In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
features = ["Age", "Workclass", "fnlwgt", "Education", "Education-Num", "Martial S
            "Race", "Sex", "Capital Gain", "Capital Loss", "Hours per week", "Cou

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

Out[2]:

	Age	Workclass	fnlwgt	Education	Education-Num	Martial Status	Occupation	Relationship	Rac
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	Whit
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	Whit
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	Whit
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Blac
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Blac
...	...	...	...	...	...	...	...	...	...
32556	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	Whit
32557	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	Whit
32558	58	Private	151910	HS-grad	9	Widowed	Adm-clerical	Unmarried	Whit
32559	22	Private	201490	HS-grad	9	Never-married	Adm-clerical	Own-child	Whit
32560	52	Self-emp-inc	287927	HS-grad	9	Married-civ-spouse	Exec-managerial	Wife	Whit

32561 rows × 15 columns

Dataset :

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

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

In [3]:
```
adult["Sex"].value_counts()
```

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

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

In [4]:
```
adult[["Sex", "Age"]].groupby("Sex").mean()
```

Out[4]:

Sex	Age
Female	36.858230
Male	39.433547

In [5]:
```
gen_fem = adult[adult["Sex"].str.contains("Female")]
gen_fem["Age"].mean()
```

Out[5]:
```
36.85823043357163
```

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

In [6]:
```
percent_germany = adult[adult['Country'].str.contains('Germany')]
propo_germany = (len(percent_germany)*100)/len(adult)
print("Proportion of German Citizens:", propo_germany)
```

```
Proportion of German Citizens: 0.42074874850281013
```

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 [7]:
```
age_more50k = adult[adult['Target'].str.contains('>50K')]
print("Mean value of Age who is having Target >50K:", age_more50k.Age.mean().round
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
```

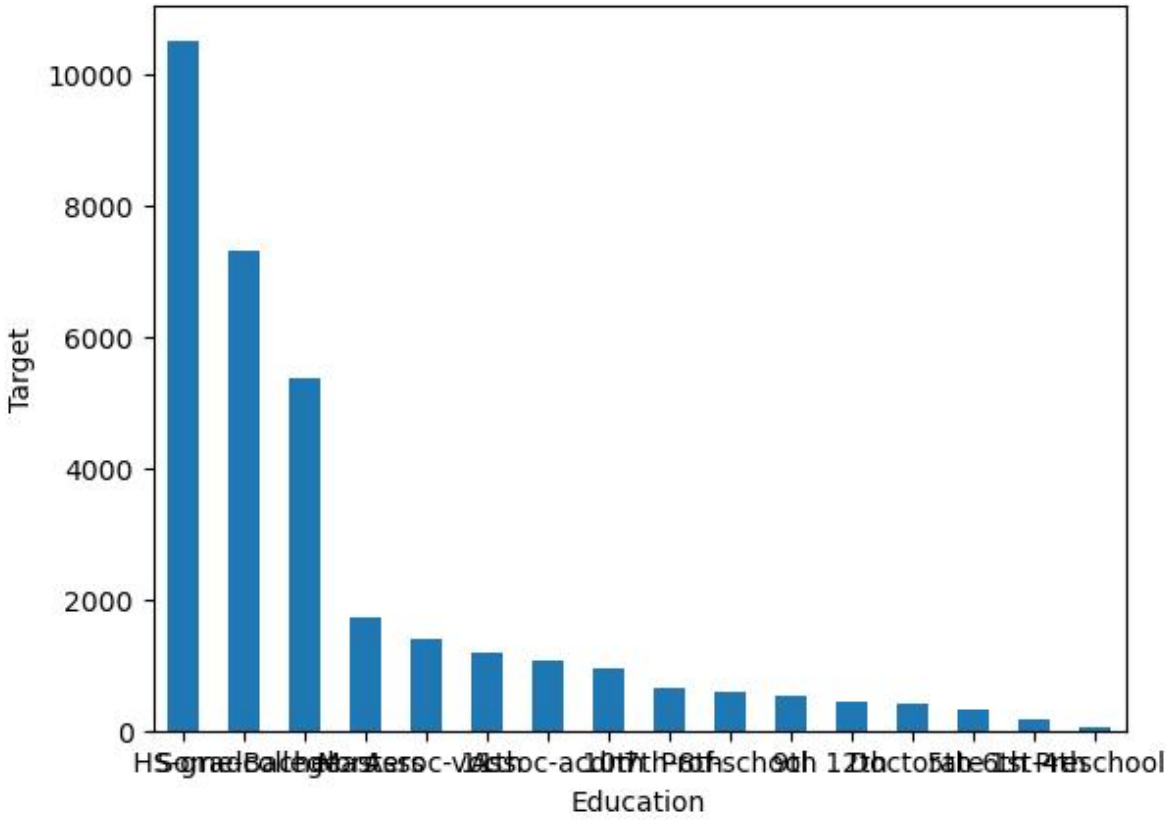
In [8]:
```
age_less50k = adult[adult['Target'].str.contains('<=50K')]
print("Mean value of Age who is having Target <=50K:", age_less50k.Age.mean().roun
print("Std value of Age who is having Target <=50K:", age_less50k.Age.std())
```

```
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 [9]:
```
adult['Education'].value_counts().plot(kind='bar',xlabel='Education', ylabel='Targ
```

Out[9]:
```
<Axes: xlabel='Education', ylabel='Target'>
```


10000
8000
6000
4000
2000
0
Target
Education