

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC THĂNG LONG
— o0o —



BÀI TẬP LỚN

DỮ LIỆU VNDIRECT

GV hướng dẫn:

TS. Nguyễn Thị Huyền Châu

Sinh viên thực hiện:

CN. Đoàn Trung Phong

Nguyễn Thị Thanh Mai

Lê Quốc Khánh

Nhóm:

1

Môn:

Nhập môn khoa học dữ liệu

Hà Nội - 2023

Mục lục

1	Phần 1: Giới thiệu chung	4
1.1	Mô tả chung về bài toán	4
1.2	Giới hạn đề tài thực hiện	4
1.2.1	Đối tượng nghiên cứu	4
1.2.2	Khách thể nghiên cứu	4
1.2.3	Quy mô nghiên cứu	5
1.2.4	Thời gian thu thập dữ liệu	5
1.2.5	Loại dữ liệu sử dụng	5
2	Phần 2: Cơ sở lý thuyết	6
2.1	Quy trình phát triển khoa học dữ liệu	6
2.2	Định nghĩa một số khái niệm chuyên môn	7
2.2.1	Vốn hóa thị trường	7
2.2.2	KLGD	7
2.2.3	Số CP lưu hành	7
2.2.4	FREE FLOAT	8
2.2.5	BETA	8
2.2.6	P/E	8
2.2.7	P/B	8
2.2.8	Tỷ suất cổ tức	8
2.2.9	BVPS	8
2.2.10	ROAE	8
2.2.11	ROAA	9
2.2.12	EPS	9
3	Phần 3: Thu thập dữ liệu	10
3.1	Mô tả nguồn dữ liệu	10
3.2	Các thông tin liên quan đến dữ liệu sau khi thu thập . .	10
3.2.1	Các thông số	10
3.2.2	Mô tả công ty	10
3.2.3	Giá	11
4	Phần 4: Xử lý dữ liệu	13
4.1	Các vấn đề trong tập dữ liệu	13
4.2	Phương pháp kiểm tra	15
4.2.1	Thiếu dữ liệu	15

4.2.2	Trùng lặp dữ liệu	15
4.2.3	Lỗi logic	15
4.3	Nhận xét và phương pháp xử lý	17
4.3.1	Lỗi thiếu dữ liệu	17
4.3.2	Lỗi trùng lặp	17
4.3.3	Lỗi logic	17
4.3.4	Một vài xử lý khác	18
5	Phần 5: Phân tích	19
5.1	DHT	19
5.1.1	Thông tin cơ bản: Sử dụng DHT_gia.describe()	19
5.1.2	Phân tích biến động giá từ đầu năm 2023	20
5.1.3	Phân tích biến động giá trong năm 2022 và 2021	21
5.2	EID	22
5.2.1	Thông tin cơ bản	22
5.2.2	Biểu đồ giá năm 2023	23
5.2.3	Biểu đồ giá năm 2022 và 2021	24
5.3	VCM	25
5.3.1	Thông tin cơ bản	25
5.3.2	Biểu đồ giá năm 2023	25
5.3.3	Biểu đồ giá năm 2022 và 2021	26
5.4	C69	27
5.4.1	Thông tin cơ bản	27
5.4.2	Biểu đồ giá năm 2023	28
5.4.3	Biểu đồ giá năm 2022 và 2021	28
5.5	VMS	29
5.5.1	Thông tin cơ bản	29
5.5.2	Biểu đồ giá năm 2023	30
5.5.3	Biểu đồ giá năm 2022 và 2021	30
5.6	DL1	30
5.6.1	Thông tin cơ bản	30
5.6.2	Biểu đồ giá năm 2023	30
5.6.3	Biểu đồ giá năm 2022 và 2021	30
5.7	VC9	30
5.7.1	Thông tin cơ bản	30
5.7.2	Biểu đồ giá năm 2023	31
5.7.3	Biểu đồ giá năm 2022 và 2021	31

5.8	CAG	31
5.8.1	Thông tin cơ bản	31
5.8.2	Biểu đồ giá năm 2023	31
5.8.3	Biểu đồ giá năm 2022 và 2021	31
5.9	VIF	31
5.9.1	Thông tin cơ bản	31
5.9.2	Biểu đồ giá năm 2023	32
5.9.3	Biểu đồ giá năm 2022 và 2021	32
5.10	GKM	32
5.10.1	Thông tin cơ bản	32
5.10.2	Biểu đồ giá năm 2023	33
5.10.3	Biểu đồ giá năm 2022 và 2021	33
5.11	Kết luận	33
6	Phần 6: Lời cảm ơn	34

1 Phần 1: Giới thiệu chung

1.1 Mô tả chung về bài toán

Trong bối cảnh thị trường chứng khoán ngày càng trở nên phức tạp và thay đổi nhanh chóng, việc tìm kiếm cổ phiếu tiềm năng và có khả năng mang lại lợi nhuận ổn định trở thành mục tiêu quan trọng đối với các nhà đầu tư. Trên sàn giao dịch HNX (Hà Nội), có nhiều cổ phiếu đã thu hút sự quan tâm của các nhà đầu tư bởi tiềm năng phát triển và cơ hội sinh lợi nhất định.

Trong báo cáo này, chúng tôi xin giới thiệu đến quý độc giả top 10 cổ phiếu nổi bật trên sàn HNX. Với mục tiêu cung cấp thông tin chi tiết và cập nhật về sự phát triển và hiệu suất của các cổ phiếu này, các cổ phiếu đã được chúng tôi lựa chọn dựa trên nhiều yếu tố, bao gồm tăng trưởng doanh thu, lợi nhuận, tiềm năng phát triển, quỹ đất, quản trị doanh nghiệp và tầm nhìn chiến lược.

Chúng tôi hi vọng rằng báo cáo này sẽ giúp cho quý độc giả một cái nhìn tổng quan về thị trường cổ phiếu HNX và những cơ hội đầu tư tiềm năng. Tuy nhiên, việc đầu tư vào chứng khoán luôn mang lại rủi ro, chúng ta nên thực hiện sau khi đã cân nhắc kỹ lưỡng và tham khảo ý kiến của các chuyên gia tài chính. Hi vọng rằng thông tin trong báo cáo sẽ hữu ích cho nhà đầu tư và giúp họ đưa ra những quyết định đầu tư thông minh và cân nhắc.

1.2 Giới hạn đề tài thực hiện

1.2.1 Đối tượng nghiên cứu

Các cổ phiếu trong danh sách top 10 cổ phiếu của HNX.

1.2.2 Khách thể nghiên cứu

Cổ phiếu và các dữ liệu về các cổ phiếu này, bao gồm thông tin về giá cổ phiếu, vốn hóa thị trường, EPS (lợi nhuận trên cổ phiếu), P/E (tỷ lệ giá trên lợi nhuận) và các chỉ số tài chính khác. Bằng cách phân tích các thông tin này, bạn có thể hiểu sâu hơn về các cổ phiếu này và đưa ra những nhận định, dự báo về hiệu suất và tiềm năng phát triển của chúng.

1.2.3 Quy mô nghiên cứu

10 mã chứng khoán trên sàn HNX.

1.2.4 Thời gian thu thập dữ liệu

Từ 2/2021 đến 7/2023.

1.2.5 Loại dữ liệu sử dụng

Dữ liệu bảng.

2 Phần 2: Cơ sở lý thuyết

2.1 Quy trình phát triển khoa học dữ liệu

Quy trình khoa học dữ liệu thường được bắt đầu bởi một vấn đề kinh doanh. Nhà khoa học dữ liệu sẽ làm việc với các bên liên quan để tìm hiểu nhu cầu của doanh nghiệp. Một khi vấn đề đã được xác định, nhà khoa học dữ liệu sẽ giải quyết nó bằng cách sử dụng quy trình khoa học dữ liệu OSEMN:

O – Thu thập dữ liệu (Obtain data)

Dữ liệu có thể tồn tại từ trước, mới được thu thập hoặc là một kho dữ liệu có thể tải xuống từ Internet. Các nhà khoa học dữ liệu có thể trích xuất dữ liệu từ những cơ sở dữ liệu nội bộ hoặc bên ngoài, phần mềm CRM của công ty, nhật ký máy chủ web, mạng xã hội hoặc mua dữ liệu từ các nguồn bên thứ ba đáng tin cậy.

S – Làm sạch dữ liệu (Scrub data)

Làm sạch dữ liệu là quy trình chuẩn hóa dữ liệu dựa theo một định dạng được định trước. Quy trình này bao gồm xử lý dữ liệu còn thiếu, sửa lỗi dữ liệu và loại bỏ mọi dữ liệu ngoại lai. Một số ví dụ về làm sạch dữ liệu:

- Thay đổi toàn bộ các giá trị ngày thành một định dạng tiêu chuẩn phổ biến.
- Sửa lỗi chính tả hoặc thừa khoảng trống.
- Sửa lỗi tính toán không chính xác hoặc xóa dấu phẩy khỏi các số lớn.

E – Khám phá dữ liệu (Explore data)

Khám phá dữ liệu là thao tác phân tích sơ bộ dữ liệu được sử dụng để lập kế hoạch kỹ hơn cho các chiến lược mô hình hóa dữ liệu. Các nhà khoa học dữ liệu nắm được hiểu biết ban đầu về dữ liệu bằng cách sử dụng thống kê mô tả và các công cụ trực quan hóa dữ liệu. Sau đó họ khám phá dữ liệu để xác định các mẫu thú vị có thể được nghiên cứu hoặc tận dụng.

M – Mô hình hóa dữ liệu (Model data)

Phần mềm và các thuật toán máy học được sử dụng để thu thập thông tin chuyên sâu hơn, dự đoán kết quả và đề xuất hướng hành động tốt nhất. Các kỹ thuật máy học như liên kết, phân loại và phân nhóm được áp dụng cho tập dữ liệu đào tạo. Mô hình có thể được thử nghiệm so với dữ liệu thử nghiệm định trước để đánh giá độ chính xác của kết quả. Mô hình dữ liệu có thể được tinh chỉnh nhiều lần để cải thiện kết quả thu được.

N – Diễn giải kết quả (Interpret results)

Các nhà khoa học dữ liệu hợp tác cùng các chuyên gia phân tích và doanh nghiệp để chuyển đổi thông tin chi tiết về dữ liệu thành hành động. Họ tạo ra các sơ đồ, đồ thị và biểu đồ để thể hiện những xu hướng và dự đoán. Tóm tắt dữ liệu giúp các bên liên quan hiểu rõ và triển khai kết quả một cách hữu hiệu. Định nghĩa khái niệm các mô hình học máy được sử dụng

⇒ Ở bài tập lớn lần này, ta chỉ làm đến phần Explore data.

2.2 Định nghĩa một số khái niệm chuyên môn

2.2.1 Vốn hóa thị trường

Là tổng giá trị hiện tại của tất cả cổ phiếu đang lưu hành trên thị trường của một doanh nghiệp.

2.2.2 KLGD

KLGD (khối lượng giao dịch) là biểu hiện cho dòng tiền đang hoạt động trên thị trường. KLGD tăng tức là dòng tiền được bơm vào thị trường, còn KLGD giảm là dòng tiền rút ra khỏi thị trường.

2.2.3 Số CP lưu hành

Là những CP đang được nắm giữ bởi tất cả cổ đông của một công ty, trong đó bao gồm cả khối cổ phần được nắm giữ bởi nhà đầu tư tổ chức; cũng như phần vốn sở hữu của các nhà đầu tư tổ chức như quỹ tương hỗ, quỹ hưu trí và quỹ dự phòng.

2.2.4 FREE FLOAT

Free float còn gọi là tỷ lệ cổ phiếu tự do chuyển nhượng. Đây chính là tỷ lệ khối lượng cổ phiếu tự do được quyền chuyển nhượng so với tổng cổ phiếu đang lưu hành ở trên thị trường của doanh nghiệp đó.

2.2.5 BETA

Beta là hệ số phản ánh sự biến động của một tài sản trong thị trường của một ngành. Hệ số Beta trong chứng khoán có thể được hiểu là mức độ biến động của một chứng khoán cụ thể hoặc một danh mục đầu tư so với thị trường chứng khoán.

2.2.6 P/E

P/E trượt 12T, chỉ số P/E là tỷ lệ giữa giá cổ phiếu và EPS (price-to-earning ratio).

2.2.7 P/B

P/B hiện tại, chỉ số P/B (Price to Book ratio) là 1 chỉ số tài chính quan trọng, được sử dụng để so sánh giá của cổ phiếu với giá trị ghi sổ của cổ phiếu đó.

2.2.8 Tỷ suất cổ tức

Tỷ suất cổ tức là tỷ lệ cổ tức nhà đầu tư nhận được so với số tiền phải trả để mua cổ phiếu.

2.2.9 BVPS

BVPS có tên tiếng Anh đầy đủ là Book Value Per Share có nghĩa là giá trị sổ sách tính trên 1 cổ phiếu.

2.2.10 ROAE

ROAE hay Return on average equity là chỉ số lợi nhuận trên vốn chủ sở hữu bình quân. Đây là một trong những chỉ số quan trọng đo lường mức độ hiệu quả trong việc sử dụng vốn chủ sở hữu của doanh nghiệp.

2.2.11 ROAA

ROAA là viết tắt của từ Return on Average Assets, được hiểu là tỷ suất lợi nhuận trên tài sản trung bình. Chỉ số này thể hiện năng lực sinh lợi của tài sản công ty, phản ánh hiệu suất doanh nghiệp đang sử dụng tài sản.

2.2.12 EPS

EPS là thu nhập (lợi nhuận ròng) của một cổ phiếu.

3 Phần 3: Thu thập dữ liệu

3.1 Mô tả nguồn dữ liệu

- Dữ liệu được thu thập trên trang DSTOCK – trang tra cứu thông tin cổ phiếu và thị trường.
- Để thu thập dữ liệu, ta phải tìm link API bằng cách: Inspect → Network
- Link API:
- + Thông số cơ bản: thông số 1, thông số 2.
- + Mô tả công ty: mô tả.
- + Giá: Giá.

3.2 Các thông tin liên quan đến dữ liệu sau khi thu thập

3.2.1 Các thông số

- code: Mã CK.
- MARKETCAP: Vốn hóa thị trường.
- NMVOLUME_AVG_CR_10D: KLGD TB 10 phiên.
- PRICE_HIGHEST_CR_52W: Cao nhất 52 tuần.
- PRICE_LOWEST_CR_52W: Thấp nhất 52 tuần.
- OUTSTANDING_SHARES: Số CP lưu hành.
- FREEFLOAT: free float.
- BETA: beta.
- PRICE_TO_EARNINGS: P/E trượt 12T.
- PRICE_TO_BOOK: P/B hiện tại.
- DIVIDEND_YIELD: Tỷ suất cổ tức.
- BVPS_CR: BVPS hiện tại.
- ROAE_TR_AVG5Q: ROAE trượt 12T.
- ROAA_TR_AVG5Q: ROAA trượt 12T.
- EPS_TR: EPS trượt 12T.

3.2.2 Mô tả công ty

- Code: Mã CK.
- Floor: Sàn.
- Logo: Logo.

- vnName: Tên viết tắt.
- enName: Tên quốc tế.
- foundDate: Ngày.
- taxCode: Mã số thuế.
- registrationCode: Mã đăng ký.
- vnAddress: Trụ sở chính.
- enAddress: Địa chỉ tiếng Anh.
- phone: Số điện thoại.
- fax: fax.
- website: Trang web.
- email: email.
- employees: Nhân viên.
- vnSummary: Tóm tắt.
- enSummary: Tên pháp định.
- companyId: ID.
- branches: Chi nhánh.

3.2.3 Giá

- code: Mã CK.
- date: Ngày.
- time: Giờ.
- floor: Sàn.
- type: Loại.
- basicPrice: Giá cơ bản.
- ceilingPrice: Giá trần.
- floorPrice: Giá sàn.
- open: Giá mở cửa.
- high: Giá cao nhất.
- low: Giá thấp nhất.
- close: Giá đóng cửa.
- average: Giá bình quân.
- adOpen: Giá mở cửa điều chỉnh.
- adHigh: Giá cao nhất điều chỉnh.
- adLow: Giá thấp nhất điều chỉnh.
- adClose: Giá đóng cửa điều chỉnh.
- adAverage: Giá bình quân điều chỉnh.

- nmVolume: KLGD khớp lệnh.
- ptVolume: KLGD thỏa thuận.
- change: Thay đổi.
- adChange: Thay đổi điều chỉnh.
- pctChange: Phần trăm thay đổi.

4 Phần 4: Xử lý dữ liệu

4.1 Các vấn đề trong tập dữ liệu

Bảng	Thuộc tính	Thiếu dữ liệu	Trùng lặp dữ liệu	Định dạng
Mô tả	Code	0	0	object
	floor	0	Cho phép trùng lặp	object
	vnName	0	0	object
	enName	0	0	object
	foundDate	0	Cho phép trùng lặp	object
	taxCode	0	0	object
	registrationCode	0	0	object
	vnAddress	0	0	object
	enAddress	0	Cho phép trùng lặp	object
	Phone	0	0	object
	Fax	0	0	object
	Website	0	0	object
	Email	1	0	object
	Employees	1	Cho phép trùng lặp	Float64
	vnSummary	0	0	Object
	enSummary	10	0	Object
	companyId	0	0	Object
	branches	8	Cho phép trùng lặp	Float64
BCTC	MARKETCAP	0	0	Float64
	NMVOLUME_AVG_CR_10D	0	0	Float64
	PRICE_HIGHEST_CR_52W	0	0	Float64
	PRICE_LOWEST_CR_52W	0	0	Float64
	OUTSTANDING_SHARES	0	0	Float64
	FREEFLOAT	0	0	Float64
	BETA	0	0	Float64

Bảng	Thuộc tính	Thiếu dữ liệu	Trùng lặp dữ liệu	Định dạng
	PRICE_TO_EARNINGS	0	0	Float64
	PRICE_TO_BOOK	0	0	Float64
	DIVIDEND_YIELD	0	0	Float64
	BVPS_CR	0	0	Float64
	ROAE_TR_AVG5Q	0	0	Float64
	EPS_TR	0	0	Float64
Giá	code	Không	Không	Object
	date	Không	Không	Object
	time	Không	Cho phép trùng lặp	Object
	floor	Không	Không	Float64
	type	Không	Không	Float64
	basicPrice	Không	Không	Float64
	ceilingPrice	Không	Không	Float64
	floorPrice	Không	Không	Float64
	open	Không	Không	Float64
	high	Không	Không	Float64
	low	Không	Không	Float64
	close	Không	Không	Float64
	average	Không	Không	Float64
	adOpen	Không	Không	Float64
	adHigh	Không	Không	Float64
	adLow	Không	Không	Float64
	adClose	Không	Không	Float64
	adAverage	Không	Không	Float64
	nmVolume	Không	Không	Float64
	nmValue	Không	Không	Float64
	ptVolume	Không	Không	Float64
	ptValue	Không	Không	Float64
	change	Không	Không	Float64
	adChange	Không	Không	Float64
	pctChange	Không	Không	Float64

- Về lỗi logic: trừ adChange và pctChange thì các thuộc tính đều không gặp vấn đề về lỗi logic.

4.2 Phương pháp kiểm tra

4.2.1 Thiếu dữ liệu

- Ta sẽ dùng cú pháp `tenDF.isnull.sum()` để đếm số dòng thiếu dữ liệu.
- Ngoài ra ở bảng mô tả, khi dùng `mota["email"].unique()` và `mota["enSummary"].unique()` thì ta thấy cột email có một giá trị là xâu rỗng và tất cả giá trị trong cột enSummary đều là xâu rỗng.

4.2.2 Trùng lặp dữ liệu

- Ta dùng `tenDF.duplicated.sum()` để kiểm tra và không thấy quan sát nào bị trùng.
- Ngoài ra, một số cột không cho phép trùng lặp như id, tên công ty, sdt, email, ... ta sẽ dùng `tenDF.duplicated(["tên cột"]).sum()` để kiểm tra và cũng không thấy giá trị nào bị trùng lặp ở những cột như vậy.

4.2.3 Lỗi logic

Bảng mô tả

Lỗi logic mà ta cần kiểm tra đơn giản là 2 cột employees và branches có giá trị âm nào hay không:

```
[30] 1 (mota["employees"] <= 0).sum()
      0
      1 (mota["branches"] < 0).sum()
      0
```

Bảng BCTC

Giá trị của bảng BCTC rất đơn giản, chỉ có 1 dòng nên ta có thể nhận thấy ngay là các giá trị cần dương đều dương như Vốn hóa thị trường, KLGD TB 10 phiên, giá cao/thấp nhất 52 tuần.

Tuy nhiên ta cần kiểm tra độ chính xác của giá trị KLDG TB 10 phiên, ví dụ với mã DHT như sau:

```
[ ] 1 DHT_coban["KLGD TB 10 phiên"] == DHT_gia.loc[1:10, "nmVolume"].sum()/10
```


Bảng giá

Ví dụ với mã DHT, kiểm tra tương tự với 10 mã còn lại.

Kiểm tra xem giá thấp nhất có bé hơn hoặc bằng giá cao nhất không:

```
1 (~(DHT_gia['low'] <= DHT_gia['high'])).sum()

0
```

Kiểm tra xem các thuộc tính về chênh lệch giá phiên trước và phiên sau đã chính xác chưa:

```
def change_check(df):
    false_count = 0
    false_count_per = 0

    for i in range(1, 600):
        diff = df.loc[i, "adClose"] - df.loc[i + 1, "adClose"]

        if (df.loc[i, "adChange"] != diff):
            false_count += 1
        if (df.loc[i, "pctChange"] != (diff/df.loc[i+1, "adClose"])*100):
            false_count_per += 1

    print(false_count)
    print(false_count_per)

[72] 1 change_check(DHT_gia)

427
428
```

Kiểm tra xem tỷ lệ tăng/giảm giá có quá 10% hay không:

```
[75] 1 (DHT_gia[['pctChange']] < -10).sum()

pctChange    0
dtype: int64

[76] 1 (DHT_gia[['pctChange']] > 10).sum()

pctChange    0
dtype: int64
```

4.3 Nhận xét và phương pháp xử lý

4.3.1 Lỗi thiếu dữ liệu

Lỗi này chỉ xuất hiện ở bảng mô tả chung.

Ở cột email, ta đơn giản là tìm email của công ty bị thiếu trên internet và replace xâu rỗng ở trong df, ở đây công ty bị thiếu là CTCP Alpha Seven.

```
[25] 1 mota["email"] = mota["email"].replace("", "info@a7group.vn")
```

Tiếp theo, với cột employees, có 1 công ty bị thiếu và công ty không public thông tin này, nên ta ước lượng bằng trung bình nhân viên các công ty khác:

```
[88] 1 mota["employees"] = mota["employees"].fillna(mota["employees"].sum()//9)
```

Ở cột branches, đơn giản là các công ty bị thiếu không có nhiều chi nhánh, nên ta điền 0 vào các hàng bị thiếu (trụ sở chính không tính là 1 chi nhánh).

```
[22] 1 mota["branches"] = mota["branches"].fillna(0)
```

Còn cột enSummary, tất cả giá trị đều bị thiếu, xét tính quan trọng, cột này cũng không quá cần thiết nên ta bỏ nó đi.

```
[28] 1 mota.drop(["enSummary"], axis = 1, inplace = True)
```

4.3.2 Lỗi trùng lặp

Dữ liệu không có lỗi trùng lặp ở các giá trị không cho phép trùng lặp.

4.3.3 Lỗi logic

Các lỗi logic ở trong tập dữ liệu không hẳn là lỗi mà là do việc làm tròn các số kiểu float.

Vậy nên, bảng có nhiều lỗi này nhất nằm ở bảng giá.

Đầu tiên ta cần sửa lại các giá trị về chênh lệch giá, tỷ lệ chênh lệch giá:

```
def change_correct(df):
    for i in range(1, 600):
        df.loc[i, "adChange"] = df.loc[i]["adClose"] - df.loc[i + 1]["adClose"]
        df.loc[i, "change"] = round(df.loc[i]["close"] - df.loc[i + 1]["close"], 1)
        df.loc[i, "pctChange"] = ((df.loc[i]["adClose"] - df.loc[i + 1]["adClose"])/df.loc[i + 1]["adClose"])*100

    return df
```

```
1 DHT_gia = change_correct(DHT_gia)
```

Tiếp theo, có 7 mã có vấn đề khi mức tăng giảm giá lớn hơn 10%. Tuy nhiên, sau khi kiểm tra, ta thấy đây là dữ liệu chính xác chứ không phải lỗi, nên ta sẽ giữ nguyên. Và việc này sẽ cho thấy mức tăng giảm giá của các công ty này đang là không hợp lệ.

4.3.4 Một vài xử lý khác

Ở hai cột branches và employees ta đưa về kiểu int để dễ tính toán và tiết kiệm bộ nhớ:

```
1 mota["employees"].astype("int")

[33] 1 mota["branches"].astype("int")
```

Ở cột date của bảng giá, ta chỉnh lại cho các giá trị này sang kiểu datetime của pandas để dễ xử lý:

```
[66] 1 DHT_gia['date'] = pd.to_datetime(DHT_gia['date'])
```

5 Phần 5: Phân tích

Ở đây ta sẽ chia thành 11 phần, với 10 phần đầu là phân tích riêng lẻ các mã, và phần cuối sẽ là phân tích chung cả 10 mã và đưa ra nhận xét, kết luận về 10 mã này trong vòng 600 phiên giao dịch gần nhất.

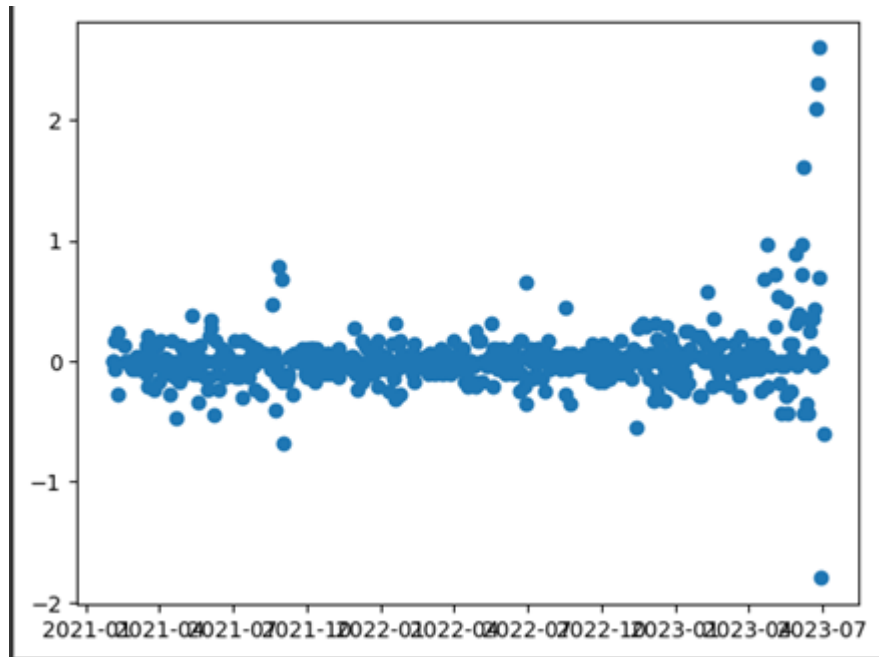
Một vài lưu ý:

- Ta sẽ xem xét dựa trên giá điều chỉnh bởi giá điều chỉnh là giá thực tế mà người giao dịch phải chi trả.
- Giá x1000.

5.1 DHT

5.1.1 Thông tin cơ bản: Sử dụng `DHT_gia.describe()`

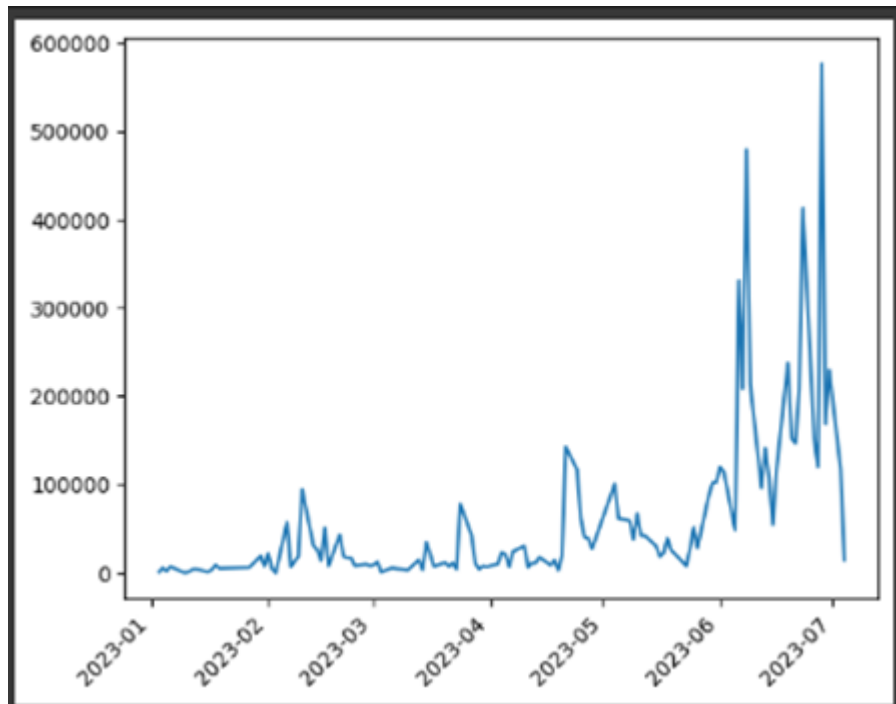
- Giá cao nhất: 32 (max adHigh).
- Giá thấp nhất: 12.832000 (min adLow).
- Lượng tăng giá cao nhất: 2.6 (max adChange).
- Lượng giảm giá nhiều nhất: -1.8 (min adChange).
- Tỷ lệ giá giảm cao nhất: -6.122449 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 9.961686 (max pctChange).
- Khối lượng giao dịch khớp lệnh lớn nhất: 576675 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 0 (min nmVolume).
- KLGD khớp lệnh trung bình: 20040.361667 (mean nmVolume).
- Lượng dao động giá không quá lớn khi mà đến 75% (tức 450 ngày) mức dao động giá nằm trong khoảng [-1.8; 0.036000]. Tuy nhiên, cần phải chú ý khi có đến 50% lượng dao động giá là giảm (< 0), nghĩa là có đến 300 ngày mà giá sẽ giảm. Ta cũng thấy lượng tăng giảm giá chỉ dao động quanh trục 0, nên có thể đây sẽ là một mã khá an toàn cho những ai sợ gặp số lỗ quá lớn, tuy nhiên lãi nhận về cũng không quá cao nhưng cần để ý khi công ty đang tăng trưởng tốt với EPS là 3896.938991.



5.1.2 Phân tích biến động giá từ đầu năm 2023



- Nhìn vào biểu đồ này, ta thấy mã DHT đang có sự hồi sinh rất mạnh mẽ kể từ thời điểm giữa tháng 5 năm nay.
- Khoảng đầu năm, giá của mã này khá thấp, chỉ khoảng dưới 17.5, tuy nhiên từ đầu tháng 5, giá đã bắt đầu tăng trở lại, và đến giữa tháng 6, giá đã vượt mốc 20.
- Đặc biệt, trong khoảng thời gian 1 tháng trở lại đây, giá tăng cực kỳ mạnh, nhất là vào khoảng đầu tháng 7 khi giá nhảy vọt từ 20 lên trên mức 25, thậm chí, giá còn nhảy lên trên cả mức 30.
- Tuy vậy, những ngày gần đây, giá đang bắt đầu có xu hướng giảm trở lại, và giảm tương đối mạnh.



- Tương tự ở biểu đồ biến động của KLGD, ta cũng thấy xu hướng giảm mạnh rất rõ rệt.
 - Từ 2 cơ sở trên, ta có thể thấy đây chính là thời điểm hợp lý và có phần hơi muộn để các nhà đầu tư sở hữu mã DHT chốt lời.
- Để tìm hiểu thêm về xu hướng giá sắp tới, ta đặt ra câu hỏi, liệu thời điểm này trong các năm trước, giá đang có xu hướng như thế nào ?

5.1.3 Phân tích biến động giá trong năm 2022 và 2021

2022:



2021:



Như ta thấy trên biểu đồ, có 2 lý do để kết luận đây là thời điểm chốt lời hợp lý:

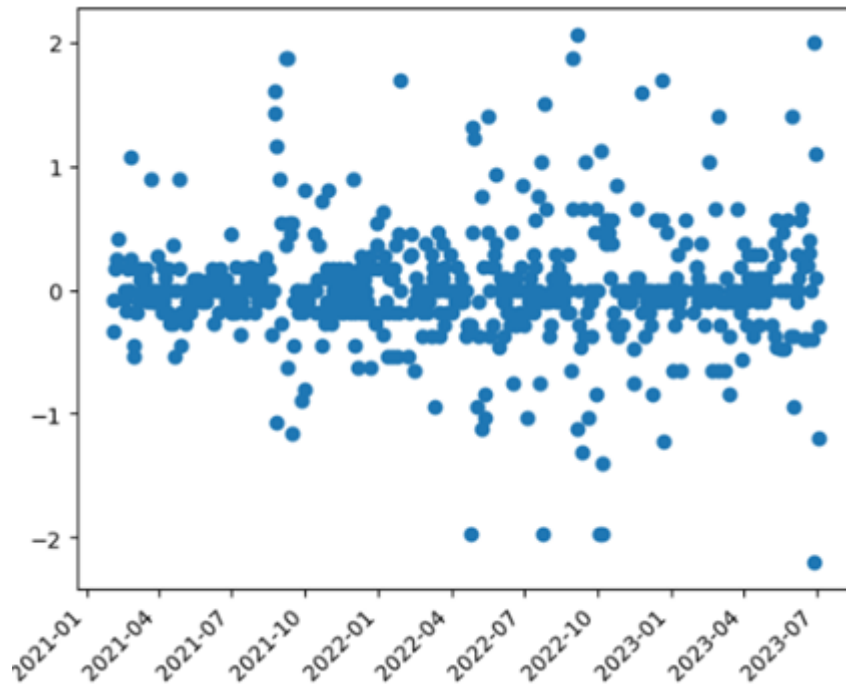
- 2 năm trước, thời điểm này giá chủ yếu là đi ngang và giảm, có thể một phần là do dịch bệnh, nhưng đây cũng là một điều đáng lưu ý.
- Có thể thấy trong 2 năm qua, giá của mã này chưa bao giờ vọt lên trên 20, nên có thể sẽ rất mạo hiểm nếu đầu tư thêm vào trong thời điểm này.

5.2 EID

5.2.1 Thông tin cơ bản

- Giá cao nhất: 25.242 (max adHigh).
- Giá thấp nhất: 12.944 (min adLow).
- Lượng tăng giá cao nhất: 2.066 (max adChange).
- Lượng giảm giá nhiều nhất: -2.2 (min adChange).
- Tỷ lệ giá giảm cao nhất: -10.001522 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 9.947968 (max pctChange).
- Khối lượng giao dịch khớp lệnh lớn nhất: 726000 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 0 (min nmVolume).
- KLGD khớp lệnh trung bình: 17799.985 (mean nmVolume).
- Khoảng dao động giá tương đối lớn, nằm trong khoảng từ tận -10.001522 đến 0.833247. Tuy nhiên, cần phải chú ý khi có đến 50% lượng dao động giá là giảm (< 0), nghĩa là có đến 300 ngày mà giá sẽ giảm. Ta thấy biến động giá của mã khá đa dạng, cũng như có phiên mà lượng biến động tương đối lớn. Vậy nên, đây sẽ là một giá mã mà nhà đầu tư cần tính toán kỹ càng, nắm bắt được xu hướng giá của mã

để đưa ra quyết định. Nếu nắm bắt được để ăn lời thì sẽ đạt được mức lời khá tốt, một phần do EPS của mã EID cũng rất ổn là 5059.355458.



5.2.2 Biểu đồ giá năm 2023



Giá chủ yếu vẫn đang đi ngang, chưa có biến động quá mạnh mẽ, chỉ có khoảng thời gian cuối tháng 3 đầu tháng 2 là giá tăng vọt. Thậm chí, thời gian gần đây, giá đang có xu hướng giảm mạnh. Nhà đầu tư sở hữu mã nên xem xét, đánh giá qua các chỉ báo để dự báo tương lai của mã để sớm cắt lỗ để giảm thiểu thiệt hại.

5.2.3 Biểu đồ giá năm 2022 và 2021

2022:



Như ta thấy, khoảng thời gian này vào năm ngoái thì giá cũng đang đi ngang khá lâu, có đôi chút giảm và tăng mạnh vào khoảng tháng 9,10 rồi lại giảm dần vào cuối năm. Đây có thể là một dấu hiệu dành cho những nhà đầu tư thích mạo hiểm để chờ tăng giá trở lại.

2021:

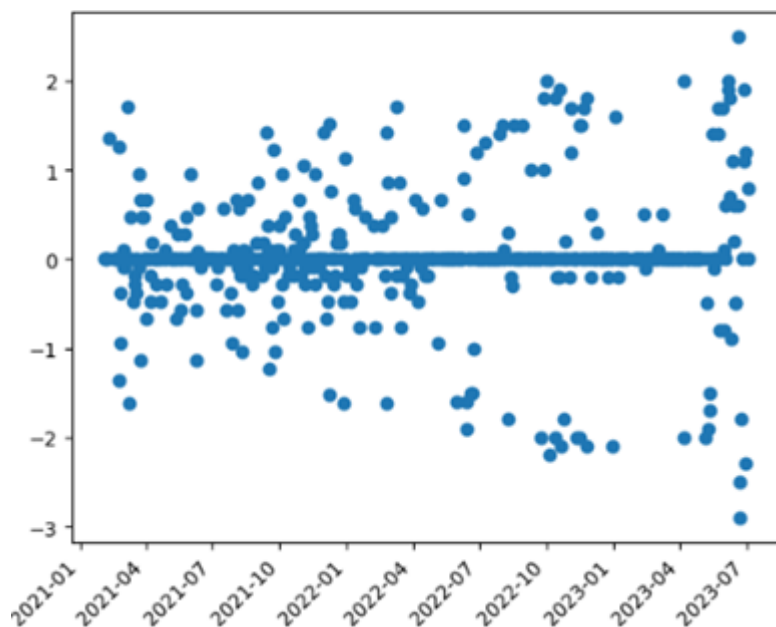


- 2021 lại là một năm khá khởi sắc cho công ty khi mà càng về những tháng cuối năm, giá mã cũng có những dấu hiệu tăng mạnh trở lại. Và cũng như năm 2022, giá mã có một đợt tăng rất mạnh vào tháng 9 tháng 10, và đi ngang vào tầm tháng 7.
- Từ đó ta thấy nếu nhà đầu tư có thể mạo hiểm ôm chứng khoán này để chờ đợt tăng vào khoảng thời gian tháng 9 tháng 10.

5.3 VCM

5.3.1 Thông tin cơ bản

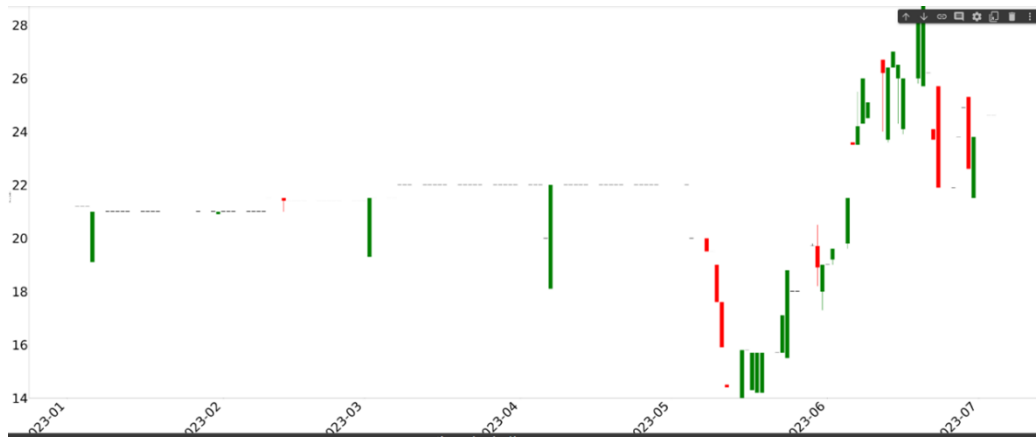
- Giá cao nhất: 29.1 (max adHigh).
- Giá thấp nhất: 12.5 (min adLow).
- Lượng tăng giá cao nhất: 2.5 (max adChange).
- Lượng giảm giá nhiều nhất: -2.9 (min adChange).
- Tỷ lệ giá giảm cao nhất: -10 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 10 (max pctChange).
- Khối lượng giao dịch khớp lệnh lớn nhất: 113300 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 0 (min nmVolume).
- KLGD khớp lệnh trung bình: 1708.173333 (mean nmVolume).
- Khoảng dao động giá lớn, tuy nhiên, có đến 75% tỷ lệ dao động giá nằm dưới 0% tức giá giảm. Ta cũng thấy là mức dao động cũng rất đa dạng. Nhưng cần lưu ý, những phiên gần đây giá dao động đặc biệt mạnh, cần phải chú ý và tính toán kỹ lưỡng nếu muốn đầu tư vào mã này. Và cũng cần để ý khi công ty này đang kinh doanh không thực sự tốt khi EPS chỉ đạt 70.646646, và nhìn vào KLGD trung bình cũng tương đối ít ỏi nên đây thực sự không phải một mã đáng để đầu tư.



5.3.2 Biểu đồ giá năm 2023

Nhìn vào biểu đồ giá này, ta dễ dàng thấy đa số đều là các cột đen điều đó có nghĩa là trong 1 phiên, giá của mã chứng khoán này rất ít khi tăng hay giảm. Và nếu có tăng/giảm thì sẽ đều sẽ rất mạnh, như từ đang 17

có thể nhảy thẳng giá lên tới gần 22. Trong những phiên gần đây việc lên xuống giá này cũng rất thất thường, nhưng giá đang đã qua đỉnh và đang có dấu hiệu đi xuống. Nhìn chung, sẽ cực kì mạo hiểm nếu đầu tư vào mã chứng khoán này bởi tính khó đoán cũng như việc tăng giảm mạnh của nó, điều mà có thể là do hiện tượng lái giá trên thị trường.



5.3.3 Biểu đồ giá năm 2022 và 2021

2022:



2021:



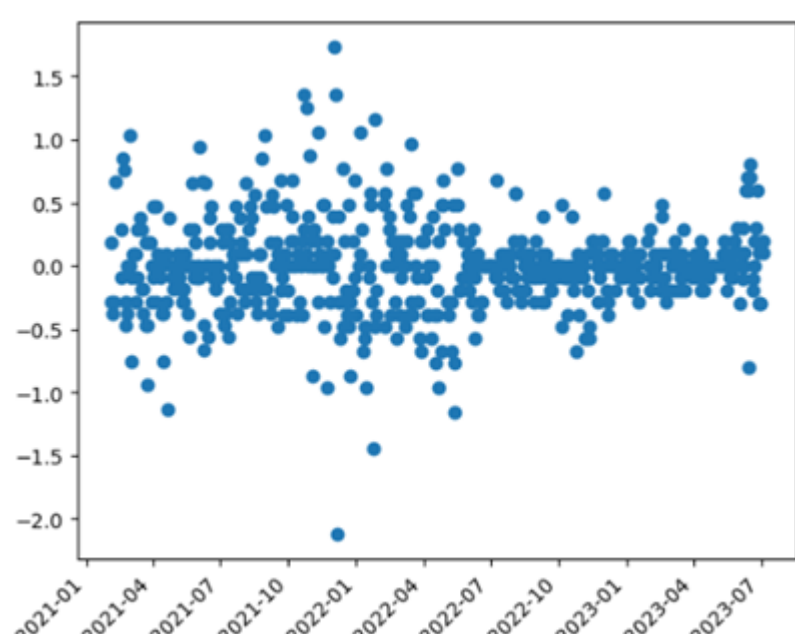
Năm 2022 cũng tương tự năm 2023, biểu đồ nhìn khá thừa thớt vì giá trong 1 phiên của mã khá “yên bình”, Và mức biến động xảy ra cũng

cực kì phức tạp cũng như là rất mạnh. Năm 2021 thì có chút khăm khá hơn, nhưng vào thời điểm này năm ấy đang là thời điểm mà giá đang giảm rất mạnh, gần đến đáy vậy nên nhìn chung, đây là một mã chứng khoán quá mạo hiểm và không đáng đầu tư.

5.4 C69

5.4.1 Thông tin cơ bản

- Giá cao nhất: 21.807 (max adHigh).
- Giá thấp nhất: 4.728 (min adLow).
- Lượng tăng giá cao nhất: 1.737 (max adChange).
- Lượng giảm giá nhiều nhất: -2.2 (min adChange).
- Tỷ lệ giá giảm cao nhất: -10.000942 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 10 (max pctChange).
- Khối lượng giao dịch khớp lệnh lớn nhất: 2.795000×10^6 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 3100 (min nmVolume).
- KLGD khớp lệnh trung bình: 264067 (mean nmVolume).
- Lại là một mã chứng khoán có khoảng dao động giá khá lớn. Tuy nhiên mật độ phân bố đã trải rộng ra rất nhiều chứ không tập trung chủ yếu ở 0. Một điều đáng chú ý nữa ở mã này đó là dường như đây là một mã khá “hot” trên thị trường khi không có ngày nào là không có giao dịch và thậm chí KLGD trung bình cũng rất cao, rất đáng để nhà đầu tư tìm hiểu. Tuy nhiên, chỉ số EPS lại gây thất vọng khi khá thấp: 403.059248.



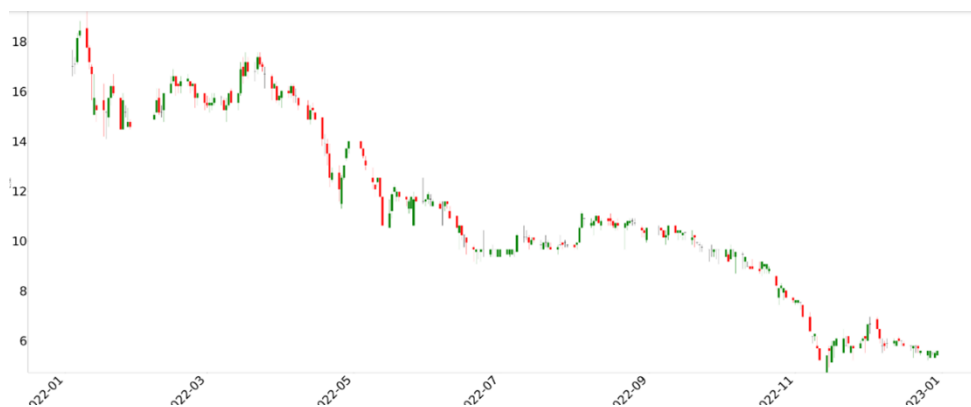
5.4.2 Biểu đồ giá năm 2023



Ta thấy được rằng, giá đang ở mức rất cao sau một thời gian suy thoái nặng nề, đặc biệt đang tăng nhanh kể từ đầu tháng 6, nhưng hiện giá đang có hiện tượng đi ngang, nhà đầu tư có thể áp dụng thêm một số chỉ báo khác để đưa ra quyết định liệu có nên chốt lời ngay không. Và hiển nhiên, không nên mua thêm nếu không có căn cứ nào rõ ràng.

5.4.3 Biểu đồ giá năm 2022 và 2021

2022:



2021:



- Chúng ta thấy rằng, 2 năm này thực sự rất trái ngược nhau, nếu 2021 giá cổ phiếu có tăng có giảm nhưng vẫn lên rất mạnh, thì năm 2022 thực sự là thảm họa khi mà giá cổ phiếu tụt cực kỳ sâu và gần như là “không phanh”.
- Và nhìn vào hai năm nay, ta thấy tháng 7 và 8 là những tháng mà cổ phiếu đi ngang, rồi trở lại xu hướng như trước.
- Và hiện năm 2023 đang có xu hướng tăng, nên ta có thể dự đoán là cổ phiếu sẽ tăng lại và nên giữ. Điều này mâu thuẫn với kết luận trên phần b, nên ta sẽ cần dùng các chỉ báo để đào sâu hơn.
- Và cũng có thể đoán rằng, việc khối lượng giao dịch cực kỳ lớn của mã có thể là do việc giá lao xuống dốc cũng như lên đỉnh một cách không phanh nên thu hút rất nhiều nhà đầu tư mua/bán.

5.5 VMS

5.5.1 Thông tin cơ bản

- Giá cao nhất: 24.5 (max adHigh).
- Giá thấp nhất: 6.565 (min adLow).
- Lượng tăng giá cao nhất: 2 (max adChange).
- Lượng giảm giá nhiều nhất: -2.4 (min adChange).
- Tỷ lệ giá giảm cao nhất: -10.001641 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 10.002215 (max pctChange).
- Khối lượng giao dịch khớp lệnh lớn nhất: 423349 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 0 (min nmVolume).
- KLGD khớp lệnh trung bình: 6205.621667 (mean nmVolume).
- Đây là một mã chứng khoán rất tiềm năng khi đa số thời gian đều tăng giá (chỉ có 25% tập dữ liệu tỷ lệ tăng/giảm giá nằm dưới 0). Khoảng tăng giảm giá cũng rất lớn. Đặc biệt, càng về gần những phiên gần đây, giá càng biến động mạnh, khi nhìn qua có thể thấy 2 lần giá tăng giảm mạnh nhất đều là những phiên mới đây. Tuy nhiên, EPS của mã này lại chưa thực sự ấn tượng khi chỉ có 995.521149, dưới mức 1000.

5.5.2 Biểu đồ giá năm 2023

5.5.3 Biểu đồ giá năm 2022 và 2021

5.6 DL1

5.6.1 Thông tin cơ bản

- Giá cao nhất: 23.863 (max adHigh).
 - Giá thấp nhất: 2.1 (min adLow).
 - Lượng tăng giá cao nhất: 1.527 (max adChange).
 - Lượng giảm giá nhiều nhất: -2.387 (min adChange).
 - Tỷ lệ giá giảm cao nhất: -10.011972 (min pctChange).
 - Tỷ lệ giá tăng cao nhất: 10 (max pctChange).
 - Khối lượng giao dịch khớp lệnh lớn nhất: 1.345 x107(max nmVolume).
 - Khối lượng giao dịch khớp lệnh ít nhất: 0 (min nmVolume).
 - KLGD khớp lệnh trung bình: 1193300 (mean nmVolume).
- Tiếp tục là một mã chứng khoán có biên độ tăng giảm giá khá rộng. Tuy nhiên, điểm nổi bật nhất của mã chứng khoán này chính là ở KLGD đặc biệt lớn, khi mà trung bình đã có tới 1193300 GD khớp lệnh. Từ biểu đồ biến động giá ta cũng có thể thấy, gần đây giá không biến động nhiều, nên ta có thể dự đoán rằng giá đang có xu hướng đi ngang. EPS của DL1 cũng không quá tốt, chỉ có 417.353415, nên đây sẽ là một mã chứng khoán mà ta cần suy xét rất nhiều.

5.6.2 Biểu đồ giá năm 2023

5.6.3 Biểu đồ giá năm 2022 và 2021

5.7 VC9

5.7.1 Thông tin cơ bản

- Giá cao nhất: 28.4 (max adHigh).
- Giá thấp nhất: 4.1 (min adLow).
- Lượng tăng giá cao nhất: 2.4 (max adChange).
- Lượng giảm giá nhiều nhất: -2.6 (min adChange).
- Tỷ lệ giá giảm cao nhất: -10 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 10 (max pctChange).
- Khối lượng giao dịch khớp lệnh lớn nhất: 1080635 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 0 (min nmVolume).
- KLGD khớp lệnh trung bình: 54698.08 (mean nmVolume).

- Tiếp tục là một mã chứng khoán có biên độ tăng giảm giá khá rộng. Nhìn vào biểu đồ ta có thể thấy được tầm giữa năm ngoái đã có một sự biến động giá vô cùng mạnh mẽ. Tuy nhiên, điều ta quan tâm nhất hiện tại là giá đang đi ngang và có vẻ không có quá nhiều biến động lớn. Và cần để ý, EPS của VC9 là rất bé, chỉ có 147.88713.

5.7.2 Biểu đồ giá năm 2023

5.7.3 Biểu đồ giá năm 2022 và 2021

5.8 CAG

5.8.1 Thông tin cơ bản

- Giá cao nhất: 36.359 (max adHigh).
- Giá thấp nhất: 6.154 (min adLow).
- Lượng tăng giá cao nhất: 3.022 (max adChange).
- Lượng giảm giá nhiều nhất: -3.588 (min adChange).
- Tỷ lệ giá giảm cao nhất: -10.001366 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 10.002206 (max pctChange).
- Khối lượng giao dịch khớp lệnh lớn nhất: 578200 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 0 (min nmVolume).
- KLGD khớp lệnh trung bình: 30027.858333 (mean nmVolume).
- Ta thấy rằng sau đợt biến động mạnh vào tháng 9-10 năm 2021 thì giá của CAG tăng/giảm không quá nhiều thất thường, phân phối tăng giảm giá cũng khá đều và xoay quanh giá trị 0. EPS của mã này cũng không đáng kì vọng cho lắm với chỉ 409.484004.

5.8.2 Biểu đồ giá năm 2023

5.8.3 Biểu đồ giá năm 2022 và 2021

5.9 VIF

5.9.1 Thông tin cơ bản

- Giá cao nhất: 22.38 (max adHigh).
- Giá thấp nhất: 9.463 (min adLow).
- Lượng tăng giá cao nhất: 1.736 (max adChange).
- Lượng giảm giá nhiều nhất: -0.091 (min adChange).
- Tỷ lệ giá giảm cao nhất: -10.001493 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 9.929078 (max pctChange).

- Khối lượng giao dịch khớp lệnh lớn nhất: 5033600 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 0 (min nmVolume).
- KLGD khớp lệnh trung bình: 11493 (mean nmVolume).
- Ta có thể thấy giá của cổ phiếu biến động khá mạnh mẽ và tần suất tăng giảm mạnh cũng diễn ra khá thường xuyên nên sẽ cần tính toán cẩn thận nếu muốn đầu tư vào mã này. Tuy gần đây, mức độ biến động gần đây đã dần quy về giá trị 0 và cũng không quá mạnh như trước, đồng thời, EPS của mã khá ổn là 1303.358422 nên đây sẽ là một chứng khoán rất tiềm năng.

5.9.2 Biểu đồ giá năm 2023

5.9.3 Biểu đồ giá năm 2022 và 2021

5.10 GKM

5.10.1 Thông tin cơ bản

- Giá cao nhất: 53.3 (max adHigh).
- Giá thấp nhất: 8.449000 (min adLow).
- Lượng tăng giá cao nhất: 4.3 (max adChange).
- Lượng giảm giá nhiều nhất: -3.639 (min adChange).
- Tỷ lệ giá giảm cao nhất: -9.698529 (min pctChange).
- Tỷ lệ giá tăng cao nhất: 10.000956 (max pctChange).
- Khối lượng giao dịch khớp lệnh lớn nhất: 1814150 (max nmVolume).
- Khối lượng giao dịch khớp lệnh ít nhất: 600 (min nmVolume).
- KLGD khớp lệnh trung bình: 304312.2(mean nmVolume).
- Đây là một mã có giá khá cao khi mà giá cao nhất lên đến tận 53 tuy nhiên lượng mua vào và bán ra cũng không phải thấp chứng tỏ mã này rất có sức hút. Nhìn qua sơ đồ biến động giá thì thấy có vẻ giá rất đều, ít khi biến động thực sự mạnh và có vẻ gần đây đang có đợt tăng giá khá mạnh nên đây sẽ chưa phải là thời điểm thích hợp để mua chứng khoán này. Tuy giá cao vậy nhưng để mà xét đến chỉ số EPS thì lại chưa được tốt chỉ ở 596.72106.

5.10.2 Biểu đồ giá năm 2023

5.10.3 Biểu đồ giá năm 2022 và 2021

5.11 Kết luận

Sau khi phân tích 10 mã chứng khoán, ta có thể chia chúng thành các nhóm như sau:

- Đầu tư dài hạn: DL1, C69
 - Chốt lời (nếu có): GKM, VC9, VMS, C69, VCM, DHT
 - Mạo hiểm: VIF
 - Các mã còn lại là không nên đầu tư, hoặc cần chờ thời gian xem xét.
- Nhìn chung, thị trường chứng khoán đa số đều mới trải qua một đợt tăng khá khá, nên ta chưa thể có được những mã để đầu tư ngắn hạn, cũng như chưa có nhiều mã để đầu tư vì giá đang ở thời kì đỉnh mà sẽ là chủ yếu bán đi để chốt lời.

6 Phần 6: Lời cảm ơn

Trước tiên với tình cảm sâu sắc và chân thành nhất, cho phép nhóm chúng em được bày tỏ lòng biết ơn đến cô Nguyễn Thị Huyền Châu và thầy Đoàn Trung Phong - những người đã truyền đạt vốn kiến thức quý báu nên đề tài của chúng em mới có thể hoàn thiện tốt đẹp. Kiến thức được tiếp thu trong quá trình học không chỉ là nền tảng cho quá trình làm bài tập lớn cho một môn học mà còn là hành trang quý báu để chúng em cũng như các nhóm khác bước vào đời một cách vững chắc và tự tin.

Tuy nhiên do vốn kiến thức còn hạn chế cũng như là lần đầu làm tài liệu, khó có thể tránh khỏi những thiếu sót, sản phẩm tuy không phải là hoàn hảo nhưng đây chính là thành quả dựa trên sự cố gắng và nỗ lực của cả nhóm. Vậy nên chúng em rất mong nhận được sự thông cảm và những ý kiến đóng góp quý báu đến từ thầy cô.

Cuối cùng, chúng em xin kính chúc thầy cô dồi dào sức khỏe và thành công trong sự nghiệp cao quý, mãi là người lái đò tận tâm tận lực đưa các thế hệ sinh viên tới bến bờ tri thức. Một lần nữa, cảm ơn thầy cô vì đã tận tình giảng dạy, cung cấp kiến thức cho chúng em.

Chúng em xin chân thành cảm ơn!