Detect the Bias

Biased data may lead to biased predictions, and AI models are only as fair as their training data. It's important to be aware of the possibility of bias, so that you can mitigate against it.

In this hands on lab, you will see how different datasets affect the predictions of an AI model. Rate the fairness of the outcomes, and see if other people agree. Then learn why reducing bias in AI models is so important.

Note: Detect the Bias is a live online demo, so you may see different models and datasets than those pictured below.

Follow these steps to explore the demo:

1. Access the demo here: Detect the Bias

2. On the **Detect the Bias** page, read the introductory text, and then click **Get Started**.

3. In the Consent message box, click **I agree**, and then click **Next**.

4. Page 1 of 3 describes the data the model uses, and how it might affect the defendant. Read the text, and then click **Next**.

5. Page 2 of 3 tells you that the data has now been preprocessed to make the decision more fair. Read the text, and then click **Next**.

6. Page 3 of 3 presents a rules-based model, and the outcome for a particular group. Read the page and then click **FAIR** or **UNFAIR** based on your evaluation of the model.

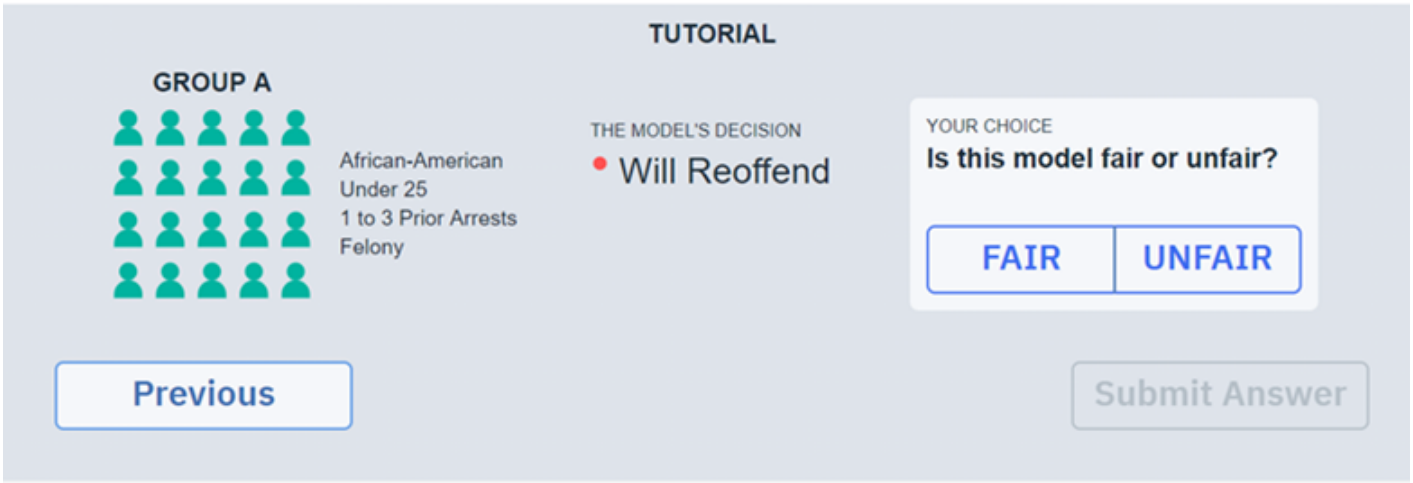
Example: This model used the **rules shown to the right** to predict that the **members of the group below will reoffend.**

Do you think this decision is fair? Make your choice below and submit your answer.

RULES CONSIDERED BY THE MODEL

If the crime is a Felony and Prior Arrests > 0,
Will Reoffend

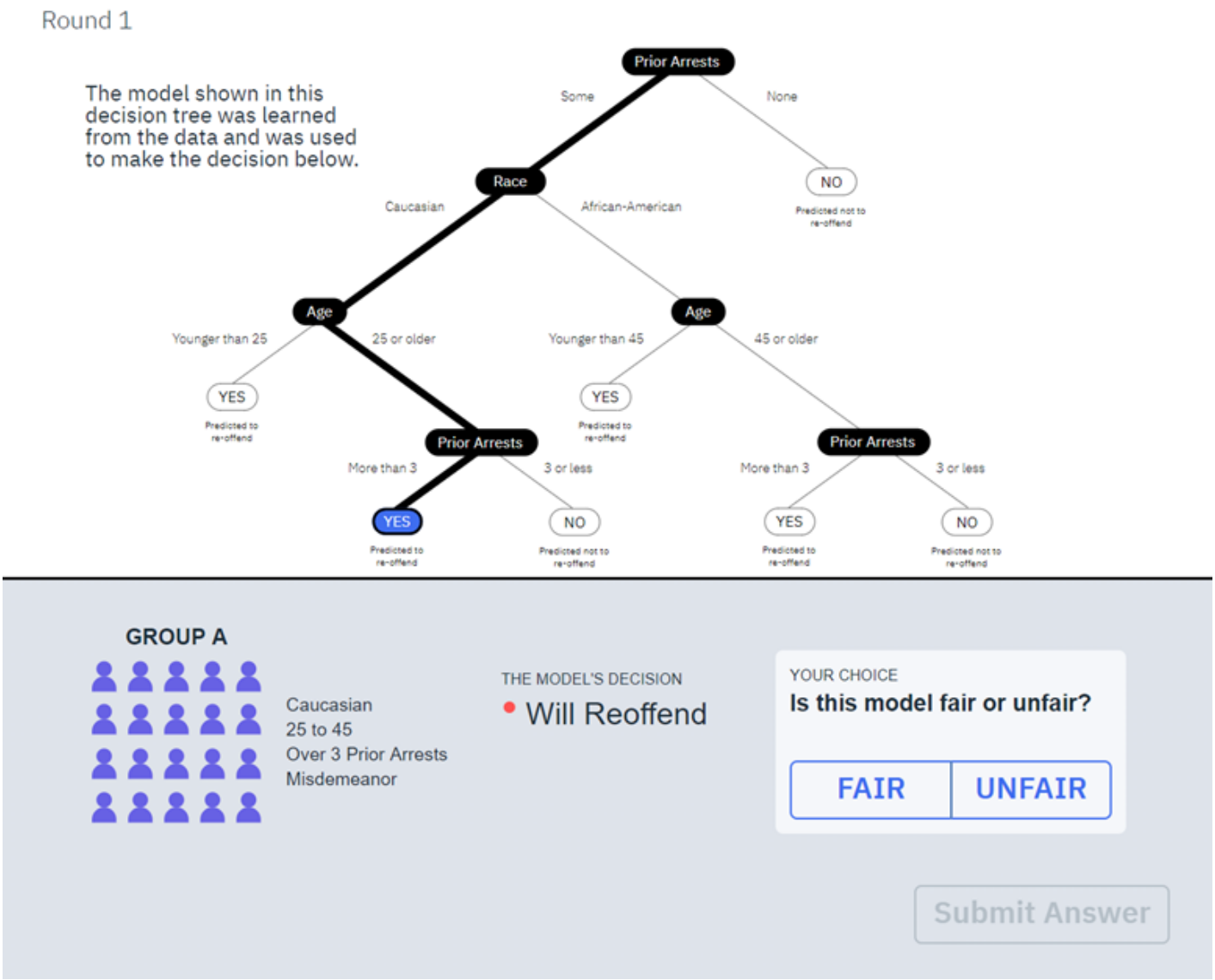
Otherwise,
Will Not Reoffend



7. Click **Submit Answer**.

The **Round 1** page shows the decision tree the model used to make the decision.

8. Read the page and then click **FAIR** or **UNFAIR** based on your evaluation of the model.

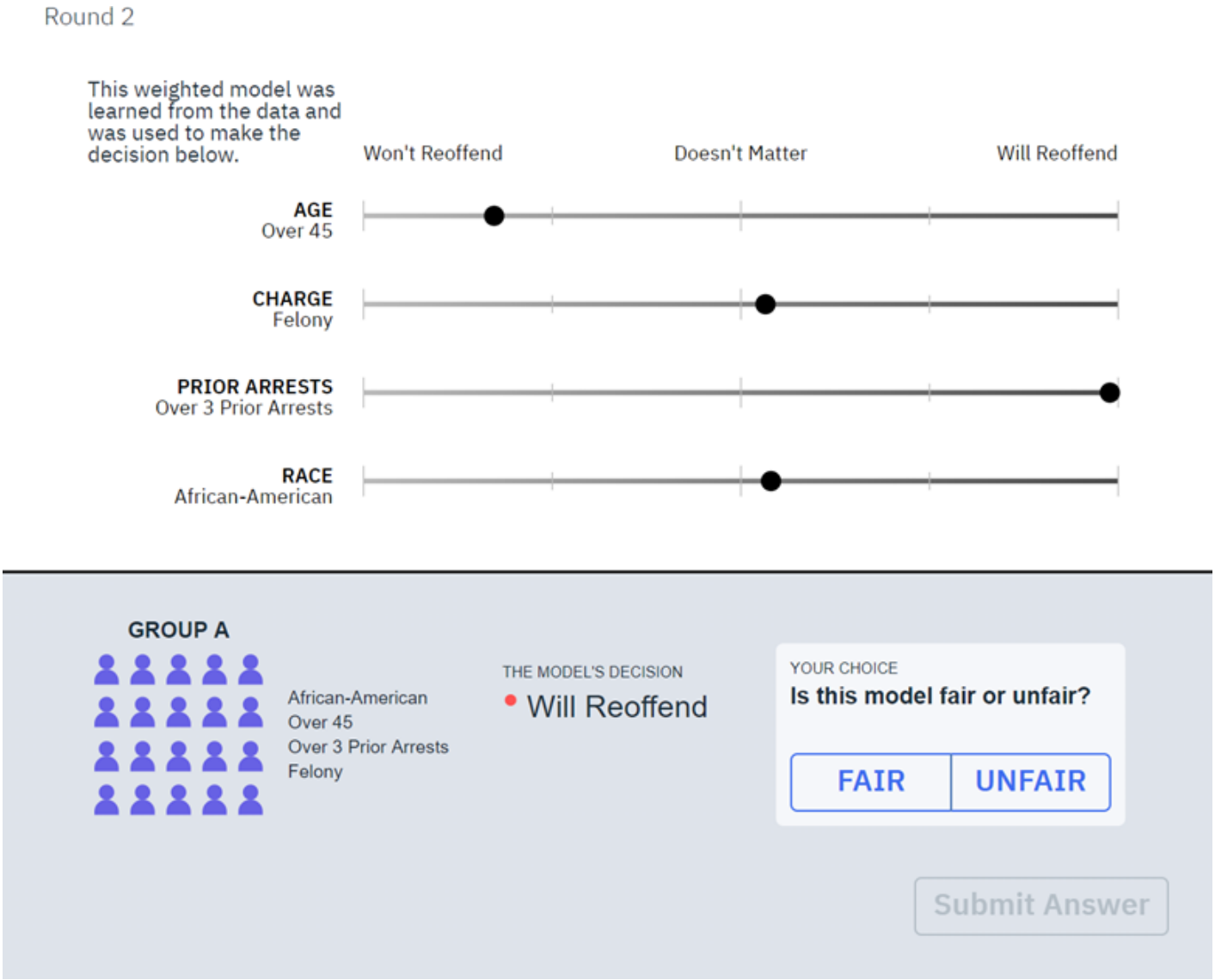


9. Click **Submit Answer**.

10. The message box confirms your choice, and tells you how many people agreed with you. Click **Next**.

The **Round 2** page shows the weighted model with the relative weights applied to key attributes that the model used to make this decision.

11. Read the page and then click **FAIR** or **UNFAIR** based on your evaluation of the model.



12. Click **Submit Answer**.

13. The message box confirms your choice, and tells you how many people agreed with you. Click **Next**.

The **Round 3** page shows the decision tree the model used to make this decision.

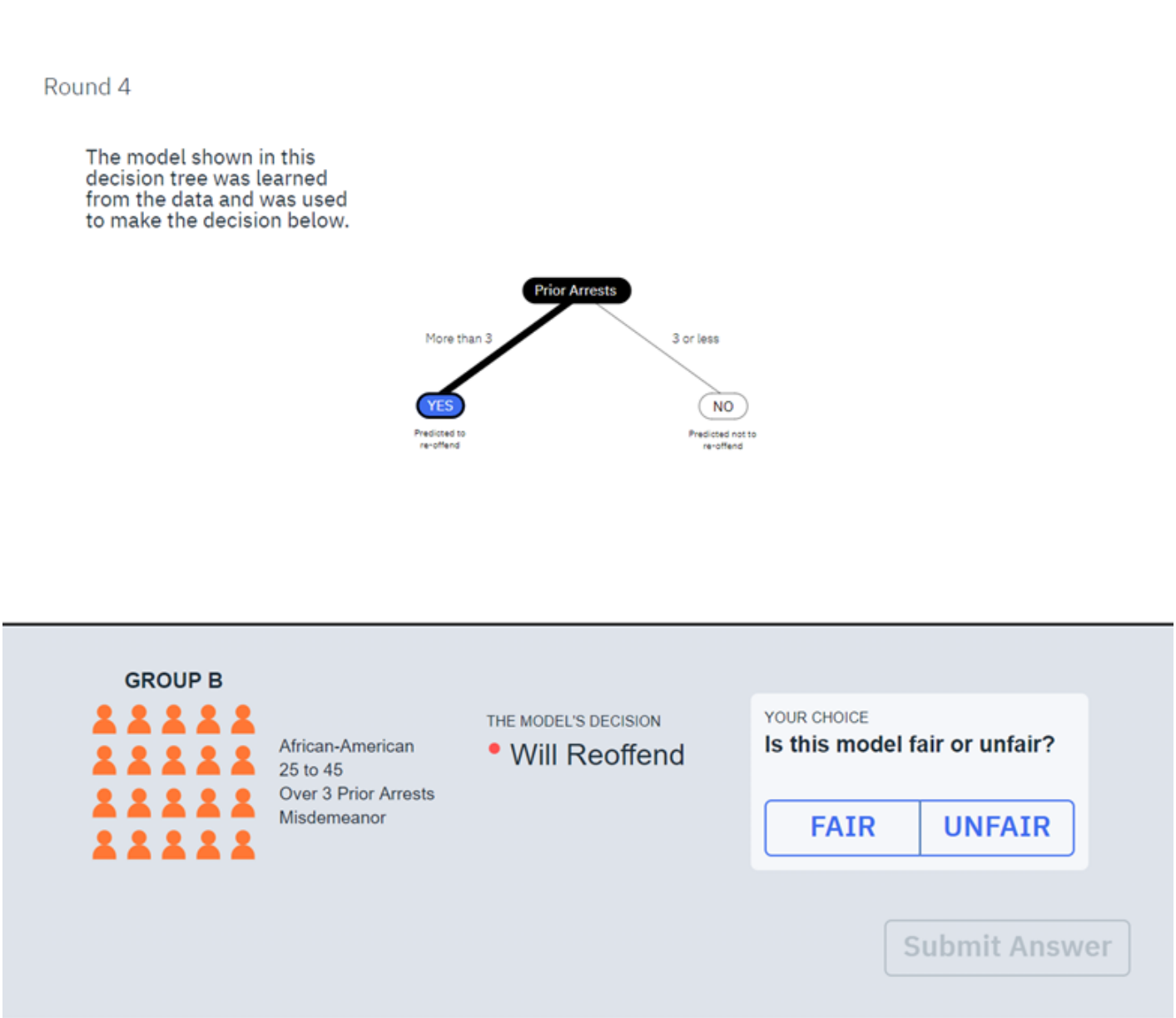
14. Read the page and then click **FAIR** or **UNFAIR** based on your evaluation of the model.

15. Click **Submit Answer**.

16. The message box confirms your choice, and tells you how many people agreed with you.Click **Next**.

The **Round 4** page shows the decision tree the model used to make this decision.

17. Read the page and then click **FAIR** or **UNFAIR** based on your evaluation of the model.



18. Click **Submit Answer**.

19. The message box confirms your choice, and tells you how many people agreed with you. Click **Next**.

The final page summarizes the results.

20. If you wish to provide your age and gender, do so. It is not required.

21. Click **Submit Answer**.

[Optional]: On the Learn More page, click the link to get more information about IBM's research on Fairness in Models.

22. Click **Exit**.

Use the Discussion Forum to talk about the models with your fellow students.

- Which model did you feel was most fair?
- What attribute seems to have most effect on re-offending?