

## Lab 2.3 – Sử dụng tool tính toán trong Query Editor Power BI

Trong Lab này các bạn sẽ được thực hành với các tool về tính toán thống kê như tính giá trị trung bình, tính giá trị lớn nhất, phép đếm không trùng nhau. Các nhiệm vụ chính như sau:

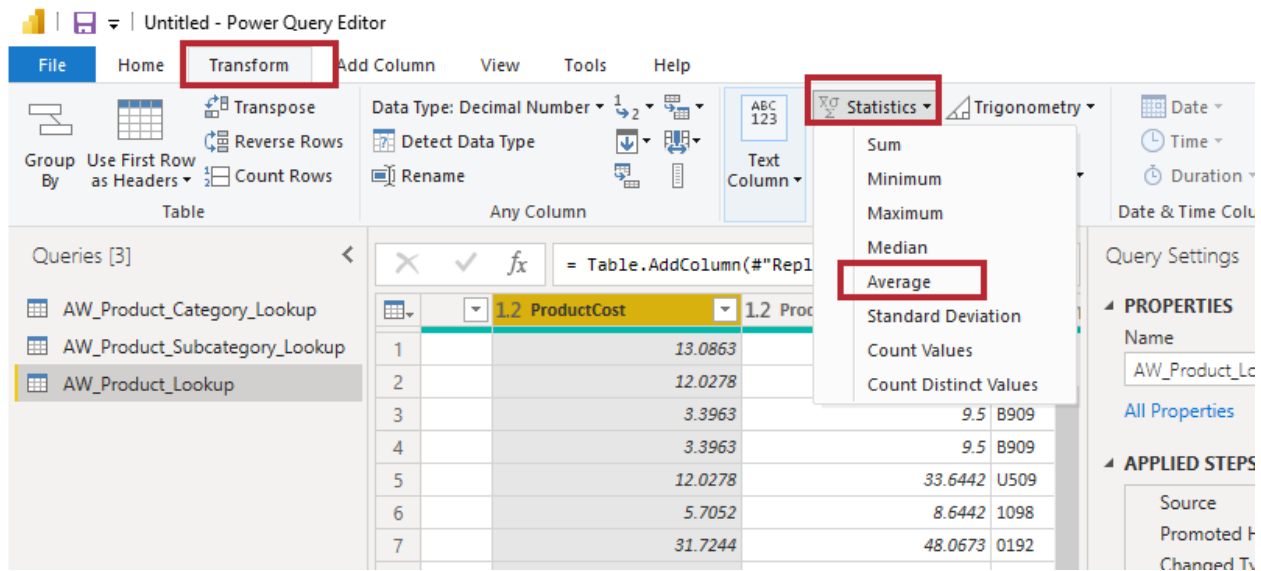
- Tính giá trị chi phí trung bình của sản phẩm trong cột **ProductCost**
- Đếm số màu không trùng nhau của sản phẩm trong cột **ProductColor**
- Import file data **AdventureWorks\_Customers.csv** vào giao diện Query Editor và đặt tên cho Query này là **AW\_Customer\_Lookup**.
- Đếm số lượng tên không trùng nhau của các khách hàng
- Tính thu nhập tối đa hàng năm của khách hàng trong cột **AnnualIncome**
- Lưu các thay đổi trong Query Editor thành file Lab 2.pbix

### 1. Tính giá trị trung bình của sản phẩm

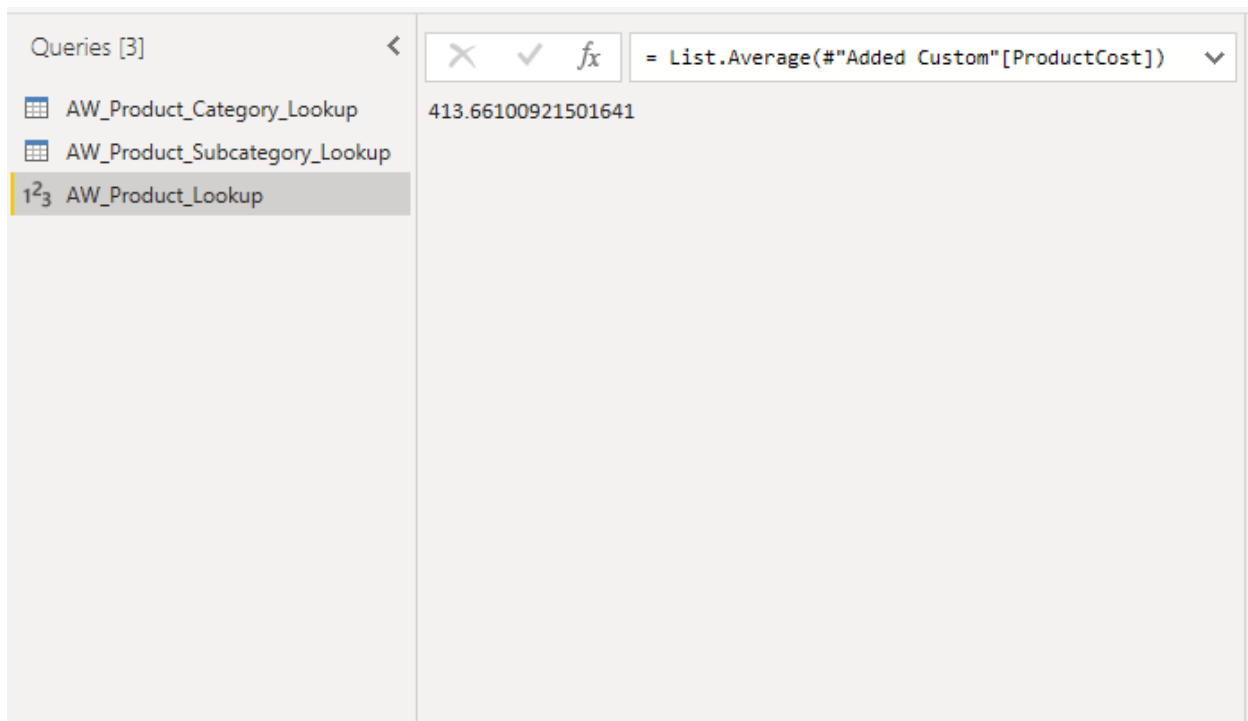
Đầu tiên các bạn chọn vào cột **ProductCost**:

	1.2 ProductCost	1.2 ProductPrice	A <sup>B</sup> <sub>C</sub> SK
	13.0863	34.99	U509
	12.0278	33.6442	U509
	3.3963	9.5	B909
	3.3963	9.5	B909
	12.0278	33.6442	U509
	5.7052	8.6442	1098
	31.7244	48.0673	0192
	31.7244	48.0673	0192
	31.7244	48.0673	0192
	31.7244	48.0673	0192
	747.9682	1263.4598	R92R
	747.9682	1263.4598	R92R
	747.9682	1263.4598	R92R
	747.9682	1263.4598	R92R
	747.9682	1263.4598	R92R
	176.1997	297.6346	R38B
	176.1997	297.6346	R38B
	176.1997	297.6346	R38B
	181.4857	306.5636	R38R
	181.4857	306.5636	R38R
	181.4857	306.5636	R38R
	181.4857	306.5636	R38R

Sau đó các bạn sẽ chọn **Transform -> Statistics -> Average**



Kết quả chúng ta tính được là **413.66**



## 2. Đếm số màu không trùng nhau của sản phẩm

Đầu tiên các bạn chọn vào cột **ProductColor**:

	ProductColor	ProductSize	
on visor.	Red	0	N
on visor.	Black	0	N
stays dry and provides just...	White	M	U
stays dry and provides just...	White	L	U
on visor.	Blue	0	N
fits all.	Multi	0	U
cling jersey	Multi	S	U
cling jersey	Multi	M	U
cling jersey	Multi	L	U
cling jersey	Multi	XL	U
ne made from the newest a...	Red	62	U

Sau đó các bạn sẽ chọn **Transform -> Statistics -> Count Distinct Values**

The screenshot shows the Power Query Editor interface. The 'Transform' tab is selected in the ribbon. The 'Statistics' dropdown menu is open, showing various statistical functions. The 'Count Distinct Values' option is highlighted with a red box. The background shows a table with columns 'ProductColor' and 'ProductSize'.

Power Query Editor - Untitled

File Home **Transform** Add Column View Tools Help

Table: Transpose, Reverse Rows, Count Rows, Use First Row as Headers, Group By

Any Column: Data Type: Text, Detect Data Type, Rename

Statistics dropdown menu:

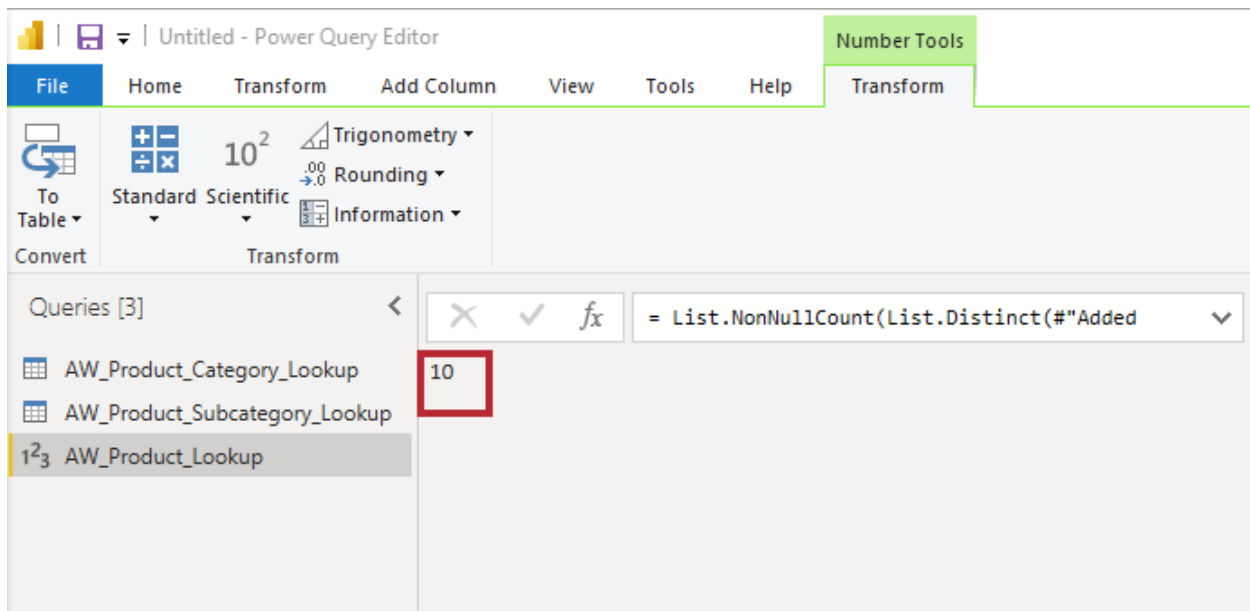
- Sum
- Minimum
- Maximum
- Median
- Average
- Standard Deviation
- Count Values
- Count Distinct Values**

Queries [3]: AW\_Product\_Category\_Lookup, AW\_Product\_Subcategory\_Lookup, AW\_Product\_Lookup

Table: ProductColor, ProductSize

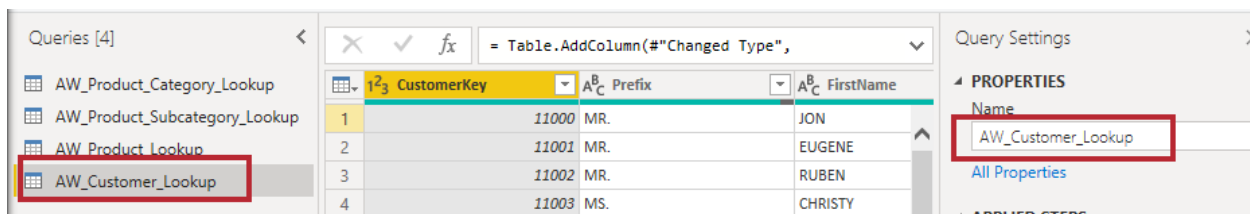
	ProductColor	ProductSize
1	Red	
2	Black	
3	White	
4	White	
5	Blue	
6	Multi	
7	Multi	
8	Multi	
9	Multi	

Kết quả là có **10** màu khác nhau:



### 3. Import file dữ liệu AdventureWorks\_Customers.csv

Để thực hiện yêu cầu này, đầu tiên các bạn phải thực hiện import file data **AdventureWorks\_Customers.csv** vào Query Editor và đặt tên cho Query này là **AW\_Customer\_Lookup**, các bước thực hiện tương tự như trong lab 2.1 hoặc 2.2.

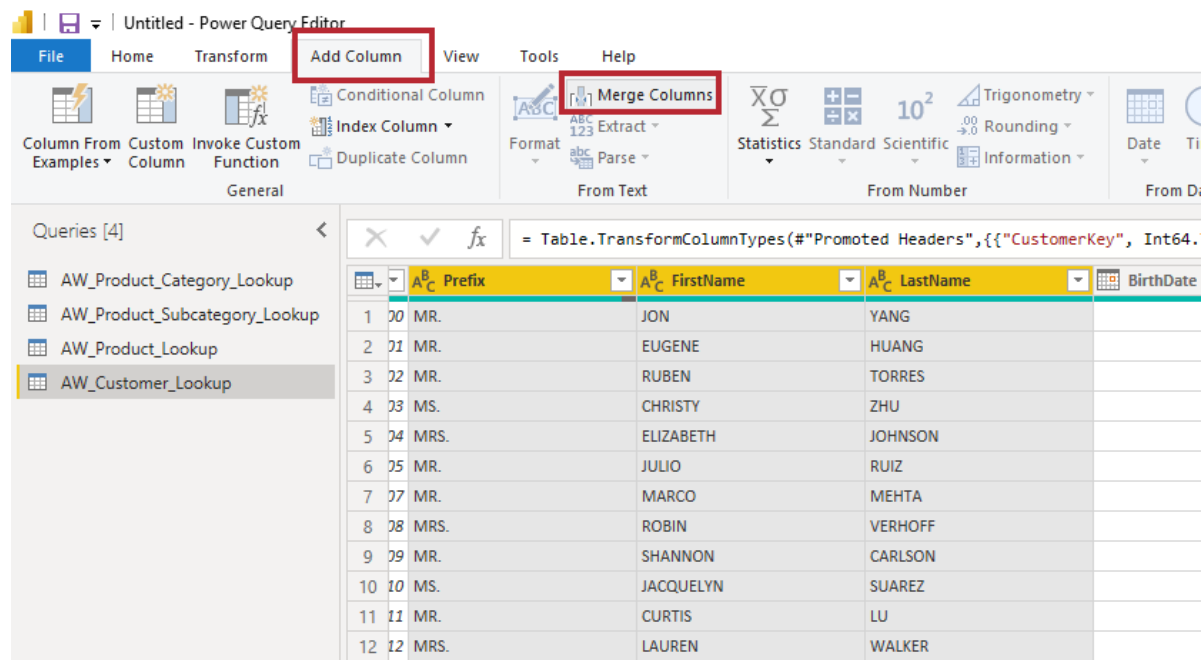


### 4. Đếm số lượng tên không trùng nhau của các khách hàng

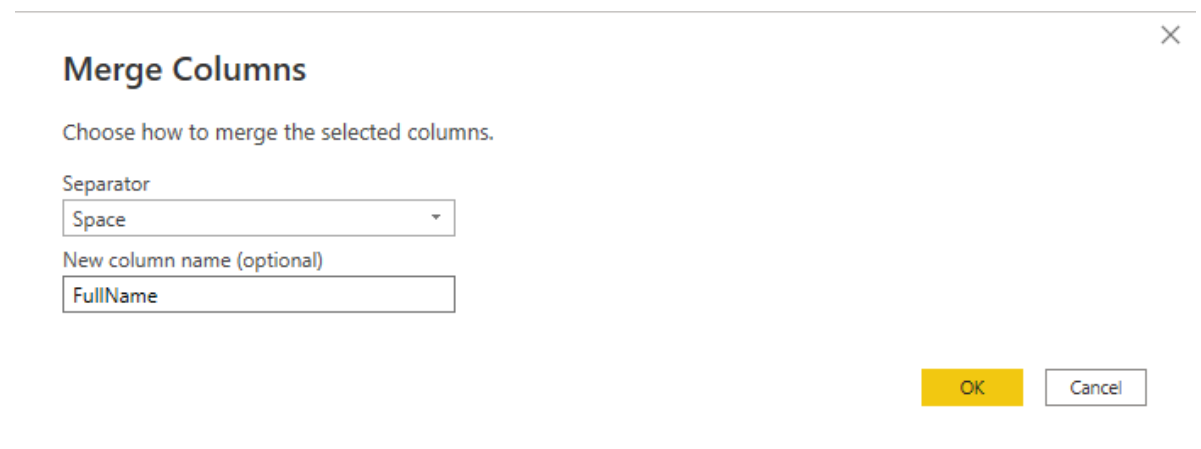
Các bạn sẽ thực hiện Merge các cột là **Prefix**, **FirstName**, **LastName** lại thành cột **FullName** bằng cách nhấn giữ Ctrl và chỉ chuột lần lượt vào ba cột muốn chọn:

CustomerKey	Prefix	FirstName	LastName	BirthDate
1	11000	MR.	JON	4/8/19
2	11001	MR.	EUGENE	5/14/19
3	11002	MR.	RUBEN	8/12/19
4	11003	MS.	CHRISTY	2/15/19
5	11004	MRS.	ELIZABETH	8/8/19
6	11005	MR.	JULIO	8/5/19
7	11007	MR.	MARCO	5/9/19

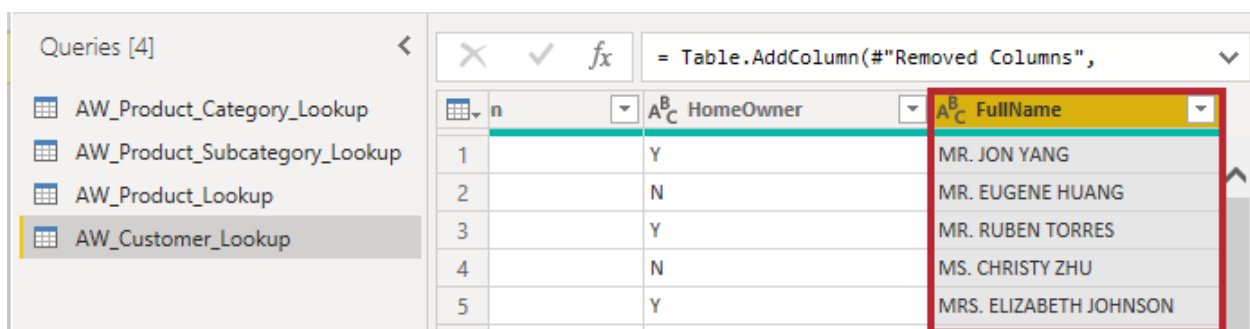
Sau đó các bạn sẽ chọn **Add Column -> Merge Columns**:



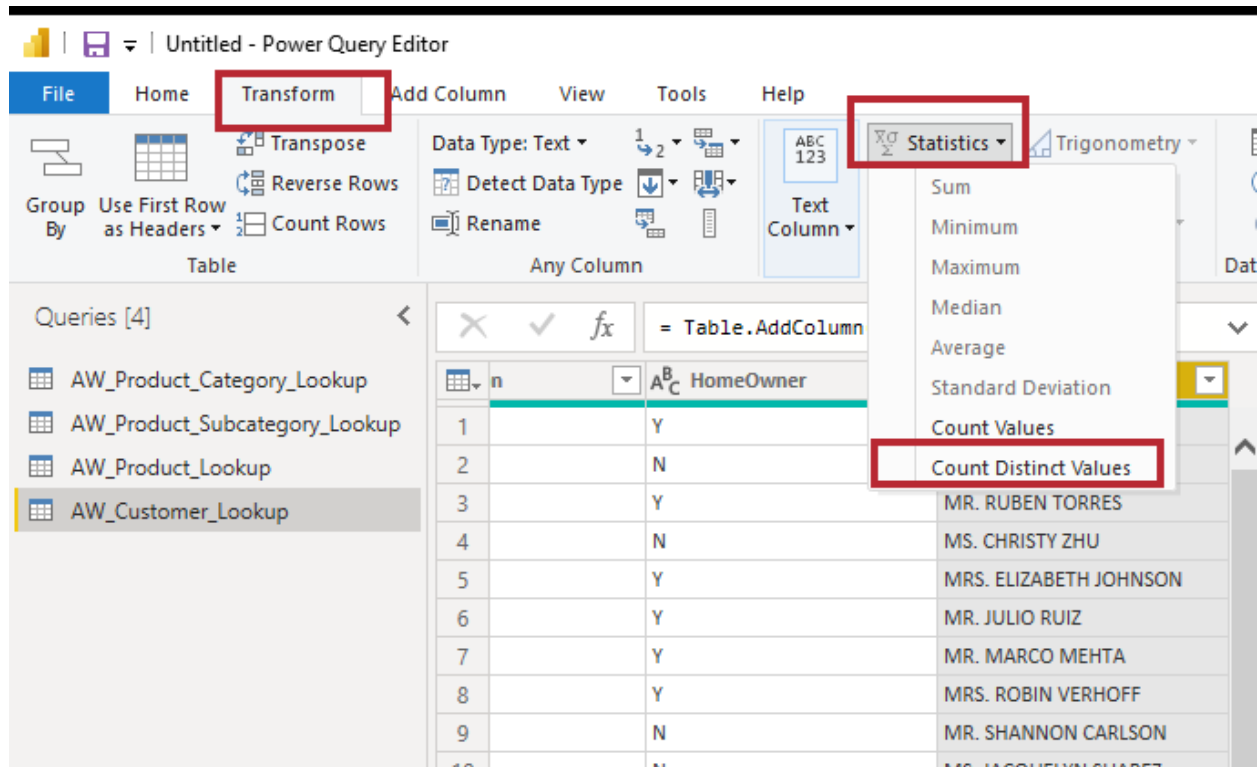
Sau đó, các bạn sẽ chọn Separator là Space và New column name là FullName.



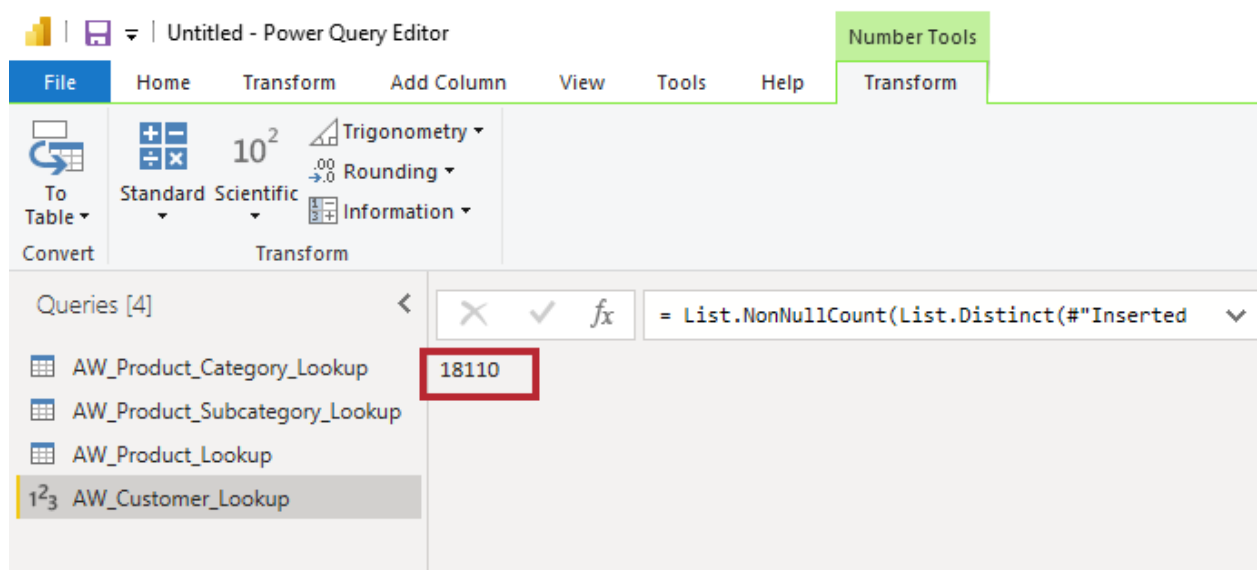
Kết quả:



Tiếp theo để đếm các tên không trùng nhau, các bạn sẽ chọn vào cột **FullName** như hình trên sau đó chọn **Transform -> Statistics -> Count Distinct Values**

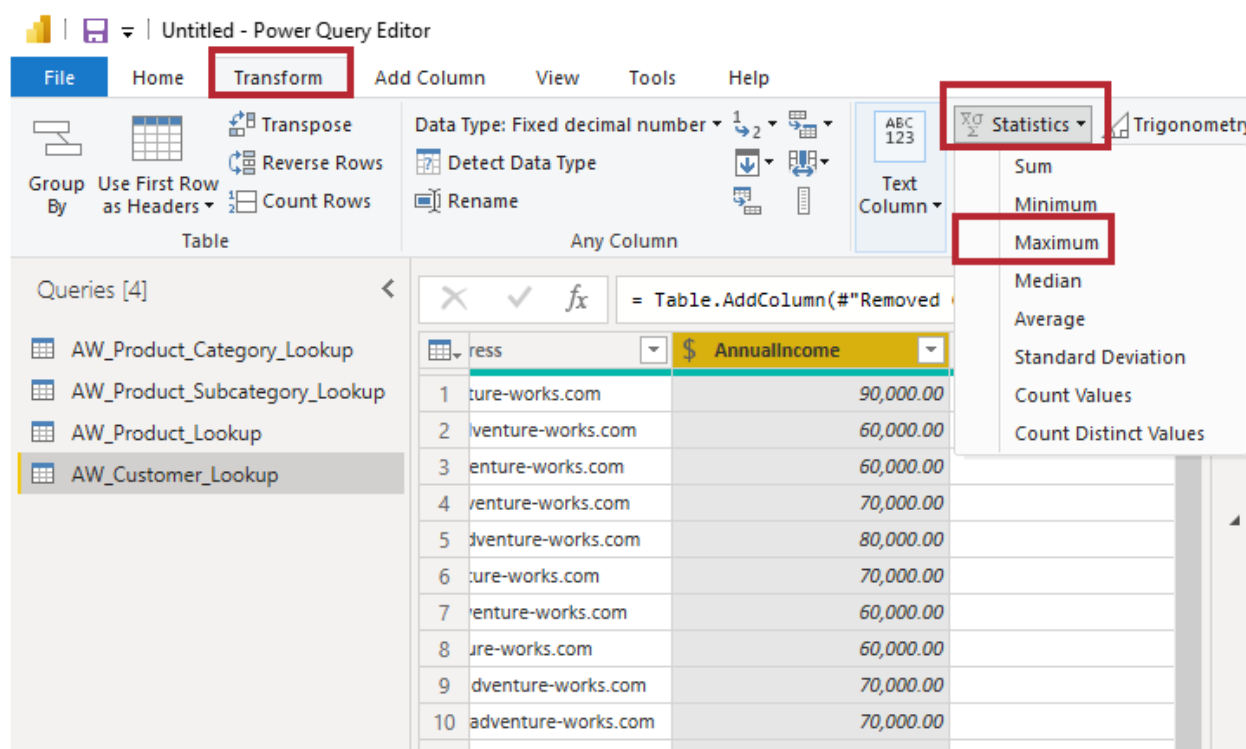


Kết quả là chúng ta tìm được **18110** tên không trùng nhau:

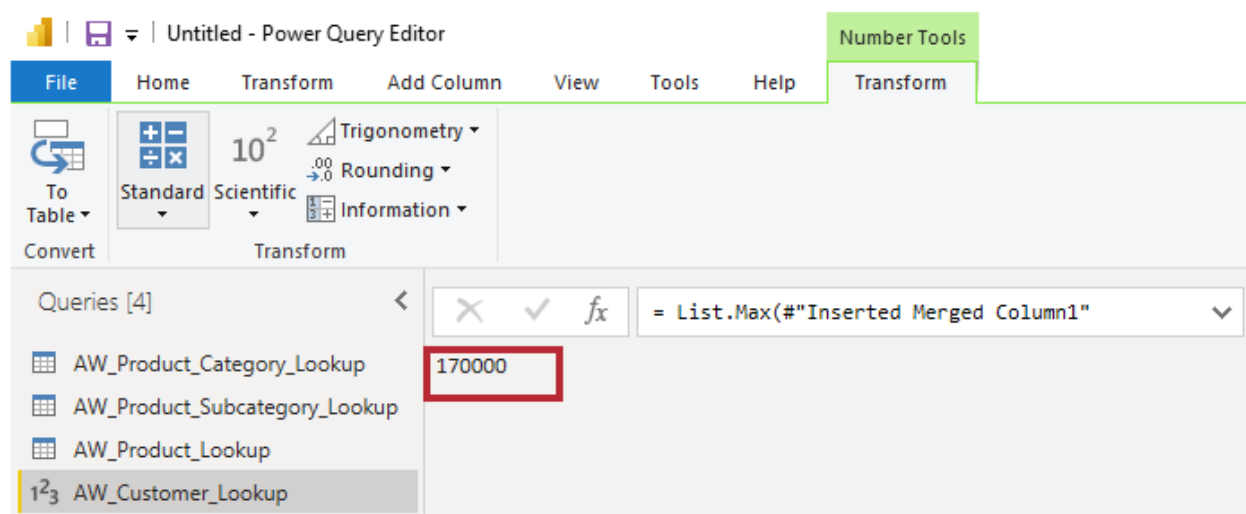


## 5. Tính thu nhập tối đa hàng năm của khách hàng

Các bạn sẽ chọn cột **AnnualIncome**, sau đó chọn **Transform -> Statistics -> Maximum**:



Kết quả chúng ta tìm được là **170000**.

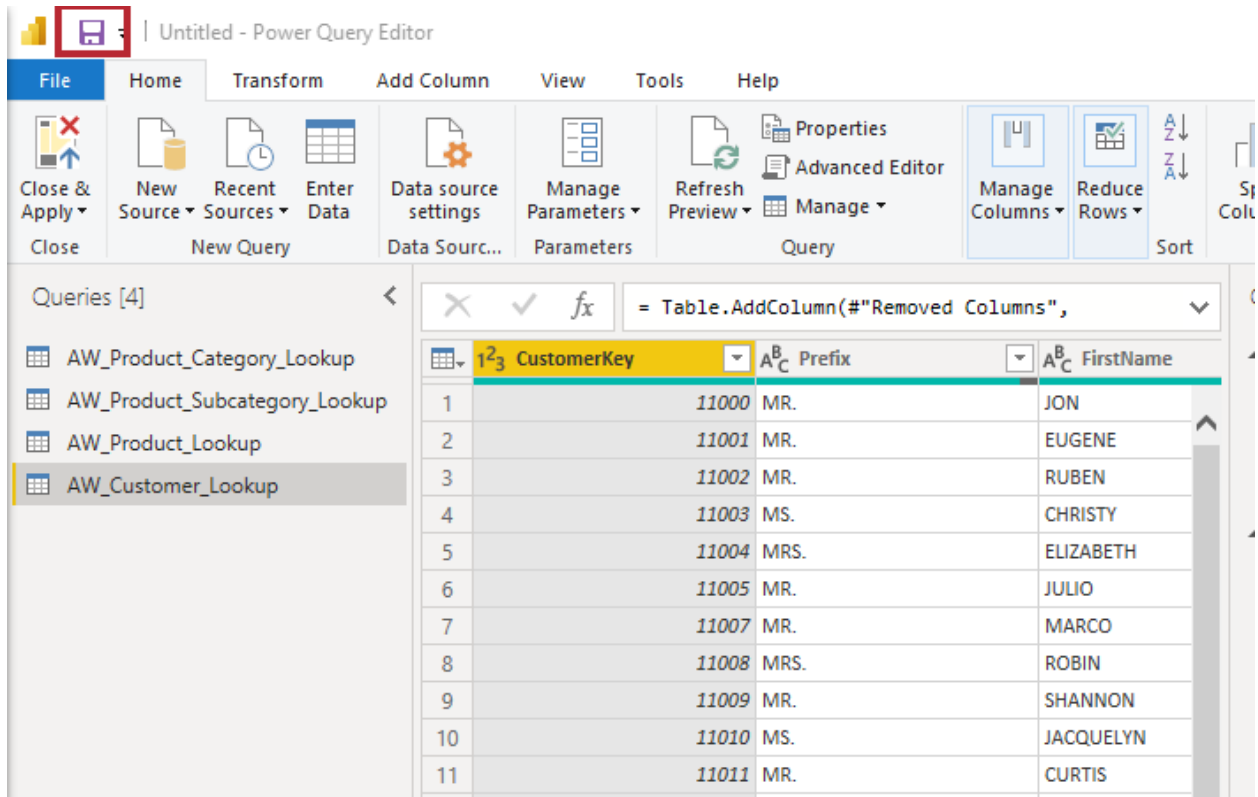


## 6. Lưu các thay đổi trong Query Editor thành file Lab 2.pbix

Sau khi tiến hành các thay đổi, các bạn sẽ lưu các thay đổi trong Query Editor. Ở Lab này, các bạn sẽ chỉ thực hiện lưu thay đổi trong Query Editor và sẽ không Apply các thay đổi (Tải dữ liệu

vào Data model), Ở Lab sau các bạn sẽ được thực hiện Apply các thay đổi này. Quy trình xử lý dữ liệu của chúng ta là sẽ sắp xếp hay làm sạch dữ liệu trước, sau khi đảm bảo giai đoạn này đã thực hiện xong chúng ta mới thực hiện tải dữ liệu vào Data Model để thực hiện xây dựng quan hệ giữa các bảng hoặc vẽ biểu đồ. Trong thực tế, công việc sắp xếp hay làm sạch dữ liệu thường là công đoạn chiếm nhiều thời gian nhất, chính vì thế khi xử lý chưa xong mà chúng ta muốn lưu lại các thay đổi để lần sau có thể mở lên để xử lý tiếp mà chưa tải vào Data Model vội, chúng ta sẽ thực hiện như sau:

Đầu tiên chọn Lưu:



Sau đó Power BI sẽ hỏi có muốn Apply các thay đổi không, các bạn sẽ chọn vào Apply later:

## Microsoft Power BI Desktop

There are pending changes in your queries that haven't been applied.  
Do you want to apply them?

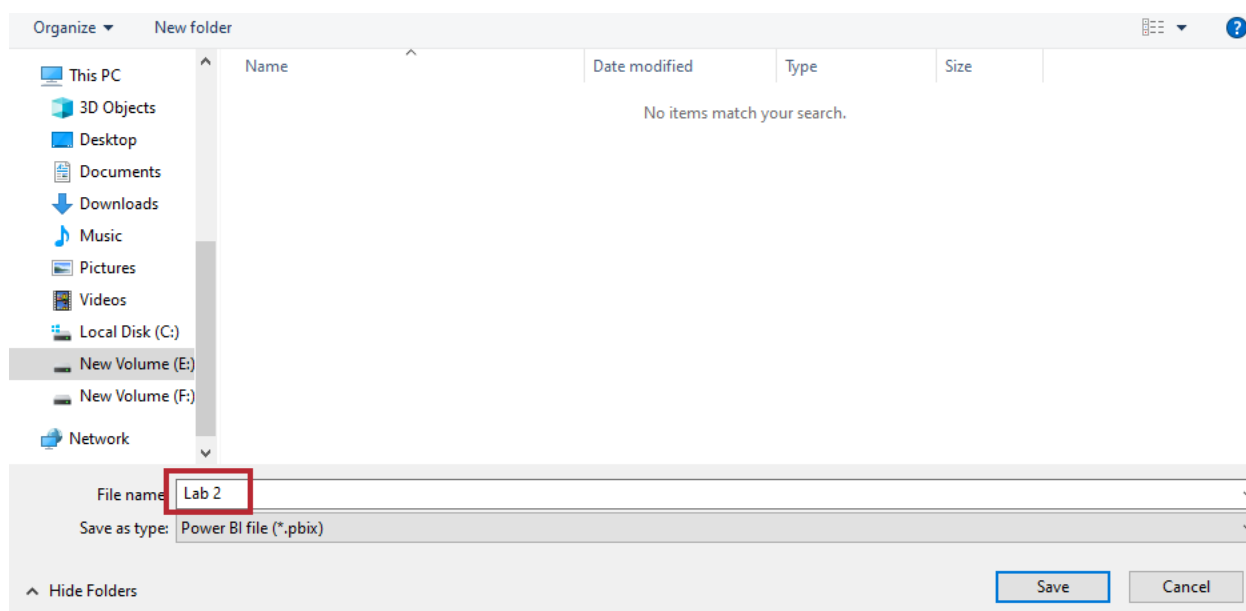
Apply

Apply later

Cancel



Sau đó chúng ta sẽ lưu với tên là Lab 2



Ở Lab sau, các bạn sẽ mở file đã lưu này để tiếp tục công việc trong Query Editor.