Assignment 2:

Group Members:
20L-1122 Muhammad Hassan Jawwad
20L-0935 AbdulRahman Nadeem
20L-0955 Nawal Fatima

DATA DESCRIPTION:

Problem Statement:

The project aims to develop a web application for predicting an individual's income level based on their personal information. This prediction will be categorized into two classes: "<=50K" and ">50K." The prediction is influenced by a set of demographic and other features, including education level, age, gender, occupation, and more.

Source of the Dataset:

The dataset used in this project, known as the "Adult Income Dataset," was sourced from kaggle (https://www.kaggle.com/datasets/wenruliu/adult-income-dataset) and originally it is from UCI Machine learning repository. This dataset contains a total of 16 columns, including one for the target variable "Income" and 14 attributes that describe an individual's personal information.

Brief Description about the Dataset:

The "Adult Income Dataset" is a collection of data that represents individuals and their annual income levels. The dataset is structured to explore the feasibility of predicting income levels based on personal attributes. Each record in the dataset corresponds to an individual and includes features such as education level, age, gender, occupation, and more. The income levels are categorized into two classes: "<=50K" (indicating an annual income of less than or equal to $50,000) and ">50K" (indicating an annual income greater than $50,000).

Description of the Attributes/Variables/Columns of the Dataset:

The dataset consists of 16 columns, including the target variable "Income" and the following 14 attributes or features:

1. Age:Age of the individual.
2. Workclass:Type of work the individual is engaged in (e.g., private, self-employed, etc.).
3. Education:Highest level of education completed.
4. Education-Num:Numeric representation of education level.
5. Marital Status: Marital status of the individual.
6. Occupation:Occupation or job type.
7. Relationship:Relationship status (e.g., husband, wife, etc.).
8. Race:Ethnicity or race of the individual.
9. Gender:Gender of the individual.
10. Capital Gain: Capital gains reported by the individual.
11. Capital Loss:Capital losses reported by the individual.
12. Hours per Week:Average number of hours worked per week.
13. Native Country: Country of origin.
14. Income:Target variable indicating income level ("<=" or ">" $50K).

These attributes will be used as input features for predicting the income level of an individual using machine learning models.