Data Extraction

Introduction

This folder got files and codes to extract data from IEX website for Day Ahead Market - Area Price for any date with interval of 15 minutes.

Give your date input and get all the datas in one single file.

This program will also convert the numbers saved as text in excel to numerical value. Also additional Time Block column has been added. The final excel sheet will look like.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Hour | Tim	Unnamed:	Unnamed:	Time Block	A1	A2	E1	E2	N1	N2	N3	S1	S2	S3	W1	W2	W3	MCP	
2	01-02-202	1	00:00 - 00:	1	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	3181.19	
3		1	00:15 - 00:	2	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	3136.24	
4		1	00:30 - 00:	3	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	3136.12	
5		1	00:45 - 01:	4	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	3058.25	
6		2	01:00 - 01:	5	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	2958.91	
7		2	01:15 - 01:	6	2958	2958	2958	2958	2958	2958	2958	2958	2958	2958	2958	2958	2958	2958	
8		2	01:30 - 01:	7	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	2910.22	
9		2	01:45 - 02:	8	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	2870.78	
10		3	02:00 - 02:	9	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	2870.43	
11		3	02:15 - 02:	10	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	2870.38	
12		3	02:30 - 02:	11	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	2860.82	

Setup

In the zip file, these are the folder and files
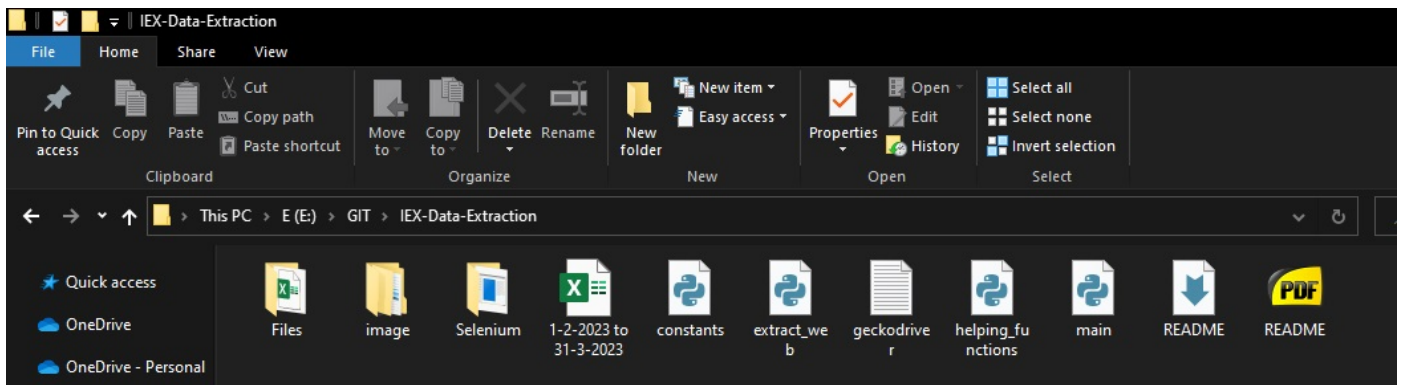
1. Extract the zip file
2. Requirement

- Python 3
  - Download Python from this website (Python 3.10) and install python.
  - This documentation can be consulted for guidance on the installation process.
- Firefox browser
  - Download firefox from here.

3. Installing python libraries After installation of python, install the required libraries. Open the Windows Command Prompt and input these commands, then press "Enter."

  - Pandas

    ```
    pip install pandas
    ```

  - selenium

    ```
    pip install selenium
    ```

    If these commands are executed correctly, it indicates the successful installation of these libraries.

Python files

The code is splitted into three:

1. main.py
2. extract_web.py
3. helping_functions.py
4. constants.py

main.py

This is the main code that needs to be run to execute the whole program.

This code will process the dates and then give this as input to other two code.

extract_web.py

This code is to work on the web and scrap the data from the website.

Here we have two input the web driver location (here Firefox web driver is used) and the path where we want to downloaded files need to be placed is also an input here.

helping_functions.py
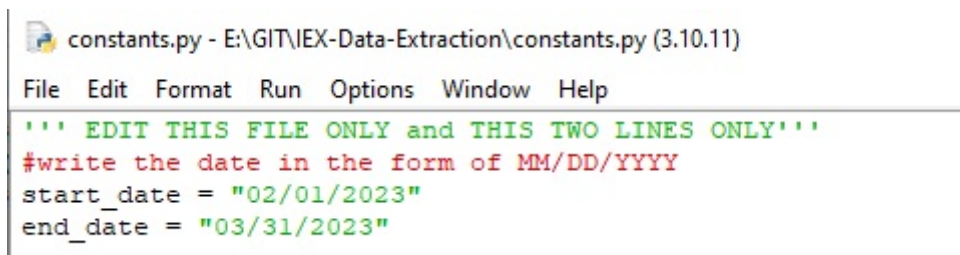
This code contains functions that we are using repeatedly for our task.

constants.py

This code contains all the input and only this file needs to be edited.

Open this file with a text editor or Python IDLE (right click on the file and click on "open with" and select this option).
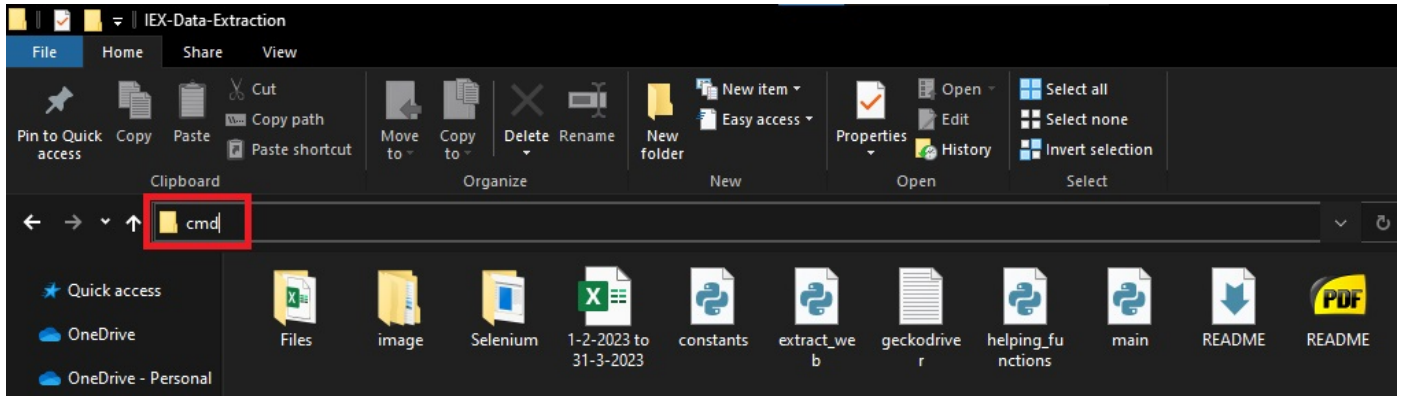
Edit the start and end date.

```
constants.py - E:\GIT\IEX-Data-Extraction\constants.py (3.10.11)
File  Edit  Format  Run  Options  Window  Help
''' EDIT THIS FILE ONLY and THIS TWO LINES ONLY'''
#write the date in the form of MM/DD/YYYY
start_date = "02/01/2023"
end_date = "03/31/2023"
```

Run the Program

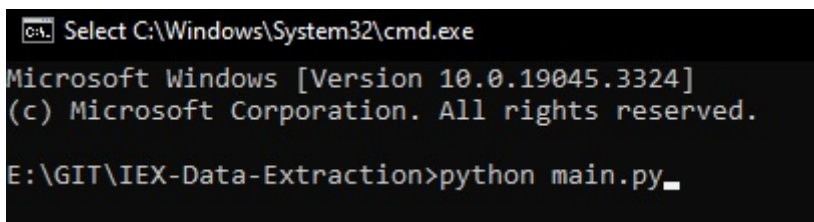To run the program, open the folder containing all the files, and type cmd in address bar.

This will open command prompt in the working folder location.
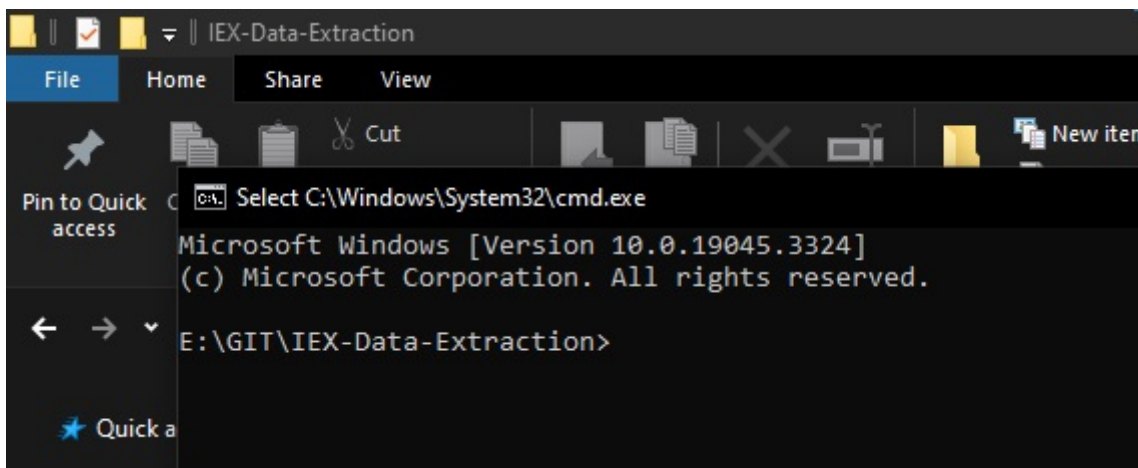
Type:

```
python main.py
```
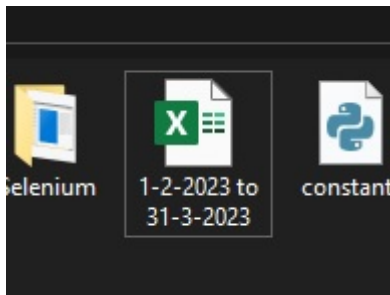

and press enter.

If the code runs successfully, it will look like this.

```
C:\Windows\System32\cmd.exe

Microsoft Windows [Version 10.0.19045.3324]
(c) Microsoft Corporation. All rights reserved.

E:\GIT\IEX-Data-Extraction>python main.py
1/Feb/2023 - 28/Feb/2023
1/Mar/2023 - 31/Mar/2023
Completed

E:\GIT\IEX-Data-Extraction>_
```

The final sheet will be saved with the title of the start and end date.

Errors

If ever a error message is shown, run the program again.