

PROGRAMMING FOR DS

PROJECT 5&6: WEB SCRAPING

Lecturer: Mr. OL Say







Start Slide

OUR TEAM(GROUP 1)





CHAN SOPHARA e20211081



DOK DOMINIQUE e20210337



CHHORN SOLITA e20210537

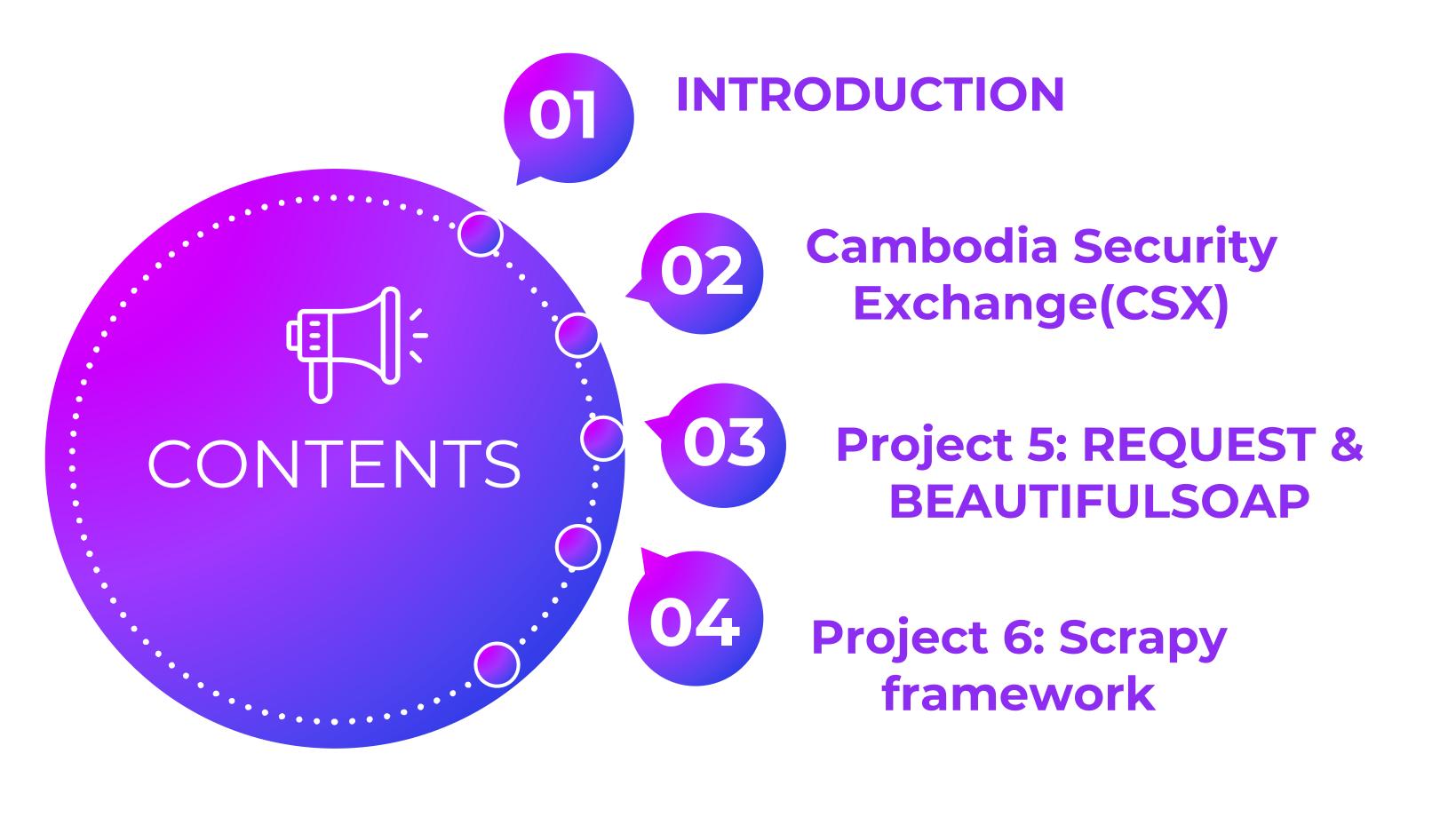


EN SREYTOCH e20210085



EN SREYTHOM e20210084





Introduction

For today subject, we are going to scrape data from website Cambodia Security Exchange(CSX) with 2 different method(Project).

- First project we work with requests and beautifulsoup4 libraries.
- Second project we work with Scrapy framework.

Detail about Project 5&6:

Scrape the historical stock prices from Cambodia Security Exchange (CSX):

A. Scrape the Phnom Penh Water Supply Authority (PWSA) historical stock price from 18th of April, 2012 to 31st May, 2024.

B. Scrape the historical stock prices of the remaining companies listed above from their joining date to the 31st of May, 2024.

Cambodia Security Exchange(CSX)

The Cambodia Securities Exchange (CSX) is the national stock exchange of Cambodia, playing a crucial role in the development of the country's capital markets. Established in 2011, CSX operates under the joint venture between the Cambodian government and the Korea Exchange, with the latter providing technical support and expertise.

Vision: To build a financial "highway" for Cambodia.



- To facilitate the raising of capital by companies in Cambodia.
- To establish investor-friendly environment for securities trading for investors in and outside Cambodia.
- To offer a variety of state-of-the-art products and services to all market participants.
- To operate a self-sustaining public enterprise under the guidance of the Royal Government of Cambodia.

LIST OF COMPANIES:

Symbol	Companies Name	Date
MJQE	MENGLY J. QUACH EDUCATION PLC.	28-Jun-2023
CGSM	CAMGSM Plc.	27-Jun-2023
JSL	JS LAND PLC	10-Feb-2022
DBDE	DBD Engineering Plc.	06-Sep-2021
PEPC	Pestech (Cambodia) Plc.	12-Aug-2020
ABC	ACLEDA Bank Plc.	25-May-2020
PAS	Sihanoukville Autonomous Port	08-Jun-2017
PPSP	Royal Group Phnom Penh SEZ Plc.	30-May-2016
PPAP	Phnom Penh Autonomous Port	09-Dec-2015
GTI	Grand Twins International (Cambodia) Plc.	16-Jun-2014
PWSA	Phnom Penh Water Supply Authority	18-Apr-2012

TYPE OF DATA WE NEED TO SCRAPE:

Trading **Trading** Full Market Cap. Market Closing Opening Date Change Value High Low Volume Cap.(Mil.KHR) Price (Mil.KHR) (KHR) (shr) 07/06/2024 7,020 **→** 20 29,415,840 7,020 7,040 7,000 610,552 3,356,080 4,192 06/06/2024 7.040 **→** 300 20,454 143,793,100 7.060 7.060 7.000 612.291 3,365,642 05/06/2024 **→** 20 20,269 148,770,280 7,360 7,360 7,300 638,383 3,509,064 7.340 04/06/2024 7,360 **20** 4,358 32,006,220 7,340 7,360 7,320 640,122 3,518,625

DATA STRUCTURE:

table=[2] div id = "Index_table" class = "summary" summary = "Index"



MAIN BOARD:

PWSA STOCK FROM 18/4/2012 TO 31/5/2024:

URL: https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH 1000010004&forma=ALL&fromDate=20120418&toDate=20240531#

• GTI STOCK FROM 16/6/2014 TO 31/5/2024:

URL: https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH_1000020003&forma=ALL&fromDate=20140616&toDate=20240531#

PPAP STOCK FROM 9/12/2015 TO 31/5/2024:

https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH 1000040001&forma=ALL&fromDate=20151209&toDate=20240531#

PPSP STOCK FROM 30/5/2016 TO 31/5/2024:

https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH 1000050000&forma=ALL&fromDate=20160530&toDate=20240531#

PAS STOCK FROM 8/JUNE/2017 TO 31/5/2024:

URL: https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH
1000060009&forma=ALL&fromDate=20170608&toDate=20240531#

ABC STOCK FROM 25/MAY/2020 TO 31/5/2024:

URL: https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH 1000100003&forma=ALL&fromDate=20200525&toDate=20240531#

• PEPC STOCK FROM 12/AUGUST/2020 TO 31/5/2024:

URL: https://csx.com.kh/data/growthdaily/listPosts.do?lang=en&MNCD=60202&board_type=G&issuecode=KH1000160007&forma=ALL&fromDate=20220210&toDate=20240531#

CGSM STOCK FROM 27/JUNE/2023 TO 31/5/2024:

URL: https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH 1000220009&forma=ALL&fromDate=20230627&toDate=20240531#

MJQE STOCK FROM 28/JUNE/2023 TO 31/5/2024:

URL: https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH 1000210000&forma=ALL&fromDate=20230628&toDate=20240531#

GROWTH BOARD:

DBDE STOCK FROM 6/SEPT/2021 TO 31/5/2024:

https://csx.com.kh/data/growthdaily/listPosts.do?lang=en&MNCD=60202&board_type=G&issuecode=KH1000150008&forma=ALL&fromDate=20210906&toDate=20240531#

• JSL STOCK FROM 10/2/2022 TO 31/5/2024:

URL: https://csx.com.kh/data/growthdaily/listPosts.do?lang=en&MNCD=60202&board_type=G&issuedecode=KH1000150008&forma=ALL&fromDate=20210906&toDate=20240531#

PROJECT 5: USING REQUEST & BEAUTIFULSOUP4

PROJECT 5

PROCESS: USING REQUESTS AND BEAUTIFULSOUP4 LIBRARIES

A. PWSA

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
URL="https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000010004&forma=ALL&fromDate=20120418&toDate=20240531#"
response=requests.get(url=URL)
soup=BeautifulSoup(markup=response.text,features="html.parser")
tables=soup.find_all(name="table")
table=tables[2]
header=table.find_all(name="th")
header_text= [th.text.strip() for th in header]
trs=table.find all(name="tr")
data=list()
for tr in trs[1::1]:
    tds=tr.find_all(name="td")
    for td in tds:
        row_text=[td.text.strip() for td in tds]
   item =dict(zip(header_text, row_text))
    data.append(item)
df=pd.DataFrame(data)
print("Company PWSA stock price from 18/4/2012 to 31/5/2024:")
```

A. PWSA OUTPUT 🔏

Company PWSA stock price from 18/4/2012 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	7,340	20	3,900	28,561,180	7,340	7,340	7,300	638,383	3,509,064
1	30/05/2024	7,320	20	5,521	40,444,260	7,340	7,360	7,320	636,644	3,499,502
2	29/05/2024	7,340	20	9,496	69,716,440	7,360	7,360	7,340	638,383	3,509,064
3	28/05/2024	7,360	20	11,257	82,751,340	7,380	7,380	7,320	640,122	3,518,625
4	27/05/2024	7,380	0	14,282	104,828,220	7,400	7,400	7,300	641,862	3,528,187
2879	24/04/2012	9,700	500	54,735	530,929,500	9,700	9,700	9,700	843,640	4,637,319
2880	23/04/2012	10,200	0	542,994	5,553,099,200	10,300	10,300	10,200	887,126	4,876,356
2881	20/04/2012	10,200	450	377,342	3,848,888,400	10,200	10,200	10,200	887,126	4,876,356
2882	19/04/2012	9,750	450	488,528	4,763,148,000	9,750	9,750	9,750	847,988	4,661,223
2883	18/04/2012	9,300	100	879,426	8,217,893,100	9,400	9,400	9,300	808,850	4,446,089

B.REMAINING COMPANIES

MAIN BOARD:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
def scrape_data(url):
    response = requests.get(url=url)
    soup = BeautifulSoup(markup=response.text, features="html.parser")
    tables = soup.find_all(name="table")
    # Assuming the table structure is consistent
    table = tables[2]
    header = table.find_all(name="th")
    header_text = [th.text.strip() for th in header]
    trs = table.find_all(name="tr")
    data = []
    for tr in trs[1:]:
       tds = tr.find_all(name="td")
        row_text = [td.text.strip() for td in tds]
       item = dict(zip(header_text, row_text))
       data.append(item)
    df = pd.DataFrame(data)
    return df
```

MAIN BOARD:

```
# URLs for the different companies

urls = {

"GTI": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000020003&forma=ALL&fromDate=20140616&toDate=20240531#",

"PPAP": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000040001&forma=ALL&fromDate=20151209&toDate=20240531#",

"PPSP": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000050000&forma=ALL&fromDate=20160530&toDate=20240531#",

"PAS": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000050000&forma=ALL&fromDate=20170600&toDate=20240531#",

"ABC": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000100003&forma=ALL&fromDate=20200525&toDate=20240531#",

"PEPC": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000140000&forma=ALL&fromDate=20200812&toDate=20240531#",

"CGSM": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000120000&forma=ALL&fromDate=20230627&toDate=20240531#",

"MJQE": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000210000&forma=ALL&fromDate=20230628&toDate=20240531#",

"MJQE": "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board_type=M&issueCode=KH1000210000&forma=ALL&fromDate=20230628&toDate=20240531#")

# Scrape data for each URL

dfs = {}

for company, url in urls.items():
    dfs[company] = scrape_data(url)
```

MAIN BOARD:

B. GTI OUTPUT:

GTI

```
# Display the DataFrame for one company as an example
print("Company GTI stock price from 16/6/2014 to 31/5/2024: ")
dfs["GTI"]
```

Company GTI stock price from 16/6/2014 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	2,350	30	705	1,650,970	2,370	2,370	2,320	94,000	94,000
1	30/05/2024	2,380	10	8,458	19,886,340	2,380	2,380	2,330	95,200	95,200
2	29/05/2024	2,370	70	7,905	18,529,040	2,300	2,400	2,300	94,800	94,800
3	28/05/2024	2,300	0	884	2,032,940	2,300	2,320	2,290	92,000	92,000
4	27/05/2024	2,300	10	377	866,130	2,290	2,300	2,290	92,000	92,000
***	***			.*** /	***				***	***
2384	23/06/2014	8,960 (SQ)	60	0	0	=	8,960	8,960	358,400	358,400
2385	20/06/2014	9,020	460	24,579	221,854,160	9,200	9,200	9,020	360,800	360,800
2386	19/06/2014	9,480 (SQ)	160	0	0	-	9,480	9,480	379,200	379,200
2387	17/06/2014	9,640	420	94,329	893,706,740	9,140	9,640	8,960	385,600	385,600
2388	16/06/2014	9,220	480	3,101	29,887,220	9,700	9,700	9,220	368,800	368,800

B. PPAP OUTPUT:

PPAP

```
# Display the DataFrame for one company as an example
print("Company PPAP stock price from 9/12/2015 to 31/5/2024: ")
dfs["PPAP"]
```

·· Company PPAP stock price from 9/12/2015 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	13,400	0	423	5,655,500	13,360	13,400	13,320	277,170	1,533,677
1	30/05/2024	13,400	20	393	5,258,660	13,380	13,400	13,200	277,170	1,533,677
2	29/05/2024	13,380	20	463	6,188,780	13,180	13,380	13,180	276,757	1,531,388
3	28/05/2024	13,400	60	664	8,860,900	13,340	13,400	13,240	277,170	1,533,677
4	27/05/2024	13,340	80	509	6,724,480	13,380	13,380	13,120	275,929	1,526,809
	•••									
2046	16/12/2015	5,380	0	4,984	26,813,920	5,380	5,380	5,380	111,282	615,760
2047	15/12/2015	5,380	80	2,396	12,706,480	5,200	5,380	5,200	111,282	615,760
2048	14/12/2015	5,300	40	28,438	146,321,520	5,220	5,320	5,120	109,627	606,603
2049	11/12/2015	5,260	120	31,929	164,041,360	5,140	5,300	5,120	108,800	602,025
2050	09/12/2015	5,140	60	13,050	67,743,780	5,200	5,200	5,140	106,318	588,291

B. PPSP OUTPUT:

PPSP

```
# Display the DataFrame for one company as an example print("Company PPSP stock price from 30/5/2016 to 31/5/2024: ") dfs["PPSP"]
```

Company PPSP stock price from 30/5/2016 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	2,350	50	68,325	160,404,950	2,380	2,380	2,330	168,906	168,906
1	30/05/2024	2,400	70	376,120	885,811,910	2,370	2,430	2,330	172,500	172,500
2	29/05/2024	2,330	60	79,341	182,938,440	2,300	2,330	2,270	167,469	167,469
3	28/05/2024	2,270	40	111,356	254,678,470	2,230	2,330	2,230	163,156	163,156
4	27/05/2024	2,230	10	10,311	22,945,950	2,220	2,230	2,220	160,281	160,281
1936	06/06/2016	2,860	0	0	0	-	2,860	2,860	165,522	205,562
1937	03/06/2016	2,860	30	2,300	6,582,000	2,880	2,880	2,860	165,522	205,562
1938	02/06/2016	2,890	0	14,000	40,326,880	2,880	2,890	2,880	167,259	207,719
1939	31/05/2016	2,890	0	500	1,445,000	2,890	2,890	2,890	167,259	207,719
1940	30/05/2016	2,890	90	24,078	71,710,090	2,980	2,980	2,890	167,259	207,719

B. PAS OUTPUT:

PAS

```
# Display the DataFrame for one company as an example
print("Company PAS stock price from 8/June/2017 to 31/5/2024: ")
dfs["PAS"]
```

Company PAS stock price from 8/June/2017 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	12,540	40	156	1,958,420	12,580	12,580	12,540	1,075,580	5,646,797
1	30/05/2024	12,500	0	285	3,564,660	12,520	12,520	12,480	1,072,150	5,628,785
2	29/05/2024	12,500	80	373	4,667,620	12,420	12,520	12,420	1,072,150	5,628,785
3	28/05/2024	12,420	120	463	5,798,480	12,520	12,540	12,420	1,065,288	5,592,761
4	27/05/2024	12,540	40	478	6,091,660	12,500	13,180	12,400	1,075,580	5,646,797
1685	14/06/2017	5,080 (SQ)	20	0	0	-	5,080	5,080	435,722	2,287,538
1686	13/06/2017	5,100	20	4,800	24,480,000	5,100	5,100	5,100	437,437	2,296,544
1687	12/06/2017	5,120	20	5,020	25,602,400	5,100	5,120	5,100	439,152	2,305,550
1688	09/06/2017	5,100	0	6,944	35,414,400	5,100	5,100	5,100	437,437	2,296,544
1689	08/06/2017	5,100	260	13,798	73,193,260	5,360	5,360	5,100	437,437	2,296,544

B. PEPC OUTPUT:

PEPC

```
# Display the DataFrame for one company as an example
print("Company PEPC stock price from 12/August/2020 to 31/5/2024: ")
dfs["PEPC"]
```

Company PEPC stock price from 12/August/2020 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	2,320	30	375	875,200	2,340	2,340	2,320	173,872	173,872
1	30/05/2024	2,350	40	121	285,330	2,380	2,380	2,310	176,121	176,121
2	29/05/2024	2,310	90	4,198	9,664,000	2,390	2,390	2,210	173,123	173,123
3	28/05/2024	2,400	110	429	988,380	2,290	2,500	2,280	179,868	179,868
4	27/05/2024	2,290	60	162	368,760	2,290	2,290	2,250	171,624	171,624
928	25/08/2020	3,450	30	18,896	64,442,810	3,420	3,450	3,380	258,560	258,560
929	24/08/2020	3,480	50	9,958	34,528,470	3,510	3,530	3,450	260,809	260,809
930	14/08/2020	3,530	30	23,058	80,710,370	3,500	3,550	3,460	264,556	264,556
931	13/08/2020	3,500	210	52,669	186,085,460	3,690	3,690	3,400	262,308	262,308
932	12/08/2020	3,710	270	220,236	817,816,860	3,980	3,990	3,590	278,046	278,046

B. CGSM OUTPUT:

CGSM

```
# Display the DataFrame for one company as an example
print("Company CGSM stock price from 27/June/2023 to 31/5/2024: ")
dfs["CGSM"]
```

Company CGSM stock price from 27/June/2023 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	2,420	10	14,977	36,405,950	2,430	2,440	2,420	4,741,436	4,741,436
1	30/05/2024	2,430	0	7,834	18,988,000	2,420	2,450	2,420	4,761,029	4,761,029
2	29/05/2024	2,430	10	17,454	42,461,740	2,440	2,440	2,430	4,761,029	4,761,029
3	28/05/2024	2,440	0	22,977	55,897,260	2,440	2,440	2,420	4,780,622	4,780,622
4	27/05/2024	2,440	20	4,879	11,865,560	2,440	2,450	2,420	4,780,622	4,780,622
						•••				
227	03/07/2023	2,410	10	61,078	147,728,870	2,400	2,440	2,400	4,721,844	4,721,844
228	30/06/2023	2,400	70	297,528	704,541,820	2,340	2,410	2,310	4,702,251	4,702,251
229	29/06/2023	2,330	30	48,564	112,422,680	2,320	2,330	2,300	4,565,102	4,565,102
230	28/06/2023	2,300	60	90,187	209,600,090	2,370	2,370	2,300	4,506,324	4,506,324
231	27/06/2023	2,360	80	481,559	1,127,039,760	2,280	2,400	2,270	4,623,880	4,623,880

B. MJQE OUTPUT:

MJQE

```
# Display the DataFrame for one company as an example
print("Company MJQE stock price from 28/June/2023 to 31/5/2024: ")
dfs["MJQE"]
```

Company MJQE stock price from 28/June/2023 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	2,040	0	14,610	29,832,760	2,040	2,050	2,040	661,081	661,081
1	30/05/2024	2,040	0	32,323	65,704,230	2,040	2,040	2,030	661,081	661,081
2	29/05/2024	2,040	0	68,163	138,544,930	2,030	2,040	2,030	661,081	661,081
3	28/05/2024	2,040	0	50,306	102,380,320	2,040	2,040	2,020	661,081	661,081
4	27/05/2024	2,040	10	37,997	77,601,670	2,050	2,050	2,040	661,081	661,081
				***	***					
226	04/07/2023	2,100	10	54,533	114,881,470	2,110	2,130	2,100	680,525	680,525
227	03/07/2023	2,110	20	64,525	135,759,270	2,130	2,130	2,090	683,766	683,766
228	30/06/2023	2,130	0	75,267	159,687,130	2,120	2,140	2,100	690,247	690,247
229	29/06/2023	2,130	70	51,530	110,060,020	2,170	2,170	2,110	690,247	690,247
230	28/06/2023	2,200	0	457,959	998,688,040	2,200	2,220	2,080	712,931	712,931

B.REMAINING COMPANIES GROWTH BOARD:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
def scrape_data(url):
    response = requests.get(url=url)
    soup = BeautifulSoup(markup=response.text, features="html.parser")
   tables = soup.find all(name="table")
   # Assuming the table structure is consistent
   table = tables[2]
    header = table.find all(name="th")
   header_text = [th.text.strip() for th in header]
   trs = table.find all(name="tr")
    data = []
   for tr in trs[1:]:
       tds = tr.find all(name="td")
        row text = [td.text.strip() for td in tds]
        item = dict(zip(header_text, row_text))
        data.append(item)
   df = pd.DataFrame(data)
   return df
```

GROWTH BOARD:

```
# URLs for the different companies
urls = {
    "DBDE": "https://csx.com.kh/data/growthdaily/listPosts.do?lang=en&MNCD=60202&board_type=G&issueCode=KH1000150008&forma=ALL&fromDate=20210906&toDate=20240531#",
    "JSL": "https://csx.com.kh/data/growthdaily/listPosts.do?lang=en&MNCD=60202&board_type=G&issueCode=KH1000160007&forma=ALL&fromDate=20220210&toDate=20240531#"
}

# Scrape data for each URL
dfs = {}
for company, url in urls.items():
    dfs[company] = scrape_data(url)
```

GROWTH BOARD

B. DBDE OUTPUT:



```
# Display the DataFrame for one company as an example
print("Company DBDE stock price from 6/Sept/2021 to 31/5/2024: ")
dfs["DBDE"]
```

Company DBDE stock price from 6/Sept/2021 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	2,230	30	33,682	74,397,380	2,200	2,250	2,190	14,409	41,169
1	30/05/2024	2,200	10	16,719	36,547,620	2,200	2,210	2,180	14,215	40,615
2	29/05/2024	2,190	0	6,563	14,317,900	2,190	2,190	2,180	14,151	40,431
3	28/05/2024	2,190	0	9,709	21,172,950	2,180	2,190	2,180	14,151	40,431
4	27/05/2024	2,190	10	2,424	5,291,070	2,190	2,190	2,180	14,151	40,431
674	10/09/2021	2,390	0	40,434	96,597,410	2,400	2,400	2,380	15,443	44,123
675	09/09/2021	2,390	10	18,974	45,507,630	2,420	2,420	2,390	15,443	44,123
676	08/09/2021	2,400	40	44,458	107,068,010	2,430	2,430	2,400	15,508	44,308
677	07/09/2021	2,440	0	52,547	127,548,650	2,450	2,450	2,410	15,766	45,046
678	06/09/2021	2,440	20	351,341	855,783,700	2,420	2,520	2,390	15,766	45,046

B. JSL OUTPUT:

JSL

```
# Display the DataFrame for one company as an example
print("Company JSL stock price from 10/2/2022 to 31/5/2024: ")
dfs["JSL"]
```

Company JSL stock price from 10/2/2022 to 31/5/2024:

	Date	ClosingPrice	Change	Trading Volume(shr)	Trading Value(KHR)	Opening	High	Low	MarketCap.(Mil.KHR)	Full Market Cap. (Mil.KHR)
0	31/05/2024	4,220	80	46	190,520	4,280	4,280	4,080	108,496	108,496
1	30/05/2024	4,300	100	900	3,788,700	4,200	4,300	4,100	110,553	110,553
2	29/05/2024	4,200	100	341	1,390,780	4,300	4,300	4,000	107,982	107,982
3	28/05/2024	4,300	20	13	55,900	4,300	4,300	4,300	110,553	110,553
4	27/05/2024	4,320	350	2,188	8,763,840	3,970	4,320	3,830	111,067	111,067
571	16/02/2022	3,530	320	680	2,400,400	3,530	3,530	3,530	90,756	90,756
572	15/02/2022	3,210	290	1,037	3,328,770	3,210	3,210	3,210	82,529	82,529
573	14/02/2022	2,920	260	3,650	10,658,000	2,920	2,920	2,920	75,073	75,073
574	11/02/2022	2,660	240	10,651	28,331,660	2,660	2,660	2,660	68,389	68,389
575	10/02/2022	2,420	220	301,445	691,387,000	2,200	2,420	2,190	62,218	62,218

PROJECT 6: USING SCRAPY FRAMEWORK

PROJECT 6

PROCESS: USING SCRAPY FRAMEWORK INSTEAD OF OF BEAUTIFULSOUP4 LIBRARY

In order to use scrapy framework we need to mport Scrapy

PS C:\Users\sopha> pip install scrapy

Start 2 projects (One for PWSA, One for Remaining Companies):
 For PWSA:

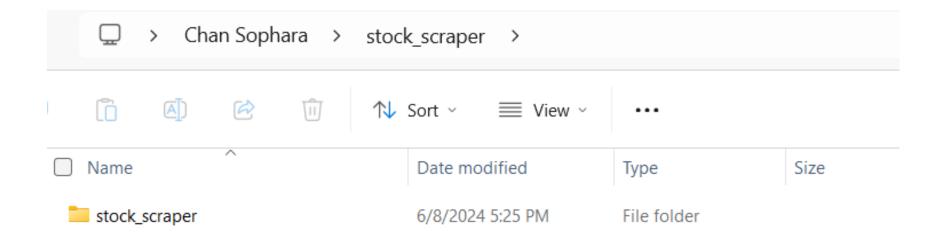
PS C:\Users\sopha> scrapy startproject stock_scraper

For Remaining companies:

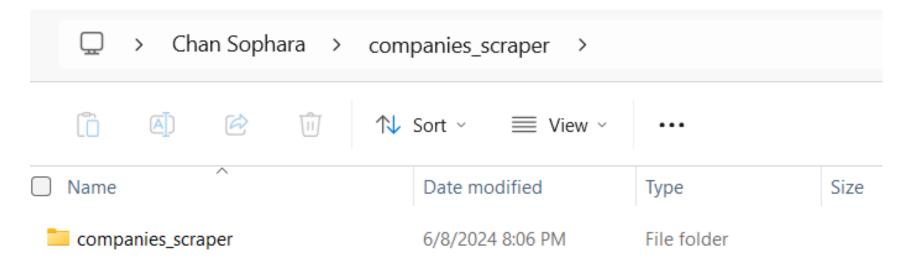
PS C:\Users\sopha> scrapy startproject companies_scraper

• After the we start project, it will generate a folder in our user directory immediatly:

For PWSA:



For Remaining companies:



• Then we need to create a ".py" file in "spiders" directory in each folder.

For PWSA: stock_spider.py

```
C: > Users > sopha > stock_scraper > stock_scraper > spiders > ♥ stock_spider.py > ધ StockSpider > ♥ parse
      import scrapy
      import pandas as pd
       class StockSpider(scrapy.Spider):
           name = "stock spider"
           allowed domains = ["csx.com.kh"]
           start urls = [
               "https://csx.com.kh/data/stock/daily.do?lang=en&MNCD=60202&board type=M&issueCode=KH1000010004&forma=ALL&fromDate=20120418&toDate=20240531"
  8
  9
 10
           def parse(self, response):
 11
               headers = [
 12
                   'Date', 'Closing Price', 'Change', 'Trading Volume (shr)',
 13
                   'Trading Value (KHR)', 'Opening', 'High', 'Low',
 14
                   'Market Cap. (Mil.KHR)', 'Full Market Cap. (Mil.KHR)'
 15
 16
 17
               # Extract the table and rows
 18
 19
               table = response.xpath('//div[@id="index table"]/table[@class="summary"]')
               rows = table.xpath('.//tr[position()>1]')
 20
 21
               data = []
 22
 23
               for row in rows:
                   # Cleaning each column to ensure there are no extra spaces or newline characters
 24
                   cols = ["".join(col.xpath('.//text()').getall()).strip() for col in row.xpath('.//td')]
 25
                   data.append(cols)
 26
 27
 28
               # Try to create DataFrame and save it
 29
                   df = pd.DataFrame(data, columns=headers)
 30
                   df.to csv('PWSA Scrapy stock prices.csv', index=False)
 31
                   self.logger.info("Stock data saved to 'PWSA Scrapy stock prices.csv'")
 32
 33
               except Exception as e:
                   self.logger.error(f"Error creating DataFrame: {e}")
 34
 35
```

For Remaining Companies: companies_stock_spider.py

```
C: > Users > sopha > companies_scraper > companies_scraper > spiders > decompanies_stock_spider.py > {} scrapy
      import scrapy
      import pandas as pd
      from datetime import datetime
  4
      class CompaniesStockSpider(scrapy.Spider):
  5
          name = 'companies stock spider'
  6
           allowed domains = ['csx.com.kh']
  7
  8
           custom settings = {
  9
               'DOWNLOAD DELAY': 1.0, # Respectful crawling by setting delay between requests
 10
 11
 12
           def start requests(self):
 13
               base url main = "https://csx.com.kh/data/stock/daily.do?MNCD=60202&forma=ALL"
 14
               base url growth = "https://csx.com.kh/data/growthdaily/listPosts.do?MNCD=60202&forma=ALL"
 15
 16
               companies = [
 17
                   # Main Board companies
 18
                   {"symbol": "PWSA", "issueCode": "KH1000010004", "joiningDate": "20120418", "url": base_url_main},
 19
                   {"symbol": "GTI", "issueCode": "KH1000020003", "joiningDate": "20140616", "url": base_url_main},
 20
                   {"symbol": "PPAP", "issueCode": "KH1000040001", "joiningDate": "20151209", "url": base url main},
 21
                   {"symbol": "PPSP", "issueCode": "KH1000050000", "joiningDate": "20160530", "url": base_url_main},
 22
                   {"symbol": "PAS", "issueCode": "KH1000060009", "joiningDate": "20170608", "url": base_url_main},
 23
                   {"symbol": "ABC", "issueCode": "KH1000100003", "joiningDate": "20200525", "url": base_url_main},
 24
                   {"symbol": "PEPC", "issueCode": "KH1000140009", "joiningDate": "20200812", "url": base url main},
 25
                   {"symbol": "MJQE", "issueCode": "KH1000210000", "joiningDate": "20230628", "url": base url main},
 26
                   {"symbol": "CGSM", "issueCode": "KH1000220009", "joiningDate": "20230627", "url": base url main},
 27
                   # Growth Board companies
 28
                   {"symbol": "DBDE", "issueCode": "KH1000150008", "joiningDate": "20210906", "url": base_url_growth},
 29
                   {"symbol": "JSL", "issueCode": "KH1000160007", "joiningDate": "20220210", "url": base url growth}
 30
 31
```

```
for company in companies:
33
                 joining_date = datetime.strptime(company['joiningDate'], '%Y%m%d').strftime('%Y%m%d')
34
                 url = f"{company['url']}&issueCode={company['issueCode']}&fromDate={joining date}&toDate=20240531"
35
                 request = scrapy.Request(url, callback=self.parse)
36
                 request.meta['company symbol'] = company['symbol']
37
                 yield request
38
39
         def parse(self, response):
40
             company symbol = response.meta['company symbol']
41
             headers =
42
                  'Date', 'Closing Price', 'Change', 'Trading Volume (shr)',
43
                 'Trading Value (KHR)', 'Opening', 'High', 'Low',
44
                 'Market Cap. (Mil.KHR)', 'Full Market Cap. (Mil.KHR)'
45
46
47
             # Extract the table rows
48
             rows = response.xpath('//div[@id="index table"]/table[@class="summary"]//tr[position()>1]')
49
50
             data = []
51
             for row in rows:
52
                 cols = row.xpath('.//td')
53
                 col texts = ["".join(col.xpath('.//text()').get()).strip() for col in cols]
54
                 if len(col texts) == len(headers):
55
                     data.append(col texts)
56
57
                 else:
                     self.logger.warning(f"Data mismatch in row: {col texts}")
58
             # Create DataFrame and save to CSV
60
             try:
61
                 if data:
62
                    df = pd.DataFrame(data, columns=headers)
63
                    file name = f'{company symbol} stock data.csv'
64
                    df.to csv(file name, index=False)
65
                     self.logger.info(f"Stock data for {company symbol} saved to '{file name}'")
66
                 else:
67
                     self.logger.error(f"No valid data found for {company symbol}")
68
             except Exception as e:
69
                 self.logger.error(f"Error creating DataFrame for {company symbol}: {e}")
70
71
```

72

 In order to scrape data we need to redirect directory in cmd or Window Powershell:

For PWSA:

```
PS C:\Users\sopha> cd stock_scraper
PS C:\Users\sopha\stock_scraper> scrapy crawl stock_spider
```

For Remaining companies:

```
PS C:\Users\sopha> cd companies_scraper
PS C:\Users\sopha\companies_scraper> scrapy crawl companies_stock_spider
```

• After scrape the data, the ".csv" file will upload to each directory:

For PWSA:	stock_scraper	6/8/2024 5:25 PM	File folder	
	PWSA_Scrapy_stock_prices	6/9/2024 2:42 PM	Microsoft Excel Co	241 KB
	scrapy	6/8/2024 7:43 PM	Configuration Sou	1 KB

For Remaining companies:

companies_scraper	6/8/2024 8:06 PM	File folder	
ABC_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	93 KB
CGSM_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	21 KB
DBDE_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	55 KB
GTI_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	183 KB
JSL_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	46 KB
MJQE_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	20 KB
PAS_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	147 KB
PEPC_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	74 KB
PPAP_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	167 KB
PPSP_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	155 KB
PWSA_stock_data	6/9/2024 2:46 PM	Microsoft Excel Co	236 KB
scrapy	6/8/2024 8:03 PM	Configuration Sou	1 KB

RESULT: stock_scraper

PWSA_Scrapy_Stock_Prices.csv

	Α	В	С	D	Е	F	G	Н	I	J
1	Date	Closing Price	Change	Trading Volume (shr)	Trading Value (KHR)	Opening	High	Low	Market Cap. (Mil.KHR)	Full Market Cap. (Mil.KHR)
2	30/05/2024	7,320	20	5,521	40,444,260	7,340	7,360	7,320	636,644	3,499,502
3	29/05/2024	7,340	20	9,496	69,716,440	7,360	7,360	7,340	638,383	3,509,064
4	28/05/2024	7,360	20	11,257	82,751,340	7,380	7,380	7,320	640,122	3,518,625
5	27/05/2024	7,380	0	14,282	104,828,220	7,400	7,400	7,300	641,862	3,528,187
6	24/05/2024	7,380	100	12,478	91,669,260	7,300	7,440	7,300	641,862	3,528,187
7	23/05/2024	7,280	100	1,600	11,641,140	7,200	7,280	7,200	633,165	3,480,379
8	21/05/2024	7,180	120	12,195	87,687,880	7,260	7,280	7,180	624,467	3,432,572
9	20/05/2024	7,300	40	411	2,981,060	7,220	7,300	7,220	634,904	3,489,941
10	17/05/2024	7,260	0	13,518	98,323,140	7,220	7,380	7,220	631,425	3,470,818
11	16/05/2024	7,260	80	6,083	44,048,420	7,200	7,260	7,200	631,425	3,470,818
12	15/05/2024	7,180	40	6,136	43,907,480	7,140	7,180	7,120	624,467	3,432,572
13	13/05/2024	7,140	20	9,063	64,481,920	7,120	7,140	7,100	620,988	3,413,449
14	10/5/2024	7,120	20	5,520	39,213,600	7,100	7,140	7,100	619,249	3,403,888
15	9/5/2024	7,100	0	1,500	10,632,620	7,100	7,100	7,060	617,509	3,394,326
16	8/5/2024	7,100	0	1,452	10,306,240	7,100	7,100	7,060	617,509	3,394,326
17	7/5/2024	7,100	20	888	6,253,340	7,140	7,140	7,020	617,509	3,394,326
18	6/5/2024	7,080	0	6,580	46,133,680	7,020	7,100	7,000	615,770	3,384,765
19	3/5/2024	7,080	80	11,479	80,493,160	7,100	7,100	7,000	615,770	3,384,765
20	2/5/2024	7,000	0	10,289	72,218,240	7,100	7,100	6,960	608,812	3,346,519

GTI_stock_data.csv

	Α	В	С	D	E	F	G	Н	I	J
1	Date	Closing Price	Change	Trading Volume (shr)	Trading Value (KHR)	Opening	High	Low	Market Cap. (Mil.KHR)	Full Market Cap. (Mil.KHR)
2	30/05/2024	2,380		8,458	19,886,340	2,380	2,380	2,330	95,200	95,200
3	29/05/2024	2,370		7,905	18,529,040	2,300	2,400	2,300	94,800	94,800
4	28/05/2024	2,300		884	2,032,940	2,300	2,320	2,290	92,000	92,000
5	27/05/2024	2,300		377	866,130	2,290	2,300	2,290	92,000	92,000
6	24/05/2024	2,290		2,977	6,793,190	2,290	2,290	2,270	91,600	91,600
7	23/05/2024	2,290		234	536,010	2,280	2,290	2,260	91,600	91,600
8	21/05/2024	2,280		973	2,219,220	2,290	2,300	2,280	91,200	91,200
9	20/05/2024	2,300		2,659	6,074,750	2,280	2,300	2,270	92,000	92,000
10	17/05/2024	2,290		13,303	30,016,790	2,250	2,300	2,250	91,600	91,600
11	16/05/2024	2,270		1,220	2,732,650	2,280	2,280	2,200	90,800	90,800
12	15/05/2024	2,280		21,164	47,636,520	2,250	2,320	2,250	91,200	91,200
13	13/05/2024	2,330		174	404,370	2,330	2,330	2,320	93,200	93,200
14	10/5/2024	2,330		158	367,360	2,340	2,340	2,320	93,200	93,200
15	9/5/2024	2,340		2,042	4,689,370	2,340	2,340	2,290	93,600	93,600
16	8/5/2024	2,350		630	1,474,400	2,350	2,350	2,330	94,000	94,000
17	7/5/2024	2,350		170	399,330	2,350	2,350	2,340	94,000	94,000
18	6/5/2024	2,350		4,376	10,202,030	2,320	2,350	2,310	94,000	94,000
19	3/5/2024	2,320		7,131	16,323,190	2,260	2,320	2,260	92,800	92,800
20	2/5/2024	2,260		1,462	3,291,660	2,250	2,260	2,240	90,400	90,400

PPAP_stock_data.csv

	Α	В	С	D	Е	F	G	Н	1	J
1	Date	Closing Price	Change	Trading Volume (shr)	Trading Value	Opening	High	Low	Market Cap. (Mil.KHR)	Full Market Cap. (Mil.KHR)
2	30/05/2024	13,400		393	5,258,660	13,380	13,400	13,200	277,170	1,533,677
3	29/05/2024	13,380		463	6,188,780	13,180	13,380	13,180	276,757	1,531,388
4	28/05/2024	13,400		664	8,860,900	13,340	13,400	13,240	277,170	1,533,677
5	27/05/2024	13,340		509	6,724,480	13,380	13,380	13,120	275,929	1,526,809
6	24/05/2024	13,420		517	6,849,480	13,400	13,420	13,240	277,584	1,535,966
7	23/05/2024	13,400		616	8,204,200	13,460	13,540	13,240	277,170	1,533,677
8	21/05/2024	13,460		1,910	25,630,860	13,360	13,480	13,260	278,412	1,540,544
9	20/05/2024	13,360		258	3,414,540	13,200	13,360	13,200	276,343	1,529,099
10	17/05/2024	13,400		367	4,899,940	13,360	13,480	13,240	277,170	1,533,677
11	16/05/2024	13,360		71	938,760	13,220	13,360	13,220	276,343	1,529,099
12	15/05/2024	13,280		501	6,596,940	13,320	13,340	13,080	274,688	1,519,942
13	13/05/2024	13,320		158	2,101,620	13,300	13,380	13,120	275,516	1,524,520
14	10/5/2024	13,180		355	4,682,600	13,200	13,220	13,100	272,620	1,508,497
15	9/5/2024	13,180		1,548	20,256,560	13,200	13,200	12,900	272,620	1,508,497
16	8/5/2024	13,180		685	8,998,580	13,220	13,220	12,720	272,620	1,508,497
17	7/5/2024	13,200		1,241	16,216,160	13,200	13,200	12,420	273,034	1,510,786
18	6/5/2024	13,180		1,424	18,411,320	13,500	13,500	12,380	272,620	1,508,497
19	3/5/2024	13,160		541	7,054,360	13,260	13,260	13,000	272,206	1,506,208
20	2/5/2024	13,000		414	5,379,100	13,060	13,080	12,980	268,897	1,487,895

PPSP_stock_data.csv

	Α	В	С	D	Е	F	G	Н	I	J
1	Date	Closing Price	Change	Trading Volume	Trading Value (KHR)	Opening	High	Low	Market Cap. (Mil.KHR)	Full Market Cap. (Mil.KHR)
2	30/05/2024	2,400		376,120	885,811,910	2,370	2,430	2,330	172,500	172,500
3	29/05/2024	2,330		79,341	182,938,440	2,300	2,330	2,270	167,469	167,469
4	28/05/2024	2,270		111,356	254,678,470	2,230	2,330	2,230	163,156	163,156
5	27/05/2024	2,230		10,311	22,945,950	2,220	2,230	2,220	160,281	160,281
6	24/05/2024	2,220		6,603	14,627,500	2,220	2,220	2,210	159,562	159,562
7	23/05/2024	2,220		6,952	15,339,560	2,200	2,220	2,190	159,562	159,562
8	21/05/2024	2,200		20,978	46,228,960	2,210	2,230	2,200	158,125	158,125
9	20/05/2024	2,230		48,450	107,717,100	2,230	2,280	2,200	160,281	160,281
10	17/05/2024	2,230		198,592	442,826,230	2,230	2,230	2,200	160,281	160,281
11	16/05/2024	2,030		478	967,220	2,030	2,030	2,020	145,906	145,906
12	15/05/2024	2,030		563	1,141,840	2,020	2,040	2,020	145,906	145,906
13	13/05/2024	2,030		517	1,046,460	2,020	2,030	2,020	145,906	145,906
14	10/5/2024	2,030		1,354	2,742,430	2,030	2,030	2,020	145,906	145,906
15	9/5/2024	2,030		44,992	90,070,750	2,060	2,060	1,990	145,906	145,906
16	8/5/2024	2,050		21,373	42,919,710	2,050	2,050	2,000	147,344	147,344
17	7/5/2024	2,050		777	1,589,460	2,050	2,050	2,020	147,344	147,344
18	6/5/2024	2,050		3,018	6,154,380	2,060	2,060	2,010	147,344	147,344
19	3/5/2024	2,050		3,949	8,070,550	2,060	2,060	2,030	147,344	147,344
20	2/5/2024	2.060		3.539	7.242.080	2.050	2.060	2.030	148.062	148.062

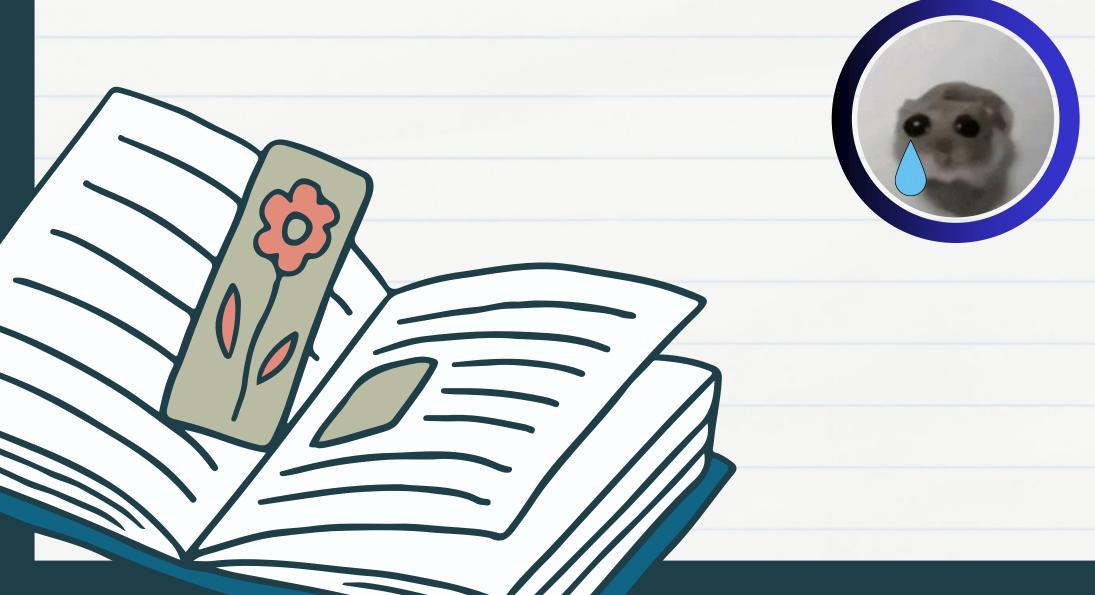
PAS_stock_data.csv

	Α	В	С	D	Е	F	G	Н	I	J
1	Date	Closing Price	Change	Trading Volume	Trading Value (KHR)	Opening	High	Low	Market Cap. (Mil.KHR)	Full Market Cap. (Mil.KHR)
2	30/05/2024	2,400		376,120	885,811,910	2,370	2,430	2,330	172,500	172,500
3	29/05/2024	2,330		79,341	182,938,440	2,300	2,330	2,270	167,469	167,469
4	28/05/2024	2,270		111,356	254,678,470	2,230	2,330	2,230	163,156	163,156
5	27/05/2024	2,230		10,311	22,945,950	2,220	2,230	2,220	160,281	160,281
6	24/05/2024	2,220		6,603	14,627,500	2,220	2,220	2,210	159,562	159,562
7	23/05/2024	2,220		6,952	15,339,560	2,200	2,220	2,190	159,562	159,562
8	21/05/2024	2,200		20,978	46,228,960	2,210	2,230	2,200	158,125	158,125
9	20/05/2024	2,230		48,450	107,717,100	2,230	2,280	2,200	160,281	160,281
10	17/05/2024	2,230		198,592	442,826,230	2,230	2,230	2,200	160,281	160,281
11	16/05/2024	2,030		478	967,220	2,030	2,030	2,020	145,906	145,906
12	15/05/2024	2,030		563	1,141,840	2,020	2,040	2,020	145,906	145,906
13	13/05/2024	2,030		517	1,046,460	2,020	2,030	2,020	145,906	145,906
14	10/5/2024	2,030		1,354	2,742,430	2,030	2,030	2,020	145,906	145,906
15	9/5/2024	2,030		44,992	90,070,750	2,060	2,060	1,990	145,906	145,906
16	8/5/2024	2,050		21,373	42,919,710	2,050	2,050	2,000	147,344	147,344
17	7/5/2024	2,050		777	1,589,460	2,050	2,050	2,020	147,344	147,344
18	6/5/2024	2,050		3,018	6,154,380	2,060	2,060	2,010	147,344	147,344
19	3/5/2024	2,050		3,949	8,070,550	2,060	2,060	2,030	147,344	147,344
20	2/5/2024	2.060		3.539	7.242.080	2.050	2.060	2.030	148.062	148.062

ABC_stock_data.csv

	Α	В	С	D	E	F	G	Н	I	J
1	Date	Closing Price	Change	Trading Volume (shr)	Trading Value (KHR)	Opening	High	Low	Market Cap. (Mil.KHR)	Full Market Cap. (Mil.KHR)
2	30/05/2024	7,620		21,428	163,832,920	7,660	7,660	7,620	3,300,702	3,300,702
3	29/05/2024	7,660		36,986	282,009,160	7,660	7,680	7,600	3,318,029	3,318,029
4	28/05/2024	7,660		36,015	275,524,460	7,640	7,680	7,620	3,318,029	3,318,029
5	27/05/2024	7,640		30,536	233,537,660	7,640	7,700	7,620	3,309,365	3,309,365
6	24/05/2024	7,620		44,684	339,653,920	7,600	7,620	7,580	3,300,702	3,300,702
7	23/05/2024	7,580		91,945	694,774,420	7,520	7,600	7,500	3,283,376	3,283,376
8	21/05/2024	7,540		108,162	823,701,220	7,760	7,760	7,540	3,266,049	3,266,049
9	20/05/2024	7,760		52,866	409,088,400	7,760	7,780	7,700	3,361,345	3,361,345
10	17/05/2024	7,780		55,587	432,094,340	7,760	7,800	7,740	3,370,008	3,370,008
11	16/05/2024	7,800		36,857	286,595,600	7,800	7,800	7,740	3,378,672	3,378,672
12	15/05/2024	7,800		91,832	715,736,240	7,760	7,820	7,760	3,378,672	3,378,672
13	13/05/2024	7,740		40,010	309,874,600	7,720	7,800	7,720	3,352,682	3,352,682
14	10/5/2024	7,720		94,134	722,237,120	7,680	7,780	7,640	3,344,019	3,344,019
15	9/5/2024	7,680		237,562	1,840,072,040	7,640	7,940	7,620	3,326,692	3,326,692
16	8/5/2024	7,560		87,818	655,378,660	7,400	7,580	7,400	3,274,712	3,274,712
17	7/5/2024	7,420		74,447	548,629,460	7,300	7,420	7,300	3,214,070	3,214,070
18	6/5/2024	7,420		105,975	786,751,040	7,500	7,500	7,320	3,214,070	3,214,070
19	3/5/2024	7,600		173,480	1,316,897,480	7,620	7,620	7,560	3,292,039	3,292,039
20	2/5/2024	7,620		302,282	2,292,919,500	7,600	7,620	7,460	3,300,702	3,300,702

Thank you!!!





Resource page

