

## HƯỚNG DẪN THỰC HÀNH LẤY VÀ XỬ LÝ SỐ LIỆU

Một trong những bước quan trọng để chạy mô hình hồi quy thành công là việc phải tạo được một bộ dữ liệu đầu vào chất lượng. Thông qua các hoạt động sau, anh/ chị sẽ thực hành các kỹ năng liên quan tới tìm kiếm, tải và định dạng dữ liệu để chuẩn bị cho case study môn Quantitative Methods.

### HOẠT ĐỘNG 1: TÌM KIẾM VÀ THU THẬP DỮ LIỆU

Trong hoạt động này, anh/ chị cần tìm kiếm và tải dữ liệu của:

#### Chỉ số VN30:

- Ngày giao dịch
- Lịch sử điểm
- Phần trăm thay đổi (%)

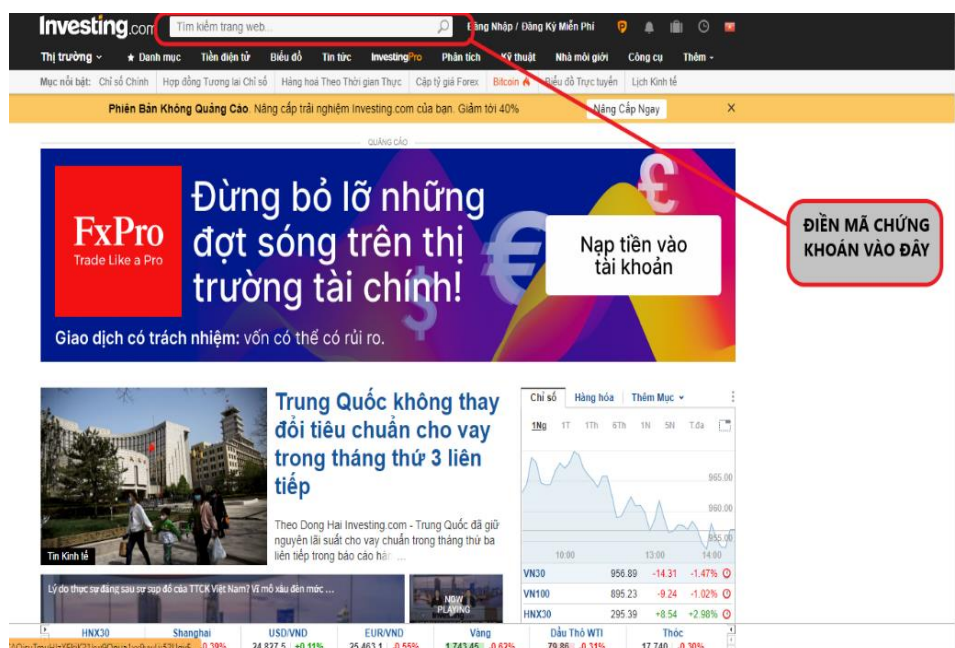
#### Cổ phiếu CTCP Tập đoàn Masan (MSN):

- Ngày giao dịch
- Giá đóng cửa
- Phần trăm thay đổi (%)

Bước 1: Truy cập vào đường link sau đây để lấy số liệu:

<https://vn.investing.com/>

Bước 2: Điền mã chứng khoán “VN30” vào mục tìm kiếm.

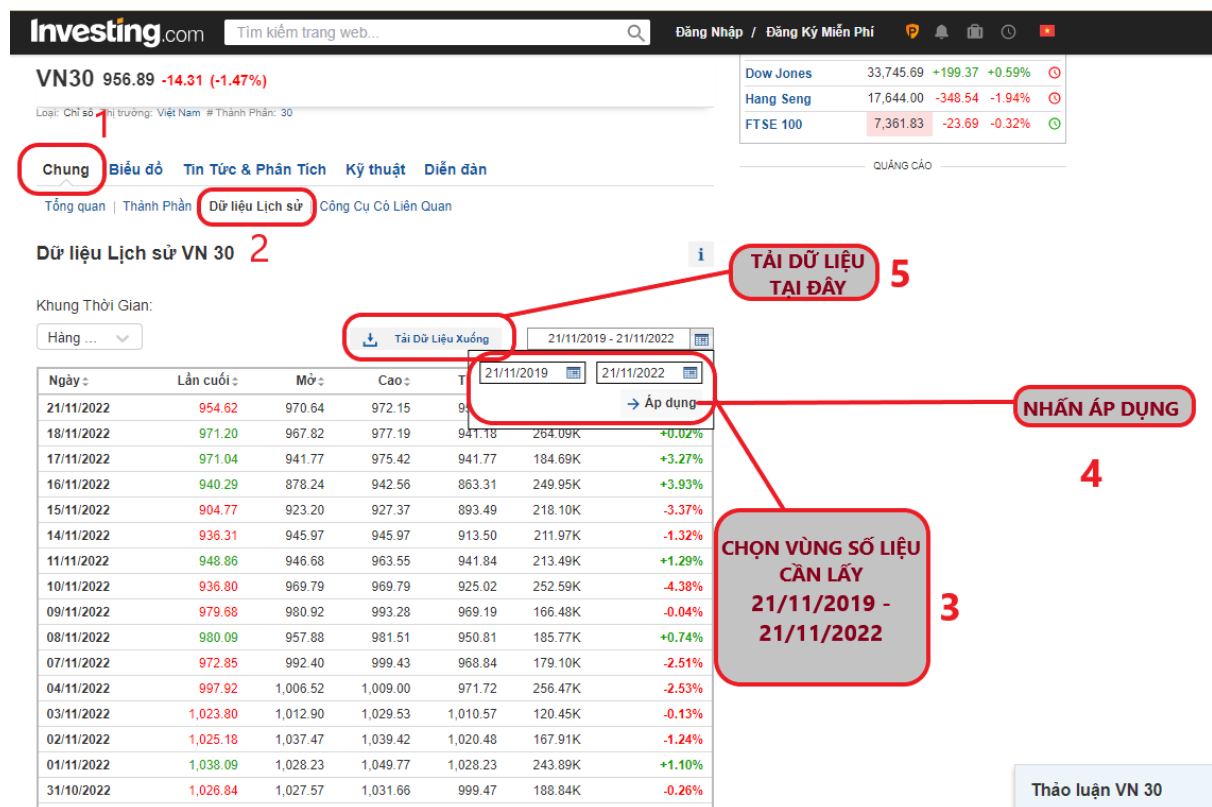


**Bước 3:** Chọn lần lượt theo thứ tự sau đây: **Chung > Dữ liệu lịch sử**.

**Bước 4:** Tiến hành chọn thời gian cần lấy dữ liệu: Chọn giai đoạn **21/11/2019 – 21/11/2022**.

**Bước 5:** Nhấn “Tải Dữ Liệu Xuống”.

**Bước 6:** Thực hiện tương tự các bước trên với dữ liệu của **cổ phiếu MSN**.



**Investing.com** | Tìm kiếm trang web... | Đăng Nhập / Đăng Ký Miễn Phí

**VN30** 956.89 -14.31 (-1.47%)

Loại: Chỉ số | Thị trường: Việt Nam | Thành Phần: 30

**Chung** | Biểu đồ | Tin Tức & Phân Tích | Kỹ thuật | Diễn đàn

Tổng quan | Thành Phần | **Dữ liệu Lịch sử** | Công Cụ Có Liên Quan

**Dữ liệu Lịch sử VN 30**

Khung Thời Gian: Hàng ...

**Tải Dữ Liệu Xuống** | 21/11/2019 - 21/11/2022

Ngày	Lần cuối	Mở	Cao	Thấp	Thay đổi
21/11/2022	954.62	970.64	972.15	941.18	-0.02%
18/11/2022	971.20	967.82	977.19	941.77	+3.27%
17/11/2022	971.04	941.77	975.42	941.77	+3.27%
16/11/2022	940.29	878.24	942.56	863.31	+3.93%
15/11/2022	904.77	923.20	927.37	893.49	-3.37%
14/11/2022	936.31	945.97	945.97	913.50	-1.32%
11/11/2022	948.86	946.68	963.55	941.84	+1.29%
10/11/2022	936.80	969.79	969.79	925.02	-4.38%
09/11/2022	979.68	980.92	993.28	969.19	-0.04%
08/11/2022	980.09	957.88	981.51	950.81	+0.74%
07/11/2022	972.85	992.40	999.43	968.84	-2.51%
04/11/2022	997.92	1,006.52	1,009.00	971.72	-2.53%
03/11/2022	1,023.80	1,012.90	1,029.53	1,010.57	-0.13%
02/11/2022	1,025.18	1,037.47	1,039.42	1,020.48	-1.24%
01/11/2022	1,038.09	1,028.23	1,049.77	1,028.23	+1.10%
31/10/2022	1,026.84	1,027.57	1,031.66	999.47	-0.26%

**Tải Dữ Liệu Xuống** | **ÁP DỤNG**

**CHỌN VÙNG SỐ LIỆU CẦN LẤY 21/11/2019 - 21/11/2022**

**NHẤN ÁP DỤNG**

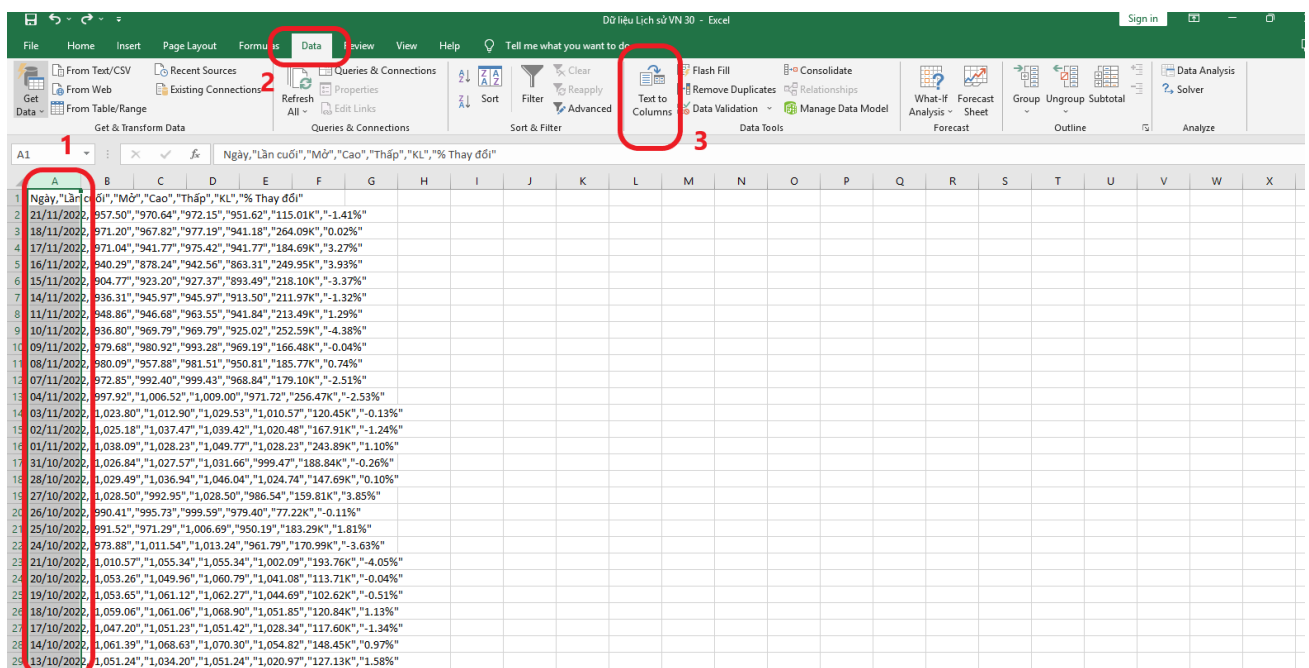
Thảo luận VN 30

## HOẠT ĐỘNG 2: XỬ LÝ SỐ LIỆU

Sau khi đã hoàn tất tải hai bộ dữ liệu thô dưới định dạng file excel, tiến hành thực hiện các bước sau để xử lý số liệu. Với bộ dữ liệu của **cổ phiếu MSN**, áp dụng Bước 1, 2 và 5. Với bộ dữ liệu của **chỉ số VN30**, áp dụng từ Bước 1 đến 5.

**Bước 1:** Trên giao diện ban đầu của dữ liệu, tiến hành format lại dữ liệu như sau:

**Chọn toàn bộ cột A > Data > Text to columns**

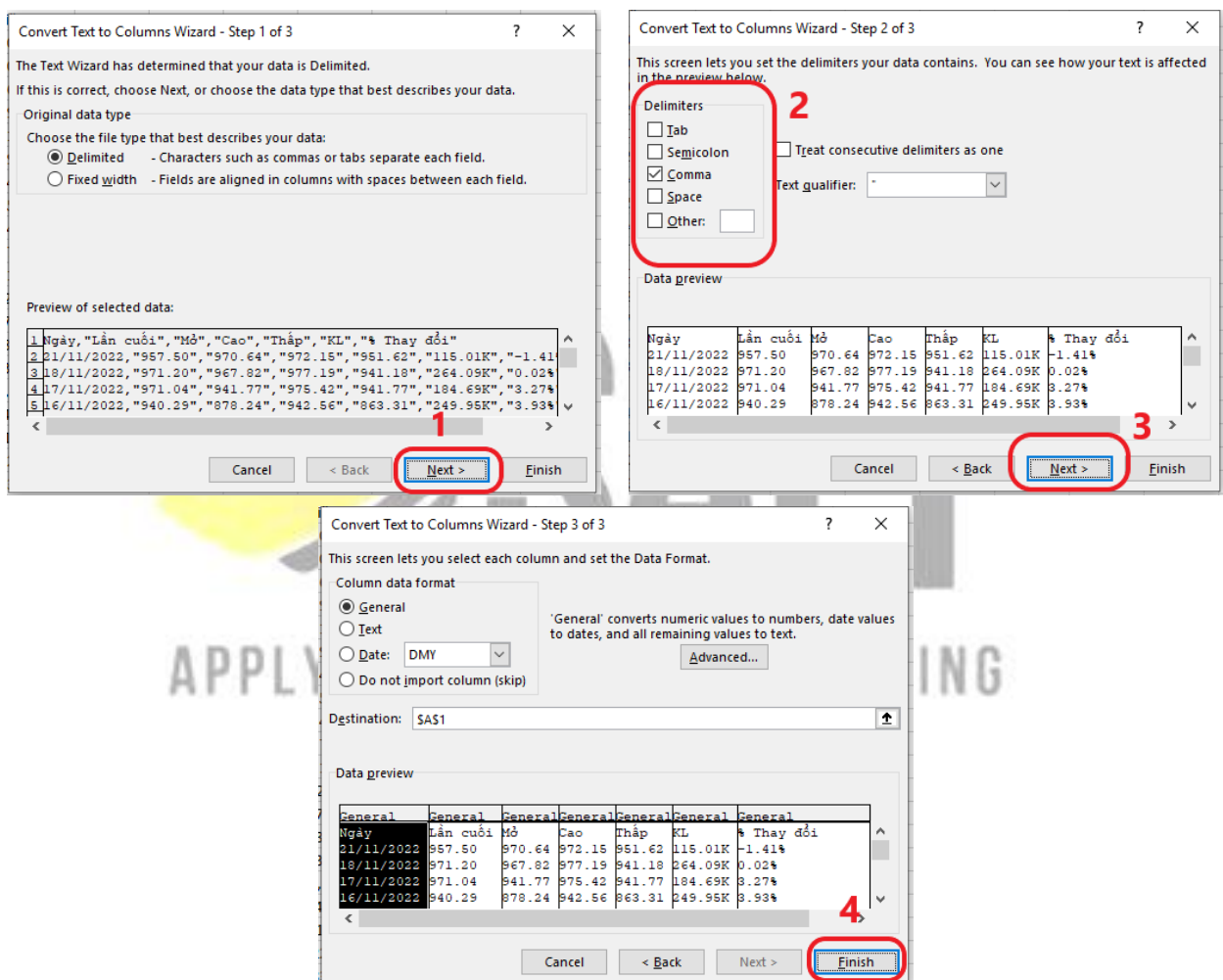


Xuất hiện hộp thoại “**Convert text to columns wizards**”.

**Bước 2:** Trên giao diện hộp thoại “**Convert text to columns wizards**”, thực hiện các thao tác sau:

- Step 1: Chọn **Next**.
- Step 2: Tick vào ô **Comma** và chọn **Next**.
- Step 3: Chọn **Finish**.

Khi đó, bộ dữ liệu sẽ được tách ra thành 7 cột dữ liệu khác nhau.



**Convert Text to Columns Wizard - Step 1 of 3**

The Text Wizard has determined that your data is Delimited.  
If this is correct, choose Next, or choose the data type that best describes your data.

Original data type  
Choose the file type that best describes your data:  
☒ Delimited - Characters such as commas or tabs separate each field.  
☐ Fixed width - Fields are aligned in columns with spaces between each field.

Preview of selected data:

1	Ngày	"Lần cuối"	"Mở"	"Cao"	"Thấp"	"KL"	"% Thay đổi"
2	21/11/2022	"957.50"	"970.64"	"972.15"	"951.62"	"115.01K"	"-1.41%
3	18/11/2022	"971.20"	"967.82"	"977.19"	"941.18"	"264.09K"	"0.02%
4	17/11/2022	"971.04"	"941.77"	"975.42"	"941.77"	"184.69K"	"8.27%
5	16/11/2022	"940.29"	"878.24"	"942.56"	"863.31"	"249.95K"	"3.93%

**1** (Next > button)

**Convert Text to Columns Wizard - Step 2 of 3**

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters  
☐ Tab  
☐ Semicolon  
☒ Comma  
☐ Space  
☐ Other:

**2** (Comma checkbox)

Treat consecutive delimiters as one  
☐

Text qualifier:

Data preview

Ngày	Lần cuối	Mở	Cao	Thấp	KL	% Thay đổi
21/11/2022	957.50	970.64	972.15	951.62	115.01K	-1.41%
18/11/2022	971.20	967.82	977.19	941.18	264.09K	0.02%
17/11/2022	971.04	941.77	975.42	941.77	184.69K	8.27%
16/11/2022	940.29	878.24	942.56	863.31	249.95K	3.93%

**3** (Next > button)

**Convert Text to Columns Wizard - Step 3 of 3**

This screen lets you select each column and set the Data Format.

Column data format  
☒ General  
☐ Text  
☐ Date:  DMY  
☐ Do not import column (skip)

'General' converts numeric values to numbers, date values to dates, and all remaining values to text.

Destination:  \$A\$1

Data preview

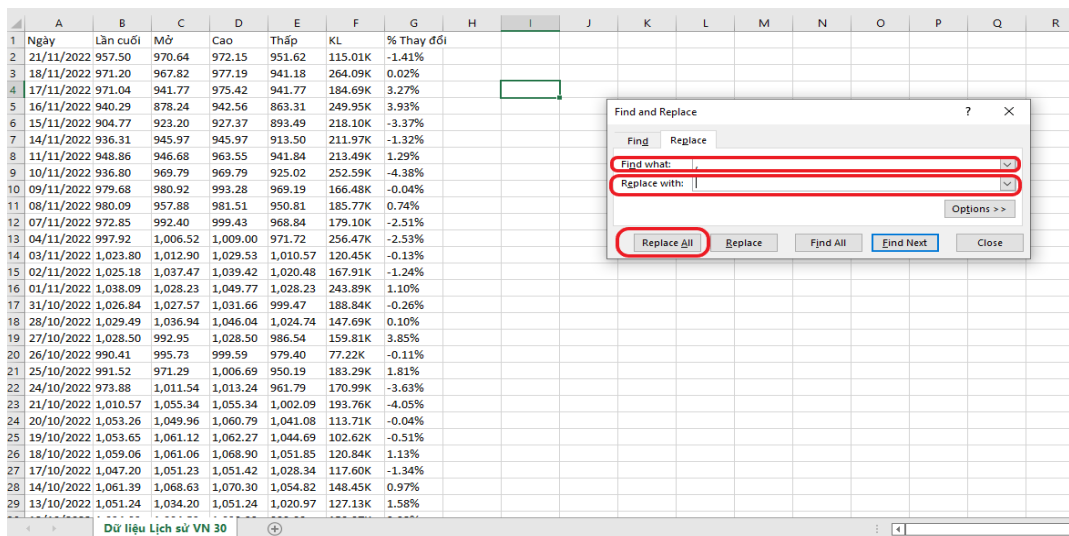
General	General	General	General	General	General	General
Ngày	Lần cuối	Mở	Cao	Thấp	KL	% Thay đổi
21/11/2022	957.50	970.64	972.15	951.62	115.01K	-1.41%
18/11/2022	971.20	967.82	977.19	941.18	264.09K	0.02%
17/11/2022	971.04	941.77	975.42	941.77	184.69K	8.27%
16/11/2022	940.29	878.24	942.56	863.31	249.95K	3.93%

**4** (Finish button)

**Bước 3:** Nhấn tổ hợp phím **Ctrl + H** để xuất hiện hộp thoại **“Find and Replace”**.

**Bước 4:** Chọn thẻ **Replace** và thực hiện các thao tác sau:

- Step 1: **Find what** nhập **“,”** và **Replace with** để trống > **Replace All**.
- Step 2: **Find what** nhập **“.”** và **Replace with** nhập **“,”** > **Replace All**.



**Bước 5:** Tiến hành copy cột dữ liệu **“Ngày”**, **“Lần cuối”** và **“% Thay đổi”** qua một sheet mới và trình bày theo mẫu dưới đây.

	A	B	C	D	E	F	G
1	<b>Dữ liệu lịch sử của các Input đầu vào</b>						
2							
3	<b>Ngày</b>	<b>Dữ liệu giá của MSN (ĐVT: Nghìn VNĐ)</b>		<b>Dữ liệu lịch sử chỉ số VN30</b>			
4		<b>Lần cuối</b>	<b>% Thay đổi</b>	<b>Lần cuối</b>	<b>% Thay đổi</b>		
5	21/11/2022	94,30	-0,95%	957,50	-1,41%		
6	18/11/2022	95,20	-1,86%	971,20	0,02%		
7	17/11/2022	97,00	6,71%	971,04	3,27%		
8	16/11/2022	90,90	1,68%	940,29	3,93%		
9	15/11/2022	89,40	0,34%	904,77	-3,37%		
10	14/11/2022	89,10	3,01%	936,31	-1,32%		
11	11/11/2022	86,50	6,79%	948,86	1,29%		
12	10/11/2022	81,00	-6,90%	936,80	-4,38%		
13	09/11/2022	87,00	1,16%	979,68	-0,04%		
14	08/11/2022	86,00	1,78%	980,09	0,74%		
15	07/11/2022	84,50	0,24%	972,85	-2,51%		
16	04/11/2022	84,30	-1,29%	997,92	-2,53%		
17	03/11/2022	85,40	5,43%	1023,80	-0,13%		
18	02/11/2022	81,00	-6,36%	1025,18	-1,24%		
19	01/11/2022	86,50	1,17%	1038,09	1,10%		
20	31/10/2022	85,50	1,91%	1026,84	-0,26%		