

## **SAIT Podcast: Navigating the Labyrinth of Careers in Data Episode 8**

[00:00:00] ANNCR: The Best Careers You Never Knew Existed Podcast. Sparked by SAIT and co-hosted by CITI, the podcast that helps you navigate jobs, learn about new careers and industries.!

[00:00:12] Lora: Hello and welcome to the podcast. In this episode, we're talking to Ben Reeves about data. Stick around for the resources and advice after the interview.

[00:00:22] ANNCR: Now, here's a career you never knew existed.

[00:00:25] Lora: Thanks so much for coming, Ben. Could you please introduce yourself?

[00:00:30] Ben: Sure, my name's Ben, and I work as a data scientist and software engineer. And I am a community builder who's passionate about helping Calgary realize its technology potential. Yay, community.

[00:00:41] Zach: Yay, community, love it.

[00:00:43] Lora: Today, we're here to talk about data. So why is data so important?

[00:00:47] Ben: I mean, a good way to think about data is it's just a quantified piece of information, right? And we as humans use information to make decisions. So, data is really about optimizing the decision-making process, which you can think that touches all aspects of industry, academia, etc.

[00:01:06] Ben: So, kind of what we're looking to do here in the data field is you know, one thing, make that collection of information easier and better, right? So that's when the, that's in the data collection, data curation space of things. Then there's also, okay, now we want to extract some type of insight from that data.

[00:01:25] Ben: We want to turn that data, which is a quantified piece of information, probably not very useful on its own, you know, you've got some piece of information like, oh, IP address, 10. 4. 13, you know, logged into Facebook at 10:43am, right? Doesn't really mean a lot on its own, but whenever you put a lot of those small pieces of information together, you can start to glean some type of insight from that data, and then you can take that insight and turn it into an actionable decision.

[00:01:55] Ben: And then, you know, Say rounding out the process or rounding out the data pipeline, you then go to use something maybe on the data science side of things where you're trying to turn those insights into automated decisions or, you know, using statistics, mathematics in order to help you make better decisions with the empirical data as opposed to just maybe examining or looking at the data and then having a human come up with an insight from examining all of that at once.

[00:02:24] Zach: Do you believe the notion that data is the new oil?

[00:02:26] Ben: Maybe, I would say that data permeates everything that we do. It's more of a catalyst, right? Like, oil has a very immediate and direct feedback mechanism for the impact that it has on society. It's called oil, an analogy for the energy industry as a whole, right?

[00:02:48] Ben: Energy makes things tick. Data makes things tick better, faster to go on the oil side of things. We think about an internal combustion engine, right? The original Model Ts were like 15% efficiency at turning gasoline into usable energy. Data is what lets us go from that 15% to 60, 70, 80. It lets us accelerate the processes that we would have done otherwise.

[00:03:16] Zach: What are the top skills and knowledge data professionals need to know in order to move it from... The 15 to the 70%?

[00:03:23] Ben: Yeah. So, the data space, um, maybe I'm going to take a little step back and kind of explain what the data space is a little bit so that that'll provide a bit more context into what we're thinking about.

[00:03:35] Ben: Right? So, I usually I explain the data industry as a generalist Industry and there's really three main pillars. So, you have the business insights, that information gathering, you have the technology side of things, building the pipelines, and then you have the predictive analytics turning those insights into automated decisions.

[00:04:01] Ben: And across those three pillars, the data industry kind of encapsulates all of them. And you can pick and choose to which extent you want to direct towards one of those pillars or another. And so, with those three pillars, you know, you might have somebody like a data engineer, right? Who's really focused on building pipelines, on gathering information, on making that data accessible to other people.

[00:04:26] Ben: They're gonna sit a lot closer to that software engineering side of things, that software engineering space. Then you could have a data analyst who's going to mostly take data that somebody else has gathered and look for patterns in it, look to figure out, okay, we've launched a new widget, a new, um, a new t shirt and we want to figure out which demographic we've sold the most of these t shirts with so we can create a targeted ad campaign, right?

[00:04:54] Ben: And the data analyst might take all of the sales information that's been gathered by the company so far. Look at that and pick out some patterns or some geographic dispersions and try to come up with some of those actionable insights. And that's going to be more on that business analytics side of things that takes a deep understanding of the business of the domain.

[00:05:11] Ben: And those people who are working in that space are going to be much more attuned to industry, let's say. And then you've kind of got the predictive analytics. This is also where maybe machine learning and AI starts to come into things a bit more. And that's going to be focused on the turning data into repeated, automated decisions.

[00:05:29] Ben: And so that's going to be where you're going to have a lot more of statistics, mathematics. Um, you're going to be really focused on model building, where you're taking that

clean, prepared data that you've got and using statistical models to create automated decisions. So, when we're talking about the skills that are required in that data industry, it sort of depends on where you're sitting within that three pronged spectrum.

[00:05:56] Ben: Obviously, for all of them, curiosity is going to be an absolute must. Resiliency, because as much as we like to say, oh yeah, it's easy, you're turning data into information and then using that information to make decisions, right? Sounds very nice and easy on paper, the reality is you deal with a lot of, there's a lot of cleaning of the data, dealing with corner cases with ambiguity, you know, humans don't really produce the cleanest of data or measurement systems, fail temperature pressure gauges in an oil well or on an assembly line, malfunction, and so you have to do a lot of massaging of the systems. So, definitely having a good aptitude for resiliency and really feeling that satisfaction with like a job that's been completed to be able to push you through whenever you're working on things that aren't going quite the way that you want them to is a good skill.

[00:06:54] Ben: And then, yeah, everything within that STEM field. So, you know, an analytical mind is important because you are at the core breaking things down into their constituent pieces, but you've got to keep that high level. Heuristic or that high level understanding of the, you know, not losing the forest for the trees because we're trying to make decisions, right?

[00:07:17] Ben: And we're trying to improve the decision-making process. So, you really need to understand the problem space that you're operating within in order to optimize those decisions and make sure that the map doesn't lead you if data is the map, you need to take a look and view the landscape so you can make an accurate, an accurate map.

[00:07:35] Ben: So yeah, I think to summarize, you need that curiosity. Resiliency, an appreciation and hopefully a love of the STEM areas. And yeah, if you put all that together, then you're probably going to have a pretty good skill set for the data world.

[00:07:51] Lora: Would you say that almost in any company, organization, and industry, that they need data professionals in some capacity?

[00:07:58] Ben: I would say that every company should have a data strategy and should have somebody within their organization who understands or at least has a vision of where data can be used to augment, to catalyze their business. But to have a full-time data professional, I think there's a lot of cases where you need to generate the data, you need to have a plan and you need to have sufficient scale in order to be able to pay a relatively high paid professional to take all of the information that your organization is currently generating and improve that by 10 percent, 20 percent, 2 percent, whatever, whatever that number is.

[00:08:43] Ben: So, I think having data understanding and a vision and an understanding of where it can be useful is important. I would say, depending on the industry, if you're trying to start an e commerce shop, yeah, you should know how that data could be used in the future, but you probably don't need a data professional right away. If you're a company of 150 people, probably no matter what you're doing, you could have a small data team in order to help out and improve.

[00:09:15] Lora: From my perspective, it seems like everybody starts with Excel. We love Excel, I think, in a lot of industries in Alberta. I'm curious about how widely Excel is used and what other tools are being used to analyze and convey data and tell the stories behind it.

[00:09:28] Ben: It's funny, I don't use Excel anymore personally. I mean, that's not true. I do still use it, but I will defer or I will start with other tools, but that being said, I always continue to sing the praises of Excel because it's low barrier to entry. It makes sense, you can see what's going on, you can understand it, you can see the data, you can make a change, you can watch a transform.

[00:09:52] Ben: And it's a tool that most of us grow up working with and feel very comfortable working with that. So, Excel is kind of this, maybe call it a gateway drug into the world, into the world of data. You know, if you've ever been working on an Excel, Excel sheet, and you're like, Oh, let me just create this tab, and ooh, what if I just make a chart here, and I add a little macro so I can press this button to update all of my charts.

[00:10:17] Ben: Excel is constrained by... your personal computer that you're working on and the data that you can put into Excel. And that's where we start to need other tools. You've probably heard a lot about big data. You start to deal with data sets that are on the gigabyte, terabyte, petabyte. The big players are dealing with exabyte scales sometimes.

[00:10:40] Ben: And so, it's one of those aspects where Excel just can't handle that. And then we get into other tools and technologies. So, call Excel, the gateway drug, then you've got SQL, S Q L. This permeates everything. It's, you got to know, you got to know SQL. And that is a database. When people talk about a database, hopefully they're referring to a SQL database and not an Excel table.

[00:11:06] Ben: I don't know if you guys heard the, the story from back in COVID, I think it was that a hospital lost a whole bunch of patient data because they had stored everything in an Excel table, which then whenever they lost that, they don't have backups. And that's why you want to store your data in a SQL database or some equivalent database technology, as opposed to a, as opposed to an Excel table.

[00:11:30] Ben: So, SQL and databases become a, call them the bread and butter, meat and potatoes. Then you're going to have lots of different programming languages. Um, Python and R are probably going to be the two that you hear the most. And then as you move up into different areas of specialization, if you're working with really big data, you might start to hear things like Apache Spark and PySpark and Databricks.

[00:11:53] Ben: Those are going to be tools and technologies for working with big data. And then if you're working, you know, maybe more on the inference side of things on that business prong, that business intelligence prong of the data sphere, you might be hearing things like Tableau or Power BI, really tools that are the next generation of Excel or maybe Excel on steroids.

[00:12:16] Ben: They kind of offer some of that familiar interface, but they've got the

opportunity and the power to work on larger data sets to perform more complex visualizations. To create more interactive tools and systems that can be shared and distributed.

[00:12:30] Zach: What is a data lake?

[00:12:31] Ben: So back in the day we used to have databases all this time. And everyone said, okay, listen, like we've got all of this structured data. We're paying data professionals, database analysts in order to define, think of an Excel table, right? And you've got a header row and then your rows of data and everything fits really neatly and nicely into that table. You put one number or one word in each cell.

[00:12:55] Ben: And then the Internet happened. And then IOT things happened. And all of a sudden we're producing all of this Data, and there's so much of it, and it's happening so fast, and some of it's structured and it fits nicely into a cell in an Excel sheet, but a lot of it doesn't, and it's, you know, internet posts, what we call unstructured data, like articles and news and images, people are like, I don't know what to do with all of it, just dump it in folders.

[00:13:26] Ben: And that is a data lake. It's a really, really fancy cloud folder. So, it's a, it's a place where you can store all sorts of different data, whether it fits nicely into a, you know, table at some type of structured data, or if it doesn't. What the difference between, you know, maybe the unspoken question, the difference between a data lake and a database is a database is very, it's very structured.

[00:13:48] Ben: You can look at it and you know where to find everything. If you're looking for the number of logins to your app, you're going to have a table with all of the logins to your app. If you're talking about a data lake, things are going to be, I don't mean this in a bad way, but it's more of a dumping ground.

[00:14:05] Ben: You know, you're going to need to know where that data is to go and find it. It's not going to be in any pre specified structure or format. And This structure works really well for a lot of big data applications where things are being generated really, really quickly and at a really large scale, which conventional databases might struggle with, and you just need to pick up large amounts of data at a time and perform some type of analytics or inference on that in order to make a decision, train a machine learning model, et cetera.

[00:14:37] Lora: So, I was going to ask about Alberta, where many traditional industries have historical data and healthcare and energy, as an example. If you have a data set that contains a bunch of potentially old spreadsheets, Excel databases, SQL databases, whose job is it to unravel this new data, old data, historical data? Is that the responsibility of a data engineer?

[00:14:59] Ben: Yeah, it sounds like you've done your homework. I think you hit the nail on the head. And it's funny because a lot of times people ask me how to break into the data industry in Calgary, and my go to response is actually to look into data engineering, and it's for that specific reason.

[00:15:15] Ben: I'm going to make a contrast to some of like the, you know, the tech companies

out of the valley or out of Tel Aviv and you've got your Airbnbs and your Spotifys and your Metas and you hear lots of stories, you read a lot of stories about their fancy machine learning models and their AI tools that they have.

[00:15:32] Ben: The reason that they're able to build those tools is because they started as tech companies and they, from the beginning, had a vision for how that data might be used and because they were tech companies with, you know, back when they were just software engineers with specializations in data systems, they built all of the pipelines to gather, label, tag, Organize, sort, collate the data into reusable, parsable, understandable structures.

[00:16:05] Ben: And it took a lot of work and they put a lot of investment dollars into doing that. And after doing that for 10 years, now they're able to reap some of those benefits and turn that data, those pieces of information into actionable insights and do it at scale. Then we come to Calgary, which has just great Opportunities for data like healthcare systems, of course, always great, that's kind of across the globe.

[00:16:29] Ben: We've got the energy industry here in our backyard as well, which provides a really, really great rich opportunities to catalyze operations using data and extracting that information, but you have to build the pipelines. You have to do the plumbing, do the infrastructure.

[00:16:48] Ben: These algorithms, these, you know, inference, being able to get information from data. Computers are really dumb, you know, and you really have to spell everything out for a computer. And you can't label one column pressure and the other column press. Without the rest of the UR and maybe a human could look at it and be like, Oh, these are the same things I can understand that.

[00:17:11] Ben: A computer is going to be like, I don't know what that is. These are two completely different. One of them says pressure, one of them says pressur. Different. So, there is a lot of work that needs to go into, as you're saying, as you're hinting to, Lora, being able to put all of that data together, being able to take some of those legacy systems and extract that, combine it, put it together into one big kind of master consistent data set.

[00:17:38] Ben: And that's the job of data engineers. There's a lot of opportunity for that because like we said, if you want to make those decisions off your data, first you need to be able to work with it. And the data engineers are the ones who are going to be responsible for taking what exists on all of the old legacy systems putting them all together and then being able to take actionable insight off of those.

[00:17:58] Zach: Ben, you started your career as a software engineer?

[00:18:01] Ben: Sort of. I mean, everyone in the tech industry here locally was an engineer originally. So, I mean, my, my traditional schooling was in materials engineering, actually.

[00:18:12] Ben: Which is, um, you know, the science of metals and ceramics and semiconductors and polymers and plastics and stuff like that. But I started my real getting paid

career as a software engineer. Yeah. Okay, so tell us a little bit more about the journey yeah.

[00:18:30] Zach: Okay, so tell us a little bit more about the journey from studying materials engineering to software engineering. And now it seems. although I think you do both like software engineering and data, but you are quite the community builder in the data space. And, you know, we've invited you here to talk about data. Talk to us a little bit more about that, like, arc of how you've walked through and how you've gained such a strong interest and knowledge in the data field.

[00:18:52] Ben: Yeah, I think earlier, Lora, you were saying that the title for this podcast might have something to do with a labyrinth. I feel like that's a very fitting title for the wandering, meandering, and oftentimes aimless journey that I've taken in order to wind up where I've wound up. So, yeah, I mean, I guess to give the whole spiel I kind of first fell in love with software while I was still in school, trying to figure out if I wanted to do a masters.

[00:19:19] Ben: We were heating up titanium alloys really, really quickly and then cooling them really quickly to see how they formed crystal structures. Metals, you might not know, but metals also have a crystal structure just like any other crystal that you see, crystal rocks. Metals have crystal structures too, and we would heat them up really quickly, cool them down to see how that happened.

[00:19:39] Ben: And while doing that, we worked with Fortran and Python applications, and I thought it was just the most beautiful thing to be able to first create some theoretical model as to how things should work, how these crystal structures of these phase transformations should occur, model it out, And then go and actually run the real experiment and then look at the results in an electron microscope.

[00:20:03] Ben: And so that was kind of my foyer into into software and seeing how having a model or a theory of the world and then embedding that into an automated decision making process, which is what software is, and then being able to compare that to what empirically you observe in reality, doing an experiment and collating those two things was able to have meaningful impact and guide the journey.

[00:20:28] Ben: So that's kind of like where I fell in love with software and then graduating university, good jobs for material engineers at the time all looked like looking at pipeline cracks all day, which I wasn't very enthusiastic about that. So, I took the software practices that I learned in school and then moved right into the software industry in the oil and gas space.

[00:20:49] Ben: So, I then started working with temperature, pressure, flow rates, with SCADA systems, so you've got a natural gas field, and you are, or an oil field, and you're grabbing all of the natural gas in the oil, collecting it from all of the different wells, Putting it through smaller pipelines and then bigger pipelines, then ultimately a processing plant.

[00:21:11] Ben: And I was working with devices that measured all of that. And so all of a sudden I was seeing across all of North America, you could see how the energy industry was going from each individual little oil well on Farmer John's field, getting gathered and then pushed into, you

know, these larger pipelines and seeing everything at once.

[00:21:35] Ben: And that piqued the question of, oh, hey, we have all of this information. I can see the beating heart of the North American energy industry. What kind of information can you find from that? You know, oh, hey, what are auction prices on natural gas pipelines doing right now. And if I can see the broad data of what's happening in all of the different distribution lines, could I predict what auction prices are doing on the pipeline?

[00:22:04] Ben: Turns out that you can. And oh, are we able to detect whenever we get an alert for some anomaly in the data pressure temperature? Can we actually automatically label that as being a device malfunction? Or is it something, a real alert that needs to be addressed and that was kind of at the time where that interest of mine started at about the same time that some of the first real advancements in some of like the deep learning models were really starting to take off.

[00:22:35] Ben: Um, we were starting to get really consistent results in the computer vision space. So having an algorithm be able to look at a picture of a cat and say, this is a picture of a cat. Something that's really easy for humans to do, but really hard for computers to do. And so kind of given the confluence of those two things of like my interest in being able to see how the data from a software system could be used to understand the energy industry and some advancements and excitement that was just starting to get moving in kind of the MLAI space, I started to pivot more into the data science world within the company that I was working for.

[00:23:10] Ben: And we were lucky enough to have a 20 year data science veteran from, you know, the days when they were just called statisticians before we called them data scientists. And so started working, working on some like intrapreneurship projects within that firm.

[00:23:27] Ben: And then as that interest in really that predictive analytics side of things from the, we talk about those three prongs, you've got the business intelligence, the data engineering, software engineering, and then the predictive inference. So having a strong interest in all three of those, but especially kind of like at the time, drawing interest in that predictive inference led me more and more down that data science path, which happened to align quite nicely with a Local Calgary company that was trying to use a lot of those tools and technologies, had an interest in using those tools and technologies in the financial industry.

[00:24:09] Ben: And so that's whenever I got involved with Viewpoint Investment Partners, so we're a quantitative investment management company. We're basically using financial data from global financial markets in order to better understand the risk and reward profiles across the investment industry. So, I joined up with them very, very early on in the process to kind of take that software engineering data science and be able to apply that into the financial industry, into the mutual fund or the investment world.

[00:24:41] Ben: And so, yeah, that's kind of the journey. There was never any plan or any overarching flashbulb in the sky. It was more of a just sort of being led around by your nose, right? Oh, that smells interesting. Wonder what I can find out from there. Oh, that smells interesting. Let's see. What's going on over here and just kind of following that path of curiosity led me into the data world.



[00:25:06] Lora: Thanks for validating my podcast name. I'm curious about how you learned all the things you needed to know about data and programming.

[00:25:12] Ben: I was very fortunate to have both great mentorship as well as being given way too much responsibility way too early in my career. Really, really terrible decisions by my mentors.

[00:25:26] Ben: Yeah, it's funny, like, at the time, and still, there's everything that you could ever need to learn exists on the internet, especially in the technology spaces, which, you know, data engineering, data science, data analytics, those are within that tech space. So, there's everything that you need to know exists out there.

[00:25:46] Ben: So that provides a great tool and a great resource. The reality for me at least has been that most of those tools and resources have been almost secondary to learning on the job. Yeah, so the primary motivation has always been having a strong mentor who can provide guidance or ask the right questions and then being able to turn to the wealth of information that exists on the internet and as well textbooks.

[00:26:18] Ben: I've got a big shelf of very nerdy textbooks that you know you have to go through sometimes. And there's one thing about people who have pivoted into the tech industry. And most people, at least locally in Calgary, who are involved in the data industry are, to some extent or another, people who have pivoted, like this field as we know it right now hasn't really existed for that long, like the field of data science is kind of a relatively new term, so there's not that many people who have been conventionally trained in it.

[00:26:50] Ben: And so that self-starting ability, the ability to like, be curious about a problem and then go through and find the resources that you need to teach yourself, validate it against a real world, validate it within like a network of peers or mentors is probably the most crucial thing. So yeah, I don't know that that answer is like a, Hey, there's a really clear and distinct path, though there are a ton more resources now.

[00:27:17] Ben: Um, and we can talk about some of those resources that exist in Calgary for people who are looking for a more structured approach to educate themselves in this field. But even if you're looking for a more structured approach, I think it's great to have that natural curiosity and just have the belief that most of the people who are in the space have had to teach themselves to some varying degree, often quite a bit, and you can do it. You can just go through and find the information, and it'll be overwhelming at first, and you'll be like, holy cow, like, I don't remember linear algebra, what are they talking about, matrix multiplication, but eventually, you, you get it, and you get it.

[00:27:54] Ben: We were joking before that, um, she was going to call the podcast the Data Quagmire and like, oh, yeah, of course, just getting lost in the mucky swamp. You're up to your knees. You can't get out here. Well, that's what I kind of feel it's like.

[00:28:08] Lora: For anybody that wants to get into data. Like, where do I start?

[00:28:11] Lora: How do I find the data to start with? How do I find context? How do I find projects? How do I find a mentor? Because you're not already working for someone. Where you could leverage that, I think it's really, really hard to figure out how to, how to get started.

[00:28:25] Ben: Yeah, and my general advice that I give to people is to start on a project.

[00:28:32] Ben: There are lots of datasets, like we're living in maybe the golden age of datasets, the golden age of free datasets, where You know, we haven't really locked a ton of them down yet. There's a lot that are locked down, but there's a lot that are available and open. There's places like Kaggle, which has its pros and cons being able to solve.

[00:28:54] Ben: And Kaggle, for those who don't know, is a place where you can work on a data set, and you build a machine learning model, and then everybody competes to have the best machine learning model. And whoever has the best performing machine learning model that, you know, can count the number of galaxies in an image, they win a cash prize.

[00:29:12] Ben: It's a very good place to gain some basics around the field. The real world doesn't look like Kaggle competitions where you have a really clean data set and you need to be the best performer if you're 97% accurate versus 96% accurate, but that is a very good place to go and start and it's a good community to see what people are doing.

[00:29:33] Ben: But yeah, just finding a data set and you know, starting off with, oh, hey, you want to analyze the, the distribution of where people in the NHL are taking shots from and what the scoring probability is based off of where you're shooting from, from the blue line or wherever. So, getting started with something, building a project is a great way to just get your feet wet.

[00:29:55] Ben: But yeah, if we want to talk about some of the other kind of structured places, we can, we can definitely do that. I've got some, some recommendations.

[00:30:02] Lora: Yeah, that would be great for a resources list.

[00:30:06] Zach: Ben, you know at Careers in Tech, we have a lot of people getting into data. And one of the conversations is how do I decide between being a data engineer, data analyst, data scientist? How do I figure out what's the best fit and how transferable are the skill sets that if I wanted to change later in my career, how possible is that?

[00:30:24] Ben: Yeah, that's a good question. I'm going to speak specifically to the local Calgary or maybe local Alberta economy because it is different in different geographic regions. Based off of just how mature that Calgary is in the process of maturing, which means that there are a lot of opportunities for you to join a firm and not have the company know the difference between a data scientist and a data analyst and a data engineer, and you just have to do it all.

[00:30:58] Ben: And honestly, I like being in that spot. Our firm's 26 now. We have real full time data scientists that just do data science. I still do everything, and I think that's a good spot to be

in. So I'm a little bit biased with like, Oh yeah, no, just like go somewhere where you have to do the job of five people instead of one person's job.

[00:31:16] Ben: It's a great place. Great way to learn. Yeah, not, not stressful at all. Not stressful at all. But yeah, so I would kind of go back to if you're trying to figure out, trying to decide, I would go back to those three pillars, and we can just think a little bit more about what it means for each one on the data analyst side that's closest to the business.

[00:31:35] Ben: That's probably if you're really interested in, you know, health care and actually like the health care industry really gets you going and you want to just make better health care decisions, and you don't really care so much about the fancy maths and the statistics and the linear algebra that data analyst space might be better where you're going to be making reports kind of like using your human neural network to examine the data and come up with interesting insights, creating charts and graphs and visualizations.

[00:32:07] Lora: In my experience, storytelling has been key skill relative to data, because if you can't really kind of convey and put it in context, doesn't give you the bigger picture.

[00:32:15] Ben: Yeah, and that's probably if I could take, roll back the tape here by about 15 minutes and talk about the skills that are required for a data professional. Yeah, storytelling, like caring about the story and being able to communicate that, I would put it just at the top of the list as well for the data scientists that are kind of, I guess, to transition that into the data science or more of that, like predictive, trying to create a good automated decisions. That storytelling actually is important as well, because generally you need to have a story about how the data is being generated or what the underlying theory is.

[00:32:53] Ben: That's why they're called the data scientists. You think about the scientific process, you have a theory, you have a hypothesis, you gather empirical information about what happened, and then you use that to adjust or confirm your hypothesis. So, you need to have a story about how the data is being generated, the data generation process, and then you're trying to build a model that fits with that story.

[00:33:15] Ben: And if your model and your story fit well together, then you're going to have a successful product. So, on the skills for the data scientist side of things, that's going to be people that enjoy the STEM side of things more. You don't need to be an expert mathematician, but you should be at least willing, you should be able to read math, um, you should be comfortable with it, you should have some basics in your linear algebra, you don't need to be an expert, but you need to understand the tools that you're working with so that you don't misapply them, and that's going to be people that are really interested in it.

[00:33:53] Ben: Yeah, making kind of those repeatable decisions, interested in building statistical models, really liking to see that. Oh, hey, I can predict if somebody gives me just going back to the computer vision space, right? If somebody gives me a picture of a dog, I'm able to identify what breed of dog it is and something like that is exciting.

[00:34:12] Ben: That's going to be more on the data science space. And then And people who

care about probability, probability, chance, those things. And then the data engineering, that's going to be if you've got that mindset, or you've got the personality trait, where you really like processes. that work correctly, that are repeatable.

[00:34:31] Ben: You really like to build the engineering solutions. You like to build something up that can work at scale, like watching gigabytes of data flow through a real time streaming algorithm, and you're spinning up cloud servers and watching all of this information get processed, stored, and collated, and organized.

[00:34:55] Ben: And so the data engineering is going to be the closest to software. So, if you like building systems, that problem solving, then I would recommend, yeah, take a look at the data engineering side of things. It's very topical right now.

[00:35:10] Zach: So, I feel like we have to ask this question, how is generative AI going to impact these roles in the future, you think?

[00:35:14] Ben: Yeah, so I currently, use a fair amount of generative AI to write code. I think the good way to understand generative AI is it produces the most common answer to the most common question. So, and it that's very different from providing the best answer to a very specific question. So, a lot of the common use cases, because in any project where you're trying to convert data into decisions there's a lot of repetitive things that you do in every single project where honestly just doing the average thing really well is great. And so that's where generative AI is going to come in and, you know, say, hey, here's a good way to lay out your project. Here's a bunch of the boilerplate code.

[00:36:03] Ben: These are the current best practices of what people are using to design their data pipelines. Here's some of the best practices for setting up a training machine learning model or whatever. So you'll get a lot of the baseline foundations can be Leveraging generative AI to help you do that is great.

[00:36:22] Ben: And then you get into the specific nuances of the exact problem that you're working on, where there's not a hundred exabytes of internet information that's already been generated. around that specific topic. And you might be the first person doing it or doing it in a different way. And that's where your human neural network, your human generative model needs to come in.

[00:36:46] Ben: So yeah, the efficiency and efficacy generative AI will improve the efficacy of everybody who's working on this in order to double or one point or 50% like increases between, you know, 10 and 100% probably in overall efficiency, and then also helping people learn faster, quickly. I mean, a lot of generative AI is a great search engine, despite the fact that it hallucinates.

[00:37:13] Ben: It's still a great search engine, and you can use it to extract information and really get up to speed quickly on a topic. Oh, hey, Apache Airflow 2.0 just came out. What are the major changes associated with that? And being able to kind of distill some of that information down quickly will be quite beneficial.

[00:37:30] Ben: As well. And of course, it's generated a ton of hype and excitement in the industry. There's debates on whether we got like a stock boom just from AI, like any company that mentions AI enough times gets a boost to their stock price. So that excitement around the AI world and what generative AI has done for AI.

[00:37:54] Ben: The understanding or the familiarity with AI in the, you know, general people, um, the common parlance has been very accretive to the industry because now everyone wants it. So you're seeing a lot more opportunities, um, a lot of companies that maybe wouldn't have considered it before. Asked ChatGPT to write them a poem for their dog's birthday.

[00:38:16] Ben: Got something amazing out there, like holy cow, this is magic, I need to get in on this. And so it's also creating a lot more opportunities. So yeah, I guess I'm, to summarize, I see it as Improving efficiency, allowing people to quickly understand and get the best practices of the current standard practices together and built out into their systems, and then creating a lot more demand for these types of roles. So yeah, I see it as highly accretive to the industry.

[00:38:44] Lora: On the theme of future projecting, there's been a lot of talk in Alberta recently around energy expansion. So, I'm curious because of your background, and what your thoughts are about the role of energy in the future.

[00:38:55] Ben: It's been a while since I've been in deeply in the energy industry so I'm trying to remember the term for like tier two emissions, tier three emissions. Is that what they? There's a, the government of Canada has, like, for emissions tracking, there's different tiers or phases of, um, emissions tracking, and Some of them are like, okay, you want to figure out what the emissions are for the entire life cycle of a product or the emissions of a product over, not just it's, it's life cycle, but if you're using a drill press to punch a hole, what is the emission associated with even like that hole being punched in the manufacturing process.

[00:39:39] Ben: So if you want to Understand emissions. Just think about how hard that question is to know what are the for every single component in a brand new vehicle. What are the emissions that go into gathering the source materials, putting it together, assembly, manufacturing, shipping, the use of that decommissioning of that car over time.

[00:40:02] Ben: That is an incredibly complex problem, and the only way to solve it is with Data and with having data systems from lots of different companies and organizations that talk to each other, that obey the same principles, the same schemas that follow the same patterns. And so from a emissions tracking standpoint, it is absolutely critical.

[00:40:24] Ben: There's no way to do it without without these tools and technologies. And then from a actual emissions reduction. There's a lot of benefit to call it more of those predictive analytics, like true data science and helping us to better understand ways to call it reduce emissions or improve efficiency of energy.

[00:40:44] Ben: Alternative energy producing assets. There's lots of think about, wind turbines, right? And predictive maintenance on wind turbines are knowing how much like how much of the break to apply during a given windstorm or what are the probability of gusts coming through that might decommission your wind turbine.

[00:41:01] Ben: There's so much complexity that goes into that. And if you want to seek out every bit of additional power that you could, if you had all the information, if you made the optimal decisions about how to run that wind turbine farm, you can generate more green, clean energy. And so that's another case where that data science, data analytics, data engineering is going to be a catalyst that helps us actually improve our ability to both create like both have green energy, a meaningful part of the grid as well as tracking and reducing emissions in conventional sources.

[00:41:40] Lora: Can you tell us to kind of do a little bit of a pivot, but a little bit about the community building you've been doing?

[00:41:45] Ben: Yeah, so, you know, maybe kind of like as I was telling my story, I neglected to give credit to how important the local data community was in my journey and in my transition, because I still remember the first community meetup that I went to with, uh, with a friend and we had to use some deep learning tools to try and guess how old somebody was based off of a picture of their face.

[00:42:17] Ben: And just like how much I wasn't expecting such a hard problem to have to solve. It was like my first meetup that I went to, and you know, the nights that my friend and I spent, like he would bring, because I didn't have a computer with a GPU, so he would bring his computer over to my house, and we would hack away, trying to train a machine learning model.

[00:42:36] Ben: Through this thing, that aspect of community was so important to me getting started in my journey and transition. And so, I've been involved with the data community. I went from an attendee of that meetup to one of the organizers and then call it seven years ago or six years ago or so, there was maybe eight different data related meetups as I was involved with one of them, but none of us were talking to each other.

[00:43:03] Ben: None of us were communicating. We might communicate within an organization or within a community, but not across communities. We would schedule meetups at the same time as each other. And so that's whenever some of the community leaders said, okay, well, let's have a meetup of meetup organizers, right? And let's all get together in a room and figure out, okay, how can we work together?

[00:43:23] Ben: How can we collaborate? How can we actually start to build a data related ecosystem within Calgary so that we're all rowing in the same direction so that we actually have these network effects that make great ecosystems are built off of network effects. And so how can we do that for the data ecosystem?

[00:43:40] Ben: So that was the start of the YYC Data Society. So as that matured and we had more and more ambitious ideas like running a conference and creating mentorship programs and

all that stuff, as our ambitions grew, we're like, okay, We got to form an official nonprofit society, we need a bank account, we got to take sponsorship dollars, et cetera.

[00:44:02] Ben: And so that's where the, you know, really for our most ambitious project, which was the YYC DataCon, which is an annual conference we've run for three years now. This past year in March, we brought about 600 people together over two days, 40 speakers. Yeah, it was a pretty huge event. And so... Whenever we started to do the Data Con, that's whenever we really institutionalized as a community.

[00:44:29] Ben: And so the mission for the Data Society is to, you know, first and foremost, support all of the other grassroots communities. We are grassroots and we believe in the importance of grassroots communities for good, strong, healthy ecosystems. Because for me, It wasn't the government of Canada pumping a hundred million dollars into a venture capital fund that's going to invest in Calgary startups that led me into the data industry.

[00:44:57] Ben: It was someone saying, okay, here's a project for you to work on, go and predict the age of somebody with a face and then working on that with other people in the community that allowed me to move in. And so that kind of, that story, that experience that I had has made me want to be able to share that and really focus on that grassroots aspect so that the data society and all of our other communities like Data for Good, like Untapped Energy, like Woman in Data, like Woman in AI, we can all share work together to create this ecosystem that helps to not just bring new people into the industry, but also accelerate their journey once they're there, so they can have those peers, that they can have that support network, so they can find mentors, and so that we can also provide information about what types of programs are available, who's hiring, who you might want to go and talk to.

[00:45:54] Ben: So, yeah, that's um, we think that it's incredibly important for that space to exist in order to foster a healthy ecosystem. There's more and more excitement every day about the space for, for data and AI and in Calgary, and we're just happy to, happy to support those looking to transition or looking to start a community, it's great.

[00:46:14] Zach: Yeah, Ben, thanks so much for being here today.

[00:46:17] Ben: Yeah, you guys are welcome. It's been a ton of fun.

[00:46:20] Zach: We really enjoyed our conversation with Ben. Thank you for listening to our listeners and to Lora, not only for being part of the podcast, but for the work you're doing. at SAIT and helping people evolve their careers in the careers that you didn't know that existed.

[00:46:35] Lora: Well, Zach, and thank you for all the work that you do in the community. It was really fun talking to Ben as a community builder to hear about all the things that he's doing, building community around data. It is, you know, it really is a confusing, complex area. The roles are confusing and complex. And I really appreciated his Lens on what the roles are, the different categories of roles and the skills needed in each of those areas.

[00:47:05] Zach: Likewise, thanks for joining us and make sure to check out our website for the

resources that Ben will share.

[00:47:10] ANNCR: The Best Careers You Never Knew Existed Podcast, sparked by SAIT and CITI, funded by the government of Alberta. Have a career suggestion or want to appear as a guest? Get in touch- [SAIT.ca/careerspodcast](http://SAIT.ca/careerspodcast).

[00:47:27] ANNCR: Rate and review this podcast and you might find your review on a future episode. Please subscribe to The Best Careers You Never Knew Existed, wherever fine podcasts are downloaded. With Lora Bucsis and Zach Novak. Produced by Terran Anthony Allen and Jenna Smith. Executive produced by Lora Bucsis. Voice over by me dun dun dun. Alright. Special thanks to SAIT Radio for their support and the use of their studios. And most of all, thank you for listening.