



# TÀI LIỆU DỰ ÁN

XỬ LÝ DỮ LIỆU CHỨNG KHOÁN - StockReport



**GVHD: Thầy Lại Minh Đăng**

**HSSV: Nguyễn Đình Mai Vy**

**LỚP: DP20302**

# MỤC LỤC

1	Giới thiệu dự án .....	4
1.1	Giới thiệu công ty H2NStock .....	4
1.2	Yêu cầu của công ty .....	4
1.3	Lập kế hoạch dự án.....	5
2	Phân tích yêu cầu khách hàng .....	6
2.1	Phân tích yêu cầu.....	6
2.2	Câu chuyện dữ liệu.....	7
2.2.1	Đặt vấn đề .....	7
2.2.2	Xác định câu chuyện .....	8
2.2.3	Xác định rõ đối tượng .....	9
2.2.4	Xác định câu chuyện chi tiết.....	10
2.2.5	Trình bày dữ liệu.....	10
2.2.6	Những điều cần lưu ý.....	13
2.3	Kiến trúc hệ thống .....	14
2.3.1	Kiến trúc.....	14
2.3.2	Giải thích.....	15
2.4	Giải thích về bộ dữ liệu khách hàng.....	15
2.4.1	Các khái niệm .....	15
2.4.2	Các trường dữ liệu .....	15
3	Làm sạch và chuyển đổi dữ liệu .....	19
3.1	Chuẩn bị dữ liệu .....	19
3.1.1	Giải pháp lưu trữ dữ liệu.....	19
3.1.2	Giải pháp phân bố dữ liệu.....	20
3.2	Làm sạch dữ liệu .....	28
3.2.1	Các vấn đề ảnh hưởng tới dữ liệu .....	28
3.2.2	Các tiêu chí đánh giá chất lượng dữ liệu .....	29

3.2.3	Các bước làm sạch dữ liệu .....	30
3.3	Chuyển đổi dữ liệu .....	35
3.3.1	Các trường hợp cần chuyển đổi .....	36
3.3.2	Các kỹ thuật chuyển đổi.....	36
3.3.3	Trình bày các phép chuyển đổi trong dự án .....	38
4	Xử lý dữ liệu .....	46
4.1	Chuẩn hóa dữ liệu.....	46
4.1.1	Trình bày các bước chuẩn hóa trong dự án.....	46
4.2	Mô hình hóa dữ liệu .....	55
4.2.1	Các loại mô hình hóa .....	55
4.2.2	Các tiêu chí đánh giá mô hình dữ liệu .....	56
4.2.3	Trình bày các bước mô hình hóa .....	57
4.2.4	Trình bày các bước tạo bảng dữ liệu .....	64
4.3	Xử lý dữ liệu DAX .....	78
4.3.1	Measure.....	78
4.3.2	Calculated column .....	90
4.3.3	Filter .....	93
5	Trực quan hóa dữ liệu .....	95
5.1	Các kỹ thuật trực quan hóa.....	95
5.2	Các nguyên tắc trực quan hóa .....	96
5.3	Trình bày cách thêm visual mới.....	97
5.4	Trình bày tạo các report cho dự án.....	100
5.4.1	Tạo visual thống kê chi tiết.....	100
5.4.2	Tạo visual thống kê tổng thể.....	109
6	Xây dựng báo cáo .....	124
6.1	Dashboard và report .....	124
6.2	Xây dựng báo cáo.....	125

6.2.1	Dashboard vs Report.....	125
6.2.2	Dashboard .....	127
6.2.3	Report.....	127
6.2.4	Bookmark.....	129
7	KẾT LUẬN.....	133
7.1	Báo cáo .....	133
7.1.1	Các bước viết báo cáo .....	133
7.1.2	Tổng hợp .....	133
7.2	Khó khăn .....	136
7.3	Thuận lợi.....	136
7.4	Hướng phát triển.....	136

# 1 GIỚI THIỆU DỰ ÁN

## GIỚI THIỆU CÔNG TY H2NSTOCK

**Công ty H2NSTOCK** là một doanh nghiệp hoạt động trong lĩnh vực chứng khoán, chuyên cung cấp dịch vụ thống kê, phân tích và tư vấn đầu tư cho khách hàng cá nhân và tổ chức. Với thế mạnh trong việc thu thập, xử lý khối lượng lớn dữ liệu giao dịch từ nhiều sàn như HOSE, HNX, và UPCOM, OTC nhằm đưa ra góc nhìn tổng quan & dễ hiểu hơn cho người đọc.

### Thông tin chung về bộ dữ liệu:

- ❖ Tập dữ liệu liên quan đến các mã cổ phiếu đang giao dịch trên thị trường chứng khoán Việt Nam.
- ❖ Được thu thập từ nhiều website khác nhau (có trả phí và miễn phí)
- ❖ Trong một khoảng thời gian nhất định
- ❖ Chỉ là những con số khô khan
- ❖ Thông tin phẳng, không theo cấu trúc.

### 1.1 YÊU CẦU CỦA CÔNG TY

#### DỮ LIỆU

- ✓ Làm sao truyền đạt và diễn giải những thông tin này tới nhân viên và sếp.
- ✓ Thông điệp muốn nói một cách rõ ràng và hấp dẫn.
- ✓ Data Insights một cách sống động, dễ hiểu, thuyết phục
- ✓ Đề hiểu được vấn đề nhanh nhất và chính xác nhất.

#### QUẢN LÝ VÀ LUU TRỮ

- ✓ Giảm không gian lưu trữ, quản lý tài liệu khoa học, tiết kiệm thời gian tìm kiếm.
- ✓ Tăng cường khả năng bảo mật thông tin.

- ✓ Chia sẻ thông tin một cách nhanh chóng.
- ✓ Chi phí vận hành và quản lý thấp nhưng vẫn mang lại hiệu quả cao.

## MỤC TIÊU

- ✓ Công nghệ phổ biến.
- ✓ Chi phí hợp lý.
- ✓ Có thể mở rộng nhanh chóng.

## CÔNG NGHỆ

- ✓ Microsoft Power BI: phiên bản Desktop và Service
- ✓ Hệ quản trị CSDL: SQL Server 2016+
- ✓ Microsoft Office: Ms PowerPoint và Ms Word

- ✓ *Đánh giá tính khả thi (năng lực, khả năng của mình có thể thực hiện dự án này)*
  - Năng lực (skill đang có): ...
  - Năng lực (skill sẽ học thêm cho dự án): ...

## 1.2 LẬP KẾ HOẠCH DỰ ÁN

- ✓ *Lập kế hoạch kế hoạch dự án theo mẫu sau*

TT	HẠNG MỤC	BẮT ĐẦU	KẾT THÚC	KẾT QUẢ
1	Giới thiệu dự án			
1.1	Giới thiệu công ty			
1.2	Yêu cầu công ty			
	...			
2	Phân tích yêu cầu			
2.1	Phân tích yêu cầu KH			
2.2	Câu chuyện dữ liệu			
	...			
3	Làm sạch và chuyển đổi dữ liệu			
3.1	Chuẩn bị dữ liệu			
	...			
4	Xử lý dữ liệu			

4.1	Chuẩn hóa dữ liệu			
	...			
5	Trực quan hóa dữ liệu			
5.1	Các kỹ thuật trực quan hóa			
	...			
6	Xây dựng báo cáo			
6.1	Dashboard và Report			
	...			

## 2 PHÂN TÍCH YÊU CẦU KHÁCH HÀNG

### 2.1 PHÂN TÍCH YÊU CẦU

#### ❖ Dữ liệu

- ❖ Làm sao truyền đạt và diễn giải những thông tin này tới nhân viên và sếp.
- ❖ Thông điệp muốn nói một cách rõ ràng và hấp dẫn
- ❖ Data Insights một cách sống động, dễ hiểu, thuyết phục.
- ❖ Để hiểu được vấn đề nhanh nhất và chính xác nhất.

#### ❖ Quản lý và lưu trữ

- ❖ Giảm không gian lưu trữ, quản lý tài liệu khoa học, tiết kiệm thời gian tìm kiếm
- ❖ Tăng cường khả năng bảo mật thông tin.
- ❖ Chia sẻ thông tin một cách nhanh chóng
- ❖ Chi phí vận hành và quản lý thấp nhưng vẫn mang lại hiệu quả cao

#### ❖ Công nghệ

- ❖ Công nghệ phổ biến
- ❖ Chi phí hợp lý.
- ❖ Có thể mở rộng nhanh chóng Microsoft Office (PowerPoint, Word) → tạo tài liệu báo cáo và trình bày kết quả.

✓ *Quyết định dùng công nghệ*

## **❑Công nghệ**

- ❖ Trong phạm vi dự án và môn học Dự Án Mẫu
- ❖ Microsoft Power BI: phiên bản Desktop và Service
- ❖ Hệ quản trị CSDL: SQL Server 2016+
- ❖ Microsoft Office: Ms PowerPoint và Ms Word

## **2.2 CÂU CHUYỆN DỮ LIỆU**

- ✓ *Hãy mô tả các bước cụ thể để xây dựng câu chuyện, bằng cách đặt câu hỏi:*
  - *Ai đọc báo cáo này?*
  - *Giải thích vấn đề là gì? Ở mức độ nào?*
  - *Sử dụng phương pháp gì?*

### **❑Bạn đang cố gắng giải thích điều gì từ dữ liệu ?**

=> Tôi đang muốn giải thích các mã chứng khoán thay đổi theo ngày, theo ngành nghề và theo sàn giao dịch.

### **❑Mục tiêu cụ thể của bạn khi kể câu chuyện dữ liệu này là gì ?**

=> Mục tiêu cụ thể là tách và phân nhóm dữ liệu thành các đối tượng cụ thể.

=> Sau đó sẽ hiện sự thay đổi trên từng đối tượng này thông qua hình ảnh là các dạng biểu đồ.

### **❑Bạn muốn đề xuất giải pháp gì từ những phân tích này ?**

=> Trong phạm vi dự án này, tôi chỉ muốn trình bày các báo cáo tổng hợp dữ liệu rõ ràng và chính xác về quy mô thị trường với từng mã chứng khoán.

=> KHÔNG đề xuất MUA hoặc BÁN chứng khoán.

=> KHÔNG dự báo thị trường tăng hay giảm trong tương lai.

=> KHÔNG khuyến nghị hay đề xuất người xem (học viên) tham gia thị trường chứng khoán.

### **2.2.1 ĐẶT VẤN ĐỀ**

- ✓ **MÔ TẢ THỰC TRẠNG:**

Hiện nay rất nhiều doanh nghiệp hoặc cá nhân ứng dụng phân tích dữ liệu vào quá trình quản lý, kinh doanh, đầu tư. Công ty H2NStock là một trong số đó.

- ❑ Đối với hầu hết các cá nhân và doanh nghiệp, thiếu dữ liệu không phải là vấn đề.
- ❑ Họ có thể lấy, mua bán hoặc trao đổi thông tin dữ liệu từ nhiều nguồn khác nhau.
- ❑ Các dữ liệu không đồng nhất về cấu trúc định dạng.
- ❑ Chỉ là dữ liệu thô, không theo tiêu chuẩn nào.
- ❑ Có quá nhiều thông tin không được diễn giải để giúp họ đưa ra một quyết định

✓ **DỮ LIỆU LIÊN QUAN:**

- Dữ liệu sau khi xử lý, kết quả là hàng loạt con số khô khan.  
Làm thế nào để diễn giải và truyền đạt kết quả này tới cấp trên và đồng nghiệp?

Ba yếu tố chính để kể chuyện là:

- ❖ Xây dựng dữ liệu (Data)
- ❖ Sử dụng hình ảnh để làm sáng tỏ (Visual)
- ❖ Hiển thị dữ liệu để hỗ trợ (Narrative)

✓ **MỤC TIÊU:**

Truyền đạt thông điệp dữ liệu rõ ràng, hấp dẫn giúp người nghe hiểu được vấn đề nhanh chính xác.  
Thể hiện sinh động, dễ hiểu, thuyết phục.

## 2.2.2 XÁC ĐỊNH CÂU CHUYỆN

- ✓ *Bằng cách đặt câu hỏi, hình thành giả thuyết, tổng hợp đào sâu dữ liệu để tìm câu trả lời:*
- *Bạn đang có gǎng giải thích điều gì từ dữ liệu?*

→ Xác định xu hướng biến động giá và khối lượng giao dịch theo từng mã cổ phiếu, từng ngành.

- *Mục tiêu cụ thể của bạn khi kể câu chuyện dữ liệu này là gì?*

→ Giúp lãnh đạo nhận biết mã cổ phiếu tiềm năng, ngành tăng trưởng mạnh, và thời điểm mua bán hợp lý.

- *Bạn muốn đề xuất giải pháp gì từ những phân tích này?*

→ Tối ưu danh mục đầu tư, phân bổ vốn theo ngành/nhóm mã tăng trưởng tốt.

- ✓ *Một số cách tiếp cận dữ liệu:*

- *Tìm kiếm mối tương quan?*

- Sàn Giao Dịch >> Mã Chứng Khoán
  - Ngành nghề >> Mã Chứng Khoán

- *Xác định xu hướng?*

- Trong phạm vi bộ dữ liệu được cung cấp, giá Mã Chứng Khoán thay đổi theo ngày (TĂNG, GIẢM, KHÔNG ĐỒI).
  - Trong thực tế, giá Mã Chứng Khoán thay đổi theo thời gian thực trong phiên giao dịch.

- *Rút ra so sánh?*

- So sánh sự thay đổi giá trong phiên giao dịch giữa các ngày trên cùng một Mã Chứng Khoán.

### 2.2.3 XÁC ĐỊNH RÕ ĐỐI TƯỢNG

- ✓ *Trả lời các câu hỏi:*

- *Ai cần nghe câu chuyện này?*

→ Nhân viên các phòng ban và sếp của cty H2NStock.

- *Họ đã biết đến lĩnh vực này chưa?*

→ Nhân viên chưa biết và nhân viên mới tiếp xúc trong thời gian ngắn.

=> Vì vậy nên trình bày rõ ràng, dễ hiểu.

- *Thứ họ quan tâm nhất là gì?*

→ Xu hướng giá, mã cổ phiếu nổi bật, ngành tăng trưởng.

- *Thứ họ quan tâm nhất là gì?*

=> Trong phạm vi dự án này, chỉ cung cấp kiến thức, khái niệm cơ bản về thị trường chứng khoán.

=> Hiểu được một số nghiệp vụ giao dịch cơ bản

=> Biết cách nhìn biểu đồ giao dịch.

#### 2.2.4 XÁC ĐỊNH CÂU CHUYỆN CHI TIẾT

- ✓ *Giải thích dữ liệu rõ ràng và cần trình tự logic hợp lý:*
  - *Bối cảnh?*

=> Tại Việt Nam, chứng khoán đang là một kênh đầu tư tiềm năng với khả năng sinh lời hấp dẫn trong thời gian ngắn. Trước khi quyết định tham gia vào thị trường này, bạn cần nắm rõ các kiến thức về thị trường chứng khoán là gì, đặc điểm và chức năng ra sao.

- *Ai sẽ xem báo cáo này?*

→ Nhân viên chưa biết và nhân viên mới tiếp xúc trong thời gian ngắn.

- *Vấn đề?*

=> Có hiểu được rủi ro khi tham gia thị trường không?

- *Giải pháp?*

=> Cần phải có kiến thức cơ bản về thị trường chứng khoán, hệ thống tài chính, cổ phiếu.

#### 2.2.5 TRÌNH BÀY DỮ LIỆU

- ✓ *Dùng hình ảnh để trình bày:*
  - *Sử dụng biểu đồ gì?*

❑ Hình ảnh là cách hiệu quả để thu hút người xem

=> Khi người xem không phải là người có chuyên môn sâu về dữ liệu hoặc lĩnh vực đang đề cập (về chứng khoán).

❑ Trình bày đơn giản, ngắn gọn, làm nổi bật dữ liệu quan trọng, đáng được lưu ý)

=> Tổng giá trị giao dịch, tổng khối lượng cổ phiếu được giao dịch trong ngày.

❑ Sử dụng các dạng biểu đồ, bộ lọc hiệu quả trong lĩnh vực chứng khoán

=> Card



=> Line chart: Theo dõi xu hướng giá điều chỉnh theo thời gian



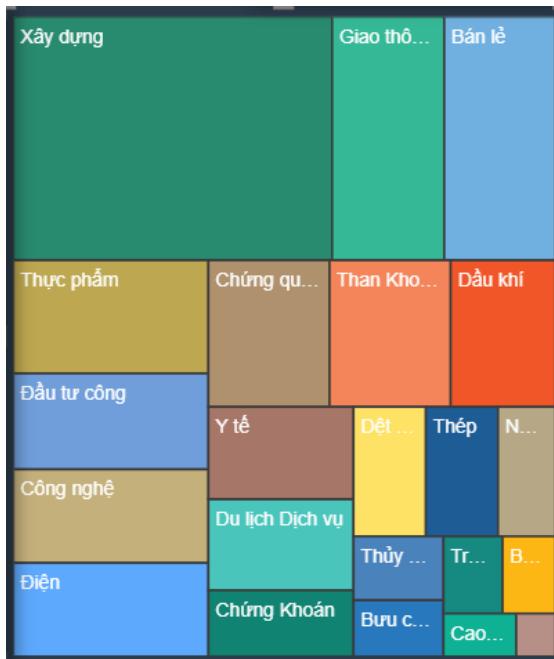
=> Table

Mã CK	Mở cửa	Trần (CE)	Sàn	Cao nhất	Thấp nhất	Đóng cửa	Trạng thái	Khối lượng
VND	54,15 ₫	19,86 ₫	16,25 ₫	58,00 ₫	53,85 ₫	57,90 ₫	▲ TĂNG	156.666.295,00
SSI	72,30 ₫	25,79 ₫	22,41 ₫	76,00 ₫	71,70 ₫	75,45 ₫	▲ TĂNG	102.972.342,00
SHB	36,60 ₫	13,42 ₫	10,98 ₫	37,60 ₫	36,60 ₫	37,50 ₫	▲ TĂNG	92.859.569,00
SHS	23,80 ₫	13,09 ₫	10,71 ₫	25,50 ₫	23,60 ₫	25,40 ₫	▲ TĂNG	82.820.373,00
MBB	59,70 ₫	21,29 ₫	18,51 ₫	61,45 ₫	59,40 ₫	61,35 ₫	▲ TĂNG	68.384.650,00
VND	19,20 ₫	21,12 ₫	17,28 ₫	19,50 ₫	18,10 ₫	18,10 ₫	▼ GIẢM	66.280.900,00
NVL	29,60 ₫	15,84 ₫	13,76 ₫	30,00 ₫	28,40 ₫	28,80 ₫	▼ GIẢM	64.834.900,00
DIG	63,45 ₫	22,63 ₫	19,67 ₫	64,75 ₫	62,40 ₫	64,05 ₫	▲ TĂNG	64.039.593,00
VIX	29,91 ₫	10,97 ₫	8,97 ₫	31,45 ₫	29,76 ₫	31,35 ₫	▲ TĂNG	61.964.840,00
HSG	47,70 ₫	17,01 ₫	14,79 ₫	48,70 ₫	47,10 ₫	48,30 ₫	▲ TĂNG	56.863.391,00
GEX	51,75 ₫	19,84 ₫	14,66 ₫	54,55 ₫	51,00 ₫	54,00 ₫	▲ TĂNG	54.325.522,00
DXG	43,80 ₫	15,62 ₫	13,58 ₫	44,80 ₫	43,35 ₫	44,70 ₫	▲ TĂNG	51.900.790,00
STB	84,15 ₫	30,01 ₫	26,09 ₫	85,00 ₫	83,25 ₫	84,90 ₫	▲ TĂNG	51.553.901,00
SHB	12,30 ₫	13,53 ₫	11,07 ₫	12,35 ₫	12,05 ₫	12,20 ₫	▼ GIẢM	45.928.500,00
GEX	19,15 ₫	22,02 ₫	16,28 ₫	20,10 ₫	17,90 ₫	17,90 ₫	▼ GIẢM	45.727.500,00
SHB	12,25 ₫	13,48 ₫	11,03 ₫	12,40 ₫	12,10 ₫	12,35 ₫	▲ TĂNG	44.911.100,00
VPB	59,25 ₫	23,70 ₫	15,80 ₫	60,10 ₫	59,10 ₫	59,55 ₫	▲ TĂNG	44.845.778,00
NVL	40,50 ₫	14,45 ₫	12,56 ₫	41,05 ₫	40,20 ₫	40,80 ₫	▲ TĂNG	43.481.275,00
VND	18,30 ₫	20,13 ₫	16,47 ₫	19,00 ₫	18,10 ₫	18,05 ₫	▲ TĂNG	43.027.600,00
HPG	22,70 ₫	24,29 ₫	21,11 ₫	23,15 ₫	22,70 ₫	22,80 ₫	▲ TĂNG	42.521.400,00
VND	17,25 ₫	18,98 ₫	15,53 ₫	17,45 ₫	17,05 ₫	17,40 ₫	▲ TĂNG	40.638.900,00
SHB	24,50 ₫	13,48 ₫	11,03 ₫	24,55 ₫	24,00 ₫	24,35 ₫	▼ GIẢM	38.873.600,00
VND	18,60 ₫	20,46 ₫	16,74 ₫	18,60 ₫	17,95 ₫	18,05 ₫	▼ GIẢM	38.173.900,00
Total	215.854,1 9 ₫			219.237, 77 ₫	212.854,00 đ	216.439,1 2 ₫		10.549.208.092,00

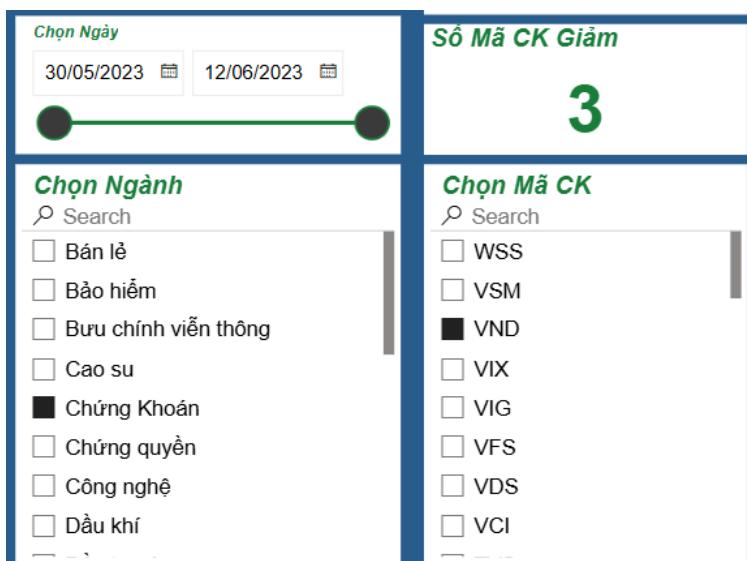
=> Gauge



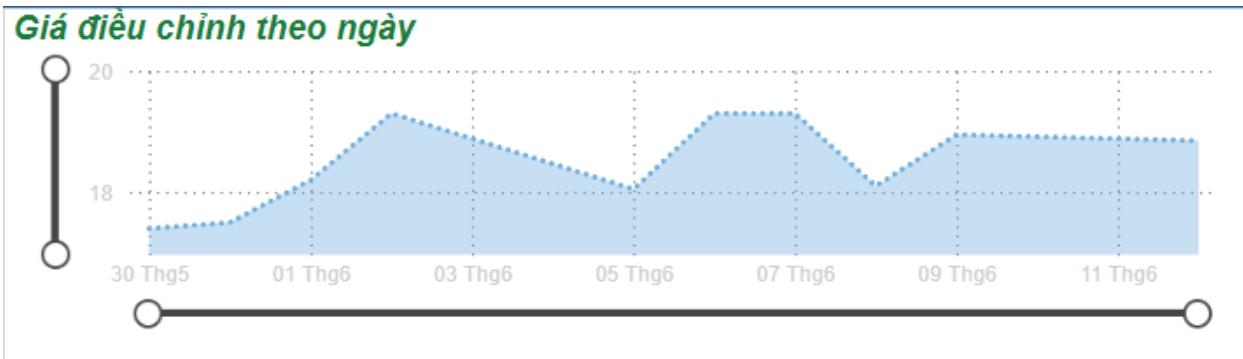
=> Treemap



=> Slicer (Ngày, Ngành, Mã CK)



=> Stacked area chart



=> Candlestick (biểu đồ nến đặc trưng trong giao dịch)



=> Line and stacked column chart



## 2.2.6 NHỮNG ĐIỀU CẦN LUU Ý

Những điều cần lưu ý khi kể chuyện bằng dữ liệu :

- Không nên chọn các giá trị tùy ý để làm cơ sở cho thang đo và đơn vị cho các biểu đồ

=> Giá trị được chọn phải có tính đại diện giúp người xem thấy được toàn cảnh vấn đề.

❑ Không nên chỉ thể hiện dữ liệu cho một góc nhìn

=> Cần cho người xem thấy bức tranh toàn cảnh của mọi vấn đề

❑ Không nên thay đổi màu sắc, nhãn và quy ước giữa các hình ảnh. Sự không thống nhất có thể gây nhầm lẫn, khó theo dõi chính xác vấn đề

=> Đỏ là giảm

=> Xanh lá cây là tăng

=> Vàng là tham chiếu (không đổi)

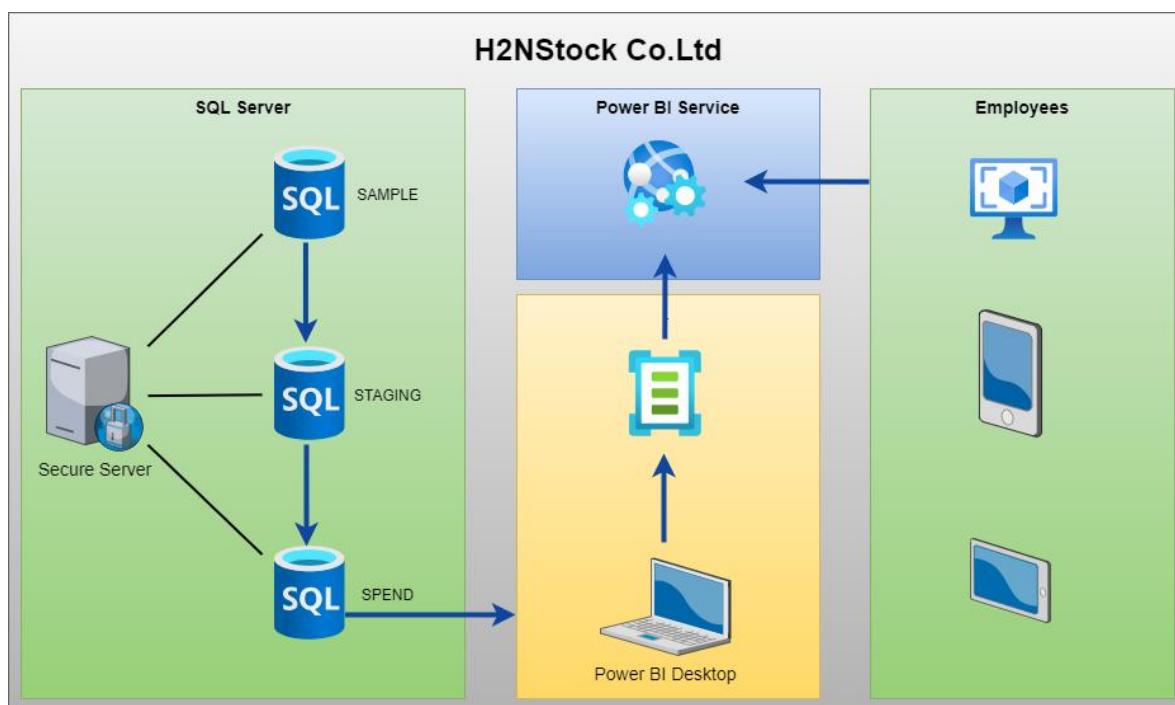
=> Tím là trần

=> Xanh lơ là sàn

## 2.3 KIẾN TRÚC HỆ THỐNG

### 2.3.1 KIẾN TRÚC

✓ *Trình bày kiến trúc hệ thống của dự án*



Hệ thống H2NStock được thiết kế với luồng xử lý và phân tích dữ liệu như sau:

- **Tầng dữ liệu:** SQL Server (SAMPLE → STAGING → SPEND).
- **Tầng xử lý/báo cáo:** Power BI Desktop + Power BI Service.
- **Tầng người dùng:** Truy cập qua nhiều thiết bị.

### 2.3.2 GIẢI THÍCH

- ✓ *Giải thích cách thiết kế.*

Hệ thống được thiết kế theo mô hình nhiều tầng nhằm đảm bảo **bảo mật, tính toàn vẹn dữ liệu và hiệu suất phân tích**:

#### 1. Tầng dữ liệu: SQL Server (SAMPLE → STAGING → SPEND)

- Gồm 3 tầng chính:
  - **SAMPLE:** Lưu dữ liệu mẫu ban đầu
  - **STAGING:** Lưu dữ liệu trung gian sau bước làm sạch và xử lý.
  - **SPEND:** Lưu dữ liệu đã sẵn sàng để phân tích.

#### 2. Xử lý & trực quan hóa (Power BI)

- **Power BI Desktop** kết nối trực tiếp với cơ sở dữ liệu **SPEND** để xây dựng báo cáo và dashboard.
- Các báo cáo sau đó được xuất bản lên **Power BI Service** để chia sẻ nội bộ.

#### 3. Truy cập & sử dụng (Employees)

- Nhân viên có thể truy cập báo cáo qua máy tính, máy tính bảng hoặc điện thoại, đảm bảo tính linh hoạt và tiện lợi trong việc theo dõi dữ liệu chứng khoán.

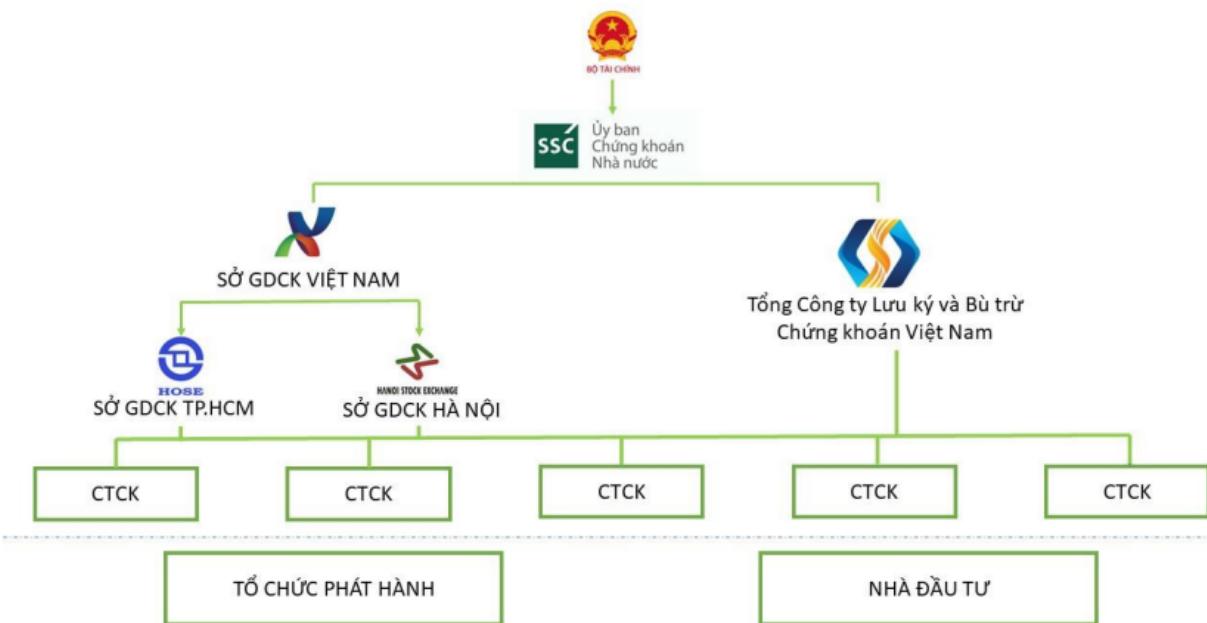
## 2.4 GIẢI THÍCH VỀ BỘ DỮ LIỆU KHÁCH HÀNG

- ✓ *Cung cấp kiến thức cơ bản về bộ dữ liệu cung cấp*

### 2.4.1 CÁC KHÁI NIỆM

- ✓ *Giải thích các khái niệm và nghiệp vụ liên quan*

### CẤU TRÚC CÁC CHỦ THỂ THAM GIA TRÊN TTCK VIỆT NAM



□ Việt Nam hiện tại có các sàn giao dịch là:

- ❖ HOSE: Sở Giao dịch Chứng khoán Thành phố Hồ Chí Minh
- ❖ HNX: Sở Giao dịch Chứng khoán Hà Nội
- ❖ Upcom: (Unlisted Public Company Market) là nơi tập trung các cổ phiếu chưa niêm yết và là nơi để các doanh nghiệp cổ phần “quá độ” lên các sàn giao dịch như HOSE, HNX. Sàn giao dịch Upcom được quản lý chính thức bởi sở giao dịch chứng khoán Hà Nội HNX.
- ❖ OTC: thị trường phi tập trung (Over The Counter), việc giao dịch chứng khoán không thông qua mặt bằng giao dịch như sàn chứng khoán được niêm yết. Các chứng khoán thuộc sàn OTC thực hiện việc định giá, chào bán cổ phiếu dựa vào thỏa thuận giữa các bên mua và bên bán.

□ Chứng khoán:

- ❖ Chứng khoán là một tài sản, bằng chứng xác nhận quyền sở hữu hợp pháp của người sở hữu đó với tài sản hoặc phần vốn của công ty hay tổ chức đã phát hành.

❖ Chứng khoán có thể giao dịch mua bán trên thị trường. Các công ty phát hành chứng khoán để huy động vốn từ thị trường. Nhà đầu tư sẽ rót tiền mua chứng khoán, đồng thời là chủ sở hữu một phần tài sản/ vốn của doanh nghiệp phát hành.

❖ Cổ phiếu: đây là sản phẩm giao dịch phổ biến nhất trên thị trường. Cổ phiếu được phát hành bởi các công ty Cổ phần, người sở hữu cổ phiếu được gọi là cổ đông và cũng là đồng sở hữu của công ty phát hành. Mã cổ phiếu có ký hiệu tối đa 3 chữ cái, là tên viết tắt của công ty phát hành.

❖ Trái phiếu: Nhà đầu tư trái phiếu đóng vai trò như người cho vay của doanh nghiệp phát hành. Do đó, doanh nghiệp có nghĩa vụ trả cho nhà đầu tư khoản tài sản xác định theo cam kết trên hợp đồng.

Ngành nghề

❖ Có rất nhiều ngành nghề trên thị trường

❖ Mỗi mã cổ phiếu thuộc 1 ngành nghề

Cổ phiếu

❖ Mỗi mã cổ phiếu chỉ được niêm yết trên 1 sàn giao dịch

❖ Giá cổ phiếu sẽ thay đổi theo thời gian thực trong phiên giao dịch

❖ Khi muốn chuyển sàn giao dịch phải làm thủ tục chuyển sàn

Phiên giao dịch

❖ Từ thứ Hai đến thứ Sáu (trừ ngày Lễ, thứ Bảy, CN)

❖ Bắt đầu từ 9:00 đến 14:45

❖ Lệnh ATO là lệnh giao dịch tại giá khớp lệnh xác định giá mở cửa, chỉ áp dụng cho sàn chứng khoán TP HCM (9:00 đến 9:15)

❖ Lệnh ATC (viết tắt của At The Close) là lệnh mua hoặc bán chứng khoán tại mức giá đóng cửa dù đó là bất kỳ giá nào (14:30 đến 14:45)

## 2.4.2 CÁC TRƯỜNG DỮ LIỆU

✓ *Giải thích các trường dữ liệu có trong bộ dữ liệu*

STT	Tên trường dữ liệu	Nội dung
1	MaCK	Mã chứng khoán (Ticker symbol) hay Mã cổ phiếu (Stock symbol) là một ký tự viết tắt được sử dụng để xác định duy nhất (đặc định) các cổ phiếu được giao dịch công khai của một cổ phiếu cụ thể trên một thị trường chứng khoán cụ thể.
2	NGAYGIAODICH	Ngày giao dịch của thị trường (từ thứ Hai đến thứ Sáu).
3	GIAMOCUA	Giá mở cửa hay còn gọi là giá tham chiếu, giá khi khởi đầu của ngày giao dịch.
4	GIACAONHAT	Giá cao nhất của một mã chứng khoán trong ngày giao dịch.
5	GIATHAPNHAT	Giá thấp nhất của một mã chứng khoán trong ngày giao dịch.
6	GIADONGCUA	Giá đóng cửa hay còn gọi là giá cuối ngày, khi thị trường đóng cửa.
7	KHOILUONGGIAODICH	Khối lượng giao dịch là tổng số lượng cổ phiếu được mua bán trong ngày, của một mã cổ phiếu.
8	TENNHOMEANGANH	Tên nhóm ngành trên thị trường, mỗi một công ty chỉ thuộc một ngành.
9	MANHOMNGANH	Mã nhóm ngành.
10	THONGTINCONGTY	Chứa đường dẫn giới thiệu thông tin chi tiết về công ty niêm yết trên thị trường.
11	SAN	Sàn giao dịch chứng khoán (stock exchange, securities exchange hoặc bourse) là một hình thức sàn giao dịch cung cấp cho những người môi giới cổ phiếu và người mua bán cổ phiếu một nền tảng và phương tiện để trao đổi và mua bán

		các cổ phiếu, trái phiếu cũng như các loại chứng khoán khác. Việt Nam hiện tại có các sàn giao dịch chính là HOSE, HNX, Upcom, OTC (thị trường phi tập trung).
12	CONGTY	Tên công ty niêm yết trên thị trường.
13	TENSAN_VIET	Tên sàn giao dịch tiếng Việt.
14	TENSAN_ANH	Tên sàn giao dịch tiếng Anh.
15	BIENDODAODONG	Biên độ giao động là tỷ lệ tăng hoặc giảm tối đa của một mã chứng khoán trên sàn trong một ngày: HOSE: $\pm 7\%$ , HNX: $\pm 10\%$ , Upcom: $\pm 15\%$ , OTC: $\pm 20\%$ .
16	GIADIEUCHINHDONGCUA	Giá đóng cửa điều chỉnh (Adjusted Closing Price) là giá đóng cửa của cổ phiếu được điều chỉnh để phản ánh chính xác giá trị của cổ phiếu đó sau khi hạch toán cho bất kỳ hành động nào của công ty.

### 3 LÀM SẠCH VÀ CHUYỂN ĐỔI DỮ LIỆU

- ✓ *Bao gồm 2 bước:*
  - *Làm sạch dữ liệu*
  - *Chuyển đổi dữ liệu*

#### 3.1 CHUẨN BỊ DỮ LIỆU

- ✓ *Quá trình chuẩn bị dữ liệu thường bao gồm việc chuẩn hóa các định dạng dữ liệu, kết hợp các bộ dữ liệu khác nhau để tăng độ phong phú của dữ liệu nguồn và loại bỏ các giá trị ngoại lai (outliers)*

##### 3.1.1 GIẢI PHÁP LUU TRỮ DỮ LIỆU

- ✓ *So sánh giải pháp nền tảng đám mây (Cloud-based solution) hay các ứng dụng tại chỗ (on-premise).*
- ✓ *Quyết định chọn giải pháp nào cho dự án? Lý do?*

Khi dữ liệu và quy trình được chuyển lên nền tảng đám mây (cloud native) sẽ đem lại nhiều lợi ích

- ❖ Khả năng mở rộng vượt trội. Việc chuẩn bị dữ liệu trên đám mây có thể bắt kịp với tốc độ của doanh nghiệp.
- ❖ Việc chuẩn bị dữ liệu đám mây sẽ tự động nâng cấp để có thể bật các tính năng mới hoặc các bản sửa lỗi ngay sau khi chúng được phát hành.
- ❖ Thúc đẩy sử dụng và hợp tác về dữ liệu. Chuẩn bị dữ liệu trên đám mây đồng nghĩa với trạng thái luôn luôn hoạt động, không yêu cầu bất kỳ cài đặt kỹ thuật nào và cho phép các nhóm cùng cộng tác đầy nhanh tiến độ và hiệu quả.

Phạm vi dự án

- ❖ Trong phạm vi môn học Dự Án Mẫu, dữ liệu sẽ không lưu trữ trên Cloud.
- ❖ Để tiết kiệm chi phí, dữ liệu sẽ được lưu trữ trên server nội bộ (On-premise).

### 3.1.2 GIẢI PHÁP PHÂN BỐ DỮ LIỆU

- ✓ *Nhân bản là một kỹ thuật quan trọng và hữu hiệu trong việc phân bố cơ sở dữ liệu (CSDL) và thực thi các Stored procedure. Đây là chủ đề rất rộng, chỉ tập trung chọn giải pháp trong phạm vi dự án.*

#### 3.1.2.1 Ý nghĩa việc phân bố dữ liệu

- Giảm tải cho việc báo cáo, các truy vấn đang chạy trên CSDL nhân bản có thể giải phóng tài nguyên cho CSDL chính, giúp tăng hiệu năng của hệ thống.
- Bảo vệ dữ liệu khỏi các lỗi quản trị: trong trường hợp xảy ra lỗi trong quá trình thao tác, bạn có thể dùng CSDL đã nhân bản trước đó để copy dữ liệu qua lại.
- Bảo vệ dữ liệu trước lỗi người dùng: giảm thiểu tác động các lỗi do người dùng chính, chẳng hạn như lỗi xóa bảng, cập nhật field nhầm ...
- Quản lý các dữ liệu thử nghiệm (database for QC testing): trong môi trường thử nghiệm, sẽ hữu ích khi liên tục chạy một giao thức thử nghiệm cho cơ sở dữ liệu để chứa các dữ liệu giống hệt nhau khi bắt đầu mỗi vòng thử nghiệm.
- Tăng độ tin cậy (tính sẵn sàng) cho hệ thống: trong quá trình đọc hoặc ghi dữ liệu, nếu dữ liệu đó bị lỗi hay vì một nguyên nhân nào đó mà không thể dùng được, ta có thể dùng ngay bản sao dữ liệu đó để hệ thống không phải dừng lại và tránh được tình trạng sử dụng các dữ liệu không chính xác.

### 3.1.2.2 Trình bày cách phân bổ dữ liệu

✓ *Trình bày cách phân bổ CSDL trong phạm vi dự án này.*

❑ *Trong phạm vi dự án và môn Dự Án Mẫu*

❖ Để tiết kiệm chi phí.

❖ Quy mô dự án nhỏ.

❖ Để thao tác.

❖ **Sẽ chọn phương án nhân bản CSDL trên cùng một server.**

❑ *Cách nhân bản CSDL*

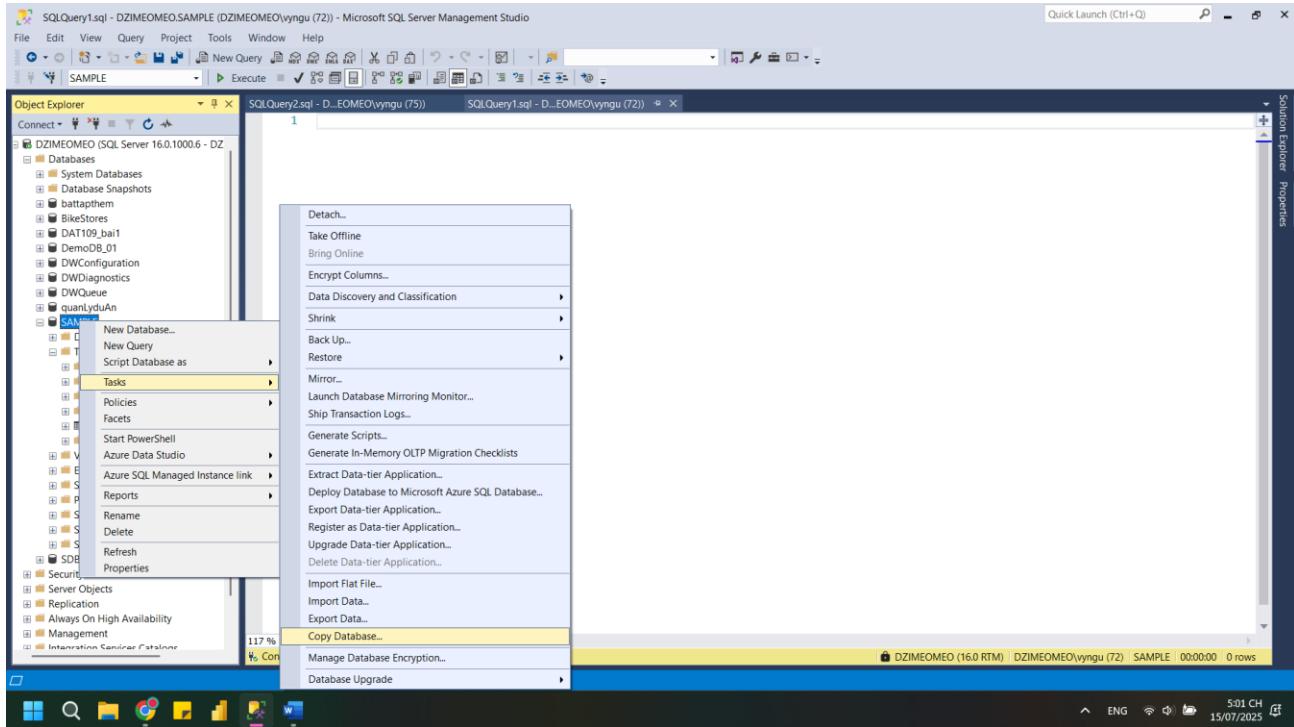
❖ Tạo Staging database từ SAMPLE database ở Lab1 bằng trình wizard.

❖ Tạo Staging database từ SAMPLE database ở Lab1 bằng Script.

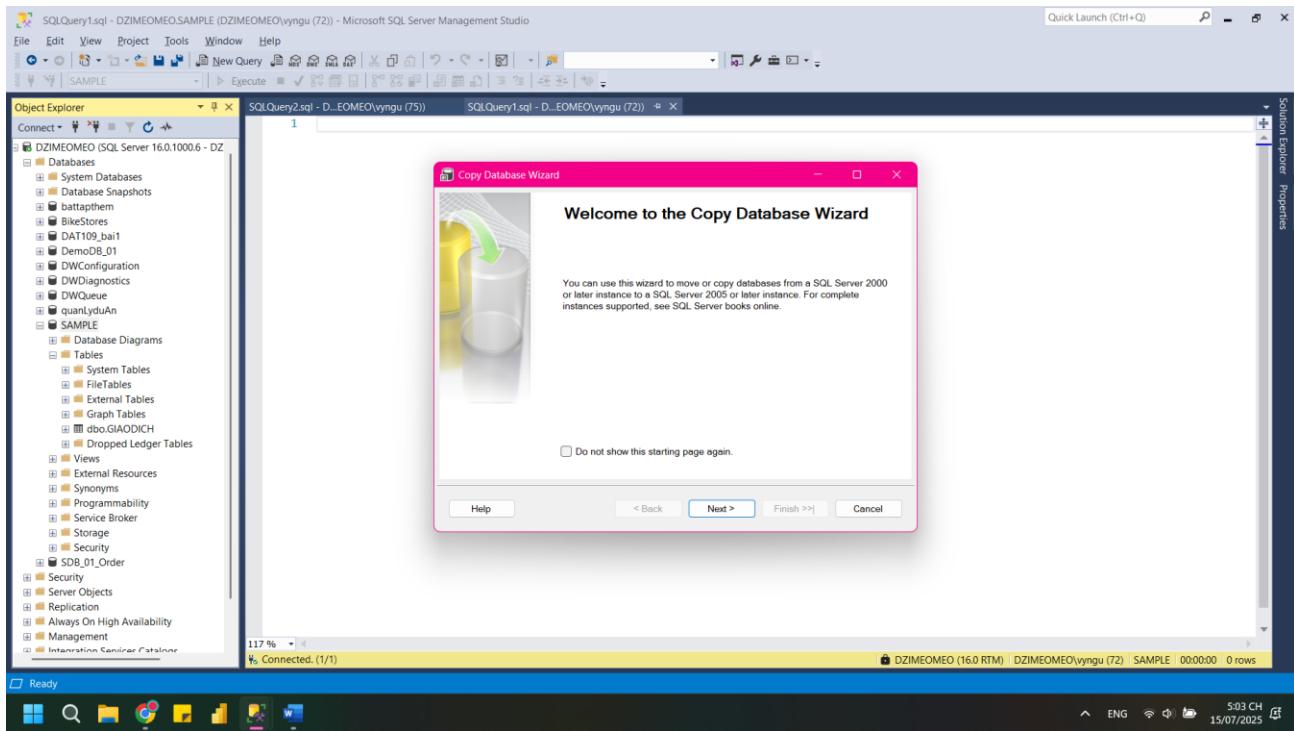
#### Cách 1:

Tạo Staging database từ SAMPLE database ở Lab1 bằng trình wizard.

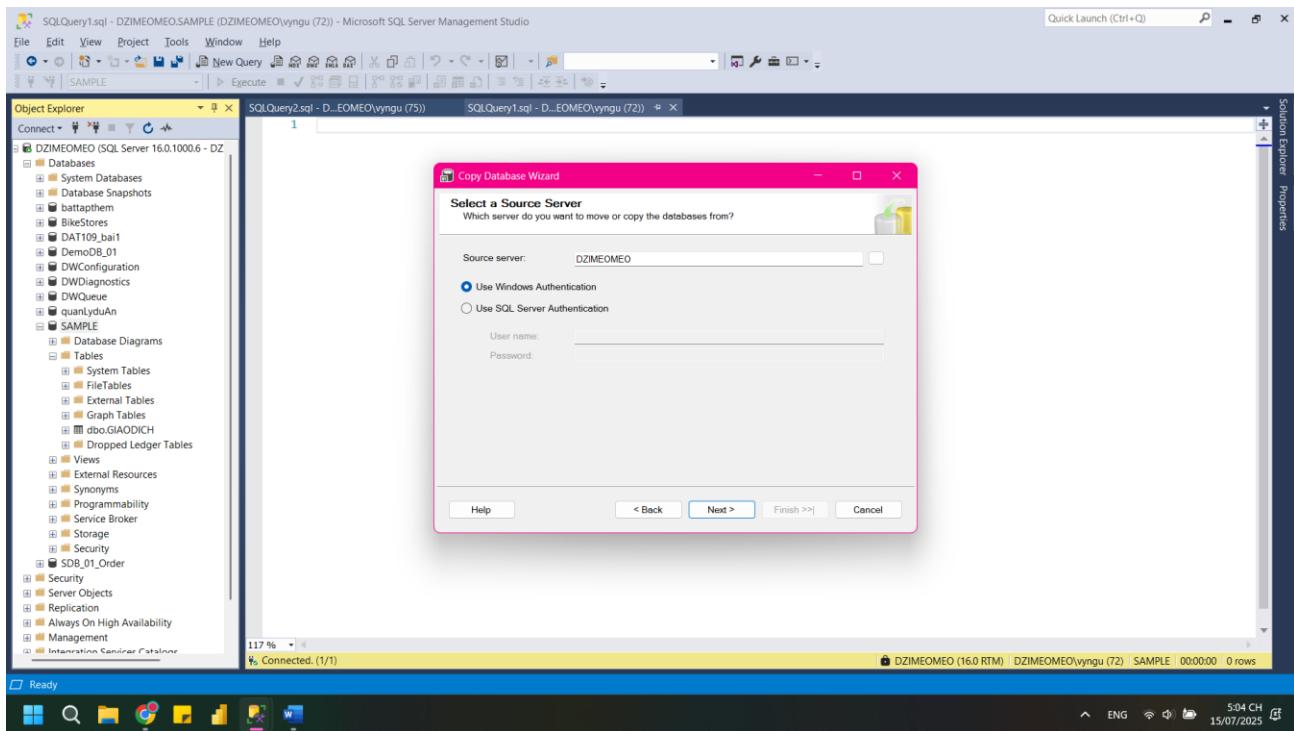
- Chọn database SAMPLE cần nhân bản. Chọn Task >> Copy Database.



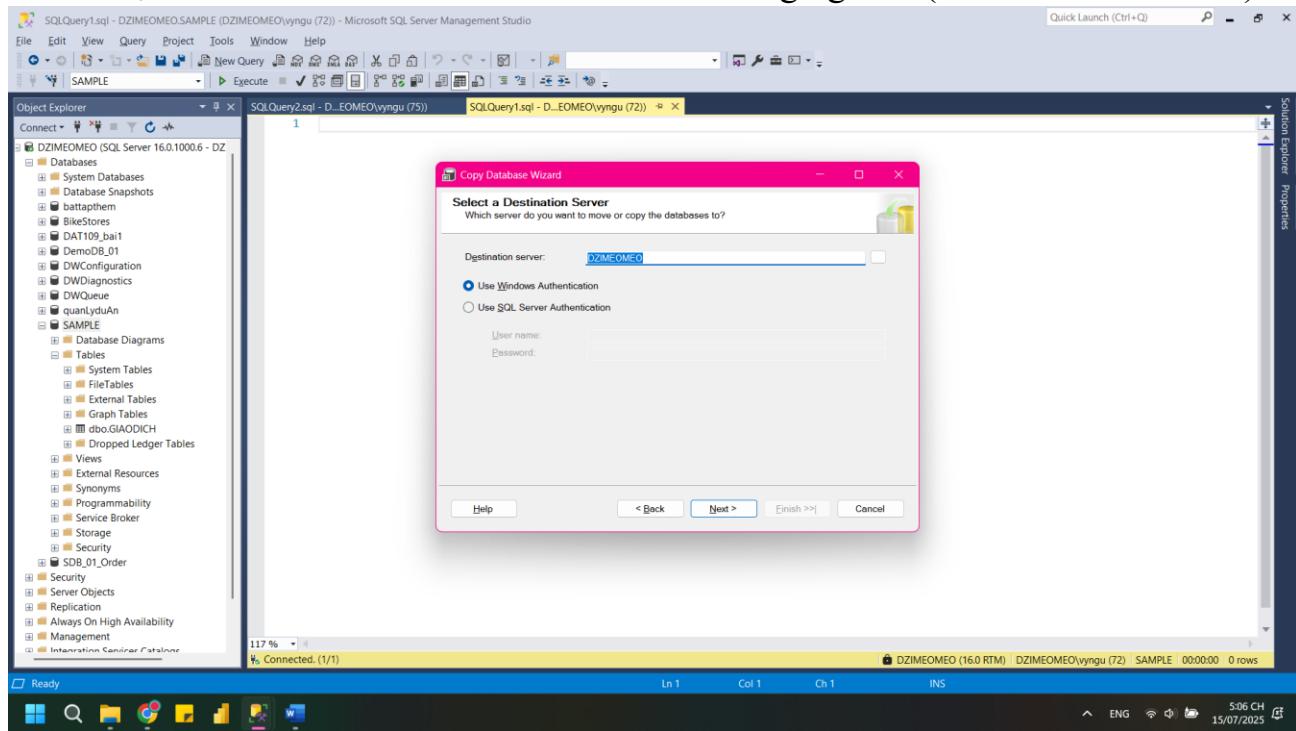
- Chọn Next ở màn hình Welcome.



- Chọn server chứa database SAMPLE. ( source )

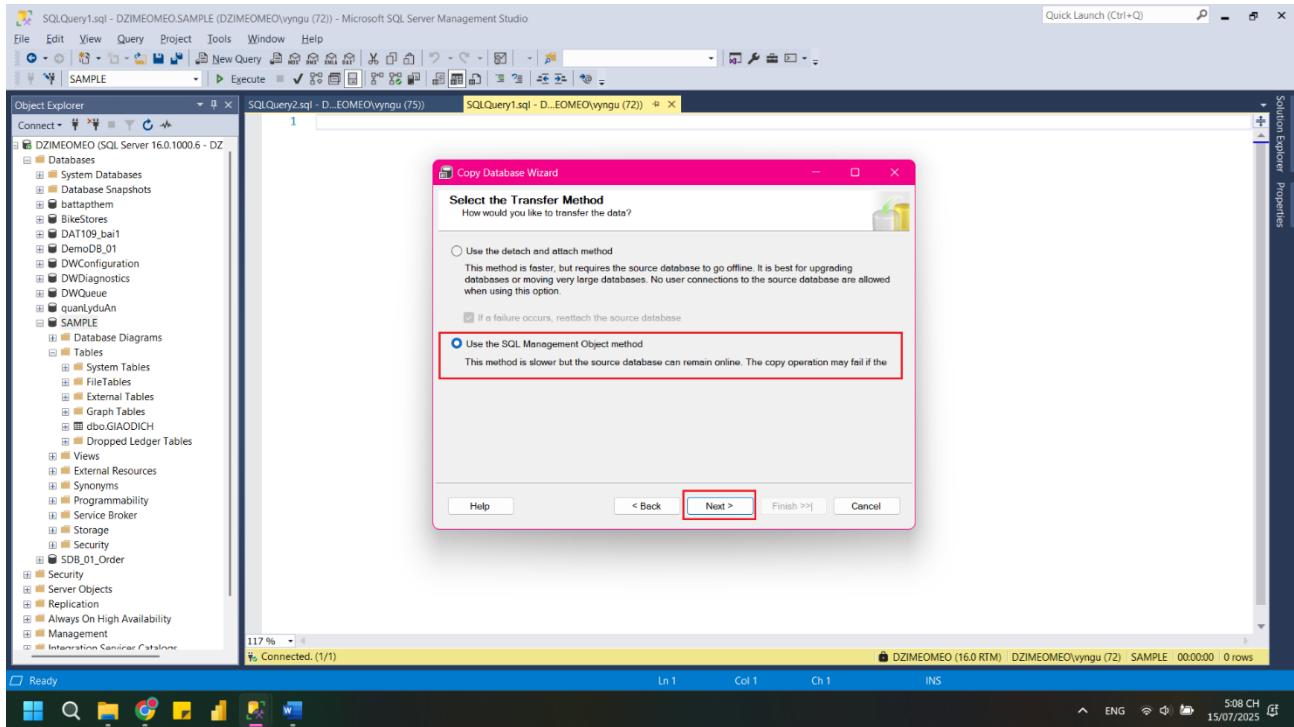


- Chọn server chứa database Staging. ( Destination )

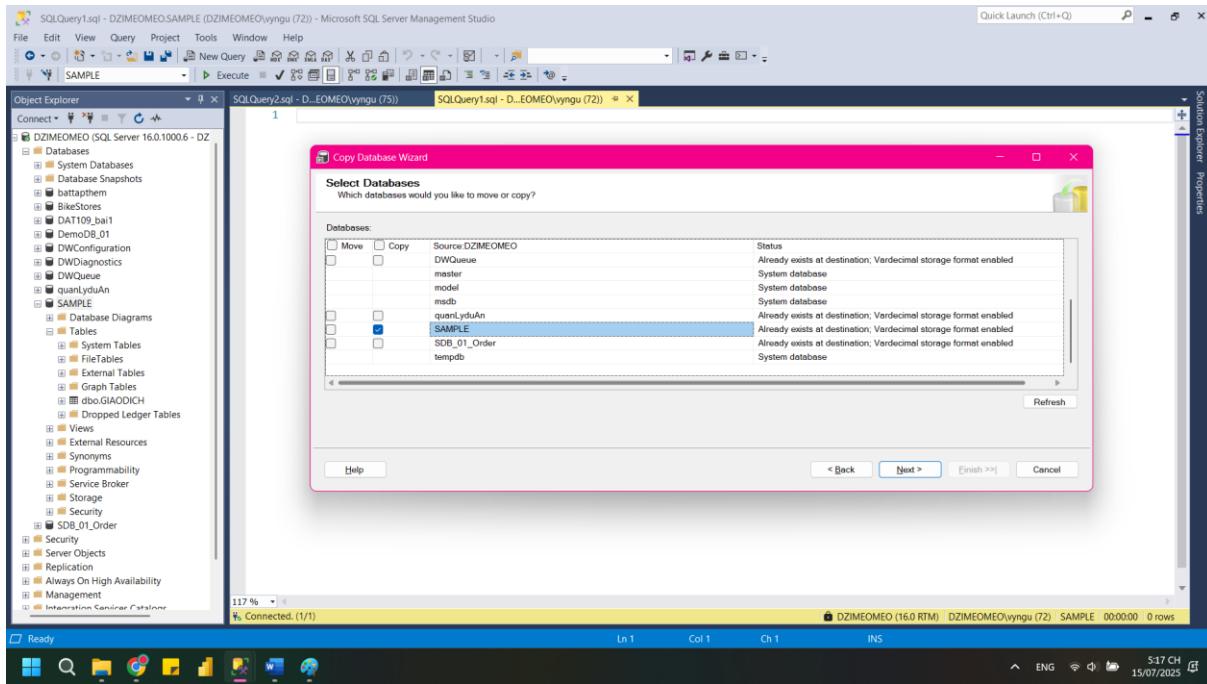


- Chọn Phương Thức chuyển dữ liệu.

- Chọn option 1: dành cho database có dữ liệu lớn, database sẽ offline trong lúc tiến trình thực thi.
- **Chọn option 2:** tốc độ chậm hơn nhưng database vẫn online trong lúc tiến trình thực thi.  
⇒ Trong phạm vi môn học, **khuyến nghị** cho option 2. **Nếu chọn option 1 sẽ ảnh hưởng tới các connection đang dùng database này.**



- Chọn database cần copy. Nhấn Next.

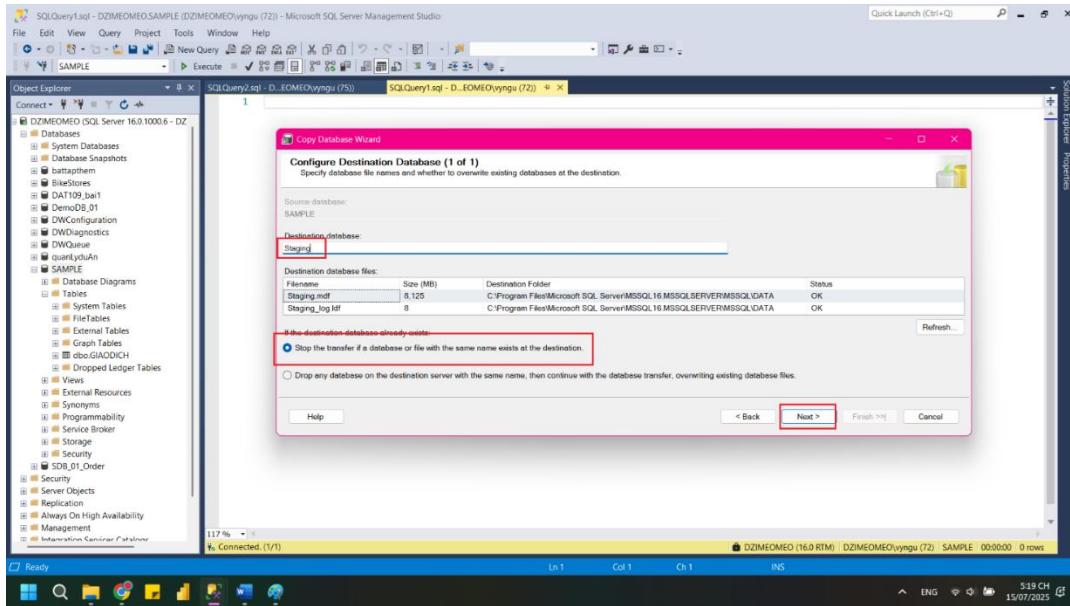


- Đặt tên cho database mới là Staging. Kiểm tra đường dẫn database Staging.

**Chọn option 1: dùng ngay tiến trình nếu Staging đã tồn tại.**

**Chọn option 2: ghi đè dữ liệu nếu Staging đã tồn tại.**

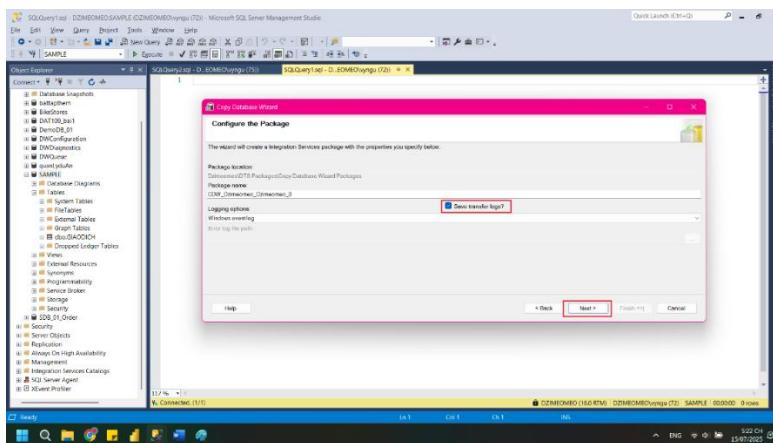
⇒ Trong phạm vi môn học, khuyến nghị cho option 1. Thực tế khi gặp bất cứ issues gì trong quá trình copy thì nên dừng lại để kiểm tra xem xét, tránh mất dữ liệu, ghi đè, ảnh hưởng tới công việc của nhóm khác.



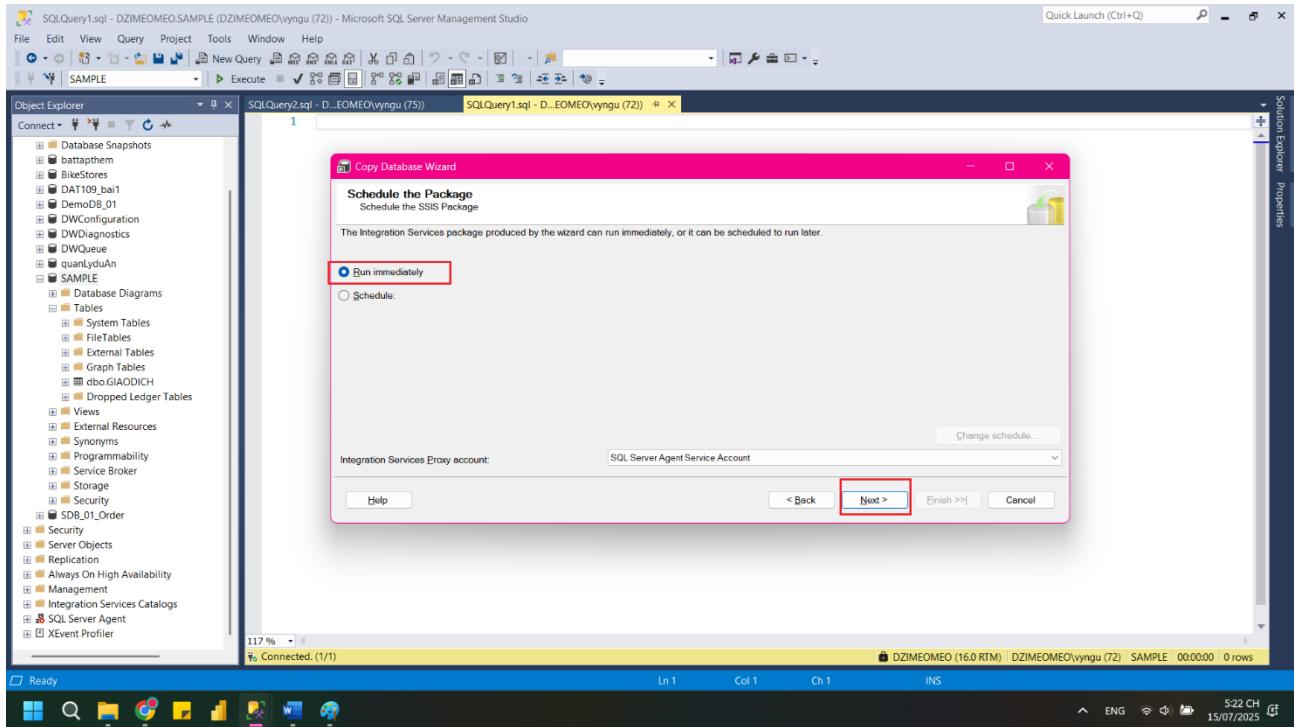
- Hệ thống sẽ tự động tạo ra 1 package để transfer dữ liệu (không cần sửa tên package mà hệ thống tự sinh ra)

Chọn Save Transfer Log nhằm mục đích lưu lại log quá trình transfer, giúp dễ truy vết khi xảy ra sự cố.

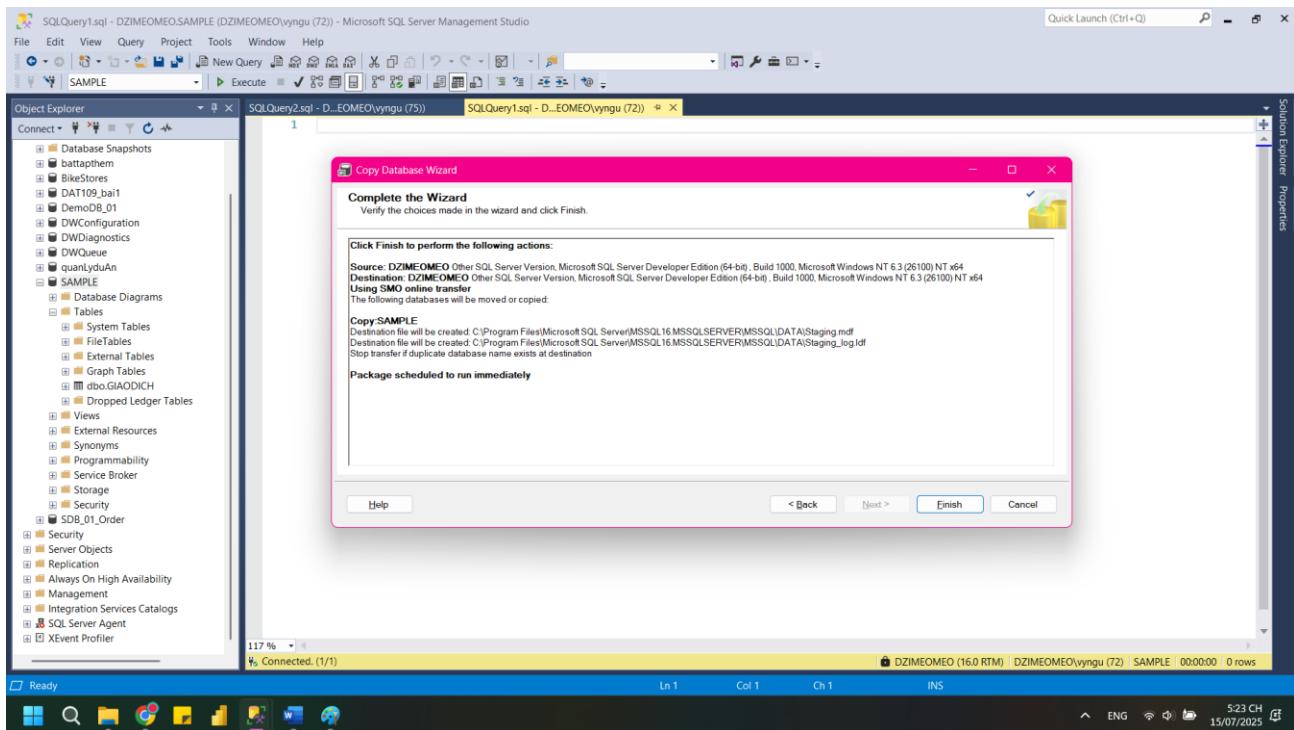
Nhấn Next.



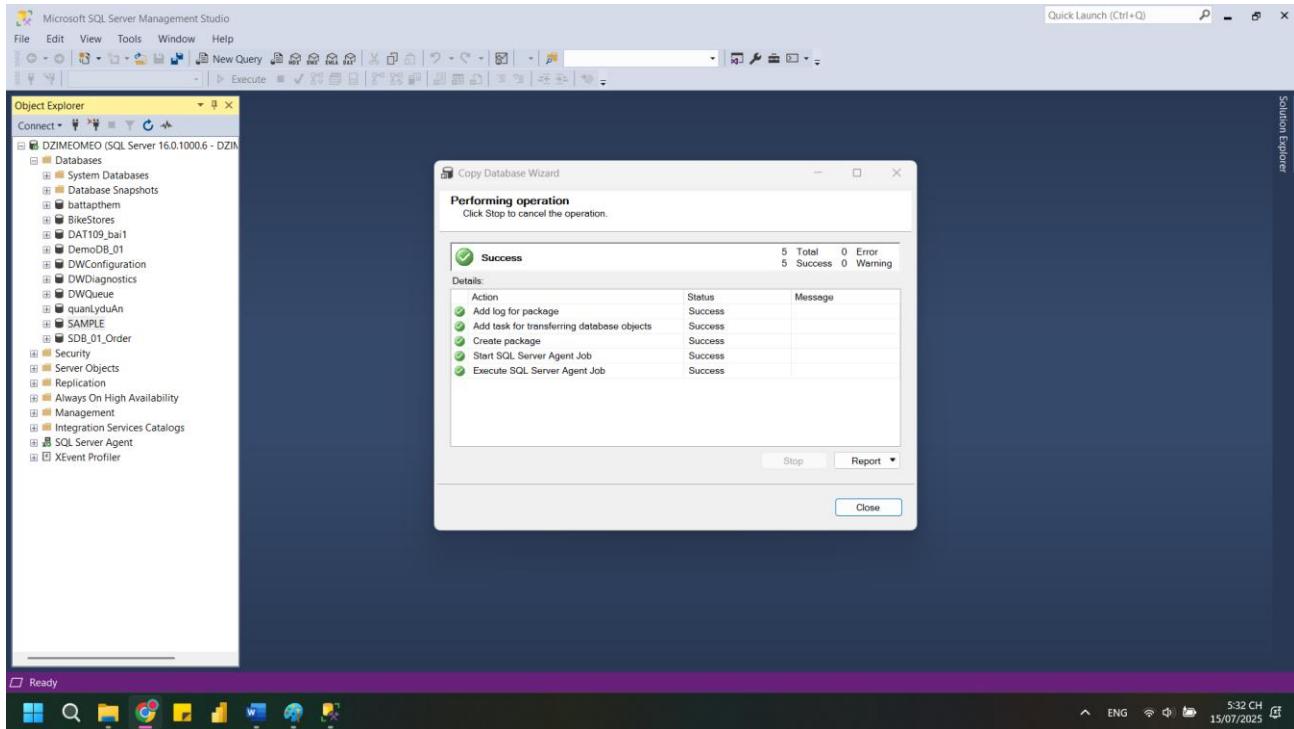
- Chọn option Run immediately. Nhấn Next.



- Kiểm tra lại thông tin 1 lần nữa. Nhấn Finish.



- Nếu không có lỗi xảy ra thì database Staging sẽ được tạo.



SQLQuery3.sql - DZIMEOME0.Staging (DZIMEOME0\vyngu (75)) - Microsoft SQL Server Management Studio

```

SELECT TOP (1000) [MaCK]
      ,[NGAYGIAODICH]
      ,[GIAMOCUA]
      ,[GIACAOINHAT]
      ,[GIATHAPINHAT]
      ,[GIADONGCUA]
      ,[KHOLUONGGIAODICH]
      ,[TENHOMMINGANH]
      ,[MAHOMMINGANH]
      ,[THONGTINCONGTY]
      ,[SAN]
      ,[CONGTY]
      ,[TENSAN_VIET]
      ,[TENSAN_ANH]
      ,[BENDODAODONG]
  FROM [Staging].[dbo].[GIAODICH]

```

MaCK	NGAYGIAODICH	GIAMOCUA	GIACAOINHAT	GIATHAPINHAT	GIADONGCUA	KHOLUONGGIAODICH	TENHOMMINGANH	MAHOMMINGANH	THONGTINCONGTY	SAN	CONGTY
A32	20230530	30.1	11.1	10.95	11	5553000	Dệt may mặc	DETMAC	http://cafed.vn/vnupcom/A32-cong-ty-co-phan-32.chm	Upcom	Công ty cổ phần 32
AAA	20230530	11.1	11.55	11.55	11.4	60211000	Cao su	CAOSU	http://cafed.vn/vnhouse/AAA-cong-ty-co-phan-nhua...	HOSE	Công ty Cổ phần Nhựa và Môi trường
AAA	20230530	11.55	11.55	11.25	11.55	6809100	Cao su	CAOSU	http://cafed.vn/vnhouse/AAA-cong-ty-co-phan-nhua...	HOSE	Công ty Cổ phần Nhựa và Môi trường
AAA	20230601	11.45	11.6	11.25	11.55	6809100	Cao su	CAOSU	http://cafed.vn/vnhouse/AAA-cong-ty-co-phan-nhua...	HOSE	Công ty Cổ phần Nhựa và Môi trường
AMM	20230530	10.9	10.9	10.8	10.85	3700	Thủy hải sản	THUHAI SAN	http://cafed.vn/vnhouse/AMM-cong-ty-co-phan-thuy...	HOSE	Công ty Cổ phần Thủ sản Mekong
AMM	20230531	10.75	10.85	10.75	10.85	4200	Thủy hải sản	THUHAI SAN	http://cafed.vn/vnhouse/AMM-cong-ty-co-phan-thuy...	HOSE	Công ty Cổ phần Thủ sản Mekong
AMM	20230601	10.65	10.9	10.25	10.85	7400	Thủy hải sản	THUHAI SAN	http://cafed.vn/vnhouse/AMM-cong-ty-co-phan-thuy...	HOSE	Công ty Cổ phần Thủ sản Mekong
AMM	20230530	12.6	13	12.6	12.9	777807	Chứng khoán	OK	http://cafed.vn/vnstock/OK-cong-ty-co-phan-chung...	OTC	Công ty Cổ phần Chứng khoán Á Âu
AMM	20230531	12.8	13.8	12.8	13.8	1821345	Chứng khoán	OK	http://cafed.vn/vnstock/OK-cong-ty-co-phan-chung...	OTC	Công ty Cổ phần Chứng khoán Á Âu
AAS	20230601	13.5	14.1	13.5	14	1039910	Chứng khoán	OK	http://cafed.vn/vnstock/AAS-cong-ty-co-phan-chung...	OTC	Công ty Cổ phần Chứng khoán Á Âu
ATT	20230530	5.95	6	5.7	5.82	2821200	Bán lẻ	BANLE	NULL		
ATT	20230531	5.82	6.22	5.76	6.22	2709400	Bán lẻ	BANLE	NULL		
ATT	20230601	6.5	6.65	6.31	6.42	3059600	Bán lẻ	BANLE	NULL		
AVV	20230530	5.8	5.8	5.3	5.7	1169269	Bán lẻ	BANLE	https://cafed.vn/hastc/AVV-cong-ty-co-phan-ava...	HNX	Công ty Cổ phần AAV/
AVV	20230531	5.7	5.9	5.6	5.7	1339831	Bán lẻ	BANLE	https://cafed.vn/hastc/AVV-cong-ty-co-phan-ava...	HNX	Công ty Cổ phần AAV/

## Cách 2: Tạo Staging database từ SAMPLE database ở Lab1 bằng Script.

- Viết script copy SAMPLE database to Staging database như hình bên dưới.

Tiến trình thực hiện:

- Hệ thống sẽ backup SAMPLE database ở thời điểm hiện tại.
  - Sau đó sẽ restore tại đường dẫn mới và đổi tên với option Replace.
  - Đổi đường dẫn file lưu trữ file Sample.bak ( D:\FPT Polytechnic\DAT109\4.Resources )
  - Đổi đường dẫn file cài đặt sql ( C:\SQL\ )

SQLQuery2.sql - DZIMEOME0.master (DZIMEOME0\vyngu (59)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Object Explorer

DZIMEOME0 (SQL Server 16.0.1000 - DZIM)

Databases

Staging

Views

External Resources

Synonyms

Programmability

Query Store

Service Broker

Storage

Security

Server Objects

Replication

Always On High Availability

Management

Integration Services Catalogs

SQL Server Agent

XEvent Profiler

SQLQuery2.sql - D...EOME0\vyngu (59)\* X SQLQuery1.sql - D...EOME0\vyngu (79)\*

```
12 DECLARE @destDb nvarchar(50);
13 DECLARE @destMdf nvarchar(100);
14 DECLARE @destLdf nvarchar(100);
15 DECLARE @sqlServerDbFolder nvarchar(100);
16
17 SET @sourceDb = 'SAMPLE'
18 SET @sourceDb_log = @sourceDb + '_log'
19 SET @backupPath = 'D:\FPT Polytechnic\DAT109\4. Resources' + @sourceDb + '.bak'
20 SET @sqlServerDbFolder = 'C:\SQL\'
21 SET @destDb = 'Staging'
22 SET @destMdf = @sqlServerDbFolder + @destDb + '.mdf'
23 SET @destLdf = @sqlServerDbFolder + @destDb + '_log' + '.ldf'
24
25 BACKUP DATABASE @sourceDb TO DISK = @backupPath
26
27 RESTORE DATABASE @destDb FROM DISK = @backupPath
WITH REPLACE,
MOVE @sourceDb TO @destMdf,
MOVE @sourceDb_log TO @destLdf
31
```

106 %

Messages

```
WITH REPLACE,
MOVE @sourceDb TO @destMdf,
MOVE @sourceDb_log TO @destLdf
...
Processed 1035 pages for database 'SAMPLE', file 'SAMPLE' on file 2.
Processed 2 pages for database 'SAMPLE', file 'SAMPLE_log' on file 2.
BACKUP DATABASE successfully processed 1034 pages in 0.039 seconds (207.031 MB/sec).
Processed 1032 pages for database 'Staging', file 'SAMPLE' on file 1.
Processed 2 pages for database 'Staging', file 'SAMPLE_log' on file 1.
RESTORE DATABASE successfully processed 1034 pages in 0.028 seconds (278.421 MB/sec).
```

Completion time: 2025-07-21T18:12:13.0809060+07:00

106 %

Query executed successfully.

DZIMEOME0 (16.0 RTM) DZIMEOME0\vyngu (59) master 00:00:00 0 rows

Ready

612 CH 21/07/2025

⇒ Thành công copy SAMPLE database to Staging database

### 3.2 LÀM SẠCH DỮ LIỆU

- ✓ *Kỹ thuật làm sạch, tầm quan trọng, lợi ích và tiêu chí đánh giá dữ liệu.*

### 3.2.1 CÁC VẤN ĐỀ ẢNH HƯỞNG TỚI DỮ LIỆU

- ✓ *Trình bày các vấn đề ảnh hưởng.*
  - ✓ *Vấn đề nào đang tồn tại trong dự án này.*

- ❖ Dữ liệu bị trùng lặp (duplicate data): Có hai hoặc nhiều record có thông tin giống hệt nhau. Điều này có thể gây ra sai lệch về số lượng hàng tồn kho, trùng lặp tài sản thế chấp.

- ❖ Dữ liệu bị mâu thuẫn (conflict data): Khi cùng một record nhưng lại tạo ra hai dòng dữ liệu với hai thuộc tính khác nhau, dữ liệu lúc này đang ở trạng thái bị mâu

thuần. Ví dụ, một công ty chỉ có một trụ sở nhưng trong database lại có hai địa chỉ khác nhau, điều này có thể gây khó khăn cho shipper.

- ❖ Dữ liệu không đầy đủ (incomplete data): Dữ liệu bị thiếu một số thuộc tính. Ví dụ, bảng lương của nhân viên không có dữ liệu về số an sinh xã hội cá nhân.
- ❖ Dữ liệu không hợp lệ (invalid data): Một thuộc tính dữ liệu không phù hợp với các tiêu chuẩn dữ liệu đã đặt ra. Ví dụ: số điện thoại được lưu với 9 chữ số thay vì 10 chữ số.

Trong dự án này gặp các vấn đề sau:

- Dữ liệu trùng lặp
- Dữ liệu không hợp lệ (các ngày sai hoặc các giá trị âm)

### **3.2.2 CÁC TIÊU CHÍ ĐÁNH GIÁ CHẤT LƯỢNG DỮ LIỆU**

- ✓ *Trình bày các tiêu chí.*
- ✓ *Tiêu chí nào sẽ áp dụng trong dự án này.*

#### Tính chuẩn xác (validity)

- ❖ Mandatory constraints (Bắt buộc về tính đầy đủ của dữ liệu): Một số trường thông tin trong bảng bắt buộc không được để trống.
- ❖ Data-type constraints (Thống nhất về định dạng dữ liệu): Các giá trị trong một cột phải thuộc một dạng dữ liệu nhất định.
- ❖ Range constraints (Giới hạn phạm vi dữ liệu): Giá trị tối thiểu và tối đa cho dạng dữ liệu số hoặc ngày tháng.
- ❖ Foreign-key constraints (Thống nhất về foreign-key): đảm bảo mỗi liên kết giữa các bảng với nhau, giá trị của cột foreign-key phải tương ứng với giá trị primary-key ở bảng khác.
- ❖ Unique constraints (Yêu cầu về tính duy nhất): Một số cột trong bảng phải là duy nhất trong bảng dữ liệu, ví dụ như cột customer\_ID trong bảng dim\_customer.

❖ Regular expression patterns (Ràng buộc về cách thức diễn đạt): Các trường dữ liệu có định dạng free-text cần có quy tắc xác thực theo cách này, để dữ liệu được lưu trữ một cách thống nhất.

❖ Cross-field validation (Xác thực chéo giữa các trường dữ liệu): Một số cột dữ liệu lặp lại giữa các bảng khác nhau cần được xác thực chéo để đảm bảo tính đúng đắn về các giá trị dữ liệu.

❑ Tính chính xác (accuracy): Phản ánh mức độ trung thực của dữ liệu, các giá trị đo được trong bảng dữ liệu có gần với giá trị đích/ giá trị thực tế hay không.

❑ Tính đầy đủ (completeness): Mức độ chi tiết và toàn diện của dữ liệu về một vấn đề, chủ đề.

❑ Tính nhất quán (consistency): Mức độ thống nhất của dữ liệu khi đổi chiếu trên toàn bộ hệ thống dữ liệu và công cụ lưu trữ khác nhau.

❑ Tính đồng nhất (uniformity): Đơn vị đo lường giống nhau được sử dụng trong tất cả các hệ thống dữ liệu.

❑ Tính minh bạch của nguồn dữ liệu (traceability): Dữ liệu cần đảm bảo có thể truy tìm lại nguồn thu thập.

❑ Tính kịp thời (timeliness): Dữ liệu phải được thu thập và cập nhật từ các nguồn mới nhất, “up-to-date” nhất.

❑ Đảm bảo đầy đủ những tiêu chí này có thể giúp doanh nghiệp có dữ liệu chất lượng tốt và sử dụng cho nhiều mục đích khác nhau.

### 3.2.3 CÁC BƯỚC LÀM SẠCH DỮ LIỆU

✓ *Quá trình làm sạch dữ liệu bao gồm một số bước xác định và sửa chữa các mục bị lỗi.*

#### 3.2.3.1 Trình bày các bước làm sạch

Quá trình làm sạch dữ liệu bao gồm một số bước xác định và sửa chữa các mục bị lỗi:

❑ Dữ liệu trùng lặp

=> Loại bỏ thông tin trùng lặp.

Dữ liệu không liên quan

=> Xác định các trường quan trọng đối với phân tích và loại bỏ dữ liệu không liên quan khỏi phân tích.

Dữ liệu ngoại lai

=> Dữ liệu ngoại lai có thể ảnh hưởng đáng kể đến hiệu suất của mô hình, vậy nên cần phải xác định các dữ liệu ngoại lai và tiến hành biện pháp thích hợp.

Dữ liệu bị thiếu

=> Gắn cờ và loại bỏ hoặc thay thế dữ liệu bị thiếu.

Lỗi cấu trúc

=> Sửa lỗi đánh máy và các điểm không nhất quán khác, đồng thời khiến dữ liệu tuân theo mẫu hoặc quy ước chung.

### 3.2.3.2 Trình bày các bước làm sạch trong phạm vi dự án

- ✓ *Trình bày cụ thể các bước tác động cụ thể đến bộ dữ liệu dự án (từng cột, từng dòng).*
- ✓ *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

Làm sạch dữ liệu trong Staging database.

• **GIAODICH\_CLEAN\_1**

4.1 Làm sạch dữ liệu thủ công (có rất nhiều cách để thực hiện và đây chỉ là một trong số các cách đó):

1. Xóa các giao dịch có ngày không hợp lý.
2. Xóa các giao dịch có khối lượng không đúng.
3. Đánh số thứ tự dòng dữ liệu.
4. Insert dữ liệu vào table mới GIAODICH\_CLEAN\_1

(mục đích của việc tạo ra table mới là không ảnh hưởng tới dữ liệu ban đầu table GIAODICH). Nếu có “N” thao tác làm sạch khác thì chúng ta sẽ tạo ra nhiều table GIAODICH\_CLEAN\_N

- Cách 1: Tạo table GIAODICH\_CLEAN\_1 trước (cùng cấu trúc với table GIAODICH).

## Dùng công thức:

```
select top 0 * [top 0 giúp tạo bảng mới với cấu trúc cột giống hệt mà không copy dữ liệu.]  
into staging.dbo.giaodich_clean_1  
from staging.dbo.giaodich;
```

kết quả:

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the 'Staging' database is selected. In the center pane, a query window displays the following T-SQL code:

```
1  SELECT TOP (1000) [MaCK]  
2      ,[NGAYGIAODICH]  
3      ,[GIAMOCUA]  
4      ,[GIACAONHAT]  
5      ,[GIAATHAPNHAT]  
6      ,[GIADONGCUA]  
7      ,[KHOILUONGGIAODICH]  
8      ,[TENNHHOMNGAH]  
9      ,[MAHHOMNGAH]  
10     ,[THONHTINCONGTY]  
11     ,[SAN]  
12     ,[CONGTY]  
13     ,[TENSAN_VIET]  
14     ,[TENSAN_ANH]  
15     ,[BIENDODAODONG]  
16     ,[ID]  
17  FROM [Staging].[dbo].[GIAODICH_CLEAN_1]
```

The results pane shows the output of the query, which is empty (0 rows). A status bar at the bottom indicates "Query executed successfully." and "0 rows".

- Sau đó dùng lệnh INSERT INTO ..... SELECT .....
- Thêm dữ liệu từ bảng gốc giao dịch -> vào bảng mới clean 1
- Điều kiện: loại bỏ các giao dịch có ngày & khối lượng không đúng

SQLQuery7.sql - DZIMEOMEStaging (DZIMEOME\vyngu (74)) - Microsoft SQL Server Management Studio

```

1: INSERT INTO STAGING.DBO.GIAODICH_CLEAN_1 ([ID]
2: , [NGAYGIAODICH]
3: , [GIAODICH_NHAT]
4: , [GIAODICH_HAT]
5: , [GIAODICH_NHAT]
6: , [GIAODICH_NHAT]
7: , [GIAODICH_NHAT]
8: , [KHOLIULONGGIAODICH]
9: , [THONGTINCONGTY]
10: , [TENHOMNGBANH]
11: , [MANHMONBANH]
12: , [PHONHTHOCONGTY]
13: , [SAN])
14: , [CONGTY]
15: , [TENHOMNGBANH]
16: , [MANHMONBANH]
17: , [PHONHTHOCONGTY]
18: , [SAN])
19: , [NGAYGIAODICH]
20: , [GIAODICH_NHAT]
21: , [GIAODICH_HAT]
22: , [GIAODICH_NHAT]
23: , [GIAODICH_NHAT]
24: , [GIAODICH_NHAT]
25: , [KHOLIULONGGIAODICH]
26: , [THONGTINCONGTY]
27: , [TENHOMNGBANH]
28: , [MANHMONBANH]
29: , [PHONHTHOCONGTY]
30: , [SAN])
31: , [CONGTY]
32: , [TENHOMNGBANH]
33: , [MANHMONBANH]
34: , [PHONHTHOCONGTY]
35: FROM STAGING.DBO.GIAODICH
WHERE KHOILOUNGKHOILOUNG > @ AND ISDATE(NGAYGIAODICH) > @;

```

80 %

Messages

(10952 rows affected)

Completion time: 2023-07-21T20:00:10.2244851+07:00

80 %

Query executed successfully.

LN 19 Col 14 Ch 14 INS

DZIMEOME (16.0 RTM) | DZIMEOME\vyngu (74) Staging | 00:00:00 | 0 rows

Ready ENG 803 CH 21/07/2025

SQLQuery13.sql - DZIMEOMEStaging (DZIMEOME\vyngu (77)) - Microsoft SQL Server Management Studio

```

1: SELECT *
2: , [ID]
3: , [NGAYGIAODICH]
4: , [GIAODICH_NHAT]
5: , [GIAODICH_HAT]
6: , [GIAODICH_NHAT]
7: , [GIAODICH_NHAT]
8: , [GIAODICH_NHAT]
9: , [KHOLIULONGGIAODICH]
10: , [THONGTINCONGTY]
11: , [TENHOMNGBANH]
12: , [MANHMONBANH]
13: , [PHONHTHOCONGTY]
14: , [SAN])
15: , [CONGTY]
16: , [TENHOMNGBANH]
17: , [MANHMONBANH]
18: , [PHONHTHOCONGTY]
19: FROM [Staging].[dbo].[GIAODICH_CLEAN_1]

```

80 %

Messages

ID	MÃ-SỐ	KHOILOUNG	GIAODICH_NHAT	GIAODICH_HAT	GIAODICH_NHAT	GIAODICH_NHAT	KHOLIULONGGIAODICH	TENHOMNGBANH	MANHMONBANH	PHONHTHOCONGTY	SAN	CONGTY	
1	A32	20230530	30	30.1	30	30	100				ĐỀ MÃY MẶC	DETMAXMAC	http://cafeif.vn/ccc/42/cong-ty-co-phap-nhua-v...
2	A32	20230606	30	30	30	30	200				ĐỀ MÃY MẶC	DETMAXMAC	http://cafeif.vn/ccc/42/cong-ty-co-phap-nhua-v...
3	A32	20230609	30	30	30	30	400				ĐỀ MÃY MẶC	DETMAXMAC	http://cafeif.vn/ccc/42/cong-ty-co-phap-nhua-v...
4	A32	20230607	30	30.1	30	30	1157				ĐỀ MÃY MẶC	DETMAXMAC	http://cafeif.vn/ccc/42/cong-ty-co-phap-nhua-v...
5	A32	20230612	31	31	31	31	102				ĐỀ MÃY MẶC	DETMAXMAC	http://cafeif.vn/ccc/42/cong-ty-co-phap-nhua-v...
6	A32	20230605	30	30	29.5	29.5	400				ĐỀ MÃY MẶC	DETMAXMAC	http://cafeif.vn/ccc/42/cong-ty-co-phap-nhua-v...
7	A32	20230606	11.7	11.8	11.45	11.45	7264800	Cao su	CAOSU				
8	A32	20230609	11.5	11.6	11.15	11.45	5323100	Cao su	CAOSU				
9	A32	20230612	11.45	11.6	11.35	11.55	5029600	Cao su	CAOSU				
10	A32	20230605	11.45	11.6	11.35	11.55	6292700	Cao su	CAOSU				
11	A32	20230530	11.1	11.3	10.95	11	5553000	Cao su	CAOSU				
12	A32	20230605	11.85	11.9	11.45	11.45	4830100	Cao su	CAOSU				
13	A32	20230607	11.45	11.7	11.35	11.55	6293188	Cao su	CAOSU				
14	A32	20230602	11.45	11.6	11.35	11.55	6292700	Cao su	CAOSU				
15	A32	20230531	11.05	11.55	10.9	11.4	10211000	Cao su	CAOSU				

80 %

Messages

(10952 rows affected)

Completion time: 2023-07-21T20:00:10.2244851+07:00

80 %

Query executed successfully.

LN 1 Col 8 Ch 8 INS

DZIMEOME (16.0 RTM) | DZIMEOME\vyngu (77) Staging | 00:00:00 | 10.952 rows

Ready ENG 808 CH 21/07/2025

⇒ Kết quả: bảng Giaodich\_clean\_1 có tổng 10.952 dòng

CÁCH 2: Dùng lệnh SELECT .... INTO .... FROM .... WHERE như hình bên dưới. mục đích:

- Xóa các giao dịch có ngày không hợp lý.
- Xóa các giao dịch có khối lượng không đúng.

### 3. Đánh số thứ tự dòng dữ liệu.

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery1.sql - DZIMEOMEOM.SAMPLE (DZIMEOMEOM\vyngu (111)) - Microsoft SQL Server Management Studio". The left pane displays the "Object Explorer" tree, which includes nodes for the database "DZIMEOMEOM", the schema "dbo", and various system objects like "Tables", "Views", and "Procedures". A red box highlights the "Tables" node under "dbo". The main pane shows a query window with the following T-SQL code:

```
1 --SELECT ROW_NUMBER() OVER(ORDER BY Mack ASC) AS ID, *
2 INTO [Staging].[dbo].[GIAODICH_CLEAN_1]
3 FROM [Staging].[dbo].[GIAODICH]
4 WHERE [KHOIQUONGGIAODICH] > 0 AND ISDATE([NGAYGIAODICH]) > 0;
```

The status bar at the bottom indicates "106 %", "Query executed successfully.", and the date "18/07/2023".

### **GIAODICH CLEAN 2**

4.2 Làm sạch dữ liệu thủ công (tiếp theo) (có rất nhiều cách để thực hiện và dưới đây chỉ là một trong số các cách đó):

1. Loại bỏ các dòng dữ liệu bị trùng (duplicate data)
  2. Duplicate data (trong ngữ cảnh bộ dataset này) khi giá trị ở các Cột trên Dòng giống nhau là ([MaCK],[NGAYGIAODICH],[GIAMOCUA],[GIACAONHAT],[GIATHAPNHAT],[GIADONGCUA],[KHOILUONGGIAODICH])
  - 3.Các dòng trùng sẽ bị gán giá trị > 1 trên cột DuplicateRow
  - 4.Insert các dòng không trùng vào table GIAODICH CLEAN 2

```

SELECT * INTO [Staging].[dbo].[GIAODICH_CLEAN_2]
FROM (
    SELECT *
    , DuplicateRow = ROW_NUMBER() OVER(
        PARTITION BY
            [MaCK]
            ,[NGAYGIAODICH]
            ,[GIAMOCUA]
            ,[GIACONNHAT]
            ,[GIATHAPNHAT]
            ,[GIADONGCUA]
            ,[KHOILUONGGIAODICH]
        ORDER BY (SELECT NULL)
    )
    FROM [Staging].[dbo].[GIAODICH_CLEAN_1]
) TEMP
WHERE TEMP.DuplicateRow = 1;

```

```

SELECT *
FROM [Staging].[dbo].[GIAODICH_CLEAN_1]

```

Results (10952 rows affected)

MACK	NGAYGIAODICH	GIAMOCUA	GIACONNHAT	GIATHAPNHAT	GIADONGCUA	KHOILUONGGIAODICH	TEN
A32	20230505	30	30	29.5	400	400	DH
AAA	20230612	11.45	11.6	11.35	11.55	5029600	Cao
AAM	20230601	10.65	10.9	10.25	10.85	7400	Thứ
AAT	20230607	6.38	6.5	6.3	6.39	1061170	Bán
AAV	20230605	5.6	5.8	5.3	5.4	866213	Bán
ABB	20230606	8.8	8.9	8.5	8.7	5019925	Ngã
ABC	20230530	18.3	18.4	18.2	18.3	28900	Bùi
ABI	20230607	35.1	36.6	35	35.5	78550	Bảo
ABS	20230609	6.9	7.06	6.74	7.06	2472900	Chủ
ABT	20230602	31.4	31.4	31.4	31.4	200	Thờ
ACC	20230602	21.8	21.85	21.7	21.8	9941200	Ngã
ACZ	20230609	13.75	13.8	13.6	13.7	64400	Xây
ACE	20230531	35	35	35	35	100	Xây
ACL	20230605	13	13.3	12.65	12.7	180200	Thờ
ACD	20230531	9	9	8.6	8.6	200	Xây

Kết quả: Từ bảng [GIAODICH\_CLEAN\_1] (10952 dòng) -> lọc không trùng thêm vào bảng mới [GIAODICH\_CLEAN\_2] (còn 10915 dòng)

```

SELECT * FROM [Staging].[dbo].[GIAODICH_CLEAN_2]
ORDER BY ID;

```

Results (10915 rows)

ID	MACK	NGAYGIAODICH	GIAMOCUA	GIACONNHAT	GIATHAPNHAT	GIADONGCUA	KHOILUONGGIAODICH	TENNHHOMNGANH	MANHOMNGANH	THONGTINCONGTY	SAN	CONC
1	A32	20230530	30.1	30.1	30	30	1000	Dết may mè	DET MAY MAC	http://cafevn/upcom/A32-cong-ty-co-phan-32.chn	Upcom	Công
2	A32	20230605	30	30	29.5	400	Dết may mè	DET MAY MAC	http://cafevn/upcom/A32-cong-ty-co-phan-32.chn	Upcom	Công	
3	A32	20230612	31	31	31	31	102	Dết may mè	DET MAY MAC	http://cafevn/upcom/A32-cong-ty-co-phan-32.chn	Upcom	Công
4	A32	20230607	30	30.1	30	30	1157	Dết may mè	DET MAY MAC	http://cafevn/upcom/A32-cong-ty-co-phan-32.chn	Upcom	Công
5	A32	20230605	30	30	30	30	200	Dết may mè	DET MAY MAC	http://cafevn/upcom/A32-cong-ty-co-phan-32.chn	Upcom	Công
6	A32	20230609	30	30	30	30	400	Dết may mè	DET MAY MAC	http://cafevn/upcom/A32-cong-ty-co-phan-32.chn	Upcom	Công
7	AAA	20230608	11.7	11.8	11.45	11.45	7264800	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công
8	AAA	20230603	11.1	11.3	11.25	11	5523000	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công
9	AAA	20230605	11.05	11.9	11.45	11.45	4830100	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công
10	AAA	20230607	11.45	11.7	11.35	11.55	6263168	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công
11	AAA	20230609	11.5	11.6	11.15	11.45	5323100	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công
12	AAA	20230605	11.45	11.6	11.35	11.55	6262700	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công
13	AAA	20230602	11.45	11.6	11.35	11.55	5029600	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công
14	AAA	20230612	11.45	11.6	11.35	11.55	10211000	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công
15	AAA	20230531	11.05	11.55	10.9	11.4	10211000	Cao su	CAOSU	http://cafevn/hose/AAA-cong-ty-co-phan-hose-V...	HOSE	Công

### 3.3 CHUYỂN ĐỔI DỮ LIỆU

- ✓ Các trường hợp và kỹ thuật chuyển đổi dữ liệu.

### **3.3.1 CÁC TRƯỜNG HỢP CẦN CHUYỂN ĐỔI**

✓ *Nêu các trường hợp cần chuyển đổi dữ*

#### **❑ Analytics (Phân tích)**

- Phân tích để hỗ trợ các quyết định bắt đầu với các chỉ số.
- Đôi khi, các chỉ số có thể được tính toán từ một nguồn duy nhất và chỉ cần một lượng biến đổi nhỏ.
- Những lần khác, cách duy nhất để tính toán số liệu là kết hợp dữ liệu từ nhiều nguồn và sau đó tổng hợp lại.

#### **❑ Machine learning (Máy học)**

- Máy học là nhận dạng mẫu tự động.
- Các ứng dụng kinh doanh của máy học bao gồm dự báo doanh thu và lợi nhuận, mô hình dự đoán để hỗ trợ các quyết định chính, hệ thống giới thiệu sản phẩm cho khách hàng và tất cả các loại tự động hóa quy trình kinh doanh.

#### **❑ Regulatory compliance (Tuân thủ quy định)**

- Lưu trữ không cần thiết thông tin nhận dạng cá nhân (PII) khiến dữ liệu dễ bị lỗi.
- Vi phạm dữ liệu làm tổn hại đến quyền riêng tư của dữ liệu và tạo ra các vấn đề nghiêm trọng cho cả bạn và khách hàng nếu như dữ liệu bị hack.
- Không lưu trữ các thông tin nhạy cảm (sensitive data) như thông tin thẻ tín dụng, thanh toán ...

### **3.3.2 CÁC KỸ THUẬT CHUYỂN ĐỔI**

✓ *Trình bày các kỹ thuật chuyển đổi*  
✓ *Kỹ thuật nào được chọn áp dụng cho dự án.*

Dữ liệu thô hiếm khi được cấu trúc hoặc định dạng theo cách phục vụ các trường hợp sử dụng được liệt kê ở trên.

❖ Sau đây là một số các phép biến đổi dữ liệu phổ biến.

- **Sửa đổi (Revising):** là một hình thức sửa đổi dữ liệu bằng cách giảm mô hình dữ liệu về dạng “bình thường” mà không có dư thừa hoặc một-nhiều giá trị

trong một cột. Chuẩn hóa làm giảm nhu cầu lưu trữ và làm cho mô hình dữ liệu ngắn gọn hơn và dễ đọc hơn đối với các nhà phân tích.

- **Làm sạch (Data cleansing):** dữ liệu chuyển đổi các giá trị dữ liệu để tương thích với định dạng.
- **Sửa đổi/chuyển đổi (format revision/conversion):** định dạng thay thế các ký tự không tương thích, chuyển đổi đơn vị, chuyển đổi định dạng ngày tháng và thay đổi kiểu dữ liệu.
- **Tái cấu trúc khóa (key restructuring):** tạo ra các số nhận dạng chung ngoài các giá trị có ý nghĩa tích hợp, vì vậy chúng có thể được sử dụng như các khóa cố định, duy nhất trên các bảng.
- **Loc dữ liệu trùng lặp (Deduplication):** có nghĩa là xác định và loại bỏ các bản ghi trùng lặp.
- **Xác thực dữ liệu (Data validation):** đánh giá tính hợp lệ của một bản ghi bằng tính đầy đủ của dữ liệu, thường bằng cách loại trừ các bản ghi không đầy đủ.
- **Việc loại bỏ các cột không sử dụng và lặp lại (removing unused and repeated columns):** cho phép bạn chọn các trường bạn muốn sử dụng làm tính năng, tức là các biến đầu vào cho mô hình dự đoán. Nó cũng có thể cải thiện hiệu suất và tính dễ đọc tổng thể của một mô hình.
- **Tính toán chéo (derivation):** bao gồm các phép tính cột chéo đơn giản.
- **Tóm tắt (summarization):** bao gồm việc sử dụng các hàm tổng hợp để tạo ra các giá trị tóm tắt.
- **Xoay vòng (pivoting):** biến các giá trị hàng thành cột và ngược lại.
- **Sắp xếp và lập chỉ mục (sorting, ordering and indexing):** tổ chức các bản ghi theo một số thứ tự để cải thiện hiệu suất tìm kiếm.
- **Chia tỷ lệ và chuẩn hóa (scaling, standardization and normalization):** đặt các con số trên một thang đo nhất quán, chẳng hạn như các phân số của độ lệch chuẩn trong chuẩn hóa điểm Z. Điều này cho phép các con số khác nhau được so sánh với nhau.
- **Vector (vectorization):** hóa chuyển đổi dữ liệu không phải số thành mảng số. Có rất nhiều ứng dụng học máy của những chuyển đổi này, chẳng hạn như để xử lý ngôn ngữ tự nhiên (NLP) và nhận dạng hình ảnh.
- **Tách biệt (splitting):** việc phân tách bao gồm việc phân chia các giá trị thành các phần cấu thành của chúng. Các giá trị dữ liệu thường được kết hợp trong cùng một trường vì tính riêng trong thu thập dữ liệu, nhưng có thể cần được tách riêng để thực hiện phân tích chi tiết hơn.

- **Lọc loại trừ (filtering):** dữ liệu trên cơ sở các giá trị hàng hoặc cột nhất định.
- Kết nối (joining): là hành động liên kết dữ liệu giữa các bảng.
- **Hợp nhất (Merging)** còn được gọi là thêm hoặc kết hợp, kết hợp các bản ghi từ nhiều bảng, bằng cách kết hợp hai bảng sử dụng một cột chung.

### 3.3.3 TRÌNH BÀY CÁC PHÉP CHUYỂN ĐỔI TRONG DỰ ÁN

- ✓ *Trình bày cụ thể các bước tác động cụ thể đến bộ dữ liệu dự án (từng cột, từng dòng).*
- ✓ *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

#### ❑ Chi tiết các phép chuyển đổi

##### ❖ Giá trần (GIATRAN):

- Là mức giá cao nhất trong phiên giao dịch mà nhà đầu tư có thể đặt lệnh mua hoặc bán.
- Và ngay cả khi bạn muốn mua với giá cao hơn giá trần, lệnh sẽ không được thực hiện.

##### ❖ Giá sàn (GIASAN):

- Ngược với giá trần, đây là mức giá thấp nhất trong phiên giao dịch mà nhà đầu tư có thể mua, bán chứng khoán.
- Và bạn sẽ không thể đặt lệnh với giá thấp hơn giá sàn.

##### ❖ Tỷ lệ tăng giảm trong ngày (TILETRONGNGAY):

- Tính tỉ lệ tăng hoặc giảm của giá trong ngày so với giá mở cửa.

#### ❑ Thực hiện các phép chuyển đổi

##### ❖ Định dạng lại cột [NGAYGIAODICH].

❖ Thêm Giá trần = Giá tham chiếu (mở cửa) x (100% + Biên độ dao động).

❖ Thêm Giá sàn = Giá tham chiếu (mở cửa) x (100% – Biên độ dao động).

##### ❖ Tỷ lệ tăng giảm trong ngày:

$$TILETRONGNGAY = ([GIADONGCUA] - [GIAMOCUA]) * 100 / [GIAMOCUA].$$

❖ Insert tất cả các dòng vào table GIAODICH\_FINAL.

SQLQuery16.sql - DZIMEOMEOSTaging (DZIMEOMEO\vyngu (77)) - Microsoft SQL Server Management Studio

```

File Edit View Query Project Tools Window Help
File Connect Object Explorer Solution Explorer Properties
File Staging Execute New Query Task Copy Database
File Staging
1 use Staging
2 go
3
4 SELECT
5   [ID],
6   [MaCK],
7   ,FORMAT(CONVERT(datetime, [NGAYGIAODICH]), 'd', 'us') AS [NGAYGIAODICH]
8   ,FORMAT([GIAMOCUA])
9   ,FORMAT([GIACAOHNAT])
10  ,FORMAT([GIATHAPNHAT])
11  ,FORMAT([GIADONGCUA])
12  ,FORMAT([KHOILUONGGIAODICH])
13  ,FORMAT([GIATRAN])=(1-[BIENDODAODONG]), 'F', 'en-us') AS GIATRAN
14  ,FORMAT([GIAMOCUA])*(1-[BIENDODAODONG]), 'F', 'en-us') AS GIASAN
15  ,FORMAT(([GIAMOCUA]-[GIADONGCUA])*100/[GIAMOCUA], 'F', 'en-us') AS TILETRONGNGAY
16  ,FORMAT([TENNHOMNGANH])
17  ,FORMAT([MANHOMNGANH])
18  ,FORMAT([THONGTINCONGTY])
19  ,FORMAT([SAN])
20  ,FORMAT([CONGTY])
21  ,FORMAT([TENSAN_VIET])
22  ,FORMAT([TENSAN_ANH])
23  ,FORMAT([BIENDODAODONG])
24  ,FORMAT([DuplicateRow])
25  INTO Staging.dbo.GIAODICH_FINAL
26  FROM Staging.dbo.GIAODICH_CLEAN_2;

```

117 % Messages (10915 rows affected)

117 % Query executed successfully.

DZIMEOMEOSTaging (DZIMEOMEO\vyngu (77)) Staging 00:00:00 0 rows

Ready 8:56 CH 21/07/2025

SQLQuery17.sql - DZIMEOMEOSTaging (DZIMEOMEO\vyngu (79)) - Microsoft SQL Server Management Studio

```

File Edit View Query Project Tools Window Help
File Connect Object Explorer Solution Explorer Properties
File Staging Execute New Query Task Copy Database
File Staging
1 SELECT [ID]
2   , [MaCK]
3   , [NGAYGIAODICH]
4   , [GIAMOCUA]
5   , [GIACAOHNAT]
6   , [GIATHAPNHAT]
7   , [GIADONGCUA]
8   , [KHOILUONGGIAODICH]
9   , [GIATRAN]
10  , [GIASAN]
11  , [TILETRONGNGAY]
12  , [TENNHOMNGANH]
13  , [MANHOMNGANH]
14  , [THONGTINCONGTY]
15  , [SAN]
16  , [CONGTY]
17  , [Result] , [Messages]
18
19 ID MaCK NGAYGIAODICH GIAMOCUA GIACAOHNAT GIATHAPNHAT GIADONGCUA KHOILUONGGIAODICH GIATRAN GIASAN TILETRONGNGAY TENNHOMNGANH MANHOMNGANH THONGTINCONGTY
2 5374 LCK 06/01/2023 13.5 13.6 13.3 13.5 11065700 14.45 12.56 0.00 Đầu tư công DAUTUCONG http://cafeit.vn/host/LCK-cong-ty-co-phan-lloogi...
3 5396 LCK 05/09/2023 2.8 2.8 2.6 2.8 797164 3.00 2.60 0.00 Thien Khong san KHAHKHONG http://cafeit.vn/host/LCK-cong-ty-co-phan-lloogi...
4 5378 LCK 06/01/2023 3.1 3.2 3.1 3.2 678791 3.25 2.88 -3.23 Thien Khong san KHAHKHONG http://cafeit.vn/host/LCK-cong-ty-co-phan-khai-t...
5 5403 LCK 06/01/2023 3.0 7 6.8 6.9 802525 756 6.24 0.00 VTE http://cafeit.vn/host/VTE-cong-ty-co-phan-lloogi...
6 5432 LCK 06/01/2023 5.17 5.2 4.9 5.07 886249 4.93 4.29 -6.07 Xay dung XAYDUNG http://cafeit.vn/host/LCK-cong-ty-co-phan-dau-t...
7 5442 LCK 06/08/2023 49.2 49.3 48.9 49.3 19400 54.12 44.28 -0.20 Xay dung XAYDUNG http://cafeit.vn/host/LCK-cong-ty-co-phan-dau-t...
8 5443 LCK 06/08/2023 49.3 49.3 48.9 48.9 7100 54.23 44.37 0.81 Xay dung XAYDUNG http://cafeit.vn/host/LCK-cong-ty-co-phan-dau-t...
9 5449 LCK 06/08/2023 29.55 30 29 29.1 280900 31.82 27.48 1.52 Xay dung XAYDUNG http://cafeit.vn/host/LCK-cong-ty-co-phan-long-i...
10 5465 LCK 05/01/2023 19.5 19.7 18.2 18.9 5200 22.43 16.50 3.08 Xay dung XAYDUNG https://cafeit.vn/boom/LCK-cong-ty-co-koop-o...
11 5460 LCK 06/08/2023 19.5 20.3 18.7 19.2 16000 22.43 16.50 1.54 Xay dung XAYDUNG https://cafeit.vn/boom/LCK-cong-ty-co-koop-o...
12 5471 LCK 06/05/2023 5.1 5.3 5 5.2 1659415 5.61 4.59 -1.96 Xay dung XAYDUNG http://cafeit.vn/host/LCK-cong-ty-co-phan-lloogi...
13 5472 LCK 06/05/2023 5.4 5.6 5.2 5.2 2785705 5.94 4.86 3.70 Xay dung XAYDUNG http://cafeit.vn/host/LCK-cong-ty-co-phan-lloogi...
14 5489 LLK 06/07/2023 7.9 8 7.9 7.9 241 9.09 6.72 0.00 Bán lẻ BANLE https://cafeit.vn/boom/LLK-long-cong-ty-lap-mi...
15 5494 LLK 06/09/2023 7.8 7.8 7.8 7.8 2500 8.97 6.63 0.00 Bán lẻ BANLE https://cafeit.vn/boom/LLK-long-cong-ty-lap-mi...

```

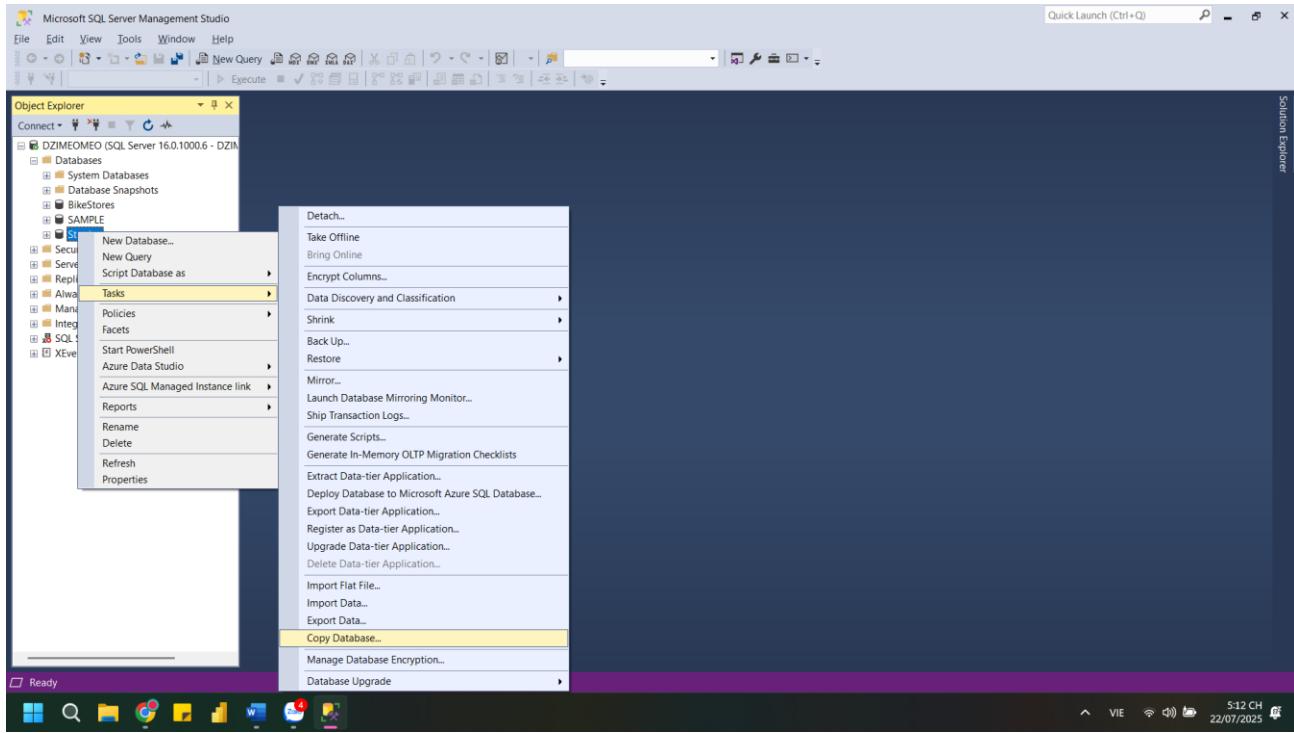
117 % Messages (10915 rows affected)

DZIMEOMEOSTaging (DZIMEOMEO\vyngu (79)) Staging 00:00:00 10,915 rows

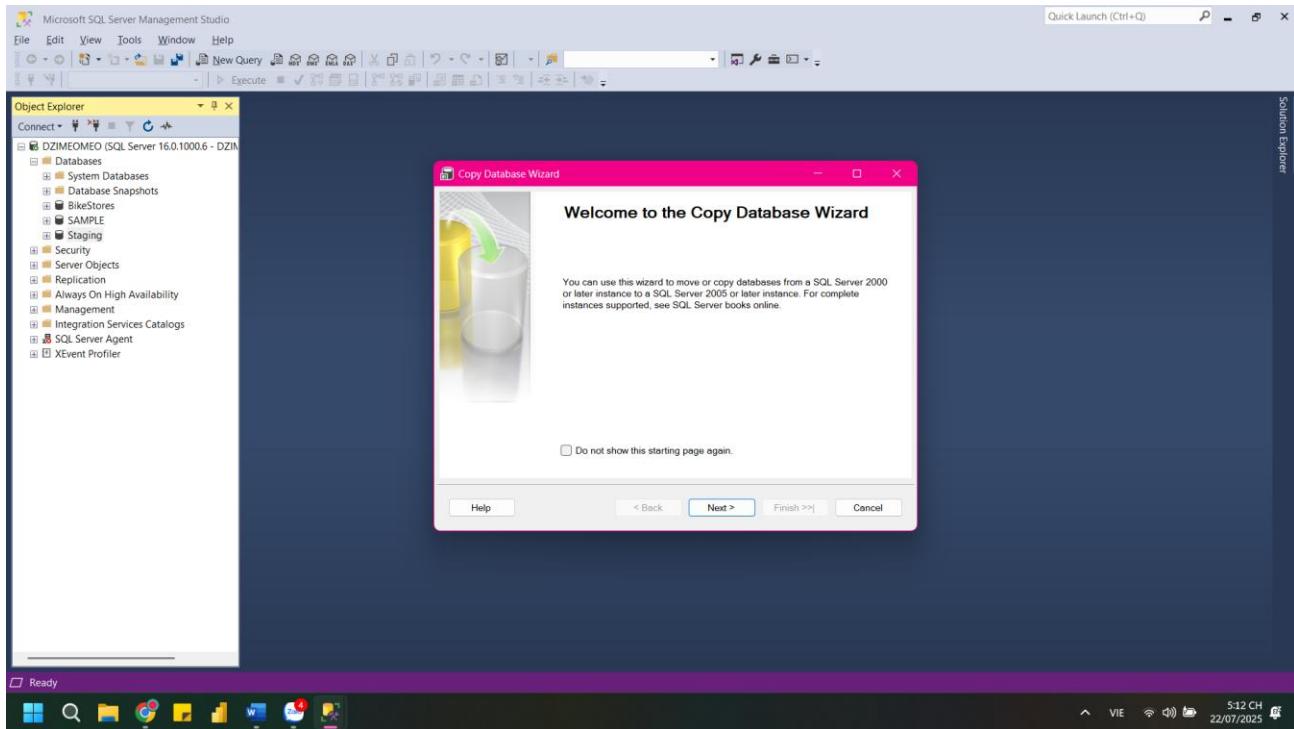
Ready 8:56 CH 21/07/2025

Tạo Spend database từ Staging database bằng trình wizard.

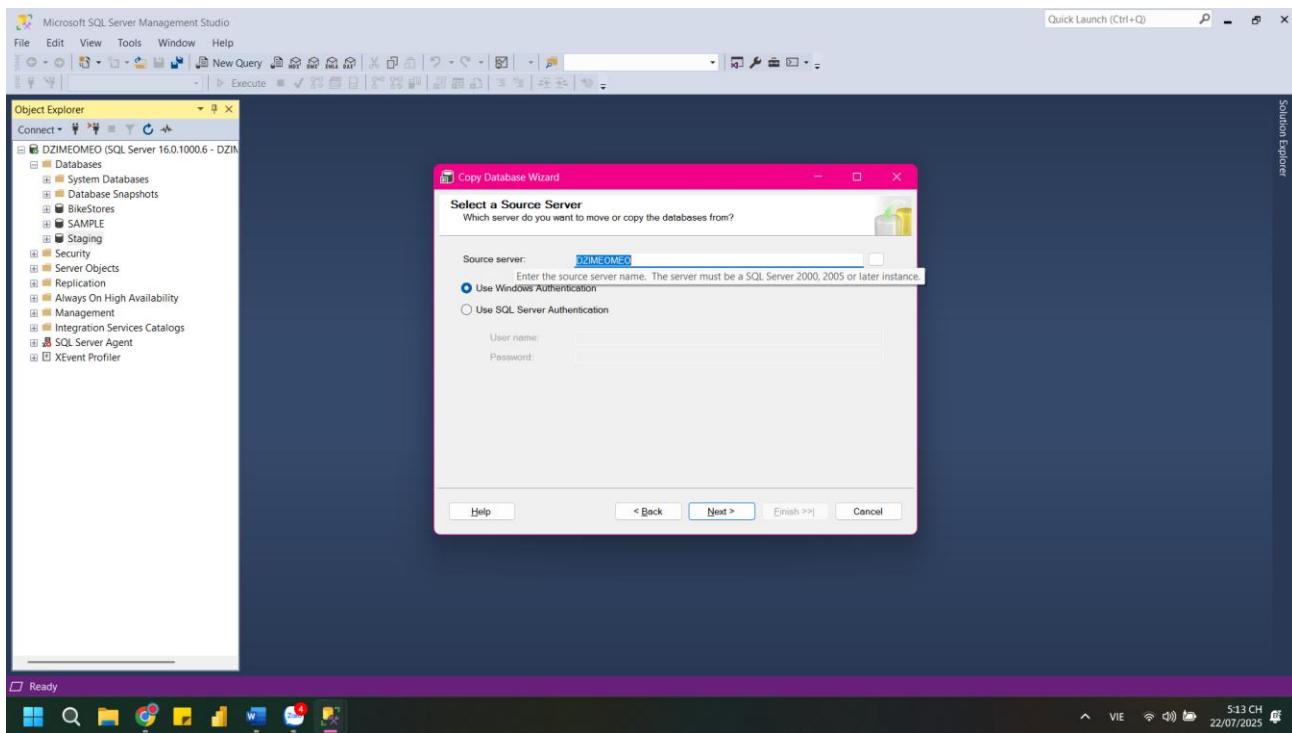
- Chọn database Staging cần nhân bản. Chọn Task >> Copy Database.



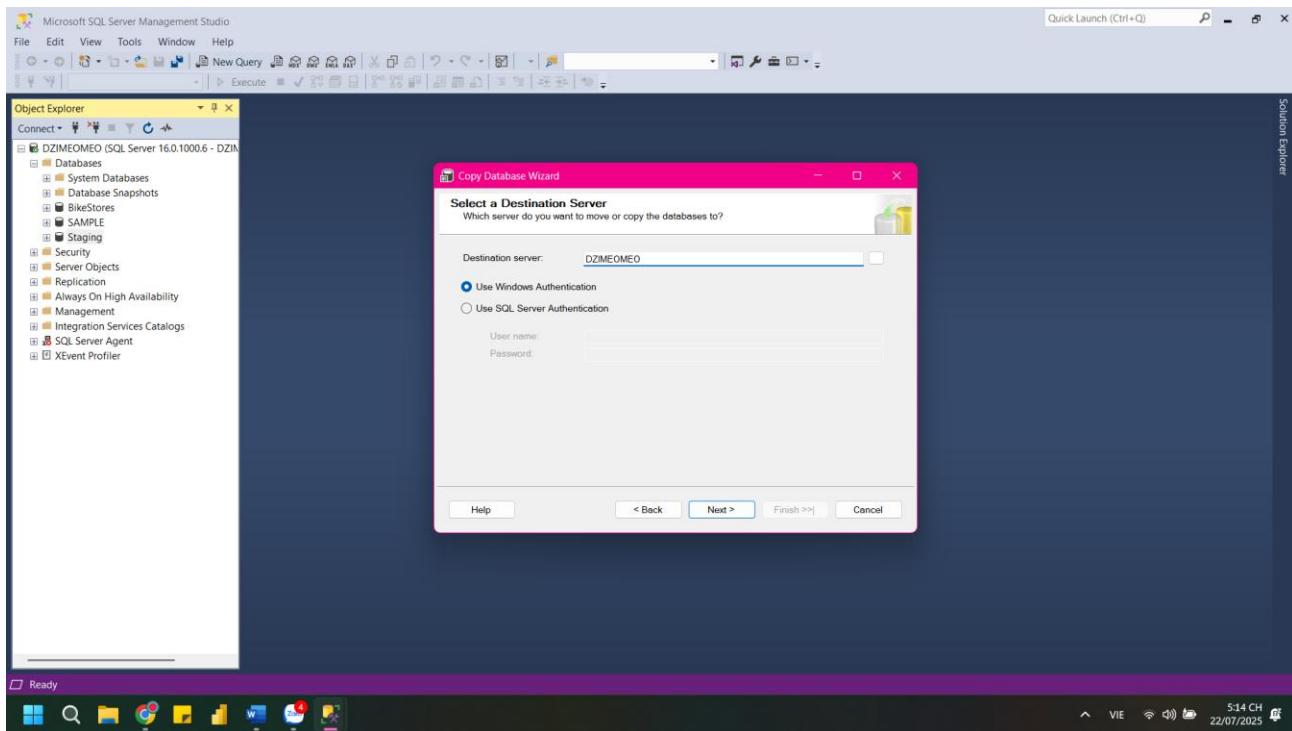
- Chọn Next



- Chọn server chứa database Staging.



- Chọn server chứa database Spend.

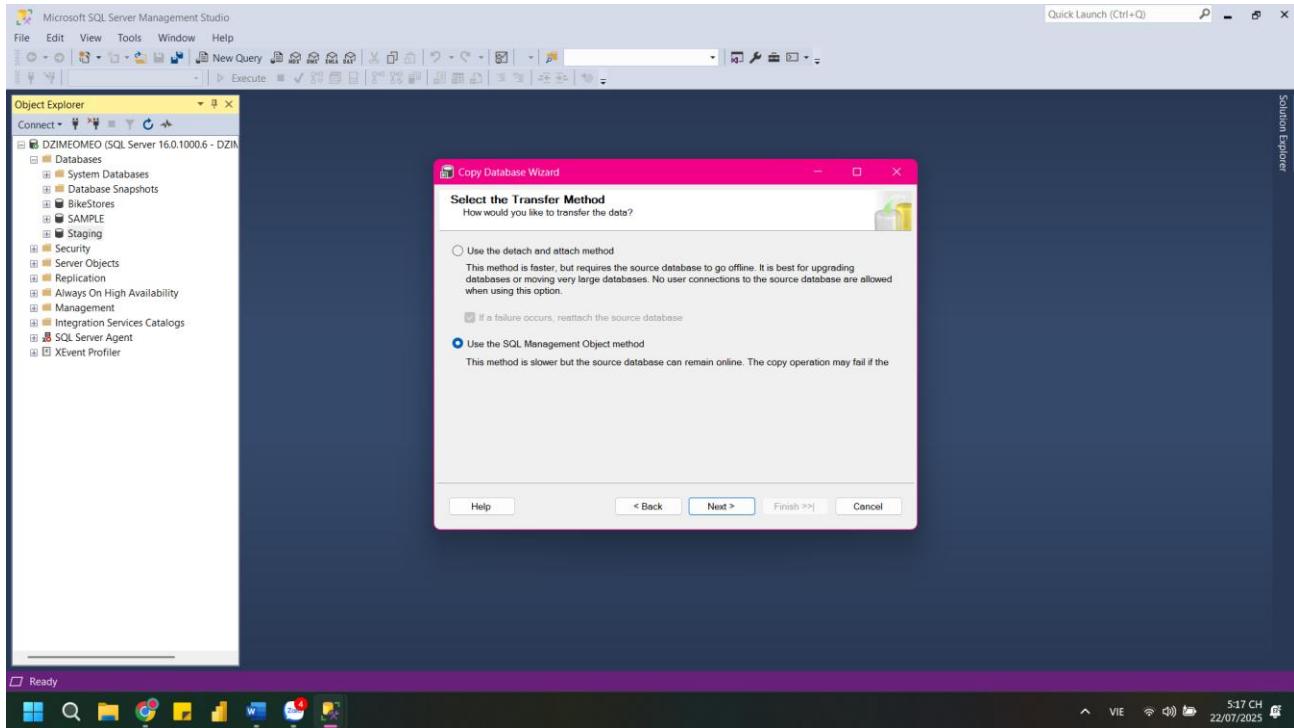


- Chọn Phương Thức chuyển dữ liệu.

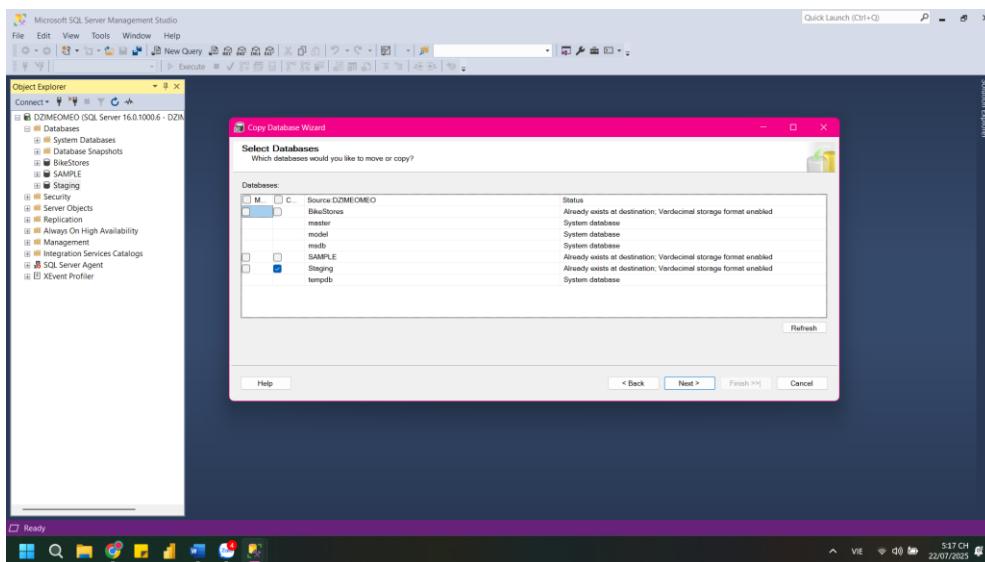
Chọn option 1: dành cho database có dữ liệu lớn, database sẽ offline trong lúc tiến trình thực thi.

Chọn option 2: tốc độ chậm hơn nhưng database vẫn online trong lúc tiến trình thực thi.

=> Trong phạm vi môn học, khuyến nghị cho option 2. Nếu chọn option 1 sẽ ảnh hưởng tới các connection đang dùng database này.



- Chọn database cần copy. Nhấn Next.

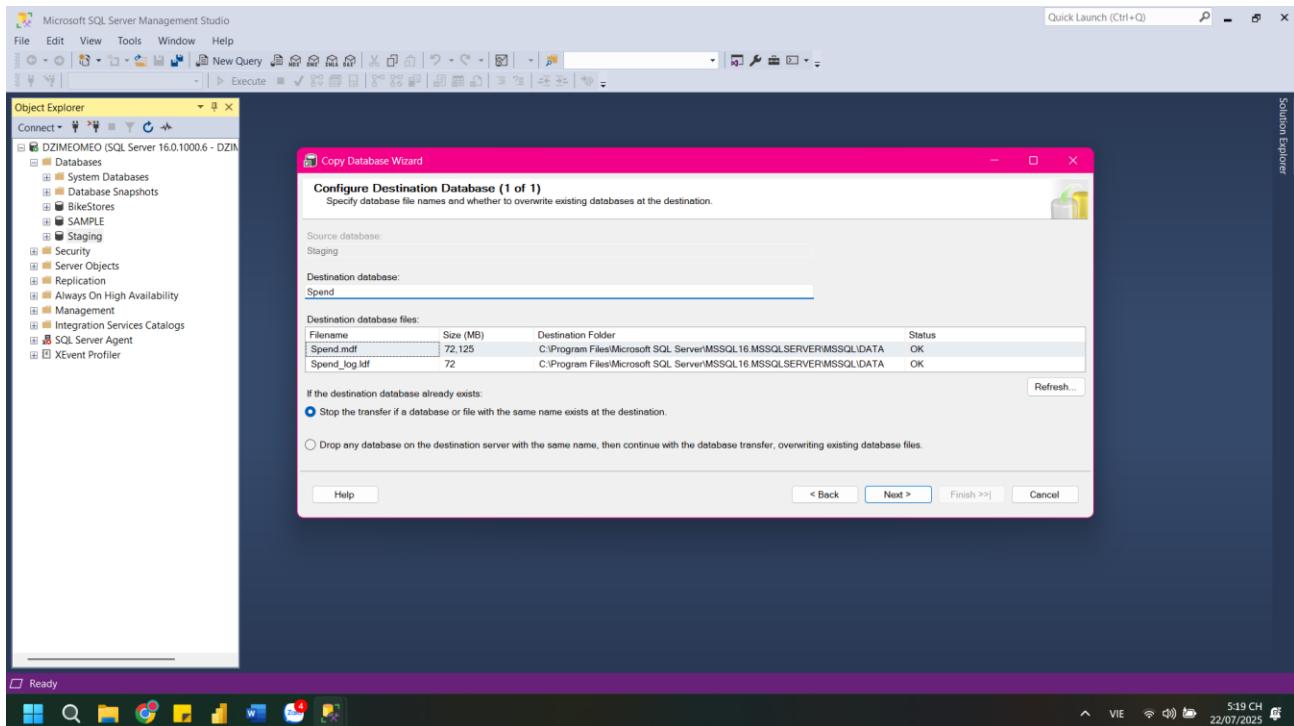


- Đặt tên cho database mới là **Spend**. Kiểm tra đường dẫn database Spend.

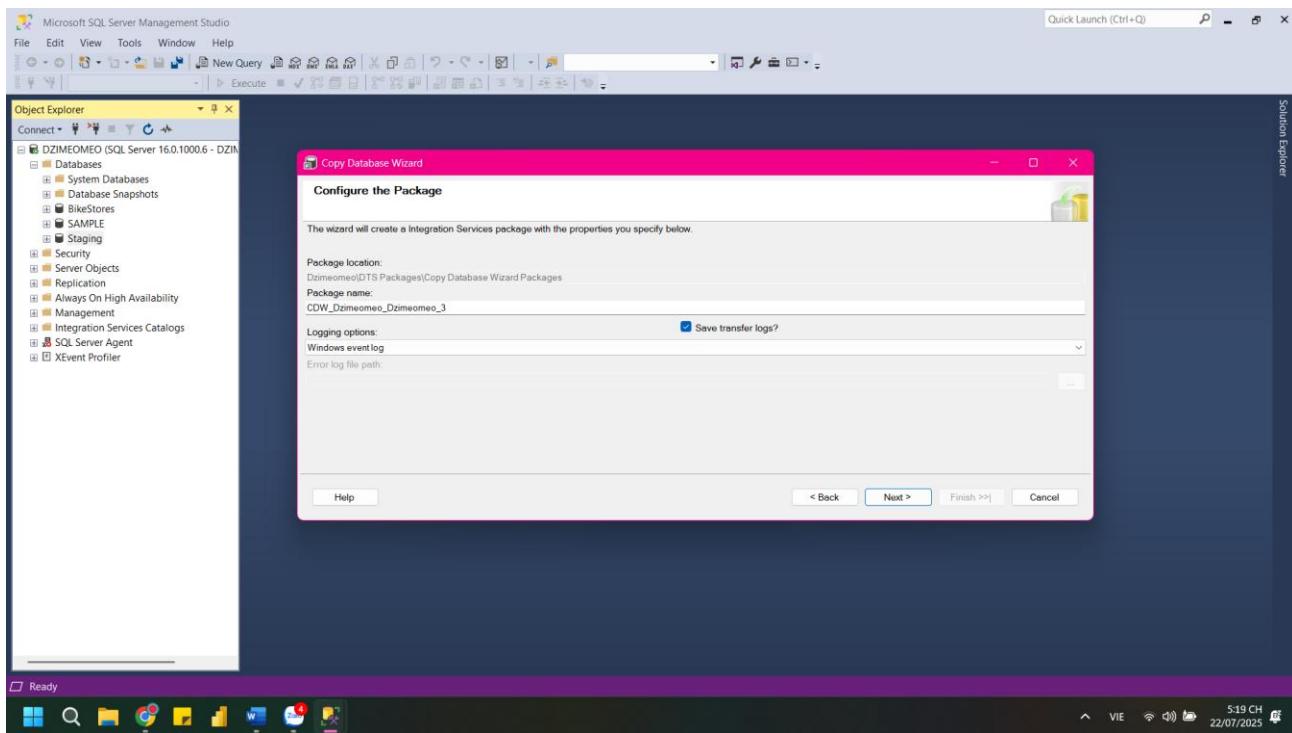
**Chọn option 1:** dừng ngay tiến trình nếu Staging đã tồn tại.

**Chọn option 2:** ghi đè dữ liệu nếu Staging đã tồn tại.

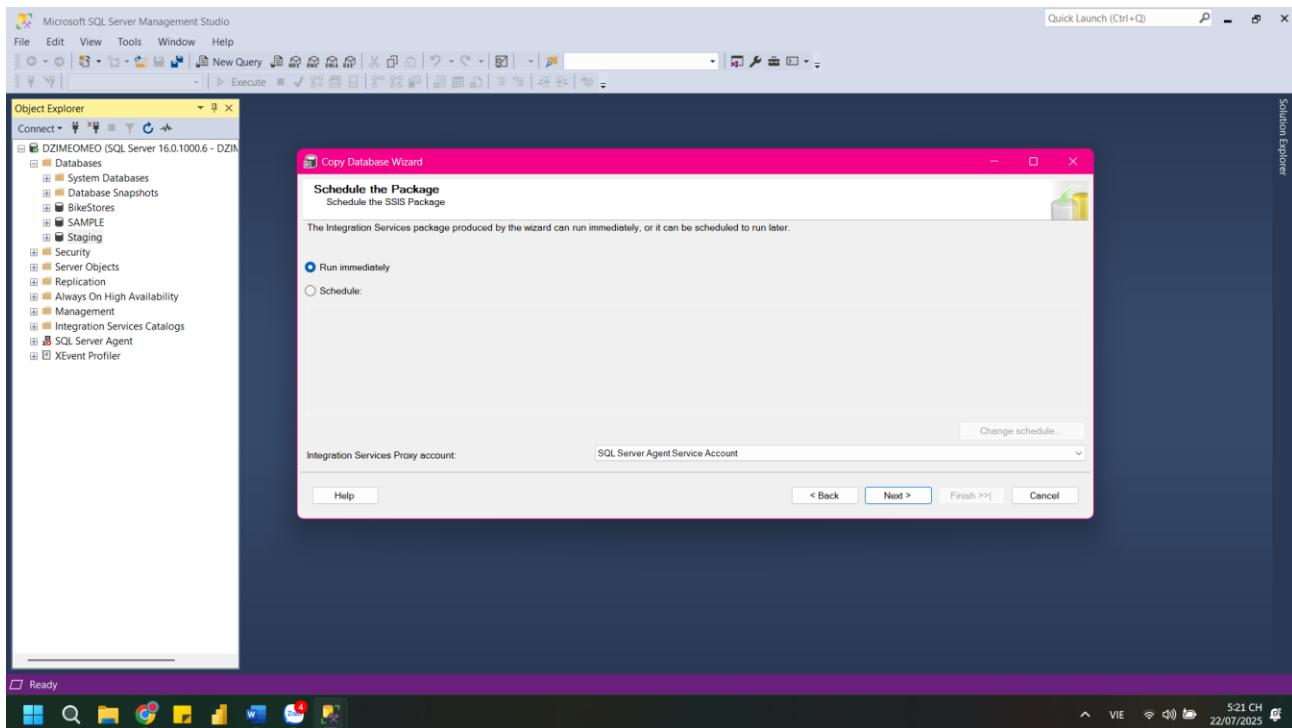
- ⇒ Trong phạm vi môn học, khuyến nghị cho option 1. Thực tế khi gặp bất cứ issues gì trong quá trình copy thì nên dừng lại để kiểm tra xem xét, tránh mất dữ liệu, ghi đè, ảnh hưởng tới công việc của nhóm khác.



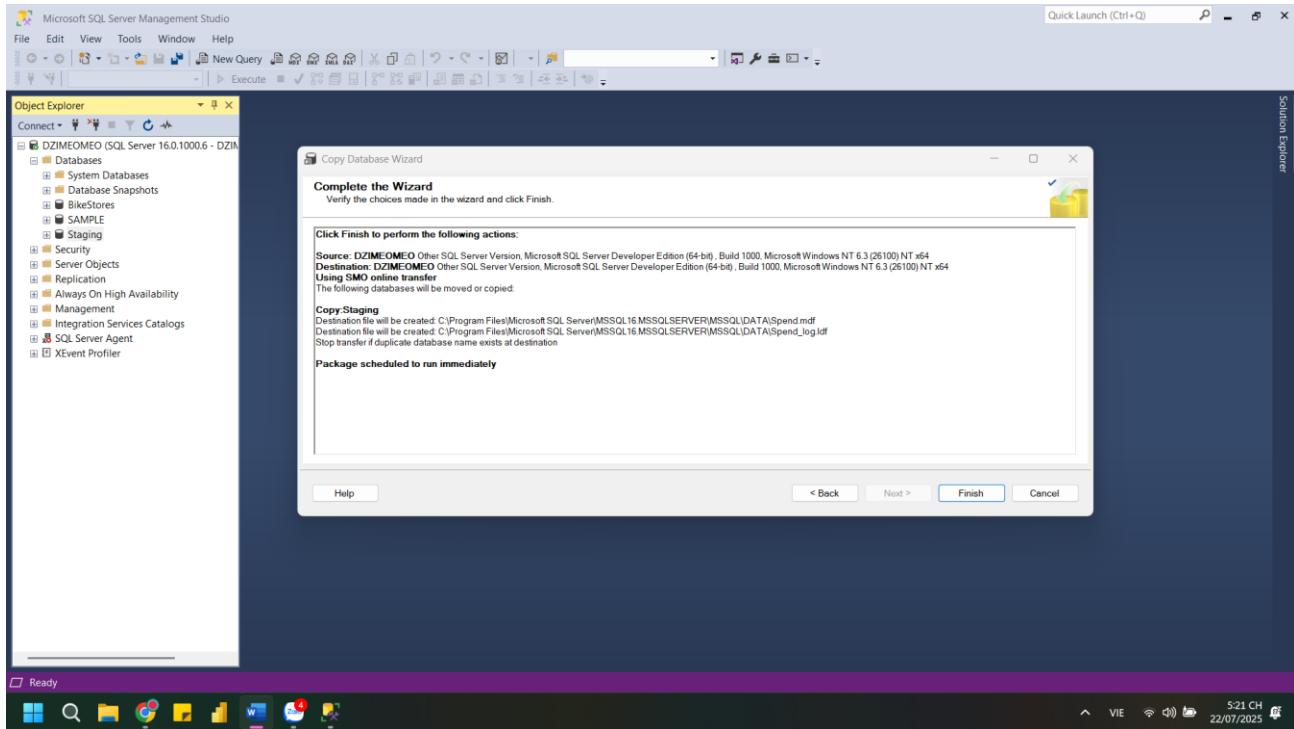
- Hệ thống sẽ tự động tạo ra 1 package để transfer dữ liệu (không cần sửa tên package mà hệ thống tự sinh ra)
- Chọn Save Transfer Log nhằm mục đích lưu lại log quá trình transfer, giúp dễ truy vết khi xảy ra sự cố.
- Nhấn Next.



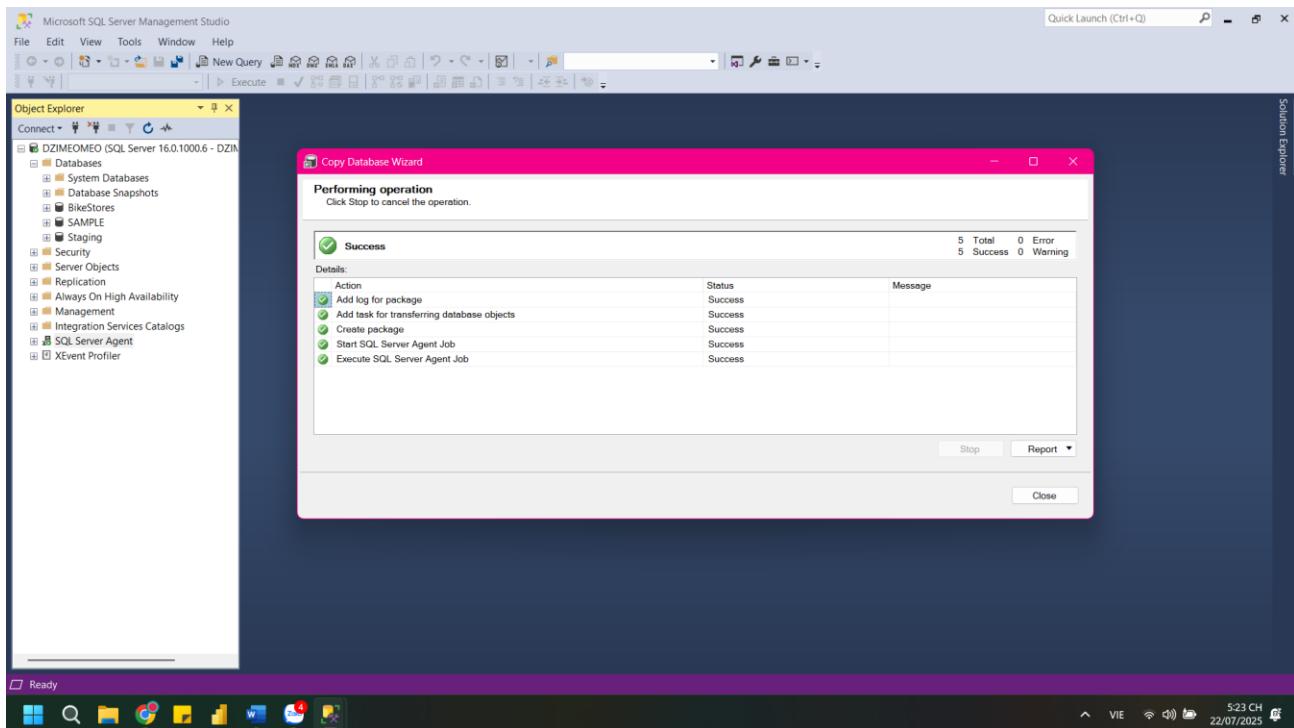
- Chọn option Run immediately. Nhấn Next.



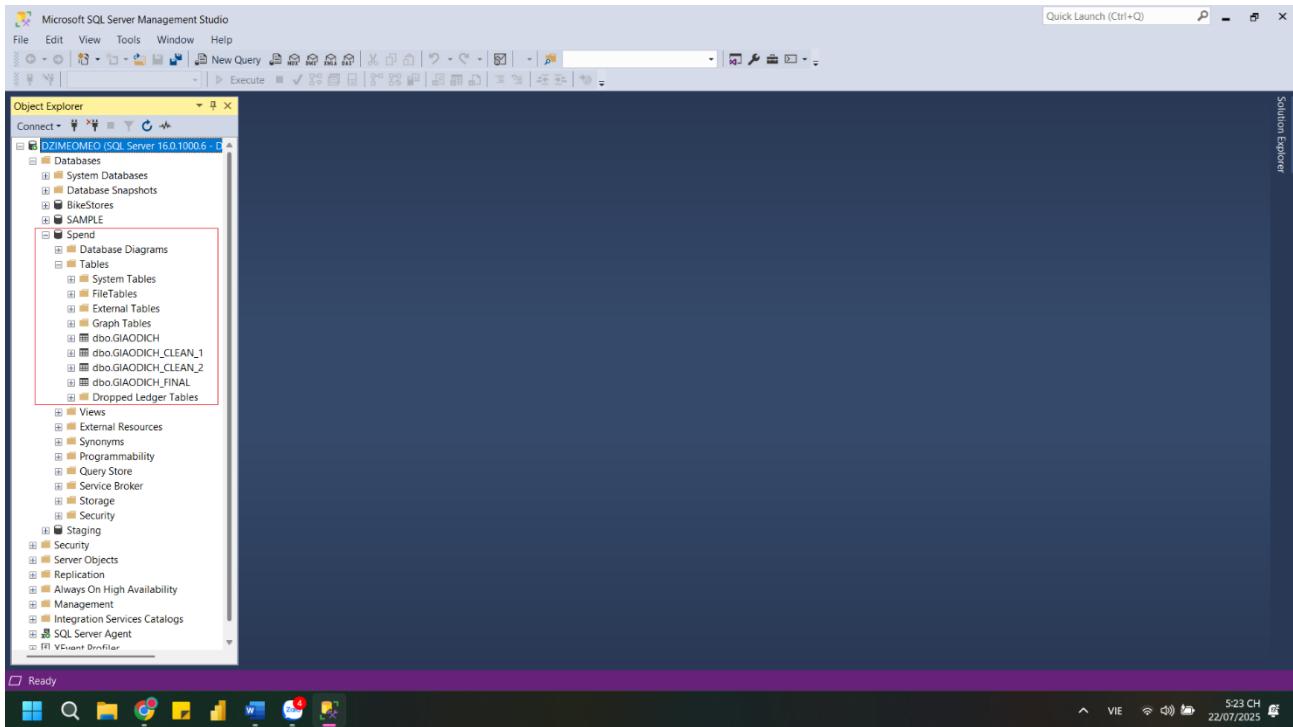
- Kiểm tra lại thông tin 1 lần nữa. Nhấn Finish.



- Nếu không có lỗi xảy ra thì database Staging sẽ được tạo.



- Refresh database kiểm tra Spend đã được tạo chưa.



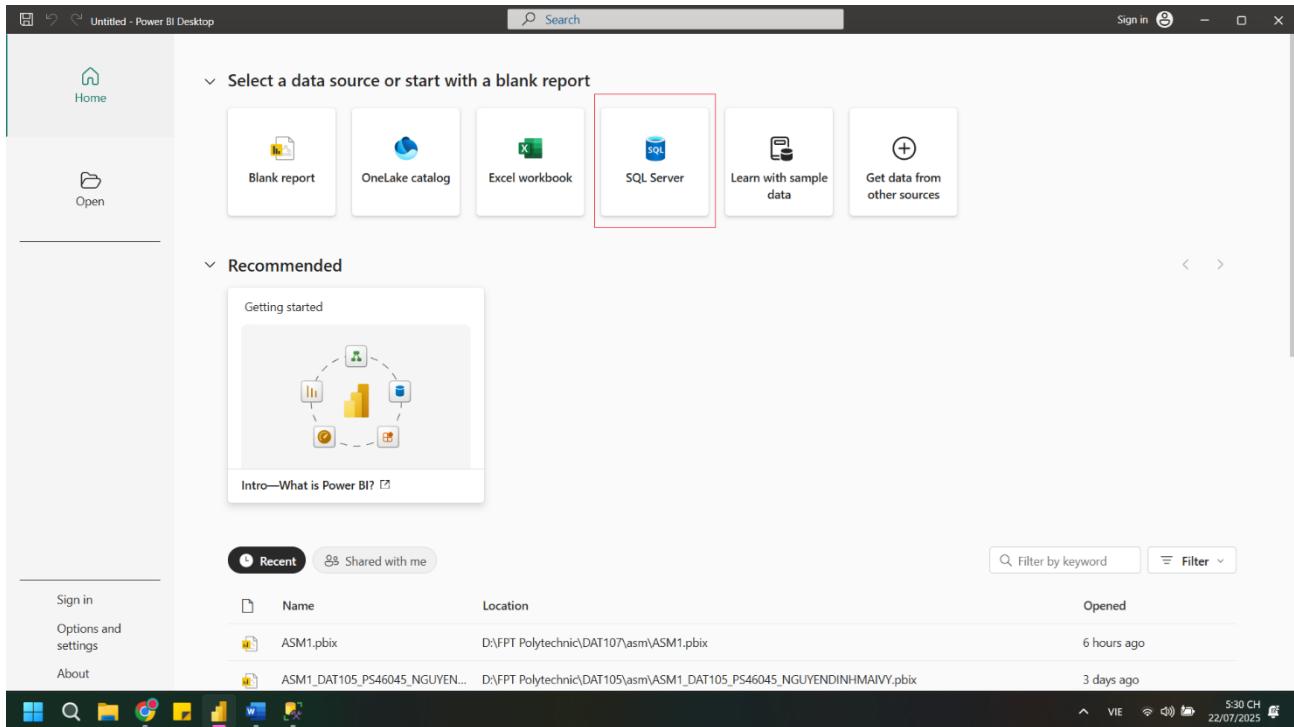
## 4 XỬ LÝ DỮ LIỆU

### 4.1 CHUẨN HÓA DỮ LIỆU

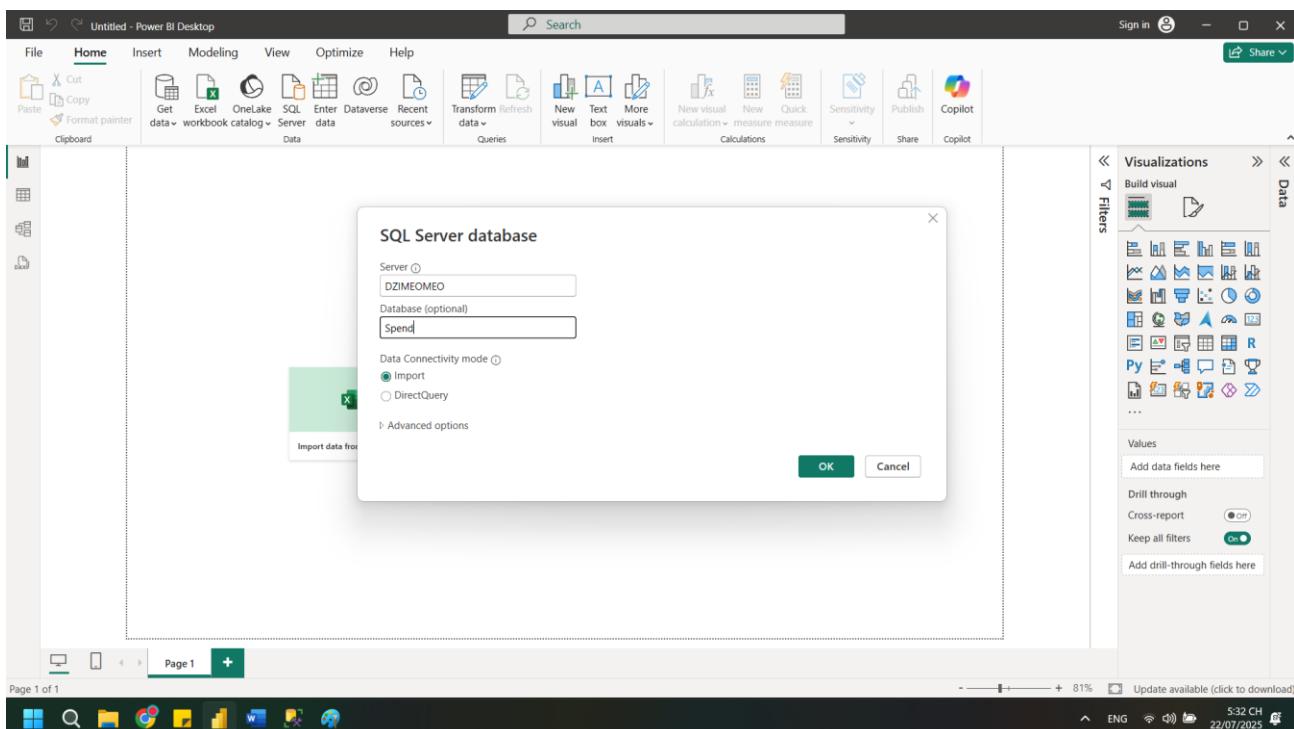
- ✓ Chuẩn hóa dữ liệu:
  - Power Query
  - Chính sửa table, column, data type

#### 4.1.1 TRÌNH BÀY CÁC BƯỚC CHUẨN HÓA TRONG DỰ ÁN

- ✓ Trình bày cụ thể các bước chuẩn hóa dữ liệu bằng Power Query tác động cụ thể đến bộ dữ liệu dự án (từng cột, từng dòng).
- ✓ Việc này giúp lưu vết các chỉnh sửa thay đổi trên.
  - Mở Power Bi Desktop. Chọn get data
  - Chọn Import data from SQL Server



- Nhập thông tin server và tên Database (trong trường hợp này là Spend). Nhấn OK.



- Các table trong database Spend đã được load vào như hình. Chọn table GIAODICH\_FINAL. Nhấn nút Transform Data.

The screenshot shows the Power BI Desktop interface. The main area displays a table titled "GIAODICH\_FINAL" with 999 rows and 12 columns. The columns are: ID, MaCK, NGAYGIAODICH, GIAMCUA, GIACAONHAT, GIATHAPNHAT, GIADONGCUA, KHOIQUONGGIAODICH, GIATRAN, GIASAN, TILETRONGNGAY, and TENNHOMNGA. The data includes various numerical values and some text entries like "Đầu tư công", "Thân Khoản", etc. On the left, the "Navigator" pane shows a tree structure with "DZIMEOME: Spend [4]" expanded, showing "GIAODICH", "GIAODICH\_CLEAN\_1", "GIAODICH\_CLEAN\_2", and "GIAODICH\_FINAL" (which is selected and highlighted with a red border). At the bottom right of the main area, there are buttons for "Load", "Transform Data", and "Cancel". The status bar at the bottom indicates "Page 1 of 1", "81%", "Update available (click to download)", "ENG", "5:33 CH", and "22/07/2025".

- Chọn Load table GIAODICH\_FINAL để bắt đầu xử lý.

The screenshot shows the Power Query Editor interface. The main area displays the "GIAODICH\_FINAL" query in preview mode, showing the same 12 columns and 999 rows as the table in Power BI. The "Transform" tab is currently selected. The "Properties" pane on the right shows the query name is "GIAODICH\_FINAL". The "Applied Steps" pane shows a single step named "Navigation". The status bar at the bottom indicates "PREVIEW DOWNLOADED AT 5:33 CH", "20 COLUMNS, 999+ ROWS", "Column profiling based on top 1000 rows", "ENG", "5:33 CH", and "22/07/2025".

- Rename GIAODICH\_FINAL thành [NGUON]

Power Query Editor - Untitled

File Home Transform Add Column View Tools Help

Queries [1] Source[[Schema="dbo",Item="GIAODICH\_FINAL"]][Data]

NGUON

	ID	MaCK	NGAYGIAODICH	GIAMOCUA	GIACONHAT	GIATHAPNHAT	GIADONGCUA												
1	5374 LCG	06/01/2023	100% Valid	100% Valid	100% Valid	100% Valid	100% Valid	13,6	13,3										
2	5382 LCM	05/30/2023	2,8	2,9															
3	5379 LCM	06/12/2023	3,1	3,2															
4	5403 LDP	06/12/2023	6,9	7															
5	5432 LGL	06/07/2023	4,61	5															
6	5434 LGL	06/12/2023	5,17	5,2															
7	5442 LHC	06/06/2023	49,2	49,3															
8	5443 LHC	06/09/2023	49,3	49,3															
9	5449 LHG	06/08/2023	29,55	30															
10	5465 LIC	05/30/2023	19,5	19,7															
11	5460 LIC	06/09/2023	19,5	20,3															
12	5471 LIG	06/05/2023	5,1	5,3															
13	5472 LIG	06/08/2023	5,4	5,6															
14	5496 LLM	06/07/2023	7,9	8															
15	5494 LLM	06/09/2023	7,8	7,8															
16	5518 LMC	05/30/2023	7	7															
17	5517 LMC	06/02/2023	7,9	8															
18	3389 FPT	06/02/2023	85,3	85,3															
19	3393 FPT	06/07/2023	85,3	85,4															
20	3411 FPT	06/08/2023	57,5	57,9															
21	3426 FTS	06/07/2023	28,75	29,9															

20 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 5:35 CH

- Remove cột [DuplicateRow] (vì không còn cần dùng nữa).

Power Query Editor - Untitled

File Home Transform Add Column View Tools Help

Queries [1] Source[[Schema="dbo",Item="GIAODICH\_FINAL"]][Data]

NGUON

	NGTY	TENSAN_VIET	TENSAN_ANH	BIENDODAODONG	DuplicateRow
1	y cổ phần LICOGI 16	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
2	y cổ phần Khai thác và Chế biến Khoáng sản Lào Cai	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
3	y cổ phần Khai thác và Chế biến Khoáng sản Lào Cai	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
4	y cổ phần Dược Lâm Đồng - Ladophar	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,1	
5	y cổ phần Đầu tư và Phát triển Đô thị Long Giang	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
6	y cổ phần Đầu tư và Phát triển Đô thị Long Giang	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
7	y cổ phần Đầu tư và Xây dựng Thủy lợi Lâm Đồng	Sở Giao dịch Chứng khoán Hà Nội	Ha Noi Stock Exchange	0,1	
8	y cổ phần Đầu tư và Xây dựng Thủy lợi Lâm Đồng	Sở Giao dịch Chứng khoán Hà Nội	Ha Noi Stock Exchange	0,1	
9	y cổ phần Long Hải	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
10	Ông ty Licoji	Công ty đại chúng	Unlisted Public Company Market	0,15	
11	Ông ty Licoji	Công ty đại chúng	Unlisted Public Company Market	0,15	
12	y cổ phần Licoji 13	Sở Giao dịch Chứng khoán Hà Nội	Ha Noi Stock Exchange	0,1	
13	y cổ phần Licoji 13	Sở Giao dịch Chứng khoán Hà Nội	Ha Noi Stock Exchange	0,1	
14	Ông ty Lắp máy Việt Nam	Công ty đại chúng	Unlisted Public Company Market	0,15	
15	Ông ty Lắp máy Việt Nam	Công ty đại chúng	Unlisted Public Company Market	0,15	
16	y cổ phần Khoáng sản Latca	Công ty đại chúng	Unlisted Public Company Market	0,15	
17	y cổ phần Khoáng sản Latca	Công ty đại chúng	Unlisted Public Company Market	0,15	
18	y cổ phần FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
19	y cổ phần FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
20	Án lê Kỹ thuật số FPT-HOSE	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	
21	y cổ phần Chứng khoán FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07	

20 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 5:35 CH

- Chuyển DataType [BIENDODAODONG] thành Percentage. Rename [BIENDODAODONG] thành [BIENDODAODONGSAN]

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply New Recent Sources Data Data source settings Manage Parameters Refresh Preview Advanced Editor Choose Columns Columns Keep Remove Rows Rows Split Column Group By Data Type: Decimal Number Use First Row as Headers Merge Queries Text Analytics

New Query Data Sources Parameters Query Manage Columns Remove Rows Sort Transform Combine AI Insights

Queries [1] Table.RemoveColumns(dbo\_GIAODICH\_FINAL, {"DuplicateRow"})

NGUON

	N	A <sub>C</sub> CONGTY	A <sub>C</sub> TENSAN_VIET	A <sub>C</sub> TENSAN_ANH	1.2 BIENDODAODONG
1		100% • Valid 0% • Error 0% • Empty	100% • Valid 0% • Error 0% • Empty	100% • Valid 0% • Error 0% • Empty	100% • Valid 0% • Error 0% • Empty
2		Công ty cổ phần Khai thác và Chế biến Khoáng sản Lào Cai	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	
3		Công ty Cổ phần Khai thác và Chế biến Khoáng sản Lào Cai	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	
4		Công ty Cổ phần Được Lâm Đồng - Ladophar	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	
5		Công ty cổ phần Đầu tư và Phát triển Đô thị Long Giang	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	
6		Công ty cổ phần Đầu tư và Phát triển Đô thị Long Giang	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	
7		Công ty Cổ phần Đầu tư và Xây dựng Thủ lợt Lâm Đồng	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	
8		Công ty Cổ phần Đầu tư và Xây dựng Thủ lợt Lâm Đồng	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	
9		Công ty Cổ phần Long Hậu	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07
10		Tổng Công ty Licogi 13	Công ty đại chúng	Unlisted Public Company Market	0,15
11		Tổng Công ty Licogi	Công ty đại chúng	Unlisted Public Company Market	0,15
12		Công ty Cổ phần Licogi 13	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	0,1
13		Công ty Cổ phần Licogi 13	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	0,1
14		Tổng Công ty Lắp máy Việt Nam	Công ty đại chúng	Unlisted Public Company Market	0,15
15		Tổng Công ty Lắp máy Việt Nam	Công ty đại chúng	Unlisted Public Company Market	0,15
16		Công ty Cổ phần Khoáng sản Latca	Công ty đại chúng	Unlisted Public Company Market	0,15
17		Công ty Cổ phần Khoáng sản Latca	Công ty đại chúng	Unlisted Public Company Market	0,15
18		Công ty Cổ phần FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07
19		Công ty Cổ phần FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07
20		CTCP Bán lẻ Kỹ thuật số FPT - HOSE	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07
21		Công ty Cổ phần Chứng khoán FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	0,07

19 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 5:35 CH

File Home Transform Add Column View Tools Help

Close & Apply New Recent Sources Data Data source settings Manage Parameters Refresh Preview Advanced Editor Choose Columns Columns Keep Remove Rows Rows Split Column Group By Data Type: Percentage Use First Row as Headers Merge Queries Text Analytics

New Query Data Sources Parameters Query Manage Columns Remove Rows Sort Transform Combine AI Insights

Queries [1] Table.RenameColumns(#"Changed Type", {"BIENDODAODONG", "BIENDODAODONGSAN"})

NGUON

	N	A <sub>C</sub> CONGTY	A <sub>C</sub> TENSAN_VIET	A <sub>C</sub> TENSAN_ANH	% BIENDODAODONGSAN
1		100% • Valid 0% • Error 0% • Empty	100% • Valid 0% • Error 0% • Empty	100% • Valid 0% • Error 0% • Empty	100% • Valid 0% • Error 0% • Empty
2		Công ty cổ phần Khai thác và Chế biến Khoáng sản Lào Cai	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%
3		Công ty Cổ phần Khai thác và Chế biến Khoáng sản Lào Cai	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%
4		Công ty Cổ phần Được Lâm Đồng - Ladophar	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	10,00%
5		Công ty cổ phần Đầu tư và Phát triển Đô thị Long Giang	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%
6		Công ty cổ phần Đầu tư và Phát triển Đô thị Long Giang	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%
7		Công ty Cổ phần Đầu tư và Xây dựng Thủ lợt Lâm Đồng	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	10,00%
8		Công ty Cổ phần Đầu tư và Xây dựng Thủ lợt Lâm Đồng	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	10,00%
9		Công ty Cổ phần Long Hậu	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%
10		Tổng Công ty Licogi	Công ty đại chúng	Unlisted Public Company Market	15,00%
11		Tổng Công ty Licogi 13	Công ty đại chúng	Unlisted Public Company Market	15,00%
12		Công ty Cổ phần Licogi 13	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	10,00%
13		Công ty Cổ phần Licogi 13	Sở Giao dịch Chứng khoán Hà Nội	Ho Nai Stock Exchange	10,00%
14		Tổng Công ty Lắp máy Việt Nam	Công ty đại chúng	Unlisted Public Company Market	15,00%
15		Tổng Công ty Lắp máy Việt Nam	Công ty đại chúng	Unlisted Public Company Market	15,00%
16		Công ty Cổ phần Khoáng sản Latca	Công ty đại chúng	Unlisted Public Company Market	15,00%
17		Công ty Cổ phần Khoáng sản Latca	Công ty đại chúng	Unlisted Public Company Market	15,00%
18		Công ty Cổ phần FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%
19		Công ty Cổ phần FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%
20		CTCP Bán lẻ Kỹ thuật số FPT - HOSE	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%
21		Công ty Cổ phần Chứng khoán FPT	Sở Giao dịch Chứng khoán Tp Hồ Chí Minh	Ho Chi Minh Stock Exchange	7,00%

19 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 5:38 CH

- Chuyển DataType cột [NGAYGIAODICH] thành kiểu Date.

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply New Recent Sources Data Data source settings Manage Parameters Refresh Preview Advanced Editor Properties Choose Columns Columns Remove Keep Rows Remove Rows Sort Split Column Group By Data Type: Text Use First Row as Headers Merge Queries Append Queries Text Analytics Vision Combine Files Azure Machine Learning AI Insights

Queries [1]

NGUON

Table.RenameColumns(#"Changed Type", {"BIENDODAODONG", "BIENDODAODONGSAN"})

NGAYGIAODICH 1.2 GIAMOCUA 1.2 GIACAOHNAT 1.2 GIATHAPNHAT 1.2 GIADONGCUA 1.2 KHK

Valid Error Empty

Date

1 5374 LCG 10 unique 639 distinct, 381 unique 13,5 2,8 13,6 13,3 13,5

2 5382 LCM 2,8 2,9 2,6 2,8

3 5379 LCM 3,1 3,2 3,1 3,2

4 5403 LDP 6,9 7 6,8 6,9

5 5432 LGL 4,61 5 4,61 4,89

6 5434 LGL 5,17 5,2 4,9 5,07

7 5442 LHC 49,2 49,3 48,9 48,9

8 5443 LHC 49,3 49,3 48,9 48,9

9 5449 LHG 29,55 30 29 29,1

10 5465 LIC 19,5 19,7 18,2 18,9

11 5460 LIC 19,5 20,3 18,7 19,2

12 5471 LIG 5,1 5,3 5 5,2

13 5472 LIG 5,4 5,6 5,2 5,2

14 5496 LLM 7,9 8 7,9 7,9

15 5494 LLM 7,8 7,8 7,8 7,8

16 5518 LMC 7 7 7 7,1

17 5517 LMC 7,9 8 7,9 8

18 3389 FPT 85,3 85,3 84,7 85,3

19 3393 FPT 85,3 85,4 84,7 85,3

20 3411 FRT 57,5 57,9 57,1 57,2

21 3426 FTS 28,75 29,9 28,4 29,75

19 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 5:38 CH

ENG ⚡ 5:38 CH 22/07/2025

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply New Recent Sources Data Data source settings Manage Parameters Refresh Preview Advanced Editor Properties Choose Columns Columns Remove Keep Rows Remove Rows Sort Split Column Group By Data Type: Date Use First Row as Headers Merge Queries Append Queries Text Analytics Vision Combine Files Azure Machine Learning AI Insights

Queries [1]

NGUON

Table.TransformColumnTypes(#"Renamed Columns", {"NGAYGIAODICH", "type date"})

NGAYGIAODICH 1.2 GIAMOCUA 1.2 GIACAOHNAT 1.2 GIATHAPNHAT 1.2 GIADONGCUA

Valid Error Empty

Date

1 5374 LCG 1000 distinct, 1000 unique 639 distinct, 381 unique 10 distinct, 0 unique 01/06/2023 13,5 13,6 13,3

2 5382 LCM 30/05/2023 2,8 2,9 2,6

3 5379 LCM 12/06/2023 3,1 3,2 3,1

4 5403 LDP 12/06/2023 6,9 7 6,8

5 5432 LGL 07/06/2023 4,61 5 4,61

6 5434 LGL 12/06/2023 5,17 5,2 4,9

7 5442 LHC 06/06/2023 49,2 49,3 48,9

8 5443 LHC 09/06/2023 49,3 49,3 48,9

9 5449 LHG 08/06/2023 29,55 30 29

10 5465 LIC 30/05/2023 19,5 19,7 18,2

11 5460 LIC 09/06/2023 19,5 20,3 18,7

12 5471 LIG 05/06/2023 5,1 5,3 5

13 5472 LIG 08/06/2023 5,4 5,6 5,2

14 5496 LLM 07/06/2023 7,9 8 7,9

15 5494 LLM 09/06/2023 7,8 7,8 7,8

16 5518 LMC 30/05/2023 7 7 7

17 5517 LMC 02/06/2023 7,9 8 7,9

18 3389 FPT 02/06/2023 85,3 85,3 84,7

19 3393 FPT 07/06/2023 85,3 85,4 84,7

20 3411 FRT 08/06/2023 57,5 57,9 57,1

21 3426 FTS 07/06/2023 28,75 29,9 28,4

19 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 5:41 CH

ENG ⚡ 5:42 CH 22/07/2025

- Thay thế giá trị null trong cột [THONGTINCONGTY] thành "Chưa cập nhật thông tin"

The screenshot shows the Power Query Editor interface with the 'Transform' tab selected. A 'Replace Values' dialog box is open over a table query named 'NGUON'. The dialog has 'Value To Find' set to 'null' and 'Replace With' set to 'Chưa cập nhật thông tin'. A red arrow points from the 'Replace With' input field towards the 'OK' button.

- Duplicate cột [NGAYGIAODICH] thành [NGAYGIAODICH-Copy]

The screenshot shows the Power Query Editor interface with the 'Transform' tab selected. A context menu is open over the 'NGAYGIAODICH' column in the 'NGUON' query. The 'Duplicate Column' option is highlighted with a red box. Other options visible in the menu include 'Copy', 'Remove', 'Remove Other Columns', 'Add Column From Examples...', 'Remove Duplicates', 'Remove Errors', 'Change Type', 'Transform', 'Replace Values...', 'Replace Errors...', 'Group By...', 'Unpivot Columns', 'Unpivot Other Columns', 'Unpivot Only Selected Columns', 'Rename...', 'Move', 'Drill Down', and 'Add as New Query'.

- Sắp xếp lại cột [NGAYGIAODICH-Copy] bên phải cột [NGAYGIAODICH]

The screenshot shows the Power Query Editor interface with the 'Reordered Columns' step selected in the 'Applied Steps' pane. The table contains 20 columns and 999+ rows. The columns are: ID, MaCK, NGAYGIAODICH, NGAYGIAODICH - Copy, GIAMOCUA, GIACAOHAT, and GIATHAPNHAT. The 'Properties' pane shows the query name is 'NGUON'. The 'Applied Steps' pane lists the steps taken: Source, Navigation, Removed Columns, Renamed Columns, Changed Type with Locale, Replaced Value, Duplicated Column, and Reordered Columns.

- Tách cột [NGAYGIAODICH-Copy] thành 3 cột [THANG], [NGAY],[NAM]
- Chọn Split Column > By delimiter

The screenshot shows the Power Query Editor interface with the 'Split Column' step selected in the 'Applied Steps' pane. The 'Properties' pane shows the query name is 'NGUON'. The 'Applied Steps' pane lists the steps taken: Source, Navigation, Removed Columns, Renamed Columns, Changed Type with Locale, Replaced Value, Duplicated Column, and Reordered Columns.

- Hiện bảng chọn tách bởi “ / ” -> nhấn OK

Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

Custom /

Split at

Left-most delimiter

Right-most delimiter

Each occurrence of the delimiter

Advanced options

Quote Character

Split using special characters

Insert special character

OK Cancel

PREVIEW DOWNLOADED AT 5:42 CH

- Đổi tên cột thành NGAY, THANG, NAM

Table.RenameColumns

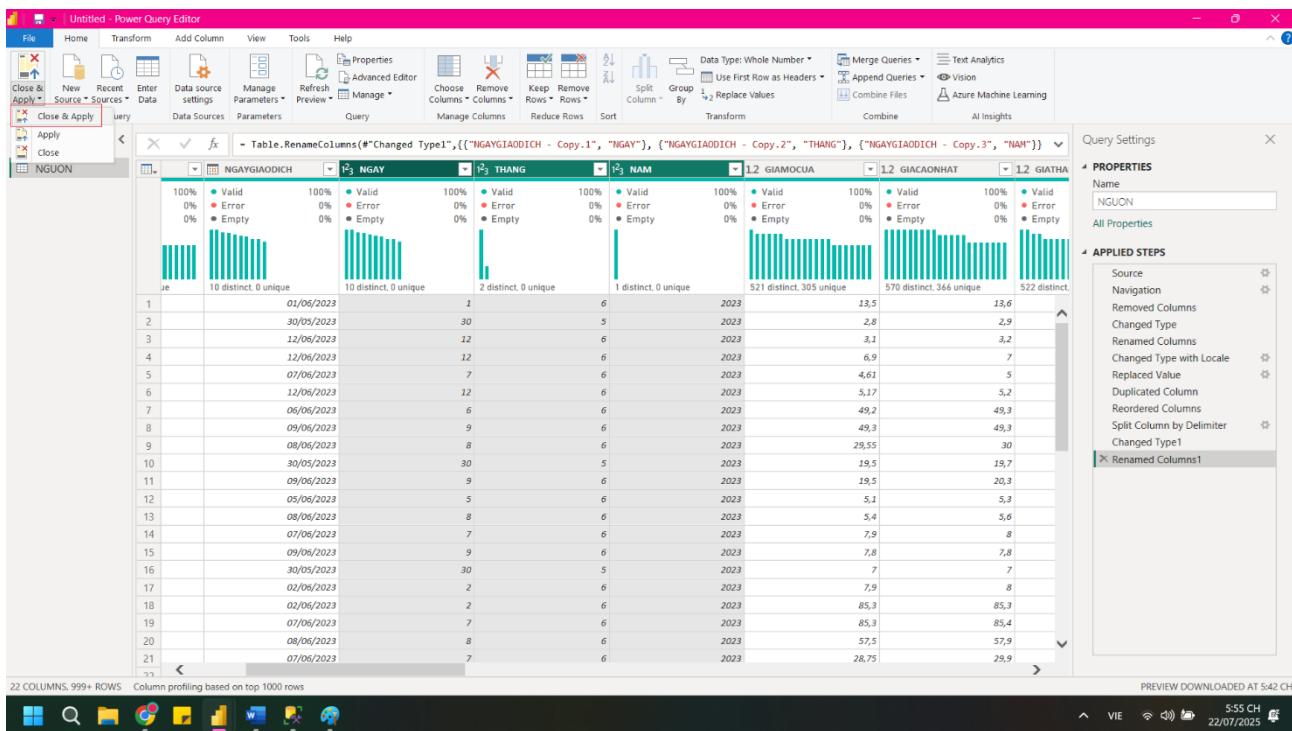
Source: NGUON

Target:

Original Column	New Column
NGAYGIAODICH	NGAY
NGAYGIAODICH - Copy.1	NGAY
NGAYGIAODICH - Copy.2	THANG
NGAYGIAODICH - Copy.3	NAM

PREVIEW DOWNLOADED AT 5:42 CH

- Close và Apply dự án.



## 4.2 MÔ HÌNH HÓA DỮ LIỆU

- ✓ *Mô hình hóa dữ liệu:*
  - *Data model*
  - *Data modeling*

### 4.2.1 CÁC LOẠI MÔ HÌNH HÓA

- ✓ *Trình bày các loại mô hình*
- ✓ *Mô hình nào đang được áp dụng cho bộ dữ liệu dự án ?*

#### ❑ *Conceptual data models (Mô hình Dữ liệu Khái niệm)*

❖ Khái niệm:

- Mô hình dữ liệu khái niệm ghi nhận các mối quan hệ cấp cao nhất giữa các thực thể khác nhau.

❖ Đặc điểm:

- Chứa các thực thể thiết yếu và các mối quan hệ giữa chúng.
- Không có thuộc tính nào được chỉ định.

- Không có khóa chính nào được chỉ định, có thể thấy rằng dữ liệu duy nhất được hiển thị qua mô hình dữ liệu khái niệm là các thực thể xác định dữ liệu và mối quan hệ giữa các thực thể đó.

- Không có dữ liệu nào khác, như được hiển thị thông qua mô hình dữ liệu khái niệm

#### **◻ Logical Data Model (Mô hình Dữ liệu Logic)**

❖ Khái niệm:

- Logical Data Model xác định thông tin theo cấu trúc mà không cần quan tâm về mặt vật lý trong cơ sở dữ liệu.

❖ Đặc điểm:

- Mục tiêu chính của mô hình dữ liệu logic là ghi lại cấu trúc dữ liệu kinh doanh, quy trình, quy tắc và mối quan hệ bằng một chế độ xem duy nhất.

#### **◻ Physical data models (Mô hình Dữ liệu Vật lý)**

❖ Khái niệm:

- Mô hình dữ liệu này mô tả cách hệ thống sẽ được triển khai bằng cách sử dụng một hệ thống quản lý cơ sở dữ liệu cụ thể.

❖ Đặc điểm:

- Mô hình này thường được tạo bởi chuyên viên quản trị dữ liệu và các nhà phát triển. Mục đích là triển khai thực tế cơ sở dữ liệu.

### **4.2.2 CÁC TIÊU CHÍ ĐÁNH GIÁ MÔ HÌNH DỮ LIỆU**

- ✓ *Trình bày các tiêu chí đánh giá mô hình dữ liệu tốt ?*
- ✓ *Mô hình đang dùng trong dự án đáp ứng các tiêu chí nào ?*

#### **◻ Clearness (Tính rõ ràng)**

❖ Sự dễ hiểu đối với những người sử dụng.

❖ Hầu hết thời gian developers đọc mã thay vì viết.

❖ Vì vậy chúng ta cần hiểu rõ ràng những gì chúng ta đang làm với dữ liệu của mình.

### **❑ Flexibility (Tính linh hoạt)**

- ❖ Khả năng phát triển của mô hình mà không cần phải tác động quá lớn đến các đoạn code.
- ❖ Công ty startup mà bạn làm việc đang phát triển, vì vậy các hệ thống sẽ thay đổi và các mô hình dữ liệu đãng sau sẽ cần phải phát triển theo thời gian.

### **❑ Performance (Hiệu suất)**

- ❖ Đây là một chủ đề rất rộng.
- ❖ Trong môn học này sẽ không nói về các nhà cung cấp cơ sở dữ liệu (database vendors) hoặc một số chỉnh sửa kỹ thuật để cải thiện tốc độ đọc và ghi dữ liệu.
- ❖ Cách thức thiết kế data model đúng đắn cũng đem lại lợi ích về hiệu suất.

### **❑ Productivity (Năng suất)**

- ❖ Dưới góc nhìn của lập trình viên (developer), chắc hẳn bạn sẽ muốn có một mô hình dữ liệu dễ làm việc mà không cần sử dụng nhiều thời gian.

### **❑ Traceability (Khả năng truy xuất nguồn gốc)**

- ❖ Cuối cùng, các công ty không chỉ muốn có dữ liệu liên quan đến người dùng của mình mà còn có dữ liệu liên quan đến chính hệ thống.
- ❖ Dữ liệu có thể cung cấp thông tin những gì đã xảy ra trong quá khứ, những giá trị công ty có tại một thời điểm nào đó.

## **4.2.3 TRÌNH BÀY CÁC BƯỚC MÔ HÌNH HÓA**

- ✓ *Trình bày cụ thể các bước mô hình hóa dữ liệu bằng Power BI trên bộ dữ liệu của dự án.*
- ✓ *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

Mô hình hóa dữ liệu theo thứ tự SAN > NGANH > CHUNGKHOAN.

### **1. SAN.**

- Chọn cột [SAN], Group by

The screenshot shows the Power Query Editor interface with the 'Transform' tab selected. A 'Group By' dialog is open, prompting the user to choose a column to group by ('SAN') and an operation ('All Rows'). The main table view shows four columns: SAN, CONGTY, TENSAN, and TENSAN\_VIET. The 'APPLIED STEPS' pane on the right lists various transformations applied to the query, including 'Renamed Columns1'.

- Thêm cột mới, đánh index, đặt tên là [SAN\_ID]

The screenshot shows the Power Query Editor interface with the 'Transform' tab selected. An 'Index Column' dialog is open, with the 'From 1' option highlighted. The main table view shows a single column named 'Group\_BY'. The 'APPLIED STEPS' pane on the right lists 'Renamed Columns1' and 'Grouped Rows'.

The screenshot shows the Power Query Editor interface. The main area displays a table with four columns: SAN\_ID, SAN, and Group\_SAN. The Group\_SAN column contains four distinct values: HOSE, HNX, Upcom, and OTC. The 'APPLIED STEPS' pane on the right lists several steps, including 'Renamed Columns2'. The status bar at the bottom indicates 'PREVIEW DOWNLOADED AT 5:42 CH'.

- Trả lại trạng thái ban đầu cho cột để tiếp tục phân tích.

The screenshot shows the Power Query Editor with the 'Group By' dialog open. The 'Group\_SAN' column is selected for grouping. The 'OK' button is highlighted with a red box. The status bar at the bottom indicates 'PREVIEW DOWNLOADED AT 5:42 CH'.

## 2. NGANH.

- Chọn cột [NGANH], Group by

The screenshot shows the Power Query Editor interface with a table named 'Table.ExpandTableColumn("Renamed Columns2", "Group\_SAN", {"ID", "MaCK", "NGAYGIAODICH", "NGAY", "THANG", "NAM", "GIANHUA", "GIACAOHAT", "MANHOMNGANH", "THONGTINCONGTY"}). The 'THONGTINCONGTY' column has a context menu open, with the 'Group By...' option highlighted by a red box.

The screenshot shows the 'Group By' dialog box in the Power Query Editor. The 'Basic' tab is selected, and the 'MANHOMNGANH' column is chosen as the grouping column. The 'New column name' field contains 'Group\_NGANH'. The 'OK' button is visible at the bottom right of the dialog.

- Thêm cột mới, đánh index

The screenshot shows the Power Query Editor interface. In the 'APPLIED STEPS' pane, the 'Grouped Rows1' step is highlighted. The 'Source' step is also listed. The 'Queries' pane contains a table named 'NGUON' with 22 rows and 2 columns. The 'Grouped Rows' step has been applied to this table.

- Đổi tên cột [NGANH\_ID]

The screenshot shows the Power Query Editor interface. In the 'APPLIED STEPS' pane, the 'Renamed Columns3' step is highlighted. The 'Source' step is also listed. The 'Queries' pane contains a table named 'NGUON' with 22 rows and 3 columns. The 'Renamed Columns' step has been applied to this table.

- Trả lại trạng thái ban đầu cho cột để tiếp tục phân tích.

The screenshot shows the Power Query Editor interface. In the center, a 'Rename Columns' dialog is open, displaying a list of columns from a table named 'Group\_NGANH'. One column, 'MANHOMNGANH', is highlighted with a red border. The 'APPLIED STEPS' pane on the right lists various actions taken on the query, including 'Renamed Columns3'.

### 3. CHUNGKHOAN

- Chọn cột [MaCK], Group by

The screenshot shows the Power Query Editor interface. A 'Group By' dialog is open, specifying 'MaCK' as the column to group by. The 'APPLIED STEPS' pane on the right lists various actions taken on the query, including 'Grouped Rows2'.

- Thêm cột mới, đánh index, đặt tên là [MaCK\_ID]

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Column From Custom Invoke Custom Examples Column Function General

Index Column **From 1**

Conditional Column Merge Columns Statistics Standard Scientific Information Date Time Duration Text Analytics Vision Azure Machine Learning

Queries [1] Custom...

Table.Group(#"Expanded Group\_NGANH", {"MaCK"}, {"Group\_MaCK", each \_, type table [NGANH\_ID=number, MANHOMNGANH=nullable text], Grouped Rows1}, Grouped Rows2)

**NGUON**

MaCK Group\_MaCK

Valid: 100% Error: 0% Empty: 0% Valid: 100% Error: 0% Empty: 0%

1000 distinct. 1000 unique

1 LCG Table  
2 GVR Table  
3 DPR Table  
4 VSI Table  
5 HUB Table  
6 SBV Table  
7 SAV Table  
8 NO1 Table  
9 DHC Table  
10 TDW Table  
11 VC9 Table  
12 VCG Table  
13 IDC Table  
14 PDB Table  
15 GTS Table  
16 H11 Table  
17 HEP Table  
18 HGW Table  
19 BDW Table  
20 BWA Table  
21 UCT Table  
22 UPA Table

2 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 5:42 CH

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Column From Custom Invoke Custom Examples Column Function General

Index Column Duplicate Column Format Parse From Text From Number From Date & Time

Conditional Column Merge Columns Statistics Standard Scientific Information Date Time Duration Text Analytics Vision Azure Machine Learning

Queries [1] Custom...

Table.RenameColumns(#"Reordered Columns3", {"Index", "MaCK\_ID"})

**NGUON**

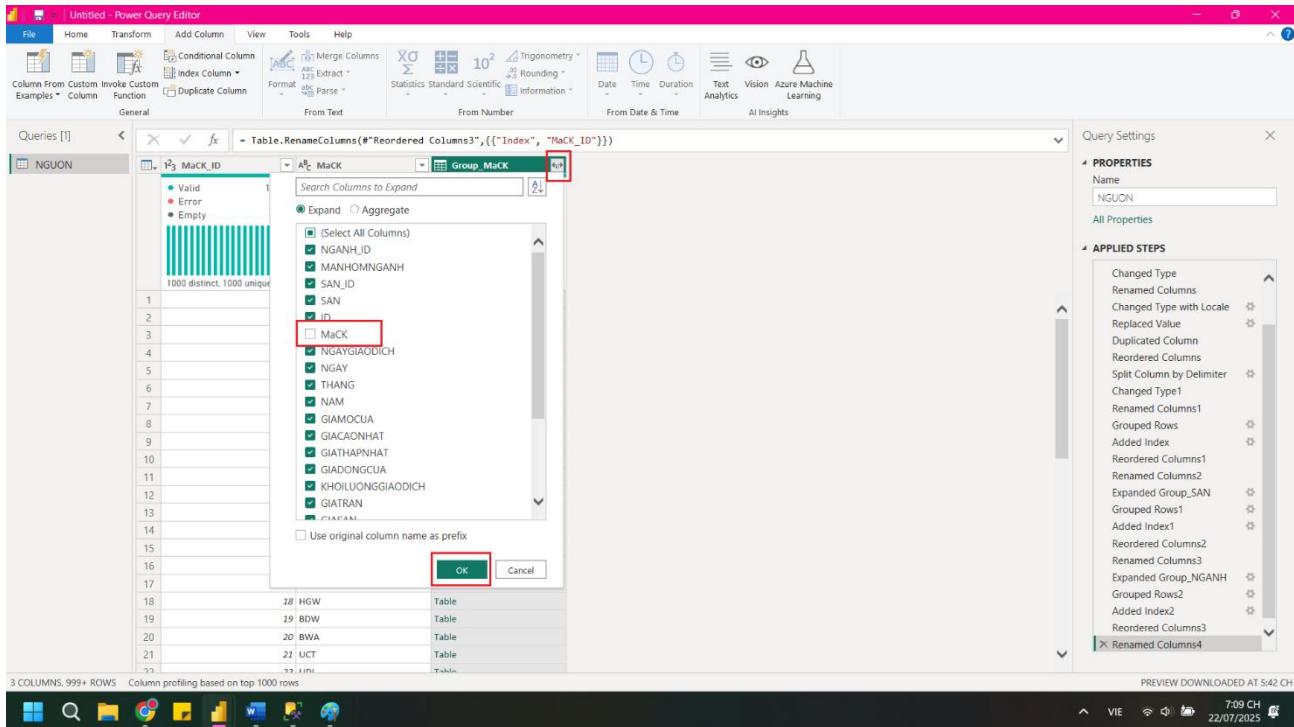
MaCK\_ID MaCK Group\_MaCK

1 LCG Table  
2 GVR Table  
3 DPR Table  
4 VSI Table  
5 HUB Table  
6 SBV Table  
7 SAV Table  
8 NO1 Table  
9 DHC Table  
10 TDW Table  
11 VC9 Table  
12 VCG Table  
13 IDC Table  
14 PDB Table  
15 GTS Table  
16 H11 Table  
17 HEP Table  
18 HGW Table  
19 BDW Table  
20 BWA Table  
21 UCT Table  
22 UPA Table

3 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 5:42 CH

- Trả lại trạng thái ban đầu cho cột để tiếp tục phân tích.



#### 4.2.4 TRÌNH BÀY CÁC BƯỚC TẠO BẢNG DỮ LIỆU

- ✓ *Trình bày cụ thể các bước tạo bảng dữ liệu bằng Power BI trên bộ dữ liệu của dự án.*
- ✓ *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

##### 1. Tạo bảng tbl\_SAN

- Tạo bảng Reference từ NGUON

- Rename thành tbl\_SAN

Untitled - Power Query Editor

Queries [2]

tbl\_SAN

Properties

APPLIED STEPS

Source

Query Settings

Name: tbl\_SAN

All Properties

APPLIED STEPS

Source

PREVIEW DOWNLOADED AT 5:42 CH

7:34 CH 22/07/2025

- Giữ lại các cột ([SAN\_ID], [SAN], [TENSAN\_VIET], [TENSAN\_ANH], [BIENDODAODONGSAN]) > nhấn OK

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Queries [2]

tbl\_SAN

Properties

APPLIED STEPS

Source

Choose Columns

Choose the columns to keep

san

(Select All Search Results)

BIENDODAODONGSAN

GIASAN

SAN

SAN\_ID

TENSAN\_ANH

TENSAN\_VIET

OK Cancel

PREVIEW DOWNLOADED AT 5:42 CH

7:35 CH 22/07/2025

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Data Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Properties Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Replace Values Data Type: Text Use First Row as Headers Merge Queries Append Queries Text Analytics Combine Files Azure Machine Learning Close New Query Queries [2]

NGUON

tbl\_SAN

1.2 SAN\_ID A[=] SAN A[=] TENSAN\_VIET A[=] TENSAN\_ANH 1.2 BIENDOADAODONGSAN

	1.2 SAN_ID	A[=] SAN	A[=] TENSAN_VIET	A[=] TENSAN_ANH	1.2 BIENDOADAODONGSAN
1	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
2	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
3	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
4	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
5	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
6	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
7	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
8	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
9	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
10	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
11	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
12	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
13	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
14	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
15	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
16	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
17	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
18	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
19	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
20	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
21	1	HOSE	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07
22	1	UNKN.	Sđ Giao dịch Chứng khoán Tp Hồ Chí Minh	Hồ Chí Minh Stock Exchange	0,07

5 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

Query Settings

PROPERTIES

Name: tbl\_SAN

All Properties

APPLIED STEPS

Removed Other Columns Reordered Columns

PREVIEW DOWNLOADED AT 5:42 CH 7:36 ENG 22/07/2025

- Remove duplicate data.

The screenshot shows the Microsoft Power Query Editor interface. The ribbon at the top includes File, Home, Transform, Add Column, View, Tools, and Help. The main area displays a query for 'tbl\_SAN'. The 'Transform' tab is selected, showing various data manipulation tools like 'Remove Rows', 'Remove Duplicates', and 'Remove Blank Rows'. A context menu is open over the 'TENSAN\_ANH' column, with 'Remove Duplicates' highlighted. The 'APPLIED STEPS' pane on the right shows a step named 'Removed Other Columns'. The bottom status bar indicates there are 5 COLUMNS, 999+ ROWS, and the preview was downloaded at 5:42 CH.

Query Settings

- PROPERTIES**
  - Name: `tbl_SAN`
  - All Properties
- APPLIED STEPS**
  - Source
  - Removed Other Columns
  - Reordered Columns
  - Removed Duplicates**

## 2. Tạo bảng `tbl_NGANH`

- Tạo bảng Reference từ NGUON

Query Settings

- PROPERTIES**
  - Name: `tbl_SAN`
  - All Properties
- APPLIED STEPS**
  - Source
  - Removed Other Columns
  - Reordered Columns
  - Removed Duplicates**

- Rename thành `tbl_NGANH`

Query Settings

- PROPERTIES**
  - Name: `tbl_NGANH`
  - All Properties
- APPLIED STEPS**
  - Source

- Giữ lại các cột (`[NGANH_ID]`, `[MANHOMNGANH]`, `[TENNTHONGNGANH]`) > OK

Choose Columns

Choose the columns to keep

Search Columns

- MaCK
- NGANH\_ID
- MANHOMNGANH
- SAN\_ID
- SAN
- ID
- NGAY
- THANG
- NAM
- GIAMOCUA
- GIACAOHAT
- GIATHAPNHAT
- GIADONGCUA
- KHOILUONGGIAODICH
- GIASAN
- TILETRONGNGAY
- TENTHONGNGANH
- THONGTINCONGTY

OK Cancel

- Remove duplicate data.

The screenshot shows the Power Query Editor interface. In the center, there is a table with three columns: 'L2 NGANH\_ID', 'A<sup>b</sup> MANHOMNGANH', and 'A<sup>b</sup> TENNHOMNGANH'. The 'MANHOMNGANH' and 'TENNHOMNGANH' columns contain identical data. A context menu is open over the first row of the 'MANHOMNGANH' column, with the 'Remove Duplicates' option highlighted.

**Query Settings**

- PROPERTIES**
  - Name: tbLNGANH
- APPLIED STEPS**
  - Source
  - Removed Other Columns

PREVIEW DOWNLOADED AT 5:42 CH  
7:44 CH 22/07/2025

The screenshot shows the Power Query Editor interface after applying the 'Removed Other Columns' step. The table now has 22 distinct rows, each containing unique values for 'MANHOMNGANH' and 'TENNHOMNGANH'. The 'APPLIED STEPS' list now includes both 'Removed Other Columns' and 'Removed Duplicates'.

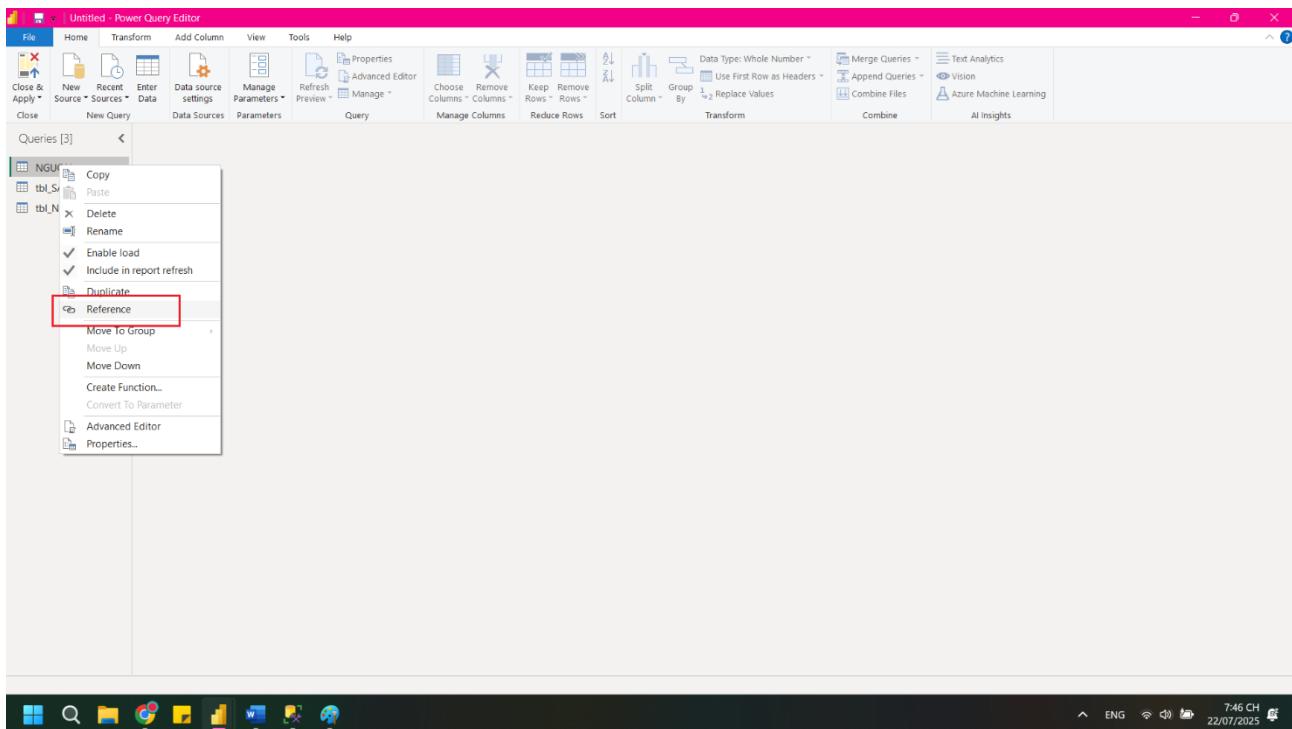
**Query Settings**

- PROPERTIES**
  - Name: tbLNGANH
- APPLIED STEPS**
  - Source
  - Removed Other Columns
  - Removed Duplicates

PREVIEW DOWNLOADED AT 5:42 CH  
7:45 CH 22/07/2025

### 3. Tạo bảng tbl\_CHUNGKHOAN

- Tạo bảng Reference từ NGUON



- Rename thành `tbl_CHUNGKHOAN`

**Query Settings**

- PROPERTIES**
  - Name: `tbl_CHUNGKHOAN`
  - All Properties
- APPLIED STEPS**
  - Source

PREVIEW DOWNLOADED AT 5:42 CH  
25 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

- Giữ lại các cột (`[MaCK_ID]`, `[MaCK]`, `[MANHOMNGANH]`, `[SAN]`, `[THONGTINCONGTY]`, `[CONGTY]`)

- Remove duplicate data.

## 4. Tạo bảng tbl\_GIAODICH

- Tạo bảng Reference từ NGUON

The screenshot shows the Power Query Editor interface with the 'tbl\_CHUNGKHOAN' query selected. A context menu is open over the 'tbl\_CHUNGKHOAN' table, with the 'Reference' option highlighted. The 'APPLIED STEPS' pane shows a step named 'Removed Duplicates'.

- Rename thành tbl\_GIAODICH

The screenshot shows the Power Query Editor interface with the 'tbl\_GIAODICH' query selected. The table structure includes columns: MaCK\_ID, NGANH\_ID, MANHOMNGANH, SAN\_ID, SAN, and ID. The 'APPLIED STEPS' pane shows a step named 'Source'.

- Giữ lại các cột ([MaCK], [NGAYGIAODICH], [THANG], [NGAY], [NAM], [GIAMOCUA], [GIACAONHAT], [GIATHAPNHAT],

## [GIADONGCUA], [KHOILUONGGIAODICH], [GIATRAN], [GIASAN], [TILETRONGNGAY])

Choose Columns

Choose the columns to keep

Search Columns

tbl\_GIAODICH

Column	Type	Valid	Error	Empty
MaCK	Whole Number	100%	0%	0%
NGANH_ID	Whole Number	100%	0%	0%
MANHONGNANH	Text	0%	0%	100%
SAN_ID	Text	0%	0%	100%
ID	Text	0%	0%	100%
NGAYGIAODICH	Text	100%	0%	0%
NGAY	Text	100%	0%	0%
THANG	Text	100%	0%	0%
NAM	Text	100%	0%	0%
GIAMOCUA	Text	100%	0%	0%
GIAONHAT	Text	100%	0%	0%
GIATHAPNHAT	Text	100%	0%	0%
GIADONGCUA	Text	100%	0%	0%
KHOILUONGGIAODICH	Text	100%	0%	0%
GIATRAN	Text	100%	0%	0%
GIASAN	Text	100%	0%	0%
TILETRONGNGAY	Text	100%	0%	0%
TENNTHONGNGAY	Text	0%	100%	0%
THONHTINCONGTY	Text	0%	100%	0%

OK Cancel

Properties

Name: tbl\_GIAODICH

Applied Steps

Source

Query Settings

PREVIEW DOWNLOADED AT 5:42 CH 22/07/2025

File Home Transform Add Column View Tools Help

Queries [5]

tbl\_GIAODICH

Table.SelectColumns(Source, { "MaCK", "NGAYGIAODICH", "NGAY", "THANG", "NAM", "GIAMOCUA", "GIAONHAT", "GIATHAPNHAT", "GIADONGCUA", "KHOILUONGGIAODICH", "GIATRAN", "GIASAN", "TILETRONGNGAY", "TENNTHONGNGAY", "THONHTINCONGTY" })

Column	Type	Valid	Error	Empty
MaCK	Whole Number	100%	0%	0%
NGAYGIAODICH	Text	100%	0%	0%
NGAY	Text	100%	0%	0%
THANG	Text	100%	0%	0%
NAM	Text	100%	0%	0%
GIAMOCUA	Text	100%	0%	0%
GIAONHAT	Text	100%	0%	0%
GIATHAPNHAT	Text	100%	0%	0%
GIADONGCUA	Text	100%	0%	0%
KHOILUONGGIAODICH	Text	100%	0%	0%
GIATRAN	Text	100%	0%	0%
GIASAN	Text	100%	0%	0%
TILETRONGNGAY	Text	100%	0%	0%
TENNTHONGNGAY	Text	0%	100%	0%
THONHTINCONGTY	Text	0%	100%	0%

Properties

Name: tbl\_GIAODICH

Applied Steps

Source

Removed Other Columns

Query Settings

PREVIEW DOWNLOADED AT 5:42 CH 22/07/2025

- Không nên dùng Remove duplicate data tại table có dữ liệu biến động.
- Unload table NGUON.

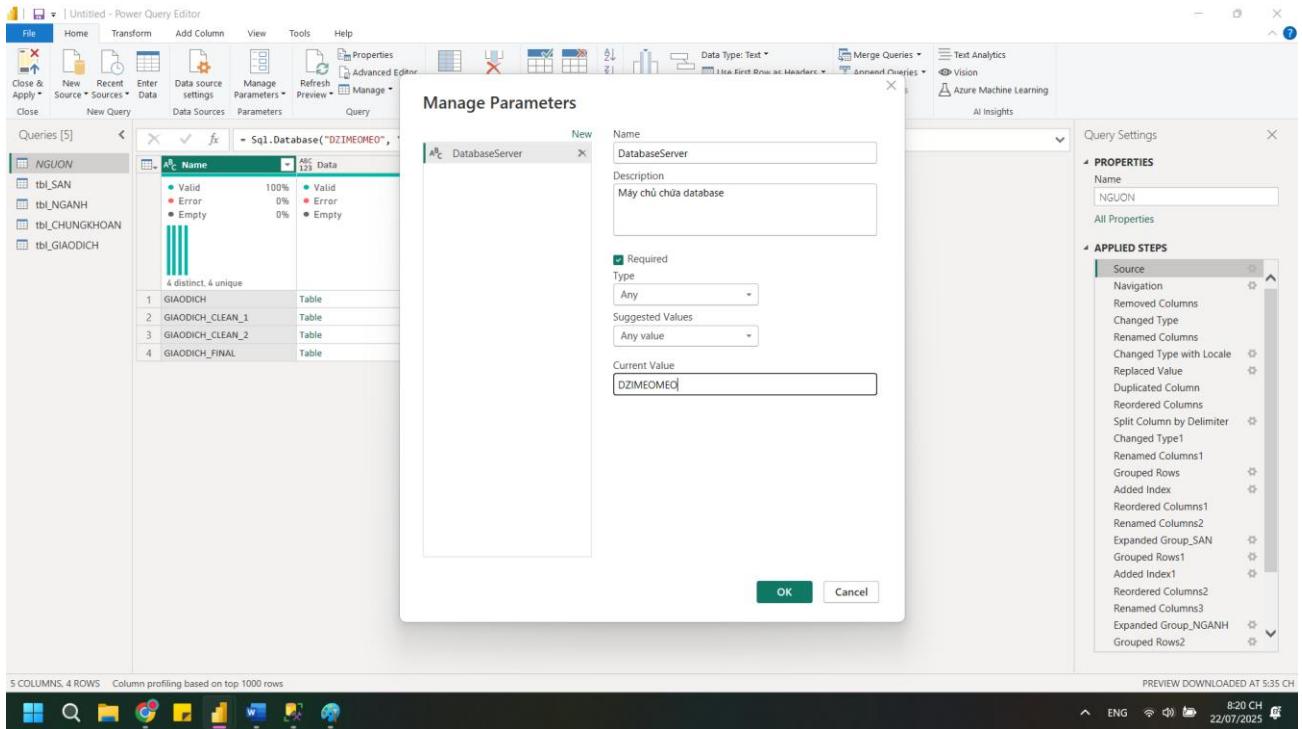
The screenshot shows the Power Query Editor interface. A query named 'tbl\_GIAODICH' is selected. The context menu for this query is open, with the 'Enable load' option highlighted. The main pane displays a table with columns: NGUO, NGAYGIODICH, NGAY, THANG, NAM, GIAMCUA, and GIAONHAT. The table has 13 columns and 999+ rows. The status bar at the bottom indicates 'PREVIEW DOWNLOADED AT 5:42 CH'.

Tạo param cho kết nối SQL database.

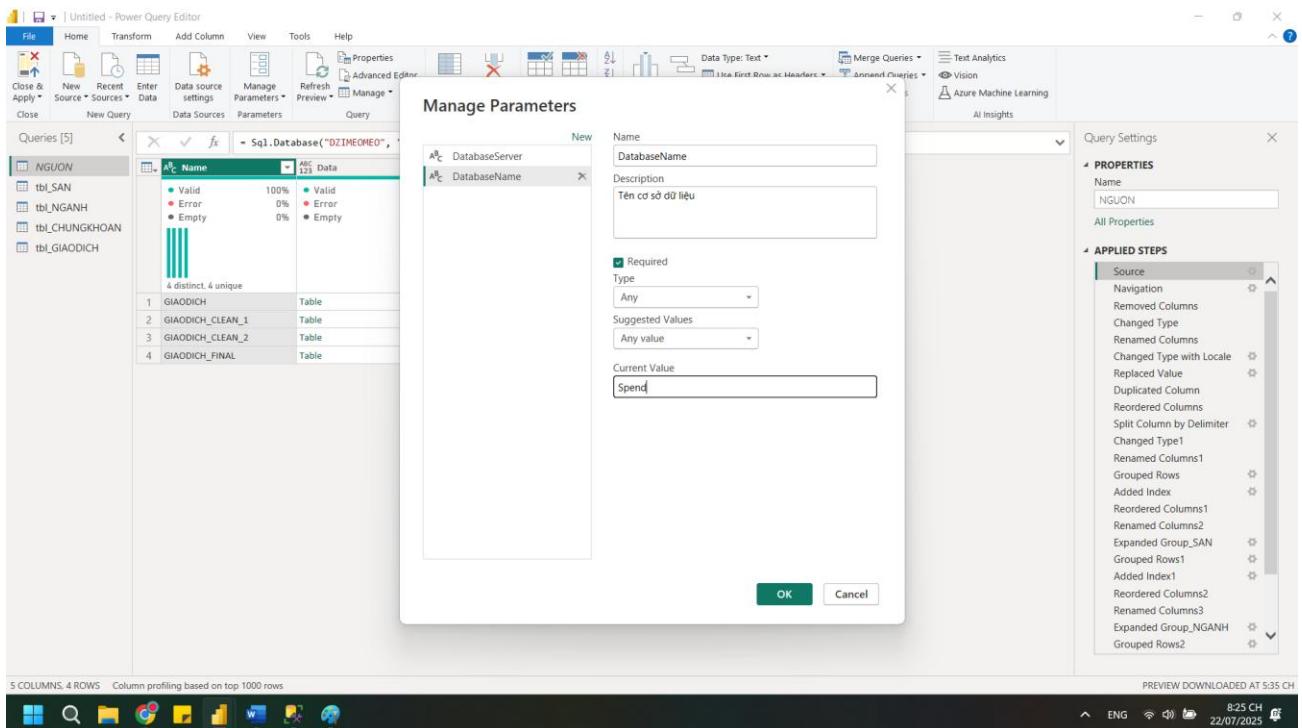
- Thêm new param.

The screenshot shows the Power Query Editor interface. A query named 'NGUON' is selected. The 'Manage Parameters' button is highlighted, and a tooltip 'New Parameter' is shown above it. The main pane displays a table with columns: Name, Data, Schema, Item, and Kind. The table has 5 columns and 4 rows. The status bar at the bottom indicates 'PREVIEW DOWNLOADED AT 5:35 CH'.

- Thêm param DatabaseServer.

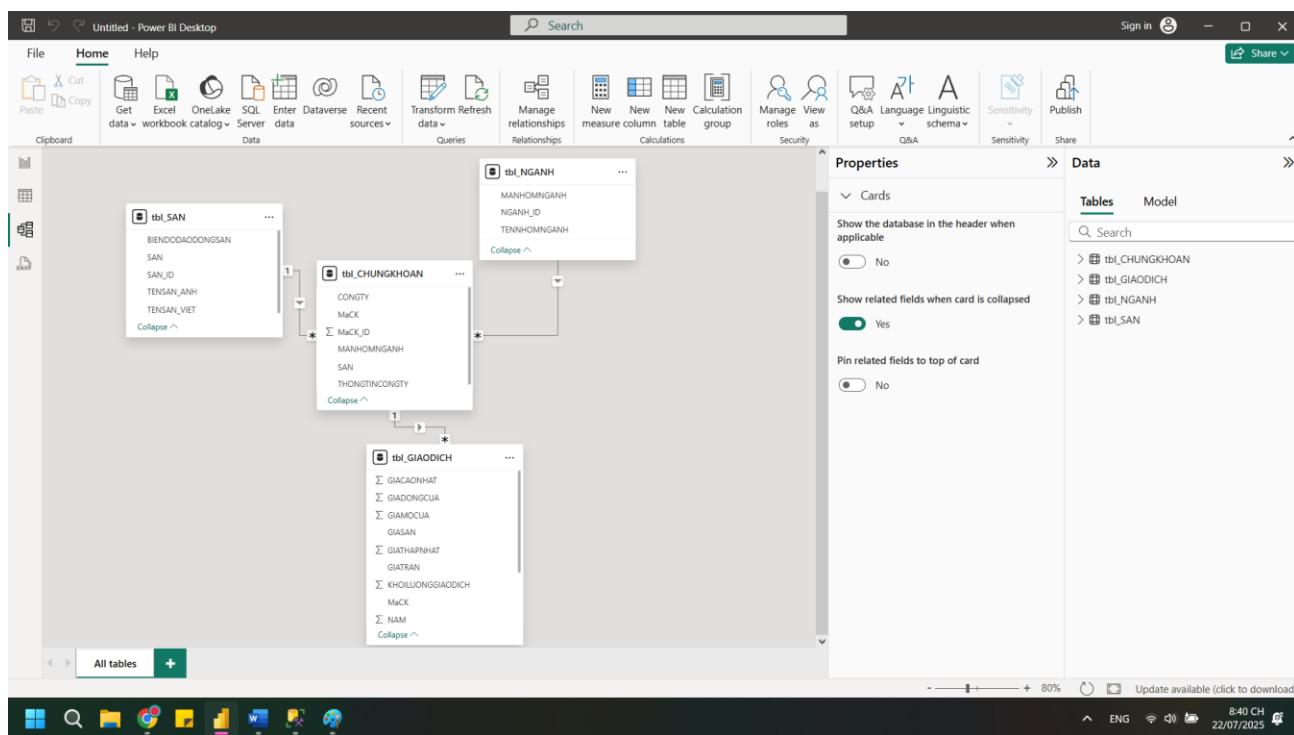
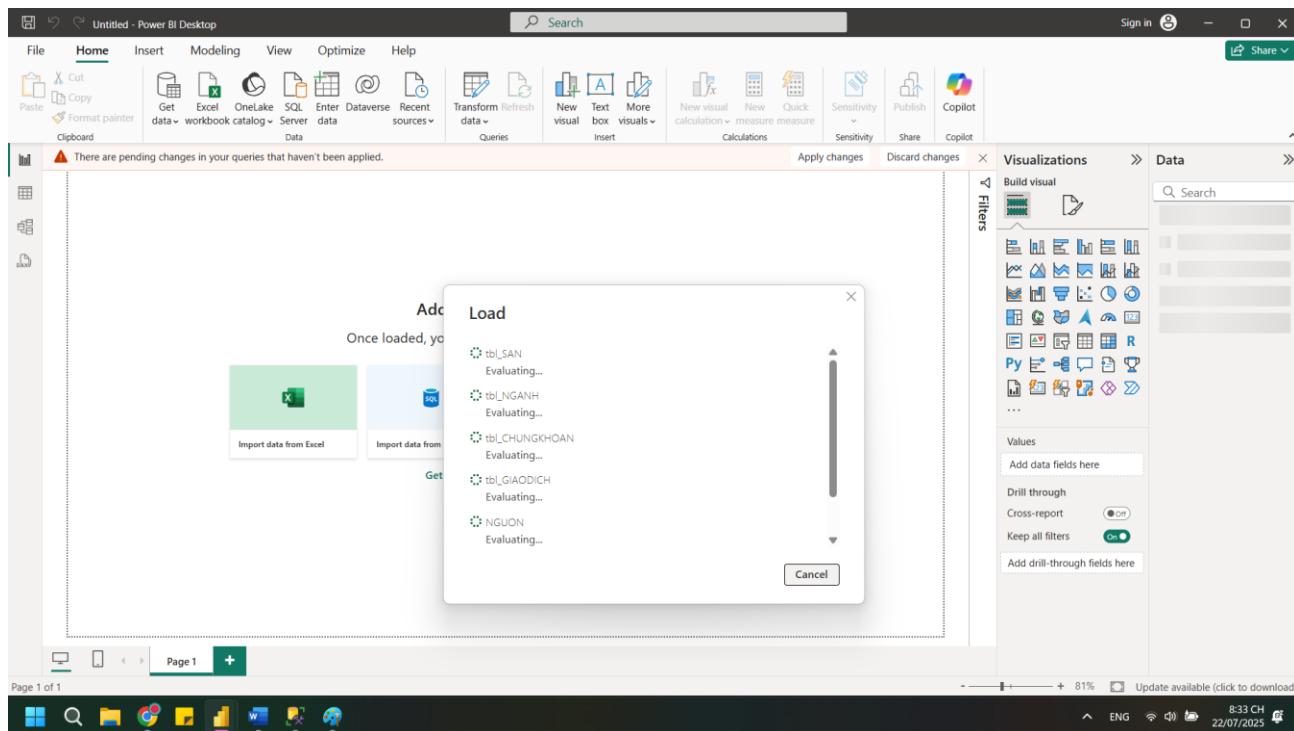


- Thêm param DatabaseName



- Thay connection source bằng hai param mới tạo

- Close và Apply dự án.



## 4.3 XỬ LÝ DỮ LIỆU DAX

- ✓ *Xử lý dữ liệu bằng ngôn ngữ DAX*

- *Measure*
- *Calculated column*
- *Filter*

### 4.3.1 MEASURE

- ✓ *Trình bày các measure được dùng trong dự án.*
- ✓ *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

#### 4.3.1.1 Tạo calendar ngày giao dịch

Cột [DATE] (từ thấp đến cao)

Cột [NAM] là năm giao dịch

Cột [THANG] là tháng giao dịch

Cột [TENTHANG] là tên tháng giao dịch

Cột [NGAYTRONGTUAN] là ngày thứ mấy

Cột [QUY] là quý 1 2 3 4 (một quý là 3 tháng)

Cột [NAMQUY] là định dạng [NAM]/Q[QUY]

Tạo liên kết với bảng tbl\_GIAODICH thông qua [NGAYGIAODICH]

The screenshot shows the Power BI Data Editor interface. In the top ribbon, the 'Table tools' tab is selected. Under the 'Calculated columns' section, a new column named 'Calendar Ngày Giao Dịch' is being defined. The formula is:

```
1 Calendar Ngày Giao Dịch = ADDCOLUMNS(
2     CALENDAR(IIF([tbl_GIAODICH[NGAYGIAODICH]]>MAX(tbl_GIAODICH[NGAYGIAODICH])), 
3     "NAM", YEAR([Date]),
4     "THANG", MONTH([Date]),
5     "TENTHANG", FORMAT([Date], "MMM"),
6     "NGAYTRONGTUAN", FORMAT([Date], "0000"),
7     "QUY", FORMAT([Date], "Q"),
8     "NAMQUY", FORMAT([Date], "YYYY") & "/Q" & FORMAT([Date], "0") )
9 )
```

The resulting table has columns: Date, NAM, THANG, TENTHANG, NGAYTRONGTUAN, QUY, and NAMQUY. The data shows dates from 30/05/2023 to 12/06/2023, with corresponding values for the other columns. The Power BI Data view pane on the right shows the relationships between tables: Calendar Ngày Giao Dịch, tbl\_GIAODICH, tbl\_NGANH, and tbl\_SAN.

#### 4.3.1.2 Tạo số lượng giao dịch trong CountTable

The screenshot shows the Power BI Data Model view. A red box highlights the 'New measure' button in the ribbon toolbar. A red arrow labeled '1' points to the formula bar at the top left, which contains the formula: `1 Số lượng Giao Dịch = COUNTROWS(tbl_GIAODICH)`. Another red arrow labeled '2' points to the 'tbl\_GIAODICH' table in the data model. A red arrow labeled '3' points to the 'CountTable' node in the 'Tables' section of the Properties pane.

#### 4.3.1.3 Tạo số lượng Mã CK trong CountTable

The screenshot shows the Power BI Data Model view. A red box highlights the 'New measure' button in the ribbon toolbar. A red arrow labeled '1' points to the formula bar at the top left, which contains the formula: `1 Số lượng Mã CK = COUNTROWS(tbl_CHUNGKHOAN)`. Another red arrow labeled '2' points to the 'tbl\_CHUNGKHOAN' table in the data model. A red arrow labeled '3' points to the 'Số lượng Mã CK' node in the 'CountTable' node of the 'Tables' section of the Properties pane.

#### 4.3.1.4 Tạo số lượng Ngành trong CountTable

The screenshot shows the Power BI Data Model view. A calculated column named 'Số lượng Ngành' is being created in the 'CountTable' table. The formula is set to `COUNTROWS(tbl1_NGANH)`. The Data Model pane on the right lists various tables and their columns, including 'tbl\_SAN', 'tbl\_CHUNGKHOAN', 'tbl\_GIAODICH', and 'Calendar Ngày Giao Dịch'. The 'CountTable' table is expanded to show its columns: MANHOMNGANH, NGANH\_ID, and TENNHOMNGANH.

#### 4.3.1.5 Tạo số lượng Sàn trong CountTable

The screenshot shows the Power BI Data Model view. A calculated column named 'Số lượng Sàn' is being created in the 'CountTable' table. The formula is set to `COUNTROWS(tbl1_SAN)`. The Data Model pane on the right lists various tables and their columns, including 'tbl\_SAN', 'tbl\_CHUNGKHOAN', 'tbl\_GIAODICH', and 'Calendar Ngày Giao Dịch'. The 'CountTable' table is expanded to show its columns: MANHOMNGANH, NGANH\_ID, and TENNHOMNGANH.

### 4.3.1.6 Tạo tổng KLGD trong Khối lượng Giao Dịch Table

The screenshot shows the Power BI Data Model view. In the ribbon, the 'Data' tab is selected. A red box highlights the 'New measure' button in the 'Calculations' group. In the main area, a DAX formula is being typed into the formula bar: `1 Tổng KLGD = SUM(tbl_GIAODICH[KHOI LUONG GIAO DICH])`. To the right, the 'Properties' pane is open, showing the 'Tables' section with a list of tables including 'tbl\_CHUNGKHOAN', 'tbl\_GIAODICH', 'tbl\_NGANH', and 'tbl\_SAN'. A red box highlights the 'Tổng KLGD' entry in this list.

### 4.3.1.7 Tạo tổng KLGD hôm qua trong Khối lượng Giao Dịch Table

The screenshot shows the Power BI Data Model view. Similar to the previous screenshot, the 'Data' tab is selected in the ribbon, and the 'New measure' button is highlighted. A red box highlights the formula in the formula bar: `1 Tổng KLGD Hôm Qua = CALCULATE([Tổng KLGD], DATEADD('Calendar Ngày Giao Dịch'[Date], -1, DAY))`. The 'Properties' pane is open, showing the 'Home table' set to 'Calendar Ngày Giao Dịch'. Under 'Synonyms', a new synonym is defined with the name '[Tổng KLGD Hôm Qua]'. A red box highlights this new synonym entry. The 'Data' pane on the right shows the table structure with 'Tổng KLGD' and 'Tổng KLGD Hôm Qua' listed under 'Khối Lượng Giao Dịch Table'. A red box highlights the 'Tổng KLGD Hôm Qua' entry in the table list.

#### 4.3.1.8 Tạo tổng KLGD theo Sản trong Khối lượng Giao Dịch Table

The screenshot shows the Power BI Data Model view. A new measure is being created with the formula: `= CALCULATE([Tổng KLGD], ALL(tbl_GIAODICH))`. The formula is highlighted in a red box. The Properties pane on the right shows the measure's name as 'Tổng KLGD theo Sản' and its home table as 'Calendar Ngày Giao Dịch'. The Data pane on the right lists various tables and measures, with 'Tổng KLGD theo Sản' selected.

#### 4.3.1.9 Tạo tổng tiền GTGD trong Khối lượng Giao Dịch Table

The screenshot shows the Power BI Measure tools view. A new measure is being created with the formula: `= SUM(TBL_GIAODICH[THUONGTIEN])`. The formula is highlighted in a red box. The Properties pane on the right shows the measure's name as 'Tổng tiền GTGD' and its data category as 'Uncategorized'. The Data pane on the right lists various tables and measures, with 'Tổng tiền GTGD' selected. A red arrow points from the formula area to the 'Tổng tiền GTGD' entry in the Data pane.

### 4.3.1.10 Tạo tỷ lệ KLGD (%) trong Khối lượng Giao Dịch Table

The screenshot shows the Power BI interface with the following details:

- Title Bar:** DAT109-DP20302-PS46145 • Last saved: Yesterday at 4:53 PM
- File, Home, Help, Table tools, Measure tools** tabs are visible.
- Measure tools ribbon:** New measure, Quick measure, Calculations.
- Data Category:** Uncategorized.
- Table Structure:** A calendar table with columns: Date, NAM, THANG, TENTHANG, NGAYTRONGTUAN, QUY, NAMQUY.
- Measure Definition:** 1 Tỷ lệ KLGD (%) = DIVIDE([Tổng KLGD], [Tổng KLGD theo Sản])
- Properties Panel:** Shows the measure's properties.
- Data View:** Shows the calendar data with a red arrow pointing to the 'Tỷ lệ KLGD (%)' column.
- Data Source:** Calendar Ngày Giao Dịch.
- Bottom Status Bar:** Update available (click to download), ENG, 1000 SA, 01/08/2025.

### 4.3.1.11 Tạo Tỷ lệ KLGD VARiable trong Khối lượng Giao Dịch Table

The screenshot shows the Power BI interface with the following details:

- Title Bar:** DAT109-DP20302-PS46145 • Last saved: Yesterday at 4:53 PM
- File, Home, Help, Table tools, Measure tools** tabs are visible.
- Measure tools ribbon:** New measure, Quick measure, Calculations.
- Data Category:** Uncategorized.
- Table Structure:** A calendar table with columns: Date, NAM, THANG, TENTHANG, NGAYTRONGTUAN, QUY, NAMQUY.
- Variable Definition:**

```

1 Tỷ lệ KLGD VARIABLE =
2 VAR __TongKLGD = SUM(tbl_GIAODICH[KHOI LUONG GIAODICH])
3 VAR __TongKLGD_SAN = CALCULATE(
4     SUM(tbl_GIAODICH[KHOI LUONG GIAODICH]),
5     ALL(tbl_GIAODICH)
6 )
7 RETURN
8 DIVIDE(
9     __TongKLGD,
10    __TongKLGD_SAN
11 )
12 
```
- Properties Panel:** Shows the variable's properties.
- Data View:** Shows the calendar data with a red arrow pointing to the 'Tỷ lệ KLGD VARIABLE' column.
- Data Source:** Calendar Ngày Giao Dịch.
- Bottom Status Bar:** Update available (click to download), ENG, 1004 SA, 01/08/2025.

#### 4.3.1.12 Tạo số lượng mã giảm trong Trạng Thái Giao Dịch Table

The screenshot shows the Microsoft Power BI interface. In the top ribbon, 'Table tools' is selected. The 'Measure tools' tab is active. A new measure is being created with the formula: `1 SỐ LƯỢNG MÃ GIẢM = COUNTROWS(FILTER(tbl_GIAODICH,tbl_GIAODICH[TRANGTHAI] = "GIẢM"))`. The formula bar and the 'Measure tools' ribbon are highlighted with red boxes. A large red arrow points from the formula area towards the data pane on the right. The data pane shows a table named 'Calendar Ngày Giao Dịch' with 14 rows of data. The right pane displays the data model with various tables listed.

#### 4.3.1.13 Tạo số lượng mã tăng trong Trạng Thái Giao Dịch Table

The screenshot shows the Microsoft Power BI interface. Similar to the previous screenshot, it's in 'Table tools' mode with 'Measure tools' selected. A new measure is being created with the formula: `1 SỐ LƯỢNG MÃ TĂNG = COUNTROWS(FILTER(tbl_GIAODICH,tbl_GIAODICH[TRANGTHAI] = "TĂNG"))`. The formula bar and the 'Measure tools' ribbon are highlighted with red boxes. A large red arrow points from the formula area towards the data pane on the right. The data pane shows the same 'Calendar Ngày Giao Dịch' table with 14 rows. The right pane shows the data model, and the 'Số Lượng Mã Tăng' measure is visible in the list of measures.

### 4.3.1.14 Tạo số lượng mã tham chiếu trong Trạng Thái Giao Dịch Table

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. A new measure named 'Số Lượng Mã THAM CHIỀU' is being created, defined by the formula: `COUNTROWS(FILTER(tbl_GIAODICH,tbl_GIAODICH[TRANGTHAI] = "THAM CHIỀU"))`. The 'Calculations' ribbon tab is highlighted with a red box. A red arrow points from the formula bar to the 'Data' pane on the right, which lists the table 'Khối Lượng Giao Dịch Table' and its measures: 'Số Lượng Mã GIẢM', 'Số Lượng Mã TĂNG', and 'Số Lượng Mã THAM CHIỀU'. The status bar at the bottom indicates it's 1009 SA on 01/08/2025.

### 4.3.1.15 Tạo tổng giá cao nhất trong Trạng Thái Giao Dịch Table

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. A new measure named 'Tổng giá cao nhất' is being created, defined by the formula: `SUM(tbl_GIAODICH[GIACAONHAT])`. The 'Calculations' ribbon tab is highlighted with a red box. A red arrow points from the formula bar to the 'Data' pane on the right, which lists the table 'Khối Lượng Giao Dịch Table' and its measures: 'Tổng giá cao nhất'. The status bar at the bottom indicates it's 1010 SA on 01/08/2025.

### 4.3.1.16 Tạo tổng giá thấp nhất trong Trạng Thái Giao Dịch Table

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. In the 'Calculations' section, a new measure named 'Tổng giá thấp nhất' is being defined with the formula `1 Tổng giá thấp nhất = SUM(tbl_GIAODICH[GIATHAPNhat])`. A red box highlights this formula. A red arrow points from the formula area to the 'Data' pane on the right, which lists the measure under 'Tổng giá thấp nhất'. The bottom status bar shows the date as 01/08/2025.

### 4.3.1.17 Tạo tổng giá đóng cửa trong Trạng Thái Giao Dịch Table

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. In the 'Calculations' section, a new measure named 'Tổng giá đóng cửa' is being defined with the formula `1 Tổng giá đóng cửa = SUM(tbl_GIAODICH[GIADONGCUA])`. A red box highlights this formula. A red arrow points from the formula area to the 'Data' pane on the right, which lists the measure under 'Tổng giá đóng cửa'. The bottom status bar shows the date as 01/08/2025.

#### 4.3.1.18 Tạo tổng giá mở cửa trong Trạng Thái Giao Dịch Table

DAT109-DP20302-PS46145 • Last saved: Yesterday at 4:53 PM

File Home Help Table tools Measure tools

Name TỔNG GIÁ MỞ CỦA

Home table Calendar Ngày Giao... ✓

Format General ✓ \$ % ⚡ Auto ✓

Data category Uncategorized

New measure Quick measure

Structure Formattting Properties Calculations

1 Tổng giá mở cửa = sum(tbl\_GIAODICH[GIAMCUA])

Date	NAM	THANG	TENTHANG	NGAYTRONGTUAN	QUY	NAMQUY
30/05/2023 12:00:00 SA	2023	5	May	Tuesday	2	2023/Q2
31/05/2023 12:00:00 SA	2023	5	May	Wednesday	2	2023/Q2
01/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
02/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
03/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
04/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
05/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2
06/06/2023 12:00:00 SA	2023	6	June	Tuesday	2	2023/Q2
07/06/2023 12:00:00 SA	2023	6	June	Wednesday	2	2023/Q2
08/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
09/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
10/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
11/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
12/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2

Data

Search

- ✓ Calendar Ngày Giao Dịch
  - > Bỏ lọc Filter
  - > CountTable
  - > Khối Lượng MÃ GIẢM
    - ✓ Số Lượng MÃ GIẢM
    - ✓ Số Lượng MÃ TĂNG
    - ✓ Số Lượng MÃ THAM CHIẾU
    - ✓ Tổng giá cao nhất
    - ✓ Tổng giá đóng cửa
    - ✓ Tổng giá mở cửa
    - ✓ Tổng giá thấp nhất
  - > tbLCHUNGHOAN
  - > tbLGIADICH
  - > tbLNGANH
  - > tbLSAN

#### **4.3.1.19 Tạo tổng giá sàn trong Trạng Thái Giao Dịch Table**

The screenshot shows the Microsoft Power BI Data view. At the top, there's a ribbon with 'File', 'Home', 'Help', 'Table tools', and 'Measure tools'. The 'Measure tools' tab is selected, with a red box highlighting the 'New measure' button. Below the ribbon is a search bar and a 'Data category' dropdown set to 'Uncategorized'. A table is displayed with columns: Date, NAM, THANG, TENTHANG, NGAYTRONGTUAN, QUY, and NAMQUY. The first row contains a formula: 'Tổng giá sản = sum(tbl1\_GIAODICH[GIASAN])'. A red box highlights this formula. To the right, the 'Data' pane shows a tree view of data sources: 'Calendar Ngày Giao Dịch', 'tbl\_CHUNGKHOAN', 'tbl\_GIAODICH', 'tbl\_LNGANH', and 'tbl\_TSAN'. A red arrow points from the 'New measure' button in the ribbon to the 'Tổng giá sản' node in the Data pane. The bottom status bar shows the table has 14 rows and 0 distinct values for the total column.

### 4.3.1.20 Tạo tổng giá trần trong Trạng Thái Giao Dịch Table

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. A new measure is being created with the formula: `1 Tổng giá trần = SUM(tbl_GIAODICH[GIATRAN])`. The 'Calculations' section is highlighted with a red box. The Data pane on the right shows the table structure and the newly created measure.

### 4.3.1.21 Tạo tỷ lệ giảm trong Trạng Thái Giao Dịch Table

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. A new measure is being created with the formula: `1 Tỷ lệ GIẢM = DIVIDE([Số Lượng MÃ GIẢM],[Số Lượng Giao Dịch])`. The 'Calculations' section is highlighted with a red box. The Data pane on the right shows the table structure and the newly created measure.

### 4.3.1.22 Tạo tỷ lệ tăng trong Trạng Thái Giao Dịch Table

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. A new measure named 'Tỷ lệ TĂNG' is being created using the formula `Tỷ lệ TĂNG = DIVIDE([Số Lượng MÃ TĂNG],[Số Lượng Giao Dịch])`. The 'Calculations' pane on the right shows the measure has been added to the table.

Date	NAM	THANG	TENTHANG	NGAYTRONGTUAN	QUY	NAMQUY
30/05/2023 12:00:00 SA	2023	5	May	Tuesday	2	2023/Q2
31/05/2023 12:00:00 SA	2023	5	May	Wednesday	2	2023/Q2
01/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
02/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
03/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
04/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
05/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2
06/06/2023 12:00:00 SA	2023	6	June	Tuesday	2	2023/Q2
07/06/2023 12:00:00 SA	2023	6	June	Wednesday	2	2023/Q2
08/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
09/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
10/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
11/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
12/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2

### 4.3.1.23 Tạo tỷ lệ tham chiếu trong Trạng Thái Giao Dịch Table

The screenshot shows the Power BI interface with the 'Measure tools' tab selected. A new measure named 'Tỷ lệ THAM CHIỀU' is being created using the formula `Tỷ lệ THAM CHIỀU = DIVIDE([Số Lượng MÃ THAM CHIỀU],[Số Lượng Giao Dịch])`. The 'Calculations' pane on the right shows the measure has been added to the table.

Date	NAM	THANG	TENTHANG	NGAYTRONGTUAN	QUY	NAMQUY
30/05/2023 12:00:00 SA	2023	5	May	Tuesday	2	2023/Q2
31/05/2023 12:00:00 SA	2023	5	May	Wednesday	2	2023/Q2
01/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
02/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
03/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
04/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
05/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2
06/06/2023 12:00:00 SA	2023	6	June	Tuesday	2	2023/Q2
07/06/2023 12:00:00 SA	2023	6	June	Wednesday	2	2023/Q2
08/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
09/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
10/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
11/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
12/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2

## 4.3.2 CALCULATED COLUMN

- ✓ Trình bày các calculated column được dùng trong dự án.
- ✓ Việc này giúp lưu vết các chỉnh sửa thay đổi trên.

### 4.3.2.1 Tạo Cột [TENNTHONNGANH] trong tbl\_CHUNGKHOAN

The screenshot shows the Microsoft Excel interface with the 'tbl\_CHUNGKHOAN' table open. In the 'Column tools' ribbon, the 'New column' button is highlighted with a red box and arrow. The formula bar displays the formula: `1 TENTHONNGANH = RELATED(tbl_NGANH[TENNTHONNGANH])`. The 'Data' pane on the right shows the new column has been successfully added with the name 'TENNTHONNGANH'.

### 4.3.2.2 Tạo Cột [BIENDOGIATRONGNGAY] trong tbl\_GIAODICH

The screenshot shows the Microsoft Excel interface with the 'tbl\_GIAODICH' table open. In the 'Column tools' ribbon, the 'New column' button is highlighted with a red box and arrow. The formula bar displays the formula: `1 BIENDOGIATRONGNGAY = (tbl_GIAODICH[GIACAOHNAT]-tbl_GIAODICH[GIATHAPNHAT])/tbl_GIAODICH[GIAMOCUA]`. The 'Data' pane on the right shows the new column has been successfully added with the name 'BIENDOGIATRONGNGAY'.

### 4.3.2.3 Tạo Cột [GIADIEUCHINHDONGCUA] trong tbl\_GIAODICH

The screenshot shows the Microsoft Power BI Data Editor interface. A red box highlights the formula in the 'New column' dialog:   
`= IF(HASONEVALUE(tbl_GIAODICH[NGAYGIAODICH]),  
 VALUES(tbl_GIAODICH[GIADONGCUA]),  
 CALCULATE(VALUES(tbl_GIAODICH[GIADONGCUA]),  
 LASTDATE(tbl_GIAODICH[NGAYGIAODICH])))`

A red arrow points from the formula area to the 'Data' pane on the right, which lists the table structure. Another red box highlights the 'tbl\_GIAODICH' entry in the tree view.

### 4.3.2.4 Tạo Cột [THANHTIEN] trong tbl\_GIAODICH

- Thành tiền = GIADONGCUA \* KHOILUONGGIAODICH

The screenshot shows the Microsoft Power BI Data Editor interface. A red box highlights the formula in the 'New column' dialog:   
`= tbl_GIAODICH[GIADONGCUA]*tbl_GIAODICH[KHOILUONGGIAODICH]`

A red arrow points from the formula area to the 'Data' pane on the right, which lists the table structure. Another red box highlights the 'tbl\_GIAODICH' entry in the tree view.

### 4.3.2.5 Tạo Cột [TRANGTHAI] trong tbl\_GIAODICH

- Trạng thái: tăng, giảm, tham chiếu, khi thị trường đóng cửa

Screenshot of Microsoft Power BI Data Editor showing the creation of a calculated column 'TRANGTHAI' in the 'tbl\_GIAODICH' table.

**Column Tools Properties:**

- Name: TRANGTHAI
- Data type: Text
- Format: Text
- Summarization: Don't summarize
- Data category: Uncategorized
- Sort by column: Sort
- Data groups: Groups
- Manage relationships: Relationships
- New column (highlighted with a red box)

**Calculated Column Formula:**

$$1 \text{ TRANGTHAI} = \text{IF}(\text{tbl\_GIAODICH}[\text{GIAMOCUA}] > \text{tbl\_GIAODICH}[\text{GIADONGCUA}], \text{"GIÁM"}, \\ \text{IF}(\text{tbl\_GIAODICH}[\text{GIAMOCUA}] = \text{tbl\_GIAODICH}[\text{GIADONGCUA}], \text{"THÀM CHIỀU"}, \text{TĂNG}))$$

**Table Data:**

GIATHAPNHAT	GIADONGCUA	KHOIQUONGGIAODICH	GATRAN	GIASAN	TILETRONGNGAY	BIENDOIGATRONGNGAY	GIADIEUCHINHDONGCUA	THANHTIEN	TRANGTHAI
9	43,8	43,8	100	46,87	40,73	0,00	0	43,8	4380 THAM CHIEU
6	33,6	33,6	100	38,64	28,56	0,00	0	33,6	3360 THAM CHIEU
8	11,8	11,8	100	13,57	10,03	0,00	0	11,8	1180 THAM CHIEU
2	5,2	5,2	100	5,98	4,42	0,00	0	5,2	520 THAM CHIEU
3	16,3	16,3	100	18,75	13,86	0,00	0	16,3	1630 THAM CHIEU
5	15,5	15,5	100	17,83	13,18	0,00	0	15,5	1550 THAM CHIEU
6	16,6	16,6	100	19,09	14,11	0,00	0	16,6	1660 THAM CHIEU
5	43,6	43,6	100	46,65	40,55	0,00	0	43,6	4360 THAM CHIEU
7	20,7	20,7	100	23,81	17,60	0,00	0	20,7	2070 THAM CHIEU
3	23	23	100	26,45	19,55	0,00	0	23	2300 THAM CHIEU
8	4,8	4,8	100	5,52	4,08	0,00	0	4,8	480 THAM CHIEU
6	6,6	6,6	100	7,59	5,61	0,00	0	6,6	660 THAM CHIEU
9	29	29	100	31,03	26,97	0,00	0	29	2900 THAM CHIEU
5	55	55	100	60,50	49,50	0,00	0	55	5500 THAM CHIEU
4	15,4	15,4	100	16,94	13,86	0,00	0	15,4	1540 THAM CHIEU
3	19,3	19,3	100	22,20	16,41	0,00	0	19,3	1930 THAM CHIEU
2	15,2	15,2	100	17,48	12,92	0,00	0	15,2	1520 THAM CHIEU
4	9,4	9,4	100	10,81	7,99	0,00	0	9,4	940 THAM CHIEU
5	16,5	16,5	100	18,98	14,03	0,00	0	16,5	1650 THAM CHIEU
9	13,9	13,9	100	15,99	11,82	0,00	0	13,9	1390 THAM CHIEU
0	760	760	100	874,00	646,00	0,00	0	760	76000 THAM CHIEU
9	9,99	9,99	100	10,69	9,29	0,00	0	9,99	999 THAM CHIEU
4	34	34	100	36,38	31,62	0,00	0	34	3400 THAM CHIEU
3	11,3	11,3	100	12,43	10,17	0,00	0	11,3	1130 THAM CHIEU
4	13,4	13,4	100	14,74	12,06	0,00	0	13,4	1340 THAM CHIEU
1	12,1	12,1	100	13,31	10,89	0,00	0	12,1	1210 THAM CHIEU

Table: tbl\_GIAODICH (10,915 rows) Column: TRANGTHAI (3 distinct values)

Update available (click to download)

#### 4.3.2.6 Cột [SoLuongMaCK] trong tbl\_GIAODICH

Screenshot of Microsoft Power BI Data Editor showing the creation of a calculated column 'SoLuongMaCK' in the 'tbl\_GIAODICH' table.

**Column Tools Properties:**

- Name: SoLuongMaCK
- Data type: Whole number
- Format: Whole number
- Summarization: Don't summarize
- Data category: Uncategorized
- Sort by column: Sort
- Data groups: Groups
- Manage relationships: Relationships
- New column (highlighted with a red box)

**Calculated Column Formula:**

$$1 \text{ SoLuongMaCK} = \text{COUNTRONS}(\text{RELATEDTABLE}(tbl\_CHUNGKHOAN))$$

**Table Data:**

NGANH_ID	MANHOMNGANH	TENNHHOMNGANH	SoLuongMaCK
1	DAUTUCONG	Đầu tư công	78
2	KHAIKHOANG	Thịn Khoáng sản	74
3	XAYDUNG	Xây dựng	324
4	CONGNGH	Công nghệ	76
5	BANL	Bán lẻ	113
6	CK	Chứng khoán	40
7	CHUNGQUYEN	Chứng quyền	74
8	DAUKHI	Dầu khí	64
9	DIEN	Điện	76
10	DETMAXMAC	Đèt may mặc	39
11	GTVT	Giao thông Vận tải	114
12	NGANHANG	Ngân Hàng	31
13	CAOSU	Cao su	13
14	THEP	Thép	39
15	THUYHAI SAN	Thủy hải sản	24
16	TTGT	Truyền thông giải trí	19
17	THUCPHAM	Thực phẩm	92
18	YTE	Y tế	56
19	HANGKHONG	Hàng không	7
20	BAOHIEM	Bảo hiểm	17
21	DICHVU	Dịch vụ	55
22	BCVT	Bưu chính viễn thông	21

Table: tbl\_NGANH (22 rows) Column: SoLuongMaCK (19 distinct values)

Update available (click to download)

#### 4.3.2.7 Cột [TongKhoiLuongGiaoDich] trong tbl\_GIAODICH

Table: tbl\_NGANH (22 rows) Column: TongKhoiLuongGiaoDich (22 distinct values)

Update available (click to download)

ENG 9:55 SA 01/08/2025

#### 4.3.3 FILTER

- ✓ Trình bày các bộ lọc filter được dùng trong dự án.
- ✓ Việc này giúp lưu vết các chỉnh sửa thay đổi trên.

##### 4.3.3.1 Tạo filter chọn Mã CK

Table: Calendar Ngày Giao Dịch (14 rows) Column: Chon Ma CK (0 distinct values)

Update available (click to download)

ENG 10:38 SA 01/08/2025

### 4.3.3.2 Tạo filter chọn Ngành

The screenshot shows the Power BI interface with the following details:

- Title:** DAT109-DP20302-PS46145 • Last saved: Today at 10:14 AM
- File menu:** File, Home, Help, Table tools, Measure tools (selected), Data.
- Formatting ribbon:** Name: Chọn Ngành, Format: Text, Data category: Uncategorized.
- Properties ribbon:** New measure, Quick measure, Calculations.
- Table structure:**

Date	NAM	THANG	TENTHANG	NGAYTRONGTUAN	QUY	NAMQUY
30/05/2023 12:00:00 SA	2023	5	May	Tuesday	2	2023/Q2
31/05/2023 12:00:00 SA	2023	5	May	Wednesday	2	2023/Q2
01/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
02/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
03/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
04/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
05/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2
06/06/2023 12:00:00 SA	2023	6	June	Tuesday	2	2023/Q2
07/06/2023 12:00:00 SA	2023	6	June	Wednesday	2	2023/Q2
08/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
09/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
10/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
11/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
12/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2
- Filter bar:** I Chọn Ngành = IF(ISFILTERED(tbl\_CHUNGKHOAN[TENHINHOMINGAIGH]), SELECTEDVALUE(tbl\_CHUNGKHOAN[TENHINHOMINGAIGH]), "Chọn Ngành")
- Data pane:** Shows the hierarchy of tables: Calendar Ngày Giao Dịch, 0.Fx, Bỏ Lọc Filter, Chọn Mã CK, Chọn Ngành, CountTable, Khối Lượng Giao Dịch Table, Trang Thái Giao Dịch Table, tbl\_CHUNGKHOAN, tbl\_GIAODICH, tbl\_NGANH, and tbl\_SAN.
- System status:** Update available (click to download), ENG, 10:42 SA, 01/08/2025.

### 4.3.3.3 Tạo filter chọn Ngày Giao Dịch

The screenshot shows the Power BI interface with the following details:

- Title:** DAT109-DP20302-PS46145 • Last saved: Today at 10:14 AM
- File menu:** File, Home, Help, Table tools, Measure tools (selected), Data.
- Formatting ribbon:** Name: Chọn Ngày Giao D... (partially visible), Format: General, Data category: Uncategorized.
- Properties ribbon:** New measure, Quick measure, Calculations.
- Table structure:**

Date	NAM	THANG	TENTHANG	NGAYTRONGTUAN	QUY	NAMQUY
30/05/2023 12:00:00 SA	2023	5	May	Tuesday	2	2023/Q2
31/05/2023 12:00:00 SA	2023	5	May	Wednesday	2	2023/Q2
01/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
02/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
03/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
04/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
05/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2
06/06/2023 12:00:00 SA	2023	6	June	Tuesday	2	2023/Q2
07/06/2023 12:00:00 SA	2023	6	June	Wednesday	2	2023/Q2
08/06/2023 12:00:00 SA	2023	6	June	Thursday	2	2023/Q2
09/06/2023 12:00:00 SA	2023	6	June	Friday	2	2023/Q2
10/06/2023 12:00:00 SA	2023	6	June	Saturday	2	2023/Q2
11/06/2023 12:00:00 SA	2023	6	June	Sunday	2	2023/Q2
12/06/2023 12:00:00 SA	2023	6	June	Monday	2	2023/Q2
- Filter bar:** I Chọn Ngày Giao Dịch = IF(ISFILTERED(tbl\_GIAODICH[NGAYGIAODICH]), SELECTEDVALUE(tbl\_GIAODICH[NGAYGIAODICH]), "Chọn Ngày Giao Dịch")
- Data pane:** Shows the hierarchy of tables: Calendar Ngày Giao Dịch, 0.Fx, Bỏ Lọc Filter, Chọn Mã CK, Chọn Ngành, Chọn Ngày Giao Dịch, CountTable, Khối Lượng Giao Dịch Table, Trang Thái Giao Dịch Table, tbl\_CHUNGKHOAN, tbl\_GIAODICH, tbl\_NGANH, and tbl\_SAN.
- System status:** Update available (click to download), ENG, 10:46 SA, 01/08/2025.

# 5 TRỰC QUAN HÓA DỮ LIỆU

## 5.1 CÁC KỸ THUẬT TRỰC QUAN HÓA

- ✓ *Trình bày các kỹ thuật trực quan hóa*
- ✓ *Kỹ thuật nào đang được áp dụng cho dự án ? Vì sao ?*

### ◻ *Temporal data visualization (Trực quan hóa dữ liệu chuỗi thời gian)*

- ❖ Được sử dụng để trình bày các đối tượng một chiều tuyến tính như đồ thị đường, biểu đồ đường hoặc dòng thời gian.
- ❖ Ví dụ: bạn có thể sử dụng biểu đồ đường để biểu thị các thay đổi xảy ra liên tục trong một khoảng thời gian nhất định.

### ◻ *Hierarchical data visualization (Trực quan hóa dữ liệu phân cấp)*

- ❖ Đề cập đến một nhóm hoặc tập hợp các mục có liên kết chung với một mục cha. Bạn có thể sử dụng những cây dữ liệu này để hiển thị các cụm thông tin.
- ❖ Ví dụ: bạn có thể biểu thị lượng dữ liệu về hàng tồn kho dưới dạng cây, trong đó có nút cha (quần áo) và các nút con (áo sơ mi, quần dài và tất).

### ◻ *Network data visualization (Trực quan hóa dữ liệu mạng)*

- ❖ Biểu đồ phân tán trình bày dữ liệu dưới dạng điểm trên một đồ thị.
- ❖ Biểu đồ bong bóng thêm một yếu tố dữ liệu thứ ba vào biểu đồ phân tán.
- ❖ Đám mây trình bày tần suất từ bằng cách sử dụng từ có các kích cỡ khác nhau.

### ◻ *Multidimensional data visualization (Trực quan hóa dữ liệu đa chiều)*

- ❖ Biểu đồ cột, biểu đồ tròn và đồ thị cột là những ví dụ phổ biến về các loại trực quan hóa này.
- ❖ Ví dụ: biểu đồ cột so sánh hai hoặc nhiều yếu tố dữ liệu và cho thấy các thay đổi của một biến trong một khoảng thời gian.
- ❖ Biểu đồ tròn trực quan hóa các phần của tổng thể thuộc từng danh mục.

## ◻ Geospatial data visualization (*Trực quan hóa dữ liệu không gian địa lý*)

- ❖ Bản đồ nhiệt, bản đồ mật độ hoặc bản đồ địa hình, trình bày dữ liệu liên quan đến các vị trí ngoài đời thực.
- ❖ Ví dụ: trực quan hóa dữ liệu biểu thị số lượng khách hàng ghé thăm các chi nhánh bán lẻ khác nhau.

## 5.2 CÁC NGUYÊN TẮC TRỰC QUAN HÓA

- ✓ *Trình bày các nguyên tắc trong trực quan hóa*
- ✓ *Nguyên tắc nào bạn cho rằng quan trọng nhất trong phạm vi dự án này? Vì sao?*

### ◻ #1 Chọn đúng loại biểu đồ

- ❖ Điều này có vẻ hiển nhiên, nhưng thực tế vẫn nhiều người chọn biểu đồ theo thói quen hoặc cảm tính.
- ❖ “form follows function” (chức năng đi trước, trình bày đi sau). Phải tự hỏi mục đích của bạn là gì? Bạn muốn sánh các giá trị gì, mối liên hệ giữa các biến nào?

### ◻ #2 Không phải tất cả data đều quan trọng

- ❖ Không nên ném toàn bộ dataset vào một biểu đồ, hãy phân tích trước để biểu diễn/ highlight những thông tin quan trọng sao cho người xem dễ nắm bắt..

### ◻ #3 Biểu đồ thể hiện đúng tương quan số liệu thực tế

- ❖ Việc đồ thị hóa dữ liệu nên giúp thể hiện sự tương quan số liệu đúng nhất.
- ❖ Không nên cố tình lợi dụng các yếu tố về khung tham chiếu, tỷ lệ hình ảnh để gây nhầm lẫn cho người xem.

### ◻ #4 Sử dụng màu sắc hợp lý khi chuyển dữ liệu sang dạng biểu đồ

❖ Việc sử dụng màu sắc mang mục đích dễ nhìn và nhấn mạnh những yếu tố quan trọng trong biểu đồ.

❖ Tuy nhiên, cần chọn lựa màu sắc thông minh, không quá thùa thiê, không gây rối và mất tập trung.

#### ◻ #5 Luôn đảm bảo dữ liệu được trình bày một cách đơn giản và hiệu quả

❖ Các công cụ trực quan hóa dữ liệu có thể giúp chúng ta tạo ra những biểu đồ đẹp mắt và phức tạp với chỉ vài cái click chuột.

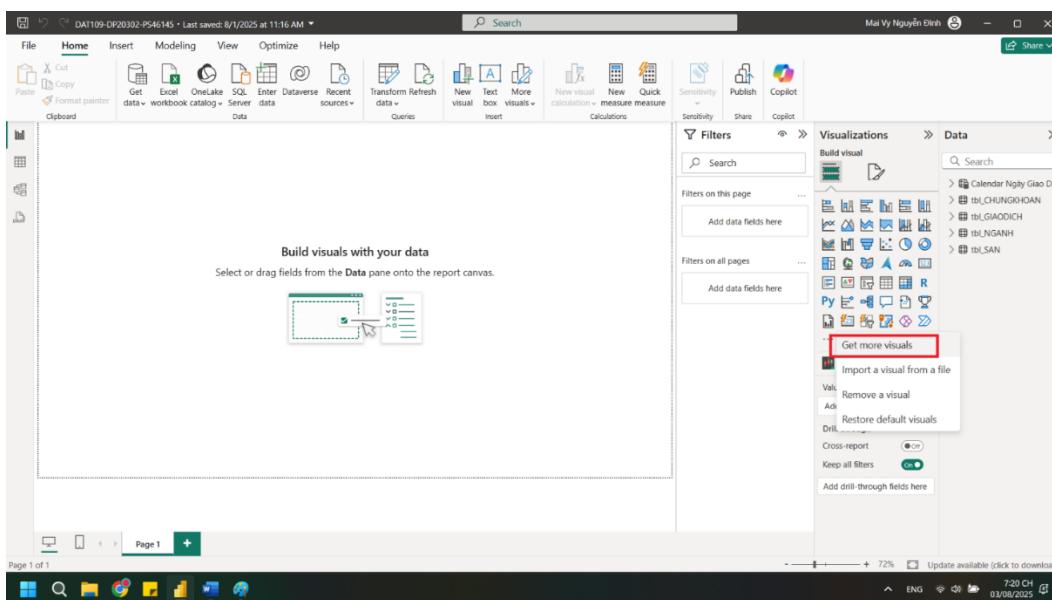
❖ Tuy nhiên, quá nhiều yếu tố về thẩm mỹ có thể làm người xem sao nhãng khỏi thông điệp chính. Hãy lược bỏ những yếu tố rườm rà không cần thiết.

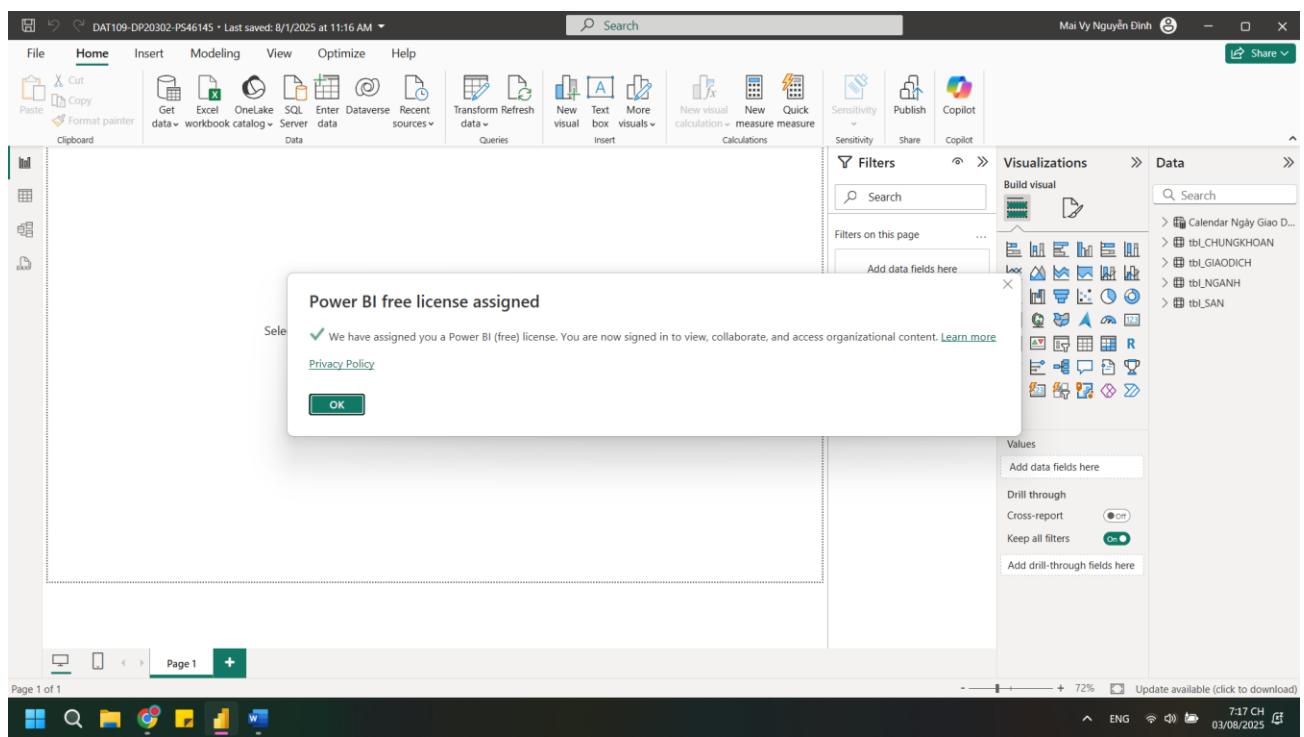
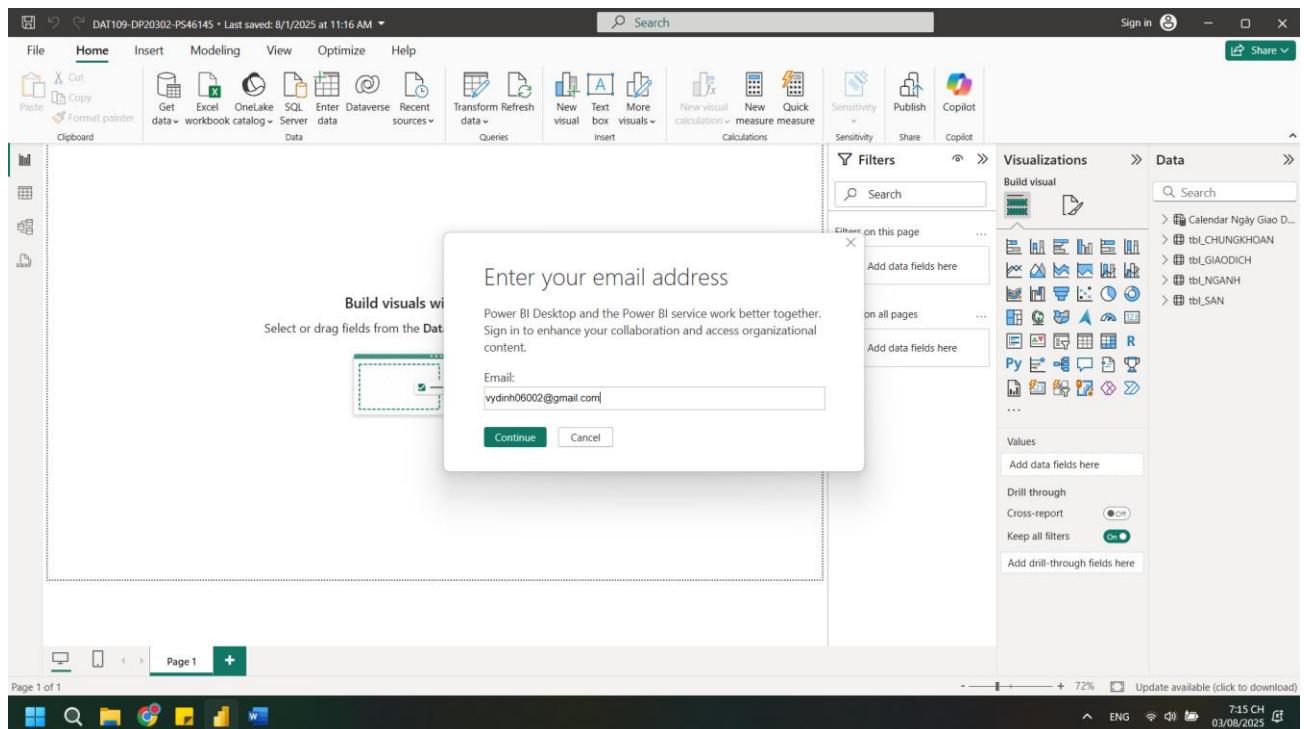
### 5.3 TRÌNH BÀY CÁCH THÊM VISUAL MỚI

- ✓ Các visual trong Power BI đôi khi không đáp ứng được nhu cầu nghiệp vụ mới.
- ✓ Thêm visual mới là điều tất yếu. Trình bày các bước thêm visual (Candlestick chart) trong dự án này.
- ✓ Khuyến khích tìm visual mới phục vụ cho business của bộ dữ liệu.

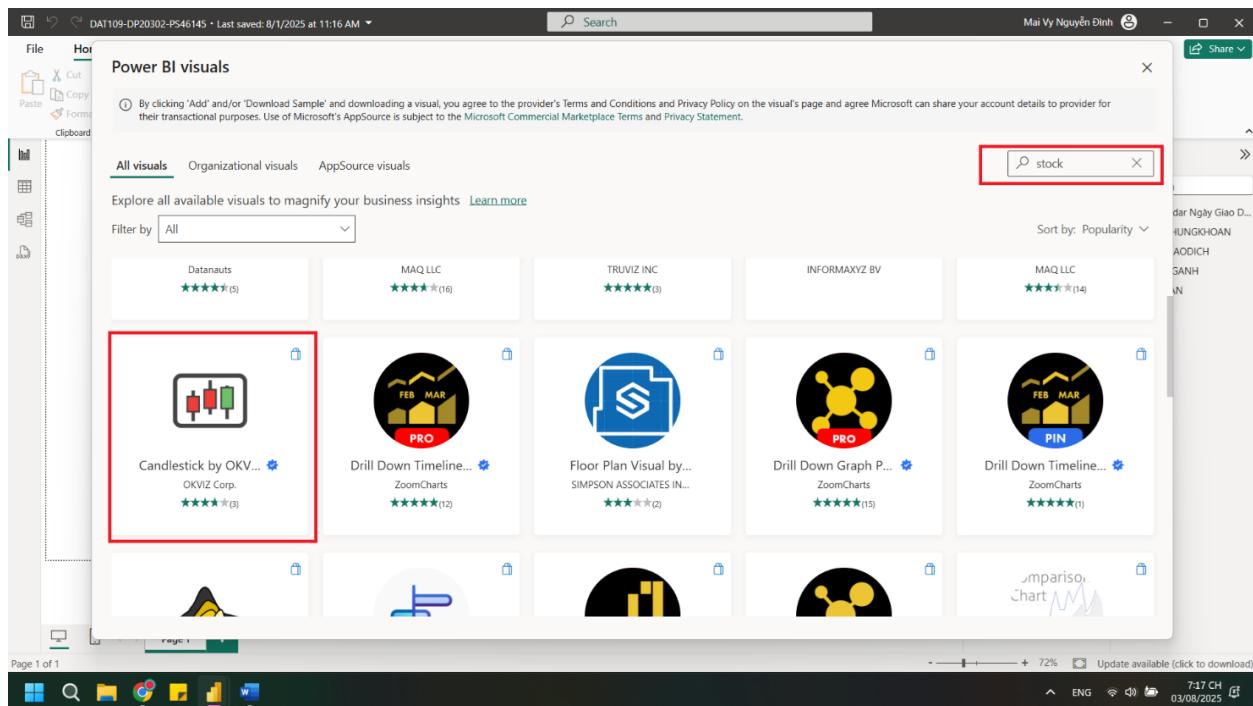
#### ◻ Các bước thêm một visual mới (Candlestick chart)

❖ Mở project. Đăng nhập bằng email FPT. Chọn Get more visuals.

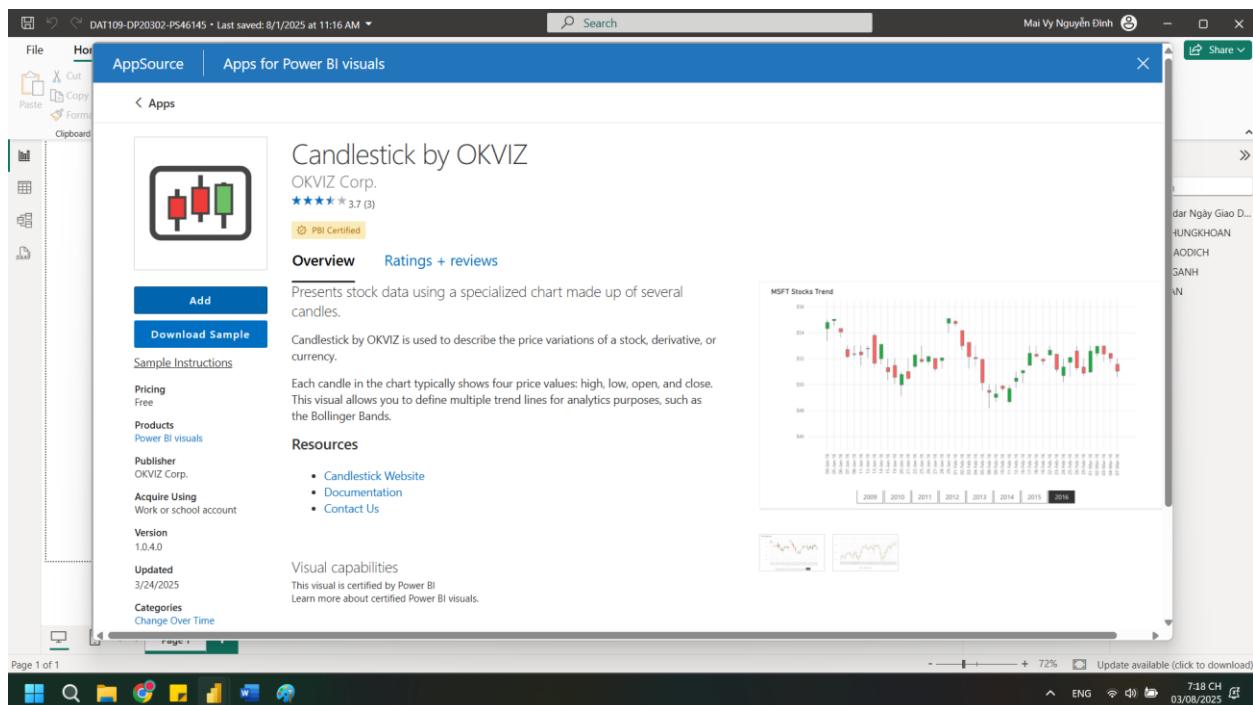




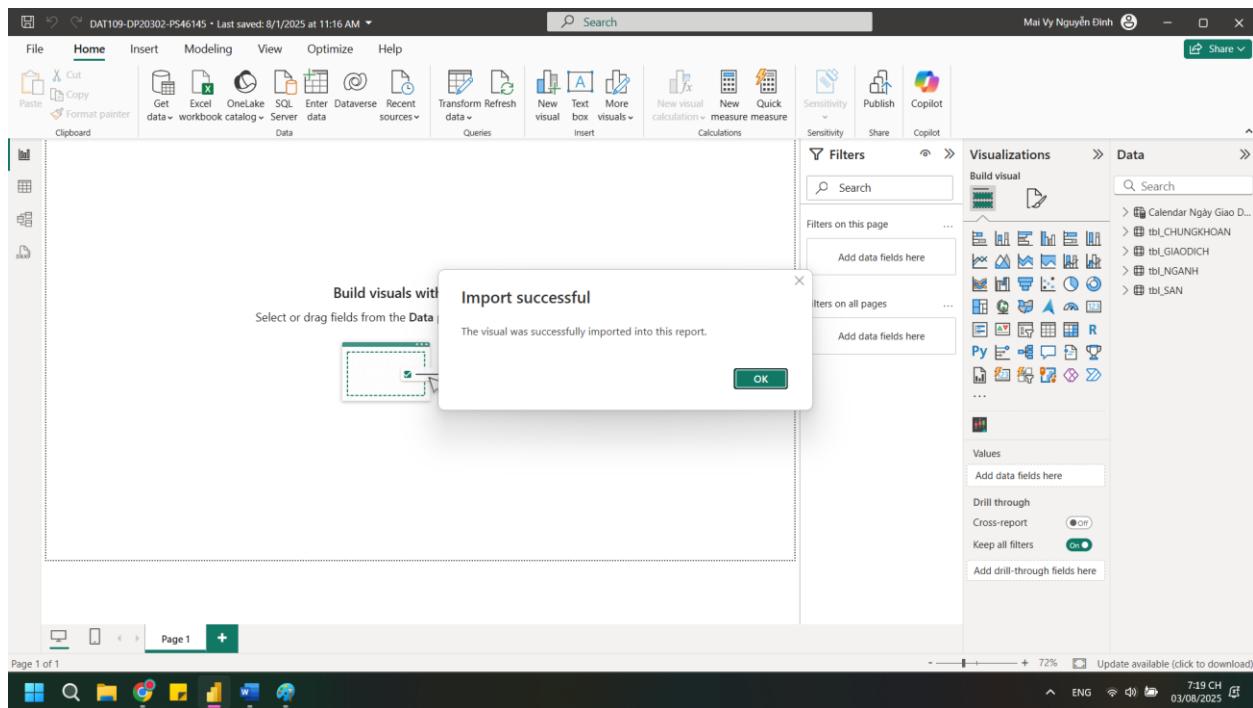
❖ Tìm kiếm từ khóa stock. Chọn Candlestick by OKViz (nên chọn visual có tick xanh).



❖ Chọn Add. Chú ý xem Price, Acquire Using, Industries.



❖ Visual đã được thêm vào thành công.



❖ Khuyến nghị download sample của visual để biết cách sử dụng.

## 5.4 TRÌNH BÀY TẠO CÁC REPORT CHO DỰ ÁN

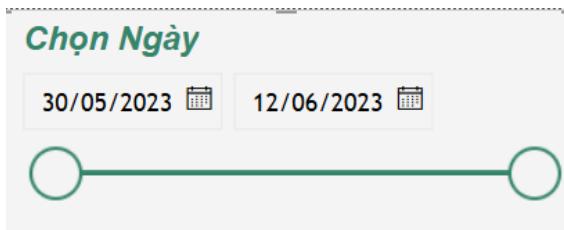
- ✓ *Tạo các bộ lọc, biểu đồ theo nhu cầu dự án.*
- ✓ *Có thể phân chia theo đối tượng trong mô hình.*
- ✓ *Có thể phân chia theo chức năng hoặc page.*
- ✓ *Trong dự án cho môn học này thì phân chia theo page.*

### 5.4.1 TẠO VISUAL THÔNG KÊ CHI TIẾT

- ✓ *Tạo các visual dùng trong trang thống kê chi tiết.*
- ✓ *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

#### 5.4.1.1 Tạo visual filter theo ngày giao dịch

- Giao diện



- Thiết đặt

The screenshot shows the Microsoft Power BI Data Editor interface. On the left, there's a data grid with a column labeled 'Date' containing various dates from 04/06/2023 to 12/06/2023. A red arrow points from the first few items in this column to the 'Filters' pane on the right. The 'Filters' pane has a tree view where 'Date' is selected under 'Calendar Ngày Giao Dịch'. Other collapsed categories include 'Bộ Lọc Filter', 'Color Table', 'CountTable', 'Khối Lượng Giao Dịch Table', 'Trạng Thái Giao Dịch Table', and several tables like 'tbl\_CHUNGKHOAN', 'tbl\_GIAODICH', etc.

#### 5.4.1.2 Tạo visual filter theo ngành nghề

- Giao diện:

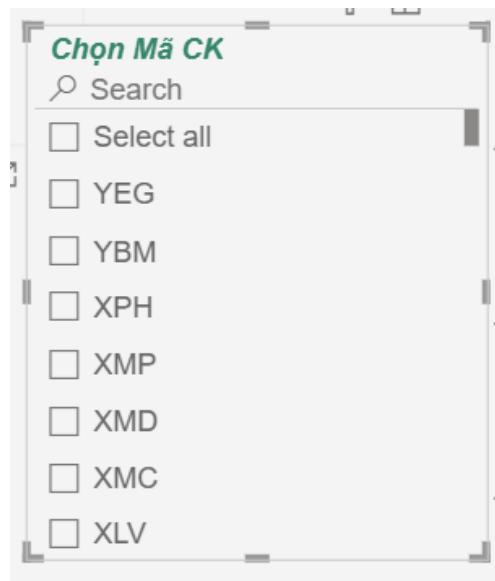
The screenshot shows a modal dialog box titled 'Chọn Ngành' (Select Industry). It includes a search bar at the top and a list of industry names below. Each name is preceded by a checkbox that is currently unchecked. The industries listed are: Bán lẻ, Bảo hiểm, Bưu chính viễn thông, Cao su, Chứng Khoán, Chứng quyền, Công nghệ, and Dầu khí.

- Thiết đặt:

The screenshot shows the Microsoft Power BI desktop interface. On the left, there is a visual titled "Chọn Ngày" with a date range selector from 30/05/2023 to 12/06/2023. To the right of the visual is the "Data" pane. The "Data" pane has three numbered callouts: 1 points to the "Visualizations" section where a "Slicer" icon is highlighted; 2 points to the "tbl\_CHUNGKHOAN" table in the data source list where the checkbox for "TENNTHONGANH" is checked; and 3 points to the "Field" dropdown menu which also contains the checked "TENNTHONGANH" field.

#### 5.4.1.3 Tạo visual filter theo mã chứng khoán

- Giao diện:



- Thiết đặt:

#### 5.4.1.4 Tạo thống kê số lượng theo sàn

- Giao diện:

Sàn	Số lượng Mã CK	Số lượng Giao Dịch
Upcom	574	3594
HOSE	458	4193
HNX	372	2816
OTC	42	312
<b>Total</b>	<b>1446</b>	<b>10915</b>

- Thiết đặt:

### 5.4.1.5 Tạo thống kê tổng giá trị giao dịch

- Giao diện:



- Thiết đặt:

The screenshot shows the Power BI desktop interface with the following details:

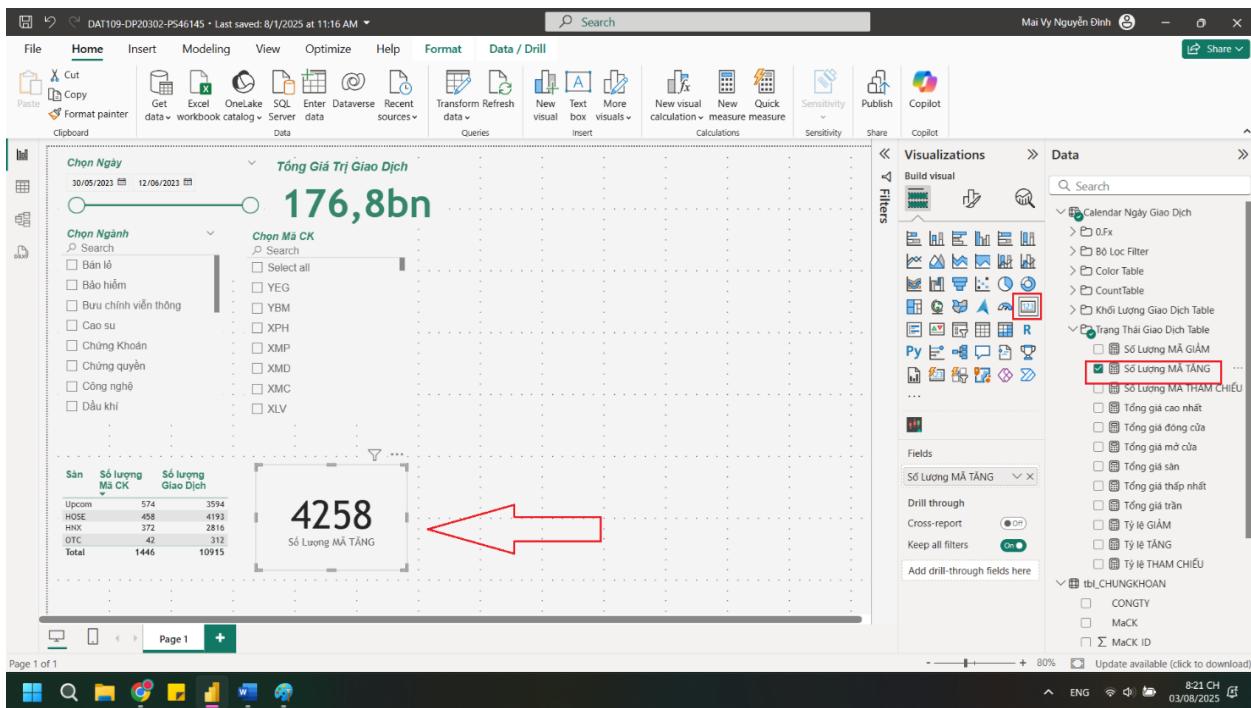
- File, Home, Insert, Modeling, View, Optimize, Help, Format, Data / Drill** tabs are visible at the top.
- Clipboard** pane on the left.
- Chọn Ngành** and **Chọn Mã CK** dropdown filters on the left.
- Data** pane on the right showing fields like 'Tổng tiền GTGD' (highlighted with a red box) and 'Tỷ lệ KLGD (%)'.
- A table visualization showing transaction data with columns 'Sàn', 'Số lượng Mã CK', 'Số lượng Giao Dịch', and a total row.
- A large green text box labeled '176,82bn' with the subtitle 'Tổng tiền GTGD' below it.
- An arrow points from the 'Tổng tiền GTGD' field in the Data pane to the green text box.

### 5.4.1.6 Tạo thống kê số mã tăng

- Giao diện:



- Thiết đặt:



### 5.4.1.7 Tạo thống kê số mã giảm

- Giao diện:

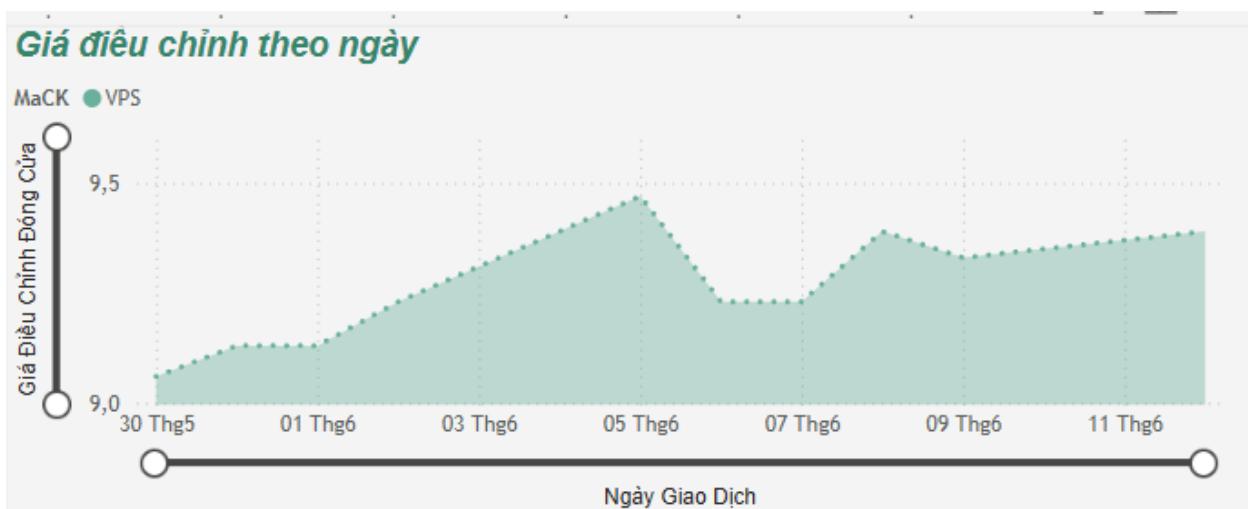


- Thiết đặt:

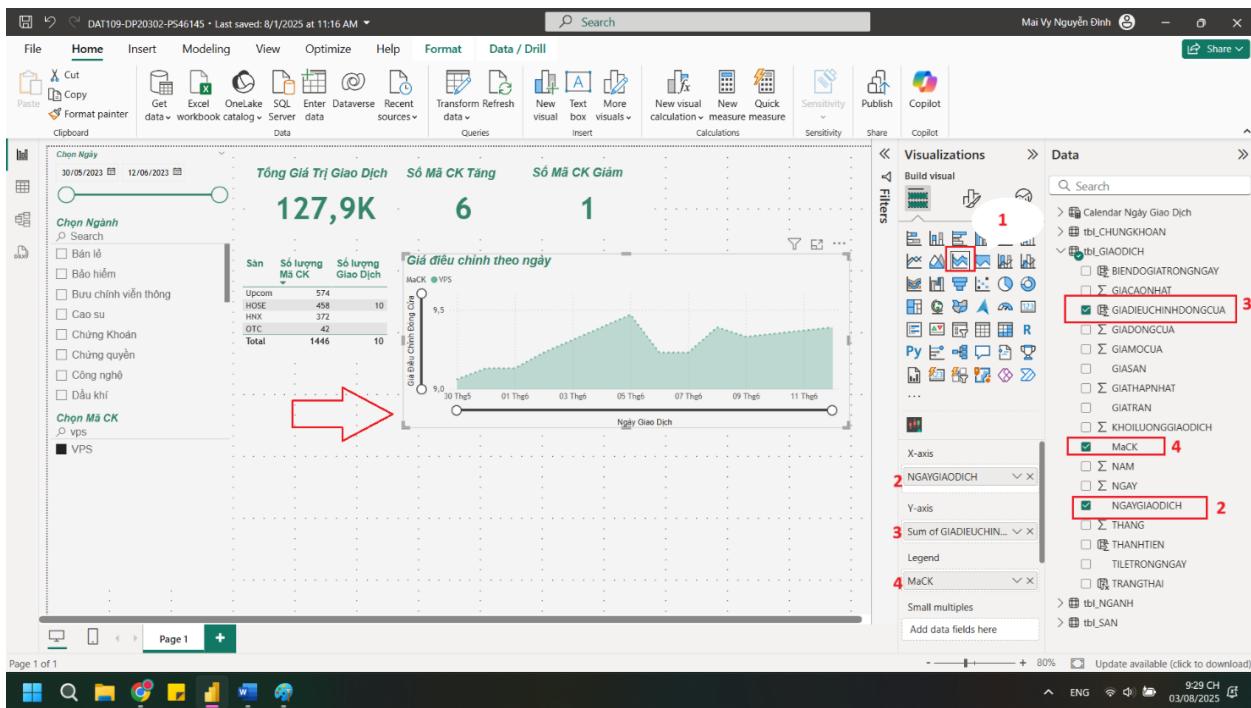
The screenshot shows a Power BI desktop interface with a dashboard containing several visualizations. On the left, there are two filter panes: 'Chọn Ngành' and 'Chọn Mã CK'. The main area features a large card visualization with the text 'Tổng Giá Trị Giao Dịch' (Total Transaction Value) and 'Số Mã CK Tăng' (Number of Increasing Stocks). Below these are two numerical values: '176,8bn' and '4258'. To the right of these is another card visualization with the text 'Số Mã CK Giảm' (Number of Decreasing Stocks) and the value '3601'. A red arrow points from the text 'Giá Điều chỉnh' in the question to this '3601' value. The dashboard also includes a table visualization showing stock information across different markets like Lipcom, HNGE, HNX, and OTC. The bottom right corner of the screen shows the Windows taskbar with icons for File Explorer, Google Chrome, and others.

### 5.4.1.8 Tạo biểu đồ giá điều chỉnh đóng cửa theo ngày

- Giao diện:



- Thiết đặt:

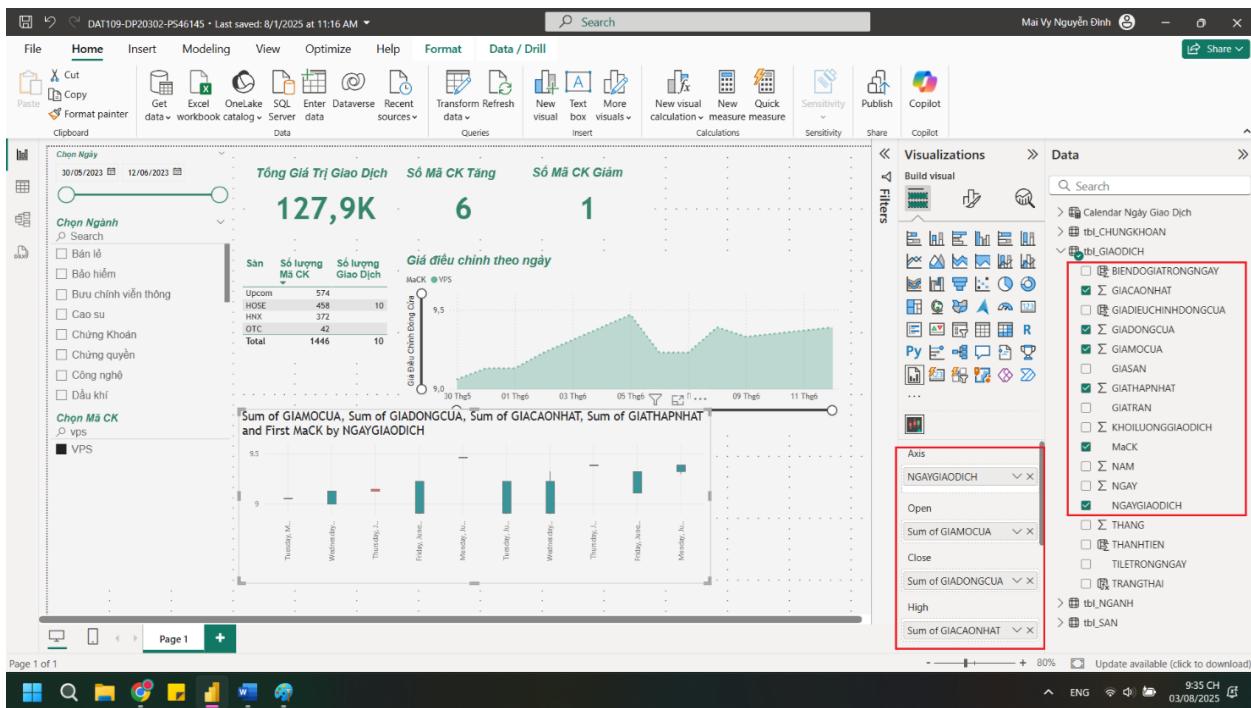


### 5.4.1.9 Tạo biểu đồ biên độ điều chỉnh giá theo ngày.

- Giao diện:

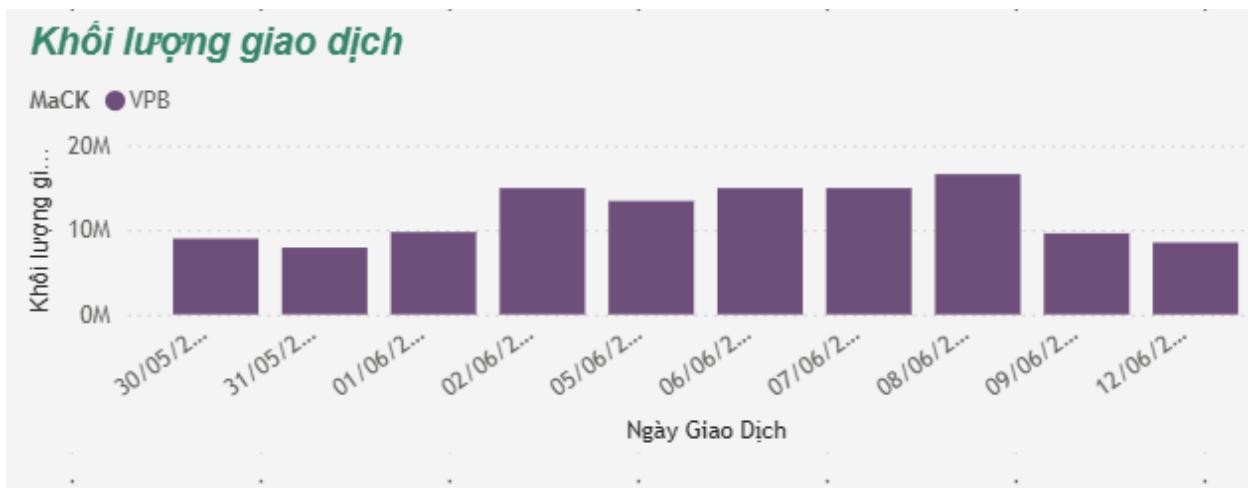


- Thiết đặt:

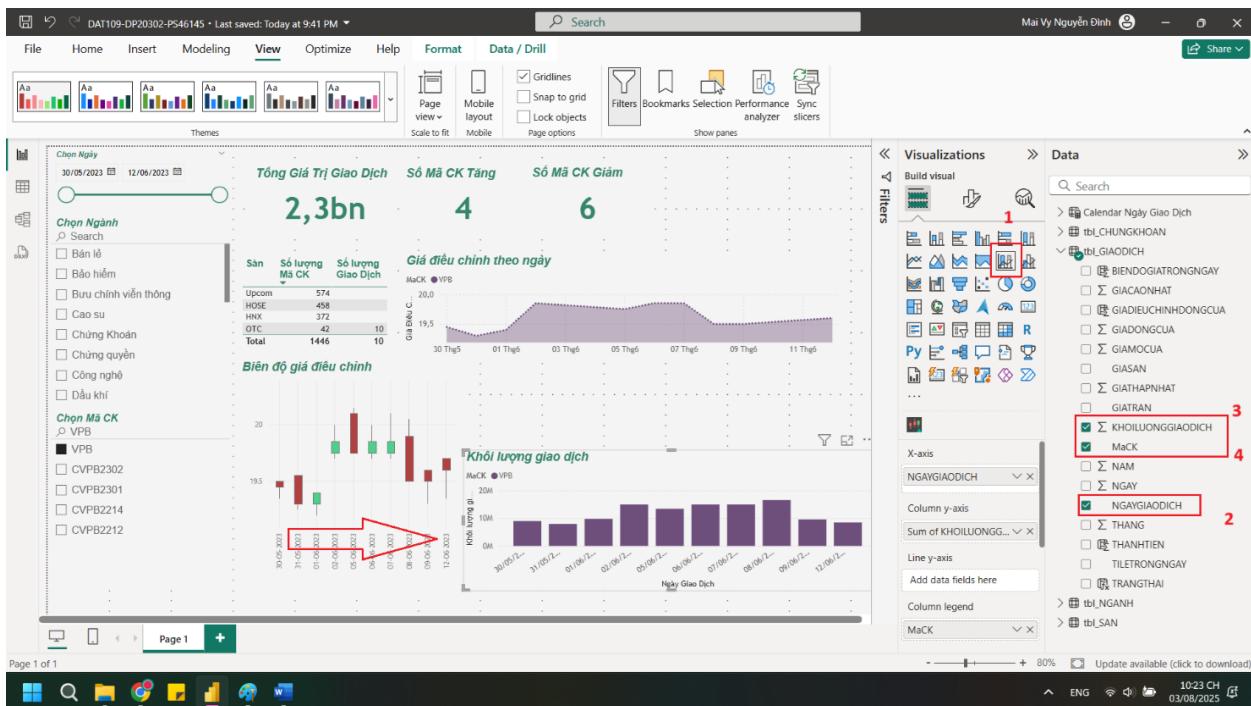


### 5.4.1.10 Tạo biểu đồ thống kê khối lượng giao dịch theo ngày

- Giao diện:



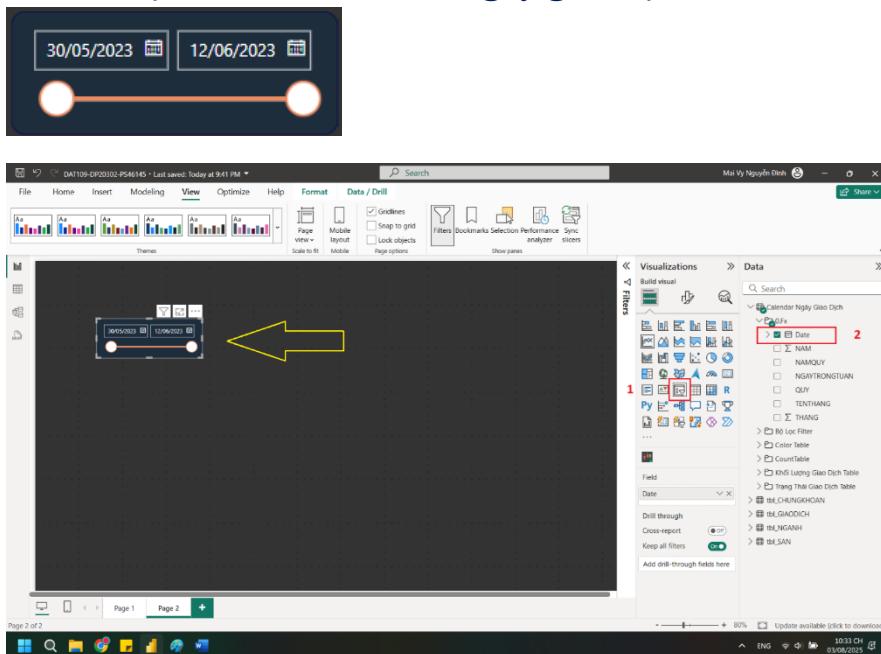
- Thiết đặt:



## 5.4.2 TẠO VISUAL THỐNG KÊ TỔNG THỂ

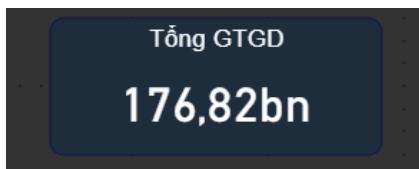
- ✓ *Tạo các visual dùng trong trang thống kê tổng thể.*
- ✓ *Việc này giúp lưu vết các chỉnh sửa thay đổi trên.*

### 5.4.2.1 Tạo visual filter theo ngày giao dịch

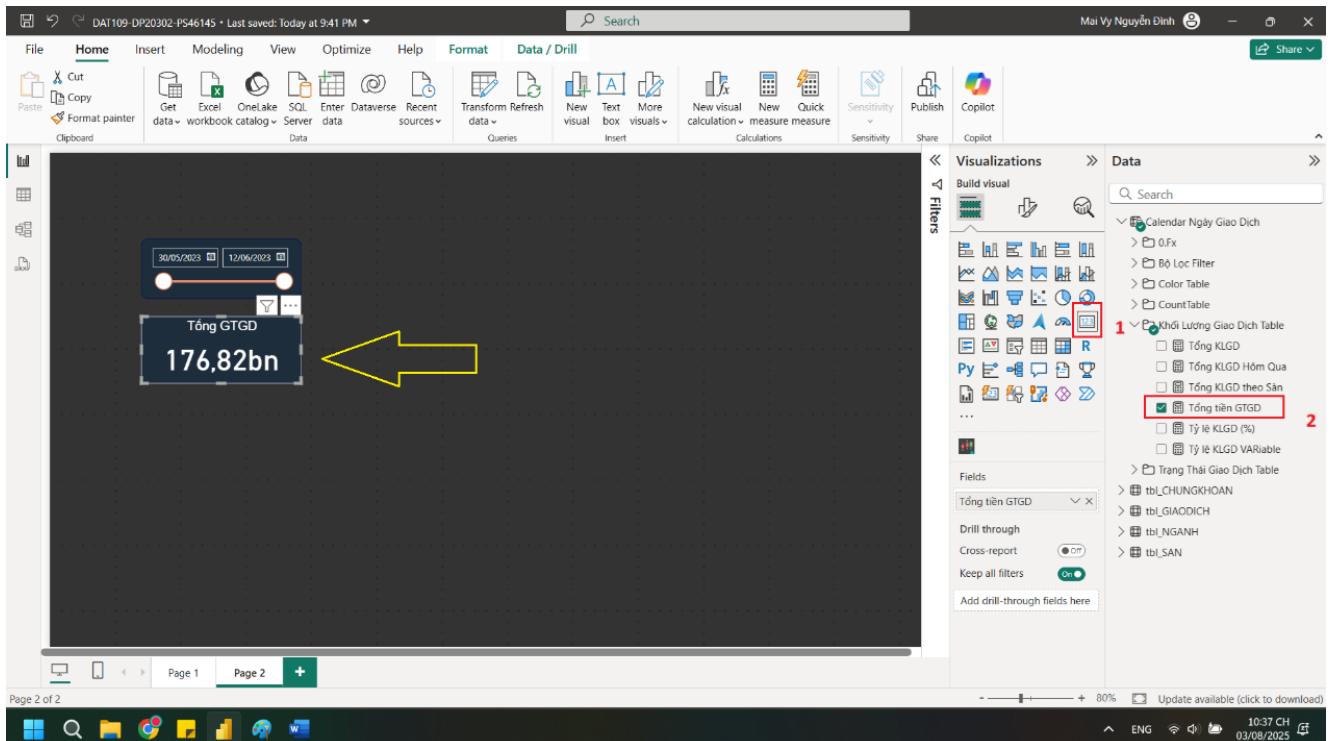


### 5.4.2.2 Tạo visual thống kê Tổng GTGD

- Giao diện:



- Thiết đặt:

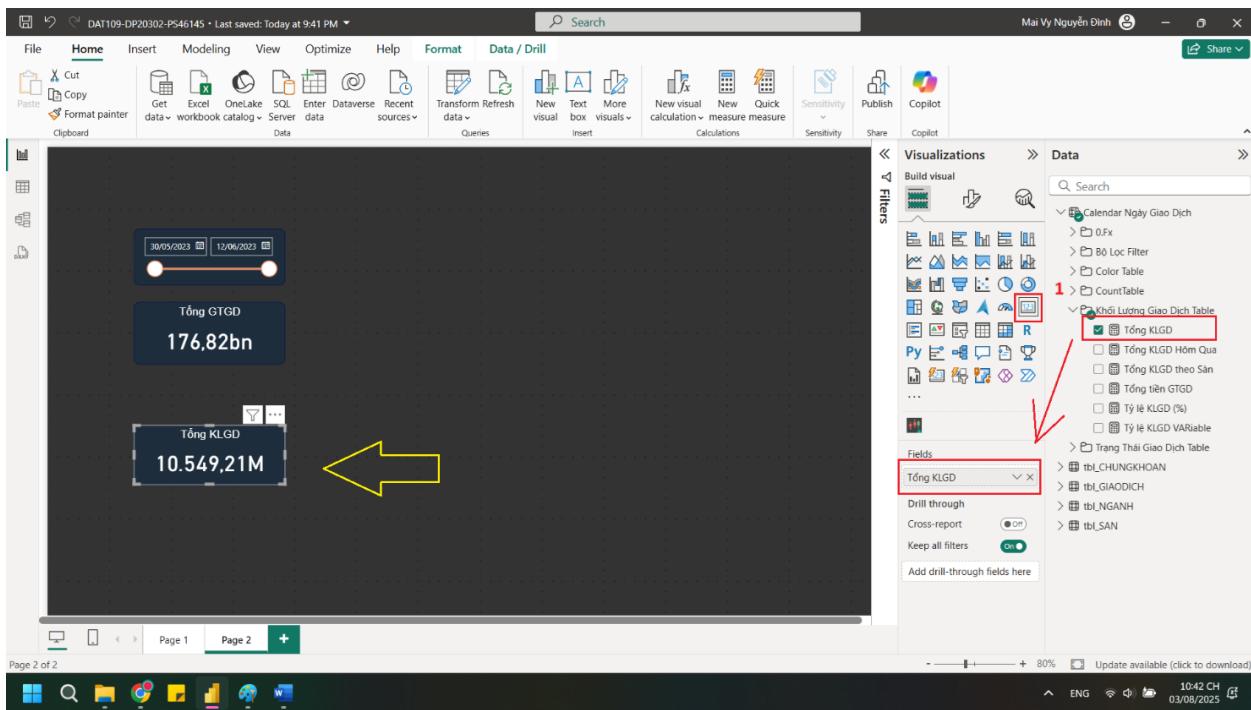


### 5.4.2.3 Tạo visual thống kê Tổng KLGD

- Giao diện:



- Thiết đặt:



#### 5.4.2.4 Tạo thống kê tổng lượng giao dịch so với hôm qua.

- Giao diện:



- Thiết đặt:

File Home Insert Modeling View Optimize Help Format Data / Drill

Clipboard Get data - workbook catalog - Data Transform refresh data - Queries New visual Text box Insert More visuals - New visual calculation - measure measure Calculations Sensitivity Publish Copilot Share

Visualizations Build visual Filters

Search

2  Tổng KLGD

4  Tổng KLGD Hôm Qua

1  Σ THANG

3  Σ THANG

Tổng KLGD

Tổng KLGD

KPI

8.472M

Goal: 7.587M (+884.14M +11.65%)

30/05/2023 12/06/2023

Tổng GTGD 176.82bn

Tổng KLGD 10.549.21M

Page 1 Page 2 +

Page 2 of 2

ENG 10:57 CH 03/08/2025

### 5.4.2.5 Tạo thống kê số mã CK giảm

- Giao diện:



- Thiết đặt:

File Home Insert Modeling View Optimize Help Format Data / Drill

Clipboard Get data - workbook catalog - Data Transform refresh data - Queries New visual Text box Insert More visuals - New visual calculation - measure measure Calculations Sensitivity Publish Copilot Share

Visualizations Build visual Filters

Search

1  Số Luong Mã Giảm

Số Luong Mã Giảm

Tổng GTGD 176.82bn

Tổng KLGD 10.549.21M

KPI

8.472M

Goal: 7.587M (+884.14M +11.65%)

30/05/2023 12/06/2023

Tổng GTGD

Tổng KLGD

KPI

Page 1 Page 2 +

Page 2 of 2

ENG 11:00 CH 03/08/2025

### 5.4.2.6 Tạo thống kê số mã CK tăng

- Giao diện:



- Thiết đặt:

The screenshot shows the Power BI desktop application. In the center, there is a dark dashboard with several cards. One card displays '4258' with the label 'Số mã Tăng'. To the left of this card is another card with '176.82bn' and 'Tổng GTGD'. Below it is a card with '10.549.21M' and 'Tổng KLOGD'. At the bottom left is a KPI card with '8.472M' and 'Goal: 7.587M (+884,14M +11.65%)'. On the right side, the 'Fields' pane is open, showing a list of fields under 'Số Lượng MÃ TĂNG'. A yellow arrow points from the 'Số mã Tăng' card on the dashboard to the 'Số Lượng MÃ TĂNG' field in the Fields pane. The 'Filters' pane is also visible on the right.

### 5.4.2.7 Tạo thống kê số lượng Mã CK giao dịch

- Giao diện:

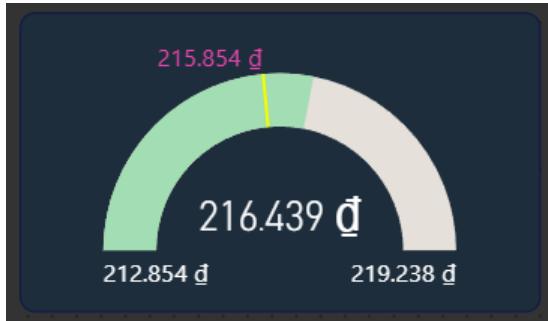
Sàn	Số lượng Mã CK	Số lượng Giao Dịch
HNX	372	2816
HOSE	458	4193
OTC	42	312
Upcom	574	3594
<b>Total</b>	<b>1446</b>	<b>10915</b>

- Thiết đặt:

The screenshot shows the Microsoft Power BI desktop interface. The main area displays a dashboard with several cards: a date range selector (30/05/2023 to 12/06/2023), a table titled 'Sàn' showing stock volumes (HNX, HOSE, OTC, Upcom), and three KPI cards for GTGD (176.82bn), KLGD (10.549.21M), and KPI (8.472M). A yellow arrow points from the table card to the Data pane on the right. The Data pane lists data sources and tables, with 'tbl\_CHUNGKHOAN' expanded to show 'Sàn' selected. Other tables like 'tbl\_GIODICH', 'tbl\_NGANH', and 'tbl\_SAN' are also listed.

#### 5.4.2.8 Tạo biểu đồ giá trong ngày của cổ phiếu so với giá tham chiếu

- Giao diện: đỏ là giảm, xanh là tăng, vàng là tham chiếu.



- Thiết đặt:

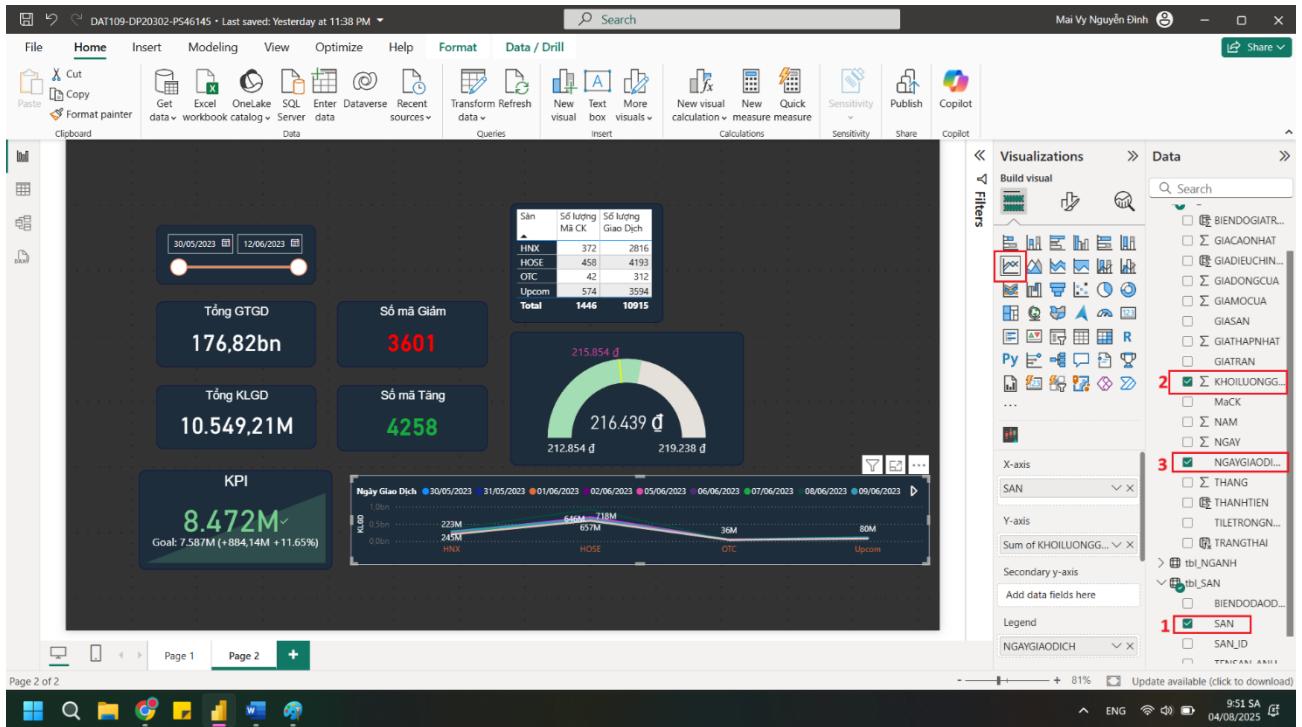
The screenshot shows a Microsoft Power BI desktop interface. The main area displays a dashboard with several visualizations: a card showing dates (30/05/2023 to 12/06/2023), a table of stock exchange volumes (HNX, HOSE, OTC, Upcom), a card for Total GTGD (176.82bn), a card for Sổ mã Giảm (3601), a card for Tổng KLGD (10.549.21M), a card for Sổ mã Tăng (4258), a KPI card (8.472M), and a gauge chart showing values from 212.85K to 216.44K. The ribbon at the top includes Home, Insert, Modeling, View, Optimize, Help, Format, and Data / Drill. The Data pane on the right lists various data sources and filters, with some items highlighted in red.

### 5.4.2.9 Tạo thống kê tổng khối lượng giao dịch trên 4 sàn

- Giao diện:



- Thiết đặt:

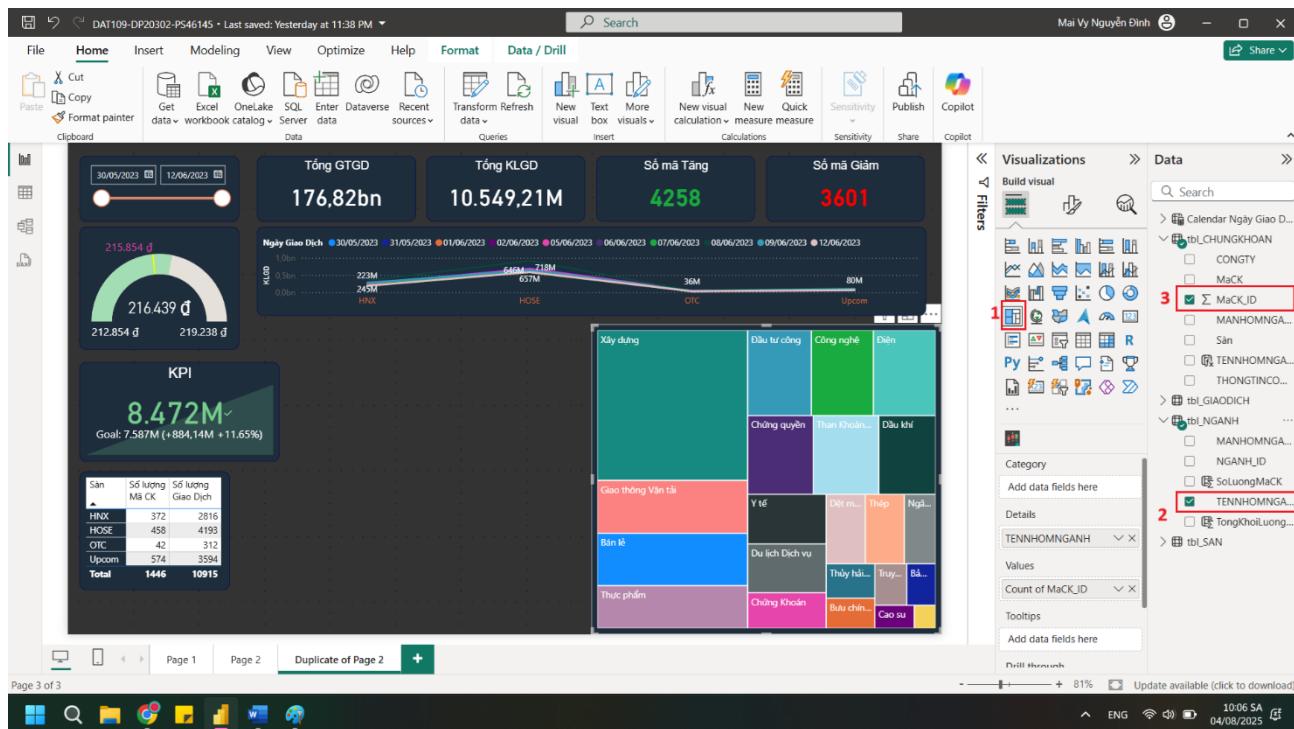


### 5.4.2.10 Tạo thống kê mã chứng khoán phân bố theo ngành

- Giao diện:



- Thiết đặt:



### 5.4.2.11 Tạo bảng thống kê trạng thái giao dịch của các mã chứng khoán

- Giao diện:

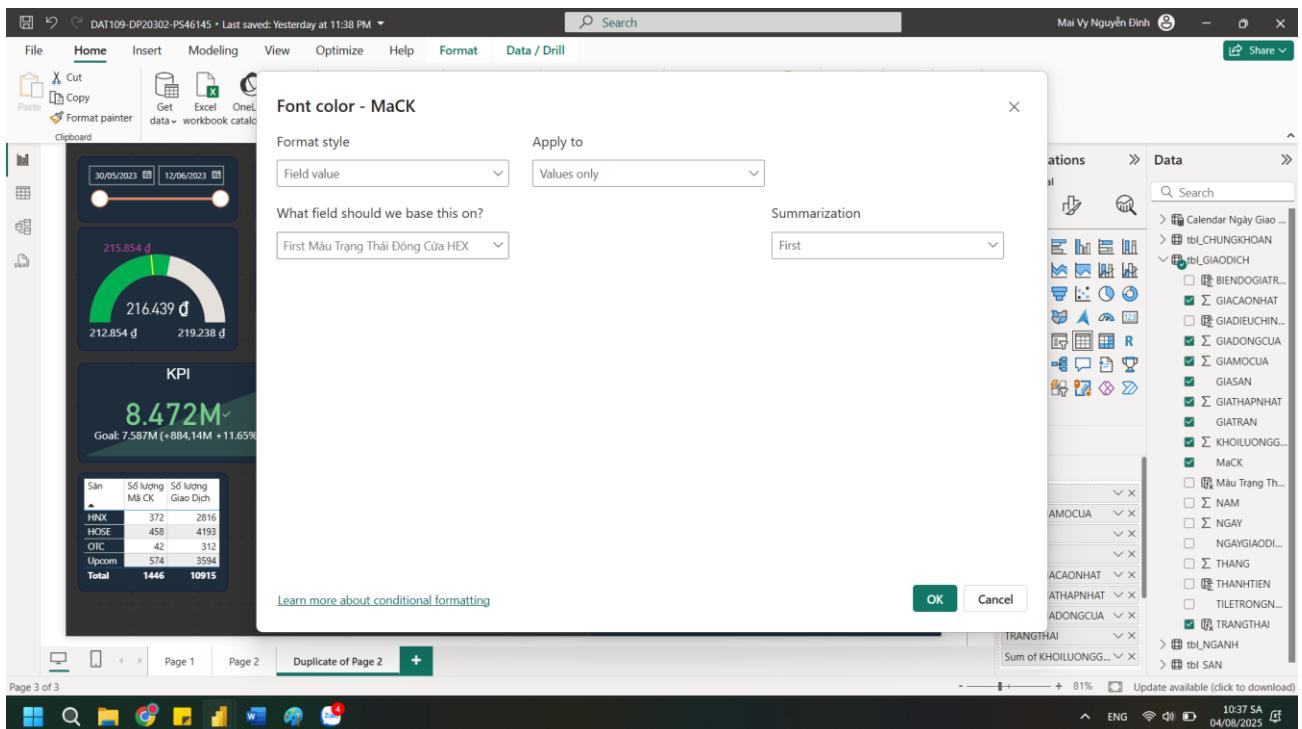
	MaCK	Sum of GIAMOCUA	GIATRAN	GIASAN	Sum of GIACAONHAT	Sum of GIATHAPNHAT	Sum of GIADONGCUA	TRANGTHAI	Sum of KHOILUONGGIAODICH
	VND	54,15	19,86	16,25	58,00	53,85	57,90	TĂNG	156.666.295,00
	SSI	72,30	25,79	22,41	76,00	71,70	75,45	TĂNG	102.972.342,00
	SHB	36,60	13,42	10,98	37,60	36,60	37,50	TĂNG	92.859.569,00
	SHS	23,80	13,09	10,71	25,50	23,60	25,40	TĂNG	82.820.373,00
	MBB	59,70	21,29	18,51	61,45	59,40	61,35	TĂNG	68.384.650,00
	VND	19,20	21,12	17,28	19,50	18,10	18,10	GIÁM	66.280.900,00
	NVL	29,60	15,84	13,76	30,00	28,40	28,80	GIÁM	64.834.900,00
	GIG	68,45	22,68	19,67	64,75	62,40	64,05	TĂNG	64.039.593,00
	VIX	29,91	10,97	8,97	31,45	29,76	31,35	TĂNG	61.964.840,00
	HSG	47,70	17,01	14,79	48,70	47,10	48,30	TĂNG	56.863.391,00
	GEX	51,75	19,84	14,66	54,55	51,00	54,00	TĂNG	54.325.522,00
	BXG	43,80	15,62	13,58	44,80	43,35	44,70	TĂNG	51.900.790,00
	STB	84,15	30,01	26,09	85,00	83,25	84,90	TĂNG	51.553.901,00
	SHB	12,30	13,53	11,07	12,35	12,05	12,20	GIÁM	45.928.500,00
	GEX	19,15	22,02	16,28	20,10	17,90	17,90	GIÁM	45.727.500,00
	SHB	12,25	13,48	11,03	12,40	12,10	12,35	TĂNG	44.911.100,00
	VPB	59,25	23,70	15,80	60,10	59,10	59,55	TĂNG	44.845.778,00
	NVI	40,50	14,45	12,56	41,05	40,20	40,00	TĂNG	43.481.275,00
	Total	215.854,19			219.237,77	212.854,00	216.439,12		10.549.208.092,00

- Thiết đặt:

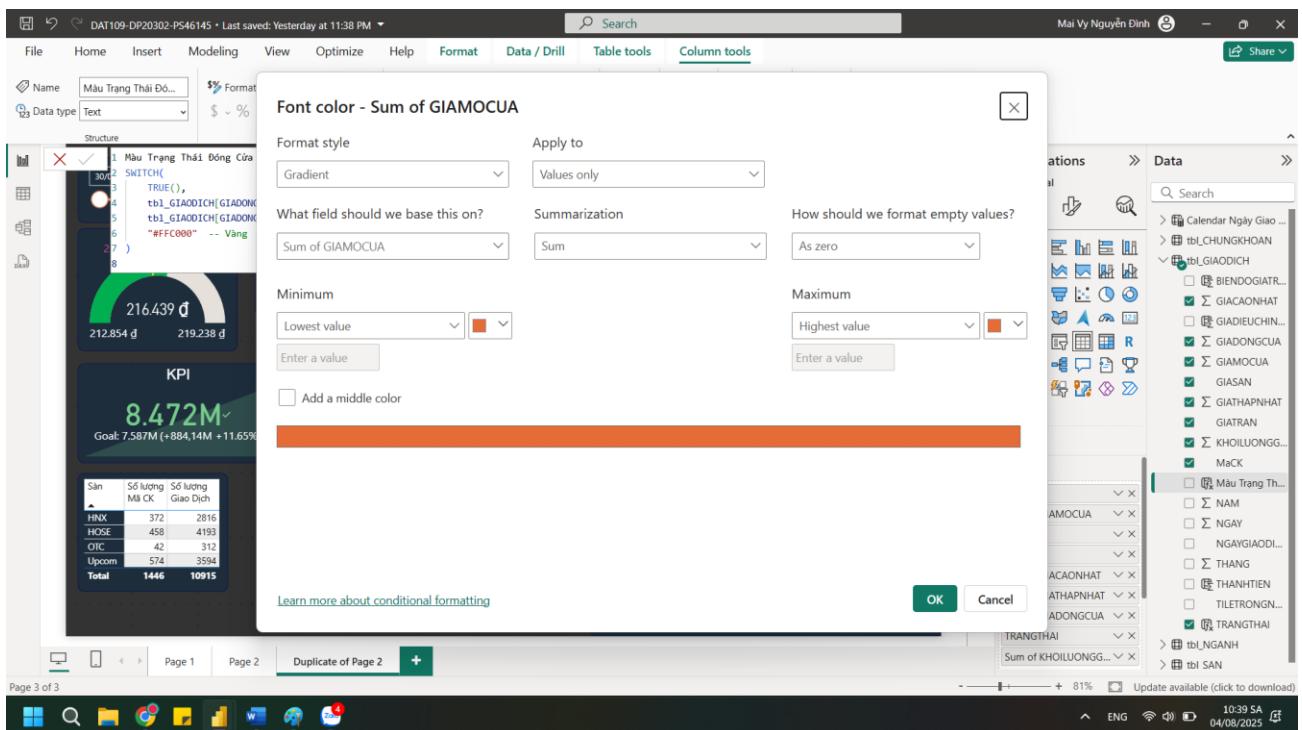
The screenshot shows a Power BI dashboard with several visualizations including KPI cards, a gauge chart, and a treemap chart. The 'Data' pane on the right displays a hierarchical view of tables and columns. A red box highlights the 'Columns' section, specifically the entry for 'MaCK'.

- Cột Mã CK: format font color

The screenshot shows the same Power BI dashboard as the previous one, but with a context menu open over the 'MaCK' column entry in the 'Columns' section of the Data pane. The menu path 'Conditional formatting > Font color' is highlighted with a red box, and the resulting color palette is shown on the right.



### - Cột Mở cửa: format font color



### - Cột Trần (CE): format font color

Font color - GIATRAN

Format style: Gradient

Apply to: Values only

What field should we base this on?: GIATRAN

Summarization: Count

How should we format empty values?: As zero

Minimum: Lowest value

Maximum: Highest value

Add a middle color:

Learn more about conditional formatting

OK Cancel

## - Cột Sàn: format font color

Font color - GIASAN

Format style: Gradient

Apply to: Values only

What field should we base this on?: GIASAN

Summarization: Count

How should we format empty values?: As zero

Minimum: Lowest value

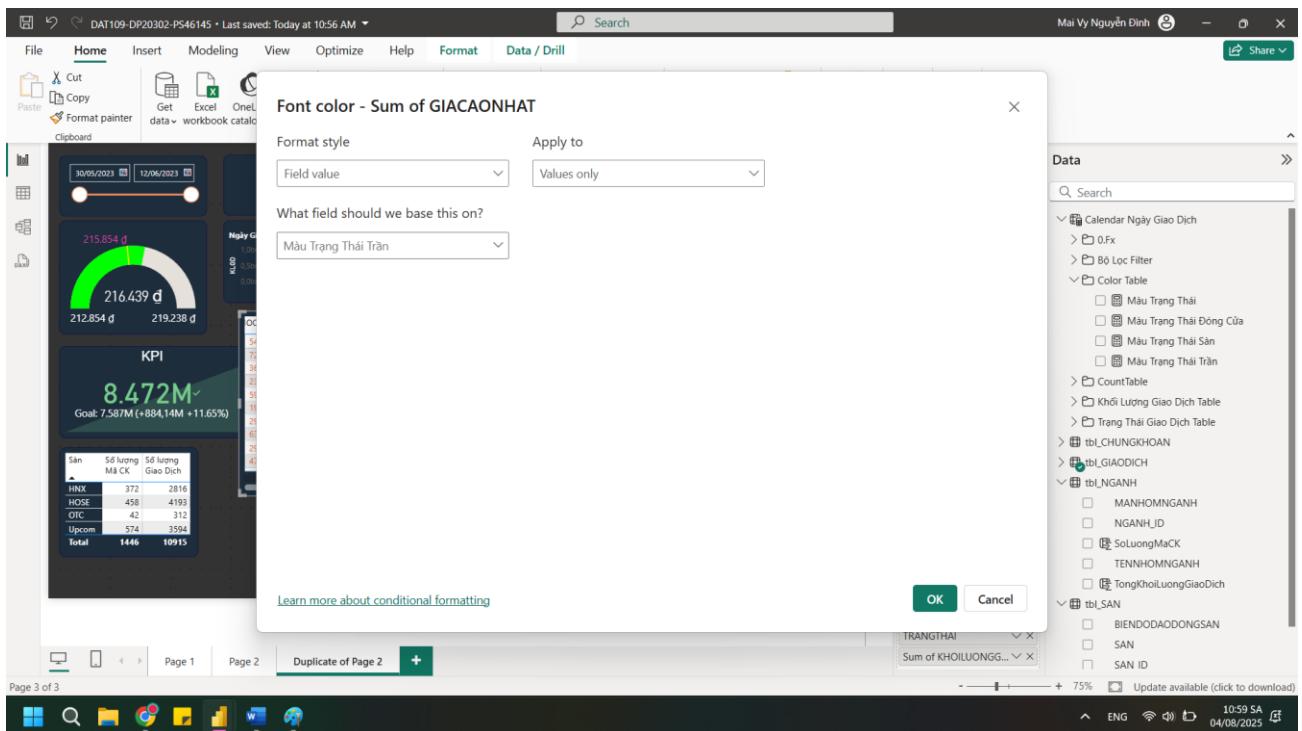
Maximum: Highest value

Add a middle color:

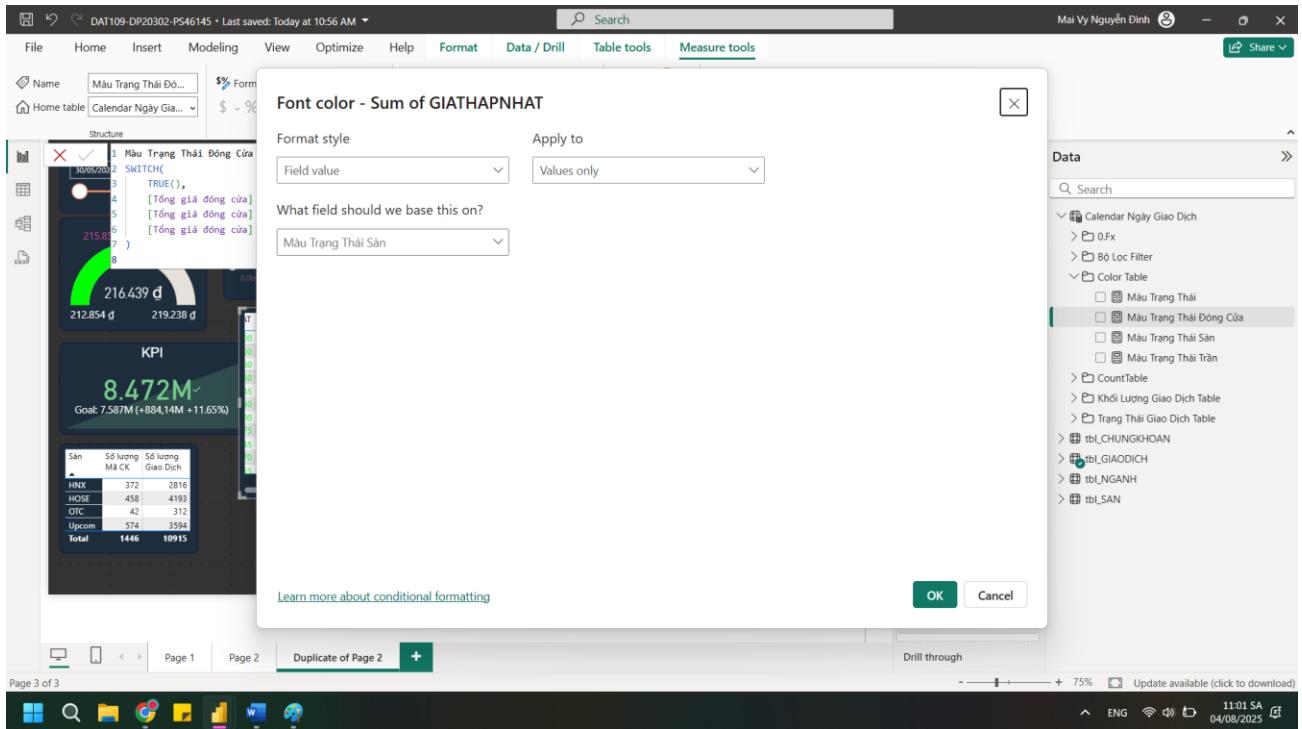
Learn more about conditional formatting

OK Cancel

## - Cột Cao nhất: format font color



- Cột Tháp nhất: format font color



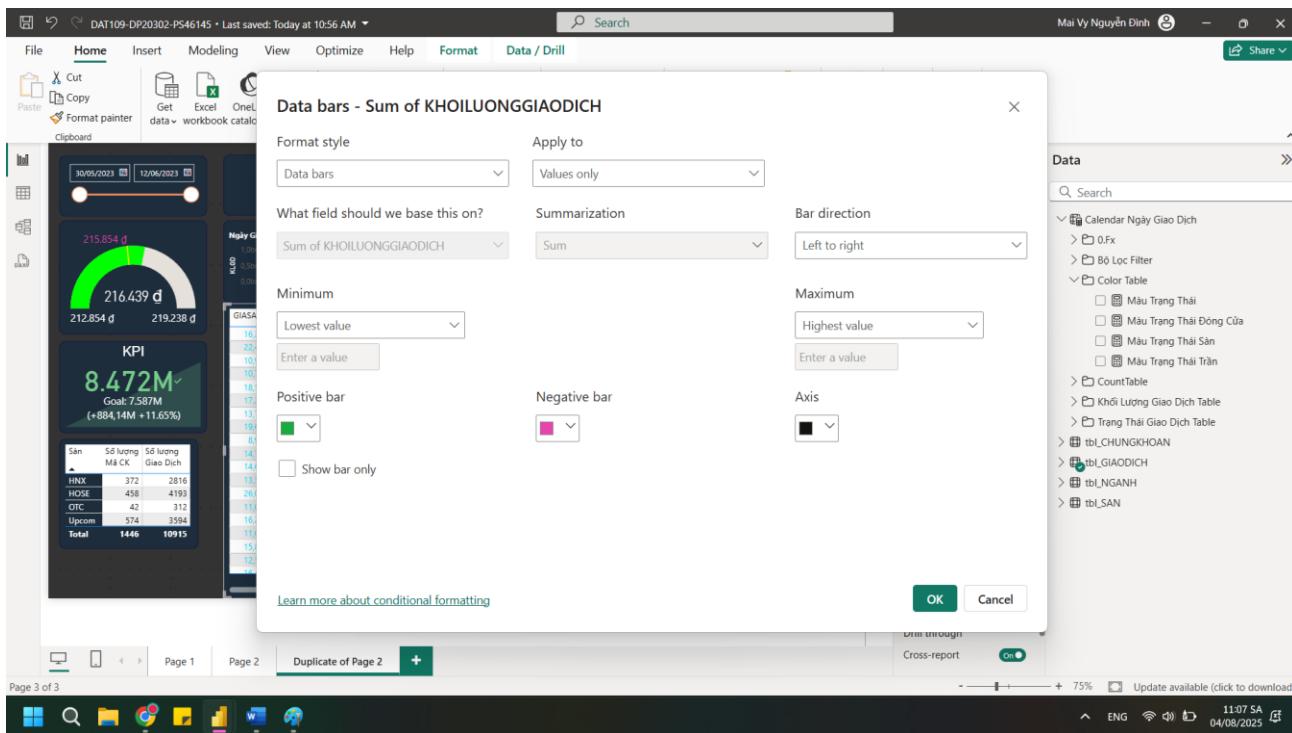
- Cột Đóng cửa: format font color

- Cột Trạng thái: format font color + icon

Screenshot of the Microsoft Power BI interface showing the 'Format' tab selected. A 'Icons - TRANGTHAI' dialog box is open, allowing users to apply icons based on specific conditions. The main workspace displays a dashboard with various KPIs and data tables.

Screenshot of the Microsoft Power BI interface showing the 'Format' tab selected. A dashboard is displayed with multiple visualizations, including a Data Bar chart for 'Khối lượng' (Volume) and a treemap chart for 'Xây dựng' (Construction). The right side shows the 'Visualizations' pane and the 'Data' pane with various data sources and filters.

- Cột Khối lượng: format Data Bar



## 6 XÂY DỰNG BÁO CÁO

### 6.1 DASHBOARD VÀ REPORT

✓ *Cách tối ưu hóa:*

- *Dashboard*
- *Report*

❑ *Áp dụng các bộ lọc để giới hạn số lượng dữ liệu*

❖ Ít dữ liệu hơn sẽ giúp cải thiện tốc độ tải hình ảnh.

❖ Khi áp dụng bộ lọc, hãy đặt câu hỏi: Người xem có cần thấy tất cả những dữ liệu này không?

❖ Còn cách nào để giảm thiểu số lượng dữ liệu được hiển thị, và hạn chế tối đa tác động xấu đến trải nghiệm của người dùng không?

❑ *Kiểm soát số lượng hình ảnh trên trang báo cáo*

❖ Nên giới hạn số lượng hình ảnh trên một trang báo cáo ở mức cần thiết và hợp lý.

- ❖ Điều này sẽ giúp người xem dễ hiểu hơn khi đọc báo cáo, và giúp báo cáo vận hành ổn định hơn.

#### ***❑ Đảm bảo hiệu suất cao của các hình ảnh tùy chỉnh***

- ❖ Lưu ý là hình ảnh trực quan có tác động rất lớn, ảnh hưởng đến toàn bộ báo cáo.

- ❖ Chúng ta nên đánh giá chất lượng và số lượng của các hình ảnh một cách chính xác.

#### ***❑ Hiểu rõ ý nghĩa về hiệu suất của dashboard***

- ❖ Khi chúng ta truy xuất dữ liệu từ bộ nhớ cache, dashboard sẽ hoạt động tốt hơn và nhanh hơn so với khi chỉ dựa vào nguồn dữ liệu.

#### ***❑ Tối ưu hóa môi trường vận hành***

- ❖ Có thể tối ưu hóa môi trường Power BI bằng cách định cấu hình cài đặt dung lượng, định cỡ công dữ liệu và giảm độ trễ mạng.

## **6.2 XÂY DỰNG BÁO CÁO**

- ✓ *Xây dựng báo cáo:*
  - *Dashboard*
  - *Report*

### **6.2.1 DASHBOARD VS REPORT**

- ✓ *Trình bày sự khác nhau giữa Dashboard và Report.*
- ✓ *Giải thích vì sao làm dashboard, report trong dự án này ?*

Item	Dashboard	Report
Pages	Chỉ một trang.	Có thể tạo một hoặc nhiều trang.

Data sources	Tạo từ nhiều datasets hoặc reports.	Tạo từ một dataset.
Visualization	Xây dựng thông tin chi tiết về dữ liệu bằng cách sử dụng biểu đồ, hình ảnh sinh động, lưu đồ.	Xây dựng trang tổng hợp thông tin, cung cấp góc nhìn tổng quan.
Available in Power BI Desktop	Không thể tạo.	Có thể tạo và xem.
Filters and Slicers	Không thể dùng, bị giới hạn trên trang đơn.	Có thể dùng các loại filter, highlight và slice khác nhau.
User Interactivity	Cho phép pin visuals từ report và datasets trên canvas đơn, làm cho nó đơn giản để nhóm những gì cần thiết cho người dùng.	Tập trung hơn vào khả năng trực quan hóa, áp dụng các phép biến đổi trên một tập dữ liệu.
Favourite	Có thể gán nhiều dashboards	Có thể gán nhiều dashboards
Q&A feature	Có	Có thể phân quyền chỉnh sửa report.
Alerts	Gửi email cảnh báo khi đáp ứng điều kiện, tiêu chí cụ thể hoặc vượt qua giới hạn.	Không thể
Subscribe	Có	Có
See underlying dataset tables and fields	Không thể xem dataset nhưng có thể xuất dữ liệu này.	Có thể xem dữ liệu trong tab dữ liệu.

Purpose	<p>Được sử dụng để giám sát cấp cao, thường theo thời gian thực hoặc gần thời gian thực, cung cấp cái nhìn tổng hợp về hiệu quả kinh doanh.</p>	<p>Được sử dụng để phân tích chuyên sâu và khám phá dữ liệu nhằm trả lời các câu hỏi kinh doanh phức tạp.</p>
---------	---	---

Giải thích:

#### **Dữ liệu**

- ❖ Loại 1: là dữ liệu thay đổi theo thời gian thực, giá cổ phiếu thay đổi trong phiên giao dịch, được cập nhật và hiển thị theo khung theo giây, 1 phút, 5 phút, giờ, ngày, tuần năm, 5 năm.
- ❖ Loại 2: dữ liệu lúc kết thúc phiên giao dịch, được chốt vào cuối ngày giao dịch.

#### **Dự án**

- ❖ Do bộ dữ liệu cung cấp (free) trong phạm vi môn học Dự án mẫu thuộc loại 2 được chốt theo ngày, giá trị chứng khoán không thay đổi theo thời gian thực.

⇒ Vì vậy sẽ tạo báo cáo report dạng tổng thể và chi tiết, chưa dùng tới dashboard (option).

*Nếu muốn mở rộng dự án, bạn có thể mua dữ liệu real-time từ sàn hoặc các công ty chứng khoán*

### **6.2.2 DASHBOARD**

- ✓ *Trình bày các loại dashboard có trong dự án.*
- ✓ *Giải thích vì sao làm các dashboard này ?*

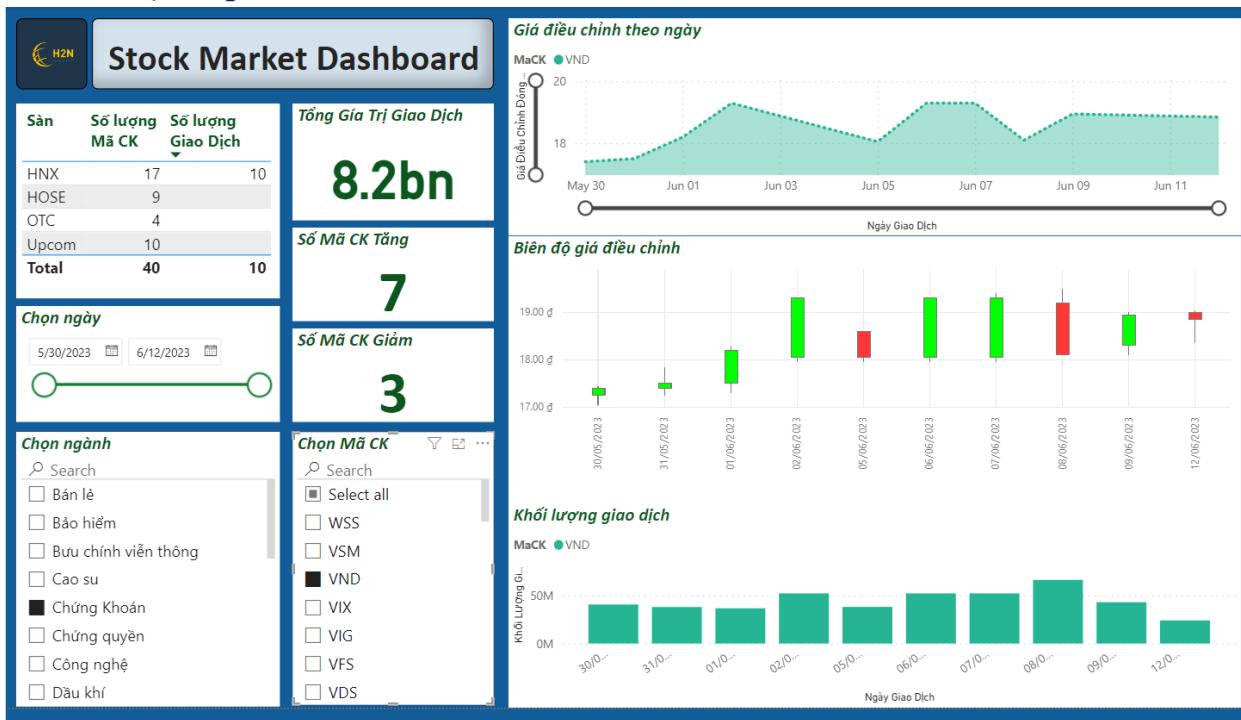
...

Giải thích: ...

### **6.2.3 REPORT**

- ✓ *Trình bày các loại report có trong dự án.*
- ✓ *Giải thích vì sao làm các report này ?*

### 6.2.3.1 Tạo report Detail



#### Giải thích:

Báo cáo này giúp người xem theo dõi **chi tiết biến động giá cổ phiếu theo từng ngày**, bao gồm:

- Giá điều chỉnh theo ngày của từng mã cổ phiếu.
- Biên độ dao động giá (candlestick chart) giúp nhận diện xu hướng tăng/giảm trong ngắn hạn.
- Khối lượng giao dịch để đánh giá mức độ quan tâm của thị trường. Ngoài ra, bộ lọc theo **ngày, ngành và mã cổ phiếu** cho phép người dùng tùy chỉnh phạm vi phân tích, hỗ trợ **đánh giá hiệu quả giao dịch trong khoảng thời gian cụ thể**.

### 6.2.3.2 Tạo report Overview



Giải thích:

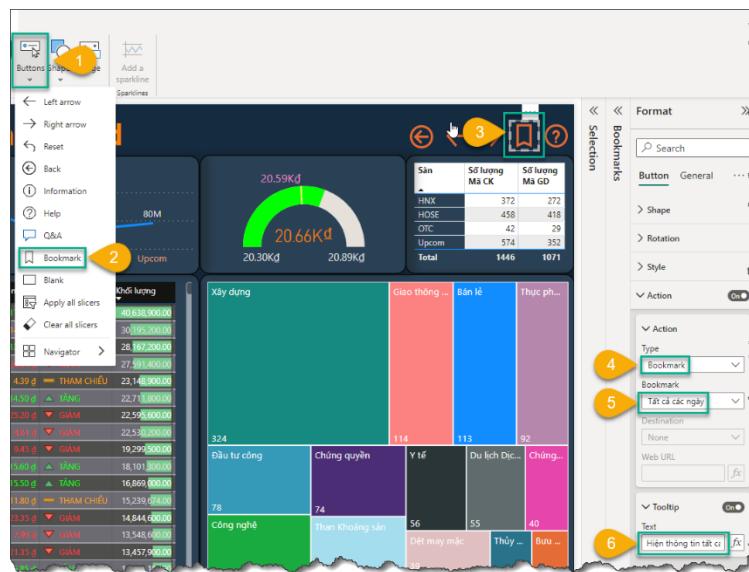
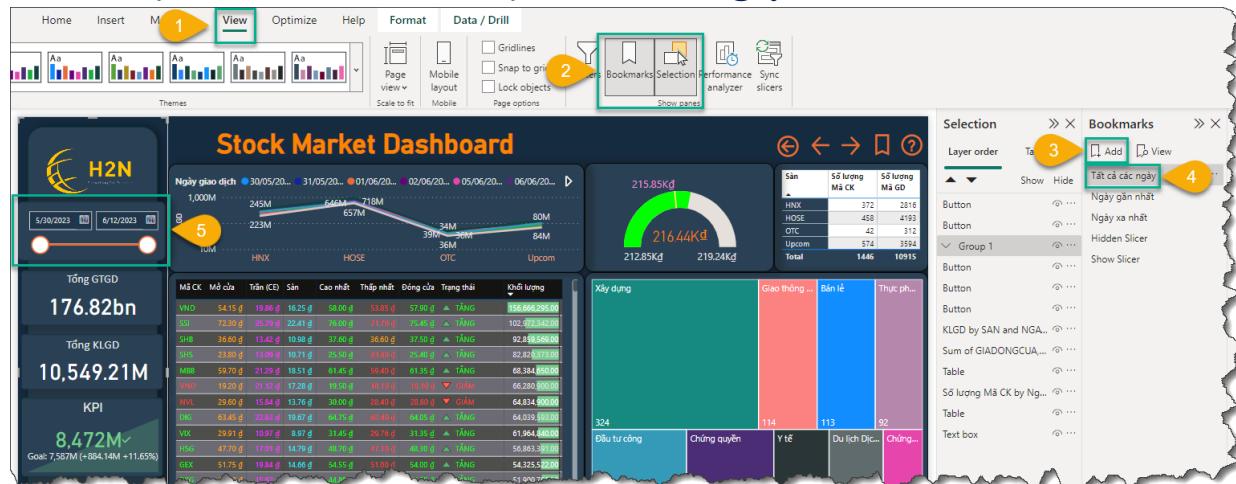
Báo cáo này cung cấp **bức tranh tổng thể** của thị trường chứng khoán tại một thời điểm hoặc trong một khoảng thời gian nhất định, bao gồm:

- Tổng giá trị và khối lượng giao dịch trên từng sàn (HNX, HOSE, OTC, Upcom).
  - Số lượng mã cổ phiếu tăng/giảm và mức thay đổi KPI so với mục tiêu.
  - Bảng dữ liệu chi tiết theo từng mã (giá mở cửa, đóng cửa, cao nhất, thấp nhất, trạng thái, khối lượng).
  - Biểu đồ treemap phân bố số lượng giao dịch theo ngành, giúp nhận diện **ngành đang dẫn dắt thị trường**.
- Báo cáo này phù hợp cho việc **nắm bắt xu hướng chung, đánh giá sức khỏe thị trường và so sánh hiệu suất giữa các sàn/nhóm ngành**.

### 6.2.4 BOOKMARK

- ✓ *Trình bày các loại bookmark có trong dự án.*
- ✓ *Giải thích vì sao làm các bookmark này ?*

### 6.2.4.1 Tạo bookmark slicer chọn tất cả các ngày



Giải thích: Giúp người xem có thể quay về trạng thái hiển thị thông tin tất cả các ngày (trạng thái ban đầu) một cách nhanh nhất.

## 6.2.4.2 Tạo bookmark slicer chọn ngày gần nhất

The screenshot shows a Power BI desktop interface with a 'Stock Market Report' dashboard. The dashboard includes a line chart showing price trends from 12/06/2022 to 12/06/2023, a gauge chart for '21.10Kd', and various tables and charts for market data. A 'Ngày Giao Dịch' (Trading Date) slicer is present. On the right, the 'Bookmarks' pane is open, showing a 'Bookmarks' section with a 'Tất cả các ngày' (All days) item and a 'Bookmark 2' item. The 'Format button' pane is also visible.

This screenshot shows the same Power BI desktop interface after selecting the 'Ngày gần nhất' (Recent day) bookmark. The 'Bookmarks' pane now highlights this selection. The dashboard content remains the same, displaying the latest market data and trends.

Giải thích: Giúp người xem có thể quay về trạng thái hiển thị thông tin ngày gần nhất một cách nhanh nhất

### 6.2.4.3 Tạo bookmark slicer chọn ngày xa nhất

The screenshot shows the Power BI interface with a 'Stock Market Report' dashboard. On the right, the 'Bookmarks' pane is open, displaying a list of bookmarks. A red box labeled '1' highlights the 'Add' button at the top of the list. Another red box labeled '2' highlights the 'Button' item in the list. A third red box labeled '3' highlights the 'Ngày xa nhất' (Most Recent Date) option under the 'Button' item. An arrow points from the 'Ngày xa nhất' option towards the 'Add' button.

Giải thích: Giúp người xem có thể quay về trạng thái hiển thị thông tin ngày xa nhất một cách nhanh nhất

### 6.2.4.4 Tạo bookmark hidden slicer

The screenshot shows the Power BI interface with a 'Stock Market Report' dashboard. On the right, the 'Bookmarks' pane is open, displaying a list of bookmarks. A red box highlights the 'Hidden Slicer' option under the 'Slicer' item. An arrow points from the 'Hidden Slicer' option towards the 'Add' button in the 'Bookmarks' pane.

Giải thích: Giúp người xem có thể nhanh chóng ẩn toàn bộ slicer, tập trung vào nội dung báo cáo chính mà không bị phân tán bởi các bộ lọc.

## 7 KẾT LUẬN

### 7.1 BÁO CÁO

#### 7.1.1 CÁC BƯỚC VIẾT BÁO CÁO

✓ *Trình bày các bước viết báo cáo phân tích dữ liệu*

- Consider Your Audience (Xem xét đối tượng)
- Draft Out the Sections (Soạn thảo nội dung)
- Develop The Body (Phát triển nội dung)
- Write the Introduction (Viết phần giới thiệu)
- Add a Short Conclusion (Thêm một kết luận ngắn)
- Include Data Visualization Elements (Thêm các đối tượng trực quan hóa dữ liệu)
- Proofread & Edit Before Submission (Chỉnh sửa trước khi gửi)

#### 7.1.2 TỔNG HỢP

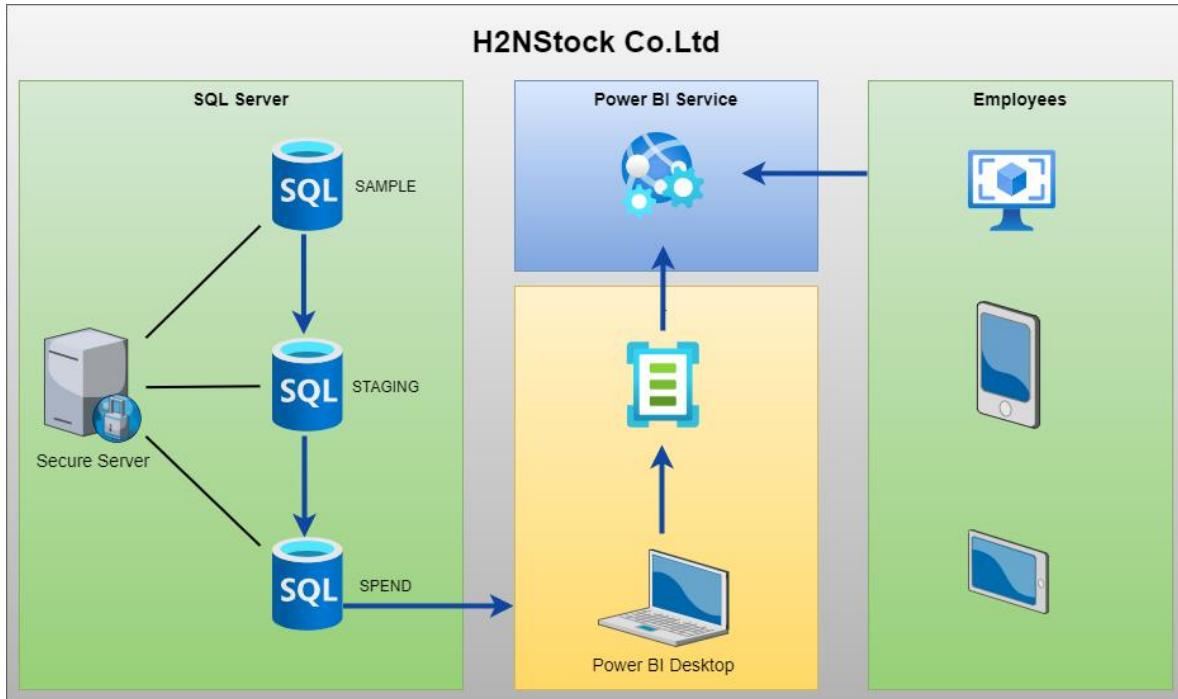
✓ *Tổng hợp các câu trả lời trong Câu Chuyện Dữ Liệu ở Phân Tích Khách Hàng, kết hợp với các bước viết báo cáo phân tích, để tạo thành bản báo cáo hoàn chỉnh.*

Gợi ý:

Dựa vào yêu cầu ban đầu của công ty H2NStock:

- Chọn công nghệ
  - ❖ Microsoft Power BI: phiên bản Desktop và Service
  - ❖ Hệ quản trị CSDL: SQL Server 2016+
  - ❖ Microsoft Office: Ms PowerPoint và Ms Word

- Chọn kiến trúc



Hệ thống H2NStock được thiết kế với luồng xử lý và phân tích dữ liệu như sau:

- **Tầng dữ liệu:** SQL Server (SAMPLE → STAGING → SPEND).
- **Tầng xử lý/báo cáo:** Power BI Desktop + Power BI Service.
- **Tầng người dùng:** Truy cập qua nhiều thiết bị.

- Cách thu thập dữ liệu, nguồn

(trong phạm vi dự án bộ dữ liệu được cung cấp sẵn, không thu thập)

- Phạm vi của bộ dữ liệu:

Dữ liệu giao dịch theo ngày, không phải dữ liệu thời gian thực.

- Làm thế nào nhân viên và quản lý vào xem báo cáo

Báo cáo được triển khai trên nền tảng **Power BI**. Nhân viên và quản lý có thể truy cập bằng hai cách:

- **Qua Power BI Desktop:** Dùng khi muốn mở file báo cáo trực tiếp trên máy tính, phục vụ chỉnh sửa hoặc phân tích chi tiết.
- **Qua Power BI Service (trực tuyến):** Truy cập thông qua đường link hoặc tài khoản nội bộ của công ty H2NStock. Cách này phù hợp cho việc xem nhanh, tương tác với biểu đồ và lọc dữ liệu mà không cần cài phần mềm.
- Báo cáo cung cấp thông tin gì ?

Báo cáo cung cấp cái nhìn tổng hợp về **sự biến động của các mã chứng khoán** dựa trên dữ liệu đã được phân nhóm theo **ngày giao dịch, ngành nghề và sàn giao dịch**. Cụ thể:

- Diễn biến giá và khối lượng giao dịch của từng mã cổ phiếu.
- So sánh sự thay đổi giữa các ngành nghề khác nhau.
- So sánh sự thay đổi giữa các sàn giao dịch.
- Thể hiện xu hướng biến động thông qua biểu đồ trực quan, giúp người xem dễ nhận biết tình hình thị trường.
- Hướng dẫn đọc thông tin báo cáo về chứng khoán ?

Khi truy cập báo cáo, người xem cần:

1. **Chọn bộ lọc (Slicer)** để lọc dữ liệu theo ngày, ngành nghề hoặc sàn giao dịch.
2. Quan sát các **biểu đồ cột, biểu đồ đường** để nhận biết biến động giá và khối lượng giao dịch.
3. Sử dụng chức năng **Drill-through** để xem chi tiết về một mã chứng khoán cụ thể.
4. Đọc chú thích màu sắc hoặc ký hiệu trên biểu đồ để hiểu ý nghĩa của từng cột hoặc đường biểu diễn.
- Các lưu ý cần thiết
  - Báo cáo được xây dựng nhằm cung cấp thông tin mô tả và phân tích dữ liệu lịch sử.

- Không bao gồm các khuyến nghị **mua** hoặc **bán** chứng khoán.
- Không đưa ra dự báo về xu hướng tăng hoặc giảm của thị trường trong tương lai.
- Không khuyến khích hoặc kêu gọi người xem (đặc biệt là học viên) tham gia đầu tư chứng khoán.
- Các số liệu trong báo cáo được trích xuất từ bộ dữ liệu đã được xử lý và đảm bảo tính chính xác, nhưng vẫn cần được xem là thông tin tham khảo, không dùng để ra quyết định tài chính.

## 7.2 KHÓ KHĂN

- Dữ liệu nguồn không đồng nhất: Bộ dữ liệu thu thập từ nhiều nguồn khác nhau (miễn phí và trả phí) nên định dạng, cấu trúc và chất lượng không đồng nhất, mất thời gian làm sạch và chuẩn hóa.
- Chưa quen với viết báo cáo nên còn rối thông tin & mất nhiều thời gian
- Thiếu kinh nghiệm: Chưa quen nhiều với SQL nên lúc đầu thao tác còn chậm, dễ làm sai, code không chạy.

## 7.3 THUẬN LỢI

- Có hướng dẫn từ giảng viên: Được hỗ trợ, giải thích các bước nên dễ hiểu hơn.
- Công cụ dễ tiếp cận: Power BI, SQL Server, Word, PowerPoint quen thuộc và có nhiều tài liệu tham khảo.
- Dữ liệu gần gũi: Là dữ liệu chứng khoán Việt Nam nên dễ liên hệ thực tế.

## 7.4 HƯỚNG PHÁT TRIỂN

- **Bổ sung thêm dữ liệu:** Thêm nhiều ngày giao dịch hơn để phân tích được xu hướng dài hơn.
- **Hoàn thiện báo cáo:** Trình bày visual đẹp hơn, thêm chú thích để người xem dễ hiểu.
- **Học thêm tính năng mới:** Thử các loại biểu đồ khác trong Power BI và các công thức DAX đơn giản hơn.
- **Thêm bộ lọc nâng cao:** Cho phép lọc dữ liệu theo nhiều tiêu chí như ngành, sàn, thời gian.