Effective Programming Practices for Economists

Basic Python

File paths with pathlib

Janoś Gabler and Hans-Martin von Gaudecker

Topics

- How to get a path to the current directory
- Working with pathlib `` `Path` ``s
- Rules for working with file paths

Motivation

- There are many ways to work with file paths in Python
- Some are not portable
- We want to give you one way that is guaranteed to work!

An example

- Consider the following directory structure

```
epp_project/
  exercises/
    exercise_1.py
    exercise_2.ipynb

  datasets/
    data.csv
```

- `epp_project`, `exercises` and `datasets` are directories
- `exercise_1.py`, `exercise_2.ipynb` and `data.csv` are files
- Want to load `data.csv` in two different scenarios:
  - from `exercise_1.py`
  - from `exercise_2.ipynb`

How not to do it

```
# don't do this!
import pandas as pd
path = `"C:\Users\MyName\Documents\epp_projects\datasets\data.csv"`
data = pd.read_csv(path)
```

- This only works on one Computer
- Backslashes (`` `\` ``) only work on windows
- Warning: This is what you get when you copy a path from your file explorer

Goal

- Start paths relative to the root folder of the project
- Only make assumptions about directory structure inside the project
- Define the path in a way that is portable across operating systems

```
from pathlib import Path

# get a path to the current directo
this_dir = Path()
print(this_dir)

# make it absolute for readability
this_dir = this_file.resolve()
print(this_file)

# move up to epp_project/
root = this_file.parent
print(root)

---
.
/home/janos/Dropbox/epp_project/exe
```

```
from pathlib import Path

# get a path to the current file
this_file = Path(__file__)
print(this_file)

# move up to epp_project/
root = this_file.parent.parent
print(root)

---

/home/janos/Dropbox/epp_project/exe
/home/janos/Dropbox/epp_project
```

From the project root to the data file

```
from pathlib import Path
# redo the .py file version in one
root = Path(__file__).parent.parent
print(root)

# go to data file
data_path = root / "datasets" / "da
print(data_path)
print(data_path.exists())

---
/home/janos/Dropbox/epp_project
/home/janos/Dropbox/epp_project/dat
True
```

- Once `` `root` `` is defined, the rest works the same in notebooks and `` `.py` `` files
- Different path snippets are defined using `` `/` ``
- Resulting path works on all platforms!

Debugging tips

- Use `path.resolve` to get full information about your path
- Use `list(path.iterdir())` to list all files in your current directory
- Build up paths one folder at a time and use `path.exists()` to catch typos

File path rules

1. Always use pathlib Paths instead of strings
2. Do not hardcode Paths outside of shared folders
3. Always concatenate paths with `` `/` ``

Remember: If you copy paste a path from your windows file explorer, all three rules are violated!