Task:
Demographic Factors and Income Analysis

Overview:
This project aimed to analyze the relationship between various **demographic factors** and **income levels** using a detailed dataset. Key demographic variables such as age, education, occupation, gender, and marital status were examined to provide actionable insights for improving income equity and guiding policy decisions. The project used statistical analysis, hypothesis testing, and data visualization to explore these relationships and make recommendations.

Key Components:

Dataset Overview:

The dataset included demographic, employment, and income-related variables such as:

- **Age**
- **Workclass** (employment sector)
- **Education** (level and years of education)
- **Marital Status**
- **Occupation**
- **Race**
- **Sex**
- **Capital Gain/Loss**
- **Hours per Week**
- **Income** (categorized as above or below $50,000)

Analysis Approach:

- **Data Preprocessing:**
  The data was cleaned to remove missing values, inconsistencies, and outliers. Categorical variables were transformed into numerical codes, and variables were normalized as necessary for analysis.
- **Exploratory Data Analysis (EDA):**
  - **Univariate Analysis:** Summary statistics and visualizations were generated for individual variables such as age, education, and income.
  - **Bivariate Analysis:** Relationships between pairs of variables (e.g., age vs. income, education vs. income) were explored using scatter plots, box plots, and correlation analysis.
  - **Multivariate Analysis:** Cross-tabulations, heatmaps, and pair plots were used to understand how multiple factors jointly influence income.

- **Hypothesis Testing:**
  Statistical tests (t-tests, ANOVA) were conducted to test hypotheses such as the effect of education on income and marital status on work hours. Regression analysis was used to model the impact of multiple factors on income.

Key Insights:

- **Age vs. Income:**
  Employees aged 35-55 in federal government positions worked the most hours (48.82 hours/week) and showed higher income levels. This age group was found to be in peak earning years, benefiting from experience and promotions.
- **Capital Gain by Country:**
  Canada led in capital gains, followed by the Dominican Republic, reflecting robust financial markets and favorable investment conditions in these regions.
- **Net Capital vs. Occupation:**
  Professional specialties (e.g., doctors, engineers) and executive managers showed the highest net capital gains, benefiting from additional income streams like bonuses and consulting.
- **Hours Per Week vs. Marital Status:**
  Married males worked the most hours (56.56 hours/week), likely driven by additional financial responsibilities and societal expectations, whereas unmarried individuals worked fewer hours on average.

Recommendations:

- **Short-Term Recommendations:**
  - **Targeted Training Programs:** Develop training programs to upskill employees in high net capital gain professions such as professional specialties and executive management.
  - **Work-Life Balance Initiatives:** Encourage employers, especially in government, to promote work-life balance through flexible work arrangements.
- **Long-Term Recommendations:**
  - **Economic Policies for Capital Growth:** Implement policies to support capital growth across a broader range of occupations, reducing income disparities.
  - **Equal Pay Policies:** Advocate for equal pay for similar work across demographic groups by conducting wage audits and enforcing equal reward structures.

Skills Utilized:

- **Data Cleaning** (Preprocessing, Handling Outliers)
- **Exploratory Data Analysis** (Univariate, Bivariate, Multivariate)
- **Hypothesis Testing** (T-tests, ANOVA, Regression Analysis)
- **Data Visualization** (Bar Charts, Stacked Bar, Map, Area chart)

- **Statistical Modelling** (Income Predictions, Correlation Analysis)

References:

https://public.tableau.com/views/S24_Group-6__INFO8677_Tableau_Group_Project_docx/Dashboard1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link