1. **Ethan** — 1/26/2026 6:57 PM

When mod 2 group work starts we can discuss here.

2. **Donovan** — 1/26/2026 7:12 PM

aight

3. **Ethan** — 1/29/2026 6:57 PM

Module 2 Group Project - Bias

- Set up your group GitHub and document – establish a communications plan
- Identify a data set with known biases
  - The bias should focus on an ethical issue with real-world consequences
  - You may need to do some research to find a good example
- Confirm the biases with your own data analysis
  - This will involve some coding and analysis
  - Since this is a group project, different students can take different approaches
- Bring the work together into a final document

19

4. **Christian** — 1/31/2026 4:25 PM

I haven't taken the module 1 exam yet but I'm gonna spend a little time looking for some potential datasets

5. Did he say whether or not synthetic datasets are ok for this project?

6. **Ethan** — 1/31/2026 4:30 PM

@jfwallin ^

7. **Donovan** — 1/31/2026 8:24 PM

Repo has been made, please send your GitHub names or emails so I can invite you please
https://github.com/DonovanManogue/Data-6550---Data-Ethics-and-Responsibility

GitHub

[DonovanManogue/Data-6550---Data-Ethics-and-Responsibility](#)

Here lies the semester of Data Ethics and Responsibility Repository for all projects and assignments within the class - DonovanManogue/Data-6550---Data-Ethics-and-Responsibility

8. Ethan you have been invited already

9. **Ethan** — 2/1/2026 9:20 AM

I think we need a repo for each module. Document says this one should be called 'DATA6550-Bias'

10. **Donovan** — 2/1/2026 11:24 AM

What if we did, folders for each one? Or we can go ahead and have like 4 repos too, which ever makes sense I'm down for whatever

11. **Ethan** — 2/1/2026 11:25 AM

i like the folder idea. however, i think he might want separate repos so it's easier to look at commit history

12. **Donovan** — 2/1/2026 11:25 AM

Ahhh REAL

13. I'm down for that

14. **Christian** — 2/4/2026 5:56 PM

https://www.kaggle.com/datasets/danofer/compass

COMPAS Recidivism Racial Bias

Racial Bias in inmate COMPAS reoffense risk scores for Florida (ProPublica)

15. ok i took the reading quiz without doing any of the readings and saw that this dataset was referenced so ill probably look for another one lol

16. **Ethan** — 2/4/2026 6:03 PM

reading mainly was centered around woman of color not being presented nearly as much as men. if that gives you any idea. i dont know if we are supposed to find a dataset that specifically is labeled as bias or if we should find one that they dont know is bias but is

17. Donovan — 2/4/2026 6:44 PM

I was gonna say I saw some aids research that does the same for people of color and women

18. With mostly white males being the majority making the Imbalanced datasets showing less of the true picture

19. I didn't find the actually dataset set tho, only the paper over it

20. Christian — 2/4/2026 6:52 PM

This has happened to me a couple times so far

21. Daniel — 2/5/2026 3:43 PM

https://github.com/guthriedaniel80

GitHub

guthriedaniel80 - Overview

guthriedaniel80 has one repository available. Follow their code on GitHub.

22. Sorry, I didn't realize we were doing the discussion in a different room and I wasn't getting notifications for some reason

23. https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge

Toxic Comment Classification Challenge

Identify and classify toxic online comments

24. Daniel — 2/5/2026 3:52 PM

Models looking to classify toxic language use are known to disproportionately flag language predominantly used by certain cultural groups/demographics

25. Ethan — 2/5/2026 4:17 PM

All good

26. Donovan — 2/5/2026 4:18 PM

https://www.kaggle.com/datasets/mosapabdelghany/adult-income-prediction-dataset/data This one is pretty good Gender Biased, Racial Imbalance, and some age skewing

Adult Income Prediction Dataset

Census-based dataset for income classification tasks

27. Ethan — 2/5/2026 5:05 PM

https://www.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page This one is pretty heavy on the person of color male frisks, searches, and use of force. More than white males, white females, and person of color females.

Publications, Reports - NYPD

28. Donovan — 2/5/2026 6:51 PM

@Christian what is your github g

29. Ethan — 2/5/2026 6:52 PM

@Donovan @Daniel @Christian what's your google emails for google doc?

30. Donovan — 2/5/2026 6:52 PM

dmanogue42@gmail.com

31. Christian — 2/5/2026 6:58 PM

jenner.adcock@gmail.com

32. My github is just ChristianAdcock

33. Daniel — 2/5/2026 7:01 PM

guthriedaniel80@gmail.com

34. Ethan — 2/5/2026 7:01 PM

Okay cool you guys should have access to the google doc

35. Ethan — 2/5/2026 7:02 PM

I think this one would be a cool one to explore the different kinds of bias

36. Ethan — 2/6/2026 9:14 AM

i got the repo structure setup for us. make sure to pull if you have already cloned it

37. Donovan — 2/6/2026 11:00 AM

Sounds good, so no meeting tonight?

38. I'll go ahead and do some analysis this weekend. Not too sure how we split this up into 4 lol

39. I guess meet Monday and discus

40. Ethan — 2/6/2026 11:02 AM

we could have two people do coded analysis and two people write the report based on the coded analysis? and then each week we rotate?

41. i could code or write this week. either or is fine with me

42. i think meeting tonight would just be discussing who is doing what but we could settle that over text and then just meet monday night

43. coders should do EDA and output graphs to support their findings so we can have plenty to write about in the report

44. @Daniel @Christian thoughts?

45. Donovan — 2/6/2026 11:11 AM

I call data analysis

46. Daniel — 2/6/2026 3:20 PM

I would be down to write part of the report

47. Ethan — 2/6/2026 4:07 PM

Cool. @Christian I'll take whatever role you dont want for this week then

48. So meetup Monday 8pm instead of tonight? And then we can just update each other daily with the work that's been done?

49. Christian — 2/7/2026 9:50 AM

I'll write this week

50. Sounds like a good plan

51. Ethan — 2/7/2026 9:51 AM

solid

52. Ethan — 2/7/2026 1:26 PM

If any of you two want to start writing, I pushed some analysis to the repo. My sandbox notebook as an analysis based on income > 50k for race & sex and an analysis based on income >50k for education, race, & sex. I have notes throughout the notebook. If you'd prefer I write a summary at the end I could to make it easier for you. I also saved 2 graphs under the analysis folder that we can use for the presentation.

53. This is the rubric:

54. No clue how much we have to write but I guess when we get together on monday we should make a small presentation? Wish we had specifics. That's assuming we get coded analysis done today and written report done tomorrow.

55. Daniel — 2/7/2026 1:48 PM

@Christian is there any specific way you would want to split up the writing portion?

56. And I'll try to get the writing done tomorrow so we can discuss on Monday. I appreciate you doing all that @Ethan

57. Christian — 2/7/2026 2:49 PM

I'm good with whatever

58. Ethan — 11:20 AM

are we still on for 8pm tonight? has anyone started writing, I see Donovan did some work in his notebook too

59. Donovan — 11:20 AM

I'm down for 8 easily

60. I might add some more tonight, but not too sure if I need to or not

61. Daniel — 11:21 AM

8 is good. I'm about to start writing

62. Daniel — 1:23 PM

@Christian I'm splitting the writing portion into explicitly encoded bias (the dataset containing protected categories, race and sex, as variables) and proxy variables (education, occupation, and work class). Is that okay, and if so, would you want to take the proxy variable portion?

63. Ethan — 1:41 PM

my notebook is high income earners based on race, sex, and education. @Donovan did you touch on occupation and work class? if not then we can probably cut that from the report

64. Christian — 1:45 PM

ok

65. Daniel — 1:49 PM

@Ethan Looks like he mentions occupation and work class as potential proxy variables, but focuses on race, sex, and education. We can stick with those

66. Donovan — 1:52 PM

Yes. I was gonna add a part 2 for those, but I was debating if we really needed that much stuff tbh

67. Daniel — 1:52 PM

DId he ever mention how long the paper needed to be?

68. Donovan — 1:53 PM

Probably in the d2l requirements or like grading guide

69. Ethan — 1:54 PM

This is all the rubric says

70. i would assume a page to two pages

71. Christian — 3:44 PM

"You will also prepare a final 5-page report as a shared Google document."

72. from the project doc

73. Ethan — 3:45 PM

good catch

74. Ethan — 7:36 PM

Ethan Stanks is inviting you to a scheduled Zoom meeting. Topic: Group 1 | Module 2 Discussion Time: Feb 9, 2026 08:00 PM Central Time (US and Canada) Join Zoom Meeting https://mtsu.zoom.us/j/89696073052 Meeting ID: 896 9607 3052 --- One tap mobile +13092053325,,89696073052# US +13126266799,,89696073052# US (Chicago) Join instructions https://mtsu.zoom.us/meetings/89696073052/invitations?signature=PdksrCKgErzshTcbJEy 5bQmYqdTp_Juix9YjkuBupQk

Zoom

Join our Cloud HD Video Meeting

Zoom is the leader in modern enterprise cloud communications.

75. Ethan — 8:37 PM

Added occupation to the end of my notebook @Christian