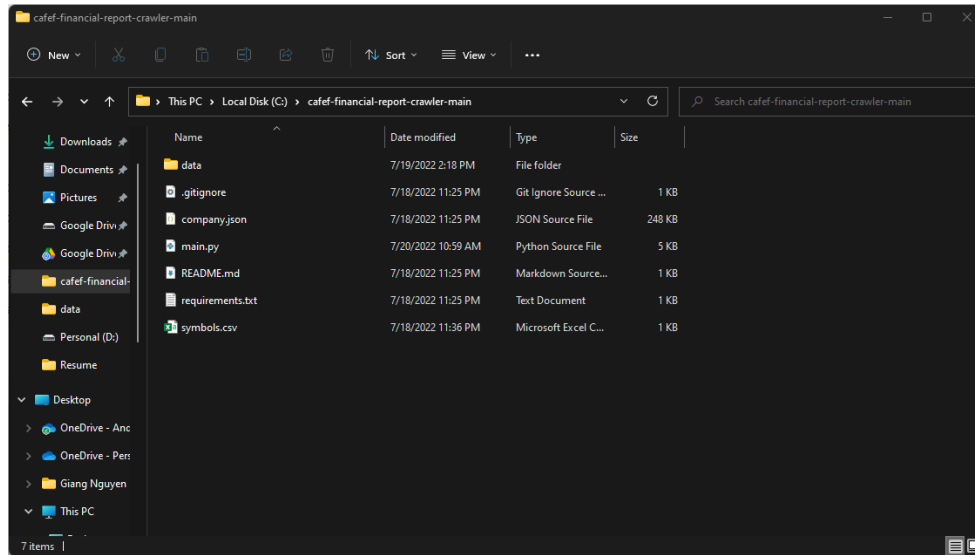


# CafeF crawler manual

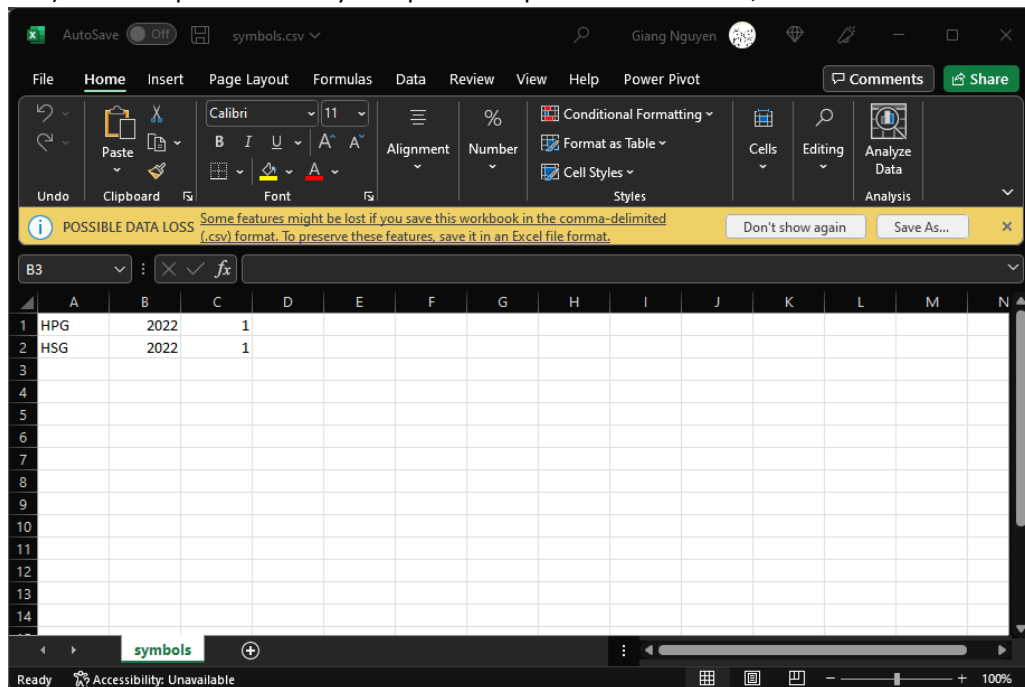
**Prerequisite:** You must install Python on your machine. [Download it here.](#)

## How to use the crawler:

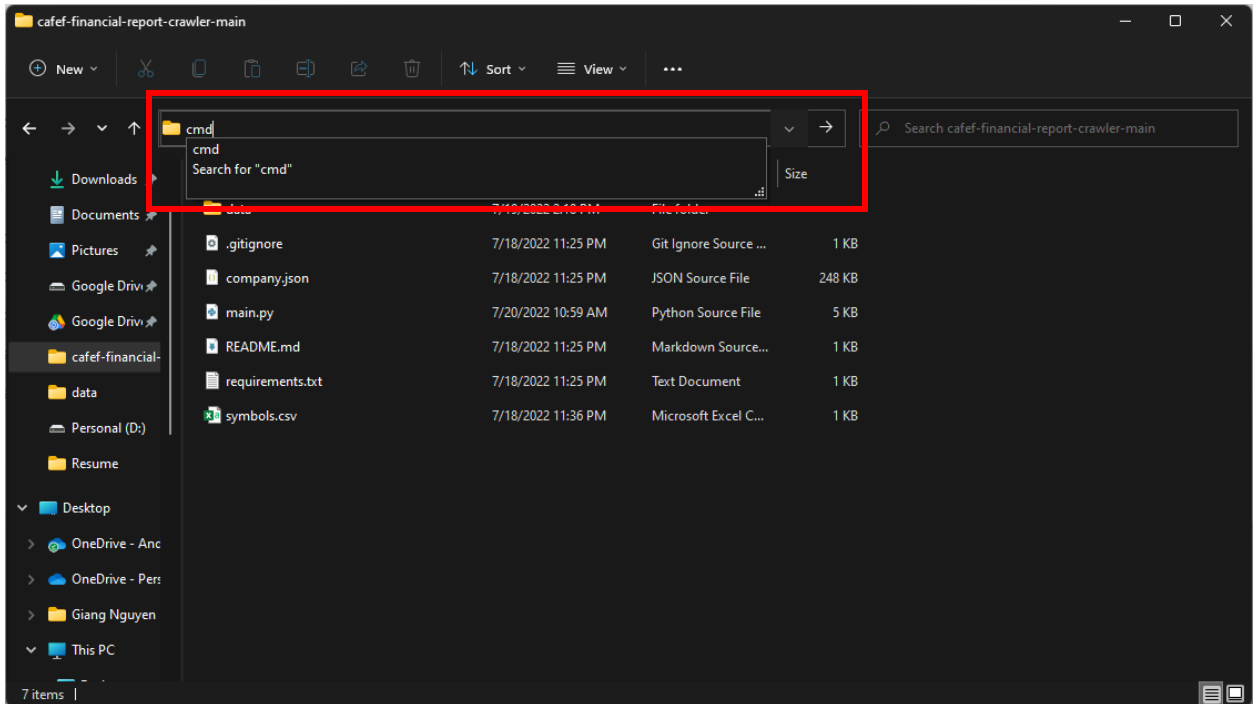
1. Download the .rar file. Extract it into a folder named: **cafef-financial-report-crawler-main**



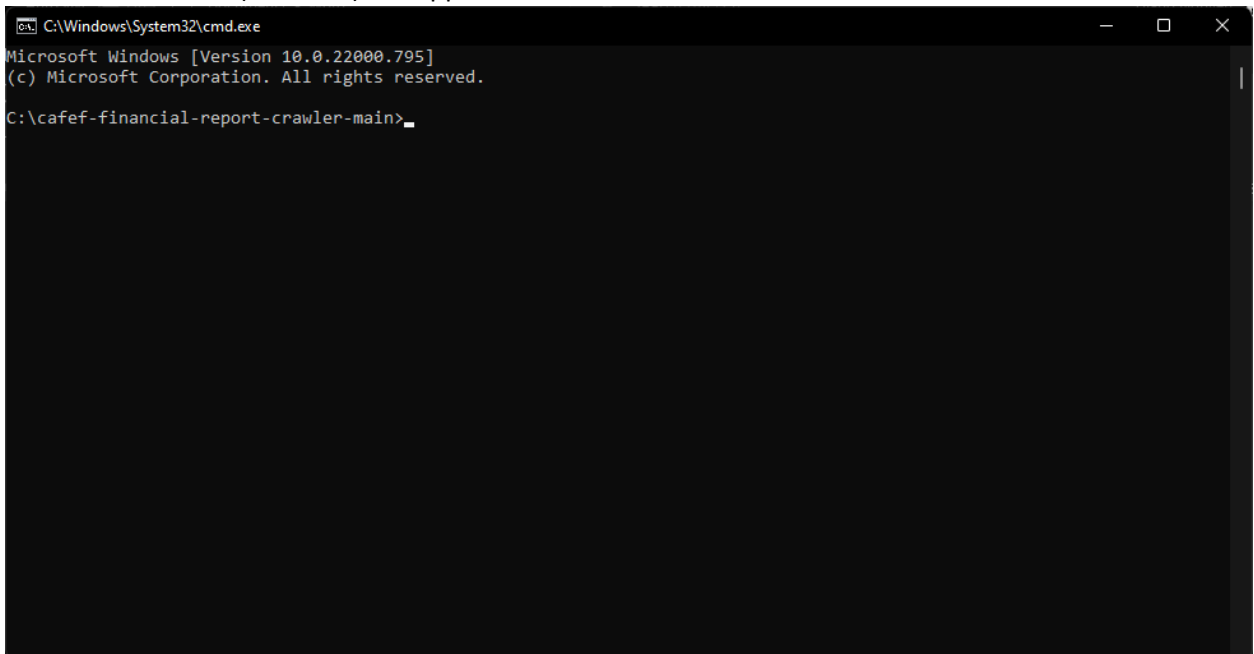
2. Open the **symbols.csv** file. Put in the company code, latest year and quarter of the FS you want to crawl (example: HPG 2022 1 if I want to get the FS of HPG in the last 4 quarters starting from Q1/2022). You can put in as many companies as possible. After that, save the file.



3. Click on the address bar (within the mentioned folder) and type **cmd**. Press **Enter**.



4. A terminal window (as below) will appear.



5. Type **pip install -r requirements.txt** then press Enter. Wait until the installation process is completed, and the path to the folder appear again.

```
C:\Windows\System32\cmd.exe
Requirement already satisfied: httpcore<0.16.0,>=0.15.0 in c:\users\admin6\appdata\local\programs\python\python310\lib\site-packages (from http[http2]==0.23.0->-r requirements.txt (line 2)) (0.15.0)
Requirement already satisfied: h2<5,>=3 in c:\users\admin6\appdata\local\programs\python\python310\lib\site-packages (from http[http2]==0.23.0->-r requirements.txt (line 2)) (4.1.0)
Requirement already satisfied: et-xmlfile in c:\users\admin6\appdata\local\programs\python\python310\lib\site-packages (from openpyxl==3.0.10->-r requirements.txt (line 4)) (1.1.0)
Requirement already satisfied: hyperframe<7,>=6.0 in c:\users\admin6\appdata\local\programs\python\python310\lib\site-packages (from h2<5,>=3->httpx[http2]==0.23.0->-r requirements.txt (line 2)) (6.0.1)
Requirement already satisfied: hpack<5,>=4.0 in c:\users\admin6\appdata\local\programs\python\python310\lib\site-packages (from h2<5,>=3->httpx[http2]==0.23.0->-r requirements.txt (line 2)) (4.0.0)
Requirement already satisfied: anyio==3.* in c:\users\admin6\appdata\local\programs\python\python310\lib\site-packages (from httpcore<0.16.0,>=0.15.0->httpx[http2]==0.23.0->-r requirements.txt (line 2)) (3.6.1)
Requirement already satisfied: h11<0.13,>=0.11 in c:\users\admin6\appdata\local\programs\python\python310\lib\site-packages (from httpcore<0.16.0,>=0.15.0->httpx[http2]==0.23.0->-r requirements.txt (line 2)) (0.12.0)
Requirement already satisfied: idna>=2.8 in c:\users\admin6\appdata\local\programs\python\python310\lib\site-packages (from anyio==3.*->httpcore<0.16.0,>=0.15.0->httpx[http2]==0.23.0->-r requirements.txt (line 2)) (3.3)
WARNING: You are using pip version 22.0.4; however, version 22.1.2 is available.
You should consider upgrading via the 'C:\Users\Admin6\AppData\Local\Programs\Python\Python310\python.exe -m pip install --upgrade pip' command.

C:\cafef-financial-report-crawler-main>
```

wait until this line appears again

6. Type **python main.py symbols.csv** then press Enter. A notification will come up to notice that the crawling process is completed.

```
C:\Windows\System32\cmd.exe
2022-07-20 11:19:18.719 | INFO | __main__:main:149 - ['HPG', '2022', '1']
Traceback (most recent call last):
  File "C:\cafef-financial-report-crawler-main\main.py", line 155, in <module>
    asyncio.run(main())
  File "C:\Users\Admin6\AppData\Local\Programs\Python\Python310\lib\asyncio\runners.py", line 44, in run
    return loop.run_until_complete(main)
  File "C:\Users\Admin6\AppData\Local\Programs\Python\Python310\lib\asyncio\base_events.py", line 646, in run_until_complete
    return future.result()
  File "C:\cafef-financial-report-crawler-main\main.py", line 150, in main
    await crawl_financial_report(symbol[0], symbol[1], symbol[2], 0, 0)
  File "C:\cafef-financial-report-crawler-main\main.py", line 125, in crawl_financial_report
    data = await crawl_financial_report_by_type(symbol, report_type, year, quarter, show_type, idx)
  File "C:\cafef-financial-report-crawler-main\main.py", line 84, in crawl_financial_report_by_type
    url=build_url(
  File "C:\cafef-financial-report-crawler-main\main.py", line 78, in build_url
    f"/bao-cai-tai-chinh-{symbol_to_slug(symbol)}.chn"
  File "C:\cafef-financial-report-crawler-main\main.py", line 72, in symbol_to_slug
    return unicode_to_ascii_with_score(get_company()[symbol])
  File "C:\cafef-financial-report-crawler-main\main.py", line 27, in get_company
    with open('company.json') as f:
FileNotFoundError: [Errno 2] No such file or directory: 'company.json'

C:\cafef-financial-report-crawler-main>python main.py symbols.csv
2022-07-20 11:19:29.359 | INFO | __main__:main:149 - ['HPG', '2022', '1']
2022-07-20 11:19:32.153 | SUCCESS | __main__:main:151 - ['HPG', '2022', '1']
2022-07-20 11:19:33.166 | INFO | __main__:main:149 - ['HSG', '2022', '1']
2022-07-20 11:19:36.064 | SUCCESS | __main__:main:151 - ['HSG', '2022', '1']

C:\cafef-financial-report-crawler-main>
```

7. Your crawled FS is in the **data** folder of **cafef-financial-report-crawler-main**.

