



# **SDS PODCAST**

## **EPISODE 187**

### **WITH**

## **FAVIO VAZQUEZ**



- Kirill Eremenko: This is episode number 187 with Principal Data Scientist at OXXO, Favio Vazquez.
- Kirill Eremenko: Welcome to the Super Data Science podcast. My Name is Kirill Eremenko, data science coach and lifestyle entrepreneur. Each week we bring 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: Hello and welcome back to the Super Data Science Podcast ladies and gentlemen. Today we've got a very exciting guest on this show Favio Vazquez. Favio is originally from Venezuela and he lives in Mexico where he works for a company called OXXO, where he is the Principal Data Scientist.
- Kirill Eremenko: If you're not familiar with OXXO, then this is a chain of convenient stores, and actually a very, very massive chain of convenient stores. Think of it as in, it's a massive competitor to 711. They have over 14000 stores in Latin America. We discussed with Favio quite a lot of interesting topics, and in addition to being a principle data scientist in OXXO, Favio is also a influencer, a thought leader in the space of data science.
- Kirill Eremenko: He's got over 37000 followers in LinkedIn and that's why I was very excited to bring him on the show, because you may already know him, may already have seen his work and followed him and heard of these things he's being doing.
- Kirill Eremenko: In this podcast we discussed quite a lot of cool and interesting things. We talked about contributing to the

community. We talked about sharing ideas on LinkedIn. We talked about why data science is becoming a science, an actual science and what that means. We talked about GitHub and we discussed the five characteristics of a data science project.

Kirill Eremenko: Those are just some of those that we talked about in this podcast. Of course there's lots more and I'm sure you're going to enjoy this journey and get some very valuable insights from Favio. Without further ado, let's dive straight into it. I bring to you Favio Vazquez, Principal Data Scientist at OXXO.

Kirill Eremenko: Welcome ladies and gentlemen back to the Super Data Science Podcast and today I've got an exciting guest on this show Favio Vazquez. Favio welcome. How are you doing today?

Favio Vazquez: Hi, hello everyone. I'm very happy to be here, very excited and just hoping to just have a great talk with you.

Kirill Eremenko: Yeah, yeah, for sure, very excited as well and where are you calling from? Tell us, such an unusual occasion, I haven't heard anybody from this country in the podcast yet.

Favio Vazquez: Okay. I'm calling from Mexico right now from Monterrey.

Kirill Eremenko: Mexico.

Favio Vazquez: Yeah, in the North of the country.

Kirill Eremenko: That's awesome, but originally you're from Venezuela, correct?



- Favio Vazquez: Yeah, that's right from Maracaibo.
- Kirill Eremenko: Maracaibo.
- Favio Vazquez: Very close to Colombia
- Kirill Eremenko: Yeah. Okay, awesome. Well, very excited about today's podcast. It was actually yesterday looking at a webinar you were having with Randall Lau, so there's quite a few things that I picked up from there that I'd love to talk about. Of course, as we just discussed, it will be cool to talk about the situation of data science in Latin America, which is quite different to the US. To get started, tell us a bit about yourself. If somebody off the street were to ask you, "Favio, what do you do?" What would you say?
- Favio Vazquez: All right. I think my first answer will be, I'm a data scientist and then I'll try to explain what's data science in a language that's very easy to understand. Normally this is not that easy, but what I normally can say is, I take data and I conform it into actions for companies.
- Kirill Eremenko: Okay, so you get insights to help companies act on their data?
- Favio Vazquez: Yeah, and normally then there will be a short discussion on AI or maybe what's machine learning and how very common it is to have these tools in our hand, like in our phones and we don't even know it.
- Kirill Eremenko: Okay, yeah, for sure. Where do you currently work?
- Favio Vazquez: Right now I'm working for different companies. My main work is at OXXO. OXXO is a commodity is to be one in Mexico and have like 17000 stores here. It's very big and I work with Raken Data Science group



and it's like a consultant firm for data science. They have a partner in a firm called Iron AI, where we do consulting by ourself with data science. I also teach some classes on data science in Python, R, Spark and these kind of things.

Kirill Eremenko: Wow. That's so many different things. You have a full time job. I'm looking through your LinkedIn if you don't mind, I'll read them out. You're a Principal Data Scientist at OXXO. You're a Senior Data Scientist at Raken Group, that's the consulting. You're a Chief Data Scientist at Iron AI. I'm guessing this is the entrepreneurial venture and startup.

Favio Vazquez: Yup.

Kirill Eremenko: You're a data science lecturer at Afi Escuela de Finanzas. That's crazy man.

Favio Vazquez: Yeah. It's a school for, I mean it's from Spain, but they have like a new program here in Mexico. I just finished that last week in Mexico city, where I teach some things about data science or Python's Park. I think it's very interesting, because we don't have that many things here in America for starting data science, like in a serious way.

Kirill Eremenko: That's so cool that you're teaching and you're helping people. It's also very, it's insane that you have four occupations at the same time. How do you get by? How do you find the time?

Favio Vazquez: All right, so what I think is that life is not work. I normally have a lot of time for being with my family, my girlfriend and have fun and stuff. I mean because I really love what I do, I'm not like overwhelmed. Like oh

man, I have so many things to do and this is kind of weird and stuff. What I normally do is that I do my best every time to do most of the work very fast, and then just tweak the things I do.

Favio Vazquez:

Normally I'll start working with some model or something and I'll recommend it right away so I have to do it later. I have my whole setup for doing data science that's very easy. What I normally do is that, I create like blogs informational webinars for people to understand what I'm doing and also for me for later to re-visit the topic. Normally you'll try like a 1000 libraries in one year. In the end, you don't remember hey, what did I do with this library? This is why I create some blog posts and stuff for me to remember what I'm doing and for people to have a peak of what's the life of contributing in open source in data science, something I'd really like. I'm a contributor for several libraries too, like Spark.

Kirill Eremenko:

Yeah, okay. That's like a lot of things and you're right. When you don't feel like that your work is work, you don't get overwhelmed. In addition to all of that, you actually also post quite a few things on LinkedIn, right, like as you said to help you remember things so that you can get back to them. Is that like, does that all tie in together? That's a common thread that I've been seeing when people are very excited about the work and they have, on the side they teach as well. At the same time they post articles on LinkedIn to help people and to help themselves remember. Do you feel that these things feed each other and makes it easier for you so there's like synergies in between them?

Favio Vazquez: Yeah. I really think that my point, that the point where I realized that LinkedIn was a great platform for sharing ideas and also getting ideas from a lot of different people, I think my life changed. From that point, I realized that, okay, so data science is not only about working in a place here. It is about also explaining the work of data science and this is what I think data science is, because it's very new. It's a very new topic.

Favio Vazquez: If we only do it by ourselves and we're in a closed environment, it's very hard to know people that are doing the same kind of things that we're doing or maybe understand some things that is not that easy to understand. Right now I think we're in the era of the blogs and books from data scientists. This is very interesting, because if you thought about it before, when we were doing physics in the 1900s, I mean just like a very small group of people that actually develop the physics that we do right now. It's like Max Planck and you have Einstein and you have Heisenberg, you have Stoellinger.

Favio Vazquez: All of these different guys wrote books. They wrote articles and also they didn't have like Medium to create blogs, but they were doing articles in the format of a blog too. Nowadays I think it's very easy to understand hard concepts as truth blogs. I mean, like five years ago when I wanted to understand a thing, I just had to go maybe to an article or maybe just a book. Right now, every day you get hundreds of amazing blog posts or only like videos, webinars where you can get all of these different kind of information



really fast. I think it all comes together like doing data science and also helping people because right now, I am where I am right now because people helped me. I think they indirectly help me and directly help me too.

Favio Vazquez: You don't know the power of the words you're saying, because sometimes there can be a post with two likes you created. For someone who says, "Oh yeah, this is like an amazing post that has changed my life." I have that experience. That's something that encouraged me to continue, because sometimes when I was writing LinkedIn post and stuff I was not getting that much attention in the beginning. People started to write me, "Hey, thank you very much for doing this post. It was very helpful." Or maybe people just wrote me like, "Hey, if you want to know more of the things that you were writing, I wrote a blog post yesterday." There was a big combination of things that made me the person I'm right now.

Kirill Eremenko: No, that's very inspiring and I love that you mentioned even if your post has like one or two likes, you never know how that impacted those one or two people that liked it. It might have just changed somebody's life.

Favio Vazquez: Yeah.

Kirill Eremenko: Definitely I've been in those situations as well, so it's definitely inspiring to hear you say that. Yeah, that's a very description and congratulations you're on the, what, you're like at 36000 followers on LinkedIn right now, what a journey. Tell us a bit about that. How did that happen? How long did that take?



- Favio Vazquez: I mean, I think my experience with LinkedIn was very similar with my experience with GitHub and I'm going to explain why. When I was starting with GitHub, I was studying computer engineering back in Venezuela. I used it as a source for a code to help me to do my homework. That's it. I mean I didn't even understand, it was like a platform for building open source codes. I didn't know it was these different kind of things.
- Favio Vazquez: I used it like, I search in the internet and all the answers were stuck overflow in GitHub. I started as a user in GitHub, but in the end I realized hey, this is people like me, that they're just posting their ideas in a code style. I started to contribute into my projects in GitHub and also on parts like Spark and things in Julia, Python and R scholar, so I got very excited. It was the same for me in LinkedIn.
- Favio Vazquez: When I discovered LinkedIn, it was like, okay this is for searching for work. I worked at a time because I was right away from my masters and then I realized hey, but this is more. People are sharing here ideas, they are sharing blogs, you can create articles on LinkedIn. I started doing the same, because I was so inspired about so many people like you, building these great things, like Super Data Science and big pages for data science. Hey, I can contribute too with the little things I know.
- Favio Vazquez: That's what I started doing. I mean I didn't do it for the followers, I started sharing my ideas. Some of my posts were very popular and I think that the first time I got a lot of attention in one of my LinkedIn post was when Jeff Weiner the LinkedIn CEO liked my posts. That

was very weird, because it's like, how did he find my posts? That post got like 3000 likes. I went wow, this is amazing.

Favio Vazquez: Then I realized I had a voice on LinkedIn and this is why I started writing more things and more things and more things. Right now I think the amount of followers you can have on LinkedIn is a way for you to say that you're doing things right and you're sharing important things for those people.

Kirill Eremenko: Yeah, that's true. Interesting to hear how Jeff is, Jeff Weiner is learning data science. It looks like everybody is learning data science now.

Favio Vazquez: Yeah.

Kirill Eremenko: Yeah, okay well, that's very cool and I appreciate you saying that you're not doing it for the followers, you're doing it to share and help people. As we discussed before, it's very important, even if you have one or two people that's, you've already done a huge contribution. Okay, well, so that's a quick overview of who Favio Vazquez is. Now let's talk a bit about data science. Let's talk a bit about your work.

Kirill Eremenko: I probably would like to start with something you mentioned in your webinar with Randall Lau. I really enjoyed this so maybe you can share it here as well. You were talking about data science projects and you said that, all data science projects have to have five characteristics. I will read them out here and then maybe we can go through them one by one and you can give us your thoughts on that.

- Kirill Eremenko: You mentioned that all data science projects should be reproducible, fallible, collaborative, creative and compliant to regulations. Could you walk us through these one by one and maybe give examples where it's possible.
- Favio Vazquez: All right, so there's a lot of talk right now about reproducibility in data science because, we are doing science in the end. I think before talking all of these five points you mentioned, I think I need to start with my view point on data science. That, it is becoming a science. I think right now it has the name science in it, but it is only becoming a full science right now.
- Favio Vazquez: If data science is a science, and if this is the case, I think all of these five points should be there in every project we do. The first one is of course necessary to do science. It's impossible to create an experiment and if you're the only one who can reproduce the experiments, it's not in the end an actual contribution to the world. I mean if I say in an article, "Hey, I just discovered that water boils at 100 degrees celsius," but I'm the only one who can ever test it. This is like a mysterious article, like oh, what is this guy doing.
- Favio Vazquez: Science is based on the way that other people can reproduce your works. I think data science is very easy to do, these different kind of things because we are in an open source world. I mean most of the data science contributions are in open source, Python, Rs, Color. We all have the blog posts and information and articles on all the things we do. I think the path to reproducibility is not that far away from data science.
- Kirill Eremenko: Got you. Makes sense.

- Favio Vazquez: Yeah.
- Kirill Eremenko: Basically what's the point of insights if you're the only one who can get them? You want anybody to be able to do that.
- Favio Vazquez: Yeah, of course. I'm not saying that you should post the work you're doing for your company, because you're going to get fired. What I'm saying here is like the techniques you're using are of course sharable if your company allows it. Of course all of designs that people do in companies or schools is not fully reproducible. I mean you need to have the data they have or the equipment they have. There are paths to or several steps on reproducibility and I think data science should be at least reproducible in environments that are similar to yours. I think this is my first one.
- Kirill Eremenko: Got you.
- Favio Vazquez: The second point was collaborative right?
- Kirill Eremenko: Fallible.
- Favio Vazquez: Oh fallible. This is a very important thing here. Science is not in the look for the truth. I think this is a very important thing to say one and again and again and again. We're in the look for knowledge. If you want to read more about this, I really recommend you, you do some reading on processing or epistemology. The thing here is, we're not solving the problem forever with a solution that will hold for eternity, all right?
- Favio Vazquez: We're creating a solution for the problem with the technology, the theory and all of the apparels we have

right now. This is important in science because if I go through an article and the article says, "This is the final solution to the problem, no one else should do anything more about this," this is not going to be a thing in science. This is not a found solution.

Favio Vazquez: This is also very cool for you to explain to your company. You're creating models and models are an obstruction of reality. You're actually trying to create a vision of reality that works for you and you can understand it and test it. You're not actually finding the ultimate solution for a problem. You're finding a solution to the problem.

Favio Vazquez: Right now with data science what we should be thinking is, okay, I found a good solution to this problem, but of course it can be better in the future. That should give you a touch of being humble of what you're doing, because sometimes we think, yeah, this is the best model we can ever create for this problem. That's never be truth because science is never going to look for the hidden truth. We're looking for the knowledge. When you look for the knowledge, there's no stop.

Favio Vazquez: This is very interesting, because people sometimes ask me when I was doing science, when does it end? What is the end of science when you know everything? My answer is, we're never going to know everything. This is not the path of doing science. Science is not about the truth of the universe. It's about the knowledge we can get from the things we see, understand and can cite.

Kirill Eremenko: Yeah.

Favio Vazquez: That's the second point.

Kirill Eremenko: Totally agree. I'll probably just add to that by just defining, I'm sure people are getting an image of what data science kind of is in terms of being able to make mistakes or not being able, not having the ultimate truth right away. Just to sum it up, I will read out the definition of the word fallible, because when I saw in your presentation I realized that I didn't know what it was, so I had to Google the definition.

Kirill Eremenko: Fallible is an adjective, it means capable of making mistakes or being erroneous. That's exactly what we just discussed, that if you're not 100% like every single time you do something and make a discovery, that's the final truth. That might be correct, but might not be the full picture, right, that there might be more to and usually there is. As you say in science, usually there is.

Kirill Eremenko: I love how before you gave the examples of Newton and other physicists like Einstein and so, because like you're a physicist. It's interesting that when we started just before the podcast we were talking like we both studied physics and just different styles. Like you studied, what was it? Cosmology, right?

Favio Vazquez: Yeah. Cosmology yeah.

Kirill Eremenko: Cosmology and now you're into data science and I also studied physics coming into data science and a couple of other people like lady [inaudible 00:22:45] that was an astronomer now she's into data science. Interesting how depending on your background, depending on where you came from, you see data science from a

different perspective. I think this will be valuable for a lot of our listeners who might not be coming from a physics background to hear this perspective of data science as a science and all of these different aspects and characteristics that we're mentioning here. Let's move on to ...

Favio Vazquez: Yeah, I think that's a very good point.

Kirill Eremenko: Yeah. There's more, number three, collaborative. What do you mean by that?

Favio Vazquez: Right. Collaborative I mean here that, we exist in a team. I mentioned that in a post last week on LinkedIn because some people are thinking that you can be a data scientist by yourself without seeing anything else. That's impossible because we are an applied science. That means that we need someone else to solve problems to, because if no one else is giving you a thing to solve, or you're not having a problem to solve with data science, then you're not doing data science.

Favio Vazquez: So we exist in a team in two ways. First is your actual team like or a data scientists, your manager or some people data analysts, data engineers, people that work close to you. It's very cool to have this kind of different like perspective of what you can do with data. I think a data scientist should not be like an expert in each of these topics, but have an idea of what these people are doing.

Favio Vazquez: The second part of the theme is the business you're working with, because sometimes when you are a data scientists, you don't need to reply or to solve problems for very different kind of people. Like marketing people,



business people, people working with distribution of something in your company. You'll need to be able to talk with these people in a different way you talk with your team in data science. You'll need to have a way of understanding what they're saying and what are their requirements.

Favio Vazquez: It's not the same to hear someone talking about a business, their business, they're working if they are a business guy or they are working in the marketing department. I think this collaboration is what in the end will guide you to make a good solution that will answer the problems for the business you're working with.

Favio Vazquez: This is the same in science. I mean of course you're going to have this scientist working by himself and stuff. In the end he's trying to solve a problem for everyone. He's trying to understand the world. In the end right now I don't think that picture of the mad scientist in a laboratory by himself, that's not true at all. I mean, my experience doing science was very different. A lot of people working with me. Every Friday we got together to see what others were doing, understanding the different parts of an article or different kind of things in a subject. I think this collaboration is what makes you an effective data scientist.

Kirill Eremenko: Yeah, and just when you were saying that example of the mad scientist in the laboratory by themselves, I remembered a part of a book I'm reading now, it's called 'Sapiens'. There they talk about collaboration or the author talks about collaboration in the sense that,

during the cognitive revolution, that was the deciding point for our species as sapiens to take over the world. The fact that we can coordinate our efforts in large numbers is what distinguishes us from others.

Kirill Eremenko: Like try putting 10000 monkeys into a stadium, there's no way they're going to collaborate, there's no way they're going to be able to follow rules or to work together and watch a concert or do something together, watch a soccer game. They're limited in their capacity to collaborate 20 or so people and things like that or monkeys and things like that.

Kirill Eremenko: In our case, the fact that we can, we have this cognitive apparatus that allows us to collaborate, that is one of our biggest advantages. We should really use that, especially like in complex things like science, data science and whatever else we're doing for our profession. I totally agree with that.

Favio Vazquez: Totally agree, yeah.

Kirill Eremenko: Very important to keep that in mind.

Favio Vazquez: Okay, so next point.

Kirill Eremenko: Excellent yeah.

Favio Vazquez: I think it was creative right? Creativity is something that is like really defined in a weird way for most people. I don't know if your people listening here ever listen to people saying, "Yeah, so the creativity is in the left part of the brain. The math and language is in this part of the brain, and the artists are the people who are very creative. Science is like very in the other side with mechanisms and tools."

- Favio Vazquez: Okay, so this is why I think it's weird. I don't think that a mathematician solving a theorem is not a creative person. I think he can be the most creative person in the world right now. I think trying to solve things with science and with mathematical things and trying to understand the world, you only have to do it in a creative way. They're in the work place. When someone says, "Hi, this is the creative department," they're not the data scientists. They're people creating like images and stuff.
- Favio Vazquez: I think this should not be the way we see data science. Data science needs creativity, because some of the things we're solving are not solved. We have no answers for them and this is why we've been, I mean if you search for data science on the internet right now, you'll get a lot of different things. Data science for applied businesses, for science itself like papers, research, review, overviews, surveys. This is why we're trying to find a definition that contains all of these information we're trying to use to solve problems.
- Favio Vazquez: I think we need to be creative, because we're using things from other domains. Let me give you an example here. A lot of the things we're using right now for doing machine learning and deep learning, they come from biology or chemistry or different kind of science. I used just look for the articles of [inaudible 00:30:04] they were all published under biological science review magazines.
- Favio Vazquez: It's very creative to have found a way to apply a model that was created for a specific purpose into another place. Right now we're using random forest and GBMs

and literature rations for doing things that no one ever thought about. I think this is one of the realizations of creativity in data science, being able to use different kind of things from other domains and apply to your work.

Kirill Eremenko: I definitely agree with that as well. Data science. I find that some of the best data scientists come from creative backgrounds, whether it's arts or music or something completely unrelated to science. Simply because there is so much capacities, so much room to be creative in data science. They find how to apply their existing creatives skills to become even more successful. I think everybody in data science should.

Favio Vazquez: Yeah, no and again, my point here is trying to make people understand that science is also creative. Data science is creative because it's a way of science, so this is my view on creativity on data science.

Kirill Eremenko: Yup.

Favio Vazquez: I think the final point is very important and is compliant to revelations. I put that, I mean I work on that list like a year ago. I work in a bank for like a year and then I realized how important it is to follow the rules you have to create models and see the data. The thing is, right now we're seeing more and more updates for the things we can do and cannot do with the data of the customers or of people we're analyzing.

Favio Vazquez: I don't know if you remember but last two months, I mean my inbox from my email was full with emails. We changed our rules and agreements. It was all because in Europe there was a lot of changes on how

people can access, view and understand the data from people.

Kirill Eremenko: Oh the GDPR right?

Favio Vazquez: Yeah, the GDPR.

Kirill Eremenko: Global Data Protection Regulation.

Favio Vazquez: Yeah, and I got a lot of messages from companies that were sending to me for me to be aware that now they're following their extenders. Right now this is like the main point of revelations in the data world, but there will be more and more in the future. I mean if you think of science, I think everybody knows right now that cloning a human being is illegal.

Favio Vazquez: If you think on data science, we can do almost whatever we want with the data. I think we are very free to do whatever we want right now and I don't see that as a bad thing. I really think that relations should be in place, because data is people too. I mean when I was working in the bank, when I created models for like credits carrying model, I really had to thought about, hey, this is not only data. This is people's lives.

Favio Vazquez: I mean I'm here understanding what they buy, what they cannot buy, I mean how they spend their money and what are their emergencies. Why are they asking for this money right now? When you're working with the data from people, I think you need to think about this not only points in a spreadsheet or a data frame. This is people you're talking about here. I think recollections will be more important in the near future and also in your own company.

- Favio Vazquez: I mean there are some departments where I've worked that they said to me, "Yeah, I need a model that follow this different kind of bullets here, because we need to be able to understand the model and the full picture of it. I think if you want to do whatever you want, I think first you need to understand what you can and you cannot do in the company you're working with. Then you start working after understanding these restrictions.
- Kirill Eremenko: Yeah, I love the point. You mentioned this in the webinar as well. I was quite touched by it, that when you're working with data for us, any role is a data point, it's completely anonymous, we don't really, often we don't even know what's behind it. If you stop and you think about it every data point is actually a human being where feelings, emotion and life, their own privacy and that's important to always respect, always important to keep in mind. I think that's ...
- Kirill Eremenko: We try as a society, we're trying to put it into regulations, but if everybody just thinks of that and everybody just keeps that in mind, like half the problem will be solved. Half of it is data issues and privacy problem will go away. All right, well thank you so much for that overview of the five different characteristics of data science. I also like how you mentioned at the start, maybe let's recap this whole thing with just talking about that for a second.
- Kirill Eremenko: You mentioned that data science is starting to become a science. That's a very powerful thought, even though that it has the word science in it's name, it really hasn't been a science so far. It's been an ad hoc, put

this together, get some experience from this industry itself, borrow this from that industry and so on. Finally, it's crystallizing into a science of it's own. I will agree with you on that one.

- Kirill Eremenko: I also think that's the case. Tell us a bit more about that. What do you mean when you say data science is becoming a science?
- Favio Vazquez: What I think is exactly as you said before. This is a very new science and I really feel if we think about that. Right now we're very used to think of physics or chemistry as a science. If you go very back to Newton and Dalton and all these great scientists in the 1600s and the 1700s, they didn't know that was physics. I mean, Newton didn't call physics, physics, because there was not an established framework for doing physics. It took a lot of centuries for us to be able to say that yeah, this is physics, this is chemistry, this is biology and these are the boarders from each.
- Favio Vazquez: Right now we are in a different path, because this is very fast. I mean right now we are having a ton of papers being published every day for deep learning and machine learning and data science. At the same time people building the code for those papers at the same time too. The transformation is, I mean is very, very aggressive right now for data science. It's starting to become a science. When I say this, I'm thinking about, people are, I mean and I include myself here, we're trying to create a framework to develop data science projects with a methodology.
- Favio Vazquez: When you add methodology to one thing that is trying to understand a part of the world, I think you're trying



to do science in that moment. Right now we don't have the full picture, but we are creating it. I think Microsoft is doing a great work when they launch the model for doing data science. I don't remember their name right now. I think it's team data science workflow, I think it's a very cool thing people should read about. It's a full a dial framework for doing data science in a scientific way.

Favio Vazquez: And Matt Dancho's business science course I don't know if you've see him before on LinkedIn. He has a full picture on adding to [inaudible 00:38:46], the data mining pipeline before, the parts of data science to actually create a framework for doing data science. We're going to see more and more of these frameworks for creating data science in a scientific way. I think it's a very, it's a great opportunity for people right now to understand what science is about because not only, when I think ...

Favio Vazquez: I'm a physicist, I didn't, I mean when I was doing cosmology, I didn't sit down and say, "This is my hypothesis and I wrote it. I'm going to do observations now." I mean this is not the way we do science either, but we now, there's a method behind everything we do. If we try to apply that to data science, we're trying to transform it into a science right now. I think we're on that path.

Kirill Eremenko: That's a very odd observations. I like what you mentioned about Microsoft, I actually spoke to one of the, one the podcast before I spoke to one of the executives at IBM. He mentioned that they're building the same thing. They're building a, like end to ends

data science tool that will allow you to like from start to finish take a whole data science project. As you mentioned that's adding methodology to something that we know about the world to some way we're exploring data in the world.

- Kirill Eremenko: That is when something is becoming, starting to become a science there. There was a lot of it that way before that you know about the adding methodology, but it all makes sense. Thanks a lot for sharing that part. Before we move on to the next part of our podcast and talk about Latin America, I just wanted to make a quick announcement for our listeners, because just before we started the show I spoke to Favio and invited him to DataScienceGO and Favio said yes.
- Kirill Eremenko: Favio will, hopefully if at the start of the line you'll be joining us for DataScienceGO as a presenter and we're going to hear more from you there in October in San Diego. How are you feeling about that Favio?
- Favio Vazquez: I'm very excited. I mean I mentioned before that I haven't been to the US in like 17 years. I'm very happy to be back there and I follow a lot of people from there and they're doing amazing things, so I'm very happy to meet you in person and meet all of the people that is going to the conference.
- Kirill Eremenko: Actually you mentioned Mike Denture with how we transformed the crisp DM methodology into this, his own system that he's teaching. He's actually going to be there as well. He's going to be presenting about that, so that will be a cool opportunity for you guys to catch up. Have you met him in person before?



- Favio Vazquez: Nope, never, so that's awesome.
- Kirill Eremenko: Yeah that will be cool. Okay, so yeah, if anybody wants to meet Favio in person, that will be a great opportunity. Now let's talk a little bit about Latin America. Actually speaking of that, last year at DataScienceGO we had somebody from Brazil take a 25 hour flight to come to the conference. That was pretty, that was quite a good demonstration of how committed they are growing their career.
- Kirill Eremenko: Tell us a bit about Latin America. You mentioned that data science in Latam, I've heard that term before Latin America, Latam is different. What is your experience there?
- Favio Vazquez: I think first we need to define Latam as you mentioned. I don't know if people are familiar to where we are, but we are a setup conference that are most in South America like Venezuela, Colombia, Peru or Argentina. It also contains the countries in center of America like Panama and Nicaragua and Guatemala and also Mexico where I am right now.
- Favio Vazquez: Mostly it's Spanish speakers, but we also have Brazil, and that's a part of it this country. When you think about this we have one thing in common and it's our language. This is very different from if you think Europe. I mean, you have a lot of countries there and most of them they have their own language. They have their own religion and way of seeing the world.
- Favio Vazquez: Right now in America we're very similar. I mean I've only been to Colombia, Mexico and Venezuela where I can say that, we're not that different. That creates a

huge opportunity for data science, because if you solve a problem to one of our countries, maybe just open it for the whole region and this is not very common in all of these countries. If you solve a problem for Germany, imagine solving that for Sweden I don't know, just making an example.

Kirill Eremenko: Yeah or in like Switzerland, you solve the problem for Northern Switzerland, they speak German there. In western Switzerland they speak French, in southern Switzerland they speak Italian. Even in one country, there's like three different groups.

Favio Vazquez: Yeah, so it's very peculiar, but in the other side, I mean this is like the great part of being here on the other side, it has a lot of challenges. Not all of these challenges are related to data science, but some of them are related to how we live in these countries. My experience in Venezuela is not that funny, because we are a country with a lot of issues right now, and you cannot develop a career in data science there. It is not easy, because you don't have the tools, you don't have the money, you don't have companies that are willing to pay for you to be a data scientist. You don't have the masters, the education, you have low internet, so there's a lot of drawbacks of living.

Favio Vazquez: I mean for me when I was living in Venezuela it was very hard for me to get in touch with all of these different kind of technologies, people and things. For some reason I worked as a data scientist in Venezuela with a company called Move. We were trying to understand some patterns for a little company there,



and it was my first experience with data science in the real world, like working with data science.

Favio Vazquez: Then I came here to Mexico to do my masters, and I've worked in several companies here. I think the feeling is the same and when I say the feeling is like we're doing things that normally they'll see in books or people doing in the US and things. We have a bigger challenge to make them like visible for companies, because all of them they've heard of data science, but they're not sure what's data science or they don't have a department on data science in their company. Or maybe they don't trust a data scientist because they trust their 30 year old expertise in a field.

Favio Vazquez: We are right now transitioning into data science in these countries. That presents an opportunity for you to create a path as a beginner here, but it's not that easy to be around experts or learn from great people who work in these different countries. That's why right now, because I'm very young to be a Principal Data Scientist, but I am right now creating the department here where I'm working.

Favio Vazquez: I mean I came here and there was a department for VI and for some graphics inter glowing things, but there was nothing about machine learning or deep learning or anything like this. I'm trying to make the business understand what's data science, creating the whole department on data science here. There are a lot of challenges like money. You need a cluster or maybe a cloud computing service. Then you'll need people for you to be like you're helping these projects.

- Favio Vazquez: Then you'll need to convince like the managers they should invest in these different technologies and in you too. I think it's, I don't know are the things in the US or in Europe, but these are some of the challenges we were presented when we're trying to do data science in these countries.
- Kirill Eremenko: I've generally heard that the space of data science education is not that very well developed in Latin America. Like for instance people need to take courses in the US on like say on Udemy, Audacity and so on like in English. In a lot of cases, they don't understand English, they really don't have that English skill that developed that they can understand their whole course on data science. It will be great if the courses and education in Latin America could develop more in this space. What is your experience, because you teach data science as well?
- Favio Vazquez: All right, so one thing before I mention about that, right now because of the same thing you're seeing right now, I created a publication on Medium at [inaudible 00:48:48] or like science and data and it's in Spanish. What I'm doing there is, I'm translating my own articles from English to Spanish. I'm translating others people like the articles to Spanish with their permission and their permission on their name in the publication. People are starting to create blogs in data science there.
- Favio Vazquez: I think I'm trying to do the thing you said like we don't have that much resources in Spanish and this my way of trying to create something. There's going to be an announcement very soon for different kinds of things

where we're going to do in Spanish, so I keep updated, keep posted with that.

Favio Vazquez: My experience is, when I teach data science I think we have three different kinds of people. I think the first kind is people that only heard of it and they are interested in the subject. It's like okay, so I heard of that, because I read a blog or I saw your personal LinkedIn, then I started searching in the internet and I'm here.

Favio Vazquez: The second kind of people is that, they're business analysts and they know that they want, they know they should be changing into data science. This is something very common in these countries. People are trying to migrate into data science from business analysts, business intelligence people and data miners. They have a lot of skills are very similar to data science, but they lack a lot of the machine learning tools or the deep learning tools of the advanced mathematic like information. It's not the same way you can treat these people.

Favio Vazquez: The third group of people I've seen in my classes is people actually doing data science, but they're like beginners. These people they know what's open source, they are on LinkedIn, they read the blogs on data science, so they're very well informed. They're waiting for a letter from you. Then they're waiting other from you. It's very hard because normally you have all these group of people in one class, so you see it's very complex to try to understand different like a subject for people very different in what they want to hear or understand.



- Favio Vazquez: It's normally that easy to create a good course for these different kind of audience. I do always my best and I think the best thing I found is, you can work for someone else is, I normally start with a wow. Like okay, so this is what we're going to build at the end of the course. This is like, yeah, we're going to build this model and this model do this, this and this. Don't worry about the word models right now, I'm going to explain what's a model and things, but they get inspired.
- Favio Vazquez: Okay, that's very cool. I think that can work to my work. I can apply that to where I'm working right now and then I go very slowly into the theory and then into the code. I think a lot of people right now are very focused on the code part or programming part of data science, but the theory is so important. If you don't understand what you're doing, the code you're creating is just useless.
- Favio Vazquez: All of the courses I teach are like 50% theory and 50% code, because it's just very important to understand the basics of the things that you're trying to create and build. Then when you realize you understand all these different things, it just becomes very easy as a tool. Like programming is a tool for data science, but the actual core of the field is understanding the things you're doing.
- Kirill Eremanenko: Wow, what a great answer. We went from Latin American and the state there into education and data science. I'm just sitting here like mesmerized by all these things that you mentioned it's so good. I totally agree on this theory and the practical, that's why for

instance in our courses we do intuition and then the coding part, because they have to come hand in hand. You have to understand at least on an intuitive level what exactly is going on in these algorithms in which you're creating.

**Kirill Eremenko:** Also, what you mentioned about the three different groups of data scientists, I found this very insightful right now because you said there's data scientists. They just heard about this podcast or something and they're here listening to it. There's transitioning people from other industries and beginners. I'll mention this again about, because I think you'll find this exciting on DataScienceGO is we have exactly three tracks. We have data scientists, we have beginners and transition in a sense and we have also executive and [inaudible 00:54:04].

**Kirill Eremenko:** We structure our talks in exactly that way that you can build your own experience of the conference. Just to reiterate that indeed there's data science is so broad that you need to tailor the experience of who you're going to be representing, to who you're talking to that to get the most variance. Important to recognize whether it's a conference, whether it's a podcast, whether it's a course that you're creating, a blog post you're writing. Do you experience the same thing, like when you're writing a blog post, do you keep in mind who you're writing it for? Is it like the beginners, is it the experienced data scientists, is it the people transitioning careers?

**Favio Vazquez:** Yeah, I am right now creating blogs for these three groups. Some blogs are for three, for these three of

them, but there are some advanced blogs. I am creating like tools and I don't talk about like what's Python and the things. I have, I've created some blog post on like introduction to the learning or what intelligence or this is like a high level things that I've been thinking about. I wanted to have them on paper.

Favio Vazquez: When I read your blog on data scientists should write books, I think you post that a month ago or something like that.

Kirill Eremenko: Yeah.

Favio Vazquez: I think that's just very important. It's very important that we write what we are thinking. I think when I started writing, I mean I never thought I was going to write. When I read your blog on, that data scientists should write books, I think I'm not writing, I mean I cannot say yes or no if I'm writing a book right now. You have that in your mind, but I think it's very interesting, it's an interesting topic to mention because we should be writing what we think.

Favio Vazquez: I think data scientists have more so of the people I've seen in data science for some reason they have this skill for translating their thoughts into well written blog post. This is not easy. I mean my way of writing and the way I'm writing blog posts has changed a lot in the past year and I got experience and experience on okay, so I read a blog from last year and I'm not very happy on how I wrote this. I'm going to create a new one and change it or something like that. Do you realize that in the end this is very easy to write, it's easier to talk right now because sometimes when you write you can think a lot of what you're saying and you

can re-read and re-read again until you're very satisfied with what you're writing.

Favio Vazquez: I think right now we should be posting articles, posting blog posts, if you have the time, create books. There are great people out there that they will help you with the financial things if you want to create a book. Or maybe just for yourself, you have great libraries like Book Noun in R. It was very easy to transform this on R. Some R filed into a book with no effort at all, so I think we are on a path and we're going to be seeing a lot more of publication in a reading style for data science in the next years.

Kirill Eremenko: Fantastic. It's fantastic. Thank you Favio for sharing that. That's very inspiring and I think, I totally agree with you. I think more people should create content and write books, blog posts, record videos, make, share things in GitHub to help progress this space forward and help the community.

Kirill Eremenko: What I really like is that, people in data science it's so friendly, there's no, none of this back stabbing, none of this extreme competitiveness when people are just trying to get in each others way. Everybody is so friendly, everybody is willing to help each other out. It's a great place to be and I totally appreciate you sharing all these insights today.

Kirill Eremenko: Unfortunately we're running out of time. We're coming to the end of the podcast, so I wanted to thank you a lot for coming on the show today. Before we wrap up, could you share with us, where is the best place for people to get in touch with you and follow you in your career?

- Favio Vazquez: All right, I think right now the best way you can find out about me is following me on LinkedIn. Unfortunately I cannot accept more people like connections. I'm top in connections, but you can follow me and I'll just be with my updates all the time. Also, I think you can follow me on Twitter too. It's FavioVas is my username and also on Medium. I think I'm going to be a lot of more active on Medium this year, that I've been these past months. If you want to read more like a blog style or article style things from me, I think Medium is a good way to follow me.
- Kirill Eremenko: Got you. Okay, we'll definitely include all those links in the show notes and people can find them there. Just one question before I let you go today, what's a book that you can recommend to our listeners to help them advance their careers in data science?
- Favio Vazquez: There's a book I think is very helpful and it was for me. It's not that easy to read, it's like more kind of an advanced book in some way. It's called 'Bayesian Rationing and Machine Learning' by David Barber. This is a great book. It's a book where you can find definitions and code and also examples on almost every machine learning algorithm there is.
- Favio Vazquez: Also, something about deep learning, belief networks and probabilistic games and stuff. They are very common to data science. It's not focused on data science but the machine learning and the probabilistic thinking is amazing. It's somehow explained in the book and it's free by the way. The creator have it on his web page, he's called David Barber, he is the creator. He also has a full code solving for problems in

MadLab and in Python. His migrated into Python and Julia all of the code from MadLab.

Favio Vazquez: You can find all of the code for the book in those three languages. I think it's a very good thing and it's free.

Kirill Eremenko: Fantastic. I haven't heard of that book before, but it's called 'Bayesian Reasoning and Machine Learning' by David Barber and I hope everybody is interested in that and will check it out.

Kirill Eremenko: Well on that note, thank you so much Favio. It's been a pleasure meeting you and having this amazing chat. I can't wait to see in San Diego in October of this year.

Favio Vazquez: All right, thank you Kirill for the invitation and thank all of you for listening to what I have to say. I hope to see you in San Diego very soon.

Kirill Eremenko: Okay. Bye.

Favio Vazquez: Bye, bye.

Kirill Eremenko: There you have it. That was Favio Vazquez, Principal Data Scientist at OXXO. I hope you enjoyed this conversation as much as I did and learned quite a few things from Favio, including the five characteristics of a data science project which we discussed. Personally my favorite take away was when Favio mentioned that data science is actually finally becoming a science.

Kirill Eremenko: What that means is not in a way like that is becoming like a science, a chemistry or physics. That's going to be very specific and it's going to have its own field of application rather than being applied to business in a more broader tense. It's still going to stay the same and we're going to be able to apply to businesses. It

just means that, it's finally crystallizing into a field of its own. The way that data science came to be is it came from different areas, many different sciences and business included needed to work with data.

Kirill Eremenko: Whether it's biology, chemistry, physics, business, business intelligence, astronomy and so on. Everywhere there're legal concepts of data science were growing, but now we have something more common, more general that we can describe as data science and it's actually a science of its own. Well it's becoming a science of its own. That's a very exciting time to live.

Kirill Eremenko: As usual, you can get the show notes at [www.superdatascience.com/187](http://www.superdatascience.com/187). There you'll find any materials we've mentioned in this podcast plus the transcript for this episode, plus the URL for Favio's LinkedIn. Make sure to go there and follow Favio on social media and get access to all of the wonderful things that he'll be sharing in the upcoming future.

Kirill Eremenko: Of course don't forget that Favio is going to be a DataScienceGO 2018 which is happening in October this year. Just over a month left until then. If you haven't gotten your ticket yet for DataScienceGO, make sure to go and grab it at [www.datasciencego.com](http://www.datasciencego.com) and you'll get to meet Favio and lots and lots of other great data scientists in person there.

Kirill Eremenko: On that note, hope you enjoyed today's podcast. I look forward to seeing you back here next time. Until then, happy analyzing.