

00:05

Speaker 1 Hello and welcome to the Human and Machine podcast. This is your host, Jako Markwat. I'm here with my co host, Lenny Smith. Lenny, another podcast, another week. I was listening to last week's podcast again with Travis. Really good insights in terms of the new release for Ignition Perspective 8.1. Definitely. After we spoke with them, we actually saw the release candidate come out, but there was a really good chat with Travis around some of the tech that's available right now and some of the features that they're introducing with Ignition 8.1.

00:40

Speaker 2

Definitely. What a bumper week this was. It was the launch of the first ICC virtual conference as well. We had element eight. We also had our first conference. We had a little bit of a meetup. It was a meetup. It was a little bit of a mix and match. We were obviously due to the levels that we've currently experiencing with the COVID pandemic. We were only limited to 50 people.

01:00

Speaker 1

Because you said two beers. We did help ABM Bev, who presented yesterday, Roman Ray from ABMbev presented yesterday, and we did help them consume some of their products.

01:10

Speaker 2

A little bit of that is definitely. And then as it happens, obviously our president moves us to a level lower. So we could have actually added a little bit more people to our event, but that's it. We had a great and a fantastic time. And definitely the release of version 8.1. Definitely some of the exciting highlights that came out of the ICC. We got a little bit of a sneak peek from that from Travis Lott last week in the podcast. But definitely, I would encourage all our listeners, get your hands on the release candidate, start playing with it, see the amazing features and functionality that we have with the new release of the perspective module. It opens up so much more, not only just for SCADA solutions, but for any other applications that you might think of to implement on your plant environment.

01:57

Speaker 1

Definitely. And we'll be sure to release. I think what we'll do is we'll do an entire episode on the event itself, and hopefully we can get guys like Francis from Clover and maybe even Rowan to give us a quick little update on some of the highlights.

02:09

Speaker 3

Cool.

02:09

Speaker 1

So this week is something a little bit different if you're new to the podcast, it is all about the manufacturing and production, mining landscape in South Africa. We're speaking a lot about the latest technologies, some of the challenges that people are experiencing. And on the back of that, we have something a little bit different this week. Very exciting in the world of manufacturing and just broadly, I suppose nowadays, in the world of business and value chain and supply chain. Some of the most critical challenges that we see that companies and people are facing through their value chain are fundamentally data challenges. Supply chain procurement, whether it's quality, rapid troubleshooting, and even forecasting, depends on accurate and up to date data and contextualized data. So this week we're very excited.

03:00

Speaker 1

It's actually a reference that we received to chat with Leandra Webbre, who is a data scientist and information engineer with Decision Inc. Leandra is quite a CB. Leandra, we read through your. I feel a little bit underwhelmed. Sorry, overwhelmed, looking at your credentials. So Leahndra has obtained degrees in both biomedical and electrical engineering. That's more in your field, Lenny, as well as a master's degree in biomedical engineering. And the focus of her master's was on brain computer interfaces. Fascinating stuff. I

don't think we have enough time to talk through all of that, but really interesting backgrounds. Leandra, welcome to the Human and Machine podcast.

03:46

Speaker 3

Thank you so much.

03:47

Speaker 1

Really good to have you on there and talk on the topic of data. And I see data scientist and information engineer. I would love to, first of all, understand exactly what is a data scientist.

04:05

Speaker 3

Absolutely. It's a bit of an enigma because it's not that there's kind of a degree for a data scientist, so it's not something you specifically set out to become, but, yeah, basically it is very much obviously involved in the role of data, but trying to get to the deeper answers from the data. So not just what is the data, what is it telling us, but really what are the insights we can get to and all the various technologies and tools that we use to get there, basically.

04:34

Speaker 2

Yeah, I think that's probably one of the myths about data scientists. I think when I studied, I also studied engineering, but when I studied, there wasn't a course, there wasn't a thing to say. You're now studying data scientists. I think people think that you need to be some super mathematical geek to be able to do these mathematical equations. But as you mentioned, data scientists is not just about the stats. It's a beautiful combination of technology, how to apply that technology to business, and then how to do these mathematical impact studies on how to change your business processes. Solandra, I would love to understand how you went from engineering into this data scientist world and how you found the transition into that.

05:22

Speaker 1

Did you fall into the field like so many other people did? I mean, going from biomedical to doing what you're doing now, that's quite different.

05:29

Speaker 3

Yeah, it is quite a change for me. I think it was kind of always where I was heading, but took quite a route to get there. I've always kind of had an interest in problem solving and numbers and all of that. So I think for me, this was kind of always the right direction. But in terms of deciding what to study, obviously engineering was a good choice because I enjoyed the problem solving and the numbers. So that was quite an easy choice from that perspective. And then obviously enjoyed the biomedical side of it, obviously wanting to make an impact with the problem solving. So that's where that came from.

06:05

Speaker 3

And then, as you mentioned, I went down to UCT and did my masters there, and really enjoyed moving from just the academic side to actually working with data and trying to make an impact with research. So I really loved that, but then did want to move into the working world. So then I worked for CSIR for a couple of years, doing some research there in the field of biometrics, actually, such as a bit of image processing and signal processing. So still working a lot with data there. And then I moved into a startup for a few years, really a great company called Lookseedo. So they were working on biomedical technologies to help, specifically in Africa. So were working on an app to assist users with fixing medical equipment and performing maintenance, helping them with 3d models and input from experts.

07:00

Speaker 3

So that was really exciting. And then after that I moved into the consulting world because I was basically looking more for a fast paced, kind of working on a lot of different projects and getting exposure to a lot of different things. So that's where I find myself looking at decision Inc. And it's been about three years now since

I've been there. And yeah, what I really enjoy about this environment is that we're really trying to make an impact. So we take the data and we're not just representing it to our clients, we're actually taking data and trying to help them make better decisions. So I'm really seeing the impact of what I'm doing on a daily basis, which is really exciting.

07:41

Speaker 3

So I think the whole engineering and all of that, although it might sound like it's quite opposite to what I'm doing, actually very much relates to what I'm doing, because we are constantly working with technology and constantly problem solving, and that's where engineering really gave me all those up front.

07:56

Speaker 1

Lovely.

07:56

Speaker 2

Yeah, I always like to say that engineering taught me how to teach myself.

08:02

Speaker 3

Exactly.

08:03

Speaker 2

And that's very relevant to this field as well. One of the quotes that I got, data science. It's almost like this little enigma everybody talks about it. Nobody really knows how to do it. Everybody thinks everybody else is doing it, so everyone claims that they are also doing it.

08:24

Speaker 1

There's so many buzzwords. So many buzzwords.

08:27

Speaker 2

But it definitely is one of, I think for the last three years in the job, satisfactory ranking that data scientists has been in the past three years, the highest rating job from a hottest job perspective. There seems to be a big shortage in data scientists lately, and there's some stats there that says that even though there's a big shortage, it seems that people are also two out of three people actually applying is not really qualified to do this type of work. So it's super important to have that right skills to be able to do this.

09:03

Speaker 1

Also highlights just the importance of value and how critical it's become.

09:06

Speaker 2

Exactly. So, Leandra, I don't know if you can give us a little bit of, I don't know, a day in the life of a data scientist. What do you actually do?

09:16

Speaker 1

You're a wife and a mom as well.

09:19

Speaker 3

Absolutely. And that's actually become a big part of my day, especially now during lockdown as well. So it's a bit of a combination of working and being a mom and wife. So quite an amalgamation there. But yeah, typically before lockdown. Look, a data scientist, it'll really depend on what kind of company you're working for and what role you're sitting in. So because I'm a consultant data scientist, my day is typically very different, which is

one of the things I love about my job. So I'll be working for different clients, working on different projects. So my day doesn't look the same. So I'll be working a lot with clients during the day, depending on which project I'm working on at the moment. But I'll be switching between technologies and different projects.

10:05

Speaker 3

So we use quite a range of technologies, including SQL R, Ultrix, and then displaying in things like clicksense and tableau. So my day would typically be whichever project I'm working on, I'll be working with the data directly within SQL, whichever database we have the data, then transforming that data, typically through Alteryx or azure machine learning, trying to get our insights, and then displaying those in click and tableau. So that's kind of a high level view of what I'd be doing. But as I said, I work on a lot of different projects, so it really does change day to day. Whereas if it's a data science sitting in a company where they are the in house data scientist, their day might look a bit different because they would typically be delving just into their company's data, but going maybe a much deeper lens.

10:56

Speaker 3

So that would really just vary depending on the role.

10:58

Speaker 1

Yeah, chatting with a couple of people that we engage with pretty much on a weekly basis. On the topic of data, one of the responses that you typically get is one of being overwhelmed. There is so much disparate data, just the silos of data typically across the entire supply chain, just on the plant for itself in the manufacturing world. And folks, very often, when they start getting into the process of what this journey looks like, one of the first reactions that you get is one of just being totally overwhelmed, not knowing where to start, what to do, how to put it all together, and how to make meaning of it.

11:38

Speaker 1

I would imagine that's the kind of sort of departure point that you also get with a number of the clients that you work with and one of the reasons why they ask you to help them.

11:49

Speaker 3

Absolutely. And that's, again, one of my favorite things about the job is we'll go into a client where a lot of the time, things will be all over the place, and they're not really sure how to get where they want to go. And so our job is very much bringing all that data together and then not just displaying it, but actually, what insights can we get from it?

12:08

Speaker 1

And what a lot of people seem to do really well is to collate and collect all of their data somehow. And, I mean, there's different buzzwords, data lake. There's different buzzwords and theories and methodologies to do that, and they end up collating all of that. Now it's just, instead of a data lake, it's a bit of a data swamp where everything is collated, but now it's perhaps in one place, but it's still not meaningful because there's no context for sure.

12:37

Speaker 3

Absolutely.

12:38

Speaker 1

Maybe on the buzzwords. So we hear about big data, particularly me personally, I'm not entirely sure. What exactly is big data? I don't know if you recall the film Philadelphia, where Tom Hanks and the lawyer where he said to me, explain it to me like I'm eight years old. That's usually when it comes to any new buzzword or technology. That's kind of my approach. But just on the topic of data, we've been hearing about it for quite a long time, for many years. What is big data, and how do you classify any data as big data?

13:12

Speaker 3

So typically for us, it's two things. So obviously, as the name suggests, big data means a lot of data. So, as in very high volumes can be typically millions of rows of data, but it's not just about how much. It's also about how quickly it's increasing in size. So particularly today, we see a lot of that in IoT, because we've got the devices that are continuously measuring data. They could be sensors that are taking measurements every five minutes. So you can just imagine how quickly that data is increasing. So it's not just that it's a large volume, but that it's increasing exponentially. And we see that not just in IoT, but things like in the retail space. I mean, you think of daily transactions, obviously, that's already millions. And then very quickly, it gets more and more.

14:02

Speaker 3

And over time, as we're wanting more data, those transactions are getting more and more detailed, which is also making the data bigger and bigger. So that's kind of the two aspects to big data, the volume, but also how quickly it's increasing. And then also it can be obviously structured or unstructured data. So very much what you mentioned now, there could be a lot of data, but it's not necessarily good data or well structured data, but that does all fall within big data. It can be both structured and unstructured.

14:31

Speaker 1

Okay, so we're generally talking about a lot of data, structured, unstructured, but a lot of data. I think it was a Gardner piece that showed the companies now understand the importance of data. I think we all agree on that. And nearly 80% of the respondents that they interviewed of these CIOs and other executives agreed that their companies will lose a competitive advantage if they don't effectively utilize their data in 2020 this year. Lenny, you always use the example of a car's dashboard. I think you did that again yesterday, where you indicate the linking of different alerts and the meaningful context that you can get from not only understanding the speed that you're traveling at, but also how that will impact your petrol consumption. And you don't necessarily have that link on a car's dashboard.

15:19

Speaker 2

And I think that's where the problem, Larry, really is. Yes, we've got all of this big unstructured data, a lot of this data, but there's a few things that we need to take and to action that as information. I think that process, from taking big data, unstructured data, and turning that to information that you can act on, I think that is probably the most important step in this. And to understand that data, what aggregation methods do you need to apply to that data? What technologies do you need to throw on that data to actually start seeing this correlation between the different points, as you say, I always take a car's dashboard. Now, your car generates so many signals, but for you to drive the car, there's no way that you can look at each and every one of those signals.

16:06

Speaker 2

So your dashboard has been designed to only give you the most critical things that you need to focus on to safely operate that vehicle and to make decisions very quickly and to act on those decisions. And I think that's where the beauty of this whole field is, to be able to understand those links, to figure out what is that correlation between these points. And that is, for me, is very exciting to try and understand and to create these models that lives off this data to now potentially predict what's going to happen next.

16:37

Speaker 1

I want to chat with Leandro about that second point about making it valuable. But just in terms of just processing all these vast amounts of different types of data, what are some of the challenges that you face with doing that, Leandro? And do you have a technology that can do it easily for you?

16:52

Speaker 3

Yeah, look, we face a lot of challenges from a data perspective, and it can typically be quite frustrating for us from analysis point of view, because we can go into a client and we can see there's a lot of value that can be gained, but from a data maturity level, it's not there yet. And not all the structures are in place, which can take a lot longer then to get to the answers, which is unfortunate, but is changing quickly. But yeah, a lot of the things

we would typically deal with, firstly, different source systems. So we might have some data sitting in a SQL database, some sitting in an oracle database. They're all legacy setups that have been there for years. So it's unlikely they're going to change overnight to one system. So it's difficult bringing in a lot of those into our analysis.

17:38

Speaker 3

And then there is typically as well, floating Excel files, text files. I know Rowan mentioned yesterday that concept of the Excel graveyard of version one.

17:48

Speaker 1

Version two, copy, somebody called it a spaghetti spreadsheet.

17:55

Speaker 3

We deal with that on an ongoing basis. So we have some master data sitting in databases, but then a lot of floating files that we need to bring in as well.

18:06

Speaker 1

Sort of manual paper driven paper trails. Do you still find that as well?

18:10

Speaker 3

Not so much paper trails. I think we've hopefully moved past that. So it is typically digital, but not always a proper order trail of which is the correct file and who updated it last and all of that. So it can be tricky. Look, we do have a lot of tools that help us, so we use a tool called AlterX. So the name is based on alter Y and X. So that's really a great tool because it allows us to bring in various different data sources, kind of in a drag and drop interface, and then very quickly we can bring them together. So without having to do a lot of the infrastructure setup. And also we use azure machine learning as well, which does a similar thing. And that really helps bringing in those different data sources.

18:53

Speaker 3

Obviously it doesn't solve the infrastructure problem, but when we're trying to do things like pocs and quick analysis, it does really help us with that. But then even once we've got the data in, typically there's other issues. The data quality is a continuous one that we're faced with. So at a data entry level, the users are not properly trained on how to enter the data correctly. We have free text fields which have rubbish in them. We have master data issues where the master data is not properly maintained. We have ownership and accessibility issues. So especially now, companies are getting more wary about sharing their data. So we often have accessibility issues there. So yeah, definitely quite a range of challenges until we can get off the ground.

19:37

Speaker 2

Yeah, I think it's a little bit of a cliché, but definitely rubbish in, rubbish out normally with the solutions, and that's why cleaning the data firsthand, having to make sure that you've got all the outliers out of the way, that the data that you're working at is a correct subset of exactly how that piece of equipment or machine works, is critical into actually giving you a proper result. At the end of the day, just on the Excel spreadsheet, there's a source of data that normally is on a plant in Leandro, you talked about master data and all of that. And normally there's some sort of manufacturing execution system that lives on plant floors, that actually drives all of that. It's an acronym, Mes. And nine times out of ten that MEs is actually not a manufacturing execution system, it is a Microsoft Excel spreadsheet.

20:28

Speaker 2

So yes, that is still very relevant and very prevalent. And I'm sure as data scientists, you guys fall in that data. And the problem that I have with spreadsheet is that you mentioned floating spreadsheets. The problem with floating spreadsheet is now you potentially have the same data point, the same calculation done in five different spreadsheets. Which one do you believe and which one is correct?

20:51

Speaker 1

Leandro, you probably find tons of that.

20:53

Speaker 3

Absolutely. And especially when we speak to different departments, even within the same company, they'll say, oh, but I was using this version. Oh, I was using this version.

21:02

Speaker 1

I just feel overwhelmed just by that alone. Somebody mentioned yesterday that Excel, it's a beautiful tool, but it's almost like the rental car of the car industry. It's the best four x four, the best Formula one car. It's the all in one. That's actually amazing, but certainly not best for everything.

21:22

Speaker 3

Yeah, absolutely. Yeah. I mean, it's easy, so that's why I think we use Excel so much. It's such an easy data entry method, but unfortunately, just the maintenance of it is just difficult.

21:34

Speaker 1

Yeah. All right, so now you've collated all this data, you use some tools to do that for you. You've found some method in the madness. It is collated, aggregated and effectively now it's at a point where you can start to derive value from this data. It's not raw anymore. Are there any kind of real world examples or stories that you can share of companies and clients that have done that really well and how they've derived value from that and applied that in their value chain?

22:04

Speaker 3

Yes, absolutely. So we work with a variety of clients. It's across the manufacturing, mining, retail, healthcare and financial services spaces. So a lot of real success stories there. If I think of one in the mining sector, I know you guys work a lot in that field. So we've done some work with Sibaniir, Stillwater. They've been doing some regression analysis of what are some of the drivers of the incidents that happen in the mines. So that was some extremely interesting analysis work we did there. And then I work a lot with retail clients. So one client we have is Woolworths. Our team in Cape Town is doing a lot of really exciting stuff with them there and things like basket analysis and what are the shoppers buying. And then I also work a lot with the Simba group.

22:53

Speaker 3

So within Pepsico and really enjoy working with that client, they really are one of our more forward thinking clients in terms of analytics. And I think the reason for this is there's three things that they've done really well. The one is that they've got a business analyst in place who's driving all these initiatives with us, and that person already has buy-in from upper management and the C suite, which is so important with these initiatives because it starts from the top down. So that's something we've really appreciated within Simba to have that support. And then secondly, they're kind of operating our data analytics as a service model. So we do a lot of fixed cost projects for them, but also we always have at least one consultant permanently there.

23:38

Speaker 3

And this is so important because I think a lot of companies think that you can go and put in analytics and then you just leave it to run. And that's really not the case. These solutions need to be put in and then really monitored over time, continuously meet with the business to see how is this working, how is it not working, and then make improvements to it. So that's another thing that they've done really well and adopted nicely. And then also they've continuously got a finger on the pulse in terms of what their pain points are. And this is across all their different departments. So we will work on a variety of different things for the different departments, and it's because the business analyst is in contact with these people saying, what is your biggest issue at the moment?

24:18

Speaker 3

And then we come there, in there and see where we can assist. So we've worked on a range of things for them, starting from things like credit limit analysis. So which of their customers are using the credit limit? Well, which customers are potential customers where they should increase the credit limit to get more sales, or versus which other customers? If they increase the credit limit, it's not really going to make a difference to their spend. Then also things like route planning. So obviously, Simba doesn't deliver all orders themselves, but for the ones they do deliver, we've done a lot of grouping of location analysis to say which orders should be delivered together and analysis like that. We've also looked at sales reduction analysis. So things like what are the key drivers that's causing stales? Because Simba does accept sales back from their customers.

25:07

Speaker 3

If they don't manage to sell it, they will buy it back from them, which is a great model to have. But obviously it's important to understand what are the drivers causing those sales and things like where the wrong products were put into the wrong stores and that then resulted in them coming back because it wasn't the right market for those products.

25:25

Speaker 1

You can actually quite easily get some kind of a fairly simple and quick roi kind of prediction or calculation on that.

25:32

Speaker 3

Absolutely, yeah. I mean, we see percentage reduction in that consistently, and it's very easy to then calculate that their investment in the analytics was very much worth it.

25:42

Speaker 1

The retail space specifically, we're a little bit more familiar with manufacturing and mining, but I can imagine the retail space that is incredibly powerful. Any business. One of the key things that you're trying to understand is who your actual target audience is, who your perfect customers retail. I can imagine if you can combine all of this data that you have. It can just give you incredible insights in terms of exactly who is likely to do what, when they're likely to do it, and how you can engage them. That is so powerful.

26:17

Speaker 2

Yeah, especially now with the IoT sensors that's available, just giving us the capability to actually have all of this information. I know, in the retail space, and when you hear it, you think, geez, that's such a simple solution. I mean, they put up infrared cameras as an example, and they just analyze the heat maps, and with that heat map analyst, they can easily identify where's the hotspots and the cold spots in the shop. So where do I need to pack product that I want to move? You move it to, obviously, where the people go. It's quite interesting that, as you mentioned, we are from manufacturing. We think apply this to big pieces of equipment to determine failures and maintenance. But the broad applications that we can do from just mining information is incredible.

27:06

Speaker 2

I do also like that what Leandra said, and I think it's something that I said on the podcast a lot of times is just because someone else is doing it doesn't mean you need to go on this journey if you do not have buy in and not have a clear user case on what you actually want to achieve with this. Just because big data is now a big word, a buzzword in the industry, like we've seen with mobility and digital transformation and digitization strategies, you must make sure that you couple this with a strategic initiative that's going to give you ROI, and it's actually going to solve a business problem. And Leandro, you probably have some horror stories as well know. If you don't have that, the project is pretty much a failure, or the ownership is not taken off completely.

27:53

Speaker 2

And rightfully, as you say, you think you install this thing, it's something you buy off the shelf, you install it,

you leave it, you get your results, and then you leave it, and that's who it should be. But surely there's some horror stories as well about people not getting the right names, getting it right, and it's not the technology's fault, it's not the data's.

28:13

Speaker 1

Fault, and I think we typically defend the technology.

28:16

Speaker 2

Yes, exactly. And I think technology gets the blame quite a lot, but I don't think it's always the case.

28:23

Speaker 3

No, absolutely. I definitely won't mention names on this one, but we definitely have seen some cases of that. And it's so sad because the investment is put in and we'll go in and do something but if their buy in isn't there and we're not brought in on an ongoing basis to maintain that solution, a few people might look at it, but they'll see something wrong with it, and then they'll say, oh, no, I don't trust this, and then they'll just stop using it. And then there's definitely no ROI because it was an investment, but now there's no improvement on it, there's no maintenance, and so it unfortunately will fall short.

28:57

Speaker 1

You've just mentioned something, Leandra, that I thought of. Now they see something. Do you find that sometimes when the truth is exposed, do you sometimes find that there is a reluctance to commit or go further once the truth is exposed? Or sometimes there's, oh, we didn't realize this is actually what it is. We have to actually manipulate it. We can't show this. Have you had some of those examples?

29:24

Speaker 3

Sure. Yeah. And there's also a lot of where people are basically scared for their own jobs based on what they're going to reveal. So instead of looking at it as, okay, wait, we can reveal these things so that you can improve them and you can then spend time on other things. Unfortunately, a lot of people are nervous for their job security. So, yes, they are concerned if we are opening up too much insights from the data that might reveal that there are inconsistencies, inefficiencies. So certainly that's, again, why we need that buy in and that confidence from upper management to remind the employees that all this analytics is here to do good for them and to help them. It's really not to put them out of a job.

30:04

Speaker 1

Yeah. Rosa, we've spoken a little bit about the ownership piece, the human change management, if I suppose we can call it that, for organizations and companies, and typically the customers that you help that are looking to become more data driven. What should some of their strategies that they apply to start with this journey include? If you can maybe summarize that for us, I think that'll be really valuable.

30:28

Speaker 3

Sure. Yeah. So basically, it's not a simple thing that happens overnight. It is a journey. So that's what we always tell our clients, that this isn't going to happen quickly. We need to plan for it and implement it over a long period. But the typical advice is, let's start small with analytics to prove the concept, while at the same time planning for the bigger future. So there's a lot of hype around a lot of these analytics concepts, but when you go in blind as a company, often it will fall short of those expectations. If you're trying to apply analytics to everything all at once, what we say is let's not try and go for everything. Let's start with something simple. So what is a typical pain point that you have in your everyday business?

31:07

Speaker 3

So, for example, maybe you're trying to understand variances in unit price with your procurement. Let's do some analytics there. Let's do some regression analysis. And let's first show the business users that just a small project like this can give them the answers they need and can help them determine things like which products are

affected by the vendor, which products are affected by the month, which products are affected by the quantities that I buy. So start with something small like that, and immediately once the business users see that actually they can really gain something from it, then it will start to pick up. So it's really important to start small, but then at the same time plan for the bigger picture.

31:46

Speaker 3

So you need to have the people, the process and the technology, so the people to drive it, the process in place, and then the technology. So from a technology standpoint, that infrastructure is really so important. So some inputs here from a colleague of mine is that when we're setting up infrastructure, there's kind of five things that companies need to plan for in their strategy. So a data lake, we spoke about that a bit earlier. So having a data lake there for the initial storage of mostly unstructured data, then a modern database to start structuring that data, then a data movement and preparation tool, we can manipulate and cleanse the data. A data exploration tool where we do our data science and our more in depth analytics. So that would be like the Alteryx or Azure machine learning studio.

32:34

Speaker 3

And then finally a business intelligence tool where we can bring everything together and visualize it, because it's great to have very exciting inputs and exciting outcomes from a data point of view. But if we don't present those correctly to business, it's also not going to be accepted. So in the planning, in getting that structures in place, it's important to start with something small, but get those five technology pieces in place so that then when you are ready and you mature, those things are already there and make the analytics a lot easier.

33:04

Speaker 1

Definitely when that maturity is there. I also love the fact that it's a journey. It absolutely is a journey. It's not something that's going to be solved for or something that you're going to see the immediate value after a month or two months. It's definitely a journey. And it's important to understand the milestones and the points on that journey, that's going to get you to a point where you can extract that value.

33:25

Speaker 2

And I think these days it's actually been made quite, I wouldn't say quite simple, but it's made a little bit more tangible. Cloud technologies allow you to have that data lake up in the cloud. You can inject that with your plant information or your plant data. A lot of the technologies are really spoiled for choice. We are spoiled for choice. I do believe that. And a lot of these companies are geared up to do these analytics and provide you with the tools. Leandro, like you mentioned, I mean, there's a whole bunch of subsets of tools. You don't have to go and write the queries and the regression analytics yourself to actually do that. We are spoiled and these tools are available. It's literally for us to have that data in place to feed into these systems, to actually start making these decisions for us.

34:11

Speaker 1

Andrew, I love what you mentioned about the, you mentioned people, process and technology. What I love about that is that you start with the people, I think, very often, and you'll probably find that there's an expectation that you start with the technology and that's all you need. I love that you start with the people and then the process and then the technology.

34:32

Speaker 3

Absolutely. We've definitely found that's critical to success. It's very much about the people and when they want the analytics and they're continuously asking us what more we can do, then that's when we know we've got the right client that's ready for the analytics.

34:46

Speaker 2

Yeah, I think that's the important part. That's right. For the analytics part. I think even after the people and the process steps, you might have already identified something that you can improve or processes that you need to tweak a bit, and then, literally to move into that next step within the analytics is to now start using machine

learning, et cetera, to drive that. And I think I might be wrong, but I think when people think of this, they immediately go that route. Oh, I must include machine learning and AI to my journey. And as you mentioned, start small, just clean the data and do some very simple analytics on that data, and that will already point you in a direction of where to go. Right.

35:26

Speaker 1

Leandro, data scientist, information engineer I'm going to put you on the spot. What are some of the trends for data and data management that you predict over the next few years? And I suppose from your point of view, there's perhaps a pre Covid view on what that will be, and now there's a post Covid view. Maybe there's even some of that.

35:47

Speaker 3

Absolutely. Maybe not question no, that's absolutely true. For me, what's really exciting as a data scientist is that the role of data is really changing. So we can see that trend definitely happening. That a few years ago there was a lot of data, but it was just seen as a byproduct, whereas now we're starting to hit the trend where data is being seen as an asset and this is only going to pick up in the years to come. So we are going to get to those answers a lot quicker because people are realizing that and structuring their data better to get to those answers. So initially there was this massive peak about AI and all of that, but no real value gained.

36:26

Speaker 3

I think we're now in the more the dip after that where it's going to be implemented correctly and then getting to the real value, which is really exciting. And as you say, Covid has really speeded up that process because firstly, a lot of people have seen how valuable data analysis can be. So just with analyzing the pandemic, it's been so incredible, the things that data has shown us about it and that we've used data to really analyze it. But on top of that, people have been working from home as well. So the data infrastructure has needed to already incorporate that and very rapidly move to that. So we've seen a move to cloud much quicker and definitely accelerated by the pandemic. That's very exciting. And then some other thoughts, and this is not all from me.

37:11

Speaker 3

There is a colleague of mine who's helped me as well with some of these inputs, but some other trends we're looking at seeing is a lot of moves to one stop. Analytical tools. So things like Snowflake, click suite, Azure, sign apps, things like that, where everything can be done in one tool. There's also some really incredible things coming out, like digital twins, IoT and sensor data. That's all very much on the rise. And then also we're seeing a big trend to more data governance. So that idea that we need to have correct and clean data, but also that we need to manage who is seeing it. So privacy is definitely becoming more of an issue. So we're going to see where I think at the moment we're seeing a lot of people giving access quite freely to data.

37:54

Speaker 3

I think that is going to change and people are going to become a lot more concerned about privacy. But then, as well, the things we've spoken about already. So the machine learning, forecast modeling, algorithm driven recommendations and decision making, that's only going to increase in the years to come, which is really exciting for us. We're really moving now from a reactive phase to a proactive phase. So before were starting to look at exactly why things happened in the past, but now we're actually starting to say, okay, well, we know why it happened, so let's now use that to predict what's going to happen going forward and plan better, instead of just understanding why something happened. Those are all really exciting for us to see.

38:37

Speaker 1

Definitely is. I want to ask you about the poppy act. So the poppy act in South Africa, it's an effect. Have you seen a little bit of a different approach towards strategies around data? Maybe a little panic from some people in terms of how that will affect them and what they have to put in place?

38:55

Speaker 3

Yes, certainly some clients are getting there, and we do tend to have to go around the security protocols first,

whereas definitely a few years back, the clients were maybe more willing to give us access, give us database admin so that we can access all the data. There's definitely starting to be second thoughts about that now, and not always from the analysts necessarily, because they would want us to have access to the data. It's typically more from your IT security aspects, which makes sense. Those are the roles that would be more affected by the security.

39:28

Speaker 1

Yeah, it does, definitely. And tying in with that, I think, let's call it a data breach. It's probably one of the more prevalent sort of occurrences that we see in the world nowadays is just data security and data breaches and call it cybersecurity. That is maybe at an alarming rate, seemingly increasing a little bit, as all over the. Over the last little while, and people looking to protect their data. Are you guys in any way involved in potentially what that looks like in some of those processes?

40:02

Speaker 3

So not so much from the security point of view. We're more on the data management side. So typically the security clients would prefer to do that in house via their own IT departments, which obviously makes complete sense that they want ownership of that. But, yeah, so we wouldn't necessarily be involved with implementing it, but we would definitely encourage it. And I think that's where we're also getting to more now, is that even if the client is willing to give us access to the data, we the ones who sort of question it and say, well, are you sure we should have this level of access?

40:36

Speaker 1

Yeah, can imagine. I think we've got to cover cybersecurity. In one of the upcoming podcasts. Definitely. You mentioned Iot a couple of times. Do you see sort of an increase in data from IoT? I mean, Iot can be many different things. Your phone as an IoT device, your fitbit that you're wearing is potentially an Iot device. So Iot as a word and a concept has actually been around for a very long time. But now that we understand the structure behind Iot, it seemed like there was a lot of promise around Iot over the last couple of years. Maybe there was a challenge with devices and networks and costs and scaling, but do you see some more IoT type data entering the data sets and then data environments with some of the customers that you're working with?

41:27

Speaker 3

Absolutely. And especially within the manufacturing fields and in plants and things like that, these tools are going to be so important. And actually, the startup I worked previously was also looking into some IoT devices and IoT solutions, especially when we're looking at things like into Africa, where we maybe don't have access to get directly to the plant, but we still want to know what's happening there on a continuous basis. And that's where the Iot devices are so important. And for things like proactive maintenance and things like that, it's really important that we can get that information. It's definitely possible. So it's just a case of putting all the technology and things into place.

42:07

Speaker 3

I think what has been hard with Iot is that it is so much data, and what maybe a lot of companies have been wary of is then, okay, this is now this huge data dump. They're paying to store this data, but what are they actually getting out of it? So, again, it's very much about, there was this huge hype at the beginning, but now we're more getting to the part of, okay, well, let's implement it properly, and that way we can get to the true value of it. But there's definitely, I would say, a lot more to come on the Iot side.

42:34

Speaker 1

Yeah, that's exciting.

42:36

Speaker 2

I'm glad you said that. There was a lot of hype, because there was a lot of hype. And obviously, with any new technology being big data, mobility, IoT, there's always, and Gartner's got this thing, they call it the hype cycle. And I'm glad that you mentioned that we over the hype. So that means we've smacked bang through, down on

our faces, through the trout of disillusionment about this whole thing, and then we're actually moving into the plateau of productivity with the hype cycle. So I'm very glad that went through that cycle and we can actually now start employing that.

43:07

Speaker 2

Last thing I want to comment is that you mentioned that, yes, people are starting to send more clean data to their warehouses and I think that's very super important is try and get your data as clean as possible from the source before you go and deploy it to these upper technologies that you employ. It will definitely just make your whole process and your whole data analytics solution so much easier and so much more effective if you can start cleaning your data from a lower level before you put it into your data lake or whatever that case is. So, yeah, it sounds like there's some interesting times ahead. Definitely.

43:41

Speaker 3

No, absolutely. Very exciting.

43:44

Speaker 1

Andrew, thank you very much for chatting to us about this journey of information and transformation, and let's just call it the information value chain. I wanted to a little bit more personal. Given your background and your experience, you're very passionate about what you do. Very clearly. You very well do it very well. Any advice for anyone considering a career in data science analytics? Given how broad and wide it has become, I think it is definitely a hot item as far as careers and jobs go. Any advice for young people perhaps still studying, just finishing up their studies, considering a career in the field of data?

44:29

Speaker 3

Absolutely. So starting from the degree side of it, obviously a degree in something technical helps. But as we discussed initially, there's no degree in data science. So it really can be a range of different degrees. The ones we would typically look at is engineering, bses, maths and statistics. Obviously actuarial science is a good one because there's that real stats core. So there is quite a range of degrees you can study, but ultimately something technical does help because that's giving you the right skills that you need to do the deeper analysis. But also it shows that you already have an interest in the technical side. To have chosen a degree like that is important and we do look at that, but it's not just that. It's very much more about the wider skill set.

45:13

Speaker 3

So we often talk in data science about a t skill set, so it needs to be wide. So you need to have a lot of different skills, ranging from technical, academic presentation, storytelling is actually extremely important.

45:27

Speaker 1

Very good point that.

45:28

Speaker 3

Yeah, and I think that's where a lot of people might have the technical, but they can't necessarily communicate the insights either to their own company or to the clients, which is unfortunate. So it's really important to have that wide set of skills. But then also deep within, specifically the technical side of it. So do you have those deeper technical analytic skills to actually be able to perform that side of it? But then also, what have you done on your own to kind of self teach yourself? So, because it's not a degree in data science, if you come to us with saying you've got an engineering degree, that's great, but what have you done in your own time to investigate data science? Have you done any projects of your own? What skills have you self taught yourself?

46:15

Speaker 3

And we talk a lot about a citizen data scientist as well, which is becoming much more prevalent due tools such as alteryx, where it is very much these drag and drop interfaces. You can do things very quickly. So have you played around with those tools and tried to become a citizen data scientist first? Have you done courses on data

camp, things like that? So we can do further training once the data scientist is with us. But it's important that we see that they've had that initiative on their own and really shown an interest in the field already.

46:49

Speaker 2

I think that's probably one of the myths we spoke about right at the beginning. A data scientist is just a math geek that you put in a corner.

46:57

Speaker 1

But if you lock away in an office.

47:01

Speaker 2

Slide a pizza underneath the door every now and then, you definitely need to have that combination of technology. So you need to be interested in new technology trends, understand how to apply those technologies, but also have that little bit of business skills to do these type of consultings, to go out to interact with people, to understand the processes and draw up those processes. And then obviously, you need still a little bit of math number crunching, but definitely it's a very broad skill set that you would require.

47:30

Speaker 1

Yeah, I think everything that Leandra pointed out there, super great. And maybe the one that wasn't obvious to me, Leandro, was the ability to tell a story and articulate it very well. That's one that wasn't so obvious to me. So super happy that you mentioned that. And it's actually, to your point, it's actually super critical, it is to be able to do that effectively.

47:49

Speaker 2

I mean, the regression analysis can spit out an equation for you.

47:52

Speaker 1

Right?

47:53

Speaker 2

What is the r square for linear regression? But if you can't tell the story about what that's telling you, it's super important to relate that back to actual business value.

48:03

Speaker 1

Yeah. Leandro, fascinating stuff. It sounds really exciting, actually, I'm quite interested in understanding a little bit more, a little bit better. Maybe I should have made different career choices, but fascinating stuff. I love the summary about the skills needed and I think that's a very good summary for somebody considering data. Maybe also just some of the myths around. Typically what you do and what the job entails is definitely a lot more exciting I think than people would expect it to be.

48:35

Speaker 3

Absolutely. Yeah. I will just mention here as well, we are always looking so if there are any data scientists or someone interested who is listening, take a look at Decision Inc. So that's decisioninc.com. We do definitely always have openings for data scientists, so anyone who's interested can just contact us.

48:55

Speaker 1

Definitely. We're going to share your email address if that's okay, as well as the website. We'll share it as part of the description of the podcast. But amazing chatting to you landra, thank you so much. We can probably chat with you for another couple of hours as such a fascinating topic and such a relevant topic.