

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:

columns:

```
['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 x 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      23  
2      11  
3      11  
4       6  
5       4  
  
..  
620    1  
634    1  
737    1  
1770   1  
1850   1  
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 COLUMNS:

```
Index(['fnlwgt', 'education_num', 'relationship', 'race', 'sex',  
      'capital_gain', 'capital_loss', 'hours_per_week', 'income',  
      'age_generalized', 'education_generalized', 'occupation_generalized',  
      'workclass_generalized', 'marital_generalized', 'native_generalized',  
      'suppressed'],  
      dtype='object')
```

	fnlwgt	education_num	relationship	race	sex	capital_gain	capital_loss	hours_per_week	income	age_generalized	education_generalized	occupation_generalized	workclass_generalized	marital_generalized	native_generalized	suppressed
1	83311	13	Husband	White	Male	0	0	13	<=50K	40+	HigherEd	White-collar	Self-Employed	Married	USA	False
3	234721	7	Husband	Black	Male	0	0	40	<=50K	40+	Secondary	Blue-collar	Private	Married	USA	False
4	338409	13	Wife	Black	Female	0	0	40	<=50K	<40	HigherEd	White-collar	Private	Married	Other	False
5	284582	14	Wife	White	Female	0	0	40	<=50K	<40	HigherEd	White-collar	Private	Married	USA	False
6	160187	5	Not-in-family	Black	Female	0	0	16	<=50K	40+	Secondary	Service	Private	Married	Other	False
...
32552	84661	11	Husband	White	Male	0	0	45	<=50K	40+	Secondary	Service	Private	Married	USA	False
32554	321865	14	Husband	White	Male	0	0	40	>50K	40+	HigherEd	White-collar	Private	Married	USA	False
32556	257302	12	Wife	White	Female	0	0	38	<=50K	<40	Secondary	White-collar	Private	Married	USA	False
32557	154374	9	Husband	White	Male	0	0	40	>50K	40+	Secondary	Blue-collar	Private	Married	USA	False
32560	287927	9	Wife	White	Female	15024	0	40	>50K	40+	Secondary	White-collar	Self-Employed	Married	USA	False

15387 rows x 16 columns

education_num	relationship	race	sex	capital_gain	capital_loss	hours_per_week	income	age_generalized	education_generalized	occupation_generalized	workclass_generalized	marital_generalized	native_generalized	suppressed
13	Husband	White	Male	0	0	13	<=50K	40+	HigherEd	White-collar	Self-Employed	Married	USA	False
7	Husband	Black	Male	0	0	40	<=50K	40+	Secondary	Blue-collar	Private	Married	USA	False
13	Wife	Black	Female	0	0	40	<=50K	<40	HigherEd	White-collar	Private	Married	Other	False
14	Wife	White	Female	0	0	40	<=50K	<40	HigherEd	White-collar	Private	Married	USA	False
5	Not-in-family	Black	Female	0	0	16	<=50K	40+	Secondary	Service	Private	Married	Other	False
...
11	Husband	White	Male	0	0	45	<=50K	40+	Secondary	Service	Private	Married	USA	False
14	Husband	White	Male	0	0	40	>50K	40+	HigherEd	White-collar	Private	Married	USA	False
12	Wife	White	Female	0	0	38	<=50K	<40	Secondary	White-collar	Private	Married	USA	False
9	Husband	White	Male	0	0	40	>50K	40+	Secondary	Blue-collar	Private	Married	USA	False
9	Wife	White	Female	15024	0	40	>50K	40+	Secondary	White-collar	Self-Employed	Married	USA	False

quasi identifiers= ["age","education","occupation","workclass","marital_status","native_country"]

generalized_quasi_identifiers=["age_generalized","education_generalized","occupation_generalized","workclass_generalized","marital_generalized","native_generalized"]

sensitive attributes : ["income"]