29-10-2024

Training Day – 29

- *Topic:* Pandas Data Manipulation
  - Performed filtering, sorting, and adding new columns.
  - Example: Filtered rows where column values exceeded a threshold.

```
# Import necessary libraries
import pandas as pd

# Create a sample dataset
data = {
    'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eve'],
    'Age': [25, 32, 18, 45, 22],
    'Salary': [50000, 60000, 32000, 78000, 45000],
    'Department': ['HR', 'IT', 'Finance', 'Marketing', 'HR']
}

# Convert to a DataFrame
df = pd.DataFrame(data)

# Display the original dataset
print("Original Dataset:")
display(df)

# 1. Filtering rows where Salary > 45000
filtered_df = df[df['Salary'] > 45000]
print("\nFiltered Dataset (Salary > 45000):")
display(filtered_df)

# 2. Sorting by Age (ascending)
sorted_df = df.sort_values(by='Age')
print("\nDataset Sorted by Age:")
display(sorted_df)

# 3. Adding a new column 'Seniority'
# Seniority is 'Senior' if Age > 30, otherwise 'Junior'
df['Seniority'] = ['Senior' if age > 30 else 'Junior' for age in df['Age']]
print("\nDataset with New Column 'Seniority':")
```

```
display(df)

# Exporting to CSV for further analysis if needed
df.to_csv('pandas_manipulation_example.csv', index=False)
```

1. **Original Dataset**:

markdown Copy code

```
    Name  Age  Salary Department
0   Alice   25  50000         HR
1     Bob   32  60000         IT
2 Charlie   18  32000    Finance
3   David   45  78000  Marketing
4     Eve   22  45000         HR
```

2. **Filtered Dataset** (`Salary > 45000`):

markdown Copy code

```
    Name  Age  Salary Department
0   Alice   25  50000         HR
1     Bob   32  60000         IT
3   David   45  78000  Marketing
```

3. **Sorted Dataset** (by `Age`):