How Statistical Paradoxes Challenge Traditional Gender Pay Gap Analysis

Yahel Ivgi and Tamar Michelson

August 2025

Abstract

This study examines gender-based income disparities using the Adult Income Dataset (N=32,561) to investigate the relationship between gender, marital status, and high income achievement ($<50K$). While confirming the existence of a substantial overall gender gap (19.7 percentage points), our analysis reveals a surprising finding: this disparity completely disappears among married individuals, with married women actually showing slightly higher income rates than married men (45.5% vs 44.6%, $p = 0.486$). Through comprehensive statistical analysis including z-tests, chi-square tests, Mann–Whitney U tests, and logistic regression modeling, we demonstrate that the gender gap is concentrated entirely among single individuals. Age-stratified analysis reveals persistent gender gaps across all age groups, creating an apparent contradiction that we resolve through the identification of Simpson's Paradox. The phenomenon occurs due to dataset-specific biases, including markedly different marriage rates between genders (61.2% of men vs 15.5% of women married). Machine learning validation using logistic regression confirms marital status as a strong predictor of income level, which motivated a deeper analysis of marital patterns in the dataset.

Keywords: Gender gap, Simpson's Paradox, Non-parametric methods, marital status, subgroup analysis

1 Introduction

Gender-based income disparities have been a persistent concern in socioeconomic research, with numerous studies documenting significant differences in earnings between men and women across various demographics and industries. Traditional analyses of gender pay gaps typically focus on overall population comparisons, often revealing substantial disparities that have important policy implications for workplace equality and economic justice.

However, the relationship between gender and income is complicated by multiple confounding variables, including marital status, age, education, and career choices. Recent research has suggested that these factors may interact in complex ways that are not immediately apparent from simple demographic comparisons. The phenomenon known as *Simpson's Paradox*, where trends observed in individual groups reverse when groups are combined, has been documented in various fields but remains underexplored in gender income research.

1.1 Research Objectives

This study aims to examine gender income disparities through a comprehensive statistical lens, specifically investigating:

1. The existence and magnitude of overall gender income gaps
2. Whether these gaps persist when controlling for marital status

3. The distribution of income disparities across different demographic subgroups
4. The potential presence of statistical paradoxes in gender income data

1.2 Research Hypotheses

H1: Significant gender income disparities exist in the general population

H2: Gender income gaps persist among married individuals

H3: Gender income gaps are concentrated among unmarried individuals

The Adult Income Dataset, with its extensive demographic and socioeconomic variables for over 32,000 individuals, offers a comprehensive foundation for examining the complex interplay between gender, marital status, age, and income.

2 Results

2.1 Overview of Dataset and Initial Findings

Our analysis was conducted on the Adult Income Dataset from Kaggle, comprising 32,561 individuals with 15 demographic and socioeconomic variables. The dataset showed a baseline high income rate ($<50K$) of 24.1%, with notable variations across demographic groups. Initial exploration revealed significant disparities that warranted deeper statistical investigation.

2.2 Preliminary Statistical Analysis

Prior to examining specific gender-income relationships, we conducted comprehensive preliminary analyses to assess the distributional properties of our variables and identify factors influencing income levels. Normality testing revealed that none of the dataset features followed normal distributions, necessitating the use of non-parametric statistical methods for subsequent analyses. We employed Chi-square tests for categorical variables and Mann-Whitney U tests for continuous variables to examine the relationship between each demographic characteristic and income status. Additionally, Pearson correlation analysis was conducted to quantify the strength of linear relationships between features and income. While all correlations were statistically significant, they were generally weak in magnitude, suggesting that individual features alone have limited predictive power for income status. The results demonstrated that all examined characteristics statistically significant ($p < 0.05$ for all non-parametric tests), even after applying multiple comparisons correction to control for Type I error inflation. This confirmed the multifactorial nature of income determination and justified our detailed investigation into specific demographic interactions, particularly the complex relationship between gender, marital status, and income achievement.

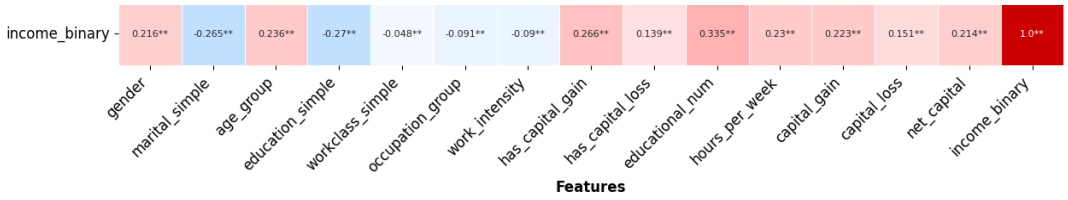

Figure 1: Pearson Correlation Between all Features and Income (* p<0.05, ** p<0.01)

2.3 Claim 1: There Exists a Significant Gender Gap in Income

Hypothesis: Male and female populations differ significantly in their likelihood of earning high income.

Our analysis strongly supports this claim. The observed high income proportions are:

- **Male**: 0.306 (6,662 out of 21,790)
- **Female**: 0.109 (1,179 out of 10,771)
- **difference**: 0.197

We conducted a one-proportion z-test to assess the significance of this difference, yielding:

- **Z-statistic**: 38.973
- **P-value**: < 0.001 (highly significant)
- **Cohen's d**: 0.47 (small to medium effect size)

The Chi-square test for gender and income association produced:

$$\chi^2 = 1,517.81, \quad p < 0.001, \quad \text{Cramer's} V = 0.216$$

demonstrating a moderate association between gender and income level.

Having established the existence of a substantial gender gap, we sought to understand whether this disparity persists across different demographic subgroups. Marriage represents a critical life factor that may influence income patterns, making it essential to examine whether the observed gender gap remains consistent among married individuals. In addition, according to the feature importance results from our linear regression model, marital status ranked as the second most influential predictor of income. We did not focus on the top-ranked feature, capital gain, as we consider it more a consequence of high income than a cause. We used a linear regression model rather than a Random Forest because it performed better on our dataset. Given the high ranking of marital status in the model's coefficients, we decided to examine this factor more closely, particularly its interaction with gender.

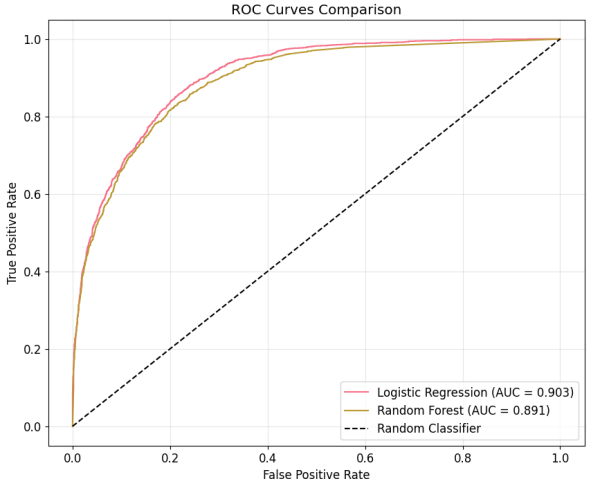

Figure 2: Roc curve for income forecasting models (Random forest and Logistic regression)

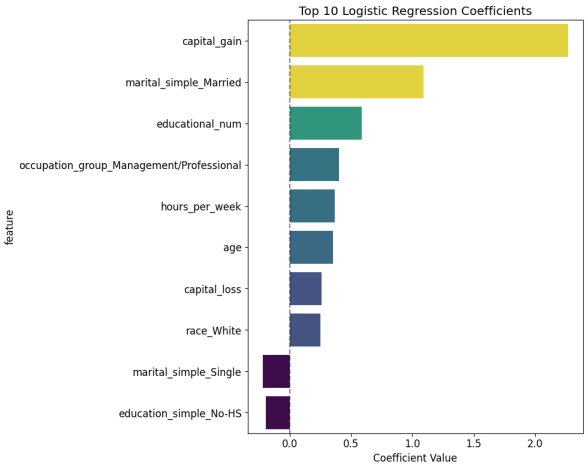

Figure 3: Logistic regression (Better model) coefficient

Chi-square tests for marital status and income association show:

$$\chi^2 = 6,517.26, \quad p < 0.001, \quad \text{Cramer's} V = 0.447$$

indicating a strong association.

2.4 Claim 2: The Gender Gap Persists Among Married Individuals

Hypothesis: Married men earn significantly more than married women.

Surprisingly, this hypothesis is **rejected** by our data. Among married individuals, the observed high income proportions are:

- **Married males**: 0.446 (5,942 out of 13,328)
- **Married females**: 0.455 (760 out of 1,671)
- **Difference**: -0.009 (favoring married women)

Statistical validation using a one-proportion z-test yields:

- **Z-statistic**: -0.697
- **P-value**: 0.757 (not statistically significant)
- **95% CI for difference**: [-0.034, 0.016] (indicating that the difference in proportions is not statistically significant since the interval includes zero)

This finding indicates that among married individuals, there is no significant gender gap in high income achievement. In fact, married women show a slightly higher (though not statistically significant) rate of high income earning than married men.

The results from Claims 1 and 2 present an intriguing pattern. Although a significant overall gender gap exists in the general population, this disparity is completely disappears among married individuals. A possible conclusion from this is that if the overall gap is substantial but absent among married individuals, then the gender disparity is likely concentrated in another segment of the population.

2.5 Claim 3: The Gender Gap is Concentrated Among Single Individuals

Hypothesis: The overall gender gap is primarily driven by differences among unmarried individuals.

This claim is strongly supported by our analysis.

Evidence from Single Population Analysis: Analysis of the single population reveals substantial gender disparities. Among single individuals in the dataset:

- **Proportion of single males with high income**: 0.085 (720 out of 8,462)
- **Proportion of single females with high income**: 0.046 (419 out of 9,100)
- **Difference**: 0.039

A one-proportion z-test indicates a significant difference between genders among single individuals:

- **Z-statistic**: 10.498
- **p-value**: < 0.001 (highly significant)
- **Cohen's d**: 0.159 (small effect size)

Having confirmed that the gender gap is concentrated among single individuals, we next sought to understand how this pattern manifests across different life stages. Age represents a crucial factor in career development and income progression, and examining age-specific patterns could reveal whether the gender gap among singles is consistent across all ages or concentrated in particular age groups. This analysis would also help validate our findings by testing whether the marriage-related patterns we observed hold true when we control for age effects.

2.6 Age-Stratified Analysis

To further investigate the interaction between age, gender, and income, we conducted age-stratified analyses. Our expectation, based on Claim 2, was to find at least some age groups where gender differences were minimal or non-significant.

Surprising Findings: Gender gaps persist across **all** age groups examined:

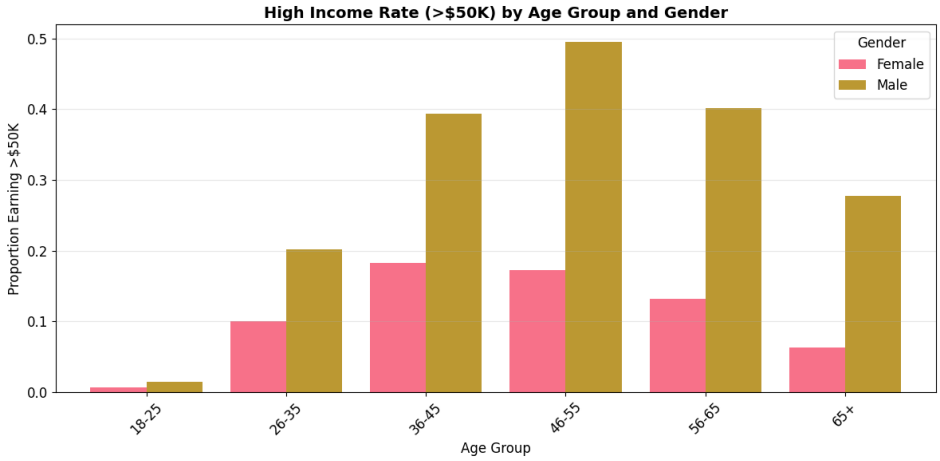

Figure 4: High Income Rate ($50K) by Age Group and Gender

Age Group	Male HIP	Female HIP	Gap	One Proportion Z Test
18–25	0.035	0.012	0.023	2.708 (p-value = 0.007)
26–35	0.238	0.094	0.144	11.742 (p-value $<$ 0.001)
36–45	0.429	0.271	0.158	18.503 (p-value $<$ 0.001)
46–55	0.510	0.386	0.124	22.85 (p-value $<$ 0.001)
56–65	0.418	0.264	0.154	14.697 (p-value $<$ 0.001)
65+	0.245	0.151	0.094	9.069 (p-value $<$ 0.001)

Table 1: Age-stratified gender gaps in HIP(high income proportions)

All age-gender combinations showed statistically significant differences using individual z-tests for proportions (all $p < 0.05$).

2.7 The Simpson's Paradox Explanation

The seemingly contradictory findings between Claims 2 and the age-stratified analysis represent a classic case of **Simpson's Paradox**. This statistical phenomenon occurs when trends that appear in different groups of data reverse when the groups are combined.

Key Insight: The composition of married vs. single populations differs dramatically across age groups and between genders:

1. **Unequal Marriage Rates**: In our dataset, 61.2% of men are married compared to only 15.5% of women, creating a fundamental compositional bias.

2. **Marriage Selection Effects**: Marriage appears to be associated with higher income potential for both genders, but the dramatic difference in marriage rates between men and women creates the paradoxical results.

The weighted average effect shows that while married individuals show gender parity, the overall population statistics are dominated by the large single population where significant gender gaps exist.

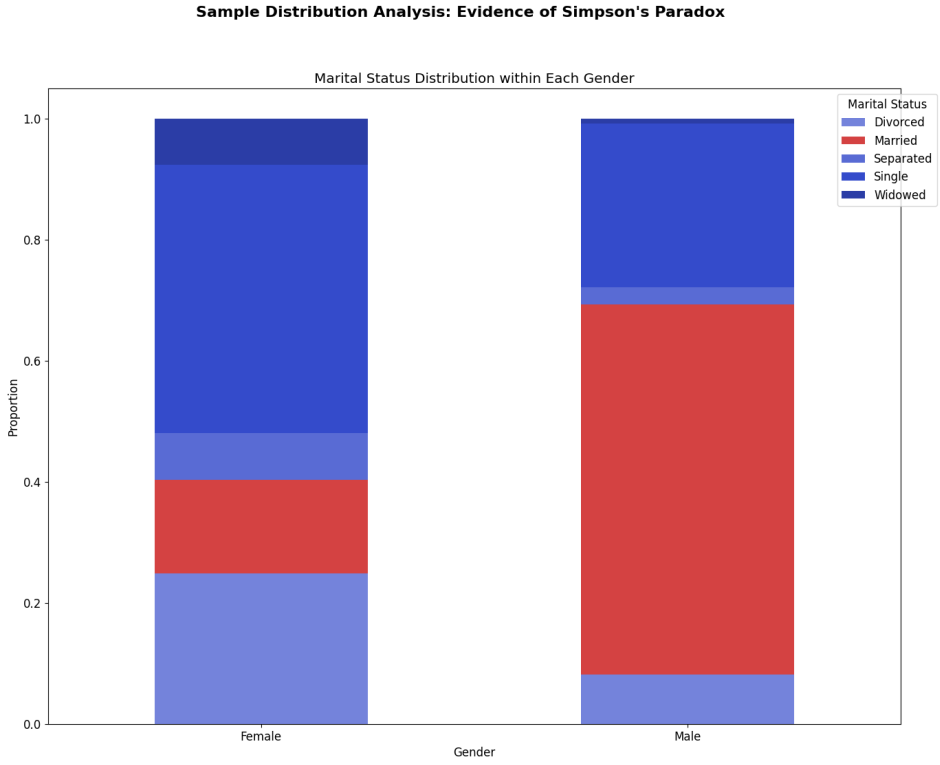

Figure 5: Marital Status Distribution within Each Gender

3 Methods

3.1 Dataset and Variables

Our analysis utilized the Adult Income Dataset from Kaggle, comprising 32,561 individuals with 15 demographic and socioeconomic variables. The dataset contains no missing values, ensuring complete case analysis. The primary outcome variable was binary high-income status (>\$50K annually), with key predictor variables including gender, marital status, age, education level, occupation, and work hours.

Variable preprocessing involved several standardization steps to facilitate analysis. Gender was coded as a binary variable (Male or Female), while income was dichotomized as high (>\$50K) versus low ($\leq$\$50K). Marital status was simplified from the original seven categories to four meaningful groups: married, single, separated, and widowed. Age was grouped into six categories spanning the adult working population: 18–25, 26–35, 36–45, 46–55, 56–65, and 65+. Education levels were consolidated from the original 16 categories into five meaningful tiers, ranging from no high school diploma to advanced degrees, to reduce complexity while maintaining interpretability.

3.2 Statistical Analysis Methods

Our statistical approach employed multiple complementary methods to ensure robust findings. Primary statistical tests included one-proportion z-tests for comparing income rates across gender, marital

status, and age groups; chi-square tests of independence for examining categorical associations; and Mann–Whitney U tests for non-parametric group comparisons. To address multiple comparisons, we applied the Benjamini–Hochberg false discovery rate correction where appropriate.

Effect size measurements were calculated to assess practical significance alongside statistical significance. We computed Cohen's d for the magnitude of differences between groups, Cramer's V for the strength of categorical associations, and confidence intervals for proportion differences to provide a comprehensive understanding of effect magnitudes.

Our advanced analysis strategy was designed specifically to investigate the apparent contradictions in our initial findings. Age-stratified analysis involved systematically examining income patterns within each age group separately, allowing us to control for age effects while investigating gender–income relationships. This approach was crucial for identifying whether the overall gender gap persisted uniformly across age groups or varied by life stage. The Simpson's Paradox investigation employed subgroup decomposition methods, where we systematically examined how overall population statistics related to subgroup patterns. This involved calculating weighted averages across demographic categories and testing whether aggregate trends matched subgroup trends. We also conducted sensitivity analyses by examining different combinations of demographic variables to understand which factors contributed most strongly to the paradoxical findings.

Machine learning validation supplemented our traditional statistical approach by using logistic regression and random forest models to identify the most important predictors of high-income status. These models helped validate our focus on marital status as a key moderating variable and provided feature importance rankings that guided our analytical priorities.

3.3 Software and Reproducibility

All analyses were conducted using Python with `scipy.stats` for statistical testing, `pandas` for data manipulation, and `scikit-learn` for machine learning implementations. Statistical significance was set at $\alpha = 0.05$, with multiple comparisons correction applied where appropriate to maintain Type I error control. All code and processed data are available for reproducibility verification, with detailed documentation of preprocessing steps and analytical procedures.

4 Discussion

This analysis of the Adult Income Dataset reveals a compelling example of Simpson's Paradox in the context of gender pay gap research. While the dataset shows a substantial overall gender gap (19.7 percentage points), this disparity disappears among married individuals, where women show slightly higher high-income rates than men (45.5% vs. 44.6%, $p = 0.486$). This apparent contradiction is driven by the dataset's demographic structure—specifically, the unusually large difference in marriage rates (61.2% among men vs. only 15.5% among women). These marriage rates should be interpreted as features of the dataset rather than as representative of the real population, as they may reflect sampling design, historical context, or data collection artifacts.

The presence of Simpson's Paradox here demonstrates how sample composition can fundamentally alter statistical interpretation. The overall gender gap in this dataset is largely a product of the high proportion of single individuals, among whom significant disparities persist. However, when married individuals are examined separately, gender parity emerges. This suggests that, in the dataset's internal structure, marital status acts as a moderating variable in income classification—but these findings cannot be directly generalized to the real-world population without further validation.

These results have methodological implications for the analysis of demographic disparities: aggregated statistics can obscure important subgroup patterns, and compositional bias can create misleading overall trends. At the same time, it is critical to remember that the observed patterns reflect the structure of this dataset rather than necessarily representing broader social reality.

4.1 Limitations

Several limitations must be acknowledged. First, the dataset is derived from the 1994 U.S. Census and may not reflect current socio-economic conditions. Second, the extremely low proportion of married women (15.5%) is likely a result of dataset-specific factors rather than an accurate population statistic. Third, the income variable is not the actual wage or salary, but a binary classification indicating whether annual income exceeds $50K. Therefore, our findings do not indicate that men earn more than women in absolute terms, but rather that, in this dataset, a higher proportion of men than women are classified as earning above the $50K threshold. Finally, the cross-sectional design prevents causal inference, and observed patterns may be influenced by selection bias, reverse causation, or unmeasured confounding variables.

4.2 Future Research

From a research perspective, these findings highlight the need for replication using contemporary datasets with more representative demographic structures and continuous income measures. Future studies should also examine whether similar instances of Simpson's Paradox occur in other datasets, and develop best practices for identifying and addressing compositional bias in demographic research. In addition, examining how marriage affects individual incomes could clarify the sources of the observed gender parity, showing whether it reflects income gains for women, income reductions for men, or a combination of both.

5 appendices

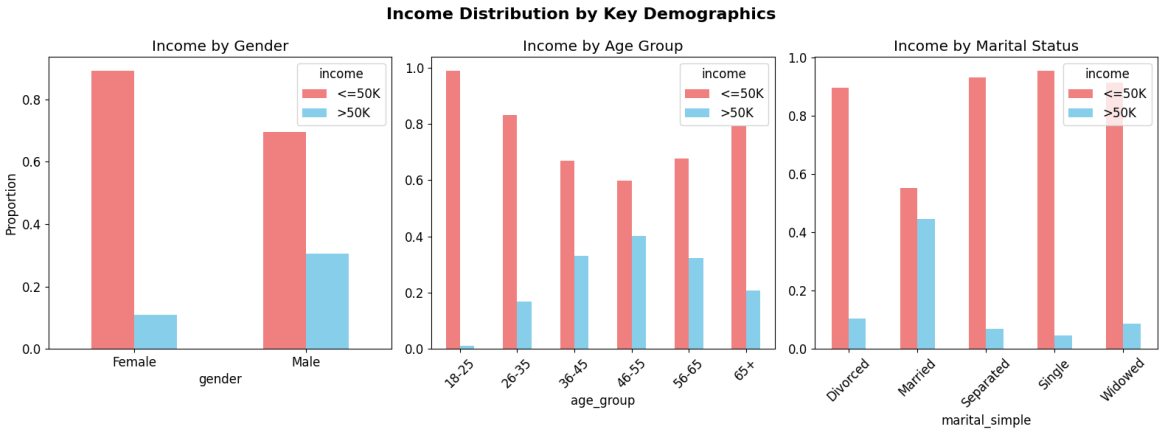

Figure 6: The distributions of each feature analyzed in this study.

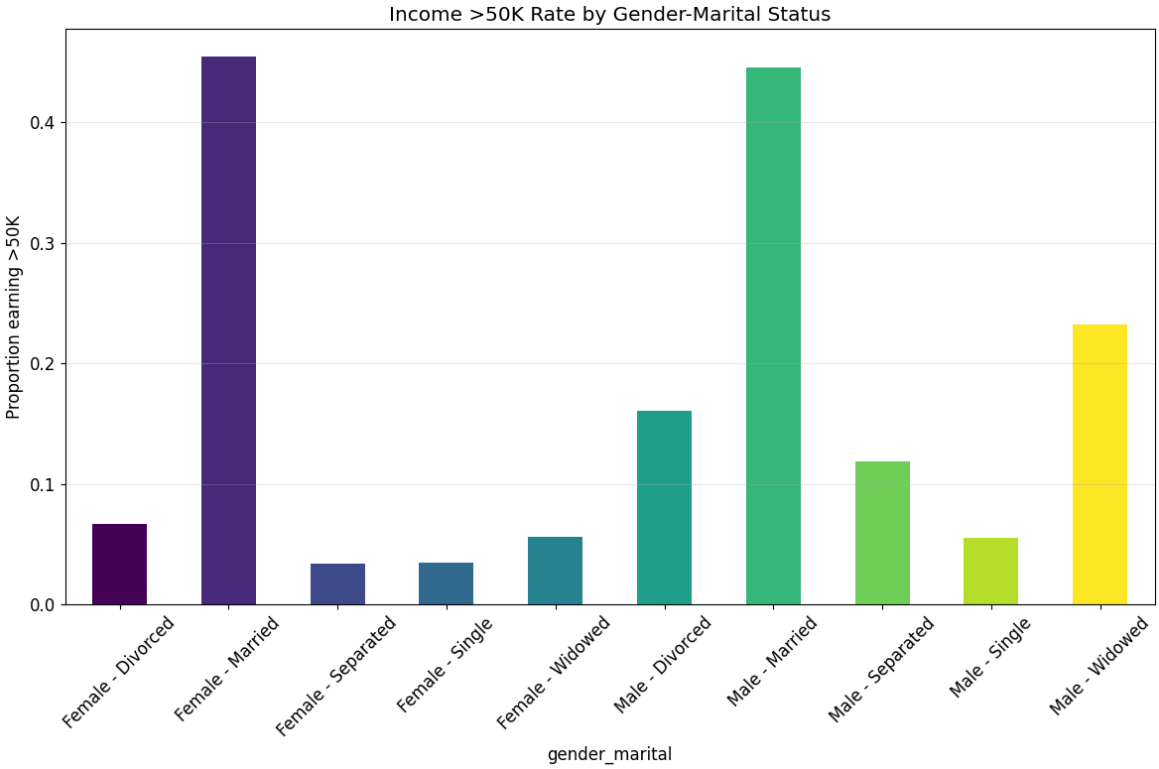

Figure 7: Income < $50K by Gender-Marital Status

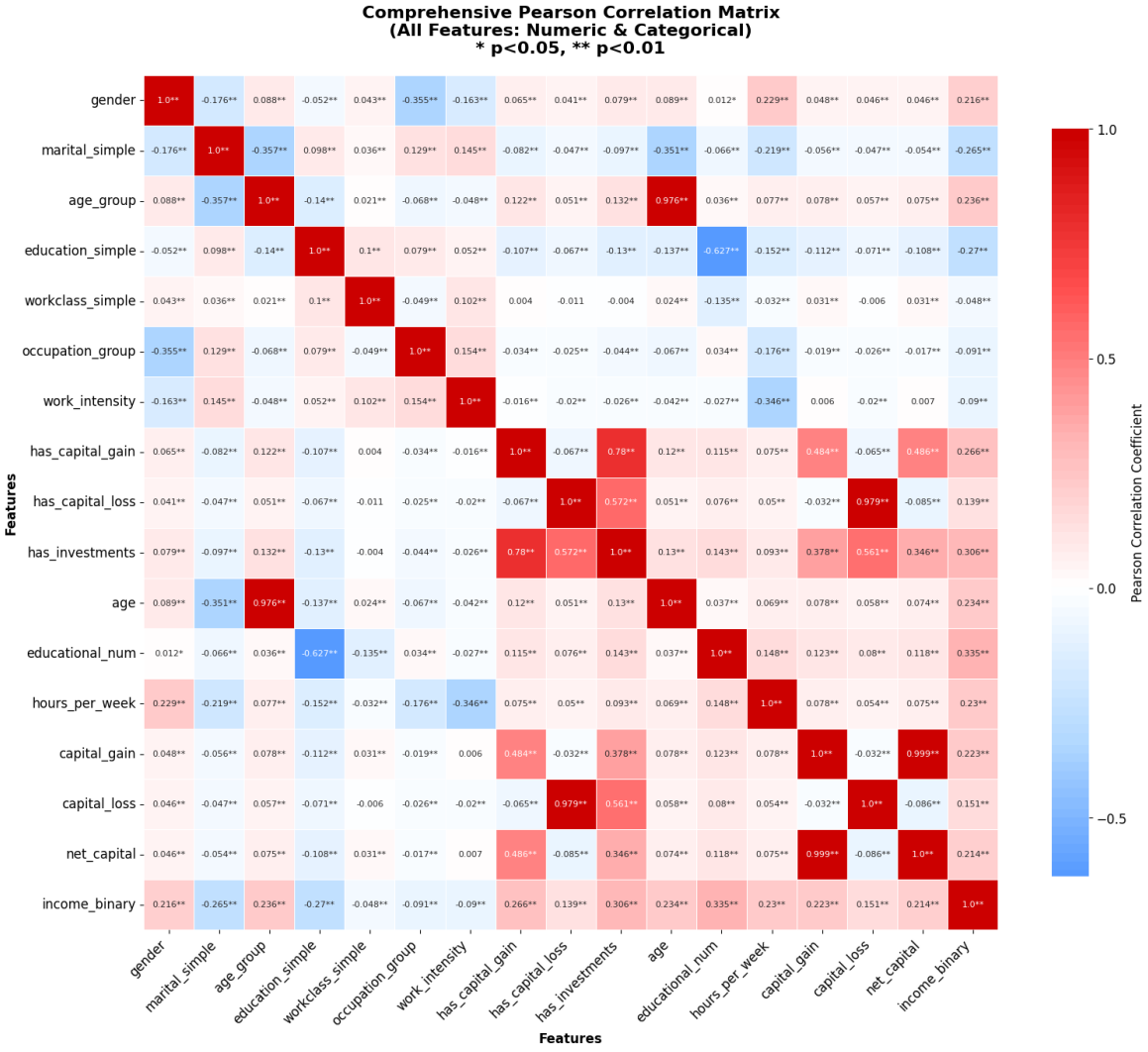

Figure 8: The correlation matrix of all features, computed using Pearson's correlation coefficient.