Okay, I'm here I want to go there what is the smoothest path?" And then the machine will tell you. So that is one project where machine learning can have social impact.

Kirill Eremenko: Wow that is amazing.

Jose Quesada: We do this as much as we can. So every batch there is at least two or three projects that are like this. So we built a malaria microscope that runs on a phone, we built a toy self driving car that goes around and picks up cigarette butts. We built all kinds of other things. A tool to help children in Africa to... Not children but the teachers do great dictation automatically because there are very few teachers in many areas in Africa and people need to learn how to write.

Jose Quesada: So often we run out of time and we don't get to finish the product and this is really sad because people graduate and they forget about their projects. That's the way it is, they just want to get a job. They get a job and they're done. So many projects are sitting there doing nothing and they could be definitely been put to production and we just don't have the capacity to do this. So you have any ideas on how to find people that

could take these projects to production do let me know.

Kirill Eremenko: Yeah, that's the tricky part really because that's... I completely agree with you it'd be really cool to have these projects become businesses but ultimately I guess that's the difference between what you're doing and what companies like Y Combinator are doing. Here you help people find jobs and I can totally see how somebody who has done this for instance even this project we just spoke about comes to a job interview puts out on the table. That's the end of the interview because it shows how passionate this person is and if somebody came to an interview with me like that I would not really not ask any questions. Maybe more about their attitude, as in how well they'll fit into the team cultural aspects of what is a good working culture that they appreciate and things like that.

Kirill Eremenko: But technical questions would stop right there because it's a great testament and the difference is I guess that is much less risky than starting a startup and launching a product because you run into all sorts of challenges. Like you have to financing business, business plan, marketing, sales, materials, competition, legislation. There's so many different aspects to running a business and I completely understand. While yes, it would be great to have these projects out there changing the world, the chances of a project like that failing are extremely high. Most startups fail within the first couple of years and I completely understand if somebody is like, "Oh, hey this was a great project but actually what I want is a

job." And hey once they have a stable job, maybe they'll have free time and they'll continue working on this project or get back to it at a later stage.

Jose Quesada: Yeah. So this is what happens. We try to build something around this to continue the projects but if there's too complicated the round way is say five, 10 years we don't have that capacity to support projects for that long.

Kirill Eremenko: Yeah. You should maybe create a little repository online where anybody can go and look at these.... I don't if the people work these projects will be fine with it, but similar to how Tesla made its patents available to everybody. You could just be like "Hey, here are the projects. Anybody can pick them up from here and continue working on them."

Jose Quesada: Yeah. We considered that. That has some downsides one is that people expect things to look better and you have to support them somehow. And in my experience when people graduate the last thing they want is to look back at their projects and support them and answer questions about, "Oh, it doesn't run on my computer. Can you help me?" They don't want that.

Kirill Eremenko: Yeah. Got you. Well, I guess that's the next step for you to solve as a business, as Data Science Retreat. That's a good challenge to have. But let's move on to the next point, that was a really cool example and that was to illustrate the point that you don't need too much data, you don't have to be Google. These guys just ran around took some videos, trained the model and then accessed the Google API. Another cool thing

in your video pointed out... So that was number one. The second one was the eyebrow test, the pretty cool filter for making sure you're picking the right project. Tell us about that.

- Jose Quesada: Yeah. So there are so many things that you can do with machine learning now that are impressive that you should not settle for anything that doesn't make the other person have the eyebrow effect. So if the eyebrows don't go up when you tell somebody about your project then find another project. Because there are so many ideas that you can do right now with machine learning that will make people go, "Wow, this is amazing." So just don't do anything that doesn't produce the eyebrow effect. Eyebrows must go up, if they don't, pick another project.
- Kirill Eremenko: And that can't be your mom. You can't be using that test on your mom.
- Jose Quesada: They cannot be used sorry?
- Kirill Eremenko: Like you said in your video, that you can't check the eyebrow test with your mother.
- Jose Quesada: Oh yeah, it cannot be your mom.
- Kirill Eremenko: It has to be someone else, okay. And another cool one you said number three was a ballpark of a good idea. So if somebody's listening to this, by the way, why is this useful to you listening to this podcast? There's 10,000 people listening to this podcast, naturally even if everybody wanted to run to Data Science Retreat they just no capacity to get everybody through. But why is this useful for you to do at home if you are

planning on growing, skyrocketing your career? It's because just follow the same principles that Jose is pointing out and ideally read the blogs once we put them in the show notes.

Kirill Eremenko: But ultimately follow the same principles and create your own projects and do them at home. And maybe through a platform like ours, like SuperDataScience, you can get together into a team and be part of maybe a group of people working together. But ultimately do those projects and then that is going to be something you can showcase on your LinkedIn, on Medium when you come to an interview where else? And that's why I really think this is very valuable. And do you think these principles are valuable to people who are going to be doing this on their own?

Jose Quesada: Absolutely. So this is why I did this series of three blog posts, it took me weeks to write it down. Everything that I've learned about picking projects is there so you can benefit from that. Just read the blog posts and see if you can come up with projects yourself that will produce the eyebrow effect and will be on your GitHub or better yet will be show-able. Will be physically present in the real world and you can put it in front of people and they will go, "Wow, this is something that you can do with machine learning today. Now I get AI, now I get why people get excited." Because you know what? There is this idea that AI is all hype that we are talking it up way too much. I see that point, I see that the expectations could be out of proportion for things like self driving cars or artificial general intelligence or even NLP, natural language processing.

- Jose Quesada: We don't want journalists writing that computers can understand your language now and produce language that is as good as a human because that's not where we are, that's not. Even the self driving car it's not there yet and I have friends who work in the industry, but what I think is not hype is this superpower that anybody in their kitchen with a crappy laptop can build things like what you will see if you read this blog post or if you just follow people doing Data Science Retreat. These things are at your fingertips and it's tragic that only a tiny percentage of humanity is being able to use this technology. For example women in machine learning with women in AI, not that many of them and we try to incentivize them.
- Jose Quesada: We have been partially successful, the batch that we had the most women was 33% I think. So it's very hard to get women into AI and that's 50% of the population of the world. What are we doing wrong? So my gut feeling after doing this for six years is that it's an embarrassment that we don't train more people to use this technology. It's like some of us discovered the fire and we didn't share it with the rest of humanity, we just kept it to ourselves.
- Kirill Eremenko: Yeah, that's a great analogy, we should share it more for sure. And to your point about the self driving cars by the way for anyone listening Jose's article, well the first blog is available on LinkedIn and then you can click the links from there to follow to the second and third. And so I'm looking at the first one here and you've got a quote from Andrew Ng saying, "If you're trying to understand AI's near term impact, don't think

sentence instead think automation on steroids."
That's a valid point that, that's where we are and
that's where the most value lies right now.

Jose Quesada: Yeah. So another quote from Andrew Ng is "Anything that takes a human less than one second to make a decision is automated or will be automated very soon." So think about boring and repetitive tasks that you do that you don't like and don't think about higher condition tasks like writing a novel or deciding whether somebody is suitable for a job that's definitely outside of the realm of possibility but there is so much low hanging fruit. And yes in manufacturing which is very strong field in Europe there's so many opportunities for improvement. So there is this idea that Europe is really behind in the AI, if you go to conferences you don't see that much action for sure from European researchers. But it's just in the culture there that people don't talk about what they're doing.

Jose Quesada: But there are companies like Audi, BMW, [Bannon 00:51:11], other who are manufacturing, Airbus doing all kinds of cool stuff they just don't tell anybody. And there is where you could have a lot of impact, the low hanging fruit is still there. If you go to the U.S for example, where there are 20 startups for every possible application of AI already the low hanging fruit may not be there, in Europe things are starting to happen now and they're happening at a scale and they are really affecting production plans big time.

Kirill Eremenko: That's a good point and speaking of conference I think it's a good time to mention that Jose might if the stars align, might be joining us for DSGO Europe, which is a

15, 16, 17th of May, 2020 so if you haven't gotten your tickets yet you can get them at DataScienceGO.com. I know Jose, you haven't been to a DSGO yet and this is our first time actually running it in Europe. We're going to be running in Berlin but how do you feel about if the stars align and you're able to come, how do you feel about it?

Jose Quesada:

Oh, I'm super excited. I would love to go to your conference and just from the things that you've told me, the way you ran conferences and your goal with it this is totally a very different conference from what you will get otherwise. Everybody has been to conferences that are very commercial where you get pitched products and so on mostly big platform, better vendors and I don't know about the rest of the conference, but I feel completely cheated when I'm in one of those.

Kirill Eremenko:

Yeah, I agree with you and definitely very excited about this upcoming event in Europe and if you again, if the stars align again in October this year you can come DSGO U.S. I think you'd enjoy talking to Gabriela de Queiroz this will be her third time coming to the event. And to your point about having or inspiring more women to get into the space of AI, she's doing a great job. She runs R-Ladies.org and they have over 100 I think chapters worldwide now with something crazy I think like 30,000 members. I might be getting numbers wrong, but she's doing a fantastic job and at the end of the day, I've talked to her about this a few times, at the end of the day, it boils down to creating those role models to encourage people.

- Jose Quesada: Absolutely. I think for women in particular, what they really benefit from is to have a role model. If they see another woman that made it big in the field, they get encouraged. I mean, that's true for everybody, right? I don't think it's special for women, but I've heard from women that they really look forward to see role models.
- Kirill Eremenko: Yeah, that's definitely. It's always helpful for any minority if you don't see anybody that you can relate to in a certain field, then you're less likely to go into it.
- Jose Quesada: Exactly.
- Kirill Eremenko: Very interesting. But hopefully we can do our parts in that and that example where you said 33% in one cohort were women in your intake I think that's setting a great example and showcasing those stories is great as well. So let's continue. So the next step, so we've talked about you don't need to be Google, you don't need a crazy amount of data to start an idea that a lot of people get put off by that. Then check the eyebrow test. This next one was very cool. Ballpark of a good idea. If you don't have a good idea, you just need to be in the ballpark of a good idea and this is how that cigarette project came to life. Can you tell us a bit about that?
- Jose Quesada: Right. So one person came to me towards the end of week three and he told me "Well Jose, I've been thinking about this for three weeks I could not come up with anything. I have zero ideas." And then I asked him, "Okay, what do you feel passionate about? What is alive inside of you?" And he said, "Well, I actually hate waste. I don't want to see things going to waste."

And then we started looking for things that had to do with recycling, with trash so he found a data set from Stanford where they have pictures of trash and then we thought "Okay, how about we do something about picking up trash?" And then I realized that we had a self driving toy car from a previous batch in batch eight. So Marcus Jones who is an incredible developer, he produced an operating system for a toy car back in 2018 before Amazon created theirs and made it open source.

Jose Quesada: And his portfolio project was a toy self driving car that was driving around in a circuit, so doing laps in a circuit. So he thought, why don't we take that car and we make a version of Wall-E do you know this Pixar movie Wall-E? It's this little robot that goes around picking up trash. So what if-

Kirill Eremenko: I love that movie.

Jose Quesada: Right, me too. So what if we take that idea and okay, we have a self driving car but it's a toy. We just have to give it some way to pick up trash and then he went out and two other team members and then they figured out that they could mount a mechanic arm, a robotic hand on top of the car but then just identifying all types of trash was not as interesting. So one different person from that team came up with the idea of picking up cigarette butts. So cigarette butts are very poisonous for the environment. One cigarette butt can contaminate 20 liters of water, birds take them and give them to their chicken and they die. They are terrible and smokers keep flipping them around and they are very pollutant aspect of what we humans do.

- Jose Quesada: And they are not easy to be picked up with a broom or with many other means because they're so tiny and so light. So then the goal was to just make a robot that goes around and stab the cigarette butts. We tried grabbing them with a pincer but it didn't work so stabbing. So this is what I mean when I say that you only need to be in the ballpark of a good idea. The initial idea was okay, I hate waste. Then another expansion is let's try to pick up trash. Then another expansion we can use a toy car on a previous batch. Another expansion is let's mount a robotic arm on top. Another expansion is let's not pick any trash, but just cigarette butts, these are high value targets. Another expansion is let's try to stab them instead of grabbing them. So this is what I mean that you need to land in a part of the space where there is a good idea, you can iterate around this. This took weeks for sure don't get me wrong, but this is how it works in general.
- Kirill Eremenko: And in your blogging video as well you have a great visualization to support this like a radar the ones that you see on a ship, on a military ship. An X in the middle and it's green and this thing is around, it's a great example. So you don't know what your idea is, but if it's somewhere on the radar it means you're nearby start somewhere get to it. I love that example. And if you still can't come up with idea, there is step number four that you can take is watering holes. What's that all about?
- Jose Quesada: Right. So this is something that I borrowed from Amy Hoy who talks about how to be pick ideas for a startup. She says that you need to pick up problem

that is a real problem, like hair on fire and not a nice to have. And the one characteristic that those problems have is that people talk about them, they bitch out them. So where do they bitch about them? In what she calls watering holes. Watering hole is a place where people gather to talk about the problems. It could be Reddit, it could be Twitter, it could be a forum, it could be an actual water cooler in an office. People hang around it and they talk about their problems there. So you should be there and you should listen to people complaining about their life because if there is something in their life that you can fix, then that's your portfolio project, that's your eyebrow effect right there.

Kirill Eremenko: Exactly and your analogy is great. The one you used in the video was if you are going to an African Safari you're not going to wait for the animals in the bush, you're going to go to the watering hole where all animals come to drink and that's where all the action happens. I think we have time for one more, one that I really liked and we can dive into this one. It's again to do with data and it is about producing your own data and you gave a very good tip, use hardware. Sometimes you can find data online, right? Great. But sometimes just go and produce your own data and we already talked about that wheelchair project. That was a great example of producing that starting data to identify the pathways or classify those walkways. But you have another fantastic example of producing your own data the malaria microscope, tell us a bit about

that. I found that project incredibly inspiring that one person was able to accomplish on their own.

Jose Quesada: That's really amazing. So there was this participant Eduardo, who woke up one day and realized that if he did the right things, he could save 600,000 lives per year. So what he wanted to do was to use deep learning to detect malaria, using a phone and a cheap microscope that he will put together with some parts that you can get from China. So you can buy a cheap microscope and kind of attach it to a phone and get enough augmentation to detect malaria parasites. So the way humans do this right now doctors is just to put a blood sample on a microscope and count the number of parasites and if there is more than a certain number per square millimeter or whatever then they declare that the person has malaria.

Jose Quesada: So counting and detecting things in any image is something that machine learning can do very well for you. So his idea was to just use machine learning to do that and it had to be very low cost. So he got to \$60 per unit and a second hand microscope sorry, a second hand phone. So when we saw this we thought we need to do more projects like this number one and number two, we have to get this project the most visibility that we can. So we paid for somebody to do a video, professional video. This video went into a crowdfunding campaign so he got I think 5,000 Euro, which he used to fly to the beginning of the Amazon river where there is plenty of malaria and then he started collecting samples there.

Jose Quesada: So the first dataset he got to do proof of concept was an online dataset for malaria that anybody could have used in the world. Then the second data set he used his hardware to collect more samples but then as this became more popular, then he started approaching hospitals that are in the areas where there is malaria. And he just asked them to give him malaria samples he will take pictures of them. After doing this for a while he has now more malaria samples than any single hospital on earth. So yeah, this is the power of using hardware. So a cheap microscope can give you leverage. That's an incredible amount of data that you can collect if people use it.

Kirill Eremenko: That's crazy.

Jose Quesada: Yeah. He's still going around giving talks about this and so on. It's a little bit of a pity because at some point he had 20 people in a WhatsApp group so it was going well, it was a nonprofit and so on. But at some point he realized that he's not going to make money out of this ever because you cannot make money out of poor people that you are trying to save. So he will never be able to run this as the single activity of his life so he kind of gave up. I mean, not completely he's still running it, but he's now a consultant that's doing... He has a full time job.

Kirill Eremenko: Yeah. That's the next real challenge after you have a good idea, what exactly do you do with it? In terms of even if you're extremely passionate about it, these things require resources and unless there's funding or unless it's like the way we run our business, you and I unless it's a bootstrapped you are able to invest some

money initially, but then you're able to generate profit and reinvest that back into the business. Unless that's the case, you're going to ultimately just work, work. You'll have to sell your house, your car, you'll start eating beans and rice and so on. And that's it. And then you'll run out of even just ways to support yourself let alone the business. So it's a real dilemma. You're right, it's a pity that sometimes some of these fantastic ideas don't get traction.

Jose Quesada: Yes. Coming up with business ideas on top of doing fancy demos that could have social impact is really a challenge. One reason I like ISA so much, income share agreements is that it's one way to have social impact and ferment social mobility that is actually a good business as well.

Kirill Eremenko: Yeah, no that's definitely true. It helps people in all walks of life. If anybody listening to this is in Berlin area or Germany or even Europe or anywhere in the world and is interested in checking out Data Science Retreat by Jose make sure to get in touch. When are your next intakes this year in 2020?

Jose Quesada: I think the next one is March 30th, I have actually go to the website to tell you exactly. So we have four intakes per year. We had one in January so the next one is March 30th, then the one after June 29th and September 21st.

Kirill Eremenko: Okay. But then how far in advance should people apply? Is it too late to apply for the March intake?

Jose Quesada: No, no. We keep running interviews until basically the day before. So if you are good and you pass the

interview, you can come anytime. So it's a lot of interviews, but yeah.

Kirill Eremenko: Okay. And what's the website to apply?

Jose Quesada: Yeah. Datasciencercetreat.com we can put it on the show notes.

Kirill Eremenko: Yeah. Datasciencercetreat.com. Is there like a retreat? Do you like go do drinks and fun stuff as well?

Jose Quesada: Yeah. People keep finding us looking for a yoga retreats and things like that. The name retreat may have been unfortunate because in North America we could not use it because it's associated to things that are not hard work. Be lazy or doing yoga or being a monk, not so in Europe. The thing that I was going for is this writer's retreat. When you are a writer and you want to finish your masterpiece novel and you go on a retreat and go to the top of a mountain and get time to be undisturbed and finish your novel, this is what I was going for. That's the meaning of retreat and it's getting some traction. So people start talking about boot camps as retreats, which is very cool. Just like in Spain less people talk about yogurts as Danone and Danone is just a brand that does the yogurt. So this is kind of happening with retreat which is very cool.

Kirill Eremenko: Awesome. Fantastic. Okay, well Jose we've actually come to end in terms of time for our podcast. Huge pleasure having you on this show, really cool-

Jose Quesada: Same here.

Kirill Eremenko: Great cool discussion. Where before we go, before we finish up where can our listeners find you? So we



already mentioned Datasciencetretreat.com anywhere or other places that is a good way to follow you and whatever else you work on?

Jose Quesada: Yeah. You can follow my Twitter or LinkedIn I post maybe two or three times per day on Twitter that's @Quesada my last name and Data Science Retreat has another Twitter.

Kirill Eremenko: Nice. Fantastic. Okay, great. We'll put all those in the show notes, make sure to connect with Jose on LinkedIn. One final question for you. What's a book that has impacted your career or something that you'd like to recommend people who are starting out into the world of creating AI projects?

Jose Quesada: Aha. So there's no book. Interesting. That could be an opportunity. There is no book for creating projects. There is one book-

Kirill Eremenko: You should write it, you should write that book.

Jose Quesada: There is one book that impacted me quite a bit called AI Superpowers, but it's very negative. So if you read that book, you will get disappointed, you will think that the end is near for anybody not in China or the U.S and that's not my thinking right now at that. That's a conversation for another day but I think because of open source and pre-trained models and so on the AI have not, the rest of the world that is not China and the U.S have a huge opportunity here. But okay, the book that you asked me is AI Superpowers by Kai-Fu Lee.



- Kirill Eremenko: AI Superpowers by Kai-Fu Lee. Make sure not to get carried away with the U.S China principle. You're right, anybody in the world can build AI things and really revolutionize this space for sure.
- Jose Quesada: Yeah.
- Kirill Eremenko: Okay. Well once again, Jose, thanks so much for coming on the show. Very cool chat and I look forward to seeing you at DataScienceGO either Europe in May or in California in October this year.
- Jose Quesada: Awesome. Thanks Kirill so much for inviting me and it was a pleasure for me to talk to you. I look forward to meet you in person in one of the conferences.
- Kirill Eremenko: So there you have it everybody, that was Jose Quesada, Founder and CEO at Data Science Retreat. A huge thank you to Jose and huge thank you to you for spending this hour with us. Data Science Retreat, looks like an amazing project, I am personally very excited to be sharing this with our audience. Don't forget that you can go and get your special discount just for SuperDataScience Podcast listeners. If you do decide to sign up, make sure to use the datascienceretreat.com/sds URL. And when you are checking out, make sure to specify that you heard about Data Science Retreat on SDS or SuperDataScience, also to get your discount. SuperDataScience does benefit from this as well.
- Kirill Eremenko: So that's one way of getting in touch with Jose, and maximizing your chances of getting hired to your dream job in data science. Definitely check it out. Otherwise, if you want to meet Jose, there's a chance

that he will join us at one of the DataScienceGO events this year, either in Europe on the 15, 16, 17 May or in the US on the 23, 24, 25 October. So have a look out at our Speaker lineup at datasciencego.com, and if you see him there, you can jump in. And speaking of DataScienceGO, it's a fantastic event, we are for the first time running it in Europe, also in Berlin, funnily enough, so if you are around if you can make it in May, make sure to come on over, there is only 200 seats this year, take yours at datasciencego.com.

Kirill Eremenko: On that note, as usual, you can get all the show notes for this episode at superdatascience.com/343. There you will find the transcript for this episode, plus any materials mentioned on the podcast, any links, and of course, Jose's LinkedIn where you can follow him and ask him any questions that you feel are relevant. So that's all at superdatascience.com/343 and that's also how to share this episode. If you know someone in Berlin, if you know someone in Germany or in Europe who is interested in growing their data science career and getting their dream job, send them this episode, superdatascience.com/343 and maybe, just maybe you'll help them get the dream data science job that they've always been looking for.

Kirill Eremenko: On that note, I'm checking out, thanks so much for being here today and I'll see you next time. Until then, happy analyzing.