Fair or not? Synthetic hiring data exploration

Hanshu Yu

1st September 2025

1. Background

We would like to have a fair world. But wait a second…what is “fair”?

If you are hiring people for a company, how could you “be fair”?

In this assignment, we make use of an open source synthetic dataset Utrecht Fairness Recruitment dataset version 4 to explore this ethical question and practice your programming skills. Use the latest edition

2. Instruction

1. Define what does “fairness” mean in this scenario. You do not have to come to a conclusion, but you do have to come up with at least one definition, but it will be nice if you have multiple.
2. Translate your fairness definition into desired scenarios that you would like to consider. (Age, gender, hobbies, etc.)
3. Explore the dataset, clean the data. Pick the data you feel relevant and need to use.
4. Develop statistical methods to evaluate fairness based on the desired scenarios your have translated from your definition.
5. Visualize your fairness evaluation and explain what you find.
6. Critically examine your fairness definition. What do you think of your initial hypothesis?

(For exploratory bonus)

Can you implement machine learning techniques in the analysis?

Explain what method/technique you used and why. What is the difference between this technique and your previous analysis?

3. Deliverables

1. A short report, clear and to the point.
    a. Min 1 page A4, max 6 pages A4 (excluding appendix)
    b. Suggested structure:
        - Explaining fairness: from theory to examples
        - Statistical evaluation of fairness
        - Code structure
        - Results:
            - ✓ How did you process the dataset?
            - ✓ What method did you use?
            - ✓ Based on your definition(s), is the hiring process fair?
        - Reflection:
            - ✓ What did we do good?
            - ✓ What can be improved?
            - ✓ What did we learn while working on this code?
            - ✓ What we know better about fairness?
            - ✓ Good and bad about statistical evaluation of fairness.
2. Your code in appendix.

3.1 Report guideline

- Very concise and straight to the point. You do not need a cover page, you do not need contents like: “The purpose of this project……” You do not need 3rd person perspective language, just say “We decided……”,”We write ……”. You do not need background and too much description in detail.
- The report must be logical and easy to read, free from grammar & spelling mistakes.
- Screenshots must be clear and readable.
- I strongly suggest you to just copy and paste your code to the appendix, do not use screenshots, do not worry about the colors and general format that make your code look good, just make sure indentations are there when needed. Do make sure that the code readable, a direct copy from a dark theme IDE may make your code unreadable.
- Recommended format normal text: A4 paper, Arial, size 12, line space: multiple 1.25.
- Submit the report in groups, remember to include your names and student numbers on the report.

3.2 Code guideline

- It should work like what is described in the assignment.(Section 1 of this document)
- You should have necessary comments to the extent that you still can understand each line after leaving it for a week. You must be able to explain your code.