

GROUP 11
DATA VISUALIZATION

STOCK VISUALIZATION

21/05/2024

THÀNH VIÊN NHÓM

19127476 - Trần Thị Huế Minh

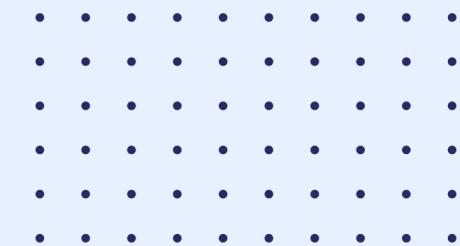
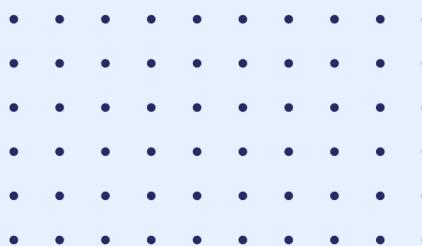
Trực quan - Phân tích hồi quy, dự đoán

19127486 - Nguyễn Lê Nguyên

Trực quan - Thống kê đơn giản

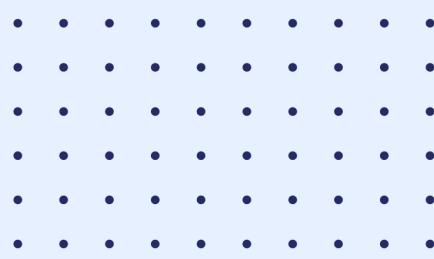
19127125 - Lê Duy Dũng

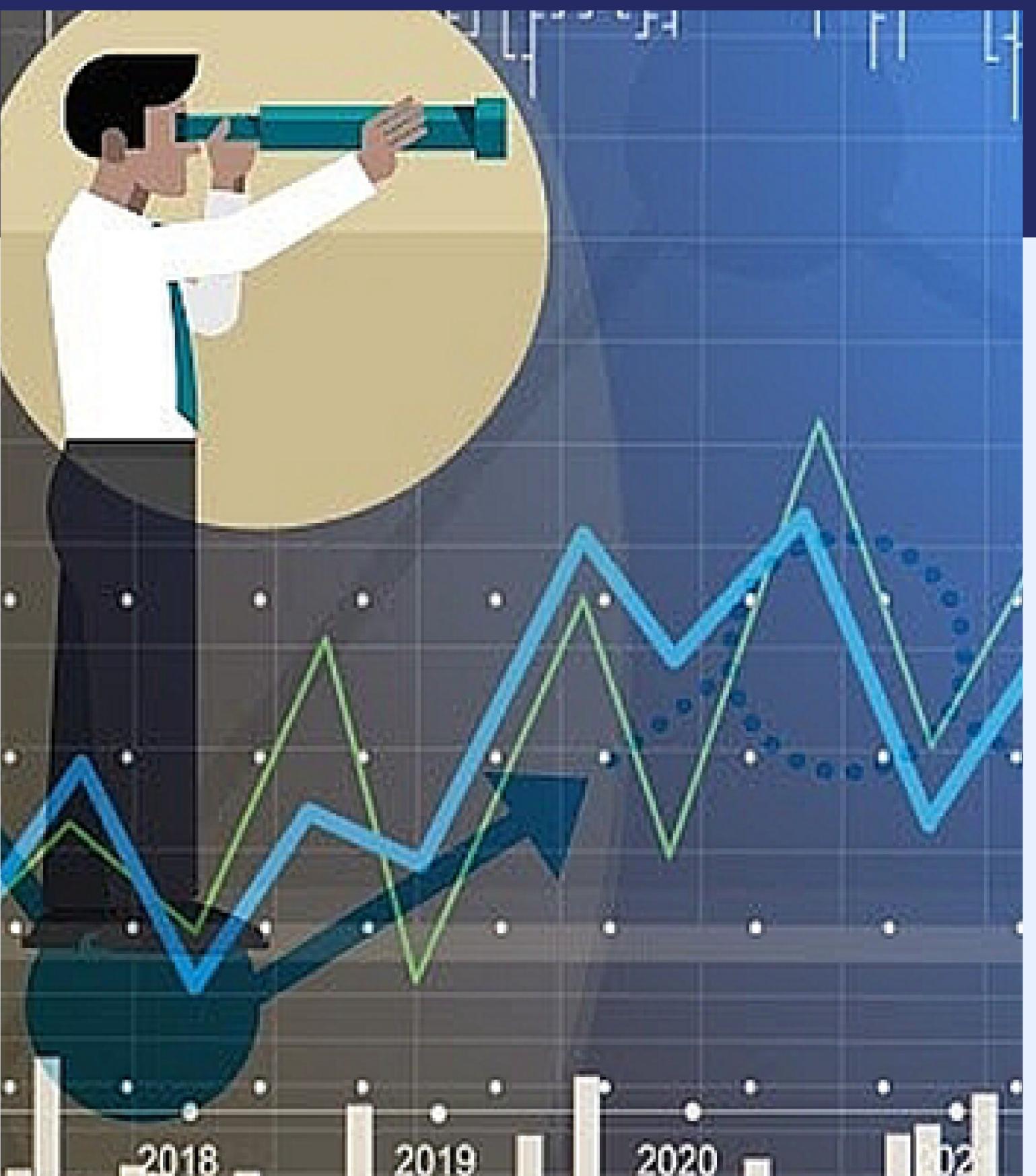
Thống kê mô tả - Website



NỘI DUNG

- Tình huống giả định
- Thống kê mô tả
- Thống kê đơn giản
- Phân tích hồi quy, dự đoán
- Trực quan
- Demo

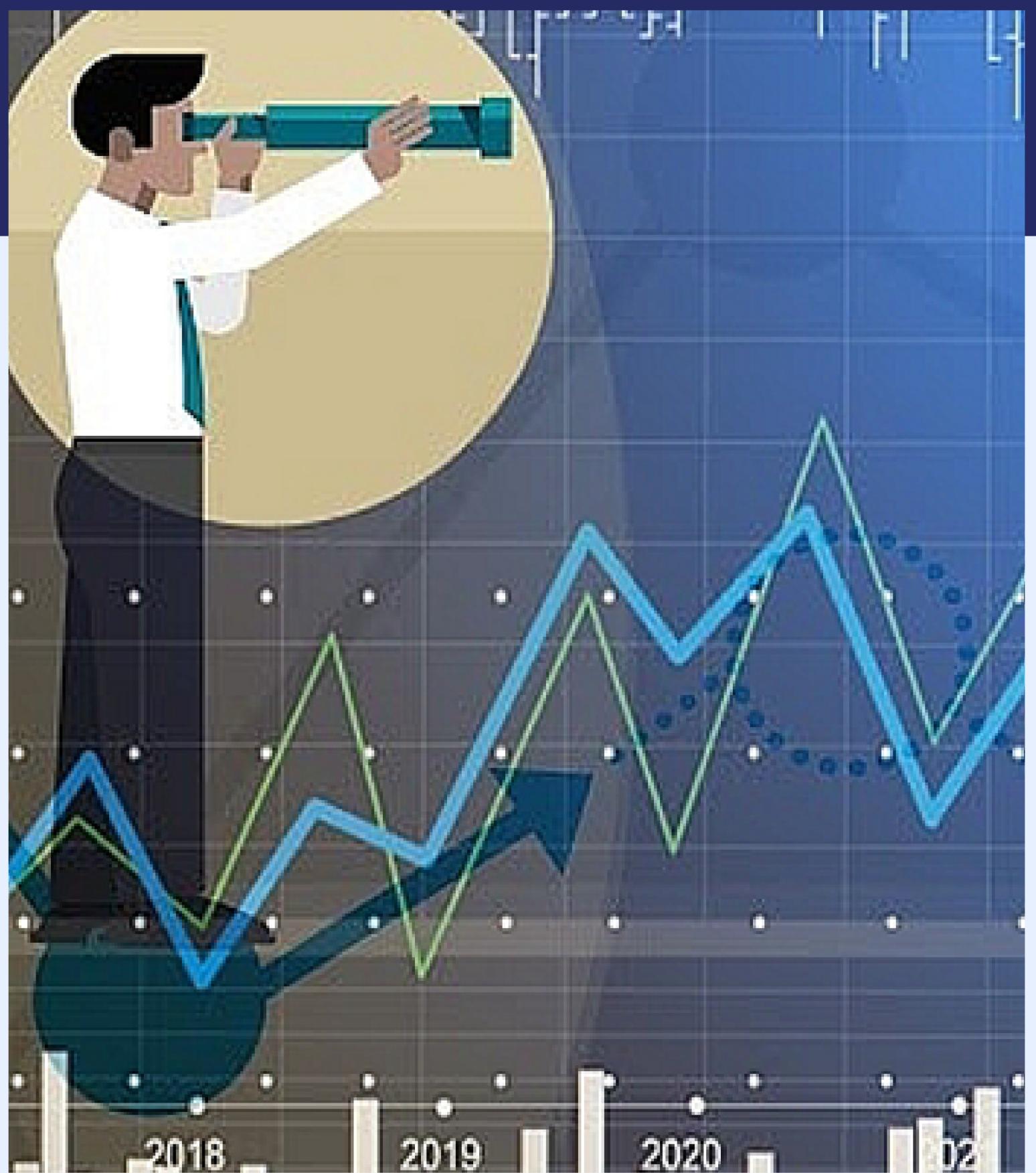




TÌNH HUỐNG GIẢ ĐỊNH

Giả sử bạn nghe ai đó nói về công việc giao dịch các mã chứng khoán có thể xem như một công việc tay trái mang lại lợi nhuận đáng kể, thậm chí là có người trở nên giàu có nhờ công việc đó. Các câu chuyện đó khiến bạn tò mò, hứng thú,... Và bạn bắt đầu tìm hiểu về chứng khoán...

Thế là bạn bắt đầu tìm hiểu, mày mò về chứng khoán. Không lâu sau đó, bạn kiếm được 1 triệu đô la đầu tiên... trên tài khoản demo thứ n. Rồi bạn bắt đầu tham gia vào thị trường và cháy tài khoản không lâu sau đó. Khi đó bạn nhận ra rằng muốn kiếm tiền bằng công việc này thì việc dự đoán theo quán tính hay nghe lời người khác là hoàn toàn không thể



TÌNH HUỐNG GIẢ ĐỊNH

Bạn phải thật sự hiểu về chứng khoán, về các con số, về các quy luật ẩn sâu bên trong các đồ thị... Chỉ khi đó việc $\times 10$ $\times 100$ tài khoản mới có thể xảy ra trước khi tài khoản của bạn bị cháy lần tiếp theo.

Ở đồ án này, nhóm sẽ tìm hiểu về các con số, các dữ liệu liên quan đến một số mã chứng khoán ở Việt Nam. Từ đó dựng lên các trực quan hóa dữ liệu, các đồ thị mà chỉ cần nhìn vào 5 phút là bạn đã có thể trở thành một nhà giao dịch chuyên nghiệp.

THỐNG KÊ MÔ TẢ

Nhóm sử dụng dữ liệu về lịch sử giá và khối lượng giao dịch của một số cổ phiếu thuộc các ngân hàng Việt Nam trên các sàn chứng khoán HOSE và HNX

	<Ticker>	<DTYYYYMMDD>	<OpenFixed>	<HighFixed>	<LowFixed>	<CloseFixed>	<Volume>
0	VCB	20170103	26.1847	27.5823	26.1847	27.3984	2018750
1	VCB	20170104	27.5823	27.6558	27.0674	27.2145	1339740
2	VCB	20170105	27.1409	27.4351	27.1042	27.3248	855250
3	VCB	20170106	27.3616	28.2442	27.2880	27.8765	2495910
4	VCB	20170109	28.0971	28.6488	28.0236	28.4649	2379510
...
5378	TCB	20230417	29.0000	29.4500	28.7500	29.4500	3014200
5379	TCB	20230418	29.4000	29.4000	28.9000	29.1500	2553900
5380	TCB	20230419	29.2000	29.4000	29.0000	29.0000	2611900
5381	TCB	20230420	28.9000	29.0000	28.5500	28.8000	2124700
5382	TCB	20230421	29.0000	29.0000	28.7000	28.7000	2354900

- Số lượng dòng và cột của tập dữ liệu

Số lượng dòng: 15383
Số lượng cột: 7

- Thông tin các cột của dữ liệu

0	<Ticker>	15383	non-null	object
1	<DTYYYYMMDD>	15383	non-null	int64
2	<OpenFixed>	15383	non-null	float64
3	<HighFixed>	15383	non-null	float64
4	<LowFixed>	15383	non-null	float64
5	<CloseFixed>	15383	non-null	float64
6	<Volume>	15383	non-null	int64

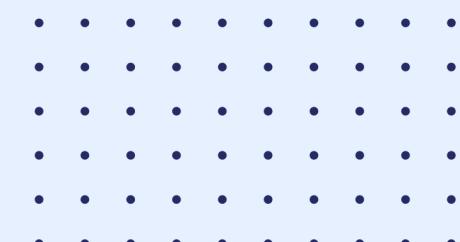
THỐNG KÊ MÔ TẢ

- Dữ liệu có 7 cột, trong đó:
 - <Ticker> : Mã cổ phiếu của các ngân hàng
 - <DTYYYYMMDD> : Ngày được ghi nhận.
 - <OpenFixed> : Giá mở là giá của một cổ phiếu, chứng khoán hoặc sản phẩm tài chính khác khi phiên giao dịch bắt đầu. Đây là giá đầu tiên được đưa ra trong ngày giao dịch
 - <HighFixed> : Giá đỉnh là giá cao nhất mà một cổ phiếu, chứng khoán hoặc sản phẩm tài chính khác đạt được trong một phiên giao dịch.
 - <LowFixed> : Giá sàn là giá thấp nhất mà một cổ phiếu, chứng khoán hoặc sản phẩm tài chính khác đạt được trong một phiên giao dịch.
 - <CloseFixed> : Giá đóng là giá của một cổ phiếu, chứng khoán hoặc sản phẩm tài chính khác khi phiên giao dịch kết thúc. Đây là giá cuối cùng được đưa ra trong ngày giao dịch.
 - <Volume> : Khối lượng giao dịch là tổng số lượng cổ phiếu, chứng khoán hoặc sản phẩm tài chính khác được giao dịch trong một phiên giao dịch. Đây là một chỉ số quan trọng để đo lường sự sôi động của thị trường và khối lượng lớn nhỏ của giao dịch có thể ảnh hưởng đến giá của : một sản phẩm tài chính.

THỐNG KÊ MÔ TẢ

- Nhập xét dữ liệu

- Số lượng cổ phiếu: 10
 - ACB: Ngân hàng thương mại cổ phần Á Châu
 - BID: Ngân hàng TMCP Đầu tư và Phát triển Việt Nam
 - FPT: Tập đoàn FPT
 - GAS: Tổng Công ty Khí Việt Nam - CTCP
 - HPG: Tập đoàn Hòa Phát
 - MBB: Ngân hàng TMCP Quân Đội
 - MSN: CTCP Tập đoàn Masan
 - SAB: Tổng Công ty cổ phần Bia - Rượu - Nước giải khát Sài Gòn
 - TCB: Ngân hàng Thương mại cổ phần Kỹ Thương Việt Nam
 - VCB: Ngân hàng thương mại cổ phần Ngoại thương Việt Nam
- Số ngày lấy dữ liệu: 1575
 - Từ ngày 3/1/2017 đến ngày 21/4/2023

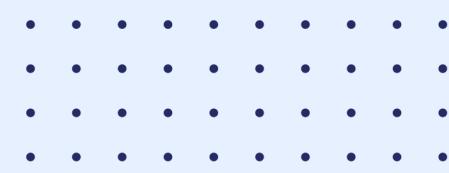


THỐNG KÊ MÔ TẢ

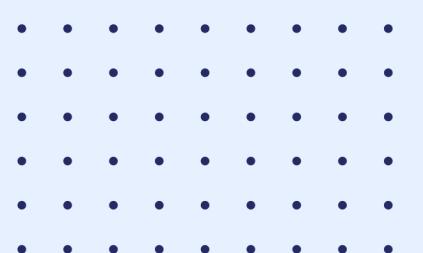
Các đại lượng thống kê

Xét toàn tập dữ liệu

- Trong TH này, ta tập trung vào giá đóng và khối lượng giao dịch vì đây là dữ liệu quan trọng trong việc phân tích dữ liệu giá cổ phiếu:
 - Giá đóng trung bình là 57.015905 VNĐ
 - Khối lượng giao dịch trung bình mỗi ngày là 3,902,440
- Variance của dữ liệu cho từng cột là rất cao, suy ra dữ liệu phân tán rộng và không gần giá trị trung bình.



	<OpenFixed>	<HighFixed>	<LowFixed>	<CloseFixed>	<Volume>
count	15383.000000	15383.000000	15383.000000	15383.000000	1.538300e+04
mean	57.018182	57.779522	56.255356	57.015905	3.902440e+06
std	54.520037	55.178253	53.776732	54.492360	6.584123e+06
min	5.298800	5.688000	5.298800	5.688000	0.000000e+00
25%	20.400000	20.681150	20.112000	20.400000	6.459000e+05
50%	34.670900	35.149300	34.200000	34.662600	1.543500e+06
75%	77.032100	77.883700	76.000000	76.954600	3.916535e+06
max	305.390400	310.036500	294.847400	302.888700	9.965880e+07
Variance	2972.434434	3044.639589	2891.936907	2969.417313	4.335067e+13
Median	34.670900	35.149300	34.200000	34.662600	1.543500e+06



THỐNG KÊ MÔ TẢ

Các đại lượng thống kê

Xét từng mã cổ phiếu

VCB

	<code><OpenFixed></code>	<code><HighFixed></code>	<code><LowFixed></code>	<code><CloseFixed></code>	<code><Volume></code>
count	1573.000000	1573.000000	1573.000000	1573.000000	1.573000e+03
mean	59.958217	60.698851	59.226411	59.992508	1.366186e+06
std	19.102602	19.321334	18.829756	19.096270	8.621718e+05
min	25.596300	25.890500	25.522800	25.633100	1.414500e+05
25%	44.036300	44.923600	43.585400	44.222200	7.757700e+05
50%	63.737800	64.121800	63.123500	63.661000	1.151200e+06
75%	76.100000	76.954600	75.174000	76.025600	1.701570e+06
max	95.800000	96.300000	94.000000	96.000000	6.724710e+06
Variance	364.909404	373.313929	354.559703	364.667535	7.433401e+11
Median	63.737800	64.121800	63.123500	63.661000	1.151200e+06

ACB

	<code><OpenFixed></code>	<code><HighFixed></code>	<code><LowFixed></code>	<code><CloseFixed></code>	<code><Volume></code>
count	1575.000000	1575.000000	1575.000000	1575.000000	1.575000e+03
mean	16.051324	16.273762	15.834504	16.053778	4.259974e+06
std	6.828868	6.890952	6.744937	6.814420	3.712935e+06
min	5.298800	5.688000	5.298800	5.688000	0.000000e+00
25%	11.009400	11.171600	10.880000	11.027700	1.799313e+06
50%	12.648500	12.837900	12.455400	12.610700	3.249267e+06
75%	23.424000	23.775000	23.040000	23.412000	5.450320e+06
max	30.400000	30.560000	29.160000	30.360000	3.515600e+07
Variance	46.633443	47.485222	45.494179	46.436317	1.378588e+13
Median	12.648500	12.837900	12.455400	12.610700	3.249267e+06

BID

	<code><OpenFixed></code>	<code><HighFixed></code>	<code><LowFixed></code>	<code><CloseFixed></code>	<code><Volume></code>
count	1573.000000	1573.000000	1573.000000	1573.000000	1.573000e+03
mean	29.417805	29.898439	28.987809	29.433877	2.145729e+06
std	8.737613	8.865694	8.605696	8.724978	1.517622e+06
min	10.642400	11.235700	10.642400	11.235700	1.808000e+05
25%	24.528900	24.911600	24.261100	24.567200	1.105060e+06
50%	31.013100	31.378700	30.662900	31.000000	1.779600e+06
75%	35.100000	35.600000	34.450000	35.050000	2.731500e+06
max	49.500000	49.950000	48.100000	49.000000	1.570104e+07
Variance	76.345874	78.600532	74.058007	76.125244	2.303178e+12
Median	31.013100	31.378700	30.662900	31.000000	1.779600e+06

FPT

	<code><OpenFixed></code>	<code><HighFixed></code>	<code><LowFixed></code>	<code><CloseFixed></code>	<code><Volume></code>
count	1573.000000	1573.000000	1573.000000	1573.000000	1.573000e+03
mean	43.584032	44.084420	43.117423	43.595040	1.595809e+06
std	24.898220	25.204546	24.587137	24.886886	1.069939e+06
min	15.370100	15.440300	15.247300	15.370100	1.414800e+05
25%	22.850200	23.171100	22.697800	22.900900	8.243000e+05
50%	32.476800	32.767300	32.193200	32.499500	1.307200e+06
75%	74.311400	75.099100	73.466500	74.230600	2.078100e+06
max	95.098400	97.057400	94.037100	95.180000	8.399400e+06
Variance	619.921361	635.269116	604.527307	619.357104	1.144770e+12
Median	32.476800	32.767300	32.193200	32.499500	1.307200e+06

THỐNG KÊ MÔ TẢ

Các đại lượng thống kê

Xét từng mã cổ phiếu

GAS					
	<OpenFixed>	<HighFixed>	<LowFixed>	<CloseFixed>	<Volume>
count	1573.000000	1573.000000	1573.000000	1573.000000	1.573000e+03
mean	84.289081	85.511283	83.108822	84.236026	6.849744e+05
std	20.239401	20.655171	19.813342	20.190912	5.072365e+05
min	41.904200	42.694800	41.904200	41.904200	3.086000e+04
25%	69.358600	70.237400	68.414900	69.358600	3.369000e+05
50%	87.004900	87.854200	85.872600	86.717200	5.660500e+05
75%	99.943900	101.744900	98.784400	100.265800	8.766100e+05
max	127.522500	130.442800	124.894100	130.442800	4.515800e+06
Variance	409.633348	426.636086	392.568502	407.672908	2.572889e+11
Median	87.004900	87.854200	85.872600	86.717200	5.660500e+05

HPG					
	<OpenFixed>	<HighFixed>	<LowFixed>	<CloseFixed>	<Volume>
count	1573.000000	1573.000000	1573.000000	1573.000000	1.573000e+03
mean	17.752543	18.024331	17.507434	17.755091	1.346355e+07
std	9.942664	10.093734	9.810086	9.933114	1.222955e+07
min	6.244400	6.314600	6.150700	6.244400	9.073400e+05
25%	10.379000	10.492800	10.265200	10.379000	4.328790e+06
50%	13.656500	13.883000	13.516400	13.656500	8.995460e+06
75%	22.800000	23.231100	22.450000	22.900000	1.989610e+07
max	43.895800	44.198600	43.366100	43.895800	9.965880e+07
Variance	98.856575	101.883466	96.237779	98.666754	1.495619e+14
Median	13.656500	13.883000	13.516400	13.656500	8.995460e+06

MBB					
	<OpenFixed>	<HighFixed>	<LowFixed>	<CloseFixed>	<Volume>
count	1573.000000	1573.000000	1573.000000	1573.000000	1.573000e+03
mean	16.609396	16.843148	16.392679	16.610412	8.179114e+06
std	7.115125	7.213165	7.005515	7.102149	6.887198e+06
min	5.835800	5.879100	5.835800	5.835800	1.628400e+05
25%	11.764800	11.966200	11.551900	11.764800	3.525870e+06
50%	13.681400	13.913400	13.530400	13.690000	6.240080e+06
75%	20.296300	20.592600	19.925900	20.259300	1.059790e+07
max	34.600000	34.900000	34.150000	34.400000	5.162240e+07
Variance	50.625005	52.029754	49.077237	50.440525	4.743350e+13
Median	13.681400	13.913400	13.530400	13.690000	6.240080e+06

MSN					
	<OpenFixed>	<HighFixed>	<LowFixed>	<CloseFixed>	<Volume>
count	1573.000000	1573.000000	1573.000000	1573.000000	1.573000e+03
mean	72.757376	73.926568	71.643829	72.815599	1.082943e+06
std	27.067465	27.613078	26.546866	27.032186	1.003069e+06
min	32.529500	33.096000	32.286800	32.893700	5.748000e+04
25%	48.551500	49.360700	47.661400	48.551500	4.840800e+05
50%	69.105000	70.157000	68.376800	69.186000	8.183800e+05
75%	90.000000	92.000000	87.948400	89.400000	1.318800e+06
max	142.286300	144.768100	140.218200	142.286300	9.870200e+06
Variance	732.647665	762.482081	704.736069	730.739093	1.006147e+12
Median	69.105000	70.157000	68.376800	69.186000	8.183800e+05

THỐNG KÊ MÔ TẢ

Các đại lượng thống kê

Xét từng mã cổ phiếu

SAB

	<OpenFixed>	<HighFixed>	<LowFixed>	<CloseFixed>	<Volume>
count	1573.000000	1573.000000	1573.000000	1573.000000	1.573000e+03
mean	191.601930	193.867358	189.099220	191.532469	9.944209e+04
std	33.863115	34.109966	33.300412	33.819837	9.188694e+04
min	105.501500	112.848900	105.030500	108.798500	1.140000e+03
25%	165.411400	167.354500	163.304100	165.702400	3.413000e+04
50%	184.949800	187.296100	182.883400	185.307100	7.510000e+04
75%	218.644200	220.575000	215.452000	217.350800	1.402000e+05
max	305.390400	310.036500	294.847400	302.888700	9.500000e+05
Variance	1146.710558	1163.489783	1108.917446	1143.781359	8.443209e+09
Median	184.949800	187.296100	182.883400	185.307100	7.510000e+04

TCB

	<OpenFixed>	<HighFixed>	<LowFixed>	<CloseFixed>	<Volume>
count	1224.000000	1224.000000	1224.000000	1224.000000	1.224000e+03
mean	32.850041	33.285335	32.392361	32.817443	6.785990e+06
std	14.253653	14.467792	13.968657	14.202612	7.512943e+06
min	15.000000	15.500000	14.000000	14.900000	3.287100e+05
25%	23.000000	23.200000	22.800000	22.950000	1.803458e+06
50%	27.100000	27.500000	26.750000	27.200000	3.357770e+06
75%	40.012500	40.462500	39.375000	39.875000	9.432225e+06
max	108.000000	109.000000	104.500000	109.000000	5.823550e+07
Variance	203.166637	209.317011	195.123386	201.714187	5.644431e+13
Median	27.100000	27.500000	26.750000	27.200000	3.357770e+06

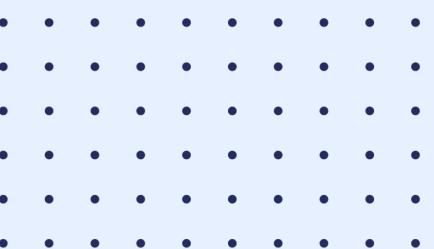
Nhận xét:

- Nhìn chung giá đóng trung bình của cổ phiếu SAB là cao nhất so với các cổ phiếu còn lại
- Tuy nhiên về độ lệch về giá và volume của một số cổ phiếu là khá cao như VCB, FPT, GAS, MSN, SAB, TCB

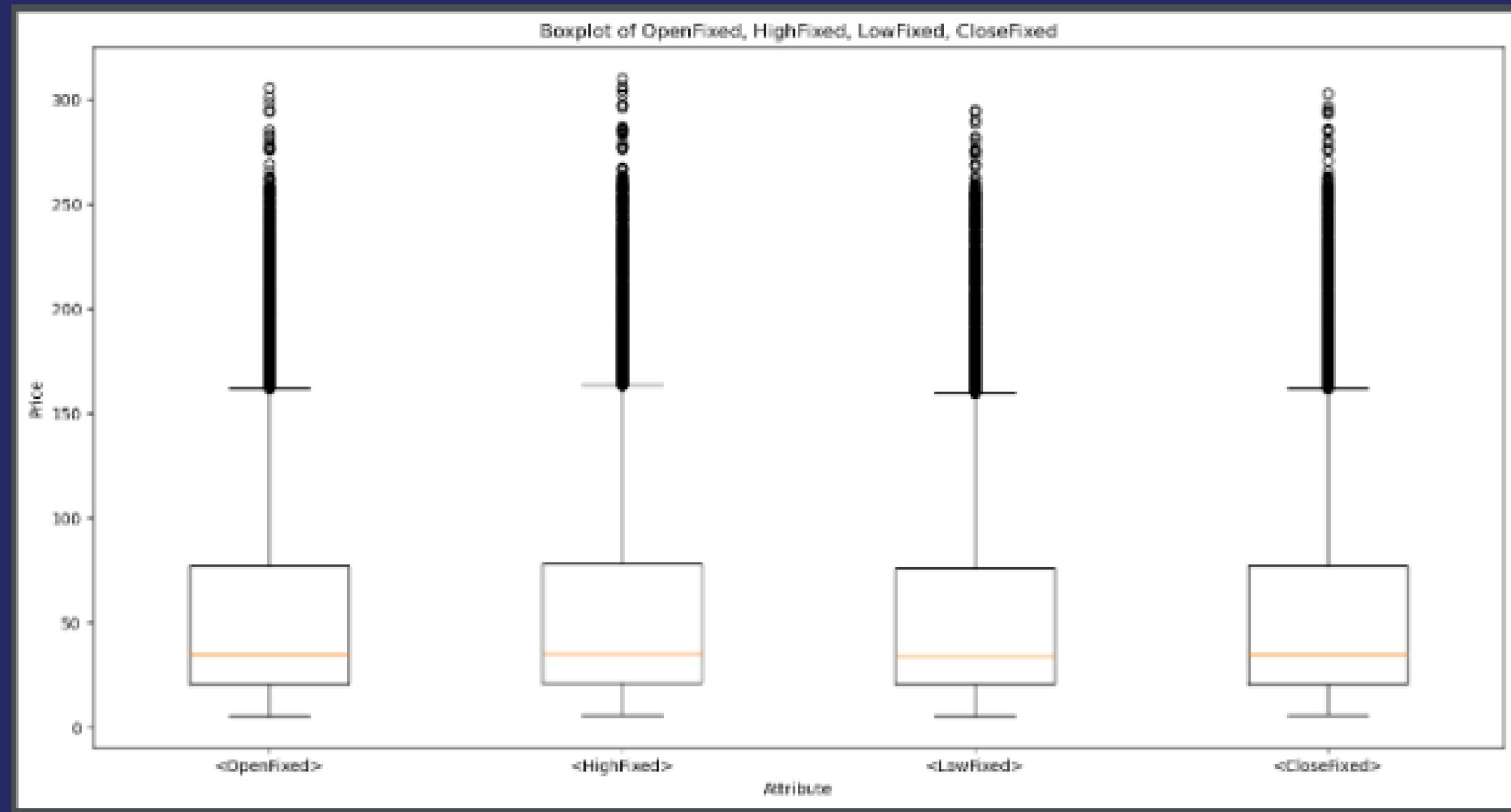
THỐNG KÊ MÔ TẢ



BOXPLOT BIỂU ĐỒ PHÂN BỐ DỮ LIỆU



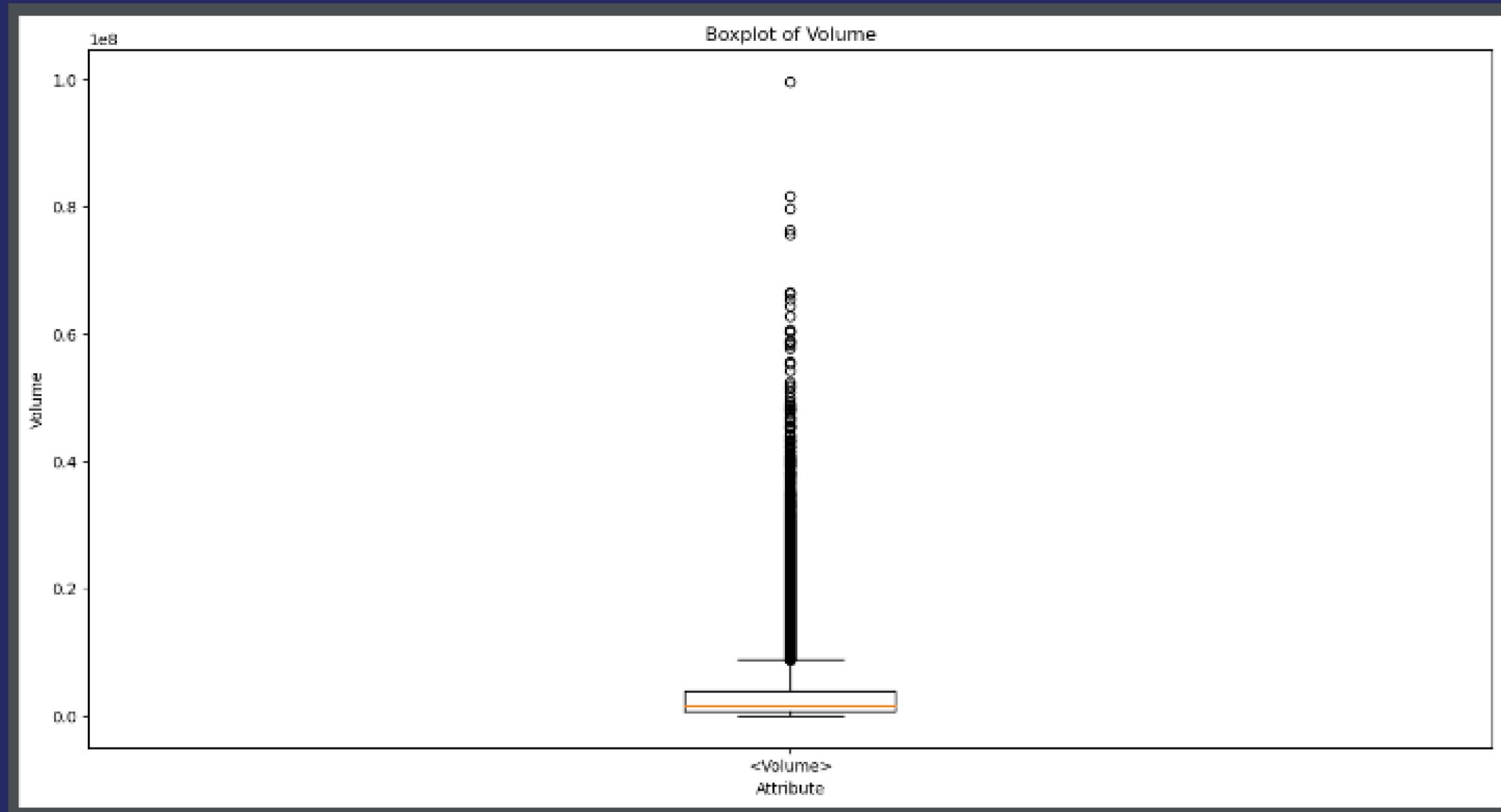
QUAN SÁT ĐỘ PHÂN BỐ DỮ LIỆU CỦA CÁC CỘT GIÁ MỞ, GIÁ ĐÓNG, GIÁ CAO, GIÁ THẤP CỦA TOÀN BỘ DỮ LIỆU



Nhận xét:

- Về các loại giá của cổ phiếu có sự tập trung từ khoảng 25 đến 80 ngàn đồng
- Tuy nhiên, có rất nhiều điểm ngoại lệ xuất hiện cao trên mức 150 ngàn đến hơn 300 ngàn đồng
- Không có quá nhiều sự chêch lệch giữa trung vị của các loại giá, nghĩa là giữa giá đóng, giá mở, và giá cao - thấp trong ngày ít có sự thay đổi lớn.

QUAN SÁT ĐỘ PHÂN BỐ DỮ LIỆU VOLUMN CỦA TOÀN BỘ DỮ LIỆU

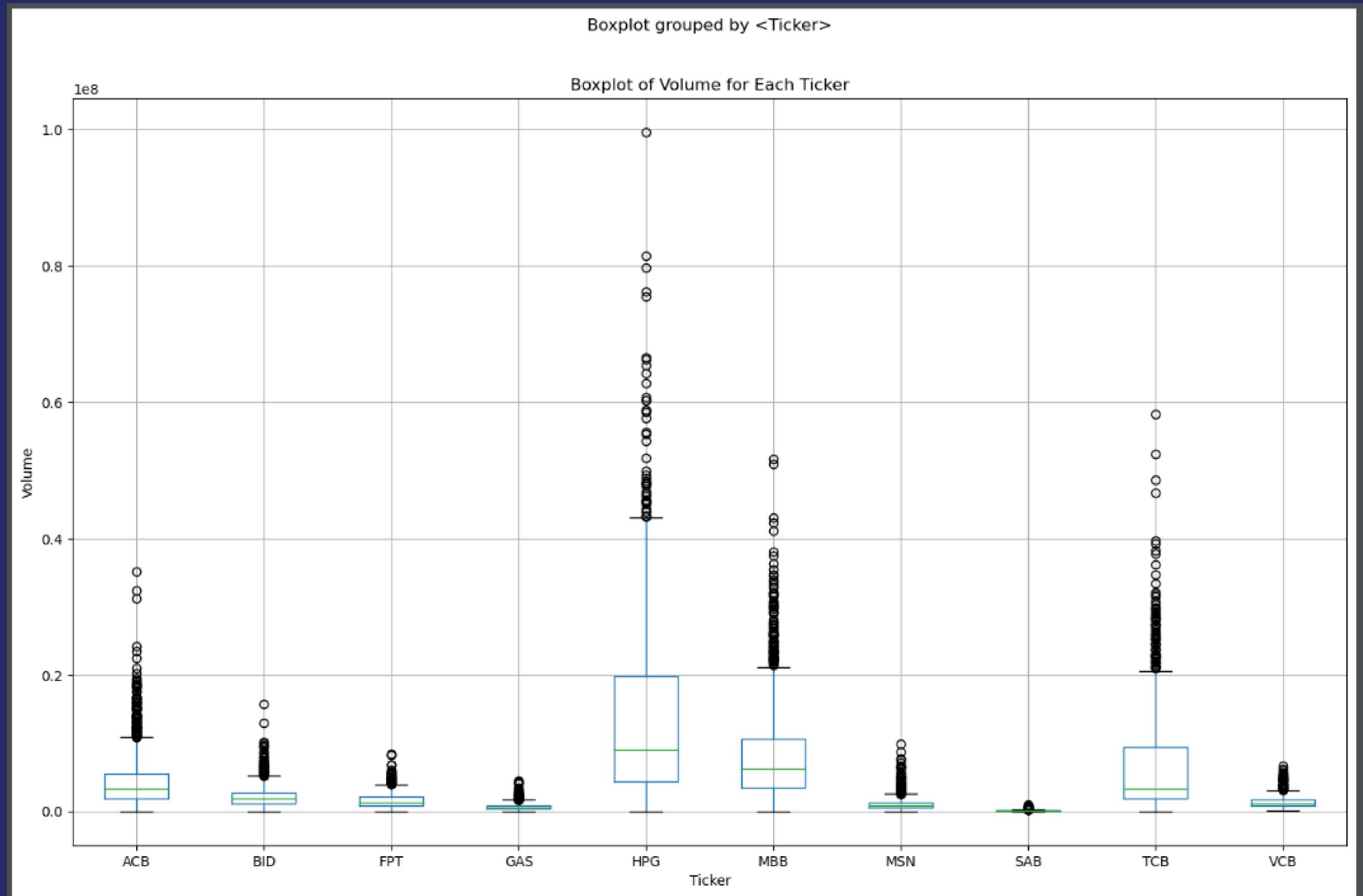


Nhận xét:

- Khối lượng giao dịch phần lớn là dưới 10.000.000 lượt giao dịch, nhưng các điểm ngoại lệ cũng rất nhiều: có những ngày giao dịch của một cổ phiếu lên đến 100.000.000, gấp 10 lần so với lượt giao dịch trung vị

QUAN SÁT ĐỘ PHÂN BỐ DỮ LIỆU CỦA CÁC CỘT THEO TỪNG CỔ PHIẾU

• VOLUME

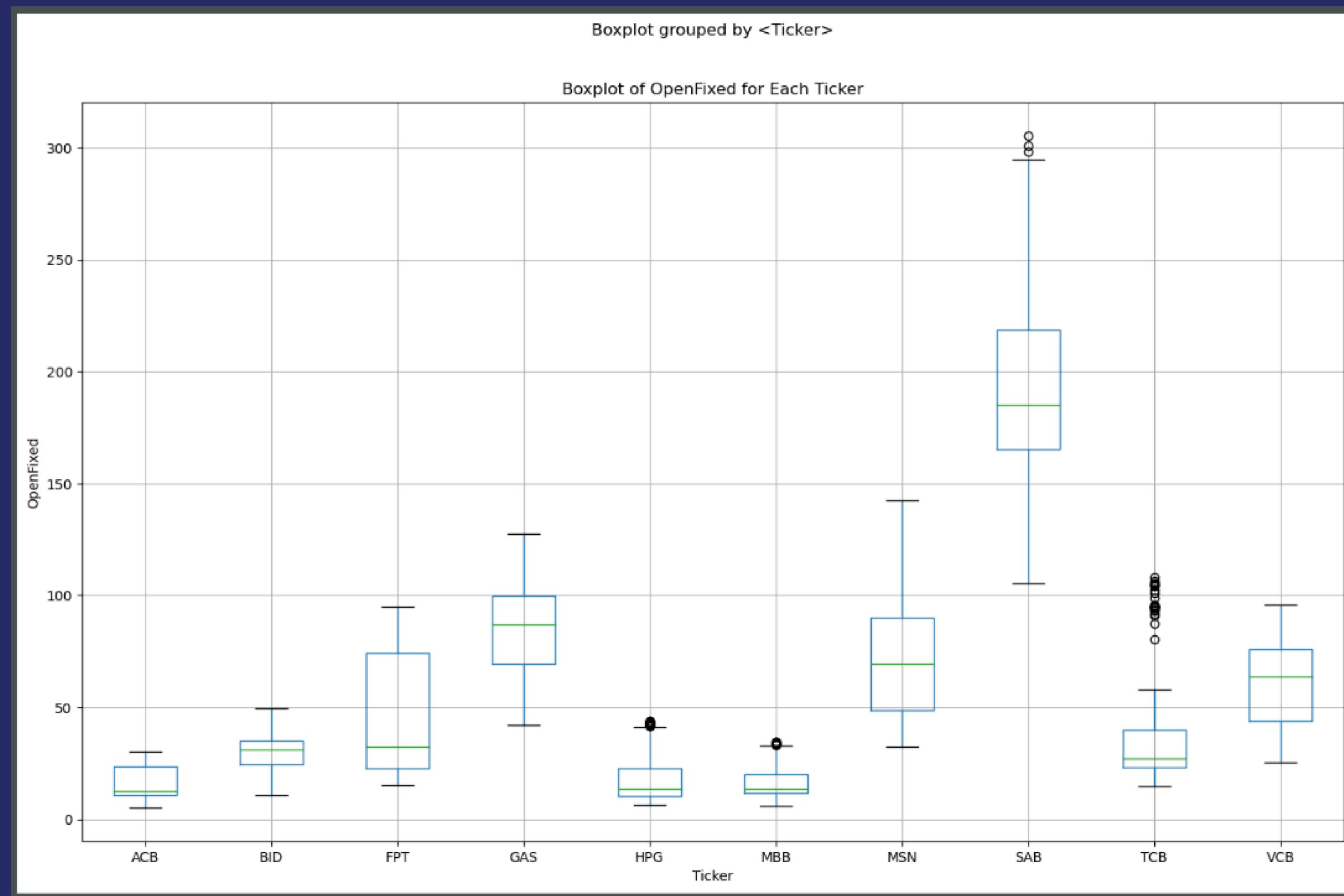


Nhận xét:

- Quan sát ta thấy HPG có khối lượng giao dịch cao hơn hẳn so với các loại cổ phiếu còn lại và khối lượng giao dịch cao nhất cũng đến từ cổ phiếu HPG

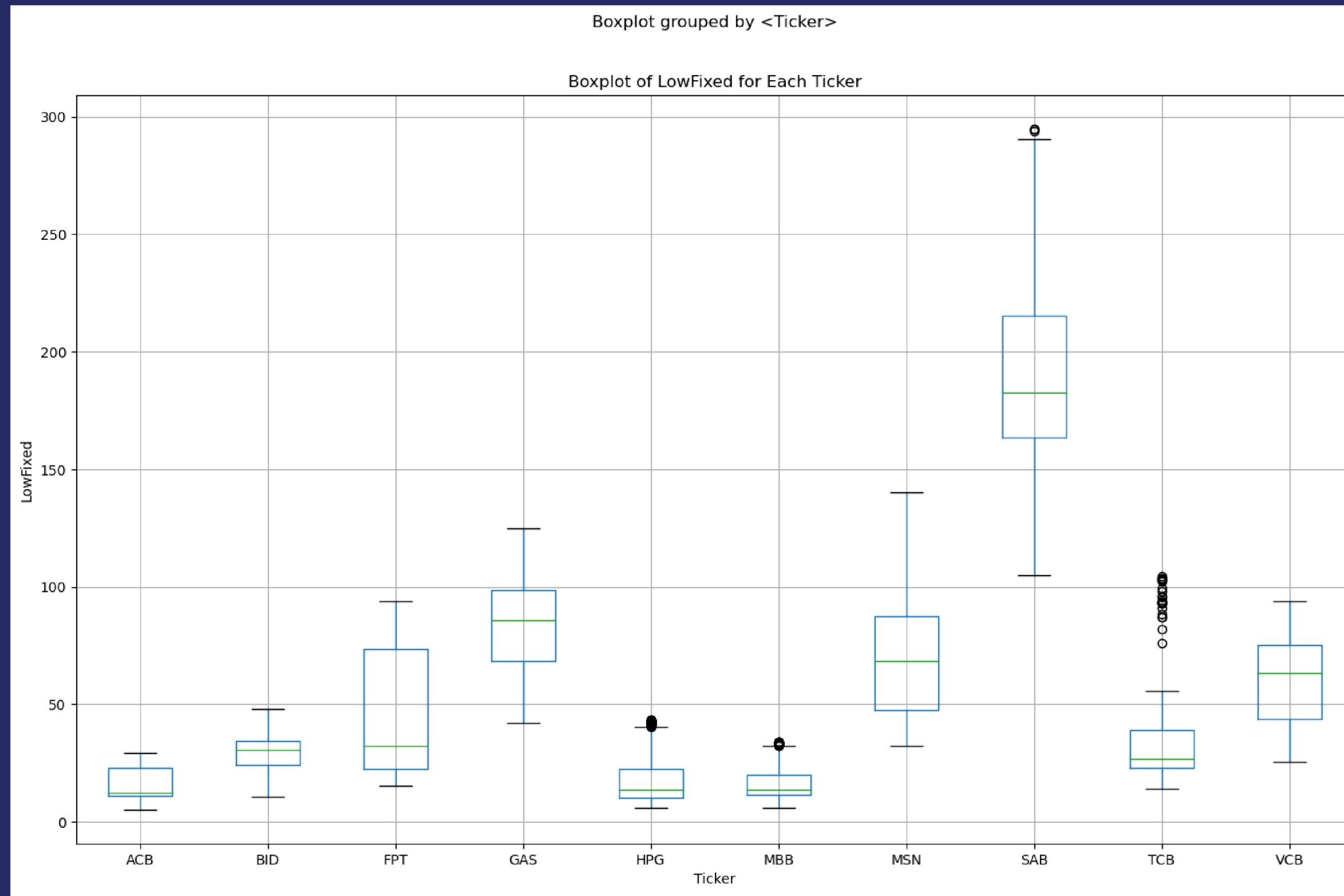
QUAN SÁT ĐỘ PHÂN BỐ DỮ LIỆU CỦA CÁC CỘT THEO TỪNG CỔ PHIẾU

- OPENFIXED



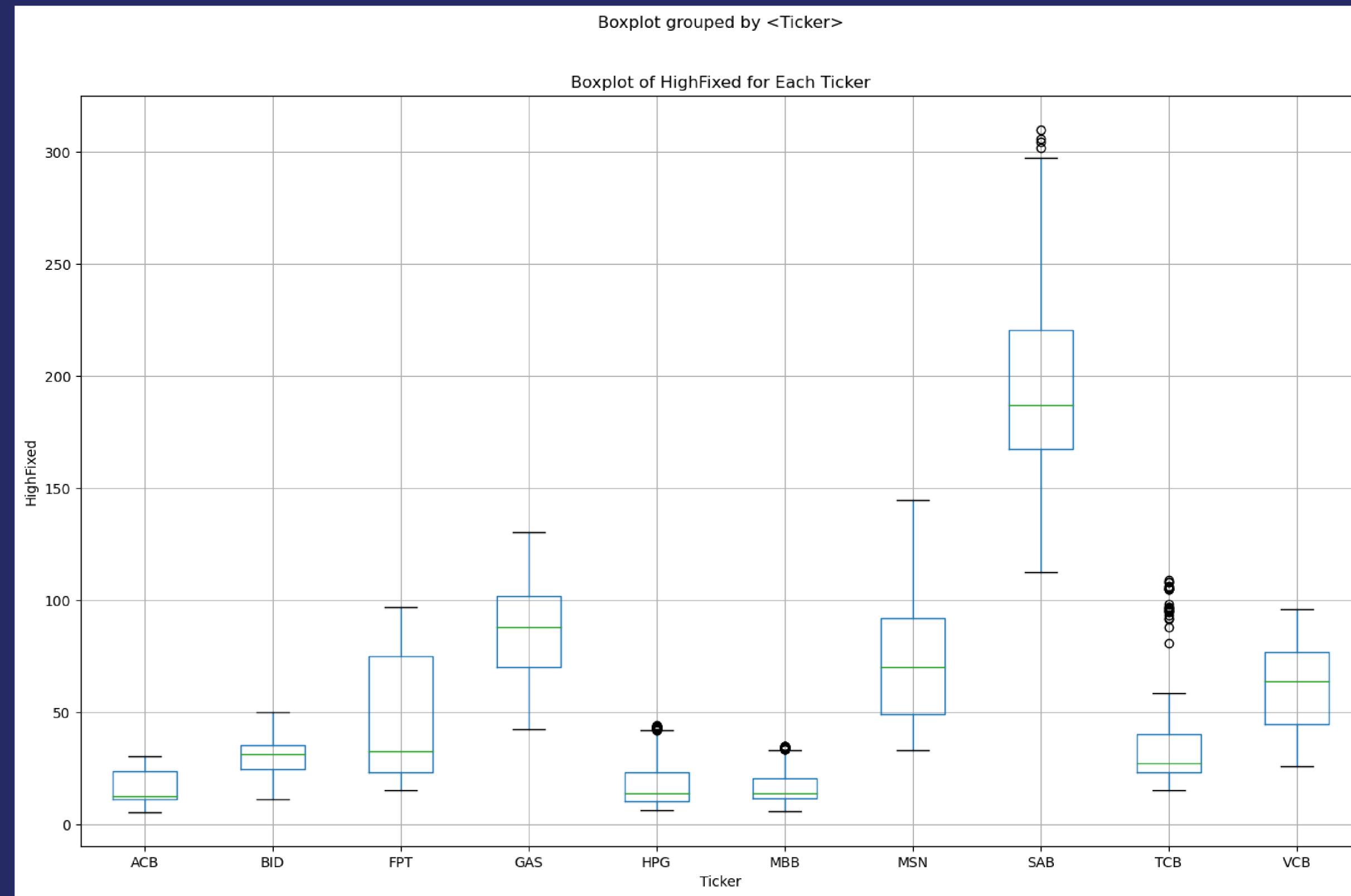
QUAN SÁT ĐỘ PHÂN BỐ DỮ LIỆU CỦA CÁC CỘT THEO TỪNG CỔ PHIẾU

• LOWFIXED



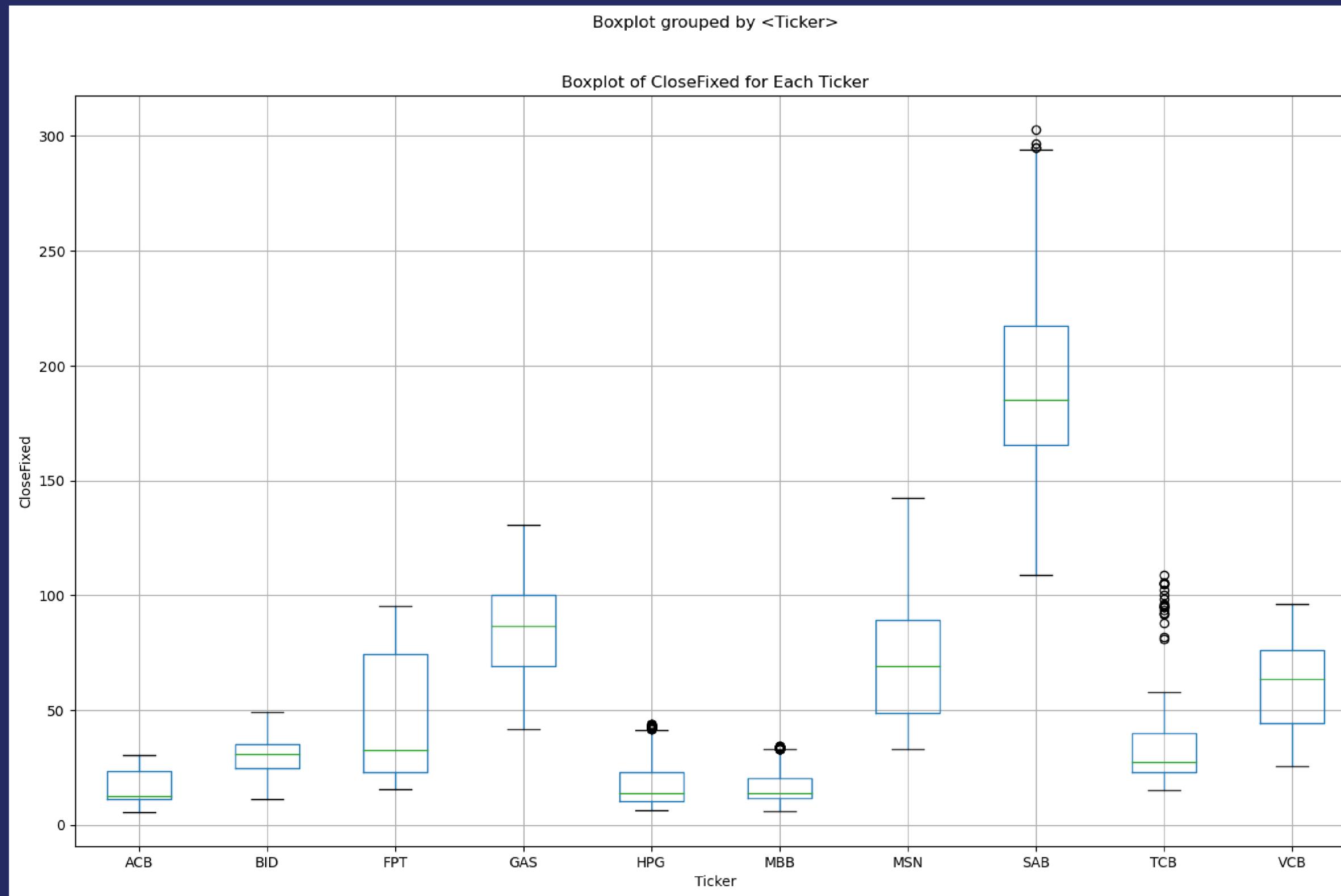
QUAN SÁT ĐỘ PHÂN BỐ DỮ LIỆU CỦA CÁC CỘT THEO TỪNG CỔ PHIẾU

- HIGHFIXED



QUAN SÁT ĐỘ PHÂN BỐ DỮ LIỆU CỦA CÁC CỘT THEO TỪNG CỔ PHIẾU

• CLOSEFIXED



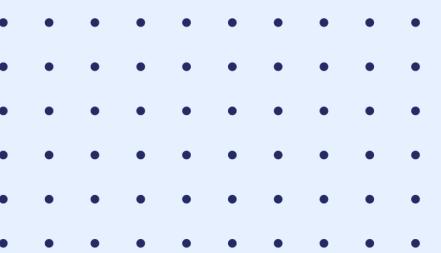
Nhận xét:

- Về giá, SAB là cổ phiếu có giá cao nhất so với các loại cổ phiếu còn lại

THỐNG KÊ MÔ TẢ

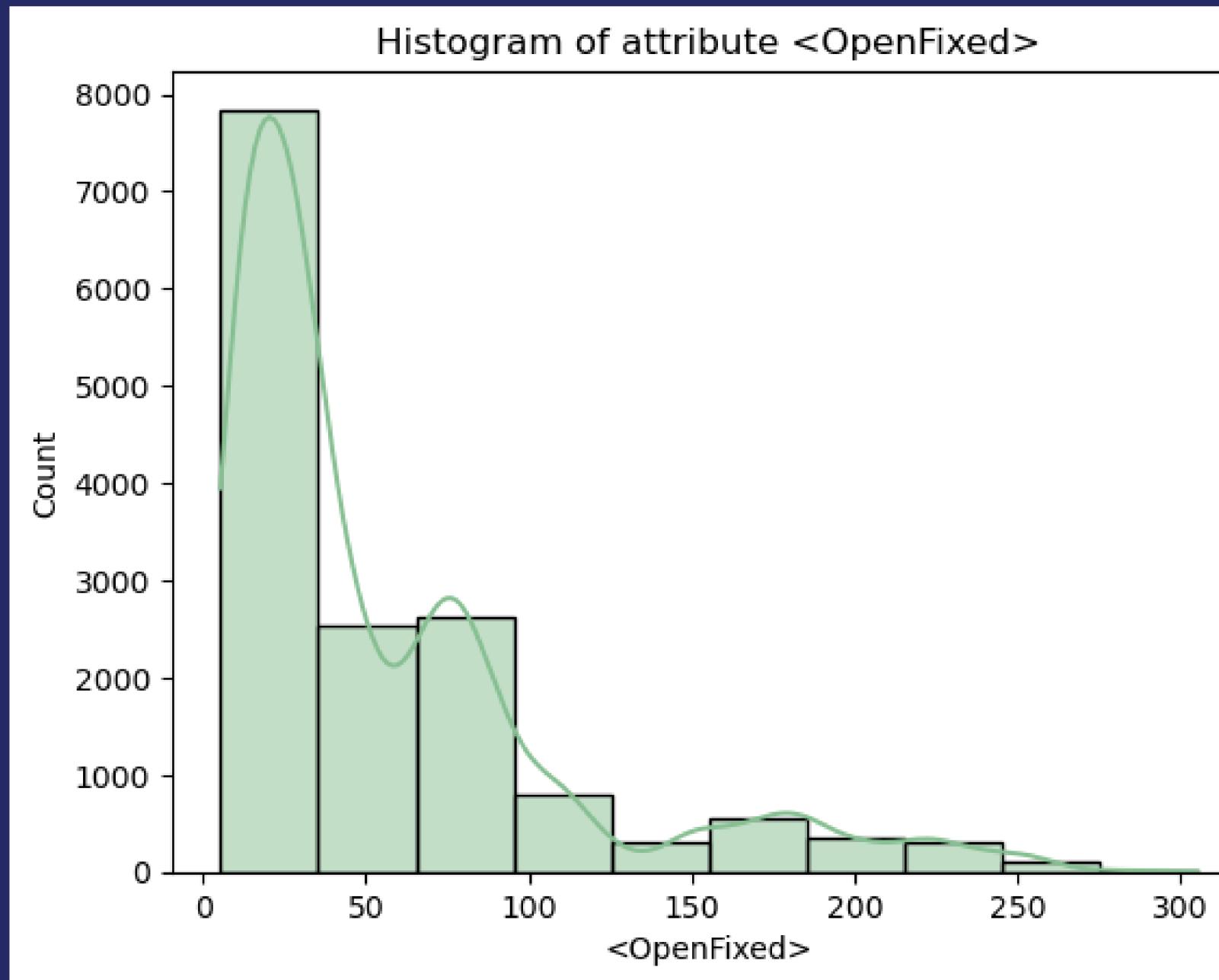


HISTOGRAM: BIỂU ĐỒ TẦN XUẤT

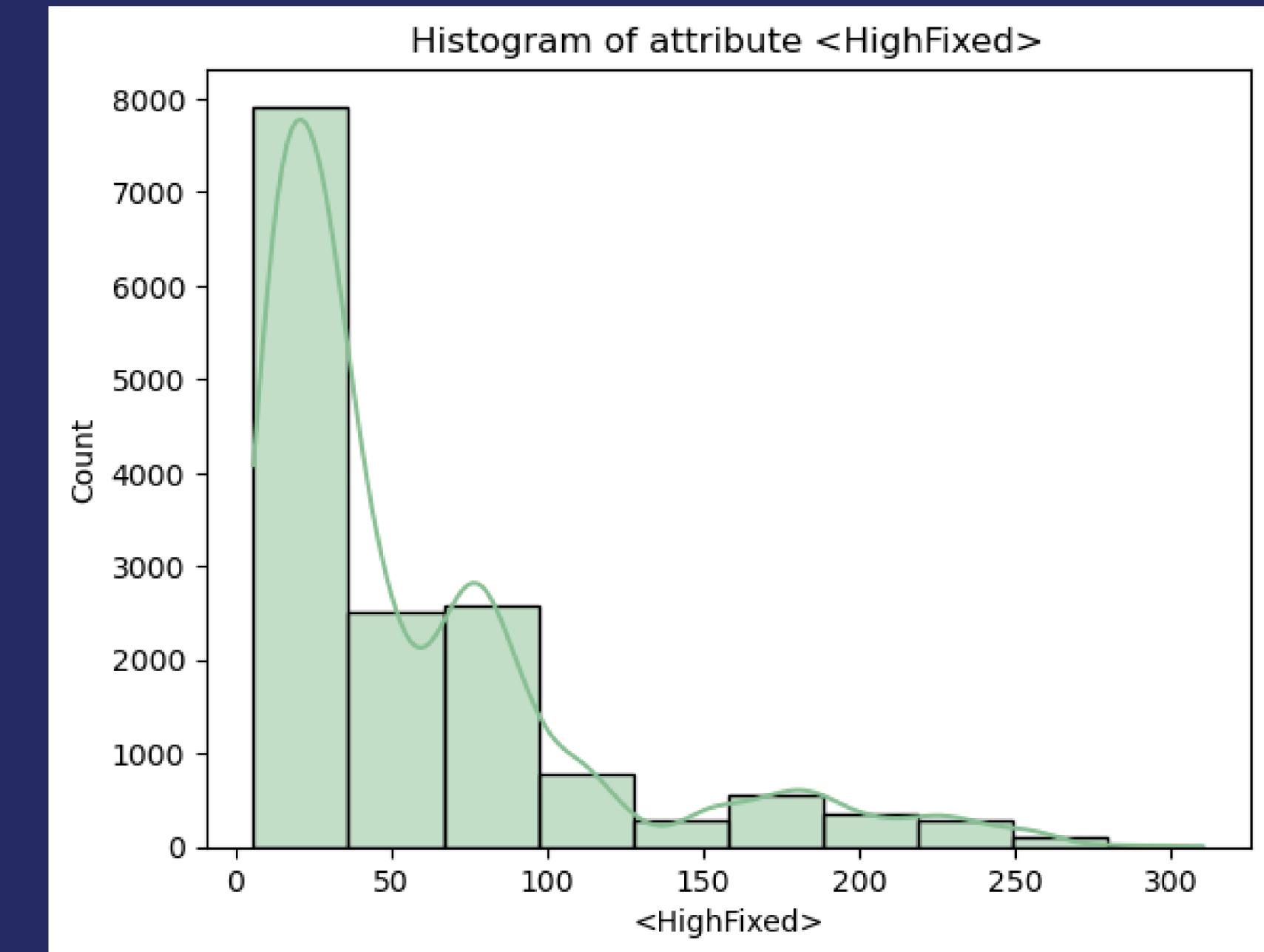


HISTOGRAM: BIỂU ĐỒ TẦN SUẤT

- OPENFIXED



- CLOSEFIXED

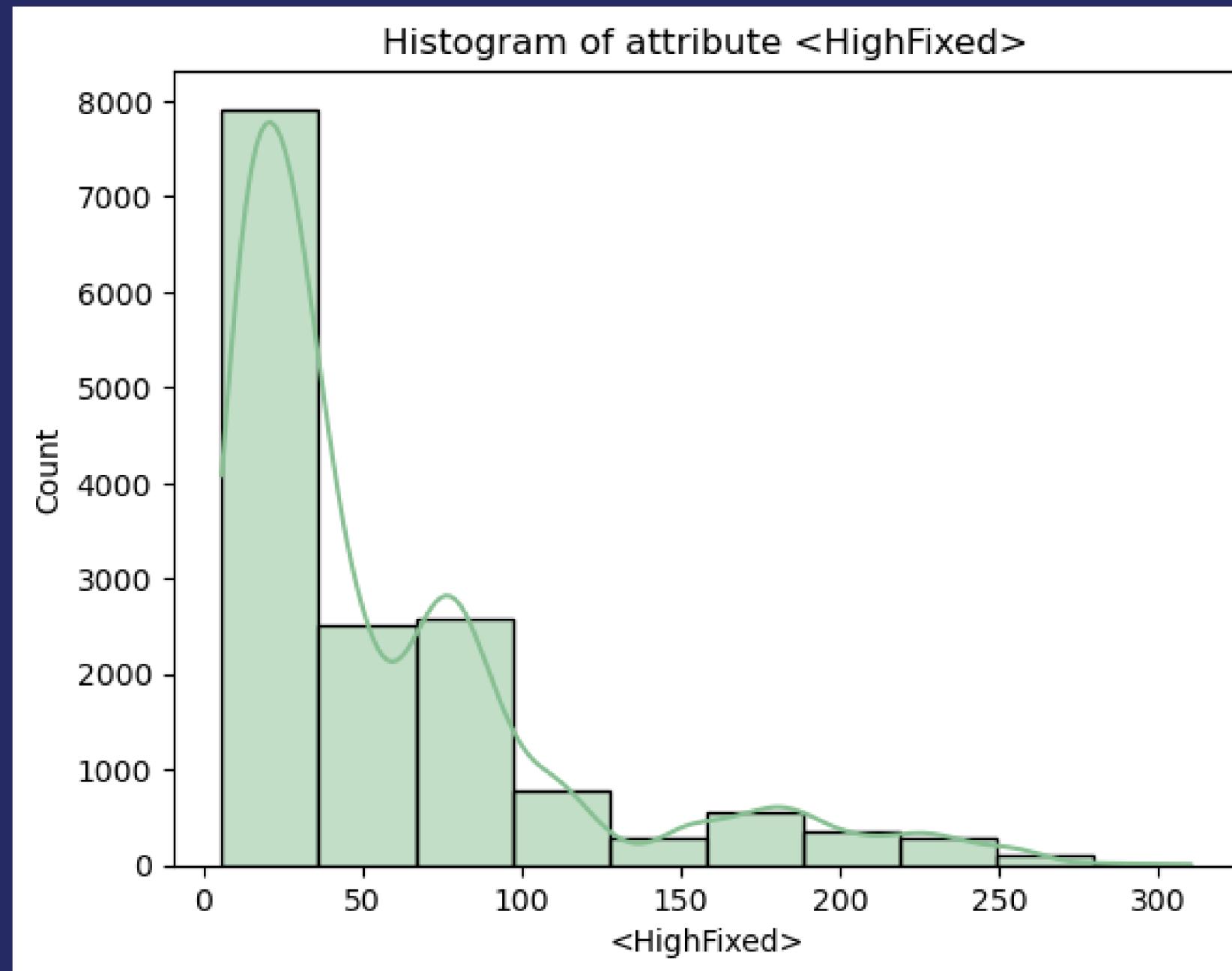


- Nhận xét: Đồ thị có phân phối không chuẩn, và lệch sang trái

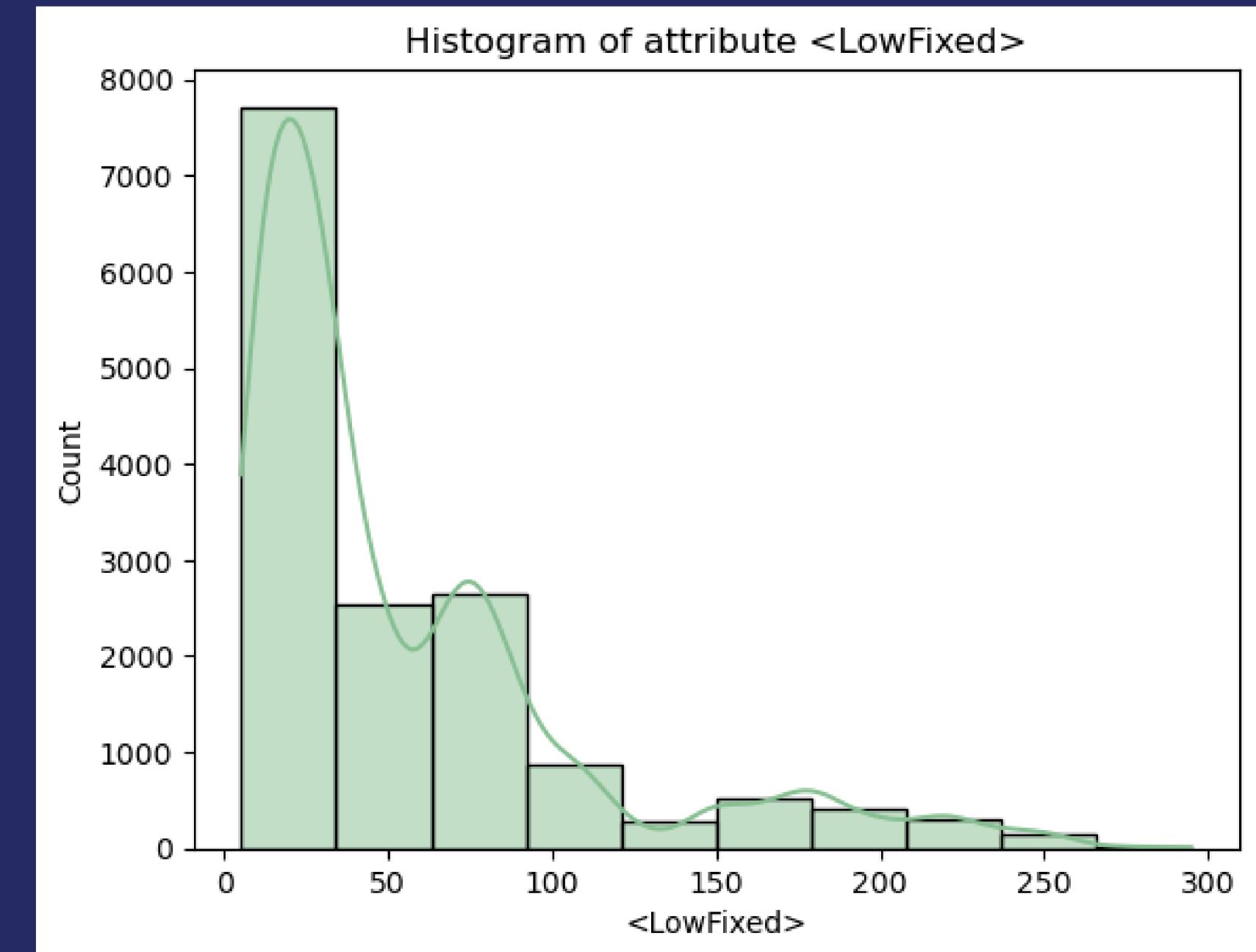
- Đối với giá mở mỗi ngày của tất cả cổ phiếu thường từ 10K đến 40K, cũng có một số ngày cổ phiếu có giá từ 50-100K.
- Tương tự với giá đóng và giá cao, giá thấp vì sự chênh lệnh về các loại giá này của từng cổ phiếu trong một ngày thì không đáng kể

HISTOGRAM: BIỂU ĐỒ TẦN XUẤT

- HIGHFIXED

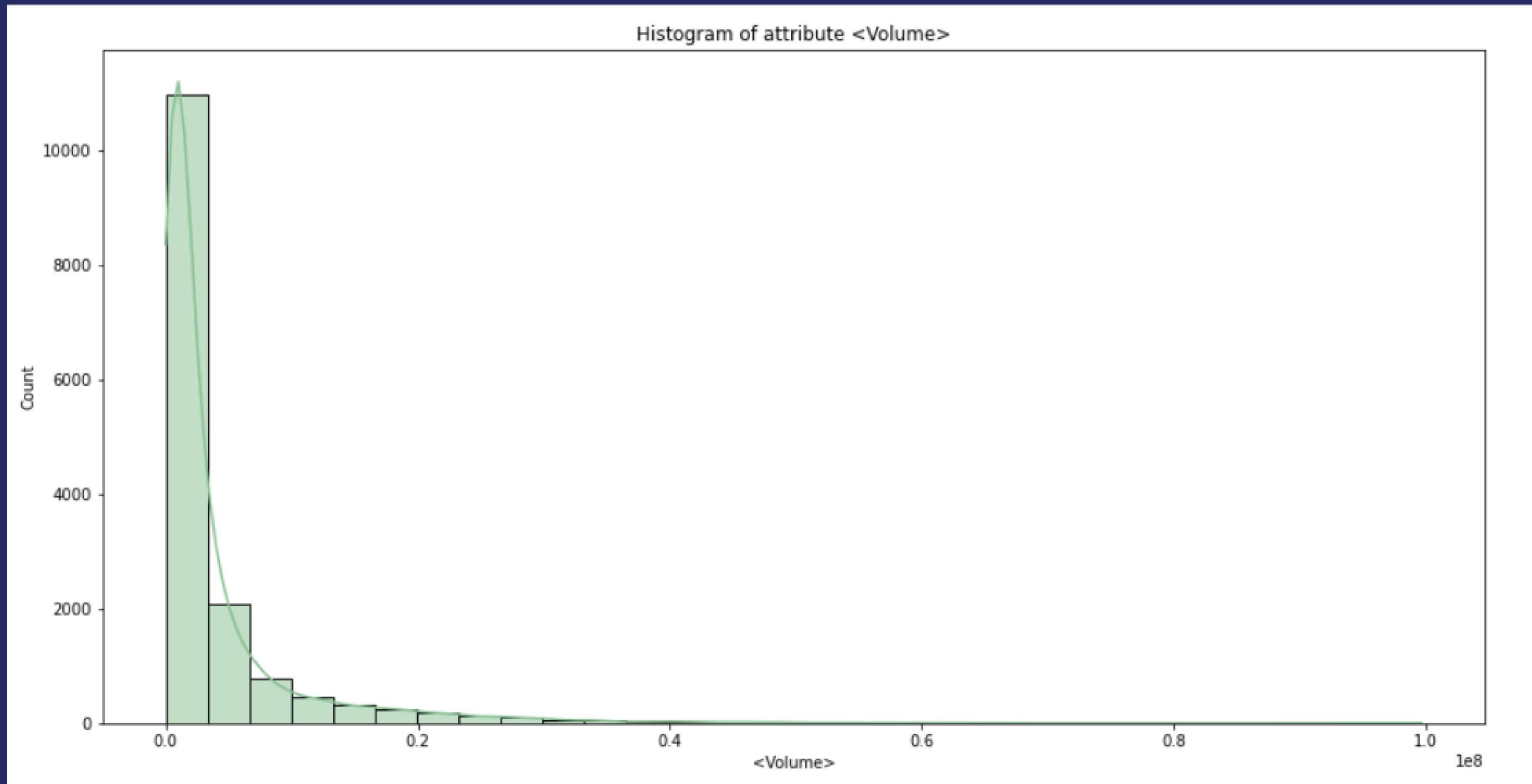


- LOWFIXED



HISTOGRAM: BIỂU ĐỒ TẦN XUẤT

- VOLUMN

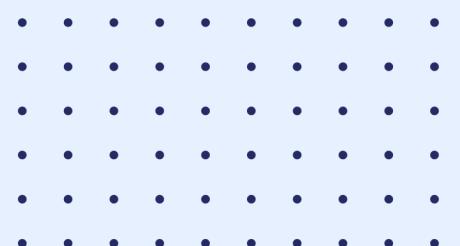


- **Nhận xét:** Đồ thị có phân phối không chuẩn và lệch sang trái
 - Khối lượng giao dịch cổ phiếu trong ngày thường ở mức dưới 40000000 giao dịch.
 - Tuy nhiên, khối lượng giao dịch có thể lên đến 100000000 giao dịch trong một ngày nhưng tần suất rất thấp



PHÂN TÍCH DỮ LIỆU

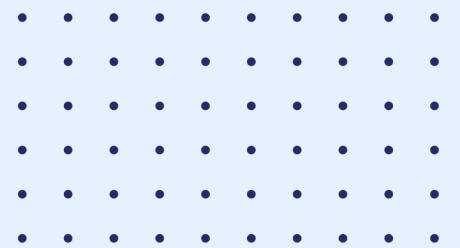
ĐƠN GIẢN



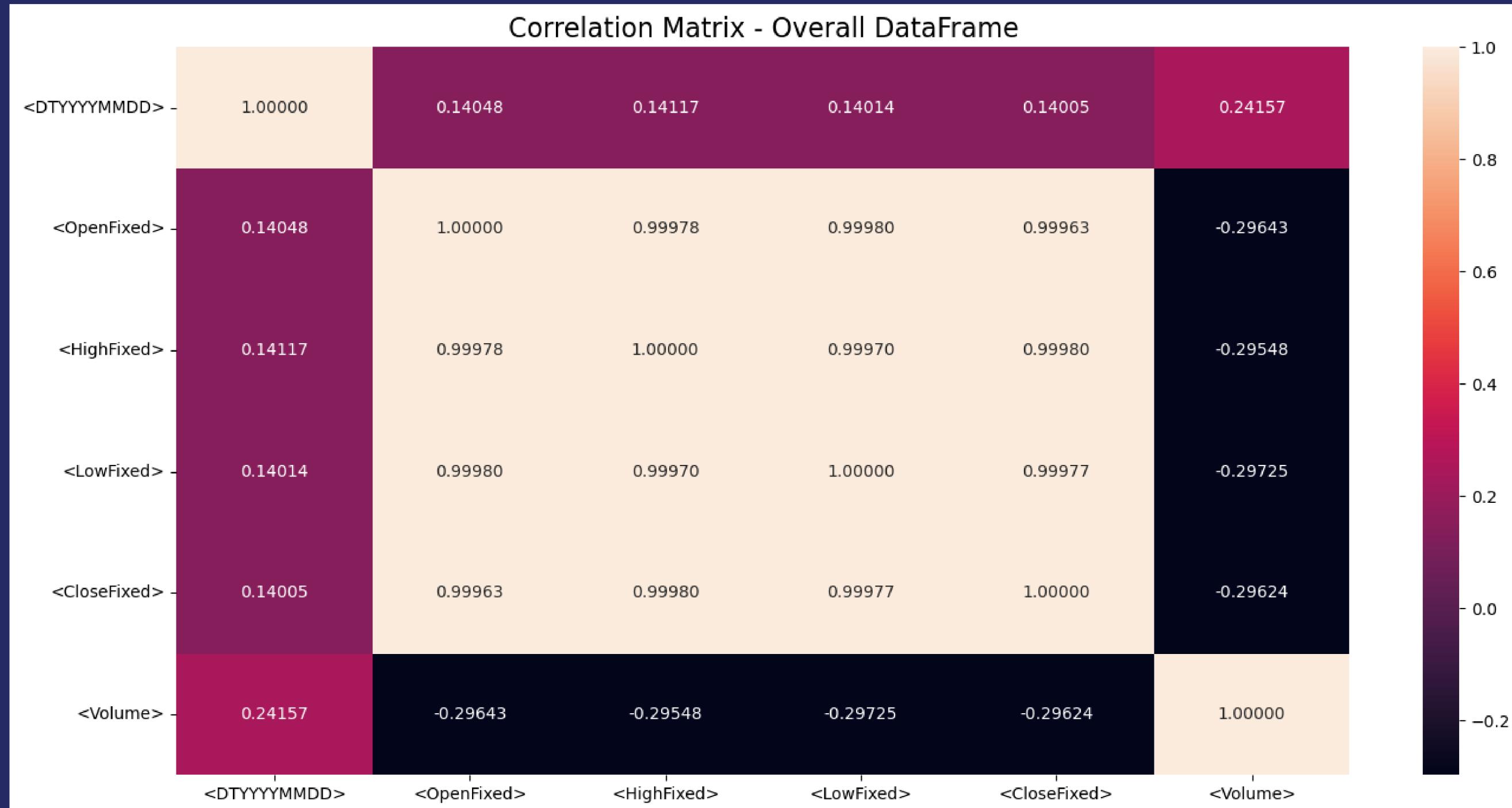
PHÂN TÍCH DỮ LIỆU



HEATMAP
MỐI QUAN HỆ TƯƠNG QUAN



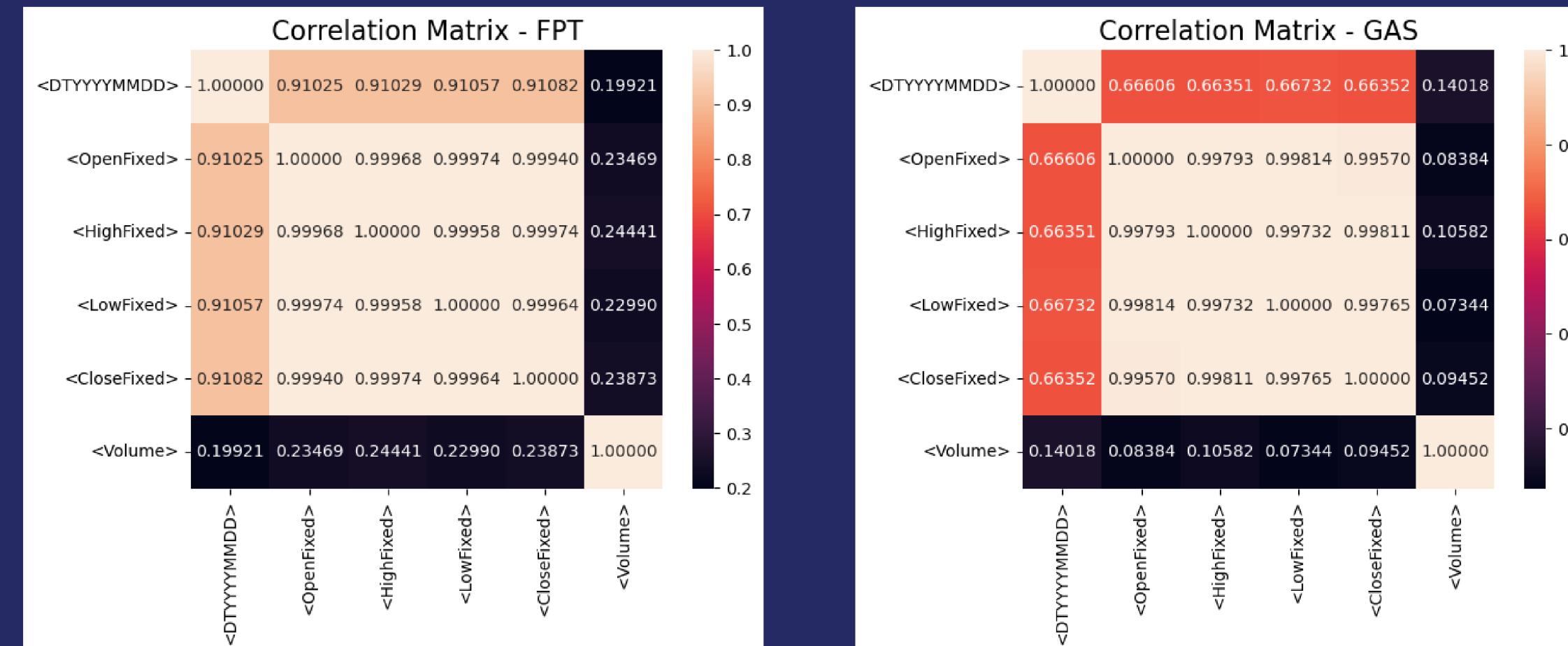
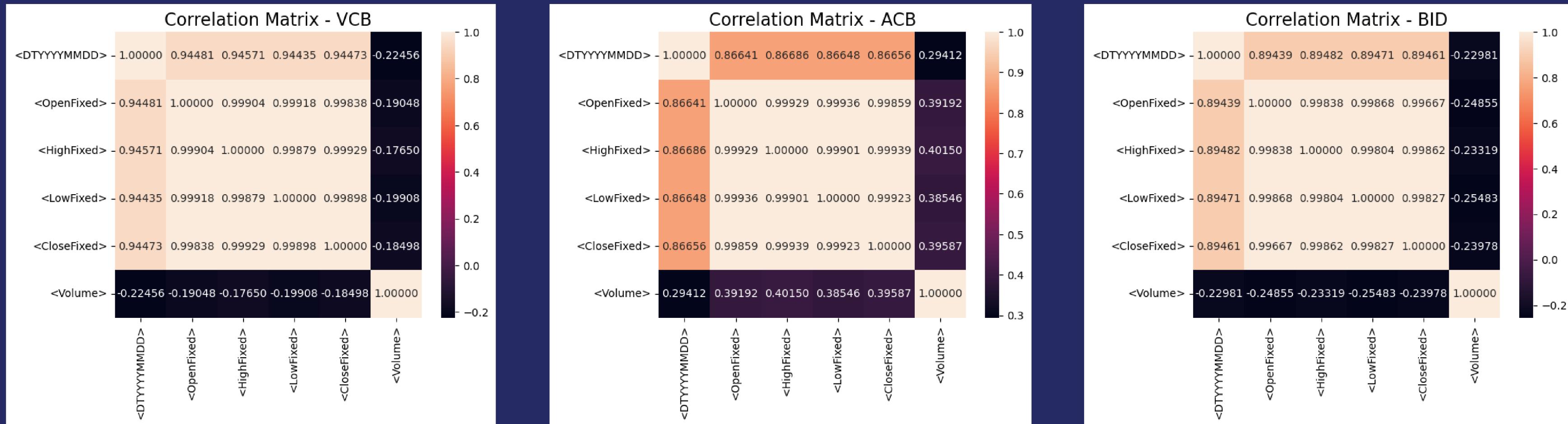
QUAN SÁT MỐI TƯƠNG QUAN GIỮA CÁC BIẾN TRONG TOÀN BỘ DỮ LIỆU



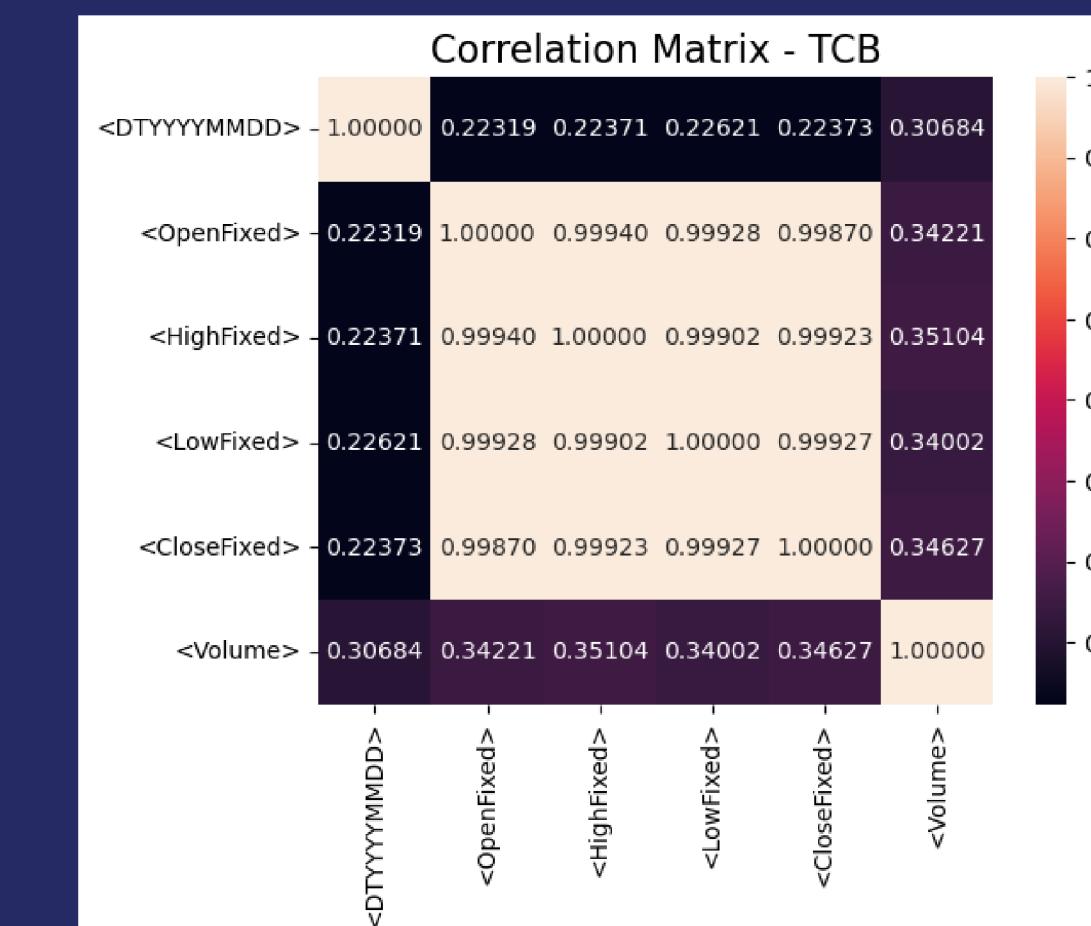
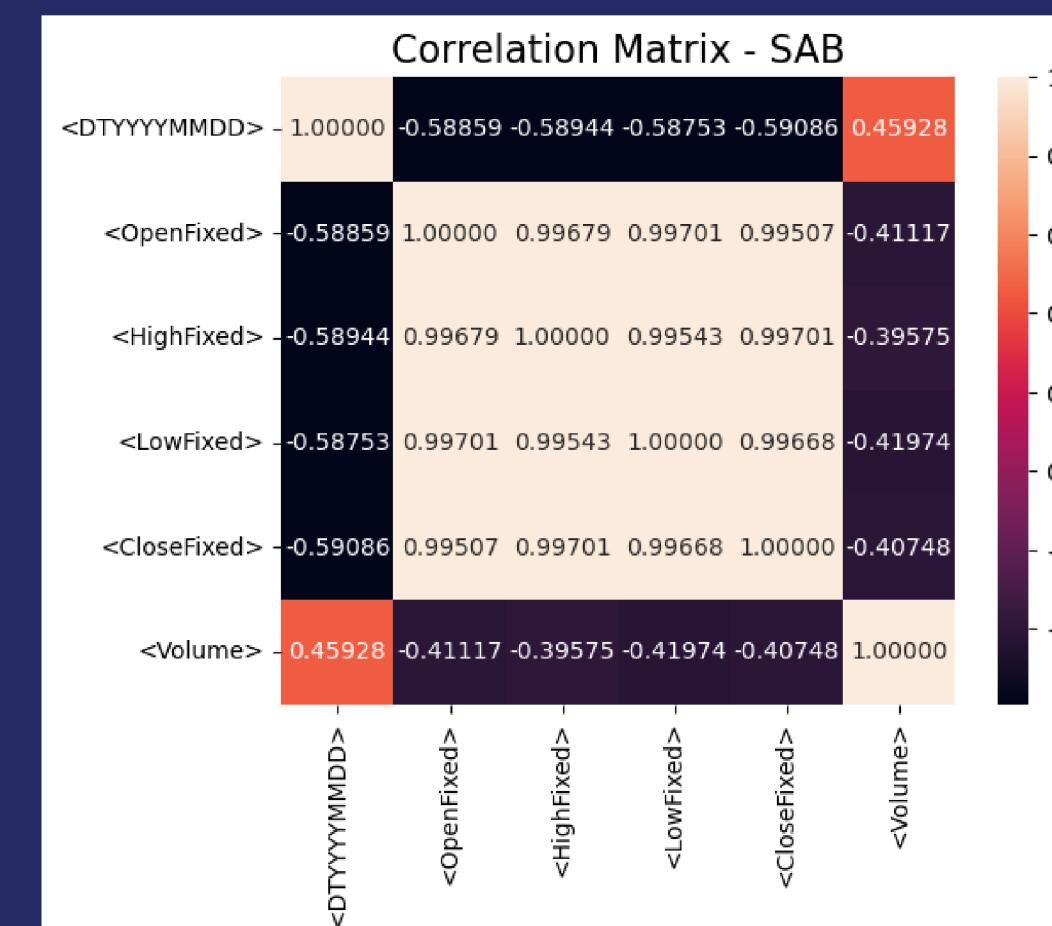
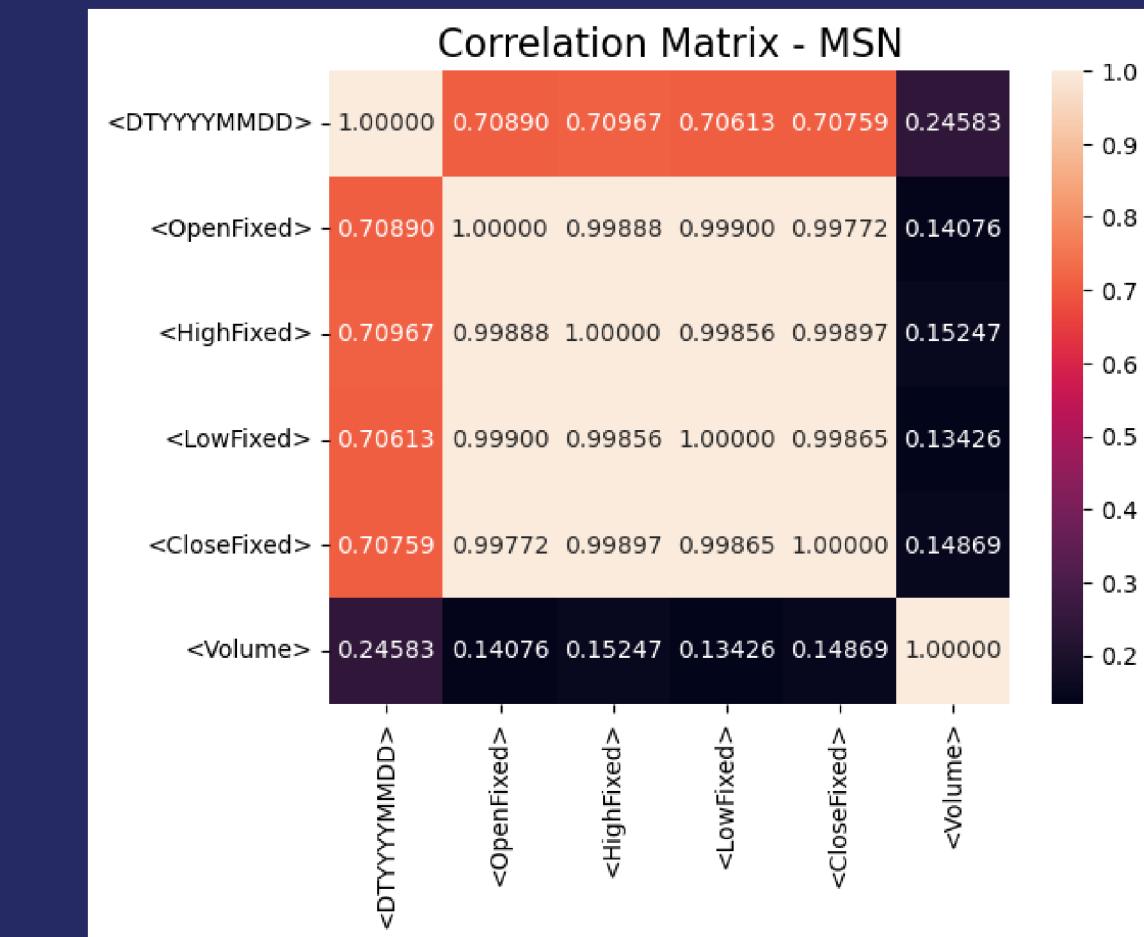
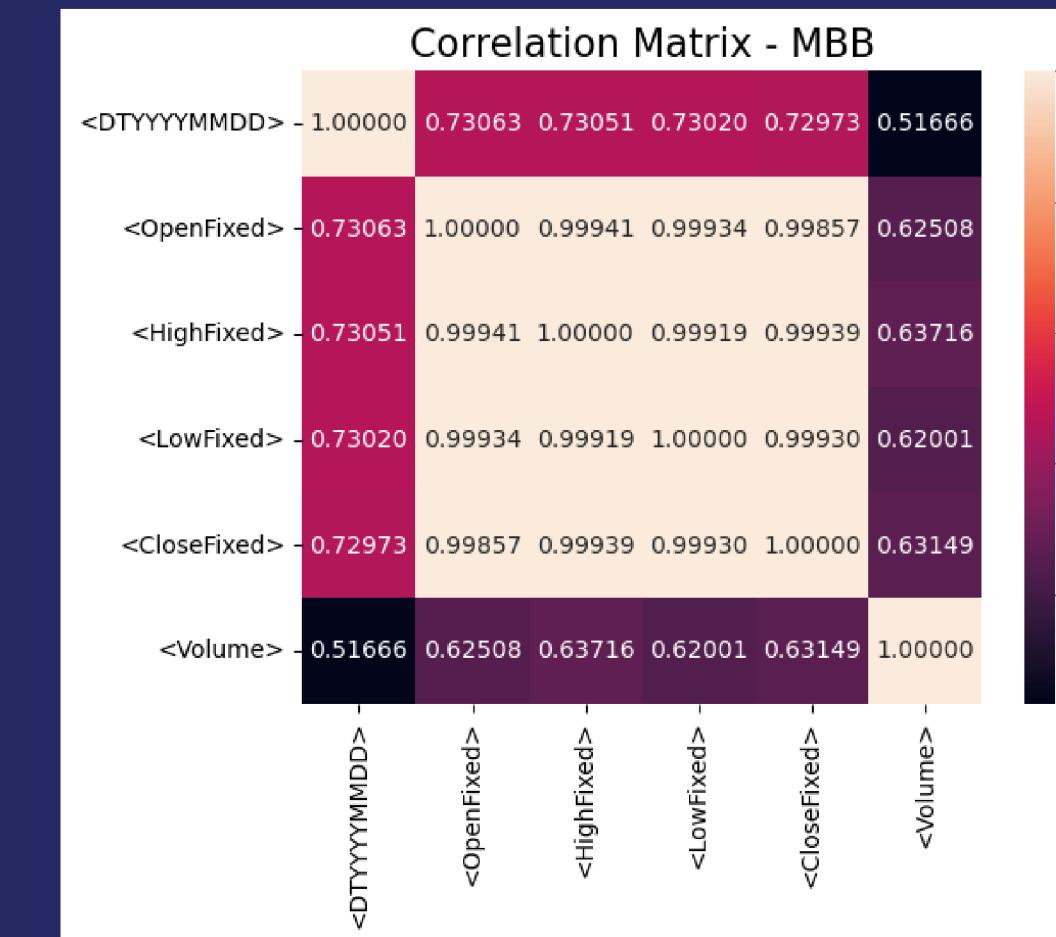
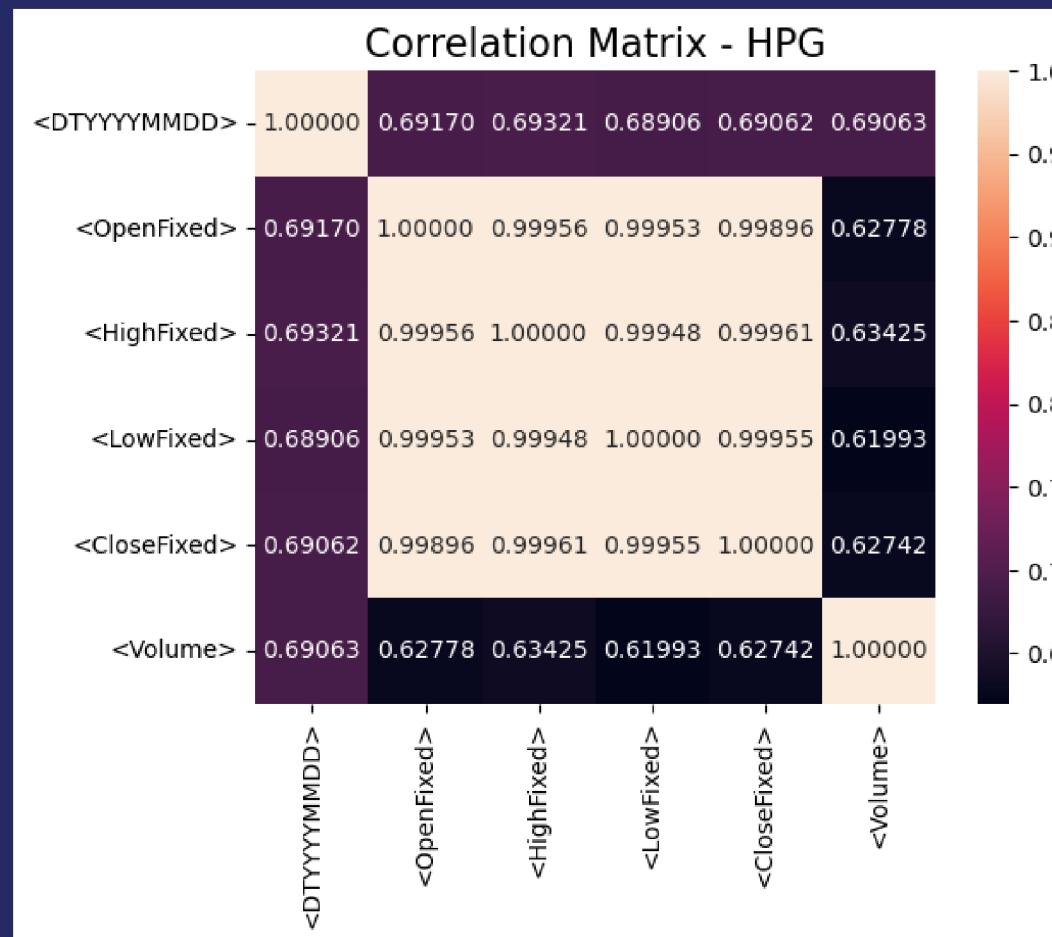
Nhận xét

- Nhìn vào biểu đồ ta thấy **OpenFixed**, **HighFixed**, **LowFixed**, **CloseFixed** có sự tương quan dương cao
- Volume nhìn chung có sự tương quan âm đối với các loại giá nhưng không lớn

QUAN SÁT MỐI TƯƠNG QUAN GIỮA CÁC BIẾN ĐỔI VỚI TỪNG MÃ CỔ PHIẾU



QUAN SÁT MỐI TƯƠNG QUAN GIỮA CÁC BIẾN ĐỔI VỚI TỪNG MÃ CỔ PHIẾU



QUAN SÁT MỐI TƯƠNG QUAN GIỮA CÁC BIẾN ĐỔI VỚI TỪNG MÃ CỔ PHIẾU

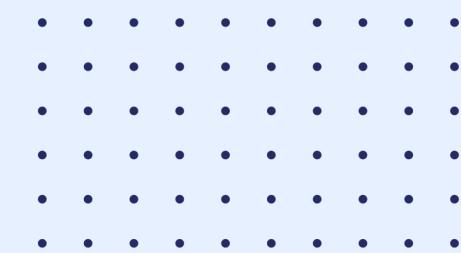
Nhận xét

- Đối với một số cổ phiếu cũng có sự tương quan cao giữa ngày và các loại giá. Tuy nhiên không phải toàn bộ loại cổ phiếu đều có mối quan tương quan cao giữa giá và thời gian
- Độ tương quan giữa ngày và giá đóng của từng cổ phiếu
 - VCB: 0.94473
 - ACB: 0.86633
 - BID: 0.89405
 - FPT: 0.91053
 - GAS: 0.66432
 - HPG: 0.69175
 - MBB: 0.73126
 - MSN: 0.70993
 - SAB: 0.59082
 - TCB: 0.22578

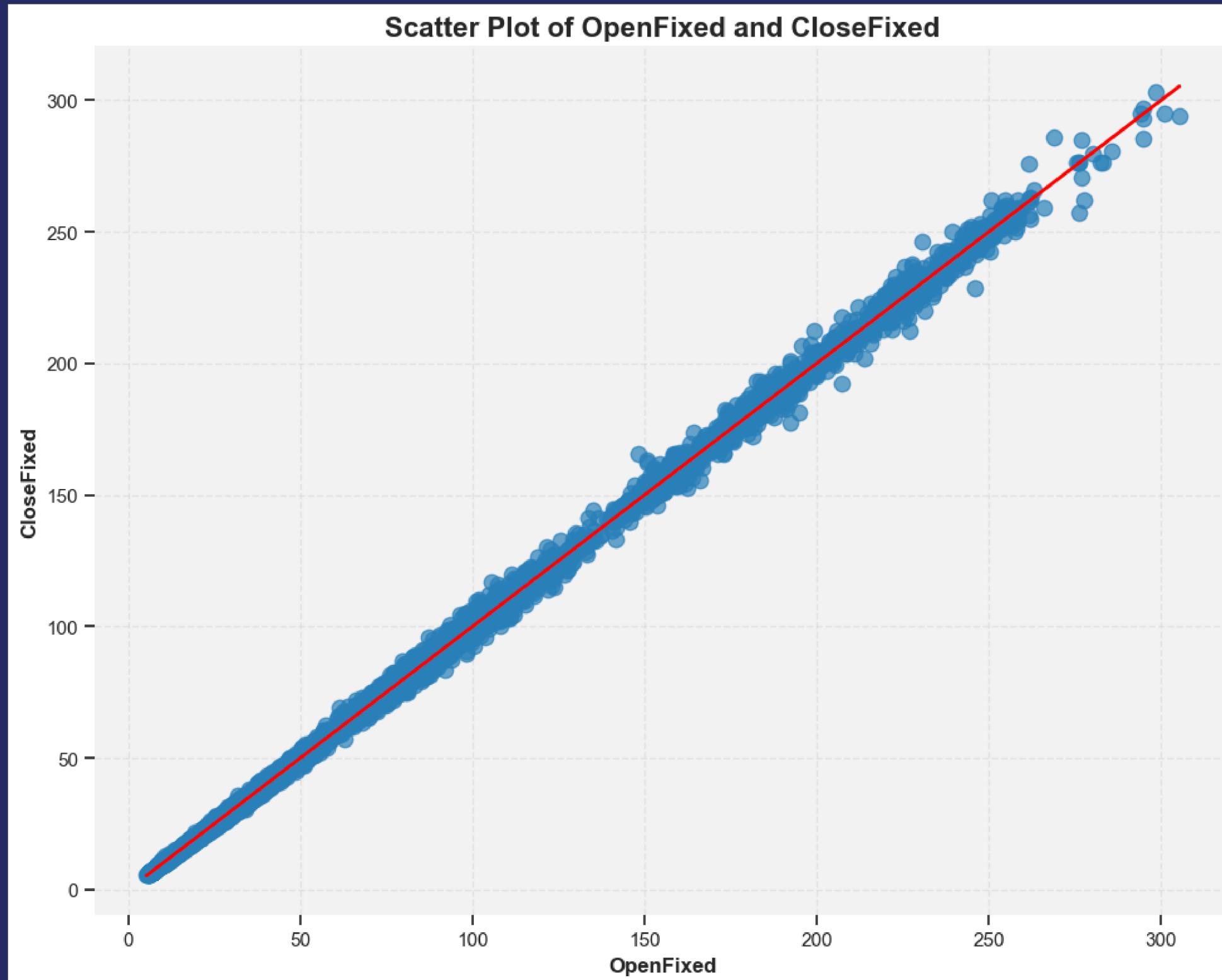
PHÂN TÍCH DỮ LIỆU



**SCATTERPLOT
MỐI TƯƠNG QUAN GIỮA HAI BIẾN**



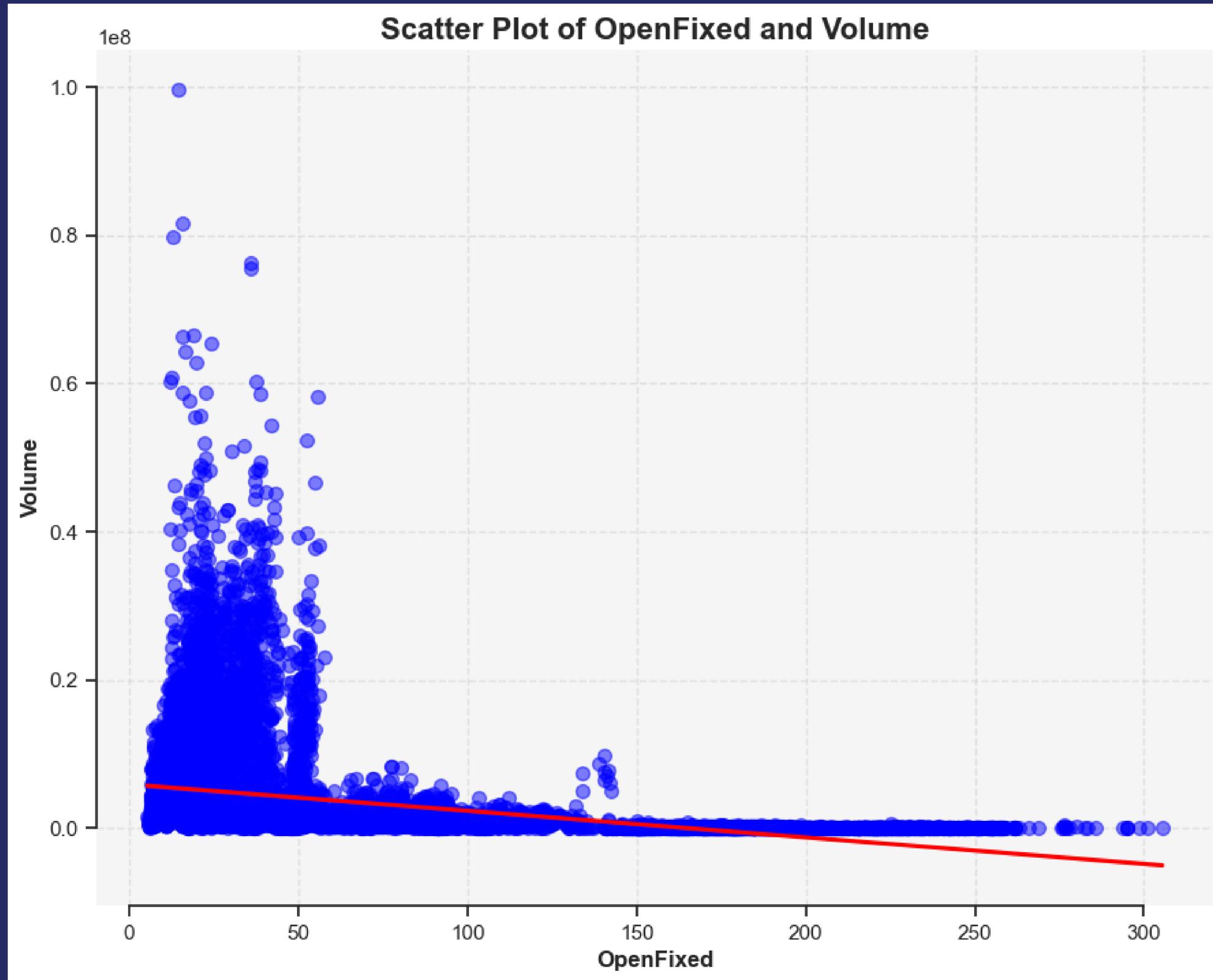
MỐI QUAN HỆ CỦA GIÁ ĐÓNG VÀ GIÁ MỞ



Nhận xét

- Giá đóng và giá mở tương quan đồng biến
- Trong thực tế, giá mở và giá đóng trong ngày không thực sự ảnh hưởng nhau, vì giá đóng phụ thuộc vào các yếu tố thị trường khác và có sự chênh lệch với giá mở hoặc không, tuy nhiên giá mở của ngày hôm nay có thể được xem xét bởi giá đóng của ngày hôm trước.

MỐI QUAN HỆ GIỮA GIÁ MỞ VÀ KHỐI LƯỢNG GIAO DỊCH

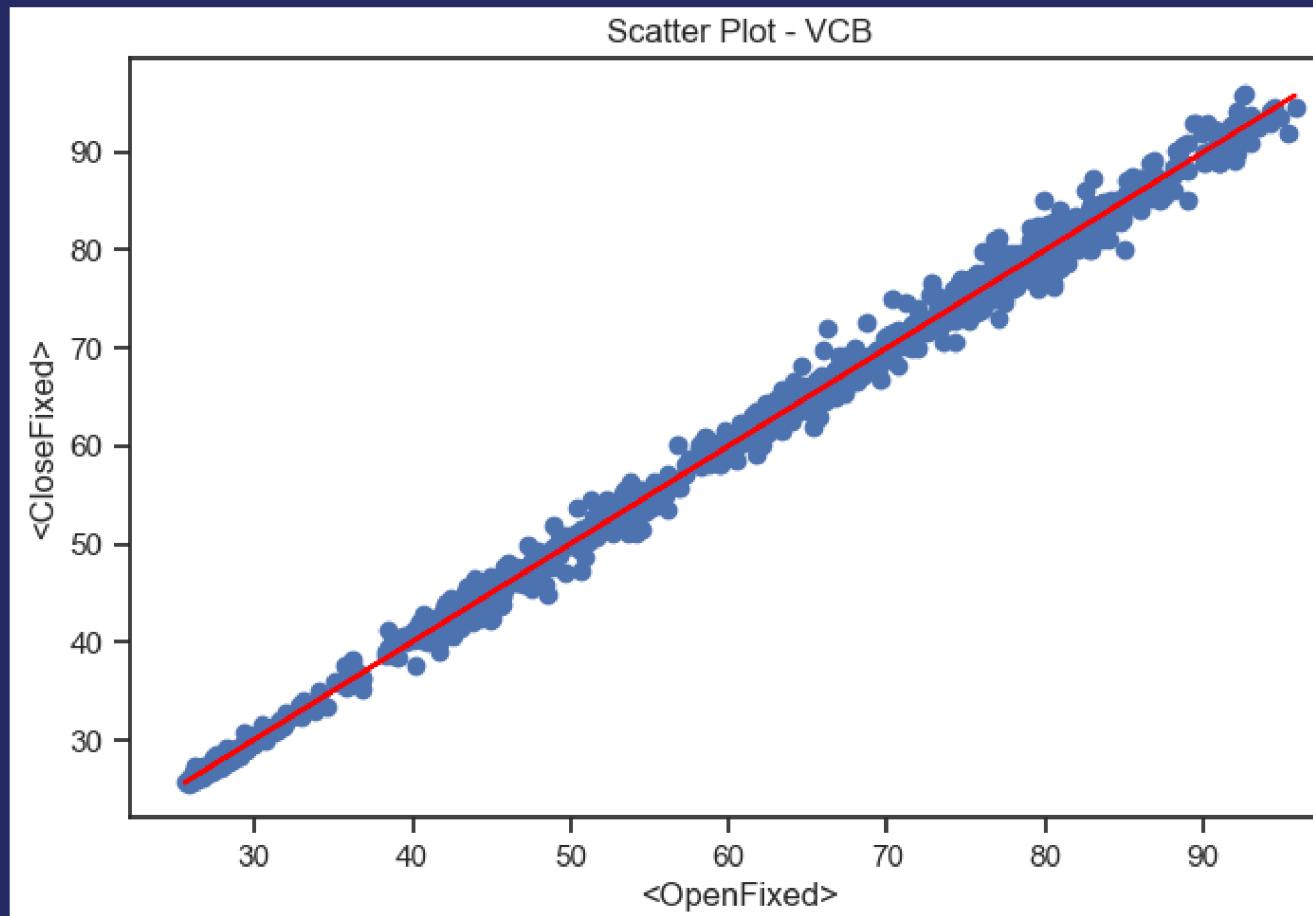


Nhận xét

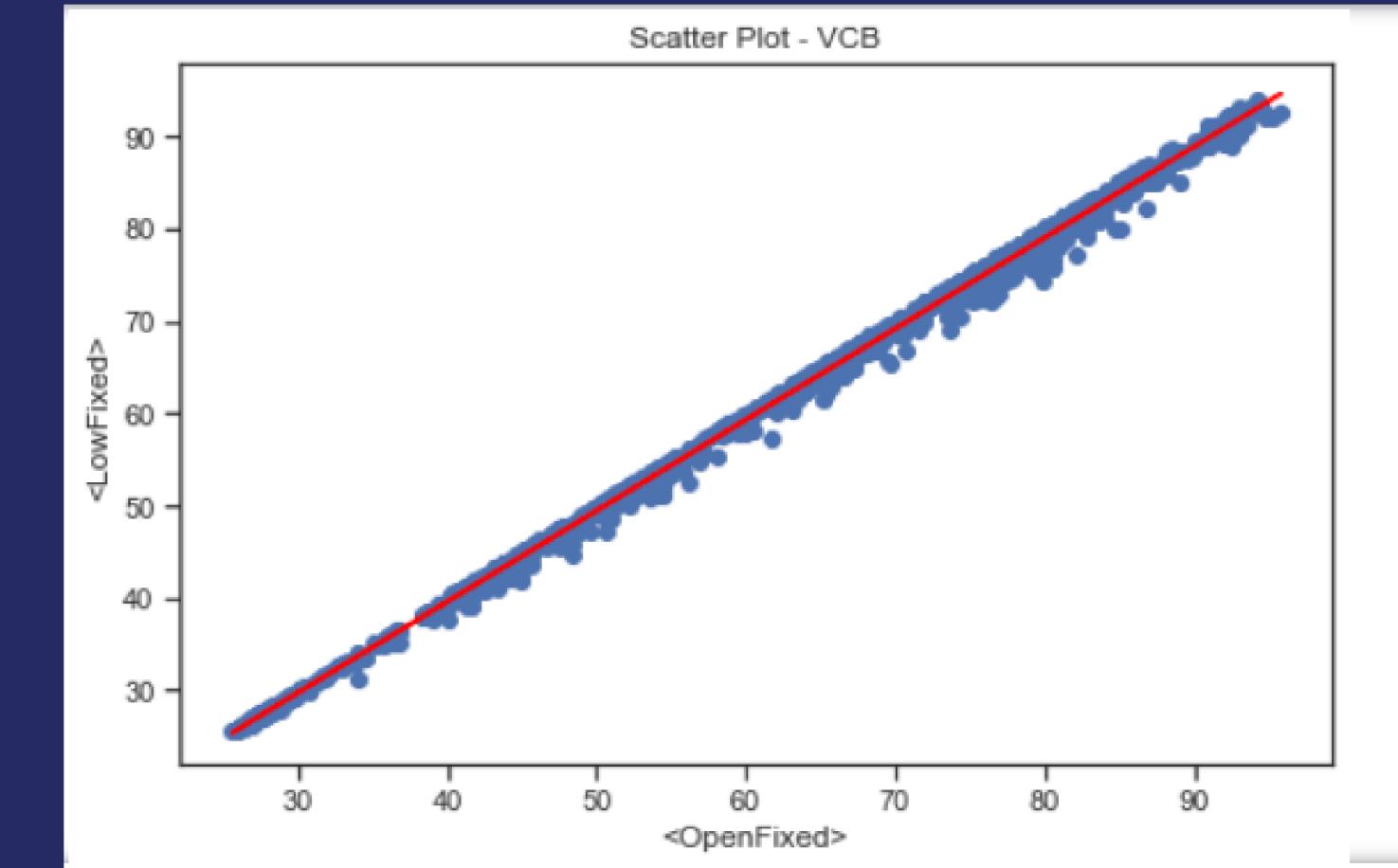
- Giá mở và khối lượng giao dịch có xu hướng tương quan nghịch biến
- Thực tế thì, giá mở có thể hấp dẫn nhà đầu tư tham gia giao dịch và tăng khối lượng giao dịch. Nếu giá mở cao hơn hoặc thấp hơn so với giá đóng của phiên trước đó, điều này có thể tạo ra sự tăng lên hoặc giảm đi trong khối lượng giao dịch.

MỐI QUAN HỆ CỦA GIÁ MỞ VỚI CÁC LOẠI GIÁ TỪNG CỔ PHIẾU - VCB

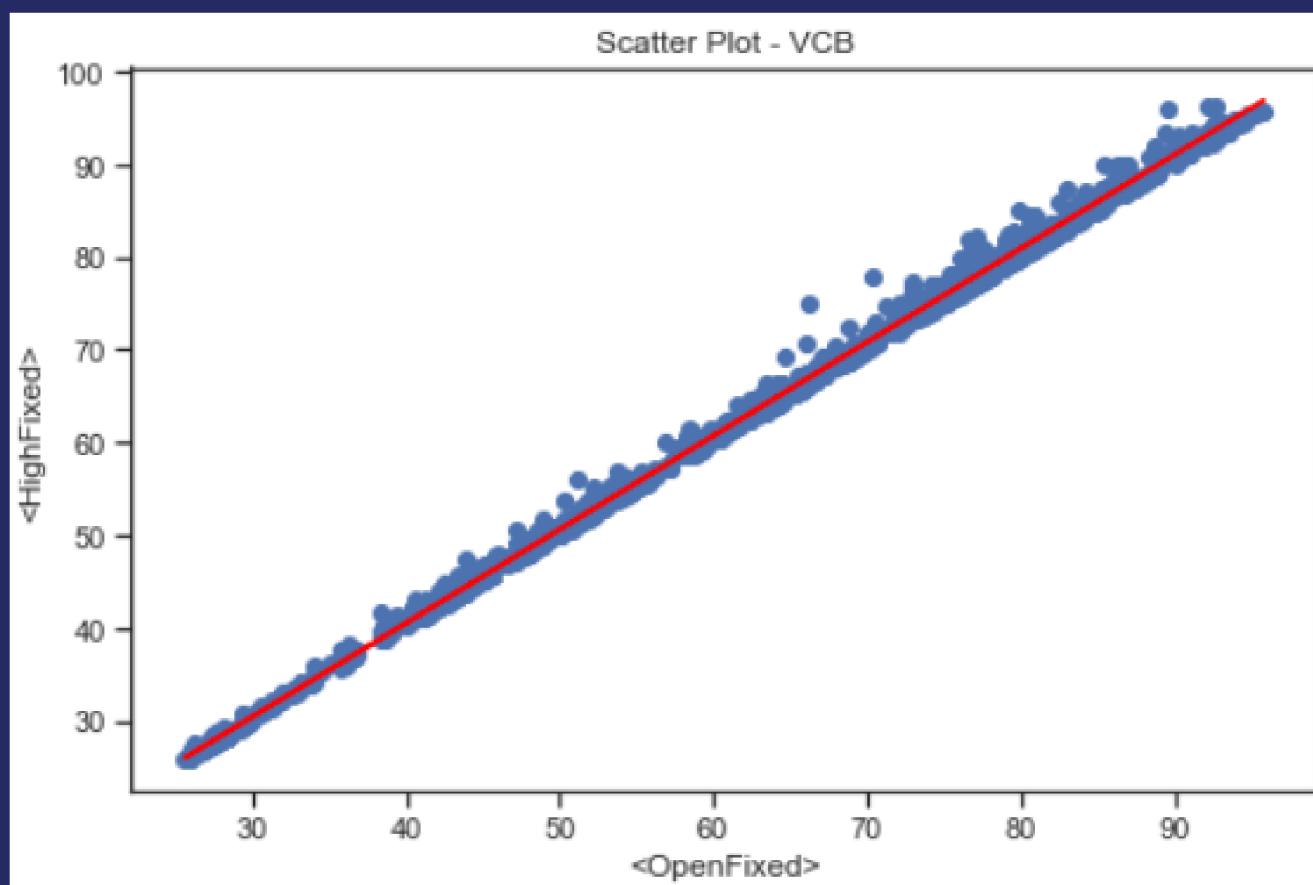
Giá mở -
Giá đóng



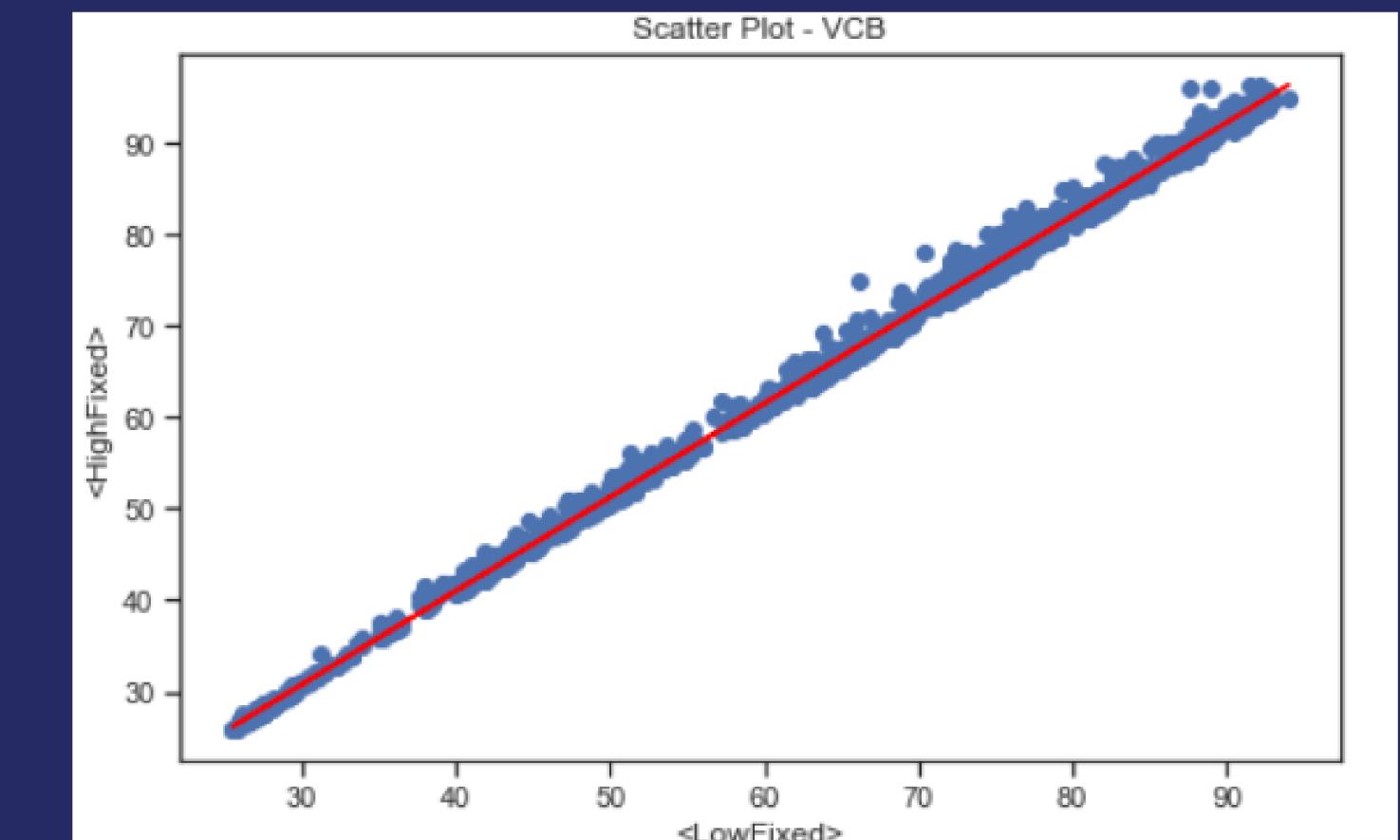
Giá mở -
Giá thấp



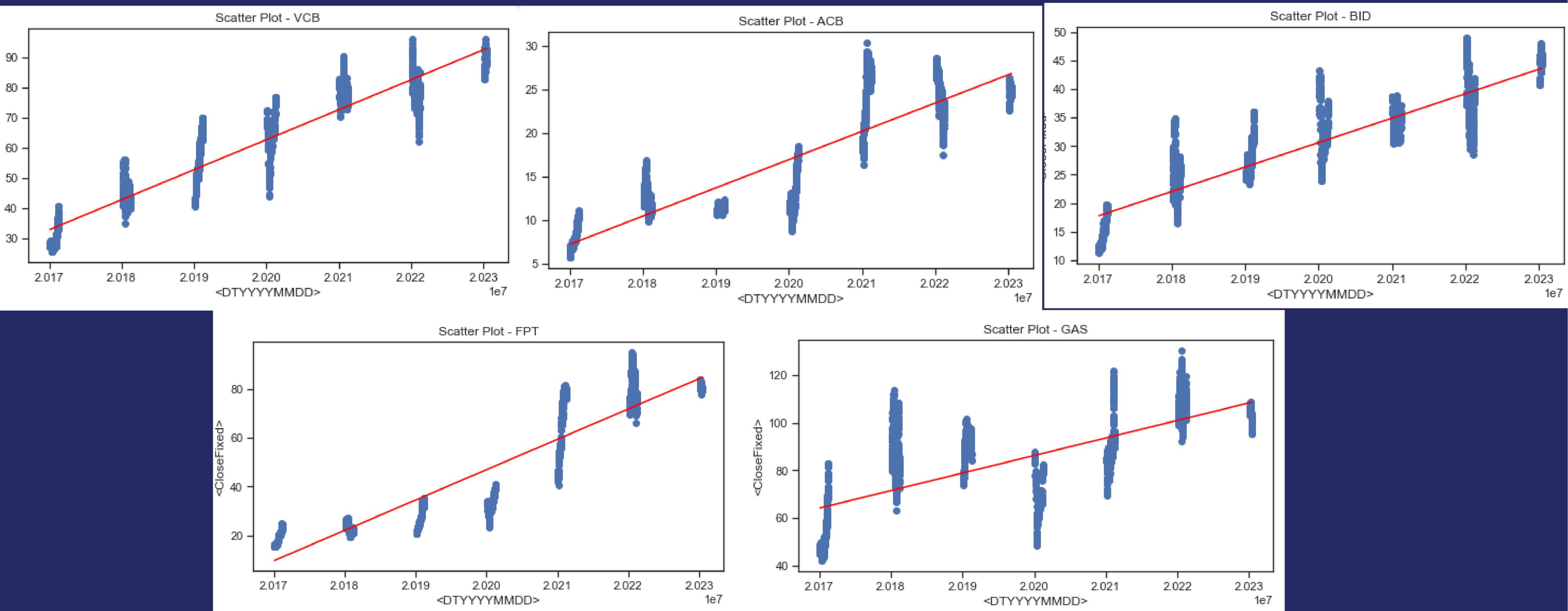
Giá mở -
Giá cao



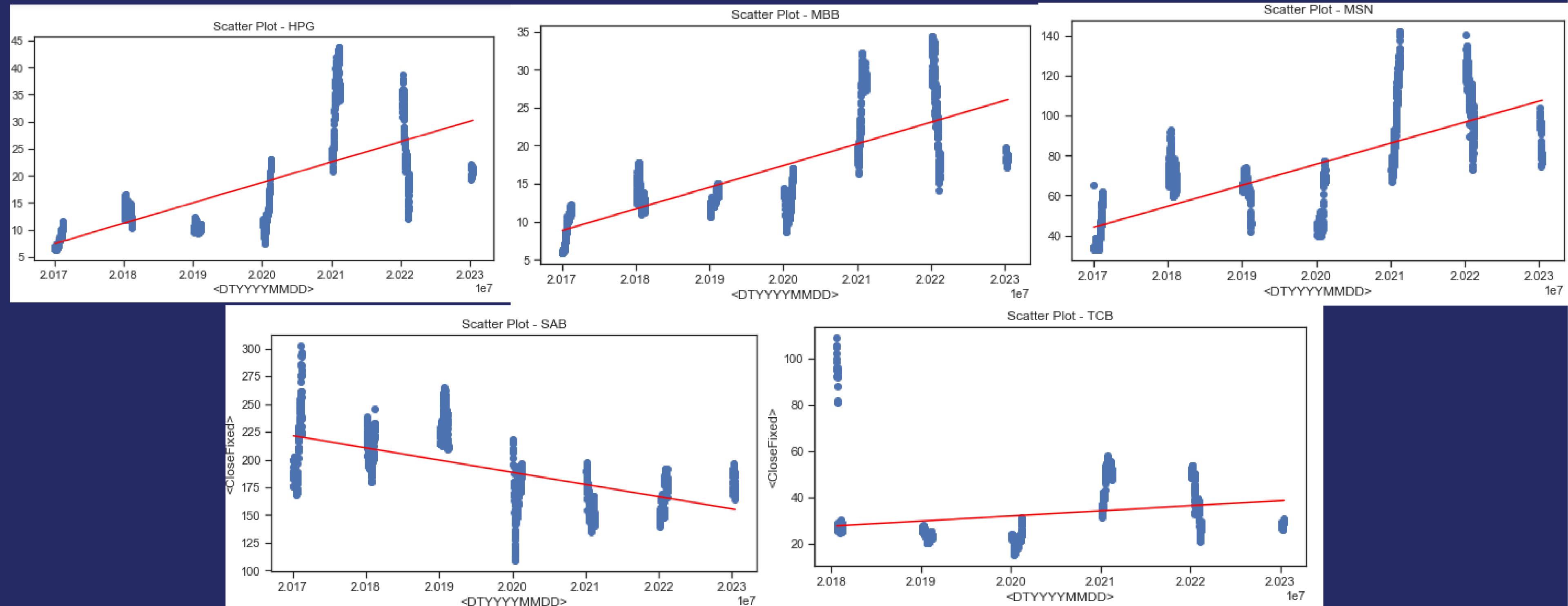
Giá thấp-
Giá cao



MỐI QUAN HỆ CỦA GIÁ ĐÓNG VÀ NGÀY TỪNG CỔ PHIẾU



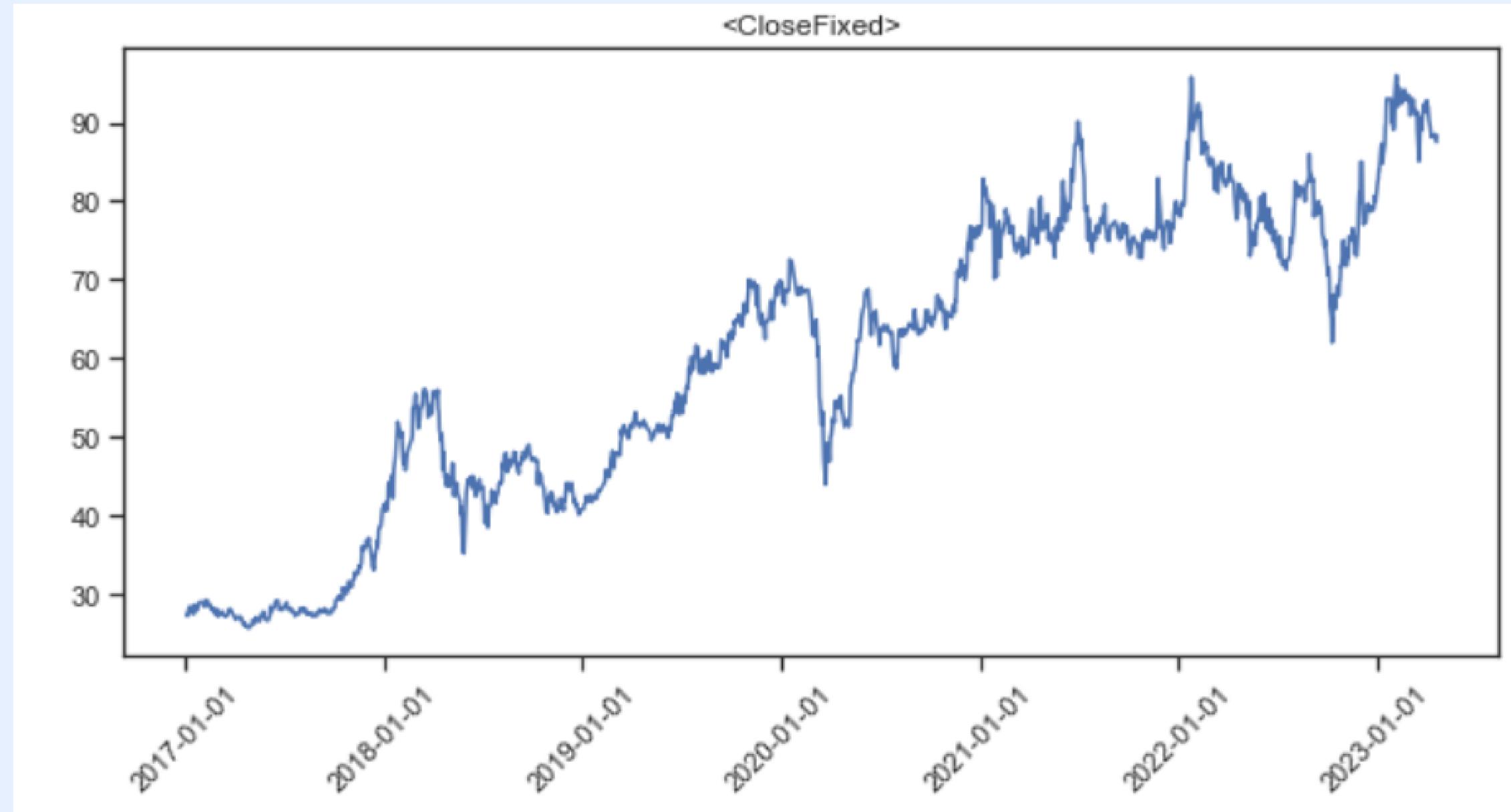
MỐI QUAN HỆ CỦA GIÁ ĐÓNG VÀ NGÀY TỪNG CỔ PHIẾU



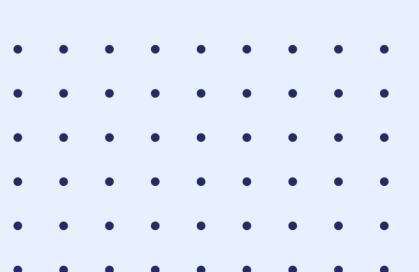
Một số cổ phiếu có sự tương quan đồng biến cao giữa giá mở và ngày như: VCB, BID, FPT.
Cổ phiếu có sự giảm giá mở theo thời gian là SAB

PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

Dựa vào độ tương quan cao giữa giá đóng và ngày của tập dữ liệu của cổ phiếu VCB, nhóm thực hiện xây dựng mô hình chuỗi thời gian trên tập dữ liệu này



BIỂU ĐỒ BIẾN ĐỘNG GIÁ ĐÓNG CỦA VCB THEO THỜI GIAN



PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

Kiểm tra các yếu tố mùa vụ

Trong một số chuỗi thời gian thường xuất hiện yếu tố mùa vụ. Việc tìm ra chu kỳ và qui luật mùa vụ sẽ giúp cho mô hình dự báo chuẩn xác hơn.

Nhận xét:

- Nhìn chung về dữ liệu không có chu kỳ theo thời gian trong năm, hay theo xu hướng. Vì vậy dữ liệu không phụ thuộc vào mùa vụ

• • • • • • •
• • • • • • •
• • • • • • •
• • • • • • •
• • • • • • •
• • • • • • •



PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

MÔ HÌNH HỒI QUY ARIMA

Mô hình Autoregressive Integrated Moving Average (ARIMA) là một mô hình thống kê sử dụng để phân tích và dự đoán các chuỗi dữ liệu có xu hướng và mô hình ngẫu nhiên. Nó kết hợp các thành phần chính bao gồm tự hồi quy tự động (AR), và trung bình trượt (MA). ARIMA dựa vào thông tin quá khứ để dự đoán giá trị hiện tại và có ứng dụng rộng rãi trong dự báo tài chính, tiêu thụ năng lượng và phân tích chuỗi thời gian.

Phương trình hồi qui ARIMA(p, d, q) có thể được biểu diễn dưới dạng:

$$\Delta x_t = \phi_1 \Delta x_{t-1} + \phi_2 \Delta x_{t-2} + \dots + \phi_p \Delta x_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Trong đó Δx_t là giá trị sai phân bậc d và ϵ_t là các chuỗi nhiễu trắng.

- Auto regression (AR) là thành phần tự hồi qui bao gồm tlop hợp các độ trễ của biến hiện tại. Độ trễ bậc p chính là giá trị lùi về quá khứ p bước thời gian của chuỗi. Độ trễ dài hoặc ngắn trong quá trình AR phụ thuộc vào tham số trễ p .
- Moving average: Quá trình trung bình trượt được hiểu là quá trình dịch chuyển hoặc thay đổi giá trị trung bình của chuỗi theo thời gian

PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

MÔ HÌNH HỒI QUY ARIMA

- Chọn tập train, test để xây dựng mô hình hồi quy:
 - Train: Là tập dữ liệu giá đóng của VCB từ đầu đến ngày 31/3/2023
 - Test: Là tập dữ liệu giá đóng của VCB từ ngày 1/4/2023 đến ngày 21/4/2023

Ta có:

- Tập train

<CloseFixed>	
<DTYYYYMMDD>	
2017-01-03	27.3984
2017-01-04	27.2145
2017-01-05	27.3248
2017-01-06	27.8785
2017-01-09	28.4649
---	---
2023-03-27	91.0000
2023-03-28	91.8000
2023-03-29	92.3000
2023-03-30	92.0000
2023-03-31	91.4000

1558 rows x 1 columns

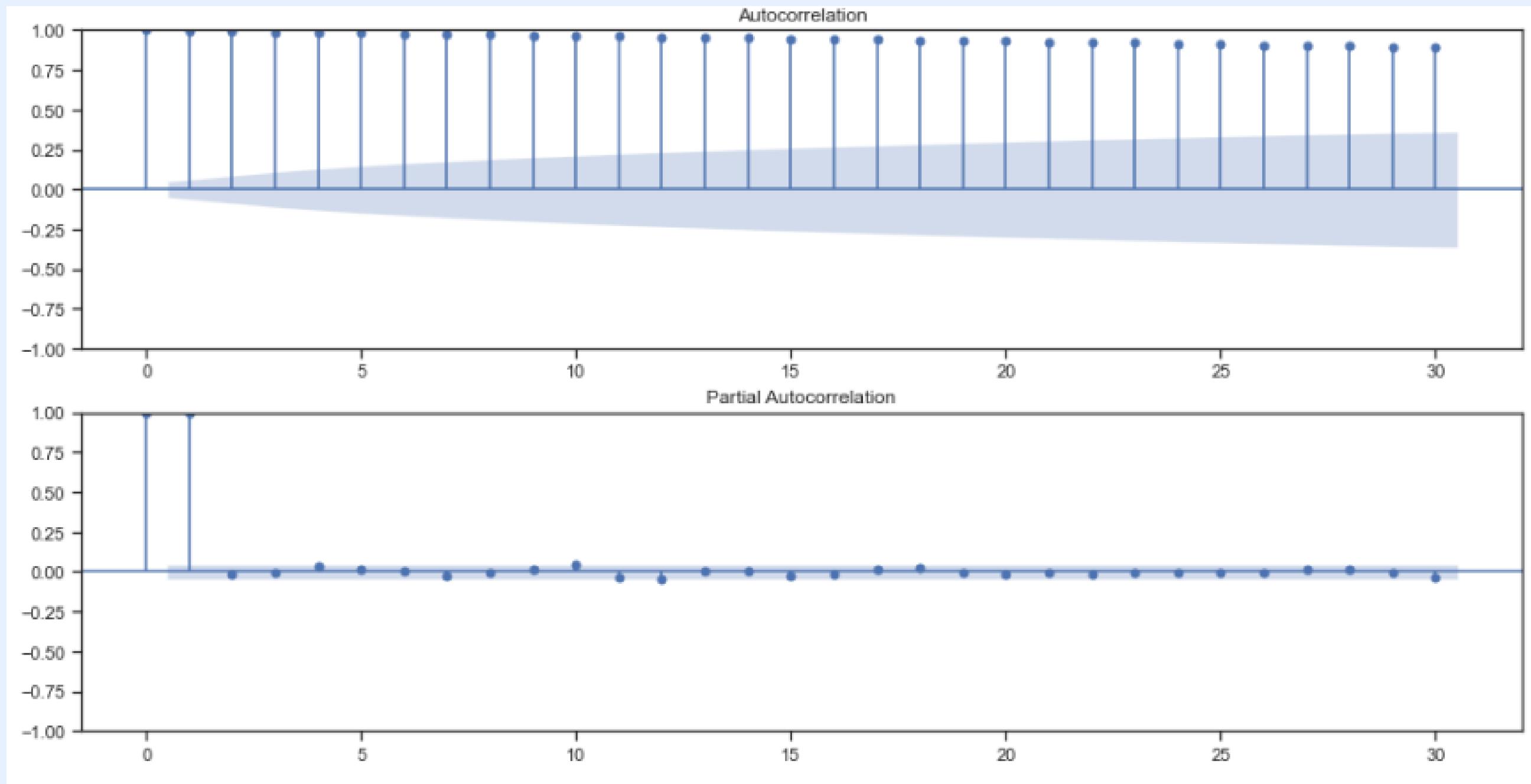
- Tập test

<CloseFixed>	
<DTYYYYMMDD>	
2023-04-03	92.8
2023-04-04	92.0
2023-04-05	92.1
2023-04-06	91.2
2023-04-07	90.4
2023-04-10	89.0
2023-04-11	88.2
2023-04-12	88.3
2023-04-13	88.2
2023-04-14	88.4
2023-04-17	88.4
2023-04-18	88.0
2023-04-19	88.5
2023-04-20	88.4
2023-04-21	87.5

PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

MÔ HÌNH HỒI QUY ARIMA

Vẽ biểu đồ ACF, PACF xác định các chỉ số p,d,q trong mô hình ARIMA



- ACF giúp xác định chỉ số q bằng số lượng lag đến khi xuất hiện một điểm tương quan nằm trong khoảng tin cậy
- Tương tự, PACF giúp xác định chỉ số p

PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

MÔ HÌNH HỒI QUY ARIMA

Phương pháp auto arima để tự xác định chỉ số p,d,q tốt nhất cho mô hình

```
model_arima = auto_arima(train, start_p=0, start_q=0,  
                         max_p=5, max_q=5, m=12,  
                         start_P=0, seasonal=False,  
                         d=0, D=0, trace=True,  
                         error_action='ignore',  
                         suppress_warnings=True,  
                         stepwise=True)
```

```
Performing stepwise search to minimize aic  
ARIMA(0,0,0)(0,0,0)[0] : AIC=17315.923, Time=0.02 sec  
ARIMA(1,0,0)(0,0,0)[0] : AIC=inf, Time=0.31 sec  
ARIMA(0,0,1)(0,0,0)[0] : AIC=15201.036, Time=0.31 sec  
ARIMA(1,0,1)(0,0,0)[0] : AIC=4857.331, Time=0.55 sec  
ARIMA(2,0,1)(0,0,0)[0] : AIC=4859.582, Time=0.50 sec  
ARIMA(1,0,2)(0,0,0)[0] : AIC=4859.268, Time=0.59 sec  
ARIMA(0,0,2)(0,0,0)[0] : AIC=13301.697, Time=0.74 sec  
ARIMA(2,0,0)(0,0,0)[0] : AIC=inf, Time=0.50 sec  
ARIMA(2,0,2)(0,0,0)[0] : AIC=4861.330, Time=0.66 sec  
ARIMA(1,0,1)(0,0,0)[0] intercept : AIC=4856.918, Time=0.60 sec  
ARIMA(0,0,1)(0,0,0)[0] intercept : AIC=11573.639, Time=0.33 sec  
ARIMA(1,0,0)(0,0,0)[0] intercept : AIC=inf, Time=0.32 sec  
ARIMA(2,0,1)(0,0,0)[0] intercept : AIC=4858.977, Time=0.45 sec  
ARIMA(1,0,2)(0,0,0)[0] intercept : AIC=4858.838, Time=0.57 sec  
ARIMA(0,0,0)(0,0,0)[0] intercept : AIC=13593.776, Time=0.04 sec  
ARIMA(0,0,2)(0,0,0)[0] intercept : AIC=9984.163, Time=0.73 sec  
ARIMA(2,0,0)(0,0,0)[0] intercept : AIC=inf, Time=0.60 sec  
ARIMA(2,0,2)(0,0,0)[0] intercept : AIC=4860.205, Time=0.69 sec
```

Best model: ARIMA(1,0,1)(0,0,0)[0] intercept

Total fit time: 8.515 seconds

4856.918123256406

- Từ đây ta có mô hình ARIMA tốt nhất là:

$$Y(t) = c + \phi_1 Y(t-1) + \theta_1 \varepsilon(t-1) + \varepsilon(t)$$

Trong đó:

- $Y(t)$ là giá trị của dữ liệu tại thời điểm t .
- c là hệ số chặn (intercept).
- ϕ_1 là hệ số hồi quy tự hồi quy (AR).
- $Y(t-1)$ là giá trị của dữ liệu tại thời điểm $t-1$.
- θ_1 là hệ số hồi quy trượt (MA).
- $\varepsilon(t-1)$ là nhiễu trắng tại thời điểm $t-1$.
- $\varepsilon(t)$ là nhiễu trắng tại thời điểm t .

PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

MÔ HÌNH HỒI QUY ARIMA

Đọc hiểu mô hình

- Số lượng quan sát trong dữ liệu: 1558
 - AIC: Đánh giá hiệu suất mô hình là 4856.918
 - intercept: 0.0667, cho biết giá trị dự báo ban đầu của chuỗi dữ liệu khi không có giá trị đầu vào khác.
 - ar.L1: 0.9989, cho biết độ tương quan giữa giá trị hiện tại và giá trị trước đó của chuỗi dữ liệu. AR rất gần với 1, cho thấy mối tương quan mạnh giữa các giá trị liền kề.
 - ma.L1: 0.0135, cho biết độ tương quan giữa nhiều trăng hiện tại và nhiều trăng trước đó của chuỗi dữ liệu. Tuy nhiên, giá trị này không đạt mức ý nghĩa thống kê ($P>|z| > 0.05$), cho thấy hệ số MA không có ảnh hưởng đáng kể đến chuỗi dữ liệu.
 - sigma2: Phương sai ước lượng của mô hình,: 1.3106, cho biết mức độ biến động trong chuỗi dữ liệu, càng cao thì chuỗi dữ liệu càng biến đổi mạnh.

```

SARIMAX Results
=====
Dep. Variable:                  y      No. Observations:                 1558
Model: SARIMAX(1, 0, 1)      Log Likelihood:                -2424.459
Date: Sun, 21 May 2023        AIC:                         4856.918
Time: 14:21:08                 BIC:                         4878.323
Sample: 0 - 1558               HQIC:                        4864.877
Covariance Type: opg
=====
            coef    std err          z      P>|z|      [0.025      0.975]
-----
intercept  0.0667    0.085      0.780      0.435     -0.101      0.234
ar.L1       0.9989    0.001    696.027      0.000      0.996      1.002
ma.L1       0.0135    0.019      0.703      0.482     -0.024      0.051
sigma2      1.3106    0.031    42.409      0.000      1.250      1.371
=====
Ljung-Box (L1) (Q):             0.00  Jarque-Bera (JB):                472.13
Prob(Q):                      0.96  Prob(JB):                     0.00
Heteroskedasticity (H):         2.83  Skew:                       -0.10
Prob(H) (two-sided):           0.00  Kurtosis:                   5.69
=====

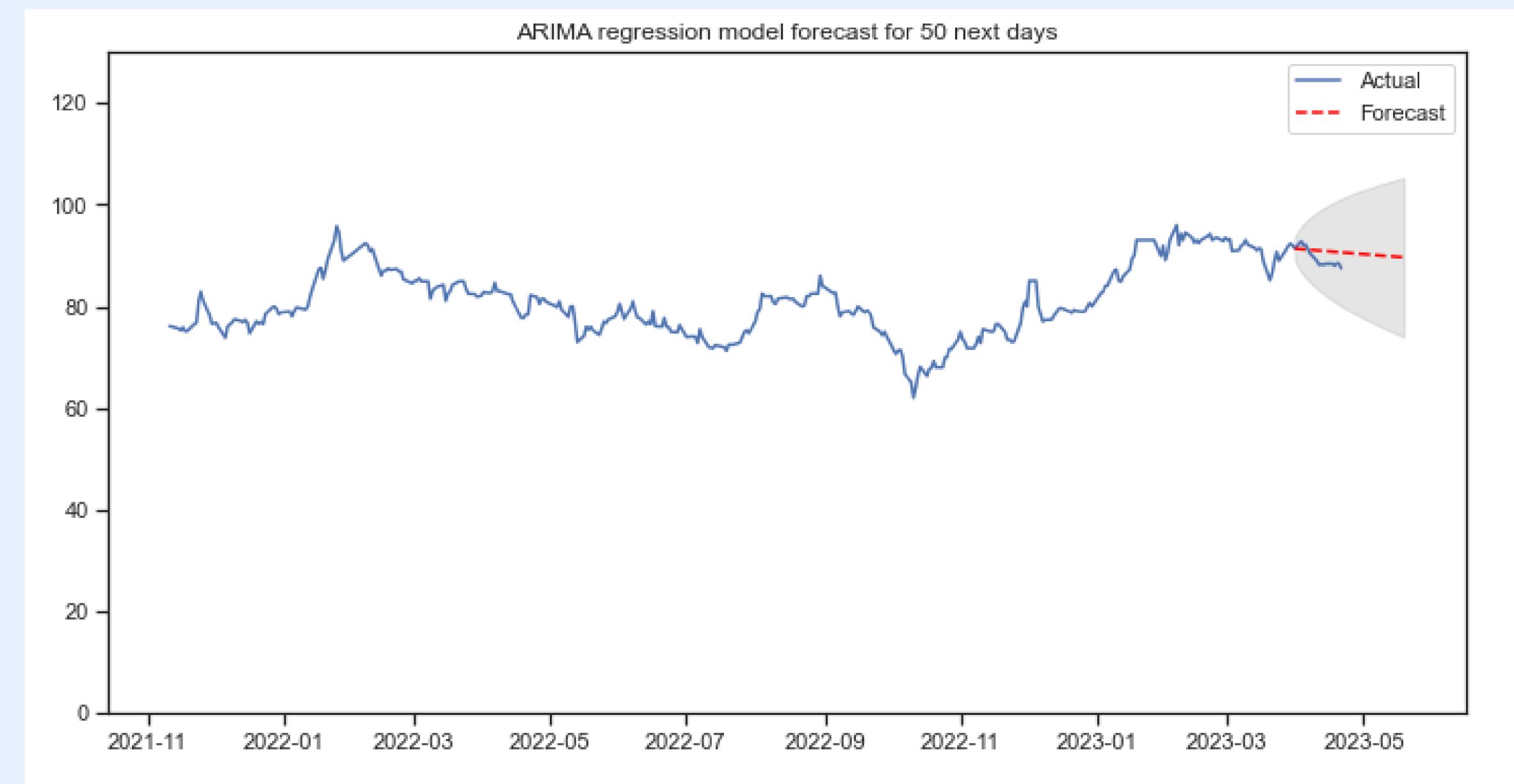
```

PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

DỰ ĐOÁN GIÁ ĐÓNG CỦA CỔ PHIẾU VCB TRONG 50 NGÀY TIẾP THEO

Kết quả dự đoán:

- Kết quả dự đoán giá cổ phiếu của VCB trong 50 ngày kể từ ngày 1/4/2023 có xu hướng giảm (đường nét đứt màu đỏ)



PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

DỰ ĐOÁN GIÁ ĐÓNG CỦA CỔ PHIẾU VCB TRONG 50 NGÀY TIẾP THEO

Kết quả dự đoán:

2023-03-31	91.356775	2023-04-26	90.443184
2023-04-01	91.321143	2023-04-27	90.408574
2023-04-02	91.285550	2023-04-28	90.374004
2023-04-03	91.249998	2023-04-29	90.339472
2023-04-04	91.214485	2023-04-30	90.304978
2023-04-05	91.179012	2023-05-01	90.270524
2023-04-06	91.143578	2023-05-02	90.236108
2023-04-07	91.108185	2023-05-03	90.201730
2023-04-08	91.072830	2023-05-04	90.167391
2023-04-09	91.037516	2023-05-05	90.133090
2023-04-10	91.002241	2023-05-06	90.098828
2023-04-11	90.967006	2023-05-07	90.064604
2023-04-12	90.931810	2023-05-08	90.030419
2023-04-13	90.896653	2023-05-09	89.996271
2023-04-14	90.861536	2023-05-10	89.962162
2023-04-15	90.826458	2023-05-11	89.928091
2023-04-16	90.791419	2023-05-12	89.894059
2023-04-17	90.756420	2023-05-13	89.860064
2023-04-18	90.721460	2023-05-14	89.826108
2023-04-19	90.686539	2023-05-15	89.792189
2023-04-20	90.651657	2023-05-16	89.758309
2023-04-21	90.616814	2023-05-17	89.724466
2023-04-22	90.582010	2023-05-18	89.690661
2023-04-23	90.547245	2023-05-19	89.656895
2023-04-24	90.512519		
2023-04-25	90.477832		

PHÂN TÍCH HỒI QUY VÀ DỰ ĐOÁN

DỰ ĐOÁN GIÁ ĐÓNG CỦA CỔ PHIẾU VCB TRONG 50 NGÀY TIẾP THEO

Kiểm tra tính chính xác của mô hình

```
def _measure_metric(y, yhat):
    e = y-yhat
    mse=np.mean(e**2)
    rmse=np.sqrt(mse)
    mae=np.mean(np.abs(e))
    mape=np.mean(e/y)
    print('Mean Square Error: {}'.format(mse))
    print('Root Mean Square Error: {}'.format(rmse))
    print('Mean Absolute Error: {}'.format(mae))
    print('Mean Absolute Percentage Error: {}'.format(mape))
    return mse, rmse, mae, mape

mse, rmse, mae, mape=_measure_metric(test.values[0], fitted)
```

```
Mean Square Error: 5.544218000048013
Root Mean Square Error: 2.3546163169501764
Mean Absolute Error: 2.300782589185094
Mean Absolute Percentage Error: 0.024792915831735925
```

Dựa vào kết quả ta có:

- RMSE: Biên độ giao động của giá trị dự báo xung quanh giá trị thực tế là 2.35
- MAE: Trung bình sai số giữa giá trị dự báo và giá trị thực tế là 2.3
- MAPE: Sai số giữa giá trị dự báo so với giá trị thực tế bằng 2.48% giá trị thực tế.

TỔNG KẾT



TỔNG KẾT

- Thông qua việc tìm hiểu về chứng khoán và thực hiện xây dựng các trực quan hóa về chứng khoán, ta cũng đã hiểu phần nào về bộ môn này. Tuy nhiên các con số, các trực quan cho dù ta phân tích chuẩn đến đâu thì vẫn có tỉ lệ sai số.
 - Chính vì vậy, để có thể thực sự kiếm tiền được từ chứng khoán, ta phải am hiểu sâu sắc về nó, phải tích lũy kinh nghiệm trong rất nhiều năm để có thể hiểu và xử lý các tình huống sai số xảy ra.
 - Việc giao dịch, mua bán chứng khoán đã có thâm niên hàng trăm năm. Có người giàu nhờ chứng khoán, có người phá sản vì chứng khoán. Nếu không đủ kiến thức hoặc kinh nghiệm về chứng khoán, hãy tích lũy thêm trước khi nghĩ đến việc kiếm tiền với nó.

A uniform grid of 100 blue circular dots, arranged in a perfect 10 by 10 square pattern. The dots are evenly spaced both horizontally and vertically, creating a clean, geometric background.

THANK YOU