# DATA 6550 - Group 6 - Module 2 Discussion Summary

## Summary of Discussion

This document summarizes the discussion held on February 2, 2025, by Group 6 of the DATA 6550 course. The team focused on analyzing recidivism data, identifying biases, and using GitHub for collaboration.

## Project Planning & Task Division

- The group discussed the dataset, which contains known biases related to race, gender, and age.  
- Tasks were divided among members, including statistical analysis, visualization, and bias identification.

## Biases in Data

- Race: Black defendants were more likely to be misclassified as high risk.  
- Gender: The dataset consists of 82% males and 18% females.  
- Age: The majority of subjects fall in the 21-25 age range.  
- Charge Descriptions: Potential bias in how law enforcement categorizes charges.

## GitHub Collaboration

- The team set up GitHub for version control and collaboration.  
- Members were guided on cloning repositories, committing changes, and pushing updates.  
- Individual folders were created for organizing Jupyter notebooks.

## Data Preparation & Analysis

- The ‘parsed filtered’ dataset version was chosen for analysis.  
- A utility script (utils.py) was introduced to streamline data loading and package installation.

## Statistical Approach

- The team explored controlling for prior crimes when evaluating recidivism risk scores.  
- Consideration of whether to analyze factors independently or together.

## Learning & Technical Challenges

- Some members were new to GitHub and required assistance.  
- Issues with file visibility and repository setup were addressed.

## Next Steps

- Completing individual bias analysis and integrating findings into a report.  
- Ensuring all members can efficiently use GitHub.  
- Conducting controlled statistical analyses to validate biases in recidivism data.