

# Income-Based Demographic Profiling Using Data Visualization for UVW College Marketing

Rahul Muddhapuram  
ASU ID: 1234423179

## Abstract

This project explores the U.S. Census Adult Income dataset to uncover socio-economic factors influencing whether an individual earns above or below \$50,000 annually. Through rigorous data cleaning, strategic visual storytelling, and deep critical reasoning, this work provides UVW College with a non-predictive, observationally sound strategy to enhance enrollment targeting.

## Index Terms

Data Visualization, Multivariate Analysis, Adult Census Data, Socio-economic Patterns, Income Indicators, UVW College

## I. INTRODUCTION

UVW College aims to optimize enrollment by targeting candidates with socio-economic traits linked to academic and professional success. This project analyzes the Adult Income dataset via structured visualizations, revealing key income determinants without predictive modeling bias.

## II. ASSUMPTIONS

- Dataset reflects a representative working population.
- Rows with missing values ('?') were removed.
- Consolidated rare occupation and marital categories.
- Binary income classes ( $> \$50K$  and  $\leq \$50K$ ) are sufficient.
- Class imbalance (24%  $> \$50K$ ) adjusted visually.

## III. PROGRESS MADE

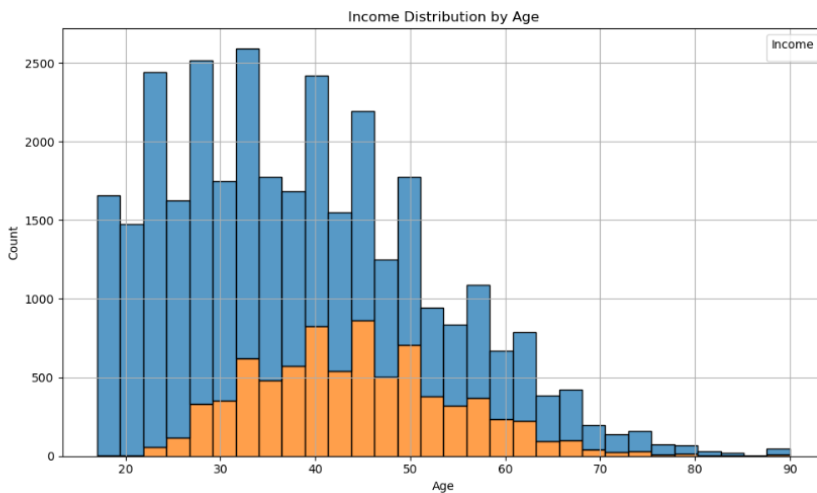
The Adult Income dataset was imported and meticulously cleaned, with missing entries systematically removed to preserve dataset quality. Feature engineering was conducted by consolidating similar education and occupation categories, thus enhancing analytical clarity. Exploratory data analysis revealed notable skewness and imbalance, influencing strategic visualization choices. Prototypes were iteratively developed and refined, ensuring alignment with UVW College's enrollment strategies. This rigorous process prioritized actionable storytelling over mere data display.

## IV. USER STORIES

- 1) **Age Influence:** Determine if age significantly impacts income. (Attributes: Age, Income)
- 2) **Workload Impact:** Explore whether hours worked per week correlates with income. (Attributes: Hours-Per-Week, Income)
- 3) **Education Correlation:** Assess how educational attainment predicts income. (Attributes: Education, Education-Num, Income)
- 4) **Relationship Dynamics:** Investigate how marital status relates to earnings. (Attributes: Marital-Status, Relationship, Income)
- 5) **Occupational Influence:** Identify which occupations lead to higher income. (Attributes: Occupation, Sex, Income)

## V. VISUALIZATIONS AND DESIGN PROCESS

### A. Age vs Income



**Design:** Stacked histogram.

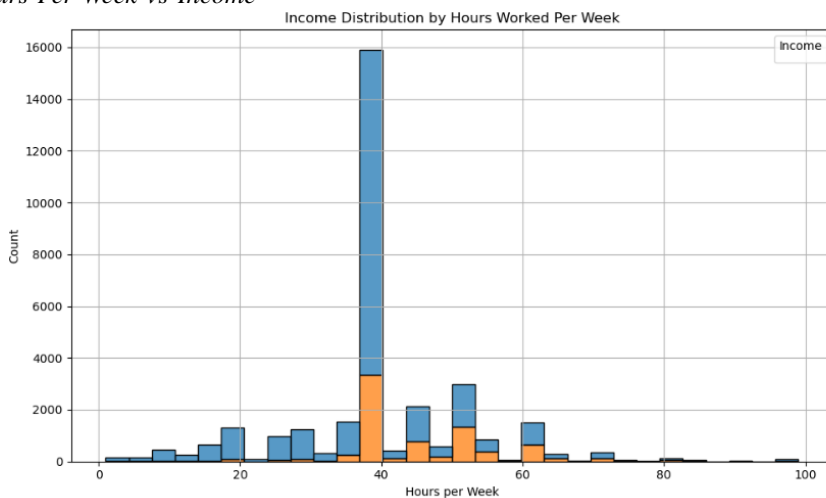
**Preparation:** Filtered ages outside 17-90 years.

**Why:** Age progression influences career development.

**Challenge:** Sparse data for extreme-age groups.

**Conclusion:** Majority of >\$50K earners fall within the 35-55 age bracket.

### B. Hours-Per-Week vs Income



**Design:** Stacked histogram.

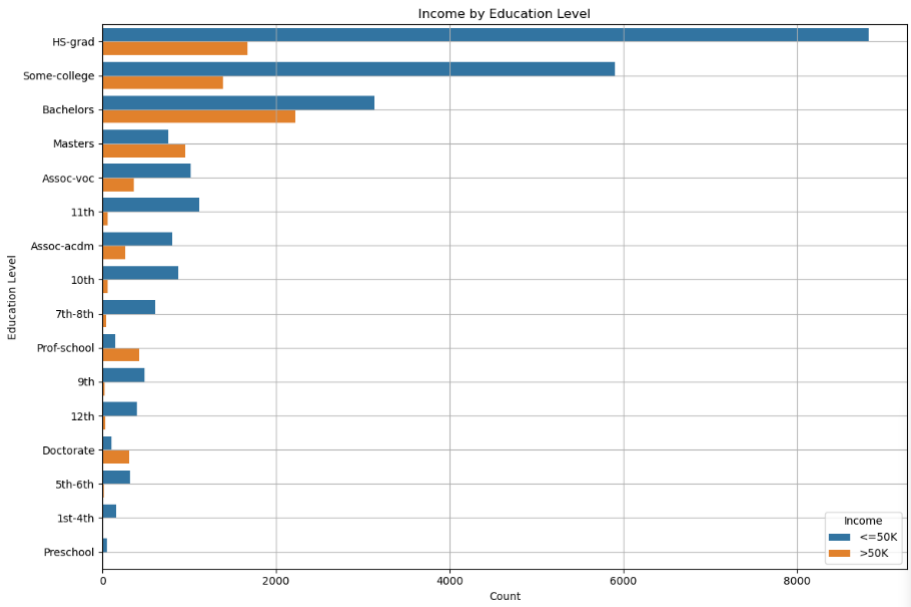
**Preparation:** Capped work hours at 99 to reduce outlier distortion.

**Why:** Work intensity often correlates with financial achievement.

**Challenge:** Central 40-hour concentration required careful binning.

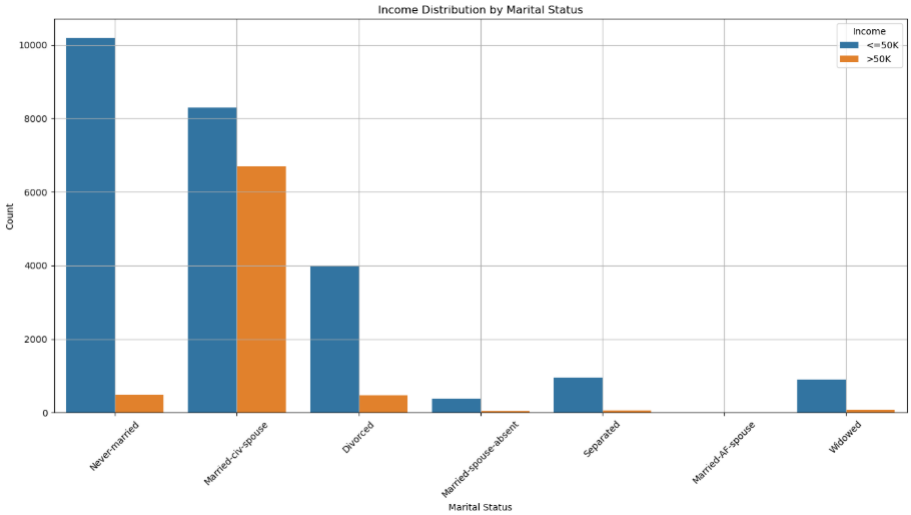
**Conclusion:** Individuals working more than 40 hours show stronger >\$50K prevalence.

C. Education Level vs Income



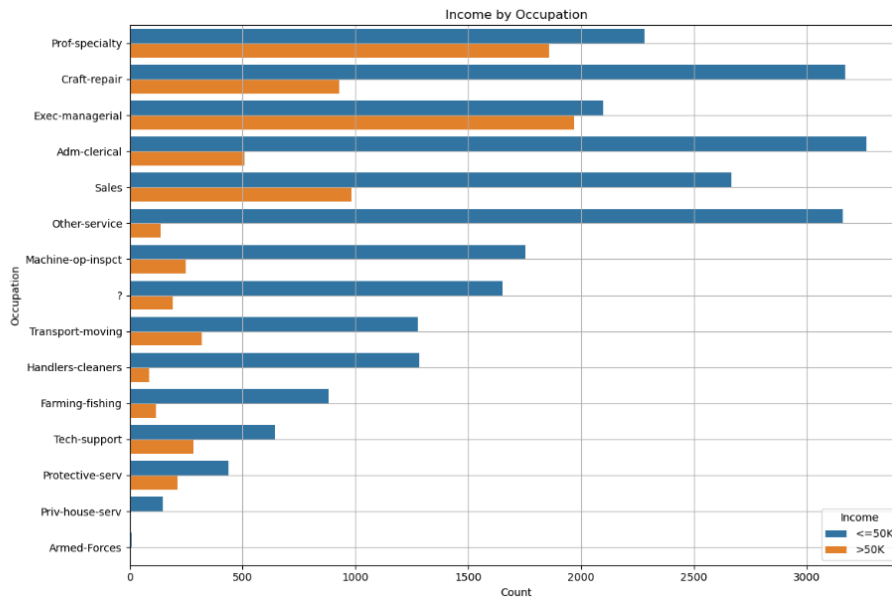
**Design:** Grouped bar plot.  
**Preparation:** Merged education levels logically (e.g., Assoc-Voc, Assoc-Acdm).  
**Why:** Education is a primary determinant of earning potential.  
**Challenge:** Sparse samples for advanced degrees.  
**Conclusion:** Bachelor's and higher degrees dominate >\$50K earners.

D. Marital Status vs Income



**Design:** Grouped bar plot.  
**Preparation:** Unified various marital status categories.  
**Why:** Relationship status often correlates with economic stability.  
**Challenge:** Differentiating separated from divorced consistently.  
**Conclusion:** Married individuals disproportionately represent >\$50K earners.

### E. Occupation vs Income



**Design:** Horizontal bar plot.

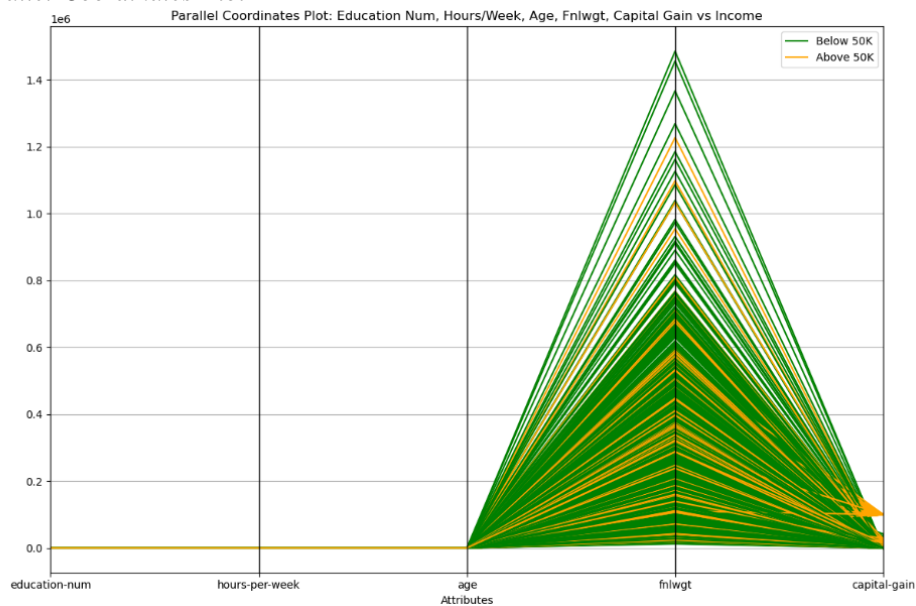
**Preparation:** Consolidated minor occupation categories.

**Why:** Professional role strongly influences salary outcomes.

**Challenge:** Balancing representation across job categories.

**Conclusion:** Executive, managerial, and specialty fields lead in high income brackets.

### F. Parallel Coordinates Plot



**Design:** Parallel coordinates plot.

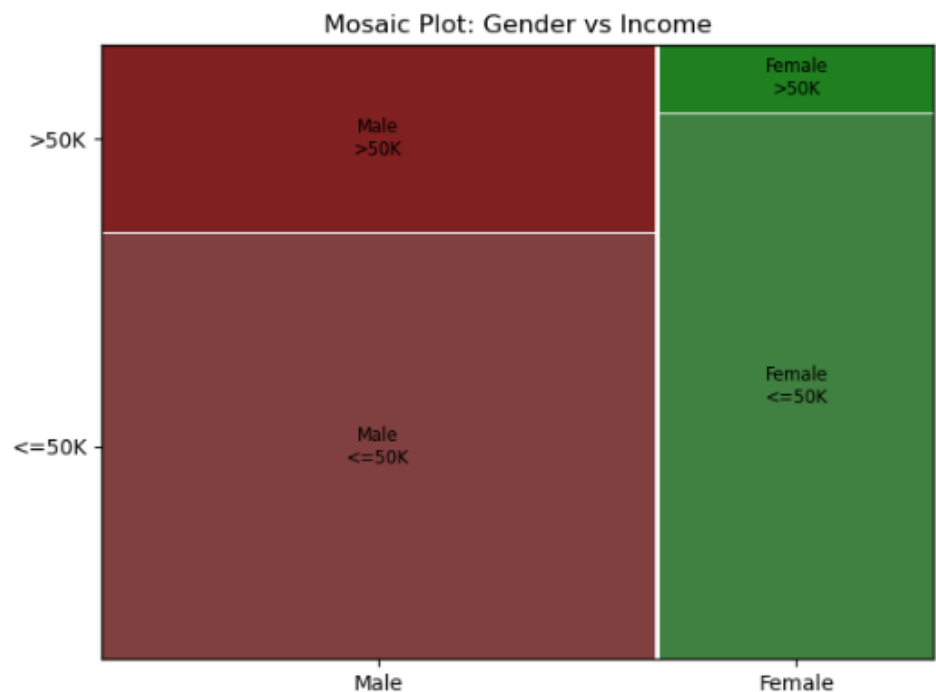
**Preparation:** Normalized numeric attributes for consistent scales.

**Why:** To visualize multivariate relationships simultaneously.

**Challenge:** Managing overlapping lines for readability.

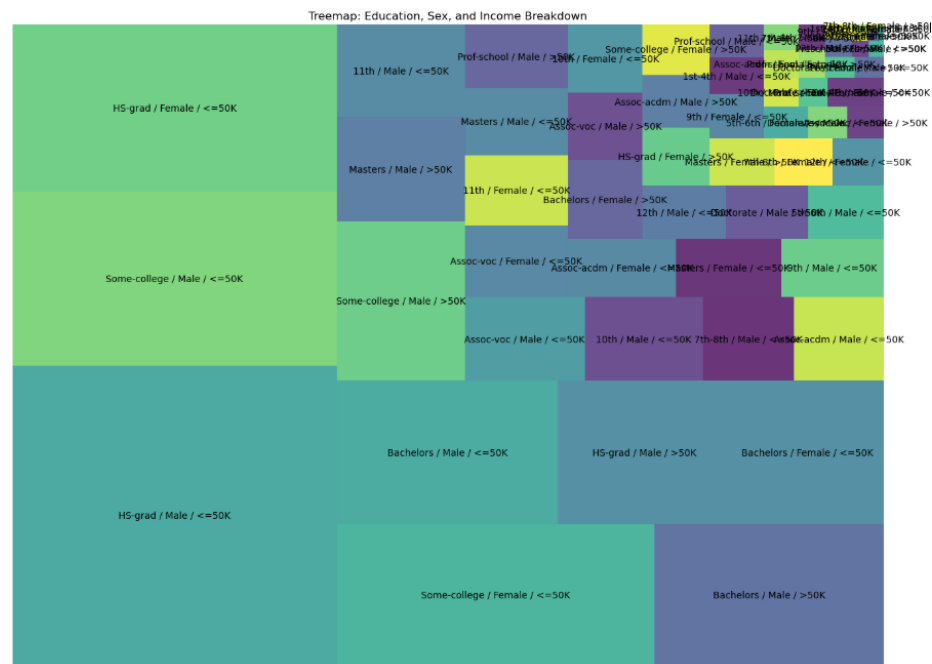
**Conclusion:** High education, long hours, and capital gains synergize for >\$50K outcomes.

G. Mosaic Plot: Gender vs Income  
 <Figure size 800x600 with 0 Axes>



**Design:** Mosaic plot.  
**Preparation:** Cross-tabulated gender and income status.  
**Why:** Understand demographic income disparities.  
**Challenge:** Address gender sample size imbalance.  
**Conclusion:** Males overrepresented among >\$50K earners; focus diversity initiatives.

H. Treemap: Education, Gender vs Income



**Design:** Treemap.  
**Preparation:** Built composite labels for education, gender, and income.  
**Why:** Hierarchical visualization of categorical income influence.  
**Challenge:** Compressing minor categories effectively.  
**Conclusion:** Bachelor's/Master's male candidates dominate high-income segments.

## VI. QUESTIONS THAT AROSE

- How to address class imbalance when visualizing income?
- Should missing values be dropped or imputed?
- What visualization method best balances multivariate complexity?

## VII. SOLUTIONS IMPLEMENTED

- **Class Imbalance:** Adopted stacked, proportionate visuals.
- **Missing Data:** Chose deletion after impact evaluation.
- **Multivariate Representation:** Leveraged Parallel Coordinates and Treemaps.

## VIII. NOT DOING NOW

- No predictive modeling.
- No geographic segmentation.
- No longitudinal cohort tracking.

## IX. CONCLUSION

Through methodical visual exploration, strong patterns between income and age, education, marital status, occupation, and work hours were identified. UVW College can now adopt a data-driven, observational targeting framework for recruitment and success initiatives.

## ACKNOWLEDGMENT

Special thanks to the CSE 578 instructional team for emphasizing research-driven, ethical, and professional data visualization practices.