# **Dataset Overview :**For this task, we’ll use the **Adult Census Income Dataset**, which contains various demographic attributes like age, education level, occupation, etc., and a target label indicating whether an individual earns more than $50K annually.

# The dataset includes the following features:

1. Age
2. Work class
3. Education
4. Marital status
5. Occupation
6. Race
7. Sex
8. Hours per week
9. Native country
10. Salary (target variable: >50K or <=50K)

**Step 1: Loading and Exploring the Data:**

Import pandas' library ,load and read the dataset and explore the data using describe, info and columns function

# **Step 2: Analyze Data through Visualization:**

Check relationship between data to each other, check percentage of some value in some column and checking factor that affecting to salary in visualization.

Import mataplotlib.pyplot and seaborn library. Plot the data using pie chart, column chart, scatterplot, histogram etc.

# **Step 3: Data Preprocessing**

Machine learning models often perform better when the data is cleaned and preprocessed. We’ll take the following steps:

**Handle Missing Data**: If any missing values exist, impute or remove them. Use different pandas function like isnull, fillna and remove.

**Encode Categorical Features**: Many machine learning algorithms don’t work with categorical data directly, so we need to convert these into numerical representations. For this import Labelencoder library and use fit\_transform() function of Labelencoder.

**Feature Scaling**: Normalize or standardize numerical features. For this import StandardScaler library and use fit\_transform() function of StandardScaler. Before Scaling the data, split it into train and test set using train\_test\_split library and scale the train set.

# **Step 4: Model Selection and Training**

# Experiment with models such as Logistic Regression or Support Vector Machines (SVM). For this import sklearn library.

# **Step 5: Model Evaluation**

To evaluate the performance of our model, we’ll use common classification metrics such as accuracy, precision, recall, and F1-score.

# **Conclusion**

We successfully built a machine learning model using the Adult Census Income dataset to predict whether an individual’s salary exceeds $50,000 per year. We concluded that Random Forest give better result than logistic regression ,SVM ,Decision Tree Classifier and Kneighbour Classifier.