

## **“PYTHON DATA ANALYTICS: INVESTIGATING ANNUAL INCOME”**

I utilized Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn to manipulate, clean, transform, visualize, and extract meaningful insights from the ‘adult.csv’ dataset sourced from Kaggle, the world’s largest data science community.

### **1. Data Exploration and Coding Operations:**

- Performed operations like head, tail, info, describe, and shape to understand the data structure.
- Explored the dataset to gain insights into the type of data being worked with.

### **2. Data Cleaning:**

- Replaced missing values (denoted by ‘?’) with NaN.
- Removed rows with missing data.
- Eliminated duplicated records.

### **3. Age Distribution:**

- 75% of the population falls within the age range of 17 to 47 years.
- The minimum age is 17, and the maximum age is 90.
- Total individuals aged 17 to 48: 34,858.

### **4. Workclass Categories:**

- There are 7 unique workclass categories:
  - Private
  - Local government
  - Self-employed (not incorporated)
  - Federal government
  - State government
  - Self-employed (incorporated)
  - Without pay
- Majority work in the private sector, followed by self-employed (not incorporated).

### **5. Education Levels:**

- The total number of people with bachelor’s and master’s degrees: 10,072.

### **6. Salary by Workclass:**

- Self-employed (incorporated) and federal government workers have the highest salaries.

### **7. Gender and Salary:**

- Males have a better chance of earning > \$50k annually.