Project-2 (Pandas' CSV reader & basic analysis)

Name: Suheal Ahmad

Role: Data Science Intern

Objective:

To demonstrate basic data analysis and CSV file handling using the Pandas library in Python.

Codes:

```
import pandas as pd

# Creating a sample dataset
data = {
    "Name": ["Amit", "Riya", "Sohan", "Neha", "Rahul"],
    "Age": [23, 28, 25, 30, 27],
    "Department": ["HR", "IT", "Finance", "IT", "HR"],
    "Salary": [35000, 55000, 48000, 60000, 42000]
}

df = pd.DataFrame(data)

# Save dataset to CSV
df.to_csv("employee_data.csv", index=False)

print("Sample dataset created successfully!")
```

```
Sample dataset created successfully!
```

```
# Read CSV file
df = pd.read_csv("employee_data.csv")

# Inspect data
print("First 5 rows:\n", df.head())
print("\nLast 5 rows:\n", df.tail())
print("\nData Types:\n", df.dtypes)

# Summary statistics
print("\nSummary Statistics:\n", df.describe())

# Filtering rows
filtered_data = df[df["Age"] > 25]

# Selecting columns
selected_columns = df[["Name", "Salary"]]

# Slicing data
subset = df.iloc[0:3, 0:3]

# Save results
filtered_data.to_csv("filtered_employee_data.csv", index=False)
selected_columns.to_excel("selected_columns.xlsx", index=False)

print("\nFiles saved successfully!")
```

Output:

```
First 5 rows:
     Name  Age Department  Salary
0   Amit   23         HR   35000
1   Riya   28         IT   55000
2  Sohan   25    Finance   48000
3   Neha   30         IT   60000
4  Rahul   27         HR   42000

Last 5 rows:
     Name  Age Department  Salary
0   Amit   23         HR   35000
1   Riya   28         IT   55000
2  Sohan   25    Finance   48000
3   Neha   30         IT   60000
4  Rahul   27         HR   42000

Data Types:
 Name           object
Age             int64
Department     object
Salary          int64
dtype: object

Summary Statistics:
              Age         Salary
count   5.000000       5.000000
mean   26.600000   48000.000000
std     2.701851    9974.968672
min    23.000000   35000.000000
25%    25.000000   42000.000000
50%    27.000000   48000.000000
75%    28.000000   55000.000000
max    30.000000   60000.000000

Files saved successfully!
```

Summary:

- *Created a sample dataset using Pandas and saved it as a CSV file.*
- *Successfully read the CSV file into a DataFrame.*
- *Inspected the dataset using head(), tail(), and dtypes().*
- *Generated descriptive statistics including mean, minimum, maximum, and count.*
- *Filtered rows based on conditions and selected specific columns.*
- *Extracted subsets of data using slicing.*
- *Exported processed data into CSV and Excel formats.*

Result: *Demonstrated complete understanding of CSV handling and basic data analysis using Pandas.*