In [3]:
```
import pandas as pd
import numpy as np
```

In [4]:
```
adult = pd.read_csv('adult.data', names=features)
adult
```

Out[4]:

	Age	Workclass	fnlwgt	Education	Education-Num	Martial Status	Occupation	Relationship	Race	Sex	Capital Gain	Capital Loss	Hours per week	Country
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
32556	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	0	38	United-States
32557	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	0	40	United-States
32558	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	0	40	United-States
32559	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	White	Male	0	0	20	United-States
32560	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	White	Female	15024	0	40	United-States

32561 rows × 15 columns

https://rstudio-pubs-static.s3.amazonaws.com/538563_85cb2b4cd06b4dc48d33de73fa97a297.html

https://archive.ics.uci.edu/dataset/2/adult

Dataset :

http://archive.ics.uci.edu/dataset/2/adult

Question: Do data analysis using Pandas and answer following questions?

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

In [5]:
```
adult["Sex"].value_counts()
```

Out[5]:
```
 Male      21790
 Female    10771
Name: Sex, dtype: int64
```

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

In [6]:
```
adult[["Sex", "Age"]].groupby("Sex").mean()
```

Out[6]:

Sex	Age
Female	36.858230
Male	39.433547

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

In [7]:
```
country_germany = adult[adult['Country'].str.contains('Germany')]
```

In [8]:
```
country_germany.describe()
```

Out[8]:

	Age	fnlwgt	Education-Num	Capital Gain	Capital Loss	Hours per week
count	137.000000	137.000000	137.000000	137.000000	137.000000	137.000000
mean	39.255474	189325.313869	10.985401	887.094891	77.978102	41.014599
std	12.962065	100809.067728	2.370112	3627.385181	371.502899	12.328223
min	18.000000	21306.000000	4.000000	0.000000	0.000000	6.000000
25%	29.000000	116391.000000	9.000000	0.000000	0.000000	40.000000
50%	36.000000	178322.000000	10.000000	0.000000	0.000000	40.000000
75%	47.000000	231604.000000	13.000000	0.000000	0.000000	50.000000
max	74.000000	606111.000000	16.000000	27828.000000	1977.000000	70.000000

In [9]:
```
adult["Country"].value_counts()
```

Out[9]:
```
United-States                 29170
Mexico                          643
?                               583
Philippines                     198
Germany                         137
Canada                          121
Puerto-Rico                     114
El-Salvador                     106
India                           100
Cuba                             95
England                          90
Jamaica                          81
South                            80
China                            75
Italy                            73
Dominican-Republic               70
Vietnam                          67
Guatemala                        64
Japan                            62
Poland                           60
Columbia                         59
Taiwan                           51
Haiti                            44
Iran                             43
Portugal                         37
Nicaragua                        34
Peru                             31
France                           29
Greece                           29
Ecuador                          28
Ireland                          24
Hong                             20
Cambodia                         19
Trinadad&Tobago                  19
Laos                             18
Thailand                         18
Yugoslavia                       16
Outlying-US(Guam-USVI-etc)       14
Honduras                         13
Hungary                          13
Scotland                         12
Holand-Netherlands                1
Name: Country, dtype: int64
```

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 than 50K per year?

In [10]:
```
age_more50k = adult[adult['Target'].str.contains('>50K')]
print("Mean value of Age who is having Target >50K:", age_more50k.Age.mean().round(2))
print("Std value of Age who is having Target >50K:", age_more50k.Age.std())
```

```
Mean value of Age who is having Target >50K: 44.25
Std value of Age who is having Target >50K: 10.51902771985177
```

In [11]:
```
age_less50k = adult[adult['Target'].str.contains('<=50K')]
print("Mean value of Age who is having Target <=50K:", age_less50k.Age.mean().round(2))
print("Std value of Age who is having Target <=50K:", age_less50k.Age.std())
```

```
Mean value of Age who is having Target <=50K: 36.78
Std value of Age who is having Target <=50K: 14.020088490824813
```

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, Masters or Doctorate feature)

In [4]:
```
import matplotlib.pyplot as plt
%matplotlib inline
```

In [ ]:
```
adult.plot.bar(x = "Education", y = "Target"==">50")
```