**National Institute of Technology Warangal**

**Lab 5 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**

# Initial dataset state:

**colums:**

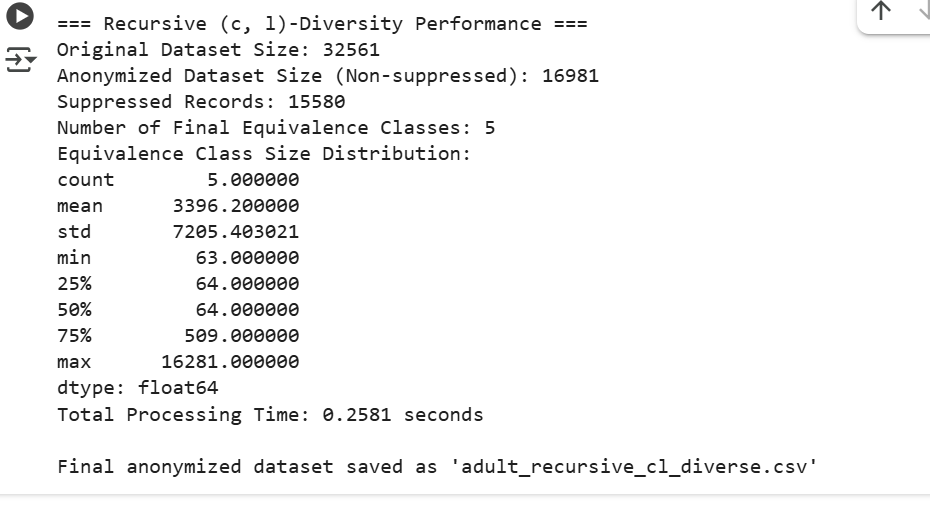
['age', 'workclass', 'fnlwgt', 'education', 'education\_num', 'marital\_status', 'occupation', 'relationship', 'race', 'sex',

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

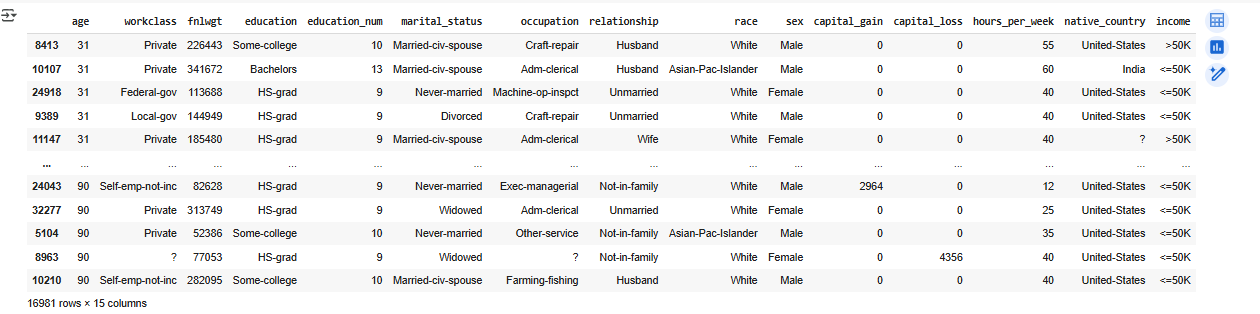
'occupation\_generalized', 'workclass\_generalized', 'marital\_generalized', 'native\_generalized', 'suppressed'], dtype='object')



# Final modified data (statistical analysis):

****

**Output COLUMS**:



**we assume that the :**

**sensitive attribute** = income.

**splitting attribute** = age.