

Python for spreadsheet work: total, combine and clean exports with pandas

The three modules of the free Python for Spreadsheet Work Jump Start, practised there on real Google Colab screens.

Colab basics

Start

colab.research.google.com, **New notebook**.

Run

Shift+Enter runs a cell and opens a new one.

Load

```
import pandas as pd
```

Read

`sales = pd.read_csv('sales.csv')` (upload it in the Files panel).

Total and rank

- `sales['amount'].sum()` adds one column.
- `totals = sales.groupby('customer')['amount'].sum()`
- `totals.sort_values(ascending=False)` biggest first.

Combine monthly exports

- Label first: `july['month'] = 'Jul'` (same for each month).
- `quarter = pd.concat([july, august, september], ignore_index=True)`
- `quarter.groupby('month', sort=False)['amount'].sum()`

Clean before you total

- `df['client'] = df['client'].str.strip().str.title()`
- `df['amount'] = df['amount'].str.replace(',', '').astype(int)`
- `df = df.drop_duplicates()` after the clean-up, not before.

Good habits

- Show the table after every change: its name on the last line.
- Clean, then de-duplicate, then total.
- Keep the notebook: next month only the file name changes.

Remember: Write the steps once, run them on every export: same result every time, no missed rows and no double counts.

Practice it free: Python: Total a Sales Export by Customer in Google Colab at learn.ynteractive.com/modules/python-sales-report. This sheet: learn.ynteractive.com/content/cheat-sheets/python-spreadsheet-work. Lakeside Office Supply is a made-up business from the lesson. By Robert Breen, October 24, 2026.