FOUNDATION LESSON

WEEK 5 · LESSON 3

Tidy column names and groupby

One row has a stated grain; grouping partitions those rows by a key.

THE MENTAL MODEL

MENTAL MODEL

The mental model

One row has a stated grain; grouping partitions those rows by a key.

RENAME A COLUMN

CODE WALKTHROUGH W05-L03-S03

```
01  import pandas as pd
02  df = pd.DataFrame({"Sales EUR": [10, 20]})
03  clean = df.rename(columns={"Sales EUR": "sales_eur"})
04  print(clean.columns.tolist())
05  print(clean["sales_eur"].sum())
```

WHAT TO TRACK

Rename a column

PREDICT THE RESULT

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

RENAME A COLUMN · RESULT

CODE WALKTHROUGH W05-L03-S04

```
01  import pandas as pd
02  df = pd.DataFrame({"Sales EUR": [10, 20]})
03  clean = df.rename(columns={"Sales EUR": "sales_eur"})
04  print(clean.columns.tolist())
05  print(clean["sales_eur"].sum())
```

WHAT TO TRACK

Rename a column

VALIDATE THE RESULT

```
['sales_eur']
30
```

REVENUE BY REGION

CODE WALKTHROUGH W05-L03-S05

```
01  import pandas as pd
02  df = pd.DataFrame({"region": ["East", "West", "East"],
03                     "revenue_eur": [100, 80, 50]})
04  summary = df.groupby("region")["revenue_eur"].sum()
05  print(summary.to_string())
06  print("Control:", summary.sum(), df["revenue_eur"].sum())
```

WHAT TO TRACK

Revenue by region

PREDICT THE RESULT

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

REVENUE BY REGION · RESULT

CODE WALKTHROUGH W05-L03-S06

```
01  import pandas as pd
02  df = pd.DataFrame({"region": ["East", "West", "East"],
03                     "revenue_eur": [100, 80, 50]})
04  summary = df.groupby("region")["revenue_eur"].sum()
05  print(summary.to_string())
06  print("Control:", summary.sum(), df["revenue_eur"].sum())
```

WHAT TO TRACK

Revenue by region

VALIDATE THE RESULT

```
region
East    150
West     80
Control: 230 230
```

AN UNKNOWN REGION

CODE WALKTHROUGH W05-L03-S07

```
01  import pandas as pd
02  df = pd.DataFrame({"region": ["East", None],
03                     "amount": [10, 20]})
04  print(df.groupby("region", dropna=False)["amount"].sum())
```

WHAT TO TRACK

An unknown region

PREDICT THE RESULT

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

AN UNKNOWN REGION · RESULT

CODE WALKTHROUGH W05-L03-S08

```
01  import pandas as pd
02  df = pd.DataFrame({"region": ["East", None],
03                     "amount": [10, 20]})
04  print(df.groupby("region", dropna=False)["amount"].sum())
```

WHAT TO TRACK

An unknown region

VALIDATE THE RESULT

```
region
East    10
NaN     20
Name: amount, dtype: int64
```

GUIDED PRACTICE

LEARNER PRACTICE

Guided practice

Group A, B, A with amounts 5, 7, 8. Print group totals and the overall total.

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

Rename Amount EUR to amount_eur and group by region, retaining a missing region.

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: Does a regional sum retain one row per order? No. It reduces the table to one summary per region.

RECAP AND NEXT STEP

Recap and next step

Name variables clearly, state the row grain, choose the grouping key, and reconcile the result.

UNDERSTAND APPLY VALIDATE EXPLAIN