Bias Detection Report

Detection Report Summary

Dataset Analyzed

- **File Path**: source_files/Adult.csv

- **Features**: "race" and "income"

Types of Bias Detected

- **Correlation Bias** between two categorical features.

Tools and Methods Used

1. **Chi-Square Test**: Indicated significant bias with a Chi-Square statistic of 330.92 and a p-value of 2.31e-70.

2. **Wasserstein-2 Distance**: Measured significant distributional differences with a value of 3.43.

3. **Maximum Mean Discrepancy (MMD)**: Resulted in a negative value (-5.91), suggesting negligible bias or requires re-evaluation.

Evaluation of Bias Levels

- The Chi-Square and Wasserstein-2 results point towards a significant bias between "race" and "income."

- MMD requires careful interpretation due to its negative value.

Visualizations

- Grouped Bar Chart and Box Plot illustrate the income distribution across different racial groups.

Recommendations

- **Significant Bias**: Caution is advised when using this dataset for analysis that involves race and income. Consider addressing the bias or adjusting analyses accordingly.

Additional Explanation

The Chi-Square test and Wasserstein-2 distance both indicate substantial bias. The MMD result may require further examination or adjustment in methodology to align with the other findings.