0:00

So far we've covered the different roles data analysts play in business environments, and the kinds of tasks that come with those roles. But data analysts have another important responsibility, making sure their analyses are fair. Now I know what you're probably thinking, data is based on collected facts. How can it be unfair? Well, that's a good question. So let's learn what fairness means when we talk about data analysis and why it's important for you as an analyst to keep in mind. Fairness means ensuring that your analysis doesn't create or reinforce bias. In other words, as a data analyst, you want to help create systems that are fair and inclusive to everyone. Sounds simple enough? Well, here's the tough part about fairness in data analytics. There isn't one standard definition of it. But hopefully the way we've just described it can give you one way to think about fairness for right now, but it's about to get a bit trickier. Sometimes conclusions based on data can be true and unfair. What can you do then? Well, let's find out with an example. Let's say we have a company that's notorious for being kind of a boys club. It's very male dominated, and there aren't many women employees. This company wants to see which employees are doing well so they start gathering data on employee performance and their own company culture. The data shows that women just aren't succeeding as often as men in their company. Their conclusion, that they should hire fewer women. After all, women are doing poorly here, right? But that's not a fair conclusion for a couple of reasons. First, it doesn't even consider all of the available data on company culture, so it paints an incomplete picture. Second, it doesn't think about the other surrounding factors that impact the data. Or in other words, the conclusion doesn't consider the difficulties women have trying to navigate a toxic work environment. If the company only looks at this conclusion, they won't acknowledge and address how harmful their culture is, and they won't understand why women are set up to fail within it. That's why it's important to keep fairness in mind when analyzing data. The conclusion that women aren't succeeding in this company is true, but it ignores the other systemic factors that are contributing to this problem. But don't worry, there's a way to make a fair conclusion here. An ethical data analyst could look at the data gathered, and conclude that the company culture is preventing women from succeeding, and the company needs to address these problems to boost performance. See how this conclusion paints a much more complete and fair picture? It recognizes the fact that women aren't doing as well in this company, and factors in why that could be instead of discriminating against women applicants in the future. As a data analyst, it's your responsibility to make sure your analysis is fair, and factors in the complicated social context that could create bias in your conclusions. It's important to think about fairness from the moment you start collecting data for a business task to the time you present your conclusions to your stakeholders. We'll learn more about bias in the data analysis process later on in another course. For now, let's check out an example of a data analysis that does a good job of considering fairness in its conclusion. A team of Harvard data scientists were developing a mobile platform to track patients at risk of cardiovascular disease in an area of the United States called the Stroke Belt. It's important to call out that there were a variety of reasons people living in this area might be more at risk. With that in mind, these data scientists recognize that fairness needed to be a priority for this project, so they built fairness into their models. The team took several fairness measures to make sure they were being as fair as possible when examining sensitive and potentially biased data. First, they teamed analysts with social scientists who could provide insights on human bias and the social context that created them. They also collected self reported data in a separate system to avoid the potential for racial bias, which might skew the results of their study and unfairly represent patients. And to make sure their sample population was representative, they oversampled non-dominant groups to ensure their model was including them. It's clear that the team made fairness a top priority every step of the way. This helped them collect data and create conclusions that didn't negatively impact the communities they were studying. Hopefully these examples have given you a better idea of what fairness means in data analysis. But we're going to keep building on your understanding of fairness throughout this program, and you'll get to practice with some activities.