IMPORTING DATA SET

```
import pandas as pd
import numpy as np

features = ["Age", "WClass", "fnlwgt", "Edu", "EduNum", "Martial
Status", "Occupation", "Relationship",
            "Race", "Sex", "CapGain", "CapLoss", "HoursPerWeek",
"Country", "Salary"]

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

```
       Age             WClass  fnlwgt          Edu  EduNum  \
0       39          State-gov   77516    Bachelors      13
1       50   Self-emp-not-inc   83311    Bachelors      13
2       38            Private  215646      HS-grad       9
3       53            Private  234721         11th       7
4       28            Private  338409    Bachelors      13
...    ...                ...     ...          ...     ...
32556   27            Private  257302   Assoc-acdm      12
32557   40            Private  154374      HS-grad       9
32558   58            Private  151910      HS-grad       9
32559   22            Private  201490      HS-grad       9
32560   52       Self-emp-inc  287927      HS-grad       9

            Martial Status          Occupation    Relationship    Race
\
0            Never-married        Adm-clerical   Not-in-family   White

1       Married-civ-spouse     Exec-managerial         Husband   White

2                 Divorced   Handlers-cleaners   Not-in-family   White

3       Married-civ-spouse   Handlers-cleaners         Husband   Black

4       Married-civ-spouse      Prof-specialty            Wife   Black

...                    ...                 ...             ...     ...

32556   Married-civ-spouse        Tech-support            Wife   White

32557   Married-civ-spouse   Machine-op-inspct         Husband   White

32558              Widowed        Adm-clerical       Unmarried   White

32559        Never-married        Adm-clerical       Own-child   White

32560   Married-civ-spouse     Exec-managerial            Wife   White
```

```
           Sex  CapGain  CapLoss  HoursPerWeek         Country  Salary
0         Male     2174        0            40   United-States   <=50K
1         Male        0        0            13   United-States   <=50K
2         Male        0        0            40   United-States   <=50K
3         Male        0        0            40   United-States   <=50K
4       Female        0        0            40            Cuba   <=50K
...        ...      ...      ...           ...             ...     ...
32556   Female        0        0            38   United-States   <=50K
32557     Male        0        0            40   United-States    >50K
32558   Female        0        0            40   United-States   <=50K
32559     Male        0        0            20   United-States   <=50K
32560   Female    15024        0            40   United-States    >50K

[32561 rows x 15 columns]
```

ANSWER 1

```
print(adData['Sex'].tolist().count(' Female'))
print(adData['Sex'].tolist().count(' Male'))
```

```
10771
21790
```

ANSWER 2

```
print ("Avg age of Female: ")
adData.loc[adData['Sex'] == ' Female', 'Age'].mean()
```

```
Avg age of Female:

36.85823043357163
```

ANSWER 3

```
print("Proportion of German citizens: ")
float((adData['Country'] == ' Germany').sum()) / adData.shape[0]
```

Proportion of German citizens:

0.004207487485028101

ANSWER 4-5

```
adData = adData.fillna(0)
adData
age1 = adData.loc[adData['Salary'] == ' >50K', 'Age']
age2 = adData.loc[adData['Salary'] == ' <=50K', 'Age']
print("The average age of the people with > 50k salary: {0} +- {1}
years, and <= 50k - {2} +- {3} years.".format((age1.mean()),
(age1.std(), 1), (age2.mean()), (age2.std(), 1)))
```

```
The average age of the people with > 50k salary: 44.24984058155847 +-
(10.51902771985177, 1) years, and <= 50k - 36.78373786407767 +-
(14.020088490824813, 1) years.
```

ANSWER 6

```
print("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)")
edlist = list(adData.loc[adData['Salary'] == ' >50K', 'Edu'].unique())
all(x in [' Bachelors', ' Prof-school', ' Assoc-acdm', ' Assoc-voc', '
Masters', 'Doctorate'] for x in edlist)
```

```
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)

False
```