In [1]:
```
import pandas as pd
df=pd.read_csv('adult_dataset.csv')
df
```

Out[1]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	E
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	W
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	W
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	E
4	18	?	103497	Some-college	10	Never-married	?	Own-child	W
...	...	...	...	...	...	...	...	...	
48837	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	W
48838	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	W
48839	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	W
48840	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	W
48841	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	W

48842 rows × 15 columns

In [2]:
```
df.describe()
```

Out[2]:

	age	fnlwgt	educational-num	capital-gain	capital-loss	hours-per-week
count	48842.000000	4.884200e+04	48842.000000	48842.000000	48842.000000	48842.000000
mean	38.643585	1.896641e+05	10.078089	1079.067626	87.502314	40.422382
std	13.710510	1.056040e+05	2.570973	7452.019058	403.004552	12.391444
min	17.000000	1.228500e+04	1.000000	0.000000	0.000000	1.000000
25%	28.000000	1.175505e+05	9.000000	0.000000	0.000000	40.000000
50%	37.000000	1.781445e+05	10.000000	0.000000	0.000000	40.000000
75%	48.000000	2.376420e+05	12.000000	0.000000	0.000000	45.000000
max	90.000000	1.490400e+06	16.000000	99999.000000	4356.000000	99.000000

In [4]:
```
# Data SUBSETS

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48842 entries, 0 to 48841
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   age              48842 non-null  int64 
 1   workclass        48842 non-null  object
 2   fnlwgt           48842 non-null  int64 
 3   education        48842 non-null  object
 4   educational-num  48842 non-null  int64 
 5   marital-status   48842 non-null  object
 6   occupation       48842 non-null  object
 7   relationship     48842 non-null  object
 8   race             48842 non-null  object
 9   gender           48842 non-null  object
 10  capital-gain     48842 non-null  int64 
 11  capital-loss     48842 non-null  int64 
 12  hours-per-week   48842 non-null  int64 
 13  native-country   48842 non-null  object
 14  income           48842 non-null  object
dtypes: int64(6), object(9)
memory usage: 5.6+ MB
```

In [5]:
```
# First subset: Like and Share
df_subset_1 = df[['workclass','education','capital-gain']]
df_subset_1
```

Out[5]:

	workclass	education	capital-gain
0	Private	11th	0
1	Private	HS-grad	0
2	Local-gov	Assoc-acdm	0
3	Private	Some-college	7688
4	?	Some-college	0
...	...	...	...
48837	Private	Assoc-acdm	0
48838	Private	HS-grad	0
48839	Private	HS-grad	0
48840	Private	HS-grad	0
48841	Self-emp-inc	HS-grad	15024

48842 rows × 3 columns

In [6]:
```
# second subset: Comment and Type
df_subset_2 = df[['race','native-country']]
df_subset_2
```

Out[6]:

	race	native-country
0	Black	United-States
1	White	United-States
2	White	United-States
3	Black	United-States
4	White	United-States
...	...	...
48837	White	United-States
48838	White	United-States
48839	White	United-States
48840	White	United-States
48841	White	United-States

48842 rows × 2 columns

In [7]:
```
#MERGE DATA

merged_data  = pd.concat([df_subset_1,df_subset_2],axis=1)
merged_data
```

Out[7]:

	workclass	education	capital-gain	race	native-country
0	Private	11th	0	Black	United-States
1	Private	HS-grad	0	White	United-States
2	Local-gov	Assoc-acdm	0	White	United-States
3	Private	Some-college	7688	Black	United-States
4	?	Some-college	0	White	United-States
...	...	...	...	...	...
48837	Private	Assoc-acdm	0	White	United-States
48838	Private	HS-grad	0	White	United-States
48839	Private	HS-grad	0	White	United-States
48840	Private	HS-grad	0	White	United-States
48841	Self-emp-inc	HS-grad	15024	White	United-States

48842 rows × 5 columns

In [8]:
```
#SORT DATA

# Sorting merged_data in descending order wrt 'capital-gain'
merged_data.sort_values(by=['capital-gain'],ascending=False)
```

Out[8]:

	workclass	education	capital-gain	race	native-country
28936	Self-emp-inc	HS-grad	99999	White	?
18384	Self-emp-inc	Some-college	99999	White	United-States
34689	Self-emp-inc	Prof-school	99999	White	United-States
34744	Private	Prof-school	99999	White	United-States
48519	Private	Prof-school	99999	White	United-States
...	...	...	...	...	...
16967	?	7th-8th	0	White	United-States
16968	Local-gov	Bachelors	0	White	United-States
16969	Private	Masters	0	White	Dominican-Republic
16970	Private	HS-grad	0	White	United-States
24421	Private	HS-grad	0	White	United-States

48842 rows × 5 columns

In [9]:
```
#TRANSPOSING DATA

merged_data.T
```

Out[9]:

	0	1	2	3	4	5	6	7	8	9
workclass	Private	Private	Local-gov	Private	?	Private	?	Self-emp-not-inc	Private	Private
education	11th	HS-grad	Assoc-acdm	Some-college	Some-college	10th	HS-grad	Prof-school	Some-college	7th-8th
capital-gain	0	0	0	7688	0	0	0	3103	0	0
race	Black	White	White	Black	White	White	Black	White	White	White
native-country	United-States	United-States	United-States	United-States	United-States	United-States	United-States	United-States	United-States	United-States

5 rows × 48842 columns

In [10]:
```
#SHAPE AND RESHAPE DATA

df
```

Out[10]:

	age	workclass	fnlwgt	education	educational-num	marital-status	occupation	relationship	
0	25	Private	226802	11th	7	Never-married	Machine-op-inspct	Own-child	E
1	38	Private	89814	HS-grad	9	Married-civ-spouse	Farming-fishing	Husband	W
2	28	Local-gov	336951	Assoc-acdm	12	Married-civ-spouse	Protective-serv	Husband	W
3	44	Private	160323	Some-college	10	Married-civ-spouse	Machine-op-inspct	Husband	E
4	18	?	103497	Some-college	10	Never-married	?	Own-child	W
...	...	...	...	...	...	...	...	...	
48837	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	W
48838	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	W
48839	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	W
48840	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	W
48841	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	W

48842 rows × 15 columns

In [11]:
```
# Reshape
pd.melt(df, id_vars =['education'], value_vars =['capital-gain'])
```

Out[11]:

	education	variable	value
0	11th	capital-gain	0
1	HS-grad	capital-gain	0
2	Assoc-acdm	capital-gain	0
3	Some-college	capital-gain	7688
4	Some-college	capital-gain	0
...	...	...	...
48837	Assoc-acdm	capital-gain	0
48838	HS-grad	capital-gain	0
48839	HS-grad	capital-gain	0
48840	HS-grad	capital-gain	0
48841	HS-grad	capital-gain	15024

48842 rows × 3 columns

In [ ]: