

Find Empty PDFs Using Pathlib

2025-11-20

Table of contents

Today I Learned added on 2025-11-20, learned on 2025-11-19; edited on 2026-06-14.

In generating routine forecast visualizations yesterday, I received an error thrown by `pypdf`, which indicated that the a PDF I was processing was empty (the forecast visualizations are aggregated into PDFs). After checking the usual things (i.e. that the internal R package was in fact up-to-date an installed and that I was logged-in and authenticated by Azure), I determined that I needed to actually find the empty PDF to resolve the error.

Since this seems common enough an issue, I expected there to be some one-line Python solutions. The solution I came up with may not be the best in terms of length or speed, but it worked for my use case:

```
import pathlib

pdf_files = [f for f in pathlib.Path(".").rglob("*.pdf") if
↳ f.stat().st_size == 0]
```

This line recursively gets all PDF files in the current (".") directory and all its sub-directories and adds them to the list if the size of the file is 0 (`if f.stat().st_size == 0`).