After analyzing the COMPAS dataset using IBM's AI Fairness 360 toolkit, we observed significant racial bias in the risk prediction scores. African-American individuals showed a higher false positive rate compared to White individuals, indicating they were more often incorrectly labeled as high-risk.

The disparate impact metric was below the 0.8 fairness threshold, confirming systemic bias. To mitigate this, we applied the **Reweighing** algorithm, which adjusts instance weights to ensure fairness during training.

Post-mitigation, fairness improved, but residual differences remained. This highlights the importance of combining technical remediation with systemic changes in data collection and societal context.

Our visualizations provided a clear comparison of metrics before and after mitigation, emphasizing the ethical need to audit and correct AI systems regularly.

We recommend continued bias audits, inclusive data sourcing, and stakeholder transparency as long-term strategies for deploying fair AI systems in criminal justice.