Outcome Distribution by Sex

Table 1: Outcome distribution by sex and income (all data)

Sex	≤50K	>50K	Total
Female	13026	1669	14695
Male	20988	9539	30527
Total	34014	11208	45222

Metrics

- Mean difference in positive outcome rate (Male − Female): 0.198901432678815
- Correlation between target (y) and sex_male indicator: 0.21576045786566342

Discussion of Potential Bias Sources

The observed positive outcome rate (income >50K) is substantially higher for males than females (difference $\approx$ 0.20), and the moderate positive correlation between the binary target and the sex_male indicator (0.216) suggests that sex is associated with earning above $50K. This disparity may reflect underlying structural factors encoded in other correlated variables (e.g., occupation, hours-per-week, capital gains) that themselves carry historical gender imbalances. Because the dataset uses existing census information, selection mechanisms (who is employed, occupational segregation, differing work hours) can embed systemic biases; models trained directly on these features risk learning patterns that amplify inequities if sex-correlated proxies (like certain occupations or relationship statuses) drive predictions. Additionally, the class imbalance (fewer high-income females) can lead to poorer calibration for female subgroups. These preliminary statistics motivate further fairness-aware evaluation when building predictive models on this dataset.