US Income Data Analysis using SQL

* By Devanshi Shah

Problem Statement

**Executive Summary:**  
The US Income Data Analysis project aimed to explore trends and changes in household income in the United States. The goal was to clean and analyze the dataset containing income data for US households across various years. The project involved two primary steps: first, cleaning the dataset by removing duplicates and ensuring data integrity; second, conducting an exploratory data analysis to identify income distribution patterns, changes over time, and potential factors contributing to shifts in income levels. The findings from this analysis were intended to offer insights into economic disparities, identify trends in income inequality, and provide recommendations for addressing economic challenges at the national level.  
  
**Data Description:** The dataset used for this analysis was sourced from national income surveys, which includes data for households across the United States over several years. The dataset features columns such as State, Year, Household Income, and various demographic variables. Data cleaning steps ensured that any duplicate entries were removed and that erroneous data points were corrected. The key data columns are:

* **State**: The state where the household income data was collected.

## Year: The year in which the data was recorded.

## Household\_Income: The median or average household income for that year in the specific state.

## Additional demographic and economic indicators.

## The analysis focused on identifying income trends across different states and years, analyzing the impact of factors such as state policies, economic growth, and inflation on household income levels.

**Findings and Recommendations:**

Based on data cleaning and exploratory data analysis, the following findings and recommendations were made:

1. **Data Integrity:** It was critical to remove duplicate entries from the dataset to ensure that the analysis was accurate. The cleaning process identified and eliminated duplicates based on combinations of State and Year, as well as eliminating erroneous or inconsistent data points.
2. **Income Trends:** The analysis revealed varying trends in household income across different states and years. Nationally, income growth showed fluctuations depending on economic factors, with some regions experiencing faster growth than others. The analysis highlighted how income disparities exist between states and within different socioeconomic groups.
3. **State-Specific Variations:** The analysis found that states with higher income levels often exhibited stronger economic conditions, such as lower unemployment rates, higher educational attainment, and more robust job markets. These states showed a steadier upward trend in household income over time, while states with lower growth faced challenges like economic downturns and a higher cost of living.
4. **Economic and Social Impacts:** The project also pointed to a growing concern with income inequality in certain regions, particularly in urban vs. rural areas. A widening gap between the highest and lowest income households was observed. It is recommended that policies be developed to address income disparities through investments in education, job training, and healthcare.
5. **Further Analysis:** To gain deeper insights, further analysis could be conducted to correlate income trends with factors such as education levels, healthcare access, inflation rates, and social policies. This will provide actionable insights for policymakers aiming to reduce economic inequality and improve overall household income in the country.