In [2]:
```
import numpy as np
import pandas as pd
data = pd.read_csv(r'C:\Users\DELL\Desktop\AIML-JNTU\adult.data.csv')
data.head()
```

Out[2]:

	age	workclass	number	education	rollno	martial_status	occupation	relationship	race	sex	gain	loss	hoursperweek	native-country	salary
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

In [3]:
```
data['sex'].value_counts()
```

Out[3]:
```
 Male      21790
 Female    10771
Name: sex, dtype: int64
```

In [84]:
```
data.loc[data['sex'] == ' Female', 'age'].mean()
```

Out[84]:
```
36.85823043357163
```

In [83]:
```
float((data['native-country'] == ' Germany').sum()) / data.shape[0]
```

Out[83]:
```
0.004207487485028101
```

In [85]:
```
ages1 = data.loc[data['salary'] == ' >50K', 'age']
ages2 = data.loc[data['salary'] == ' <=50K', 'age']
print("The average age of the rich: {0} +- {1} years, poor - {2} +- {3} years.".format
      (round(ages1.mean()), round(ages1.std(), 1),
       round(ages2.mean()), round(ages2.std(), 1)))
```

```
The average age of the rich: 44 +- 10.5 years, poor - 37 +- 14.0 years.
```

In [86]:
```
data.loc[data['salary'] == ' >50K', 'education'].unique() #No
```

Out[86]:
```
array([' HS-grad', ' Masters', ' Bachelors', ' Some-college',
       ' Assoc-voc', ' Doctorate', ' Prof-school', ' Assoc-acdm',
       ' 7th-8th', ' 12th', ' 10th', ' 11th', ' 9th', ' 5th-6th',
       ' 1st-4th'], dtype=object)
```

In [ ]:
```

```