Hands-on: Bias Testing Code

Python Bias Testing Example

```
from sklearn.metrics import confusion_matrix import pandas as pd # Evaluate performance by group
def evaluate_fairness(y_true, y_pred, sensitive_attr): results = {} for group in
sensitive_attr.unique(): mask = sensitive_attr == group tn, fp, fn, tp = confusion_matrix(
y_true[mask], y_pred[mask] ).ravel() results[group] = { 'TPR': tp / (tp + fn), # Recall 'FPR': fp
/ (fp + tn), 'PPV': tp / (tp + fp) # Precision } return pd.DataFrame(results).T
```

Testing Tools

- Fairlearn (Microsoft)
- AI Fairness 360 (IBM)
- What-If Tool (Google)
- Aequitas

Visualization

- Performance comparison charts by group
- Confusion matrix heatmap
- Fairness metrics dashboard
- ROC curve comparison

Key Practice Points

Calculate TPR, FPR, and PPV for various demographic groups and verify that disparities are within acceptable ranges (e.g., 10%)