**Title: Adult Census Income Prediction Analysis**

1. **Introduction**  
   The Adult Census Income dataset from Kaggle is used to predict whether an individual's income exceeds $50,000 per year based on census data. This binary classification problem involves analyzing various demographic and employment-related features to determine income levels.
2. **Data Exploration and Preprocessing**

**Dataset Overview:**

* Initial shape: 32,561 rows and 15 columns
* Features include age, workclass, education, marital status, occupation, race, gender, hours per week, and more
* Target variable: income (>50K or <=50K)

**Data Cleaning:**

* Handled missing values represented as '?' in workclass, occupation, and native country columns
* Replaced '?' with mode values: 'Private' for workclass, 'Prof-specialty' for occupation, and 'United-States' for native country
* Removed 24 duplicate rows, resulting in 32,537 unique entries

**Feature Engineering:**

* Simplified education categories (e.g., combined 'Masters', 'Doctorate' into 'graduate')
* Grouped marital status categories (e.g., combined different marriage types into 'married')
* Renamed some columns for clarity (e.g., 'sex' to 'gender', 'fnlwgt' to 'sampleweight')

1. **Exploratory Data Analysis (Insights and Observation)**

* Age distribution shows most people are between 19-50 years old, with a peak around 50
* Most people work 35-40 hours per week
* Majority people earn less than $50K per year
* Outliers observed most in capital gain and loss variables
* Education: Most common are high school graduates and those with some college or associate degrees who have higher counts in both earning less and more than 50 K
* Gender: More males than females across all education levels
* Workclass: Private sector employees are the majority, followed by self-employed and government workers
* Race: White individuals are the most represented group and earned more than black and Asian
* Marital Status: Married individuals are more likely to earn >$50K

1. **Key Findings**

* Higher education levels (undergraduate and above) correlate with higher incomes
* Males are more represented in both income categories, especially in the >$50K group
* Married individuals are more likely to earn >$50K compared to other marital statuses
* Private sector employees have the highest representation in both income categories
* White individuals have the highest representation in both income categories
* Professional and executive occupations are associated with higher incomes