ASSIGNMENT-4

1)
```
import pandas as pd
url = 'http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data'
data = pd.read_csv(url, header=None)
gender_counts = data[9].value_counts()
print(gender_counts)
```

Output - The Dataset contains 21,790 men and 10,771 women.

1)
```
import pandas as pd

url = 'http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data'
data = pd.read_csv(url, header=None)
women_data = data[data[9] == 'Female']
average_age_women = women_data[0].mean()
print("Average age of women:", average_age_women)
```

Output - Proportion of German citizens: 0.004207487485028101

2)
```
import pandas as pd

url = 'http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data'
data = pd.read_csv(url, header=None)
native_country_counts = data[13].value_counts()
proportion_german_citizens = native_country_counts['Germany'] / data.shape[0]
print("Proportion of German citizens:", proportion_german_citizens)
```

Output - Proportion of German citizens: 0.004207487485028101

3)
```
import pandas as pd
url = 'http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data'
data = pd.read_csv(url, header=None)
mean_age_high_income = data[data[14] == '>50K'][0].mean()
std_age_high_income = data[data[14] == '>50K'][0].std()
mean_age_low_income = data[data[14] == '<=50K'][0].mean()
std_age_low_income = data[data[14] == '<=50K'][0].std()

print("Mean age of high-income individuals:", mean_age_high_income)
print("Standard deviation of age for high-income individuals:", std_age_high_income)
print("Mean age of low-income individuals:", mean_age_low_income)
print("Standard deviation of age for low-income individuals:", std_age_low_income)
```

Output - Mean age of high-income individuals: 44.24984058155847
Standard deviation of age for high-income individuals: 10.51902771985177
Mean age of low-income individuals: 36.78373786407767
Standard deviation of age for low-income individuals: 14.020088490824813

4) import pandas as pd

```
url = 'http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data'
data = pd.read_csv(url, header=None)
high_income_data = data[data[14] == '>50K']
unique_education_levels = high_income_data[3].unique()
        result = all(level in ['Bachelors', 'Prof-school', 'Masters', 'Doctorate'] for level
                                                                    in  unique_education_levels)

print("Do all individuals who receive more than 50K have at least a high school
education?", result)
```

Output -

Do all individuals who receive more than 50K have at least a high school education? False