



FPT POLYTECHNIC



caodang.fpt.edu.vn

NHẬP MÔN XỬ LÝ DỮ LIỆU

BÀI 3: LÀM SẠCH DỮ LIỆU

MỤC TIÊU

- ◎ HIỂU ĐƯỢC CÁC KỸ THUẬT LÀM SẠCH DỮ LIỆU
- ◎ ÁP DỤNG CÁC HÀM ĐỂ LÀM SẠCH DỮ LIỆU
- ◎ KỸ THUẬT LÀM SẠCH DỮ LIỆU TỰ ĐỘNG

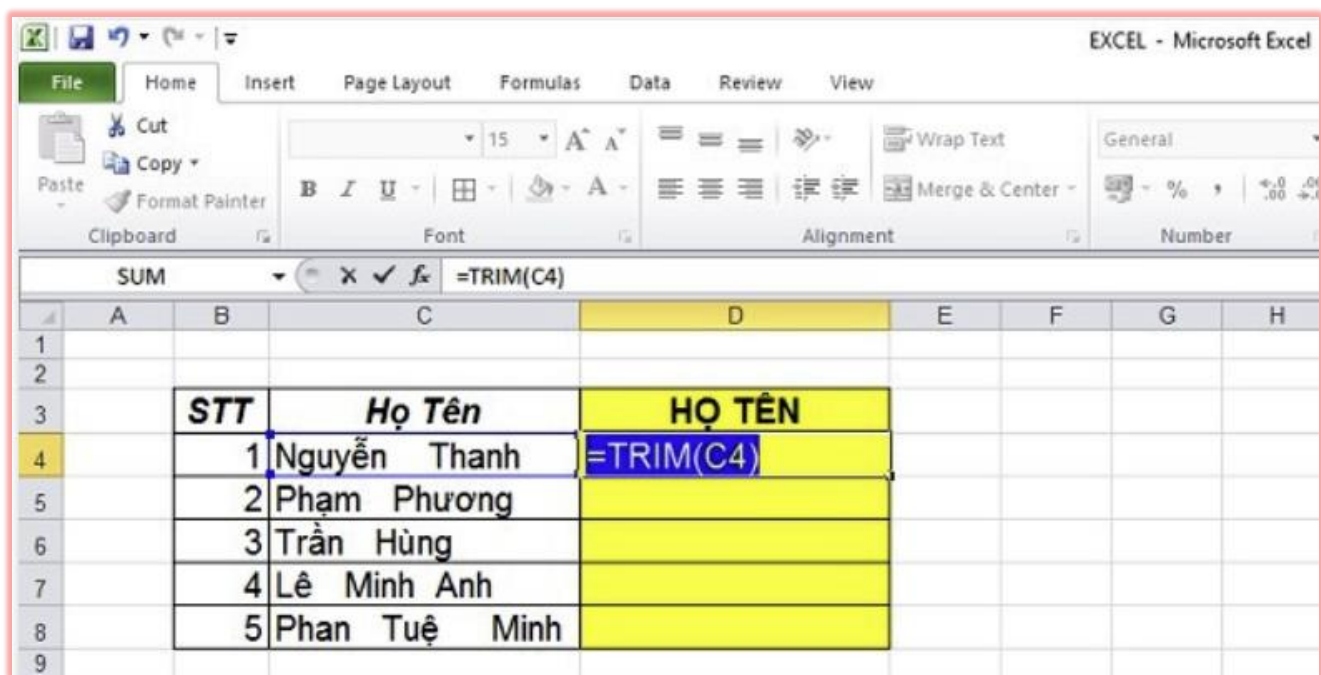


- ☐ CÁC HÀM TRONG EXCEL LÀM SẠCH DỮ LIỆU
- ☐ SỬ DỤNG MACROS LÀM SẠCH DỮ LIỆU TỰ ĐỘNG



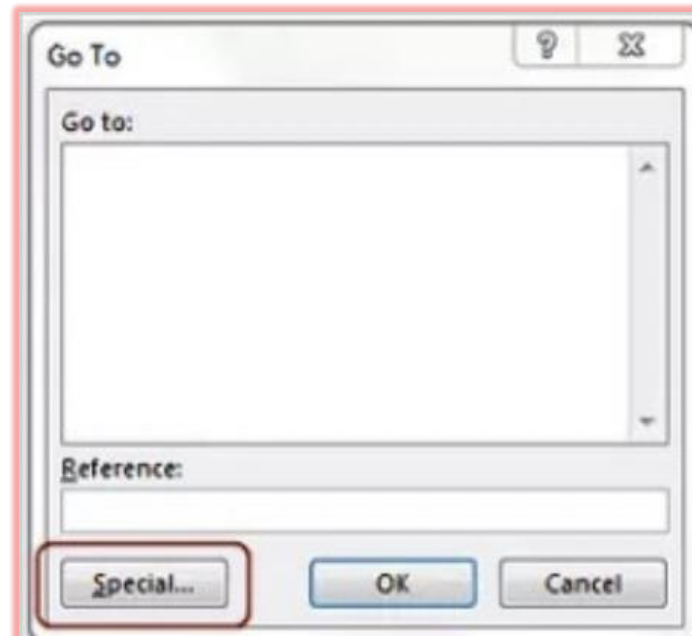
PHẦN I: LÀM SẠCH DỮ LIỆU

- ❖ Hàm **TRIM** trong Excel dùng để loại bỏ các khoảng trống ra khỏi văn bản, chỉ để lại một khoảng trống giữa các từ.



❑ Xử lý tất cả các ô trống:

- ❖ Bước 1: Chọn toàn bộ tập dữ liệu.
- ❖ Bước 2: Nhấn **F5** để mở hộp thoại **Go to**.
- ❖ Bước 3: Nhấp vào Special... để mở hộp thoại **Go To Special**.



❑ Xử lý tất cả các ô trống:

❖ Bước 4: Chọn tùy chọn Blanks > OK.

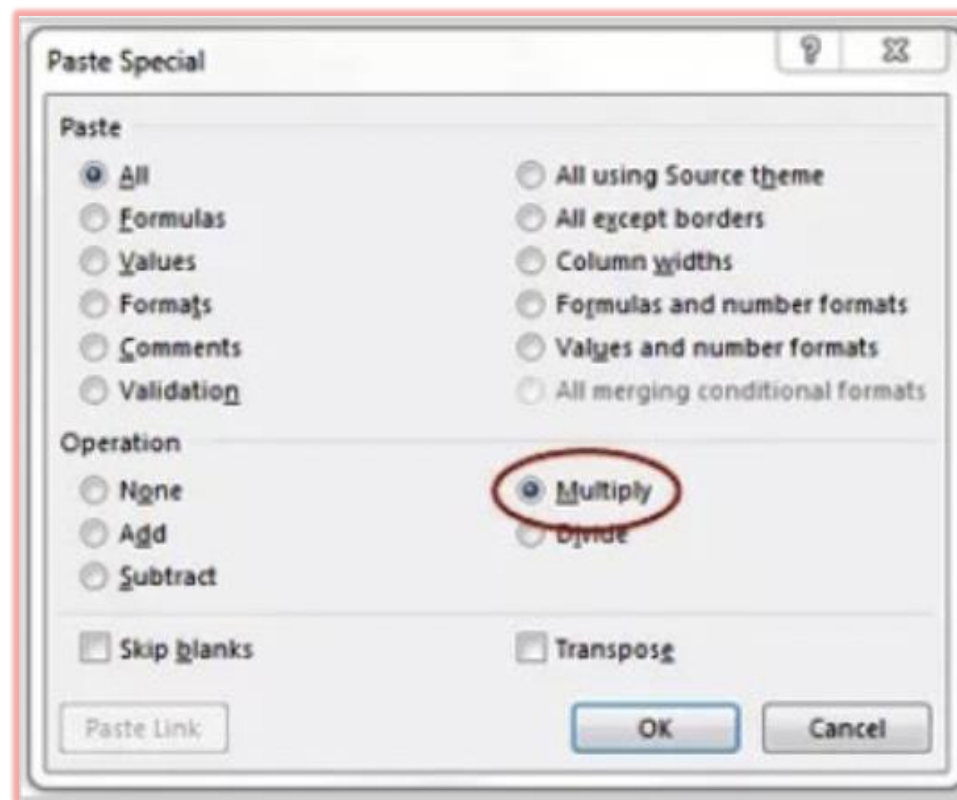


❑ Chuyển đổi số ở dạng văn bản thành số:

- ❖ Bước 1: Trong bất kỳ ô trống nào, hãy nhập 1.
- ❖ Bước 2: Chọn ô mà bạn đã nhập 1 và nhấn **Ctrl + C**.
- ❖ Bước 3: Chọn ô/dải ô mà bạn muốn chuyển đổi thành số.
- ❖ Bước 4: Chọn **Paste** và **Paste Special**.
- ❖ Bước 5: Trong hộp thoại Paste Special, chọn tùy chọn **Multiply**.

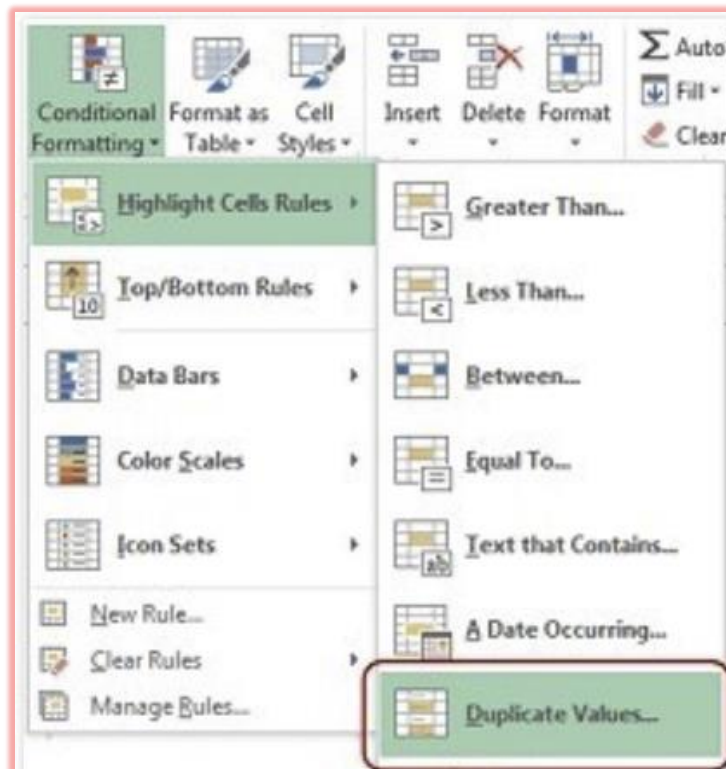
❑ Chuyển đổi số ở dạng văn bản thành số:

- ❖ Bước 6: Nhấn OK chuyển đổi tất cả các số ở định dạng văn bản trở lại số.



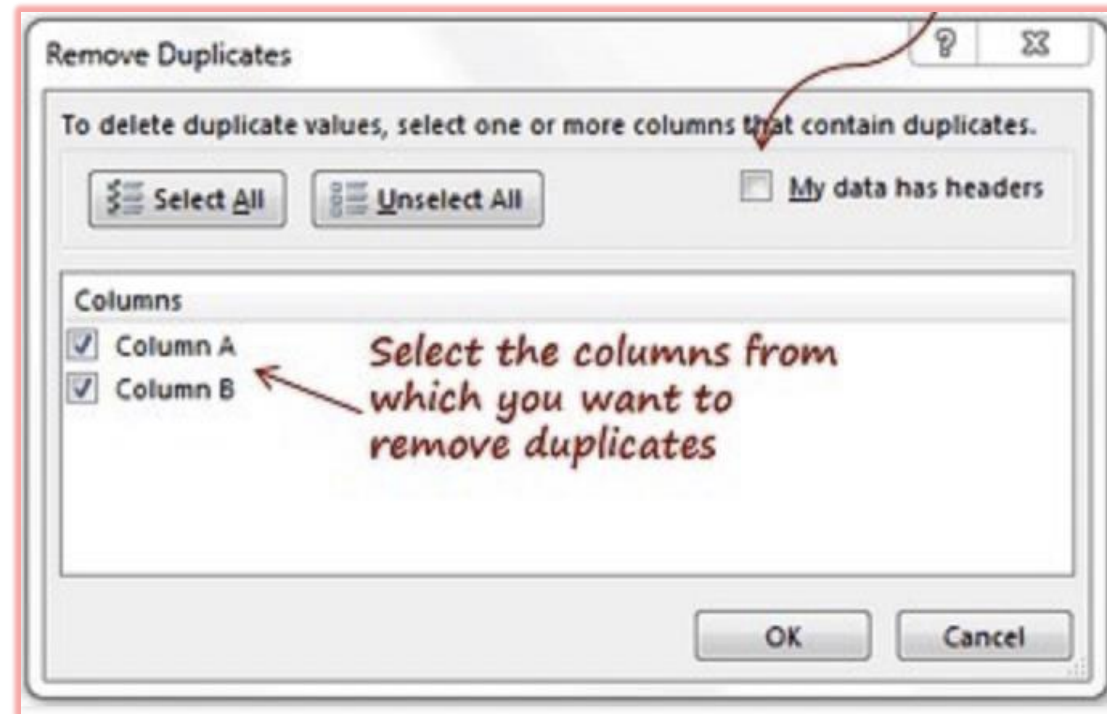
❑ Xóa các bản sao:

- ❖ Để đánh dấu dữ liệu trùng lặp: Chọn dữ liệu và truy cập vào công cụ **Home > Conditional Formatting > Highlight Cells Rules > Duplicate Values.**



❑ Xóa các bản sao:

- ❖ Chọn dữ liệu và truy cập vào công cụ **Data > Remove Duplicates**.
- ❖ Nếu dữ liệu có tiêu đề, chọn **My data has headers**.



❑ Đánh dấu lỗi:

- ❖ Bước 1: Chọn toàn bộ tập dữ liệu.
- ❖ Bước 2: Nhấn **F5** để mở hộp thoại **Go to**.
- ❖ Bước 3: Chọn **Special** ở dưới cùng bên trái.
- ❖ Bước 4: Chọn **Formulas** và bỏ chọn tất cả các tùy chọn ngoại trừ **Errors**.



❑ Đổi văn bản thành chữ thường/hoa:

- ❖ **LOWER ()** – Chuyển đổi tất cả văn bản thành **chữ thường**.
- ❖ **UPPER ()** – chuyển đổi tất cả văn bản thành **chữ hoa**.

Change Case in Excel

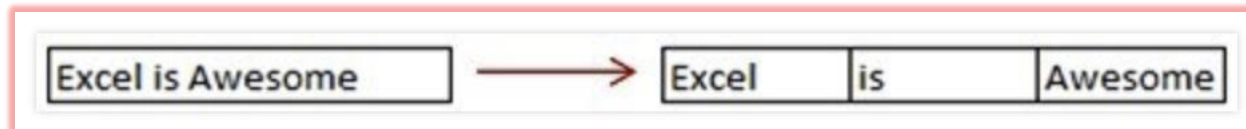
B	A	B	C
MONTH	MONTH	month	
JAN	JAN	jan	
FEB	FEB	feb	
MAR	MAR	mar	
APR	APR	apr	
MAY	MAY	may	
JUN	JUN	iun	

=UPPER()
UPPER(text)
Converts a text string to all uppercase letters

=LOWER()
LOWER(text)
Converts all letters in a text string to lowercase

❑ Phân tích cú pháp dữ liệu với tính năng Text to Column:

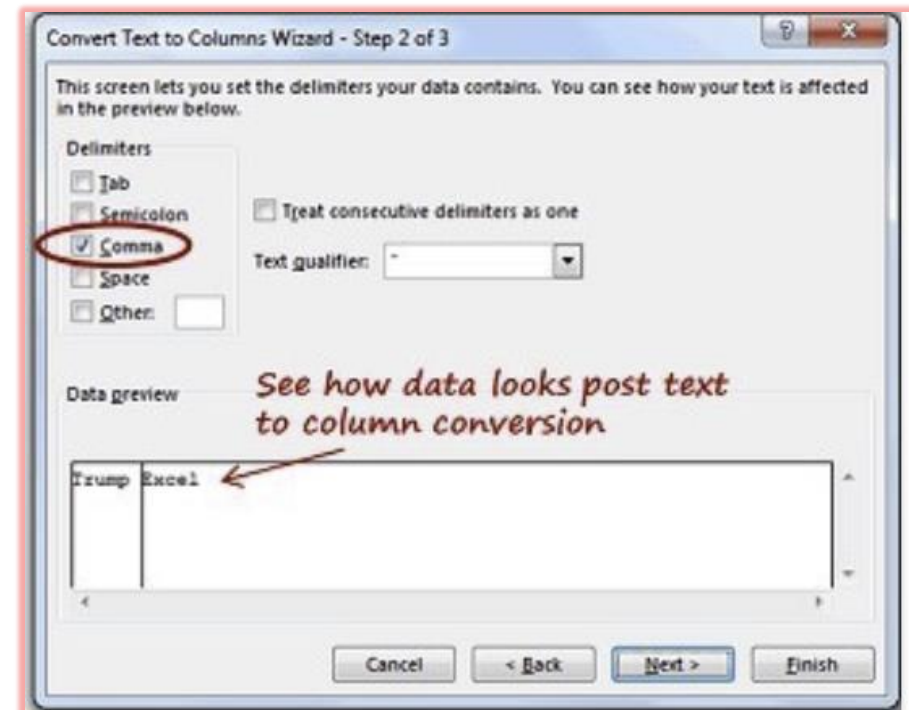
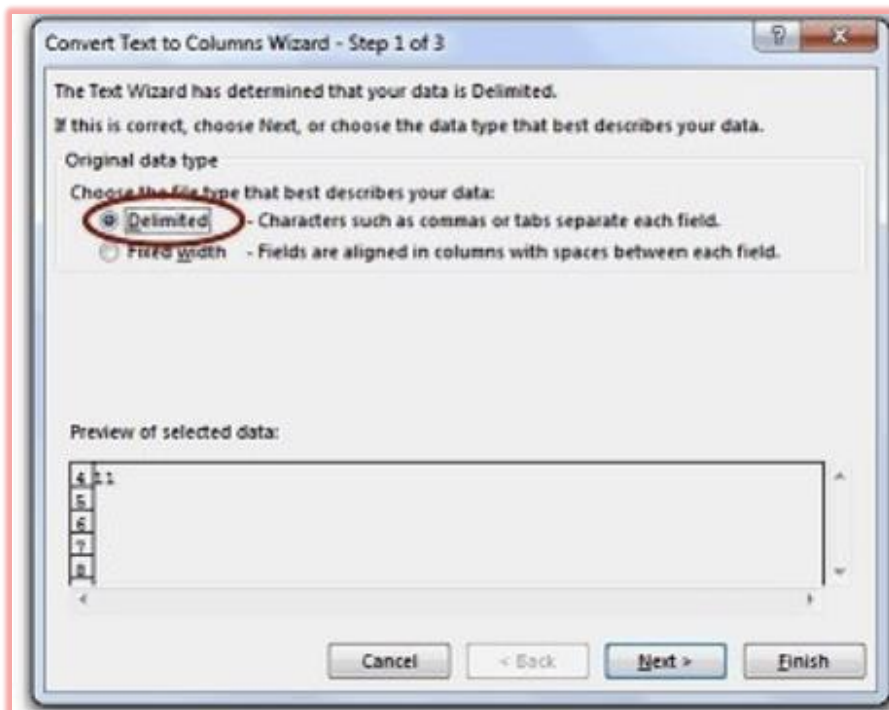
- ❖ Khi lấy dữ liệu từ cơ sở dữ liệu hoặc nhập dữ liệu từ tệp văn bản, có thể xảy ra trường hợp tất cả văn bản nằm trong một ô.
- ❖ Trong trường hợp này, phân tích cú pháp văn bản này thành nhiều ô bằng tính năng **Text to Column** trong Excel.



❑ Phân tích cú pháp dữ liệu với tính năng Text to Column:

- ❖ Bước 1: Chọn dữ liệu/văn bản bạn muốn phân tích cú pháp.
- ❖ Bước 2: Truy cập vào công cụ **Data > Text to Column** để mở trình hướng dẫn Text to Columns Wizard:
 - 2.1 Chọn kiểu dữ liệu (chọn Delimited nếu dữ liệu không có khoảng cách đều nhau và được phân tách bằng các ký tự như dấu phẩy, dấu gạch ngang, dấu chấm ..) > Nhấn **Next**.

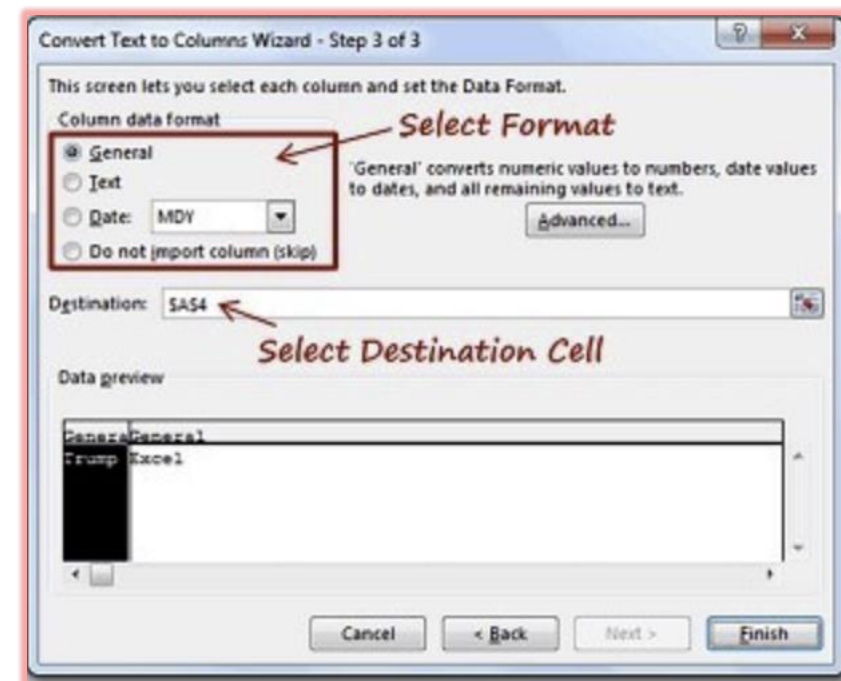
- ❑ Phân tích cú pháp dữ liệu với tính năng Text to Column:
- 2.2: Chọn **Delimiter**. Có thể chọn dấu phân cách được xác định trước hoặc bất kỳ thứ gì khác bằng cách sử dụng tùy chọn **Other**.



❑ Phân tích cú pháp dữ liệu với tính năng Text to Column:

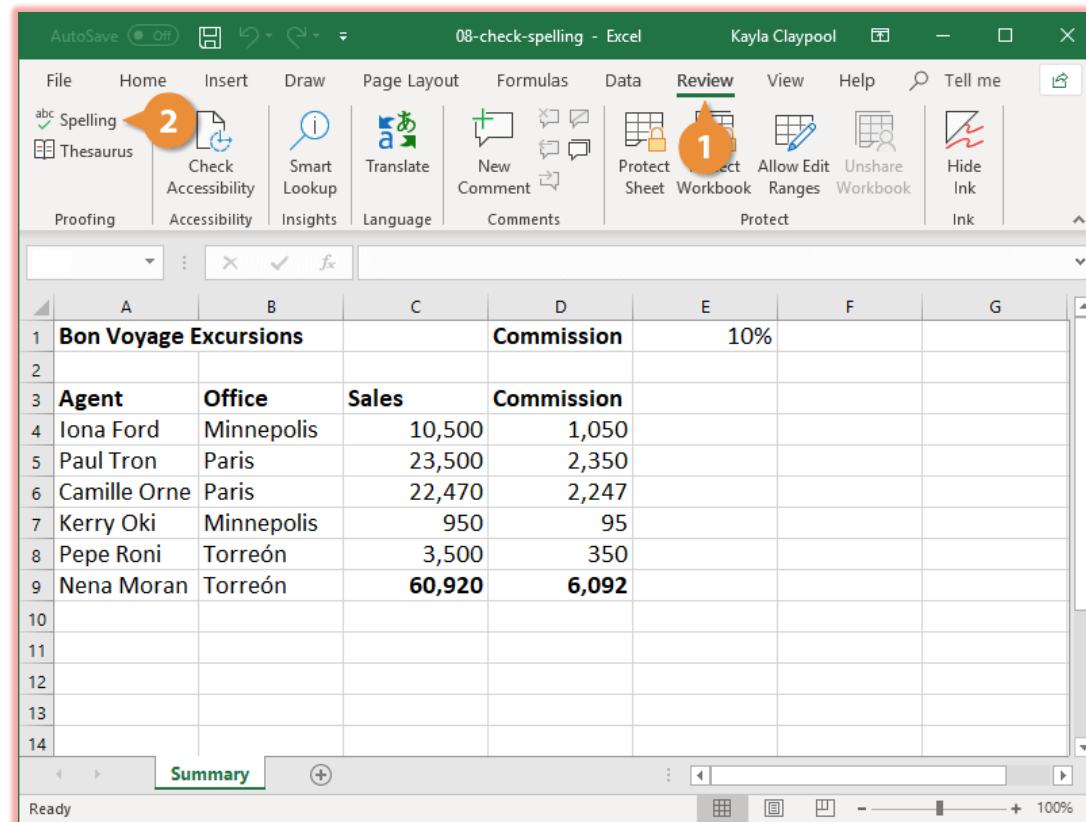
Chọn định dạng dữ liệu > chọn ô đích. Nếu ô đích không được chọn, ô hiện tại sẽ bị ghi đè lên.

❖ Bước 3: Bấm nút **Finish** để kết thúc.



❑ Kiểm tra lỗi chính tả:

- ❖ Kiểm tra lỗi chính tả hoặc sai sót trong tập dữ liệu, sử dụng phím tắt **F7**.

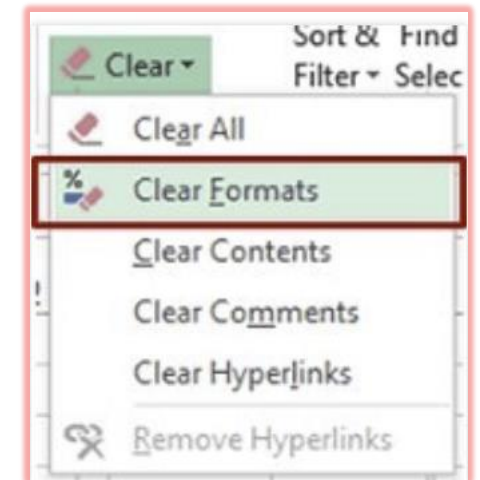




PHẦN II: LÀM SẠCH DỮ LIỆU (TT)

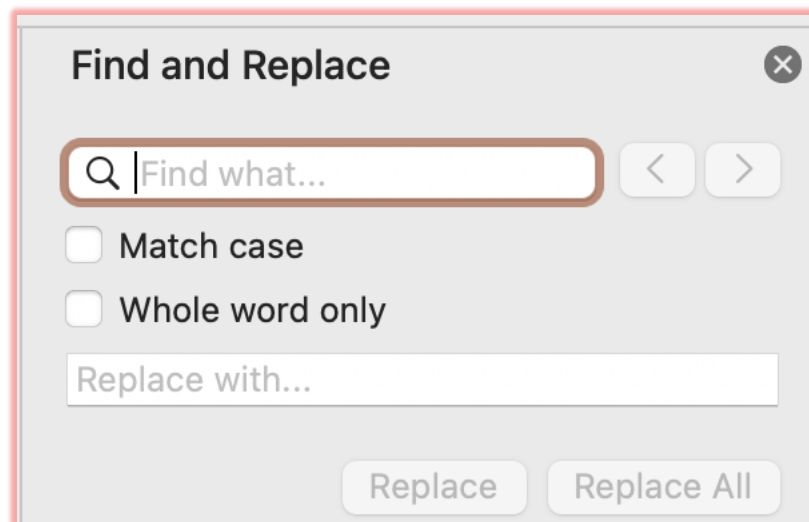
❑ Xóa hết các định dạng:

- ❖ Nếu dùng nhiều cơ sở dữ liệu để lấy dữ liệu, **format** không đồng nhất.
- ❖ Bởi mỗi cơ sở dữ liệu đều có định dạng khác nhau nên khi nhập vào sẽ rất lộn xộn.
 - Bước 1: Chọn tập dữ liệu.
 - Bước 2: Chọn công cụ **Home > Clear > Clear Format**.



❑ Sử dụng tính năng Find and Replace:

- ❖ **Find and Replace** đóng vai trò làm sạch dữ liệu Excel.
- ❖ Có thể chọn và loại bỏ tất cả các số 0.
- ❖ Thay đổi các tham chiếu trong công thức, tìm và thay đổi định dạng.



□ Tự động làm sạch dữ liệu:

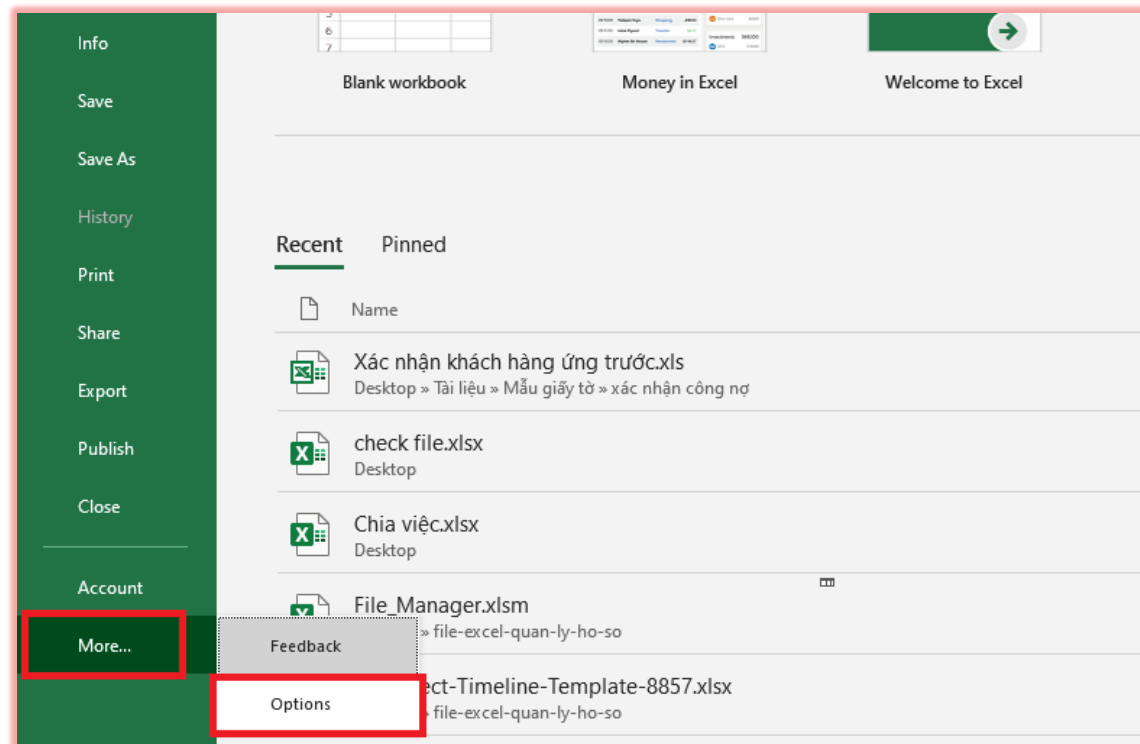
- ❖ Dữ liệu hàng ngày đều phải lặp đi lặp lại công việc chỉnh sửa format cho dữ liệu thô?
- ❖ Thao tác này sẽ rất tốn thời gian phải không?
- ❖ Excel tự động lặp lại những thao tác trước đó trên cùng một file hay khác file.

**MACRO EXCEL
LÀM SẠCH DỮ LIỆU**

TỰ ĐỘNG

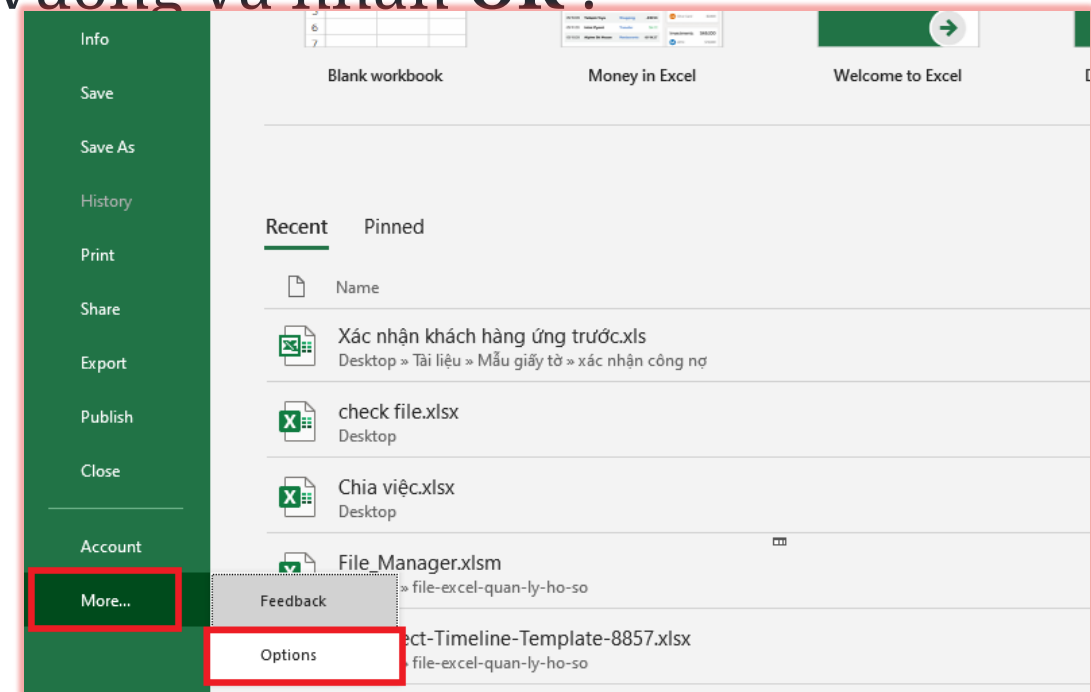
❑ Tự động làm sạch dữ liệu:

- ❖ **Bước 1:** Mở file Excel chứa dữ liệu cần làm sạch lên. Sau đó => chọn thẻ **File** trên thanh công cụ. Tiếp theo, chọn mục **Options**.



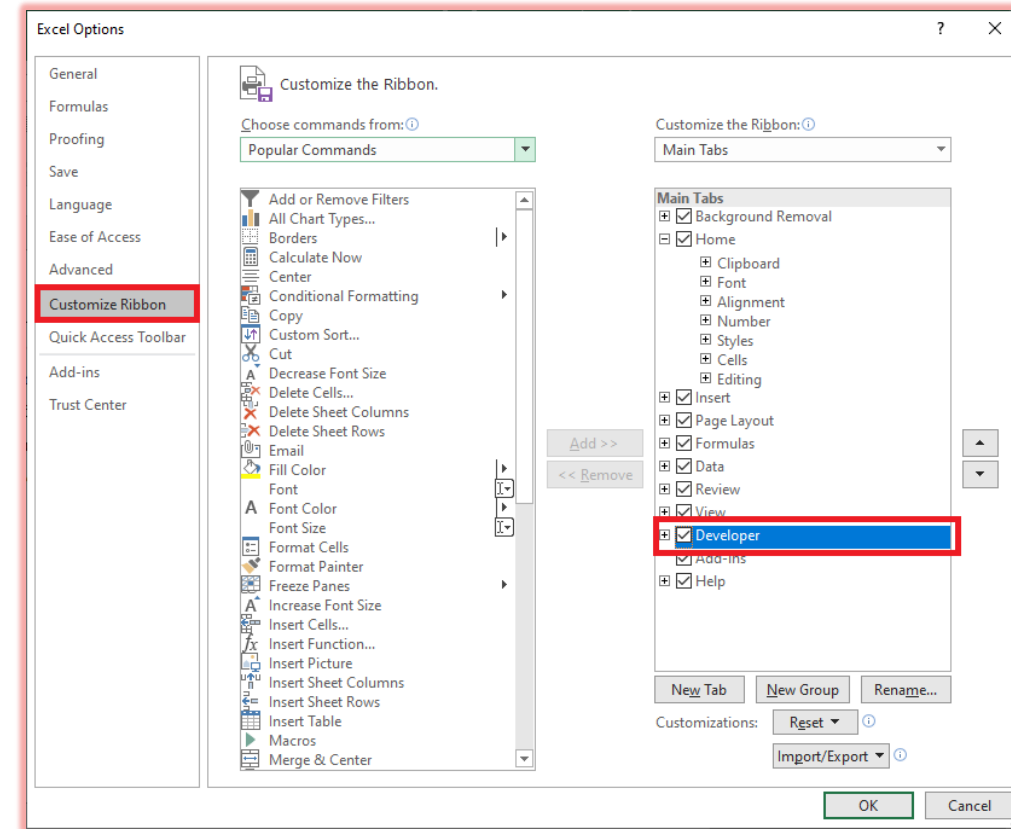
□ Tự động làm sạch dữ liệu:

- ❖ **Bước 2 (TT):** Cửa sổ Excel Options hiện ra => chọn thẻ **Customize Ribbon**. Sau đó tại mục **Main Tabs** tìm tới mục **Developer** tích chọn vào ô vuông và nhấn **OK**.



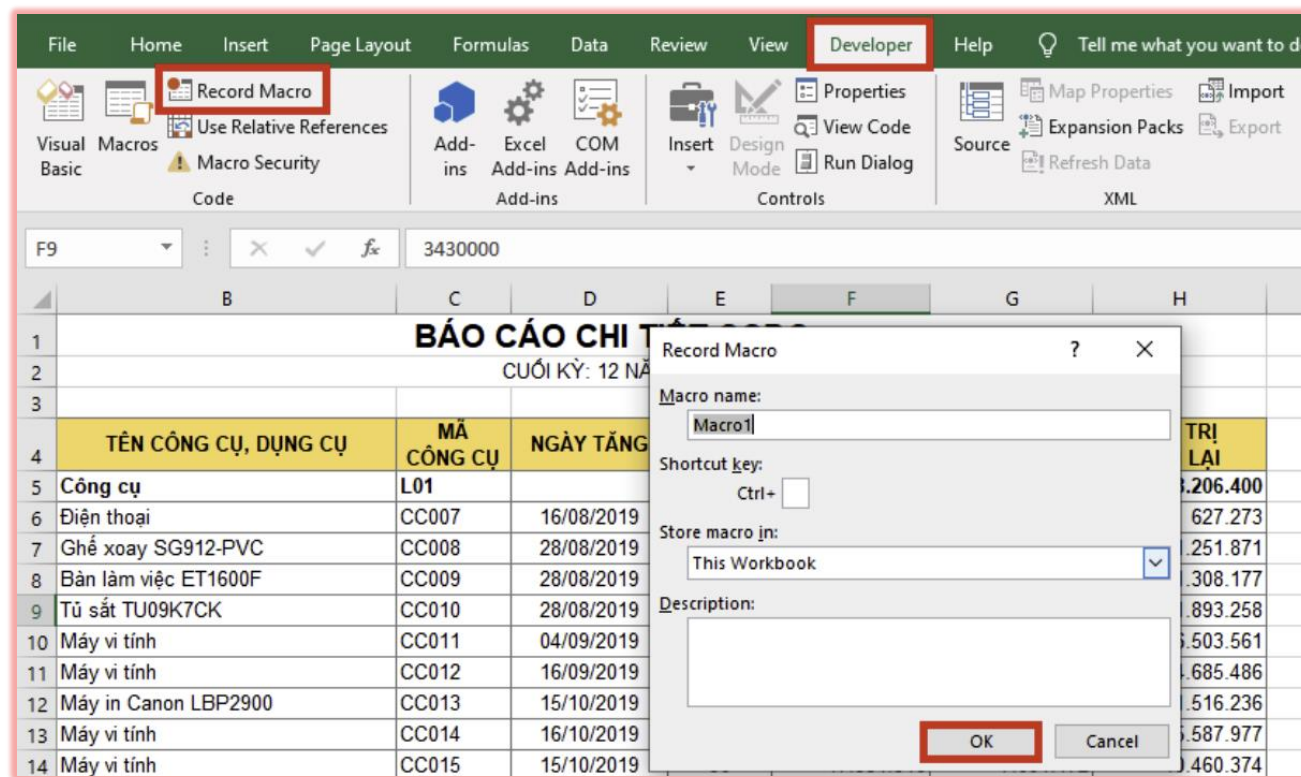
❑ Tự động làm sạch dữ liệu:

- ❖ **Bước 2:** Cửa sổ Excel Options hiện ra => chọn thẻ **Customize Ribbon**. Sau đó tại mục **Main Tabs** tìm tới mục **Developer** tích chọn vào ô vuông và nhấn **OK**.



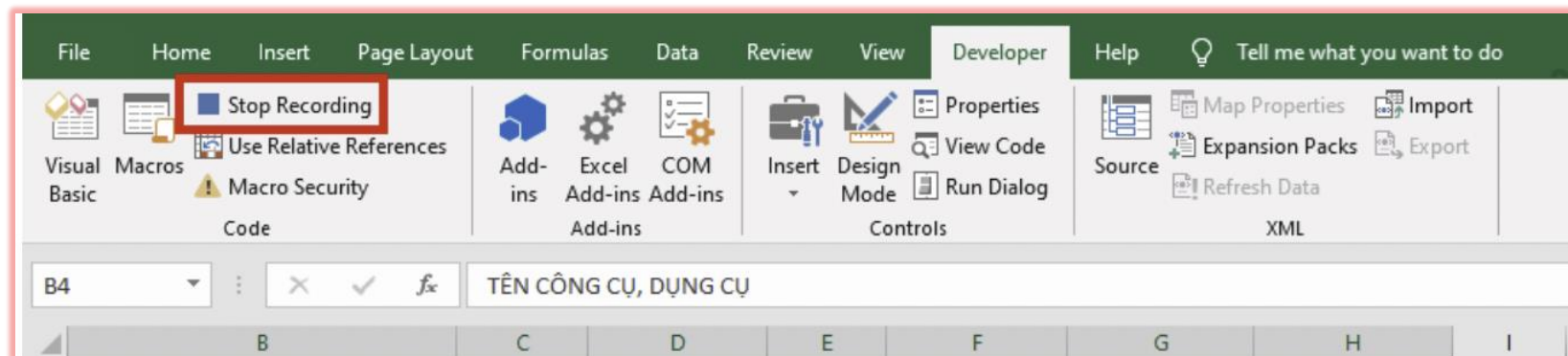
❑ Tự động làm sạch dữ liệu:

- ❖ **Bước 3:** Thẻ Developer đã xuất hiện trên thanh công cụ. Chọn thẻ **Developer**. Sau đó chọn mục **Record Macro** và nhấn OK.



□ Tự động làm sạch dữ liệu:

- ❖ **Bước 4:** Tiến thành chỉnh sửa dữ liệu theo ý muốn như chỉnh sửa kiểu chữ, size chữ, tô màu nền hay là sắp xếp theo thứ tự... Sau khi chỉnh sửa xong, nhấn chọn mục **Stop Recording** để dừng việc tạo **Record Macro**.



❑ Tự động làm sạch dữ liệu:

- ❖ **Bước 5:** Sau khi tạo Macro xong, có thể sử dụng nó để chỉnh sửa dữ liệu một cách nhanh chóng bằng cách chọn dữ liệu cần **format**.
- ❖ **Bước 6:** Chọn thẻ **Developer** => **Macros** => chọn Macro vừa tạo ở trên, sau đó nhấn **Run**. Dữ liệu sẽ tự động cập nhật theo định dạng

The screenshot shows an Excel spreadsheet with a table of data and a 'Macro' dialog box open. The table has columns 'Mã hàng' and 'Tên hàng'. The 'Macro' dialog box shows 'Macro1' selected in the list, and the 'Run' button is highlighted with a red box.

Mã hàng	Tên hàng
0501004001	Đá Thạch Anh Bellingham 2cm
0501010001	Đá Thạch Anh Harlech 2cm
0501012001	Đá Thạch Anh Kelvingrove 2cm
0501017001	Đá Thạch Anh Princetown 2cm

Macro dialog box:

- Macro name: Macro1
- Buttons: Run, Step Into, Edit, Create, Delete



- ✓ Áp dụng các hàm trong Excel để làm sạch dữ liệu
- ✓ 6 bước lọc dữ liệu tự động bằng Macros.



FPT POLYTECHNIC

Thank you