**National Institute of Technology Warangal**

**Lab 7 Assignment**

**Name: Ayush Rana**

**Roll Number: 24CSM2R05 Course Title: Data Privacy**

**Department: Computer Science and Engineering**

**Program: M. Tech in Computer Science and Information Security**

**Semester: 2**

# Assumptions:

# ****Sensitivity:**** We assume that the sensitivity Δf for the query (here, the mean of the 'income' attribute) can be approximated by the range of income values.

* **Epsilon (ε):** We set the privacy budget to 1.0. A lower ϵ would yield stronger privacy (more noise) but lower accuracy.

# ****No Built-in Laplace Function**:** Instead of using np.random.laplace, we implement our own noise generation using the inverse transform sampling method.

# Methodology:

# Laplace Noise Generation:

# 

# ****Application of DP Mechanism:**** We add the generated Laplace noise to each income value to create a differentially private version of the income attribute.

* **Empirical Simulation:**  
  The function simulate\_dp\_income\_custom runs multiple trials (1000) to compute the DP mean for the full dataset and for a neighboring dataset (one record removed). The near-identical distributions of the DP means demonstrate that the mechanism is robust to small changes, fulfilling the differential privacy guarantee.

**Proof of Protection:**

* The theoretical guarantee of differential privacy ensures that for any two neighboring datasets, the probability distributions of the outputs are nearly identical.
* The simulation shows that the DP means from the full and neighboring datasets are almost the same, which empirically confirms that the influence of any single record is effectively masked by the added noise.

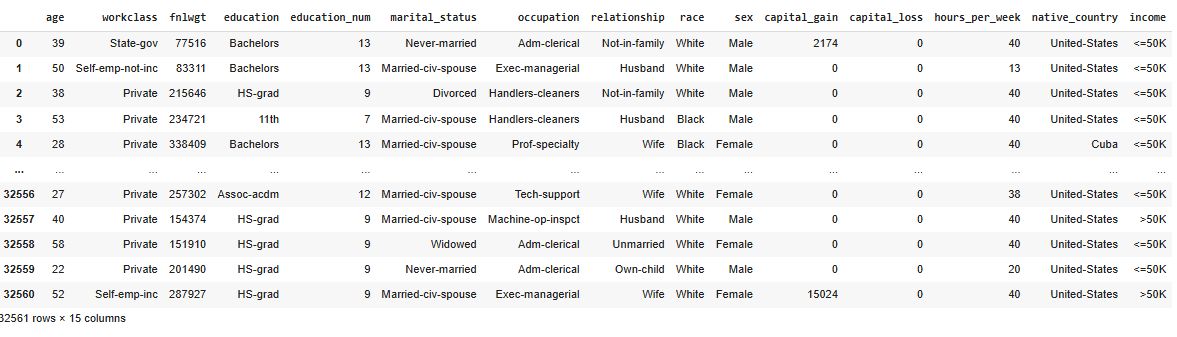
# Initial dataset state:

**colums:**

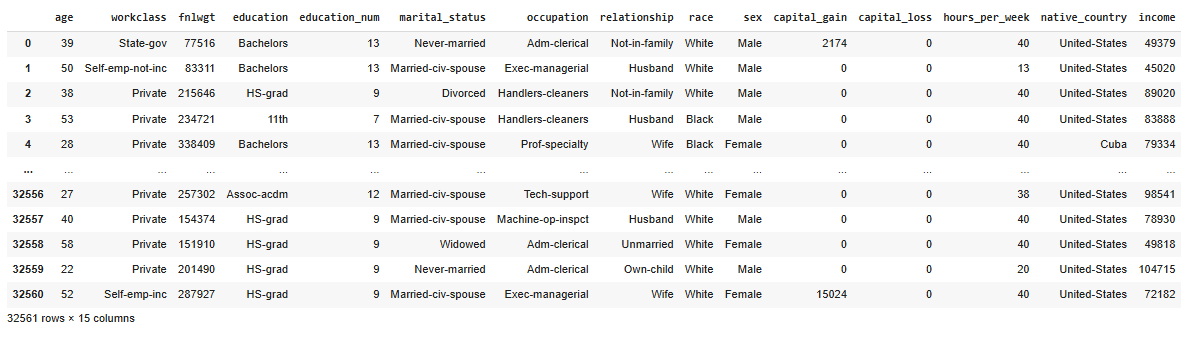
['age', 'workclass', 'fnlwgt', 'education', 'education\_num', 'marital\_status', 'occupation', 'relationship', 'race', 'sex',

'capital\_gain', 'capital\_loss', 'hours\_per\_week', 'native\_country', 'income', 'age\_generalized', 'education\_generalized',

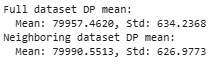
'occupation\_generalized', 'workclass\_generalized', 'marital\_generalized', 'native\_generalized', 'suppressed'], dtype='object')

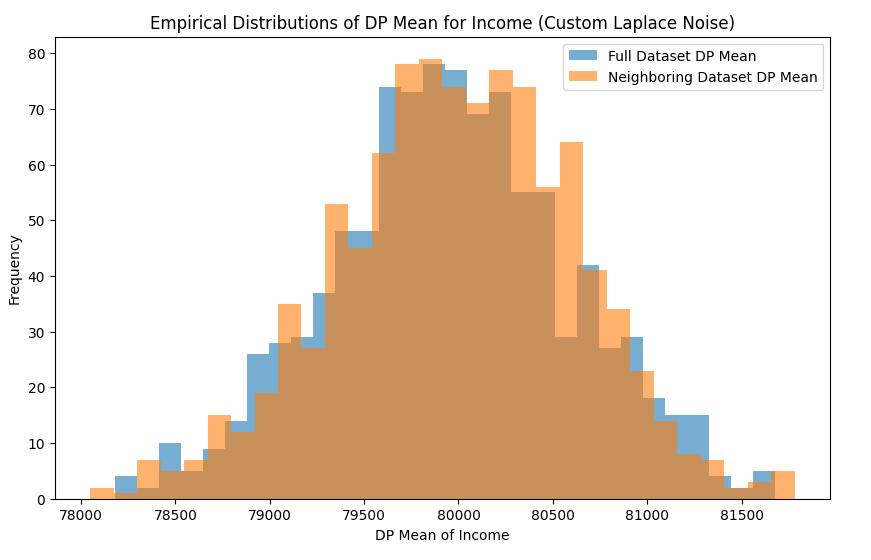


**Modified data:**



# statistical analysis:

****

****