> Based on the code and analysis in the comprehensive EDA notebook at adult_eda.ipynb, What features show the
  strongest correlation with earning >$50K?

Based on the comprehensive EDA analysis in the notebook, the features showing the strongest correlation with
  earning >$50K are:

  Top predictive features:

  1. Education level - Advanced degrees show dramatically higher rates of >$50K income:
    - Doctorate: 74.1% earn >$50K
    - Professional school: 73.4% earn >$50K
    - Masters: 55.7% earn >$50K
    - Bachelors: 41.5% earn >$50K
  2. Age - Mean age for >$50K earners is 44.2 vs 36.8 for ≤$50K earners
  3. Hours per week - >$50K earners work 45.5 hours/week vs 38.8 for ≤$50K earners
  4. Capital gains - Mean capital gains are $4,006 for >$50K vs $149 for ≤$50K earners

  The analysis shows education level has the strongest predictive power, with clear income thresholds at different
  educational attainment levels.

> How does income vary with education, marital status, or hours worked per week?

 Education Level

  Strong positive correlation with income:
  - Doctorate: 74.1% earn >$50K
  - Professional school: 73.4% earn >$50K
  - Masters: 55.7% earn >$50K
  - Bachelors: 41.5% earn >$50K
  - Associates: ~25% earn >$50K
  - High school or less: 5-16% earn >$50K

  Marital Status

  From the categorical analysis (adult_eda.ipynb:cell-10), married individuals dominate high-income brackets. The
  data shows:
  - Married-civ-spouse: 14,976 individuals (largest group)
  - Never-married: 10,683 individuals
  - Divorced: 4,443 individuals

  Hours Per Week

  Clear difference in work hours:
  - >$50K earners: Average 45.5 hours/week
  - ≤$50K earners: Average 38.8 hours/week
  - Difference: 6.7 more hours per week for high earners

  The analysis shows education has the strongest predictive power, followed by work hours, with marital status
  (particularly being married) also being a significant factor for higher income levels.

> Are there disparities across race, sex, or native country?

Based on the comprehensive analysis of the Adult dataset, there are significant disparities across race, sex, and
native country:

Race Disparities (17.8% range)

- Asian-Pac-Islander: 26.5% earn >$50K (highest)
- White: 25.6% earn >$50K
- Black: 12.3% earn >$50K
- Amer-Indian-Eskimo: 11.6% earn >$50K
- Other: 8.7% earn >$50K (lowest)

Sex Disparities (19.6% gap)

- Male: 30.6% earn >$50K
- Female: 10.9% earn >$50K
- Gender Gap: Males are 2.8x more likely to earn >$50K

Native Country Disparities (37.1% range)

Highest earners:
- India: 40.0% earn >$50K
- Taiwan: 39.2% earn >$50K
- Japan: 38.7% earn >$50K

Lowest earners:
- Dominican-Republic: 2.9% earn >$50K
- Columbia: 3.4% earn >$50K
- Mexico: 5.1% earn >$50K

Key Findings

1. Native country shows the largest disparities (37.1% range)
2. Gender is the most consistent predictor across all groups
3. Intersectional effects compound disadvantages - Asian-Pac-Islander males (33.5%) vs Other females (4.9%)
4. Systematic inequalities exist across all demographic dimensions

The analysis reveals persistent and significant income disparities that compound when multiple demographic
factors intersect.

> Do capital gains/losses strongly impact the income label?

Yes, capital gains and losses strongly impact the income label.

Key Findings:

Predictive Power

- Capital gains: Having any capital gains increases probability of earning >$50K by 37.8 percentage points (from

24.1% to 61.8%)
- Capital losses: Having any capital losses increases probability by 26.8 percentage points (from 24.1% to 50.9%)

Income Group Differences

- High earners (>$50K): Mean capital gains of $4,006 vs $149 for low earners (26.9x higher)
- High earners (>$50K): Mean capital losses of $195 vs $53 for low earners (3.7x higher)

Prevalence

- Only 8.3% of people have non-zero capital gains
- Only 4.7% of people have non-zero capital losses
- High earners are 5.1x more likely to have capital gains (21.4% vs 4.2%)
- High earners are 3.3x more likely to have capital losses (9.9% vs 3.0%)

Statistical Significance

- Capital gains correlation: 0.22 (medium effect)
- Capital losses correlation: 0.15 (small-medium effect)
- Both highly significant ($p < 0.001$)

Conclusion

Capital gains and losses are excellent predictors of high income, indicating investment activity which strongly
correlates with higher earnings. The binary presence/absence of these features would be particularly valuable in
income prediction models.