Consider fairness

Previously, you learned that part of a data professional's responsibility is to make certain that their analysis is fair. **Fairness** means ensuring your analysis doesn't create or reinforce bias. This can be challenging, but if the analysis is not objective, the conclusions can be misleading and even harmful. In this reading, you're going to explore some best practices you can use to guide your work toward a more fair analysis!

Consider fairness

Following are some strategies that support fair analysis:

Best practice	Explanation	Example
Consider all of the available data	Part of your job as a data analyst is to determine what data is going to be useful for your analysis. Often there will be data that isn't relevant to what you're focusing on or doesn't seem to align with your expectations. But you can't just ignore it; it's critical to consider all of the available data so that your analysis reflects the truth and not just your own expectations.	A state's Department of Transportation is interested in measuring traffic patterns on holidays. At first, they only include metrics related to traffic volumes and the fact that the days are holidays. But the data team realizes they failed to consider how weather on these holidays might also affect traffic volumes. Considering this additional data helps them gain more complete insights.
Identify surrounding factors	As you'll learn throughout these courses, context is key for you and your stakeholders to understand the final conclusions of any analysis. Similar to considering all of the data, you also must understand surrounding factors that could influence the insights you're gaining.	A human resources department wants to better plan for employee vacation time in order to anticipate staffing needs. HR uses a list of national bank holidays as a key part of the data-gathering process. But they fail to consider important holidays that aren't on the bank calendar, which introduces bias against employees who celebrate them. It also gives HR less useful results because bank holidays may not necessarily apply to their actual employee population.

Include self-reported data	**Self-reporting** is a data collection technique where participants provide information about themselves. Self-reported data can be a great way to introduce fairness in your data collection process. People bring conscious and unconscious bias to their observations about the world, including about other people. Using self-reporting methods to collect data can help avoid these observer biases. Additionally, separating self-reported data from other data you collect provides important context to your conclusions!	A data analyst is working on a project for a brick-and-mortar retailer. Their goal is to learn more about their customer base. This data analyst knows they need to consider fairness when they collect data; they decide to create a survey so that customers can self-report information about themselves. By doing that, they avoid bias that might be introduced with other demographic data collection methods. For example, if they had sales associates report their observations about customers, they might introduce any unconscious bias the employees had to the data.
Use oversampling effectively	When collecting data about a population, it's important to be aware of the actual makeup of that population. Sometimes, oversampling can help you represent groups in that population that otherwise wouldn't be represented fairly. **Oversampling** is the process of increasing the sample size of nondominant groups in a population. This can help you better represent them and address imbalanced datasets.	A fitness company is releasing new digital content for users of their equipment. They are interested in designing content that appeals to different users, knowing that different people may interact with their equipment in different ways. For example, part of their user-base is age 70 or older. In order to represent these users, they oversample them in their data. That way, decisions they make about their fitness content will be more inclusive.

Think about fairness from beginning to end	To ensure that your analysis and final conclusions are fair, be sure to consider fairness from the earliest stages of a project to when you act on the data insights. This means that data collection, cleaning, processing, and analysis are all performed with fairness in mind.	A data team kicks off a project by including fairness measures in their data-collection process. These measures include oversampling their population and using self-reported data. However, they fail to inform stakeholders about these measures during the presentation. As a result, stakeholders leave with skewed understandings of the data. Learning from this experience, they add key information about fairness considerations to future stakeholder presentations.

Key takeaways

As a data professional, you will need to ensure you always consider fairness. This will allow you to avoid creating or reinforcing bias or accidentally drawing misleading conclusions. Using these best practices can help guide your analysis and make you a better data professional!

1. Which of the following statements accurately describe fairness considerations in data analysis? Select all that apply.

   - [x] Fairness practices should begin during the analyze phase of the data analysis process.
   - [ ] A data professional may include self-reported data when prioritizing fairness.
   - [x] Best practices for fairness in data analysis include identifying surrounding factors.
   - [x] Effective data analysts help create systems that are fair and inclusive to everyone.

2. A ballet company needs more dancers, so they announce upcoming auditions in online ads. Research reveals that 85% of dancers are under age 34, so the data team decides that young people are more likely to be successful. Therefore, they target the ads to young job seekers. What should they have done instead?

   - ( ) Only show ads for the upcoming auditions to people who are 35 and older.
   - (•) Conduct more research to understand the surrounding factors of this situation.
   - ( ) Ask the executive team to decide which type of applicants to target with the ads.
   - ( ) Find an additional data source that supports the strategy to target young people.

3. Which fairness best practice is intended to help data teams understand circumstances that could have an impact on their data insights?

   - ( ) Use oversampling
   - ( ) Include self-reported data
   - (•) Identify surrounding factors
   - ( ) Consider relevant data

4. A political candidate surveys voters to learn about their opinions on key issues. The politician's data team notices that most respondents are people who have already voted in the current election. The fairness of the survey could be improved by over-sampling people from which group?

- ○ First-time voters who already voted in the current election
- ○ People who have not yet voted in the election
- ◉ People who voted for the candidate in the current election
- ○ People who voted against the candidate in the current election

5. Fill in the blank: An executive might ask data team members to work on a _____ in order to address a particular question or problem using data analysis.

- ◉ business task
- ○ measurable outcome
- ○ relevant process
- ○ stated objective

6. A data analyst at a public library considers fairness when collecting data. Rather than asking librarians to share observations, they create a survey that asks library users to provide information about their own experiences. This helps avoid any unconscious bias that might be introduced by the librarians. Which fairness best practice does this scenario describe?

- ○ Using all available data
- ○ Self-reporting
- ○ Oversampling
- ◉ Considering context

7. A data analyst at a hospital system researches factors that contribute to low blood pressure. Rather than using only the medical records of patients with low blood pressure, the analyst considers patient demographics, medical history, lifestyle, and more. This enables them to achieve more insightful results. Which fairness best practice does this scenario describe?

- ○ Oversampling
- ○ Including self-reported data
- ◉ Considering all available data
- ○ Guiding business strategy

8. Fill in the blank: A data professional considers fairness from the start of a project to the point when their organization _____ in order to ensure their data analysis is fair.

- ○ presents findings to stakeholders
- ○ cleans and organizes the data
- ○ collects data for the project
- ◉ acts on the data insights