# KẾT NỐI DỮ LIỆU VÀ XỬ LÝ DỮ LIỆU TRONG POWERBI

# KẾT NỐI DỮ LIỆU VÀ XỬ LÝ DỮ LIỆU TRONG POWER BI

01. Các nguồn dữ liệu

02. Thao tác với Power BI Query

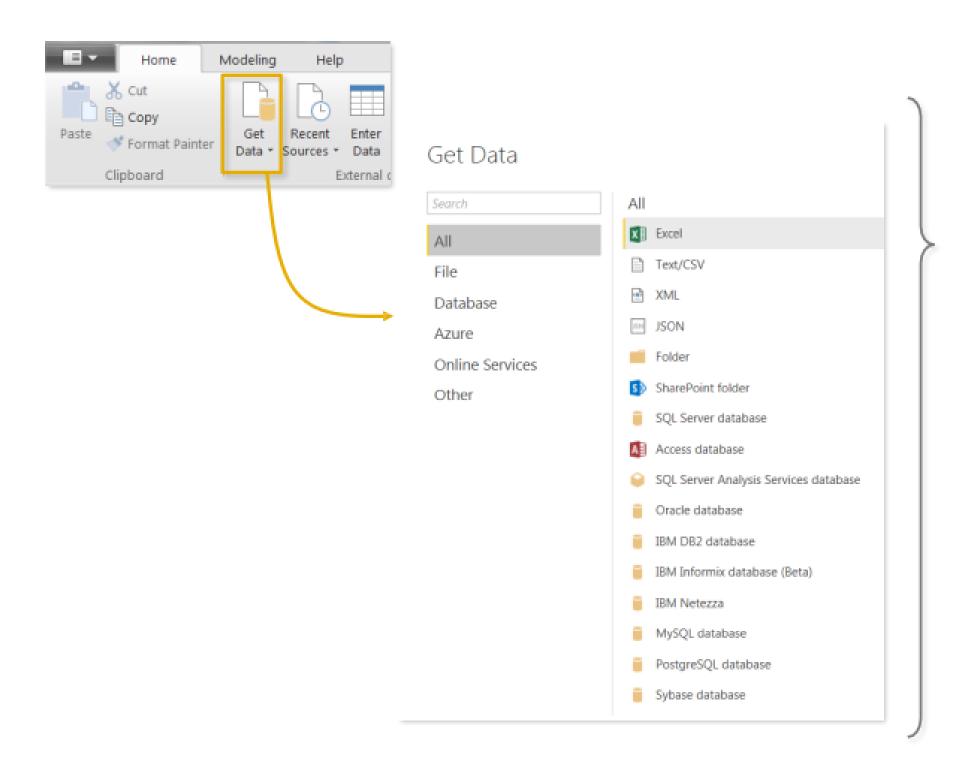
03. Các chức năng xử lý dữ liệu

04. Các chức năng khác

## 01. Các nguồn dữ liệu



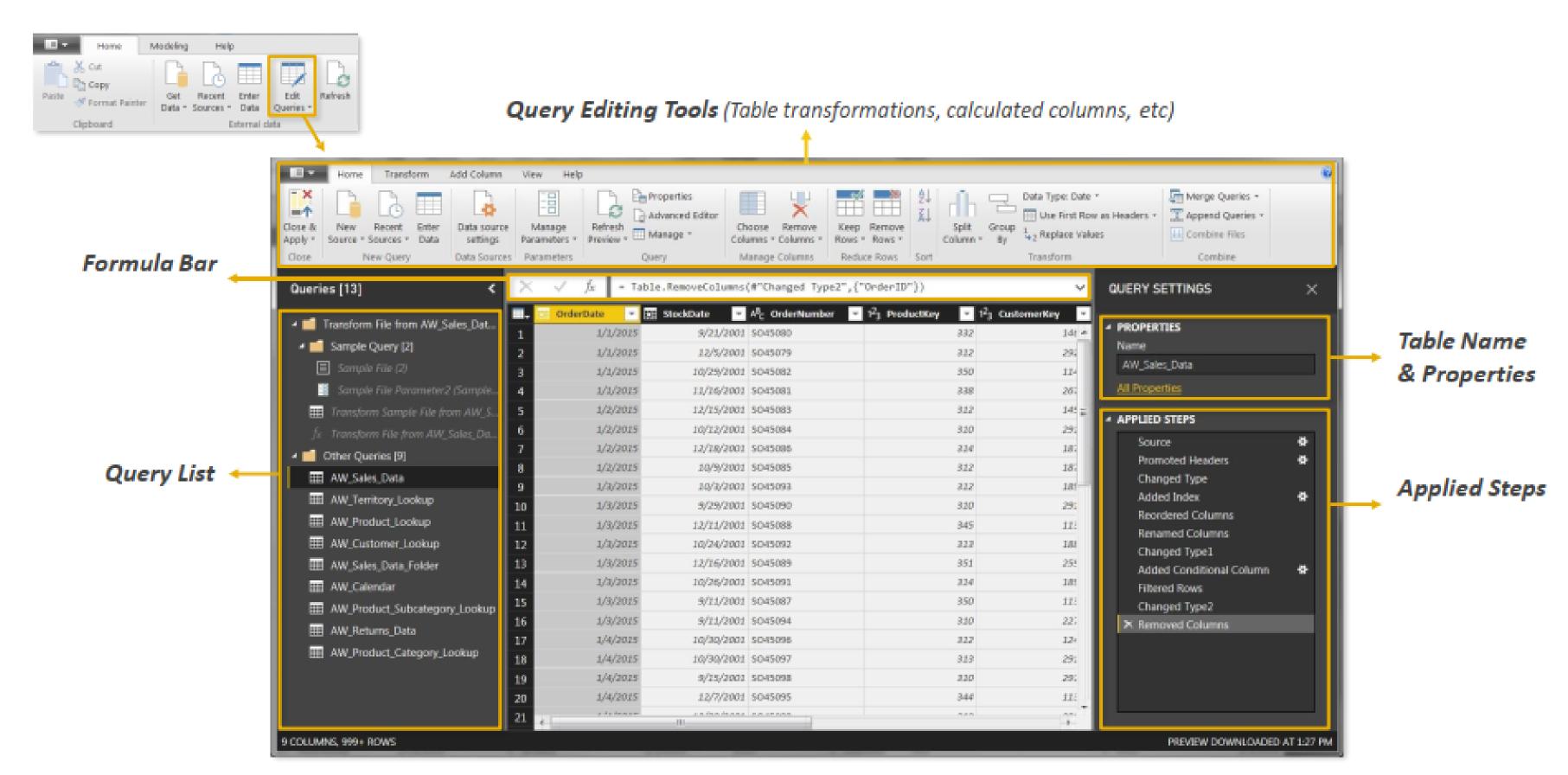
### Lấy Data từ nhiều nguồn dữ liệu



Power BI có thể kết nối tới rất nhiều nguồn data, bao gồm (nhưng không hạn chế):

- Flat files & Folders (csv, text, xls, etc)
- **Databases** (SQL, Access, Oracle, IBM, Azure, etc)
- Online Services (Sharepoint, GitHub,
   Dynamics 365, Google Analytics,
   Salesforce, Power BI Service, etc)
- Others (Web feeds, R scripts, Spark, Hadoop, etc)

# **Query Editor**



Data Sources

Parameters

Query



Close

#### **Query Editing Tools**

New Query

Tab Home Transform Add Column View Help

Home Transform Add Column View Help

Close & New Recent Enter Data source Manage Refresh Apply \* Source \* Source \* Data Source settings Parameters \* Preview \* Manage \* Rows \* Row

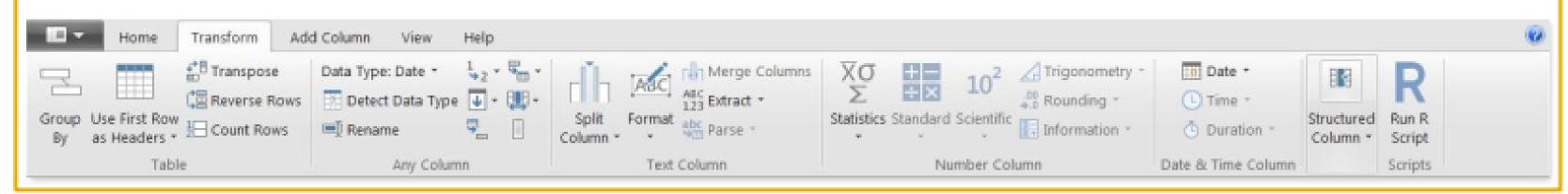
Manage Columns

Reduce Rows | Sort

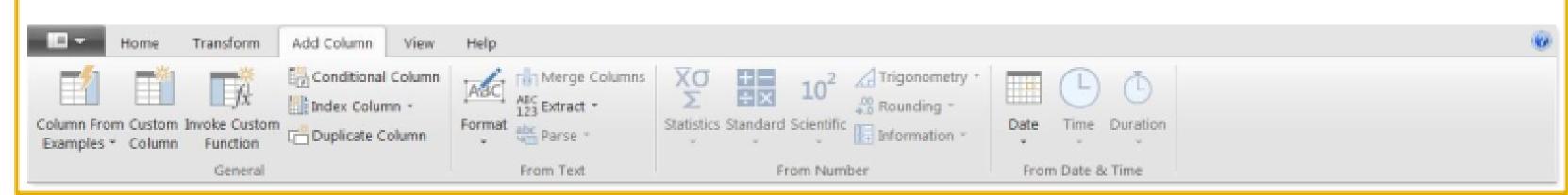
Transform

Combine

-Tab TRANSFORM bao bồm các công cụ chỉnh sửa các **bảng sẵn có đã tồn tại** (chia/gộp, chỉnh sửa, nhóm xử lý text ...)

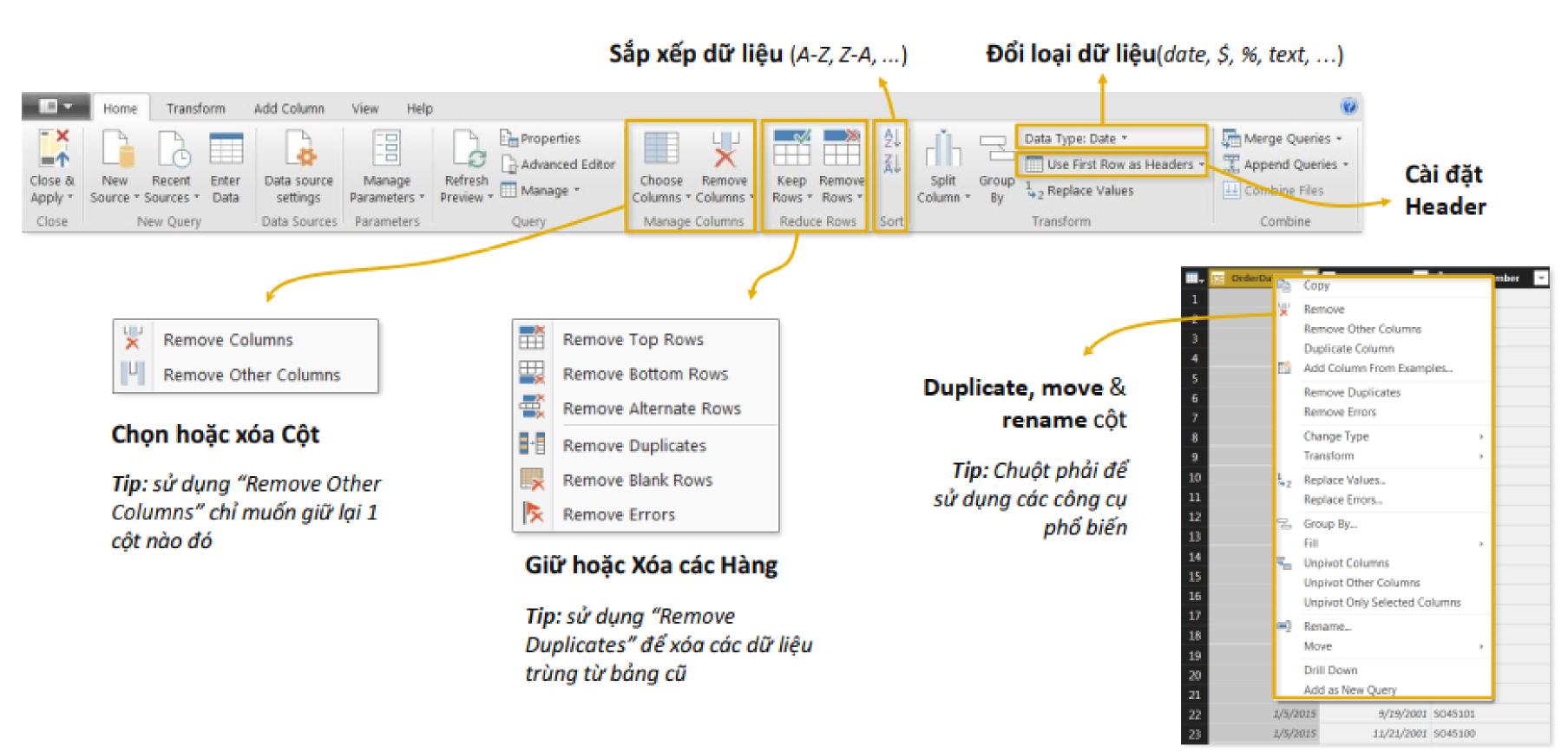


–Tab ADD COLUMN bao gồm các công cụ **tạo bảng mới** (dựa trên một điều kiện, tính toán, công thức, ngày ...)



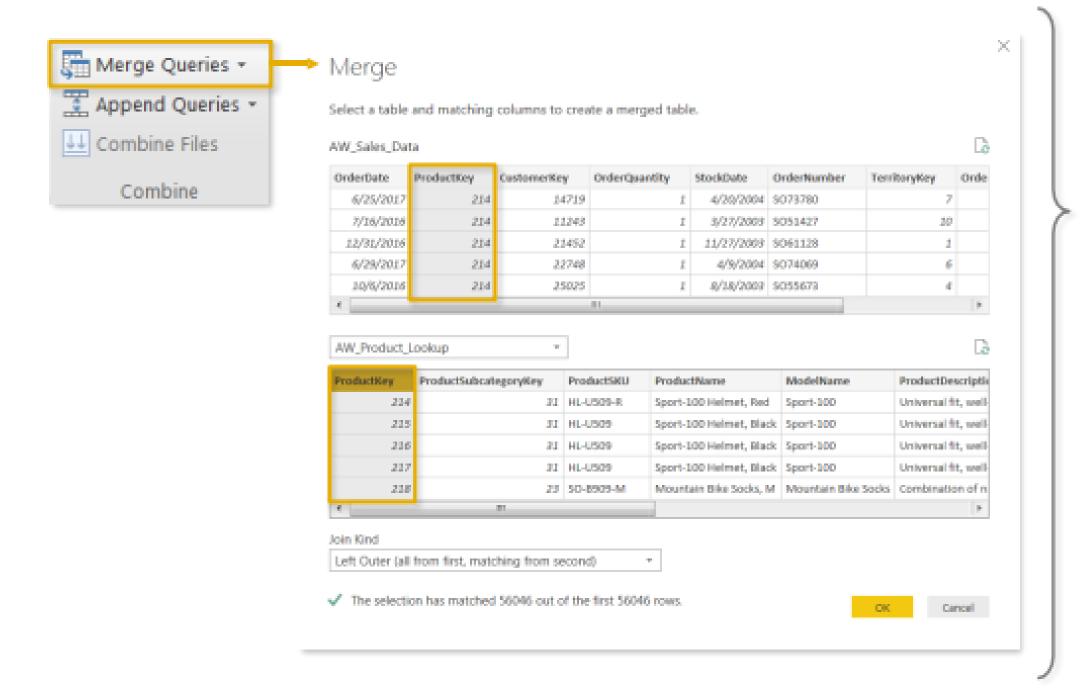
# 9

#### Chỉnh sửa bảng cơ bản



## 9

#### **Merging Queries**



Merging queries cho phép bạn **gộp bảng** dựa trên những cột chung (giống hàm VLOOKUP trong Excel)

**NOTE:** Merging queries **thêm cột** tới một bảng đã có sẵn

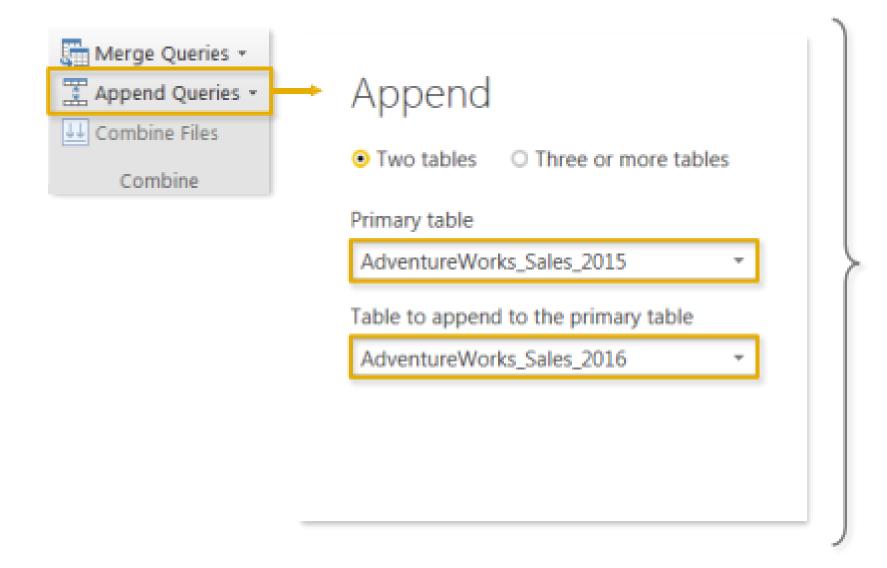
#### NOTE!

Không bởi vì chúng ta có thể gộp bảng, mà chúng ta luôn gộp để có số lượng bảng ít.

Thực tế trong PBI, sẽ tốt hơn nếu chia ra thành các bảng và tạo relationships cho chúng(chi tiết hơn ở phần sau khóa học!)



## **Appending Queries**



Appending queries cho phép bạn combine tables có cùng kiểu cấu trúc và kiểu định dạng

NOTE: Appending thêm dòng tới một bảng đã có sẵn

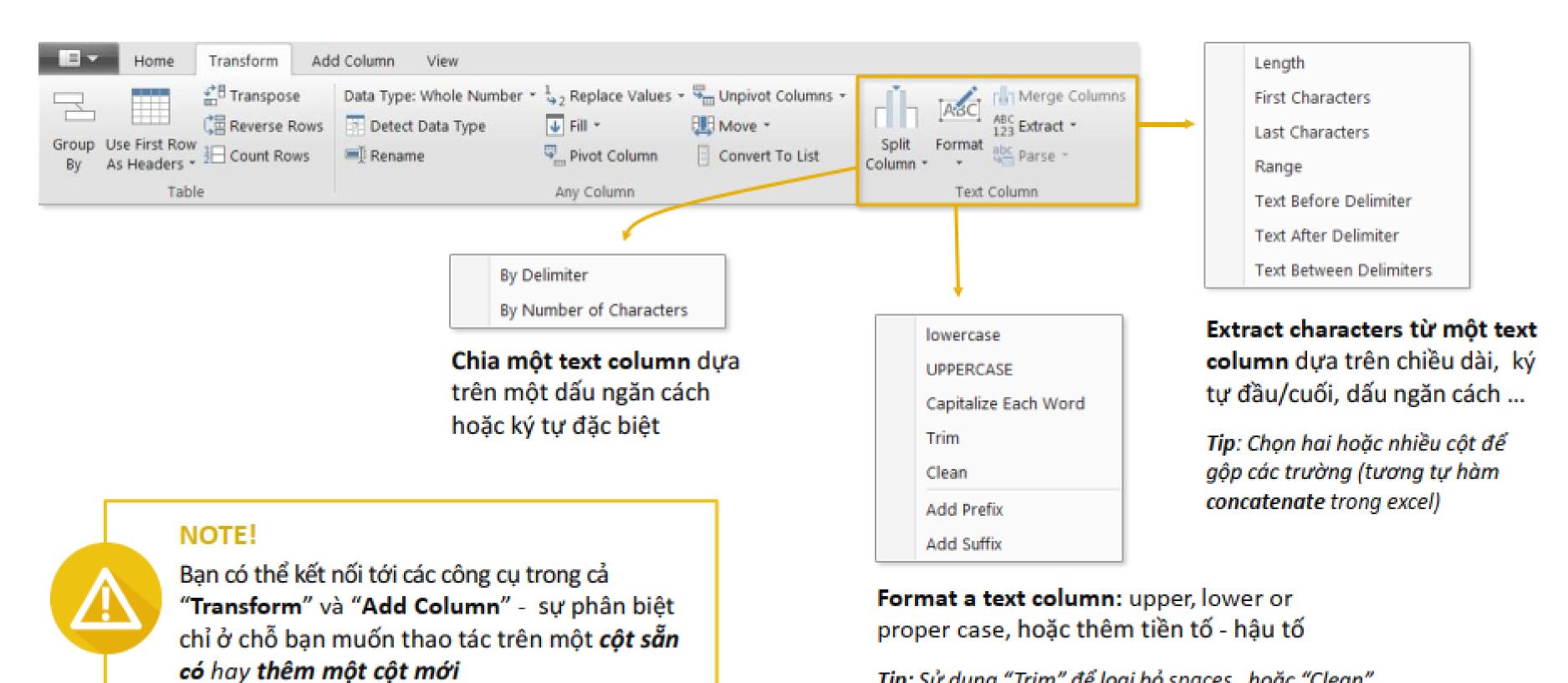


#### PRO TIP:

Sử dụng "**Folder**" option (Get Data > More > Folder) để gộp các file trong một folder (nhưng phải có chung cấu trúc); nếu bạn thêm 01 file mới, bạn chỉ cần refresh queries và mọi thứ sẽ tự động update!



#### **Text Tools**

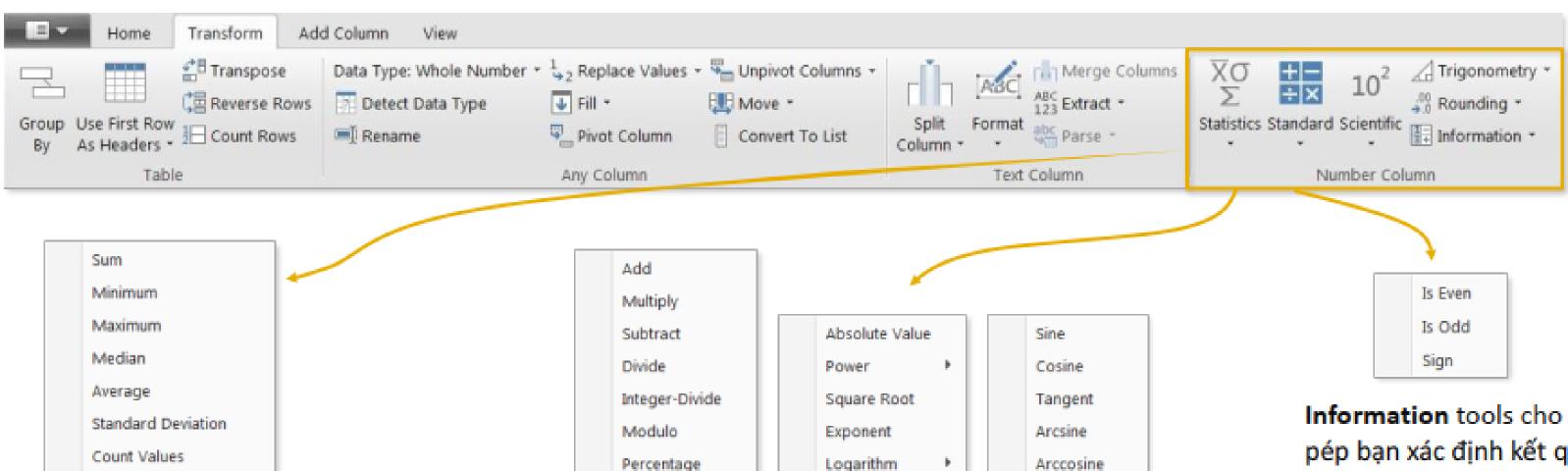


Tip: Sử dụng "Trim" để loại bỏ spaces, hoặc "Clean"

để xóa các ký tư không thể in



#### **Number Tools**



Percent Of

Standard

Statistics functions cho phép bạn thực hiện một số tính toán cơ bản từ cột được chọn (sum, min/max, average, count, ...)

Count Distinct Values

Note: Những tools này sẽ trả về một Kết quả đơn (SINGLE value), và thường sử dụng để tính toán trên bảng thay vì sử dụng ở bước xử lý data (load data) Standard, Scientific và Trigonometry tools cho phép bạn apply các công thức ở mức độ cơ bản (cộng, trừ, nhân, chia, ...) hoặc nâng cao (power, logarithm, sine, tangent, ...) cho các giá trị ở trong cột

Arctangent

Trigonometry

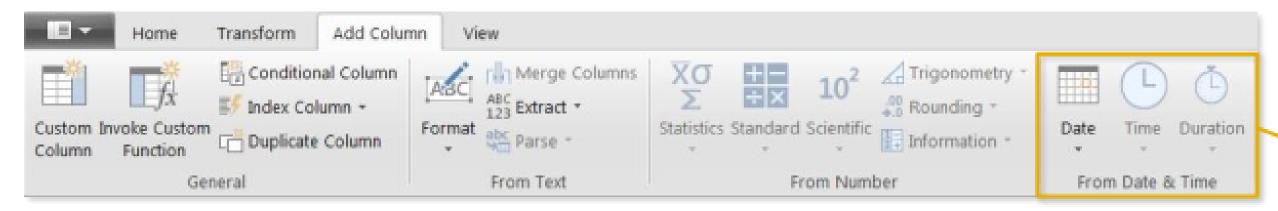
Factorial

Scientific

Note: Khác với Statistics options, các công cụ này áp dụng cho các giá trị trong cột chứ không trả về một Kết quả đơn Information tools cho pép bạn xác định kết quả Logic hoặc Nhị phân (TRUE/FALSE or 1/0) dựa trên giá trị Chẵn/Lẻ



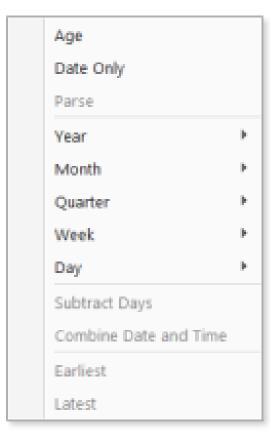
#### **Date Tools**



Date & Time tools tương đối đơn giản, bao gồm một số tùy chọn như:

- Age: sự khác nhau giữa thời gian hiện tại và thời gian trong cột
- Date Only: Bổ đi các thành phần khác ngoài định dạng Date trong cột
- Year/Month/Quarter/Week/Day: Xuất ra các thành phần từ trường Date
- Earliest/Latest: Trả về kết quả về dữ liệu Date sớm nhất hoặc muộn nhất từ Date Column dưới dạng Kết quả đơn

Note: Với dữ liệu dang Date, chúng ta nên thực hiện "Add Column" để thêm một trường/cột mới, hơn là chỉnh sửa trực tiếp dữ liệu trong "Transform"



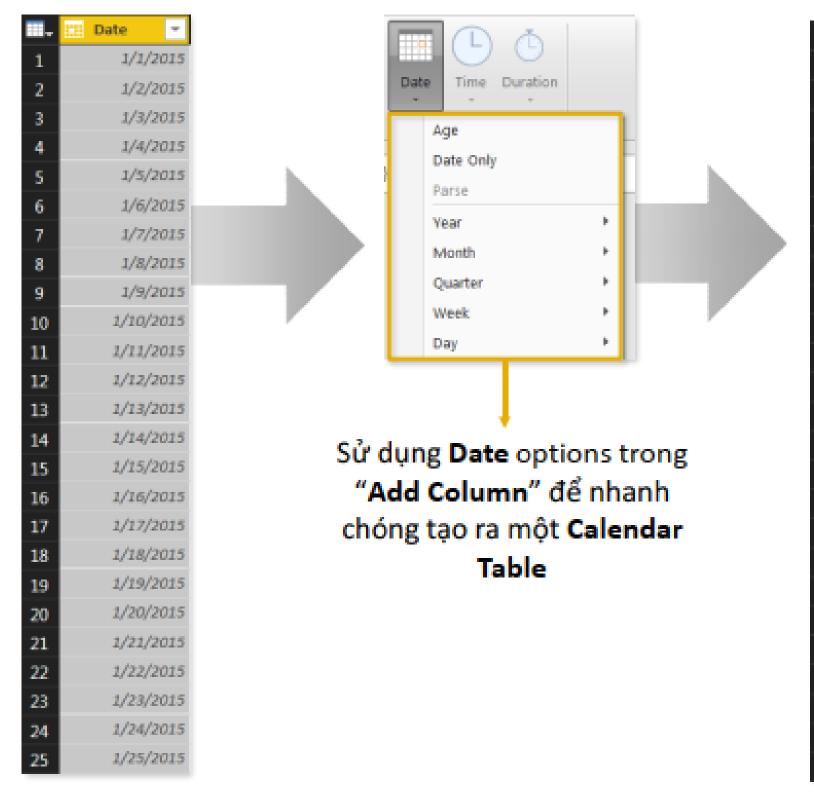


#### PRO TIP:

Nên tải một bảng chỉ chứa một cột **date column** và sử dụng Date tools để tạo ra một **calendar table** – sẽ cực kỳ hữu ích khi làm việc trong một file PowerBI có nhiều bảng



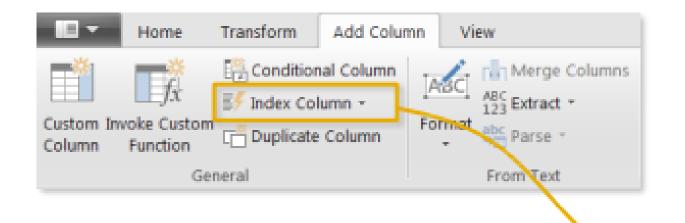
## Cách tạo một Calendar Table



-	W Date	1 <sup>2</sup> 3 Day ▽	ile Daniel Mark	all Desilvens	TIT Chart of Work	1 <sup>2</sup> <sub>3</sub> Month $=$
<b>□</b>	Date v			A <sup>II</sup> C Day Name		
1.	1/1/2015	1		Thursday	12/28/2014	1
2	1/2/2015	2		Friday	12/28/2014	1
3	1/3/2015	3		Saturday	12/28/2014	1
4	1/4/2015	4		Sunday	1/4/2015	1
5	1/5/2015	5	1	Monday	1/4/2015	1
6	1/6/2015	6	2	Tuesday	1/4/2015	1
7	1/7/2015	7	3	Wednesday	1/4/2015	1
8	1/8/2015	8	4	Thursday	1/4/2015	1
9	1/9/2015	9	5	Friday	1/4/2015	1
10	1/10/2015	10	6	Saturday	1/4/2015	1
11	1/11/2015	11	0	Sunday	1/11/2015	1
12	1/12/2015	12	1	Monday	1/11/2015	1
13	1/13/2015	13	2	Tuesday	1/11/2015	1
14	1/14/2015	14	3	Wednesday	1/11/2015	1
15	1/15/2015	15	4	Thursday	1/11/2015	1
16	1/16/2015	16	5	Friday	1/11/2015	1
17	1/17/2015	17	6	Saturday	1/11/2015	1
18	1/18/2015	18	0	Sunday	1/18/2015	1
19	1/19/2015	19	1	Monday	1/18/2015	1
20	1/20/2015	20	2	Tuesday	1/18/2015	1
21	1/21/2015	21	3	Wednesday	1/18/2015	1
22	1/22/2015	22	4	Thursday	1/18/2015	1
23	1/23/2015	23	5	Friday	1/18/2015	1
24	1/24/2015	24	6	Saturday	1/18/2015	1
25	1/25/2015	25	0	Sunday	1/25/2015	1



#### Thêm Index Columns



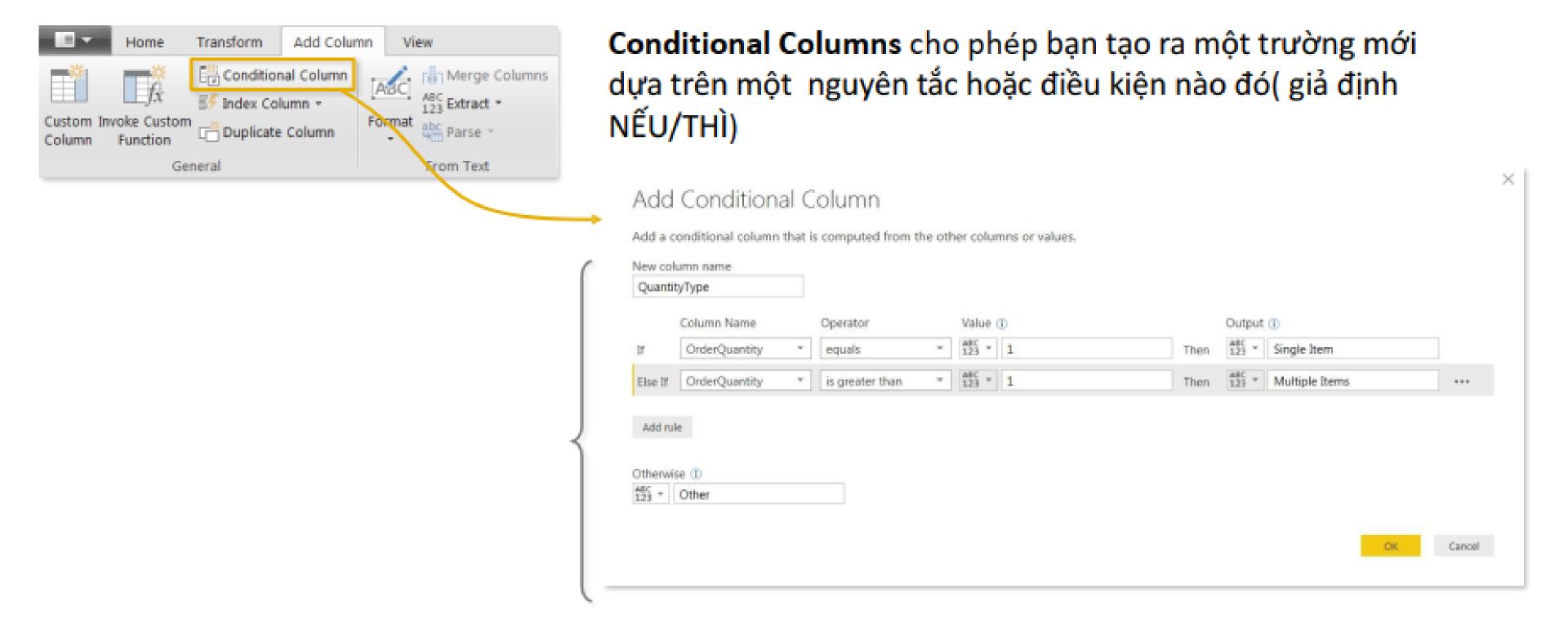
Index Columns chứa một list giá trị để xác định vị trí trong cột hoặc trong bảng (điển hình thường bắt đầu từ 0 hoặc 1)

Cột này cũng thường được dùng để tạo ra các **IDs** để có thể xác định quan hệ giữa các bảng(chi tiết ở phần sau của khóa học!)

	1 <sup>2</sup> 3 Index	OrderDate ~	StockDate 🔻	A <sup>B</sup> C OrderNumber	1 <sup>2</sup> <sub>3</sub> ProductKey	1 <sup>2</sup> <sub>3</sub> CustomerKey
1.	1	1/1/2015	9/21/2001	SO45080	332	14657
2	2	1/1/2015	12/5/2001	SO45079	312	29255
3	3	1/1/2015	10/29/2001	SO45082	350	11455
4	4	1/1/2015	11/16/2001	5045081	338	26782
5	5	1/2/2015	12/15/2001	SO45083	312	14947
6	6	1/2/2015	10/12/2001	5045084	310	29143
7	7	1/2/2015	12/18/2001	SO45086	314	18747
8	8	1/2/2015	10/9/2001	SO45085	312	18746
9	9	1/3/2015	10/3/2001	SO45093	312	18906
10	10	1/3/2015	9/29/2001	SO45090	310	29170
11	11	1/3/2015	12/11/2001	SO45088	345	11398
12	12	1/3/2015	10/24/2001	SO45092	313	18899
13	13	1/3/2015	12/16/2001	SO45089	351	25977
14	14	1/3/2015	10/26/2001	SO45091	314	18909
15	15	1/3/2015	9/11/2001	SO45087	350	11388
16	16	1/3/2015	9/11/2001	SO45094	310	22785
17	17	1/4/2015	10/30/2001	SO45096	312	12483
18	18	1/4/2015	10/30/2001	SO45097	313	29151

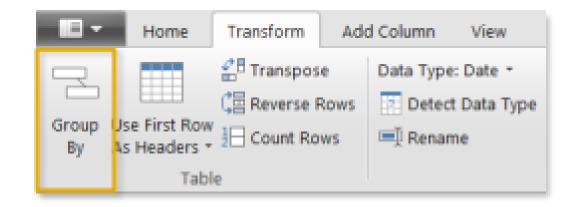


#### Thêm Conditional Columns

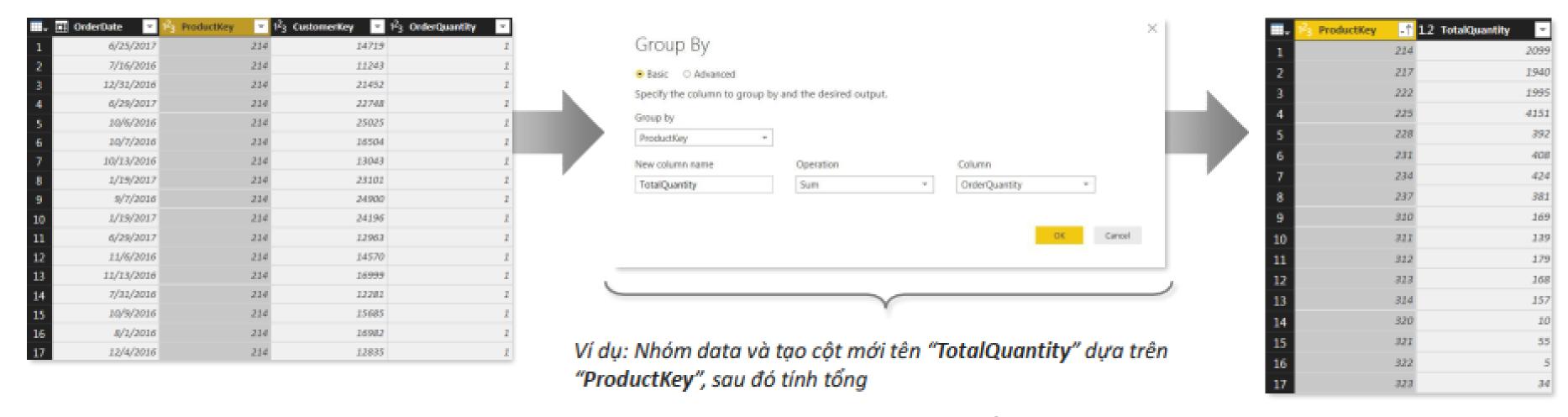




#### **Grouping Data**

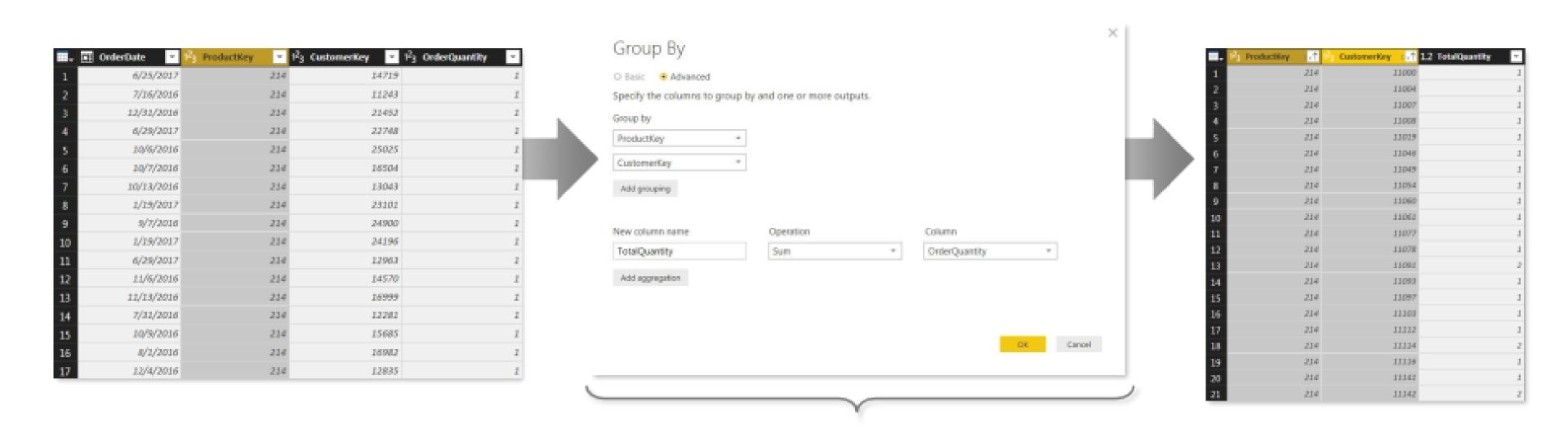


**Group By** cho phép bạn sắp xếp data theo các cấp độ hay các nhóm (data từ chi tiết ngày về theo tháng, giao dịch theo cửa hàng, ...)



NOTE: những trường không thuộc trường điều kiện nhóm hoặc trường kết quả sẽ bị xóa

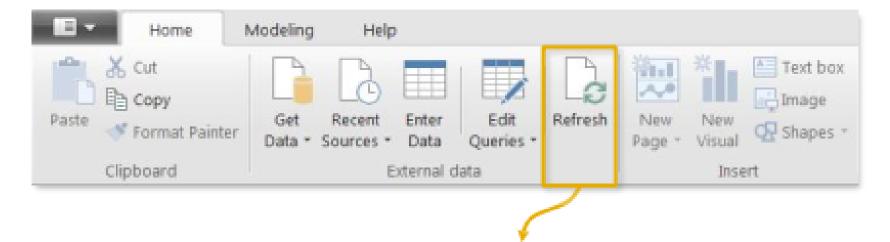




## 04. Các chức năng khác

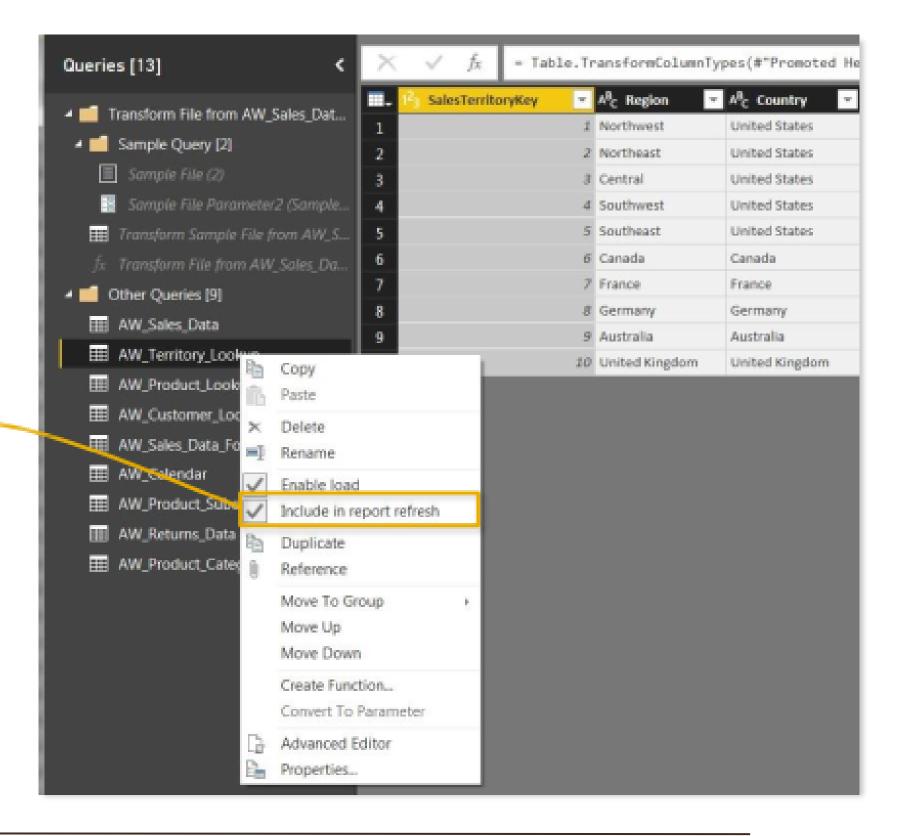


#### **Refreshing Queries**



Các queries trong model sẽ được refresh khi bạn sử dụng command "Refresh" từ tab Home

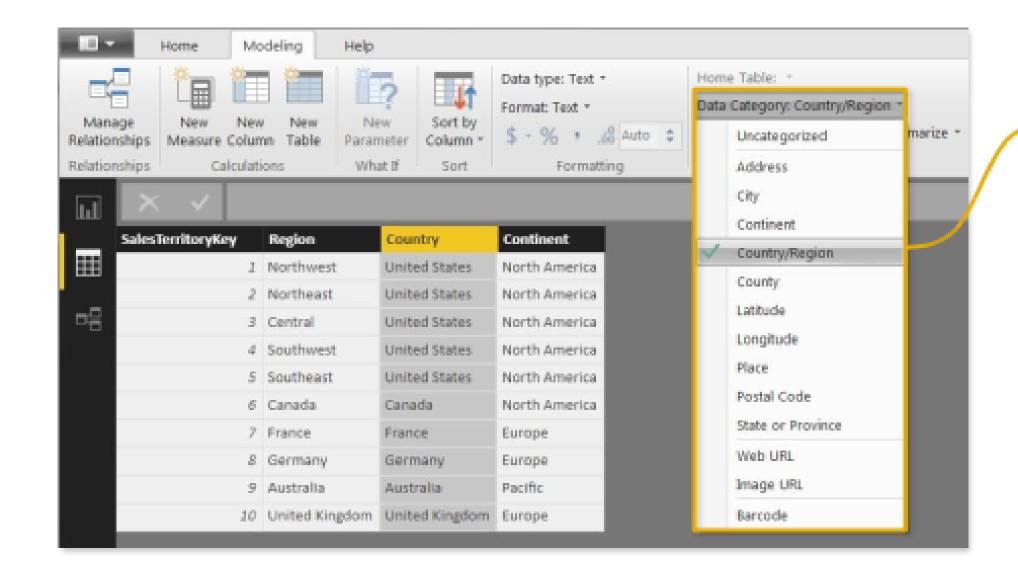
Có thể cài đặt việc Refresh thông qua phần "Include in report refresh" khi bấm vào các Queries



## 04. Các chức năng khác



#### **Defining Data Categories**



Từ Tab "Modeling" trong Data view, bạn có thể chỉnh sửa thuộc tính các trường theo một số categories đặc biệt

Thường sử dụng để xác định các trường liên quan đến addresses, countries, cities, zip codes, ... để có thể kết nối với Map



## 04. Các chức năng khác

# Một số lưu ý

- Luôn cần xử lý Data trong Queries Editor, trước khi load data vào trong Power BI
  - Cần làm sạch Data khi bắt đầu; việc nghĩ rằng có thể cập nhật sau sẽ tiêu tốn nhiều thời gian hơn.
  - Tổ chức file/folder theo cấu trúc, để tránh chỉnh sửa nguồn data nhiều lần.
- Vô hiệu hóa Report Refresh cho các nguồn bảng tĩnh
  - Không cần refresh nếu không cần update thường xuyên, ví dụ như bảng là kết quả của quá trình lookup hoặc bảng tĩnh (không cần thay đổi giá trị); chỉ refresh cho bảng có sự thay đổi thường xuyên.
- Khi làm việc với bảng dữ liệu lớn, chỉ load các data bạn cần
  - Không cần Data nào thì nên bỏ ra khỏi Bảng dữ liệu, chiết xuất quá nhiều Data có thể khiến Bảng, Model và file Power BI của bạn bị chậm trong xử lý.