In [7]:
```
import pandas as pd
import numpy as np
```

In [8]:
```
features=["Age","Workclass","fnlwgt","Education","Education-Num","Martial Status","Occupation","Relationship",
          "Race","Sex","Capital Gain","Capital Loss","Hour per week","Country","Target"]

df = pd.read_csv('adult.data',names=features)
df
```

Out[8]:

	Age	Workclass	fnlwgt	Education	Education-Num	Martial Status	Occupation	Relationship	Race	Sex	Capital Gain	Capital Loss	Hour per week	Country	Target
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

In [9]:
```
df.to_csv('adult.csv',sep=',',index=False)
```

In [ ]:
```
'''Question: Do data analysis using Pandas and answer following questions?

1. How many men and women (sex feature) are represented in this dataset?

2. What is the average age (age feature) of women?

3. What is the proportion of German citizens (native-country feature)?

4-5. What are mean value and standard deviation of the age of those who recieve more than 50K per year (salary feature) and those who receive less t

6. Is it true that people who receive more than 50k have at least high school education? (education - Bachelors, Prof-school, Assoc-acdm, Assoc-voc,
```

1st Question

In [11]:
```
df['Sex'].value_counts()
```

Out[11]:
```
 Male      21790
 Female    10771
Name: Sex, dtype: int64
```

2nd Question

In [12]:
```
female_data=df[df['Sex'].str.contains('Female')]
female_data['Age'].mean()
```

Out[12]:
```
36.85823043357163
```

3rd Question

In [13]:
```
var=df['Country'].value_counts('Germany')
var[4:5]
```

Out[13]:
```
 Germany    0.004207
Name: Country, dtype: float64
```

4th Question

In [14]:
```
sal_high=df[df['Target'].str.contains('>50K')]
```

In [15]:
```
sal_high['Age'].mean()
```

Out[15]:
```
44.24984058155847
```

In [16]:
```
sal_high['Age'].std()
```

Out[16]:
```
10.51902771985177
```

5th Question

In [17]:
```
sal_low=df[df['Target'].str.contains('<=50K')]
```

In [18]:
```
sal_low['Age'].mean()
```

Out[18]:
```
36.78373786407767
```

In [19]:
```
sal_low['Age'].std()
```

Out[19]:
```
14.020088490824813
```

6th Question

In [28]:
```
hgh_edu=df[df['Target'].str.contains('>50K')]
a=hgh_edu['Education-Num'] >=11
a
```

Out[28]:
```
7        False
8         True
9         True
10       False
11        True
         ...
32539     True
32545     True
32554     True
32557    False
32560    False
Name: Education-Num, Length: 7841, dtype: bool
```

In [27]:
```
a.value_counts()
```

Out[27]:
```
True     4535
False    3306
Name: Education-Num, dtype: int64
```

In [ ]:
```

```