# Actuaries in Data Science with Genevieve Hayes

## Interview Transcript

**Julia Lessing:** Hi, everyone. Today we're talking about opportunities for actuaries in data science and to have that discussion. I'm delighted to be joined by our special guest today, Dr. Genevieve Hayes. Genevieve is the Director of Genevieve Hayes Consulting. And not only is she a fully qualified actuary, she has a Master's in computer science and a Ph.D. in statistics. She's the host of podcast Value Driven Data Science, and she has a wide range of experience in technical, leadership and academic roles. Genevieve, thank you so much for joining us today.

**Genevieve Hayes:** Thanks for inviting me on the show, Julia.

**Julia Lessing:** So, I thought we might start there's this sort of age old debate around actuaries versus data scientists. And are actuaries data scientists, could data scientists be actuaries? I know we've talked about this a little bit about how, you know, there are some clear differences, but also some overlap. And I wondered if you could start by telling us a little bit about the data science workforce.

I know that's your special area now, even though you are also a qualified actuary. What does that universe look like and how do actuaries fit into that?

**Genevieve Hayes:** Well, before I answer that, what exactly do you mean by the data science workforce?

**Julia Lessing:** Thank you, Genevieve. There was a recent report from Deloitte around the data science workforce. We talk a lot about data science being a very trendy and popular and fast-growing workforce. As technology improves and as the volume of data that we have at our fingertips is increasing. And so, in that report, they talk about the data science workforce having a very small proportion being actuaries or statisticians, analytical professionals similar to actuaries, I suppose, but also a lot of other different roles within there too.

So, software developers and software engineers and I wondered if you could tell us a little bit about your experience about that diversity of skillset within the workforce, that data science workforce, because it's more than just actuaries or analytical professionals, isn't it?

**Genevieve Hayes:** Yeah, well, data scientists are data scientists, so there is diversity among their roles, but there are other roles that they work with in order to achieve their job and their task. And I think that's what you're asking here.

**Julia Lessing:** Yes.

**Genevieve Hayes:** So, the range of data roles you're going to encounter in any particular organization that's going to vary depending on the size and level of data maturity of that organization. And you'll find with some smaller organizations, you've got a data scientist who also has to be doing three or four other roles in order to get their job done. It's very common for data scientists to also have to do their own data engineering in less mature organizations.

But if you have an organization that has a reasonable level of data maturity in an organisation that's big enough to support a data science team that's got more than one person in it, you're going to get a variety of roles. And those are some of the roles that I've encountered in a previous job that I had where I think we had 6000 employees in total, although most of them were not data professionals.

And we had a pretty good level of data maturity. So, we had our I.T. team. So, this is your standard corporate I.T. who provide us with the core infrastructure. So those are the people that you find when your computer completely collapses.

**Julia Lessing:** Yes. They're my best friends when my computer collapses.

**Genevieve Hayes:** Yeah.

**Julia Lessing:** Not my strength at all.

**Genevieve Hayes:** They provided us access to the tools and the platforms we needed to do our job. Then we had a team of data engineers and they developed the pipelines. We needed to ingest data into these platforms and into our databases and to get that data into a form that was suitable for further analysis and modelling. So that was just getting it into a form that wasn't complete garbage.

It was getting it into a database. We had data analysts who were there to build business intelligence type dashboards. So, our audience would probably be most familiar with Power BI, those sorts of dashboards, except we were using different tools, but same principle. And we had our data scientists who would build more advanced data products. Then we had software engineers and software developers to develop software applications that could be used to gather more data.

So, for example, a front end to a database that members of staff of this organization could enter data into. We had cybersecurity people to make sure everything was safe from hackers. And as I was leaving, we were in the process of hiring a data architect. So, they designed the architecture of future databases.

**Julia Lessing:** Right. So, it sounds like in that situation, you actually had quite a wide variety of different professionals. So as the data scientist, you might have been processing the data and deriving insights and using your technical data science skills to look at that data and to derive insights for the business. But to do that, you still needed to have I.T. people and you needed to have data engineers and you needed to have data analysts who could do other parts of that process so that together you could work and do the data science and the insights that you know and provide those insights that you needed.

**Genevieve Hayes:** Yes, that's right.

**Julia Lessing:** And you were saying that that was in an organization where you had a relatively high level of data maturity. But if you are in maybe a smaller organization or a less mature data environment, you might find that you're doing some of those multiple roles yourself. Is that right?

**Genevieve Hayes:** So I've spoken to people who are working in those sorts of organizations where they are effectively doing their own data engineering. I would be surprised if you had a data scientist who was also doing something like cybersecurity, because most organizations understand you do not want to get that wrong.

**Julia Lessing:** It's very specialized, isn't it? That cybersecurity space.

**Genevieve Hayes:** Oh Yeah. Because we hadn't hired a data architect at that point in time, I was actually doing a lot of the data architecture work in that organization. So that's something that you could end up being having to do. I doubt anyone would have to do their own I.T. infrastructure because usually organizations do have some sort of I.T.

**Julia Lessing:** But it's interesting that when you think about data scientists or data science as a role, that it's not just saying, well, as actuaries, if we're wanting to move into this space, you know, maybe we move from Excel and we're doing a little bit more programing. And all of a sudden where we're data scientists. Actually working in that data science space is a lot broader than just the analysis piece.

It's about having the tools and the support in the I.T., the data engineering and the software behind it as well.

**Genevieve Hayes:** And it's also about being able to speak to those other people. So, because I have a computer science background, it made it very easy for me to speak to the software engineers and the data engineers. But if you didn't have that same background, it might be a lot harder to integrate into that sort of environment.

**Julia Lessing:** It is interesting when we're talking to people from different disciplines, and we all sort of speak a slightly different language, don't we? So, actuaries talking to actuaries.

**Genevieve Hayes:** If you're doing programing, you might literally be speaking a different language.

**Julia Lessing:** Okay, so you've talked a bit about what a data science workforce might look like in a business, but data science is also an increasing area of interest within academia as well. And I know you've done a lot of work in the teaching space. Really curious if you can tell us a little bit about how roles for data scientists might differ between academia and in that business context?

**Genevieve Hayes:** Well, so I work in the teaching space at the University of Melbourne. So I will be teaching the data science principles course in second semester. And so that's one role that data scientists can have in the teaching in the academic space, just teaching data science to students. But I think what you're getting at in the academic side of things, is data science researchers.

So these are the people who are developing the new data science algorithms and their goal is basically to produce cutting edge algorithms. There are usually several tasks that academics in data science are trying to optimize. So, for example, how can I create a model that will make a better chat bot so something like a ChatGPT type chat bot.

And there are certain test datasets that these academics will be trying to build their models using so that they can benchmark their models against other academics. Mm hmm. And that's one particular discipline. And people who specialize in that area, often they're picked up by big organizations like your Googles and Microsofts to build things like ChatGPT.

**Julia Lessing:** Right. Because it's so cutting edge. So they're building, you know, new tools and new things right at the forefront.

**Genevieve Hayes:** Yeah. I remember writing an article saying that all of the AI academics in America, the best ones, had been poached by the big tech companies to build their models. And yeah, because the big tech companies could pay them so much more than the university of wherever. And that's great. I mean, if it wasn't for those people, we wouldn't have ChatGPT.

But in many ways, that sort of data science in academia is a lot like a Kaggle competition. So you've got a nice, clean dataset that isn't going to change. All you care about is getting the most performant algorithm. You don't care about having to integrate it into a production system. You don't have to worry about ethics because this data has been around for 20 years, and it's…

**Julia Lessing:** It's open source, and everyone's happy for you to use it in whatever way you want.

**Genevieve Hayes:** Yeah, it's not someone's personal banking records or something like that. Whereas in the real world, you have to start dealing with those things. Your data is constantly changing. It's messy. You've got to worry about ethics. If, depending on your application, you might have to be able to explain how your model works to someone who doesn't have a technical background.

And you're going to have to integrate it into some sort of IT system. This is the reason why many organizations are still making use of linear regression models, even though they've been around since the 19th century, I think. But they're easy to explain. Pretty much everyone understands them. They're fast, they're reliable, and they're really easy to deploy.

So people go with them. That's why actuaries still use GLMs in insurance. So it's like there are two different data science disciplines, but a lot of people, thanks to things like Kaggle, think that data science is a lot like that sort of academic data science discipline when it's not.

**Julia Lessing:** Right. So it sounds like it's actually quite different and quite different considerations when you're doing the work. So if you're working in academic data science research and really improving those tools, you're focusing on sharpening the tools without having to worry too much about the context of data changing and implementation considerations and privacy and all of that. Stuff that comes with looking after data and working with data that probably most actuaries in a business context are quite used to.

So, you know, if you're in an actuarial team, you're looking at policyholder information or client information. So some of those considerations are probably quite familiar for actuaries. But as you said, that's probably why actuaries are leaning on some of those older, more tried and true sorts of methods that are very well known, very easy to explain, but less cutting edge as well.

**Genevieve Hayes:** Yeah, exactly. Yeah.

**Julia Lessing:** And I like your analogy of the Kaggle competition, because I think that that is often sort of something that actuaries think about when they're thinking, what would data science be like? What would a career in data science be like? And it sounds like perhaps if you're going into a business data science role, it wouldn't really be like that.

**Genevieve Hayes:** Yeah. And if you look at some of those Kaggle competitions, the winning entry, they are so ridiculously complicated. I've heard of situations where you've got three or four leading teams who all come up with some horribly complicated model, and then they all group together to form an ensemble learner, of those three or four complicated models in order to get it even more performant.

There is no way on earth anyone would use that in real life.

**Julia Lessing:** So more accurate or faster or more performant, as you say, but not necessarily as practically implementable or explainable from a practical business context.

**Genevieve Hayes:** Yeah, exactly.

**Julia Lessing:** Yeah. So does that mean that when you're when you're teaching at university, what happens when you sort of got that overlap? Because you said you do have some teaching roles as well and you're teaching data scientists. How do you sort of bridge the difference between those two sort of applications of data science?

**Genevieve Hayes:** Well, the way I'm framing my course is looking at it as model driven data science versus data driven data science. So start with the models and look at what makes for a effective model. So the Kaggle approach. But then take a step back and look at, okay, the Kaggle approach is one way of looking at data science, but you also have to, in the real world, take into account the data and then that brings in a whole extra number of complications.

And I think that's a good way of looking at that. Actuaries, data scientists, etc. can look at things.

**Julia Lessing:** So really having an understanding of both, really being able to apply some of those cutting edge techniques, having a really good understanding of the different tools that are available and different techniques that we can use. But also being able to really just think about the context as well.

**Genevieve Hayes:** Yeah, exactly.

**Julia Lessing:** So, Genevieve, you span actuarial and data science and you've got a Master's in computer science. Can you tell us a little bit about the skills that actuaries who've come up just through that actuarial professional qualification, what sort of skills do they need to add to their toolkits to be effective data scientists?

**Genevieve Hayes:** Well, before we started recording this, we're talking about the Venn diagram that describes data scientists. And for any of your listeners who isn't familiar, who aren't familiar with it, it's a Venn diagram that shows the core skills needed to be a data scientist. So you've got your domain knowledge, your maths and science skills, or math and statistics skills and your computer science or programing skills.

Now, actuaries are very good at their domain knowledge in the financial services type disciplines. They might not be so good at domain knowledge in other disciplines. So if they want to move outside the financial services industry, that would be something they would need to develop. But let's say that they do have the relevant domain knowledge or that the employer that's hiring them is willing to train them in that.

That's then all you have to do is focus on your maths and your computer science. With data science roles, my experience has been they sort of sit on a spectrum between ones that are very statistically skewed, and ones that are more skewed towards computer science. So the statistically focused ones, they tend to be things like reporting and bespoke statistical analysis and those sorts of roles align very well with the actuarial skill set because actuaries receive a lot of training in statistics during their core actuarial training, and you probably don't have to do too much extra work in order to get to a point where you're effective at doing that sort of statistical analysis.

So if you did want to do that, you'd just brush up on your statistics a bit. At the other end of the spectrum, you have the data scientists who are more focused on the computer science side of things. So with these data scientists, they're often fitting machine learning models, and they're also often working very closely with software developers.

And a lot of people who go into that side of data science did originally train as software engineers or software developers. So if you're wanting to go into that side of things, you would have to spend a lot of time building your computer science or software development skills. And that was an area that I wanted to go in.

And that was part of the motivation for me doing my Master's in computer science. So for actuaries who don't have those computer science skills, it would be harder to succeed in that side. It could be done, but they would need to build up those skills.

**Julia Lessing:** So it sounds like most actuaries it would be quite a transition to go into that sort of statistical end of data science. But in order to get to do some of the more computer science side of things is a bit more training to happen. That it would be required. And so, Genevieve, you said that was one of the things that you wanted to do that you wanted to move into that computer science side. And that's what prompted your Master's. What was it that attracted you to that kind of data science work?

**Genevieve Hayes:** Well, in a role I had about ten years ago, I was I was actually leading, I guess it was called the pricing and analytics team at that time. And my team did insurance pricing, which is very statistical, as you'd know. And we also did a lot of the what I would now call data analysis work for the area of the business I was in.

And we were a SAS programing team. And I really loved just that whole statistical SAS programing side. And then I learned there was another side of data science. And I thought, I want that. I want to add that to my skill set. I really loved the SAS programing, so I wanted to be able to build more programing skills.

And when I'd become an actuary, when I signed up to do the actuarial degree out of high school. I was also tossing up becoming an engineer. So it was something that brought me closer to that engineering skill set, although I couldn't do what a lot of these engineers that I know can do. But it allowed me to learn a lot of things that I would have liked to have learned at university had I made a different choice.

**Julia Lessing:** So it was a way to kind of bring back that one of those original options that you had coming out of school and add that to your toolkit. And what a big data toolkit that you have generated with your actuarial qualification and also your computer science Master's and your Ph.D. in statistics.

So for actuaries who are maybe qualified but feeling like they don't want to take on a Master's or a Ph.D. on top of their existing study, but wanting to move into more of a data science role, what would you suggest that they do?

**Genevieve Hayes:** I think a good starting point would be to learn some basic programing skills. So I've worked with actuaries who are doing a lot of their work in Excel, and that's fine. But if you're going to go into the data science side of things, the first thing you need to do is be able to program. So just start by doing even just a free online course on how to program in Python.

And that's not going to make you a data scientist, but it's at least giving you a taste of what's involved with programing. And then start trying to use those skills to create some projects. So that was how I learned a lot of these things. I did the things that I learned at university, and then I'd start creating my own projects that would allow me to experiment with those skills.

**Julia Lessing:** How do you mean? You created your own projects? Can you give us an example?

**Genevieve Hayes:** Well, one of the things for my masters, one of the projects we had involved using this, well, we had to implement these various algorithms and do an analysis of the output and the only package that existed to implement some of these algorithms was the Java package. And I remember thinking I remember I had the choice between either learning how to programing Java or trying to rewrite these algorithms in Python.

And I did a very quick job of rewriting them in Python. I figured it was fast learning Java this assignment and I came up with a good enough version of these algorithms in order to do the assignment. I passed the assignment, obviously. But after I finished that, I thought, Well, this is something that a lot of people could benefit from because I think we had about a thousand people doing this course per semester.

So I took those algorithms and created a Python package that is now available on PyPI. And that allowed me to learn a lot about software development. And so that's that was a project that just sprang up out of it was something that I needed and I knew that other students would need. And it produced something that taught me a lot.

And I think that's a good way of coming up with projects, just thinking about what is something that you need. And then build something that will solve whatever your problem is.

**Julia Lessing:** And having that real example, that real life example that you've just described, where you've been asked to do something, and you can see an opportunity to do it a better way or a different way, but also to use it as an opportunity to build your skills. That can be a much more effective way of getting things done than to say, Oh, I'm going to sign up to Data Camp and do all of the lessons.

I'm going to, you know, and sort of just have that theoretical learning without anything practical to apply it to be all the exercises within the course. So having a real example like that, I can see, would have been a really effective way to, to not only build skills, but to also build your profile by the sounds of it and to be able to share that with others as well.

**Genevieve Hayes:** Yeah. And with a lot of the datasets you encounter in courses such as the Data Camp one, they're just designed for educational purposes. I use these datasets in my teaching because they're very easy to use to teach a basic concept, and there's nothing wrong with that. But they're not teaching you how to deal with large datasets, messy datasets, complications, and they don't actually solve any problems that you would really want to solve in the real world.

**Julia Lessing:** Looking at the heights of basketball players!

**Genevieve Hayes:** Classifying iris flowers into species of iris!

**Julia Lessing:** Yeah, lots of simple illustrative examples, but not necessarily with the same sort of impact that some of our more practical data science projects can deliver.

**Genevieve Hayes:** Yeah.

**Julia Lessing:** So, Genevieve, you've talked to us about the Venn diagram of actuaries and data scientists, but also that workforce of data science and how that can be different in different organizations, how it's not just about having data scientists within that team, but the data scientists are also supported by I.T. professionals and data analysts and software developers and software engineers, and a range of different people who can make sure that the data is available and protected and safe in a business context and ready to do your analysis on and your data science on to derive some insights.

You've talked about how applying data science in academia can be quite different because you're really wanting to refine the tools and the techniques and optimize some of those algorithms and improve some of those algorithms rather than focusing on dealing with the practicalities of messy data and real-world data that has a whole lot of other considerations as well.

But that data science in a business context is possibly a bit more aligned to what most actuaries are doing in a practical day to day professional capacity. Because we're thinking about everything that goes with the data and making sure that we're addressing the gaps in data or we're addressing messy data or we're making sure that the data is protected and privacy is protected, and dealing with the practical implementation of using data in a business context.

And you've talked about how there's a different way of building skills for actuaries who are wanting to move into a data science space. So I guess I'd just like to finish off with a final question and just to ask your top tip for actuaries who are listening, actuaries who may be there in a traditional area, maybe they're maybe they are starting to build their Python skills, maybe they are just wondering what else is out there for them in terms of their career and thinking about how they could apply their skills in a different area. What's your top tip for those actuaries who are considering a move into data science?

**Genevieve Hayes:** Well, don't try and be everything everywhere, all at once. I think that's a big mistake. A lot of people make, not just actuaries in that they want to be everything and you can't be you just find an area to specialize in. Find a domain that you want to specialize in. And for actuaries, the financial services domain is probably a very good choice.

And then within that, find a toolkit of skills to focus on learning to begin with. So it might be get your statistics really, really good, or focus on getting really, really good at applying machine learning to a particular type of problem. So that might be, well, pricing insurance contracts, for example, but just focus on getting really good at one or two things at once rather than trying to learn a little bit about a lot of stuff because if you're trying to be everything, then there's always going to be someone who's better than you at pretty much everything.

**Julia Lessing:** That's great advice, Genevieve. So pick a domain, pick a toolbox, to really improve and focus on enhancing and. Go deep.

**Genevieve Hayes:** Yeah.

**Julia Lessing:** Excellent. Such great advice. Thank you, Genevieve, for joining us for this conversation today. I'm sure that your advice and experience and reflections will be very helpful for a lot of actuaries listening. So thank you for your time.

**Genevieve Hayes:** Thank you for having me.