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')

	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 rows × 15 columns

Final modified data (statistical analysis):

```
=== Recursive (c, l)-Diversity Performance ===
Original Dataset Size: 32561
Anonymized Dataset Size (Non-suppressed): 16981
Suppressed Records: 15580
Number of Final Equivalence Classes: 5
Equivalence Class Size Distribution:
count          5.000000
mean        3396.200000
std         7205.403021
min           63.000000
25%           64.000000
50%           64.000000
75%          509.000000
max        16281.000000
dtype: float64
Total Processing Time: 0.2581 seconds

Final anonymized dataset saved as 'adult_recursive_cl_diverse.csv'
```

Output COLUMS:

	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	hours_per_week	native_country	income
8413	31	Private	226443	Some-college	10	Married-civ-spouse	Craft-repair	Husband	White	Male	0	0	55	United-States	>50K
10107	31	Private	341672	Bachelors	13	Married-civ-spouse	Adm-clerical	Husband	Asian-Pac-Islander	Male	0	0	60	India	<=50K
24918	31	Federal-gov	113688	HS-grad	9	Never-married	Machine-op-inspct	Unmarried	White	Female	0	0	40	United-States	<=50K
9389	31	Local-gov	144949	HS-grad	9	Divorced	Craft-repair	Unmarried	White	Male	0	0	40	United-States	<=50K
11147	31	Private	185480	HS-grad	9	Married-civ-spouse	Adm-clerical	Wife	White	Female	0	0	40	?	>50K
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
24043	90	Self-emp-not-inc	82628	HS-grad	9	Never-married	Exec-managerial	Not-in-family	White	Male	2964	0	12	United-States	<=50K
32277	90	Private	313749	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0	25	United-States	<=50K
5104	90	Private	52386	Some-college	10	Never-married	Other-service	Not-in-family	Asian-Pac-Islander	Male	0	0	35	United-States	<=50K
8963	90	?	77053	HS-grad	9	Widowed	?	Not-in-family	White	Female	0	4356	40	United-States	<=50K
10210	90	Self-emp-not-inc	282095	Some-college	10	Married-civ-spouse	Farming-fishing	Husband	White	Male	0	0	40	United-States	<=50K

16981 rows × 15 columns

we assume that the :

sensitive attribute = income.

splitting attribute = age.