

Assignment 1: Data Processing with Pandas & Matplotlib

Overview

This assignment focuses on data preprocessing and visualization using Pandas and Matplotlib libraries with the Adult Income dataset.

Score: 95/100 ★

Objectives

- Process the Adult Income dataset using Pandas
- Handle missing values appropriately
- Visualize data distributions using Matplotlib

Tasks

Step 1: Read Adult Income Dataset

Load the Adult Income dataset into a Pandas DataFrame.

Step 2: Identify Missing Values

Print all rows containing missing values.

- **Expected output:** 3,620 rows with missing values

Step 3: Remove Missing Values

Clean the dataset by removing rows with missing values.

Step 4: Verify Data Shape

Print the shape of the cleaned dataset.

- **Expected shape:** (45,222, 15)

Step 5: Visualize Age Distribution

Plot the distribution of the age feature using Matplotlib.

Dataset

The Adult Income dataset (also known as Census Income dataset) contains demographic information to predict whether income exceeds \$50K/year.

Requirements

```
python  
pandas  
matplotlib
```

Installation

```
bash  
pip install pandas matplotlib
```

Usage

Run the Jupyter notebook or Python script to execute all steps sequentially.

Results

Successfully processed the dataset with proper handling of missing values and created visualization of age distribution.

Assignment completed as part of Deep Learning coursework