**ĐẠI HỌC HUẾ**

**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**



🙠🙟🕮🙝🙢

**Icon

Description automatically generated**

**BÁO CÁO**

**ĐỒ ÁN**

**Học kỳ I, năm học 2021-2022**

**Học phần:**

**Phân tích dữ liệu với R**

**Họ và tên sinh viên: Nguyễn Tiến Thịnh**

**Lớp: K1- Khoa học dữ liệu và Trí tuệ nhân tạo**

|  |
| --- |
| Số phách  *(Do hội đồng chấm thi ghi)* |

**Thừa Thiên Huế, ngày 13 tháng 12 năm 2021**

**ĐẠI HỌC HUẾ**

**KHOA KỸ THUẬT VÀ CÔNG NGHỆ**



🙠🙟🕮🙝🙢

**Icon

Description automatically generated**

**(Bìa phụ 2)**

**BÁO CÁO**

**ĐỒ ÁN**

**Học kỳ I, năm học 2021-2022**

**Học phần:**

**Phân tích dữ liệu với R**

**Giảng viên hướng dẫn: Hồ Quốc Dũng**

**Lớp: K1- Khoa học dữ liệu và Trí tuệ nhân tạo**

**Sinh viên thực hiện: Nguyễn Tiến Thịnh**

*(ký tên và ghi rõ họ tên)*

|  |
| --- |
| **Số phách**  *(Do hội đồng chấm thi ghi)* |

**Thừa Thiên Huế, ngày 13 tháng 12 năm 2021**

Mục lục

[I. Mô tả dữ liệu chứng khoán 5](#_Toc91446330)

[1. Cách lấy dữ liệu: 5](#_Toc91446331)

[2. Mô tả dữ liệu đã tải về: 6](#_Toc91446332)

[a. Số lượng dòng dữ liệu: 6](#_Toc91446333)

[b. Danh sách các trường dữ liệu: 7](#_Toc91446334)

[II. Phân tích thống kê mô tả dữ liệu (Sử dụng R) 7](#_Toc91446335)

[1. Tính các chỉ số thống kê mô tả 7](#_Toc91446336)

[2. Vẽ các biểu đồ để trực quan hóa dữ liệu: 7](#_Toc91446337)

[3. Một số nhận xét về số liệu: 11](#_Toc91446338)

[III. Phân tích dữ liệu qua Dải Bollinger Band với Python 12](#_Toc91446339)

[1. Giới thiệu về Bollinger Band 12](#_Toc91446340)

[2. Triển khai bằng Python 13](#_Toc91446341)

[a. Import Thư viện và Đọc dữ liệu 13](#_Toc91446342)

[b. Tính toán Bollinger Band 13](#_Toc91446343)

[c. Vẽ các giá trị dải Bollinger 14](#_Toc91446344)

[d. Tạo chiến lược giao dịch 14](#_Toc91446345)

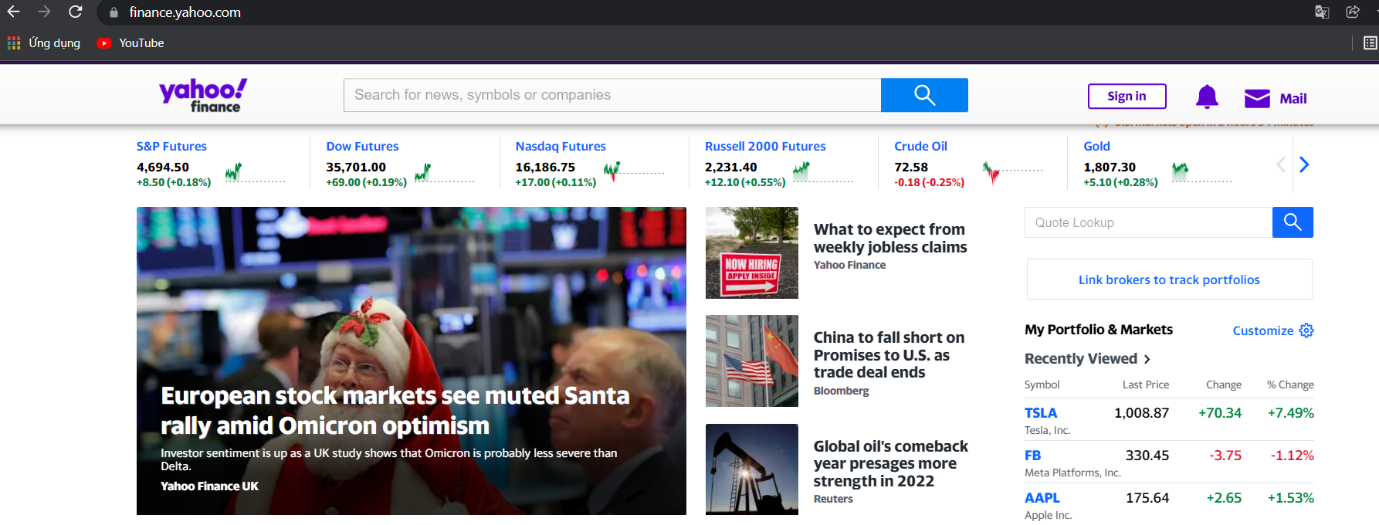
[e. Triển khai đầu tư 16](#_Toc91446346)

[f. Backtesting 16](#_Toc91446347)

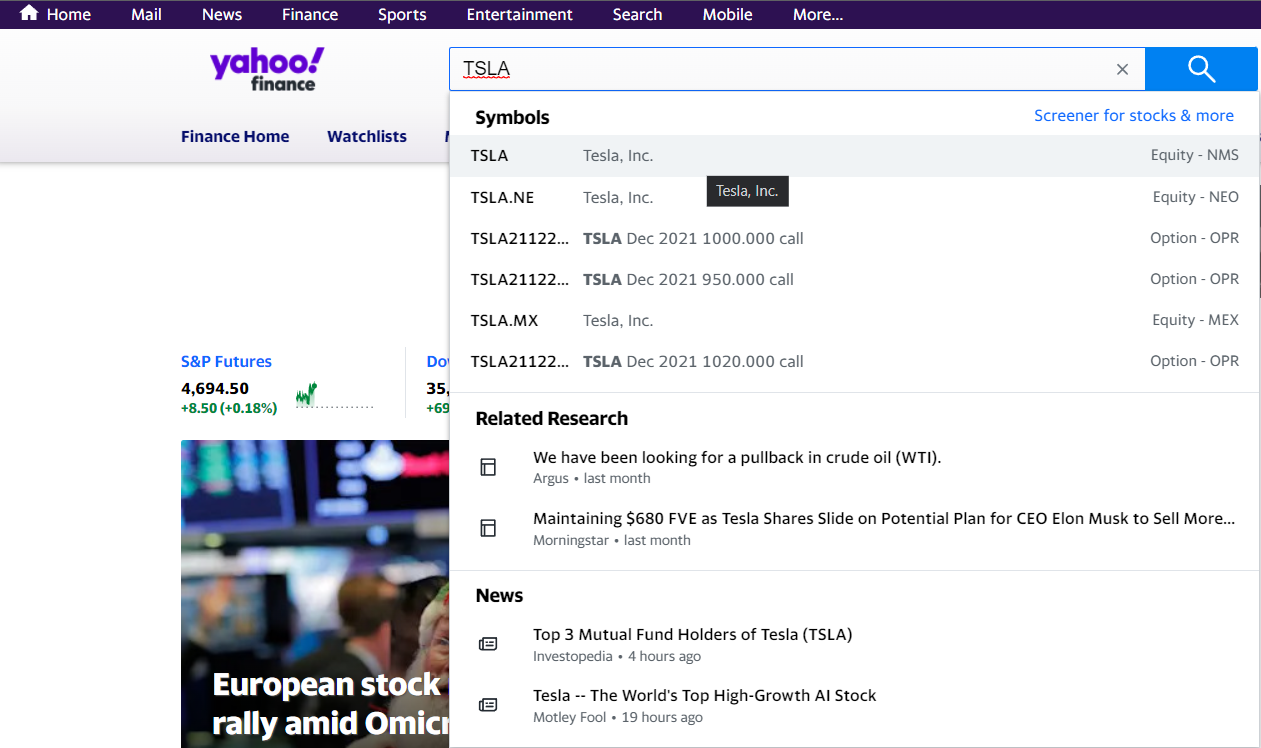
# Mô tả dữ liệu chứng khoán

## Cách lấy dữ liệu:

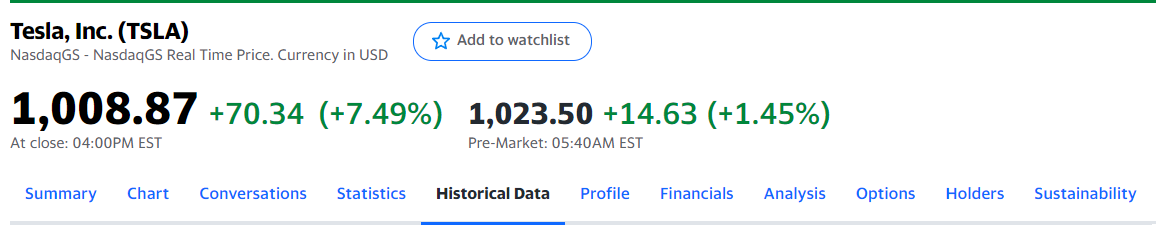
B1: Truy cập trang https://finance.yahoo.com/



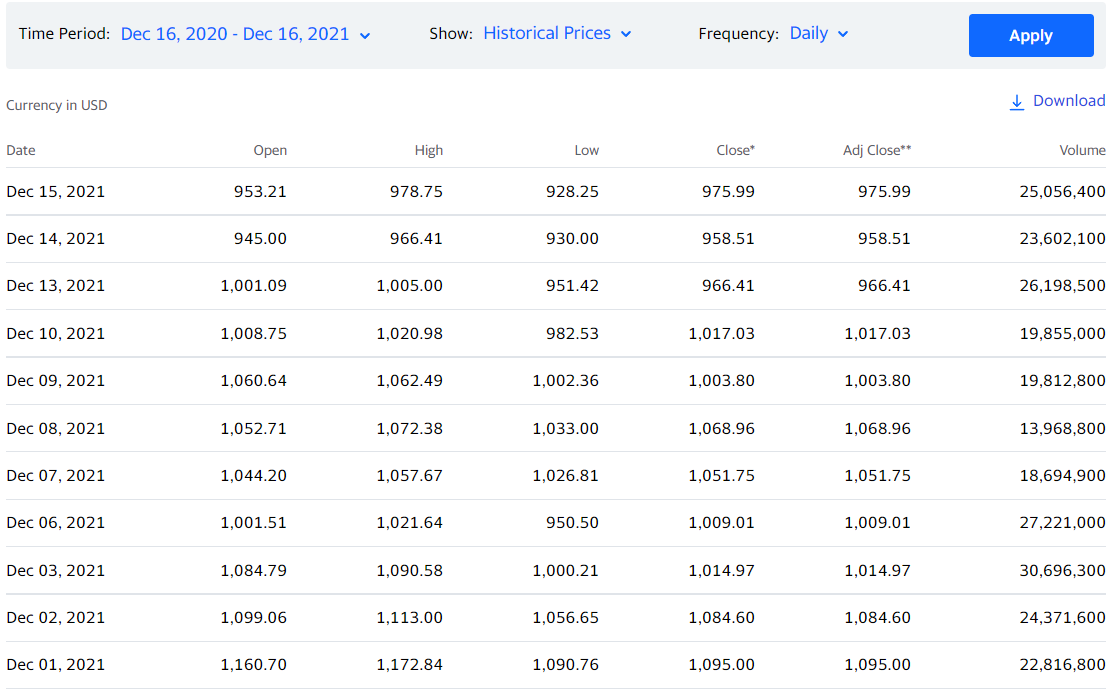
B2: Nhập tên công ty muốn lấy dữ liệu:



B3: Click vào Historical Data



B4: Chọn khoảng thời gian rồi nhấn Download



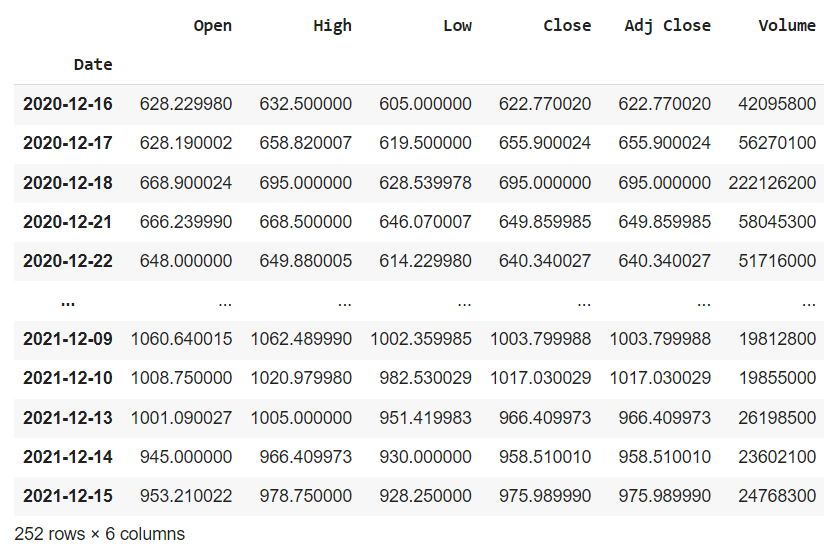
Link github: https://github.com/thinh2904/Do\_An\_PTDL\_R/blob/main/TSLA.csv

## Mô tả dữ liệu đã tải về:

### Số lượng dòng dữ liệu:

Dữ liệu gồm 252 dòng và 7 cột

Bắt đầu từ ngày 16/12/2020 đến ngày 15/12/2021 (1 năm)



### Danh sách các trường dữ liệu:

Cột Date (Ngày): Kiểu dữ liệu DateTime (Ngày giờ)

Cột Open (Giá mở cửa): Kiểu dữ liệu Float (Số thực)

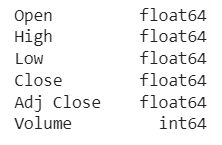
Cột High (Giá cao nhất trong ngày): Kiểu dữ liệu Float (Số thực)

Cột Low (Giá thấp nhất trong ngày): Kiểu dữ liệu Float (Số thực)

Cột Close (Giá đóng cửa): Kiểu dữ liệu Float (Số thực)

Cột Adj Close (Giá đóng cửa hiệu chỉnh): Kiểu dữ liệu Float (Số thực)

Cột Volume (Số lượng giao dịch): Kiểu dữ liệu Int (Số nguyên)



# Phân tích thống kê mô tả dữ liệu (Sử dụng R)

## Tính các chỉ số thống kê mô tả

Sử dụng hàm summary có trong R kết hợp tạo hàm funtction để tính toán thêm các chỉ số còn thiếu:

Code:

tkmt <- function(x){

av <- mean(x)

sd <- sd(x)

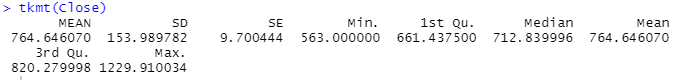
se <- sd/sqrt(length(x))

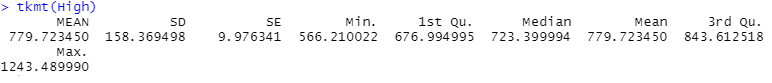
c(MEAN=av, SD=sd, SE=se, summary(x))

}

Kết quả:











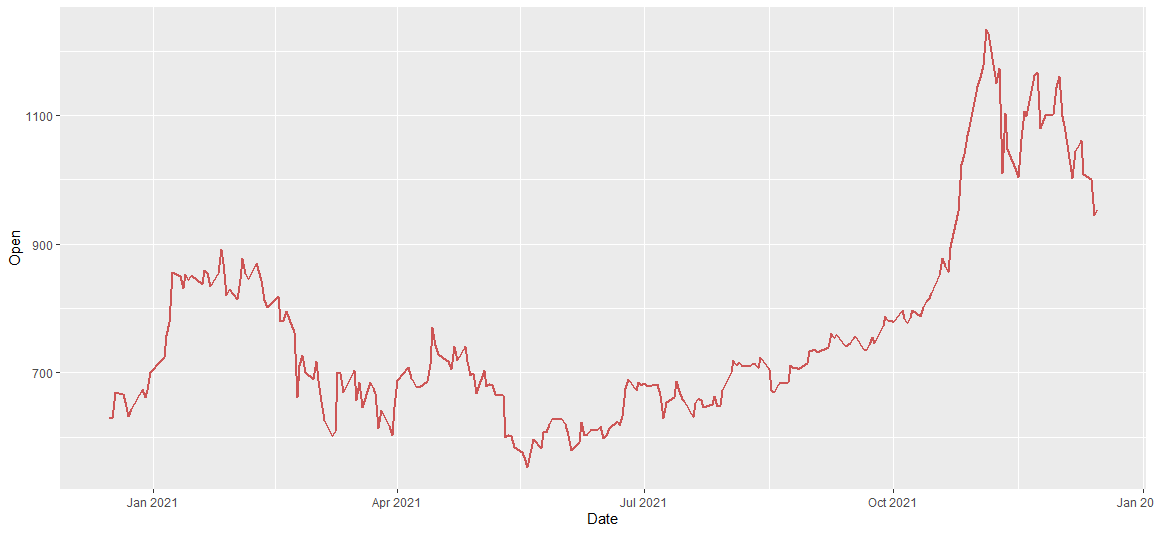
## Vẽ các biểu đồ để trực quan hóa dữ liệu:

Sử dụng thư viện ggplot2 để vẽ biểu đồ Line Plot của các cột dữ liệu

ggplot(df, aes(x = Date, y = Open)) +

geom\_line(color = "indianred3",

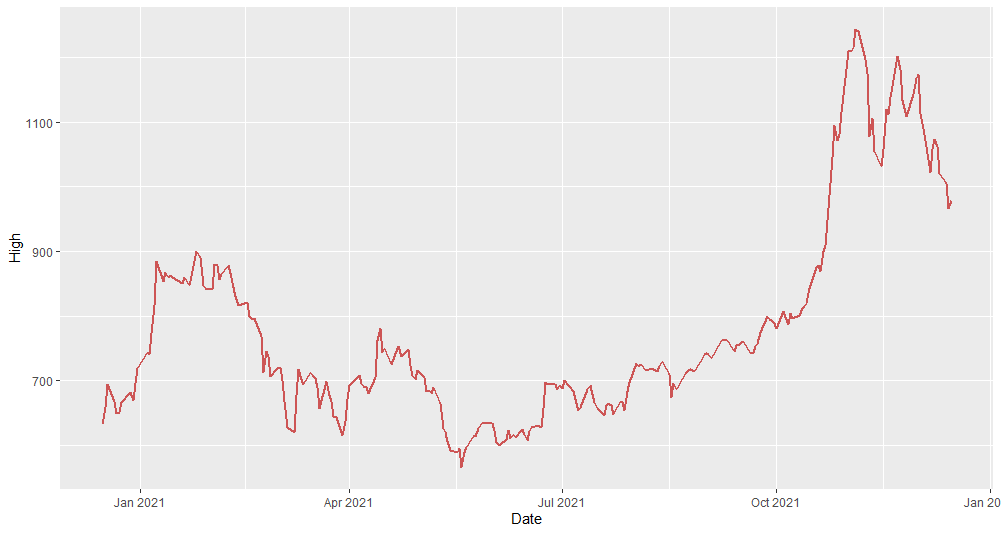
size=1 )



ggplot(df, aes(x = Date, y = High)) +

geom\_line(color = "indianred3",

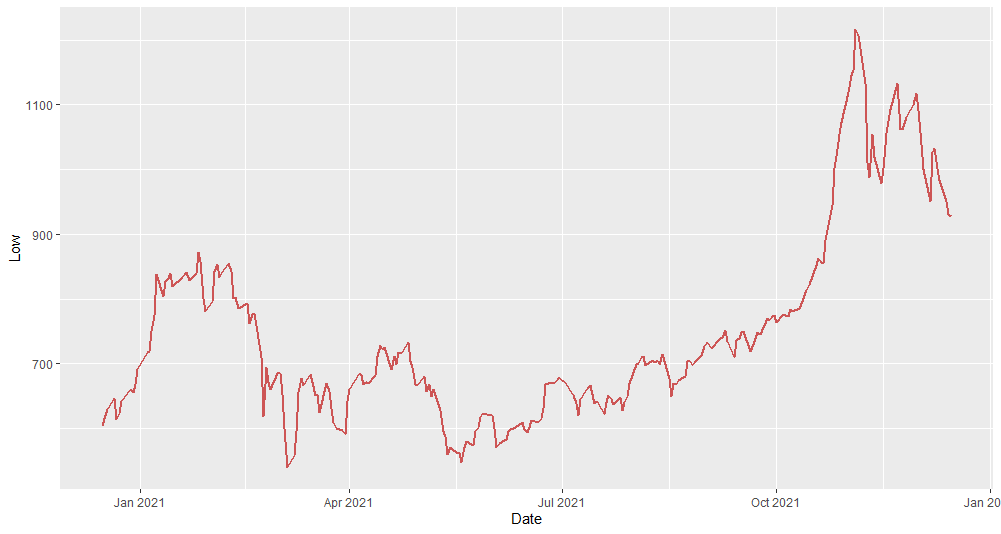
size=1 )



ggplot(df, aes(x = Date, y = Low)) +

geom\_line(color = "indianred3",

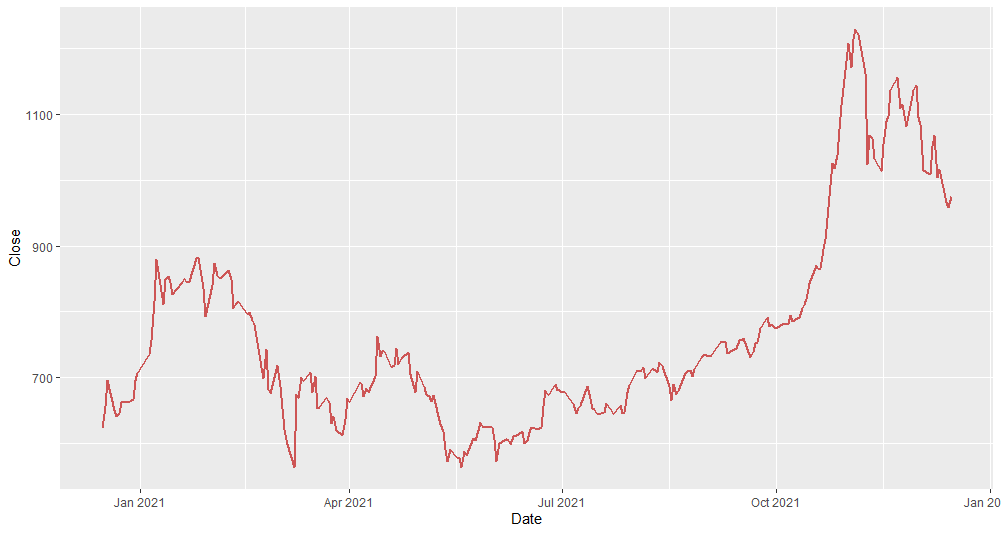
size=1 )



ggplot(df, aes(x = Date, y = Close)) +

geom\_line(color = "indianred3",

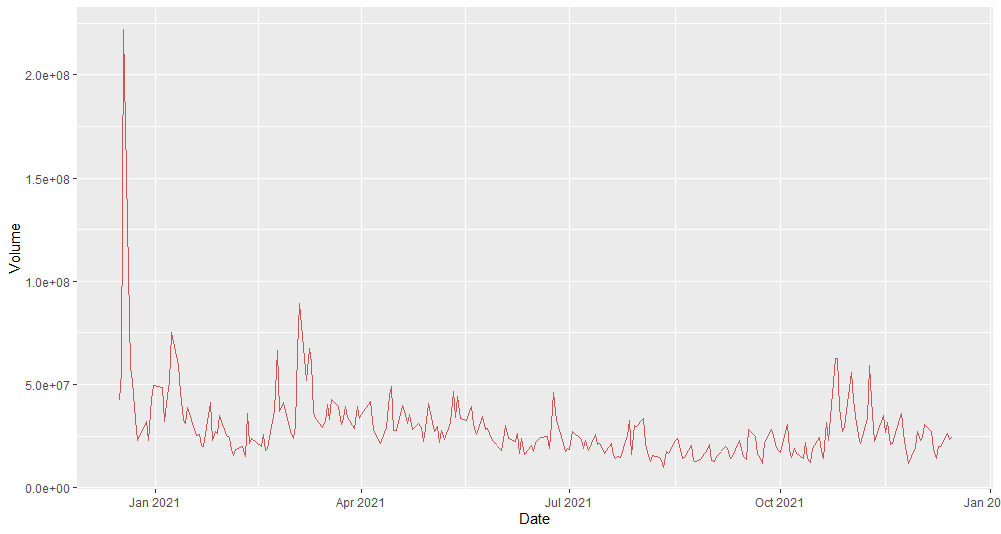
size=1 )



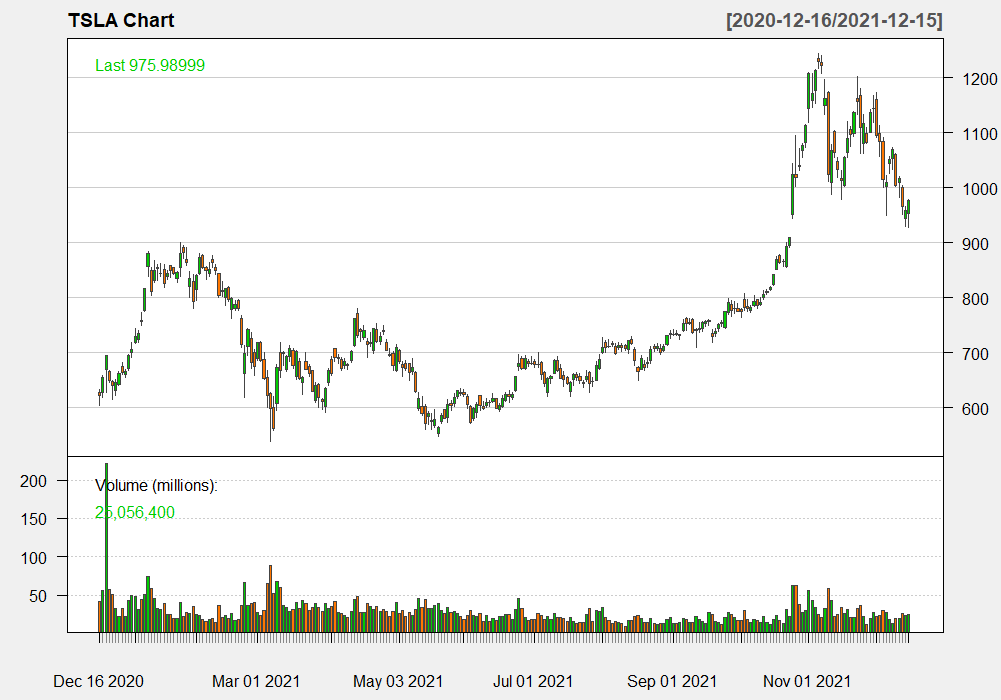
ggplot(df, aes(x = Date, y = Volume)) +

geom\_line(color = "indianred3",

size=0.5 )



Biểu đồ Nến:



Sử dụng thư viện highcharter để vẽ biểu đồ Nến kèm theo đường SMA:

highchart(type="stock") %>%

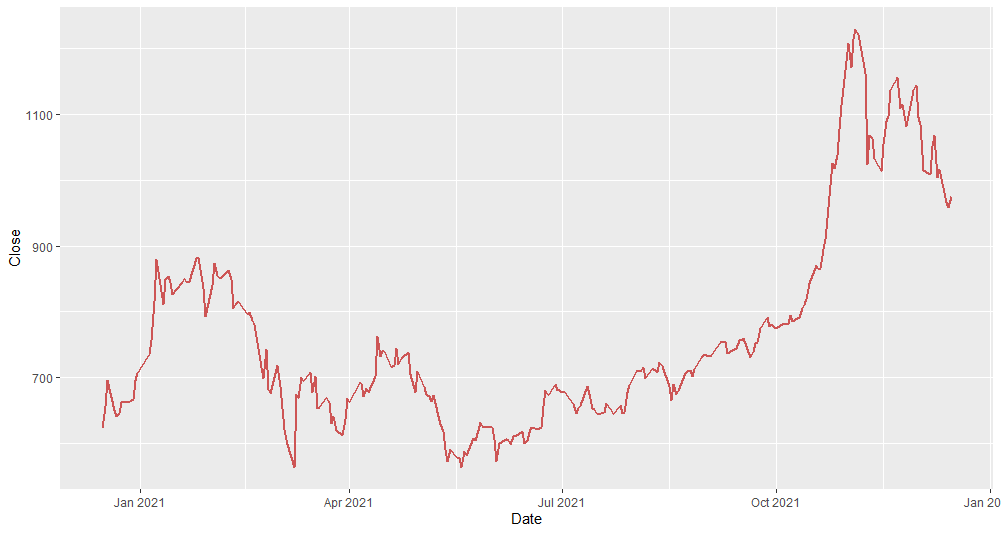
hc\_add\_series(tsla) %>%

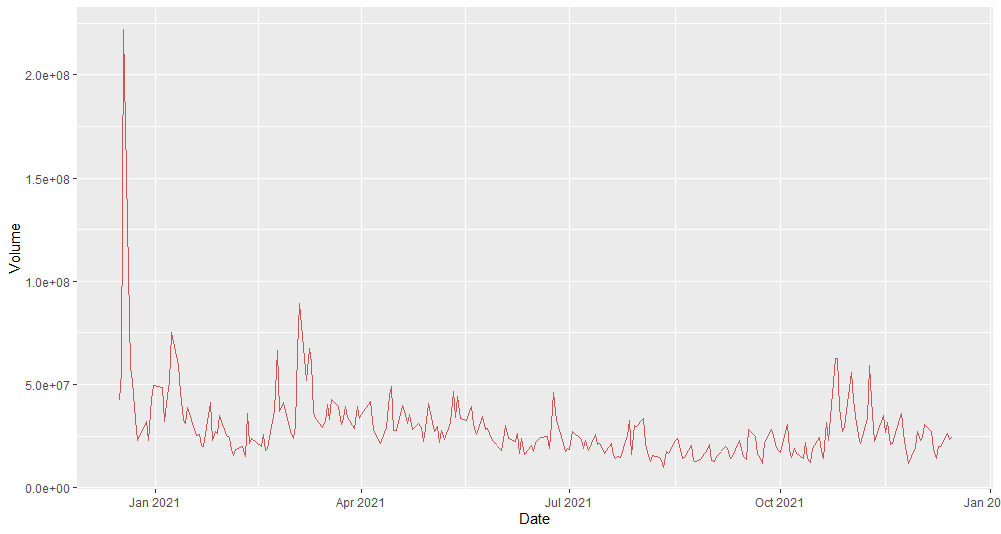
hc\_add\_series(SMA(na.omit(Cl(tsla)),n=20),name="SMA(20)") %>%

hc\_title(text="<b>TSLA</b>")



## Một số nhận xét về số liệu:





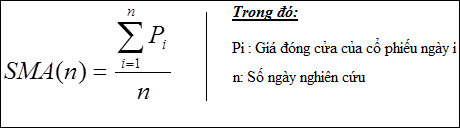
* Giá chứng khoán của TSLA nhìn chung trong năm 2021 qua là tăng rõ rệt (Thấp nhất 563, Cao nhất 1229 – Lấy giá Close làm chuẩn).
* Trong giai đoạn từ tháng 2/2021 đến giữ tháng 3/2021 thì giá giảm liên tục rồi sau đó có tăng nhẹ nhưng không thể vượt đỉnh ban đầu. Số lượng giao dịch trong khoảng thời gian này cũng tăng giảm liên tục.
* Tuy nhiên trong giai đoạn từ giữa tháng 5/2021 đến đầu tháng 11/2021 giá lại tăng tiên lục và liên tục phá đỉnh và chạm đến 1229.
* Từ tháng 11/2021 đến nay thì có dấu hiệu giảm nhẹ.

# Phân tích dữ liệu qua Dải Bollinger Band với Python

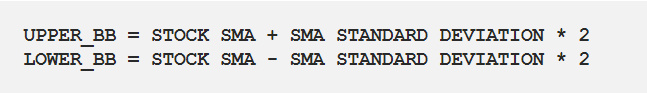
## Giới thiệu về Bollinger Band

Bollinger Band là một công cụ phân tích kỹ thuật xác định bởi đường trung bình đơn giản (Simple Moving Average – SMA) ở giữa, dải trên và dải dưới. Dải Bollinger Band sẽ tự điều chỉnh mở rộng trong các giai đoạn thị trường biến động và thu hẹp trong các giai đoạn thị trường ít biến động hơn.

Công thức tính SMA:



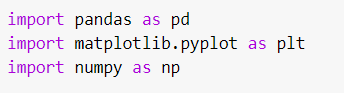
Công thức tính để tính dải trên và dải dưới của chứng khoán như sau:



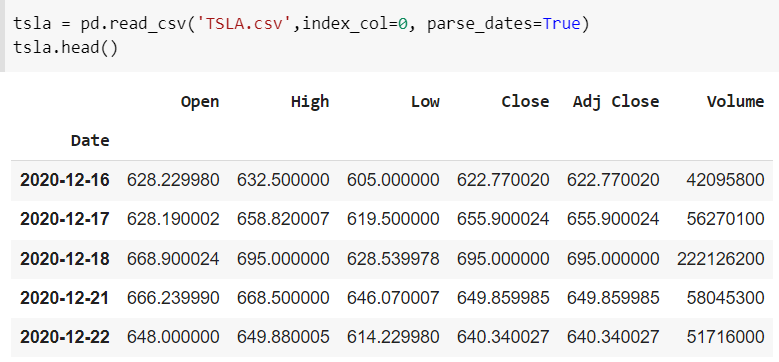
## Triển khai bằng Python

### Import Thư viện và Đọc dữ liệu

Các thư viện chính sẽ là Pandas để thao tác dữ liệu, Matplotlib cho mục đích vẽ biểu đồ và Numpy để tính toán.

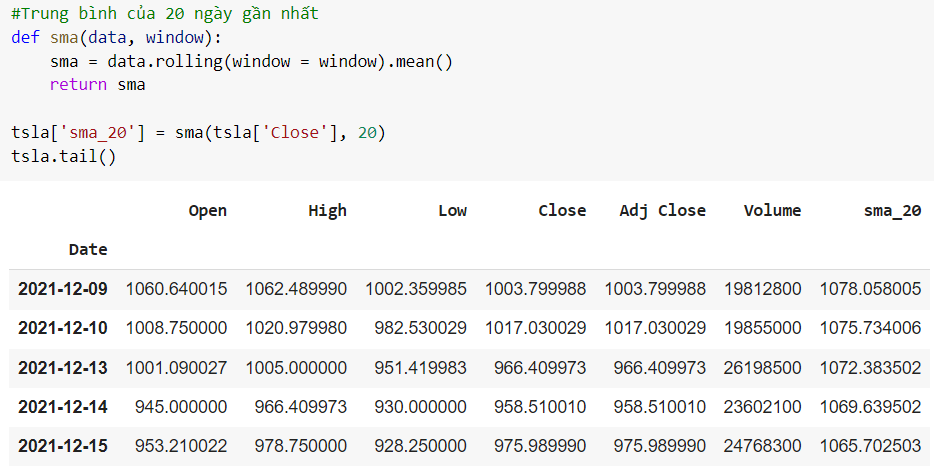


Đọc dữ liệu từ file TSLA.csv bằng thư viện Pandas

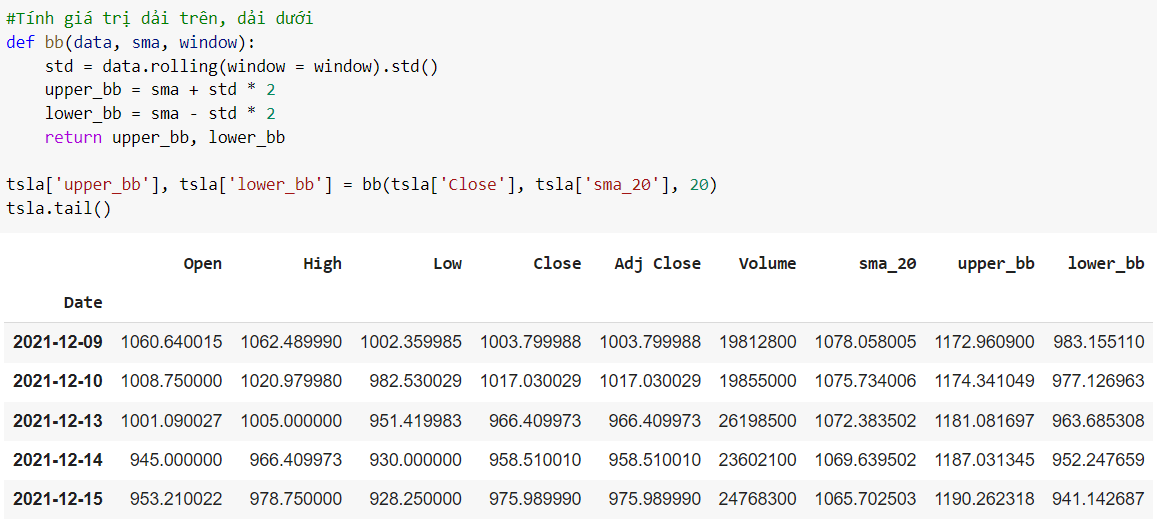


### Tính toán Bollinger Band

* Tính toán các giá trị SMA: Tính các giá trị SMA của Tesla với số chu kỳ là 20

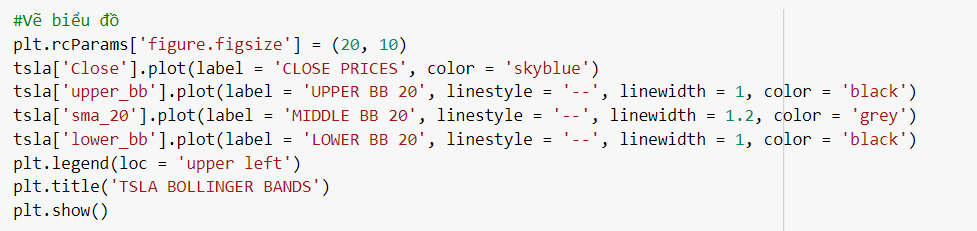


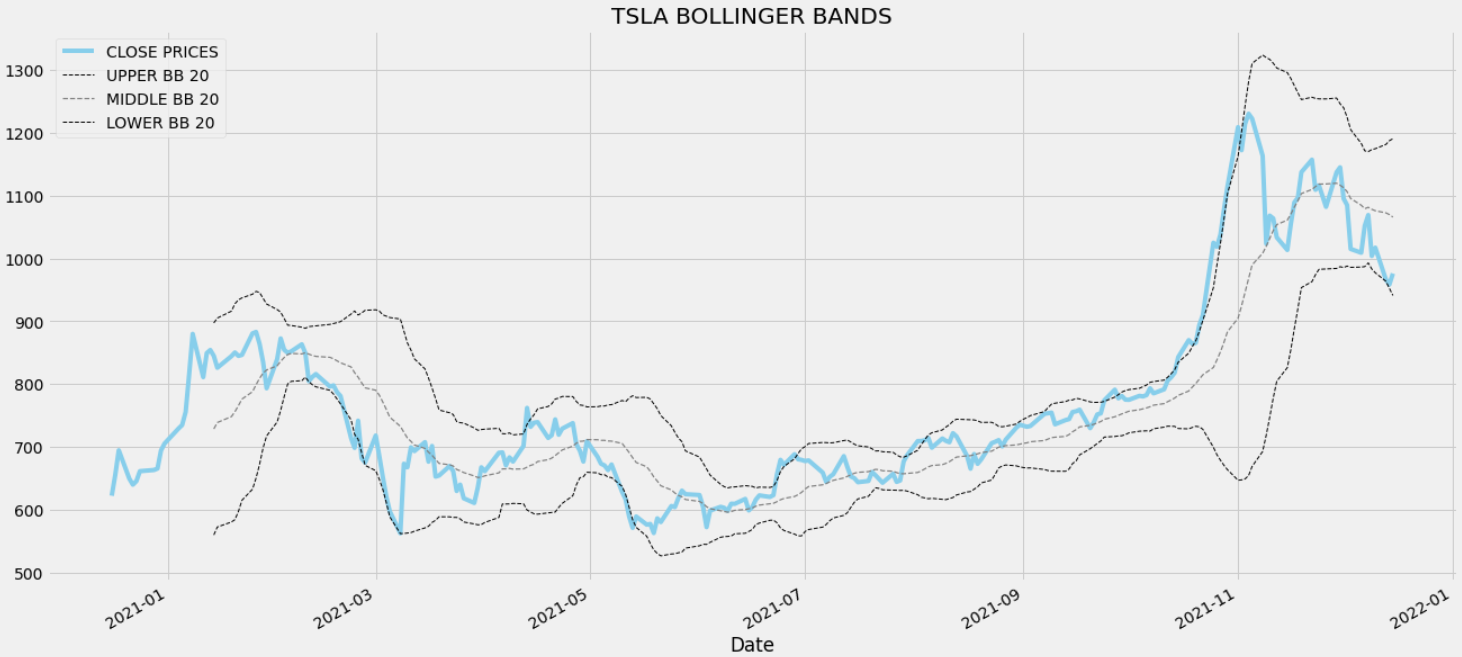
* Tính toán dải trên và dải dưới: Tính toán dải Bollinger bằng cách sử dụng các giá trị SMA mà chúng ta đã tạo trước đó.



### Vẽ các giá trị dải Bollinger

Sử dụng hàm plot được cung cấp bởi thư viện Matplotlib để vẽ biểu đồ giá trị Dải Bollinger cùng với Giá đóng cửa (tsla[‘Close’]) của Tesla.



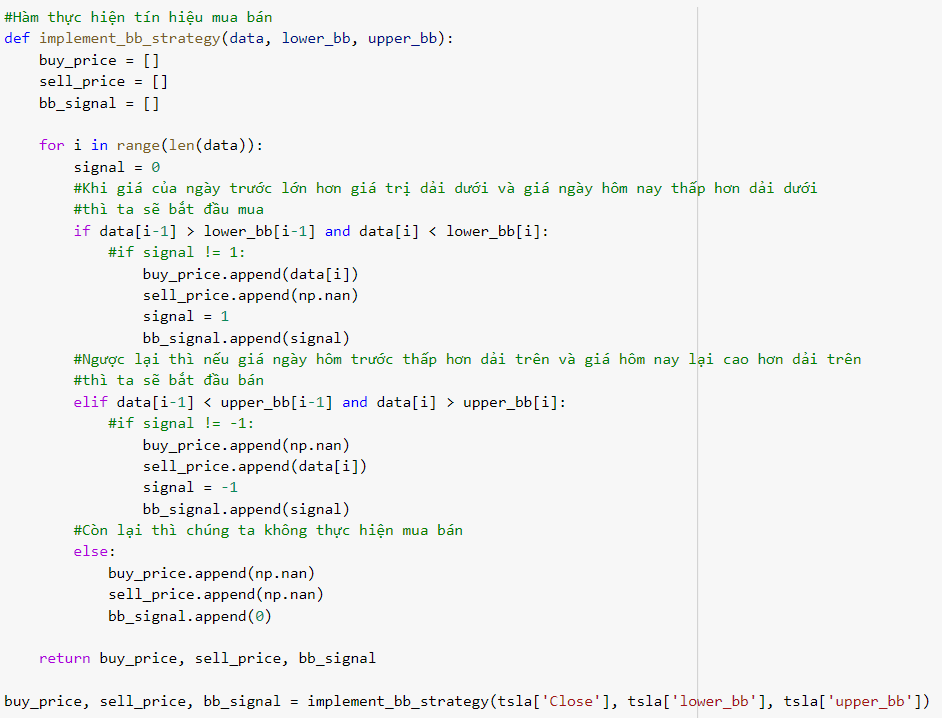


Đường màu xanh nhạt đại diện cho giá đóng cửa của Tesla và đường chấm màu đen đại diện cho các dải dưới và trên. Đường chấm xám đại diện cho dải giữa hoặc các giá trị SMA 20 của Tesla. Khoảng cách giữa dải trên và dải dưới đang trở nên hẹp hoặc rộng hơn vào những thời điểm khác nhau thể hiện sự biến động của cổ phiếu.

### Tạo chiến lược giao dịch

Tạo hàm có tên là implement\_bb\_strategy lấy giá cổ phiếu (‘data’), giá trị dải dưới (‘lower\_bb’) và giá trị dải trên (‘upper\_bb’) làm tham số.

Tạo 3 list trống buy\_price, sell\_price và bb\_siggnal, các giá trị sẽ được thêm vào khi tạo chiến lược giao dịch.



Biểu đồ giá trị của danh sách giao dịch đã tạo.

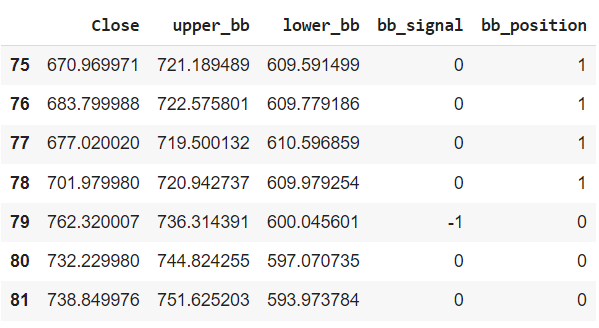


Bất cứ khi nào giá đóng cửa của cổ phiếu (đường màu xanh nhạt) bên dưới dải dưới (đường chấm đen phía dưới), tín hiệu mua được vẽ bằng màu xanh lá, tương tự như vậy, bất cứ khi nào giá đóng cửa cổ phiếu vượt lên trên dải trên (đường chấm đen phía trên), một tín hiệu bán được vẽ bằng màu đỏ.

### Triển khai đầu tư

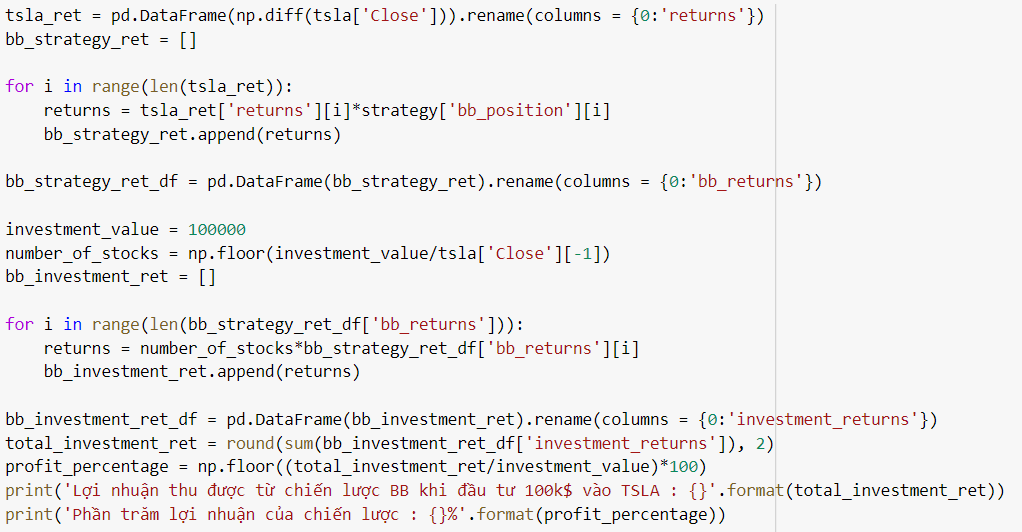
Tạo một list tên postion chỉ ra 1 nếu ta nắm giữu cổ phiếu hoặc 0 nếu ta không sở hữu cổ phiếu.





### Backtesting

Backtest là quá trình kiểm định phương pháp/hệ thống giao dịch dựa trên dữ liệu từ quá khứ.



Kết quả:

