

Wrangling a dataset **with Python**

```
import csv
from collections import defaultdict

with open("loans.csv", newline="", encoding="utf-8") as f:
    raw_rows = list(csv.DictReader(f))

rows, report = clean_rows(raw_rows)

totals = defaultdict(int)
for row in rows:
    totals[row["branch"]] += row["loans"]

ordered = sorted(totals.items(), key=lambda kv: -kv[1])
for branch, total in ordered:
    print(branch, f"{total:,}")
```

Cleaning report

Rows read	125
Duplicates removed	5
Dropped (no loans)	4
Rows clean	116

Total loans by branch

Central	15,749
Riverside	13,239
Harbor	11,950
Hillside	9,322

wrangle.py uses only the standard library. **wrangle_pandas.py** gives the same numbers in fewer lines.

Talking to a **local model**

```
from ollama import chat

response = chat(
    model="olmo-3:7b-instruct",
    messages=[{"role": "user",
                "content": "Explain a dictionary."}],
)
print(response.message.content)
```

SET UP ONCE

```
pip install ollama
ollama pull olmo-3:7b-instruct
```

- Messages have roles: system, user, assistant, tool.
- The model runs on your machine through the Ollama server.
- Nothing leaves your computer.
- Do not paste a whole table into the prompt: it may not fit, and models miscount. Let Python do the arithmetic.

Building a **local agent**

You ask

Model replies

Script runs tool

Result goes back

Model answers

```
for step in range(1, max_steps + 1):
    response = client.chat(model=model,
                           messages=messages,
                           tools=tools)

    message = response.message
    messages.append(message)
    if not message.tool_calls:
        return message.content
    for call in message.tool_calls:
        tool = registry.get(call.function.name)
        result = tool(**call.function.arguments)
        messages.append({"role": "tool", "content": result})
```

Simplified from local_agent.py. The real script also checks arguments and logs every call.

Safe by design

- Narrow, read-only tools
- Allowlist every argument
- Step limit and a log
- Ask a human before any write
- Treat data text as untrusted

Scripts and agents **checklist**

- ☐ Is the raw data left untouched?
- ☐ Do I check types, blanks and duplicates?
- ☐ Can I run it twice and get the same result?
- ☐ Did I check a few rows by hand?
- ☐ Are the tools narrow and read-only?
- ☐ Do I check arguments before running anything?
- ☐ Is there a step limit and a log of every call?
- ☐ Did I verify answers and keep secrets out?

DISCUSSION

Which step of your own data work would you automate first? What could go wrong if the script silently skipped bad rows?

An agent can only call read-only tools. What harm is still possible, and how would you notice?