1:53

Alexey

**This week, we'll talk about data strategy. We have a special guest today, Boyan. Boyan is a data strategist with a decade of experience in various academic and commercial environments such as bioinformatics, clinical trials, HR tech, legal tech, and management consulting. He's the author of two books, one of which is Elements of Data Strategy and the other being Python and R for the Modern Data Scientist.**

**Currently, he is leading the data strategy function at Exxeta, AG, which is a large technology company that focuses on the German-speaking region of the world. He also runs a digital transformation consultancy. Welcome, Boyan!**

2:38

Boyan

Hello, I'm glad to be here. Glad to talk about data strategy – my favorite topic.

2:43

Alexey

**As usual, the questions for today's interview were prepared by Johanna Bayer. Thanks, Johanna, for your help. Now let's start.**

# Boyan's background

2:50

Alexey

**Before we go into our main topic of data strategy design, let's start with your background. Can you tell us about your career journey so far?**

2:59

Boyan

Yeah, I think that's a favorite question. My career, I think, took me to quite many different places. I think, especially for your audience, it's interesting to hear because I was a data scientist before there was such a thing, I would say. It was right on the edge of that. I actually studied biochemistry, specifically cell biology. After this, completely by accident, I started to do a Master's thesis, which was a bit more computational. This way, I started to code. Around that time I wondered, “What should I do with my life?” and being able to code, knowing computer statistics. Then I heard that in the US, they have this title of “data scientist,” so I thought “Let me try Germany.” Because I studied in Germany, normally, people really care about your background. I was very lucky that the startups didn't. [chuckles]

The startups looked at me and said, “Yeah, actually, if you can code and if you know things about data, you can be a data scientist.” I got the title, I got the job, and I spent the first half of my career being a data scientist for all those kinds of different startups, mostly. That was fun. I learned a lot. I also learned how *not* to do things, which I think is also useful. Kind of towards the end of the second half of my career, it was a bit like I had another accident happen to me. I saw myself as more of an introverted person and going into consulting wasn't an obvious thing for me, to be honest, but I tried and I loved it. There I got the very strange title of data strategist, which even at the time people didn't understand. Now, I think there's still a lot of work to be done there. Hopefully, I can help a bit to do that today.

I spent some years there. I did a data strategy for somebody (for a company) and they invited me to be the CTO after that. It was a funny thing because it shows you that if you do data strategy, then you can easily transition to a different role. Also, we have a saying in Bulgarian that means “You have to eat the soup that you cooked.” It was funny, because I had to make the data strategy, and then I had to go and implement it. Of course, it's very challenging, so we can talk about this today as well. So I was the CTO of this company that was more focused on data business. That was really obvious that a data strategist would fit.

At some point, though, we weren't successful – unfortunately, that's how many startups are. I really wanted to go back to consulting. This is around the time when I finished my second book, which was about data strategy, and I thought, “Okay, now I really have unfinished business there.” This is why I'm back in consulting at Exxeta, where, as you said, I lead the data strategy function here and work with all kinds of clients on the topic. It is by far my favorite topic – data strategy.

5:47

Alexey

**So how did it happen that you were called a data strategist? You said that (or maybe my impression was) that it happened kind of by accident?**

5:56

Boyan

Yeah. I went into consulting by accident. At that company, they had data strategist as a title. When I looked at it at first, I thought, “Well, okay. That's interesting. I'm a data scientist. Is this something for me or not?” At that point, I started to realize, “Actually, I am comfortable and enjoy the things around data science.” So I thought, “Let me try this out.” There were a lot of unfamiliar words, to be honest. This is the reason why I wrote the second book, because I got the title and I started to ask, “Yeah, what do I do now as a data strategist?” [chuckles]

You get the requirements from your company, which is something like, “Make the data project successful. Go have fun.” Which is everything without the coding part. That was challenging. The role wasn't super-defined. It was kind of this “Purple People” as you often call them. There was a mention of that. I heard this term before. There was an article, I think, from DVC (data versioning tool). One of the people there mentioned that they're these “purple people”. I don't know why that's the name, but these are people who are kind of in between technology and business.

7:14

Alexey

**Like product managers, right?**

7:15

Boyan

Yeah, kind of. Like data product managers. When I wrote my book, I asked the question, “What do you call such people? Is data strategist the right title?” And I haven't found a better one. I think it's cool to say data product manager. I think that kind of feels right. Because actually, data strategy is not all we do. That's the conclusion that most people come to. They think all we do is data strategy, while 80% of the work is actually implementing a strategy. [chuckles] But it doesn't come from the title. So that's how it happened.

That's how I became a data strategist. I became a data strategist, but the skills weren't defined. It was really a situation where the requirement was “Get value from data without coding,” basically. I had to find my way and this is actually why I wrote the book, to try to define it a bit. I actually did this Venn diagram of skills. You know the data science Venn diagram, right? That was the original thing. In data science, we had the same thing, if you remember. Many years ago.

# What is data strategy?

8:13

Alexey

**Yes, I do. So, data strategy is getting value from data without coding, right? [Boyan chuckles] Can you give a more precise definition? What is a data strategy?**

8:27

Boyan

A data strategy is technically a plan to get value from data. This plan has to be actionable, and it has to be flexible enough to be changed when in operation. This plan has to contain a lot of information. I think you often see on LinkedIn when people say “data strategy,” and you have a list of goals. A list of goals is not a strategy. This always obsessed me when I saw that. Strategy is normally a very big document. That's actually a dangerous thing. But in this document, what do you... I can talk about the phases because the book explains the process of doing data strategy in phases. But one important thing to say here is a strategy cannot be static. That's a big mistake.

You have some kind of plans for your data teams, for your data products/projects, and then things don't work out. This is a very typical scenario. Somebody creates a data strategy and nothing happens. So I'm very focused on making a data strategy that is flexible and iterative. You don't make a static document at the beginning. Normally, it's a deck, actually. You have to really be able to change this. There are a lot of artifacts that are actually connected to that deck. I can give you one example – a data dictionary, for example, would be a part of the due diligence of a data strategy. You can imagine this deck, and you have a big conference page, which contains a dictionary of the data associated with use cases. This is just an example of how specific you can get.

# Due diligence and establishing a common goal

10:13

Alexey

**It's still a bit abstract to me. Okay, we have a deck, which is connected to strategy. I know that strategy is a big document. It can't be static. Then there are some other artifacts like a data dictionary. But maybe you have a more concrete example? What exactly is in this deck?**

10:31

Boyan

Yeah. Maybe I can go into the process of data strategy. You can't tell a company what to do if you don't know where they are. That is a classic mistake. You hire somebody to give you a plan, but you can't make a plan if you don't know what's there already. This is a very classic example, especially in big companies. You have to spend a lot of time at the beginning figuring things out. As you know, when you do data science – let's say a machine learning project – you can't just say, “Yeah, we need to make a churn model,” and then wait for success. It doesn't work this way. You have to see what data you have first. That's the most important thing.

So that's the due diligence part. That takes a while. The due diligence part is not strategizing, it's really mostly about figuring out where the company is. First, you have to find out what the goals of the company are. I will have this discussion today with my team, actually, “How do you deal with your clients?” Because every business has some kind of point, right? It wouldn't be a business otherwise. Let's say that you're a company that sells socks (as a basic example) and you have a data team. You can imagine that the goal is to sell more socks faster. Taking this... [cross-talk]

11:47

Alexey

**Sorry for interrupting. So we have a company that sells socks, the goal of the company is to sell more socks [cross-talk]**

11:55

Boyan

More socks, faster. Let's make it a bit more. [chuckles]

12:01

Alexey

**That's clear. We also have a data team in this company, whose goal (because they also work at the socks company) is the same goal as the entire company, right?**

12:10

Boyan

It should be. But this is where the strategy comes in. Because normally what happens in these companies is, the C-level people think “Let's hire a bunch of data scientists and tell them 'help us sell more socks faster, at a better price.'” That's literally how it is, actually. The data scientists are then like, “Yeah, but the data and this and that and that. We need data engineers. We need some kind of cloud platform. We need a budget for this. What type of skills do we need?” This becomes a mess very quickly. This is really what's happened in Germany. After this, companies realized “Oh! We do need data engineers as well!” Because somebody needs to make the data available and somebody needs to set up the infrastructure.

And now what happens is, companies realize “Oh! We need data strategists as well.” Because somebody needs to see, “How do you go from selling more socks faster?” To, “Actually, if you do a churn model for a socks customer, that somehow contributes to that goal.” And this translation of how the more technical impact contributes to the business one – that's very hard. It's really hard to find the right thing to do. So that's that part.

# Designing a data strategy

13:28

Alexey

**I think I understand, more or less, but I'm still a bit confused. So let's say we have a goal – to sell more socks faster. What exactly goes into the strategy? Because to implement this model, we need to have the data infrastructure, the actual data, the data engineers, we need to have the platform, we need to know who to hire, etc. Does this go into the deck with a strategy?**

13:57

Boyan

Yeah, correct. I think I have to jump a bit forward and then go back again because now we're going into the design phase. Let's say that during the due diligence, you found out what use cases there are, whether they have a cloud environment, what the data looks like, what type of people there are, etc. This is seeing what's there. Let's say you know where the company is and what there is. Now we have to design the strategy. The first thing you do is use case ideation. You basically take a look at what use cases are possible. I mentioned the churn model. That's one use case. But maybe you can make some kind of recommendation engine for the online shop to sell socks. That's another use case.

You can probably find 10-15 for any business. So that's the easy part. [chuckles] Because to come up with ideas, you do need to have some experience. Again, this is actually the data strategist's work. You do need to have seen many use cases. But then, what becomes very hard... ChatGPT can actually help you a lot here – you could technically ask, “What are the use cases of selling socks”. That example may be stupid, but you can also say “What are the use cases of selling flowers, (or chairs)?” [chuckles] “How do we use data science to sell chairs?” And you can get a lot of nice ideas from that. But here comes the hard part (and explains why a data strategy needs to understand technical things) and that's the feasibility part. It's easy to come up with ideas, but understanding what's feasible is very hard. Here, you take the results from the first phase (due diligence), you look at them, because you now know what data there is, what the skills of the people are – you know all of this. We can say, “For the recommendation engine, we have the right people, we have the right data, we have the right infrastructure. Let's do it.”

There, you have the prioritization part as well. Because then you need to see, “Okay, do you make a recommendation engine for the online shop for socks, how important is this?” Then you take the feasibility, the importance, the business impact, and then you prioritize. That's a bit easier, normally, but again, here, the hard part is actually the technical understanding. This is what most companies still mess up. You create some ideas, you create a strategy, and then when the people start to implement them, it's not feasible. It is very hard to know what's feasible or not.

16:17

Alexey

**And for that, you really need to know what's there – what kind of data and all these things.**

16:21

Boyan

Yeah. This process – ideation, feasibility, and prioritization – this triad is the standard thing. To be honest, after this, you do need to do several other things that come into the data strategy. You have to set up the target architecture and technology. Let's say that now we know what the use cases are. Here, there's something I want to talk a bit about. It has the fancy name, “influence cascade”. I'll give you a concrete example. Let's say we have a churn model, and we use tabular data for that – very standard numerical stuff. We build the whole build API, we build the whole product, and it's done.

Then, for whatever reason, it's not very successful – because we didn't take the right data, the right people weren't involved, we didn't know how to measure this, whatever – there are so many reasons why it may not be successful. Then you have to go back. Let's say, we take text data – this “scope creep” in data products happens all the time, but if you take text data, suddenly you have other problems. That's why it's called “cascading” – the product people often think “A small change here. What's the big deal? It's just another dataset?” Oh, no. Now you do need more storage – maybe the data is PDF. [chuckles] You suddenly need NLP people. And then the whole thing breaks down straight from the beginning.

This is why you do the use case stuff first, and then you do target architecture, because if you change the use case, you change everything. That's why that's the first step. But this is the design process. There are a few other steps there like data governance and the operating model, which are a bit more standard. The operating model is basically how people operate, “Do you set up a hybrid team? Do you set up a centralized “Center of Excellence” team?” All of that.

# Impact assessment, portfolio management, and DataOps

18:22

Alexey

**So the goal of this deck (of having this data strategy), or the output from the data strategist is: join a company, first understand what is there, then come up with some cases, come up with this targeted architecture and all these things, like designing this thing, thinking about data governance (the structure of the people), operating mode. So we have this deck. What happens next?**

18:56

Boyan

That is the million-dollar question. Because even if you create a great data strategy, you have the right use cases, the right people, and everything is correct – it is static, while the world is very complex. By having such a deck, you have already fixed a lot of problems that would happen anyway. People are supposedly going to work on the right things – the right people at the right time. You have a roadmap as well (I forgot to mention this). You're ready to start working. But this is the whole third part of the book, actually, of my process, which is delivery. I'll be honest with you, this is the hardest part. I personally cannot figure this out completely because this is where you face reality. There are three elements here that I have added in my model. One is impact assessment. Another one is portfolio management. And the third one is DataOps.

Let's start with the first ones, because they're less controversial – impact assessment. You do need to set up some kind of a baseline. Let's take the socks again. Let's say you have a data strategy – you're in the sock-selling business. Remember the initial goal, selling more socks faster. [chuckles] That should be measurable with any good business. Unfortunately, often it's not, but let's assume it is. Let's say the business unit measures that. It's a big assumption, honestly, for many companies, but let's say they do measure that. Then you really have to have a good look at your data team from time to time. Really, every several months you need to take a look at your churn model, the recommendation – there are four or five use cases that are now being worked on and deployed. Does it move the needle somewhere? That's very hard.

But you do need to have these check-ins, because then you can adjust. Let's say, people spend a lot of time on the churn model, and then you see that it does improve the selling of socks because fewer customers churn, but it's hard to measure. Here, you can have a lot of statistics, A/B testing and all of those things you can do. But the price of us maintaining this is not worth it, so we have to refocus on something else. Therefore, we completely scrap that use case, even the part of the architecture moved to something else. This is probably the result from the impact assessment.

21:19

Alexey

**Do you actually do this after it's implemented? Let's say we implemented this churn model. We ran it, we did the A/B test, and we see that there is a positive impact. But then we also did some calculations and saw that the cost of having this architecture is outrageous. It's just too large. It's not worth the impact it gives us, right? Then we say “Okay, let's scrap it.”**

21:42

Boyan

Yeah. This is the portfolio management part. You have several use cases – that's the hard thing. You have to take a look at all of them at the same time. It's a qualitative measure, unfortunately, I think. There are hidden costs – supporting the teams is expensive and other things. You can do it, obviously, immediately after the start. But for most data projects you can do a pilot in a matter of weeks and then you can start measuring. But you do need a baseline because what happens is (people forget about this) if you don't have a baseline or you just start working on it – you didn't set up some kind of metric at the beginning. You think, “Okay, let's see. We're going to see the results.” You get the result at the end.

At some point, you measure the churn rate, but you didn't know what the churn rate was at the beginning. So you go to the business owners and tell them, “Yeah, we did this.” And they say, “So how bad is the situation compared to before?” So you have to set this up at the beginning, which is really something that is very easy to forget. Before the project, you need to look at the business metric and say, “That's the number.” And then two months later, you can really say “Yeah, we improved this by 2%.” Here, the hard part comes as, “Which business unit contributed the most to this?” But at least you have some kind of a start.

23:06

Alexey

**Yeah, it's tough.**

26:08

Boyan

Yeah. [chuckles] And DataOps, maybe we can talk about this as well, because this is very... [chuckles] I think it's controversial, because when people hear the term “data product,” people think that about a data product – Data Mesh is defined in one way, people think about the data assets. So there are very different things. What I mean by DataOps...

# Data products

23:30

Alexey

**Maybe we can start with the data product, because this is what data strategy is about, right?**

23:37

Boyan

About data products? [chuckles] Yeah. I did change my mind on this one. I wrote about data products, but to be honest, it should be data- and AI-powered products. A data product is something else. A data product should be just data. I made a bit of a circle and I just joined the Data Mesh definition of that. It's just the data itself, without the use case. My view in the book was that a data product is the data plus the use case, but I don't want to compete with the more popular definition, which is just the data and the architectural components around it. It is with the data quantum and the data mesh, where it's a bit bigger. But what I talk about is data- and AI-powered products. It's like the API endpoint for a machine learning model that predicts churn in Salesforce. That's the data- and AI-powered product with an AI feature. When I say “data product,” this is what I mean.

24:37

Alexey

**So you have some data and then there is a specific use case on top of this data.**

24:42

Boyan

Yeah, that's exactly correct. That is the data- and AI-powered product. Technically we should say “data- and AI-powered product managers,” but it's a bit of a mouthful. But regarding a data product, we should leave it at the data level, without the use cases, I think.

# DataOps, Lean, and Agile

24:57

Alexey

**Now let's talk about DataOps. But then there is also an interesting question that I want to ask you later.**

25:03

Boyan

DataOps is a tricky, tricky term. People mean different things when they say it. Again, in the book, it's a combination between Agile and Lean – it's just the methodology with a CI/CD angle to it. To be honest, there are a lot of ideas there. There's one book recommendation I have that I'll give at the end as well. The book is called Practical DataOps. The idea is – I mentioned Lean and Agile. “Does Agile work in data science?” is my favorite question. The short answer is “It doesn't.” [chuckles] But maybe it's like democracy – not great, but it's the best we've got, you could say. But Agile in data – in the book, I have these interviews with practical data leaders and they all said, “Yeah, you can do Agile within a software.” We have to be careful about requirements, because it's very hard to measure certain things, especially at the beginning.

I will leave the Agile topic a bit because the Lean part is what's fascinating. And that book, Harvinder Atwal is one of my favorite authors. That book is brilliant. He took the ideas from Lean manufacturing (cars and conveyor belts) to data. The focus there is not on doing the right things, but avoiding the wrong things. This is so amazing in data science engineering, because it's hard to know what the right thing is. But everybody knows the wrong thing. Basically, he focuses on how data people waste their time and energy and literally how much time data scientists spend waiting – really concrete, measurable things and just avoiding them. That's brilliant, I think. It's a brilliant way to manage data projects. You don't focus on what's good, you focus on what's bad.

27:07

Alexey

**By the way, I think we have two podcast episodes about DataOps, so you can check them out for more details.**

# Data Strategist vs Data Science Strategist

27:18

Alexey

**The question I see is quite an interesting one, because I also wanted to ask you, “What is the difference between a data strategist and a data science strategist?” What you describe, to me, is really data science-centric, but data is not just about data science, right? We have use cases, we have dashboards, etc.**

27:43

Boyan

Yeah. There's the idea (and some very nice articles) about how data strategy is going to split into flavors. “Data strategist” is a more general flavor. You could have an AI strategy, which is somebody who focuses on just that part. A data strategist is just the more general idea. Because then you have data platforms, data engineering – it also has value. It's a bit harder to measure the success of those, because they're kind of enablers, indirectly.

There's this hierarchy of data science needs model from Shopify, where you have the advanced use cases on the top, but you can't get to them if you don't have things like a data platform and cloud services infrastructure. The data strategist is responsible for that part, too – setting this up. Especially in bigger companies, that's a bigger problem. So I will say there's a difference. It's just the more general one. There's also business intelligence reporting – that's also there.

28:45

Alexey

**It's just a bit more difficult to measure, but I guess we need analytics to be able to measure something, right?**

28:53

Boyan

Sometimes it's a gradient. I do think that we shouldn't try to measure everything. It's impossible. You should at least have an idea of where things are going. But that's why I like this. A typical example is with these lighthouse projects, when the companies want to do something very fancy with AI. Normally, it's actually the last thing they should be doing. This idea is from Martin Szugat from Datentreiber, who I interviewed for the book. He's a good friend of mine who really has brilliant ideas on this. He told me once, about lighthouse projects, that they're really the last thing a company should do because usually, it's some very advanced use case. It's some kind of NLP, computer vision, deep learning use case, because that's what gets the media attention – it's amazing and fancy people like it. But to do this, you need crazy architecture, crazy skills, crazy everything to move beyond the pilot. What you should do instead is get the most boring use cases, like reporting, automation of reporting, etc. [chuckles] Get it done, show the number, get the money from the business owners, and *then* do the more advanced things. That's his idea, not mine, but it's a brilliant one.

# The skills one needs to be a data strategist

30:02

Alexey

**Let's come back to the time when you were hired as a data strategist. As I understood, your background was data science? [Boyan confirms] What skills did you need back then to actually be able to do this role? What kind of skills did you need to show that you have and what kind of skills did you need to develop when you joined?**

30:21

Boyan

For this, I can refer to the Venn diagram. There's an exercise I did with my colleagues back then, because not just me, but the other data strategist – we all had the same question, “What skills do we have?” We had some people who were business people – completely business – people who just never [inaudible] and they are also data strategists. That's why you should have different flavors. I group it into three groups. Imagine a Venn diagram with three circles, where you have data, and you have communication systems thinking. Data is the whole tech thing. It's data science, engineering, analytics, cloud, and communication. Everybody will tell you that the last one is *the* skill. And it is *the* skill – Can you write? Can you talk? Can you explain? Can you translate? This is the hardest skill.

The ultimate data strategist skill is translation. That's why at McKinsey they call analytics translators, which I like as a title. It doesn't sound as cool, I think, as a strategist, but you do need to translate from the requirements “sell more socks” to “churn model” and beyond. You have to operate on very different levels. Systems thinking, this is the most abstract thing, which many people would roll their eyes at when they hear this, because it sounds like a management consulting thing. In the book, I describe it a bit more, but I'll give you an example.

A typical day of mine is when I have to explain what AI is [chuckles] and all in the same day, I have to look at access permissions on Azure and explain how to do that part – in the same day. That's the hardest part about the job because you have to move between those two levels. I don't think you can have a junior data strategist, almost, because it's something you need to get from those different domains. But you do have the flavors of people.

32:24

Alexey

**Do you think it's important to have these technical skills, like explaining roles in Azure?**

32:31

Boyan

This is kind of an example I sometimes give in terms of understanding (let's use a data science example) something like support vector machines. Let's take that one. Most people wouldn't understand the math behind it, but if you go on Wikipedia and look at their nice diagram of separation of classes, where you have the decision boundaries. You read the Wikipedia article and you see “Oh, well, there's this kernel tricky thing. There's this trick to maximize this boundary.” Do you understand how support for vector machines works? No, you have no idea. You can't code it yourself. But you understand it conceptually, kind of.

This is the same level you need to be a data strategist. You need these concepts of architecture and you need to understand what access is, what pipelines are, what orchestration is, etc. You really do need to understand this. It's hard for me to say – I come from a technical background. But that was as a biologist. Yeah, you do need to have conceptual knowledge of everything in data, for sure. I think otherwise, it's very hard.

33:34

Alexey

**But you don't have to be an expert in things like orchestration or Azure?**

33:38

Boyan

No. It is frustrating. On a personal level, I can tell what's frustrating about this, because everybody now is going to think “Let me become a data strategist. It's all roses and unicorns,” You will feel stupid all the time. You will be talking to people who have 20 years of experience in data architecture in the same day with people who have 20 years of experience in sales and business. And with both groups, you feel stupid, because you're this weird jack of all trades that kind of knows a bit of everything, and you will be judged from those two groups on their level. So you have to prepare yourself mentally for that. But on the plus side, you get to do many different things, which is totally worth it, I would say.

34:17

Alexey

**Okay. So again, the skills you need are mostly communication skills – you need to be able to write, read, communicate, and translate, right? Then you also need to have some technical skills. You don't need to be an expert, but you need to know what a data pipeline is, what an orchestrator is – all these technical terms.**

**Somebody will need to implement this, so you need to know exactly what it is and how much is required for that, or you may even need to know which tool is going to be used. For example, “If we are building a data pipeline, we will use this particular tool.” Right?**

35:01

Boyan

At that point, you can get advice, I think. For example, with orchestration – Airflow versus something else – the second part, you don't need to know, I think, because tools change every day. But you do need to know how to orchestrate some scripts and what that looks like, and then you find a more competent person to help you there.

35:20

Alexey

**Okay, so a data strategist is not necessarily a data architect.**

35:24

Boyan

No. [laughs] No.

35:27

Alexey

**Somebody will work with you on the strategy to actually come up with a target architecture, right?**

35:36

Boyan

Everywhere you have somebody working with you. On the business side, you will also have somebody that will work with you – on every other side. It's a role where if you're alone, it's totally pointless. This role is only an enabler. It's completely pointless without other people.

35:50

Alexey

**Does this mean that a company that wants to hire a data strategist already needs to have all these people in place, *before* they hire a data strategist? Or does a data strategist come with a team? [chuckles]**

36:07

Boyan

[chuckles] It's hard to say. If you're a smaller company that's just starting out, you do need a data strategist, but they would be called a CTO, Chief Data Officer, or Head of Data. I think a Head of Data Science has to be a data strategist. A Team Lead of Data Science has to be a bit of a data strategist. This is kind of a more managerial role. A Team Lead has to be a data strategist, actually. And you do need to hire them at the start to know what you're gonna do, I think.

# How does one become a data strategist?

36:35

Alexey

**Okay. Let's say I work as a Senior Data Scientist. I listen to this podcast episode and I think, “Okay, this is so cool. I want to be a Data Strategist.” How do I go about that?**

36:48

Boyan

You have to learn a lot of business. You *really* have to learn a lot of business. And you have to let go of a lot of things. You have to start to understand that... Because, as I said, you will feel stupid. This will be new to you. And if you're a Senior Data Scientist, at that point in your career, your ego is already high. You likely think, “Okay, I have learned so much that I can do anything.” Well, now you start from scratch because you suddenly have to understand how businesses work. It's easy to underestimate, but you do need to start from the beginning.

First, understand how businesses work – talk to the functional leadership in your company, meaning the Head of Marketing. Go to them and ask them, “How can data help you?” This is the first start because then you start to understand what they want. Try to align your churn model to the socks selling, right? [chuckles] Talk to the people who sell the socks. You work in that area. Of course, I will say to get the book. The book won't make you a data strategist, but it will give you the typical things that you need to do – the activities. But you have to go and actually do the work. You have to be comfortable stopping being a data scientist because you will work with very vague things. Normally, these are uncomfortable for technical people who want to be specific, I would say.

The worst of it is, you have to work on different levels in the same day. This is also what's going to happen if you transition to that. Because as I said, in the morning, you're talking about business, and then you have to write some Python in the afternoon. That becomes a problem very fast, I would say.

38:36

Alexey

**Why?**

38:38

Boyan

I think because of the focus you need to prepare different things in the same day. If it's in several different days, I think it's great, because your brain is active on many different levels. But in the same day, personally, for me, it was always challenging to switch the focus that fast. Because when I think about business, I think in an abstract way. I also get interrupted operationally. But if you have to write code in the same day, I think it's a challenge. But between the days it's amazing. If you can be a data strategist for two days, and three days a data scientist, maybe that's a good idea.

# Data strategist as a translator

39:09

Alexey

**Coming back to your point of learning more about the business (or learning a lot about business). You mentioned that you need to talk to functional leaders, such as the head of marketing, the head of product, maybe – all the people who do something. You come to them and say, “How can they help you?” But they have no idea and think, “Who is this person?” So how can you make it clear what exactly you can help them with?**

39:36

Boyan

Well, then just ask them what they do and then comes the understanding. You ask them what they do – they sell socks. They start talking about customers and then you get this idea, “All right, maybe we can do a customer segmentation model.” You have to read about the use cases in your area, which is easy. If you're in a standard area, it's easy. It's hard if you do something very specific. But then, if you think about use cases and already prepare before those meetings, then you can give them those ideas. They will ask you, “Okay, tell me what a customer segmentation model is.” And you have to be able to translate that.

You will mess it up the first time. [laughs] Because just by saying “customer segmentation model,” you already messed up. You should say, “Yeah, we have this way of separating the target groups and we can identify them.” If you say this, there you go. But trust me, the moment you say something and you feel smart, they feel not so smart and they don't understand. This, actually, is the biggest challenge – communication.

40:39

Alexey

**Okay. And the only way to learn this is by doing it, right?**

40:45

Boyan

Yeah, the only way. It's like how people should learn to play guitar – you can just play the guitar, but you have to have deliberate practice. You have to do it deliberately and put yourself in uncomfortable positions. You have to translate. You really have to deliberately learn and it's not easy, but the results are amazing. To be honest, yes it's hard, but on the plus side, you will be very active in supporting the business. You work on many different projects and many different things. Your brain will always be excited about new things – new technologies. As I said, this is a gateway to other positions. I became a CTO after that, for example.

# Transitioning from a Data Strategist role to a CTO

41:31

Alexey

**By the way, how did this happen? How exactly did that transition happen for you?**

41:36

Boyan

They really liked the data strategy. That was the biggest compliment to me. If you like the data strategy – and it was just a data strategy, it was not a technology strategy. I had to do the technology after that. That's very different. But they liked my approach. They liked that it was concrete. They liked these translation skills. These translation skills are the hardest thing in technology. I think with the GPT stuff coming up, where do we go to escape from that? And I will say, we should go to those middle translating positions, because that is very hard, I think. Very few people can do this, I think – translate. That was very attractive to them, because as a CTO, that's what you do all day. You translate. [chuckles]

42:21

Alexey

**Can you tell us more about the skills you needed when you were a CTO and how it was different from being a data strategist?**

42:28

Boyan

The skills... the responsibility is much higher as a data strategist, of course. That's a bit obvious. As a data strategist, you're in your data corner, and it's kind of a consulting role. That's another thing for people transitioning to that – you're an in-house consultant, basically. You can see yourself in that way. It becomes different when you are an owner, and you own the whole topic. Because, suddenly, it becomes very operational. Everything you do becomes very operational, you kind of immediately suffer the results of your mistakes. I had to learn even more about the business, obviously.

I knew some things, but I had to learn even more. I had to learn accounting. [chuckles] You have to learn budgeting much more. There's a lot of management, of course. It is a very, very different role. But being a data strategist helped me prepare because my ego was ready. I was comfortable not knowing. So when I needed to do budgets, hire or fire people, I felt like, “Okay, I can handle this somehow.” Because as a data strategist, you already feel a bit weird – you don't know different things already. Then you can transition to do something totally different.

# Using ChatGPT as a writing co-pilot

43:46

Alexey

**Interesting. Then there's something I really wanted to talk to you about. For your book, you used GPT, and that's really nice that you explicitly acknowledge that.**

44:01

Boyan

Oh, that's necessary.

44:04

Alexey

**That's necessary, but also like I'm just thinking, like “Would I feel comfortable saying that all the content I write is generated by ChatGPT”? I don't know. Okay, but this is necessary and I wanted to talk to you about that. How did it actually happen? How did you organize this process? How did you use GPT in the book?**

44:27

Boyan

There are different things and I will link to somebody else who did something even next level to that. So, it was at the beginning of this whole craze, while I just used it in several areas, only in the sidebar. For example, at some point, I need to explain what mittelstand is in Germany, which are the small/medium-sized companies that carry the economy. When I was writing the book, I wanted to write about strategy systems and thought, “I don't want to go to Wikipedia or wherever to find a definition and paraphrase it. This is totally a complete waste of time and energy. Absolutely pointless. Zero creativity in doing that – to write the definition of mittelstand in a smart way. No.” So that's why I have these sidebars – only there. To be honest, you really have to say that you did that, because it's very obvious. I think as time goes on, it's not gonna be so obvious.

I used it just for the sidebars. I have this robot icon there. And for really boring things, which I think otherwise you just have to paraphrase. This is what it's very good at. But ethically, somebody has to write original content to train those models. Because we stopped feeding Stack Overflow. I don't remember the last time I was there. Down the line, it's gonna stop feeding the generative AI models. If nobody produces original ideas, we're gonna have a delayed problem later. That's the big ethical, number one problem of this whole technology.

The thing is that when we stop generating unique content... you can argue for the uniqueness of GPT mapping between ideas, which I do use in my daily work, as I mentioned. Let's say I want to find use cases in weird areas, maybe something like a government agency that cleans the city. How do you do data science for them? Instead of me spending one hour, I will just ask ChatGPT “Give me five use cases,” and then I go deeper. For this type of stuff, yeah. But I think really original things, we're a bit far off from there. My biggest concern there is the ethics of doing that. I can refer to Christoph Molnar, who's really big in the Explainable AI field. He recently published an article about how he used it as an editor for his book.

46:47

Alexey

**I was going to ask about that, actually.**

46:48

Boyan

Ah, okay. I didn't do this. That was before. It was a bit early. Now you can do it. I think in the future, I'm open to doing that. I mean, I do use Grammarly, for example. You could argue Grammarly is actually pretty intelligent, too. I mean, now they also have generative AI features. But you could also argue if you use Grammarly, is this ethical or not? So it's a bit of a blurry line there, but I think as an editor, as a definition engine – great. But anything else? I just don't see it happening yet. But I'm not saying it's not gonna happen. [chuckles]

# Using ChatGPT as a starting point

47:20

Alexey

**So one thing I can think of is, for example, say you want to write a chapter. For me, I always have this problem of a blank page. How do I start? You can just ask GPT or ChatGPT, “I want to write a chapter of a book about this and I need you to help me come up with an outline for the chapter. In return, it gives you “You need to have these sections and each section should have these subsections.”**

47:57

Boyan

Absolutely, absolutely.

47:59

Alexey

**And then for each subsection you have a list of bullet point lists of what you should cover in this. It's so immensely helpful, provided that it is correct, right?**

48:11

Boyan

[chuckles] For sure. I do write a lot – a lot of PowerPoint, as you can imagine. I have a blank PowerPoint, where you need to say three bullet points on a topic – it's brilliant there as a start. I haven't seen it yet where you can just copy and paste it. There's a big difference between 3.5 and 4. I have to say 4 is really getting there and Microsoft Copilot is coming soon. I think in PowerPoint, you're gonna get some really nice new things. It's as you said, at the beginning, as an outlining tool, I think it's great. But for original thought, we're a bit... I don't know, it's gonna be a very scary time.

48:47

Alexey

**Yeah, but for a technical book, how original do you need to be?**

48:52

Boyan

That's true. For a very technical book, if you want to explain it, yeah. Because at the end of the day, there's a philosophical argument, “Would you buy a book which is written just by that? A purely software-created one?” You should, because a technical book is not about convincing you that much. You already bought the book about the rest, right? But you do want to learn about the rest. If that's the best way to learn it? Come on. I would read it. I wouldn't care. Especially if you can't notice the difference in cadence, because you do notice it's a bit monotonous. The text is not exciting.

49:26

Alexey

**And you need to know that you can trust the book.**

49:29

Boyan

Yeah, that's for sure.

49:32

Alexey

**You would have to check the book, if nobody edited and nobody proofread it.**

49:35

Boyan

Yeah, that's for sure. And the reference. But that's coming, I think, with the references. Bing does that a bit as well, so I think that's alright.

49:45

Alexey

**Do you think this is the future of writing books?**

49:49

Boyan

Of co-piloting your writing? 100%. Writing them? I don't think so. We're a long way from that. Non-technical books, at least. Technical books for sure. [chuckles] I think we're gonna have a lot of this. If you just have a manual/tutorial, that's different. What is a book then? Then it becomes a tutorial, not a book, arguably – philosophically. If it's just a tutorial that's printed, then you can get this. I'm not saying it's less valuable, but in a real book, you still need a human there as a main pilot, for sure.

50:24

Alexey

**Many people. [Boyan agrees] One person is often not enough to write a book, right? [Boyan agrees] You need an editor, proofreader, whatever. By the way, speaking about Christopher Molnar, he will do an Ask Me Anything quite soon at DataTalks.Club.**

50:42

Boyan

A brilliant person. He had an article about Explainable AI, and he's done so much good work for the field. Brilliant. Underrated. I think people don't know enough about his amazing books. He has several really very strange and specific books, which every data scientist should read, I would say.

# How ChatGPT can help in data strategy

51:02

Alexey

**Coming back to data strategy. I think you mentioned at some point that it's okay to use ChatGPT or GPT-whatever for some things. So how can we use ChatGPT for a data strategy design?**

51:17

Boyan

This is a topic which you do for weeks and weeks and weeks and weeks. It's not a very concrete thing and at every company it's different. You look at the use cases, the architecture, etc. It will help you in the beginning. Let's say, on the basic level, you work in a socks company and you ask, “What are the use cases for data science in an e-commerce store that sells socks?” And it will give you those five or six things. That's the most basic way. Then you become more specific, because you know from the due diligence in that company that they use Postgres, they have a lot of text data from reviews, they're on Azure, and they have a team of those people.

Then you do some prompt engineering, I would guess and say, “Based on this data, how should they transition to this architecture from this one?” Then you get some information. It's going to be a good start. It's not going to do the work, though. No way. We're very far from there. You just get very basic things like, “You need DynamoDB for this. You need this.” But then, if you really are going to implement this, then you go to the coding level of ChatGPT. Then it's a different story. It's a different thing. You go from strategy to tech. Then it's your best friend, because then you can really move very fast and develop things. But until that point, it's a bit like brainstorming. It's the same as writing a book and outlining things.

# Pitching a data strategy to a stakeholder

52:44

Alexey

**Okay, cool. I see that we have a question. “Let's say we have to do a data strategy for a consumer B2B2C business with a stakeholder...” That's very specific. B2B2C is what? It's a business that sells to a business and then to customers?**

53:03

Boyan

Yeah, correct.

53:05

Alexey

**Okay, we have that. “There is a stakeholder. With what and where do you start?”**

53:14

Boyan

Yeah, the first one is always due diligence. Who is the stakeholder? This person, what is their background? People forget about this. Who is the audience of this? This person maybe understands a ton about technology, and then you're gonna change everything in the way you talk to this person. Normally, they don't. Normally, this is a business person who is doing this. To approach this, the first thing you should do is absolutely zero technical terms. Zero.

Immediately, you have to show that you understand what the business is doing. This doesn't happen in only one meeting. This person has to feel comfortable that you, as a data strategist or whatever, know what the business is doing. Otherwise, they will destroy you. You're not going to hit the target. You have to be able to explain to them what business is about. Sock selling, for example, you have to know it almost as well as they do. Not as good, but they will ask you about it. Once you show this, then they will listen. That's how humans work.

When they listen, then you have to come up with a very specific plan, again, with zero technical terms. This strategy should be very short – one use case – very fast and without risk. You can say “Let's take one data scientist for three months and let's do this churn model and see what happens.” So it should be something very small and you should be able to feel comfortable talking about budgets. You say “The salary of this person, times their three months at 70% capacity. This is the money I need. Can you sign?” Give them a chance to just say yes or no. You don't want him to think about AI and whatever. You can use the current hype, of course, which helps.

In the end, just give them one small thing. You don't talk about data strategy, by the way – no, no, no. If you want to do a data strategy, you don't talk about data strategy. Because data strategy sounds very static and very expensive. You first focus on, “Let's do a use case. Let's get the value first.” You do a data strategy to do the use case, but you focus on value and concrete results with a budget. Then you can say “I did the data strategy there.” Then you can move on.

# Setting baselines in a data strategy

55:32

Alexey

**We also have a question about the baseline in data strategy. The question is, “Do we need to have two baselines in a strategy? One for the initial discussion or initial benchmarking – initial understanding of what the current status is – and then the second one for after implementation and actually running the model in production.”**

55:58

Boyan

I think it will be hard to come up with the exact same thing that you're trying to measure. It depends on what your baseline is. What normally is going to happen is – you have some kind of business metric. Let's say it's to reduce customer churn. That's very concrete. You can do a churn model. But how do you measure how much that contributed? So I don't think it's going to be one-to-one.

And you definitely need two, because you do need to at least qualitatively show, “We are roughly aligned with the benchmark at the beginning,” and we can demonstrate, at least visually, “This quarter, since the module has been deployed, fewer people churned.” Again, you get the question, “How did your stuff contribute?” But at least you are trying to work on the same thing. But you definitely need something at the beginning – anything is more than a zero there, I would say.

# Boyan's book recommendations

56:56

Alexey

**Well, we should be wrapping up. Maybe one last question for you. Do you know of any books or other resources that you can recommend to listeners who want to learn more about this topic?**

57:11

Boyan

Yes. The best technology book on strategy – there's a book called Technology Strategy Patterns: Architecture as Strategy, which is a gem. Again, it's a super criminally underrated book. I can really recommend that one. It's focused a bit more on general stuff, not just that much about data, but it's also very, very concrete. I will also say Practical DataOps. I already mentioned that one. And Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage, by Doug Laney. Infonomics is great. He talks about data assets and everything. Brilliant, a brilliant book. The book by Alexander Borek and Nadine Prill. That's a really great one. It's called Driving Digital Transformation through Data and AI: A Practical Guide to Delivering Data Science and Machine Learning Products. I have to get the link, but that's a brilliant book. It's also very specific. It goes much deeper than I do, with concrete questions. For example, “What do you ask a business person?” The final book, which is very weird, I'll say, is called Secrets of Power Negotiating. A terrible name – a terrible, terrible name. It's an old 70s, American-style book about how to negotiate and convince people. But it has absolutely brilliant advice on how you do that. A lot of data strategy work is just that. So read that one as well.

58:32

Alexey

**Okay, cool. Thanks. That's a lot of books. Thanks for joining us today. And thanks, everyone, for joining us today. That was a lot of useful information. We will also include a link to your book, and I think we have like five copies of your book, which we are going to give away – probably when we are going to release the audio-only version. So keep an eye on that. Have a great weekend!**