# Income Prediction

**Group Members:** Muhammad Ahmad 23L-2511

Muhammad Daniyal 23L-2600

Muhammad Sufyan 23L-2518

# Dataset Description Document

## 1. Problem Statement

The dataset aims to predict whether an individual's income exceeds $50,000 per year based on various demographic factors. This problem has applications in fields like social sciences, economics, and policy-making, helping to understand the socioeconomic status of different groups.

## 2. Source of the Dataset

This dataset, known as **'Census Income**,' dataset, originates from the 1994 Census database. It is available on **Kaggle** (<https://www.kaggle.com/datasets/wenruliu/adult-income-dataset?resource=download>).

## 3. Brief Description of the Dataset

The dataset contains demographic information about individuals and includes both categorical and numerical features. Each row represents a different individual, with various attributes describing their demographics, work type, and income level.

## 4. Description of the Attributes/Variables

* **age:** Age of the individual (numerical).
* **workclass:** The type of work the individual is involved in (categorical: e.g., Private, Self-emp-not-inc, etc.).
* **fnlwgt:** Final weight, representing the number of people the census believes the entry represents (numerical).
* **education:** Highest level of education achieved by the individual (categorical: e.g., Bachelors, HS-grad, etc.).
* **education-num:** Numerical representation of education level (numerical).
* **marital-status:** Marital status of the individual (categorical: e.g., Married-civ-spouse, Never-married, etc.).
* **occupation:** The individual’s occupation (categorical: e.g., Tech-support, Craft-repair, etc.).
* **relationship:** Relationship status within the family (categorical: e.g., Husband, Wife, etc.).
* **race:** Race of the individual (categorical: e.g., White, Asian-Pac-Islander, etc.).
* **gender:** Gender of the individual (categorical: Male, Female).
* **capital-gain:** Capital gain during the year (numerical).
* **capital-loss:** Capital loss during the year (numerical).
* **hours-per-week:** Number of hours the individual works per week (numerical).
* **native-country:** Country of origin (categorical: e.g., United-States, Mexico, etc.).
* **income:** Income level (binary classification: '>50K' or '<=50K').