FOUNDATION LESSON

WEEK 6 · LESSON 8

README documentation and data provenance

Provenance records origin and identity; the README explains how to reproduce and interpret the analysis.

THE MENTAL MODEL

MENTAL MODEL

The mental model

Provenance records origin and identity; the README explains how to reproduce and interpret the analysis.

WRITE EXPLICIT CONVENTIONS

CODE WALKTHROUGH W06-L08-S03

```
01  from pathlib import Path
02  text="""# Fictional quarterly business analysis
03  Run from the project root with the supplied dependencies.
04  Units: EUR. Period: 2025 Q1.
05  CapEx >= 0 denotes a cash outflow; FCF = CFO - CapEx.
06  ROA = quarter profit / mean(beginning assets, ending assets).
07  ROA is quarterly. Data are synthetic, not investment evidence.
08  """
09  Path("README.md").write_text(text,encoding="utf-8")
10  print(Path("README.md").read_text(encoding="utf-8"))
```

WHAT TO TRACK

Write explicit conventions

PREDICT THE RESULT

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

WRITE EXPLICIT CONVENTIONS · RESULT

CODE WALKTHROUGH W06-L08-S04 1/2

```
01 from pathlib import Path
02 text="""# Fictional quarterly business analysis
03 Run from the project root with the supplied dependencies.
04 Units: EUR. Period: 2025 Q1.
05 CapEx >= 0 denotes a cash outflow; FCF = CFO - CapEx.
06 ROA = quarter profit / mean(beginning assets, ending assets).
07 ROA is quarterly. Data are synthetic, not investment evidence.
08 """
09 Path("README.md").write_text(text,encoding="utf-8")
10 print(Path("README.md").read_text(encoding="utf-8"))
```

WHAT TO TRACK

Write explicit conventions

VALIDATE THE RESULT

```
# Fictional quarterly
business analysis
Run from the project root
with the supplied
dependencies.
Units: EUR. Period: 2025 Q1.
CapEx >= 0 denotes a cash
outflow; FCF = CFO - CapEx.
ROA = quarter profit /
mean(beginning assets,
ending assets).
ROA is quarterly. Data are
```

WRITE EXPLICIT CONVENTIONS · RESULT

CODE WALKTHROUGH W06-L08-S04 2/2

WHAT TO TRACK

Write explicit conventions

VALIDATE THE RESULT

synthetic, not investment evidence.

HASH AN ACTUAL INPUT FILE

CODE WALKTHROUGH W06-L08-S05

```
01  from pathlib import Path
02  import hashlib,json
03  p=Path("flows.csv")
04  p.write_text("entity,revenue\nA,100\n",encoding="utf-8")
05  record={"file":p.name,"source":"synthetic teaching data",
06          "sha256":hashlib.sha256(p.read_bytes()).hexdigest()}
07  Path("w06_data_provenance.json").write_text(
08      json.dumps(record,indent=2),encoding="utf-8")
09  print("Source:",record["source"])
10  print("SHA-256 length:",len(record["sha256"]))
```

WHAT TO TRACK

Hash an actual input file

PREDICT THE RESULT

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

HASH AN ACTUAL INPUT FILE · RESULT

CODE WALKTHROUGH W06-L08-S06

```
from pathlib import Path
import hashlib,json
p=Path("flows.csv")
p.write_text("entity,revenue\nA,100\n",encoding="utf-8")
record={"file":p.name,"source":"synthetic teaching data",
        "sha256":hashlib.sha256(p.read_bytes()).hexdigest()}
Path("w06_data_provenance.json").write_text(
    json.dumps(record,indent=2),encoding="utf-8")
print("Source:",record["source"])
print("SHA-256 length:",len(record["sha256"]))
```

WHAT TO TRACK

Hash an actual input file

VALIDATE THE RESULT

```
Source: synthetic teaching
data
SHA-256 length: 64
```

A CHANGED FILE HAS A CHANGED IDENTITY

CODE WALKTHROUGH W06-L08-S07

```
01  import hashlib
02  before=b"A,100\n"
03  after=b"A,101\n"
04  same=hashlib.sha256(before).digest()==hashlib.sha256(after).dig
↪   est()
05  print("Same bytes according to hash:",same)
```

WHAT TO TRACK

A changed file has a changed identity

PREDICT THE RESULT

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

A CHANGED FILE HAS A CHANGED IDENTITY · RESULT

CODE WALKTHROUGH W06-L08-S08

```
01  import hashlib
02  before=b"A,100\n"
03  after=b"A,101\n"
04  same=hashlib.sha256(before).digest()==hashlib.sha256(after).dig
↪   est()
05  print("Same bytes according to hash:",same)
```

WHAT TO TRACK

A changed file has a changed identity

VALIDATE THE RESULT

```
Same bytes according to
hash: False
```

GUIDED PRACTICE

LEARNER PRACTICE

Guided practice

Write a README statement explaining positive-outflow CapEx and quarterly average-asset ROA.

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

Write a synthetic CSV, calculate its SHA-256, and store filename, source and hash in JSON.

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 matching SHA-256 prove the financial values are correct? No. It checks byte identity, not business truth.

RECAP AND NEXT STEP

Recap and next step

The README carries run instructions and analytical conventions. Provenance records synthetic origin and hashes of actual input bytes.

UNDERSTAND APPLY VALIDATE EXPLAIN