



SDS PODCAST

EPISODE 5

WITH

DMITRY KORNEEV



Kirill: This is episode number five with forensics investigator Dmitry Korneev.

(background music plays)

Welcome to the SuperDataScience podcast. My name is Kirill Eremenko, data science coach and lifestyle entrepreneur. And each week, we bring you inspiring people and ideas to help you build your successful career in data science. Thanks for being here today and now let's make the complex simple.

(background music plays)

Welcome everybody to episode number five. Super excited that you are on board, and that you're following these episodes. Today we've got a very special podcast. I've invited my friend, Dmitry Korneev onto the show. It was such a hassle to get Dmitry on the show because he actually works at Deloitte US, and he's constantly travelling, always in the air, always all over the place, and it took me several goes to actually get this podcast running, but finally we got there, and I was super happy about that.

So something you need to know about Dmitry is that this guy is like a just unstoppable engine. He's been working at Deloitte for 10 years now. 10 years in one of the Top 4 consulting firms. And he's worked in many different countries. So I think he's actually lived in at least 3 countries for prolonged periods of time, but he's actually been to over 30, and some of them he just travels for leisure, but a lot of them he actually went for work. He's constantly flying around the place. He worked back in Deloitte Russia for some time, then he worked in Deloitte Australia, and now he's working in Deloitte US in New York.



It is very interesting what he does. He's actually in the forensics department for computer fraud analytics. So that's not just cyber fraud, but actually Dmitry deals a lot with investigations that are Court related, or even criminal proceedings, where data has to be extracted from computers, laptops, mobile phones. It has to be analysed, and then results have to be presented in Court. And it's very interesting because not only do you need to extract the data, but also you have to go through it, you have to create algorithms, you have to find the insights that you're looking for that might help the Court case, or might help the investigator, or might even help the police, and then present it to them, and also there's a lot to do with the accuracy, as you can imagine. You have to be very accurate, you cannot just miss evidence.

So there's a lot of that involved, and that is what his job is, and he moved to New York just recently, maybe just a year ago, something like that. Before, he was one of the top-rated forensics cyber investigators in Australia. I remember working with him at Deloitte Brisbane. He would always be constantly called down to Melbourne or to Sydney to perform some data extraction and investigation there. So this is a very, very interesting person, and in this podcast, you will learn about fraud analytics. So it's a completely new area of analytics that you may have never heard of before, may have never even considered going into that path as a career, but it does exist, and it's very, very large because, as you can imagine, most of the communications now happen via SMS, via e-mail, and extracting that data, and analysing it, and presenting it in these Court cases and other places, other investigations is a very important skill, it is a very valued



skill. So this career path does exist, and it does require data science skills.

You'll learn a lot about structured versus unstructured analytics. Dmitry talks about how 80% of his work is with unstructured data, and we talk a little bit about the tools and methods that he uses there. We also talk about a few methods for structured analytics. So for example, we'll discuss Benford's law, and you'll learn what Benford's law is and how that data science law, data science principle, can be applied to find fraud in massive volumes of spreadsheets and things like that.

You will learn about the five steps that he takes when performing fraud analytics. Also we had some great casual conversations about working and travelling, and how he leverages the two, how he manages to survive at one of the top consulting firms, where the hours are just insane. Sometimes at Deloitte, you are on projects where you don't have any other choice but to work from 7 am til 2 am the next night. So sometimes the hours can be very draining, but I will talk about how he managed to survive there for 10 years, and how he actually loves what he does, how he enjoys it and how he mixes it up, how he finds new ways to challenge himself and continuously grow in that organisation.

And we also talk about relaxing and taking a break and Dmitry and I share some of the examples how we do things that help us recharge our batteries and get back straight into it. And I think that's a very important aspect for any data scientist, whether you're already in a full throttle career or you're just learning and preparing for a career, because you cannot stop. As a data scientist, you have to keep



pushing, you have to keep going, have to keep getting better and better and better all the time, have to keep up. And that's why knowing how to properly take a break, how to recharge your batteries, knowing yourself, is very important.

And so off we go. Please welcome Dmitry Korneev from Deloitte.

(background music plays)

Dmitry, what's up? Welcome to the SuperDataScience podcast. Thanks for being here.

Dmitry: Thanks, man, how you doing?

Kirill: Good, good. It's been a long time, hey? I haven't seen you in ages.

Dmitry: Yeah, I think it's been almost two years since I moved to the States, yeah. I probably should come and visit Australia. It's been quite some time.

Kirill: You haven't been back in two years?

Dmitry: I came to Australia in October for two weeks, and I think we caught up at the time, but it was a pretty brief visit, and I've been travelling around the world since, but you guys, you are so far from here! You don't quite realise how far Australia is until you leave it! Like when you live there, you don't think so, but you move to another part of the world, and you think ok, well, why would I go there?

Kirill: I know.

Dmitry: I've turned into this state. I may change one day, but that's where I am.

- Kirill: Yeah. I was in the US as well, just like a month ago, and everybody was like, where are you from? Like Australia. Like oh, that's so far away, so far away!
- Dmitry: How did you like New York?
- Kirill: New York was really good. I stayed in Harlem. It was a very nice area. So safe. Like we could walk around in the evening. So I was very surprised. The only thing about downtown New York, the main part, I didn't like is it's very narrow, and I found that the city's quite dirty. Maybe it was just the summer heat that made it feel that way, but to be honest, I liked the suburbs more than the centre. How about you?
- Dmitry: Well, I was thinking exactly the same way when I moved here first. And I live in Hoboken, which is not in Manhattan, it's a kind of area where you can escape, and the streets are wider, and it's a bit more relaxed. But I think your reaction is a typical reaction, one of the typical reactions of someone who moves here from Australia, because it's a different pace of life, and it's pretty built up, and you need time to adjust. And it took me probably about 3-4 months. And I'm thoroughly enjoying it right now, but I understand where you are. And I was thinking exactly the same way when I came here the first time.
- Kirill: Totally, yeah. I can imagine. It's definitely a different way of life. And last time I remember when we were talking quite a lot, and meeting up on a daily basis, we were working together at Deloitte, and back then, I remember you were doing a lot of -- in Australia. Deloitte Australia, Brisbane, and then you moved to Melbourne. Is that right?



Dmitry: Yeah, exactly. I moved to Melbourne for one year, and then I moved to the United States.

Kirill: Yeah. And last time when we were working together -- ages ago, when we were working together, I remember you were a fraud investigator, you do a lot of forensics work, and you were doing a lot of actually getting into -- I might even say hacking into software and hardware to get the data out to supply it to more junior data scientists for analysis. So can you tell me if that has changed, and what are you up to now?

Dmitry: No, it hasn't changed that much. I guess my role is slightly different at the moment because I'm no longer directly touching the systems and pulling data from there, but now I'm more focussed on processing and hosting and analysis of this data for actual fraud analysis, or hosting for fraud investigators. So it's probably not a part of the cycle. I mean, I still do what I was doing in Australia, just the focus might be a little different, and I'm kind of looking at the same area, like fraud investigations, forensics, but slightly from a different angle. And it's a different perspective, and actually I'm quite enjoying having the chance to see it from a different angle. Because probably I wouldn't have had this chance in Australia, just because of the maturity of the market and the demands of the clients, like what they want us to do over there was a bit different.

So yeah, although I'm still in the same firm, my role is slightly different. But I'm quite happy that I'm actually doing something different here. Because I'm probably like you, I mean, I like being everywhere and trying out different things, and some things may work out, some may not, but if they don't work out, you just move on. If you like them, you just

do them until you stop enjoying doing them, and then you move on to something else.

Kirill: Yeah, totally. And I remember you being exactly like you described. But what really surprises me and impresses me to a huge extent is how you manage to keep all this diversity of interest within one company. You've been with Deloitte for how long now?

Dmitry: Well, it's more than 9 years.

Kirill: 9 years!

Dmitry: But two things need to be taken into account. First, I worked in 5 different countries by now. And technically it's one company, but if you move to another country, you still retain certain experience and spirit, but the operational models might be different. And the markets are different. So what you do in my country may not be relevant to another. And so on.

So all my five spells have been quite different, and it allowed me to try different things. Plus, you probably know from your experience, the structure of the firm is not very rigid, because it's a professional services firm, and moving between the service lines is pretty easy and you can work on projects which are not typically handled by your department. So it's pretty flexible.

So yeah, I've been in one firm, but at no point would I say I was bored and I was feeling like ok, I've been doing the same thing for years and years and I need to try something else. I mean, if I get to the point where it becomes repetitive, then I may move to another company. But so far, it's been very diverse. And I've been thoroughly enjoying it.



- Kirill: I can imagine. Probably very challenging as well.
- Dmitry: Sometimes it can be. You would know that when you work in a professional services firm, sometimes you're busy, sometimes it might be a bit slower, because it's project based. So yeah, there have been a few bits when it was really tough and you had to manage multiple engagements, pretty complex ones. Yeah, it's a job which requires to be on top of everything. It may be chaotic sometimes, but I like it. I like that it might be a bit unpredictable, and when you can get called any time, and you sit doing nothing, and then things just blow up. Everyone is different, but I've been enjoying it so far.
- Kirill: Totally. And I remember back in Brisbane, when this part of the service line was only maturing, and you were the first and only person who was able to perform that level of hacking and getting into accounts for fraud investigation purposes. I remember that went on for like a year or two. And you were being pulled part, torn apart, by everybody. Every department wanted you and you had to fly down to Melbourne and Sydney to work for them there just because you were the only person in Australia who was capable of such a thing.
- Do you think the industry, the analytics fraud industry has moved on since then? How do you think it's developing at this stage?
- Dmitry: It's definitely evolving. Back in Brisbane, the firm I was working on was kind of a startup, the service line. So in a small startup company, you get exposure to different things, and you do this and that, which might be different when you work for a large organisation, or a large service line.

The thing about fraud analytics, the ultimate objective is probably still the same, because the way people commit fraud, it's changing in the high tech world, but within the financial industry, it might be still the same, but because of the variety of the systems we have to deal with, in terms of when you have to pull out the evidence and actually find something to prove allegations, this area's getting more and more challenging. Because right now, you know everything is moving to the Cloud. And for forensics, it represents a big challenge, because typically it's hard to get access there. So you kind of have to hack it legitimately.

You always find a way round, but the complexity of the system represents a huge challenge. And I think another part of the story is that the tools that you have at your disposal, they usually cannot do what you want to achieve. So the landscape, the technology landscape, is changing so quickly, so the vendors cannot catch up typically. So they give you a product which is expected to give you certain capabilities to pull out something, but then you find that it needs to be customised, and it has to be done usually under time pressure. And a typical investigation, it may take a few weeks, and it's not like an SAP implementation, when you can do it for years and massage it to perfection. It's really you get under pressure, you have a very short time frame, and then you find out that the tools that you have, they just cannot do what you want to do, and you kind of have to work out solutions on the fly. I've been seeing it for years, and it still remains a trend.

I guess one of the features of this industry, the typical project management skills which you will deploy in software implementation or something like that, they are not really

applicable, and if you are the project manager, you have to develop your own skill set. And it comes down to data analytics as well, because the thing about my role, because it's a mix of unstructured and structured data analytics, and it's more about unstructured analytics. Unless you investigate financial transactions or account irregularities, in most cases, you have to look, it's more about texture analytics. And sometimes it's both. And you kind of have to convince both parts. And another thing is when it comes down to finding evidence, you don't have to chase every hole. So you have to be a bit intuitive and know where to look at kind of touch different things and fuse them together. So it's a specific mindset plus specific skill set.

Kirill: Yeah, definitely. I remember the time when we were dealing with textual data. We were dealing with emails, but also with text messages. So you extract text messages from somebody's phone, and of course legitimately that has been confiscated through police findings or something like that, and then you have to go through that. So there's a lot of unstructural data, I definitely remember that. And I remember a couple of tools that we were using, especially this one when they sent it to us, I think it was from the UK, do you remember that? This box came in, and inside this box was a briefcase, and you open it up, and it's got all these little gadgets to extract data from mobile phones. It looked like a James Bond type of thing. That was really cool.

What kind of tools do you use? And I mean more like on the data side of things. What kind of software tools or programming languages do you use for your day to day job?

Dmitry: Yeah, so you refer to the specific hardware and software tools that we use for data extraction for forensic purposes.



It's slightly different to just typical data copying because we have to -- basically, you can't bring something to a Court unless you can prove that you've covered everything as an exact copy. So the tools, the hardware and software tools, they have certain algorithms which calculates the hash value of something, which is a signature, which you can show to the Court and say ok, that's an exact copy of the data.

So in terms of data analysis, there are some very specific forensic analysis products, like encase from Guiding Software, or FTK from Access Data, which are all largely focused on finding forensic artefacts in the data. In cases when you have a large scale review -- like if it's an investigation, you may need just one person to really drill into the data and pull out some piece of evidence. But sometimes you have 40 or 50 or 100 people which need to look at the data from a different perspective, because they may bring some special knowledge, like accountants, lawyers, or subject matter experts, industry experts. So when you need to have all these people looking at the data, then you have to first do forensic processing, and it wouldn't typically involve extraction of meta data, and just then hosting of the data in a special platform.

So for processing, the two biggest vendors are Ipro and Nuix. And when it comes down to large scale hosting, the industry standard is called Relativity. So the company is called Kcura, and it's pretty much a large-scale hosting platform. Having some analytics features, mostly surrounding predictive coding and textual analytics.

So there is also another system called **[19:36 Recommend]**. It's getting pretty big here in the States, and it's probably the

first kind where it's going to hit the market, and yet probably become global as well. So what I've just listed there, that's specific forensic discovery industry products. If we talk about more general analytical tools, we do sometimes do a lot of database analytics or we use pretty much the same thing, like SQL databases. Sometimes we need programmer skills as well, and there are people in my team who are pretty good with Python or C# or other programming languages. But we need them on an ad hoc basis.

So that's probably the tools that we use on a day to day basis.

Kirill: Ok. That's really cool, because all those tools that you just listed, the forensic discovery analytics tools, they're completely new to me, and I'm sure to most of our listeners, they'll be absolutely alien. But it's a very good thing to at least know about them, and know how they operate if you want to get into that space. So we'll definitely include them in the show notes. I'll flick you an email later so that you can maybe supply me this list.

But just out of interest, and so that we get to know how this whole forensics type of project operates, can you tell us about the life cycle of a forensics project? What is discovery, what comes after discovery, maybe there is something that comes before? What are the steps involved?

Dmitry: There is a term called EDRL, which is e-discovery something something. I probably should know what it means! But basically, it's a five-stage process. It used to be three only. But now it has expanded since, just because of the way the industry's growing. So typically, there is a, it starts with information governance. And it's a information governance



from a discovery perspective. What it means is that a company which operates in a highly regulated industry, and at the moment it's mostly financial institutions and pharmaceutical companies, so a company like that might be required to respond to litigation requests very quickly. Especially here in the States. If you get sued and you are requested to provide emails for 50 people, you may be required to do it within two weeks, otherwise you will be in trouble.

To do this, you may need to have special systems implemented within your IT environment, back up legal fault systems. The problem is that this wasn't the case 20 years ago, but right now it's becoming extremely important. So information governance basically all comes down to your ability to manage your unstructured data and being able to provide data at short notice. So that's the big step. It's not something that you do as a private investigation, but that's what you may want to be doing on a continuous basis just to be prepared.

So that's the first. The second step, which is the step of a typical forensic or discovery [22:49], it's called collection. And this is all about extraction of data from your system in a forensically sound manner and at this point, you start using the hardware and software tools which you saw in Brisbane, working from there.

So the next step is processing. And when we talk about data processing in the discovery sense, it's first of all about being able to pull out meta data from all the types of data you may come across. In discovery e-Discovery World, when we collect the data, we just pick up everything. So we may get databases, emails, all sorts of data which is completely

different. And in this case, you need to be able to process all of them and pull out meta data from all of them. Because for any forensic matter, it comes down to first looking at the document, and second of all looking at the meta data. And they kind of complement each other.

So when you do processing, yes. All the systems that do this kind of processing, they need to have this capability.

Kirill: Sorry, just to interrupt you there. Metadata you mean like for instance, out of a photo, you can get where it was taken or the time stamp when it was taken, things like that.

Dmitry: Yes, yes, things like that. Everything which is basically the progress of the document. But the important thing to understand is it's more than you can see if you just right click on the file in Windows. Because there are different layers of meta data. There is the data sitting within the file itself, there is a layer of data which sits within the container where the file may reside, or you have the meta data, which resides at the file level in a computer system, which you don't even see, but it's still there. So in certain cases, it's important just to find all these pieces and bring them together. Because they kind of provide the entire picture. And they may give you an idea of what has happened.

So this part is the processing. The next step is called review, or hosting. And at this point, this is where you take all the data and meta data, put it in a hosting platform, and you get 50, 100, whatever number of people you have to review it. Typically they would be subject matter experts. And I guess the specifics of a discovery review is that it's a legal role. So you may have some privileged information. For example, I might find information which one person is supposed to see

and another person is not supposed to see. So you have to make sure that they can only see what they are supposed to look at, and if they are not, there are things like redactions, where you basically can remove a certain portion of information from your file and give it to another party for review.

So at this point, it's not really analytical. You can have analytics at this point as well, but it becomes a massive project management exercise. And the last piece of the cycle is production. Typically when reviewers look at the data, they find a set of relevant documents, and it's called responsive in the discovery world. So from this set, it has to go to another party so that they can review it on their system, and there is a whole science behind how you do it to make sure that first they can read this data, and they get only what they are supposed to get.

So yeah, that is probably a brief summary of what a typical meta would involve.

Kirill: Ok, that's very interesting. So where would the data science algorithms, if they need to be applied, in which step would they come in?

Dmitry: They would come in between process and review. Because one of the key components of the review is data culling, and there are many ways you can do it. The traditional approach based on key words, they're in the past. Right now, especially if it's a large corporation, the cost is the most important factor. Because for many companies, it's a compliance issue. And if it's compliance, basically they want to do it cheap.

You mentioned textual algorithms. There are different techniques which can help you isolate all non-responsive documents straight away before even starting the review. So you basically remove all these documents, and then there are certain techniques which can help you tell which data is not responsive even if you don't look at it. Basically, you pick a sample, you find the documents which are non-relevant, then you throw it back in the entire population, and then the system uses the algorithm to propagate that code, and you can tell which other documents might be relevant or not.

In this world, you have to be extremely careful using those techniques. Because the cost is one side, but in the legal world, there is a requirement to review everything and not to miss a single bit. So there is no room for error. That's why those algorithms exist, but not every company would use them. It would depend on the case. So sometimes you just need to make sure. You need to get 100% result. You have to make sure that everything has been reviewed. And if you don't get there, you just won't use those techniques.

But there aren't many cases when it's a little bit relaxed, and you can use advanced algorithms and just get to the point quickly. There could be an error, but we are fine with that.

Kirill: Ok. So it sounds like an extreme case of a machine learning exercise, where you're doing some classification of documents and there's absolutely no room for error, because normally, you would expect some sort of level of error from a machine learning algorithm, but I guess as the legal requirements and the increase, and as it's a more serious case, the less of an error you can afford to make, and therefore some techniques you just cannot use any more.

- Dmitry: Yeah, I think it's a good summary. While it's not always the case that there is no room for error, sometimes you can have a small error. What's happening in the industry right now, typically all the tools and systems that I just told you about, they were designed to serve the needs of the industry, which is the legal review. But what many people have said since is that you can take those systems outside of the e-discovery world and use them to solve other business problems where you have a mix of structured and non-structured data. And when you put them in this world, then you have room for error. Because your purpose may be more strategic than operational. And it might be the case that you don't have to chase any hole. And I guess when we move to that part of the industry, it's not the discovery industry, but it's something else, then you have more flexibility and power to use advanced categorisation algorithms to make the review more efficient.
- Kirill: Ok. Like tell me if this is a correct example. For instance, like a huge mining company with several millions of dollars of monthly turnover, they want to investigate any duplicate payments of invoices, and so instead of looking through every invoice, which would cost them a lot of money, they launch an algorithm which picks out these duplicate invoices. And maybe it'll find only 80% of the invoices, but it'll be a cheap algorithm, it might cost several thousand dollars to implement, but it'll save them a couple of million dollars in the month. Is that a good example?
- Dmitry: Yes, it's a good example. You picked an example from a traditional transactional forensic analytics space. It's definitely what we do quite a lot, but you may not need a blend of structured and unstructured analytics to solve this

problem. In certain cases you may, but sometimes you can just do it at the database level. I guess another example could be **[31:18]** review. That's what we do a lot of for financial institutions. Because they may have a lot of contracts which are not stored in electronic form, and then when you need to have an ability to scan them, to do character recognition, and then to be able to index these data and resources. And sometimes you need to match what's in the quarter against their financial transactional data.

So that's where both things become important. Because you cannot do one without the second one.

Kirill: Ok.

Dmitry: It's probably just an expensive version of the example that you gave me. So yeah, you look at the financial transactions, but you also look at some sort of contractual background.

Kirill: Ok. And you personally, which kind of data do you work with mostly? Is it structured, or is it unstructured?

Dmitry: It's both. It's probably more unstructured, but I guess it depends on the objective of the business case. There is a business case, you have to solve it, and you pick up the data that you need to do it. I would say most of my cases, they required unstructured data. But it's different case by case.

Kirill: So would you say that unstructured data is becoming kind of the predominant medium? Because structured data, there are algorithms, and it's kind of easier to deal with. So would you say that for somebody who wants to maybe get into this space, it's a good idea to start considering and learning how to work with unstructured data early on?

Dmitry: So the stats are that the typical company would have 80% of unstructured data and 20% of structured. It may be different company by company, but that's the numbers that I've heard. And it would be fair to say, I don't know if you would agree, but we've been doing analytics on the structured data for years and years, and it's a pretty developed space. And with the unstructured data, it's still developing, let's put it this way.

In the forensic world, it's been the other way around always. But the thing is the objectives of the forensic discovery meta, they're usually quite narrow. So that's why we have the tools, but we typically use them to solve a particular problem.

So yeah, I would probably agree, it's a growing area, and it's not been developed. And one of the ways to actually make it more mature is to take the discovery tools that we have and expand them. It could be used as a [33:57] to get there.

What have you heard about unstructured analytics? Where is it? From a broader perspective, because I'm probably looking at it from my perspective, which might be a bit narrow.

Kirill: Well, with machine learning developments, and with the increasing processing power of software, natural language processing for example, which is another form of unstructured data, when you have audio recordings of phone calls. That's developing very rapidly, and even like you take your iPhone, and you want to dictate a message, now it's very spot-on. And you'll notice that if you try to dictate a voice message on your iPhone, it won't work if you're not connected to the internet. And that's because it's

using algorithms online to process what you are speaking into it. And same thing about textual data. So it's like processing scanned images is a bit of course more complex, but you can use machine learning algorithms such as decision trees and random forest, or other gradient-boosting algorithms to get that very, very spot-on and help identify exactly what's going on.

Or even some naive Bayes machine learning algorithms will do that trick. But that once the characters are recognised, identifying semantics and what's actually being said in the message, that's the fun part, I find. Trying to identify exactly what emotion is being conveyed and extract that from the textual information.

Dmitry: Yeah, this thing has been around for quite some time. How is it called when you look at the social media communication and try to figure out the mood of the audience?

Kirill: Sentiment.

Dmitry: Sentiment analysis, yeah. It's probably more applicable to web analytics as well. That's where you deal with a lot of textual.

Kirill: Yeah. Like scraping Twitter and stuff like that.

Dmitry: Yeah, I wouldn't say that it actually has any impact on what we do. You typically analyse company information. You don't go online.

Kirill: Yeah, but it's still the same algorithms, the same kind of approaches and methodologies can be applied once you have that textual information, regardless where it came from, Twitter, or from a document inside a company.

Dmitry: Yeah, absolutely. The algorithms are the same. It's just the data sources are different.

Kirill: Yeah. And so there are some great examples of some unstructured data that you deal with and probably the main goal -- I was just thinking of some outcomes that you would be looking for. But probably the main goal is to find that sentiment inside messages or look for specifics. Like you said, key words are going away. They're not as relevant any more. But at the same time, sometimes you might be looking for some key words. So searching for that type of information, synonyms or misspellings of key words, that could be helpful.

Let's have a look at some examples of techniques for structured data. Just out of curiosity, are there any specific techniques that come to mind when you're dealing with structured data in an organisation, and how those techniques can help identify fraud or just guide the investigation in certain directions.

Dmitry: Well yeah, I wouldn't say that they're that complex from a data analytics perspective, because it comes down to your understanding of the business case. And on a typical investigation, yeah. Because typically you wouldn't do much of modelling or a full-blown analysis. Your objective is to find a specific reference, or a specific piece of evidence. And to get there, you need to understand the business case, and you need to define the rules, understand the rules of how to find something to prove an allegation based on what you have in the database.

So yeah, I would say it's more about understanding the business logic and translating it into database logic, or wherever your information is coming from.

Kirill: One example that pops into my mind is Benford's law. I remember using that at Deloitte quite a lot. Were you involved in the Benford's law investigations back in Deloitte?

Dmitry: Yeah. It's an interesting observation. From my experience, it's easy from a scientific perspective, but in the real practical world, it doesn't help that much.

Kirill: Really? I would think it's such a crazy and non-intuitive but at the same time logical thing, I thought that it would come up quite a lot in investigations.

Dmitry: Well, it does. Let's put it this way. If you have an objective to give a complex review of how much risk a particular organisation may carry in terms of certain types of fraud, then it might be really good to present something like that. But if what you need to do is to find one transaction, maybe a particular person, and in many cases, it comes down like that, this chart is helpful, but it may not point you exactly to where it is. And also, sometimes there are false positives and there could be a number of explanations why you have all these anomalies.

Yeah, it's an interesting technique. It may not help as often in the investigation world as you may think it does.

Kirill: Thanks for that. That's pretty insightful. Didn't know that about Benford's law, that it has quite a lot of false positives. But just for the benefit of our listeners, I will quickly explain that. It's just a good example of a data science technique that can be applied for investigations.

Basically, in very layman terms, what Benford law says is that if you take an accounting spreadsheet, and you just take all the numbers from that accounting spreadsheet, and then you look and put them all into one bucket, and then you pick them out and look at the distribution of the very first digit in all the numbers, so you will see that number 1, the digit 1, comes up the most frequently. Digit 2 comes up second most frequently. Like digit 1 -- correct me if I'm wrong, Dmitry -- but it comes up like 36% of the time. Then digit 2 comes up like 15% of the time. Digit 3 comes up less, less, less. So it looks like a log normal type of distribution. I don't really remember which exact distribution it is.

And so if you're analysing a company in terms of fraud, then you just take a spreadsheet or something that you're analysing and you apply Benford's law, and if it doesn't apply, so if you can see that the distribution is different, that means there might be something dodgy going on.

Is that a good summary of Benford's law?

Dmitry: It's a good summary. He came up with this law based on real life observations. I can think of a few cases when it really helped to find.

Kirill: It's a good example. Even though it might be like old, and it might be not as applicable any more, or just there might be better algorithms, it's just a good example that there are some data science techniques, and that's more on the distribution side of things, or data processing side of things, that we normally don't encounter in our day to day work in insights type of data analytics. But at the same time, they're very powerful, and they're used a lot in a different space in



your world. So it's good to have this little window to see what's going on in the world of forensics analytics.

Dmitry: Absolutely. It is definitely a good thing to have, and it's a tool. It just doesn't work in isolation. It's something that you look at, and it may give you a direction. But you need to use other things to get to the bottom of the issue.

Kirill: Awesome. And next thing I would like to do is, can we get a bit into your background? So obviously now, you told us about the stuff you do at Deloitte, and it's like we just mentioned, it's a completely new world, and for most of our listeners, it will be something new to discover. Can you tell us about your background? What did you study and how did your background help you become successful in what you do now?

Dmitry: I have a degree in computer science and economics, I have a double major. And I started doing IT audit and internal audit within Deloitte, but I got bored pretty quickly, I just didn't think it was good fun. And then I had a chance to join the forensic team, and at the time, the firm was doing a large global investigation, and I joined, and I liked it, and I've been doing it since.

"Forensics" is a pretty broad term, and at the firm where I'm working, it's a mix of financial and IT analytics. When people ask me what do you do, and I say forensics, they can think whatever, like autopsy, or looking at fingerprints, and if you google forensics, you will see like 50 or 60 branches of that science. So it's a pretty broad category.

So definitely my education helped me a lot. Also, to be successful, and to enjoy, it's important to pay attention to details. But at the same time, you have to be very flexible.

And it's kind of on its own. On one hand, you need to combine those skills to be able to look at small things, but at the same time, in many cases, you don't chase every hole, and you have to use your intuition to find what you to find.

So yeah, I would say I like the occupation. It's an interesting area to be in, and you probably need to be a certain type of personality to enjoy it, but for some people I think it's something that they may want to be doing. That's their work they would enjoy all their life.

Kirill: What kind of personality would you say you need to be?

Dmitry: You need to be able to like cows.

Kirill: Cows?

Dmitry: Cows. Well, to a certain extent. I mean, it's fine to be methodical, but in many cases, a typical case may twist many times. And as I said, typical project management techniques wouldn't work. So you have to be prepared that you come to work in the morning, and everything changes in one day, and it's very difficult to predict. So you have to prepared, and don't get despair, and be able to adjust, and adopt. Yeah, I think it's very, very important for that, that type of job.

Once again, attention to details is what every data analyst would need. But probably in this world, it's extremely important. Because a very small thing may actually mean everything. As opposed to other areas of analytics or something else, where you more look at the trends, and kind of try to detach yourself from the small things and see a bigger picture.

You need to be able to see the bigger picture in the forensic world as well. But in many cases, sometimes you really have to pick up on small things and really focus on it. And kind of detach yourself from strategic trends, things like that.

So being able to go from high level to low level and back, that's also quite important.

Kirill: Sounds like there's quite a few challenges involved in this type of work.

Dmitry: Yeah, you can call it challenges. It's something which this type of work involves, and I personally enjoy it. I think many people would enjoy it. And again, it's an always-evolving area, like many others. So you have to stay current with the trends and tools. It's a special occupation!

Kirill: Yeah. Yeah, it's cool. And from two perspectives, from the extremes, what would you say has been ever your biggest challenge in your ten years with Deloitte in Forensics. What has been your biggest challenge ever? And what has been your biggest success that you think?

Dmitry: Well, are you referring to a specific project or more like long term and short term challenges that I had to overcome?

Kirill: Probably the long term and short term challenges.

Dmitry: It was a bit challenging for me to develop that kind of mentality. Because I mean, I probably used to be really, really strategic in everything I was doing. I liked looking at the trends and predictions and high level pictures. And it was difficult for me to adjust myself and sometimes go ok, stop for a second and just focus on this particular small thing and just investigate this issue and put everything else

aside. It took me a few years to develop that kind of mentality.

Because especially when you're under pressure, I was asking, why are we spending so much time on this small thing? I mean, let's look at a high level picture. But at the end of the day, it was very important. That was a big challenge for me, and I spent some time to overcome it.

In terms of my success, I like the international aspect of the work. It's probably in every area of analytics. But it's a very universal skill set, and you can apply it in any geography, and I've been travelling around the world for work and I've been quite enjoying it. And I think in this area, it's the same around the world, but fraud, by its nature, is different in each country. What you investigate in a developing country is different to what you investigate in a very much **[47:54]** country. And it's not like one is more interesting than another, it's just people coming from different ways. And sometimes it's technology. The platforms, they just allow, or do not allow, to do certain things.

It's interesting the way that it's a universal skill set, you can bring it to another geography, but you see that you have to - - it's not like a programming language. Yeah, Python is Python everywhere. Your task may be different, but it doesn't really matter where you are, and that's why you can be anywhere in the world. You can be in India, you can live in a villa there, and you work for someone in the US. You can outsource everything.

In this world, you can outsource small tasks, but in order to really do something, to achieve something, you really have to be present. To understand the nature of the fraud, the

cultural aspect, the entire picture. And I found it interesting. Yeah, as I said, I've lived in five countries. The technical skill set is very similar, but the fraud nature has been different. And again, being able to adjust, and just being able to tell ok, you take all your experience, but you can't use it the same way, that's what interesting as well. And I'm still enjoying it. Am I making sense?

Kirill: Yeah, yeah. That's totally awesome. And I can see how that would be a bit different to data science, like the core data science, where like you say, the skill is transferrable 100% and can be applied, or even outsourced, to different geographies. Whereas here, whenever you move borders, even just the legal system itself is different, right? And the way people think is different. What is available, what is not available in terms of hardware and software is also different. And it must be an interesting challenge to always be learning new stuff and exploring different ways that you can do that same job or actually the same tasks.

And you mentioned you travel quite a bit. How many countries have you been to in total, and do you generally just enjoy travelling for leisure as well?

Dmitry: Yeah, I do both. I probably haven't been to as many as you may think, because I'm not the kind of person to go somewhere for two weeks and come back. I'd rather go to one country for an extended period and just gradually go to all small corners. And that's what I'm right now doing in the United States.

Well, I lived in 5, I've probably been to between 20 to 30, but that's the way I like travelling. It's not about coming somewhere for two weeks, just going surfing, ticking the box,

then go away. I have a few other countries on my list, and I'm hoping to do the same thing with them.

Kirill: Ok. And would you say it's important in a stressful line of work like yours, because like I was in consulting, I was in Deloitte, but I only lasted for two years. And after that, it's not like I couldn't continue going, but you just realise how much pressure you are under constantly. Like constantly going home past 8 pm, 10 pm, like having 120 hour weeks, or like 112, my record was, hour weeks, or something like that. It's pretty insane.

The question is, do you think it's important to take some time to relax, like when you travel, to look around the place, the country, and you know, find ways to release that stress and have a normal life, have a social life, or have your own personal life, but then get back to work. Would you say that you have quite a bit of focus on that?

Dmitry: Yeah, I mean, work-life balance, it might be an issue if you work in consulting, and so it's a question for everyone, and all of us have different opinions on how much work you can do and what it should be. We all have different ways to recover and escape. I don't know how you do it, but some of us need more time, some of us they just need 15 minutes to do everything, and they can go back to work.

I would say I'm a visual person. Some people listen to their music and it makes them calm, relaxed. It doesn't work with me. But if I see something, and it typically would be a piece of nature, or some kind of city landscape. If I see something I like, I will recover extremely quickly. I would say in Australia, I love the beach. Here in New York, I like the New

York City skyline. And yeah, it may take me at least within 5 minutes of looking at it, and I'm good to go back to work.

So yeah, I travel a lot, and working in consulting requires a lot of travelling. I mean, it depends on your role, but I think you were travelling a lot, and I do it every week. Well, a bit less now, but I used to travel tremendous amounts of time. I like travelling, so it suits me really well. And when I'm stressed, yeah, I just find something to look at. And it's usually something which just catches my eye. It can be whatever. It's just a place, usually, with which I establish an emotional connection, and I typically know what can be that sort of a place, but sometimes I just bump into it, and I say ok, let's see it, and I get some energy from it, and I'm fully energised and go back to work.

Kirill: That's very deep. I like that. That's a very deep, profound....

Dmitry: I heard the theory that some of us are visual people like me. Some people, they need to listen to something, and it can be the sound of nature, or music, and they just put on headphones, they listen for 5 minutes, and they are fine. Some people may need to talk to someone. I found my way of recovery, and it works for me. I like that I found it. Because it worked for me really, really well.

Kirill: Yeah, that's really good. I think I'm even going to benefit even from just this part of the conversation myself a lot. Because recently, I have found that I take on a lot of stuff. Like, building courses, and doing this podcast, and answering questions, or managing projects and stuff like that. So I found that I take a lot on, and I find myself working constantly. Like this is probably my third or second week -- I'm into my second week working 12 hours every

day. And I forget about that I need to take time, and go do something. For me, it's probably riding my motorbike, or like you say, going for a walk in nature, things like that. So it's a good thing.

Taking you as an example, a person who's managed to survive for 10 years in consulting and you still love it so much, it's a good testament to the fact that we need to take care of ourselves, and not just what we eat, but also how much time we give ourselves to rest. And it shouldn't be just rest 5 hours a day, or something. It should be quality rest, like you say. Like find something that helps you relax and helps you get your mind back, and gets the energy back to keep moving forward after that. So it's very profound. Thank you for sharing that.

Dmitry: No problem. So how do you originally recover? I mean, you mentioned motorbikes, and....

Kirill: Yeah. Probably motorbike for me. Yeah. For me, it's kind of a once a week thing. If I know I'm doing a six-day week, and then on Sunday, at like 6 am, I can get on my motorbike and go for a ride to the race track, or just with some friends to the mountains, after that, I'm totally tired after that, when I come back. I can't do anything that day. I have so much adrenaline and it's like a reboot to the system, you know, and then on Monday morning, you wake up, and you're fresh again, ready to go for the week.

So I guess it's different to your method, but everybody has their own. And hopefully our listeners who are listening to this will take a minute to pause, think about it, and find out for themselves what helps you recover the most. Is it something short? Is it something long? Is it something

specific? Is it talking to someone? Think about things like are you an introvert? Are you an extrovert? Just think of that one thing that helps you recover and get back to whether it's doing data science, or learning data science, or pushing the boundaries of research further. Because whatever you do, your doing it in this world is definitely important, and we need people like you to be energised to keep doing what you're doing.

And just moving on to the closing part of our podcast today, from your perspective, where do you think, Dmitry, is the field of data science, or specifically fraud investigations going in the next 5-10 years? So people who want to establish a career, who might want to get into a career in this space, what should they look out for, what should they prepare themselves for?

Dmitry: I think they should realise that well, it's a probably general comment that the amount of data is growing exponentially, and we will be relying on data analysis methods more and more going forward. Well, I would say in the forensic and discovery world, and I mentioned before, the challenge is going to be to use analytic methods in the unstructured data world. And use it efficiently, and taking into account the objectives of why we do it. And just being able to use them together with the traditional structural analytics methods.

So in the discovery world, there is a lot of going on surrounding predictive coding, text categorisation, it's picking up quite a lot, especially for large scale reviewers, and big corporations are paying more and more attention to this.

So if you want, I would say that if you really want to get in this world and establish yourself, you really have to look into the unstructured data analytics methods.

Kirill: Ok, wonderful. Thank you for that. And any career aspirations that push you to become better and more proficient at what you do?

Dmitry: I'm enjoying what I'm doing, and I think I'll keep doing it for the foreseeable future. I definitely want to develop more industry experience, to be able to use all these investigation methods efficiently and to be able to conduct investigations, understanding of the business is crucial. And it's crucial for traditional analytics as well. But I guess in my world, you really understand all the business process, how it all works, and being able to apply your techniques knowing how it works inside is very important.

So I think that maybe over the next 3-4 years, I'll keep developing my technical skills, but I'll probably focus on one particular industry and just develop --

Kirill: Some domain knowledge, yeah?

Dmitry: Yeah, some sort of domain knowledge.

Kirill: Yeah. And that's definitely an important part, and we speak about this in some of my courses, that domain knowledge is a very important part of the work of a data scientist, because it's a completely different story when you know what you're doing analytics about and on, versus when you don't know. And there, domain knowledge can give you a massive advantage.

So thanks for that, that's some very solid advice. And our listeners would want to follow your career, or even maybe



contact you in a way, what's the best way to get in touch or just follow along and see how your career develops in the future?

Dmitry: Maybe my LinkedIn account would be the best spot to start. I probably should start a blog or something like that, but I have to confess, I haven't been doing so far. So if that changes, I'll let you know, but for now, I think probably my LinkedIn profile would be.

Kirill: Yup, wonderful. We'll include the link to Dmitry's LinkedIn. So make sure to hit Dmitry up on LinkedIn and connect so you can see how his career goes. And if Dmitry starts a blog or other social media, we'll definitely update the show notes in the future.

And one final question for you today, what is the one book that you think will help our listeners become better data scientists?

Dmitry: Yeah, that's a tricky one. Well, you might be surprised, but I would say that the book I would recommend reading is called "How Life Imitates Chess" by Gary Kasparov.

Kirill: Wow.

Dmitry: And it's not exactly about data science, but I would say more about data strategy. I used to be a chess player. I don't know if you've ever played, but I [60:34] and it's not data analytics, but typically, it's considered one of the areas where the most intelligent minds go. And he was the most intelligent one. And I'm still amazed that a human being can compete with a machine in that space.

So he's not a data scientist. But he's a kind of data strategist. And you may know that he works at MBA schools



in the States, and he delivers lectures on strategy quite a lot. So this book is pretty much -- what he's trying to do in this book is to take the world of chess and try to bring it to the business world.

And he's not the first guy from a non-business environment who tried to break into it. In my view, this attempt has been quite successful. So this book won't teach you any particular data analytics techniques. But I think it will teach you how to be a data strategist. And how to apply those techniques to solve complex strategic problems.

Kirill: Wonderful. I've never heard of that book, and it sounds very interesting, and also being a chess player myself, or when I was younger, I was very into chess as well. So there you go, guys. "How Life Imitates Chess" by Gary Kasparov.

Thank you very much, Dmitry. Really appreciate you coming onto the show and sharing your insights. I'm sure lots of our listeners will find all this useful. Thank you so much for being here today.

Dmitry: Thank you man.

Kirill: So there you have it. I hope you enjoyed today's podcast. I definitely learned a lot. Even though I knew Dmitry before, I still learned a lot about his career and about fraud analytics in general. And overall, it's good to know that this area of analytics exists and a lot of people don't even think about a possibility of a career pathway there. And maybe if you enjoyed the things we talked about, the methodologies, the type of work that he was talking about, and the different investigations that Dmitry mentioned, maybe this is something that you might want to consider for your own career.



And remember to get the show notes at www.superdatascience.com/5, and there you can get the transcript, and get all the recommended materials, and all the items that we mentioned in the podcast.

And if you enjoyed today's episode, then make sure to share it with your friends and colleagues and anybody who you know who might want to get into the space of data science.

And I can't wait to see you next time. Until then, happy analysing.