Week 1 Discussion Summary

Summary of <https://github.com/IMOsbo/DATA6550-Bias/blob/main/Collaboration/WeekA.pdf>

The discussion in the file revolves around a group project for a Data Bias course, specifically focusing on selecting a dataset for analysis.

1. **Dataset Selection**: The group considered several datasets with known biases, including the COMPAS Recidivism Dataset, UCI Adult Income Dataset, LendingClub Loan Dataset, IMDb Reviews, and various facial recognition and hiring datasets. They ultimately leaned towards using the **UCI Adult Income Dataset**, which is easier to work with and has clear biases.
2. **Task Distribution**: There was concern about whether the dataset was large enough for everyone to contribute code. Suggestions included breaking it down by variable and assigning different aspects of analysis and visualization to each group member.
3. **Project Setup**: A GitHub repository was created for collaboration, and a basic Jupyter Notebook was set up to load the dataset. Team members were encouraged to copy and rename the notebook for their work.
4. **Administrative Details**: A discussion took place about whether they needed a separate repository for each project. Initially, they hoped to keep everything in one place, but after consulting their instructor, they were required to create a new repo.
5. **Final Steps**: The team worked on a Word document summarizing their discussion and analysis. They coordinated on where to store files and how to submit their final work. A Zoom meeting was offered for last-minute coordination.

Overall, the discussion was focused on organizing their project efficiently, ensuring everyone had a meaningful contribution, and complying with their instructor’s requirements.