FOUNDATION LESSON

WEEK 5 · LESSON 6

Pivot tables and integrated cleaning

A pivot summary changes the display grain and needs an explicit aggregation rule.

THE MENTAL MODEL

MENTAL MODEL

The mental model

A pivot summary changes the display grain and needs an explicit aggregation rule.

REGION BY MONTH

CODE WALKTHROUGH W05-L06-S03

```
01  import pandas as pd
02  d=pd.DataFrame({"region":["N","N","S"],
03   "month":["Jan","Feb","Jan"],"amount":[10,20,30]})
04  p=d.pivot_table(index="region",columns="month",values="amount",
05                  aggfunc="sum",fill_value=0)
06  print(p.to_string())
07  print("Control:", p.to_numpy().sum())
```

WHAT TO TRACK

Region by month

PREDICT THE RESULT

Predict the output.
Explain each operation.
Pause before continuing.

REGION BY MONTH · RESULT

CODE WALKTHROUGH W05-L06-S04

```
01  import pandas as pd
02  d=pd.DataFrame({"region":["N","N","S"],
03   "month":["Jan","Feb","Jan"],"amount":[10,20,30]})
04  p=d.pivot_table(index="region",columns="month",values="amount",
05                  aggfunc="sum",fill_value=0)
06  print(p.to_string())
07  print("Control:", p.to_numpy().sum())
```

WHAT TO TRACK

Region by month

VALIDATE THE RESULT

```
month   Feb  Jan
region
N        20   10
S         0   30
Control: 60
```

CLEAN, JOIN, SUMMARISE

CODE WALKTHROUGH W05-L06-S05

```
01  import pandas as pd
02  o=pd.DataFrame({"customer":["A","B","A"],
03   "amount":["10","bad","20"]})
04  c=pd.DataFrame({"customer":["A","B"],"region":["N","S"]})
05  o["amount"]=pd.to_numeric(o["amount"],errors="coerce")
06  valid=o.loc[o["amount"].notna()].copy()
07  j=valid.merge(c,on="customer",how="left",validate="many_to_one"
↪   )
08  print("Rejected:", len(o)-len(valid))
09  print(j.groupby("region")["amount"].sum().to_dict())
```

WHAT TO TRACK

Clean, join, summarise

PREDICT THE RESULT

Predict the output.
Explain each operation.
Pause before continuing.

CLEAN, JOIN, SUMMARISE · RESULT

CODE WALKTHROUGH W05-L06-S06

```
01 import pandas as pd
02 o=pd.DataFrame({"customer":["A","B","A"],
03  "amount":["10","bad","20"]})
04 c=pd.DataFrame({"customer":["A","B"],"region":["N","S"]})
05 o["amount"]=pd.to_numeric(o["amount"],errors="coerce")
06 valid=o.loc[o["amount"].notna()].copy()
07 j=valid.merge(c,on="customer",how="left",validate="many_to_one"
↪ )
08 print("Rejected:", len(o)-len(valid))
09 print(j.groupby("region")["amount"].sum().to_dict())
```

WHAT TO TRACK

Clean, join, summarise

VALIDATE THE RESULT

```
Rejected: 1
{'N': 30.0}
```

AN INVALID PIVOT GRAIN

CODE WALKTHROUGH W05-L06-S07

```
01  import pandas as pd
02  d=pd.DataFrame({"region":["N","N"],
03   "month":["Jan","Jan"],"amount":[10,20]})
04  print(d.pivot(index="region",columns="month",values="amount"))
```

WHAT TO TRACK

An invalid pivot grain

PREDICT THE RESULT

Predict the output.
Explain each operation.
Pause before continuing.

AN INVALID PIVOT GRAIN · RESULT

CODE WALKTHROUGH W05-L06-S08

```
01  import pandas as pd
02  d=pd.DataFrame({"region":["N","N"],
03   "month":["Jan","Jan"],"amount":[10,20]})
04  print(d.pivot(index="region",columns="month",values="amount"))
```

WHAT TO TRACK

An invalid pivot grain

VALIDATE THE RESULT

```
ValueError: Index contains
duplicate entries, cannot
reshape
```

GUIDED PRACTICE

LEARNER PRACTICE

Guided practice

Pivot N/Jan amounts 4 and 6 using sum.
Check that the single cell is 10.

PREDICT · MODIFY · RUN

01 Restate the supplied rules.

02 Predict the result.

03 Modify and run the code.

04 Compare actual output.

05 Explain any difference.

INDEPENDENT PRACTICE

LEARNER PRACTICE

Independent practice

Clean ["8", "bad", "-2"] under the rule numeric and nonnegative. Print accepted total and rejected count.

PREDICT · MODIFY · RUN

01 Restate the supplied rules.

02 Predict the result.

03 Modify and run the code.

04 Compare actual output.

05 Explain any difference.

CHECK YOUR UNDERSTANDING

MENTAL MODEL

Check your understanding

Diagnostic: May fill_value=0 prove no orders occurred? No. That meaning must come from the dataset’s coverage rules.

RECAP AND NEXT STEP

Recap and next step

Validate before reshaping. Declare the pivot aggregation and missing-cell meaning, then reconcile the accepted total.

UNDERSTAND APPLY VALIDATE EXPLAIN