**Income\_analysis.py**

Income Data Analysis Report

1. Dataset Overview:

- Total samples: 32561

- Features processed: 108

- Training set size: 26048

- Test set size: 6513

2. Correlation Analysis:

- Correlation heatmap saved as 'correlation\_heatmap.png'

- Shows relationships between numeric features

3. Feature Distributions:

- Numeric feature distributions saved as 'numeric\_distributions.png'

- Categorical feature distributions saved as 'categorical\_distributions.png'

- Shows how different features are distributed across income categories

4. Dimensionality Reduction:

- PCA and t-SNE visualizations saved as 'dimensionality\_reduction.png'

- PCA Explained Variance Ratio:

- First component: 0.149

- Second component: 0.107

5. Feature Importance:

- Top 5 most important features:

fnlwgt 0.163778

age 0.152552

capital-gain 0.094236

hours-per-week 0.082519

marital-status\_ Married-civ-spouse 0.068388

- Complete feature importance plot saved as 'feature\_importance.png'

All visualizations have been saved as PNG files in the current directory.