Week 3: NumPy and Pandas for Data Manipulation

Objective

Learn NumPy arrays and operations for efficient data manipulation.
Work with Pandas DataFrames for data processing.
Develop a script for data cleaning and aggregation.

Tasks Completed

Basic NumPy Operations:
- Created and performed operations on NumPy arrays.
- Computed statistical values like mean and sum.

Pandas DataFrame Operations:
- Created and manipulated Pandas DataFrames.
- Handled missing values and performed dataset transformations.

Client Project:
- Developed a script to clean and aggregate a dataset by removing missing values and computing averages.

Python Scripts

1. NumPy Array Operations

```
python
import numpy as np
arr = np.array([1, 2, 3, 4, 5])
print("Array:", arr)
print("Mean:", np.mean(arr))
print("Sum:", np.sum(arr))
```

OUTPUT:

```
================ RESTART: /Users/achu/Documents/week3_task_q2.py ===============
Array: [1 2 3 4 5]
Mean: 3.0
Sum: 15
```

2. Pandas DataFrame Creation

```
python
import pandas as pd
data = {'Name': ['Alice', 'Bob', 'Charlie'],
        'Age': [25, 30, 35],
        'Salary': [50000, 60000, 70000]}
df = pd.DataFrame(data)
print(df)
```

OUTPUT:

```
>>> ========== RESTART: /Users/achu/Documents/week3_2.py ==========
      Name  Age  Salary
0    Alice   25   50000
1      Bob   30   60000
2  Charlie   35   70000
>>>
```

3. Data Cleaning Script (Handling Missing Values)

```
python
import pandas as pd
data = {'Name': ['Alice', 'Bob', 'Charlie', None],
        'Age': [25, 30, 35, None],
        'Salary': [50000, None, 70000, 60000]}
df = pd.DataFrame(data)
# Fill missing values
df.fillna({'Name': 'Unknown', 'Age': df['Age'].mean(), 'Salary': df['Salary'].mean()}, inplace=True)
print(df)
```

OUTPUT:

```
>>> ========== RESTART: /Users/achu/Documents/week3_2.py ==========
      Name   Age   Salary
0    Alice  25.0  50000.0
1      Bob  30.0  60000.0
2  Charlie  35.0  70000.0
3  Unknown  30.0  60000.0
>>>
```

Key Learnings

Used NumPy for numerical operations and efficiency.

Understood how to work with Pandas DataFrames for data analysis.

Learned how to clean and handle missing values in datasets.

Conclusion

Week 3 introduced powerful data manipulation tools through NumPy and Pandas.
The hands-on experience improved efficiency in working with datasets.