Practical Experiment: Measure Variance and Range

Objective:

To implement a Python script to measure variance and range on the columns `age`, `capital-gain`, and `capital-loss` for the UCI Adult dataset.

Steps:

1. **Read Data**:
   - Use the Pandas library to read a CSV file.
2. **Measure Variance and Range**:
   - Calculate the variance and range for the specified columns.

Data File:

The UCI Adult dataset can be found here.

Python Script:

python

```
import pandas as pd

# Step 1: Read Data
url = 'https://archive.ics.uci.edu/ml/machine-learning-
databases/adult/adult.data'
columns = ['age', 'workclass', 'fnlwgt', 'education',
'education_num', 'marital_status', 'occupation',
           'relationship', 'race', 'sex', 'capital_gain',
'capital_loss', 'hours_per_week', 'native_country', 'income']
data = pd.read_csv(url, header=None, names=columns, na_values=' ?')

# Display the first few rows of the dataset
print("Original Data:\n", data.head())

# Step 2: Measure Variance and Range
# Calculate variance and range for 'age'
age_variance = data['age'].var()
age_range = data['age'].max() - data['age'].min()

# Calculate variance and range for 'capital-gain'
capital_gain_variance = data['capital_gain'].var()
capital_gain_range = data['capital_gain'].max() -
data['capital_gain'].min()

# Calculate variance and range for 'capital-loss'
capital_loss_variance = data['capital_loss'].var()
capital_loss_range = data['capital_loss'].max() -
data['capital_loss'].min()

# Display the results
```

```
print(f"\nVariance and Range for 'age':\nVariance:
{age_variance}\nRange: {age_range}")
print(f"\nVariance and Range for 'capital-gain':\nVariance:
{capital_gain_variance}\nRange: {capital_gain_range}")
print(f"\nVariance and Range for 'capital-loss':\nVariance:
{capital_loss_variance}\nRange: {capital_loss_range}")
```

Explanation:

1. **Read Data**: The script reads the Adult dataset from a URL and loads it into a Pandas DataFrame.
2. **Measure Variance and Range**:
    - **Variance**: Measure the spread of the data from the mean. It is calculated using the `.var()` method.
    - **Range**: Measure the difference between the maximum and minimum values. It is calculated using the `.max()` and `.min()` methods.