Adult Income Classification - Major Project

# GitHub Link :-

# Step 1: Import Required Libraries

Importing essential Python libraries for data manipulation, visualization, and machine learning.

# Step 2: Load and Inspect Dataset

Loading the dataset using pandas and inspecting its structure, null values, and data types.

# Step 3: Data Cleaning and Preprocessing

Cleaning the dataset by removing whitespaces, handling missing values, and encoding categorical columns.

# Step 4: Feature Engineering

Separating features and target variable. Standardizing the data for optimal model performance.

# Step 5: Splitting the Data

Splitting the dataset into training and testing sets using an 80-20 split.

# Step 6: Data Visualization

Visualizing the data through 8 different plots using Seaborn and Plotly to understand feature distribution and relationships.

# Step 7: Model Building and Evaluation

Training and evaluating four classification models: Logistic Regression, Decision Tree, Random Forest, and K-Nearest Neighbors.

# Step 8: Summary and Conclusion

This machine learning project analyzed the UCI Adult Income dataset.  
It involved data cleaning, feature engineering, training ML models, and visualization.   
Key insights showed that Random Forest and Logistic Regression performed best. Future improvements include hyperparameter tuning, cross-validation, and exploring ensemble methods like XGBoost and LightGBM.