National Institute of Technology Warangal

Lab 2 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

1. Implement (X,Y) -Anonymity for the adult data set via code.

ort	nump	y as np	essing in	nport KBins	Discretizer								⊡ ↑ ↓ ±	무 î
mns "ag "oc "ca	= [ge", cupa pita	"workclass tion", "re l_loss", "	s", "fnlo elationsk 'hours_pe	nip", "race er_week", "	", "sex", "capi native_country'	tal_gain", ', "income")					
1														
	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	hours_per_week	native_c
0	39	State-gov	77516	Bachelors	13		Adm- clerical	Not-in- family	White	Male	2174	0	40	United
1	50	Self-emp- not-inc	83311	Bachelors	13	Married-civ- spouse	managerial		White	Male	0	0	13	United
2	38	Private	215646	HS-grad	9	Divorced	cleaners	Not-in- family	White	Male	0	0	40	United
3	53	Private	234721	11th	7	spouse	cleaners	Husband	Black	Male	0	0	40	United
4	28	Private	338409	Bachelors	13	spouse	specialty	Wife		Female	0	0	40	
	-						-							
0			77516	Bachelors	13		clerical	family	White	Male	2174	0	40	United
1	50	Self-emp- not-inc	83311	Bachelors	13	Married-civ- spouse	Exec- managerial		White	Male	0	0	13	United
2	38	Private	215646	HS-grad	9	Divorced	cleaners	Not-in- family	White	Male	0	0	40	United
3	53	Private	234721	11th	7	spouse	cleaners	Husband	Black	Male	0	0	40	Unite
4	28	Private	338409	Bachelors	13	Married-civ- spouse	Prof- specialty	Wife	Black	Female	0	0	40	
				Δssnc-		 Married-civ-	 Tech-	***			***			
				acdm		spouse	support							United
7	-			_		spouse	op-inspct						-	United
8				-			clerical							United
9	22		201490	HS-grad	9	Never-married	clerical	Own-child	White	Male	0	0	20	United
0	52	Self-emp- inc	287927	HS-grad	Q	Married-civ-	Exec-	Wife	Mhito	Famala	15024	0	40	United
	rt sk	rt nump; sklear rep 1: Li mms = [1: Li mms = [1: Li mms = [1: Li ms	### ### ### ### ### ### ### ### ### ##	rt numpy as np sklearn.preprocessing in specific in state and state age workclass", "fnlu" "cacupation", "relations! "age workclass fnlugt 1 50 Self-emp- not-inc 1 50 Self-emp- not-inc 2 38 Private 234721 2 8 Private 234721 3 8 Private 234721 4 1 2 8 Private 234721 5 1 50 Self-emp- not-inc 5 27 Private 257302 7 40 Private 154374 8 58 Private 201490 9 22 Private 201490 9 55 Self-emp- SS Se	rt numpy as np sklearn.preprocessing import KBins ep 1: Load the dataset mms = ["age", "workclass", "fnlwgt", "educ "occupation", "relationship", "race "capital_loss", "hours_per_week", " = pd.read_csv("adult.data", names= age workclass fnlwgt education 0 39 State-gov 77516 Bachelors 1 50 Self-emp- not-inc 83311 Bachelors 2 38 Private 215646 HS-grad 3 53 Private 234721 11th 4 28 Private 338409 Bachelors age workclass fnlwgt education 0 39 State-gov 77516 Bachelors 50 Self-emp- not-inc 83311 Bachelors 1 50 Self-emp- not-inc 83311 Bachelors 2 38 Private 215646 HS-grad 3 53 Private 234721 11th 4 28 Private 234721 11th 4 28 Private 234721 11th 4 28 Private 234721 11th 5 27 Private 237302 Assocation 7 40 Private 154374 HS-grad 3 58 Private 151910 HS-grad 5 22 Private 201490 HS-grad	rt numpy as np sklearn.preprocessing import KBinsDiscretizer ### ### ### ### ### ### ### ### ### #	rt numpy as np sklearn.preprocessing import KBinsDiscretizer ep 1: Load the dataset mns = ["age", "workclass", "fnlwgt", "education", "education_num", "mar "age", "workclass", "hours_per_week", "native_country", "income" e pd.read_csv("adult.data", names=columns, sep=r',\s"', engine- age workclass fnlwgt education education_num marital_status 0 39 State-gov 77516 Bachelors 13 Never-married 1 50 Self-emp- not-inc 83311 Bachelors 13 Married-civ- spouse 2 38 Private 215646 HS-grad 9 Divorced 4 28 Private 338409 Bachelors 13 Married-civ- spouse 3 53 Private 234721 11th 7 Married-civ- spouse 4 28 Private 338409 Bachelors 13 Never-married 5 Self-emp- not-inc 83311 Bachelors 13 Never-married 5 Self-emp- not-inc 83311 Bachelors 13 Never-married 2 38 Private 215646 HS-grad 9 Divorced 3 53 Private 234721 11th 7 Married-civ- spouse 3 54 Private 257302 Assoc- acdm 12 Married-civ- spouse 4 28 Private 154374 HS-grad 9 Married-civ- spouse 5 27 Private 257302 Assoc- acdm 12 Married-civ- spouse 7 40 Private 154374 HS-grad 9 Never-married 8 58 Private 201490 HS-grad 9 Never-married Married-civ- spouse Married-civ- spouse 7 40 Private 151910 HS-grad 9 Never-married Married-civ- spouse Married-civ- spouse	rt numpy as np sklearn.preprocessing import KBinsDiscretizer ### ### ### ### ### ### ### ### ### #	rt numpy as np siklearn, preprocessing import KBinsDiscretizer ### ### ### ### ### ### ### ### ### #	rt numpy as np sixlearn.preprocessing import KBinsDiscretizer ap 1: Load the dataset ms = ["age", "workclass", "finlugt", "education", "education num", "marital_status", "occupation", "relationship", "race", "sex", "capital_gain", "capital_loss", "hours_per_week", "native_country", "income" age workclass finlugt education education_num marital_status occupation relationship race pd.read_csv("adult.data", names=columns, sep=r',\s", engine='python', na_values='?') age workclass finlugt education education_num marital_status occupation relationship race pd.read_csv("adult.data", names=columns, sep=r',\s", engine='python', na_values='?') age workclass finlugt education education_num marital_status occupation relationship race pd. self-emp- not-inc assatil Bachelors 13 Mever-married Adm-clerical family private 234721 11th 7 Married-civ- spouse specialty """ "" age workclass finlugt education education_num marital_status occupation relationship race pd. self-emp- not-inc assatil Bachelors 13 Meried-civ- spouse specialty """ "" age workclass finlugt education education_num marital_status occupation relationship race pd. self-emp- not-inc assatil Bachelors 13 Meried-civ- spouse specialty """ "" age workclass finlugt education education_num marital_status occupation relationship race pd. self-emp- not-inc assatil Bachelors 13 Married-civ- spouse managerial Husband White pd. self-emp- not-inc assatil Bachelors 13 Married-civ- spouse support specialty Wife Black private 234721 11th 7 Married-civ- spouse support specialty """ "" pd. self-emp- not-inc assatil Bachelors 13 Married-civ- spouse support specialty """ "" age workclass finlugt education 14 Married-civ- spouse support specialty """ "" age assatil the self-ener spouse support specialty """ "" age assatil the self-ener spouse support suppor	### sklearn. preprocessing import KBinsDiscretizer ### sklearn. preprocessing import KBinsDiscretizer ### sklearn. preprocessing import KBinsDiscretizer #### sklearn. preprocessing import KBinsDiscretizer ##### sklearn. preprocessing import	## Number of the Address	## Number processing import KBinsDiscretizer ### 12 tood the dataset ### 21 tood the dataset ### 21 tood the dataset ### 21 tood the dataset ### 22 tood the dataset ### 22 tood the dataset ### 22 tood the dataset ### 23 tood the dataset ### 24 tood the dataset ### 25 too processing import KBinsDiscretizer ### 25 too processing import KBinsDi	## Indignals of the professional papers #88 insplicate library in the professional papers ## Indignals ## I

```
[3]: # Step 2: Define quasi-identifiers (X) and sensitive attributes (Y)
quasi_identifiers = ["age", "education", "marital_status", "race", "sex"]
sensitive_attributes = ["income"]
  [4]: # Step 3: Generalize or bin quasi-identifiers (example: binning 'age')
        def generalize_age(data, bins=10):
            binner = KBinsDiscretizer(n_bins=bins, encode='ordinal', strategy='uniform')
            data['age'] = binner.fit_transform(data[['age']]).astype(int)
            return data
        data = generalize age(data)
  [5]: 
 k Step 4: Apply k-Anonymity by grouping and suppressing low-frequency combinations k = 5 # Define the desired level of anonymity
        def apply_k_anonymity(data, quasi_identifiers, k):
                    by quasi-identifie
            grouped = data.groupby(quasi_identifiers)
             # Filter groups with fewer than k entries
            valid_groups = grouped.filter(lambda x: len(x) >= k)
            # Suppress small groups by replacing with "Suppressed"
suppressed = grouped.filter(lambda x: len(x) < k).copy()</pre>
            for qi in quasi identifiers:
                 suppressed[qi] = "Suppressed"
             # Combine valid and suppressed gro
            anonymized_data = pd.concat([valid_groups, suppressed], ignore_index=True)
return anonymized_data
        anonymized_data = apply_k_anonymity(data, quasi_identifiers, k)
[6]: # Step 5: Save the anonymized dataset
       anonymized_data.to_csv("adult_dataset_anonymized.csv", index=False)
      print("Anonymized dataset saved as 'adult_dataset_anonymized.csv'.")
      Anonymized dataset saved as 'adult_dataset_anonymized.csv'.
[7]: anonymized_data
                     age workclass fnlwgt education education_num marital_status occupation relationship
                                                                                                                                           sex capital_gain capital_loss hours_per
                                                                                                  Adm-
                                                                                                              Not-in-
           0
                       3 State-gov 77516
                                                 Bachelors
                                                                        13 Never-married
                                                                                                                            White
                                                                                                                                         Male
                                                                                                                                                       2174
                                                                                                                                                                       0
                                                                                                 clerical
                       4 Self-emp-
                                                                              Married-civ-
                                                                                                  Exec-
                                       83311
                                                 Bachelors
                                                                                                            Husband
                                                                                                                            White
                              not-inc
                                                                                    spouse
                                                                                            managerial
                                                                                              Handlers-
                                                                                                              Not-in-
           2
                              Private 215646
                                                  HS-grad
                                                                         9
                                                                                  Divorced
                                                                                                                            White
                                                                                                                                         Male
                                                                                                                                                          0
                                                                                                                                                                       0
                                                                               Married-civ-
                                                                        13
                                                                                                                                                          0
           3
                              Private 338409
                                                 Bachelors
                                                                                                                Wife
                                                                                                                            Black
                                                                                                                                       Female
                                                                                                                                                                       0
                                                                                               specialty
                                                                               Married-civ-
                                                                                                  Exec-
           4
                              Private 284582
                                                                        14
                                                                                                                                                          0
                                                                                                                                                                       0
                                                                                    spouse
                                                                                             managerial
      32556 Suppressed
                                NaN 120478 Suppressed
                                                                        11
                                                                               Suppressed
                                                                                                  NaN
                                                                                                          Unmarried Suppressed Suppressed
                                                                                                                                                          Ω
                                                                                                                                                                       ٥
                                                                                                 Other-
      32557 Suppressed
                                      199655 Suppressed
                                                                               Suppressed
                                                                                                                       Suppressed Suppressed
                                                                                                 service
                                                                                                              family
                                                                                                  Prof-
                           Self-emp-
                                                                                                              Not-in-
      32558 Suppressed
                                       99359 Suppressed
                                                                        15
                                                                               Suppressed
                                                                                                                       Suppressed Suppressed
                                                                                                                                                       1086
                                                                                                                                                                       0
                                                                                               specialty
                     age workclass fnlwgt
                                                education education num marital status occupation relationship
                                                                                                                             race
                                                                                                                                           sex capital gain capital loss hours per
                                                                                                  Adm-
           0
                                                                                                                                                                        0
                           State-gov
                                       77516
                                                 Bachelors
                                                                             Never-married
                                                                                                                                         Male
                                                                                                                                                        2174
                                                                                                 clerical
                                                                                                               family
                           Self-emp
                                                                               Married-civ-
                                                                                                  Exec-
           1
                                       83311
                                                Bachelors
                                                                        13
                                                                                                             Husband
                                                                                                                            White
                                                                                                                                         Male
                                                                                                                                                          0
                                                                                                                                                                        0
                                                                                   spouse
                                                                                             managerial
                                                                                              Handlers-
           2
                              Private 215646
                                                                         9
                                                                                                                                                                        0
                       2
                                                  HS-grad
                                                                                                                            White
                                                                                                                                         Male
                                                                                                                                                           0
                                                                                  Divorced
                                                                                               cleaners
                                                                                                               family
                                                                               Married-civ-
                                                                                                  Prof-
           3
                              Private 338409
                                                 Bachelors
                                                                        13
                                                                                                                 Wife
                                                                                                                             Black
                                                                                                                                       Female
                                                                                                                                                           ٥
                                                                                                                                                                        Ω
                                                                                   spouse
                                                                                               specialty
                                                                               Married-civ-
                                                                                                  Exec-
           4
                       2
                              Private 284582
                                                  Masters
                                                                        14
                                                                                                                Wife
                                                                                                                            White
                                                                                                                                       Female
                                                                                                                                                           0
                                                                                                                                                                       0
                                                                                             managerial
                                                                                    spouse
          ...
                                NaN 120478 Suppressed
                                                                                                                                                                        0
      32556 Suppressed
                                                                        11
                                                                                                  NaN
                                                                                                           Unmarried Suppressed Suppressed
                                                                                                                                                           0
                                                                               Suppressed
                                                                                                 Other-
                                                                                                              Not-in-
      32557 Suppressed
                              Private 199655 Suppressed
                                                                        14
                                                                               Suppressed
                                                                                                                       Suppressed Suppressed
                                                                                                                                                           0
                                                                                                                                                                       0
                                                                                                 service
                                                                                                               family
                           Self-emp-
      32558 Suppressed
                                       99359 Suppressed
                                                                                                                                                                        0
                                                                        15
                                                                                                                      Suppressed Suppressed
                                                                                                                                                        1086
                                                                               Suppressed
                              not-inc
                                                                                               specialty
                                                                                                               family
                                                                                              Handlers-
      32559 Suppressed
                              Private 34066 Suppressed
                                                                         6
                                                                                Suppressed
                                                                                                             Husband Suppressed Suppressed
                                                                                                                                                           ٥
                                                                                                                                                                        0
                                                                                               cleaners
                                                                                                  Tech-
                                                                                                              Not-in-
      32560 Suppressed
                             Private 116138 Suppressed
                                                                                                                                                                       0
                                                                        14
                                                                               Suppressed
                                                                                                                      Suppressed Suppressed
                                                                                                                                                          0
                                                                                                support
                                                                                                               family
```

2. Implement (X,Y) -Linkability for the adult data set via code.

0	39	workclass State-gov		education										
0	39			education										
		State-gov	77546		education_num	marital_status				e sex	capital_gain	capital_loss	hours_per_week	native_c
1 :			77516	Bachelors	13	Never-married	Adm- clerical	Not-in famil		e Male	2174	0	40) United
	50	Self-emp- not-inc	83311	Bachelors	13	Married-civ- spouse	Exec- managerial	Husban	d Whit	e Male	0	0	13	Unite
2	38	Private	215646	HS-grad	9	Divorced	Handlers- cleaners	Not-in famil		e Male	0	0	40) Unite
3 :	53	Private	234721	11th	7	Married-civ- spouse		Husban	d Blac	k Male	0	0	40) Unite
4	28	Private	338409	Bachelors	13	Married-civ-	Prof-	Wif	e Blac	k Female	0	0	40)
ata														
ag	ge 1	workclass	fnlwgt	education e	education_num	marital_status d	occupation re	elationship	race	sex ca	pital_gain ca	pital_loss ho	urs_per_week na	itive_co
0 3	39	State-gov	77516	Bachelors	13	Never-married	Adm- clerical	Not-in- family	White	Male	2174	0	40	United-:
1 5	50	Self-emp- not-inc	83311	Bachelors	13	Married-civ- spouse	Exec- managerial	Husband	White	Male	0	0	13	United-:
2 3	38	Private	215646	HS-grad	9	Divorced	Handlers- cleaners	Not-in- family	White	Male	0	0	40	United-:
3 5	53	Private	234721	11th	7	Married-civ- spouse	Handlers- cleaners	Husband	Black	Male	0	0	40	United-:
4 2	28	Private	338409	Bachelors	13	Married-civ- spouse	Prof- specialty	Wife	Black	Female	0	0	40	
2556 2	27	Private	257302	Assoc- acdm	12	Married-civ- spouse	Tech- support	Wife	White	Female	0	0	38	United-:
2557 4	40	Private	154374	HS-grad	9	Married-civ- spouse	Machine- op-inspct	Husband	White	Male	0	0	40	United-:
2558 5	58	Private	151910	HS-grad	9	Widowed	Adm- clerical	Unmarried	White	Female	0	0	40	United-:
2559 2	22	Private	201490	HS-grad	9	Never-married	Adm- clerical	Own-child	White	Male	0	0	20	United-:
2560 5.	52	Self-emp- inc	287927	HS-grad	9	Married-civ- spouse	Exec- managerial	Wife	White	Female	15024	0	40	United-:

```
[3]: # Step 2: Define quasi-identifiers (X) and sensitive attributes (Y)
      quasi_identifiers = ["age", "sex", "race"]
sensitive_attribute = "income"
[4]: # Step 3: Group by quasi-identifiers and calculate diversity of sensitive attributes
def calculate_1_diversity(group, sensitive_column):
          # Calculate the number of unique values in the sensitive column
           return group[sensitive column].nunique()
      # Group data by quasi-identifiers
      grouped = data.groupby(quasi_identifiers)
       # Calculate diversity for each grou
      diversity = grouped[sensitive_attribute].nunique().reset_index(name='diversity')
[5]: # Step 4: Ensure l-diversity (e.g., l = 2)
      sufficiently_diverse_groups = diversity[diversity['diversity'] >= 1]
       # Filter original data to keep only sufficiently diverse groups
      \label{linear_diverse_groups} diverse\_data = data.merge(sufficiently\_diverse\_groups[quasi\_identifiers], on=quasi\_identifiers, how='inner')
[6]: # Step 5: Save the (X, Y)-Linkable dataset
diverse_data.to_csv("adult_dataset_linkable.csv", index=False)
      print("Linkable dataset saved as 'adult_dataset_linkable.csv'.")
      Linkable dataset saved as 'adult_dataset_linkable.csv'.
                                                                                                                                                    ◎ ↑ ↓ 盐 ♀ ▮
[7]: diverse data
             age workclass fnlwgt education education_num marital_status occupation relationship race
                                                                                                                     sex capital_gain capital_loss hours_per_week native_co
                                                                                       Adm-
                                                                                                  Not-in-
                                                                                                   family White
          0 39 State-gov 77516 Bachelors
                                                              13 Never-married
                                                                                                                    Male
                                                                                                                                 2174
                                                                                                                                                 0
                                                                                                                                                                      United-
                                                                                     clerical
                                                                    Married-civ-
                                                                                       Exec-
                   Self-emp-
          1 50
                              83311
                                       Bachelors
                                                              13
                                                                                                 Husband White
                                                                                                                    Male
                                                                                                                                    0
                                                                                                                                                0
                                                                                                                                                                 13
                                                                                                                                                                      United-
                                                                                  managerial
                                                                                   Handlers
          2 38
                      Private 215646
                                        HS-grad
                                                                       Divorced
                                                                                                          White
                                                                                                                                    0
                                                                                                                                                 0
                                                                                                                                                                40
                                                                                                                                                                      United-
                                                                                                                    Male
                                                                                    cleaners
                                                                                                   family
                                                                    Married-civ-
                                                                                   Handlers-
          3 53
                      Private 234721
                                           11th
                                                                                                 Husband Black
                                                                                                                    Male
                                                                                                                                    0
                                                                                                                                                 0
                                                                                                                                                                 40
                                                                                                                                                                      United-
                                                                         spouse
                                                                                    cleaners
                                                                    Married-civ-
           4 28
                      Private 338409
                                       Bachelors
                                                              13
                                                                                                    Wife Black Female
                                                                                                                                    0
                                                                                                                                                0
                                                                                                                                                                40
                                                                         spouse
                                                                                    specialty
                                                                    Married-civ-
                                                                                       Tech-
      29443 27
                      Private 257302
                                                              12
                                                                                                    Wife White Female
                                                                                                                                    0
                                                                                                                                                 0
                                                                                                                                                                      United-
                                                                                                                                                                38
                                           acdm
                                                                         spouse
                                                                                    support
                                                                    Married-civ-
                                                                                   Machine-
                      Private 154374
                                        HS-grad
                                                                                                Husband White
                                                                                                                                                                      United-
                                                                        spouse
                                                                                   op-inspct
                                                                                       Adm-
      29445 58
                      Private 151910
                                        HS-grad
                                                                       Widowed
                                                                                               Unmarried White Female
                                                                                                                                    0
                                                                                                                                                0
                                                                                                                                                                40
                                                                                                                                                                      United-
                                                                                     clerical
      29446 22
                      Private 201490
                                        HS-grad
                                                               9 Never-married
                                                                                                Own-child White
                                                                                                                   Male
                                                                                                                                                 0
                                                                                                                                                                      United-
                                                                                                                                                                20
                                                                                     clerical
```

Married-civ-

Exec-

spouse managerial

Wife White Female

15024

0

40 United-

29448 rows × 15 columns

29447 52 Self-empinc 287927

HS-grad