National Institute of Technology

Warangal

Lab 4 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:

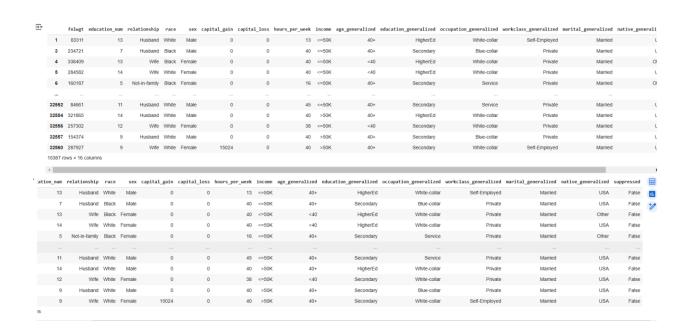
```
['age', 'workclass', 'fnlwgt', 'education', 'education_num',
    'marital_status', 'occupation', 'relationship', 'race', 'sex',
    'capital_gain', 'capital_loss', 'hours_per_week', 'native_country',
    'income', 'age_generalized', 'education_generalized',
    'occupation_generalized', 'workclass_generalized',
    'marital_generalized', 'native_generalized', 'suppressed'],
    dtype='object')
```

| | age | workclass | fnlwgt | education | education_num | marital_status | occupation | relationship | race | sex | capital_gain | capital_loss | hours_per_week | native_country | income |
|----------|------|------------------|--------|------------|---------------|--------------------|-------------------|---------------|-------|--------|--------------|--------------|----------------|----------------|--------|
| 0 | 39 | State-gov | 77516 | Bachelors | 13 | Never-married | Adm-clerical | Not-in-family | White | Male | 2174 | 0 | 40 | United-States | <=50K |
| 1 | 50 | Self-emp-not-inc | 83311 | Bachelors | 13 | Married-civ-spouse | Exec-managerial | Husband | White | Male | 0 | 0 | 13 | United-States | <=50K |
| 2 | 38 | Private | 215646 | HS-grad | 9 | Divorced | Handlers-cleaners | Not-in-family | White | Male | 0 | 0 | 40 | United-States | <=50K |
| 3 | 53 | Private | 234721 | 11th | 7 | Married-civ-spouse | Handlers-cleaners | Husband | Black | Male | 0 | 0 | 40 | United-States | <=50K |
| 4 | 28 | Private | 338409 | Bachelors | 13 | Married-civ-spouse | Prof-specialty | Wife | Black | Female | 0 | 0 | 40 | Cuba | <=50K |
| | | | | | | | | | | | | | | | |
| 32556 | 27 | Private | 257302 | Assoc-acdm | 12 | Married-civ-spouse | Tech-support | Wife | White | Female | 0 | 0 | 38 | United-States | <=50K |
| 32557 | 40 | Private | 154374 | HS-grad | 9 | Married-civ-spouse | Machine-op-inspct | Husband | White | Male | 0 | 0 | 40 | United-States | >50K |
| 32558 | 58 | Private | 151910 | HS-grad | 9 | Widowed | Adm-clerical | Unmarried | White | Female | 0 | 0 | 40 | United-States | <=50K |
| 32559 | 22 | Private | 201490 | HS-grad | 9 | Never-married | Adm-clerical | Own-child | White | Male | 0 | 0 | 20 | United-States | <=50K |
| 32560 | 52 | Self-emp-inc | 287927 | HS-grad | 9 | Married-civ-spouse | Exec-managerial | Wife | White | Female | 15024 | 0 | 40 | United-States | >50K |
| 32561 rd | ws × | 15 columns | | | | | | | | | | | | | |

Final modified data:

```
⇒ === LKC-Privacy Statistical Information ===
    Original Dataset Size: 32561
    Anonymized Dataset Size (Non-suppressed): 15387
    Suppressed Records: 17174
    Equivalence Class Size Distribution (using all generalized QIDs):
    1
            11
            11
    3
    4
            6
    5
             4
    620
    634
    737
    1770
    Name: count, Length: 82, dtype: int64
    LKC-Privacy Parameters:
      L (max adversary known QIDs): 1
      K (min group size): 2
      C (max sensitive confidence): 0.9
```

Output COLUMS:



quasi identifiers= ["age","education","occupation","workclass","marital_status","native_country"]
generalized_quasi_identifiers=["age_generalized","education_generalized","occupation_generalize
d","workclass_generalized","marital_generalized","native_generalized]

sensitive attributes: ["income"]