In [1]:

```
import pandas as pd
import numpy as np
```

In [9]:

```
df=pd.read_csv(r"C:\Users\sagar\Desktop\dataset\PARTB-DATASETS\adult_dataset.csv")
```

In [10]:

```
df.head()
```

Out[10]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	race
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	Black
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	White
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	White
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	Black
4	18	?	103497	Some-college	10	Never-married	?	Own-child	White

In [11]:

```
df.columns
```

Out[11]:

```
Index(['age', 'workclass', 'fnlwgt', 'education', 'educational-num',
       'marital-status', 'occupation', 'relationship', 'race', 'gender',
       'capital-gain', 'capital-loss', 'hours-per-week', 'native-country',
       'income'],
      dtype='object')
```

In [12]:

```
#subset creation

subset1=df[['age','workclass','education']].loc[0:15]  #0 to 15 records are extracted
```

In [13]:

```
subset1
```

Out[13]:

	age	workclass	education
0	25	Private	11th
1	38	Private	HS-grad
2	28	Local-gov	Assoc-acdm
3	44	Private	Some-college
4	18	?	Some-college
5	34	Private	10th
6	29	?	HS-grad
7	63	Self-emp-not-inc	Prof-school
8	24	Private	Some-college
9	55	Private	7th-8th
10	65	Private	HS-grad
11	36	Federal-gov	Bachelors
12	26	Private	HS-grad
13	58	?	HS-grad
14	48	Private	HS-grad
15	43	Private	Masters

In [14]:

```
subset2=df[['age','workclass','education']].loc[16:30]  #0 to 15 records are extracted
```

In [15]:

```
subset3=df[['age','workclass','education']].loc[31:50]  #0 to 15 records are extracted
```

In [16]:

```
#merging data
merging=pd.concat([subset1,subset2,subset3])
```

In [18]:

```
merging#merge all records from 1 to 50
```

Out[18]:

	age	workclass	education
0	25	Private	11th
1	38	Private	HS-grad
2	28	Local-gov	Assoc-acdm
3	44	Private	Some-college
4	18	?	Some-college
5	34	Private	10th
6	29	?	HS-grad
7	63	Self-emp-not-inc	Prof-school
8	24	Private	Some-college
9	55	Private	7th-8th
10	65	Private	HS-grad
11	36	Federal-gov	Bachelors
12	26	Private	HS-grad
13	58	?	HS-grad
14	48	Private	HS-grad
15	43	Private	Masters
16	20	State-gov	Some-college
17	43	Private	HS-grad
18	37	Private	HS-grad
19	40	Private	Doctorate
20	34	Private	Bachelors
21	34	Private	Some-college
22	72	?	7th-8th
23	25	Private	Bachelors
24	25	Private	Bachelors
25	45	Self-emp-not-inc	HS-grad
26	22	Private	HS-grad
27	23	Private	HS-grad
28	54	Private	HS-grad
29	32	Self-emp-not-inc	Some-college
30	46	State-gov	Some-college
31	56	Self-emp-not-inc	11th
32	24	Self-emp-not-inc	Bachelors
33	23	Local-gov	Some-college
34	26	Private	HS-grad
35	65	?	HS-grad
36	36	Local-gov	Bachelors

	age	workclass	education
37	22	Private	5th-6th
38	17	Private	10th
39	20	Private	HS-grad
40	65	Private	Masters
41	44	Self-emp-inc	Assoc-voc
42	36	Private	HS-grad
43	29	Private	11th
44	20	State-gov	Some-college
45	28	Private	Assoc-voc
46	39	Private	7th-8th
47	54	Private	Some-college
48	52	Private	11th
49	56	Self-emp-inc	HS-grad
50	18	Private	Some-college

In [19]:

```
#sorting the data
sort_values=df.sort_values('age',ascending=False) #to sort age attribute values in decen
```

In [20]:

```
sort_values
```

Out[20]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship
40519	90	?	166343	1st-4th	2	Widowed	?	Not-in-family
21553	90	Private	141758	9th	5	Never-married	Adm-clerical	Not-in-family
47977	90	?	313986	HS-grad	9	Married-civ-spouse	?	Husband
21651	90	Local-gov	227796	Masters	14	Married-civ-spouse	Exec-managerial	Husband
41584	90	?	175444	7th-8th	4	Separated	?	Not-in-family
...	...	...	...	...	...	...	...	...
39614	17	?	48751	11th	7	Never-married	?	Own-child
23072	17	Private	110723	11th	7	Never-married	Sales	Own-child
33275	17	Private	142457	11th	7	Never-married	Other-service	Own-child
31429	17	Private	34019	11th	7	Never-married	Other-service	Own-child
20904	17	Private	181580	11th	7	Never-married	Other-service	Own-child

48842 rows × 15 columns

In [21]:

```
#transposing data
#columns are changed to rows and rows changed to columns
```

In [22]:

```
df.transpose()
```

Out[22]:

	0	1	2	3	4	5	6	7
age	25	38	28	44	18	34	29	63
workclass	Private	Private	Local-gov	Private	?	Private	?	Self-emp-not-inc
fnlwgt	226802	89814	336951	160323	103497	198693	227026	104626
education	11th	HS-grad	Assoc-acdm	Some-college	Some-college	10th	HS-grad	Prof-school
educational-num	7	9	12	10	10	6	9	15
marital-status	Never-married	Married-civ-spouse	Married-civ-spouse	Married-civ-spouse	Never-married	Never-married	Never-married	Married-civ-spouse
occupation	Machine-op-inspct	Farming-fishing	Protective-serv	Machine-op-inspct	?	Other-service	?	Prof-specialty
relationship	Own-child	Husband	Husband	Husband	Own-child	Not-in-family	Unmarried	Husband
race	Black	White	White	Black	White	White	Black	White
gender	Male	Male	Male	Male	Female	Male	Male	Male
capital-gain	0	0	0	7688	0	0	0	3103
capital-loss	0	0	0	0	0	0	0	0
hours-per-week	40	50	40	40	30	30	40	32
native-country	United-States	United-States	United-States	United-States	United-States	United-States	United-States	United-States
income	<=50K	<=50K	>50K	>50K	<=50K	<=50K	<=50K	>50K

15 rows × 48842 columns

In [23]:

```
#shape and reshape
shaping=df.shape
```

In [24]:

```
shaping
```

Out[24]:

```
(48842, 15)
```

In [28]:

```
#reshaping
pivot_table=pd.pivot_table(df,index=['age','income'],values='educational-num')
```

In [29]:

```
pivot_table
```

Out[29]:

		educational-num
age	**income**	
17	**<=50K**	6.685714
18	**<=50K**	8.220418
19	**<=50K**	9.091429
	>50K	7.333333
20	**<=50K**	9.342626
...	**...**	...
88	**<=50K**	11.400000
	>50K	7.000000
89	**<=50K**	9.500000
90	**<=50K**	9.452381
	>50K	12.153846

142 rows × 1 columns

In [ ]:

```

```