```
import numpy as np
import pandas as pd
```

DATASET 2:Students Performance Dataset

Purpose: Sorting,value counts, apply,conditional columns

Create Dataset

```
np.random.seed(2)
rows = 150

df2 = pd.DataFrame({
    "StudentId": np.arange(1, rows + 1),
    "Age": np.random.randint(18, 30, rows),
    "Gender": np.random.choice(["male", "female"], rows),
    "Course": np.random.choice(["AI", "DS", "ML", "Python"], rows),
    "Marks": np.random.randint(35, 100, rows)
})
```

Sorting Data

```
df2.sort_values(by="Marks", ascending=False)
```

	StudentId	Age	Gender	Course	Marks
47	48	26	female	DS	99
30	31	26	female	ML	99
140	141	26	female	DS	99
145	146	25	female	ML	99
19	20	22	female	Python	98
...	...	...	...	...	...
110	111	29	female	ML	38
40	41	28	female	ML	37
144	145	18	female	Python	37
14	15	22	female	ML	35
11	12	23	female	ML	35

150 rows × 5 columns

Value Counts

```
df2["Gender"].value_counts()
```

Gender	count
female	84
male	66

dtype: int64

Apply Function

```
df2["Result"] = df2["Marks"].apply(lambda x: "Pass" if x >= 50 else "Fail")
df2.head()
```

	StudentId	Age	Gender	Course	Marks	Result
0	1	26	male	AI	65	Pass
1	2	26	female	DS	41	Fail
2	3	24	male	DS	56	Pass
3	4	29	female	ML	52	Pass
4	5	20	female	AI	72	Pass

Grouping

```
df2.groupby("Course")["Marks"].mean()
```

Course	Marks
AI	68.185185
DS	64.611111
ML	62.125000
Python	68.282051

dtype: float64