

Checklist for data loading issues

Use this quick checklist when a dataset won't load. These are the eight most common causes.

Quick checklist

- ✓ Saved the notebook or script
- ✓ Notebook/script is inside the project folder
- ✓ No typos, case-errors, and file extension included
- ✓ Using the correct pandas reader for the file type
- ✓ For text files: correct delimiter and header
- ✓ For Excel: correct sheet, range, and header
- ✓ Installed required dependencies (openpyxl/pyarrow)
- ✓ Activated the correct virtual environment

Detailed checklist

1. Have you saved the notebook or script?

Otherwise you may be working in a temporary directory and your relative path might be incorrect.

2. Is your notebook or script located in the project directory (or a subfolder)?

Otherwise relative paths might not work as expected.

3. Have you double-checked the specified path?

After the first two checkpoints you are sure where your data is located **relative** to the script or notebook.

But there are still many possible **typos** that may point to the wrong file path. In particular, capitalization and spaces in file names are prone to typos since paths are **case-sensitive** and every character (even 'invisible' ones) count.

Also, make sure to include the **file extension** (e.g. '.csv', '.xlsx')!

4. Are you using the appropriate function to read the file type?

- `pd.read_csv()` or `pd.read_table()` — text files (CSV/TSV)
- `pd.read_excel()` — Excel files
- `pd.read_parquet()` — Parquet files

5. For text files, have you specified the correct delimiter and header?

`pd.read_csv()` and `pd.read_table()` let you specify the delimiter and header row. Check the documentation for details.

6. For Excel files, have you specified the sheet name, range, and header?

`pd.read_excel()` lets you specify the sheet name, range, and header row. Check the documentation for details.

7. For Excel, Parquet, and other binary formats, have you installed the necessary dependencies?

Install any required libraries, for example: - Excel: `openpyxl` - Parquet: `pyarrow`

8. Have you selected/activated the correct virtual environment?

You may be using the global environment instead of the project's virtual environment. Activate it before running the notebook, and install project dependencies into that environment (not the global one).