

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**  
**KHOA HỆ THÔNG THÔNG TIN**



**BÁO CÁO ĐỒ ÁN**  
**XÂY DỰNG KHO DỮ LIỆU CHO HỆ THỐNG SIÊU**  
**THỊ TẠI THỔ NHĨ KỲ**  
**Nhóm 15**

**Sinh viên thực hiện:**

Hoàng Hải Anh - 21521819  
Trương Khánh Long - 21521750

**Giảng viên:**

ThS. Đỗ Thị Minh Phụng

Thành phố Hồ Chí Minh, 21 tháng 05 năm 2024

## **MỤC LỤC**

<b>LỜI CẢM ƠN .....</b>	<b>1</b>
<b>1. Tổng quan về đề tài:.....</b>	<b>2</b>
1.1. Lý do chọn đề tài: .....	2
1.2. Mô tả bộ dữ liệu: .....	2
1.3. Mô tả chi tiết các thuộc tính: .....	3
1.4. Các thao tác tiền xử lý bộ dữ liệu:.....	4
1.4.1. Xử lý các dòng dữ liệu: .....	4
1.4.1.1. Xử lý giá trị khuyết: .....	4
1.4.1.2. Thuộc tính GENDER: .....	5
1.4.1.3. Phiên dịch dữ liệu:.....	5
1.4.2. Chọn lọc thuộc tính: .....	6
1.4.3. Đổi tên cột thuộc tính:.....	6
1.5. Thiết kế kho dữ liệu:.....	8
1.5.1. Dim_Product: .....	9
1.5.2. Dim_Brand:.....	10
1.5.3. Dim_Category:.....	10
1.5.4. Dim_Customer:.....	10
1.5.5. Dim_Salesman: .....	10
1.5.6. Dim_Branch: .....	11
1.5.7. Dim_Time: .....	11
1.5.8. Fact_Supermarket: .....	12
1.6. Nội dung 15 câu truy vấn: .....	12
<b>2. Tích hợp dữ liệu vào kho dữ liệu (SSIS):.....</b>	<b>13</b>
2.1. Chuẩn bị công cụ hỗ trợ: .....	14
2.2. Khởi tạo cơ sở dữ liệu để chuẩn bị quá trình SSIS:.....	20

2.3.	Hiện thực quá trình SSIS:.....	21
2.3.1.	Tạo bảng Dim và bảng Fact: .....	23
2.3.1.1.	Bảng Dim_Time .....	27
2.3.1.2.	Bảng Dim_Customer:.....	43
2.3.1.3.	Bảng Dim_Salesman: .....	48
2.3.1.4.	Bảng Dim_Branch:.....	54
2.3.1.5.	Bảng Dim_Brand:.....	60
2.3.1.6.	Bảng Dim_Category:.....	66
2.3.1.7.	Bảng Dim_Product:.....	72
2.3.1.8.	Bảng Fact_Supermarket: .....	96
2.3.2.	Tạo khóa ngoại từ bảng Fact đến các bảng Dimension: .....	118
2.3.3.	Chạy dự án SSIS: .....	121
2.4.	Kiểm tra dữ liệu các bảng:.....	127
2.4.1.	Bảng Dim_Time:.....	127
2.4.2.	Bảng Dim_Customer:.....	127
2.4.3.	Bảng Dim_Salesman: .....	127
2.4.4.	Bảng Dim_Branch:.....	128
2.4.5.	Bảng Dim_Brand: .....	129
2.4.6.	Bảng Dim_Category: .....	130
2.4.7.	Bảng Dim_Product:.....	130
2.4.8.	Bảng Fact: .....	131
2.5.	Lược đồ sau khi hoàn thành: .....	131
<b>3.</b>	<b>Phân tích dữ liệu trong kho dữ liệu (SSAS): .....</b>	<b>132</b>
3.1.	Chuẩn bị công cụ hỗ trợ: .....	132
3.1.1.	Cài đặt Microsoft Analysis Services Projects: .....	132
3.1.2.	Cài đặt Analysis Services:.....	137
3.2.	Tạo mới project SSAS:.....	152
3.3.	Xác định dữ liệu nguồn (Data Sources): .....	154

3.4.	Xác định khung nhìn dữ liệu nguồn (Data Source Views):.....	161
3.5.	Xây dựng các khối (Cubes) và xác định các độ đo (Measures): .....	168
3.6.	Xác định các chiều (Dimensions):.....	175
3.7.	Xác định các độ đo (Measures): .....	186
3.7.1.	Đổi tên và thuộc tính các độ đo ban đầu: .....	186
3.7.2.	Tạo các độ đo mới:.....	187
3.8.	Phân cấp trong bảng chiều:.....	188
3.8.1.	Phân cấp bảng Dim_Time:.....	188
3.8.2.	Phân cấp bảng Dim_Product:.....	206
3.9.	Chạy dự án SSAS: .....	214
3.10.	Quá trình phân tích dữ liệu bằng thao tác tay trên các khối CUBE:....	216
3.10.1.	Câu truy vấn 1: .....	216
3.10.2.	Câu truy vấn 2: .....	218
3.10.3.	Câu truy vấn 3: .....	219
3.10.4.	Câu truy vấn 4: .....	221
3.10.5.	Câu truy vấn 5: .....	223
3.10.6.	Câu truy vấn 6: .....	226
3.10.7.	Câu truy vấn 7: .....	228
3.10.8.	Câu truy vấn 8: .....	230
3.10.9.	Câu truy vấn 9: .....	233
3.10.10.	Câu truy vấn 10: .....	235
3.10.11.	Câu truy vấn 11: .....	237
3.10.12.	Câu truy vấn 12: .....	238
3.10.13.	Câu truy vấn 13: .....	239
3.10.14.	Câu truy vấn 14: .....	242
3.10.15.	Câu truy vấn 15: .....	243
3.11.	Quá trình phân tích dữ liệu bằng ngôn ngữ MDX: .....	246

3.11.1.	Câu truy vấn 1: .....	246
3.11.2.	Câu truy vấn 2: .....	247
3.11.3.	Câu truy vấn 3: .....	248
3.11.4.	Câu truy vấn 4: .....	249
3.11.5.	Câu truy vấn 5: .....	250
3.11.6.	Câu truy vấn 6: .....	251
3.11.7.	Câu truy vấn 7: .....	252
3.11.8.	Câu truy vấn 8: .....	253
3.11.9.	Câu truy vấn 9: .....	254
3.11.10.	Câu truy vấn 10: .....	255
3.11.11.	Câu truy vấn 11: .....	256
3.11.12.	Câu truy vấn 12: .....	257
3.11.13.	Câu truy vấn 13: .....	258
3.11.14.	Câu truy vấn 14: .....	259
3.11.15.	Câu truy vấn 15: .....	260
3.12.	Quá trình lập báo biểu bằng công cụ Pivot Excel:.....	261
3.12.1.	Kết nối Excel với SQL Server: .....	261
3.12.2.	Thiết lập môi trường để thực hiện báo biểu bằng Excel: .....	276
3.12.3.	Thực hiện Report bằng Excel:.....	279
3.12.3.1.	Câu truy vấn 1: .....	279
3.12.3.2.	Câu truy vấn 2: .....	285
3.12.3.3.	Câu truy vấn 3: .....	291
3.12.3.4.	Câu truy vấn 4: .....	296
3.12.3.5.	Câu truy vấn 5: .....	303
3.12.3.6.	Câu truy vấn 6: .....	308
3.12.3.7.	Câu truy vấn 7: .....	313
3.12.3.8.	Câu truy vấn 8: .....	318
3.12.3.9.	Câu truy vấn 9: .....	325
3.12.3.10.	Câu truy vấn 10:.....	328

3.12.3.11. Câu truy vấn 11:.....	334
3.12.3.12. Câu truy vấn 12:.....	340
3.12.3.13. Câu truy vấn 13:.....	345
3.12.3.14. Câu truy vấn 14:.....	350
3.13. Quá trình lập báo biểu bằng công cụ Power BI: .....	354
3.13.1. Kết nối Power BI với SQL Server: .....	354
3.13.2. Thực hiện Report bằng Power BI:.....	359
3.13.2.1. Câu truy vấn 1: .....	359
3.13.2.2. Câu truy vấn 2: .....	364
3.13.2.3. Câu truy vấn 3: .....	370
3.14. Quá trình lập báo biểu bằng Visual Studio (SSRS): .....	377
3.14.1. Tạo Project SSRS, thêm data source và dataset: .....	377
3.14.2. Tạo report bảng với Grouping với Totals: .....	382
3.14.2.1. Câu truy vấn 1: .....	382
3.14.2.2. Câu truy vấn 2: .....	392
3.14.2.3. Câu truy vấn 3: .....	401
3.14.2.4. Câu truy vấn 7: .....	409
<b>4. Data Mining: .....</b>	<b>417</b>
4.1. Mô tả dataset: .....	417
4.2. Mô tả chi tiết các thuộc tính: .....	418
4.3. Quá trình đổ dữ liệu vào SQL Server bằng công cụ SSIS trên Visual Studio: .....	420
4.4. Data Mining bằng công cụ SSAS trên Visual Studio:.....	423
4.5. Nhận xét kết quả:.....	436
4.5.1. Thuật toán Microsoft Decision Tree: .....	436
4.5.2. Thuật toán Microsoft Clustering:.....	440
4.5.3. Thuật toán Microsoft Naïve Bayes: .....	444

4.5.4. Đánh giá thuật toán bằng Mining Accuracy Chart: .....	446
4.6. Dự đoán: .....	450
4.7. Phân tích Dataset gốc: .....	455
4.7.1. Thống kê mô tả: .....	455
4.7.2. Trực quan hóa dữ liệu: .....	457
4.8. Tiền xử lý dữ liệu: .....	459
4.8.1. Kiểm tra mối tương quan giữa các thuộc tính:.....	459
4.8.2. Xử lý giá trị khuyết: .....	461
4.8.3. Chuẩn hóa dữ liệu: .....	462
4.9. Ứng dụng mô hình thuật toán khai thác dữ liệu: .....	463
4.9.1. Chia dữ liệu trước khi xây dựng mô hình thuật toán: .....	464
4.9.2. Decision Tree: .....	465
4.9.2.1. Giới thiệu thuật toán: .....	465
4.9.2.2. Xây dựng mô hình: .....	466
4.9.3. Random Forest: .....	470
4.9.3.1. Giới thiệu thuật toán: .....	470
4.9.3.2. Xây dựng mô hình: .....	471
4.9.4. Đánh giá thuật toán: .....	476
4.9.4.1. Các độ đo dùng để đánh giá thuật toán: .....	476
4.9.4.2. Đánh giá, so sánh các mô hình thuật toán: .....	476
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>478</b>
<b>BẢNG PHÂN CÔNG CÔNG VIỆC .....</b>	<b>479</b>

## **DANH MỤC HÌNH ẢNH**

Hình 1: Dữ liệu gốc được hiển thị trong Jupyter Notebook.....	3
Hình 2: Thông tin các thuộc tính trong bộ dữ liệu .....	5
Hình 3: Format lại giá trị cột thuộc tính GENDER .....	5
Hình 4: Loại bỏ một số cột thuộc tính không cần thiết .....	6
Hình 5: Đổi tên một số cột thuộc tính.....	8
Hình 6: Mô hình kho dữ liệu (mô hình bông tuyết) .....	9
Hình 7: Giao diện website tải Visual Studio Community 2019 .....	14
Hình 8: Giao diện website vị trí tải Visual Studio Community 2019 .....	15
Hình 9: Giao diện download thành công file .exe .....	15
Hình 10: Giao diện sau khi mở file .exe .....	16
Hình 11: Giao diện cài đặt SQL Server Data Tools.....	16
Hình 12: Giao diện khởi động một project rỗng .....	17
Hình 13: Giao diện vào nơi download SSIS.....	17
Hình 14: Giao diện hướng dẫn tìm kiếm và tải file SSIS.....	18
Hình 15: Giao diện sau khi hoàn tất quá trình tải SSIS trong VS Community 2019.....	19
Hình 16: Giao diện hướng dẫn tiếp tục bước cuối cùng để tải SSIS.....	20
Hình 17: Kết quả sau khi hoàn tất quá trình cài đặt và thiết lập SSIS vào Visual Studio 2019 .....	20
Hình 18: Kết nối đến server trong SSMS.....	21
Hình 19: Khởi tạo một database rỗng để chuẩn bị đổ dữ liệu vào .....	21
Hình 20: Khởi tạo một dự án mới.....	22
Hình 21: Tìm kiếm và chọn loại dự án SSIS.....	22
Hình 22: Đặt tên cho dự án và thiết lập nơi lưu trữ cho dự án .....	23
Hình 23: Giao diện tạo một kết nối mới đến file chứa dữ liệu gốc (xlsx) .....	24
Hình 24: Thiết lập đường dẫn lưu trữ file dữ liệu gốc trên máy tính.....	24
Hình 25: Giao diện lưu trữ file hiện ra chọn file dữ liệu gốc để tải lên.....	25
Hình 26: Chọn sheet dữ liệu lưu trữ dữ liệu gốc .....	26
Hình 27: Kiểm tra các cột dữ liệu trong file dữ liệu gốc.....	27
Hình 28: Lấy ra cột dữ liệu cần thiết cho bảng Dim_Time.....	28
Hình 29: Công cụ chuyển đổi kiểu dữ liệu của SSIS .....	28
Hình 30: Chuyển đổi kiểu dữ liệu của cột invoice_date.....	29
Hình 31: Đặt lại output alias cho cột invoice_date.....	30

<i>Hình 32: Tạo các cột dữ liệu cho bảng Dim_Time từ cột invoice_date .....</i>	31
<i>Hình 33: Sử dụng hàm DATEPART để lấy ra ngày từ giá trị cột invoice_date .....</i>	32
<i>Hình 34: Sử dụng hàm DATEPART để lấy ra tháng, năm, giờ, phút, giây, ngày trong tuần .....</i>	33
<i>Hình 35: Kiểm tra kết quả các cột mới được tạo từ cột invoice_date .....</i>	34
<i>Hình 36: Tạo mới một OLE DB Destination để đổ dữ liệu vừa tạo .....</i>	35
<i>Hình 37: Tạo một kết nối mới đến Microsoft SQL Server .....</i>	36
<i>Hình 38: Chọn tên server lưu trữ kho dữ liệu được tạo ở Hình 19 .....</i>	37
<i>Hình 39: Kiểm tra kết nối đến SQL Server .....</i>	38
<i>Hình 40: Hoàn tất quá trình tạo mới kết nối đến SQL Server .....</i>	39
<i>Hình 41: Tạo mới một bảng Dim_Time trong cơ sở dữ liệu Turkey_MarketSales .....</i>	40
<i>Hình 42: Nhập các câu lệnh SQL tương ứng của bảng Dim_Time .....</i>	41
<i>Hình 43: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 35 .....</i>	42
<i>Hình 44: Đổi tên toàn bộ Data Flow Task (quá trình trên) thành Create Dim_Time .....</i>	43
<i>Hình 45: Tạo mới Data Flow Task cho bảng Dim_Customer .....</i>	43
<i>Hình 46: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu .....</i>	44
<i>Hình 47: Thành phần của bảng Dim_Customer .....</i>	44
<i>Hình 48: Chọn ra các cột dữ liệu cho bảng Dim_Customer .....</i>	45
<i>Hình 49: Tạo mới một OLE DB Destination đặt tên là Dim_Customer .....</i>	46
<i>Hình 50: Kết nối đến kho dữ liệu Turkey_MarketSales và tạo bảng Dim_Customer .....</i>	46
<i>Hình 51: Nhập các câu lệnh SQL tương ứng với bảng Dim_Customer .....</i>	47
<i>Hình 52: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 47 .....</i>	48
<i>Hình 53: Tạo mới Data Flow Task cho bảng Dim_Salesman .....</i>	49
<i>Hình 54: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu .....</i>	49
<i>Hình 55: Thành phần của bảng Dim_Salesman .....</i>	50
<i>Hình 56: Chọn ra các cột dữ liệu cho bảng Dim_Customer .....</i>	51
<i>Hình 57: Tạo mới một OLE DB Destination đặt tên là Dim_Salesman .....</i>	52
<i>Hình 58: Kết nối đến kho dữ liệu Turkey_Marketales và tạo bảng Dim_Salesman .....</i>	52
<i>Hình 59: Nhập các câu lệnh SQL tương ứng với bảng Dim_Salesman .....</i>	53
<i>Hình 60: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 55 .....</i>	54
<i>Hình 61: Tạo mới Data Flow Task cho bảng Dim_Branch .....</i>	55
<i>Hình 62: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu .....</i>	55
<i>Hình 63: Thành phần của bảng Dim_Branch .....</i>	56

<i>Hình 64: Chọn ra các cột dữ liệu cho bảng Dim_Branch.....</i>	57
<i>Hình 65:Tạo mới một OLE DB Destination đặt tên là Dim_Branch .....</i>	58
<i>Hình 66: Kết nối đến kho dữ liệu Turkey_MarketSales và tạo bảng Dim_Branch .....</i>	58
<i>Hình 67: Nhập các câu lệnh SQL tương ứng với bảng Dim_Branch.....</i>	59
<i>Hình 68: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 63 .....</i>	60
<i>Hình 69: Tạo mới Data Flow Task cho bảng Dim_Brand .....</i>	61
<i>Hình 70: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu .....</i>	61
<i>Hình 71: Thành phần của bảng Dim_Brand .....</i>	62
<i>Hình 72: Chọn ra các cột dữ liệu cho bảng Dim_Brand.....</i>	63
<i>Hình 73: Tạo mới một OLE DB Destination đặt tên là Dim_Brand .....</i>	64
<i>Hình 74: Kết nối đến kho dữ liệu Turkey_MarketSales và tạo bảng Dim_Brand.....</i>	64
<i>Hình 75: Nhập các câu lệnh SQL tương ứng với bảng Dim_Brand.....</i>	65
<i>Hình 76: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 71 .....</i>	66
<i>Hình 77: Tạo mới Data Flow Task cho bảng Dim_Category .....</i>	67
<i>Hình 78: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu .....</i>	67
<i>Hình 79: Thành phần của bảng Dim_Category .....</i>	68
<i>Hình 80: Chọn ra các cột dữ liệu cho bảng Dim_Category.....</i>	69
<i>Hình 81: Tạo mới một OLE DB Destination đặt tên là Dim_Category .....</i>	70
<i>Hình 82: Kết nối đến kho dữ liệu Turkey_MarketSales và tạo bảng Dim_Category.....</i>	70
<i>Hình 83: Nhập các câu lệnh SQL tương ứng với bảng Dim_Category.....</i>	71
<i>Hình 84: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 79 .....</i>	72
<i>Hình 85: Tạo mới một Data Flow Task cho bảng Dim_Product .....</i>	73
<i>Hình 86: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu .....</i>	73
<i>Hình 87: Thành phần của bảng Dim_Product .....</i>	74
<i>Hình 88: Chọn ra các cột dữ liệu cho bảng Dim_Product.....</i>	75
<i>Hình 89: Tạo mới một OLE DB Source và chọn bảng Dim_Branch làm data source.....</i>	76
<i>Hình 90: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 87 .....</i>	77
<i>Hình 91: Tạo 2 Sort và đặt tên lần lượt là Sort và Sort1 tương ứng cho Product1 và Dim_Branch.....</i>	78
<i>Hình 92: Chọn ra các cột dữ liệu cho bảng Dim_Branch.....</i>	79
<i>Hình 93: Tạo mới một Merge Join để chuẩn bị kết Dim_Product và Dim_Branch.....</i>	80
<i>Hình 94: Thực hiện phép kết trái giữa Dim_Product và Dim_Branch .....</i>	81
<i>Hình 95:Tạo mới một OLE DB Source và chọn bảng Dim_Brand làm data source .....</i>	82

<i>Hình 96: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 87 .....</i>	83
<i>Hình 97: Tạo 2 Sort và đặt tên lần lượt là Sort2 và Sort3 tương ứng cho Product2 và Dim_Branch.....</i>	84
<i>Hình 98: Chọn ra các cột dữ liệu cho bảng Dim_Brand.....</i>	85
<i>Hình 99: Tạo mới một Merge Join để chuẩn bị kết Product1 và Dim_Branch.....</i>	86
<i>Hình 100: Thực hiện phép kết trái giữa Product1 và Dim_Branch .....</i>	87
<i>Hình 101: Tạo mới một OLE DB Source và chọn bảng Dim_Category làm data source..</i>	88
<i>Hình 102: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 87 .....</i>	89
<i>Hình 103: Tạo 2 Sort và đặt tên lần lượt là Sort4 và Sort5 tương ứng cho Product2 và Dim_Category.....</i>	90
<i>Hình 104: Chọn ra các cột dữ liệu cho bảng Dim_Category.....</i>	91
<i>Hình 105: Tạo mới một Merge Join để chuẩn bị kết Product2 và Dim_Category.....</i>	92
<i>Hình 106: Thực hiện phép kết trái giữa Product2 và Dim_Category .....</i>	93
<i>Hình 107: Tạo mới một OLE DB Destination và đặt tên là Dim_Product .....</i>	94
<i>Hình 108: Nhập các câu lệnh SQL tương ứng với bảng Dim_Product.....</i>	95
<i>Hình 109: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 87 .....</i>	96
<i>Hình 110: Tạo mới một Data Flow Task cho bảng Fact_Supermarket .....</i>	97
<i>Hình 111: Tạo mới một Excel Source và một Data Conversion .....</i>	97
<i>Hình 112: Đổi kiểu dữ liệu của cột invoice_date và output alias .....</i>	98
<i>Hình 113: Tạo mới một OLE DB Source và chọn bảng Dim_Customer làm data source..</i>	99
<i>Hình 114: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 6 .....</i>	100
<i>Hình 115: Tạo 2 Sort và đặt tên lần lượt là Sort và Sort1 tương ứng cho Fact_Supermarket và Dim_Customer .....</i>	101
<i>Hình 116: Chọn ra các cột dữ liệu cho bảng Dim_Customer .....</i>	102
<i>Hình 117: Tạo mới một Merge Join để chuẩn bị kết Fact_Supermarket và Dim_Customer .....</i>	103
<i>Hình 118: Thực hiện phép kết trái giữa Fact_Supermarket và Dim_Customer .....</i>	104
<i>Hình 119: Tạo mới một OLE DB Source và chọn bảng Dim_Salesman làm data source</i>	105
<i>Hình 120: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 6 .....</i>	106
<i>Hình 121: Tạo 2 Sort và đặt tên lần lượt là Sort2 và Sort3 tương ứng cho Fact_Supermarket và Dim_Salesman.....</i>	107
<i>Hình 122: Chọn ra các cột dữ liệu cho bảng Dim_Salesman .....</i>	108
<i>Hình 123: Tạo mới một Merge Join để chuẩn bị kết Fact_Supermarket và Dim_Salesman .....</i>	109
<i>Hình 124: Thực hiện phép kết trái giữa Fact_Supermarket và Dim_Salesman.....</i>	110

<i>Hình 125: Tạo mới một OLE DB Source và chọn bảng Dim_Product làm data source..</i>	111
<i>Hình 126: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 6 .....</i>	112
<i>Hình 127: Tạo 2 Sort và đặt tên lần lượt là Sort4 và Sort5 tương ứng cho Fact_Supermarket và Dim_Product.....</i>	112
<i>Hình 128: Chọn ra các cột dữ liệu cho bảng Dim_Product.....</i>	113
<i>Hình 129:Tạo mới một Merge Join để chuẩn bị kết Fact_Supermarket và Dim_Product .....</i>	114
<i>Hình 130: Thực hiện phép kết trái giữa Fact_Supermarket và Dim_Product .....</i>	115
<i>Hình 131: Tạo mới một OLE DB Destination và đặt tên là Fact.....</i>	116
<i>Hình 132: Nhập các câu lệnh SQL tương ứng với bảng Fact_Supermarket.....</i>	117
<i>Hình 133: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 6 .....</i>	118
<i>Hình 134: Tạo mới Execute SQL Task để tạo các khóa ngoại .....</i>	119
<i>Hình 135: Tạo kết nối đến kho dữ liệu trong SQL Server .....</i>	120
<i>Hình 136: Nhập các câu lệnh tạo khóa ngoại tương ứng.....</i>	121
<i>Hình 137:Tạo mới một Execute SQL Task để đảm bảo không bị chồng chéo dữ liệu trong quá trình đổ.....</i>	122
<i>Hình 138:Tạo kết nối đến kho dữ liệu trong SQL Server .....</i>	123
<i>Hình 139: Thực hiện xóa dữ liệu cũ mỗi khi tiến hành đổ dữ liệu .....</i>	124
<i>Hình 140: Tiến hành chạy dự án SSIS.....</i>	125
<i>Hình 141: Kết quả chạy dự án SSIS lần đầu tiên .....</i>	125
<i>Hình 142: Bổ sung các câu lệnh xóa và thêm các khóa ngoại mỗi khi đổ dữ liệu .....</i>	126
<i>Hình 143: Kết quả chạy lại dự án SSIS .....</i>	126
<i>Hình 144: Dữ liệu bảng Dim_Time .....</i>	127
<i>Hình 145: Dữ liệu bảng Dim_Customer.....</i>	127
<i>Hình 146: Dữ liệu bảng Dim_Salesman.....</i>	128
<i>Hình 147: Dữ liệu bảng Dim_Branch .....</i>	129
<i>Hình 148: Dữ liệu bảng Dim_Brand .....</i>	130
<i>Hình 149: Dữ liệu bảng Dim_Category .....</i>	130
<i>Hình 150: Dữ liệu bảng Dim_Product .....</i>	131
<i>Hình 151: Dữ liệu bảng Fact_Supermarket .....</i>	131
<i>Hình 152: Lược đồ sau khi đổ dữ liệu thành công .....</i>	132
<i>Hình 153: Giao diện khởi động một project rỗng .....</i>	133
<i>Hình 154: Giao diện vào nơi download SSAS.....</i>	133
<i>Hình 155: Giao diện hướng dẫn tìm kiếm và tải SSAS.....</i>	134

<i>Hình 156: Hộp thoại download Microsoft Analysis Services Projects .....</i>	134
<i>Hình 157: Hộp thoại yêu cầu khởi động lại Visual Studio 2019 để tích hợp công cụ SSAS .....</i>	135
<i>Hình 158: Hộp thoại VSIX Installer khởi tạo .....</i>	135
<i>Hình 159: Hộp thoại yêu cầu chấp nhận điều khoản cài đặt .....</i>	136
<i>Hình 160: Hộp thoại tiếp tục hoàn thành quá trình tích hợp SSAS.....</i>	136
<i>Hình 161: Giao diện tích hợp thành công SSAS vào Visual Studio 2019 .....</i>	137
<i>Hình 162: Giao diện sau khi mở file SETUP.exe .....</i>	138
<i>Hình 163: Vị trí file SETUP.exe sau khi cài đặt SQL Server .....</i>	139
<i>Hình 164: Bắt đầu quá trình cài đặt Analysis Service .....</i>	140
<i>Hình 165: Giao diện thẻ Product Updates .....</i>	141
<i>Hình 166:Giao diện thẻ Install Rules .....</i>	142
<i>Hình 167: Giao diện thẻ Installation Type .....</i>	143
<i>Hình 168: Giao diện thẻ Product Key .....</i>	144
<i>Hình 169: Giao diện thẻ License Terms .....</i>	145
<i>Hình 170:Giao diện thẻ Feature Selection (cài đặt DES và AS).....</i>	146
<i>Hình 171: Giao diện thẻ Instance Configuration (tích hợp công cụ vào instance có sẵn) .....</i>	147
<i>Hình 172: Giao diện thẻ Server Configuration .....</i>	148
<i>Hình 173: Giao diện thẻ Database Engine Configuration.....</i>	149
<i>Hình 174:Giao diện thẻ Analysis Services Configuration .....</i>	150
<i>Hình 175: Giao diện kiểm tra toàn bộ các thiết lập trước quá trình cài đặt.....</i>	151
<i>Hình 176: Quá trình cài đặt hoàn tất và thành công .....</i>	152
<i>Hình 177: Khởi tạo một dự án mới.....</i>	153
<i>Hình 178: Giao diện hướng dẫn chọn loại dự án SSAS .....</i>	153
<i>Hình 179: Thiết lập nơi lưu trữ và đặt tên cho dự án SSAS .....</i>	154
<i>Hình 180: Giao diện tạo mới một Data Source (nguồn dữ liệu) .....</i>	155
<i>Hình 181: Hộp thoại bắt đầu quá trình tạo Data Source.....</i>	156
<i>Hình 182: Tiến hành tạo mới một kết nối đến SQL Server.....</i>	157
<i>Hình 183: Tiến hành kết nối với kho dữ liệu đã tạo trong quá trình SSIS .....</i>	158
<i>Hình 184: Tạo mới data source từ kho dữ liệu vừa kết nối .....</i>	159
<i>Hình 185: Giao diện chọn tài khoản đăng nhập .....</i>	160
<i>Hình 186: Kết quả sau khi tạo thành công data source .....</i>	161
<i>Hình 187: Giao diện tạo mới một Data Source View (khung nhìn dữ liệu) .....</i>	162

<i>Hình 188: Hộp thoại bắt đầu quá trình tạo Data Source View.....</i>	163
<i>Hình 189: Chọn data source để tạo data source view.....</i>	164
<i>Hình 190: Thêm bảng Fact vào data source view .....</i>	165
<i>Hình 191: Thêm toàn bộ các bảng chiều còn lại vào data source view .....</i>	166
<i>Hình 192: Hoàn thành quá trình tạo data source view .....</i>	167
<i>Hình 193: Kết quả data source view (lược đồ).....</i>	168
<i>Hình 194: Giao diện tạo mới một Cube (khỏi dữ liệu).....</i>	169
<i>Hình 195: Hộp thoại bắt đầu quá trình tạo Cube .....</i>	170
<i>Hình 196: Giao diện chọn phương thức khởi tạo.....</i>	171
<i>Hình 197: Giao diện phân chia các measure group.....</i>	172
<i>Hình 198: Giao diện để xuất một số độ đo trong bảng Fact.....</i>	173
<i>Hình 199: Chọn các bảng Dimensions để tạo thêm chiều cho Cubes.....</i>	174
<i>Hình 200: Giao diện hoàn tất quá trình tạo Cube.....</i>	175
<i>Hình 201: Lựa chọn các thuộc tính đưa vào Cube (Dim_Product) .....</i>	176
<i>Hình 202: Lựa chọn các thuộc tính đưa vào Cube (Dim_Customer).....</i>	176
<i>Hình 203: Lựa chọn các thuộc tính đưa vào Cube (Dim_Salesman) .....</i>	177
<i>Hình 204: Lựa chọn các thuộc tính đưa vào Cube (Dim_Time) .....</i>	178
<i>Hình 205: Hình ảnh lỗi kết nối không được thiết lập .....</i>	178
<i>Hình 206: Vị trí chứa file SQL ServerManager15.msc .....</i>	179
<i>Hình 207: Khởi động SSAS Services .....</i>	179
<i>Hình 208: Giao diện hiển thị thông tin SSAS .....</i>	180
<i>Hình 209: Kết nối vào SQL Server .....</i>	181
<i>Hình 210: Tạo login mới .....</i>	181
<i>Hình 211: Giao diện tinh chỉnh các cài đặt khi tạo login .....</i>	182
<i>Hình 212: Giao diện cấp quyền cho login mới.....</i>	182
<i>Hình 213: Trao quyền truy cập đến kho dữ liệu cho login vừa tạo.....</i>	183
<i>Hình 214: Điều chỉnh thông tin của dự án .....</i>	184
<i>Hình 215: Điện tên server dùng để kết nối vào SQL Server vào thẻ Deployment.....</i>	185
<i>Hình 216: Tiến hành chạy lại dự án lần nữa.....</i>	185
<i>Hình 217: Dự án chạy thành công.....</i>	186
<i>Hình 218: Nút hiển thị thông tin độ đo được hệ thống gợi ý.....</i>	186
<i>Hình 219: Thông tin chi tiết về các độ đo được tạo sẵn.....</i>	187
<i>Hình 220: Cập nhật lại tên và kiểu dữ liệu cho phù hợp.....</i>	187

<i>Hình 221: Tạo độ đo mới (SoLuongMuaNhiềuNhất) bằng cách sử dụng hàm Max.....</i>	188
<i>Hình 222: Thực thi lại dự án, kết quả chạy thành công .....</i>	188
<i>Hình 223: Đưa toàn bộ thuộc tính cần phân cấp của bảng Dim_Time qua cửa sổ Hierarchies .....</i>	189
<i>Hình 224: Cửa sổ sắp xếp thứ tự của các thuộc tính.....</i>	190
<i>Hình 225: Tạo lại mối quan hệ giữa các thuộc tính cho phù hợp .....</i>	190
<i>Hình 226: Chính sửa KeyColumns của thuộc tính invoice_date_second.....</i>	192
<i>Hình 227: Bổ sung các thuộc tính cấp cao hơn.....</i>	193
<i>Hình 228: Chọn lại tên thuộc tính cho phù hợp .....</i>	194
<i>Hình 229: Chính sửa KeyColumns của thuộc tính invoice_date_minute.....</i>	196
<i>Hình 230: Bổ sung các thuộc tính cấp cao hơn.....</i>	196
<i>Hình 231: Chọn lại tên thuộc tính cho phù hợp .....</i>	197
<i>Hình 232: Chính sửa KeyColumns của thuộc tính invoice_date_hour .....</i>	199
<i>Hình 233: Bổ sung các thuộc tính cấp cao hơn.....</i>	199
<i>Hình 234: Chọn lại tên thuộc tính cho phù hợp .....</i>	200
<i>Hình 235: Chính sửa KeyColumns của thuộc tính invoice_date_day .....</i>	202
<i>Hình 236: Bổ sung các thuộc tính cấp cao hơn.....</i>	202
<i>Hình 237: Chọn lại tên thuộc tính cho phù hợp .....</i>	203
<i>Hình 238: Chính sửa KeyColumns của thuộc tính invoice_date_month.....</i>	204
<i>Hình 239: Bổ sung thuộc tính cấp cao hơn .....</i>	205
<i>Hình 240: Chọn lại tên thuộc tính cho phù hợp .....</i>	206
<i>Hình 241: Đưa toàn bộ thuộc tính cần phân cấp của bảng Dim_Product qua cửa sổ Hierarchies .....</i>	207
<i>Hình 242: Cửa sổ sắp xếp thứ tự của các thuộc tính.....</i>	207
<i>Hình 243: Tạo lại mối quan hệ giữa các thuộc tính cho phù hợp .....</i>	208
<i>Hình 244: Chính sửa KeyColumns của thuộc tính Branch_Name .....</i>	210
<i>Hình 245: Bổ sung các thuộc tính cấp cao hơn.....</i>	210
<i>Hình 246: Chọn lại tên thuộc tính cho phù hợp .....</i>	211
<i>Hình 247: Chính sửa KeyColumns của thuộc tính City.....</i>	213
<i>Hình 248: Bổ sung các thuộc tính cấp cao hơn.....</i>	213
<i>Hình 249: Chọn lại tên thuộc tính cho phù hợp .....</i>	214
<i>Hình 250: Deploy lại dự án sau khi thực hiện phân cấp .....</i>	215
<i>Hình 251: Quá trình deploy thành công.....</i>	216
<i>Hình 252: Khởi tạo NameSet [Top 5 Category].....</i>	217

<i>Hình 253: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn</i>	.217
<i>Hình 254: Thêm NameSet vào ô Filter Expression và thu được kết quả câu truy vấn 1</i>	.218
<i>Hình 255: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn</i>	.218
<i>Hình 256: Đưa bảng Dim_Product và tên các nhãn hàng vào cửa sổ lọc</i>	.....219
<i>Hình 257: Đưa bảng Dim_Time bổ sung điều kiện thời gian vào cửa sổ lọc và thu được kết quả câu truy vấn 2</i>	.....219
<i>Hình 258: Khởi tạo NameSet [So Luong &gt; 15000]</i>	.....220
<i>Hình 259: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn</i>	.221
<i>Hình 260: Thêm NameSet vào ô Filter Expression và thu được kết quả câu truy vấn 3</i>	.221
<i>Hình 261: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn</i>	.222
<i>Hình 262: Đưa thuộc tính phân cấp [DiaDiemChiNhanh] và lọc ra vùng miền 'Marmara'</i>	.....222
<i>Hình 263: Đưa bảng Dim_Time vào cửa sổ lọc để bổ sung thêm điều kiện thời gian và thu được kết quả câu truy vấn 4</i>	.....223
<i>Hình 264: Khởi tạo NameSet [Top 3 Product]</i>	.....224
<i>Hình 265: Kéo thả Hierarchy [InvoiceTime] vào cửa sổ thực thi câu truy vấn</i>	.....224
<i>Hình 266: Loại bỏ một số thuộc tính không cần thiết</i>	.....225
<i>Hình 267: Tạo mới độ đo số lượng khách hàng (sử dụng aggregation là hàm DistinctCount)</i>	.....225
<i>Hình 268: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn</i>	.226
<i>Hình 269: Thêm NameSet vào ô Filter Expression và thu được kết quả câu truy vấn 5</i>	.226
<i>Hình 270: Kéo thả thuộc tính liên quan vào cửa sổ thực thi câu truy vấn</i>	.....227
<i>Hình 271: Chính sửa độ đo số lượng khách hàng (sử dụng aggregation là hàm DistinctCount)</i>	.....227
<i>Hình 272: Kéo thả độ đo liên quan và thu được kết quả câu truy vấn 6</i>	.....228
<i>Hình 273: Kéo thả Hierarchy [DiaDiemChiNhanh] vào cửa sổ thực thi câu truy vấn</i>	.....228
<i>Hình 274: Tạo mới độ đo số lượng hóa đơn (sử dụng aggregation là hàm DistinctCount)</i>	.....229
<i>Hình 275: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn</i>	.....229
<i>Hình 276: Đưa điều kiện các thành phố được nêu tên và thu được kết quả câu truy vấn 7</i>	.....230
<i>Hình 277: Tạo NameSet [Top 1 Region Revenue]</i>	.....230
<i>Hình 278: Sửa lại độ đo số lượng khách hàng (sử dụng aggregation là hàm DistinctCount)</i>	.....231
<i>Hình 279: Tạo NameSet [Filter Branch]</i>	.....231

<i>Hình 280: Đưa thuộc tính phân cấp và độ đo liên quan vào cửa sổ thực thi câu truy vấn</i>	232
<i>Hình 281: Chọn 2 NameSet vừa tạo để lọc theo điều kiện và thu được kết quả câu truy vấn 8</i>	233
<i>Hình 282: Tạo mới độ đo số lượng nhân viên (sử dụng aggregation là hàm DistinctCount)</i>	233
<i>Hình 283: Kéo thả Hierarchy [DiaDiemChiNhanh] vào cửa sổ thực thi câu truy vấn ...</i>	234
<i>Hình 284: Kéo thả Hierarchy [Invoice Time] vào cửa sổ thực thi câu truy vấn.....</i>	234
<i>Hình 285: Loại bỏ các thuộc tính không cần thiết .....</i>	235
<i>Hình 286: Kéo thả độ đo [SoLuongNhanVien] vào cửa sổ thực thi và thu được kết quả câu truy vấn 9 .....</i>	235
<i>Hình 287: Đưa thuộc tính phân cấp [DiaDiemChiNhanh] và lọc theo điều kiện tại cửa sổ lọc .....</i>	236
<i>Hình 288: Đưa độ đo [SoLuong] vào cửa sổ thực thi câu truy vấn .....</i>	236
<i>Hình 289: Đưa thuộc tính [Invoice Date Month] vào cửa sổ thực thi câu truy vấn và thu được kết quả.....</i>	237
<i>Hình 290: Giao diện chuyển sang Script Mode.....</i>	237
<i>Hình 291: Nhập câu truy vấn MDX đã chuẩn bị trước vào script .....</i>	237
<i>Hình 292: Tiến hành thực thi câu truy vấn và thu được kết quả câu truy vấn 11 .....</i>	238
<i>Hình 293: Tạo mới Measure [PHANTRAMGIAKHUYENMAI] .....</i>	238
<i>Hình 294: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn.....</i>	239
<i>Hình 295: Đưa điều kiện các tháng được yêu cầu và thu được kết quả câu truy vấn 12.239</i>	239
<i>Hình 296: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn.....</i>	240
<i>Hình 297: Giao diện chuyển sang Script Mode.....</i>	240
<i>Hình 298: Xóa đoạn code được tô đậm .....</i>	241
<i>Hình 299: Thay thế đoạn code ở bước 1 vào vị trí vừa xóa .....</i>	241
<i>Hình 300: Tiến hành thực thi câu truy vấn và thu được kết quả câu truy vấn 13 .....</i>	241
<i>Hình 301: Sửa lại độ đo số lượng hóa đơn (sử dụng aggregation là hàm DistinctCount) .....</i>	242
<i>Hình 302: Tạo NameSet [DieuKienCau14] .....</i>	242
<i>Hình 303: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn.....</i>	243
<i>Hình 304: Thêm NameSet vừa tạo vào ô Filter Expression và thu được kết quả câu truy vấn 14.....</i>	243
<i>Hình 305: Sửa lại độ đo số lượng hóa đơn (sử dụng aggregation là hàm DistinctCount) .....</i>	244
<i>Hình 306: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn.....</i>	244

<i>Hình 307: Giao diện chuyển sang Script Mode.....</i>	245
<i>Hình 308: Xóa đoạn code được tô đậm .....</i>	245
<i>Hình 309: Thay thế đoạn code ở bước 2 vào vị trí vừa xóa .....</i>	245
<i>Hình 310: Tiến hành thực thi câu truy vấn và thu được kết quả câu truy vấn 15 .....</i>	246
<i>Hình 311: Kết quả câu truy vấn 1 .....</i>	247
<i>Hình 312: Kết quả câu truy vấn 2 .....</i>	248
<i>Hình 313: Kết quả câu truy vấn 3 .....</i>	249
<i>Hình 314: Kết quả câu truy vấn 4 .....</i>	250
<i>Hình 315: Kết quả câu truy vấn 5 .....</i>	251
<i>Hình 316: Kết quả câu truy vấn 6 .....</i>	252
<i>Hình 317: Kết quả câu truy vấn 7 .....</i>	253
<i>Hình 318: Kết quả câu truy vấn 8 .....</i>	254
<i>Hình 319: Kết quả câu truy vấn 9 .....</i>	255
<i>Hình 320: Kết quả câu truy vấn 10 .....</i>	256
<i>Hình 321: Giao diện hướng dẫn vào noi kết nối SQL Server đến Excel.....</i>	262
<i>Hình 322: Hộp thoại điền thông tin Server và Database .....</i>	262
<i>Hình 323: Kết nối đến SQL Server .....</i>	263
<i>Hình 324: Giao diện chọn các bảng dữ liệu để đổ vào các sheet .....</i>	264
<i>Hình 325: Giao diện quá trình tải dữ liệu .....</i>	265
<i>Hình 326: Quá trình tải dữ liệu hoàn tất .....</i>	266
<i>Hình 327: Dữ liệu bảng Fact trong Excel .....</i>	267
<i>Hình 328: Dữ liệu bảng Dim_Time trong Excel.....</i>	267
<i>Hình 329: Dữ liệu bảng Dim_Salesman trong Excel .....</i>	268
<i>Hình 330: Dữ liệu bảng Dim_Customer trong Excel .....</i>	268
<i>Hình 331: Dữ liệu bảng Dim_Category trong Excel.....</i>	269
<i>Hình 332: Dữ liệu bảng Dim_Brand trong Excel.....</i>	269
<i>Hình 333: Dữ liệu bảng Dim_Branch trong Excel.....</i>	270
<i>Hình 334: Dữ liệu bảng Dim_Product trong Excel.....</i>	270
<i>Hình 335: Giao diện hướng dẫn vào noi tạo relationship .....</i>	271
<i>Hình 336: Tạo mới kết nối giữa các bảng .....</i>	271
<i>Hình 337: Tạo mới quan hệ giữa Dim_Product và Dim_Category .....</i>	272
<i>Hình 338: Tạo mới quan hệ giữa Dim_Product và Dim_Brand .....</i>	272
<i>Hình 339: Tạo mới quan hệ giữa Dim_Product và Dim_Branch.....</i>	272

<i>Hình 340: Tạo mối quan hệ giữa Fact và Dim_Product</i> .....	273
<i>Hình 341: Tạo mối quan hệ giữa Fact và Dim_Customer</i> .....	273
<i>Hình 342: Tạo mối quan hệ giữa Fact và Dim_Salesman</i> .....	273
<i>Hình 343: Tạo mối quan hệ giữa Fact và Dim_Time</i> .....	274
<i>Hình 344: Giao diện kiểm tra các mối quan hệ vừa thiết lập</i> .....	274
<i>Hình 345: Giao diện hướng dẫn vào nơi mở Power Pivot</i> .....	275
<i>Hình 346: Hộp thoại Bổ trợ COM</i> .....	275
<i>Hình 347: Giao diện hướng dẫn vào nơi xem lược đồ</i> .....	276
<i>Hình 348: Lược đồ sau khi hoàn tất quá trình đó dữ liệu vào tạo các mối quan hệ</i> .....	276
<i>Hình 349: Giao diện hướng dẫn vào nơi tạo kết nối với nơi chứa dữ liệu ngoài</i> .....	276
<i>Hình 350: Hộp thoại Pivot Table xuất hiện khởi tạo kết nối mới</i> .....	277
<i>Hình 351: Chọn mô hình dữ liệu số làm việc</i> .....	278
<i>Hình 352: Kiểm tra lại tên kết nối vừa thiết lập</i> .....	279
<i>Hình 353: Môi trường đã được thiết lập sẵn sàng</i> .....	279
<i>Hình 354: Kéo thả các thuộc tính thích hợp vào trường “Cột” và “Hàng” của Pivot Table</i> .....	280
<i>Hình 355: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán</i> ...	281
<i>Hình 356: Chọn lọc ra top 10 doanh số theo kho sản phẩm</i> .....	282
<i>Hình 357: Chính sửa lại chỉ lấy top 5 theo yêu cầu của câu truy vấn 1</i> .....	282
<i>Hình 358: Kết quả hiển thị câu truy vấn 1 dưới dạng matrix</i> .....	283
<i>Hình 359: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ</i> .....	283
<i>Hình 360: Giao diện chọn loại biểu đồ</i> .....	284
<i>Hình 361: Điều chỉnh lại chú giải, trực và giá trị</i> .....	285
<i>Hình 362: Kết quả câu truy vấn 1 hiển thị dưới dạng biểu đồ</i> .....	285
<i>Hình 363: Kéo thả các thuộc tính thích hợp vào trường “Cột” và “Hàng” của Pivot Table</i> .....	286
<i>Hình 364: Lọc ra tên các thương hiệu sản phẩm được đề cập trong câu truy vấn</i> .....	287
<i>Hình 365: Tiến hành chọn các kết quả nằm trong tháng 2</i> .....	288
<i>Hình 366: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán</i> ...	289
<i>Hình 367: Kết quả câu truy vấn 2 hiển thị dưới dạng table</i> .....	289
<i>Hình 368: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ</i> .....	289
<i>Hình 369: Giao diện chọn loại biểu đồ</i> .....	290
<i>Hình 370: Kết quả câu truy vấn 2 hiển thị dưới dạng biểu đồ</i> .....	291
<i>Hình 371: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table</i> .....	292

<i>Hình 372: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán ...</i>	292
<i>Hình 373: Lọc các dòng dữ liệu theo điều kiện lớn hơn.....</i>	293
<i>Hình 374: Thiết lập giá trị điều kiện của câu truy vấn.....</i>	293
<i>Hình 375: Kết quả câu truy vấn 3 hiển thị dưới dạng table.....</i>	294
<i>Hình 376: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ.....</i>	294
<i>Hình 377: Giao diện chọn loại biểu đồ .....</i>	295
<i>Hình 378: Kết quả câu truy vấn 3 hiển thị dưới dạng biểu đồ .....</i>	296
<i>Hình 379: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table .....</i>	296
<i>Hình 380: Thêm thuộc tính region vào bộ lọc.....</i>	297
<i>Hình 381: Thêm thuộc tính invoice_date_month vào bộ lọc .....</i>	298
<i>Hình 382: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán ...</i>	299
<i>Hình 383: Lọc dữ liệu theo vùng miền Marmara .....</i>	299
<i>Hình 384: Lọc dữ liệu theo tháng 04.....</i>	300
<i>Hình 385: Kết quả câu truy vấn 4 hiển thị dưới dạng table.....</i>	301
<i>Hình 386: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ.....</i>	301
<i>Hình 387: Giao diện chọn loại biểu đồ .....</i>	302
<i>Hình 388: Kết quả câu truy vấn 4 hiển thị dưới dạng biểu đồ .....</i>	303
<i>Hình 389: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table .....</i>	303
<i>Hình 390: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán ...</i>	304
<i>Hình 391: Chọn lọc ra top 10 sản phẩm theo doanh số.....</i>	305
<i>Hình 392: Chính sửa lại chỉ lấy top 3 theo yêu cầu của câu truy vấn 5 .....</i>	305
<i>Hình 393: Kết quả câu truy vấn 5 hiển thị dưới dạng table (rút gọn).....</i>	306
<i>Hình 394: Kết quả câu truy vấn 5 hiển thị dưới dạng table (đầy đủ).....</i>	306
<i>Hình 395: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ.....</i>	306
<i>Hình 396: Giao diện chọn loại biểu đồ .....</i>	307
<i>Hình 397: Kết quả câu truy vấn 5 hiển thị dưới dạng biểu đồ (đầy đủ).....</i>	308
<i>Hình 398: Kết quả câu truy vấn 5 hiển thị dưới dạng biểu đồ (rút gọn).....</i>	308
<i>Hình 399: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table .....</i>	309
<i>Hình 400: Kéo thả thuộc tính amount vào trường “Giá trị” và chọn hàm tính toán.....</i>	309
<i>Hình 401: Kéo thả thuộc tính customer_id vào trường “Giá trị” và chọn hàm tính toán .....</i>	310
<i>Hình 402: Kết quả câu truy vấn 6 dưới dạng table .....</i>	311
<i>Hình 403: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ .....</i>	311

<i>Hình 404: Giao diện chọn loại biểu đồ</i> .....	312
<i>Hình 405: Kết quả câu truy vấn 6 hiển thị dưới dạng biểu đồ</i> .....	313
<i>Hình 406: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table</i> .....	313
<i>Hình 407: Kéo thả thuộc tính customer_id vào trường “Giá trị” và chọn hàm tính toán</i> .....	314
<i>Hình 408: Lọc ra tên các thành phố được đề cập trong câu truy vấn</i> .....	315
<i>Hình 409: Kết quả câu truy vấn 7 hiển thị dưới dạng table</i> .....	316
<i>Hình 410: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ</i> .....	316
<i>Hình 411: Giao diện chọn loại biểu đồ</i> .....	317
<i>Hình 412: Kết quả câu truy vấn 7 hiển thị dưới dạng biểu đồ</i> .....	318
<i>Hình 413: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table</i> .....	319
<i>Hình 414: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán</i> ..	319
<i>Hình 415: Lọc ra top 10 vùng miền có doanh thu cao nhất</i> .....	320
<i>Hình 416: Lấy ra top 1 vùng miền có doanh thu cao nhất</i> .....	321
<i>Hình 417: Lọc các dòng dữ liệu theo điều kiện nhỏ hơn</i> .....	322
<i>Hình 418: Thiết lập giá trị điều kiện của câu truy vấn</i> .....	322
<i>Hình 419: Kết quả câu truy vấn 8 hiển thị dưới dạng table</i> .....	323
<i>Hình 420: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ</i> .....	323
<i>Hình 421: Giao diện chọn loại biểu đồ</i> .....	324
<i>Hình 422: Kết quả câu truy vấn 8 hiển thị dưới dạng biểu đồ</i> .....	324
<i>Hình 423: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table</i> .....	325
<i>Hình 424: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán</i> ..	326
<i>Hình 425: Kết quả câu truy vấn 9 hiển thị dưới dạng table</i> .....	327
<i>Hình 426: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ</i> .....	327
<i>Hình 427: Giao diện chọn loại biểu đồ</i> .....	328
<i>Hình 428: Kết quả câu truy vấn 9 hiển thị dưới dạng biểu đồ</i> .....	328
<i>Hình 429: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table</i> .....	329
<i>Hình 430: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán</i> ..	330
<i>Hình 431: Lọc dữ liệu theo vùng miền Güneydoğu Anadolu</i> .....	331
<i>Hình 432: Kết quả câu truy vấn 10 hiển thị dưới dạng table</i> .....	332
<i>Hình 433: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ</i> .....	332
<i>Hình 434: Giao diện chọn loại biểu đồ</i> .....	333
<i>Hình 435: Kết quả câu truy vấn 10 hiển thị dưới dạng biểu đồ</i> .....	334

<i>Hình 436: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table .....</i>	335
<i>Hình 437: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán ...</i>	336
<i>Hình 438: Lọc dữ liệu theo customer_id .....</i>	337
<i>Hình 439: Chọn top 1 khách hàng theo final_total_trans.....</i>	337
<i>Hình 440: Kết quả câu truy vấn 11 hiển thị dưới dạng table.....</i>	338
<i>Hình 441: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ.....</i>	338
<i>Hình 442: Giao diện chọn loại biểu đồ .....</i>	339
<i>Hình 443: Kết quả câu truy vấn 11 hiển thị dưới dạng biểu đồ .....</i>	340
<i>Hình 444: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table .....</i>	341
<i>Hình 445: Thực hiện tạo phép đo mới.....</i>	341
<i>Hình 446: Tạo độ đo Phản Trăm Giảm Giá.....</i>	341
<i>Hình 447: Kéo thả thuộc tính vào mục “Giá trị” .....</i>	342
<i>Hình 448: Lọc ra tháng 1 và tháng 4 .....</i>	343
<i>Hình 449: Kết quả câu truy vấn 12 hiển thị dưới dạng table.....</i>	344
<i>Hình 450: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ.....</i>	344
<i>Hình 451: Giao diện chọn loại biểu đồ .....</i>	345
<i>Hình 452: Kết quả câu truy vấn 12 hiển thị dưới dạng biểu đồ .....</i>	345
<i>Hình 453: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table .....</i>	346
<i>Hình 454: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán ...</i>	347
<i>Hình 455: Tiến hành lọc ra sản phẩm có số lượng mua lớn hơn 10 .....</i>	347
<i>Hình 456: Tiến hành lọc ra sản phẩm có số lượng mua lớn hơn 10 .....</i>	348
<i>Hình 457: Kết quả câu truy vấn 13 hiển thị dưới dạng table.....</i>	348
<i>Hình 458: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ.....</i>	348
<i>Hình 459: Giao diện chọn loại biểu đồ .....</i>	349
<i>Hình 460: Kết quả câu truy vấn 13 hiển thị dưới dạng biểu đồ .....</i>	349
<i>Hình 461: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table .....</i>	350
<i>Hình 462: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán ...</i>	351
<i>Hình 463: Tiến hành lọc ra vùng miền ngoại trừ 2 vùng miền ‘Doğu Anadolu’, ‘Karadeniz’ .....</i>	351
<i>Hình 464: Kết quả câu truy vấn 14 hiển thị dưới dạng table.....</i>	352
<i>Hình 465: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ.....</i>	352
<i>Hình 466: Giao diện chọn loại biểu đồ .....</i>	353
<i>Hình 467: Kết quả câu truy vấn 14 hiển thị dưới dạng biểu đồ .....</i>	354

<i>Hình 468: Giao diện hướng dẫn kết nối Excel với SQL Server.....</i>	354
<i>Hình 469: Giao diện nhập tên Server và tên cơ sở dữ liệu (warehouse) lưu trong SQL Server.....</i>	355
<i>Hình 470: Giao diện đăng nhập tài khoản để kết nối tới SQL Server.....</i>	355
<i>Hình 471: Chọn tất cả các bảng dữ liệu để đổ vào Excel .....</i>	356
<i>Hình 472: Giao diện quá trình load dữ liệu của Power BI .....</i>	357
<i>Hình 473: Giao diện hướng dẫn vào xem lược đồ cơ sở dữ liệu trong Excel .....</i>	358
<i>Hình 474: Lược đồ mối quan hệ giữa các table .....</i>	359
<i>Hình 475: Giao diện tạo mẫu báo cáo trong Power BI .....</i>	360
<i>Hình 476: Kéo các thuộc tính và độ đo thích hợp vào mục “Hàng”, “Cột” và “Giá trị” .....</i>	361
<i>Hình 477: Lọc ra điều kiện top 5 kho sản phẩm theo doanh số .....</i>	362
<i>Hình 478: Kết quả câu truy vấn 1 hiển thị dưới dạng ma trận .....</i>	362
<i>Hình 479: Tinh chỉnh cột X, cột Y và Legend của biểu đồ .....</i>	363
<i>Hình 480: Kết quả câu truy vấn 1 hiển thị dưới dạng biểu đồ .....</i>	363
<i>Hình 481: Report câu truy vấn 1 .....</i>	364
<i>Hình 482: Giao diện tạo mẫu báo cáo trong Power BI .....</i>	365
<i>Hình 483: Kéo các thuộc tính và độ đo thích hợp vào mục “Hàng”, “Cột” và “Giá trị” .....</i>	366
<i>Hình 484: Lọc ra doanh số tháng 02 của các nhãn hàng trong câu truy vấn .....</i>	367
<i>Hình 485: Kết quả câu truy vấn 2 hiển thị dưới dạng table .....</i>	367
<i>Hình 486: Chọn loại biểu đồ, tinh chỉnh cột X, cột Y và Legend của biểu đồ .....</i>	368
<i>Hình 487: Kết quả câu truy vấn 2 hiển thị dưới dạng biểu đồ .....</i>	369
<i>Hình 488: Report câu truy vấn 2 .....</i>	370
<i>Hình 489: Giao diện tạo mẫu báo cáo trong Power BI .....</i>	371
<i>Hình 490: các thuộc tính và độ đo thích hợp vào mục “Hàng”, “Cột” và “Giá trị” .....</i>	372
<i>Hình 491: Lọc ra chi nhánh có doanh số lớn hơn 15000.....</i>	373
<i>Hình 492: Kết quả câu truy vấn 3 hiển thị dưới dạng table .....</i>	374
<i>Hình 493: Chọn loại biểu đồ, tinh chỉnh cột X, cột Y và Legend của biểu đồ .....</i>	375
<i>Hình 494: Kết quả câu truy vấn 3 hiển thị dưới dạng biểu đồ .....</i>	376
<i>Hình 495: Report câu truy vấn 3 .....</i>	377
<i>Hình 496: Khởi tạo một dự án mới.....</i>	378
<i>Hình 497: Giao diện hướng dẫn chọn loại dự án SSRS .....</i>	378
<i>Hình 498: Thiết lập nơi lưu trữ và đặt tên cho dự án SSRS .....</i>	379

<i>Hình 499: Thêm nguồn dữ liệu cho Report</i> .....	380
<i>Hình 500: Giao diện tạo một kết nối mới đến kho dữ liệu</i> .....	381
<i>Hình 501: Tạo mới một Dataset có tên là MDXI</i> .....	382
<i>Hình 502: Giao diện tạo mới một Report</i> .....	383
<i>Hình 503: Hộp thoại Report Wizard xuất hiện</i> .....	384
<i>Hình 504: Giao diện chọn nguồn dữ liệu cho Report</i> .....	385
<i>Hình 505: Dán câu truy vấn MDX của câu truy vấn 1 vào ô “Query Designer”</i> .....	386
<i>Hình 506: Kiểm tra lại câu truy vấn</i> .....	387
<i>Hình 507: Giao diện chọn loại Report</i> .....	388
<i>Hình 508: Chuyển các thuộc tính thích hợp vào mục “Columns”, “Rows” và “Details”</i> .....	389
<i>Hình 509: Giao diện đặt tên cho Report</i> .....	390
<i>Hình 510: Kết quả câu truy vấn 1 hiển thị dưới dạng ma trận</i> .....	391
<i>Hình 511: Kết quả đổi tên Report</i> .....	391
<i>Hình 512: Tiến hành tạo thêm Totals có tên là TongSoLuong</i> .....	392
<i>Hình 513: Report câu truy vấn 1</i> .....	392
<i>Hình 514: Giao diện tạo mới một Report</i> .....	393
<i>Hình 515: Hộp thoại Report Wizard xuất hiện</i> .....	394
<i>Hình 516: Giao diện chọn nguồn dữ liệu cho Report</i> .....	395
<i>Hình 517: Dán câu truy vấn MDX của câu truy vấn số 2 vào ô “Query Designer”</i> .....	396
<i>Hình 518: Giao diện chọn loại Report</i> .....	397
<i>Hình 519: Chuyển các thuộc tính thích hợp vào mục “Group” và “Details”</i> .....	398
<i>Hình 520: Giao diện chọn Table Layout</i> .....	399
<i>Hình 521: Giao diện đặt tên cho Report</i> .....	400
<i>Hình 522: Report câu truy vấn 2</i> .....	400
<i>Hình 523: Giao diện tạo mới một Report</i> .....	401
<i>Hình 524: Hộp thoại Report Wizard xuất hiện</i> .....	402
<i>Hình 525: Giao diện chọn nguồn dữ liệu cho Report</i> .....	403
<i>Hình 526: Dán câu truy vấn MDX của câu truy vấn 3 vào ô “Query Designer”</i> .....	404
<i>Hình 527: Giao diện chọn loại Report</i> .....	405
<i>Hình 528: Chuyển các thuộc tính thích hợp vào mục “Group” và “Details”</i> .....	406
<i>Hình 529: Giao diện chọn Table Layout</i> .....	407
<i>Hình 530: Giao diện đặt tên cho Report</i> .....	408

<i>Hình 531: Report câu truy vấn 3 .....</i>	409
<i>Hình 532: Giao diện tạo mới một Report .....</i>	410
<i>Hình 533: Hộp thoại Report Wizard xuất hiện.....</i>	411
<i>Hình 534: Giao diện chọn nguồn dữ liệu cho Report.....</i>	412
<i>Hình 535: Dán câu truy vấn MDX của câu truy vấn 7 vào ô “Query Designer”.....</i>	413
<i>Hình 536: Giao diện chọn loại Report .....</i>	414
<i>Hình 537: Chuyển các thuộc tính thích hợp vào mục “Columns”, “Rows” và “Details” .....</i>	415
<i>Hình 538: Giao diện đặt tên cho Report.....</i>	416
<i>Hình 539: Kết quả câu truy vấn 7 dưới dạng ma trận.....</i>	416
<i>Hình 540: Tiến hành tạo thêm Totals có tên là TongSoLuongHoaDon .....</i>	417
<i>Hình 541: Report cho câu truy vấn 7 .....</i>	417
<i>Hình 542: Khởi tạo cơ sở dữ liệu rỗng.....</i>	420
<i>Hình 543: Tạo các bảng dữ liệu .....</i>	420
<i>Hình 544: Tiến hành đổ dữ liệu từ file Excel bằng công cụ SSIS.....</i>	421
<i>Hình 545: Tạo kết nối đến file Excel .....</i>	422
<i>Hình 546: Kiểm tra lại các cột thuộc tính .....</i>	423
<i>Hình 547: Giao diện thêm Datasource và Database vừa tạo.....</i>	424
<i>Hình 548: Tạo mới một Data Source View.....</i>	425
<i>Hình 549: Tạo mới Mining Structures.....</i>	425
<i>Hình 550: Chọn thuật toán Mining (thuật toán Cây quyết định) .....</i>	426
<i>Hình 551: Chọn khung nhìn dữ liệu nguồn để chạy thuật toán.....</i>	427
<i>Hình 552: Lựa chọn thuộc tính khóa và thuộc tính dự đoán(mục tiêu).....</i>	428
<i>Hình 553: Một số thuộc tính được hệ thống gọi ý làm input.....</i>	429
<i>Hình 554: Lựa chọn thuộc tính làm input.....</i>	430
<i>Hình 555: Detect các trường dữ liệu.....</i>	431
<i>Hình 556: Phân chia bộ dữ liệu thành training set và test set .....</i>	432
<i>Hình 557: Đặt tên cho Model .....</i>	433
<i>Hình 558: Thiết lập giá trị Holdoutseed.....</i>	435
<i>Hình 559; Tạo mới một mô hình datamining khác.....</i>	435
<i>Hình 560: Chọn thuật toán Mining (thuật toán phân cụm) .....</i>	435
<i>Hình 561: Chọn thuật toán Mining (thuật toán Naïve Bayes).....</i>	436
<i>Hình 562: Khởi động quá trình Mining.....</i>	436

<i>Hình 563: Tỷ lệ đánh giá của khách hàng ban đầu</i> .....	437
<i>Hình 564: Tỷ lệ đánh giá của khách hàng dựa trên “Online Boarding” và “Inflight Wifi Service”</i> .....	438
<i>Hình 565: Tỷ lệ đánh giá của khách hàng dựa trên “Online Boarding” và “Type of Travel”</i> .....	438
<i>Hình 566: Tỷ lệ đánh giá của khách hàng dựa trên “Online Boarding”, “Type of Travel”, “Baggage Handling”, “Inflight Wifi Service”, “Gate Location” và “Flight Distance”</i> .	439
<i>Hình 567: Tỷ lệ đánh giá dựa trên “Online Boarding”, “Inflight Wifi Service”, “Gate Location”, “Class” và “Age”</i> .....	439
<i>Hình 568: Sơ đồ những thuộc tính ảnh hưởng đến kết quả đánh giá của khách hàng</i> ....	440
<i>Hình 569: Sơ đồ Cluster của thuật toán</i> .....	441
<i>Hình 570: Đổi tên các cluster có độ hài lòng cao nhất và thấp nhất thành “High” và “Low”</i> .....	442
<i>Hình 571: Tập lục và kết quả của cluster có độ hài lòng cao nhất</i> .....	443
<i>Hình 572: Tập lục và kết quả của cluster có đánh giá trung lập hoặc không hài lòng cao nhất</i> .....	443
<i>Hình 573: Mối tương quan giữa hai cluster “High” và “Low”</i> .....	444
<i>Hình 574: Sơ đồ những thuộc tính ảnh hưởng đến kết quả đánh giá của khách hàng</i> ....	445
<i>Hình 575: Giá trị của thuộc tính “Class” ảnh hưởng đến kết quả đánh giá của hành khách</i> .....	445
<i>Hình 576: Một số thuộc tính còn lại ảnh hưởng đến kết quả đánh giá của hành khách.</i> ..	446
<i>Hình 577: Thiết lập giá trị thông số Input Selection</i> .....	447
<i>Hình 578: So sánh kết quả dự đoán của ba thuật toán thông qua line chart</i> .....	448
<i>Hình 579: Thiết lập giá trị thông số Input Selection</i> .....	449
<i>Hình 580: So sánh kết quả dự đoán của ba thuật toán thông qua line chart</i> .....	450
<i>Hình 581: Giao diện Mining Model Prediction</i> .....	451
<i>Hình 582: Kiểm tra ánh xạ các thuộc tính</i> .....	451
<i>Hình 583: Giao diện cửa sổ Modify Mapping</i> .....	452
<i>Hình 584: Chọn mục Design để kéo thả các thuộc tính và mô hình phục vụ cho việc dự đoán</i> .....	452
<i>Hình 585:Lựa chọn các thuộc tính và trường tính toán thích hợp</i> .....	453
<i>Hình 586:Thêm điều kiện giá trị cho biến satisfaction</i> .....	453
<i>Hình 587: Thêm các thuộc tính phục vụ cho câu truy vấn phải có ít nhất cột khóa chính</i> .....	453
<i>Hình 588: Chọn mục Query để kiểm tra câu truy vấn</i> .....	453
<i>Hình 589: Kiểm tra câu truy vấn đã đúng hay chưa</i> .....	454

<i>Hình 590: Chọn mục “Result” để kiểm tra kết quả dự đoán .....</i>	454
<i>Hình 591: Kết quả dự đoán của mô hình Cây quyết định .....</i>	455
<i>Hình 592: Lưu kết quả dự đoán.....</i>	455
<i>Hình 593: Thêm dữ liệu và các thư viện cần thiết.....</i>	456
<i>Hình 594: Lọc ra các cột dữ liệu kiểu số.....</i>	456
<i>Hình 595: Thống kê sơ bộ một số giá trị cơ bản của các cột kiểu số.....</i>	457
<i>Hình 596: Kết quả thống kê mô tả dữ liệu.....</i>	457
<i>Hình 597: Trực quan hóa dữ liệu cột Class .....</i>	458
<i>Hình 598: Trực quan hóa trung bình thời gian trễ của hai loại hình di chuyển .....</i>	459
<i>Hình 599: Ma trận tương quan giữa các thuộc tính trong bộ dữ liệu.....</i>	460
<i>Hình 600: Chọn ra các cột thuộc tính để tiến hành quá trình Mining .....</i>	461
<i>Hình 601: Bộ dữ liệu sau khi xóa bỏ các dòng dữ liệu trùng.....</i>	462
<i>Hình 602: Mã hóa các cột thuộc tính phân loại.....</i>	462
<i>Hình 603: Dữ liệu các cột phân loại sau khi được mã hóa.....</i>	463
<i>Hình 604: Chuẩn hóa các cột dữ liệu kiểu số.....</i>	463
<i>Hình 605: Tạo từ điển lưu trữ kết quả các độ đo đánh giá mô hình .....</i>	464
<i>Hình 606: Phân chia bộ dữ liệu thành hai tập (test set và validate &amp; training set) .....</i>	464
<i>Hình 607: Tiếp tục phân chia thành validate set và training set.....</i>	465
<i>Hình 608: Cây quyết định khách hàng có phản hồi email hay không.....</i>	466
<i>Hình 609: Tìm kiếm bộ tham số thích hợp nhất với mô hình Cây quyết định .....</i>	467
<i>Hình 610: Điểm đánh giá hiệu suất mô hình.....</i>	468
<i>Hình 611: DataFrame kết quả tìm kiếm dạng lướt theo thứ tự tăng dần của điểm xác thực .....</i>	468
<i>Hình 612: Ma trận nhầm lẫn khi dự đoán tập validate bằng mô hình Cây quyết định....</i>	469
<i>Hình 613: 10 thuộc tính quan trọng ảnh hưởng đến biến satisfaction (mô hình Cây quyết định) .....</i>	470
<i>Hình 614: Trực quan hóa cây quyết định .....</i>	470
<i>Hình 615: Minh họa cấu trúc rìng ngẫu nhiên.....</i>	471
<i>Hình 616: Tìm kiếm bộ tham số thích hợp nhất với mô hình rùng ngẫu nhiên .....</i>	472
<i>Hình 617: Điểm đánh giá hiệu suất mô hình.....</i>	473
<i>Hình 618: DataFrame kết quả tìm kiếm dạng lướt theo thứ tự tăng dần của điểm xác thực .....</i>	473
<i>Hình 619: Ma trận nhầm lẫn khi dự đoán tập validate bằng mô hình Rùng ngẫu nhiên</i>	474

<i>Hình 620: 15 thuộc tính quan trọng ảnh hưởng đến biến satisfaction (mô hình Rừng ngẫu nhiên) .....</i>	475
<i>Hình 621: Trục quan hóa cây quyết định đầu tiên trong rừng ngẫu nhiên.....</i>	475
<i>Hình 622: Đường cong ROC so sánh hiệu suất giữa hai mô hình.....</i>	477

## LỜI CẢM ƠN

Lời đầu tiên, nhóm chúng em xin gửi lời cảm ơn sâu sắc đến quý Thầy Cô trường Đại học Công nghệ thông tin, đặc biệt là quý thầy cô thuộc Khoa Hệ thống thông tin, đã tạo cơ hội cho chúng em tiếp thu những kiến thức quý báu trong khoảng thời gian học tập môn Kho dữ liệu và OLAP. Những kiến thức được quý Thầy cô truyền đạt trong suốt thời gian qua là một bước đệm quan trọng để giúp chúng em hoàn thành đồ án môn học một cách chỉnh chu nhất.

Chúng em cũng xin gửi lời cảm ơn sâu sắc tới ThS. Đỗ Thị Minh Phụng - giảng viên đứng lớp môn Kho dữ liệu và OLAP đã tận tình giúp đỡ, hướng dẫn và giải đáp các thắc mắc của nhóm trong suốt quá trình thực hiện đồ án môn học. Qua những lời chỉ dẫn nhiệt tình của cô, chúng em đã tiếp thu được nhiều kiến thức bổ ích cũng như trau dồi những kỹ năng cần thiết để đồ án được hoàn thiện một cách nhanh chóng và bảo đảm được chuẩn đầu ra của môn học.

Trong suốt quá trình thực hiện đồ án, nhóm đã nhận được rất nhiều sự hỗ trợ, đóng góp từ cô thông qua việc nhận xét đồ án của các bạn cùng lớp. Sự giúp đỡ ấy là một yếu tố quan trọng giúp bài báo cáo của nhóm chúng em hoàn thiện một cách tốt nhất và thông qua những kinh nghiệm mà cô nhận xét các báo cáo của nhóm khác nhưng yếu tố quan trọng hơn cả là sự nỗ lực của các thành viên trong nhóm. Chúng em đã vận dụng những kiến thức được thầy giảng dạy trên lớp cũng như kết hợp với việc học hỏi và tìm hiểu thêm những kiến thức mới thông qua tài liệu trên internet, giáo trình để đảm bảo lượng kiến thức tối thiểu trong đồ án. Tuy nhiên, trong quá trình thực hiện đồ án, nhóm có thể vẫn còn một vài sai sót không tránh khỏi. Do đó, nhóm rất mong nhận được những ý kiến đóng góp quý giá từ cô để có thể rút ra những bài học và biến nó trở thành hành trang, kinh nghiệm quý báu để nhóm thực hiện các đề tài tiếp theo trong tương lai.

Một lần nữa xin gửi đến quý thầy cô lời cảm ơn chân thành và tốt đẹp nhất.

## 1. Tổng quan về đề tài:

### 1.1. Lý do chọn đề tài:

Mua sắm được xem là nhu cầu cơ bản và quan trọng trong cuộc sống hằng ngày giúp bảo đảm các hoạt động diễn ra một cách trơn tru và hiệu quả. Mọi thứ chúng ta sử dụng đều được bày bán khắp trên các hệ thống siêu thị, các ngôi chợ truyền thống hoặc các sàn thương mại điện tử. Trong xã hội hiện đại, công nghệ phát triển, mọi thứ đều được số hóa và tích hợp chỉ trên một thiết bị thông minh giúp cho việc mua sắm diễn ra ngày một sôi động và tiết kiệm được rất nhiều thời gian của cả khách hàng và nhà phân phối sản phẩm. Một số chức năng có thể kể đến như thêm vào giỏ hàng, thanh toán trực tuyến, xem xét đánh giá của các khách hàng đã từng mua sắm tại đây.

Để có thể tăng doanh số bán hàng, các nhà kinh doanh cần phải nắm bắt được tâm lý và sở thích của các khách hàng thông qua việc tổng hợp và phân tích dữ liệu của các khách hàng đã ghé thăm cửa hàng trước đó. Từ đó, họ đưa ra các chiến lược quảng cáo và chương trình khuyến mãi phù hợp đối với từng sản phẩm đến từng đối tượng khách hàng. Có như vậy, lợi nhuận của cửa hàng mới gia tăng nhanh chóng và thu hút thêm được các khách hàng tiềm năng.

Nhận thức được vai trò và tầm quan trọng của công việc này, nhóm chúng em quyết định vận dụng những kiến thức tích góp được từ môn học IS217 để xây dựng kho dữ liệu và phân tích dữ liệu của một cửa hàng tại Thổ Nhĩ Kỳ. Nhóm hy vọng đồ án này không chỉ mang có thể trở thành một nguồn tài liệu tham khảo hữu ích cho các bạn sinh viên có đam mê với ngành khoa học dữ liệu nói chung và phân tích dữ liệu kinh doanh nói riêng.

### 1.2. Mô tả bộ dữ liệu:

Bộ dữ liệu được nhóm chúng em thu thập tại website [Kaggle](#) – một trong những website chứa nhiều bộ dữ liệu tham khảo hữu ích về nhiều đề tài, lĩnh vực khác nhau trong cuộc sống. Bộ dữ liệu bao gồm 26 cột thuộc tính trong đó có 3 cột dữ liệu ngày tháng, 6 cột dữ liệu kiểu số còn lại là các thuộc tính kiểu phân loại.

Bộ dữ liệu lưu trữ toàn bộ thông tin liên quan tới hoạt động kinh doanh của một siêu thị tại Thổ Nhĩ Kỳ trong quý I năm 2017 bao gồm: thông tin khách hàng, thông tin sản

phẩm, vị trí địa lý, thông tin về các thương hiệu, nhãn hàng tham gia mua bán trong khoảng thời gian này.

Link dataset: [Dữ liệu siêu thị 2017](#)

```
In [1]: 1 import pandas as pd  
2 import numpy as np  
3  
In [2]: 1 origin_data = pd.read_excel("MarketSales.xlsx")  
In [3]: 1 origin_data  
Out[3]:

|        | ID       | ITEMCODE | ITEMNAME                        | FICHENO | DATE_      | AMOUNT | PRICE | LINENETTOTAL | LINENET | BRANCHNR | ... | CLIENTCODE | CLIENTNAME       | BR  |
|--------|----------|----------|---------------------------------|---------|------------|--------|-------|--------------|---------|----------|-----|------------|------------------|-----|
| 0      | 11738.0  | 5863.0   | SPRITE 1 LT LIMON AROMALI GAZOZ | 18456   | 2017-01-07 | 1.0    | 2.00  | 2.00         | 1.85    | 52.0     | ... | 467369     | Sercan KIZLOK    |     |
| 1      | 10537.0  | 8.0      | TOZ SEKER                       | 18105   | 2017-01-06 | 5.0    | 2.65  | 13.25        | 12.27   | 8.0      | ... | 131464     | İsmet ŞİNGİR     |     |
| 2      | 11335.0  | 5979.0   | FALIM SAKIZ 5Lİ NANE            | 18350   | 2017-01-03 | 1.0    | 0.40  | 0.40         | 0.37    | 40.0     | ... | 656969     | Yağız KUBAL      |     |
| 3      | 11336.0  | 5979.0   | FALIM SAKIZ 5Lİ NANE            | 18350   | 2017-01-03 | 1.0    | 0.40  | 0.40         | 0.37    | 40.0     | ... | 656969     | Yağız KUBAL      |     |
| 4      | 10115.0  | 12808.0  | FALIM SAKIZ 5Lİ CILEK           | 18005   | 2017-01-05 | 1.0    | 0.40  | 0.40         | 0.37    | 41.0     | ... | 447336     | Habibe AYSAN     |     |
| ...    | ...      | ...      | ...                             | ...     | ...        | ...    | ...   | ...          | ...     | ...      | ... | ...        | ...              | ... |
| 611103 | 511703.0 | 7.0      | EKMEK 250 GR                    | 134656  | 2017-03-18 | 1.0    | 0.90  | 0.90         | 0.89    | 66.0     | ... | 374191     | Mahmut BAYATGİL  |     |
| 611104 | 511496.0 | 7.0      | EKMEK 250 GR                    | 134630  | 2017-03-18 | 1.0    | 0.90  | 0.90         | 0.89    | 7.0      | ... | 656248     | Gamze BOTUÇ      |     |
| 611105 | 492612.0 | 7.0      | EKMEK 250 GR                    | 130100  | 2017-03-15 | 1.0    | 0.90  | 0.90         | 0.89    | 59.0     | ... | 78643      | Cengiz BEREKETÇİ |     |
| 611106 | 512343.0 | 7.0      | EKMEK 250 GR                    | 134762  | 2017-03-18 | 1.0    | 0.90  | 0.90         | 0.88    | 40.0     | ... | 926041     | Elifnur EĞINKAYA |     |
| 611107 | 512179.0 | 3942.0   | ETİ 80 GR KARAM BITTER          | 134738  | 2017-03-18 | 1.0    | 1.90  | 1.90         | 1.76    | 74.0     | ... | 241748     | Hacer CANDEMİR   |     |



611108 rows x 26 columns


```

Hình 1: Dữ liệu gốc được hiển thị trong Jupyter Notebook

### **1.3. Mô tả chi tiết các thuộc tính:**

STT	Tên Thuộc Tính	Kiểu dữ liệu	Mô tả
1	ID	String	Mã định danh cho một giao dịch
2	ITEMCODE	Float	Mã định danh sản phẩm
3	ITEMNAME	String	Tên sản phẩm
4	FICHENO	Int	Số hoá đơn
5	DATE_	Datetime	Ngày thực hiện giao dịch
6	AMOUNT	Float	Số lượng sản phẩm

7	PRICE	Float	Đơn giá mỗi sản phẩm (đơn vị lira)
8	LINENETTOTAL	Float	Tổng số tiền sau khi giảm giá (nếu có) (đơn vị lira)
9	LINENET	Float	Tổng số tiền trước khi giảm giá (đơn vị lira)
10	BRANCHNR	String	Mã định danh chi nhánh
11	BRANCH	String	Tên chi nhánh
12	SALESMAN	String	Tên nhân viên bán hàng
13	CITY	String	Tên thành phố của chi nhánh
14	REGION	String	Tên khu vực của chi nhánh
15	LATITUDE	Float	Vĩ độ của chi nhánh
16	LONGITUDE	Float	Kinh độ của chi nhánh
17	CLIENTCODE	String	Mã khách hàng
18	CLIENTNAME	String	Tên khách hàng
19	BRANDCODE	String	Mã định danh thương hiệu
20	BRAND	String	Tên thương hiệu của sản phẩm
21	CATEGORY_NAME1	String	Tên loại sản phẩm cấp 1
22	CATEGORY_NAME2	String	Tên loại sản phẩm cấp 2
23	CATEGORY_NAME3	String	Tên loại sản phẩm cấp 3
24	STARTDATE	Datetime	Ngày bắt đầu hiệu lực giao dịch
25	ENDDATE	Datetime	Ngày kết thúc hiệu lực giao dịch
26	GENDER	String	Giới tính của khách hàng gồm 2 giá trị E(Erkek): Nam và K(Kadin): Nữ

#### 1.4. Các thao tác tiền xử lý bộ dữ liệu:

Toàn bộ quá trình xử lý được thực hiện trên công cụ Jupyter Notebook của Anaconda Navigator và Google Collab, sử dụng ngôn ngữ lập trình Python và thư viện hỗ trợ như Pandas.

##### 1.4.1. Xử lý các dòng dữ liệu:

###### 1.4.1.1. Xử lý giá trị khuyết:

Do dữ liệu khuyết có số lượng tương đối lớn và nằm rải rác ở nhiều cột thuộc tính, nhóm quyết định xóa toàn bộ các dòng chứa giá trị khuyết ở tất cả các cột.

```
In [5]: 1 origin_data.dropna(inplace = True)
2 origin_data.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 365889 entries, 0 to 611107
Data columns (total 26 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   ID               365889 non-null   float64
 1   ITEMCODE         365889 non-null   float64
 2   ITEMNAME         365889 non-null   object 
 3   FICHENO          365889 non-null   int64  
 4   DATE_             365889 non-null   datetime64[ns]
 5   AMOUNT            365889 non-null   float64
 6   PRICE             365889 non-null   float64
 7   LINENETTOTAL     365889 non-null   float64
 8   LINENET            365889 non-null   float64
 9   BRANCHNR          365889 non-null   float64
 10  BRANCH            365889 non-null   object 
 11  SALESMAN          365889 non-null   object 
 12  CITY              365889 non-null   object 
 13  REGION             365889 non-null   object 
 14  LATITUDE           365889 non-null   float64
 15  LONGITUDE          365889 non-null   float64
 16  CLIENTCODE         365889 non-null   object 
 17  CLIENTNAME         365889 non-null   object 
 18  BRANDCODE          365889 non-null   object 
 19  BRAND              365889 non-null   object 
 20  CATEGORY_NAME1     365889 non-null   object 
 21  CATEGORY_NAME2     365889 non-null   object 
 22  CATEGORY_NAME3     365889 non-null   object 
 23  STARTDATE          365889 non-null   datetime64[ns]
 24  ENDDATE            365889 non-null   datetime64[ns]
 25  GENDER              365889 non-null   object 

dtypes: datetime64[ns](3), float64(9), int64(1), object(13)
memory usage: 75.4+ MB
```

Hình 2: Thông tin các thuộc tính trong bộ dữ liệu

#### 1.4.1.2. Thuộc tính GENDER:

Thay đổi giá trị “E” thành “Male” và giá trị “K” thành “Female” bằng hàm replace().

```
In [6]: 1 origin_data['GENDER'].unique()
Out[6]: array(['E', 'K'], dtype=object)

In [7]: 1 origin_data['GENDER'] = origin_data['GENDER'].replace(to_replace={"E": "Male", "K": "Female"})

In [8]: 1 origin_data['GENDER'].unique()
Out[8]: array(['Male', 'Female'], dtype=object)
```

Hình 3: Format lại giá trị cột thuộc tính GENDER

#### 1.4.1.3. Phiên dịch dữ liệu:

Do dữ liệu thu thập được viết bằng Thổ Nhĩ Kỳ gây khó khăn trong quá trình phân tích nên nhóm quyết định dịch một số cột dữ liệu sang tiếng

Anh thông qua công cụ hỗ trợ của Google. Các cột dữ liệu đó bao gồm: ITEMNAME, CATEGORY\_NAME1, BRAND.

#### 1.4.2. Chọn lọc thuộc tính:

Sau quá trình tìm hiểu bộ dữ liệu, chúng em đã trao đổi với nhau và quyết định loại bỏ một vài thuộc tính không có ý nghĩa (không thực hiện câu truy vấn trên các thuộc tính đó) đối với kho dữ liệu chúng em đang xây dựng. Các thuộc tính được đề cập bao gồm: 5, 15, 16, 22, 23, 25, 26 được liệt kê ở trên.

```
In [8]: 1 columns_drop = ['DATE_', 'ENDDATE', 'LATITUDE','LONGITUDE','CATEGORY_NAME2', 'CATEGORY_NAME3']
2 origin_data.drop(columns_drop, axis=1, inplace=True)
```

```
In [9]: 1 origin_data
```

```
Out [9]:
```

	ID	ITEMCODE	ITEMNAME	FICHENO	AMOUNT	PRICE	LINETOTAL	LINETET	BRANCHNR	BRANCH	SALESMAN	CITY	REGION
0	11738.0	5863.0	SPRITE 1 LT LIMON AROMALI GAZOZ	18456	1.0	2.00	2.00	1.85	52.0	Kocaeli Subesi	Eyüp CİNE	Kocaeli	Marmara
2	11335.0	5979.0	FALIM SAKIZ 5Lİ NANE	18350	1.0	0.40	0.40	0.37	40.0	İstanbul Subesi	İsmet SARTIK	İstanbul	Marmara
3	11336.0	5979.0	FALIM SAKIZ 5Lİ NANE	18350	1.0	0.40	0.40	0.37	40.0	İstanbul Subesi	İsmet SARTIK	İstanbul	Marmara
4	10115.0	12808.0	FALIM SAKIZ 5Lİ CILEK	18005	1.0	0.40	0.40	0.37	41.0	Izmir Subesi	Gültén İŞLAK	Izmir	Ege
6	11845.0	3202.0	F NEFFIS TOZ SEKER 2 KG	18476	1.0	5.60	5.60	5.19	55.0	Malatya Subesi	Gülcan DÜZGEÇ	Malatya	Doğu Anadolu
...	...	...	...	...	...	...	...	...	...	...	...	...	...
611098	512050.0	18235.0	SULTAN TAHILLİ SANDVIC	134701	1.0	2.00	2.00	1.69	53.0	Konya Subesi	Eren KIRMIZI	Konya	İç Anadolu
611099	512106.0	18219.0	SULTAN KEPBUGDAY EKMEĞİ	134716	1.0	1.00	1.00	0.99	58.0	Mersin Subesi	Aylin ÇEVİKOĞLU	Mersin	Akdeniz
611100	512175.0	18219.0	SULTAN KEPBUGDAY EKMEĞİ	134735	1.0	1.00	1.00	0.99	53.0	Konya Subesi	Eren KIRMIZI	Konya	İç Anadolu
611101	511342.0	6480.0	UNO KEPEKLİ EKMEK 470GR.	134601	1.0	3.25	3.25	3.22	73.0	Tekirdağ Subesi	Aleyna ÇUKUREL	Tekirdağ	Marmara
611107	512179.0	3942.0	ETI 80 GR KARAM BITTER	134738	1.0	1.90	1.90	1.76	74.0	Tokat Subesi	Lütfiye DÖŞEME	Tokat	Karadeniz

365889 rows x 20 columns

Hình 4: Loại bỏ một số cột thuộc tính không cần thiết

#### 1.4.3. Đổi tên cột thuộc tính:

Nhóm tiếp tục đổi tên các cột dữ liệu phù hợp với kho dữ liệu mà nhóm đã thiết kế dựa trên ý nghĩa của chúng. Sau đây là bảng ánh xạ tên thuộc tính tương ứng.

Tên thuộc tính ban đầu	Tên thuộc tính chuẩn hóa
ID	trans_id
ITEMCODE	product_id
ITEMNAME	product_name
FICHENO	bill_id
AMOUNT	amount
PRICE	price
LINENETTOTAL	pre-promotion_total_trans
LINENET	final_total_trans
BRANCHNR	branch_id
BRANCH	branch_name
SALESMAN	salesman
CITY	city
REGION	region
CLIENTCODE	customer_id
CLIENTNAME	customer_name
BRANDCODE	brand_id
BRAND	brand_name
CATEGORY_NAME1	category_name
STARTDATE	invoice_date

```
In [10]: 1 rename_columns = {
2     "ID": "trans_id",
3     "ITEMCODE": "product_id",
4     "ITEMNAME": "product_name",
5     "FICHENO": "bill_id",
6     "STARTDATE": "invoice_date",
7     "AMOUNT": "amount",
8     "PRICE": "price",
9     "LINENETTOTAL": "pre-promotion_total_trans",
10    "LINENET": "final_total_trans",
11    "BRANCHNR": "branch_id",
12    "BRANCH": "branch_name",
13    "SALESMAN": "salesman",
14    "CITY": "city",
15    "REGION": "region",
16    "CLIENTCODE": "customer_id",
17    "CLIENTNAME": "customer_name",
18    "BRANDCODE": "brand_id",
19    "BRAND": "brand_name",
20    "CATEGORY_NAME1": "category_name"
21 }
22 origin_data= origin_data.rename(columns=rename_columns)
```

```
In [11]: 1 origin_data.info()
```

	Column	Non-Null Count	Dtype
0	trans_id	365889	non-null float64
1	product_id	365889	non-null float64
2	product_name	365889	non-null object
3	bill_id	365889	non-null int64
4	amount	365889	non-null float64
5	price	365889	non-null float64
6	pre-promotion_total_trans	365889	non-null float64
7	final_total_trans	365889	non-null float64
8	branch_id	365889	non-null float64
9	branch_name	365889	non-null object
10	salesman	365889	non-null object
11	city	365889	non-null object
12	region	365889	non-null object
13	customer_id	365889	non-null object
14	customer_name	365889	non-null object
15	brand_id	365889	non-null object
16	brand_name	365889	non-null object
17	category_name	365889	non-null object
18	invoice_date	365889	non-null datetime64[ns]
19	GENDER	365889	non-null object

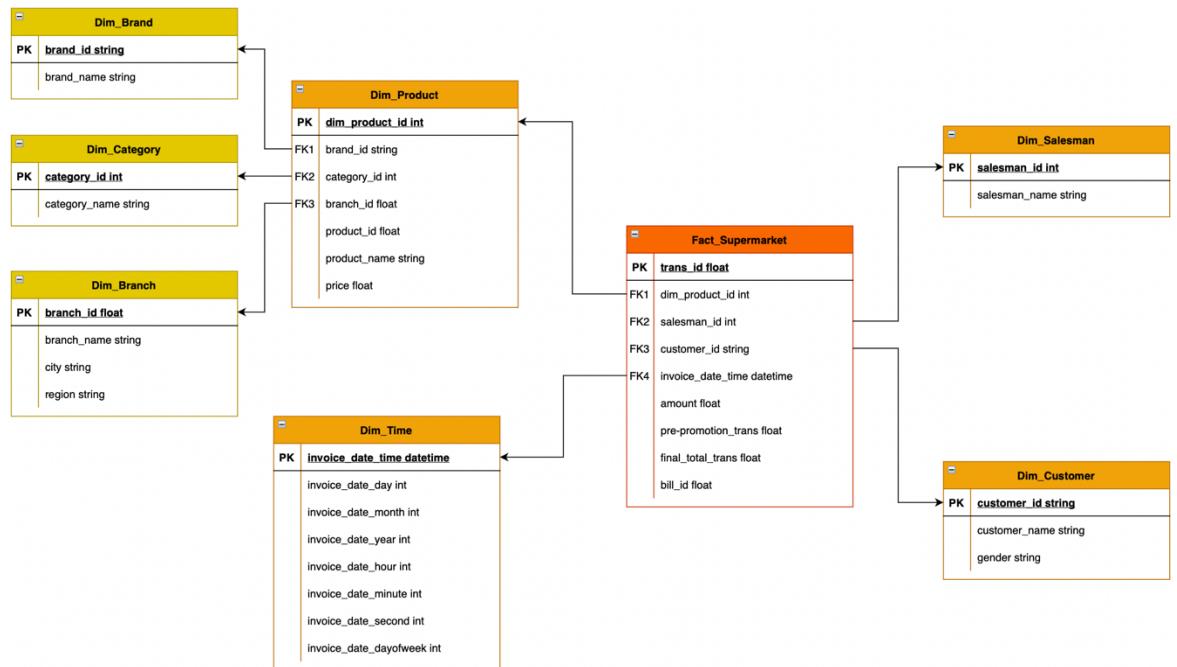
dtypes: datetime64[ns](1), float64(7), int64(1), object(11)  
memory usage: 58.6+ MB

Hình 5: Đổi tên một số cột thuộc tính

## 1.5.Thiết kế kho dữ liệu:

Bộ dữ liệu sau khi đã hoàn tất các bước xử lý trên: [Clean Data](#). Bộ dữ liệu còn lại 365889 dòng dữ liệu và 20 cột thuộc tính. Hướng chủ đề của kho dữ liệu: Quản lý hoạt động bán lẻ của một siêu thị.

Lược đồ hình bông tuyêt:



Hình 6: Mô hình kho dữ liệu (mô hình bông tuyết)

### 1.5.1. Dim\_Product:

Bao gồm 116098 dòng dữ liệu và 7 cột thuộc tính lưu trữ thông tin liên quan đến sản phẩm. Do chưa có thuộc tính có thể đảm nhiệm vai trò khóa chính, nhóm sẽ tự động phát sinh thuộc tính dim\_product\_id làm khóa chính của bảng.

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả
dim_product_id	Int	Khóa chính	Mã bảng sản phẩm
brand_id	String	Khóa ngoại liên kết đến bảng dim_brand	Mã định danh thương hiệu
category_id	Int	Khóa ngoại liên kết đến bảng dim_category	Mã loại sản phẩm
branch_id	Float	Khóa ngoại liên kết đến bảng dim_branch	Mã chi nhánh bán sản phẩm
product_id	Float		Mã định danh sản phẩm
product_name	String		Tên sản phẩm

price	Float		Đơn giá mỗi sản phẩm (đơn vị lira)
-------	-------	--	---------------------------------------

### **1.5.2. Dim\_Brand:**

Bao gồm 304 dòng dữ liệu và 2 cột thuộc tính liên quan đến các thương hiệu sản xuất và phân phối sản phẩm. Thuộc tính khóa chính của bảng là brand\_id.

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả
brand_id	String	Khóa chính	Mã định danh thương hiệu
brand_name	String		Tên thương hiệu

### **1.5.3. Dim\_Category:**

Bao gồm 8 dòng dữ liệu và 2 thuộc tính chứa các thông tin liên quan đến loại sản phẩm đang được bày bán trong hệ thống siêu thị. Do chưa có thuộc tính đảm nhiệm vai trò khóa chính, nhóm sẽ tự động phát sinh thuộc tính category\_id làm khóa chính của bảng.

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả
category_id	Int	Khóa chính	Mã loại sản phẩm
category_name	String		Tên loại sản phẩm

### **1.5.4. Dim\_Customer:**

Bao gồm 45942 dòng dữ liệu và 3 thuộc tính lưu trữ thông tin liên quan đến các khách hàng đã thực hiện giao dịch tại hệ thống siêu thị trong quý I năm 2017. Khóa chính của bảng là customer\_id.

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả
customer_id	String	Khóa chính	Mã khách hàng
customer_name	String		Tên khách hàng
gender	String		Giới tính

### **1.5.5. Dim\_Salesman:**

Bao gồm 1266 dòng dữ liệu và 2 thuộc tính chứa thông tin của các nhân viên bán hàng. Do chưa có thuộc tính khóa chính, nhóm sẽ tự động phát sinh thuộc tính salesman\_id làm khóa chính của bảng.

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả
salesman_id	Int	Khóa chính	Mã định danh nhân viên bán hàng
salesman_name	String		Tên nhân viên bán hàng

#### 1.5.6. Dim\_Branch:

Bao gồm 81 dòng dữ liệu và 3 thuộc tính chứa thông tin của các chi nhánh của hệ thống siêu thị. Khóa chính của bảng là thuộc tính branch\_id.

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả
branch_id	Float	Khóa chính	Mã định danh chi nhánh
branch_name	String		Tên chi nhánh
city	String		Tên thành phố của chi nhánh
region	String		Tên khu vực của chi nhánh

#### 1.5.7. Dim\_Time:

Bao gồm 102026 dòng dữ liệu và 8 thuộc tính chứa thông tin của các chi nhánh của hệ thống siêu thị. Khóa chính của bảng là thuộc tính invoice\_date\_time.

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả
invoice_date_time	Datetime	Khóa chính	Thời gian chi tiết của hóa đơn
invoice_date_day	Int		Ngày hóa đơn
invoice_date_month	Int		Tháng hóa đơn
invoice_date_year	Int		Năm hóa đơn
invoice_date_minute	Int		Số phút của hóa đơn
invoice_date_second	Int		Số giây của hóa đơn

invoice_date_dayof week	Int		Thứ tự ngày trong tuần của hóa đơn
----------------------------	-----	--	---------------------------------------

### 1.5.8. Fact\_Supermarket:

Bao gồm 360373 dòng dữ liệu và 9 thuộc tính chứa thông tin của các chi nhánh của hệ thống siêu thị. Khóa chính của bảng là thuộc tính trans\_id.

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc	Mô tả
trans_id	Int	Khóa chính	Mã định danh giao dịch
invoice_date_time	Datetime	Khóa ngoại liên kết đến bảng Dim_Time	Thời gian chi tiết của hóa đơn
customer_id	String	Khóa ngoại liên kết đến bảng Dim_Customer	Mã định danh khách hàng
salesman_id	Int	Khóa ngoại liên kết đến bảng Dim_Salesman	Mã định danh nhân viên bán hàng
dim_product_id	Int	Khóa ngoại liên kết đến bảng Dim_Product	Mã bảng sản phẩm
bill_id	Float		Mã hóa đơn
amount	Float		Số lượng sản phẩm
pre- promotion_total_trans	Float		Chi phí giao dịch trước khuyến mãi (đơn vị lira)
final_total_trans	Float		Chi phí giao dịch cuối cùng (đơn vị lira)

### 1.6. Nội dung 15 câu truy vấn:

- Liệt kê danh sách các thương hiệu thực hiện giao dịch trong khoảng thời gian từ ngày 15/02/2017 đến 15/03/2017.
- Liệt kê tên các chi nhánh có bán ít nhất một sản phẩm với giá trên 3 lira.
- Cho biết thông tin nhân viên và chi nhánh thực hiện giao dịch với khách hàng có tên là “Kırklareli Subesi” trong tháng 2 năm 2017.

- d) Cho biết tên thương hiệu bán được nhiều sản phẩm nhất trong tháng 1 năm 2017.
- e) Cho biết thông tin kho sản phẩm mang lại doanh thu cao nhất trong tháng 3 năm 2017 tại các chi nhánh đặt tại thành phố New York.
- f) Cho biết thông tin các nhân viên thực hiện giao dịch các sản phẩm có thương hiệu là “Sprite” tại chi nhánh “Tadim”.
- g) Cho biết thông tin khách hàng có tổng giá trị đơn hàng cao nhất trong tháng 2 và tháng 3 tại hệ thống siêu thị. Kết quả sắp xếp theo thứ tự giảm dần.
- h) Liệt kê thông tin các sản phẩm và kho sản phẩm tương ứng của các giao dịch với khách hàng có giới tính là “Male”.
- i) Tìm ra đơn hàng có số lượng sản phẩm mua lớn nhất trong tháng 01 năm 2017.
- j) Trong 3 ngày đầu của mỗi tháng, cho biết sản phẩm nào được bán ra nhiều nhất.
- k) Với mỗi tháng, cho biết tên chi nhánh và tên nhãn hàng có doanh thu thấp nhất?
- l) Liệt kê tất cả các chi nhánh và tên nhân viên đã thực hiện giao dịch với nhãn hàng “İstanbul Subesi”.
- m) Tìm khách hàng thực hiện giao dịch ở cả hai chi nhánh “AİR WICK” và “ÜLKER”.
- n) Liệt kê các khách hàng thực hiện giao dịch với nhãn hàng “Konya Subesi” nhưng không thực hiện giao dịch với nhãn hàng “Çanakkale Subesi”.
- o) Với mỗi khách hàng giao dịch trong tháng 01, cho biết top 3 khách hàng có lượng chi tiêu lớn nhất.

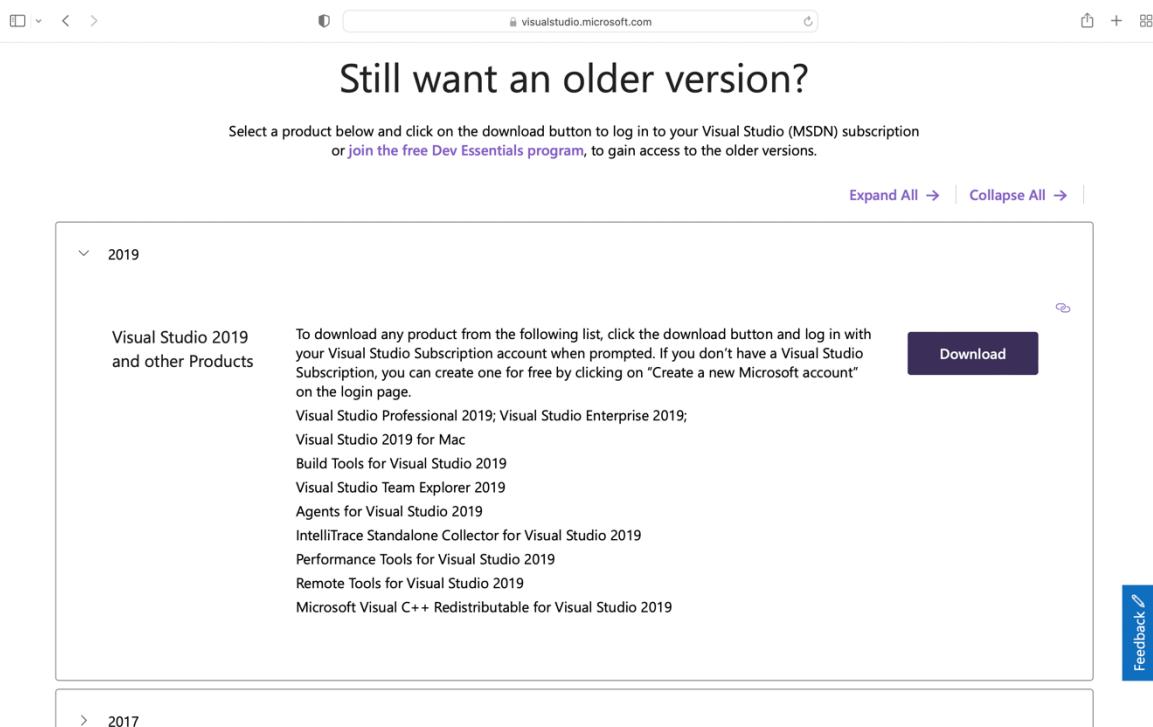
## 2. Tích hợp dữ liệu vào kho dữ liệu (SSIS):

## 2.1. Chuẩn bị công cụ hỗ trợ:

Để có thể tạo kho dữ liệu và thực hiện được quá trình đỗ dữ liệu vào trong kho dữ liệu, ta cần cài đặt một số công cụ BI hỗ trợ của SQL Server:

- + Visual Studio Community 2019.
- + SQL Server Integration Services Project.

**Bước 1:** Tải Visual Studio Community 2019 tại [link](#). Tại vị trí 2019, click Download.



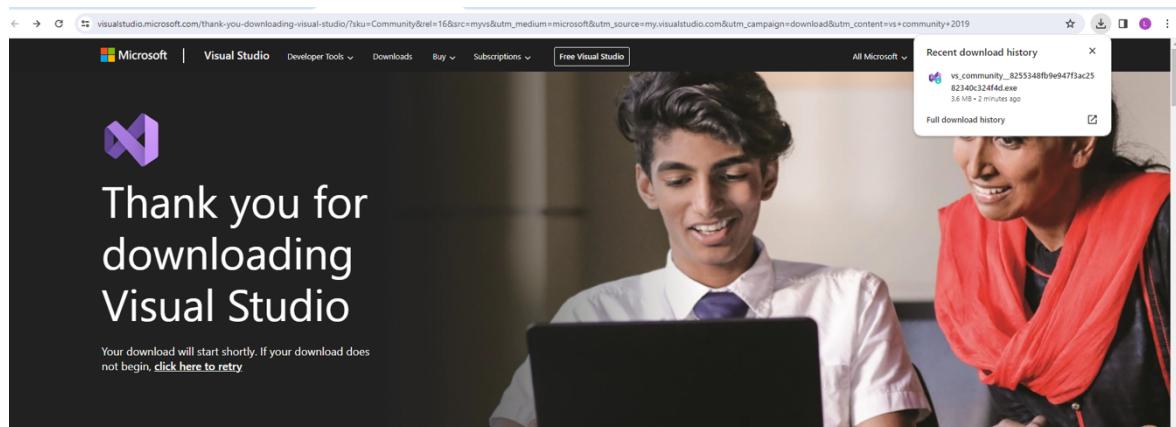
Hình 7: Giao diện website tải Visual Studio Community 2019

**Bước 2:** Tìm đến Visual Community 2019. Click “Download”.

Product	Version	Download	Keys	Language	Architecture	Type	Release Date
Visual Studio Enterprise 2019	16.11	<a href="#">Download</a>		Multiple Languages	x86	exe	13/07/2024
Visual Studio Team Explorer 2019	16.11	<a href="#">Download</a>		Multiple Languages	x86	exe	13/07/2024
Visual Studio Professional 2019	16.11	<a href="#">Download</a>		Multiple Languages	x86	exe	13/07/2024
Visual Studio Community 2019	16.11	<a href="#">Download</a>		Multiple Languages	x86	exe	13/07/2024
Agents for Visual Studio 2019	16.11	<a href="#">Download</a>		English	x64	exe	13/07/2024
Build Tools for Visual Studio 2019	16.11	<a href="#">Download</a>		Multiple Languages	x86	exe	13/07/2024
Visual C++ Redistributable for Visual Studio 2019	16.11	<a href="#">Download</a>		Multiple Languages	x64	exe	13/07/2024
Visual Studio 2019 for Mac	16.11	<a href="#">Download</a>		Multiple Languages	x64	exe	13/07/2024
Visual Studio	16.11	<a href="#">Download</a>		Multiple Languages	x64	exe	13/07/2024

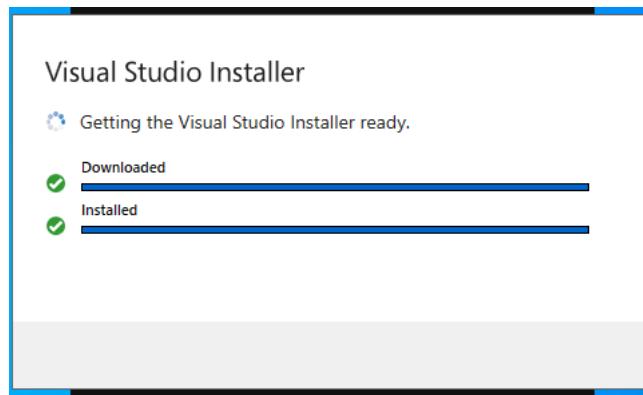
Hình 8: Giao diện website vị trí tải Visual Studio Community 2019

**Bước 3:** Sau khi quá trình tải về hoàn tất, double click vào file (.exe).



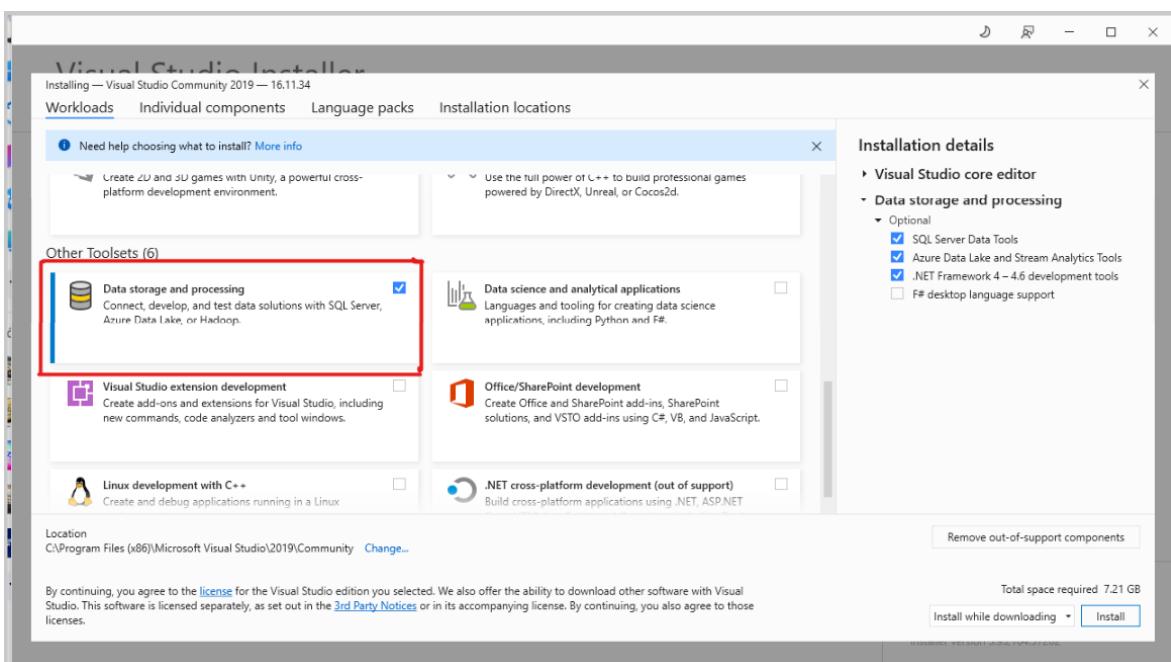
Hình 9: Giao diện download thành công file .exe

**Bước 4:** Hộp thoại download tiếp tục hiện lên. Chúng ta chờ cho đến khi mọi cài đặt được hoàn tất.



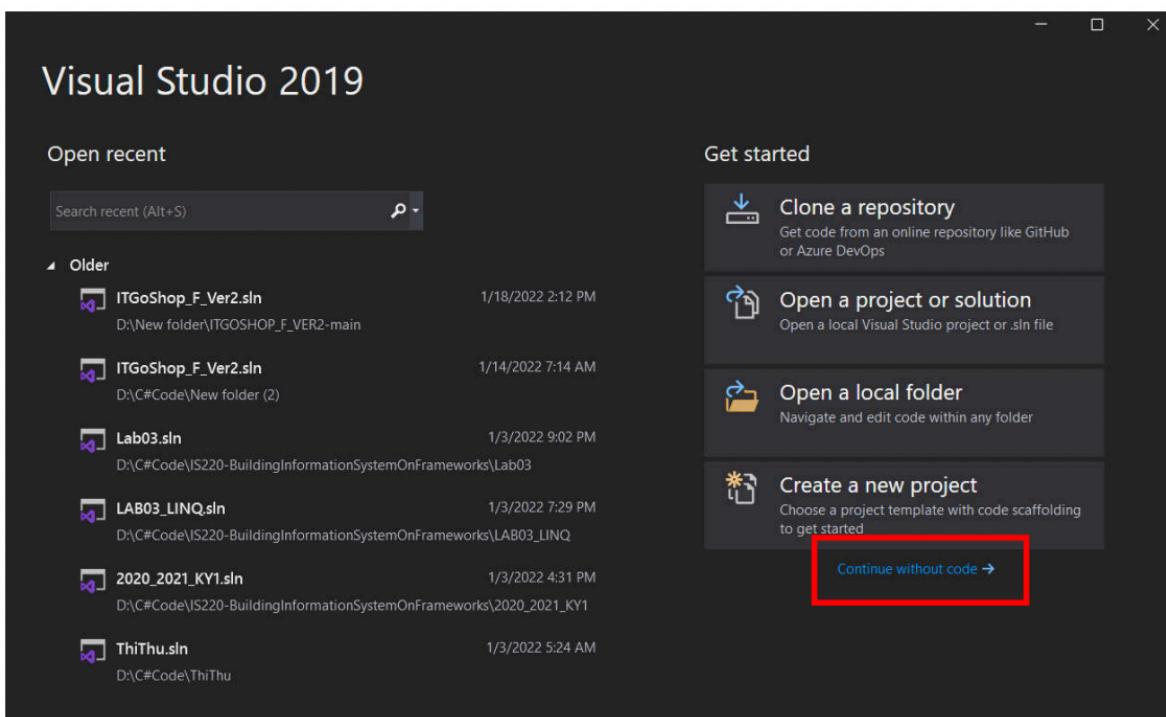
Hình 10: Giao diện sau khi mở file .exe

**Bước 5:** Tiếp tục cài đặt SQL Server Data Tools bằng cách tìm và chọn mục “**Data storage and processing**”. Sau đó click “**Install**” để tiến hành cài đặt.



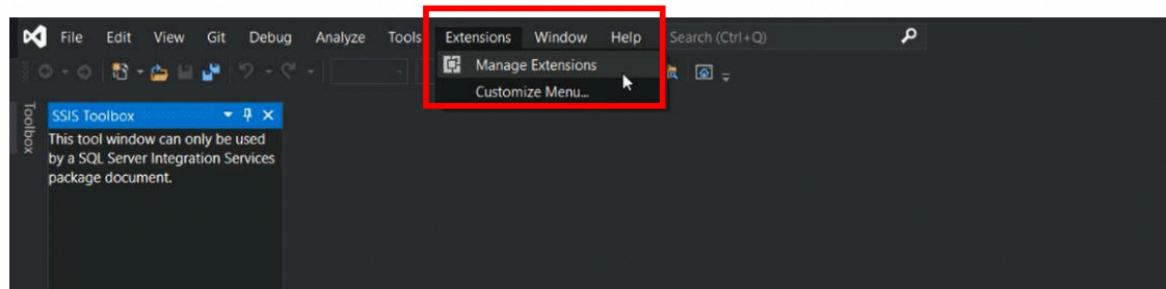
Hình 11: Giao diện cài đặt SQL Server Data Tools

**Bước 6:** Mở Visual Studio 2019 và chọn “**Continue without code**”.



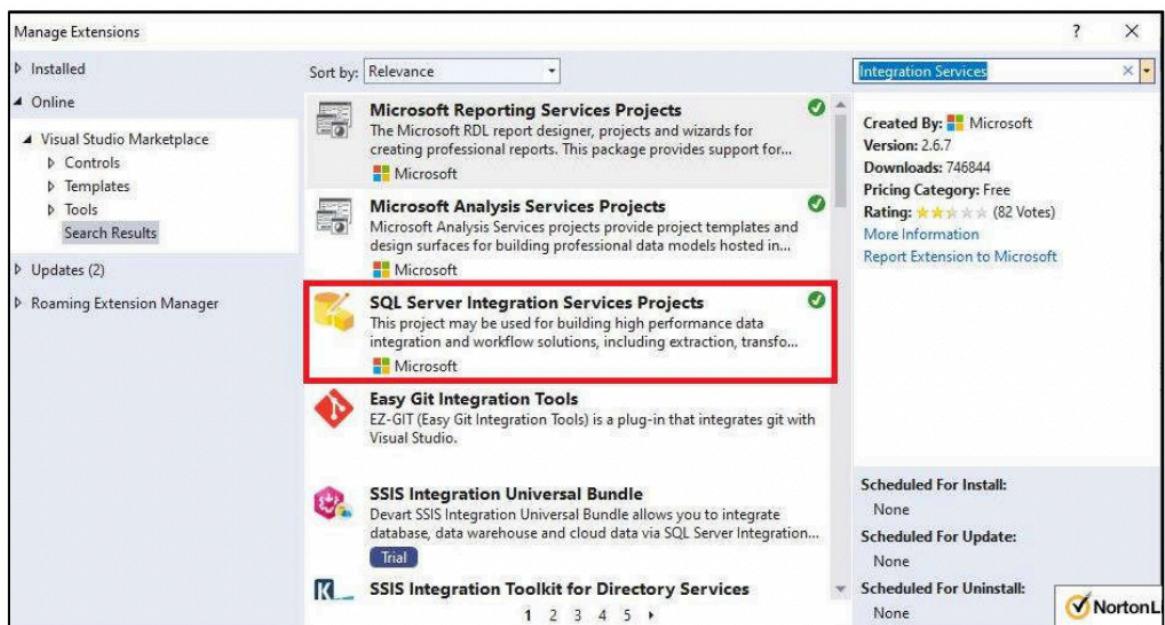
Hình 12: Giao diện khởi động một project rỗng

**Bước 7:** Trong giao diện chính, tìm tab “Extension”, chọn “Manage Extensions”.



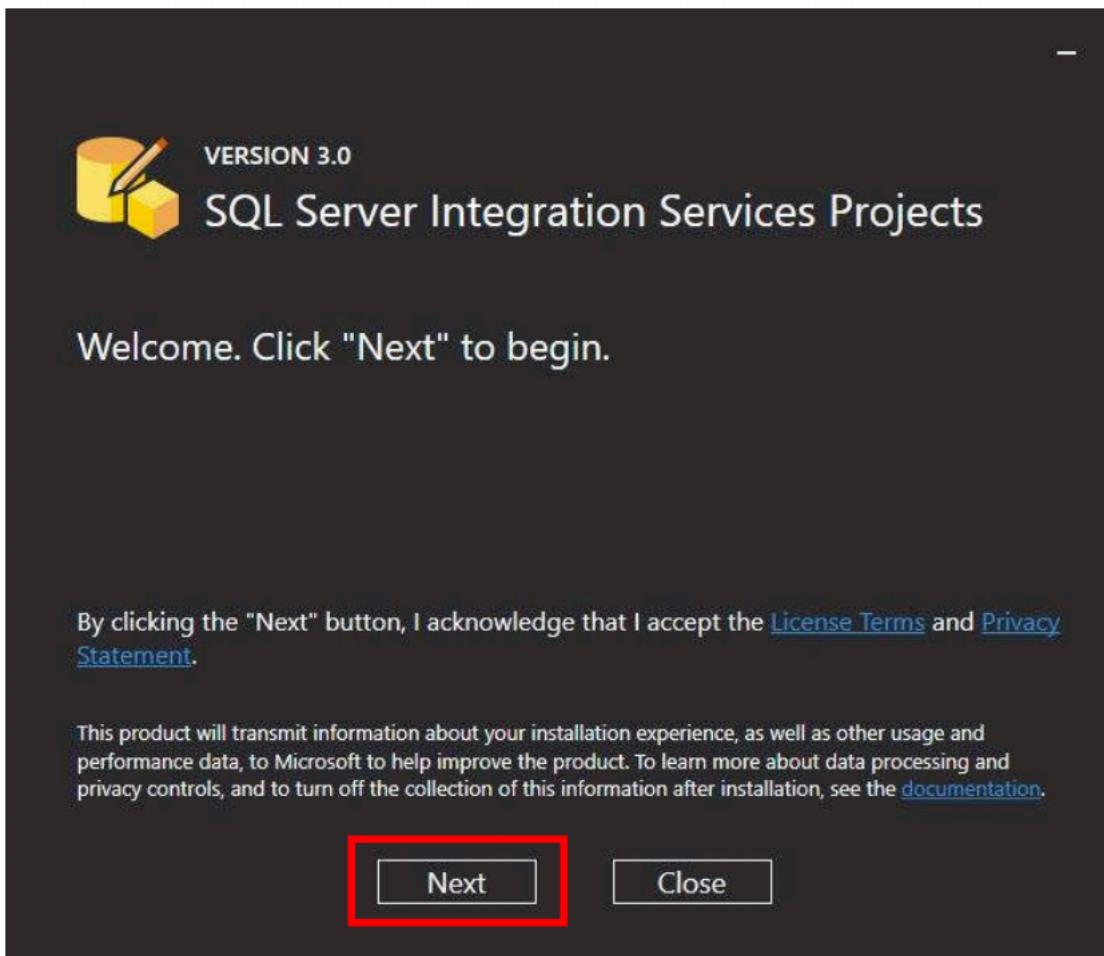
Hình 13: Giao diện vào nơi download SSIS

**Bước 8:** Tìm và tải về công cụ SQL Server Integration Services Projects.



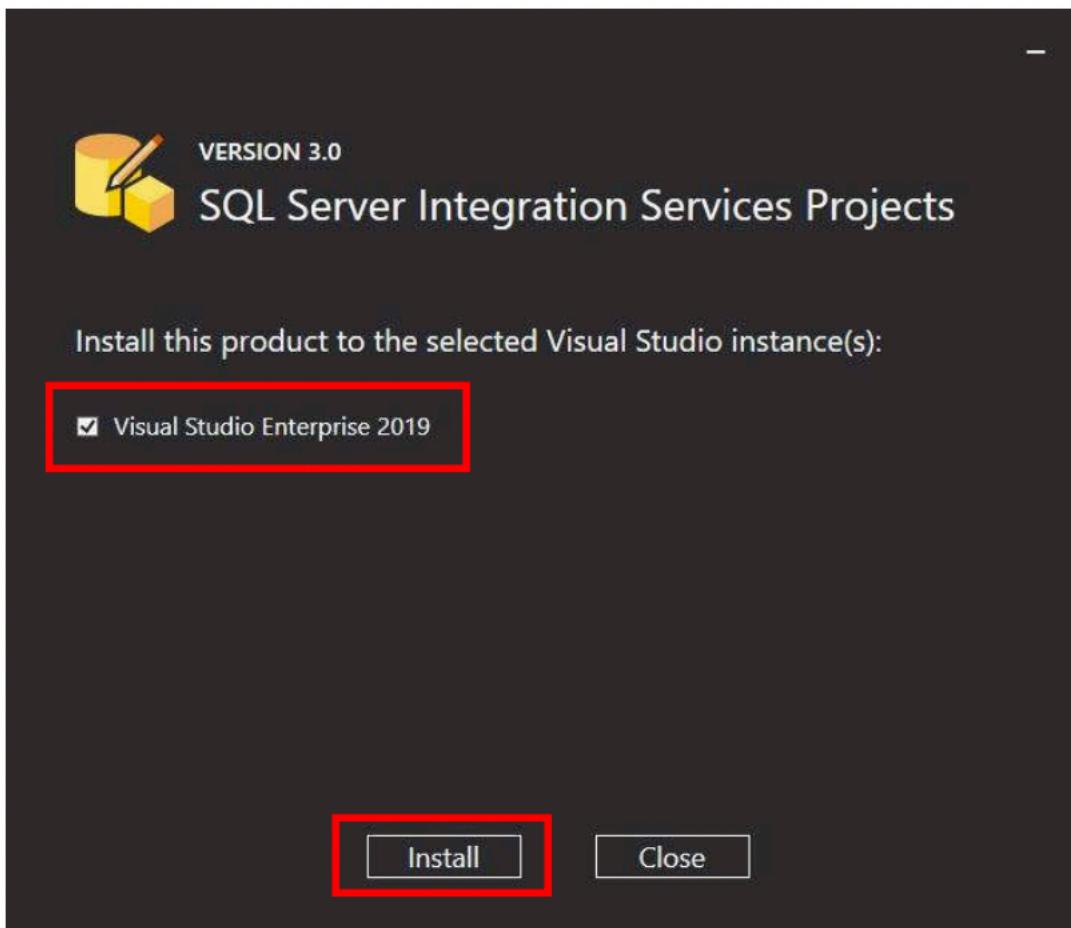
Hình 14: Giao diện hướng dẫn tìm kiếm và tải file SSIS

**Bước 9:** Mở file.exe vừa tải xuống. Chọn Next để tiếp tục.

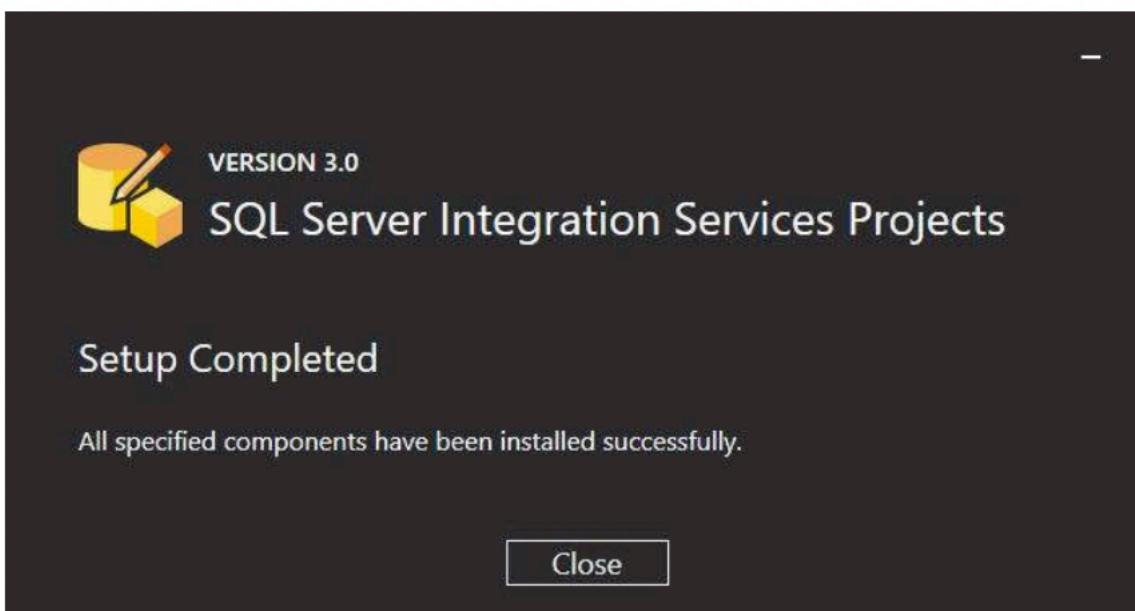


Hình 15: Giao diện sau khi hoàn tất quá trình tải SSIS trong VS Community 2019

**Bước 10:** Click chọn Visual Studio 2019 và click **Install**.



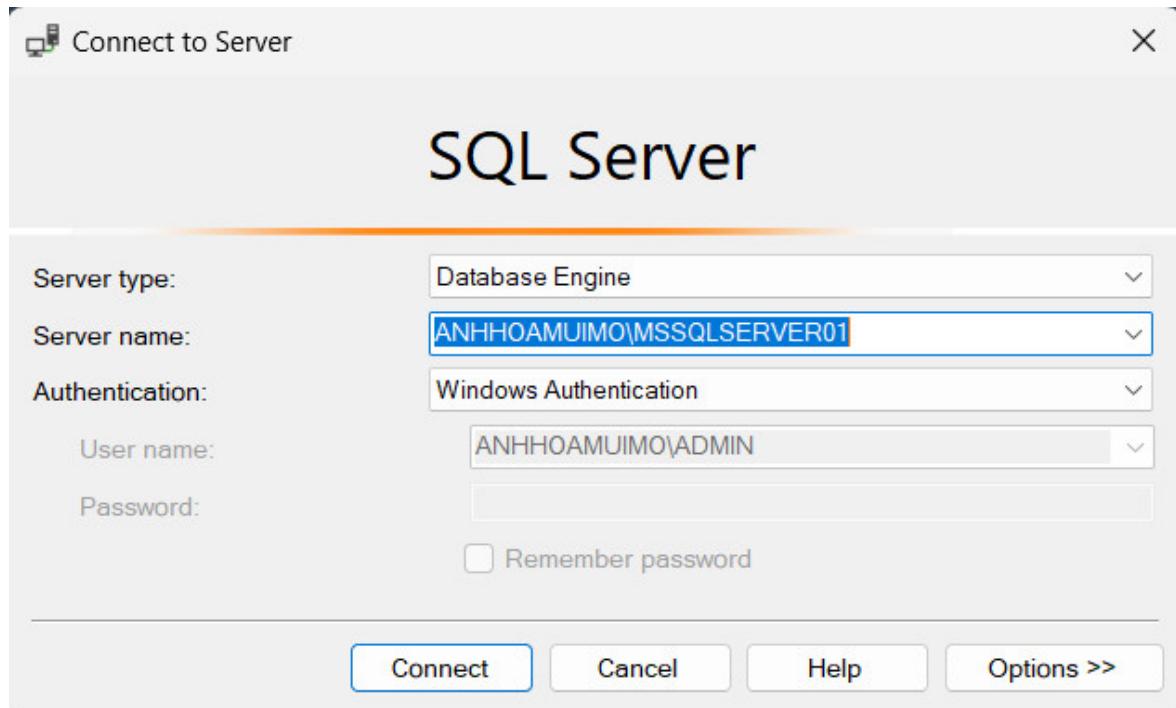
Hình 16: Giao diện hướng dẫn tiếp tục bước cuối cùng để tải SSIS



Hình 17: Kết quả sau khi hoàn tất quá trình cài đặt và thiết lập SSIS vào Visual Studio 2019

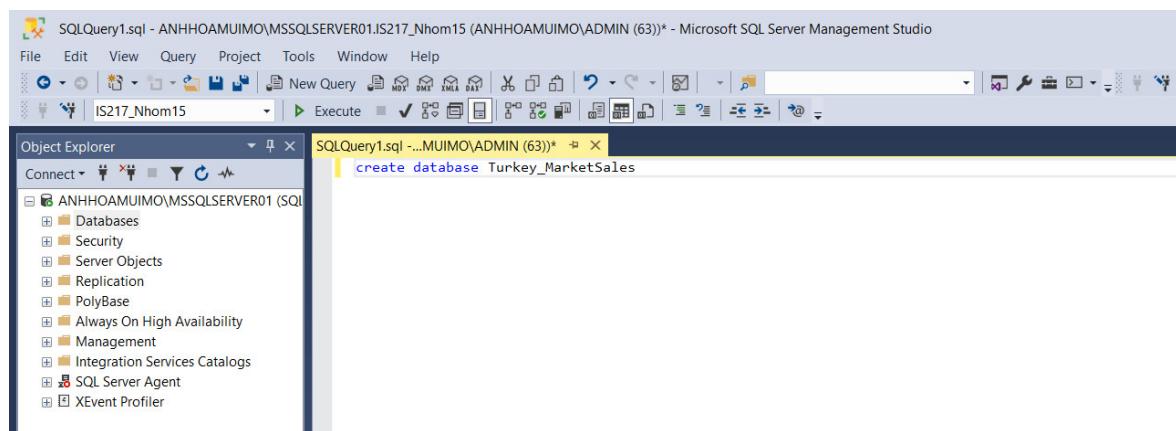
## 2.2. Khởi tạo cơ sở dữ liệu để chuẩn bị quá trình SSIS:

**Bước 1:** Khởi động phần mềm Microsoft SQL Server Management Studio 2019 và kết nối với server bằng tài khoản user của window (**Windows Authentication**).



Hình 18: Kết nối đến server trong SSMS

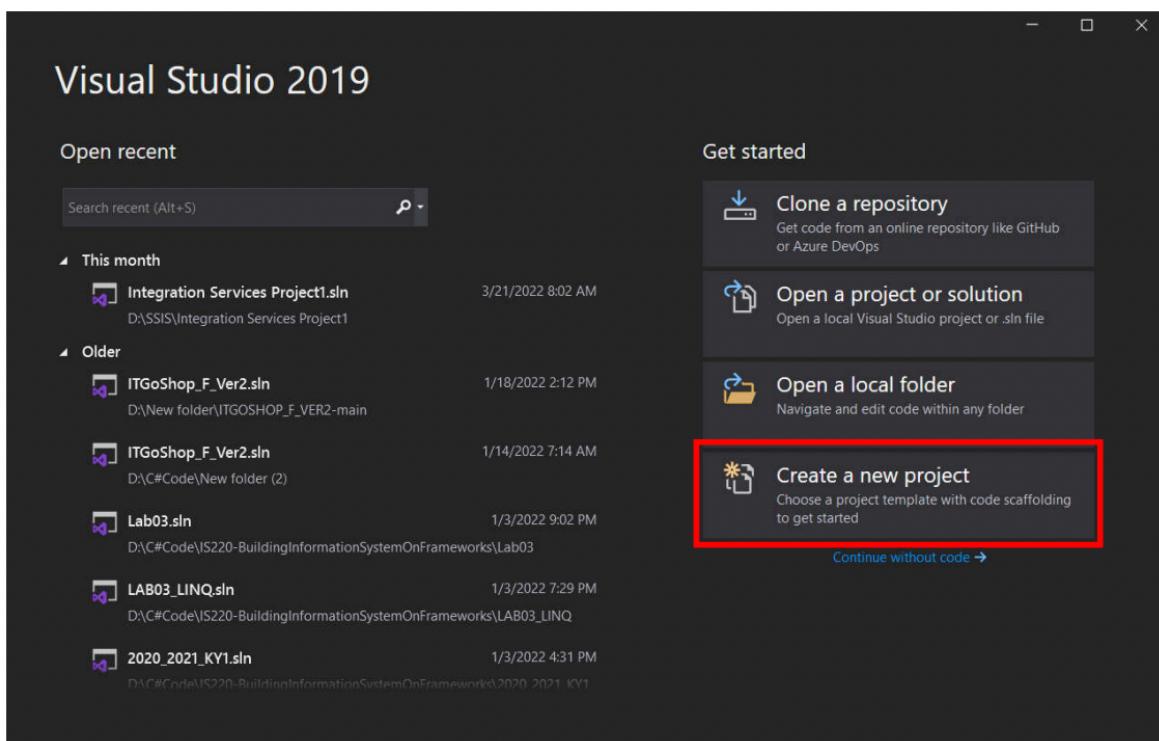
**Bước 2:** Khởi tạo một cơ sở dữ liệu có tên là: “**Turkey\_MarketSales**”, đây là kho dữ liệu – nơi lưu trữ các bảng Dim và bảng Fact với đầy đủ dữ liệu của các bảng đó.



Hình 19: Khởi tạo một database rỗng để chuẩn bị đổ dữ liệu vào

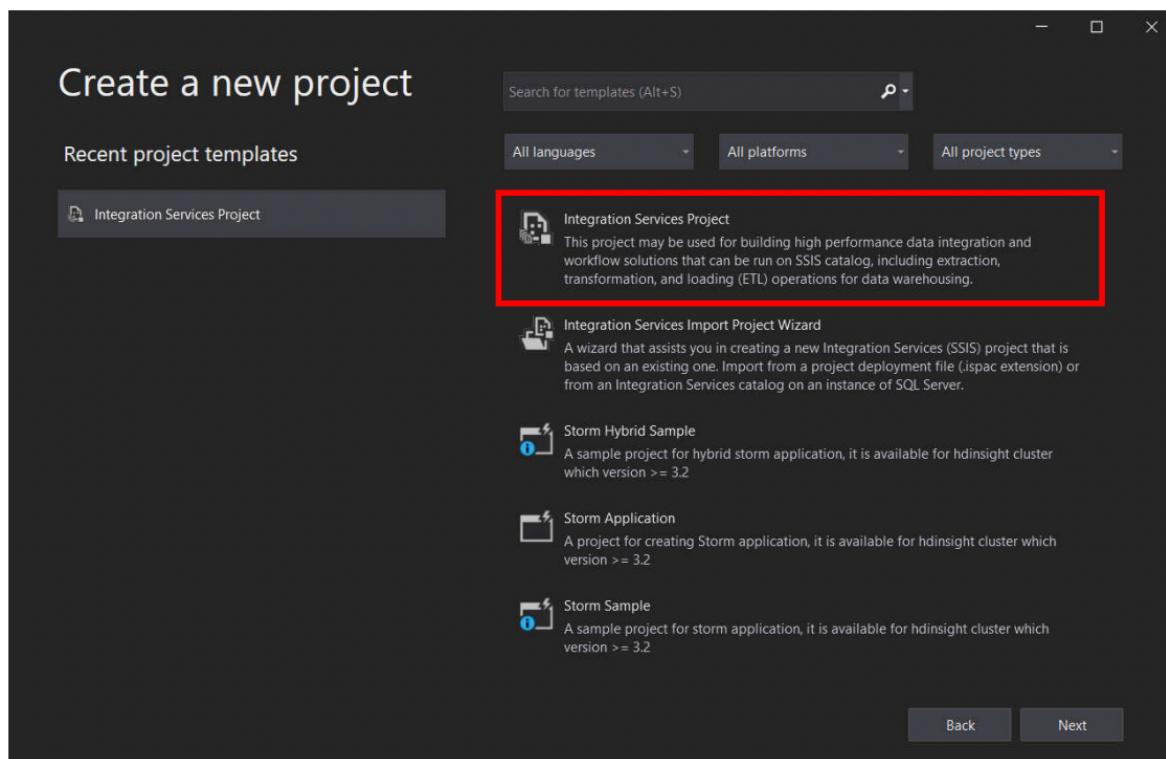
### 2.3. Hiện thực quá trình SSIS:

**Bước 1:** Khởi động Visual Studio 2019 và chọn “Create a new project”.



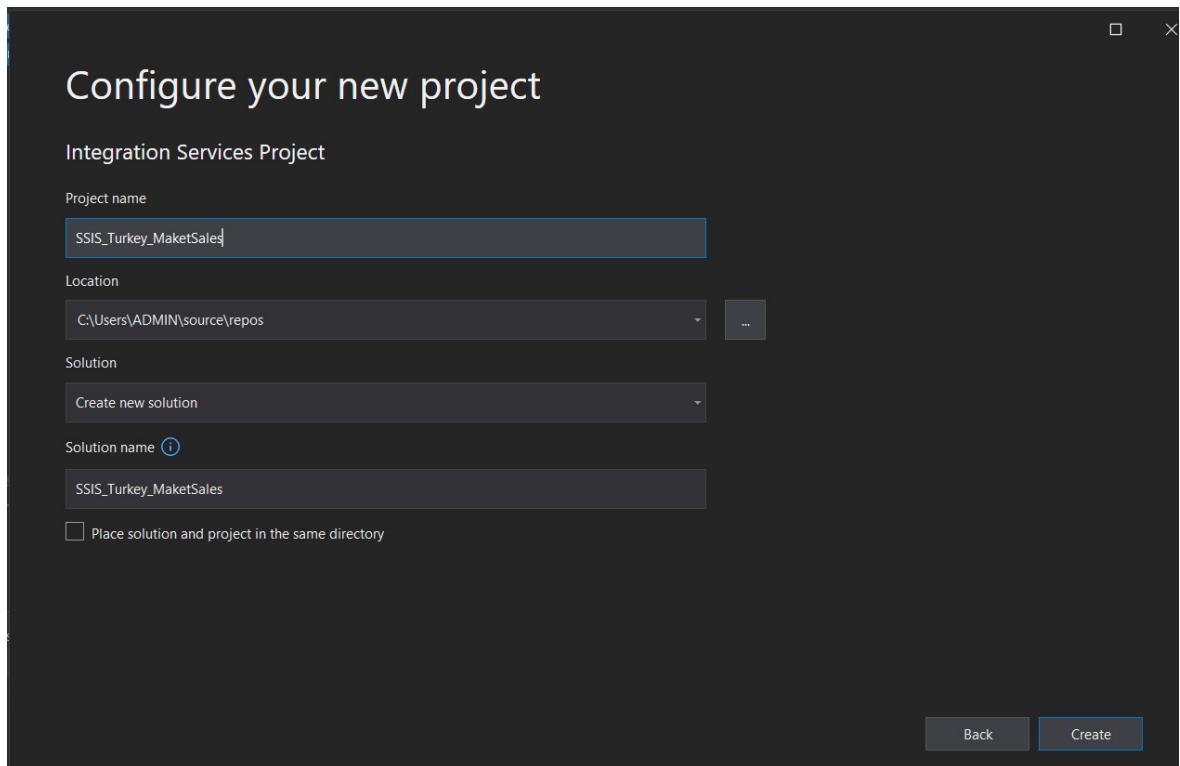
Hình 20: Khởi tạo một dự án mới

Bước 2: Chọn “Integration Service Project” và click “Next”.



Hình 21: Tìm kiếm và chọn loại dự án SSIS

**Bước 3:** Đặt tên và thiết lập đường dẫn cho Project SSIS. Sau đó click “Create”.

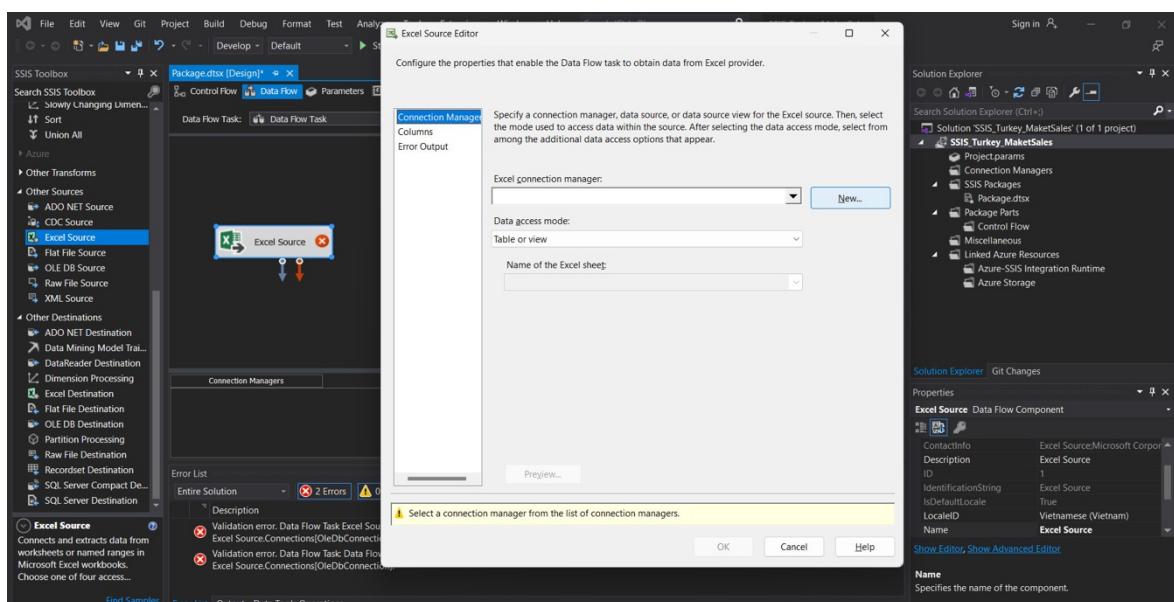


Hình 22: Đặt tên cho dự án và thiết lập nơi lưu trữ cho dự án

### 2.3.1. Tạo bảng Dim và bảng Fact:

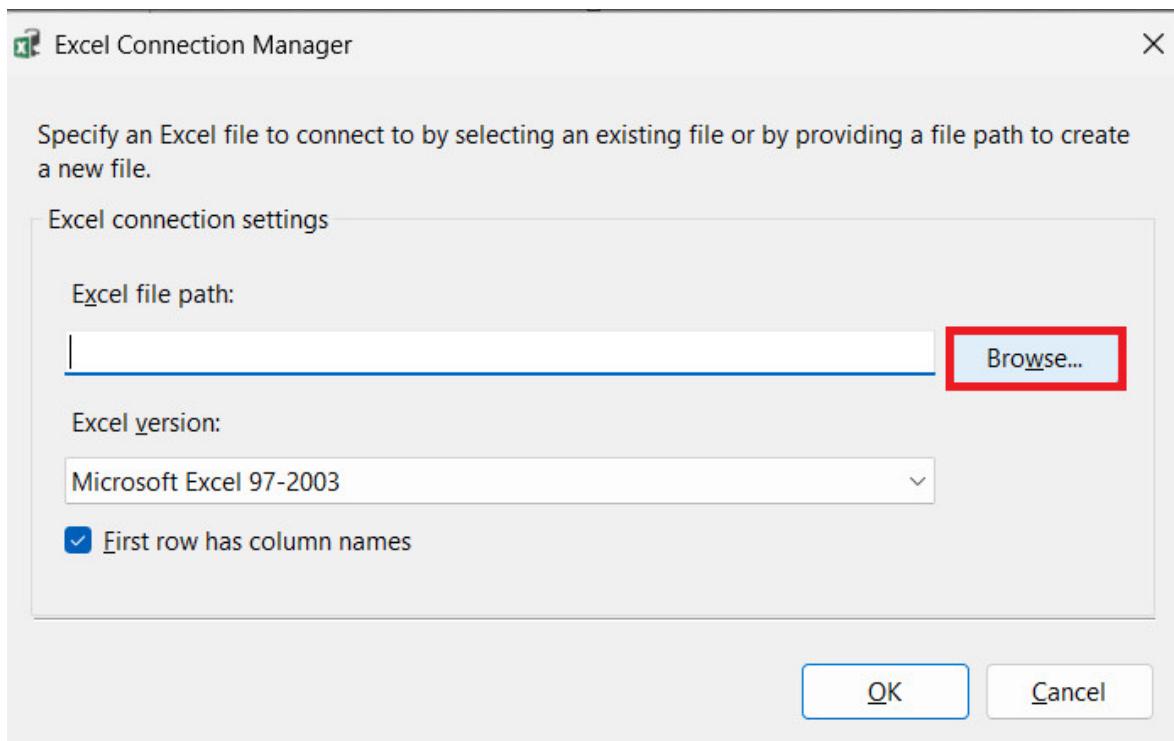
Trước khi tiến hành chia Dimension và bảng Fact, ta cần tải dữ liệu gốc từ file Excel (.xlsx) vào Data Flow:

**Bước 1:** Tạo một Data Flow Task. Trong **Data Flow**, tạo một đối tượng **Excel Source** để tải dữ liệu gốc từ file .xlsx. Click “**New**” để tạo một **Excel Connection Manager**.



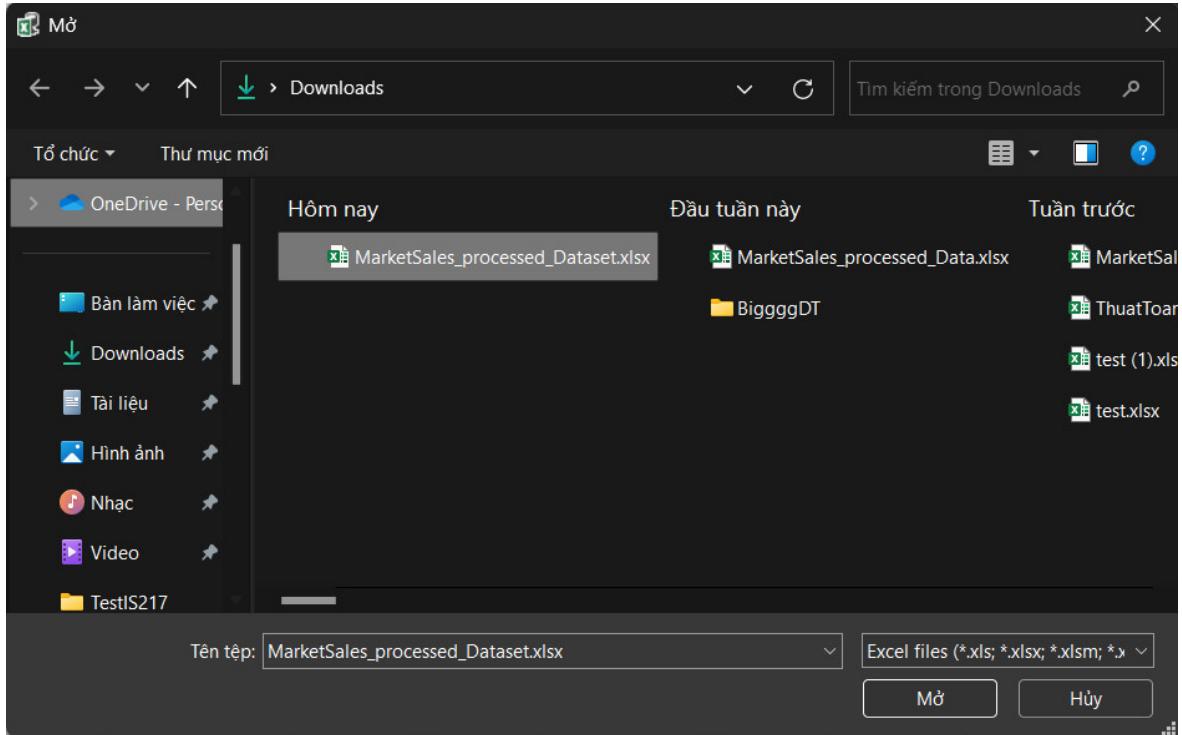
Hình 23: Giao diện tạo một kết nối mới đến file chứa dữ liệu gốc (.xlsx)

**Bước 2:** Click “Browse” để tải file dữ liệu gốc được lưu trong máy tại đường dẫn tương ứng.



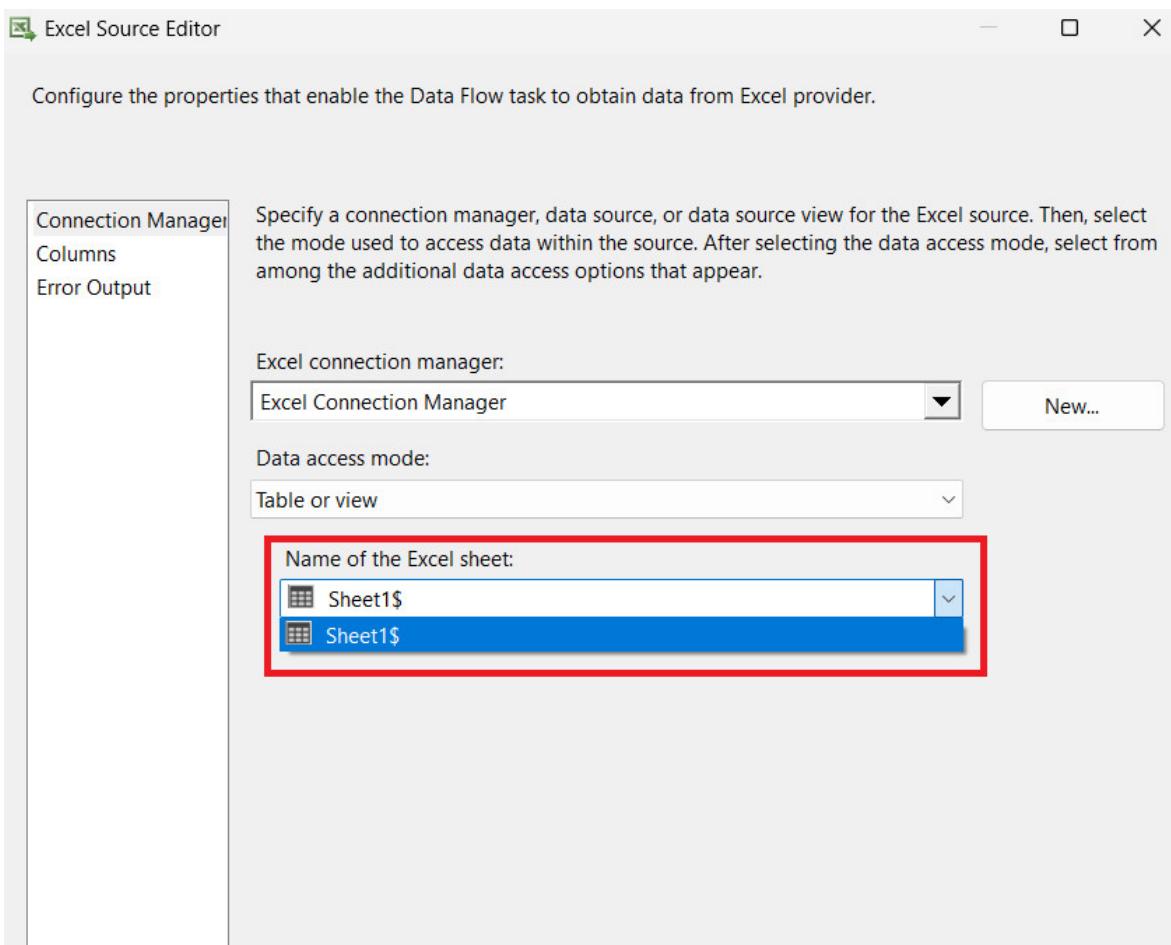
Hình 24: Thiết lập đường dẫn lưu trữ file dữ liệu gốc trên máy tính.

**Bước 3:** Chọn file excel lưu trữ dữ liệu gốc (.xlsx) và click “Open”.



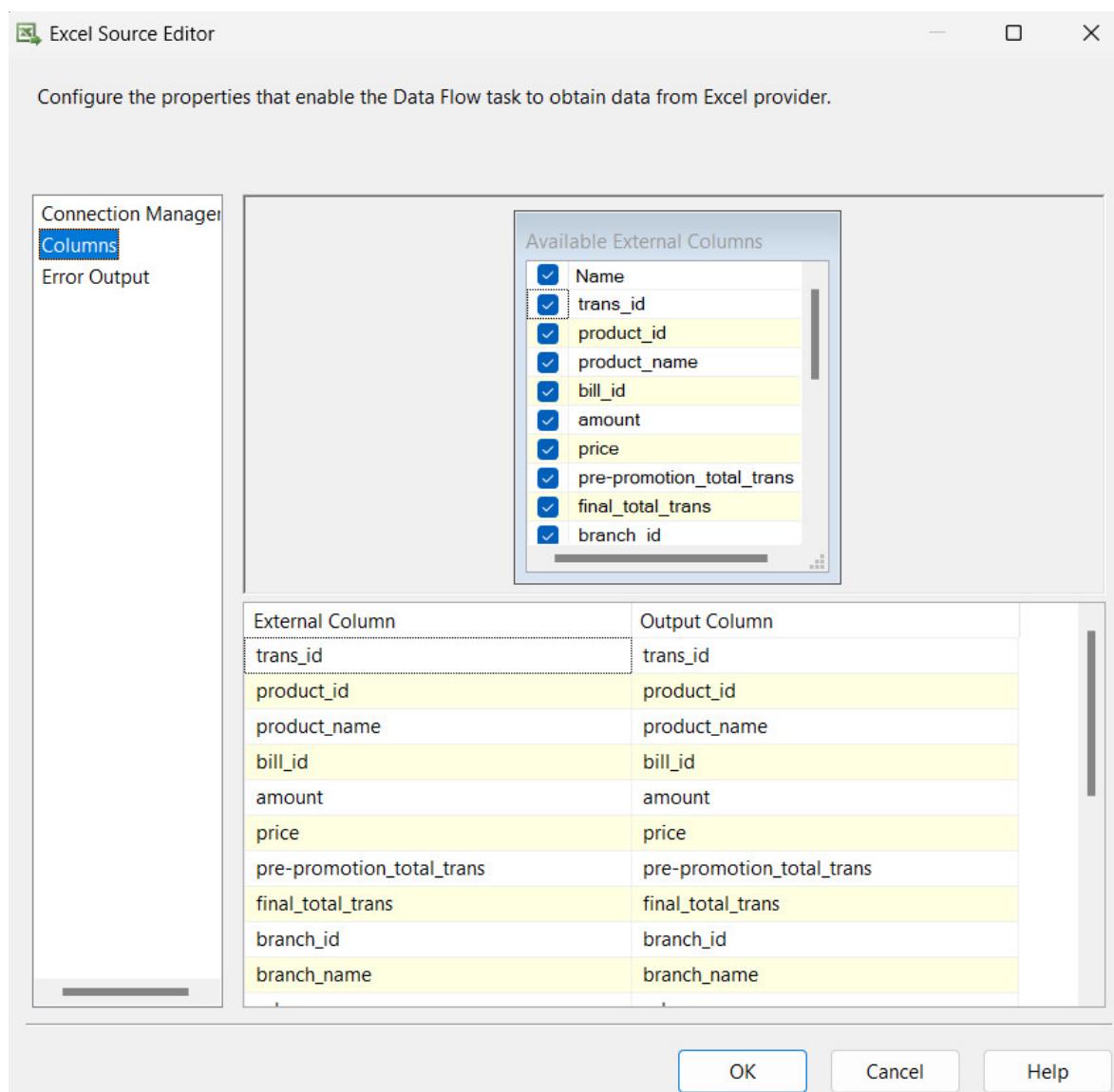
Hình 25: Giao diện lưu trữ file hiện ra chọn file dữ liệu gốc để tải lên

**Bước 4:** Chọn sheet lưu trữ dữ liệu dùng để đổ dữ liệu vào kho dữ liệu (SSIS).



Hình 26: Chọn sheet dữ liệu lưu trữ dữ liệu gốc

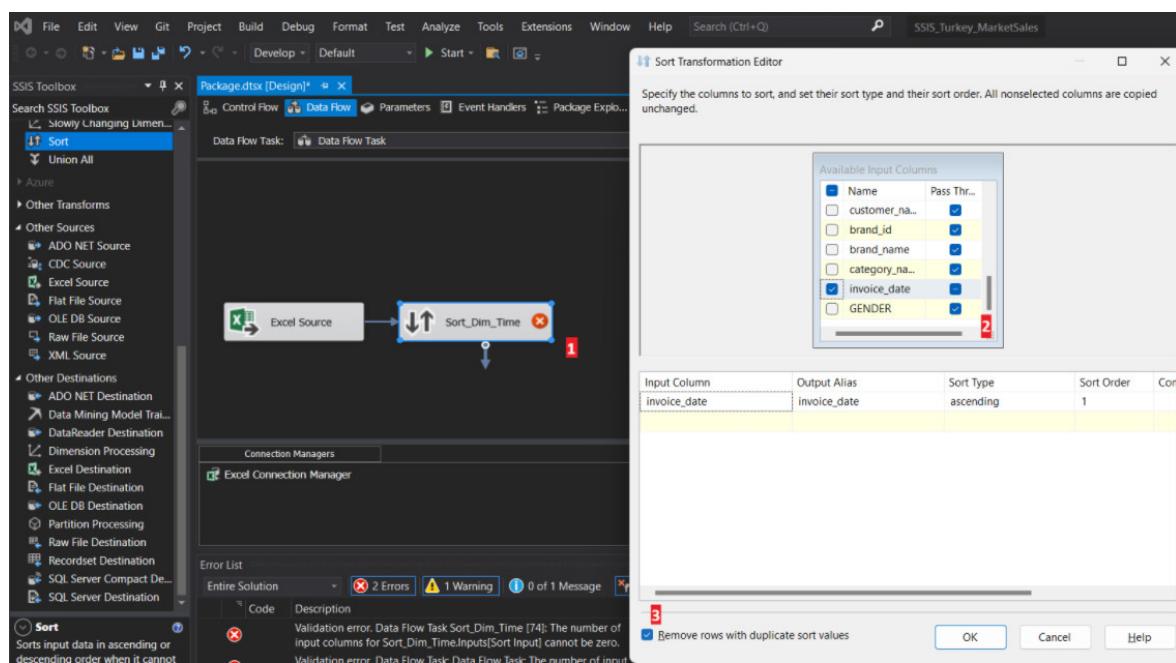
**Bước 5:** Kiểm tra các cột dữ liệu trong file dữ liệu được tải lên tại mục “**Columns**”. Click “**OK**” lần nữa để hoàn tất quá trình tải dữ liệu vào “**Excel Source**”.



Hình 27: Kiểm tra các cột dữ liệu trong file dữ liệu gốc

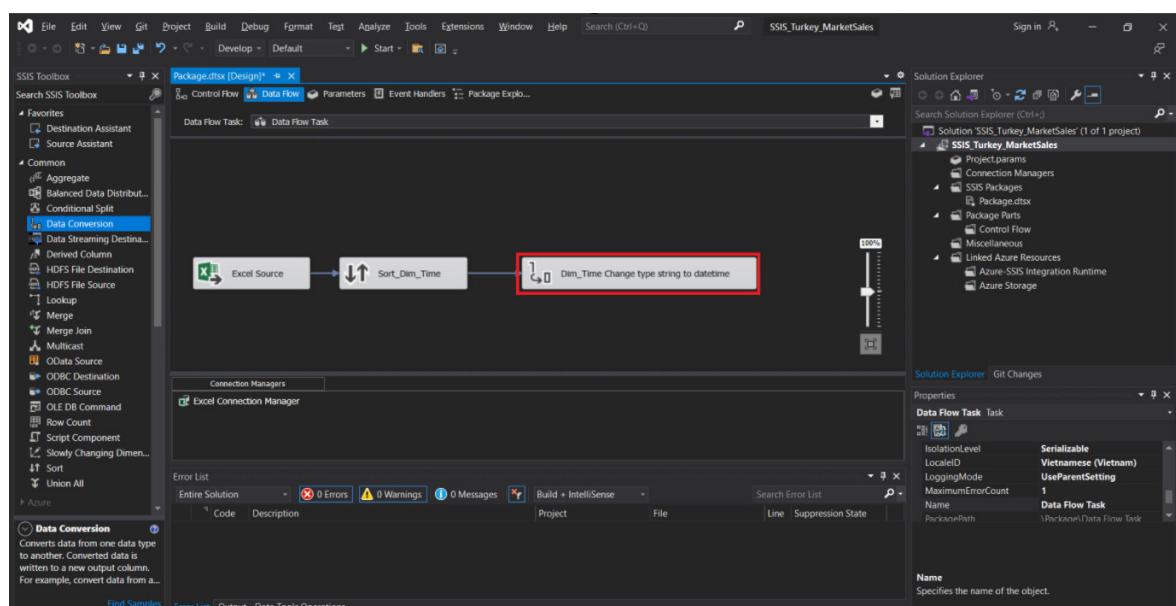
### 2.3.1.1. Bảng Dim\_Time

**Bước 1:** Tạo mới một “Sort” có tên là Sort\_Dim\_Time để lấy ra các cột dữ liệu cần thiết cho bảng Dim\_Time. Click chuột phải và chọn “Edit” để lấy cột invoice\_date làm cột dữ liệu duy nhất cho bảng Sort\_Dim\_Time. Tick chọn “Remove rows with duplicate sort values” để xóa các dòng dữ liệu trùng nhau và click “OK” để hoàn tất.



Hình 28: Lấy ra cột dữ liệu cần thiết cho bảng Dim\_Time

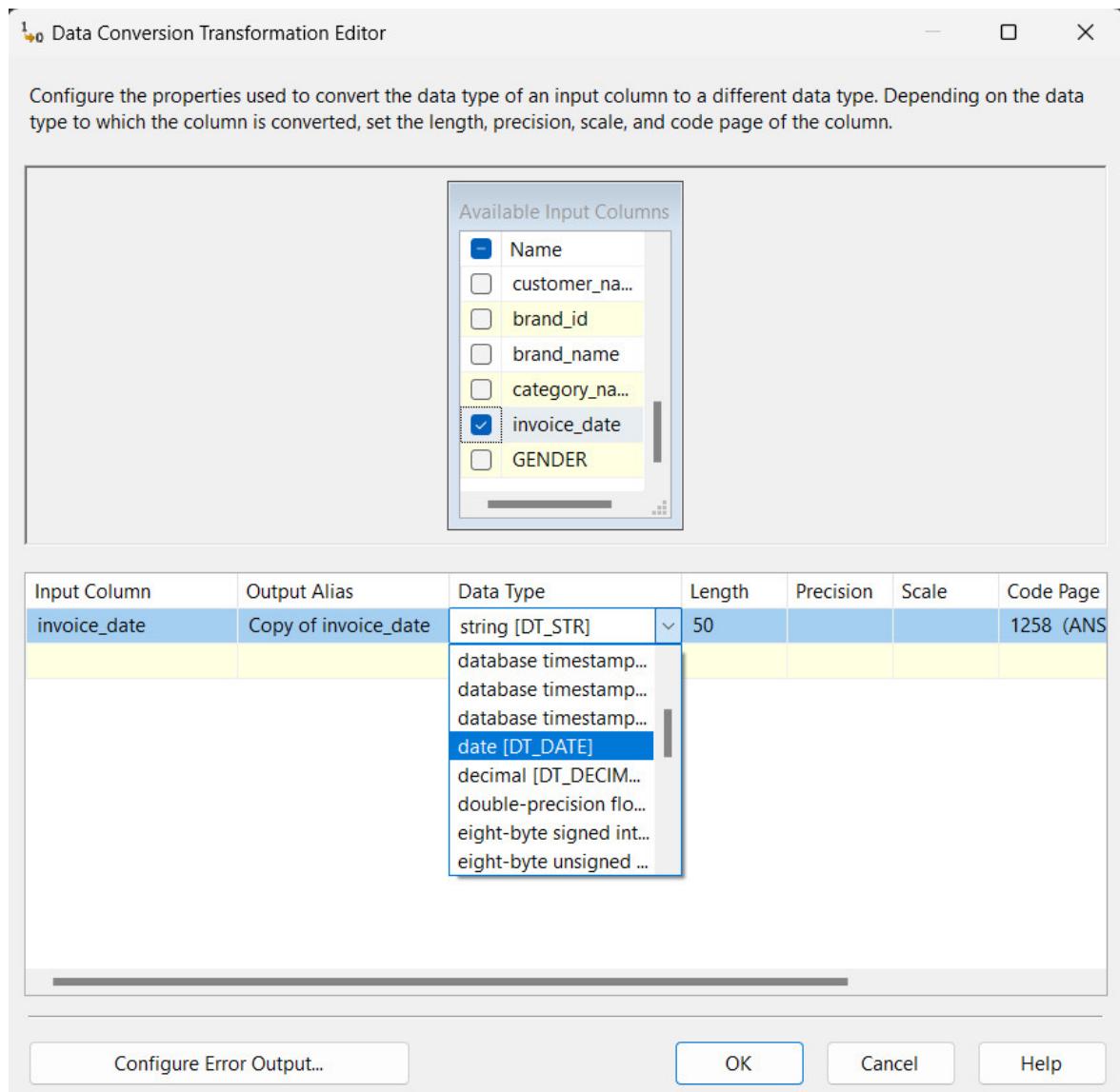
**Bước 2:** Kiểu dữ liệu DateTime khi lấy từ dữ liệu gốc mặc định là kiểu String. Ta sử dụng công cụ BI hỗ trợ là “**Data Conversion**” để chuyển đổi kiểu dữ liệu về đúng định dạng DateTime. Đặt tên cho Data Conversion là Dim\_Time Change type string to datetime.



Hình 29: Công cụ chuyển đổi kiểu dữ liệu của SSIS

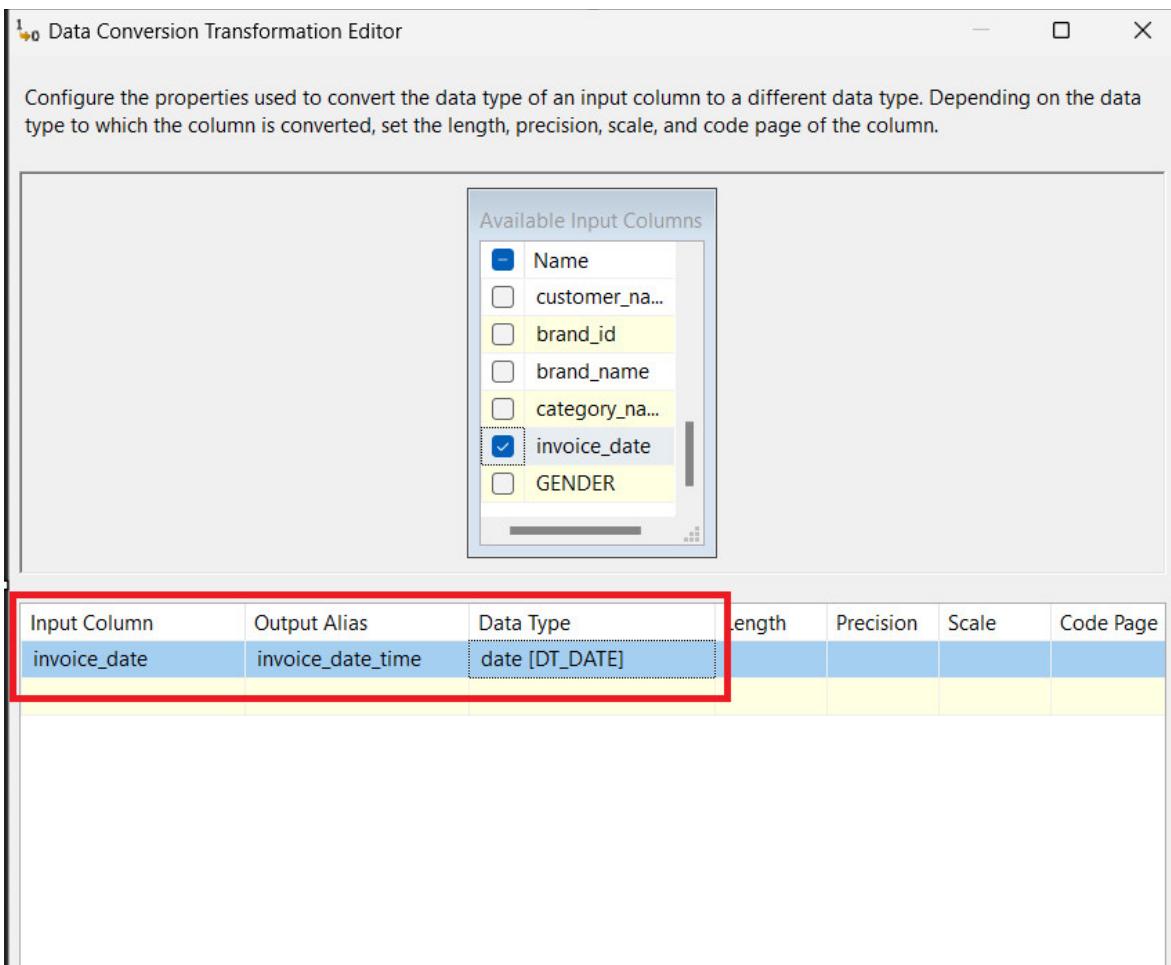
**Bước 3:** Click chuột phải vào “**Data conversion**” trên và chọn cột invoice\_date, ở cột Data Type ta nhận thấy kiểu dữ liệu mặc định là string

“string [DT\_STR]” , ta chọn lại kiểu dữ liệu cho cột là “**date [DT\_DATE]**”.



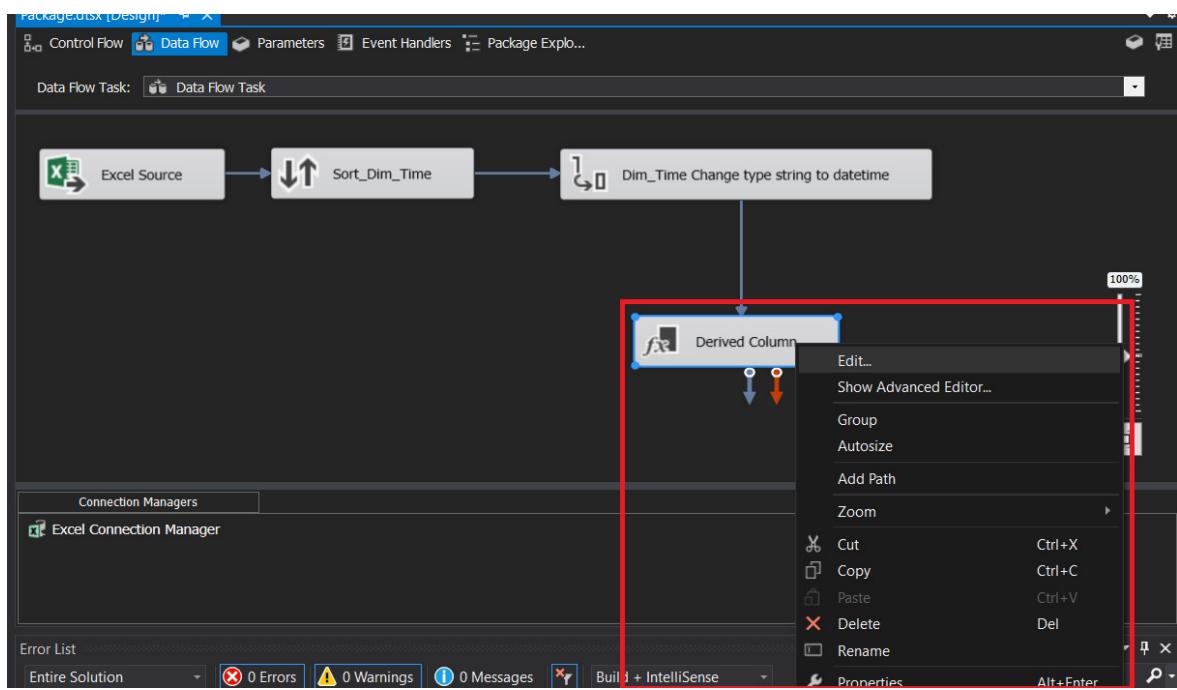
Hình 30: Chuyển đổi kiểu dữ liệu của cột `invoice_date`

**Bước 4:** Đặt lại “**Output Alias**” cho cột `invoice_date` từ “**Copy of invoice\_date**” thành “**invoice\_date\_time**”.



Hình 31: Đặt lại output alias cho cột invoice\_date

**Bước 5:** Thêm thành phần “Derived Column” và click “Edit” để tiến hành chia các cột dữ liệu cho bảng Dim\_Time.

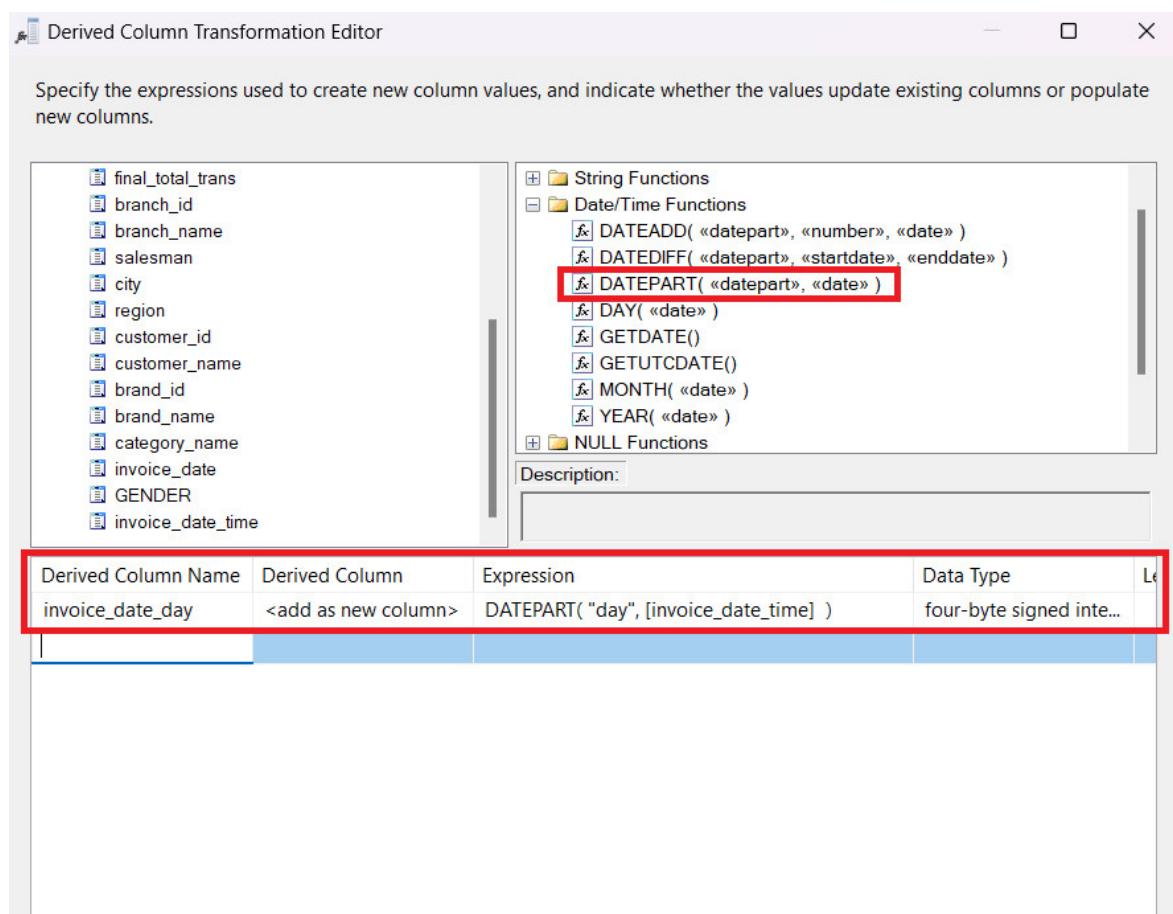


Hình 32: Tạo các cột dữ liệu cho bảng Dim\_Time từ cột invoice\_date

**Bước 6:** Chia dữ liệu từ cột **invoice\_date\_time** có kiểu dữ liệu

dd/MM/yyyy HH:mm thành các cột dữ liệu `invoice_date_hour`,  
`invoice_date_minute`, `invoice_date_day`, `invoice_date_month`,  
`invoice_date_year`, `invoice_date_dayofweek`.

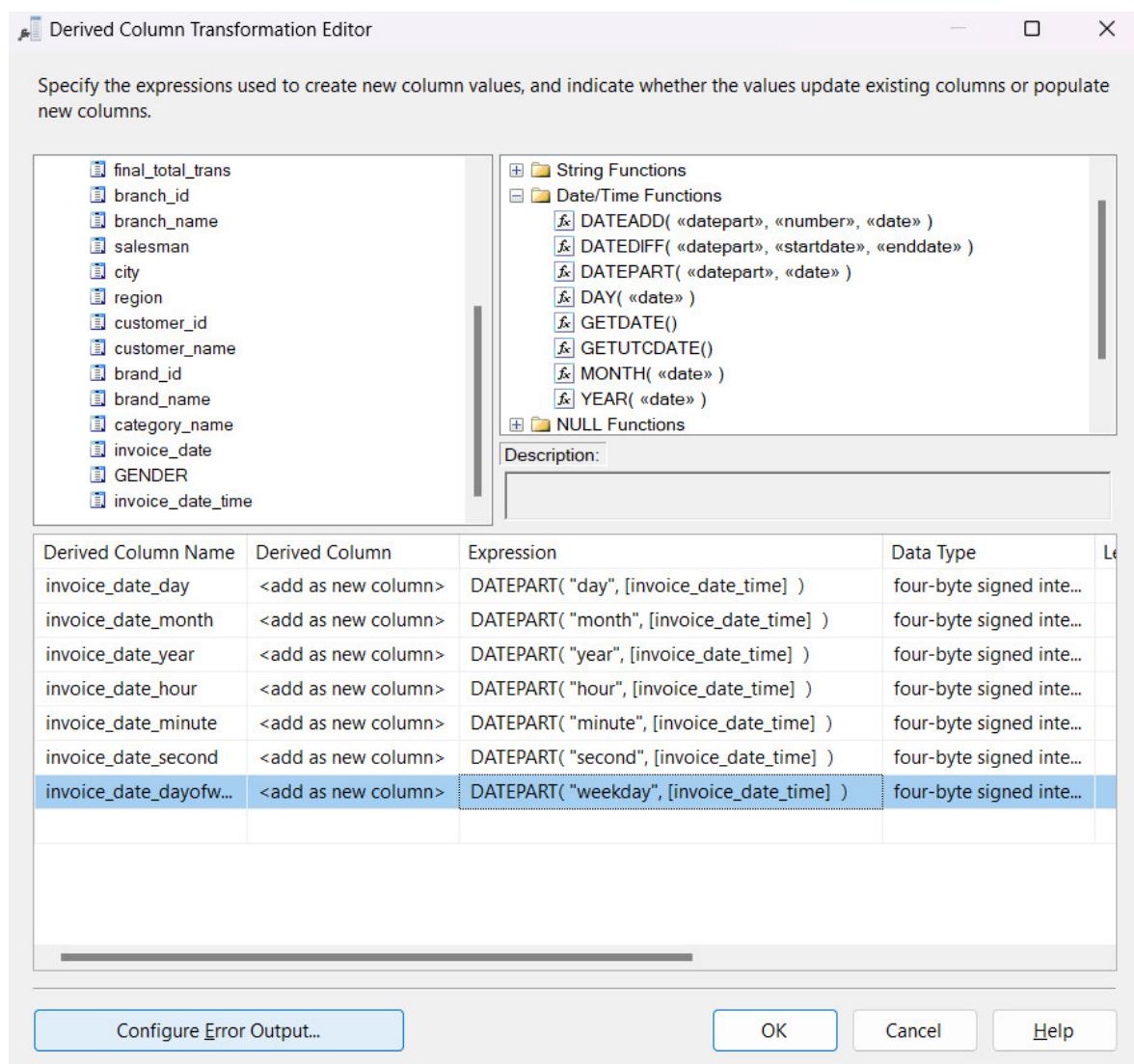
Chọn phương thức **DATEPART(<kiểu dữ liệu thời gian>, cột dữ liệu)**. Ở đây ta chia `invoice_date_time` thành cột `invoice_date_day` nên ta cài đặt: `DATEPART("day", [invoice_date_time])`.



Hình 33: Sử dụng hàm DATEPART để lấy ra ngày từ giá trị cột invoice\_date

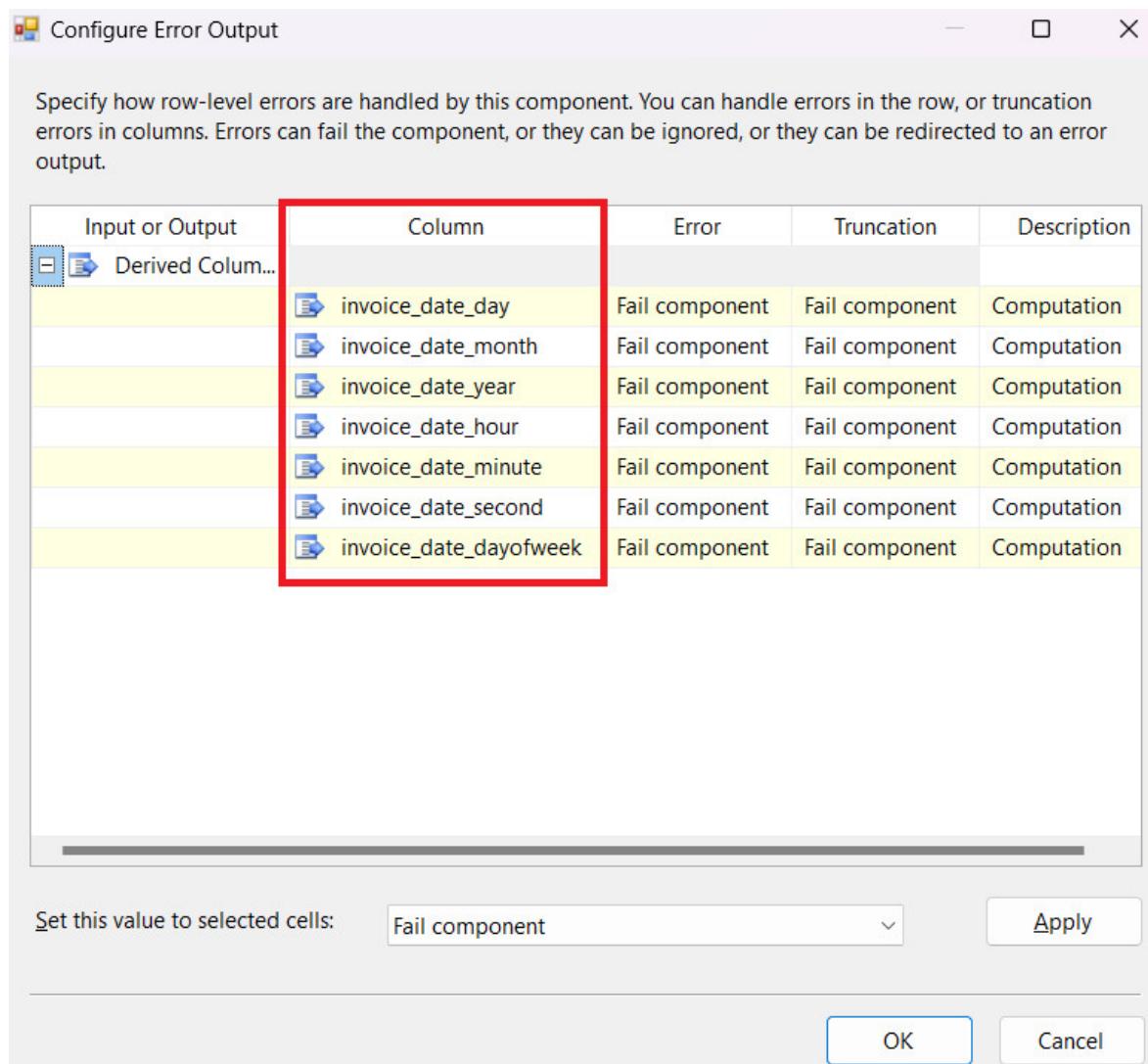
**Bước 7:** Tương tự cách tạo ra cột invoice\_date\_day, ta tiếp tục chia cột invoice\_date\_time thành các cột còn lại: invoice\_date\_hour, invoice\_date\_minute, invoice\_date\_second, invoice\_date\_month, invoice\_date\_year, invoice\_date\_dayofweek.

Tiếp theo chọn “Configure Error Output”.



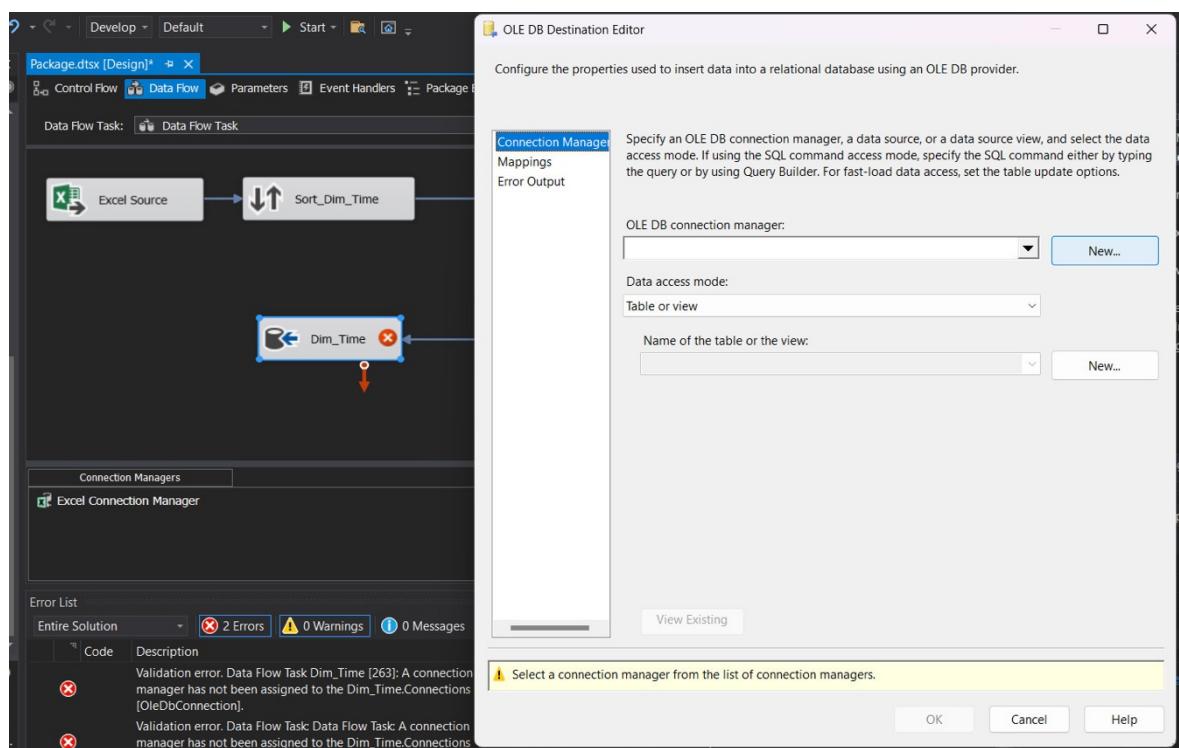
Hình 34: Sử dụng hàm DATEPART để lấy ra tháng, năm, giờ, phút, giây, ngày trong tuần

**Bước 8:** Ta kiểm tra thấy từ 1 cột dữ liệu ban đầu là invoice\_date\_time, lúc này có 7 cột được tạo ra ứng với mô hình kho dữ liệu bảng Dim\_Time. Click “OK” để hoàn tất quá trình chia các cột dữ liệu.



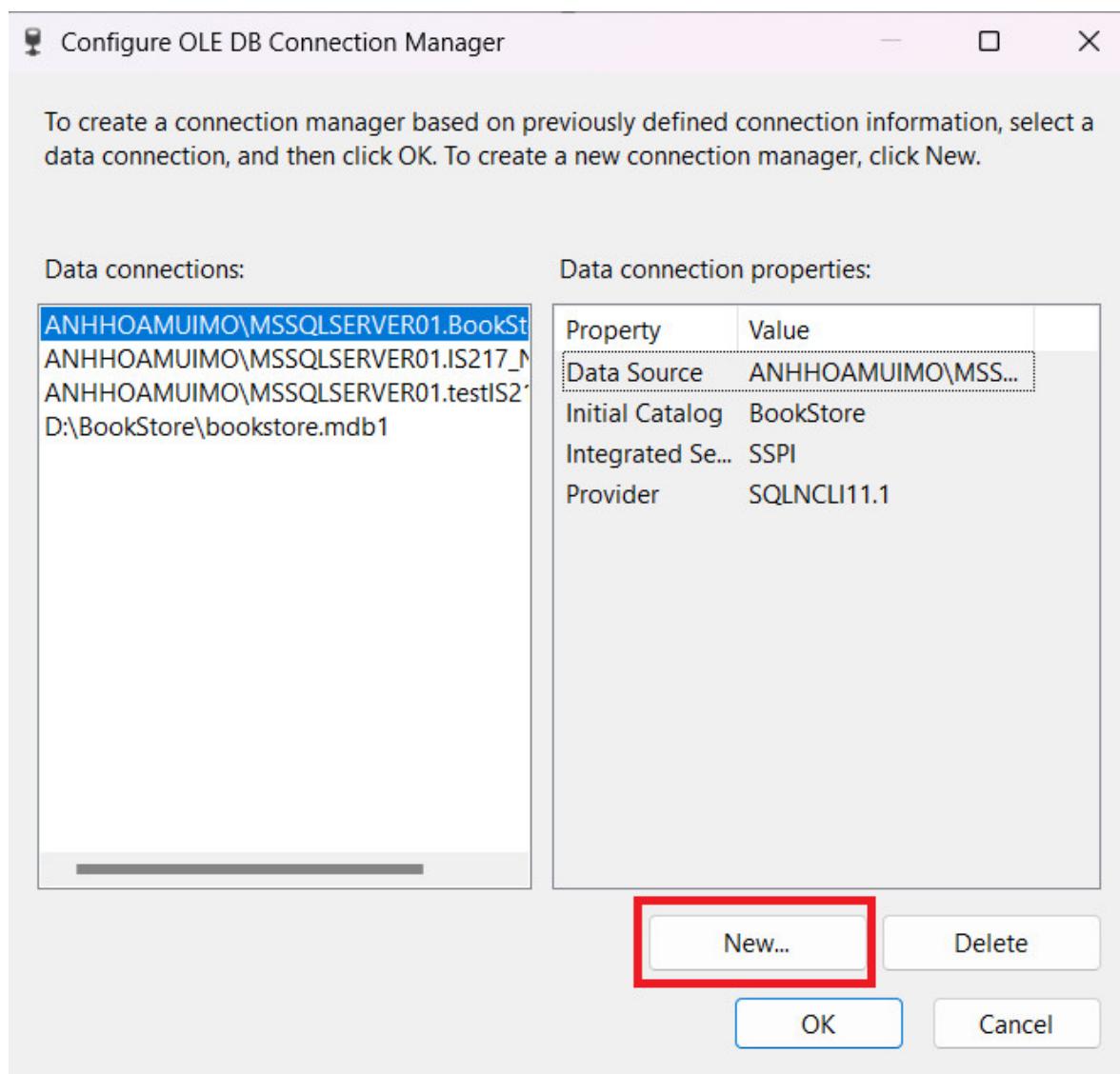
Hình 35: Kiểm tra kết quả các cột mới được tạo từ cột invoice\_date

**Bước 9:** Tạo Dim\_Time từ một “**OLE DB Destination**”. Double Click vào OLE DB Destination để tạo một kết nối mới đến Microsoft SQL Server.



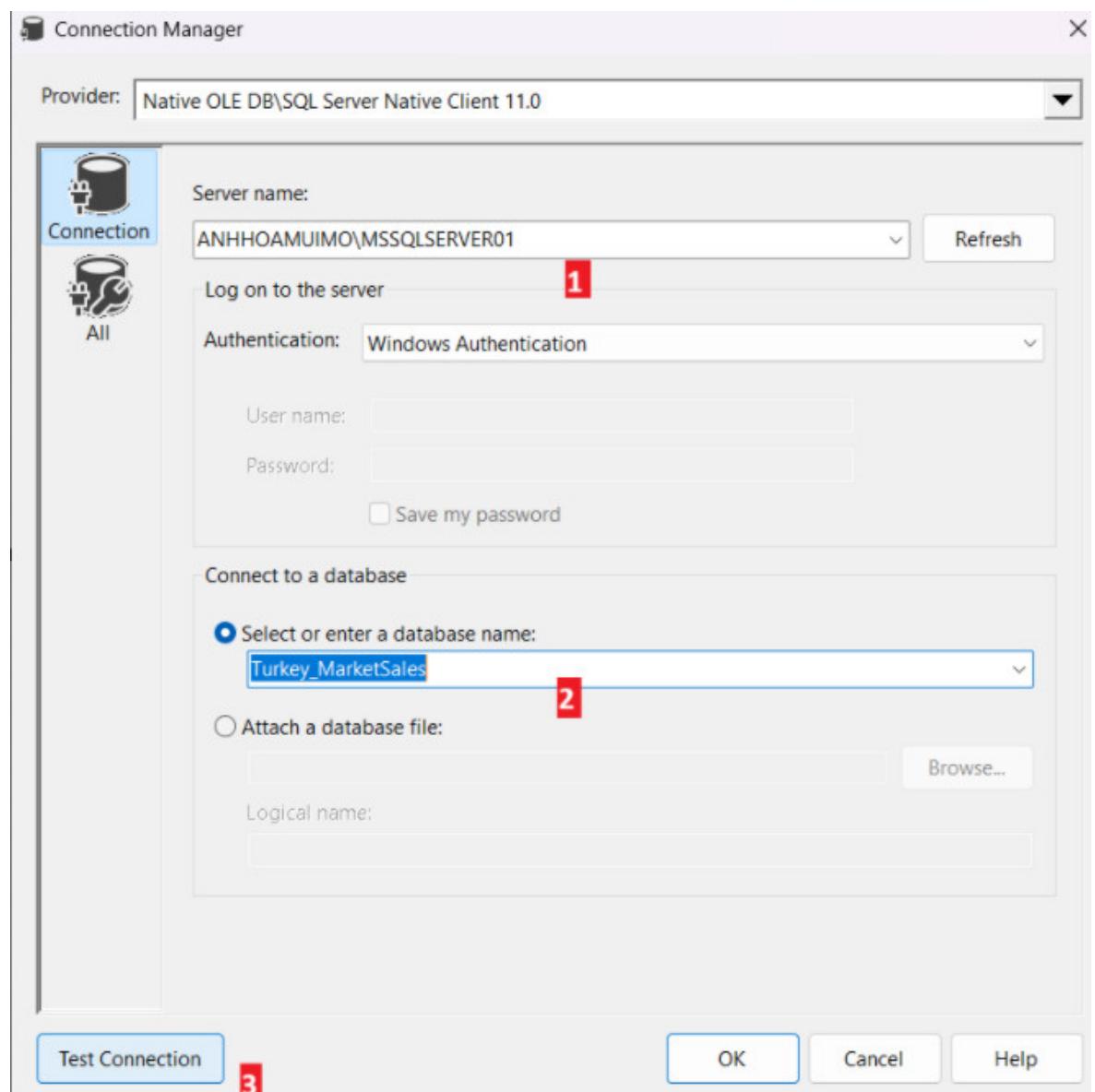
Hình 36: Tạo mới một OLE DB Destination để đổ dữ liệu vừa tạo.

Tiếp tục chọn “New....” Để tạo một kết nối mới:



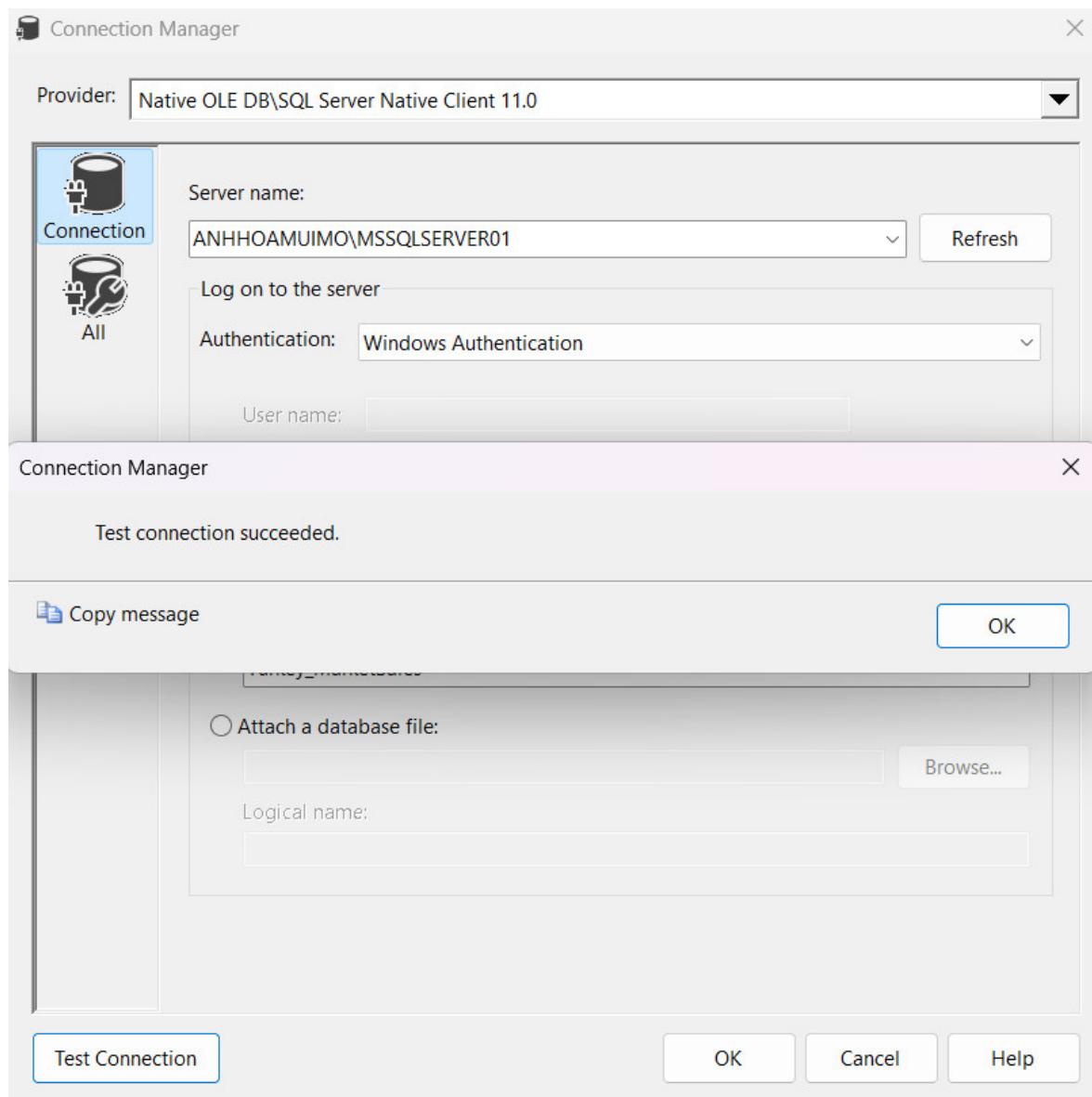
Hình 37: Tạo một kết nối mới đến Microsoft SQL Server

**Bước 10:** Chọn tên server trùng với tên server của Microsoft SQL Server để ta kết nối đến kho dữ liệu Turkey\_MarketSales vừa tạo ở trên. Kết nối đến server bằng tài khoản windows mặc định (Windows Authentication).



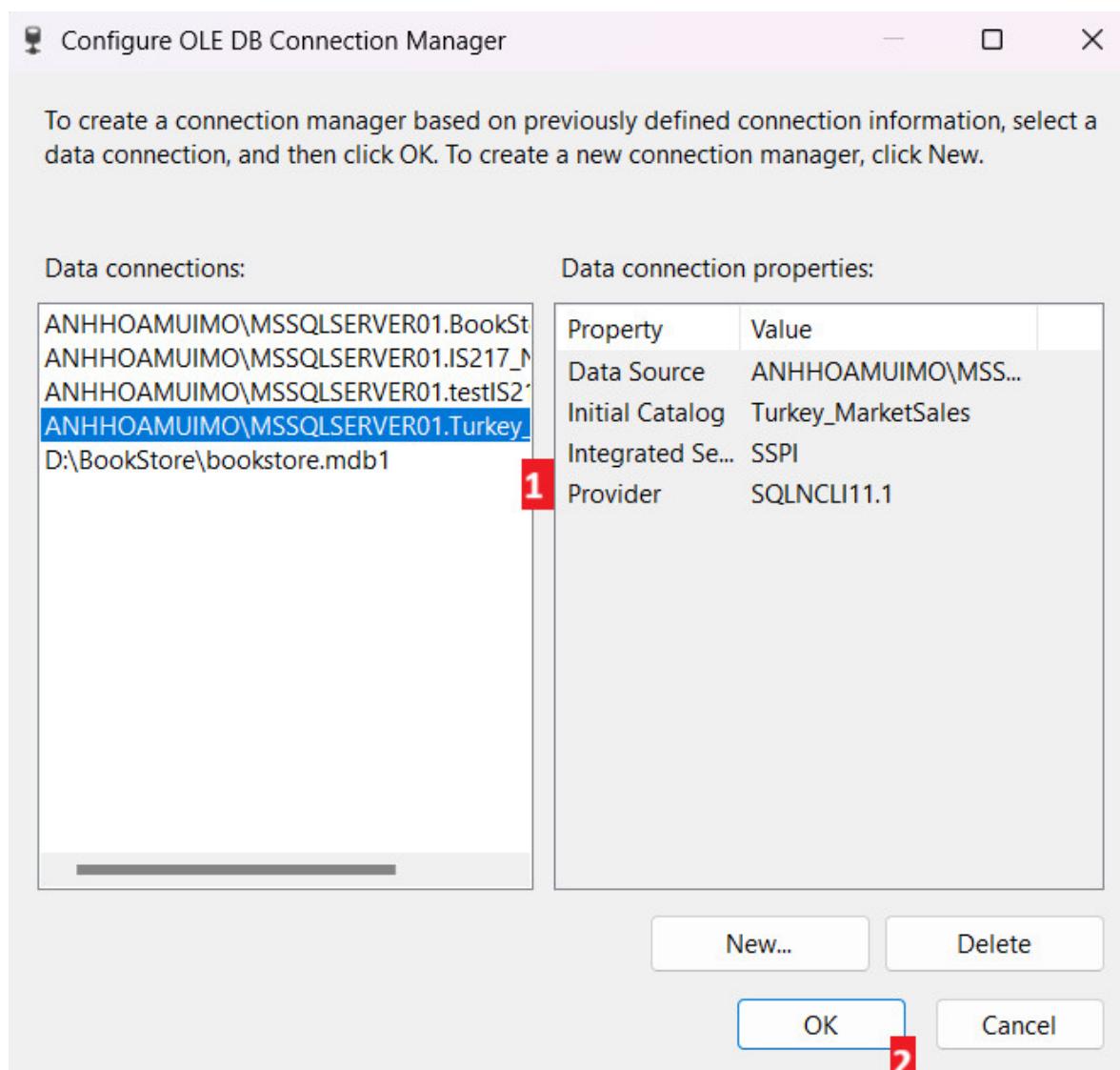
Hình 38: Chọn tên server lưu trữ kho dữ liệu được tạo ở Hình 19

Click “Test Connection” để kiểm tra kết nối.



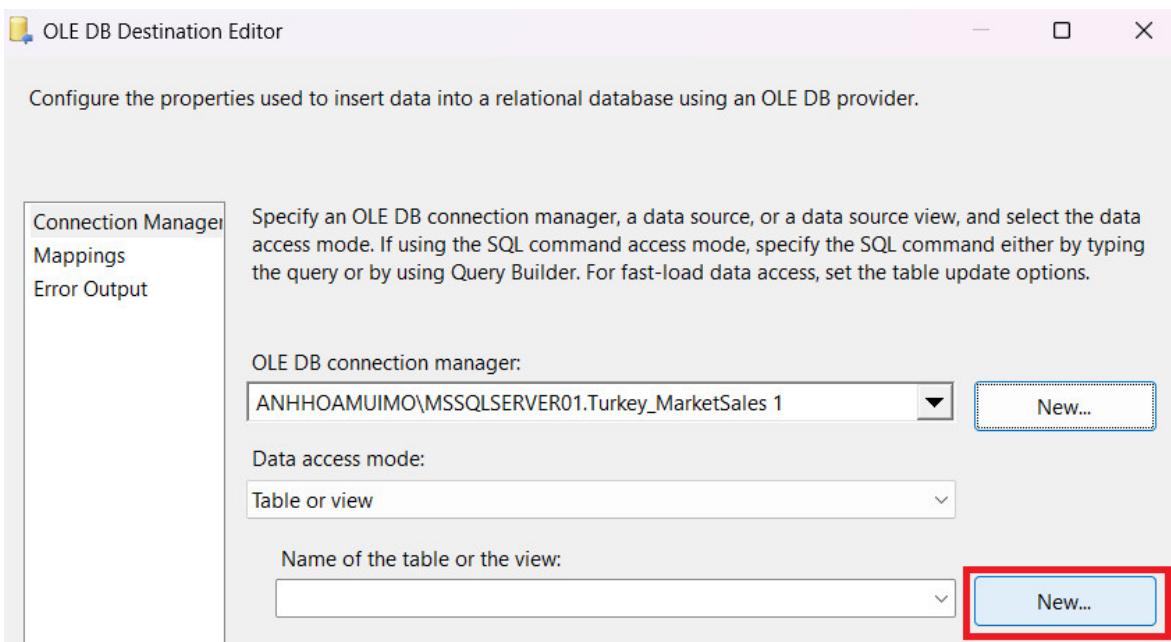
Hình 39: Kiểm tra kết nối đến SQL Server

**Bước 11:** Chọn kết nối vừa tạo đến Microsoft SQL Server và click “OK”.



Hình 40: Hoàn tất quá trình tạo mới kết nối đến SQL Server

**Bước 12:** Chọn “New....” để tạo mới bảng.



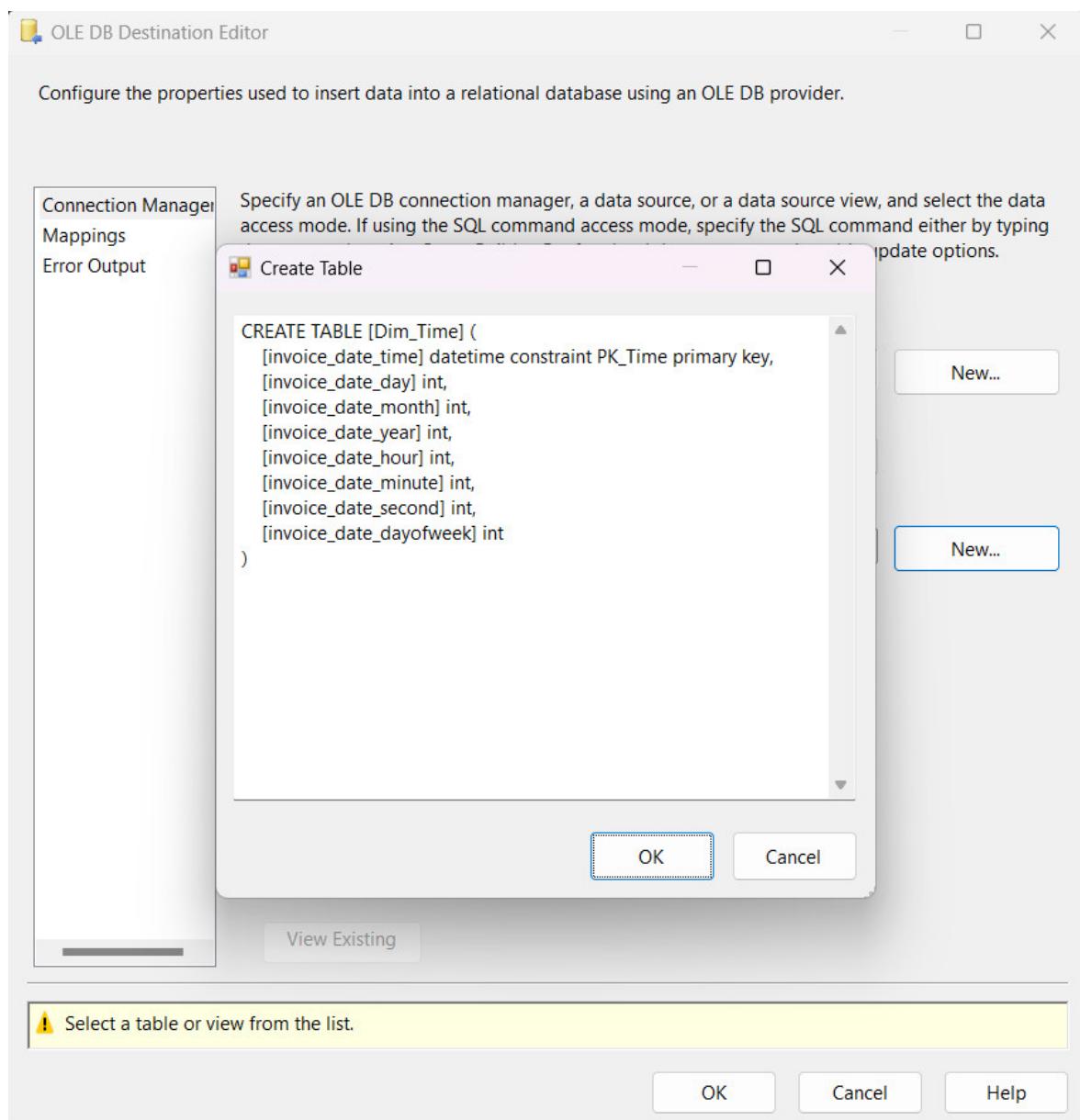
Hình 41: Tạo mới một bảng Dim\_Time trong cơ sở dữ liệu Turkey\_MarketSales

Nội dung câu lệnh SQL tạo bảng Dim\_Time như sau:

```
CREATE TABLE [Dim_Time] (
```

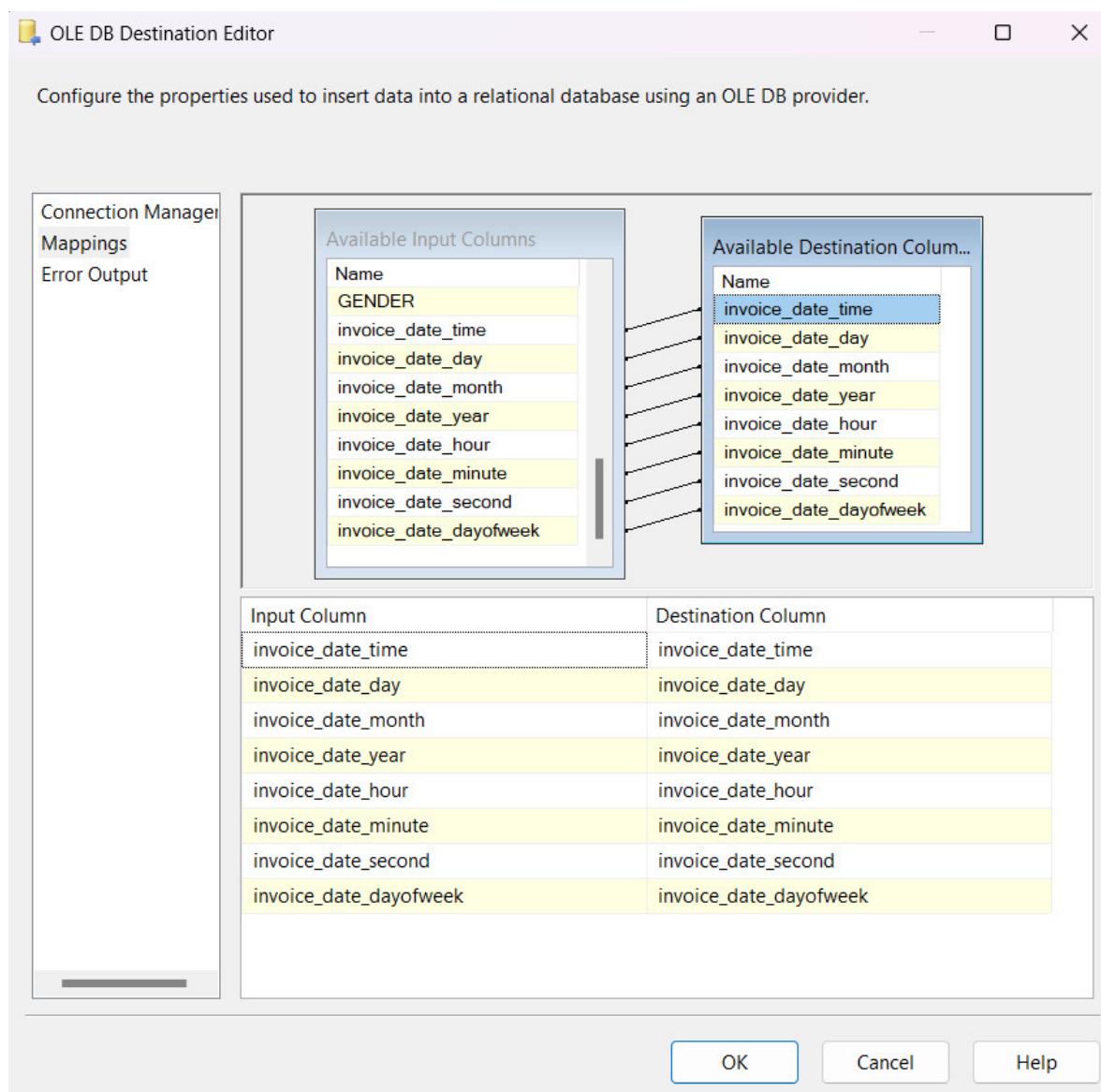
```
    [invoice_date_time] datetime constraint PK_Time primary key,  
    [invoice_date_day] int,  
    [invoice_date_month] int,  
    [invoice_date_year] int,  
    [invoice_date_hour] int,  
    [invoice_date_minute] int,  
    [invoice_date_second] int,  
    [invoice_date_dayofweek] int
```

```
)
```



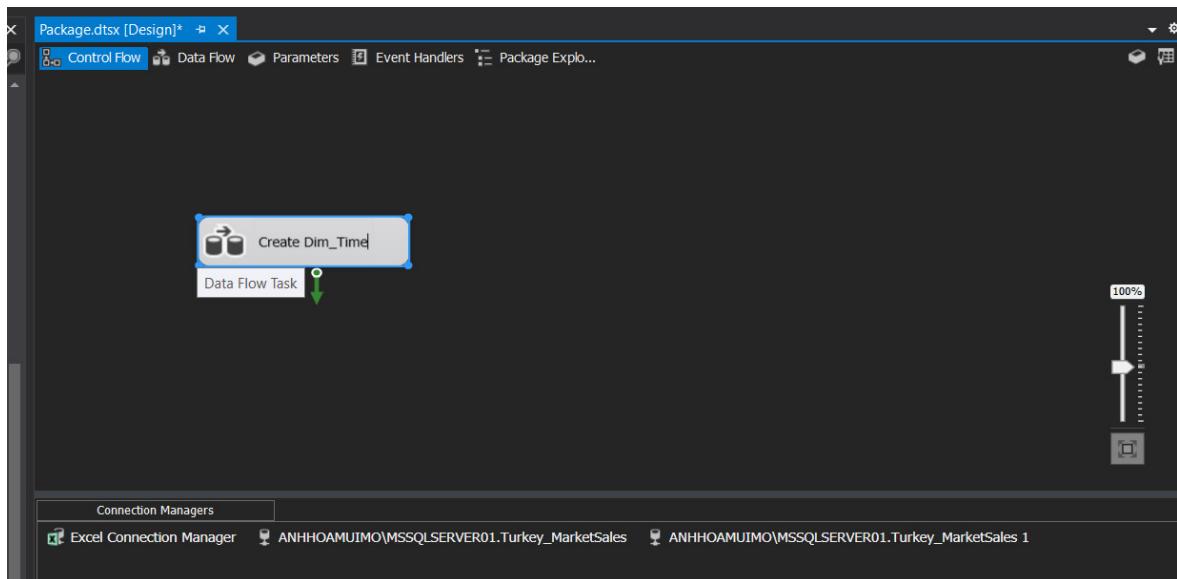
Hình 42: Nhập các câu lệnh SQL tương ứng của bảng Dim\_Time

**Bước 13:** Tiếp đến ta cần chọn mục “Mapping” để xem xét việc ánh xạ các cột dữ liệu. Click “OK” để hoàn tất thiết lập.



Hình 43: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 35.

Đổi tên “Data Flow Task” thành Create Dim\_Time.

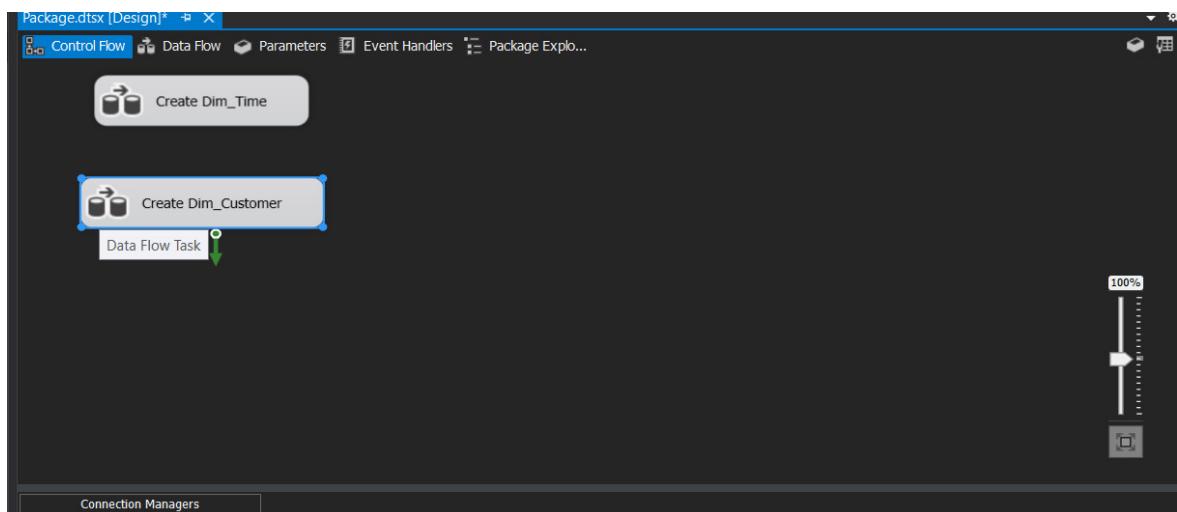


Hình 44: Đổi tên toàn bộ Data Flow Task (quá trình trên) thành Create Dim\_Time

### 2.3.1.2. Bảng Dim\_Customer:

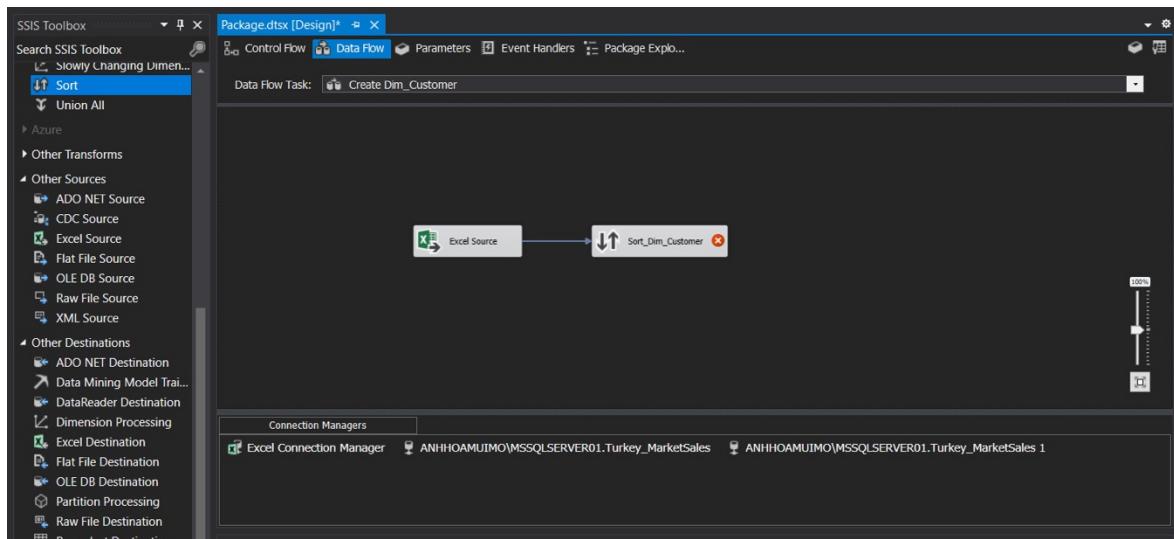
**Bước 1:** Tạo “Data Flow Task”, đổi tên thành

Create\_Dim\_Customer



Hình 45: Tạo mới Data Flow Task cho bảng Dim\_Customer

**Bước 2:** Trong Create\_Dim\_Customer, tạo “Excel Source” và chọn file dữ liệu nguồn, chọn một Sort để tạo mới Sort\_Dim\_Customer cho bảng Dim\_Customer



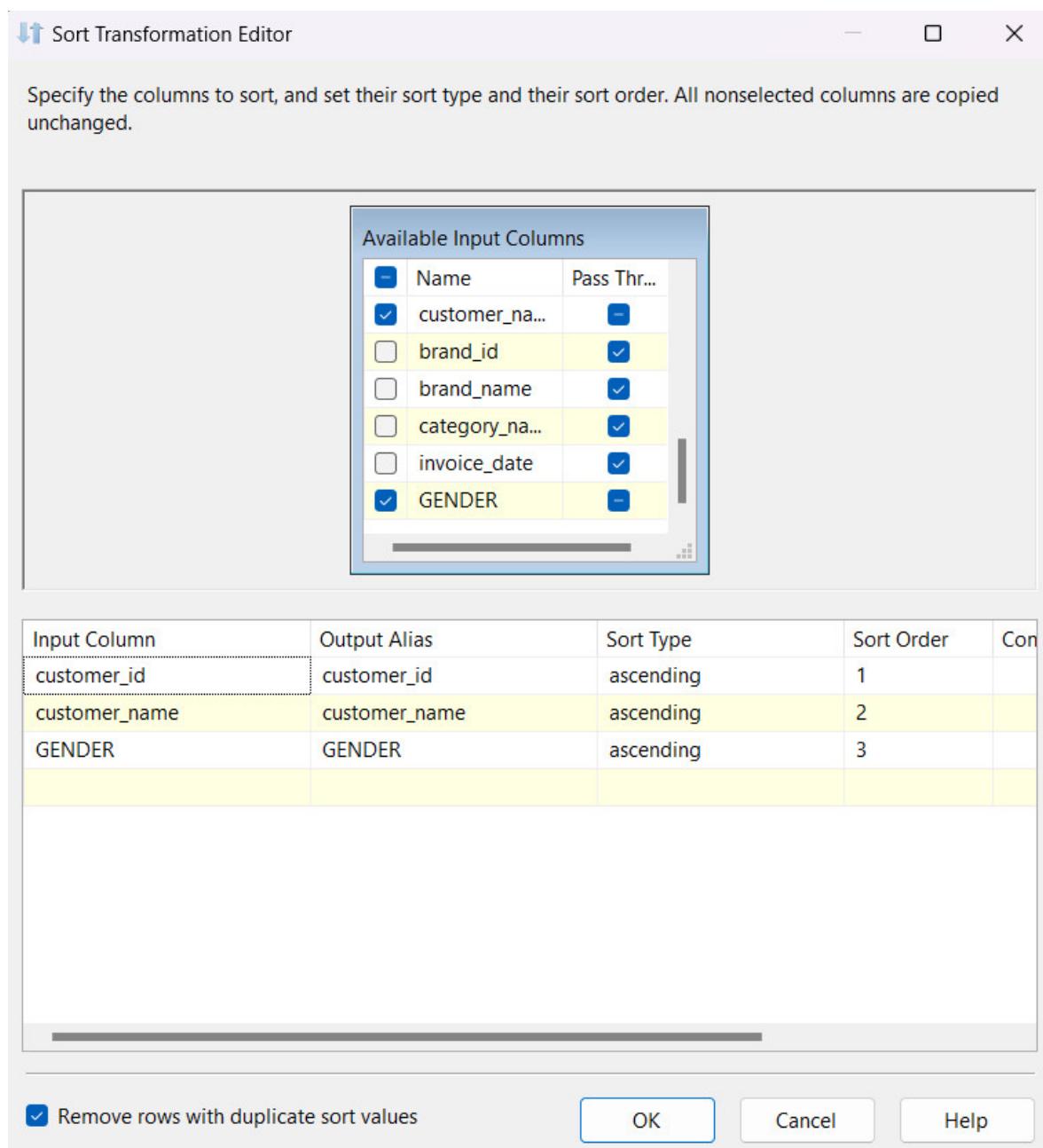
Hình 46: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu

Dim_Customer	
PK	customer_id string
	customer_name string
	gender string

Hình 47: Thành phần của bảng Dim\_Customer

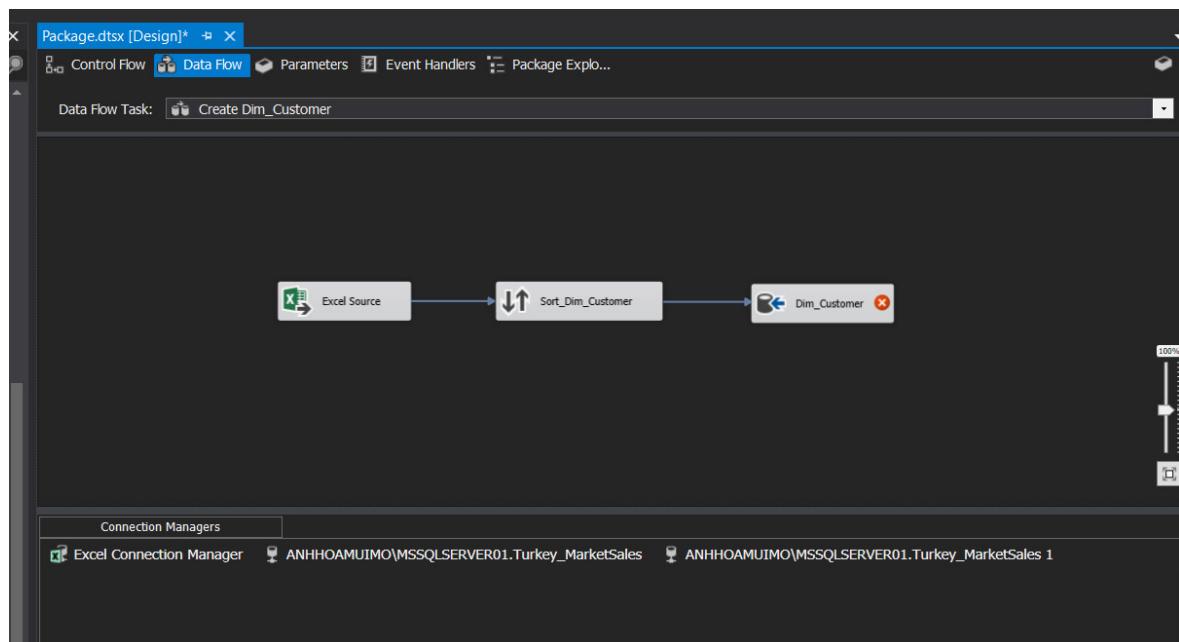
Click chuột phải vào Sort\_Dim\_Customer chọn “Edit”: lần lượt chọn các cột customer\_id, customer\_name, GENDER làm các cột để đồ dữ liệu vào Sort\_Dim\_Customer.

Tick chọn “Remove rows with duplicate sort values” xóa đi các dòng dữ liệu trùng nhau và sau đó chọn “OK”.



Hình 48: Chọn ra các cột dữ liệu cho bảng Dim\_Customer

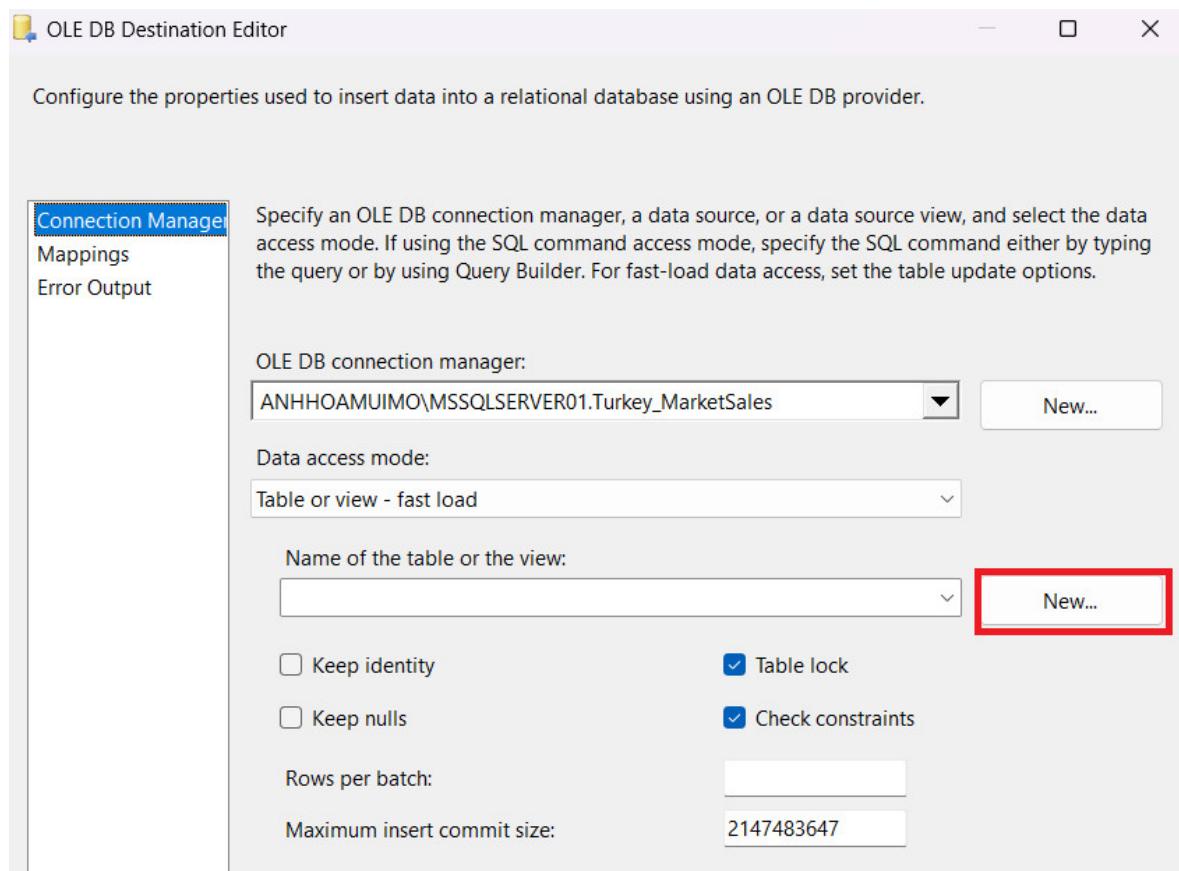
**Bước 3:** Tạo mới một “OLE DB Destination” để đổ dữ liệu gốc sau khi đã được xử lý vào trong kho dữ liệu Turkey\_MarketSales.



Hình 49: Tạo mới một OLE DB Destination đặt tên là Dim\_Customer

**Bước 4:** Connect đến kho dữ liệu đã được tạo khi tạo bảng

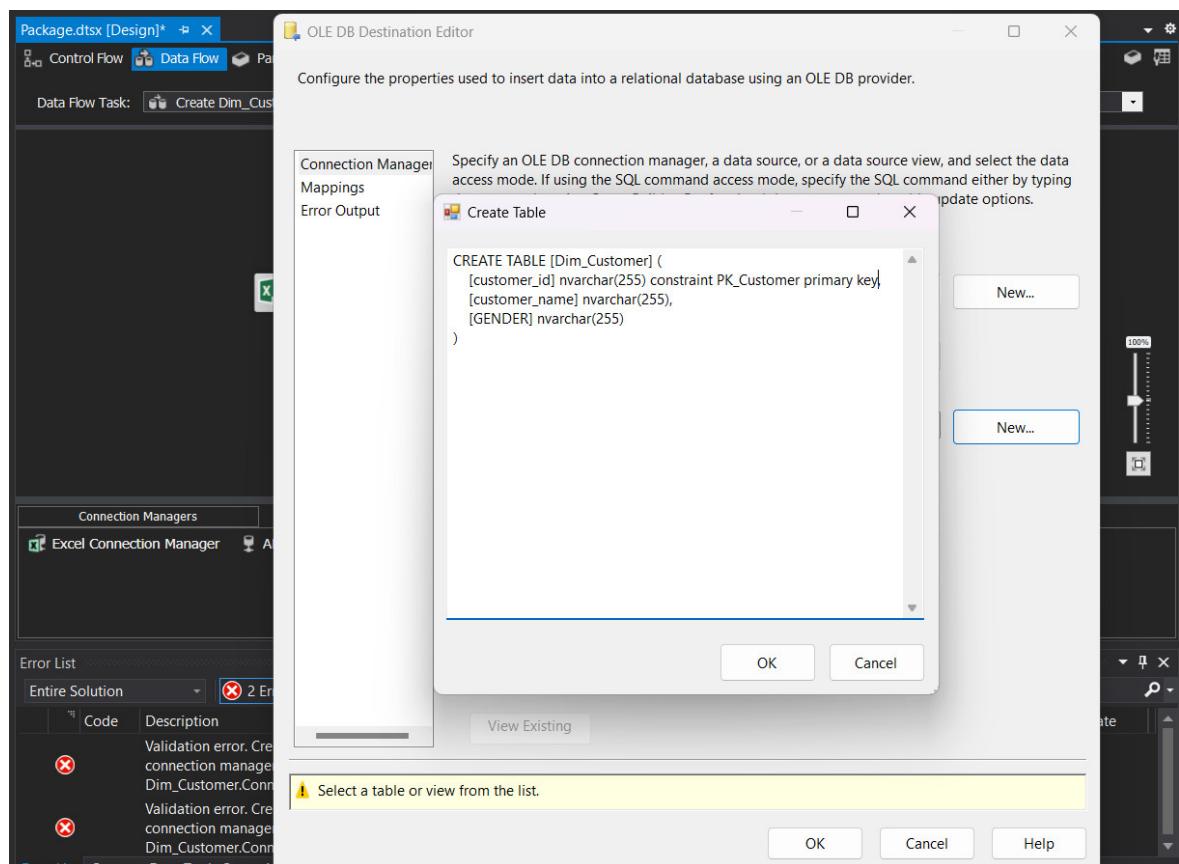
Dim\_Time, vì vậy ta chỉ cần chọn “New....” để tạo bảng Dim\_Customer.



Hình 50: Kết nối đến kho dữ liệu Turkey\_MarketSales và tạo bảng Dim\_Customer

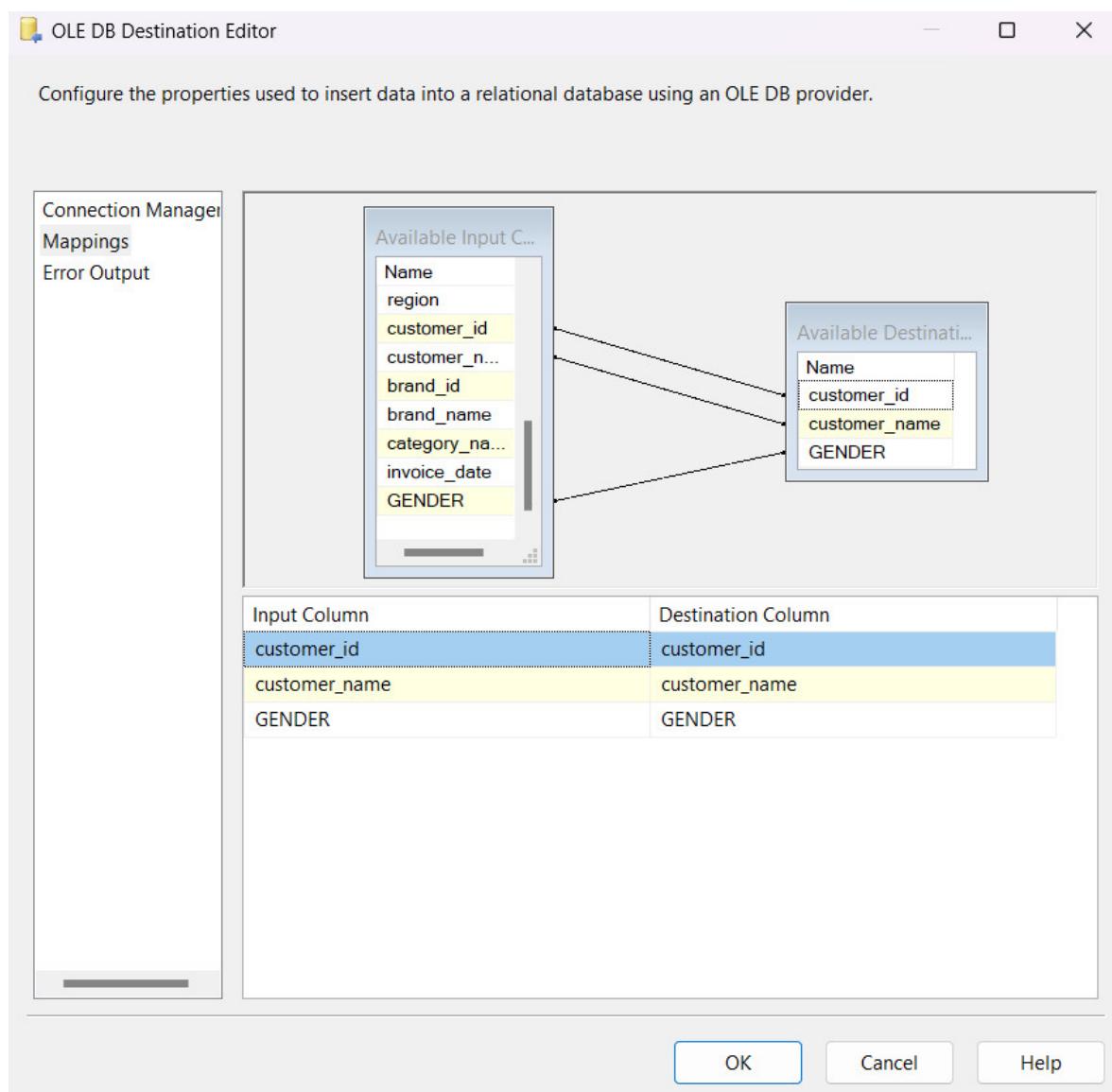
Nội dung câu lệnh SQL tạo bảng Dim\_Customer như sau:

```
CREATE TABLE [Dim_Customer] (
    [customer_id] nvarchar(255) constraint PK_Customer primary key,
    [customer_name] nvarchar(255),
    [GENDER] nvarchar(255)
)
```



Hình 51: Nhập các câu lệnh SQL tương ứng với bảng Dim\_Customer

**Bước 5:** Tiếp đến ta cần chọn mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Click “OK” để hoàn tất thiết lập.

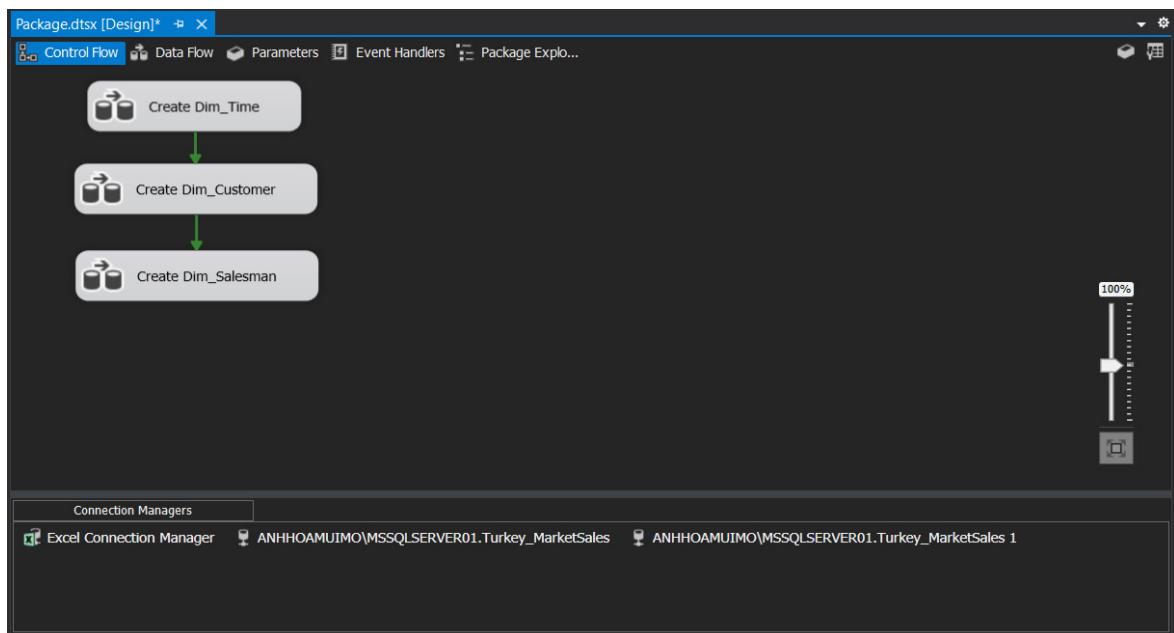


Hình 52: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 47

### 2.3.1.3. Bảng Dim\_Salesman:

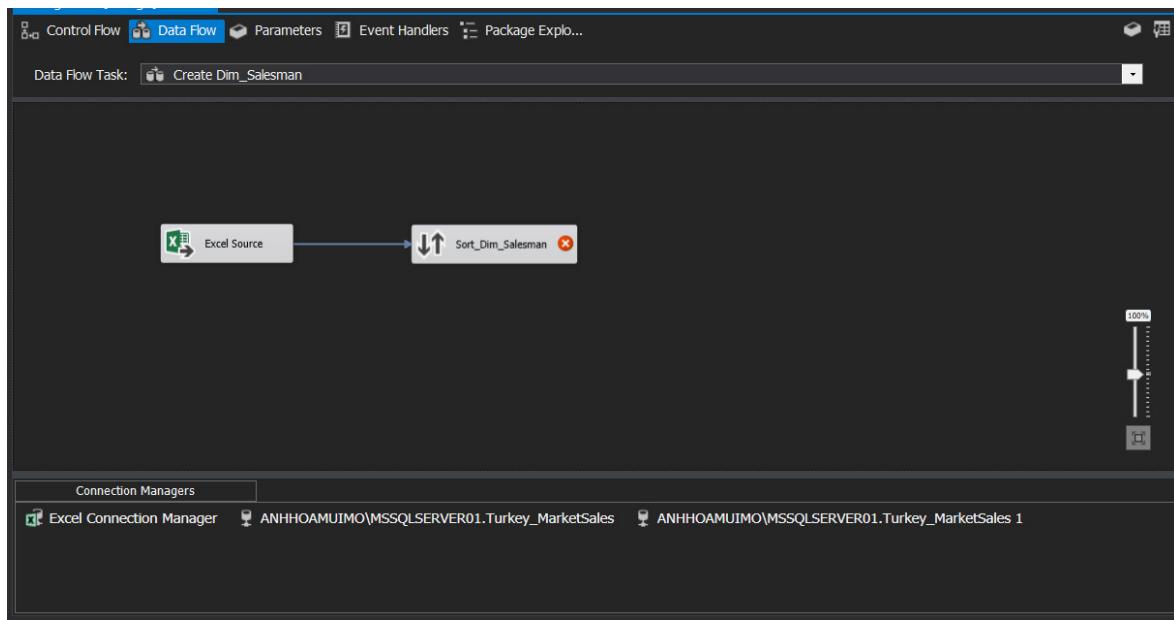
Bước 1: Tạo “Data Flow Task”, đổi tên thành

Create\_Dim\_Salesman



Hình 53: Tạo mới Data Flow Task cho bảng Dim\_Salesman

**Bước 2:** Trong Create\_Dim\_Salesman, tạo “Excel Source” và chọn file dữ liệu nguồn, chọn một Sort để tạo mới Sort\_Dim\_Salesman cho bảng Dim\_Salesman.



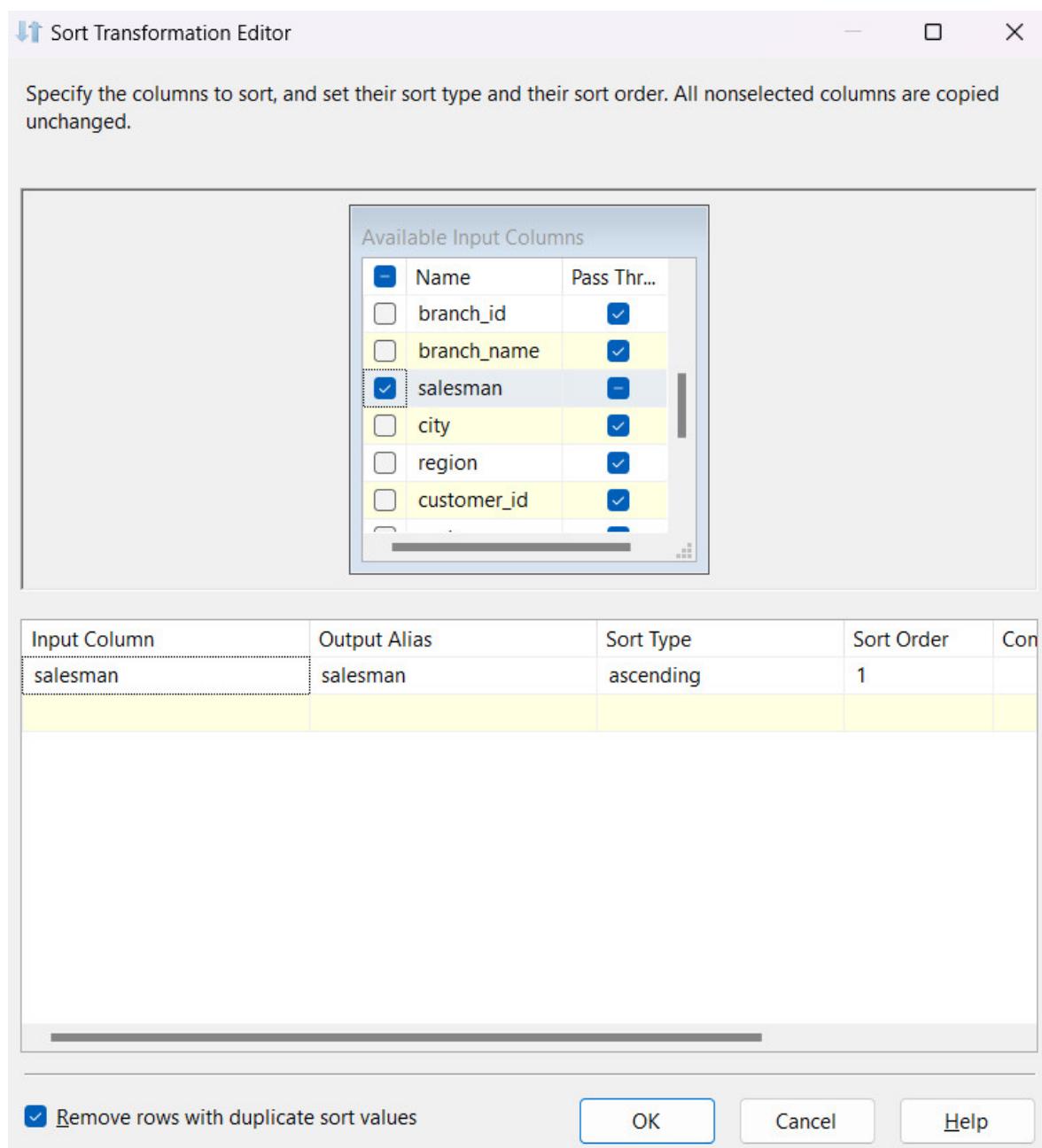
Hình 54: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu

Dim_Salesman	
► PK	<u>salesman_id int</u>
	salesman_name string

Hình 55: Thành phần của bảng Dim\_Salesman

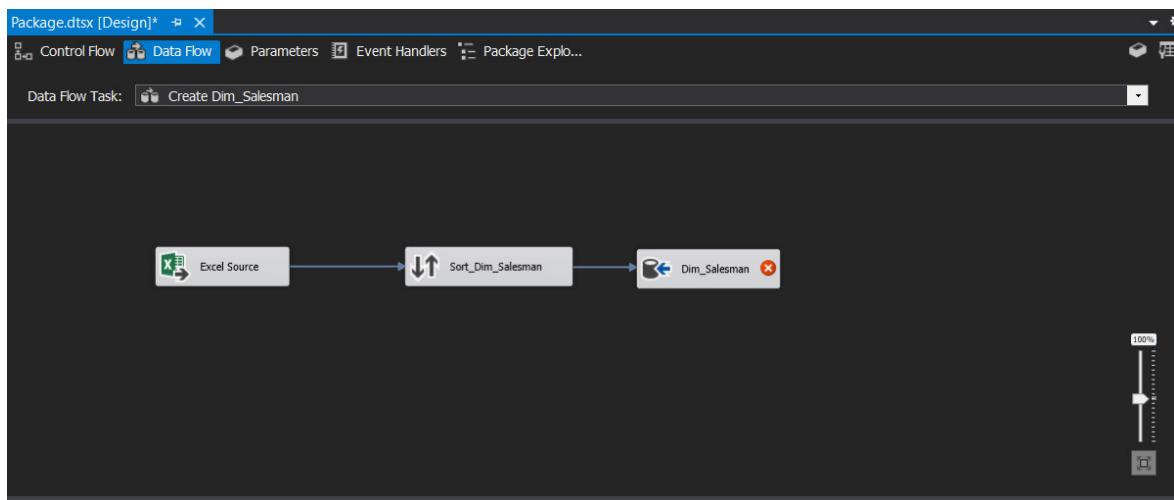
Click chuột phải vào Sort\_Dim\_Salesman chọn “Edit”: lần lượt chọn các cột salesman làm các cột để đổ dữ liệu vào Sort\_Dim\_Salesman.

Tick chọn “**Remove rows with duplicate sort values**” xóa đi các dòng dữ liệu trùng nhau và sau đó chọn “OK”.



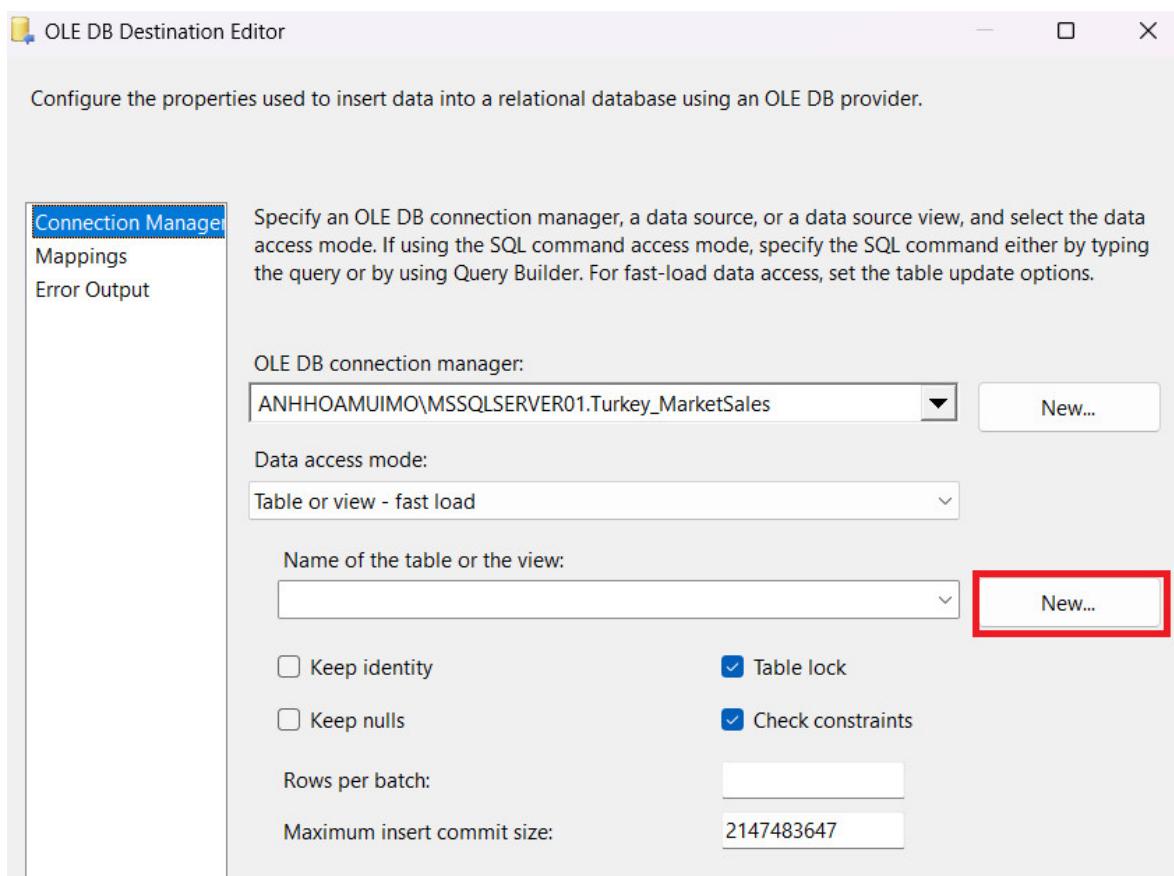
Hình 56: Chọn ra các cột dữ liệu cho bảng Dim\_Customer

**Bước 3:** Tạo mới một “OLE DB Destination” để đổ dữ liệu gốc sau khi đã được xử lý vào trong kho dữ liệu Turkey\_MarketSales.



Hình 57: Tạo mới một OLE DB Destination đặt tên là Dim\_Salesman

**Bước 4:** Connect đến kho dữ liệu đã được tạo khi tạo bảng Dim\_Time, vì vậy ta chỉ cần chọn “New....” để tạo bảng Dim\_Salesman.



Hình 58: Kết nối đến kho dữ liệu Turkey\_Marketales và tạo bảng Dim\_Salesman

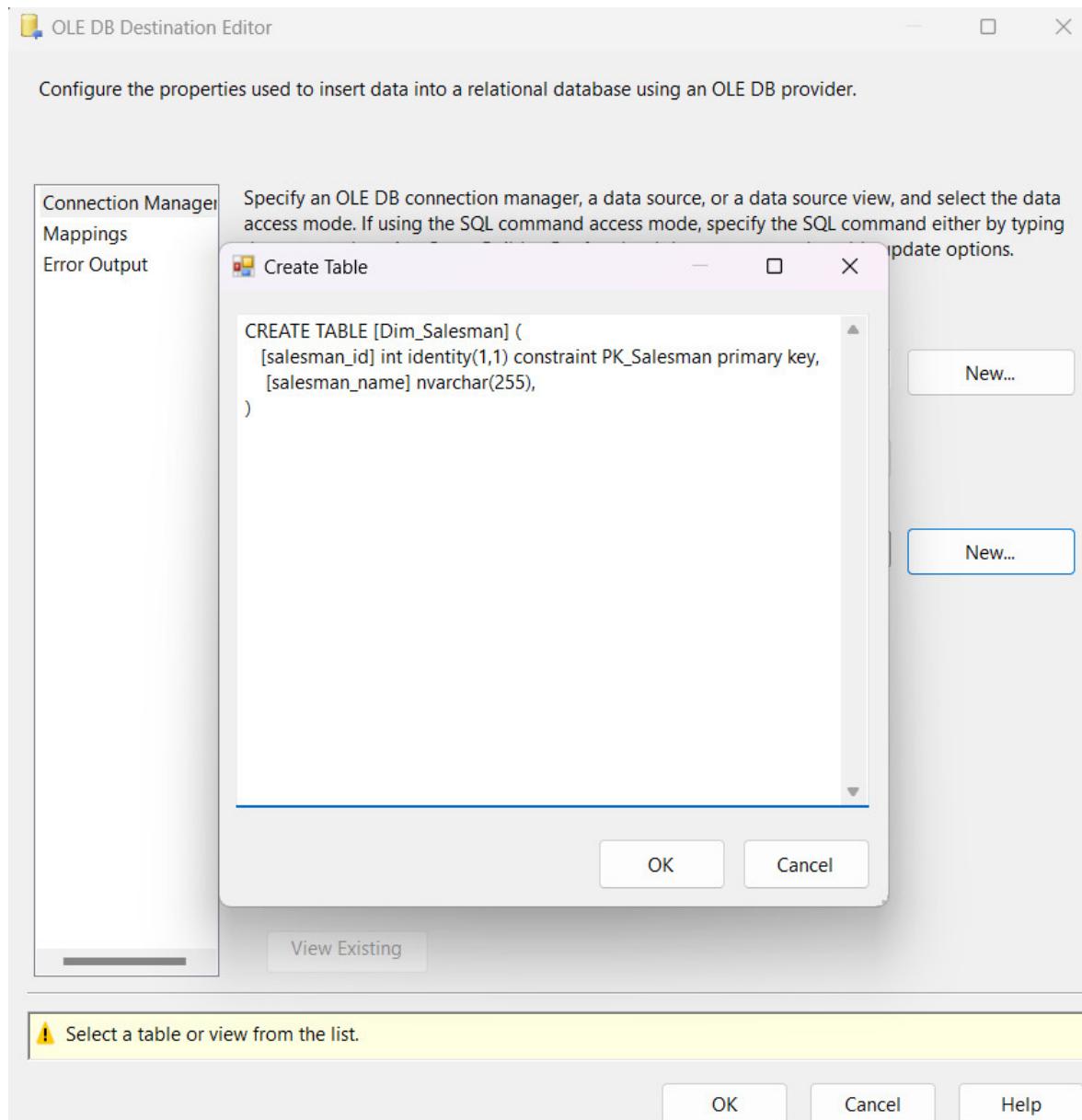
Nội dung câu lệnh SQL tạo bảng Dim\_Salesman như sau:

CREATE TABLE [Dim\_Salesman] (

[salesman\_id] int identity(1,1) constraint PK\_Salesman primary key,

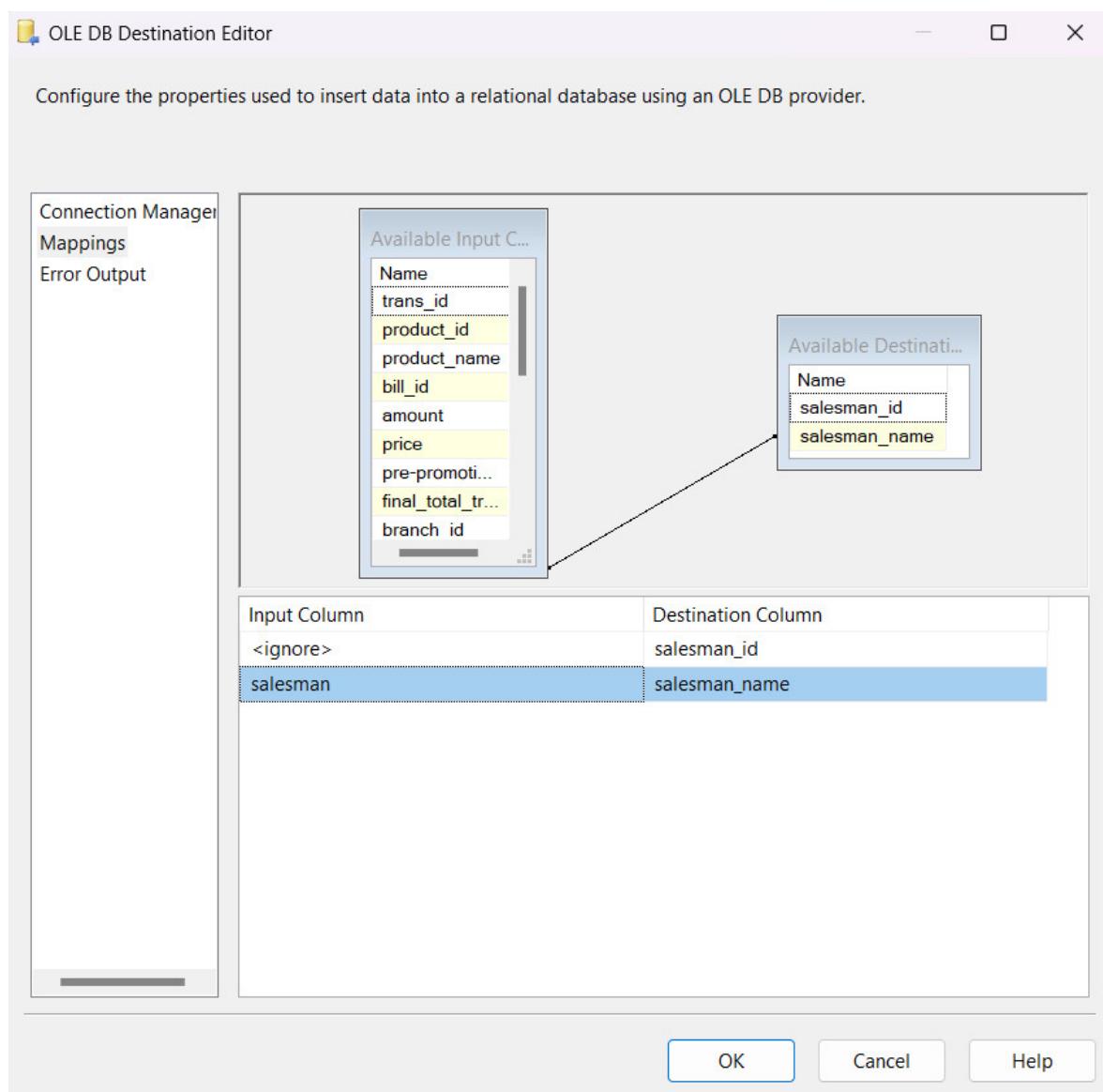
[salesman\_name] nvarchar(255)

)



Hình 59: Nhập các câu lệnh SQL tương ứng với bảng Dim\_Salesman

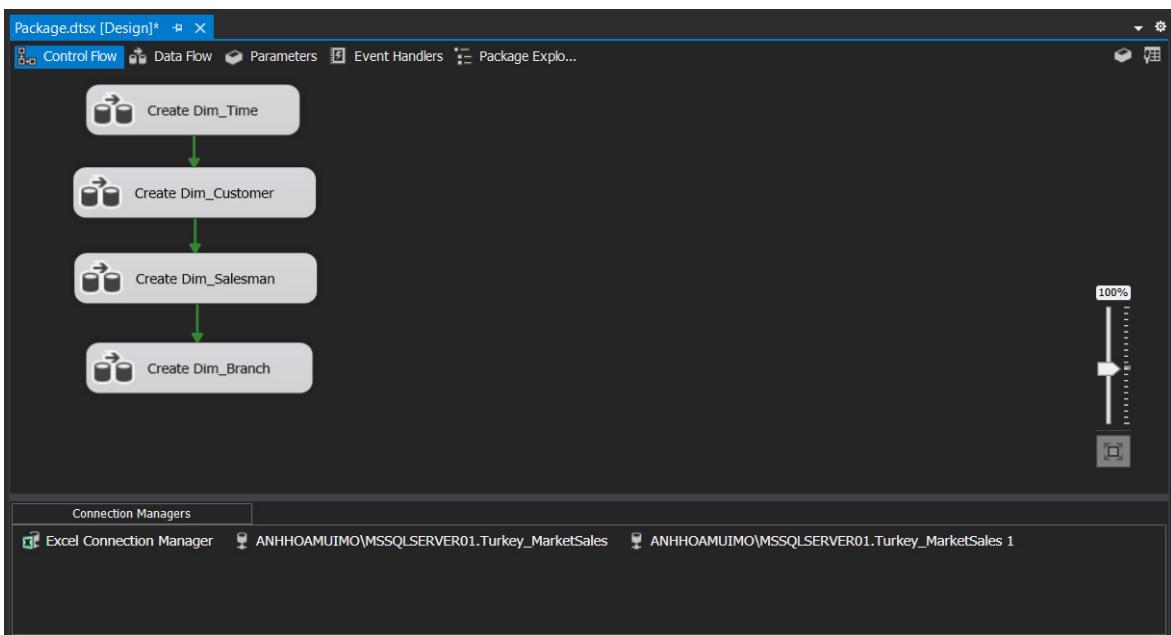
**Bước 5:** Tiếp đến ta cần chọn mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Click “OK” để hoàn tất thiết lập.



Hình 60: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 55

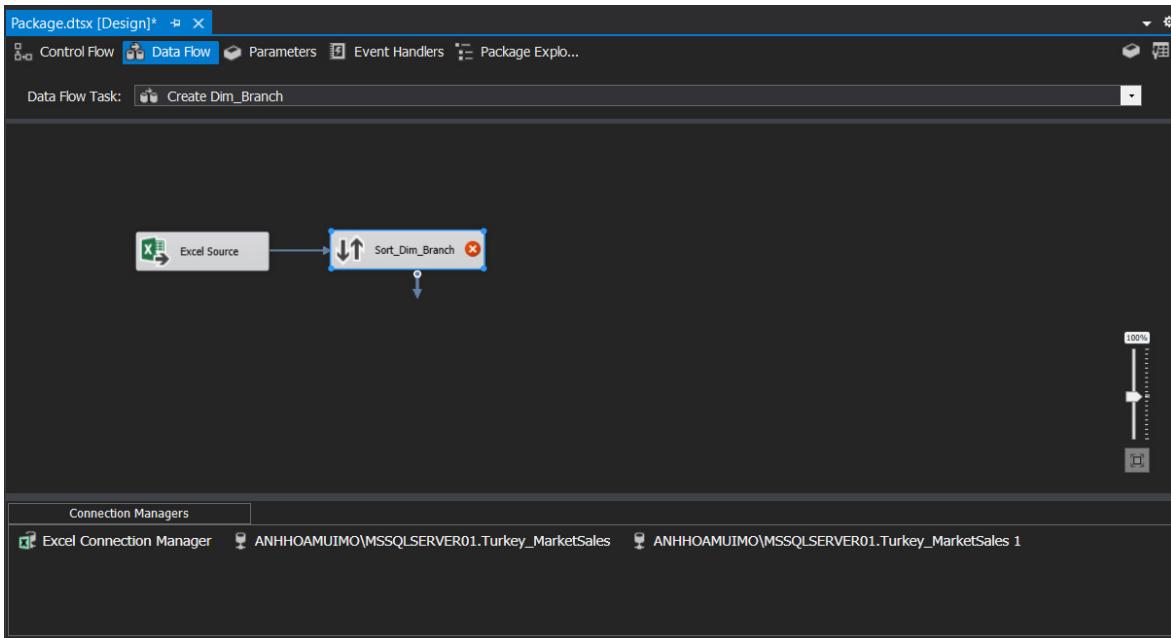
#### 2.3.1.4. Bảng Dim\_Branch:

Bước 1: Tạo “Data Flow Task”, đổi tên thành Create\_Dim\_Branch.



Hình 61: Tạo mới Data Flow Task cho bảng Dim\_Branch

**Bước 2:** Trong Create\_Dim\_Branch, tạo “Excel Source” và chọn file dữ liệu nguồn, chọn một Sort để tạo mới Sort\_Dim\_Branch cho bảng Dim\_Branch.



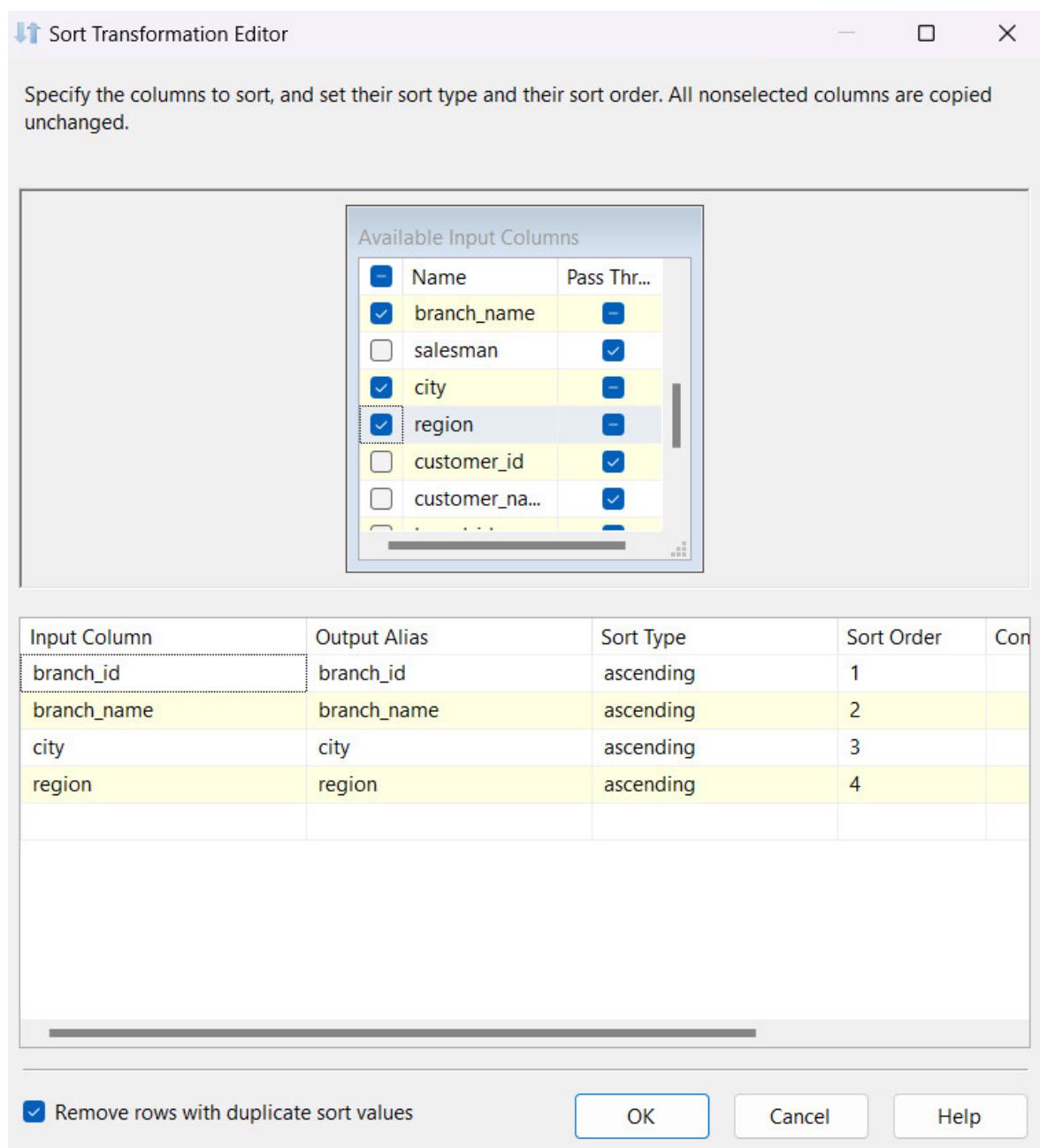
Hình 62: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu

Dim_Branch	
PK	branch_id float
	branch_name string
	city string
	region string

Hình 63: Thành phần của bảng Dim\_Branch

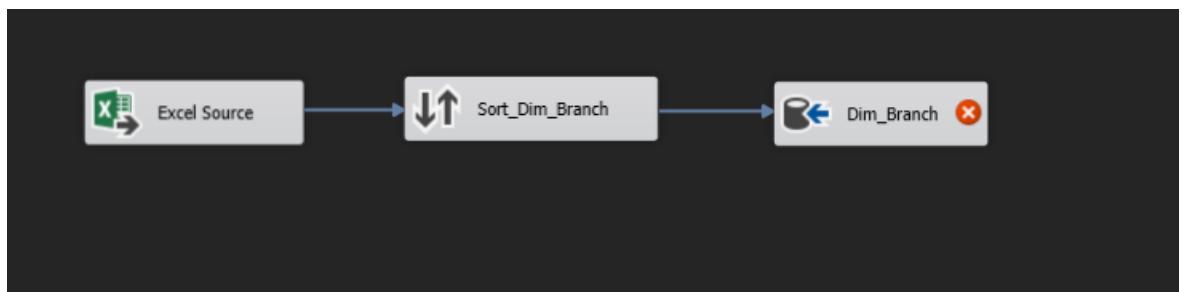
Click chuột phải vào Sort\_Dim\_Branch chọn “Edit”: lần lượt chọn các cột branch\_id, branch\_name, city, region làm các cột để đổ dữ liệu vào Sort\_Dim\_Branch.

Tick chọn “Remove rows with duplicate sort values” xóa đi các dòng dữ liệu trùng nhau và sau đó chọn “OK”.



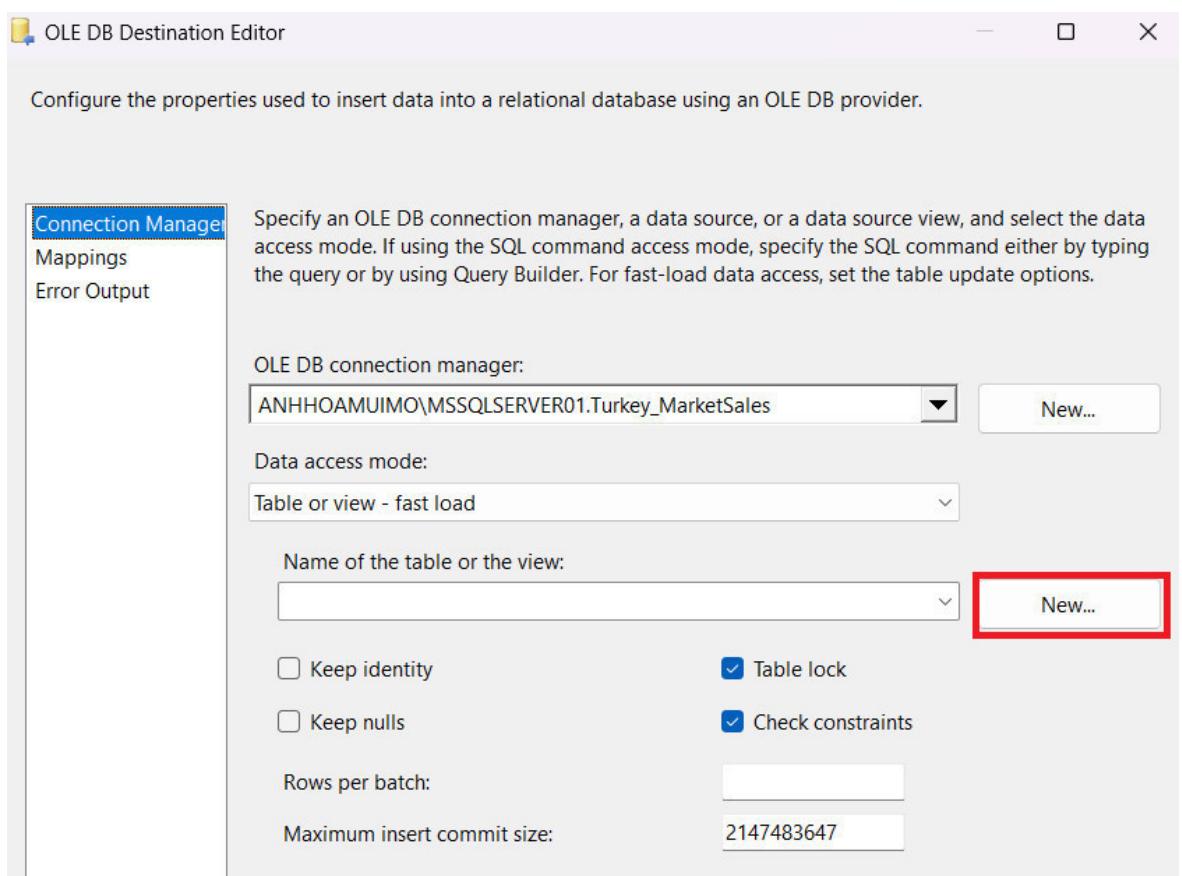
Hình 64: Chọn ra các cột dữ liệu cho bảng Dim\_Branch

**Bước 3:** Tạo mới một “OLE DB Destination” để đổ dữ liệu gốc sau khi đã được xử lý vào trong kho dữ liệu Turkey\_MarketSales.



Hình 65:Tạo mới một OLE DB Destination đặt tên là Dim\_Branch

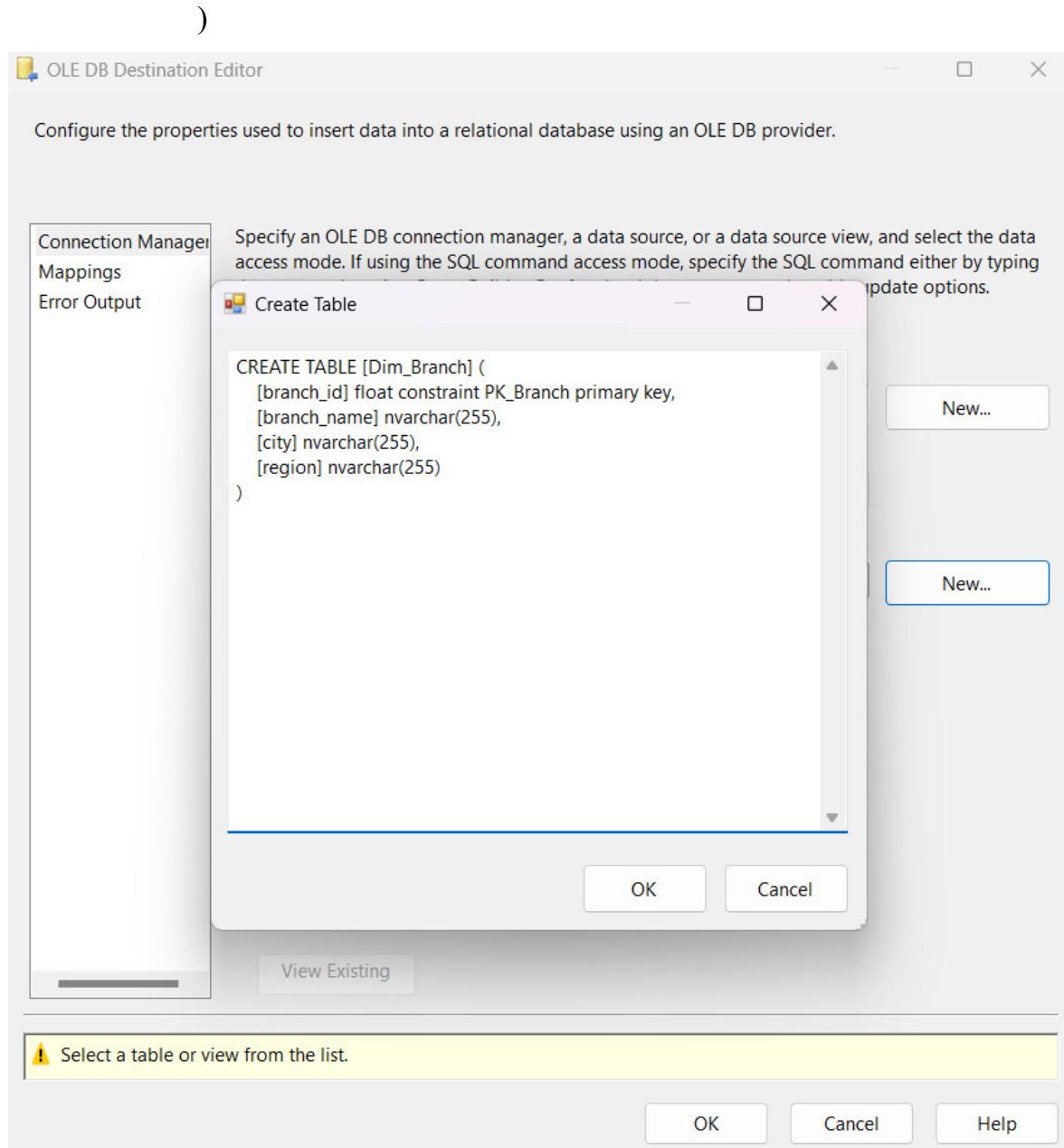
**Bước 4:** Connect đến kho dữ liệu đã được tạo khi tạo bảng Dim\_Time, vì vậy ta chỉ cần chọn “New....” để tạo bảng Dim\_Branch.



Hình 66: Kết nối đến kho dữ liệu Turkey\_MarketSales và tạo bảng Dim\_Branch

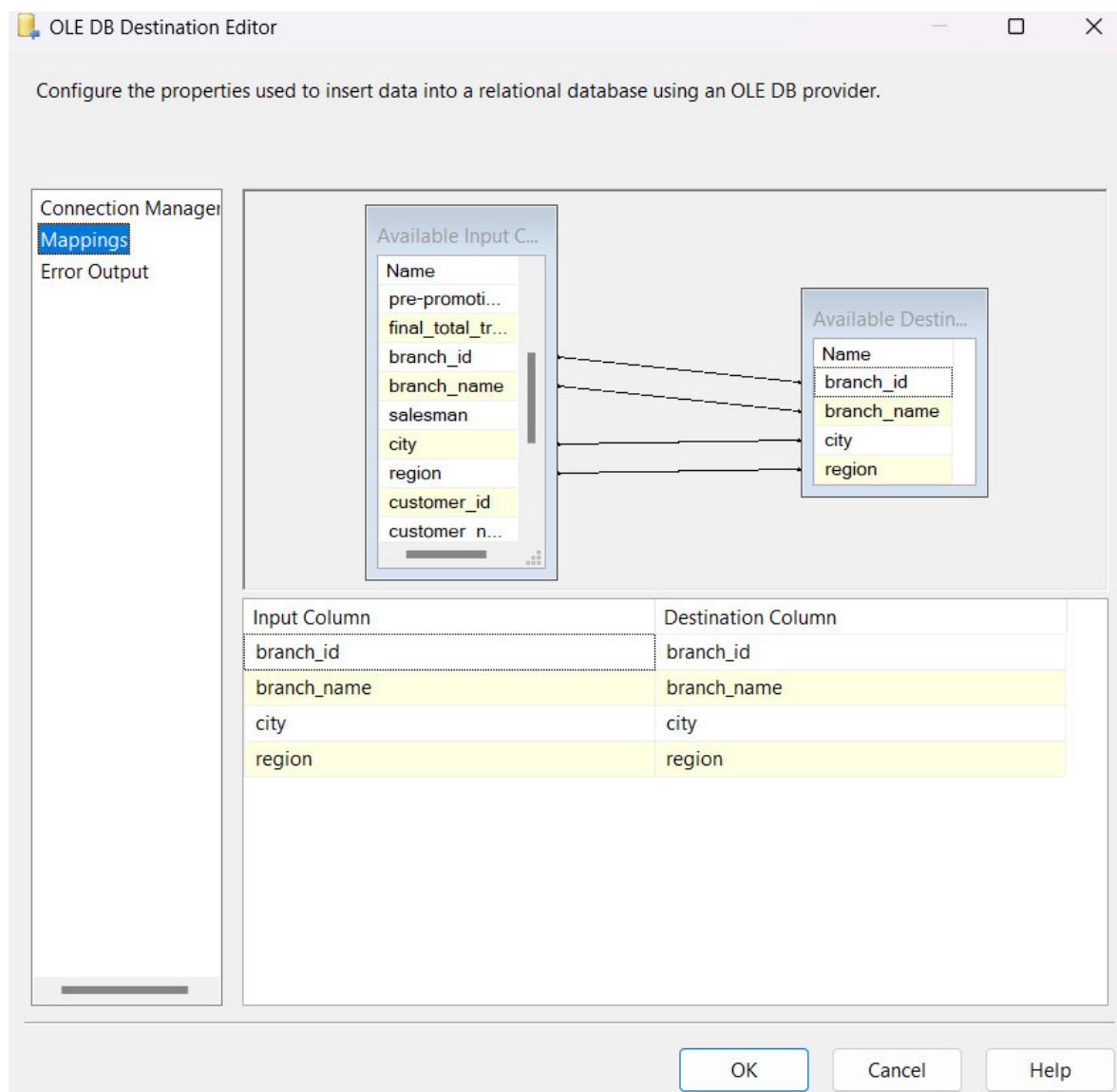
Nội dung câu lệnh SQL tạo bảng Dim\_Branch như sau:

```
CREATE TABLE [Dim_Branch] (
    [branch_id] float constraint PK_Branch primary key,
    [branch_name] nvarchar(255),
    [city] nvarchar(255),
    [region] nvarchar(255),
```



Hình 67: Nhập các câu lệnh SQL tương ứng với bảng Dim\_Branch

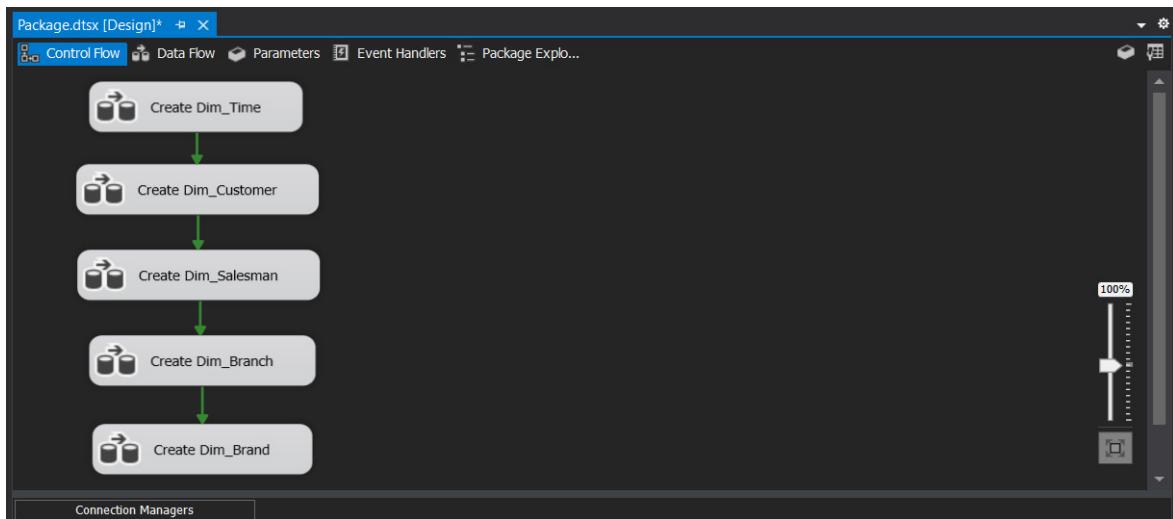
**Bước 5:** Tiếp đến ta cần chọn mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Click “OK” để hoàn tất thiết lập.



Hình 68: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 63

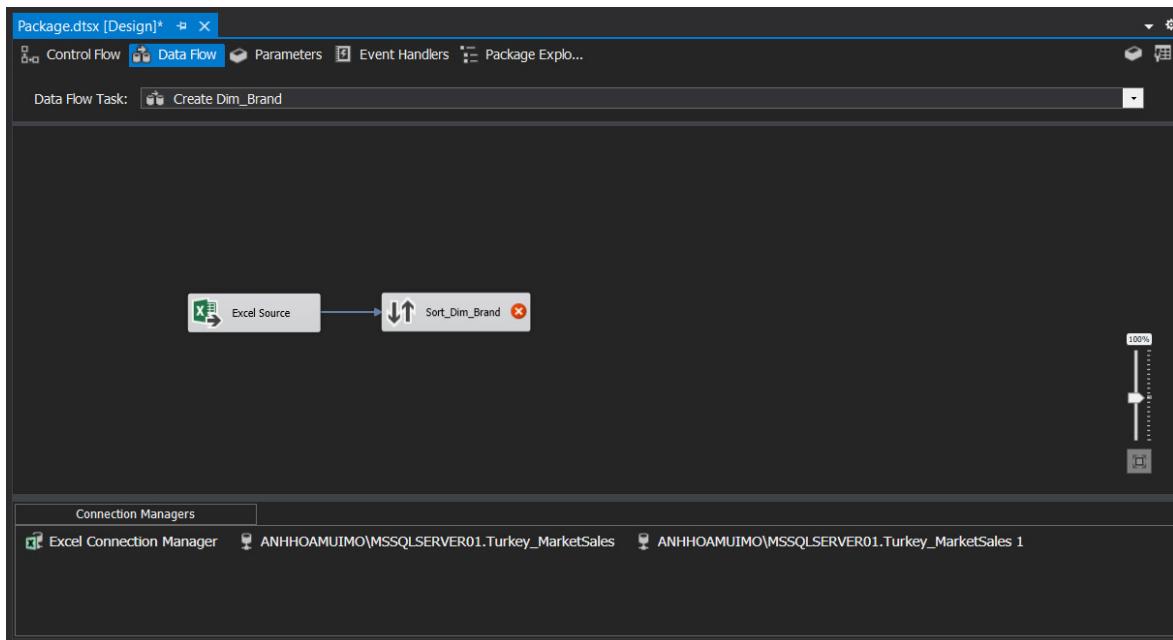
### 2.3.1.5. Bảng Dim\_Brand:

Bước 1: Tạo “Data Flow Task”, đổi tên thành Create\_Dim\_Brand.



Hình 69: Tạo mới Data Flow Task cho bảng Dim\_Brand

**Bước 2:** Trong Create\_Dim\_Brand, tạo “Excel Source” và chọn file dữ liệu nguồn, chọn một Sort để tạo mới Sort\_Dim\_Brand cho bảng Dim\_Brand.



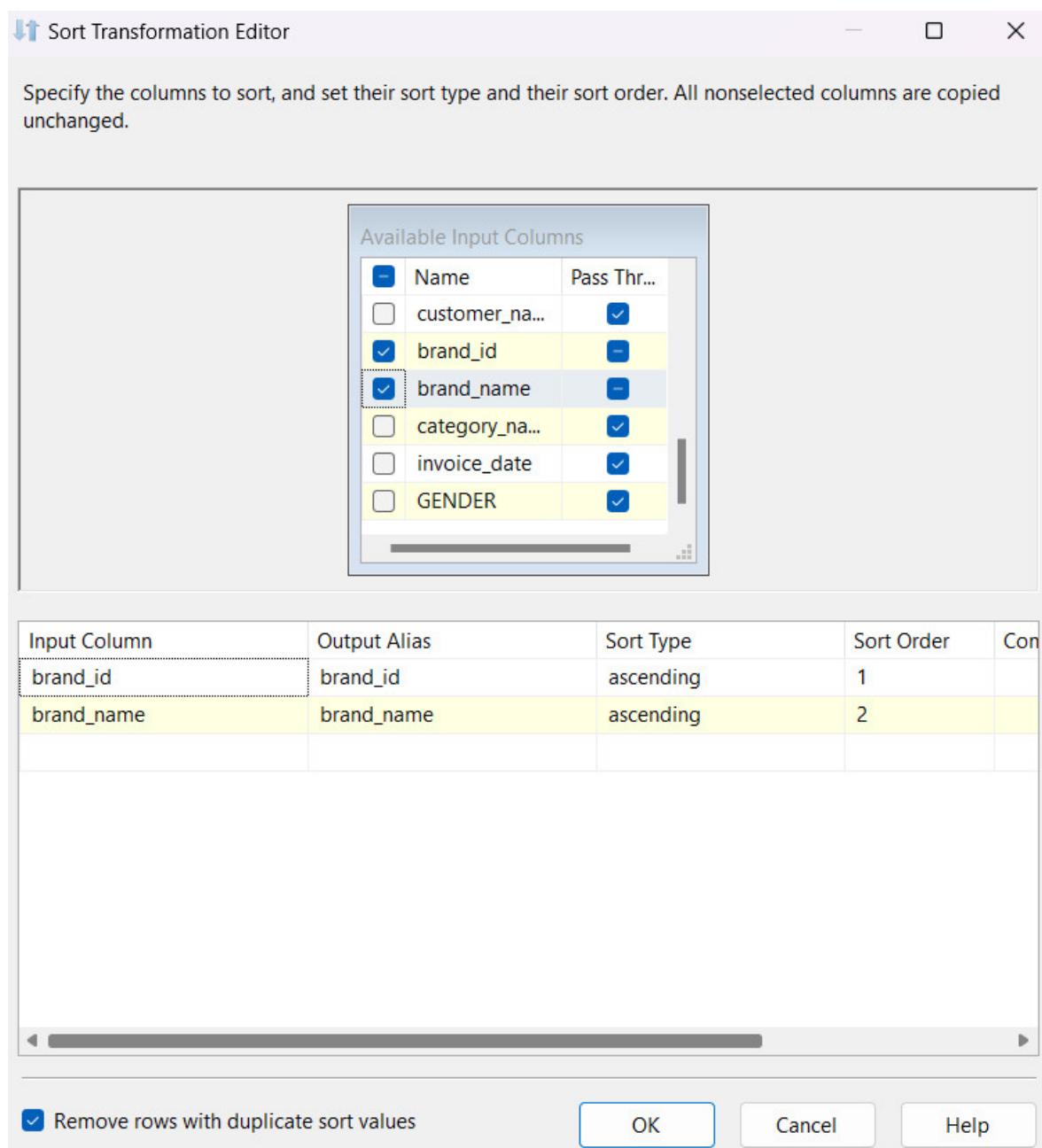
Hình 70: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu

Dim_Brand	
PK	<u>brand_id string</u>
	brand_name string

Hình 71: Thành phần của bảng Dim\_Brand

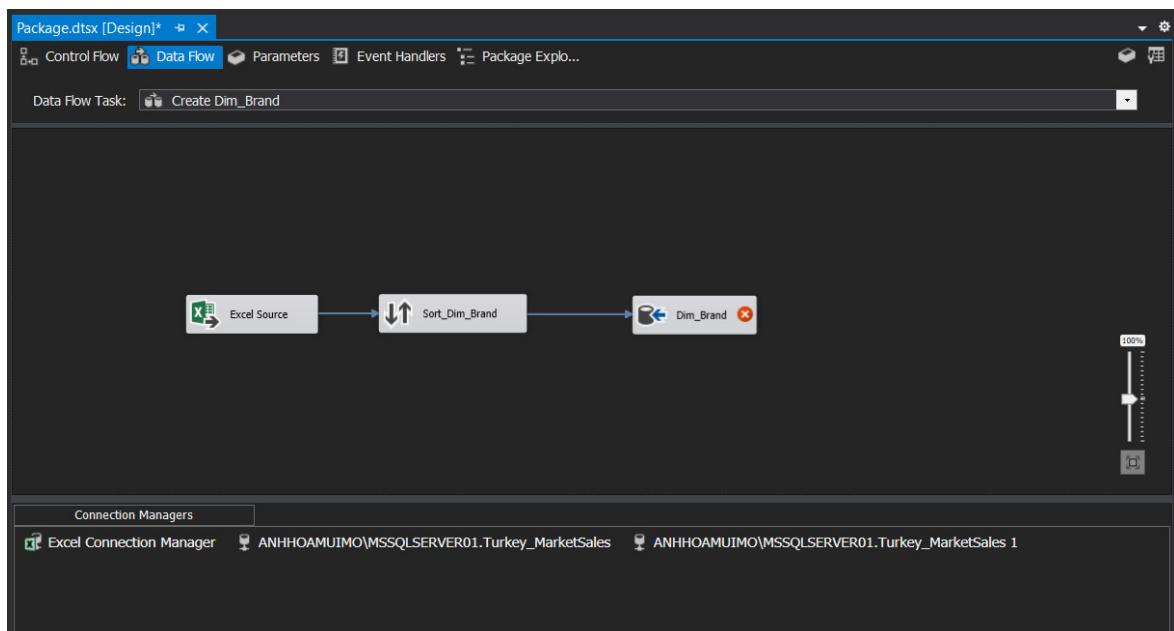
Click chuột phải vào Sort\_Dim\_Brand chọn “Edit”: lần lượt chọn các cột brand\_id, brand\_name làm các cột để đổ dữ liệu vào Sort\_Dim\_Brand.

Tick chọn “Remove rows with duplicate sort values” xóa đi các dòng dữ liệu trùng nhau và sau đó chọn “OK”.



Hình 72: Chọn ra các cột dữ liệu cho bảng Dim\_Brand

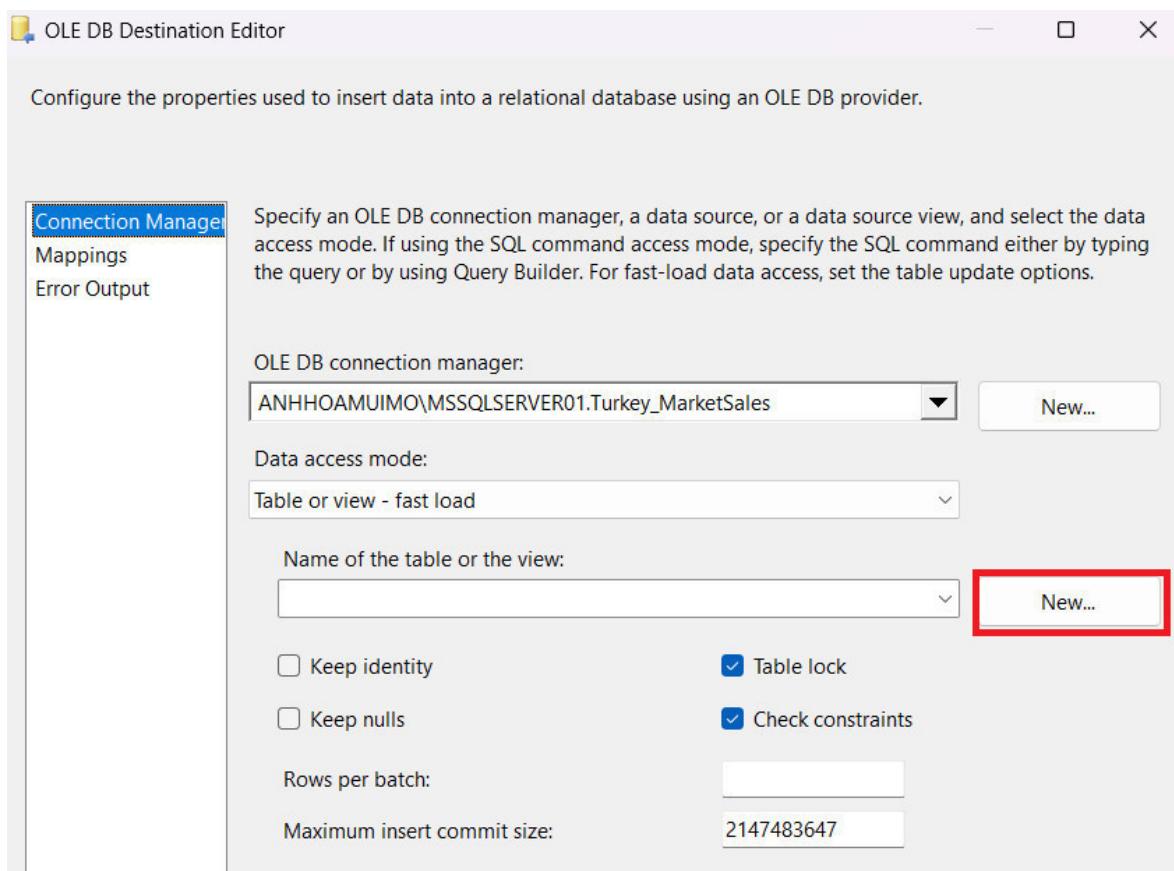
**Bước 3:** Tạo mới một “OLE DB Destination” để đổ dữ liệu gốc sau khi đã được xử lý vào trong kho dữ liệu Turkey\_MarketSales.



Hình 73: Tạo mới một OLE DB Destination đặt tên là Dim\_Brand

**Bước 4:** Connect đến kho dữ liệu đã được tạo khi tạo bảng

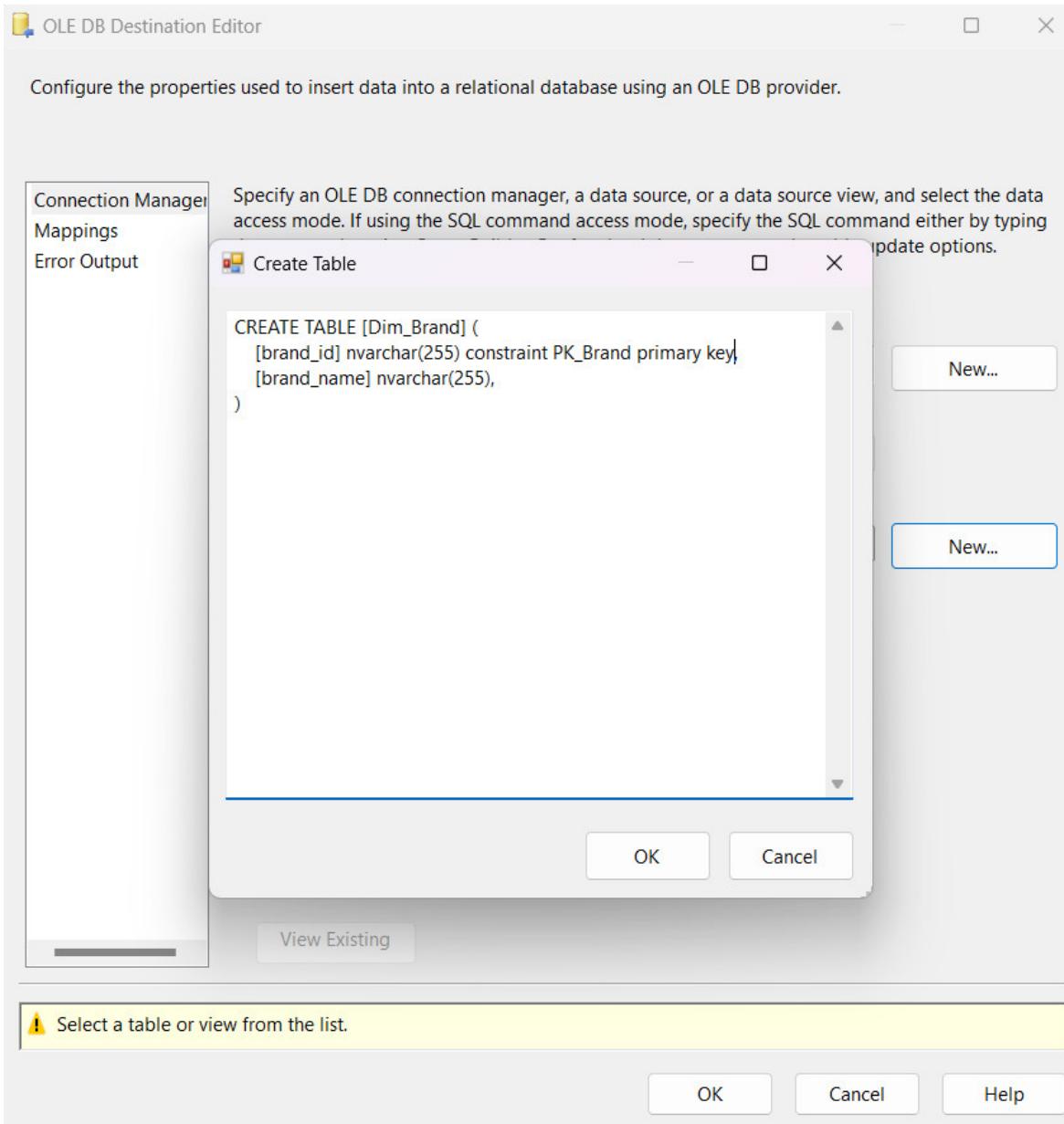
Dim\_Time, vì vậy ta chỉ cần chọn “New....” để tạo bảng Dim\_Brand.



Hình 74: Kết nối đến kho dữ liệu Turkey\_MarketSales và tạo bảng Dim\_Brand

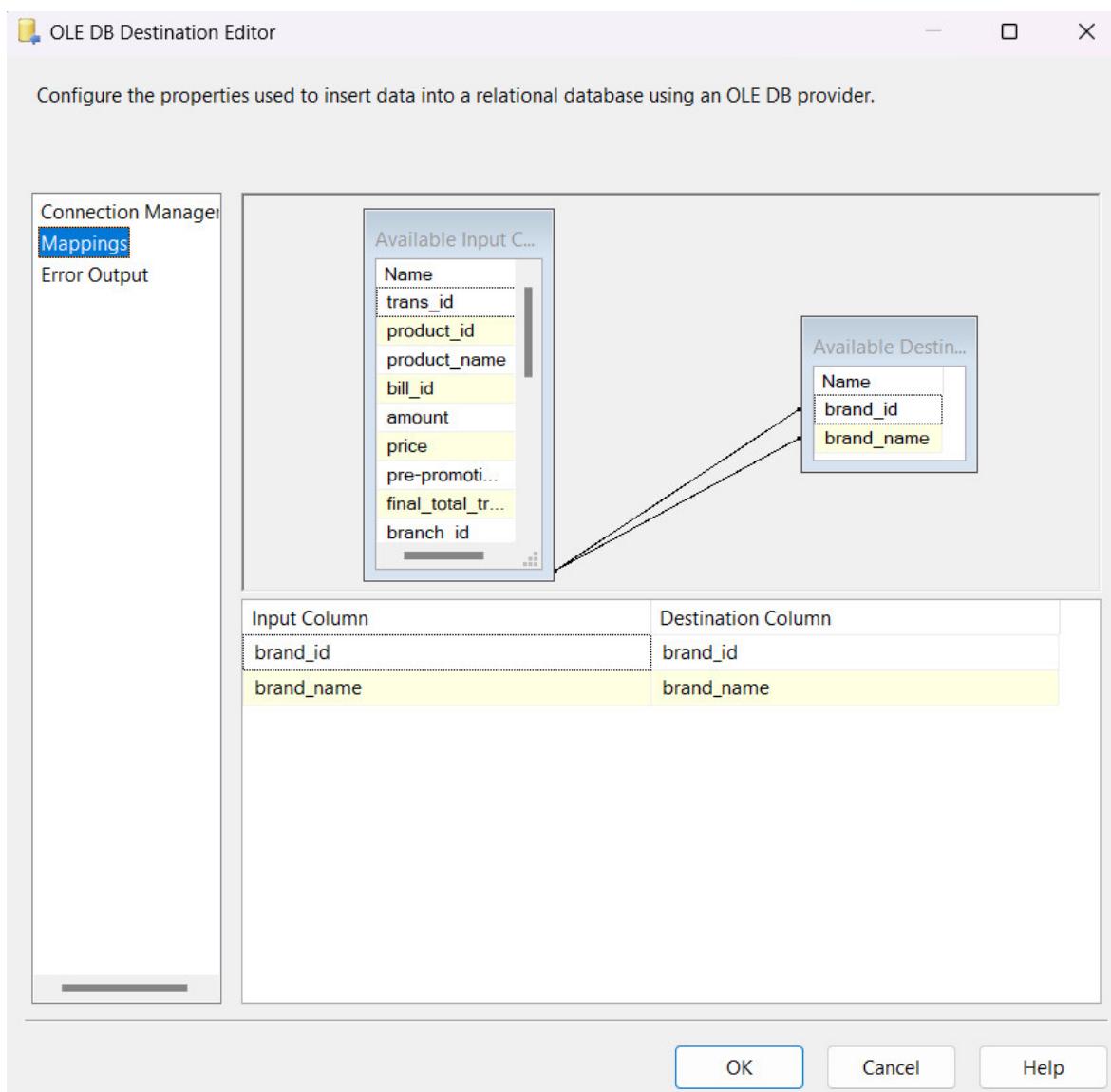
Nội dung câu lệnh SQL tạo bảng Dim\_Brand như sau:

```
CREATE TABLE [Dim_Brand] (
    [brand_id] nvarchar(255) constraint PK_Brand primary key,
    [brand_name] nvarchar(255)
)
```



Hình 75: Nhập các câu lệnh SQL tương ứng với bảng Dim\_Brand

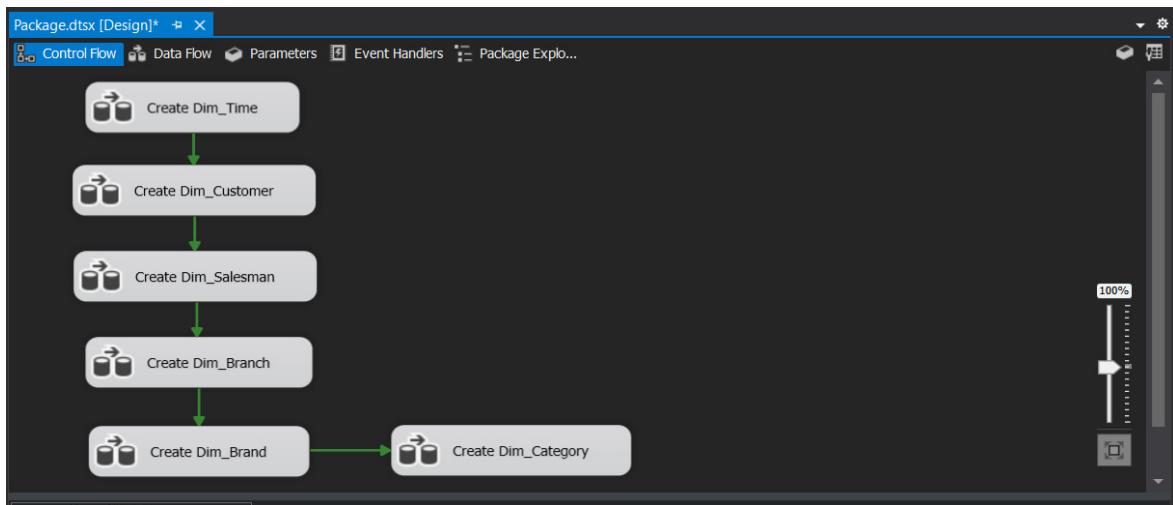
**Bước 5:** Tiếp đến ta cần chọn mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Click “OK” để hoàn tất thiết lập.



Hình 76: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 71

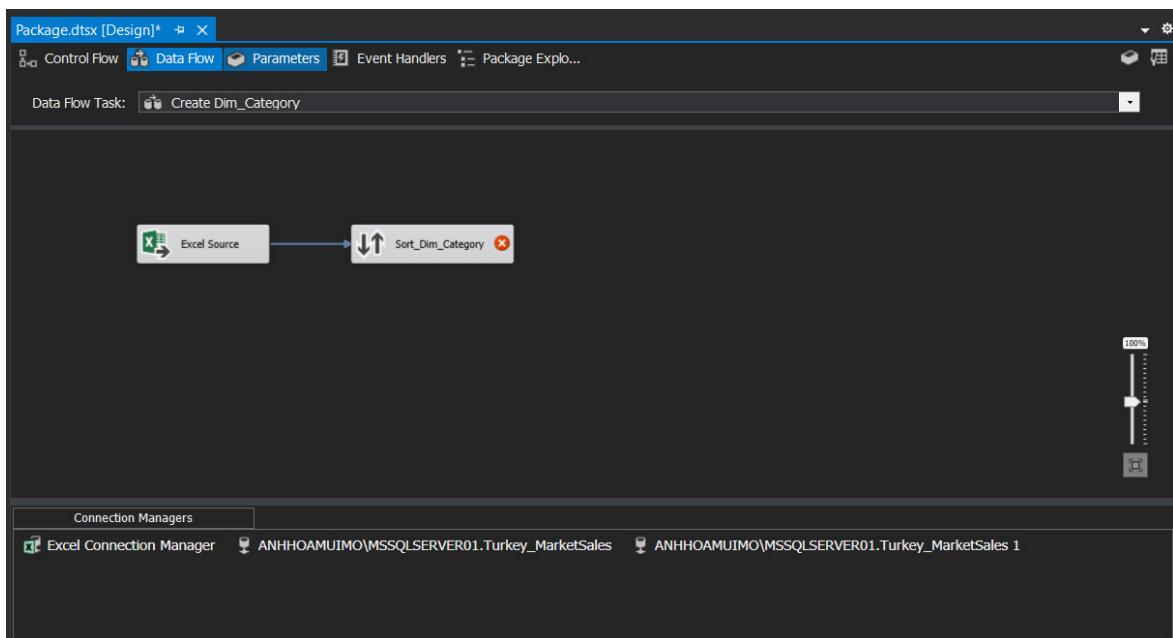
### 2.3.1.6. Bảng Dim\_Category:

Bước 1: Tạo “Data Flow Task”, đổi tên thành Create\_Dim\_Category.



Hình 77: Tạo mới Data Flow Task cho bảng Dim\_Category

**Bước 2:** Trong Create\_Dim\_Brand, tạo “Excel Source” và chọn file dữ liệu nguồn, chọn một Sort để tạo mới Sort\_Dim\_Category cho bảng Dim\_Category.



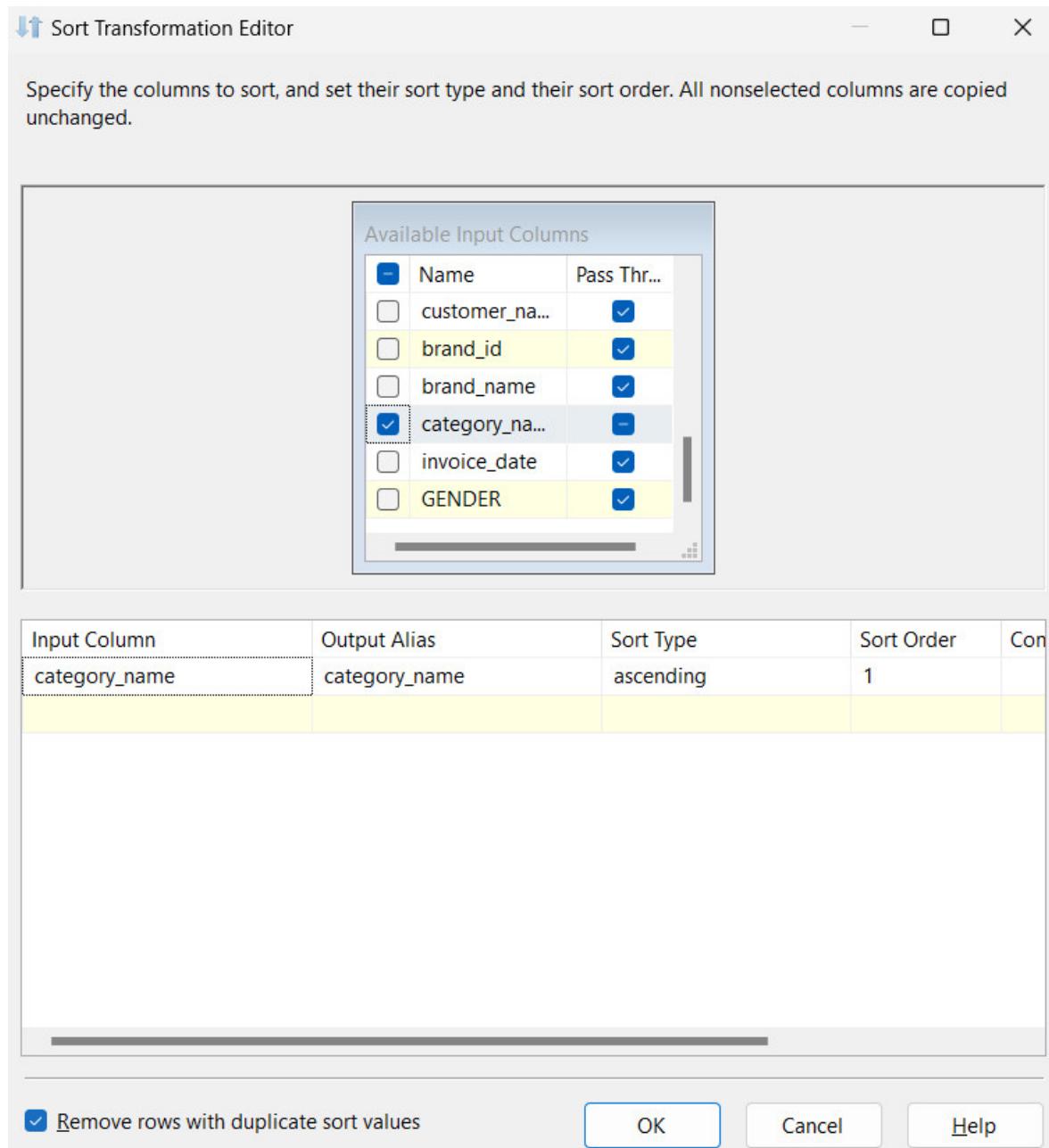
Hình 78: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu

Dim_Category	
PK	category_id int
	category_name string

Hình 79: Thành phần của bảng Dim\_Category

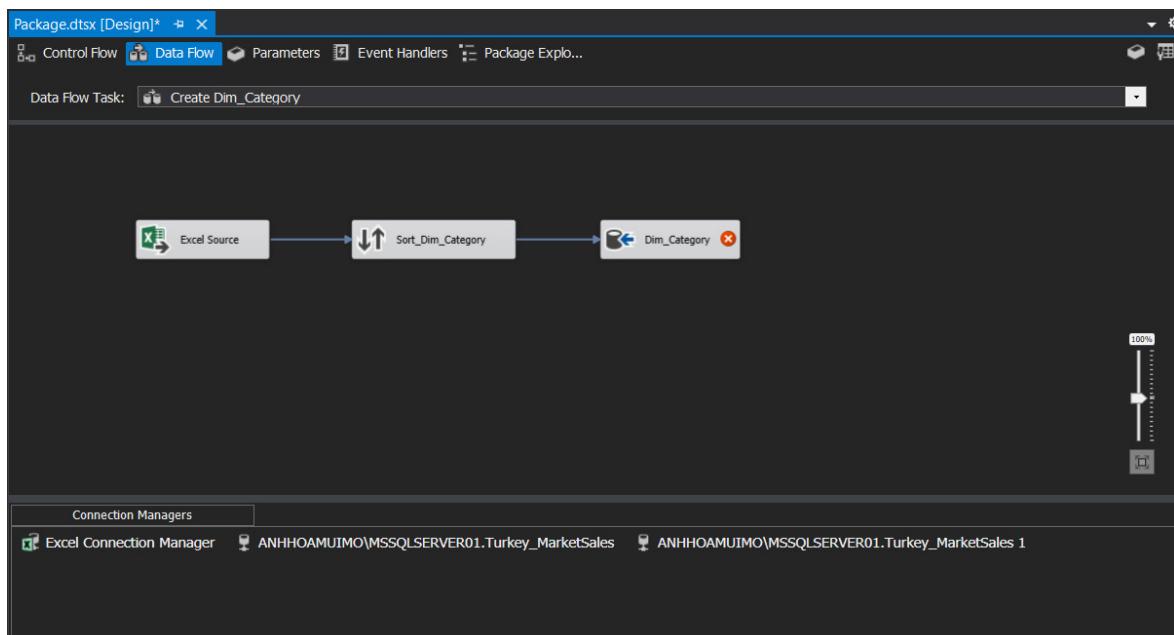
Click chuột phải vào Sort\_Dim\_Category chọn “Edit”: lần lượt chọn các cột category\_name làm các cột để đồ dữ liệu vào Sort\_Dim\_Category.

Tick chọn “Remove rows with duplicate sort values” xóa đi các dòng dữ liệu trùng nhau và sau đó chọn “OK”.



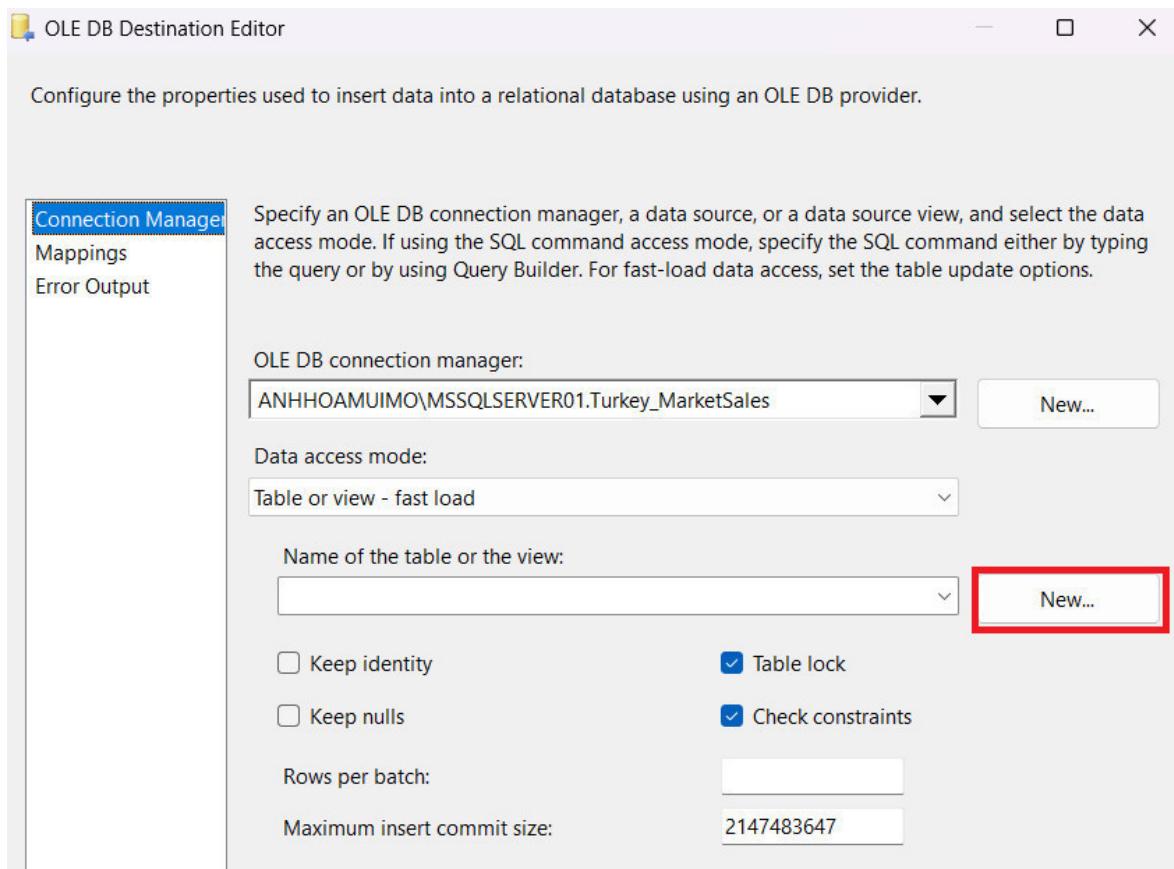
Hình 80: Chọn ra các cột dữ liệu cho bảng Dim\_Category

**Bước 3:** Tạo mới một “OLE DB Destination” để đổ dữ liệu gốc sau khi đã được xử lý vào trong kho dữ liệu Turkey\_MarketSales.



Hình 81: Tạo mới một OLE DB Destination đặt tên là Dim\_Category

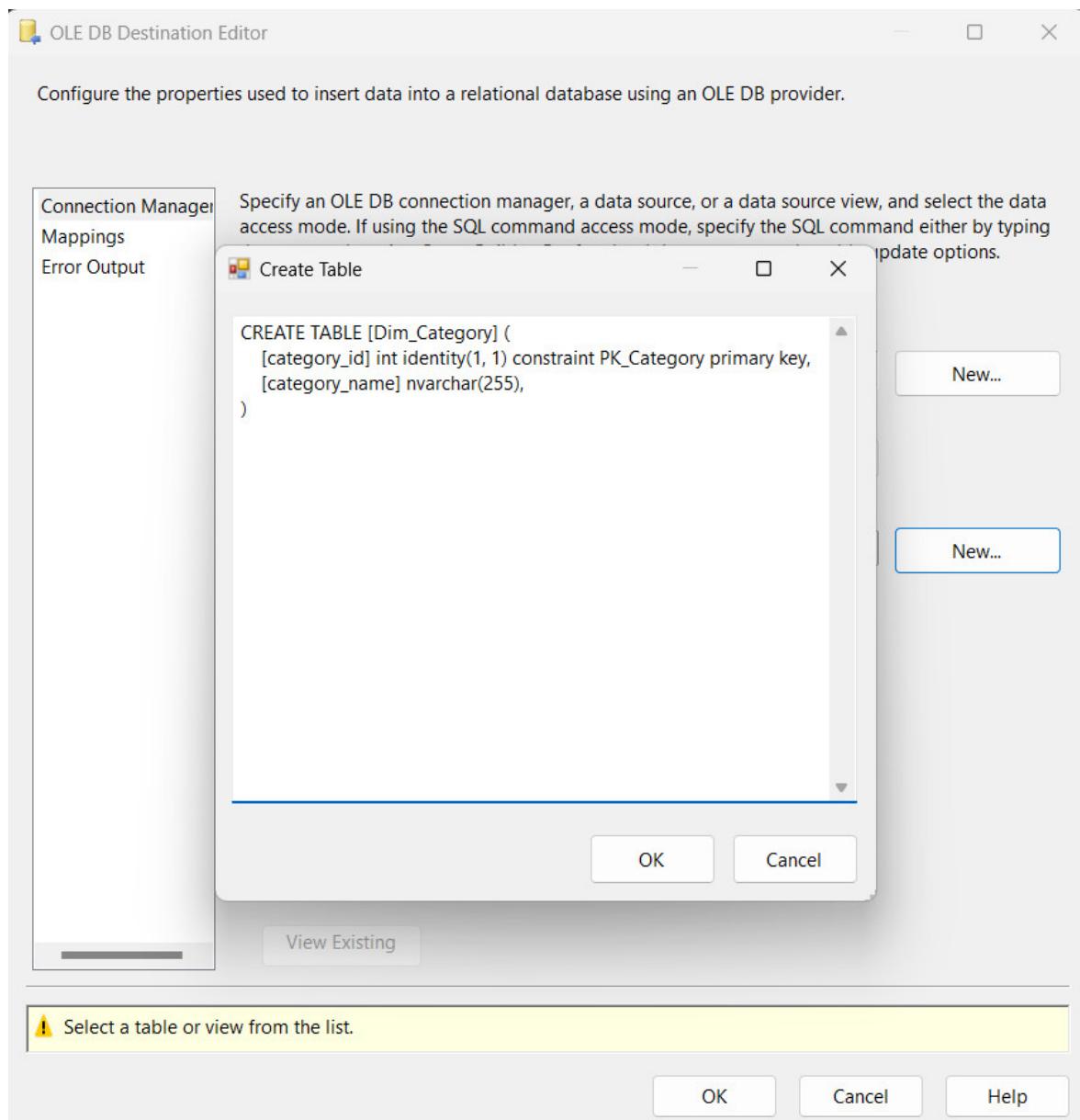
**Bước 4:** Connect đến kho dữ liệu đã được tạo khi tạo bảng Dim\_Time, vì vậy ta chỉ cần chọn “New....” để tạo bảng Dim\_Category.



Hình 82: Kết nối đến kho dữ liệu Turkey\_MarketSales và tạo bảng Dim\_Category

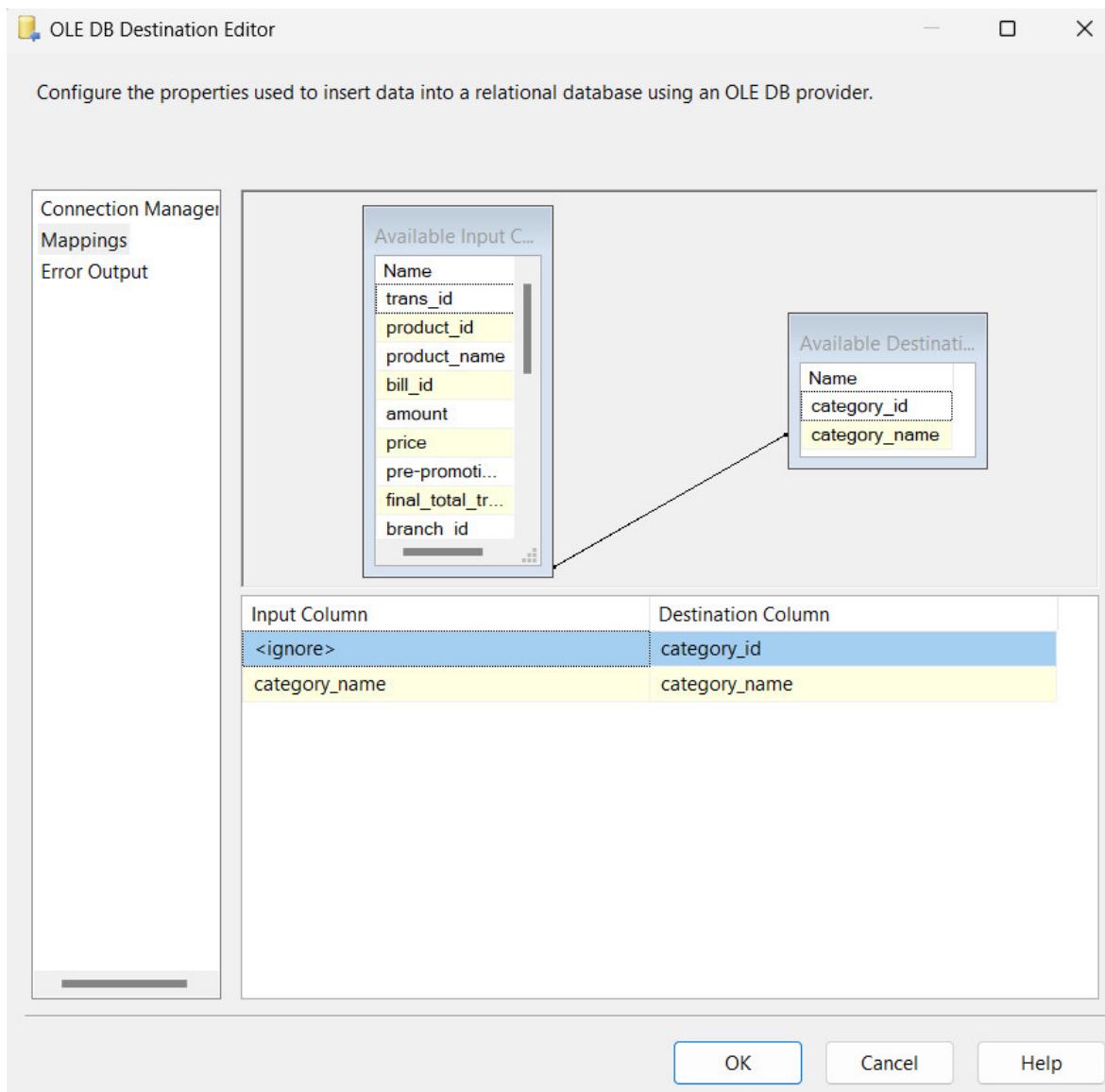
Nội dung câu lệnh SQL tạo bảng Dim\_Category như sau:

```
CREATE TABLE [Dim_Category] (
    [category_id] int identity(1,1) constraint PK_Category primary key,
    [category_name] nvarchar(255)
)
```



Hình 83: Nhập các câu lệnh SQL tương ứng với bảng Dim\_Category

**Bước 5:** Tiếp đến ta cần chọn mục “Mappings” để kiểm tra việc ánh xạ các cột dữ liệu. Click “OK” để hoàn tất thiết lập.

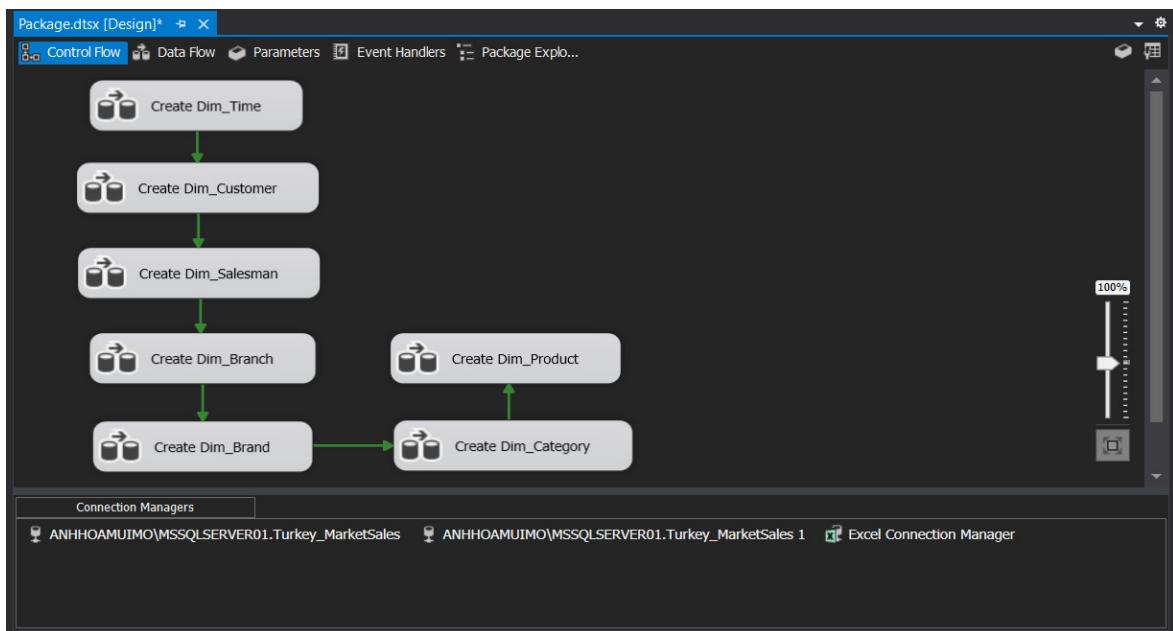


Hình 84: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 79

#### 2.3.1.7. Bảng Dim\_Product:

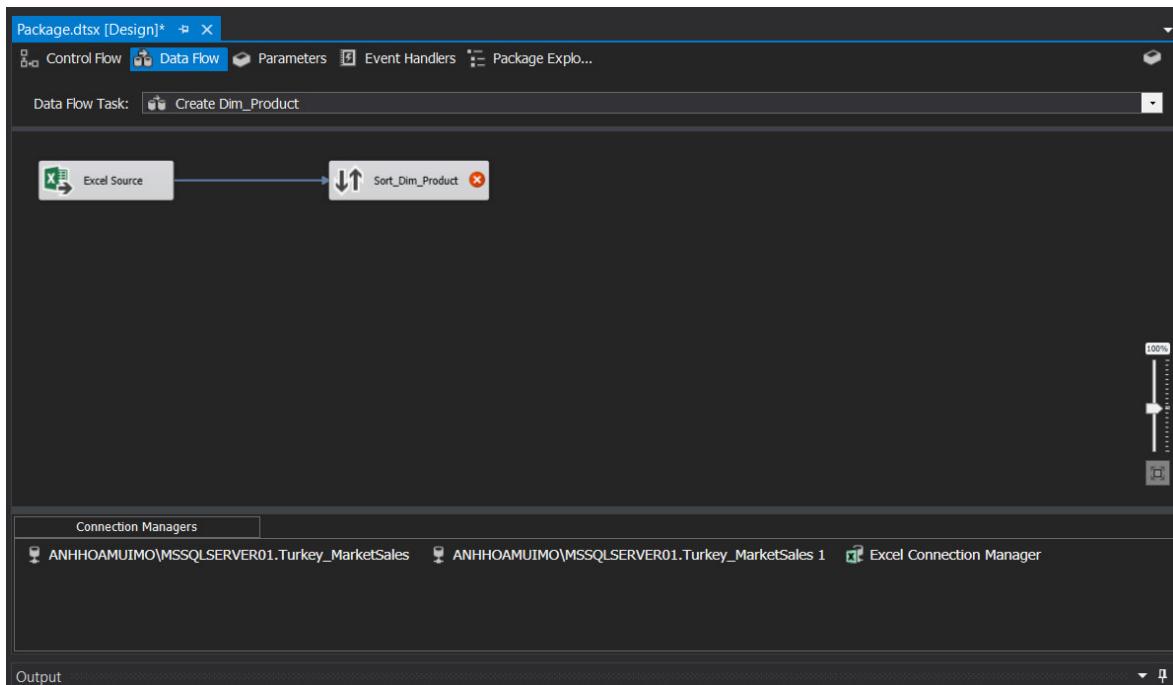
**Bước 1:** Tạo “Data Flow Task”, đổi tên thành

Create\_Dim\_Product.

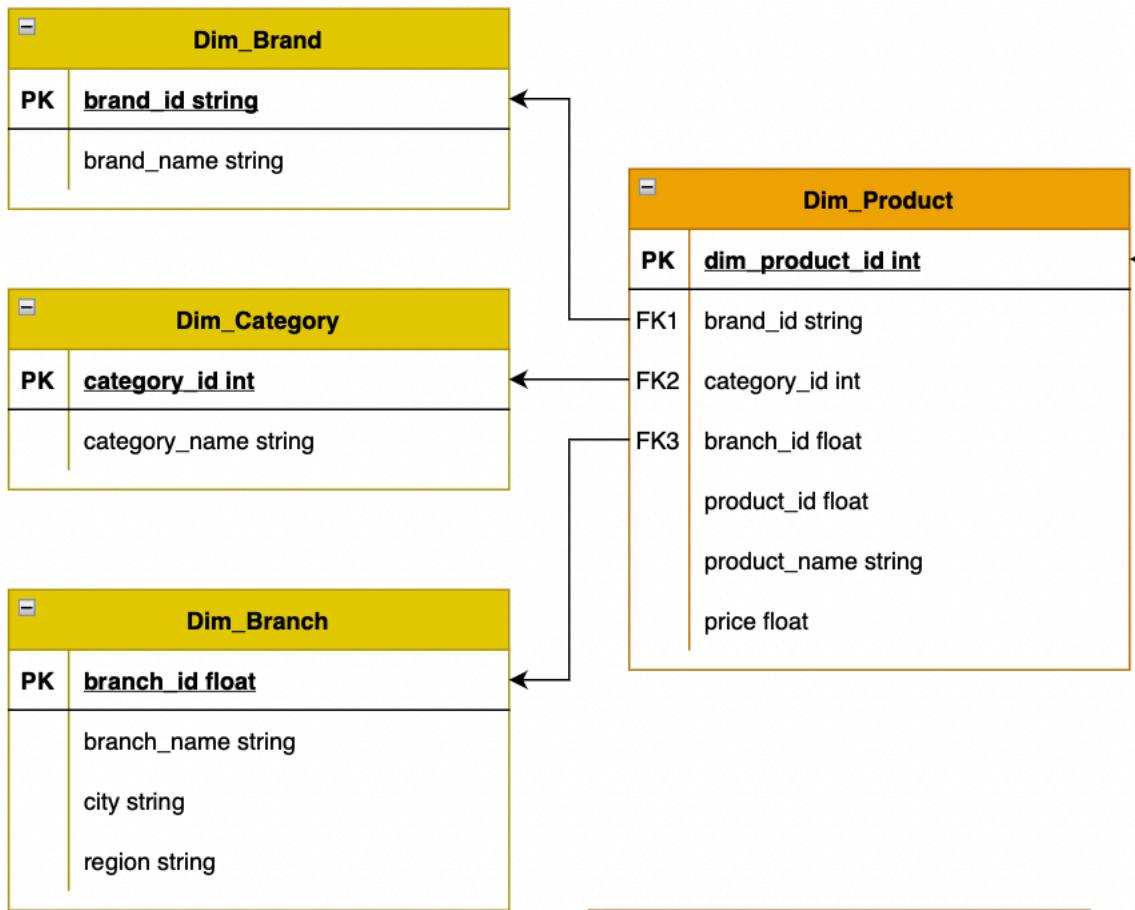


Hình 85: Tạo mới một Data Flow Task cho bảng Dim\_Product

**Bước 2:** Trong Create\_Dim\_Product, tạo “Excel Source” và chọn file dữ liệu nguồn, chọn một Sort để tạo mới Sort\_Dim\_Product cho bảng Dim\_Product.



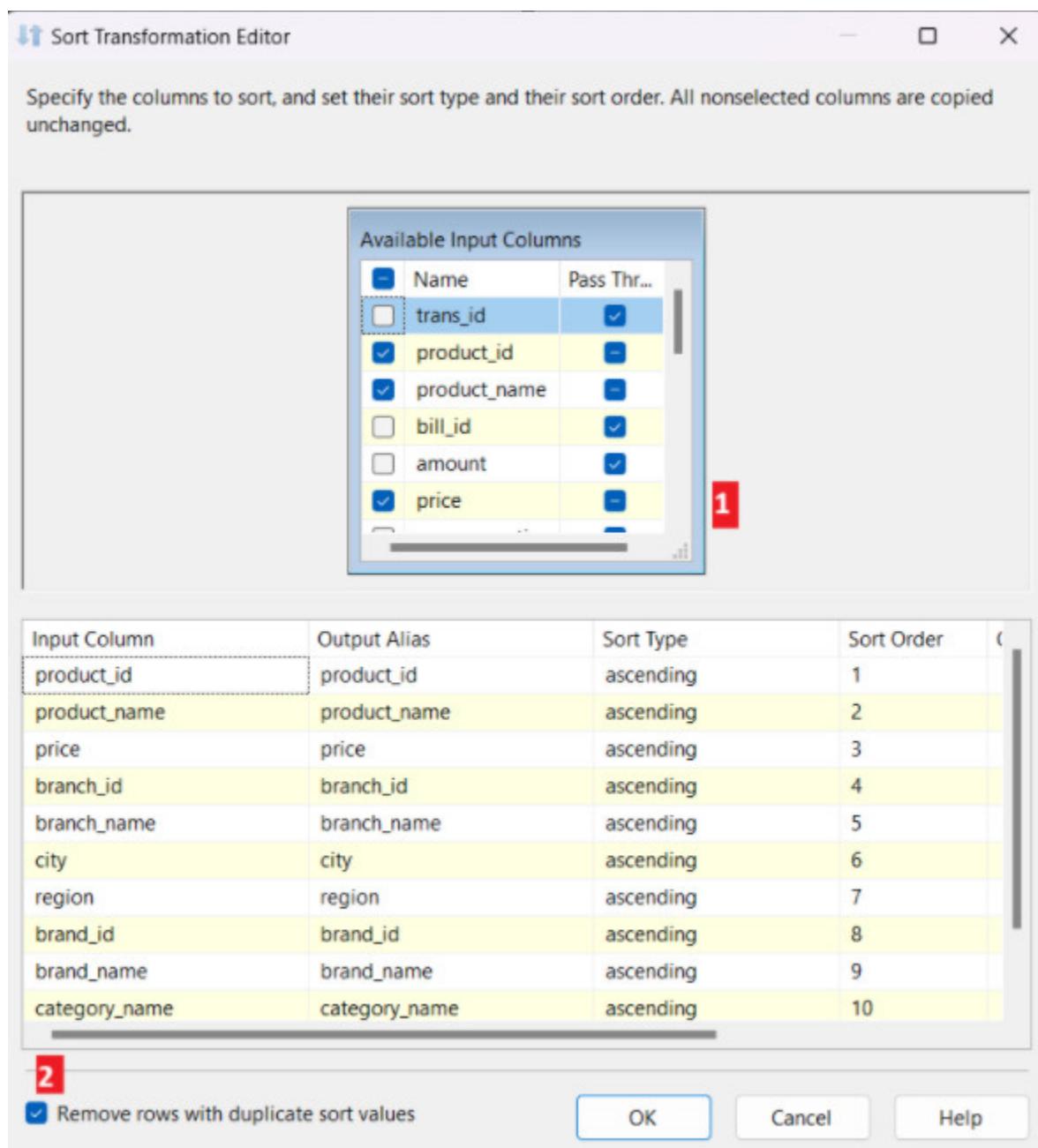
Hình 86: Tạo mới một Excel Source và một Sort để chuẩn bị dữ liệu



Hình 87: Thành phần của bảng Dim\_Product

Click chuột phải vào Sort\_Dim\_Product, chọn Edit: lần lượt chọn các cột product\_id, product\_name, price, brand\_id, brand\_name, branch\_id, branch\_name, city, region, category\_name làm các cột để tiến hành merge và đổ dữ liệu vào Sort\_Dim\_Product.

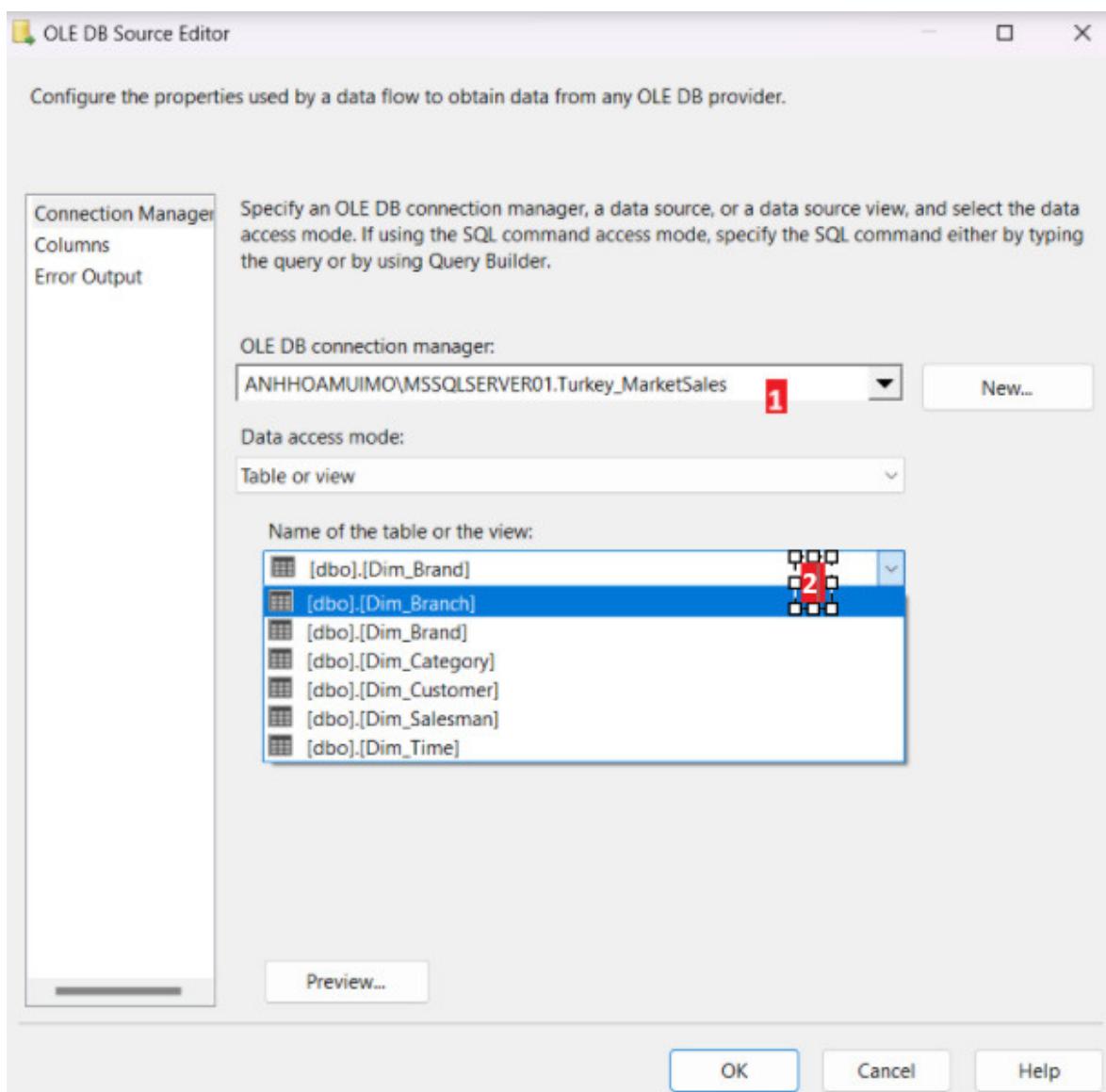
Tick chọn “**Remove rows with duplicate sort values**” xóa đi các dòng dữ liệu trùng nhau và sau đó chọn “**OK**”.



Hình 88: Chọn ra các cột dữ liệu cho bảng Dim\_Product

**Bước 3:** Tạo mới một “OLE DB Source” và đổi tên thành

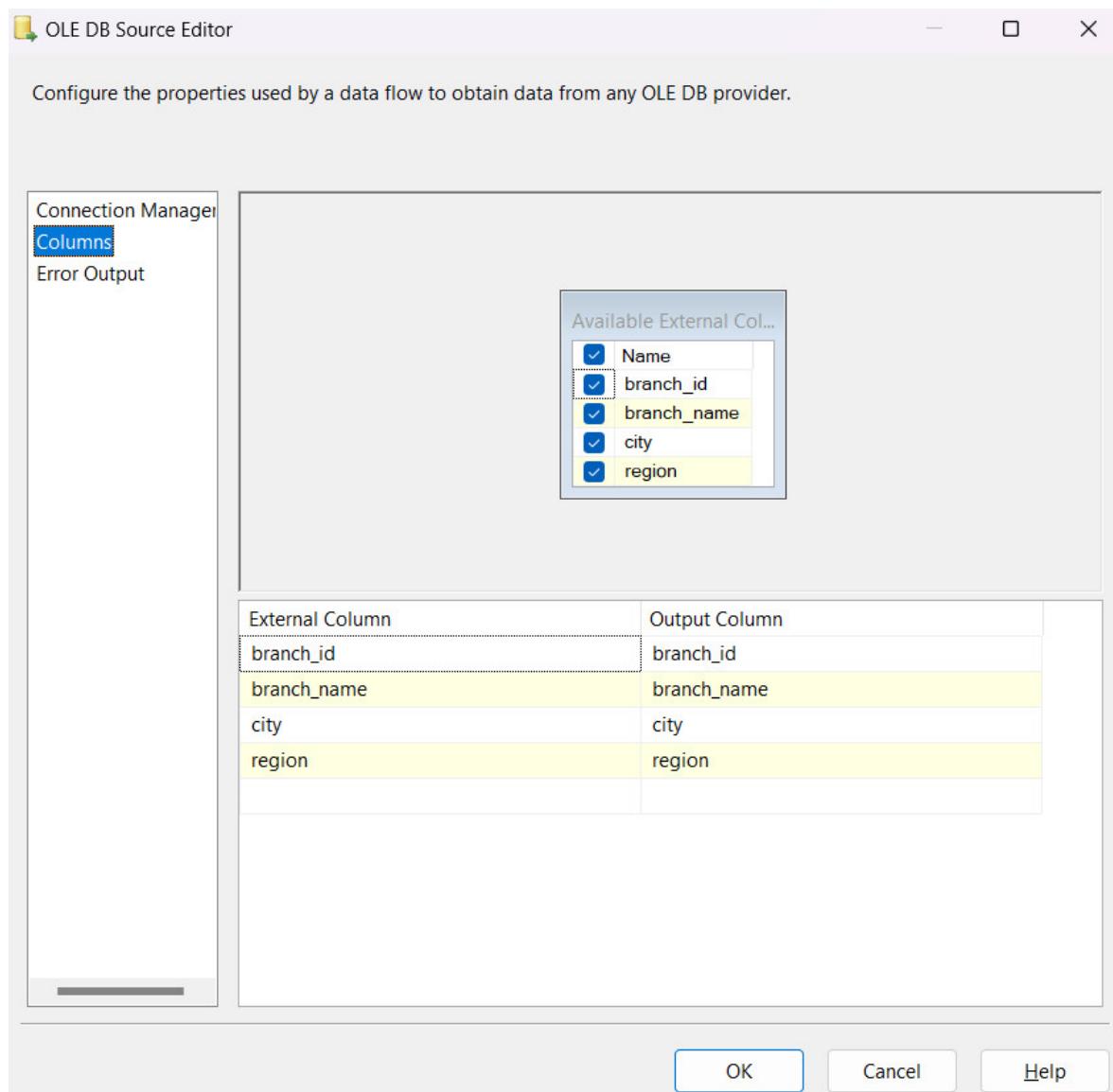
Dim\_Branch. Click chuột phải vào chọn “Edit”, sau đó chọn bảng  
Dim\_Branch đã được tạo trước đó làm data source cho bảng Dim\_Product.



Hình 89: Tạo mới một OLE DB Source và chọn bảng Dim\_Branch làm data source

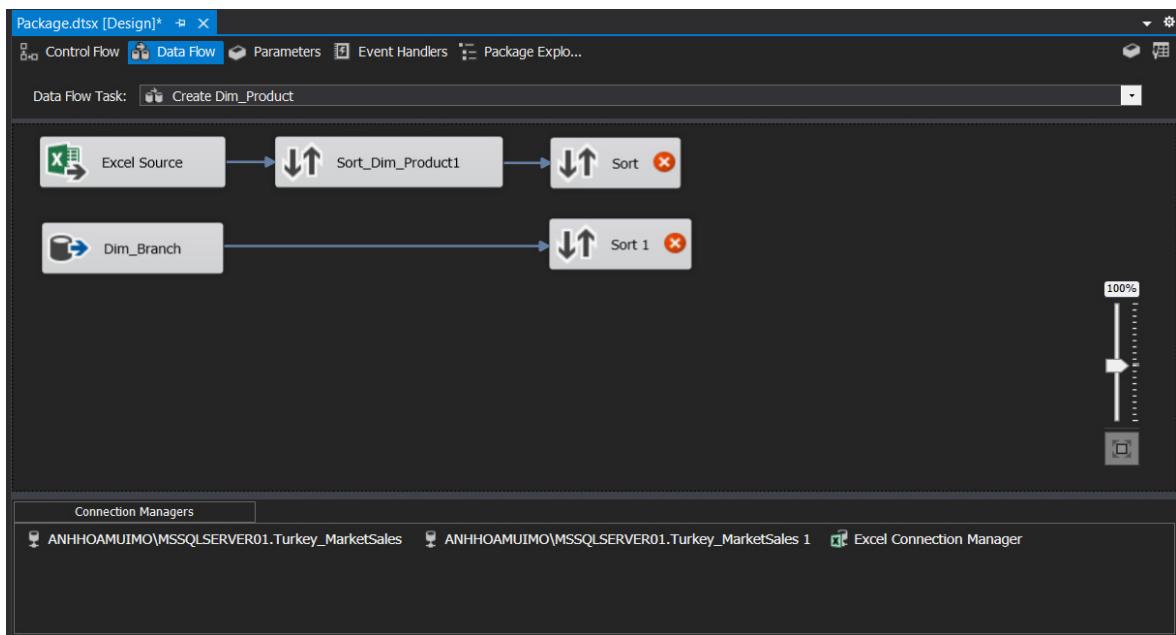
**Bước 4:** Chọn mục “**Columns**” để xem xét các cột được ánh xạ.

Sau khi kiểm tra đúng và đầy đủ, Click “**OK**”.



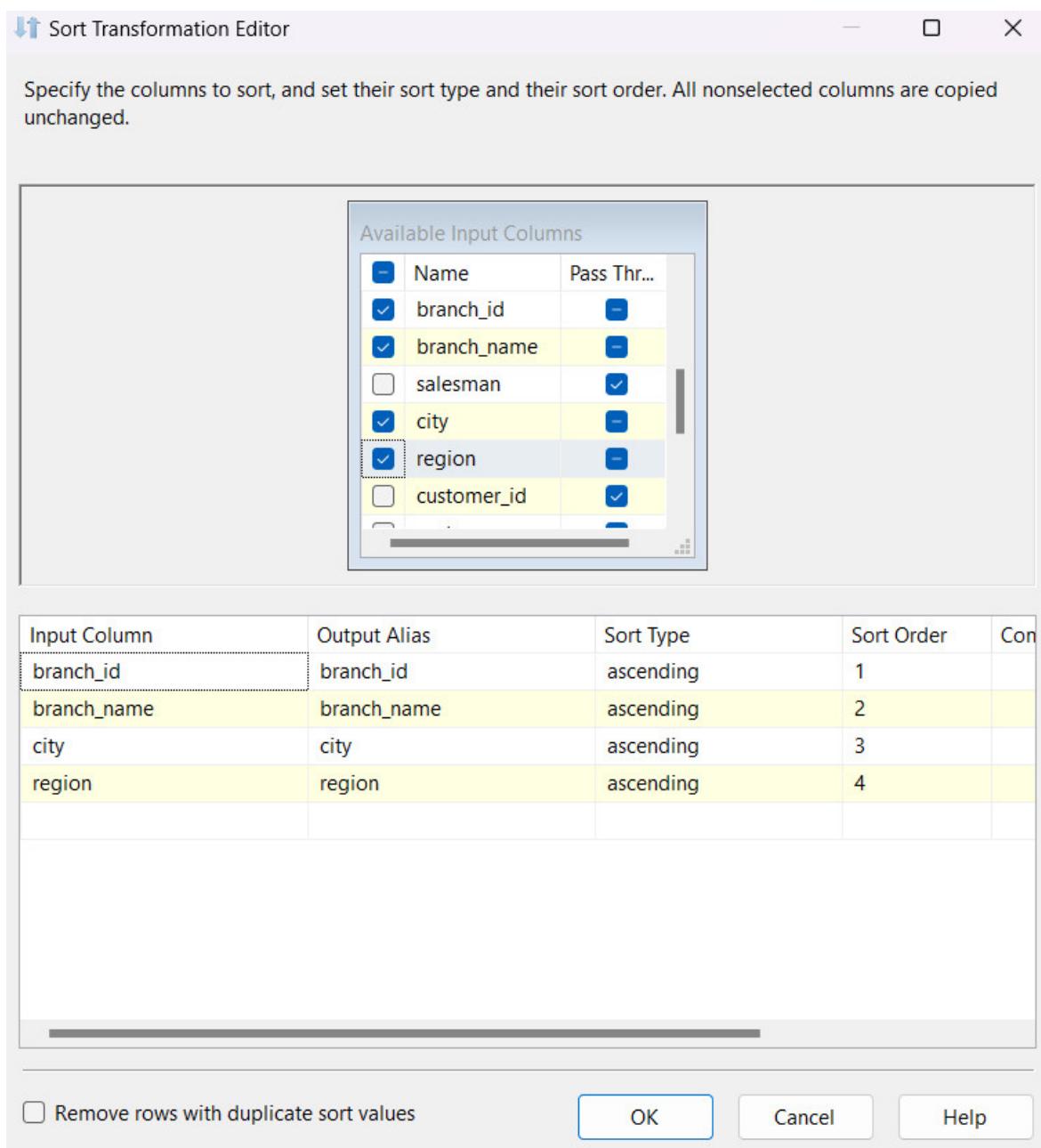
Hình 90: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 87

**Bước 5:** Tạo 2 Sort và đặt tên lần lượt là Sort và Sort1 tương ứng cho Product1 và Dim\_Branch.



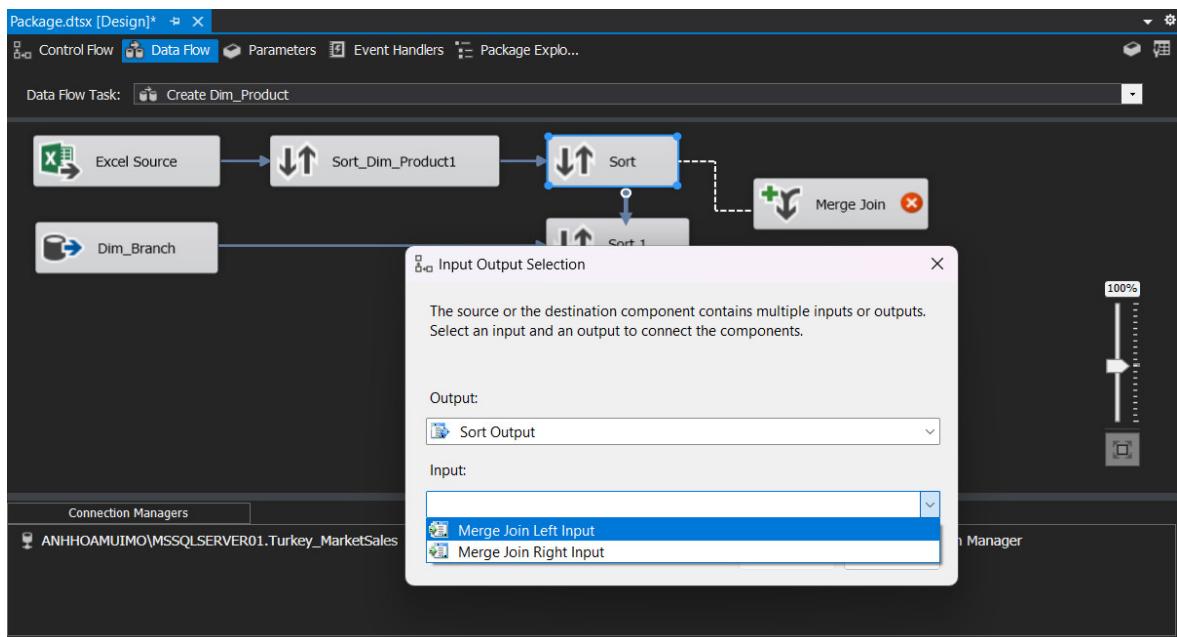
Hình 91: Tạo 2 Sort và đặt tên lần lượt là Sort và Sort1 tương ứng cho Product1 và Dim\_Branch

**Bước 6:** Ở Sort và Sort1, click chuột phải chọn “Edit” và chọn các cột theo thứ tự giống với bảng Dim\_Branch đã được thiết kế trong kho dữ liệu: branch\_id, branch\_name, city, region..



Hình 92: Chọn ra các cột dữ liệu cho bảng Dim\_Branch

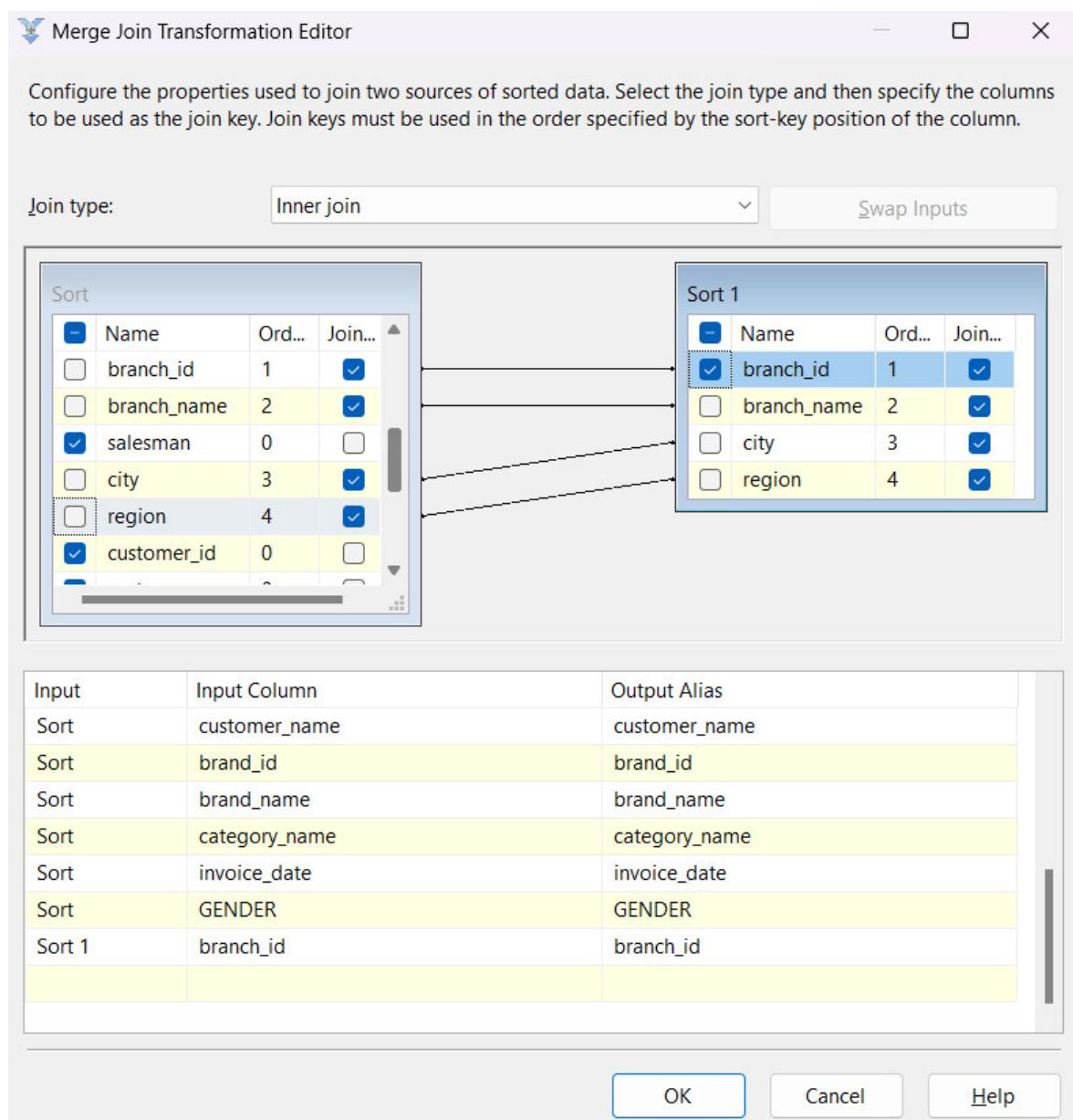
**Bước 7:** Tạo một “Merge Join” và kết nối với Sort, tiếp theo ta chọn “Merge Join Left Input” để giữ lại toàn bộ các dòng đã chọn trong Sort\_Dim\_Product1 bất kể có kết quả khi thực hiện phép kết trái với cột branch\_id của bảng Dim\_Branch hay không.



Hình 93: Tạo mới một Merge Join để chuẩn bị kết Dim\_Product và Dim\_Branch

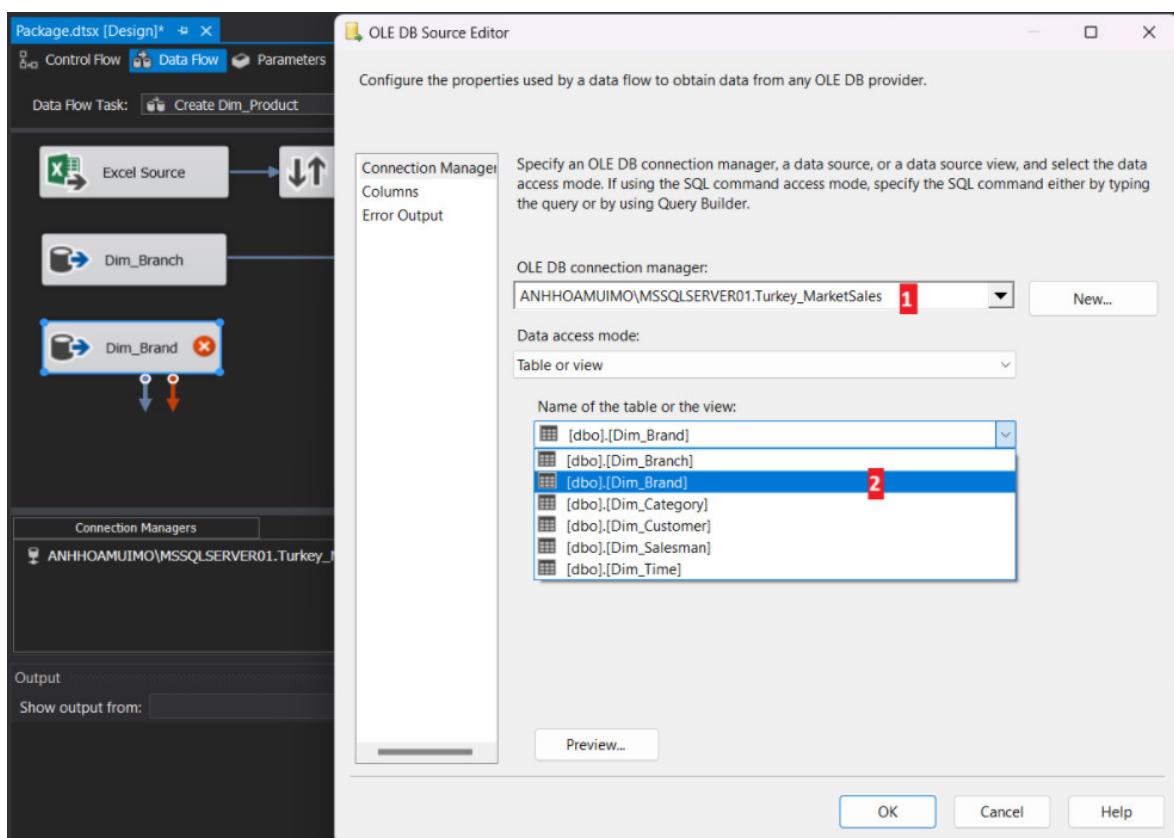
**Bước 8:** Click chuột phải vào “Merge Join” và chọn “Edit”, một hộp thoại merge editor xuất hiện. Tại vị trí này, ta tick chọn tất cả các cột của Sort nhưng không lấy 4 thuộc tính: branch\_id, branch\_name, city, region.

Tiếp theo, ta chọn thuộc tính branch\_id ở Sort1 để tiến hành merge Dim\_Branch và dữ liệu nguồn ban đầu tạo ra bảng Dim\_Product. Kết quả sau khi merge là bảng Product1 lúc này không còn 3 thuộc tính branch\_name, city, region.



Hình 94: Thực hiện phép kết trái giữa Dim\_Product và Dim\_Branch

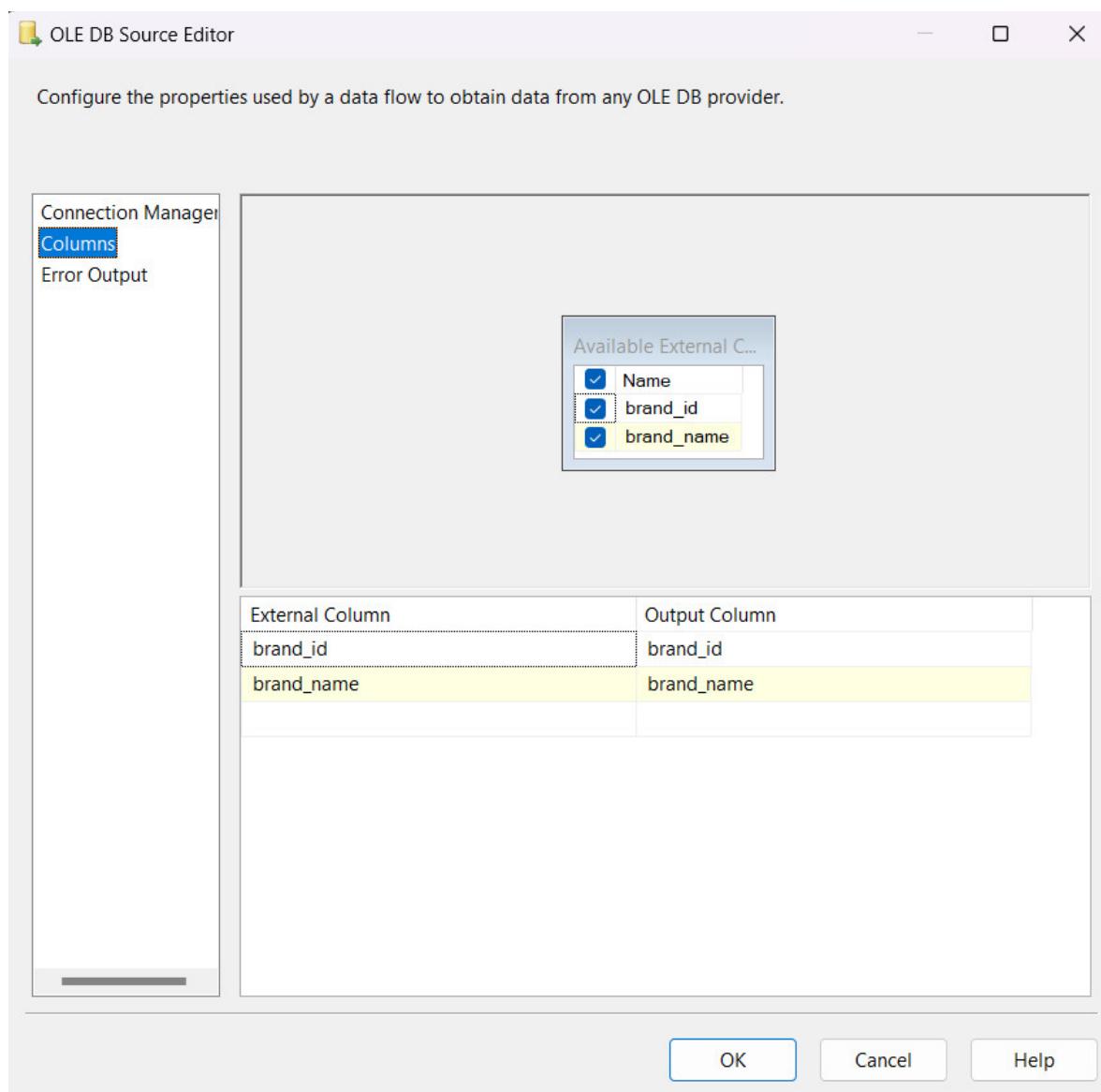
**Bước 9:** Ta tiếp tục tạo thêm “OLE DB Source” và đổi tên thành Dim\_Brand. Click chuột phải chọn “Edit”, sau đó chọn bảng Dim\_Brand đã tạo trước đó làm data source cho bảng Dim\_Product.



Hình 95:Tạo mới một OLE DB Source và chọn bảng Dim\_Brand làm data source

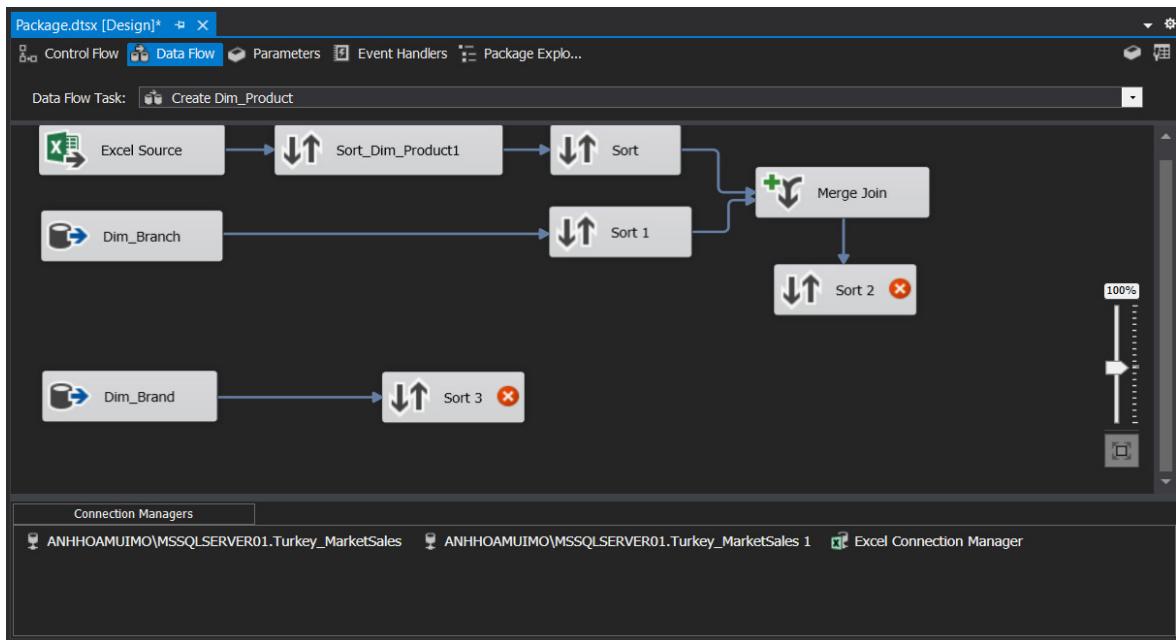
**Bước 10:** Chọn mục “**Columns**” để xem xét các cột được ánh xạ.

Sau khi kiểm tra đúng và đầy đủ, Click “**OK**”.



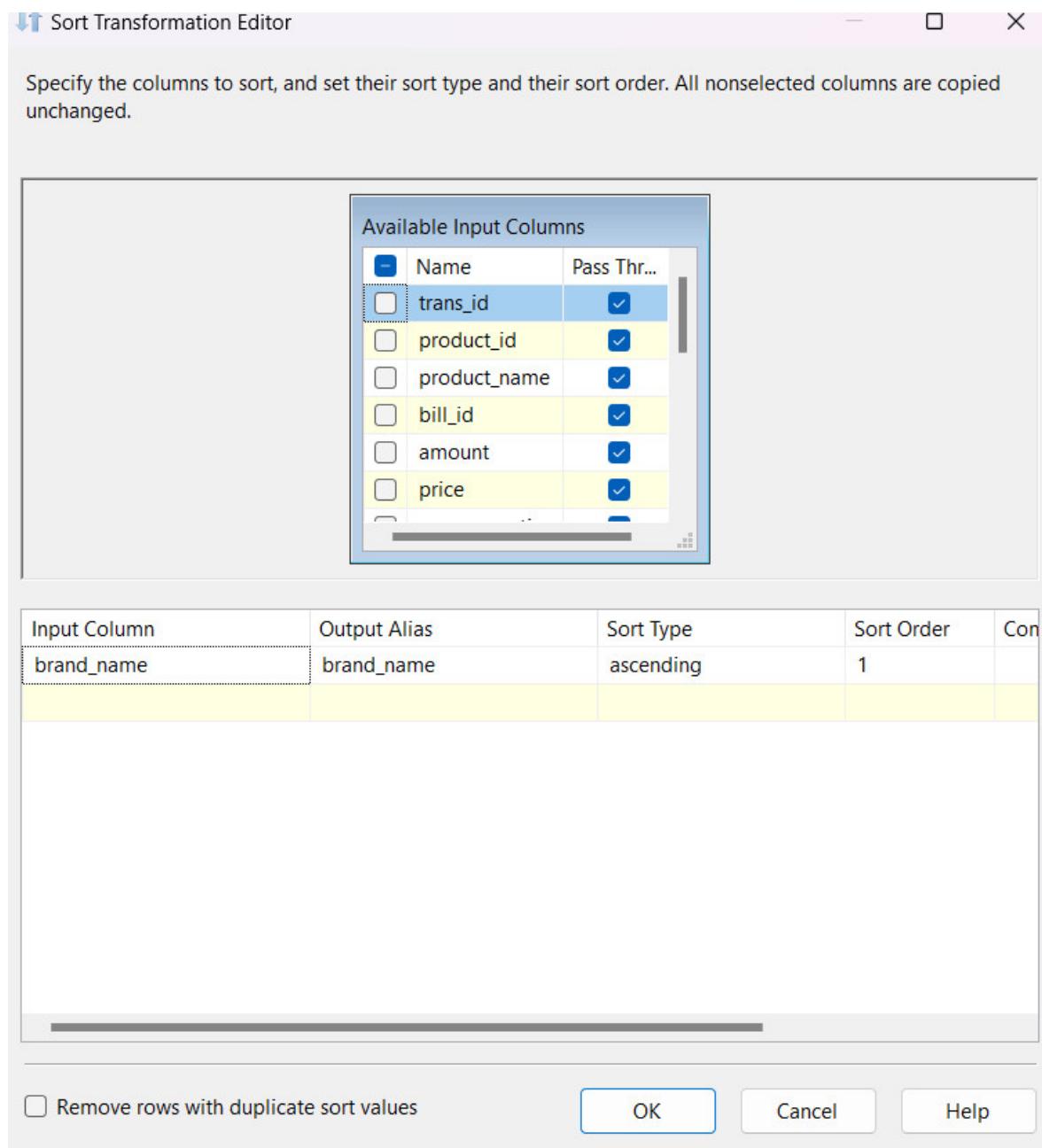
Hình 96: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 87

**Bước 11:** Lần lượt tạo 2 Sort và đặt tên là Sort2 và Sort3 tương ứng cho bảng Product2 và Dim\_Brand.



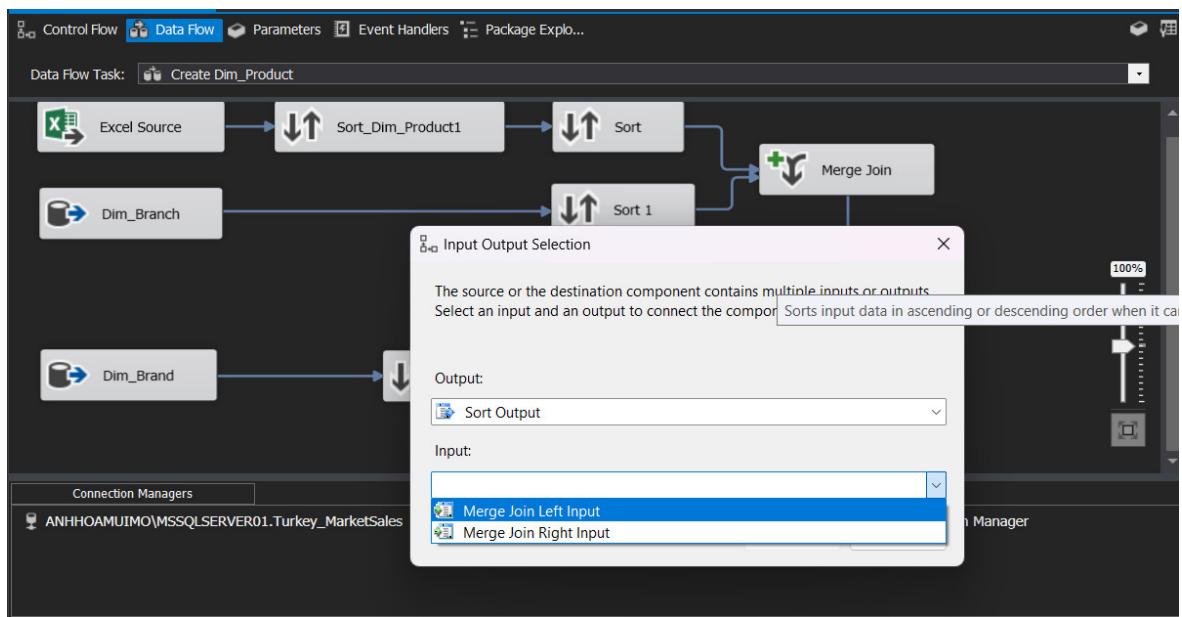
Hình 97:Tạo 2 Sort và đặt tên lần lượt là Sort2 và Sort3 tương ứng cho Product2 và Dim\_Branch

**Bước 12:** Ở Sort2 và Sort3 vừa tạo ở bước trên, click chuột phải chọn “Edit” và tick chọn cột brand\_name để chuẩn bị cho quá trình merge.



Hình 98: Chọn ra các cột dữ liệu cho bảng Dim\_Brand

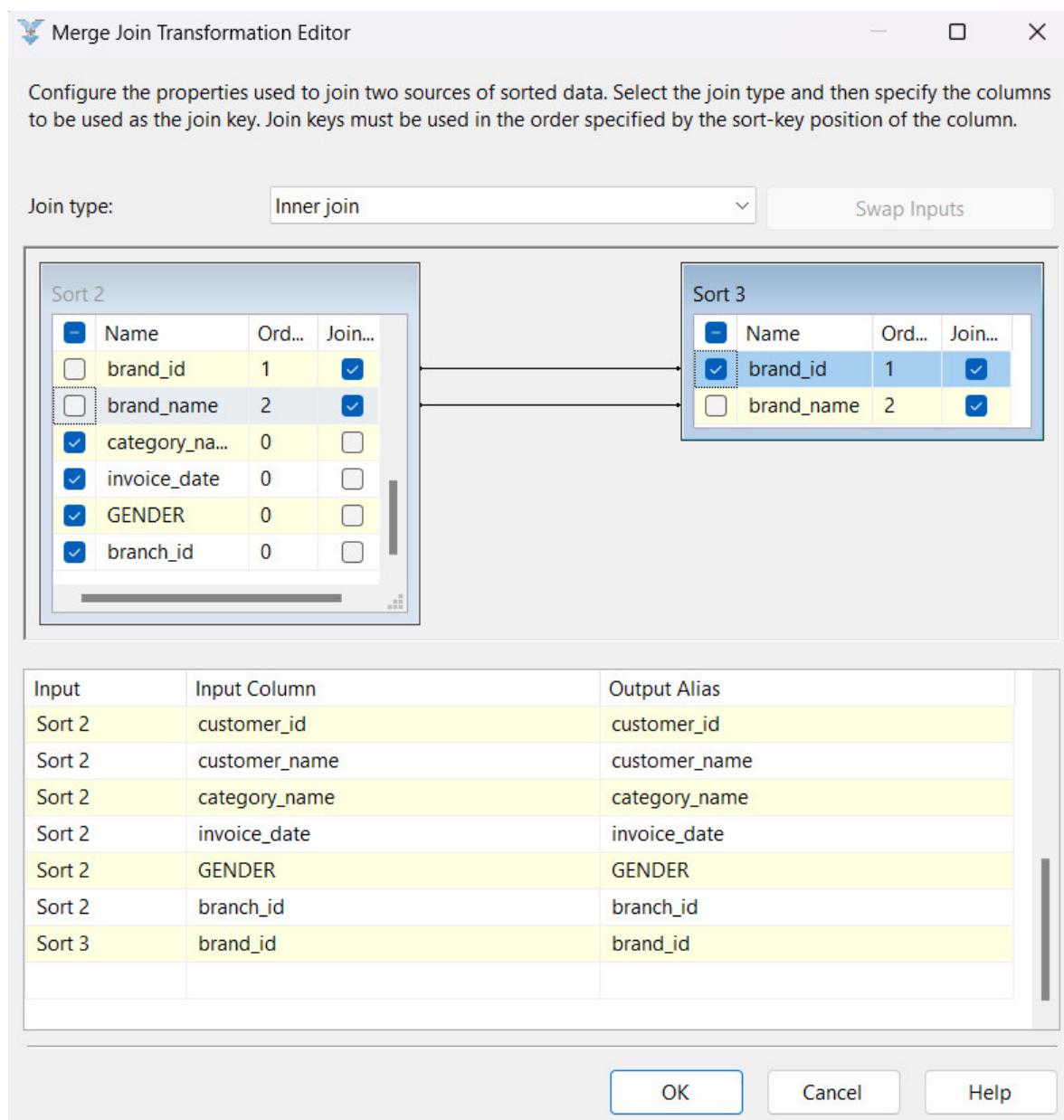
**Bước 13:** Tạo một “Merge Join” và kết nối với Sort2, tiếp theo ta chọn “Merge Join Left Input” để giữ lại toàn bộ các dòng đã merge trước đó bất kể có kết quả khi thực hiện phép kết trái với cột branch\_id của bảng Dim\_Brand hay không



Hình 99:Tạo mới một Merge Join để chuẩn bị kết Product1 và Dim\_Branch

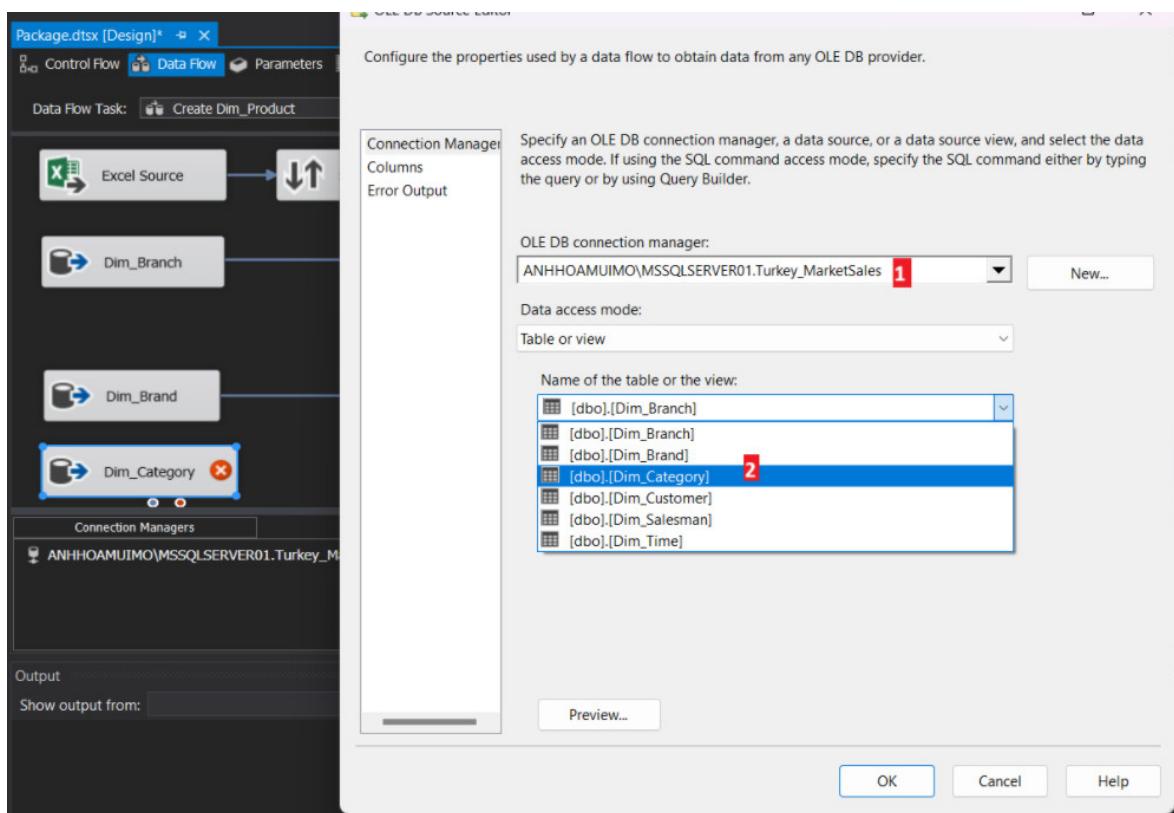
**Bước 14:** Kết nối Sort3 với “Merge Join”. Click chuột phải vào “Merge Join”, chọn “Edit”, một hộp thoại merge editor xuất hiện. Tại vị trí này, ta tick chọn tất cả các cột của Sort2 nhưng không lấy 2 thuộc tính: brand\_id và brand\_name.

Tiếp theo, ta chọn thuộc tính brand\_id ở Sort3 để tiến hành merge Dim\_Brand và Product1. Kết quả sau khi merge là bảng Product2 lúc này không còn thuộc tính brand\_name.



Hình 100: Thực hiện phép kết trái giữa Product1 và Dim\_Branch

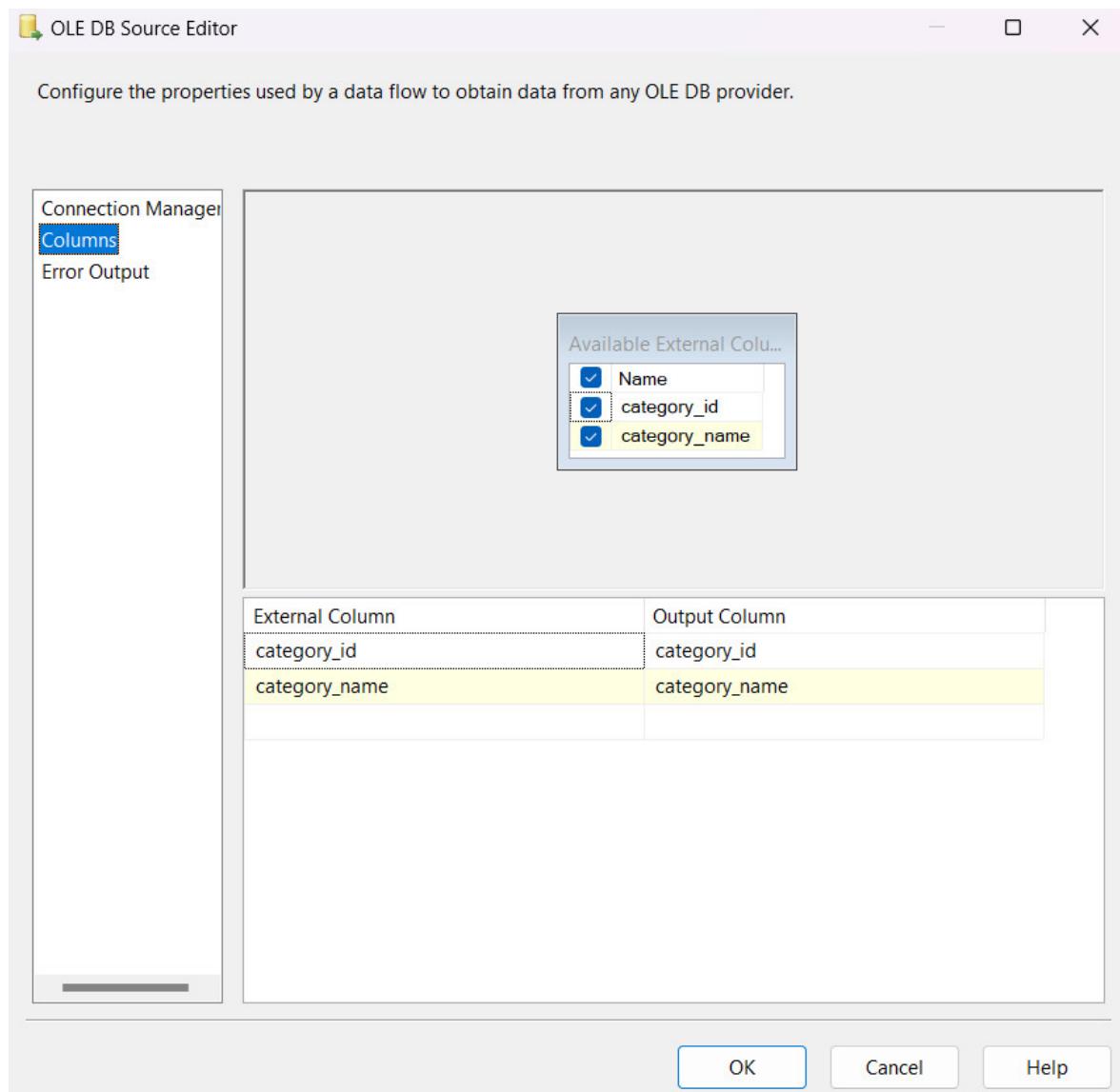
**Bước 15:** Ta tạo một “OLE DB Source” cuối cùng cho bảng Dim\_Category. Click chuột phải và chọn “Edit”, sau đó lựa chọn bảng Dim\_Category đã được tạo trước đó làm data source cho bảng Dim\_Product.



Hình 101: Tạo mới một OLE DB Source và chọn bảng Dim\_Category làm data source

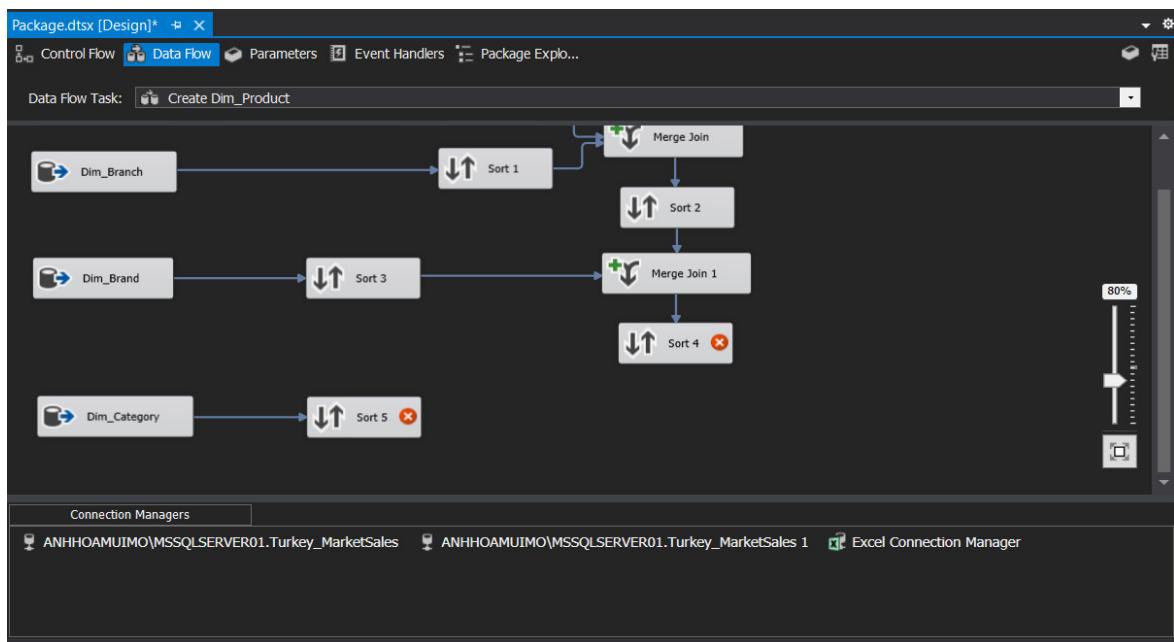
**Bước 16:** Chọn mục “Columns” để xem xét các cột được ánh xạ.

Sau khi kiểm tra đúng và đầy đủ, Click “OK”.



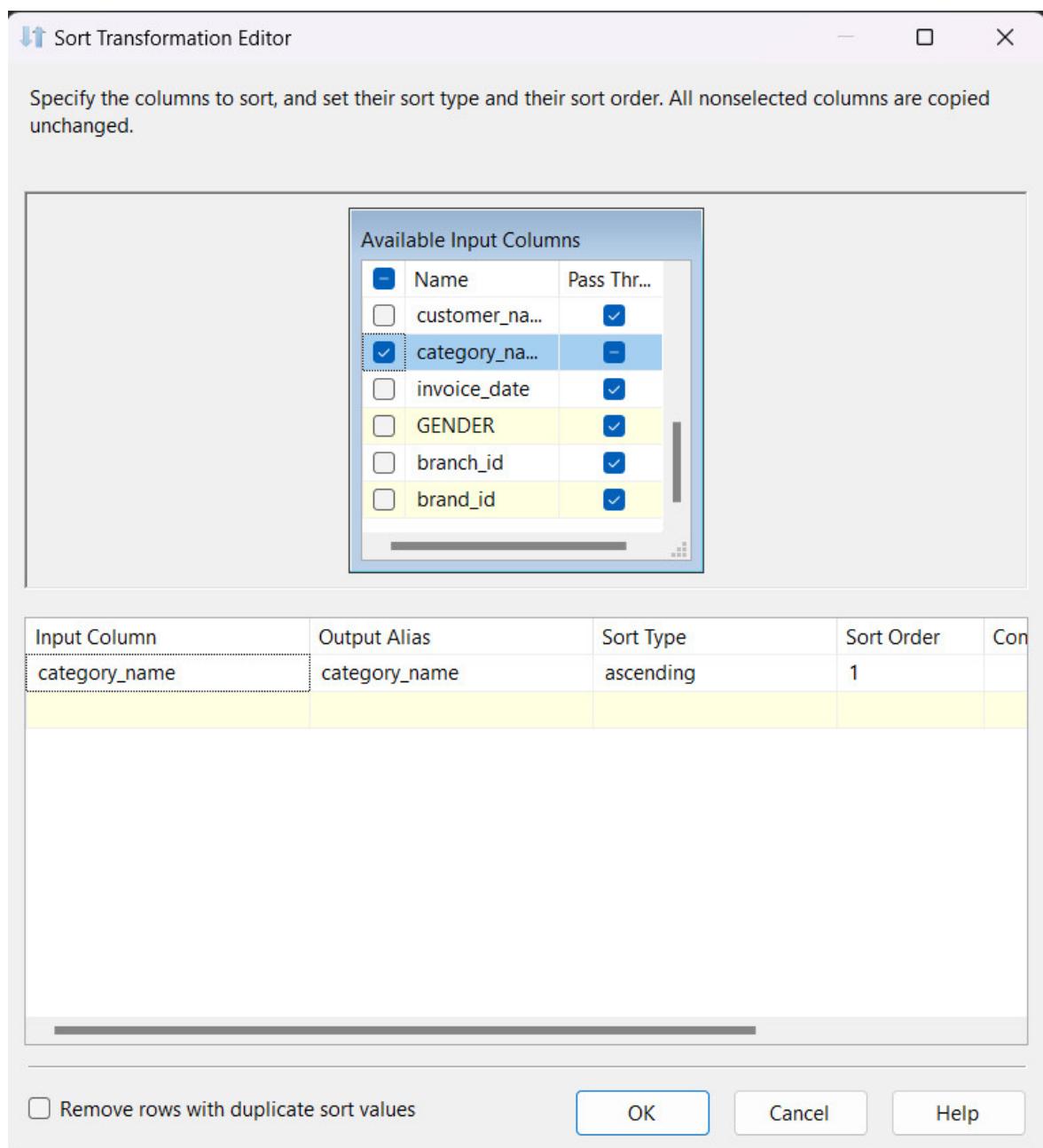
Hình 102: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 87

**Bước 17:** Tiếp tục tạo 2 Sort và đặt tên là Sort4 và Sort5 tương ứng cho Product3 và Dim\_Category.



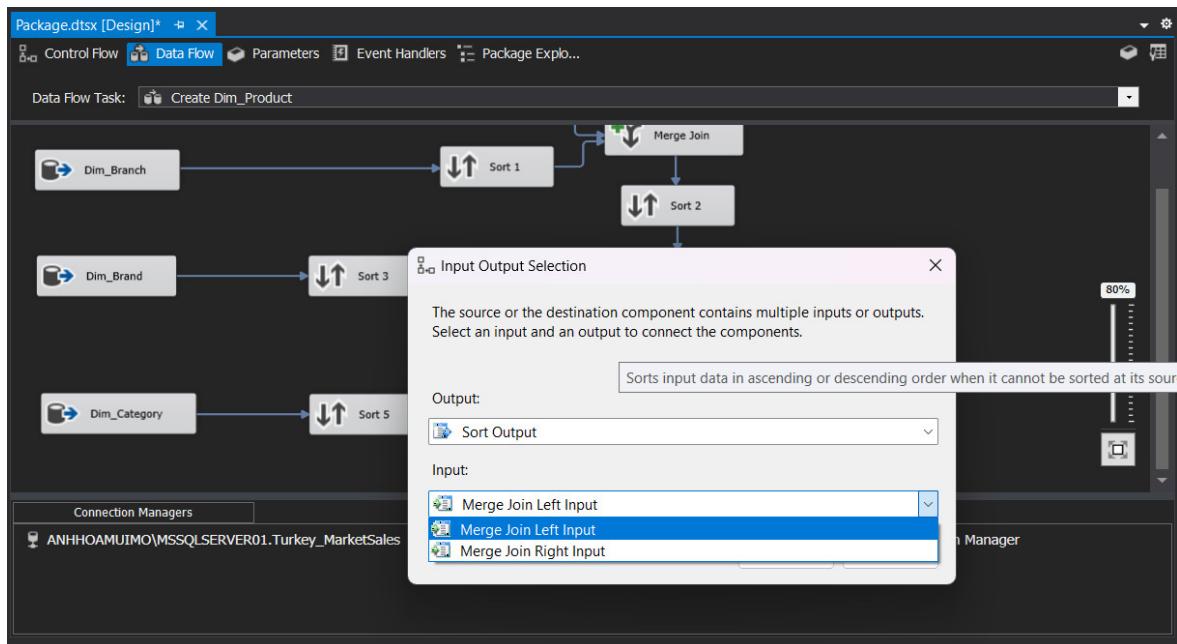
Hình 103: Tạo 2 Sort và đặt tên lần lượt là Sort4 và Sort5 tương ứng cho Product2 và Dim\_Category

**Bước 18:** Ở Sort4 và Sort5 vừa tạo ở bước trên, click chuột phải chọn “Edit” và tick chọn cột category\_name để chuẩn bị cho quá trình merge.



Hình 104: Chọn ra các cột dữ liệu cho bảng Dim\_Category

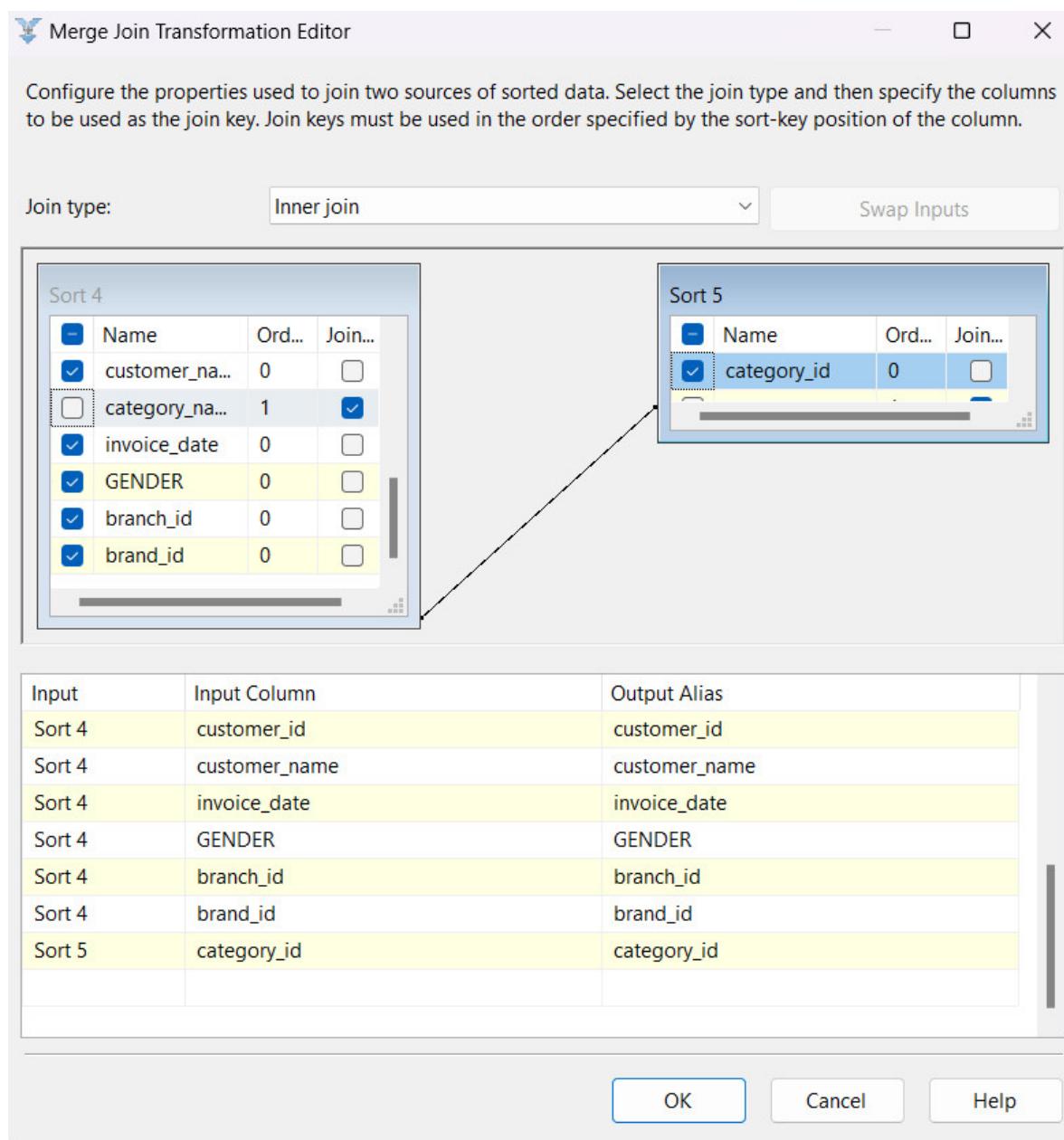
**Bước 19:** Tạo một “Merge Join” và kết nối với Sort4, tiếp theo ta chọn “Merge Join Left Input” để giữ lại toàn bộ các dòng đã merge trước đó bất kể có kết quả khi thực hiện phép kết trái với cột category\_id của bảng Dim\_Category hay không



Hình 105: Tạo mới một Merge Join để chuẩn bị kết Product2 và Dim\_Category

**Bước 20:** Kết nối Sort5 với “Merge Join”. Click chuột phải vào “Merge Join”, chọn “Edit”, một hộp thoại merge editor xuất hiện. Tại vị trí này, ta tick chọn tất cả các cột của Sort4 nhưng không lấy thuộc tính category\_name.

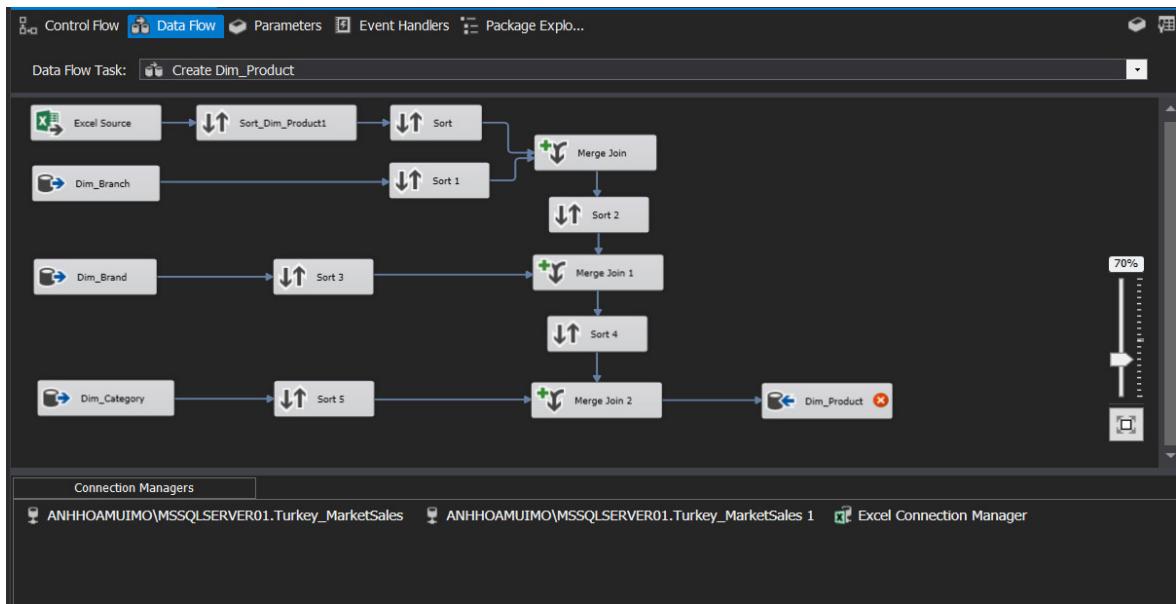
Tiếp theo, ta chọn thuộc tính category\_id ở Sort5 để tiến hành merge Dim\_Category và Product2. Kết quả sau khi merge là bảng Product3 lúc này không còn thuộc tính category\_name và thay thế bằng thuộc tính category\_id.



Hình 106: Thực hiện phép kết trái giữa Product2 và Dim\_Category

#### Bước 21: Tạo bảng Dim\_Product từ một “OLE DB Destination”

để lưu trữ toàn bộ dữ liệu đã merge trong các bước nêu trên, tạo khóa chính dim\_product\_id có kiểu int và tự động tạo giá trị tăng dần từ 1.



Hình 107: Tạo mới một OLE DB Destination và đặt tên là Dim\_Product

Nội dung câu lệnh SQL tạo bảng Dim\_Product như sau:

```
CREATE TABLE [Dim_Product] (
```

```
    [dim_product_id] int identity(1,1) constraint PK_Product primary
```

key,

```
    [product_id] float,
```

```
    [product_name] nvarchar(255),
```

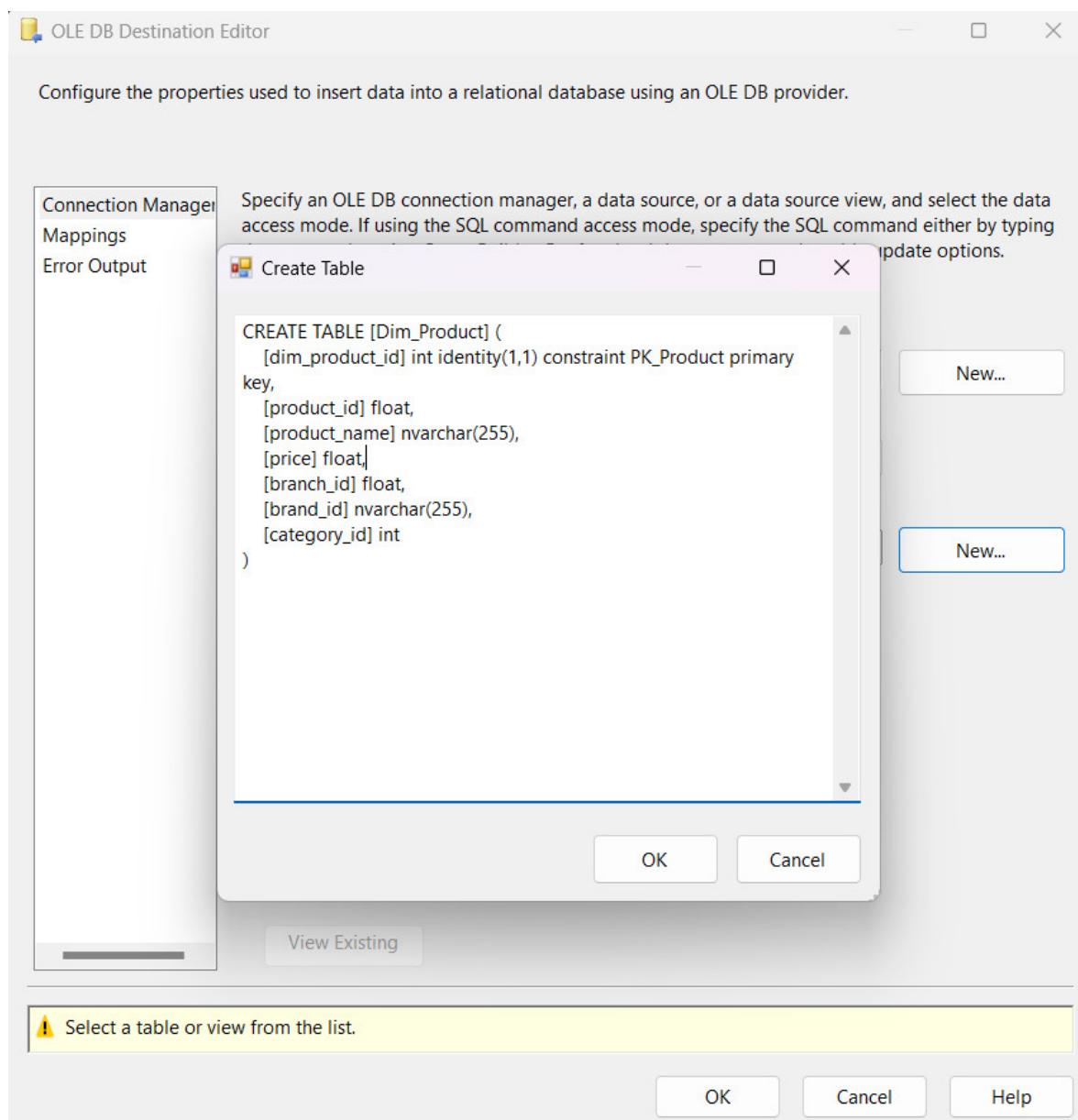
```
    [price] float,
```

```
    [branch_id] float,
```

```
    [brand_id] nvarchar(255),
```

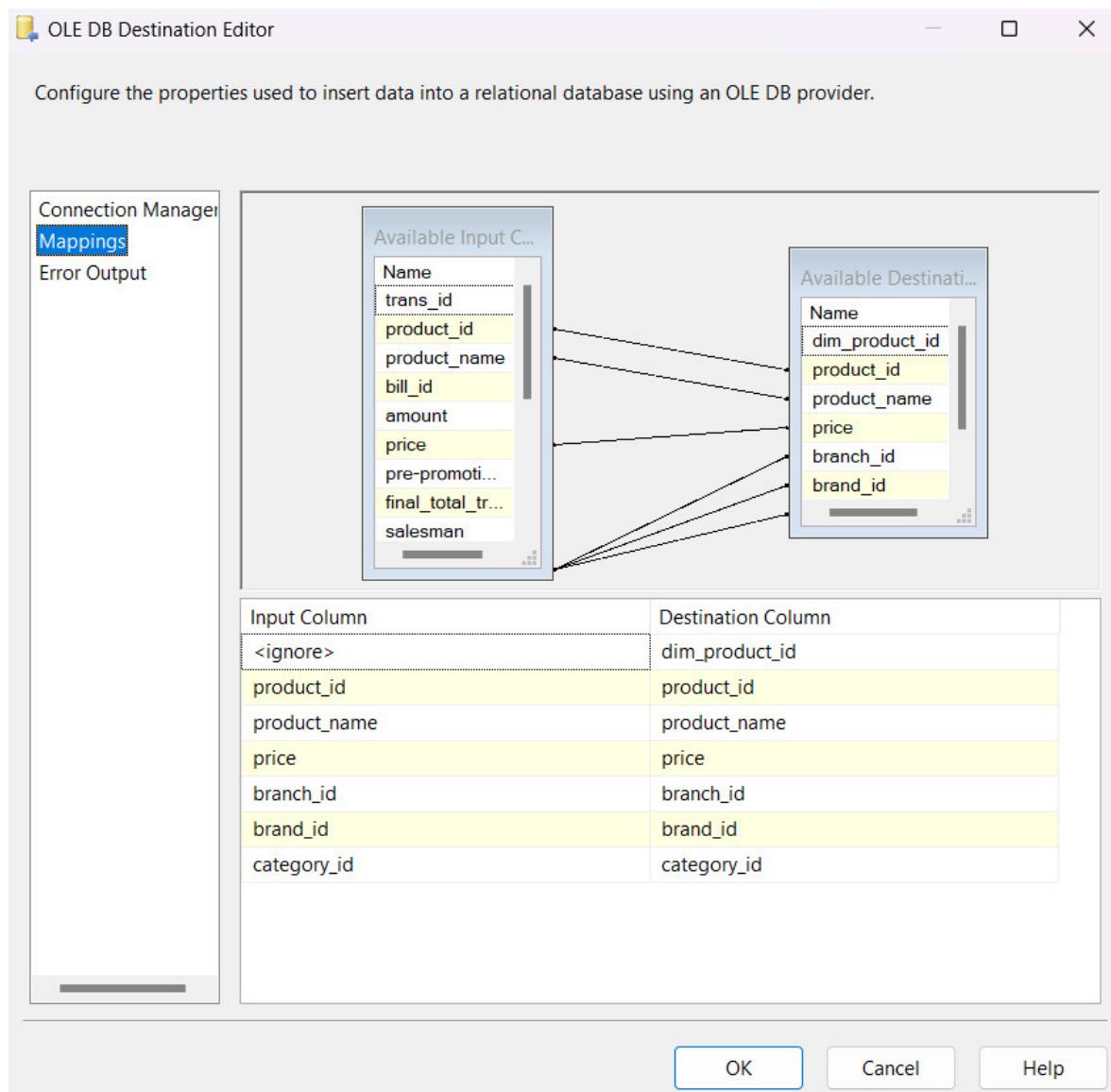
```
    [category_id] int,
```

)



Hình 108: Nhập các câu lệnh SQL tương ứng với bảng Dim\_Product

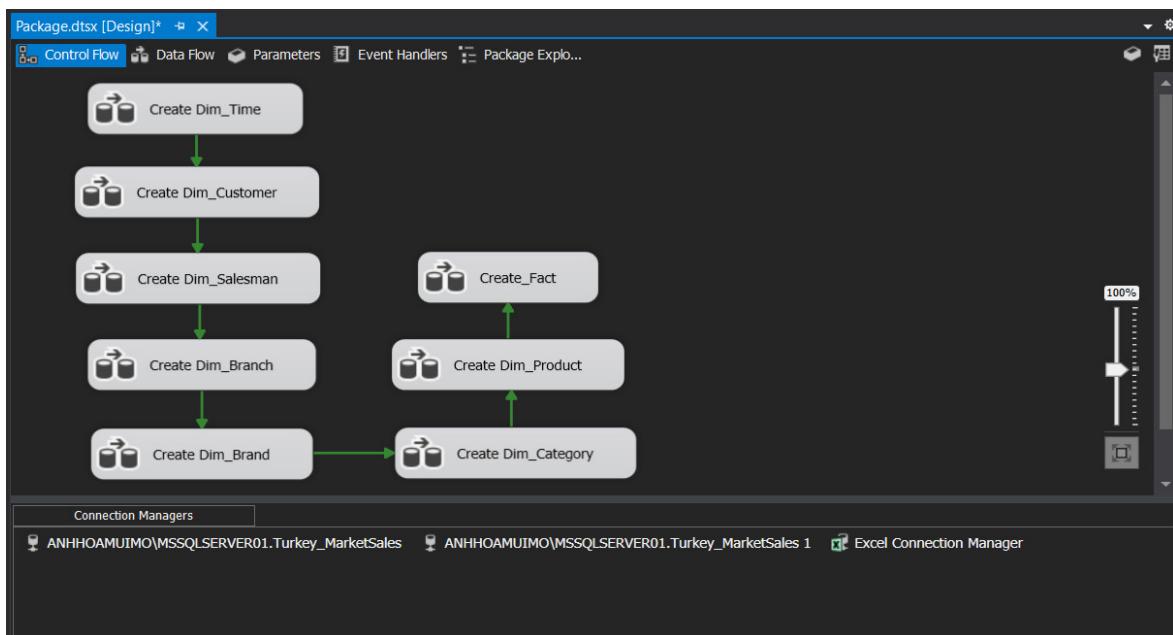
**Bước 22:** Chọn mục “**Mappings**” để xem xét việc ánh xạ các cột dữ liệu.



Hình 109: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 87

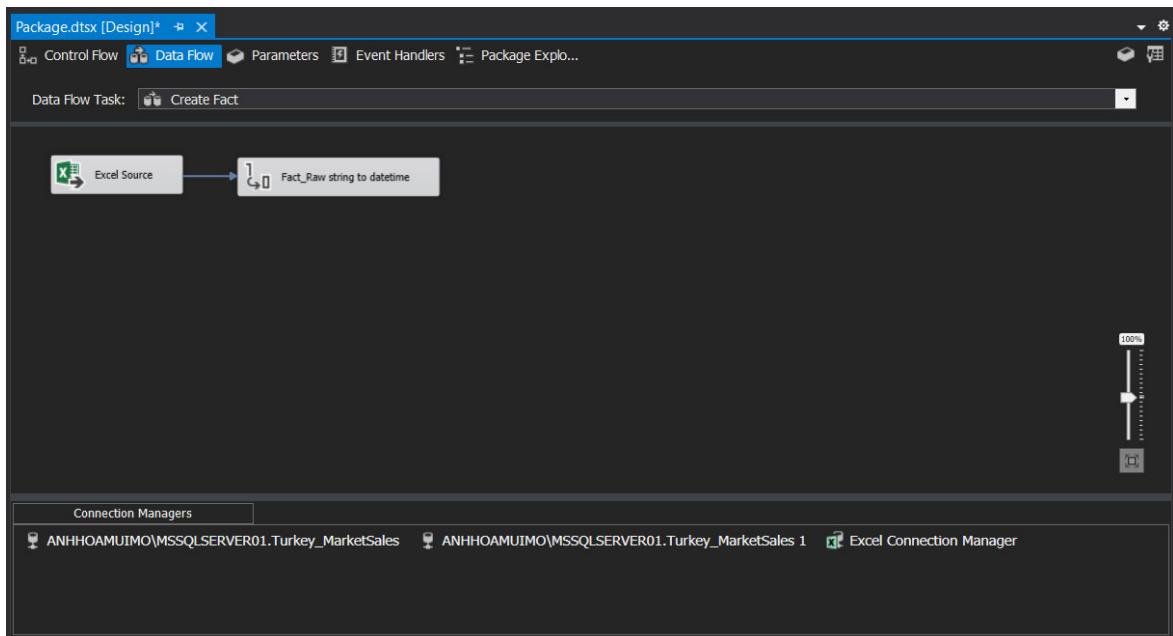
### 2.3.1.8. Bảng Fact\_Supermarket:

**Bước 1:** Tạo “Data Flow Task” đổi tên thành Create Fact.



Hình 110: Tạo mới một Data Flow Task cho bảng Fact\_Supermarket

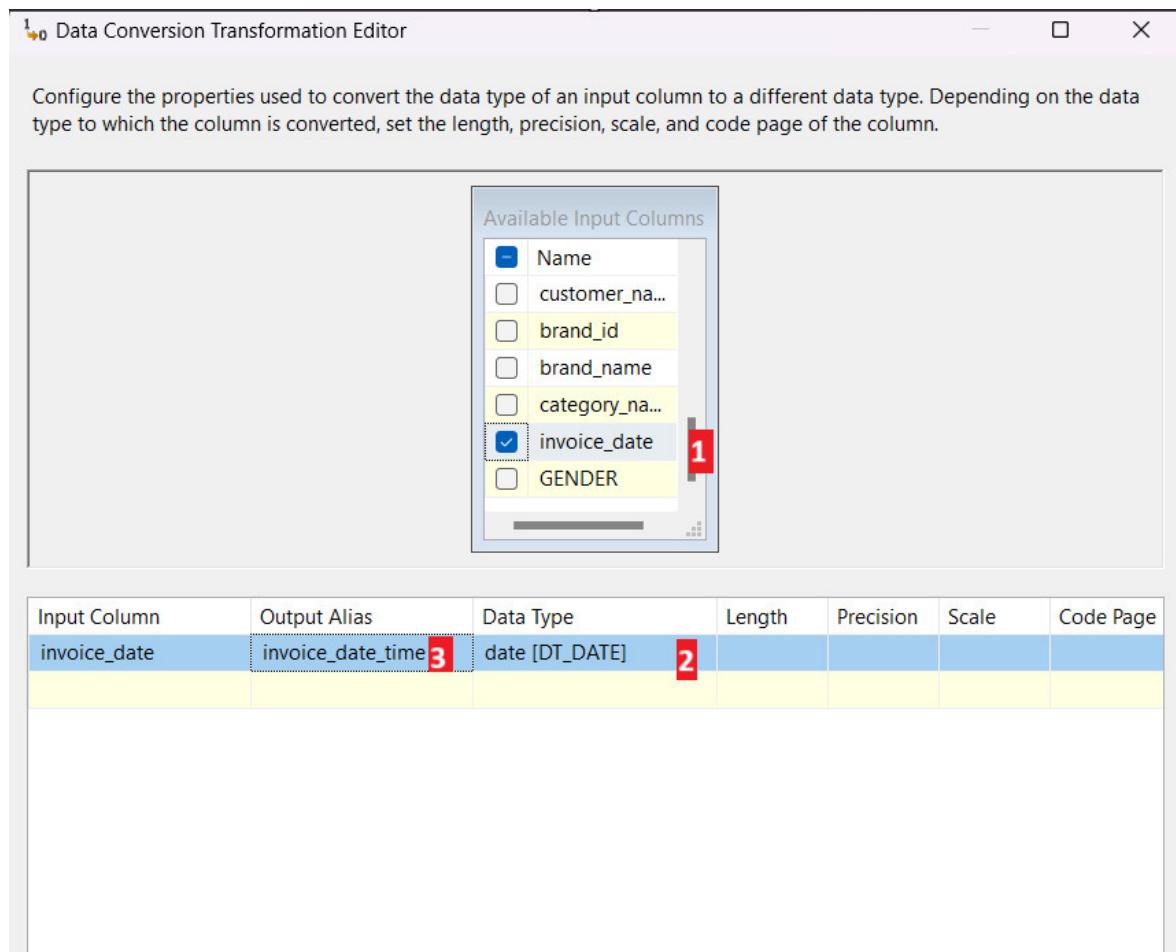
**Bước 2:** Trong “Create Fact” vừa tạo, tạo một “Excel Source” và chọn file dữ liệu nguồn, tạo một “Data Conversion” và đặt tên là “Fact\_Raw string to datetime”.



Hình 111: Tạo mới một Excel Source và một Data Conversion

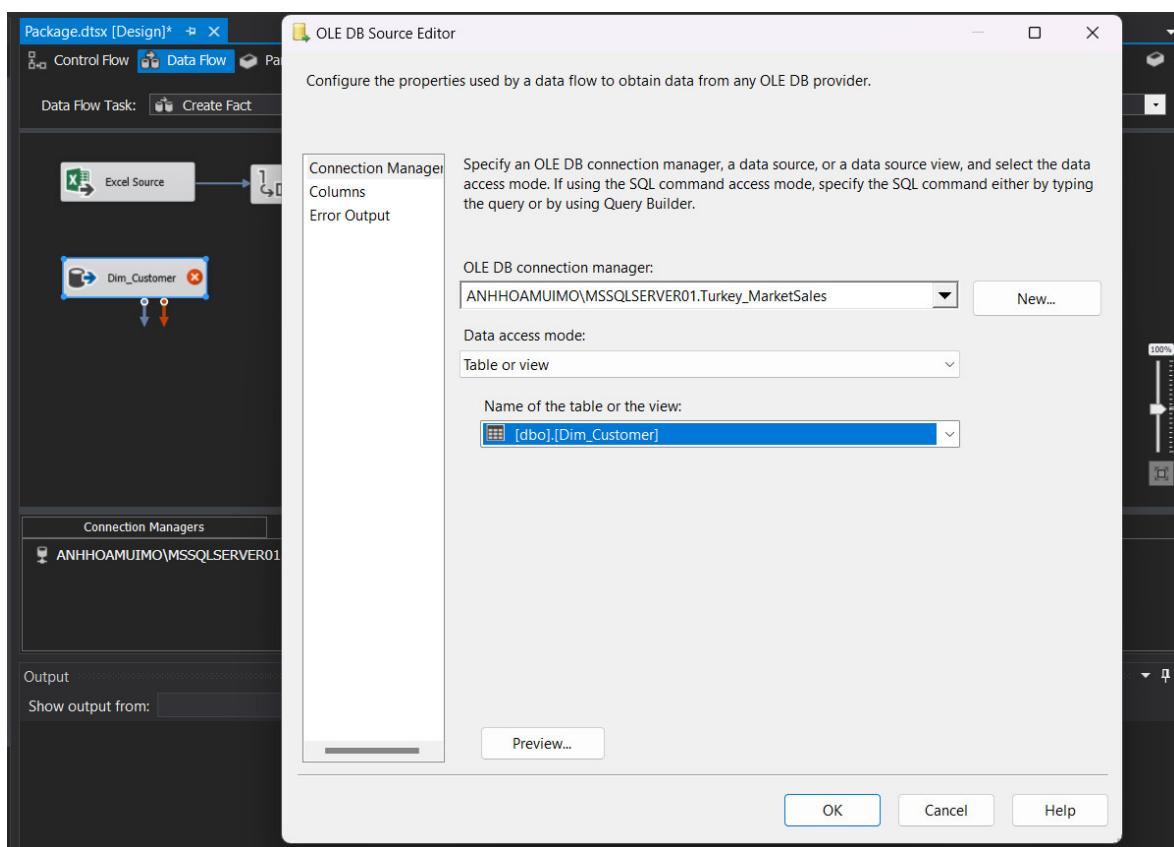
**Bước 3:** Click chuột phải vào “Data Conversion” vừa tạo ra, chọn “Edit”. Sau đó, tick chọn cột invoice\_date. Ở cột DataType ta thấy kiểu dữ liệu mặc định của cột này là string (string [DT\_STR]), ta chỉnh sửa lại kiểu

dữ liệu datetime (date [DT\_DATE]). Cuối cùng, ta đổi tên “**Output Alias**” thành invoice\_date\_time sau đó nhấn “**OK**” để hoàn tất.



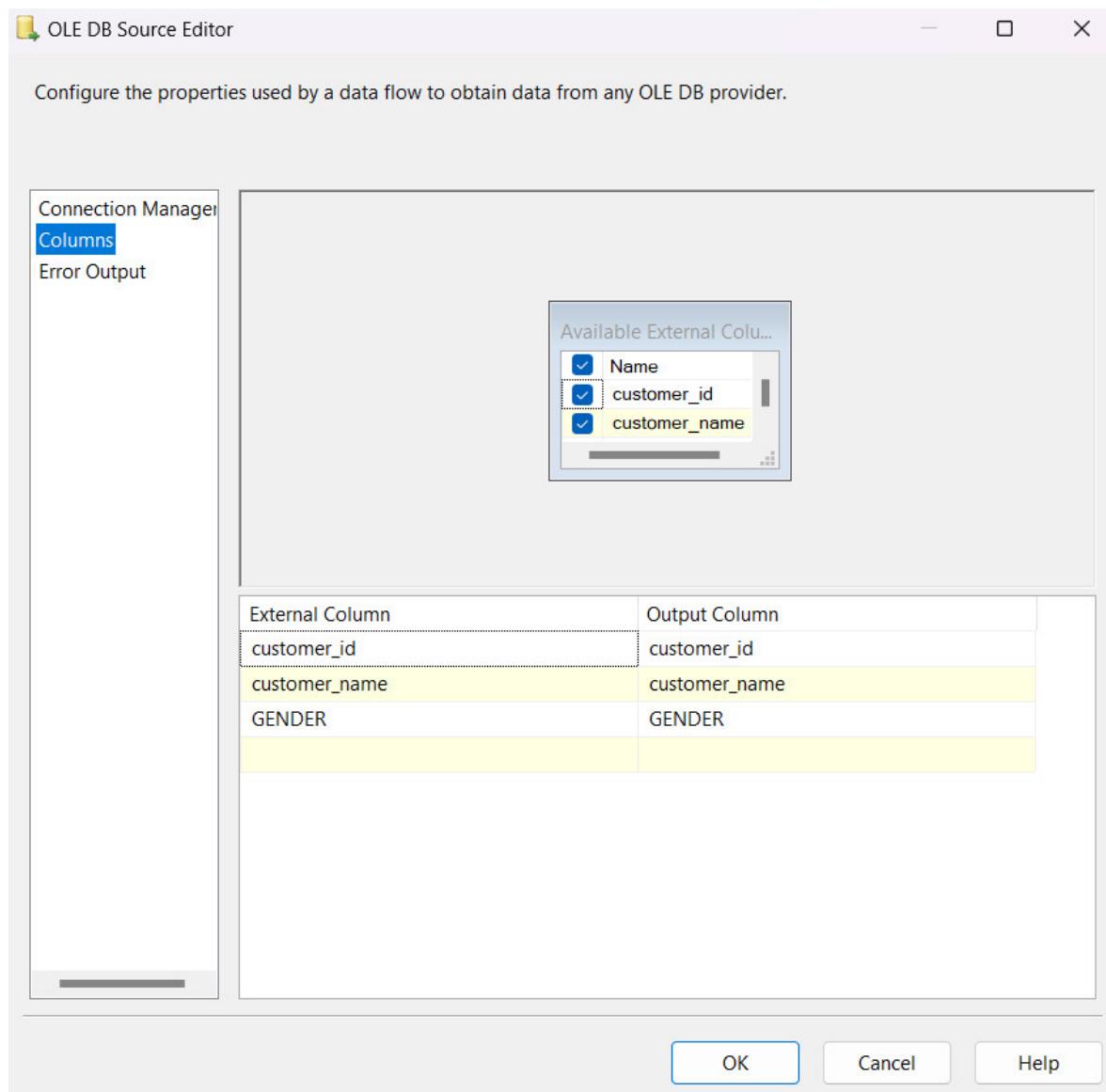
Hình 112: Đổi kiểu dữ liệu của cột invoice\_date và output alias

**Bước 4:** Ta tạo thêm “**OLE DB Source**” và đổi tên thành Dim\_Customer. Click chuột phải chọn “**Edit**”, sau đó chọn bảng Dim\_Customer đã được tạo trước đó làm data source cho bảng Dim\_Customer mới này.



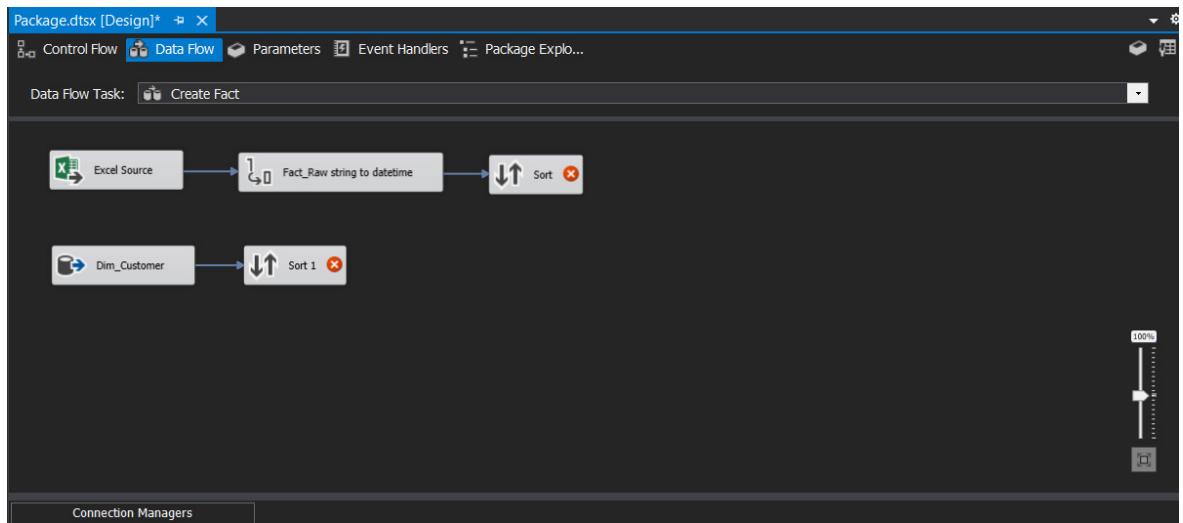
Hình 113:Tạo mới một OLE DB Source và chọn bảng Dim\_Customer làm data source

**Bước 5:** Chọn mục “**Columns**” để xem xét các cột được ánh xạ. Sau khi kiểm tra đúng và đầy đủ, Click “**OK**”.



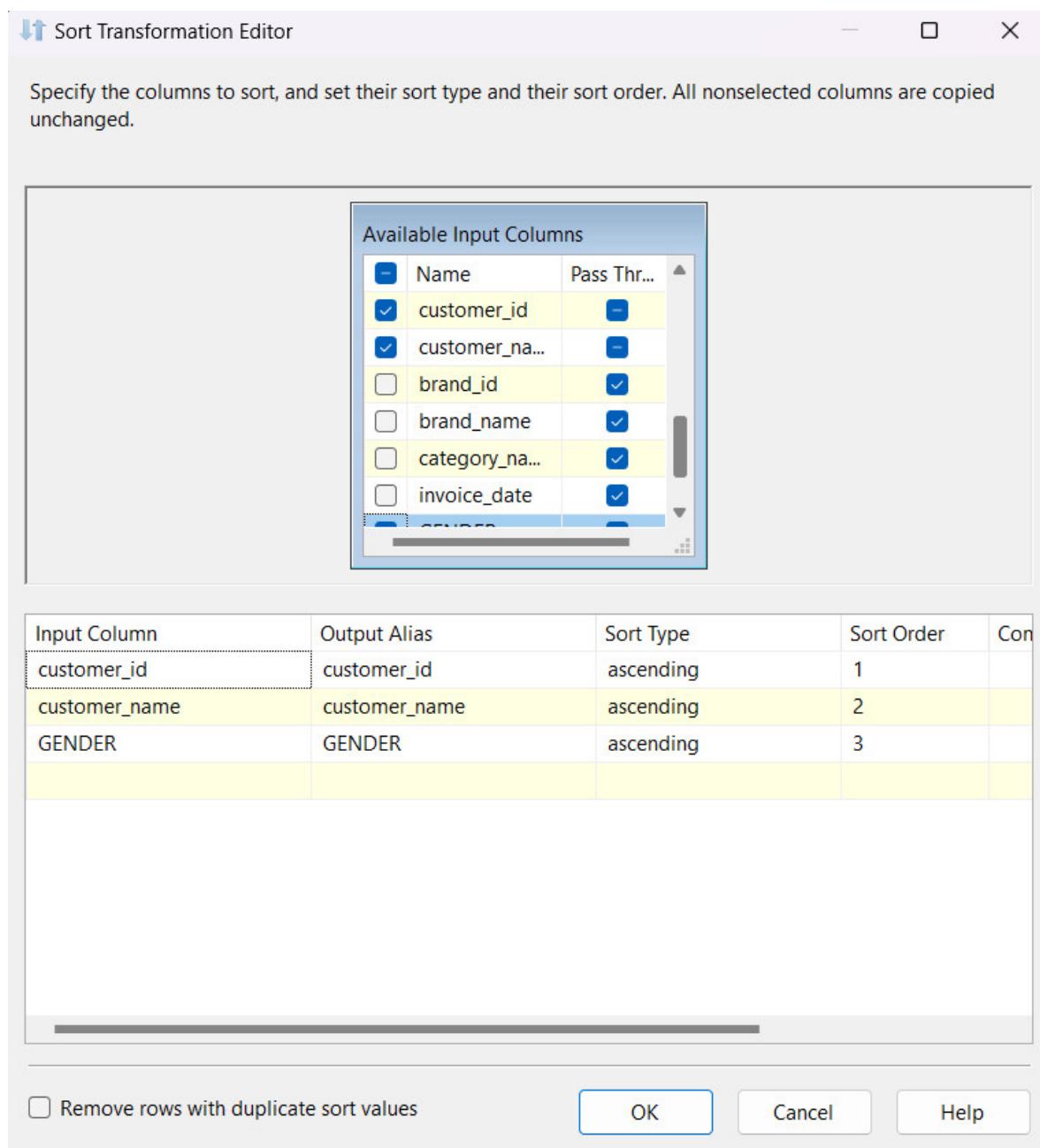
Hình 114: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 6

**Bước 6:** Tạo 2 Sort và đặt tên lần lượt là Sort và Sort1 tương ứng cho Fact và Dim\_Customer.



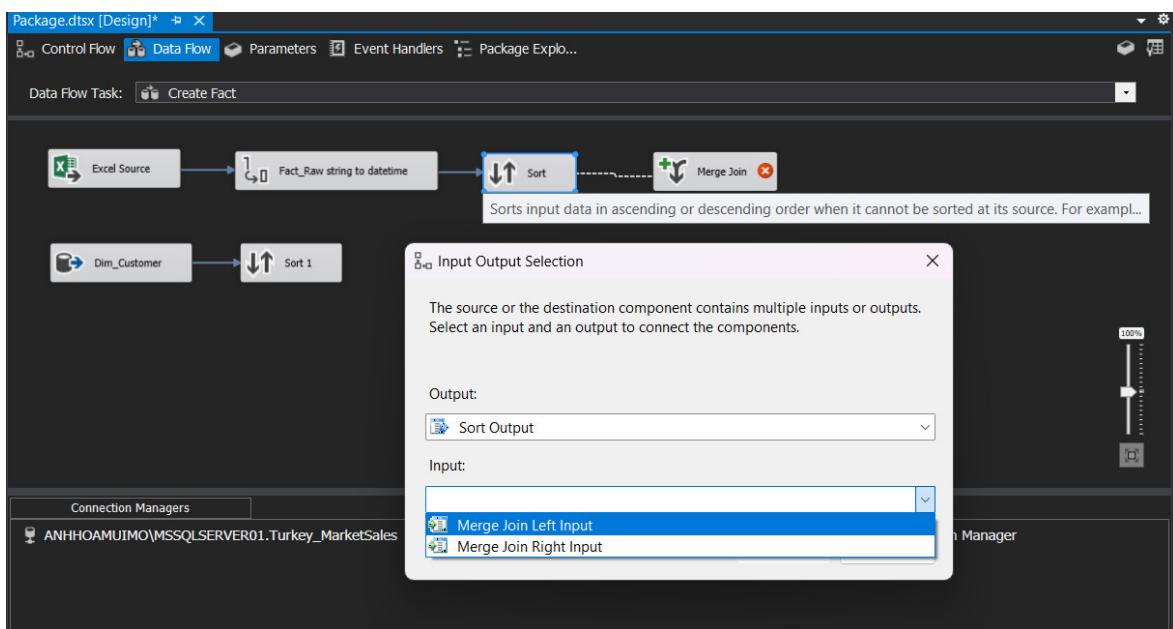
Hình 115: Tạo 2 Sort và đặt tên lần lượt là Sort và Sort1 tương ứng cho Fact\_Supermarket và Dim\_Customer

**Bước 7:** Ở Sort và Sort1, click chuột phải chọn “Edit” và tick chọn các cột customer\_id, customer\_name và GENDER để chuẩn bị cho quá trình merge.



Hình 116: Chọn ra các cột dữ liệu cho bảng Dim\_Customer

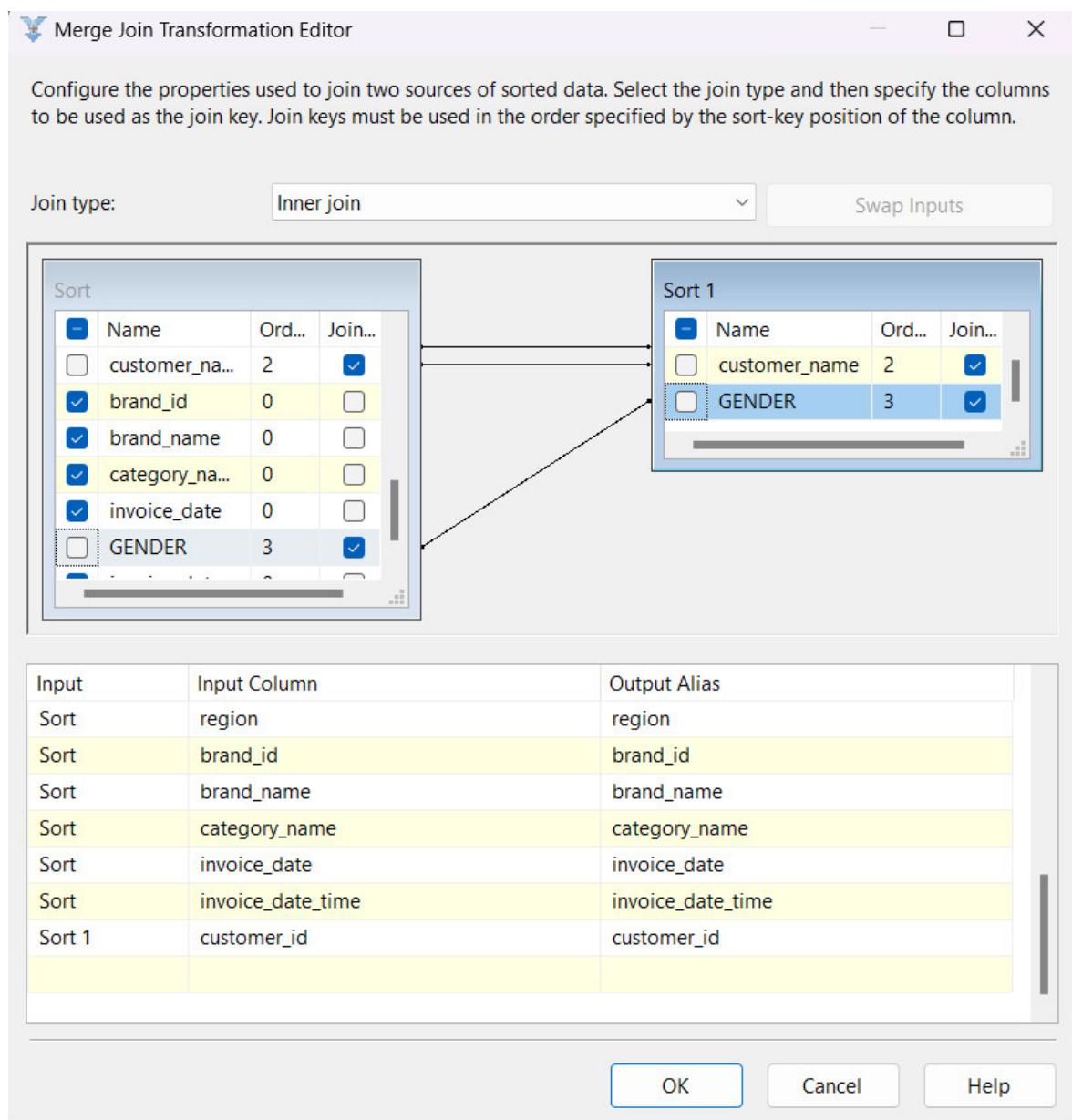
**Bước 8:** Tạo một “Merge Join” và kết nối với Sort, tiếp theo ta chọn “Merge Join Left Input” để giữ lại toàn bộ các dòng đã merge trước đó bất kể có kết quả khi thực hiện phép kết trái với cột customer\_id của bảng Dim\_Customer hay không.



Hình 117: Tạo mới một Merge Join để chuẩn bị kết Fact\_Supermarket và Dim\_Customer

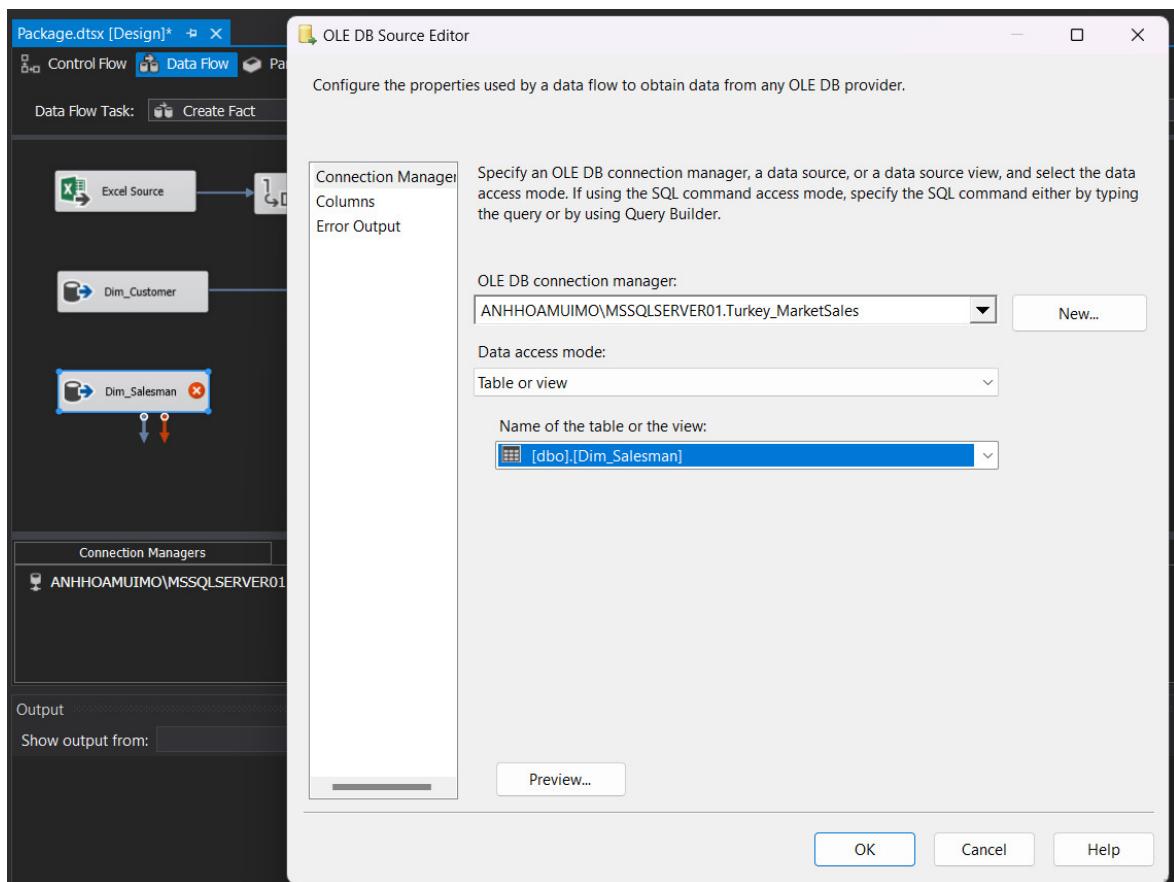
**Bước 9:** Kết nối Sort1 với “Merge Join”. Click chuột phải vào “Merge Join”, chọn “Edit”, một hộp thoại merge editor xuất hiện. Tại vị trí này, ta tick chọn tất cả các cột của Sort nhưng không 3 lấy thuộc tính: customer\_id, customer\_name và GENDER.

Tiếp theo, ta chọn thuộc tính category\_id ở Sort1 để tiến hành merge Dim\_Customer và Fact\_Supermarket. Kết quả sau khi merge là Fact\_Supermarket lúc này không còn thuộc tính customer\_name và GENDER.



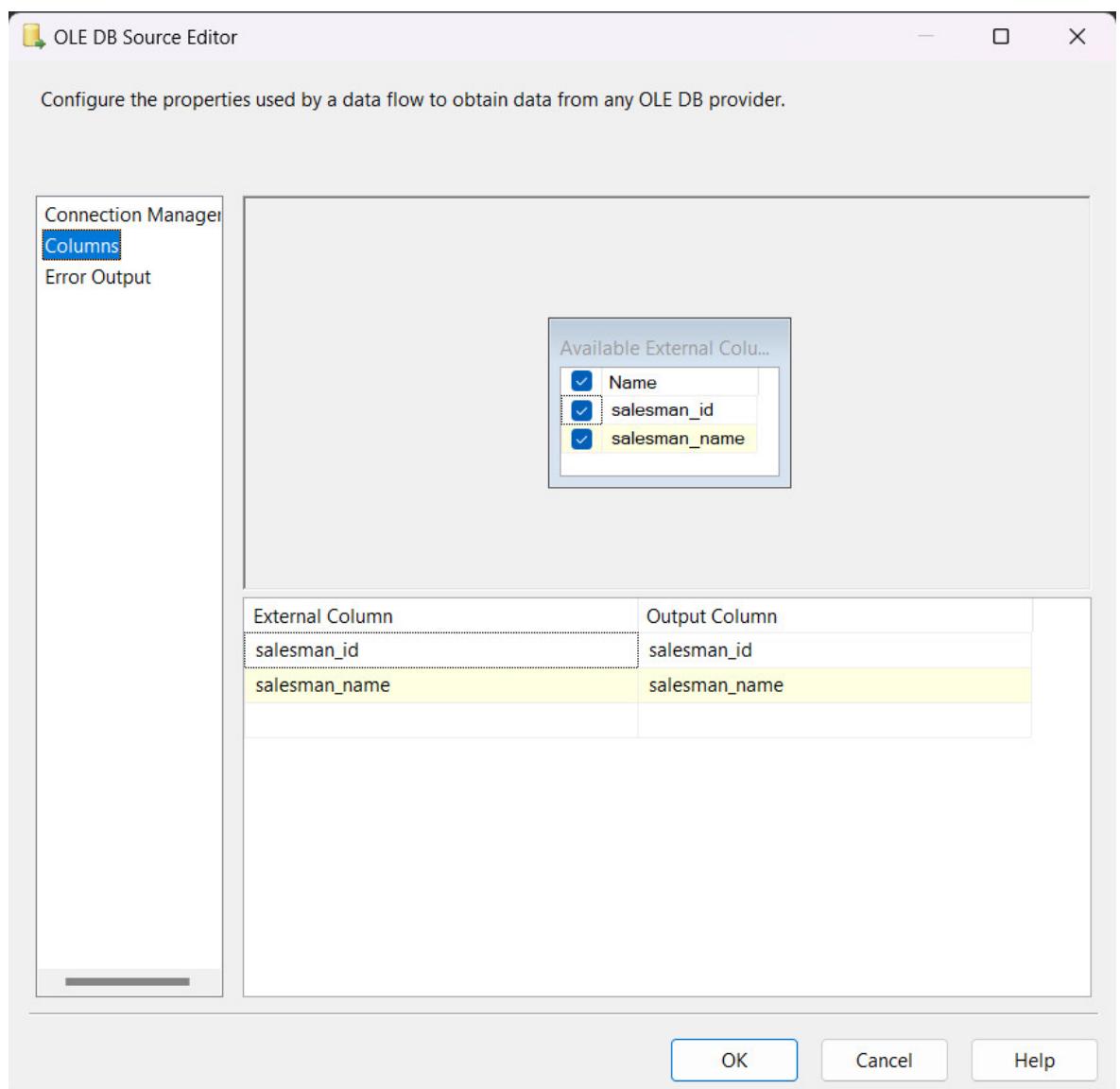
Hình 118: Thực hiện phép kết trái giữa Fact\_Supermarket và Dim\_Customer

**Bước 10:** Ta tạo thêm “**OLE DB Source**” và đổi tên thành **Dim\_Salesman**. Click chuột phải chọn “**Edit**”, sau đó chọn bảng **Dim\_Salesman** đã được tạo trước đó làm data source cho bảng **Dim\_Salesman** mới này.



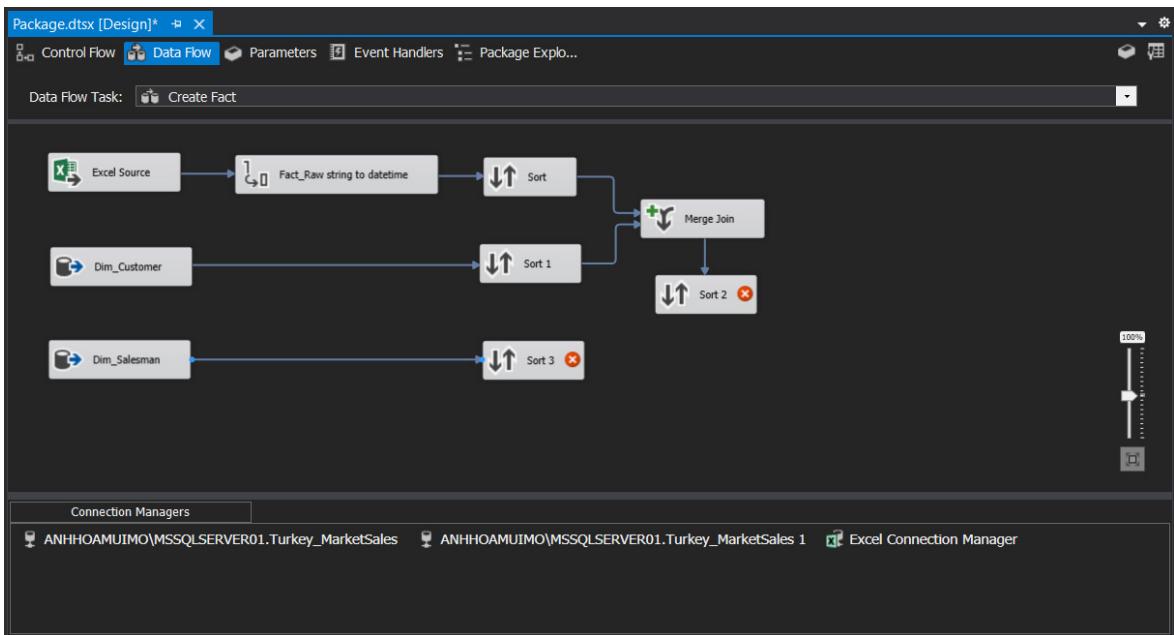
Hình 119: Tạo mới một OLE DB Source và chọn bảng Dim\_Salesman làm data source

**Bước 11:** Chọn mục “**Columns**” để xem xét các cột được ánh xạ.  
Sau khi kiểm tra đúng và đầy đủ, Click “**OK**”.



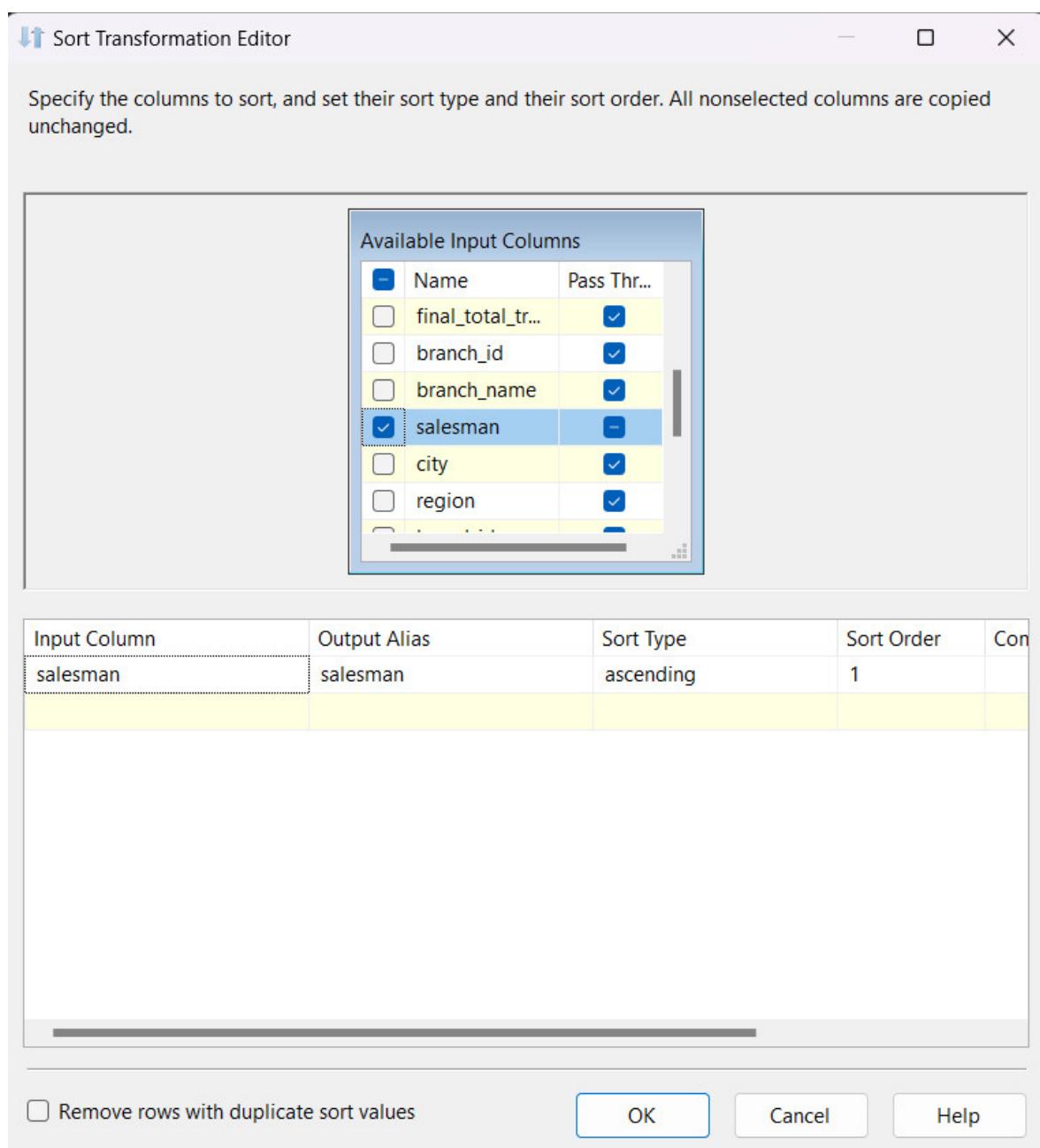
Hình 120: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 6

**Bước 12:** Tạo 2 Sort và đặt tên lần lượt là Sort2 và Sort3 tương ứng cho Fact và Dim\_Salesman.



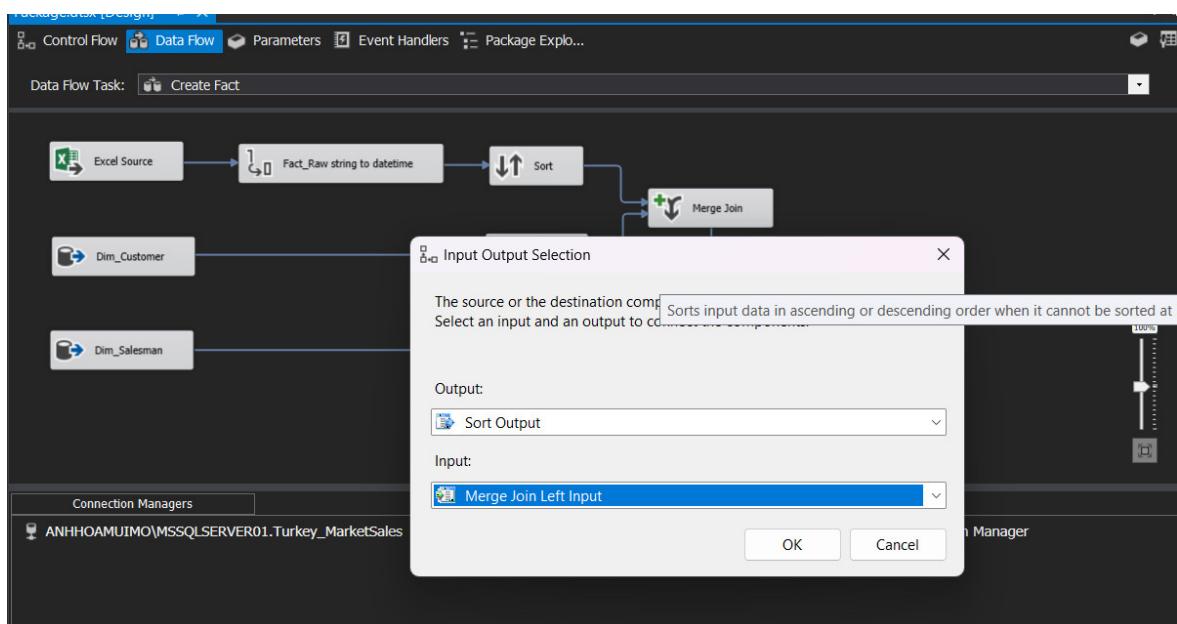
Hình 121: Tạo 2 Sort và đặt tên lần lượt là Sort2 và Sort3 tương ứng cho Fact\_Supermarket và Dim\_Salesman

**Bước 13:** Ở Sort2 và Sort3, click chuột phải chọn “Edit” và tick chọn cột salesman để chuẩn bị cho quá trình merge.



Hình 122: Chọn ra các cột dữ liệu cho bảng Dim\_Salesman

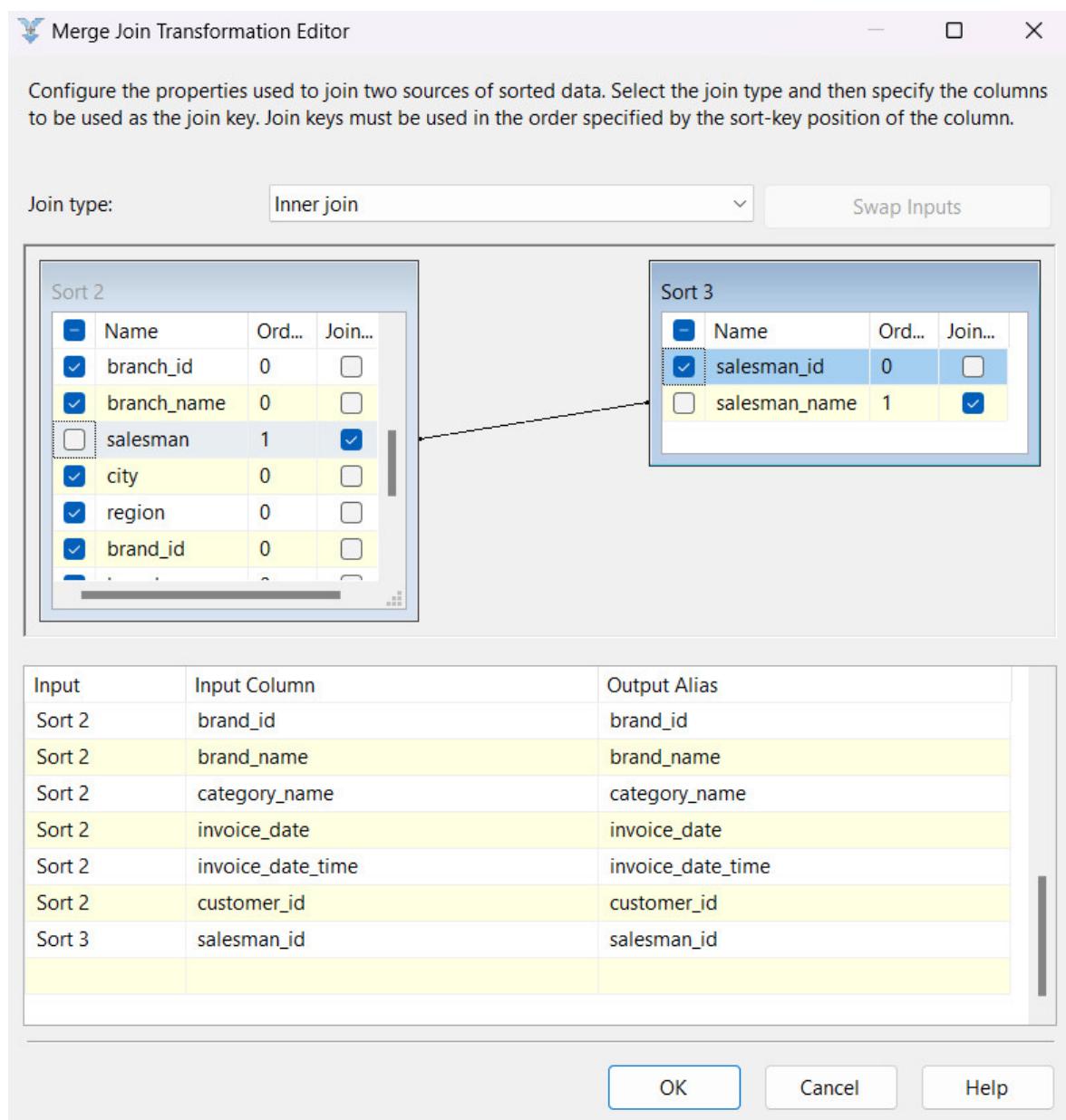
**Bước 14:** Tạo một “Merge Join” và kết nối với Sort2, tiếp theo ta chọn “Merge Join Left Input” để giữ lại toàn bộ các dòng đã merge trước đó bất kể có kết quả khi thực hiện phép kết trái với cột salesman\_id của bảng Dim\_Salesman hay không.



Hình 123: Tạo mới một Merge Join để chuẩn bị kết Fact\_Supermarket và Dim\_Salesman

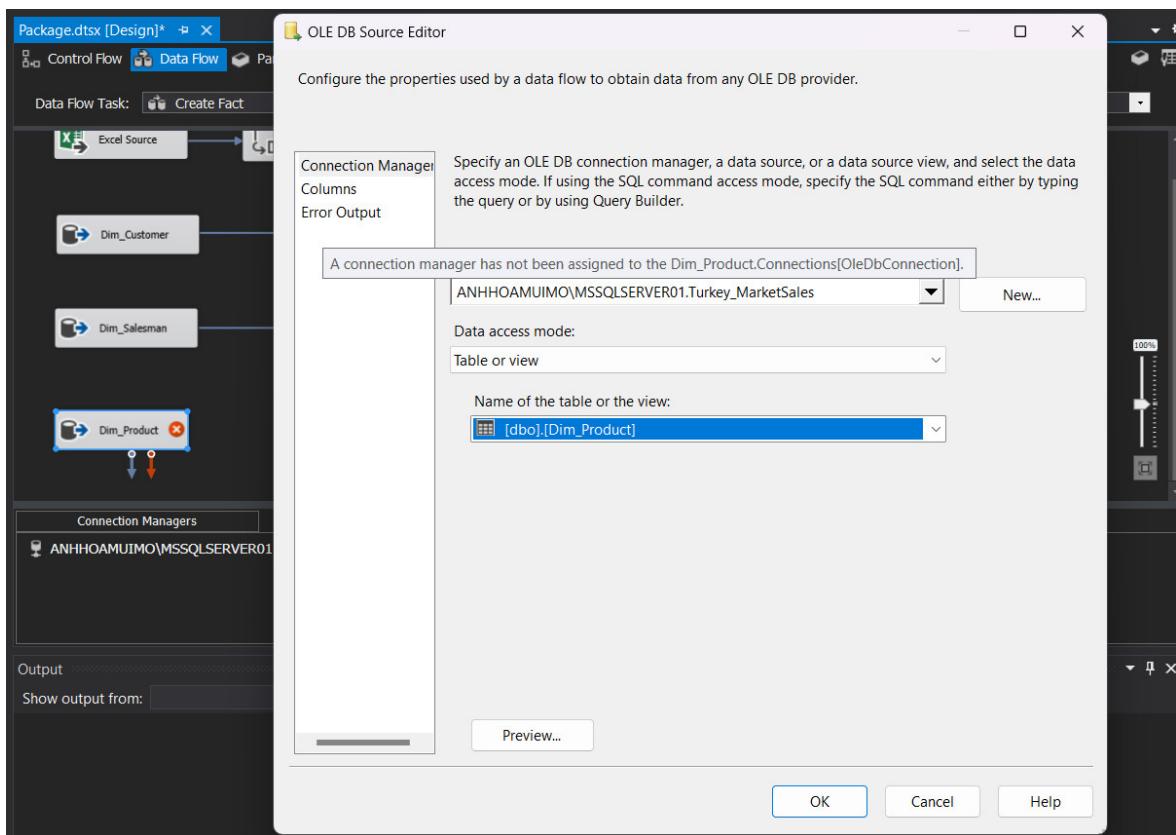
**Bước 15:** Kết nối Sort3 với “Merge Join”. Click chuột phải vào “Merge Join”, chọn “Edit”, một hộp thoại merge editor xuất hiện. Tại vị trí này, ta tick chọn tất cả các cột của Sort2 nhưng không lấy thuộc tính salesman.

Tiếp theo, ta chọn thuộc tính salesman\_id ở Sort3 để tiến hành merge Dim\_Salesman và Fact\_Supermarket. Kết quả sau khi merge là Fact\_Supermarket lúc này không còn thuộc tính salesman.



Hình 124: Thực hiện phép kết trái giữa Fact\_Supermarket và Dim\_Salesman

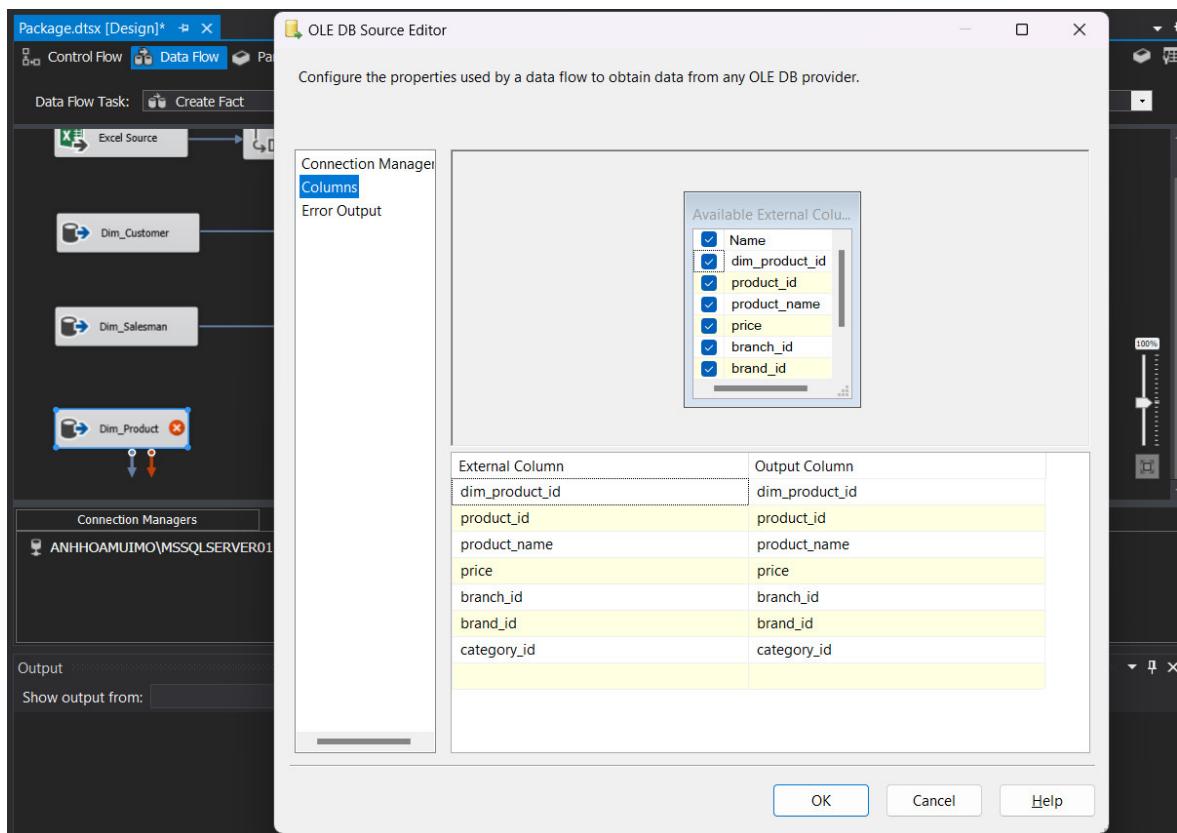
**Bước 16:** Ta tạo thêm “OLE DB Source” và đổi tên thành Dim\_Product. Click chuột phải chọn “Edit”, sau đó chọn bảng Dim\_Product đã được tạo trước đó làm data source cho bảng Dim\_Product mới này.



Hình 125: Tạo mới một OLE DB Source và chọn bảng Dim\_Product làm data source

**Bước 17:** Chọn mục “Columns” để xem xét các cột được ánh xạ.

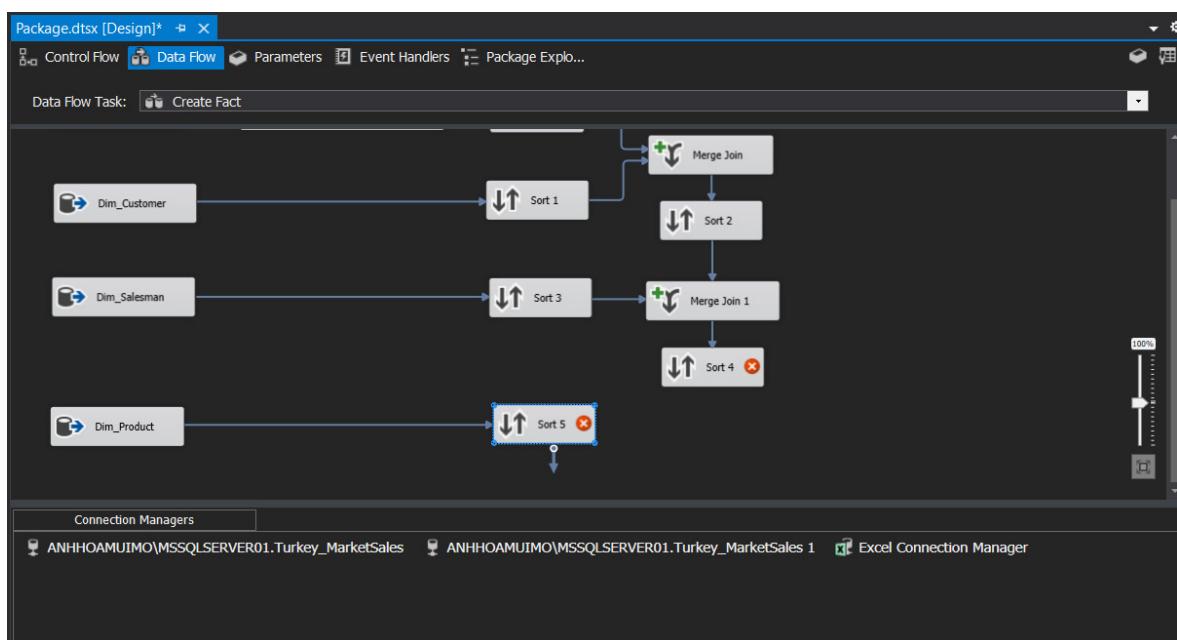
Sau khi kiểm tra đúng và đầy đủ, Click “OK”.



Hình 126: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 6

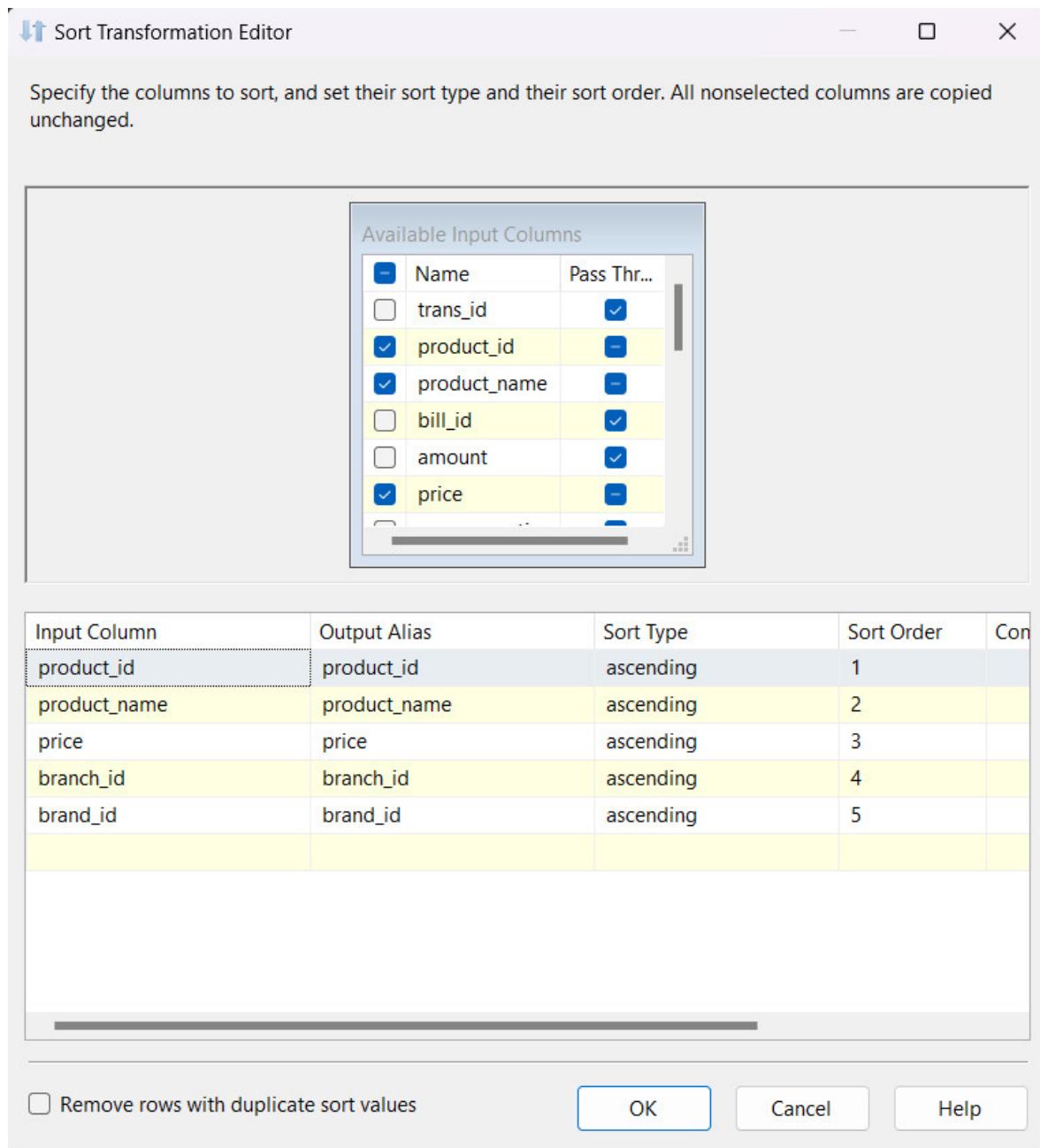
**Bước 18:** Tạo 2 Sort và đặt tên lần lượt là Sort4 và Sort5 tương ứng

cho Fact và Dim\_Product.



Hình 127: Tạo 2 Sort và đặt tên lần lượt là Sort4 và Sort5 tương ứng cho Fact\_Supermarket và Dim\_Product

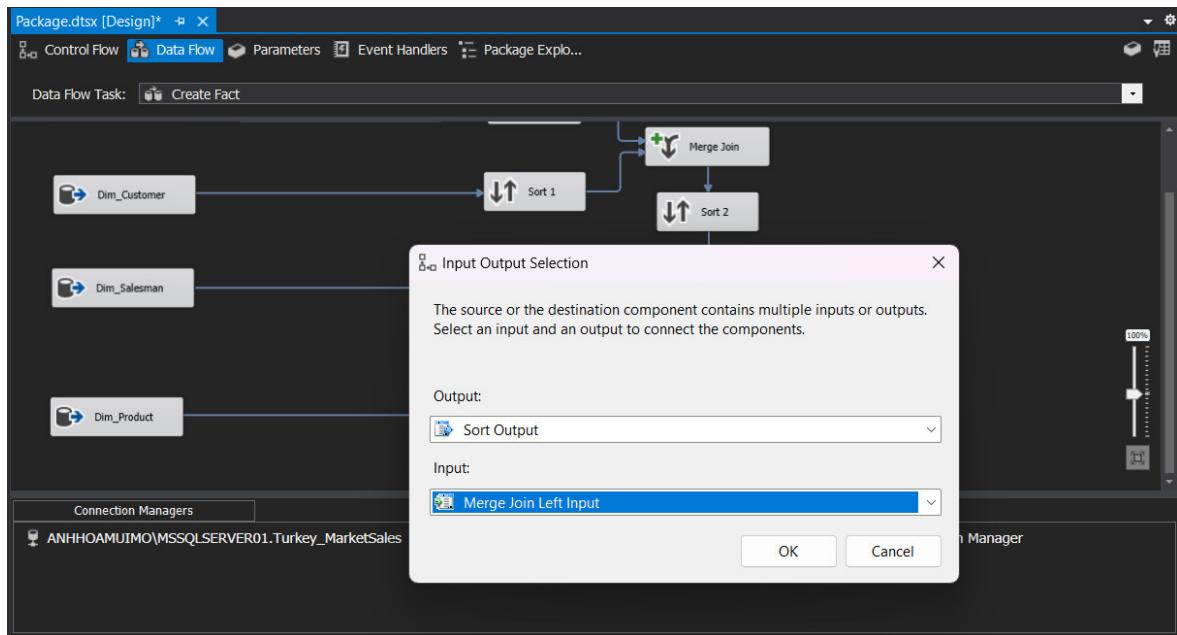
**Bước 19:** Ở Sort4 và Sort5, click chuột phải chọn “Edit” và tick chọn các cột product\_id, product\_name, price, branch\_id và brand\_id để chuẩn bị cho quá trình merge.



Hình 128: Chọn ra các cột dữ liệu cho bảng Dim\_Product

**Bước 20:** Tạo một “Merge Join” và kết nối với Sort4, tiếp theo ta chọn “Merge Join Left Input” để giữ lại toàn bộ các dòng đã merge trước

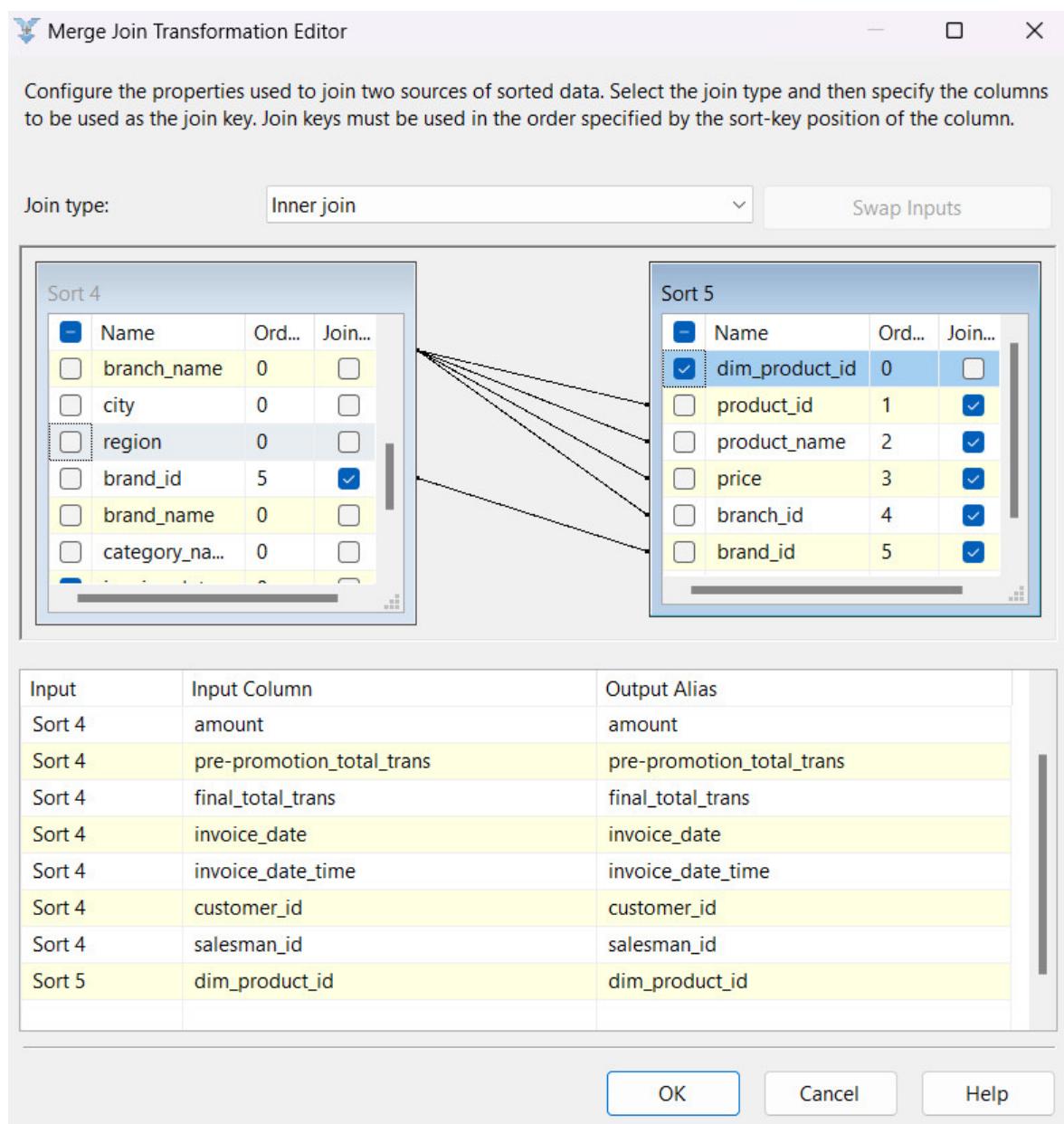
đó bất kể có kết quả khi thực hiện phép kết trái với cột dim\_product\_id của bảng Dim\_Product hay không.



Hình 129:Tạo mới một Merge Join để chuẩn bị kết Fact\_Supermarket và Dim\_Product

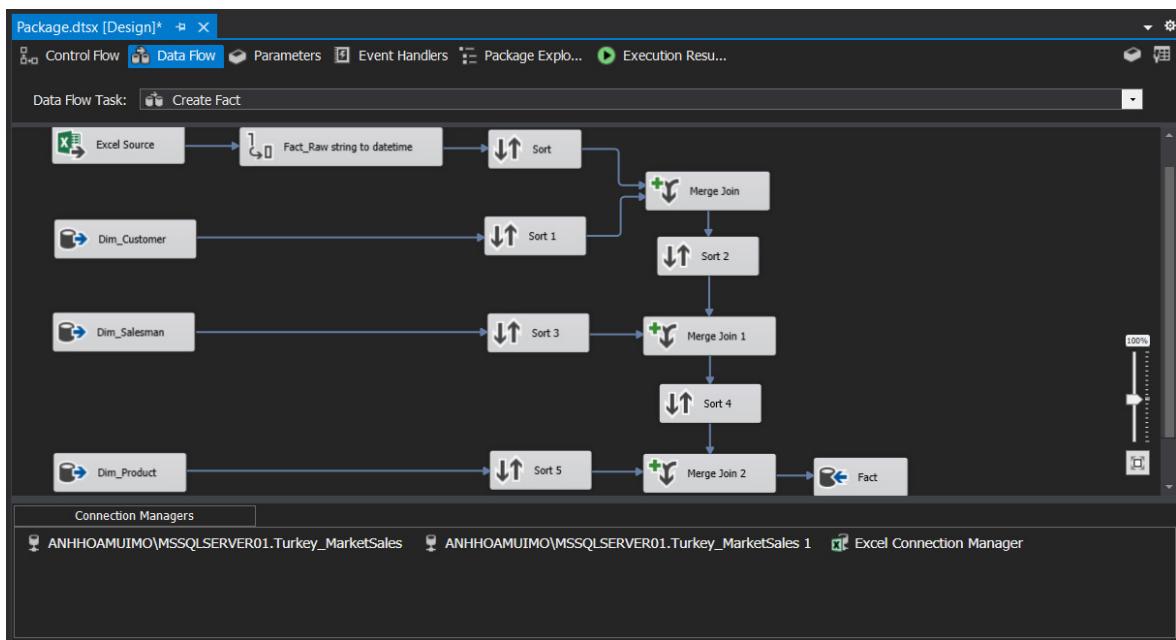
**Bước 21:** Kết nối Sort5 với “Merge Join”. Click chuột phải vào “Merge Join”, chọn “Edit”, một hộp thoại merge editor xuất hiện. Tại vị trí này, ta tick chọn tất cả các cột của Sort4 nhưng không lấy 5 thuộc tính: product\_id, product\_name, price, branch\_id và brand\_id.

Tiếp theo, ta chọn thuộc tính dim\_product\_id ở Sort5 để tiến hành merge Dim\_Customer và Fact\_Supermarket. Kết quả sau khi merge là Fact\_Supermarket lúc này không còn thuộc tính product\_id, product\_name, price, branch\_id và brand\_id.



Hình 130: Thực hiện phép kết trái giữa Fact\_Supermarket và Dim\_Product

**Bước 22:** Tạo bảng Fact từ một “OLE DB Destination” để lưu trữ toàn bộ dữ liệu đã merge trong các bước nêu trên.



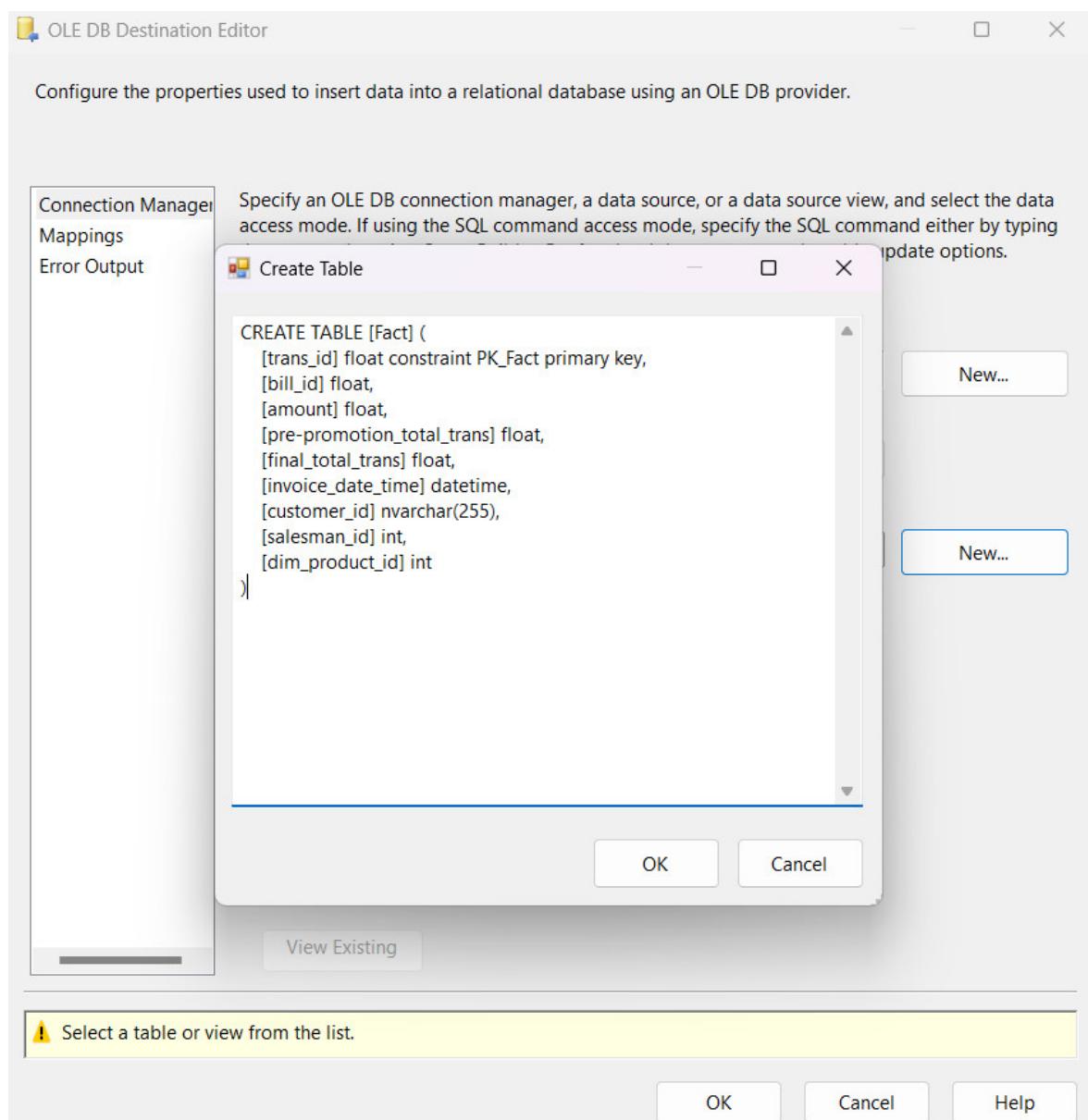
Hình 131: Tạo mới một OLE DB Destination và đặt tên là Fact

Nội dung câu lệnh SQL tạo bảng Fact1 như sau:

```
CREATE TABLE [Fact1] (
```

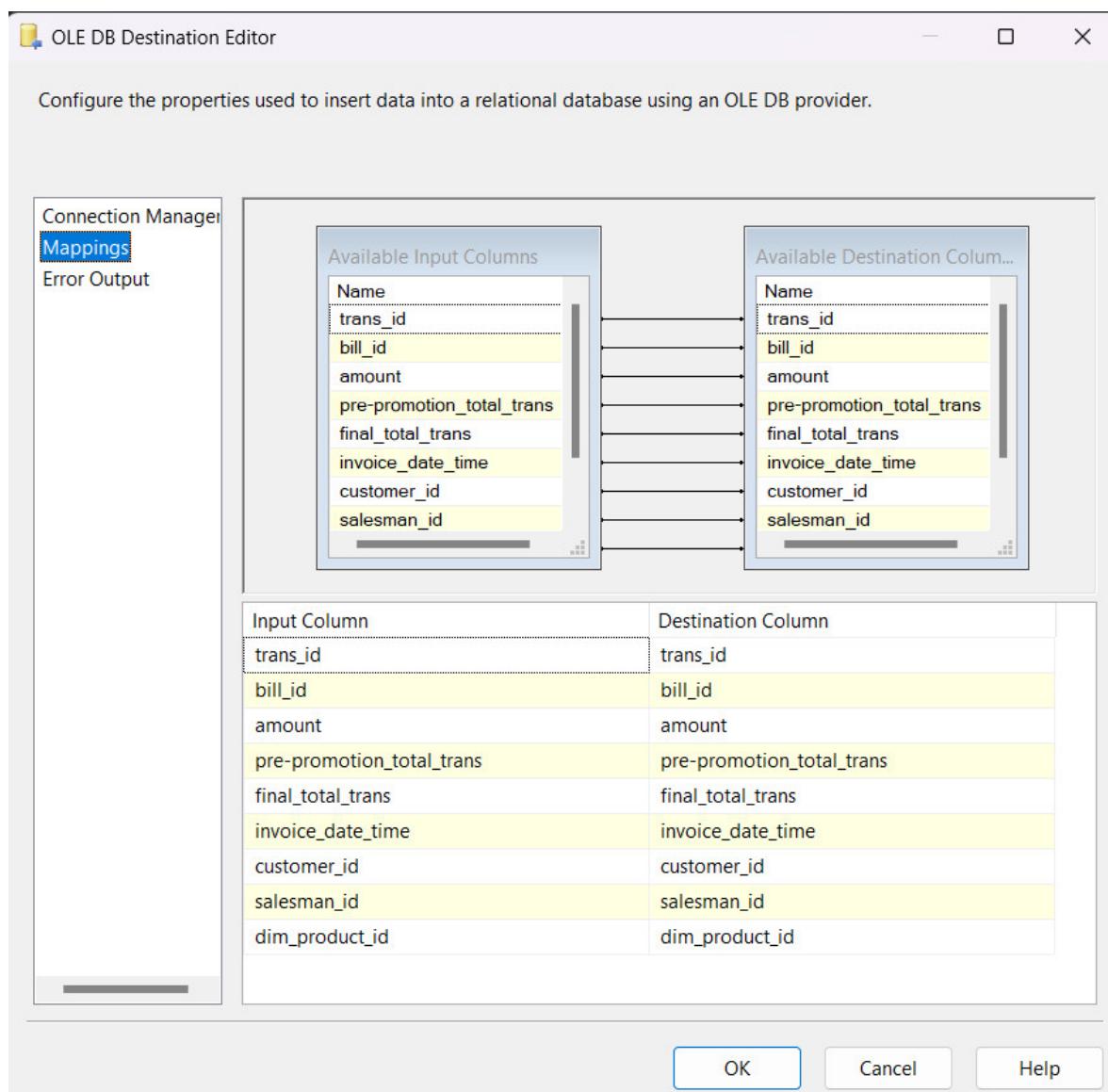
```
    [trans_id] float constraint PK_Fact1 primary key,
    [bill_id] float,
    [amount] float,
    [pre-promotion_total_trans] float,
    [final_total_trans] float,
    [invoice_date_time] datetime,
    [customer_id] nvarchar(255),
    [salesman_id] int,
    [dim_product_id] int,
```

```
)
```



Hình 132: Nhập các câu lệnh SQL tương ứng với bảng Fact\_Supermarket

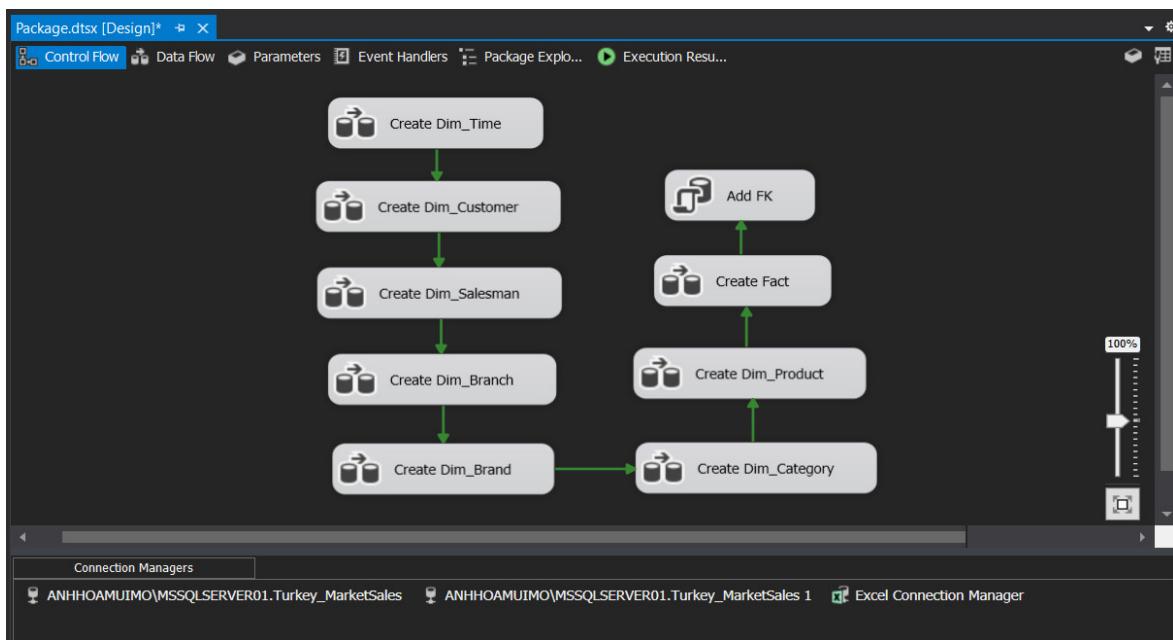
**Bước 23:** Chọn mục “**Mappings**” để xem xét việc ánh xạ các cột dữ liệu.



Hình 133: Ánh xạ các cột dữ liệu vừa tạo và cột dữ liệu ở Hình 6

### 2.3.2. Tạo khóa ngoại từ bảng Fact đến các bảng Dimension:

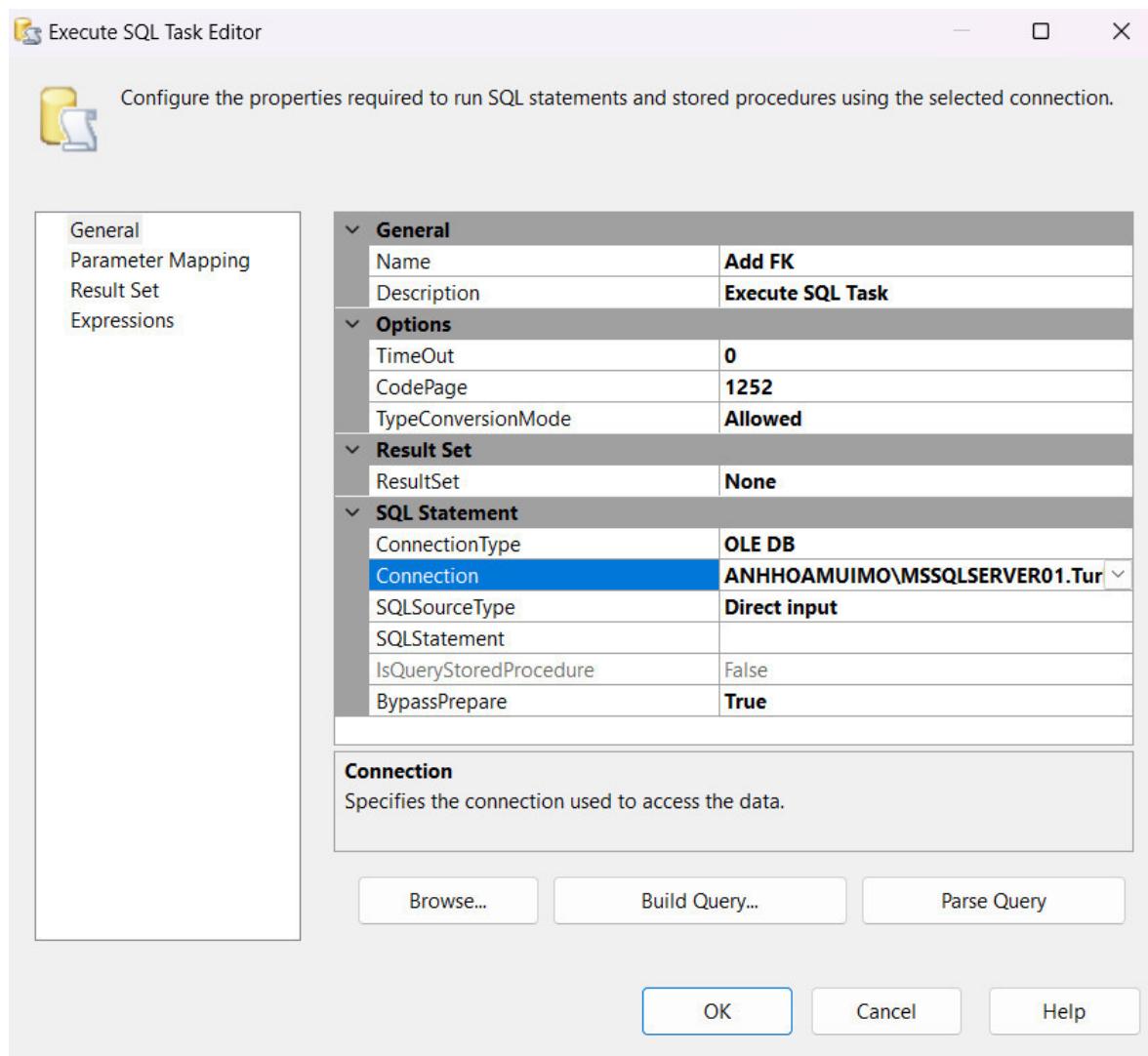
**Bước 1:** Tạo một “Execute SQL Task” và đổi tên thành “Add FK” để thực thi các câu lệnh SQL tạo các khóa ngoại từ các bảng Dimension đến bảng Fact.



Hình 134: Tạo mới Execute SQL Task để tạo các khóa ngoại

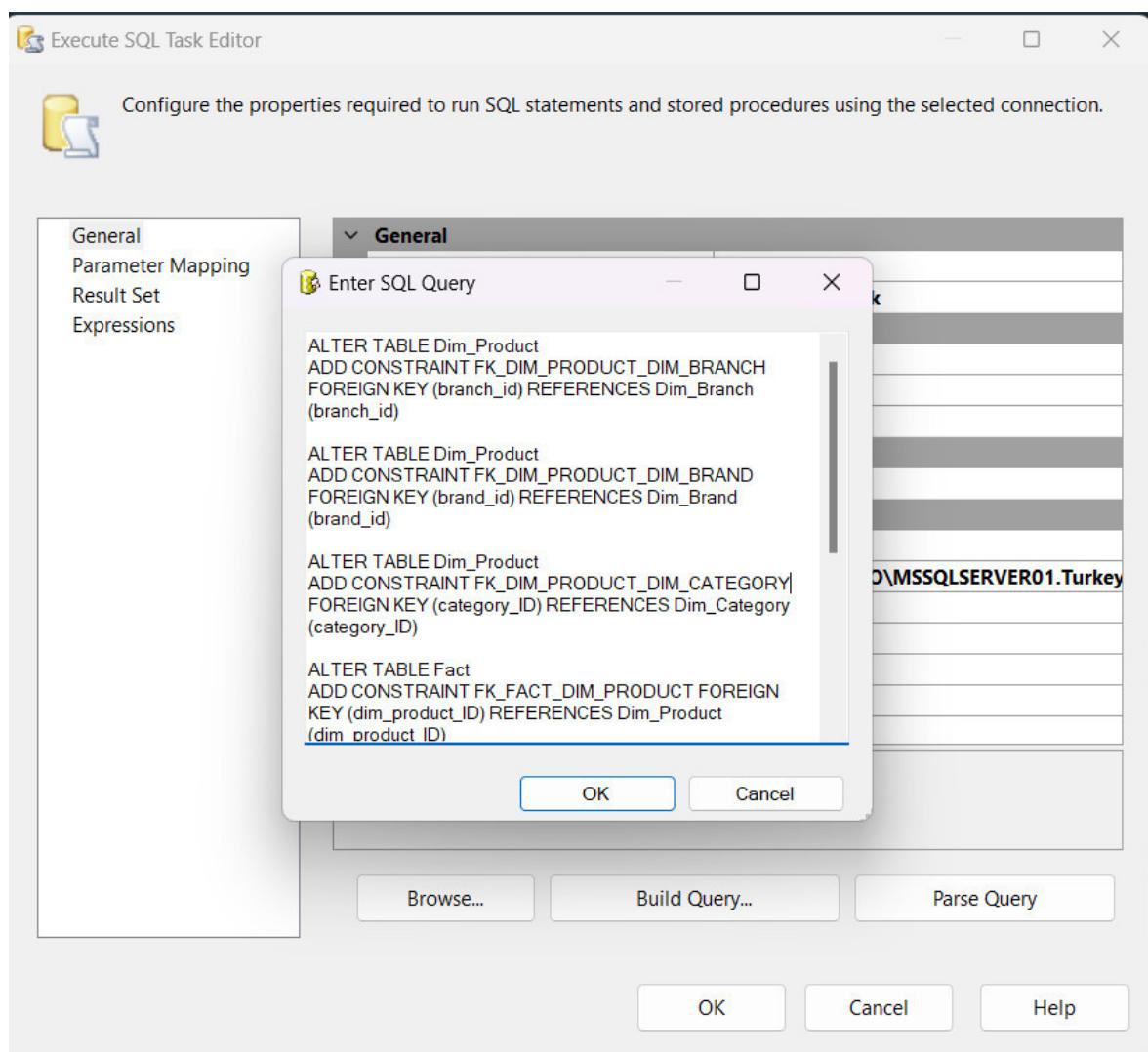
**Bước 2:** Click chuột phải vào Execute SQL Task này và chọn “Edit”.

Tại ô “Connection”, chọn kết nối đã thiết lập đến kho dữ liệu trong SQL Server.



Hình 135: Tạo kết nối đến kho dữ liệu trong SQL Server

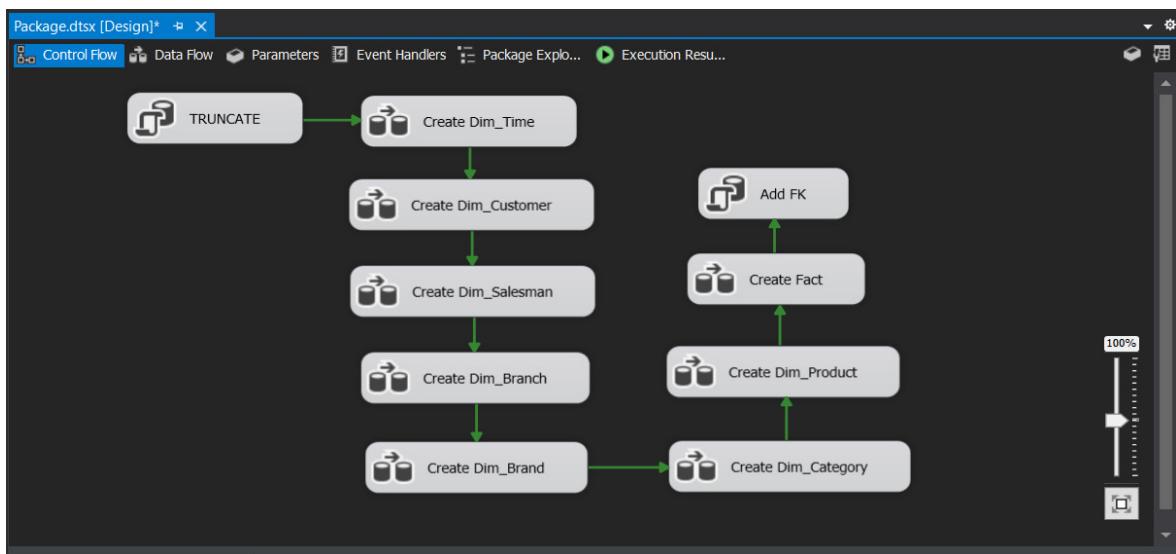
**Bước 3:** Ở ô **SQLStatement**, thêm các câu truy vấn SQL thực hiện tạo các khóa ngoại từ các bảng Dimension đến bảng Fact.



Hình 136: Nhập các câu lệnh tạo khóa ngoại tương ứng

### 2.3.3. Chạy dự án SSIS:

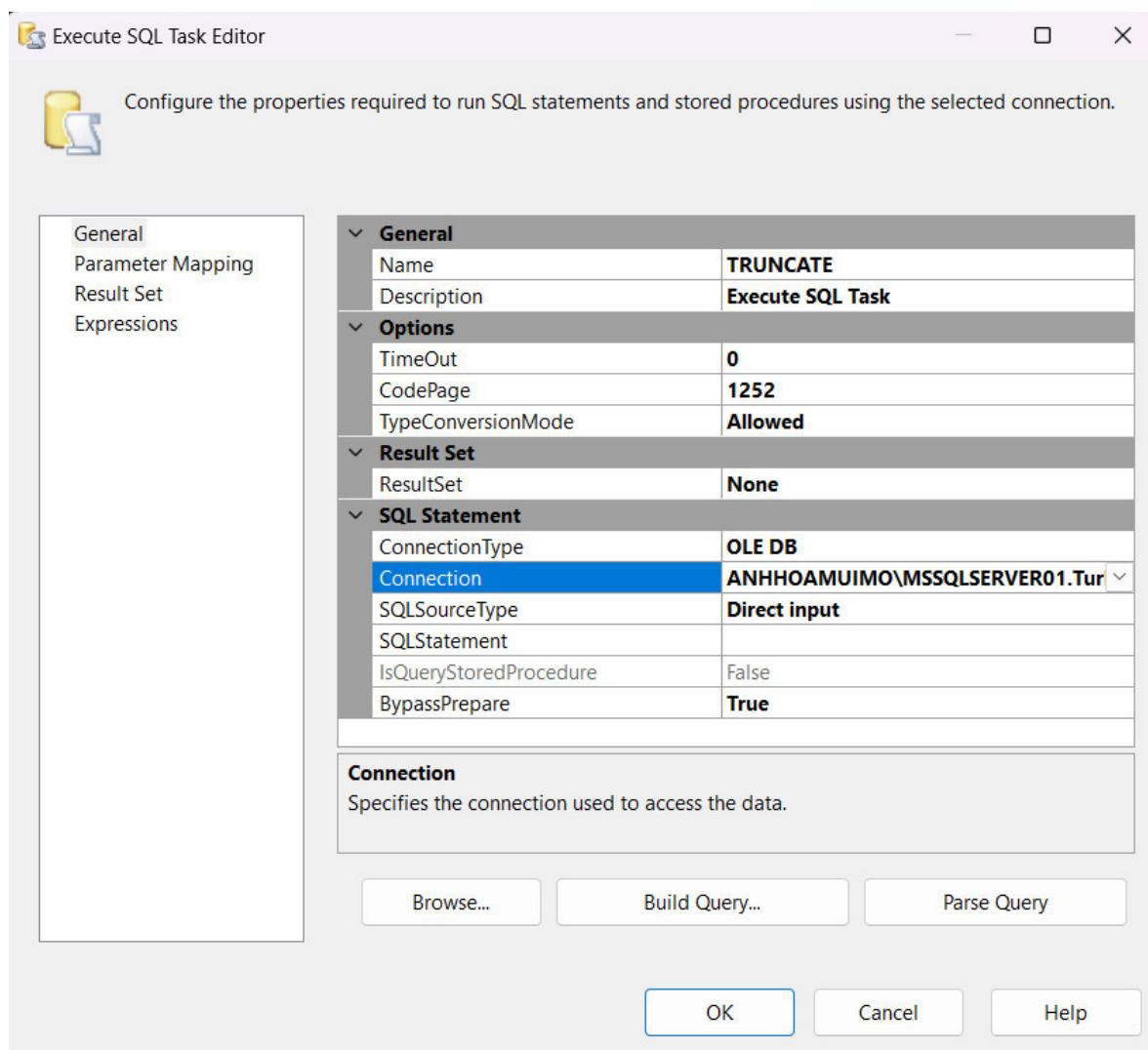
**Bước 1:** Thêm vào một “Execute SQL Task”, đổi tên thành TRUNCATE nhằm thực hiện nhiệm vụ đảm bảo dữ liệu mới hoàn toàn (không bị chồng chéo dữ liệu cũ) mỗi khi chạy project trước quá trình chia bảng Fact và các Dimension.



Hình 137:Tạo mới một Execute SQL Task để đảm bảo không bị chồng chéo dữ liệu trong quá trình đó

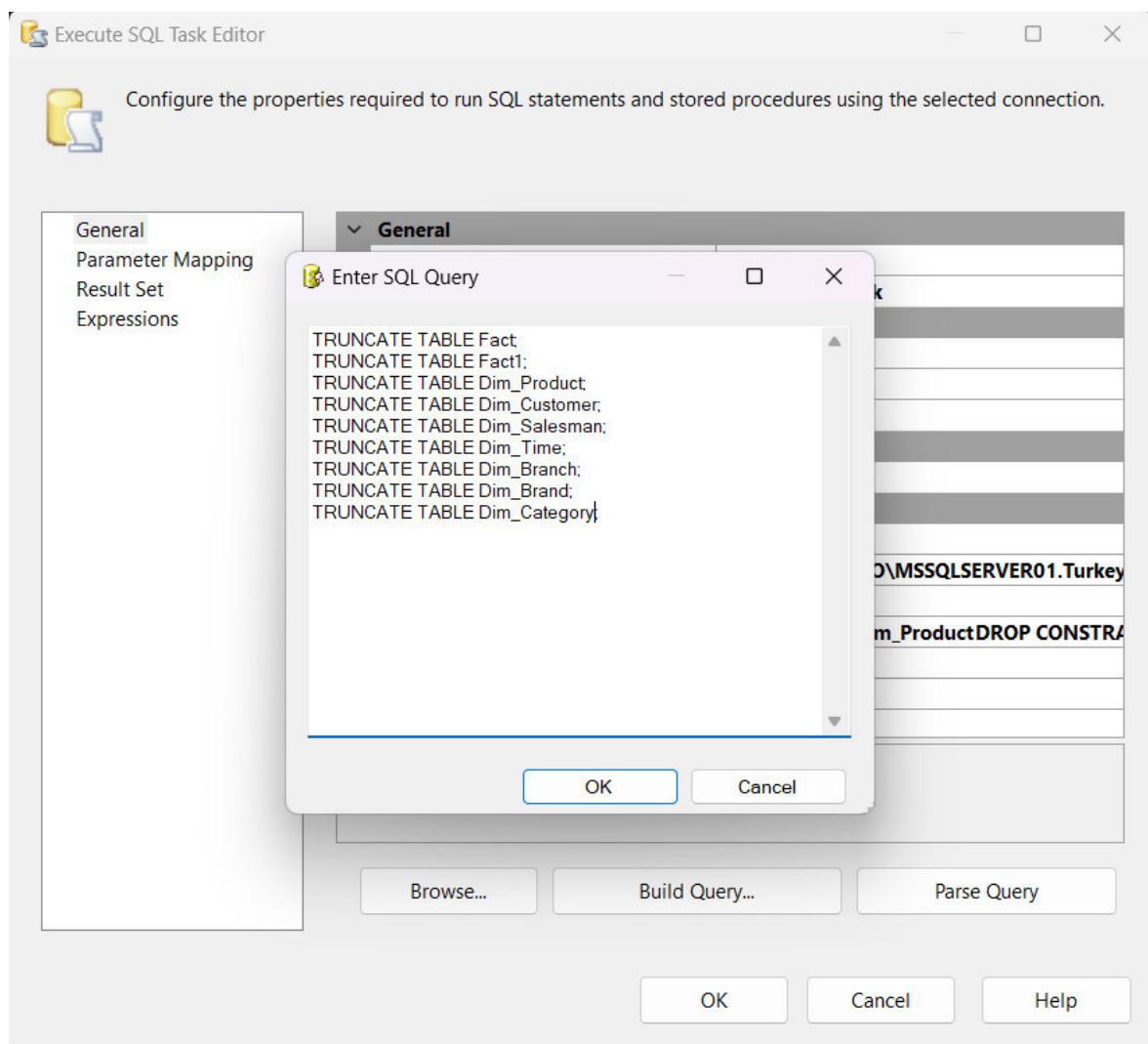
**Bước 2:** Click chuột phải vào “Execute SQL Task” và chọn “Edit”.

Ở ô “Connection”, chọn kết nối đã thiết lập đến kho dữ liệu trong SQL Server.



Hình 138:Tạo kết nối đến kho dữ liệu trong SQL Server

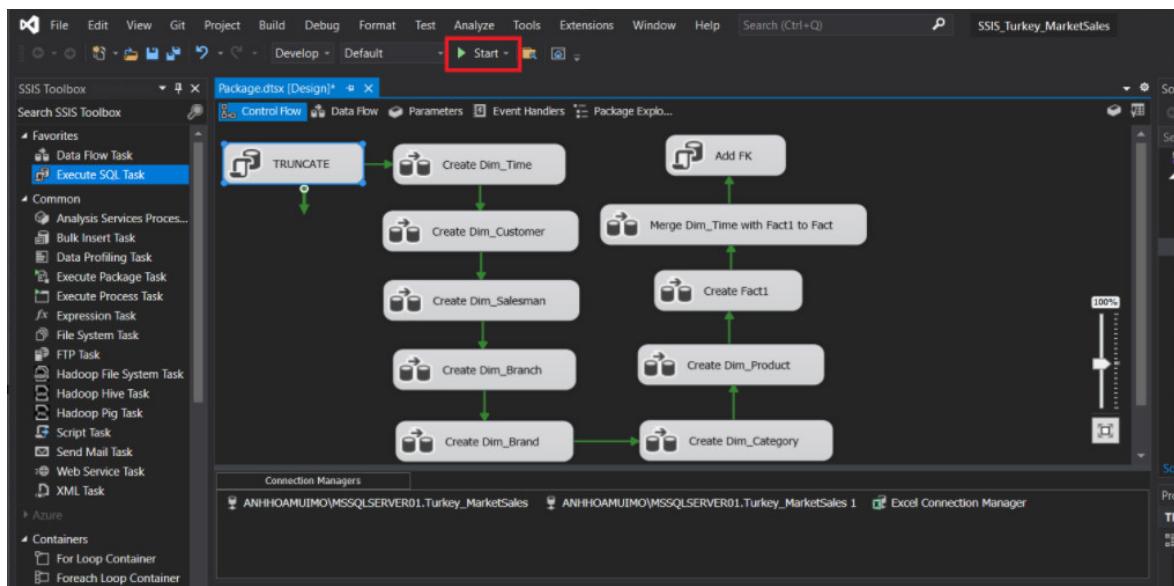
**Bước 3:** Ở ô **SQL Statement**, thêm các câu truy vấn SQL thực hiện xóa dữ liệu cũ trong các bảng mỗi khi chạy project.



Hình 139: Thực hiện xóa dữ liệu cũ mỗi khi tiến hành đổ dữ liệu

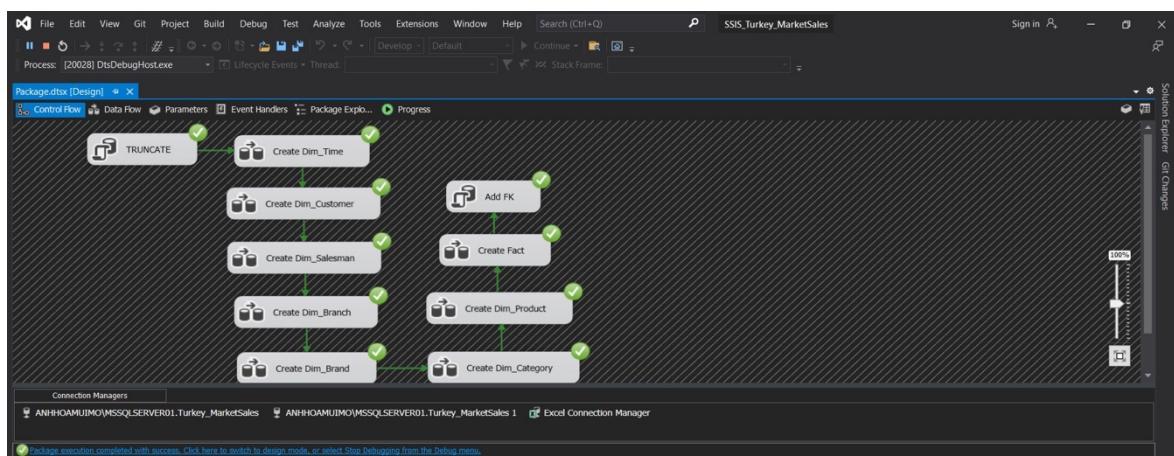
Nhấn “OK” để hoàn tất quá trình.

**Bước 4:** Nhấn nút “Start” trên thanh menu để tiến hành chạy project.



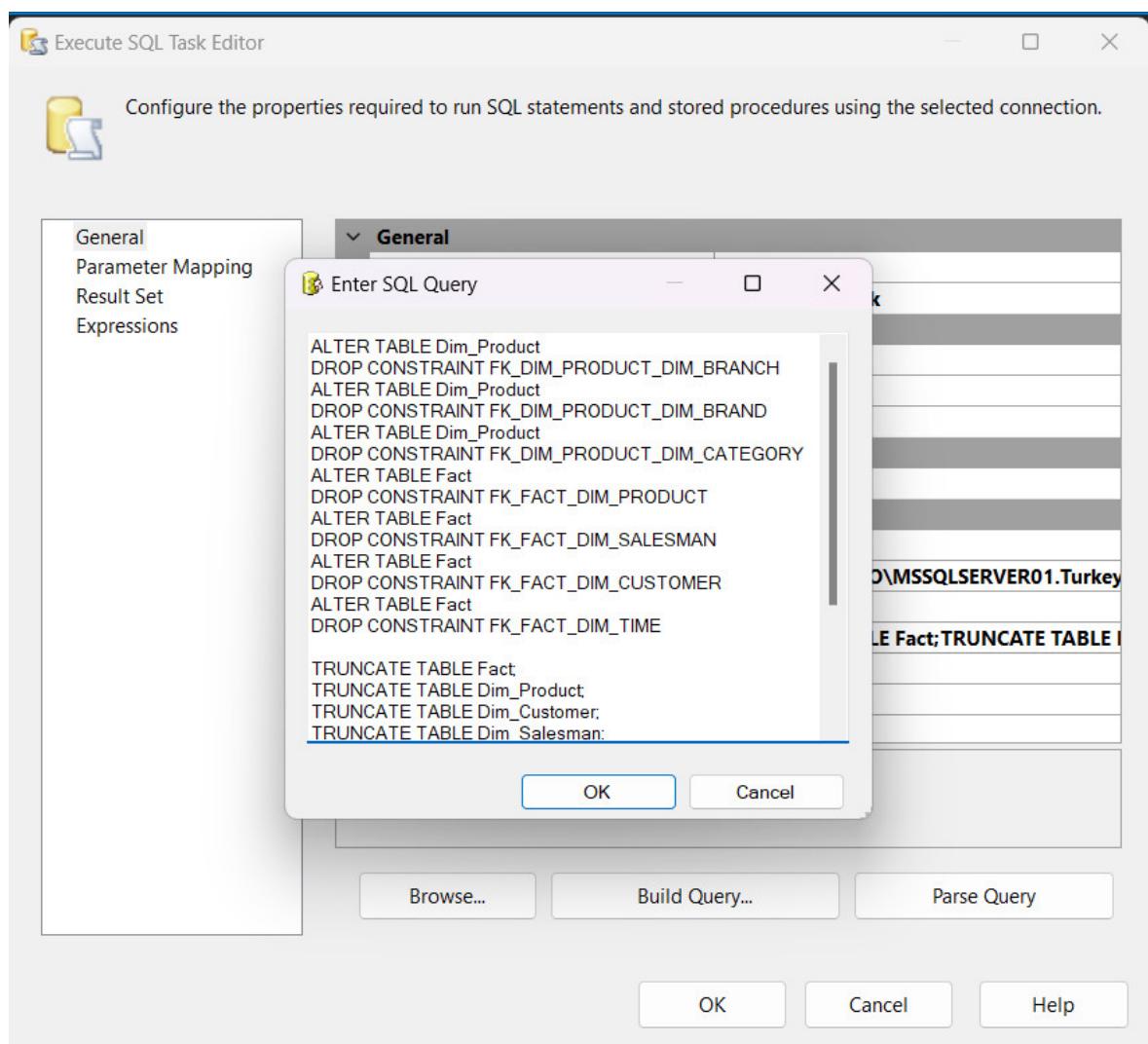
Hình 140: Tiến hành chạy dự án SSIS

Kết quả chạy project:



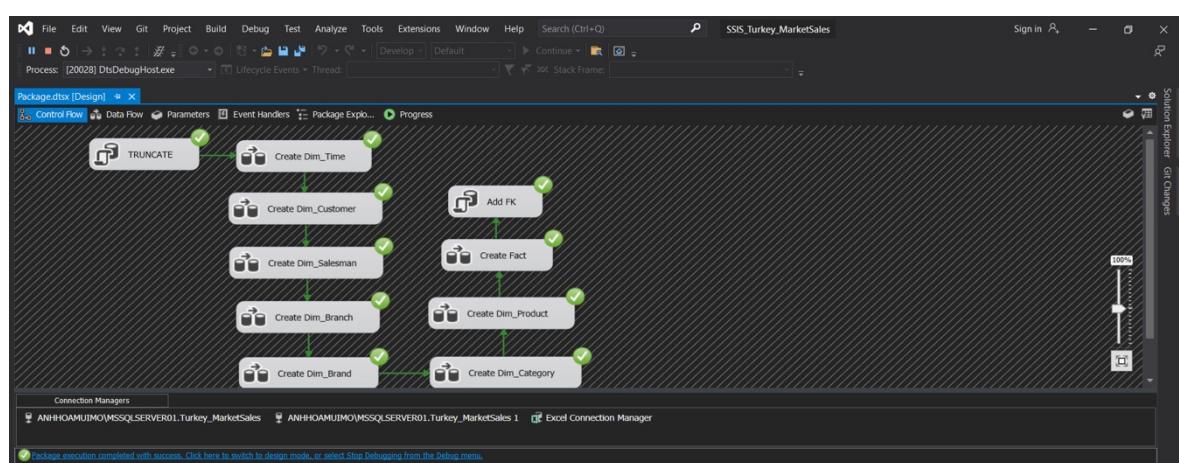
Hình 141: Kết quả chạy dự án SSIS lần đầu tiên

**Bước 5:** Sau khi đã chạy project lần đầu thành công, khóa ngoại của bảng Fact tham chiếu đến các Dimension đã được tạo. Tiến hành thêm các lệnh SQL để xóa và thêm lại các khóa ngoại vào mỗi lần khởi chạy lại project kế tiếp trong tương lai.



Hình 142: Bổ sung các câu lệnh xóa và thêm các khóa ngoại mới khi đổ dữ liệu

**Bước 6:** Tiến hành chạy lại project một lần nữa.



Hình 143: Kết quả chạy lại dự án SSIS

## 2.4. Kiểm tra dữ liệu các bảng:

### 2.4.1. Bảng Dim\_Time:

	invoice_date_time	invoice_date_day	invoice_date_month	invoice_date_year	invoice_date_hour	invoice_date_minute	invoice_date_second	invoice_date_dayofweek
1	2017-01-03 09:25:03.000	3	1	2017	9	25	3	2
2	2017-01-03 09:31:16.000	3	1	2017	9	31	16	2
3	2017-01-03 09:35:35.000	3	1	2017	9	35	35	2
4	2017-01-03 09:36:17.000	3	1	2017	9	36	17	2
5	2017-01-03 09:37:03.000	3	1	2017	9	37	3	2
6	2017-01-03 09:42:18.000	3	1	2017	9	42	18	2
7	2017-01-03 09:44:11.000	3	1	2017	9	44	11	2
8	2017-01-03 09:52:07.000	3	1	2017	9	52	7	2
9	2017-01-03 09:56:48.000	3	1	2017	9	56	48	2
10	2017-01-03 09:57:36.000	3	1	2017	9	57	36	2
11	2017-01-03 09:59:32.000	3	1	2017	9	59	32	2
12	2017-01-03 10:06:26.000	3	1	2017	10	6	26	2
13	2017-01-03 10:07:01.000	3	1	2017	10	7	1	2
14	2017-01-03 10:12:44.000	3	1	2017	10	12	44	2
15	2017-01-03 10:31:34.000	3	1	2017	10	31	34	2
16	2017-01-03 10:35:04.000	3	1	2017	10	35	4	2
17	2017-01-03 10:45:12.000	3	1	2017	10	45	12	2
18	2017-01-03 10:47:52.000	3	1	2017	10	47	52	2
19	2017-01-03 11:03:06.000	3	1	2017	11	3	6	2
20	2017-01-03 11:09:41.000	3	1	2017	11	9	41	2
21	2017-01-03 11:15:21.000	3	1	2017	11	15	21	2
22	2017-01-03 11:17:29.000	3	1	2017	11	17	29	2
23	2017-01-03 11:19:53.000	3	1	2017	11	19	53	2
24	2017-01-03 11:23:57.000	3	1	2017	11	23	57	2
25	2017-01-03 11:27:52.000	3	1	2017	11	27	52	2
26	2017-01-03 11:37:08.000	3	1	2017	11	37	8	2
27	2017-01-03 11:38:36.000	3	1	2017	11	38	36	2

Hình 144: Dữ liệu bảng Dim\_Time

### 2.4.2. Bảng Dim\_Customer:

	customer_id	customer_name	GENDER
1	1	Volkan ÇEKİP	Male
2	10	Nazlıcan ÖZSİMİTÇİ	Female
3	100	Özkan DERİLIOĞLU	Male
4	1000	Veysel ELDEM	Male
5	1000078	Dursun KENTSEVER	Male
6	1000117	Sudenur CAPAR	Female
7	1000120	Ertuğrul BEKÇE	Male
8	1000146	Yüksel BAŞBAY	Female
9	1000149	Aziz KARAÇAM	Male
10	1000228	Ekin KÜÇÜKİRİAVUL	Female
11	1000244	Gülhan ASLANHAN	Female
12	1000251	Aydın KULO	Male
13	1000288	Zilan CİVAR	Female
14	100029	Gizem TUNÇBİLEK	Female
15	1000354	Sudenaz KILINBOZ	Female

Hình 145: Dữ liệu bảng Dim\_Customer

### 2.4.3. Bảng Dim\_Salesman:

	salesman_id	salesman_name
1	1	Abdullah BENGIN
2	2	Abdulsamet TACIM
3	3	Abdurrahman CAVDARLI
4	4	Adil ONDIN
5	5	Adnan YAGIZ
6	6	Ahmet ELKATMIS
7	7	Aleyna CUKUREL
8	8	Aleyna ERCKETIN
9	9	Aleyna KASALAK
10	10	Ali Eymen GOKCESU
11	11	Ali ILTEKIN
12	12	Ali OZKOL
13	13	Aliye CUKURCAYIR
14	14	Aliye GOKDOGAN
15	15	Aliye KOZEN
16	16	Alper FETTAHOGLU
17	17	Alper INELBARS
18	18	Alper IZMIRLI
19	19	Alya BARAN
20	20	Arda CINCIN
21	21	Arife DORAP
22	22	Arin OZYUREKOGLU
23	23	Arzu GOKKUS
24	24	Arzu KARAKITAPOGLU
25	25	Arzu TUGFAN
26	26	Asel CELTIK
27	27	Asiye BUGDAYLI

Hình 146: Dữ liệu bảng Dim\_Salesman

#### 2.4.4. Bảng Dim\_Branch:

	branch_id	branch_name	city	region
1	1	Adana Subesi	Adana	Akdeniz
2	2	Adiyaman Subesi	Adiyaman	Güneydoğu Anadolu
3	3	Afyonkarahisar Subesi	Afyonkarahisar	Ege
4	4	Ağrı Subesi	Ağrı	Doğu Anadolu
5	5	Aksaray Subesi	Aksaray	İç Anadolu
6	6	Amasya Subesi	Amasya	Karadeniz
7	7	Ankara Subesi	Ankara	İç Anadolu
8	8	Antalya Subesi	Antalya	Akdeniz
9	9	Ardahan Subesi	Ardahan	Doğu Anadolu
10	10	Artvin Subesi	Artvin	Karadeniz
11	11	Aydın Subesi	Aydın	Ege
12	12	Balıkesir Subesi	Balıkesir	Ege
13	13	Bartın Subesi	Bartın	Karadeniz
14	14	Batman Subesi	Batman	Güneydoğu Anadolu
15	15	Bayburt Subesi	Bayburt	Karadeniz
16	16	Bilecik Subesi	Bilecik	Marmara
17	17	Bingöl Subesi	Bingöl	Doğu Anadolu
18	18	Bitlis Subesi	Bitlis	Doğu Anadolu
19	19	Bolu Subesi	Bolu	Karadeniz
20	20	Burdur Subesi	Burdur	Akdeniz
21	21	Bursa Subesi	Bursa	Marmara
22	22	Çanakkale Subesi	Çanakkale	Marmara
23	23	Çankırı Subesi	Çankırı	İç Anadolu
24	24	Çorum Subesi	Çorum	Karadeniz
25	25	Denizli Subesi	Denizli	Ege
26	26	Diyarbakır Subesi	Diyarbakır	Güneydoğu Anadolu
27	27	Düzce Subesi	Düzce	Karadeniz

Hình 147: Dữ liệu bảng Dim\_Branch

#### 2.4.5. Bảng Dim\_Brand:

	brand_id	brand_name
1	1	Air Wick
2	101	Nutella
3	102	Olca
4	103	Olin
5	106	Oskar
6	107	Pioneer
7	108	Öztemizel
8	109	P.Bahçe
9	11	Baghdad
10	111	Pakmaya
11	112	Palmolive
12	113	Pastavilla
13	114	Penguin
14	115	Pepsi
15	116	Peyman
16	117	Philips
17	118	Pinar
18	12	F BAKTAT
19	120	Prenses
20	121	Prima
21	122	Pringles
22	123	Raid
23	124	Reis
24	126	Saray
25	127	Sarelle
26	128	Saruhan
27	13	Balkovan

Hình 148: Dữ liệu bảng Dim\_Brand

#### 2.4.6. Bảng Dim\_Category:

	category_id	category_name
1	1	Baby
2	2	Beverage
3	3	Cosmetics
4	4	Dairy Breakfast
5	5	Detergent Cleaning
6	6	Food
7	7	Home
8	8	Meat and Poultry

Hình 149: Dữ liệu bảng Dim\_Category

#### 2.4.7. Bảng Dim\_Product:

	dim_product_id	product_id	product_name	price	branch_id	brand_id	category_id
1	1	7491	PRIMA ACT BABY ECONOMY 11-25 KG	19,95	47	121	1
2	2	1443	BIZIM RICE FLOUR 250 GR	1,95	21	146	1
3	3	23039	HERO BABY MILK 8-GRAIN 500 GR	11,5	41	146	1
4	4	23031	HERO FRUIT 8-GRAIN MILK 250GR	6,95	41	146	1
5	5	1247	HERO GOGLY RICE FLOUR PLAIN	2,99	19	146	1
6	6	23108	HERO PEACH BANANA 125 GR	1,5	19	146	1
7	7	8864	APTAMIL WITH MILK BISCUIT 2*250 GR ECONOMIC	14,9	81	238	1
8	8	8869	APTAMIL SEMILA WITH HONEY 250 GR	8,9	41	238	1
9	9	23393	WEE SILICONE TRANSPARENT PATTERNED PACIFIER WITH ...	4,95	17	270	1
10	10	23390	WEE SILICONE PATTERNED PACIFIER WITH PALATE NO:3	4	62	270	1
11	11	3708	DALIN KIDS SAC BOCUUT SAMP.300ML TROPICAL	6,95	73	318	1
12	12	21087	ETI CICI BABY 190G BANANA BAGS	1,8	23	44	1
13	13	7488	PRIMA JUMBO MAXIPLUS 4+	25,5	21	121	1
14	14	22705	ULKER BABY BISCUIT 190GR	1,85	41	146	1
15	15	23071	HERO PEAR-PINEAPPLE-PEACH	1,5	20	146	1
16	16	1250	HERO APPLE BANANA 125GR	1,25	21	146	1
17	17	22679	ULKER BABY BISCUIT 1KG -138-	6,95	4	146	1
18	18	22721	ULKER BABY 190GR. BANANA	1,85	67	146	1
19	19	8858	APTAMIL PEARL PORT. 250GR	8,9	5	238	1
20	20	9829	CANBEBE TWIN EXTRA LARGE 28PCS	16,9	52	25	1
21	21	23377	WEE SILICONE BABY BOTTLE PIECE NO:2	2,95	77	270	1
22	22	5395	UNI EAR TUBE 100 PCS	1,4	64	275	1
23	23	3715	DALIN FABRIC SOFTENER 2LT.	9,95	41	318	1
24	24	3889	ETI CICI BABY BISCUIT 190 GR.VIT.	1,8	49	44	1
25	25	7486	PRIMA ACT BABY MEGA PACK MIDI	26,95	39	121	1
26	26	1250	HERO APPLE BANANA 125GR	1,25	7	146	1
27	27	12413	HERO GOGLY RICE FLOUR APPLE	2,99	53	146	1

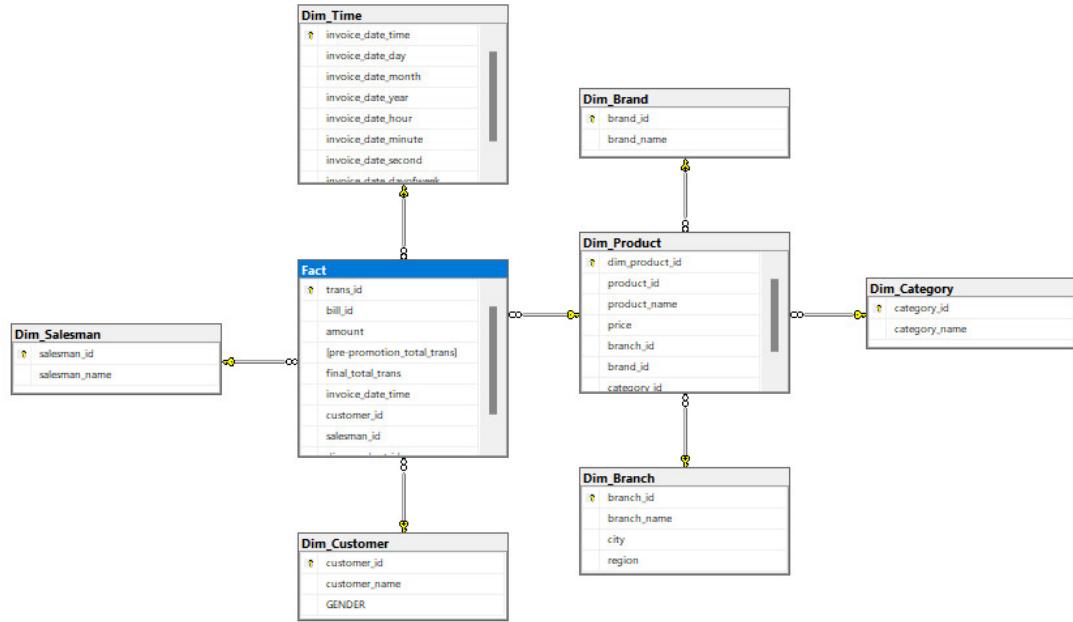
Hình 150: Dữ liệu bảng Dim\_Product

#### 2.4.8. Bảng Fact:

	trans_id	bill_id	amount	pre-promotion_total_trans	final_total_trans	invoice_date_time	customer_id	salesman_id	dim_product_id
1	2	15560	1,006	2,82	2,79	2017-01-03 09:25:03.000	6476	363	73433
2	3	15560	1	4,95	4,19	2017-01-03 09:25:03.000	6476	363	55290
3	4	15560	1	1,1	1,02	2017-01-03 09:25:03.000	6476	363	117326
4	5	15560	1	1,1	1,02	2017-01-03 09:25:03.000	6476	363	116627
5	6	15560	1	2,45	2,08	2017-01-03 09:25:03.000	6476	363	57929
6	7	15560	1	4,9	4,54	2017-01-03 09:25:03.000	6476	363	44476
7	9	15561	1,006	2,82	2,79	2017-01-03 09:25:03.000	456620	444	89650
8	10	15561	1	4,95	4,19	2017-01-03 09:25:03.000	456620	444	55590
9	11	15561	1	1,1	1,02	2017-01-03 09:25:03.000	456620	444	113521
10	12	15561	1	1,1	1,02	2017-01-03 09:25:03.000	456620	444	118164
11	13	15561	1	2,45	2,08	2017-01-03 09:25:03.000	456620	444	55446
12	14	15561	1	4,9	4,54	2017-01-03 09:25:03.000	456620	444	49697
13	16	15562	4	1,4	1,3	2017-01-03 09:31:16.000	275882	683	7519
14	21	15563	4	1,4	1,3	2017-01-03 09:35:35.000	901974	335	11141
15	22	15564	1	0,5	0,47	2017-01-03 09:36:17.000	758893	168	4692
16	23	15564	1	0,5	0,46	2017-01-03 09:36:17.000	758893	168	4692
17	24	15565	4	29,8	27,59	2017-01-03 09:37:03.000	212970	187	2753
18	27	15567	1	1,95	1,8	2017-01-03 09:42:18.000	151181	854	11923
19	28	15567	1	2,99	2,77	2017-01-03 09:42:18.000	151181	854	82846
20	29	15568	1	4,4	4,08	2017-01-03 09:44:11.000	282268	174	84965
21	30	15568	1	4,4	4,07	2017-01-03 09:44:11.000	282268	174	84965
22	34	15571	2	1	0,92	2017-01-03 09:52:07.000	775645	1	4374
23	36	15571	1	2	1,85	2017-01-03 09:52:07.000	775645	1	79578
24	37	15572	2	8,8	8,15	2017-01-03 09:56:48.000	17506	840	73454
25	38	15572	2	8,8	8,15	2017-01-03 09:56:48.000	17506	840	73454
26	40	15573	1	2,45	2,27	2017-01-03 09:57:36.000	216444	188	9543
27	42	15573	2	2,6	2,41	2017-01-03 09:57:36.000	216444	188	19205

Hình 151: Dữ liệu bảng Fact\_Supermarket

#### 2.5. Lược đồ sau khi hoàn thành:



Hình 152: Lược đồ sau khi đổ dữ liệu thành công

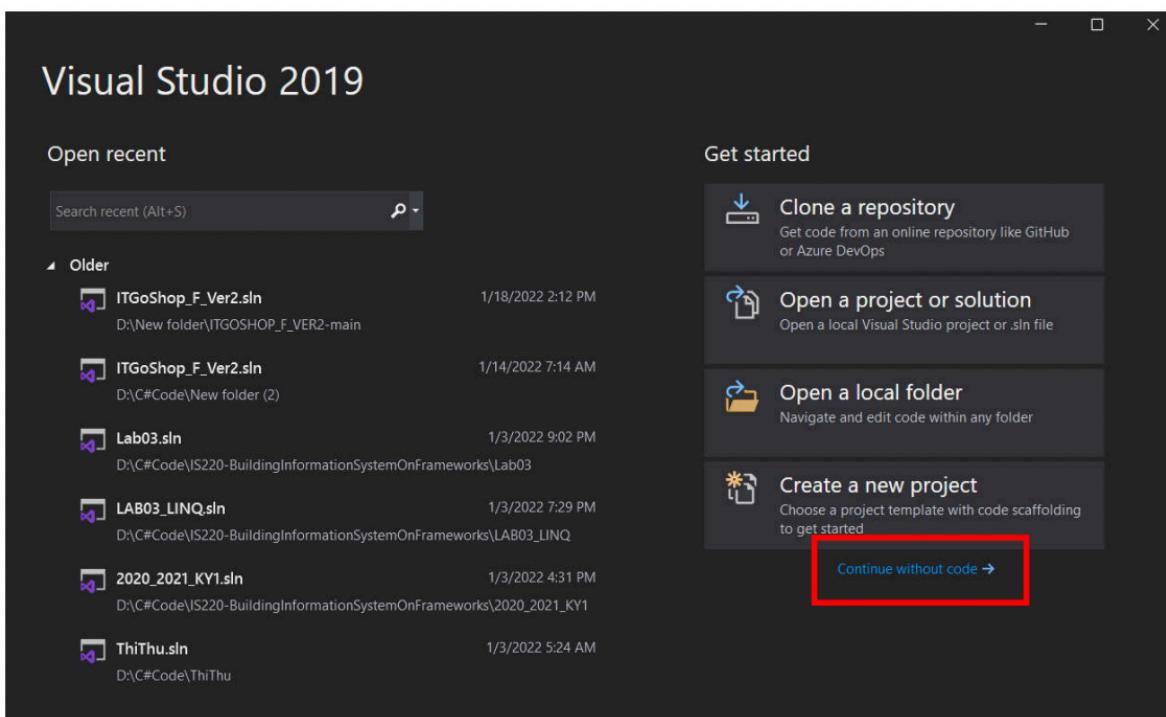
### 3. Phân tích dữ liệu trong kho dữ liệu (SSAS):

#### 3.1. Chuẩn bị công cụ hỗ trợ:

Để hiện thực quá trình SSAS, ta cần cài đặt một vài công cụ hỗ trợ từ Microsoft bao gồm: Microsoft SQL Server tích hợp Analysis Services và Microsoft Analysis Service Projects.

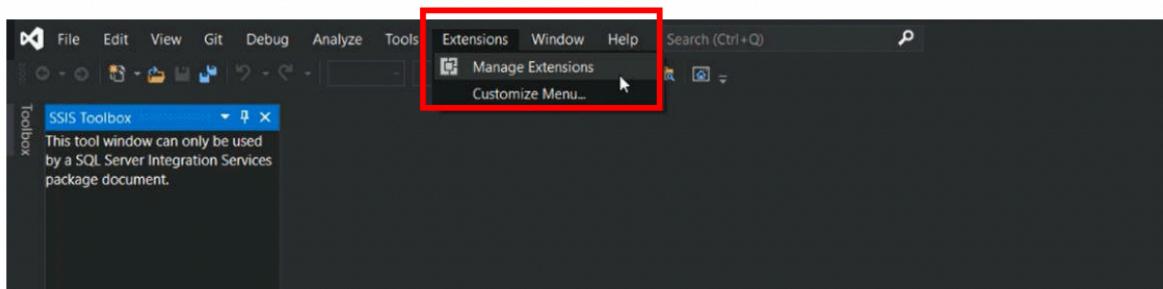
##### 3.1.1. Cài đặt Microsoft Analysis Services Projects:

Bước 1: Mở Visual Studio 2019 và chọn “Continue without code”.



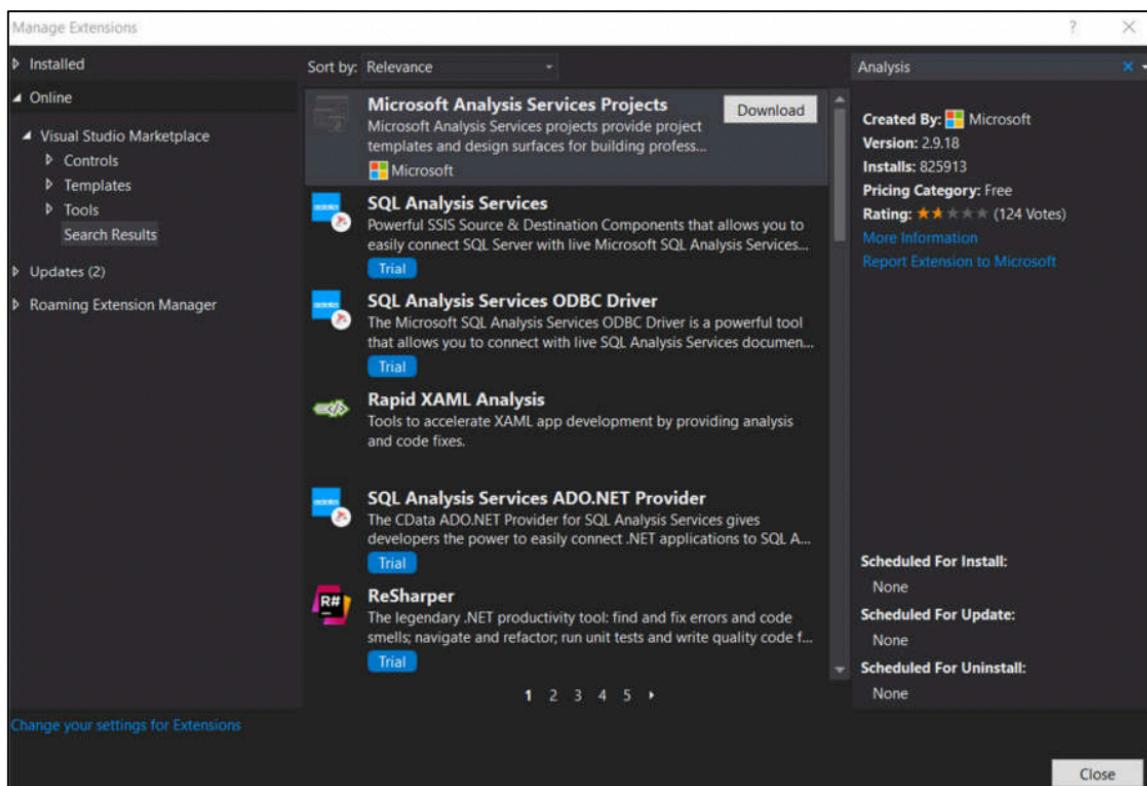
Hình 153: Giao diện khởi động một project rỗng

**Bước 2:** Trong giao diện chính, tìm tab “Extension”, chọn “Manage Extensions”.



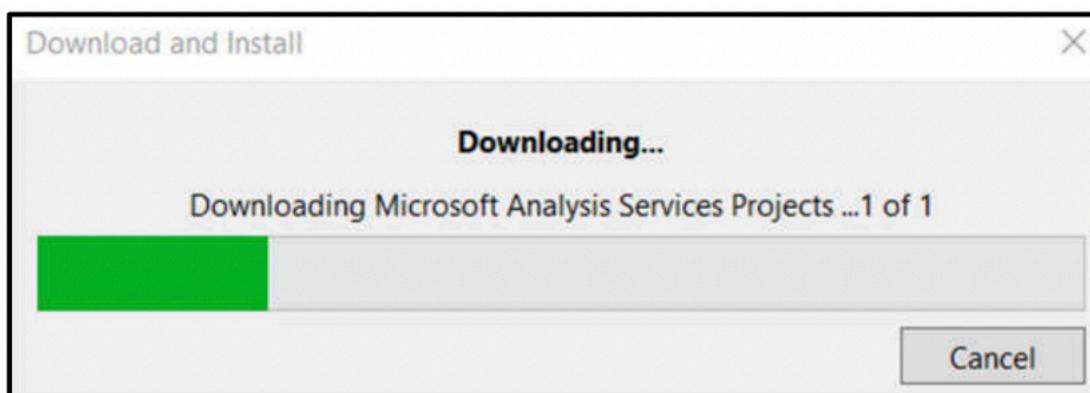
Hình 154: Giao diện vào nơi download SSAS

**Bước 3:** Tìm và tải về công cụ Microsoft Analysis Services Projects.



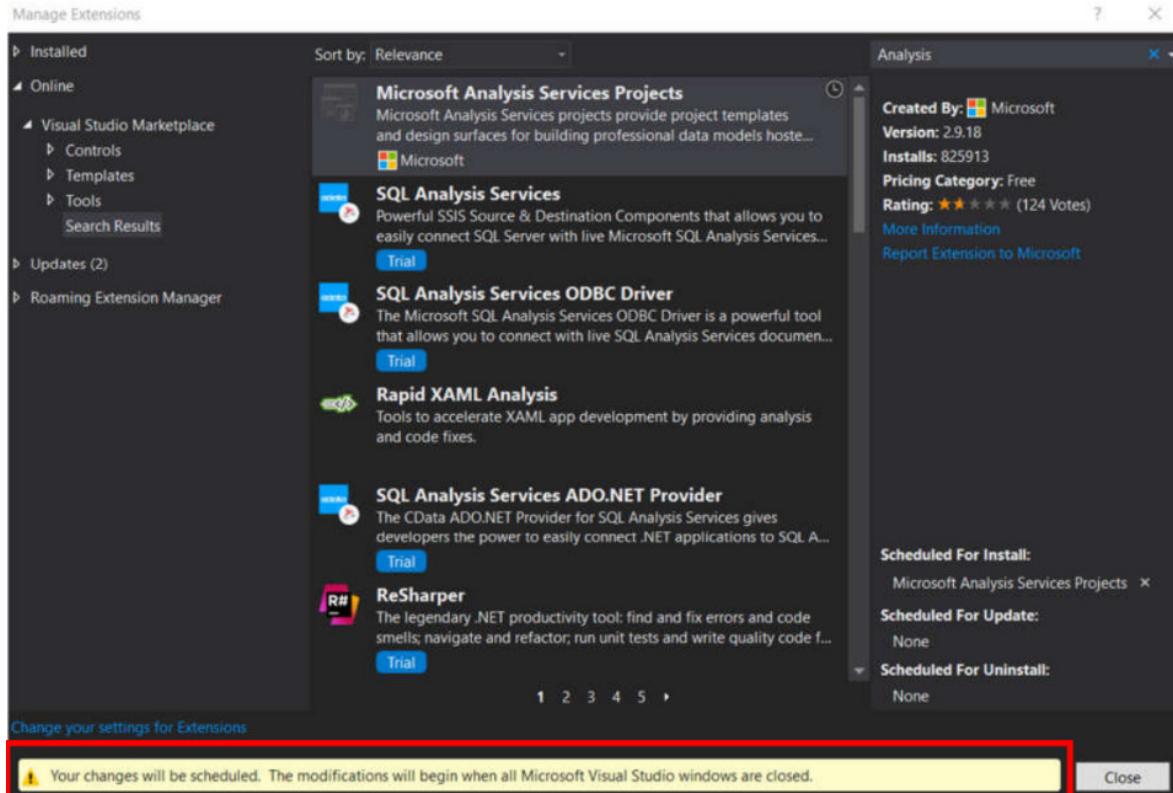
Hình 155: Giao diện hướng dẫn tìm kiếm và tải SSAS

**Bước 4:** Click vào **Download**. Hộp thoại tải về hiện lên, chờ đợi quá trình download hoàn tất.



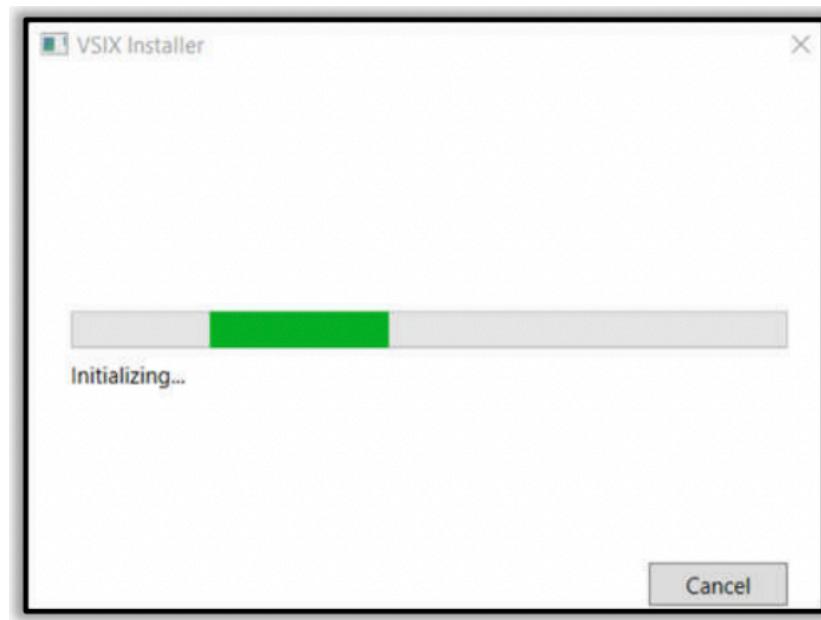
Hình 156: Hộp thoại download Microsoft Analysis Services Projects

**Bước 5:** Sau khi hoàn tất quá trình cài đặt, thoát khỏi Visual Studio 2019 để tích hợp SSAS.



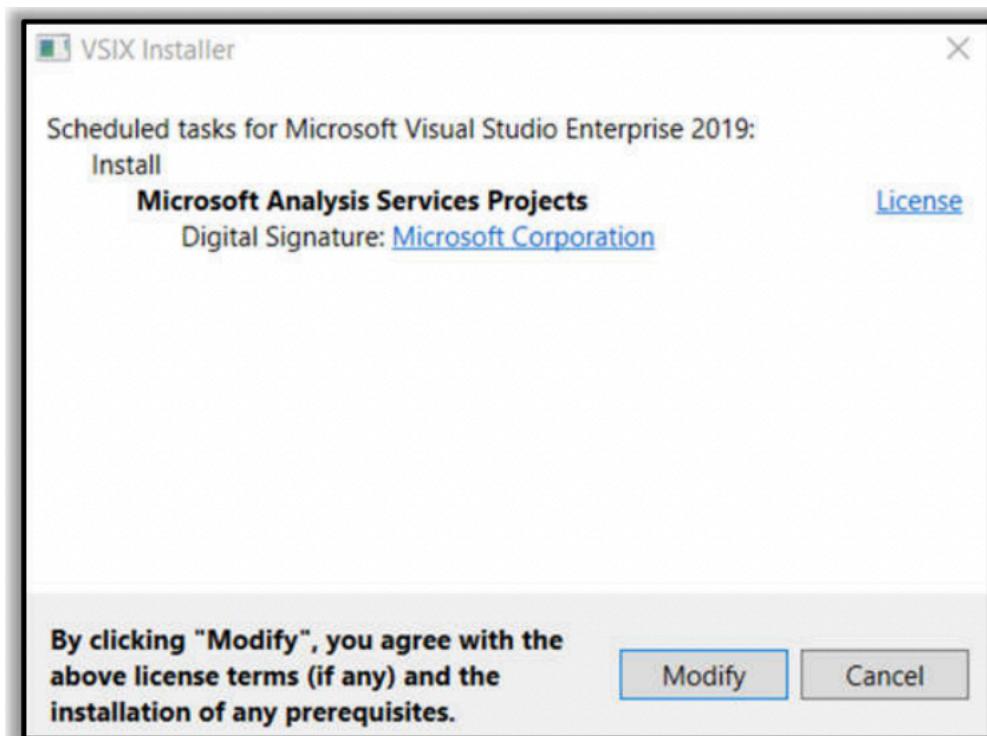
Hình 157: Hộp thoại yêu cầu khởi động lại Visual Studio 2019 để tích hợp công cụ SSAS

Sau khi thoát, hộp thoại **VSIX Installer** được khởi tạo.



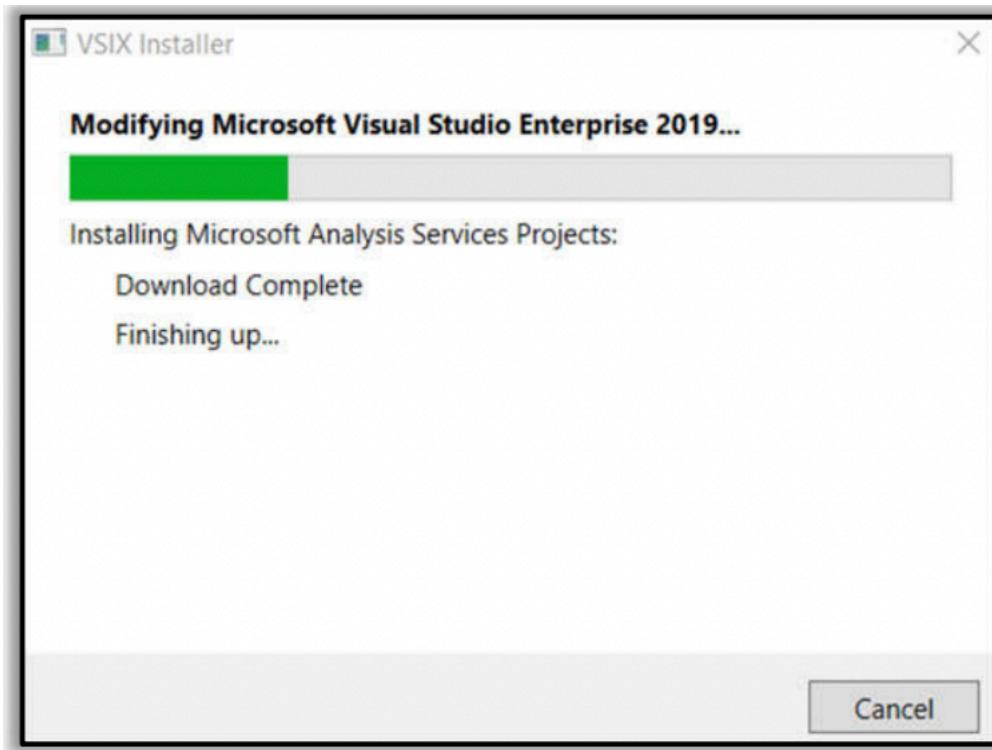
Hình 158: Hộp thoại VSIX Installer khởi tạo

**Bước 6:** Sau khi hoàn tất quá trình khởi tạo, click “**Modify**” để đồng ý với các điều khoản trước khi cài đặt **Microsoft Analysis Services Projects**.



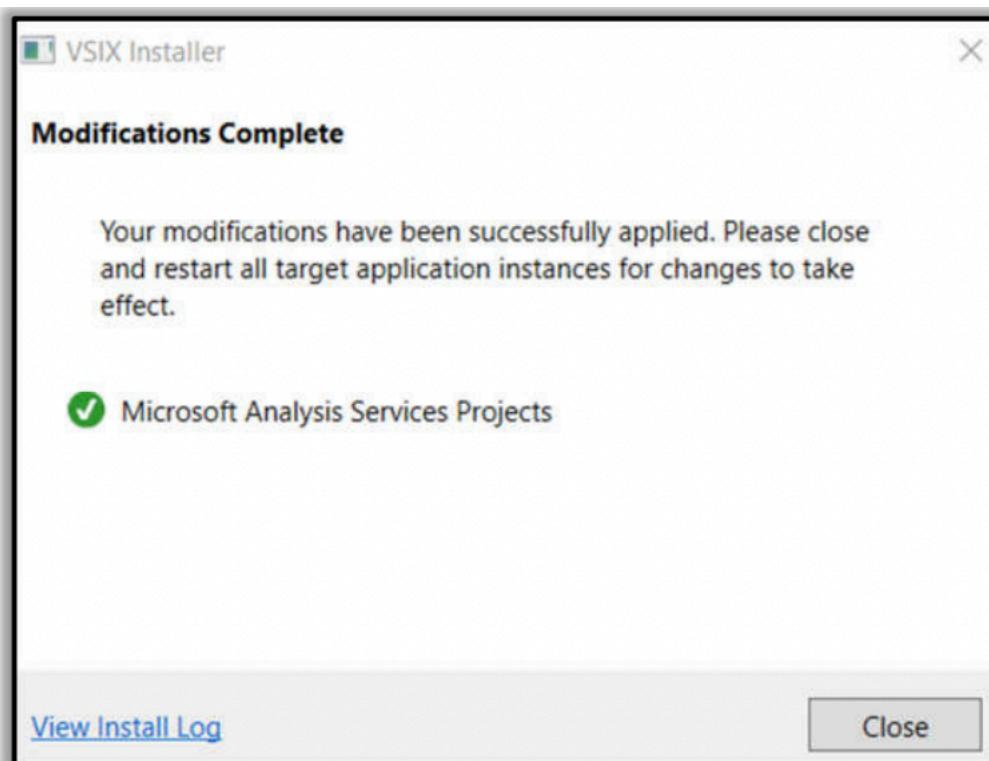
Hình 159: Hộp thoại yêu cầu chấp nhận điều khoản cài đặt

Tiếp tục quá trình cài đặt:



Hình 160: Hộp thoại tiếp tục hoàn thành quá trình tích hợp SSAS

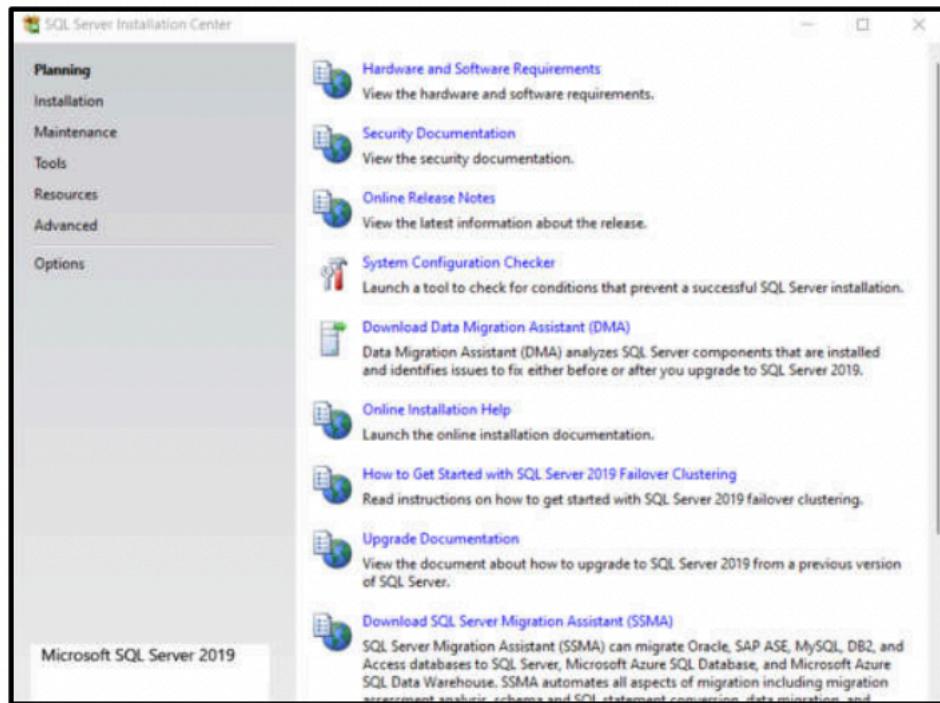
**Bước 7:** Sau khi quá trình cài đặt thành công, ta chọn “Close” để thoát khỏi cửa sổ VSIX Installer.



Hình 161: Giao diện tích hợp thành công SSAS vào Visual Studio 2019

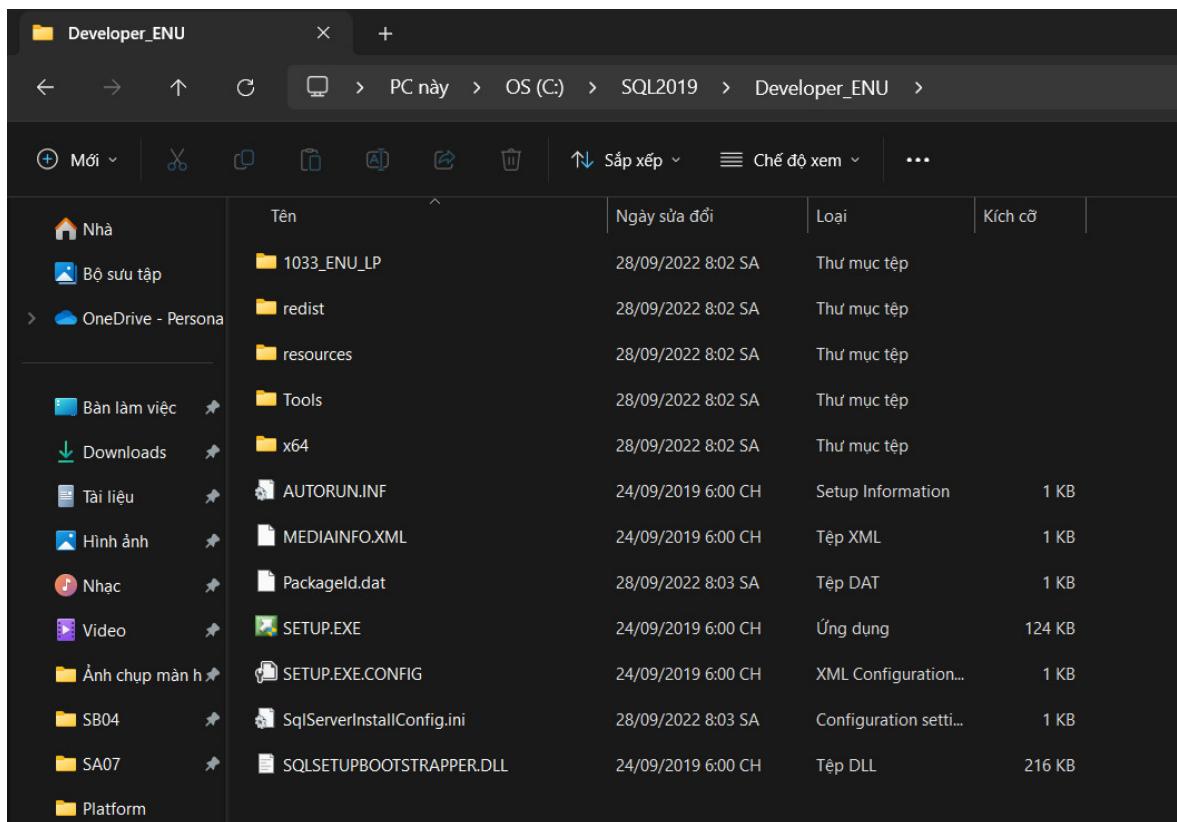
### 3.1.2. Cài đặt Analysis Services:

**Bước 1:** Khởi động SQL Service Installation Center. Trong quá trình cài đặt MS SQL Server, cửa sổ này tự động được mở.



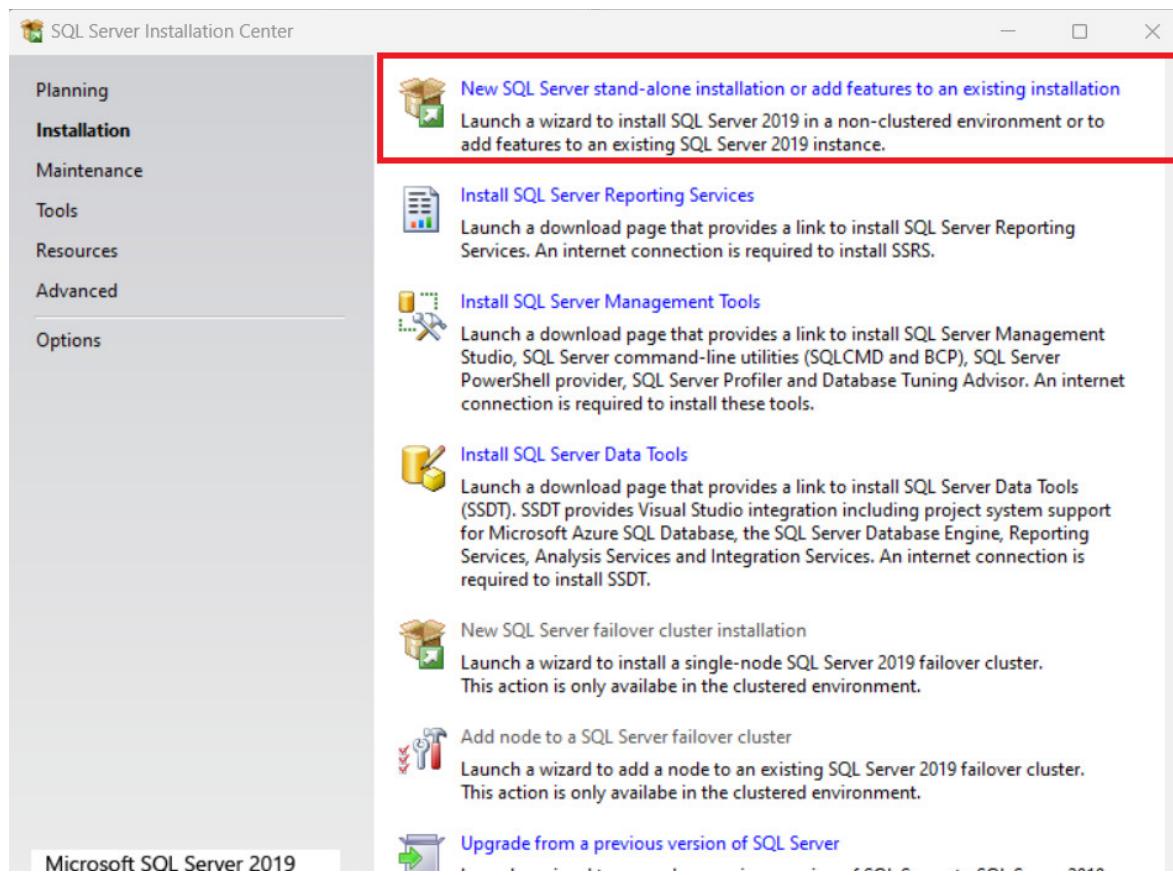
Hình 162: Giao diện sau khi mở file SETUP.exe

Nếu đã cài đặt MS SQL Server trước đó, ta vào thư mục chứa SQL Server (hiện tại là C:\SQL2019\Developer\_ENU). Sau đó khởi động “**SETUP.EXE**” để mở SQL Service Installation Center.



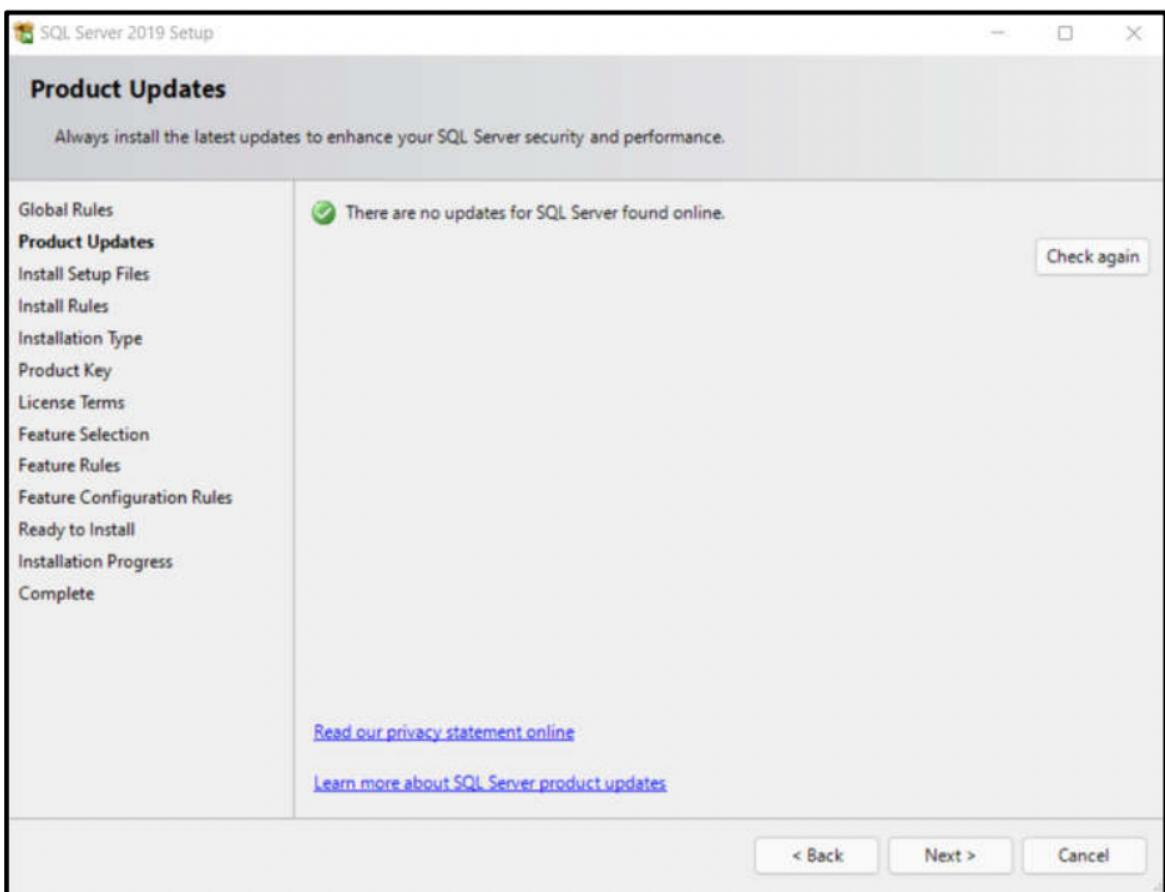
Hình 163: Vị trí file SETUP.exe sau khi cài đặt SQL Server

**Bước 2:** Chọn thẻ “**Installation**”, sau đó double click vào New SQL Server stand-alone installation or add features to an existing installation.



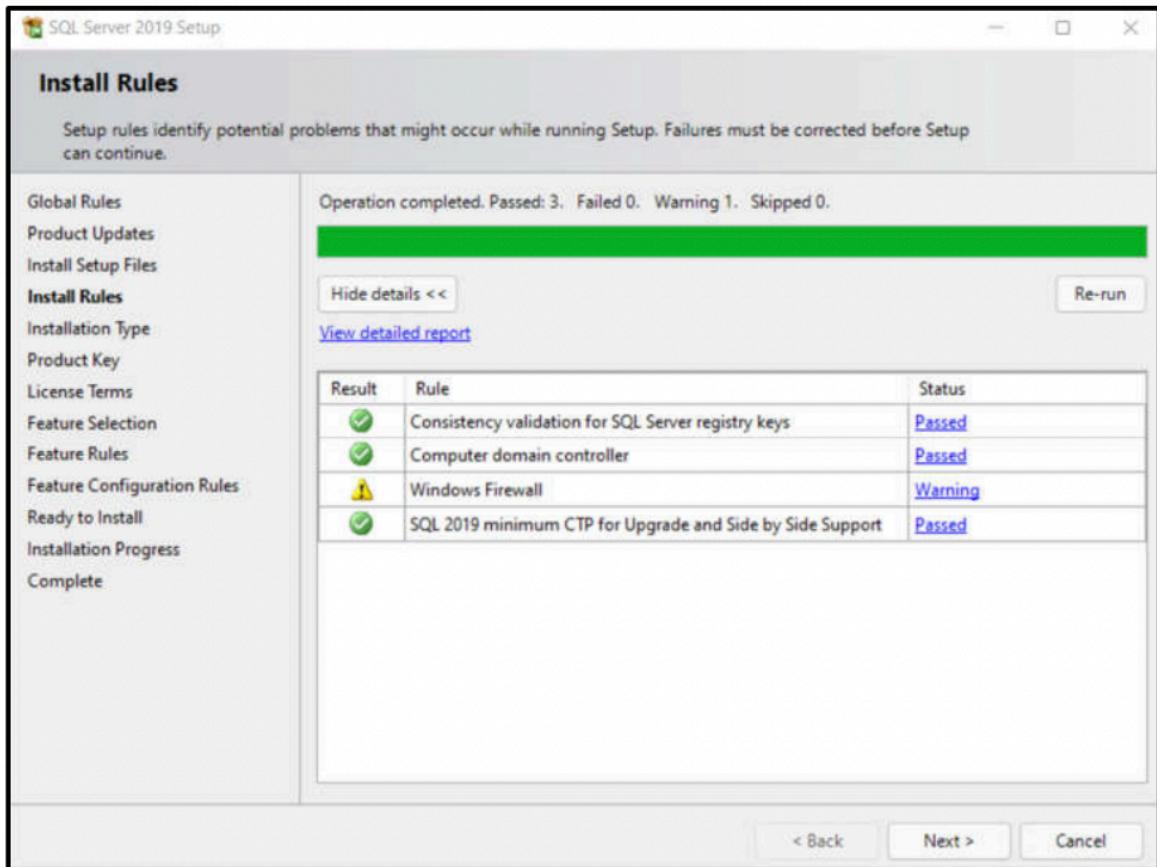
Hình 164: Bắt đầu quá trình cài đặt Analysis Service

**Bước 3:** Hộp thoại mới xuất hiện, ta chọn “Next”.



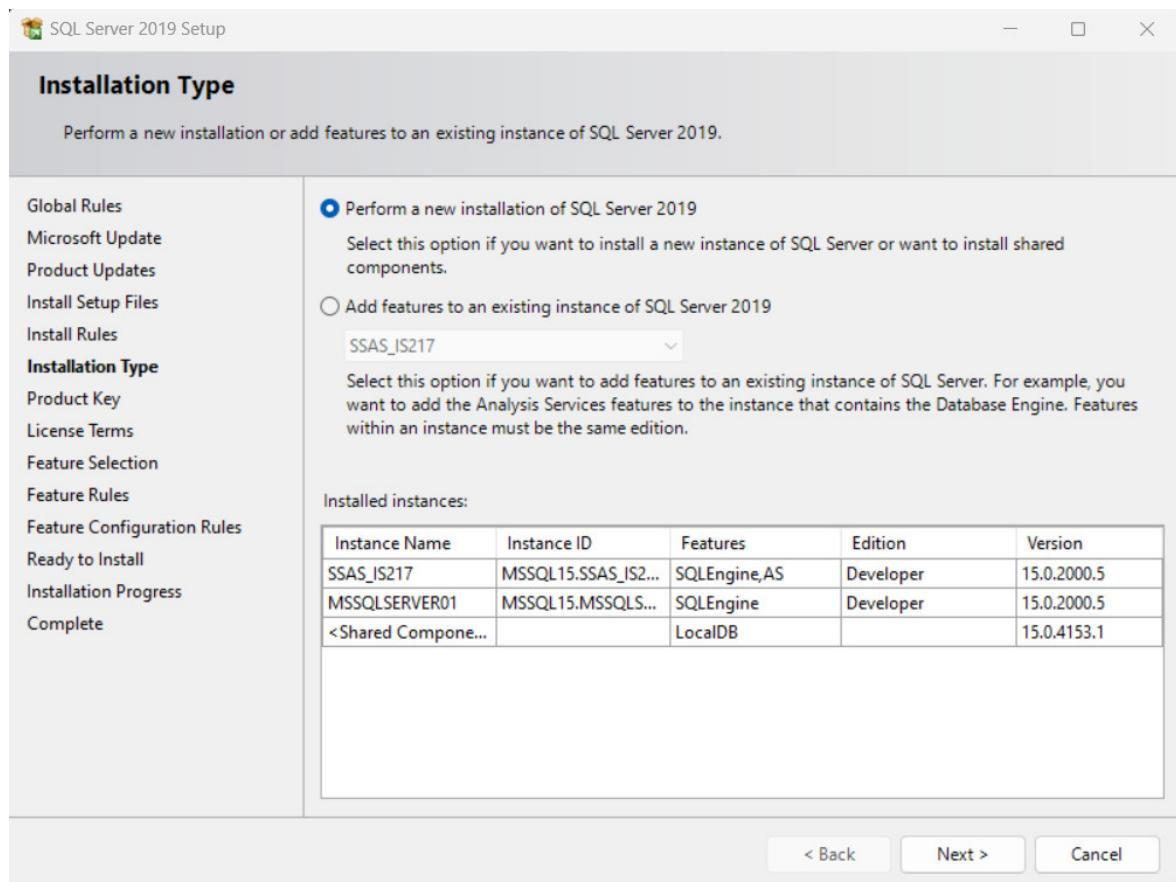
Hình 165: Giao diện thẻ Product Updates

**Bước 4:** Tiếp tục chọn “Next”.



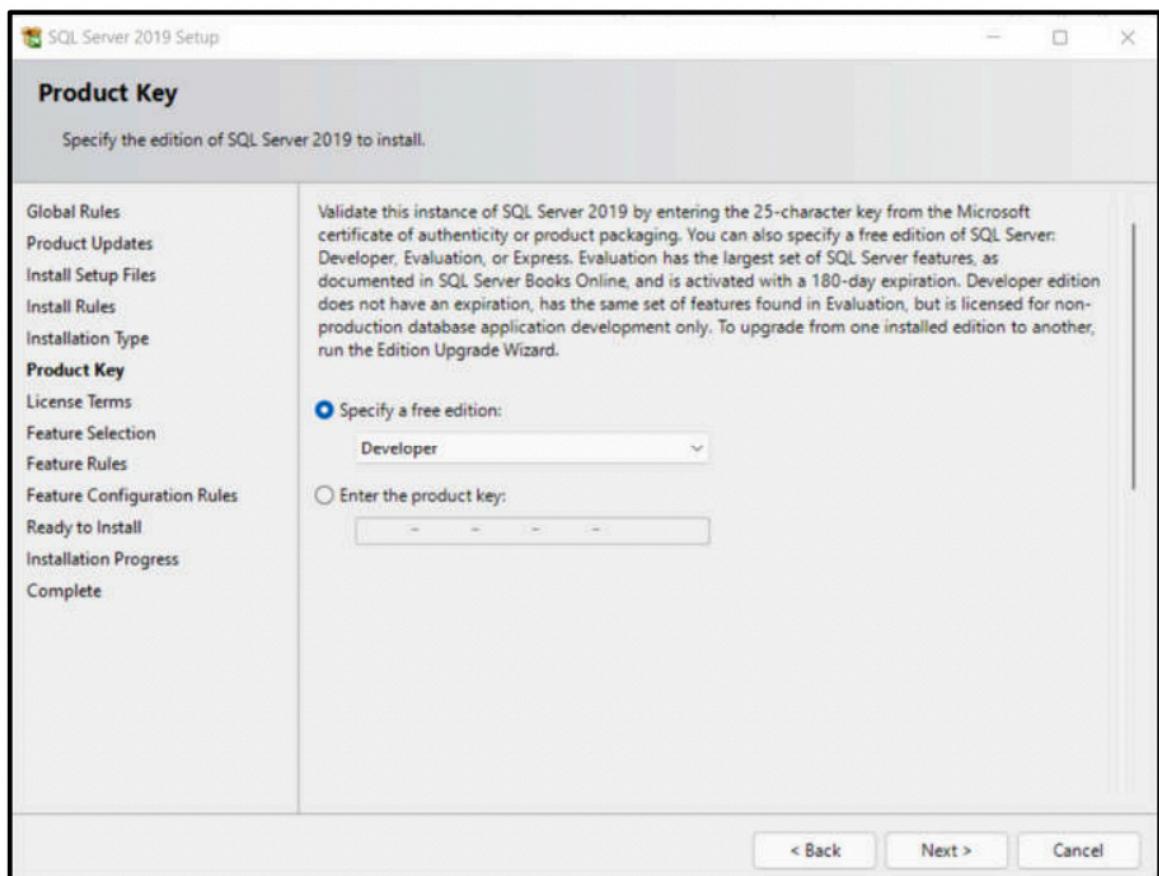
Hình 166: Giao diện thẻ *Install Rules*

**Bước 5:** Tiếp tục chọn “Next”.



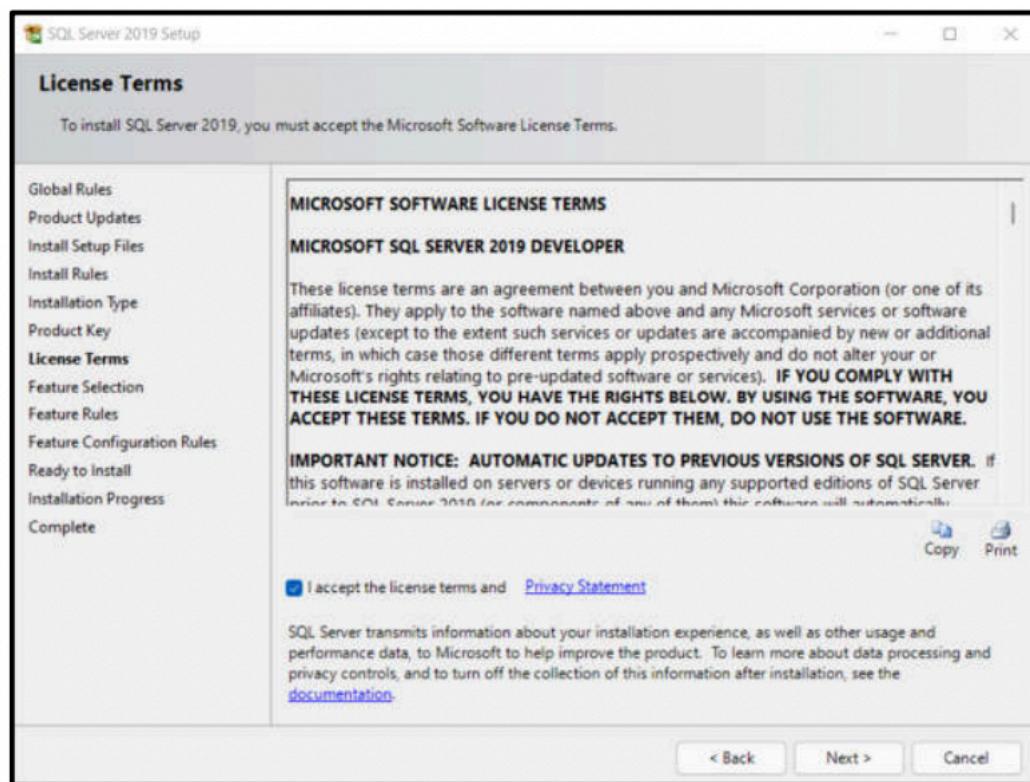
Hình 167: Giao diện thẻ Installation Type

**Bước 6:** Tiếp tục chọn “Next”.



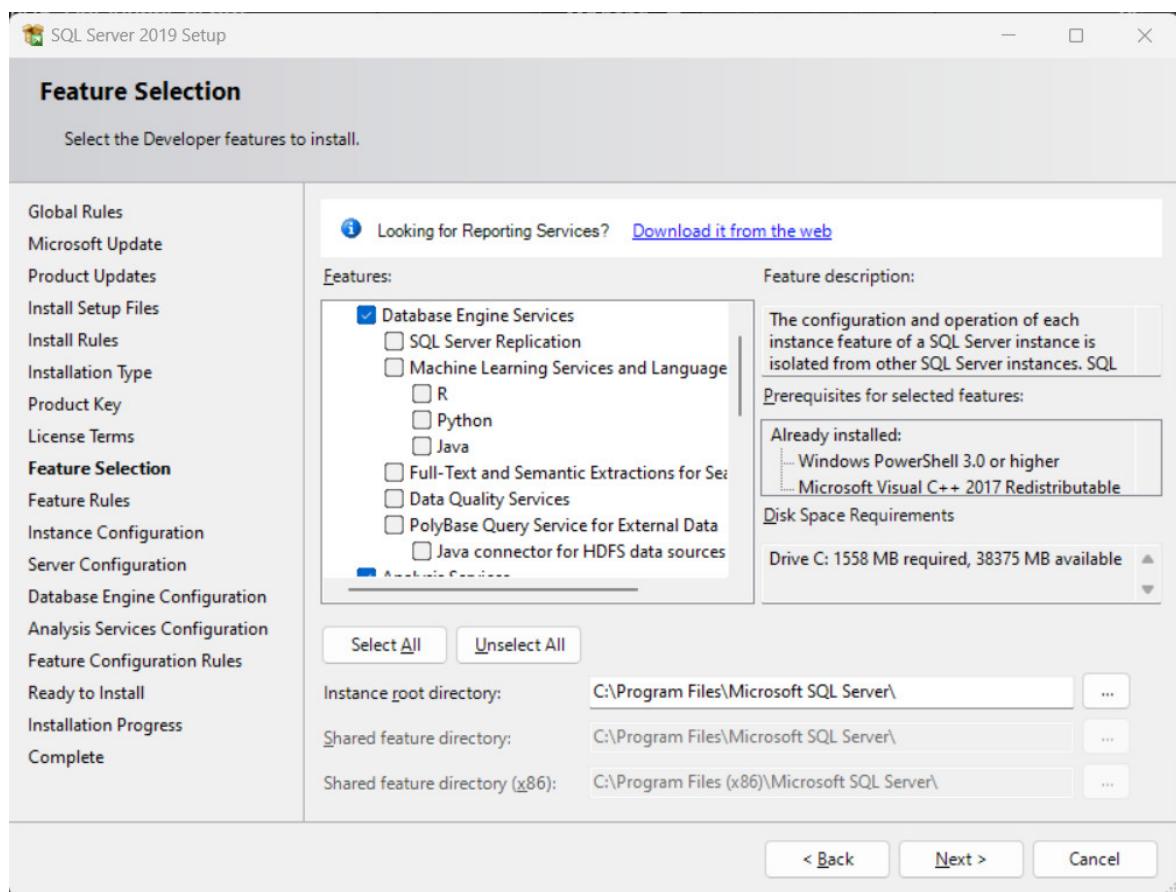
Hình 168: Giao diện thẻ Product Key

**Bước 7:** Tick chọn “I accept license terms and Privacy Statement”. Sau đó, tiếp tục chọn “Next”.



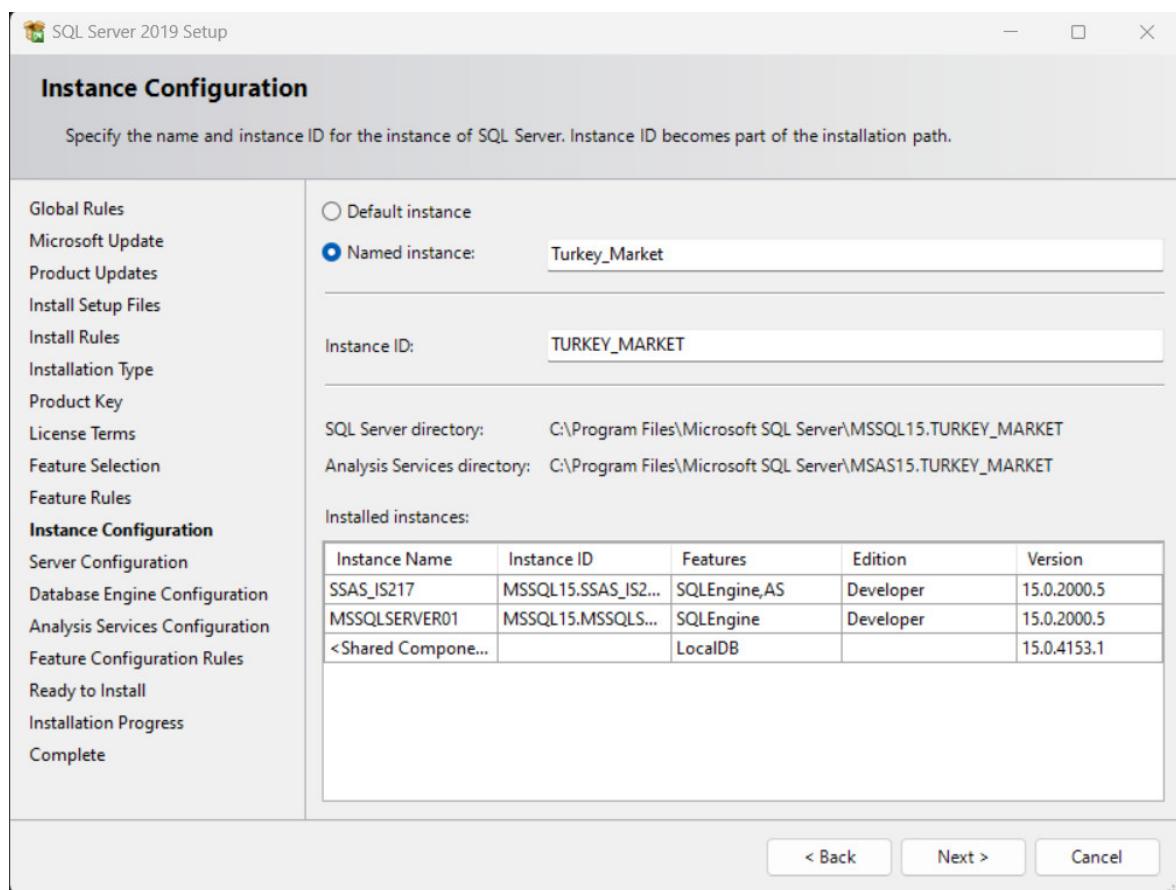
Hình 169: Giao diện thẻ License Terms

**Bước 8:** Ta chọn Database Engine Services và Analysis Services. Sau đó tiếp tục chọn “Next”.



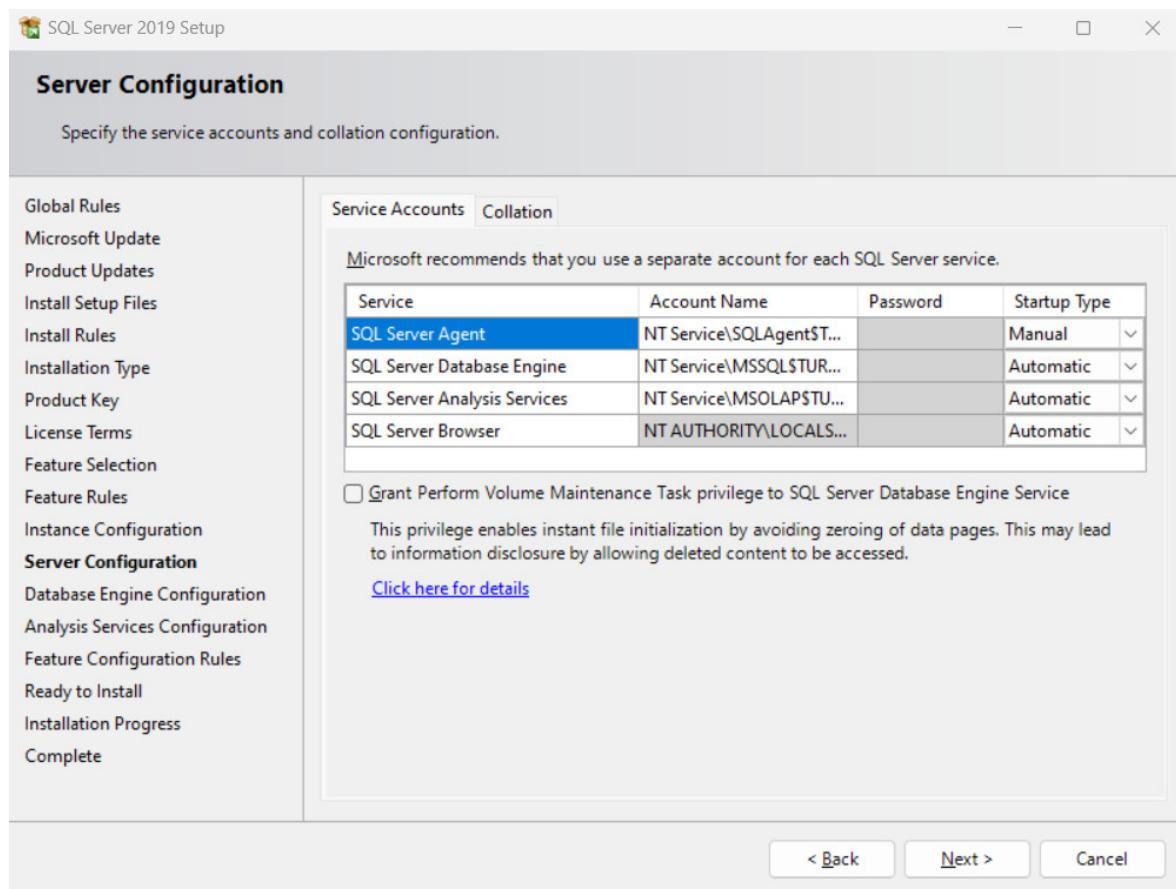
Hình 170: Giao diện thẻ Feature Selection (cài đặt DES và AS)

**Bước 9:** Chọn Named Instance và đặt tên cho instance vừa tạo.



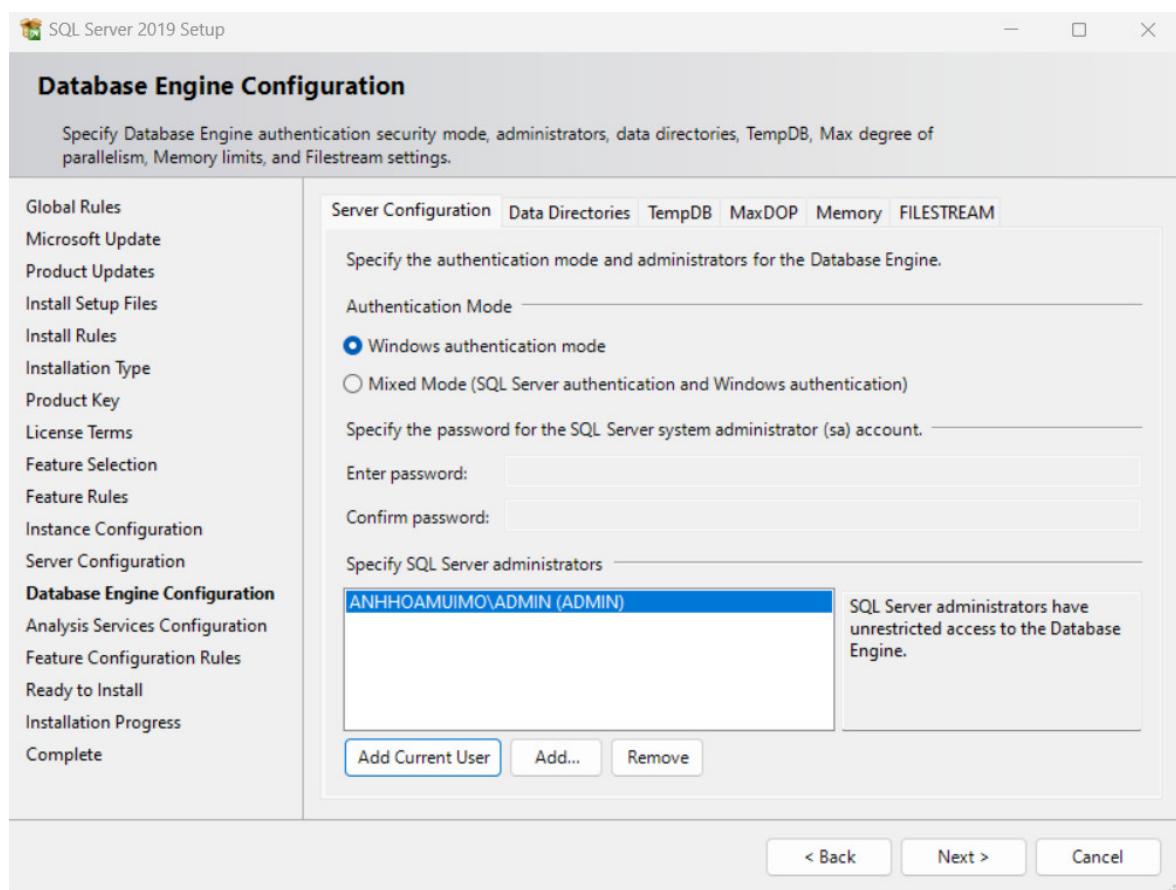
Hình 171: Giao diện thẻ Instance Configuration (tích hợp công cụ vào instance có sẵn)

**Bước 10:** Tiếp tục chọn “Next”.



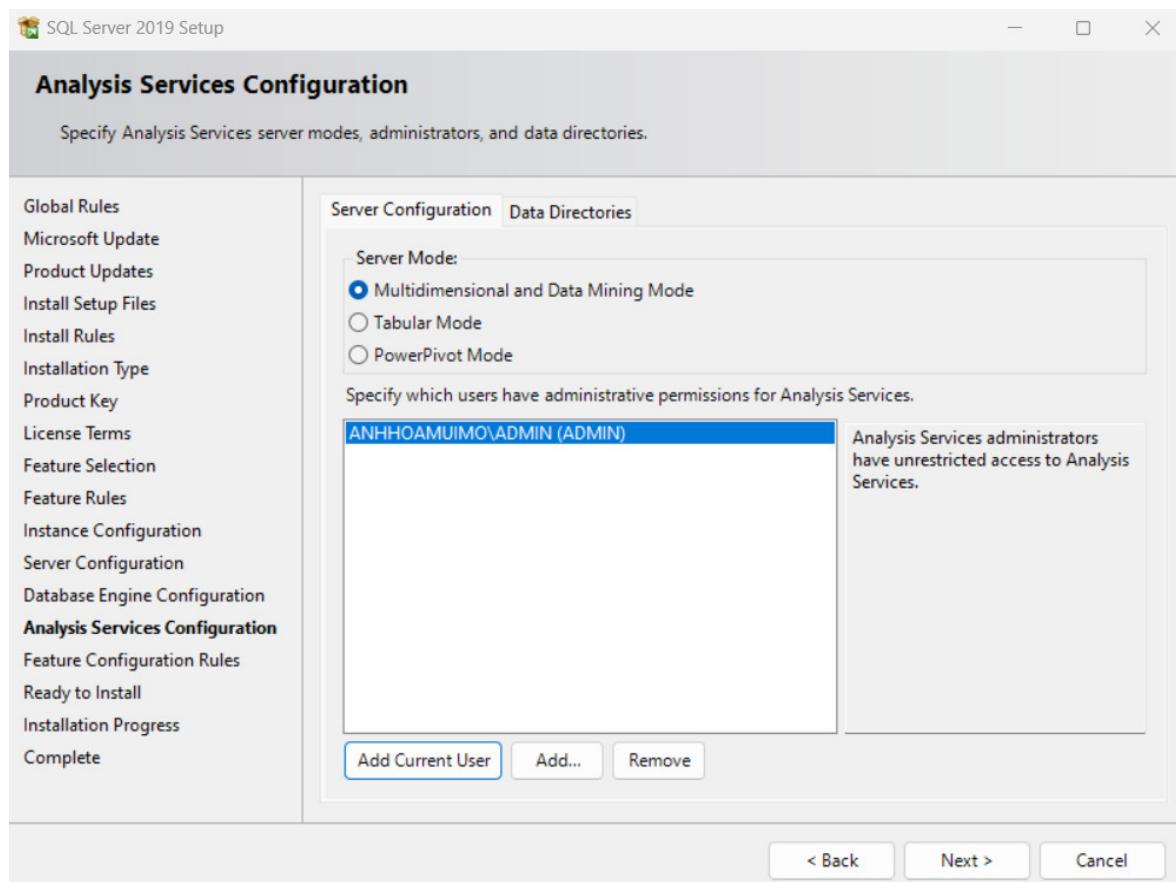
Hình 172: Giao diện thẻ Server Configuration

**Bước 11:** Chọn **Window Authentication Mode**. Tiếp theo, click “Add Current User”. Sau khi danh sách user xuất hiện, ta chọn user tương ứng và chọn “Next”.



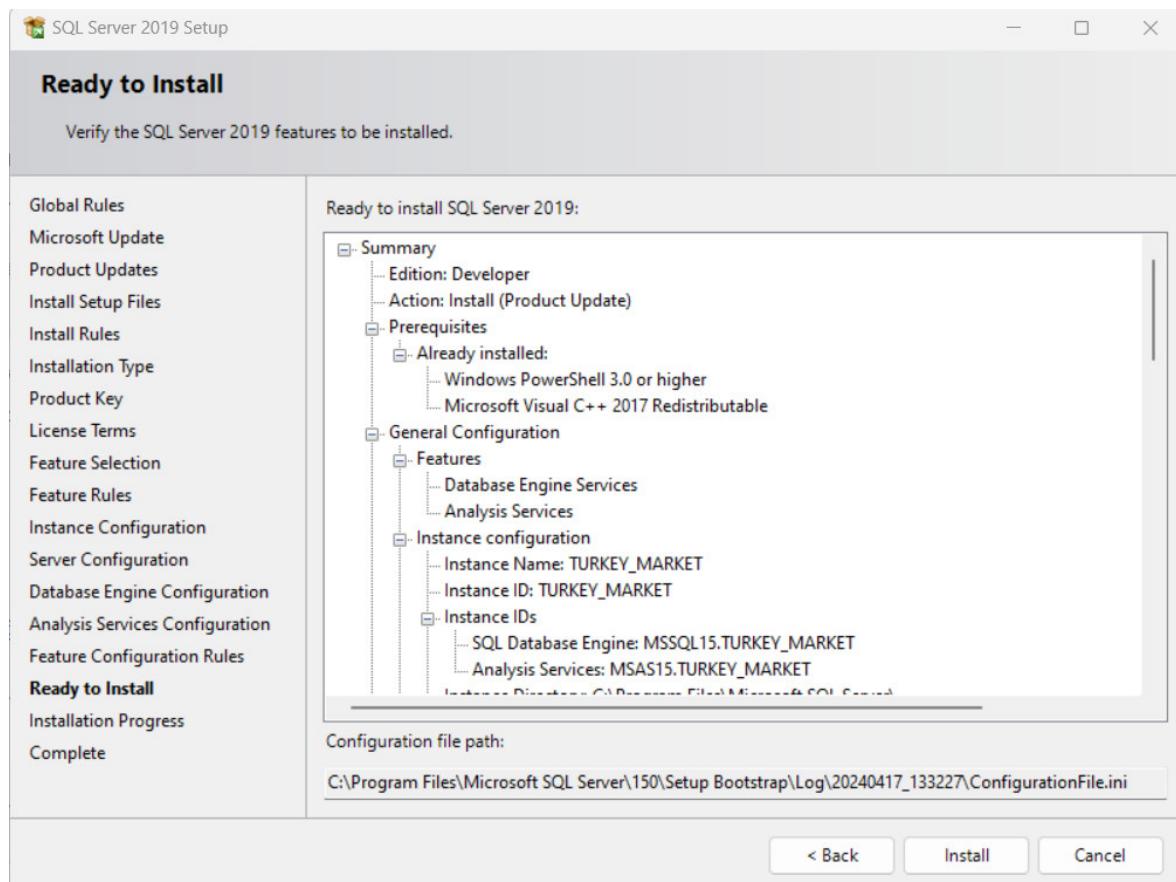
Hình 173: Giao diện thẻ Database Engine Configuration

**Bước 12:** Chọn **Multidimensional and Data Mining Mode**. Tiếp theo, chọn **Add Current User**. Sau khi user xuất hiện, ta chọn vào tên user tương ứng và chọn “**Next**”.



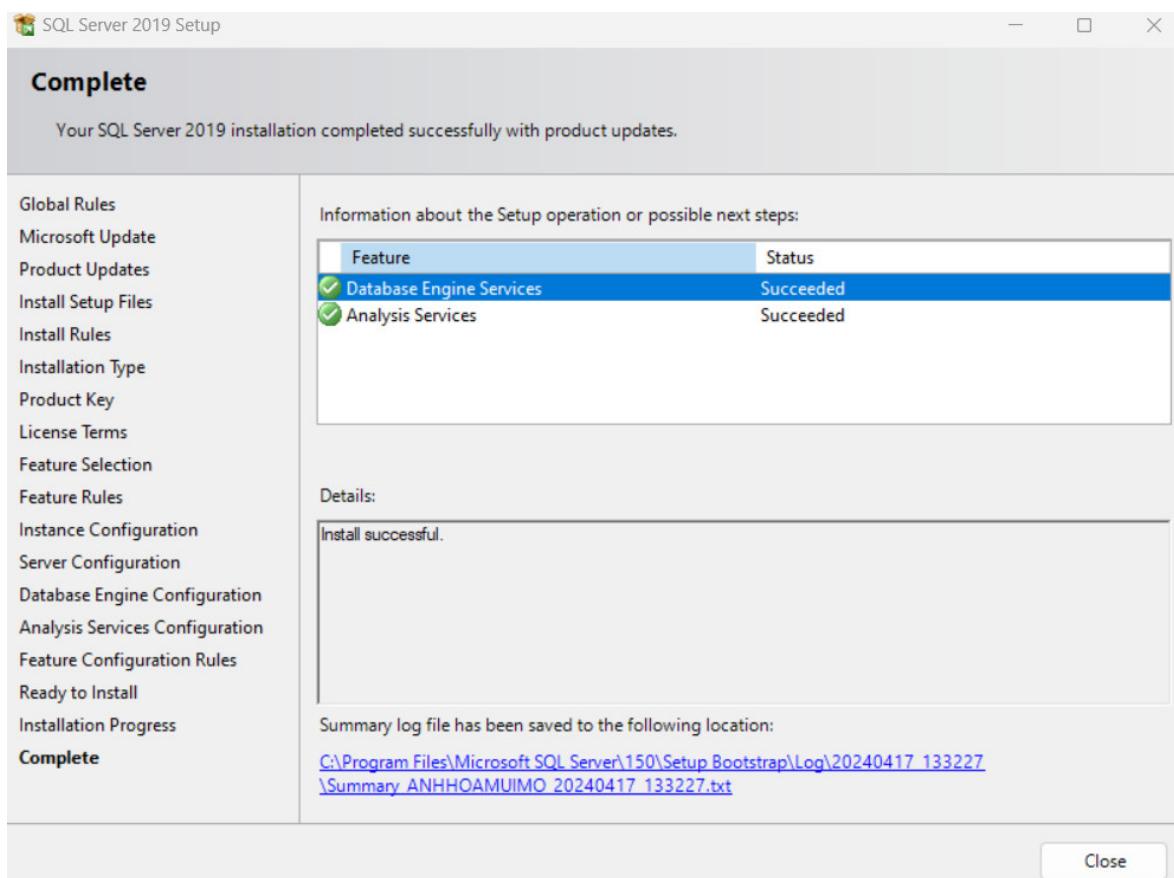
Hình 174: Giao diện thẻ Analysis Services Configuration

**Bước 13:** Hộp thoại tóm tắt những cài đặt mà chúng ta đã thực hiện ở trên xuất hiện, kiểm tra lại lần nữa và sau đó chọn “Install” để bắt đầu quá trình cài đặt.



Hình 175: Giao diện kiểm tra toàn bộ các thiết lập trước quá trình cài đặt

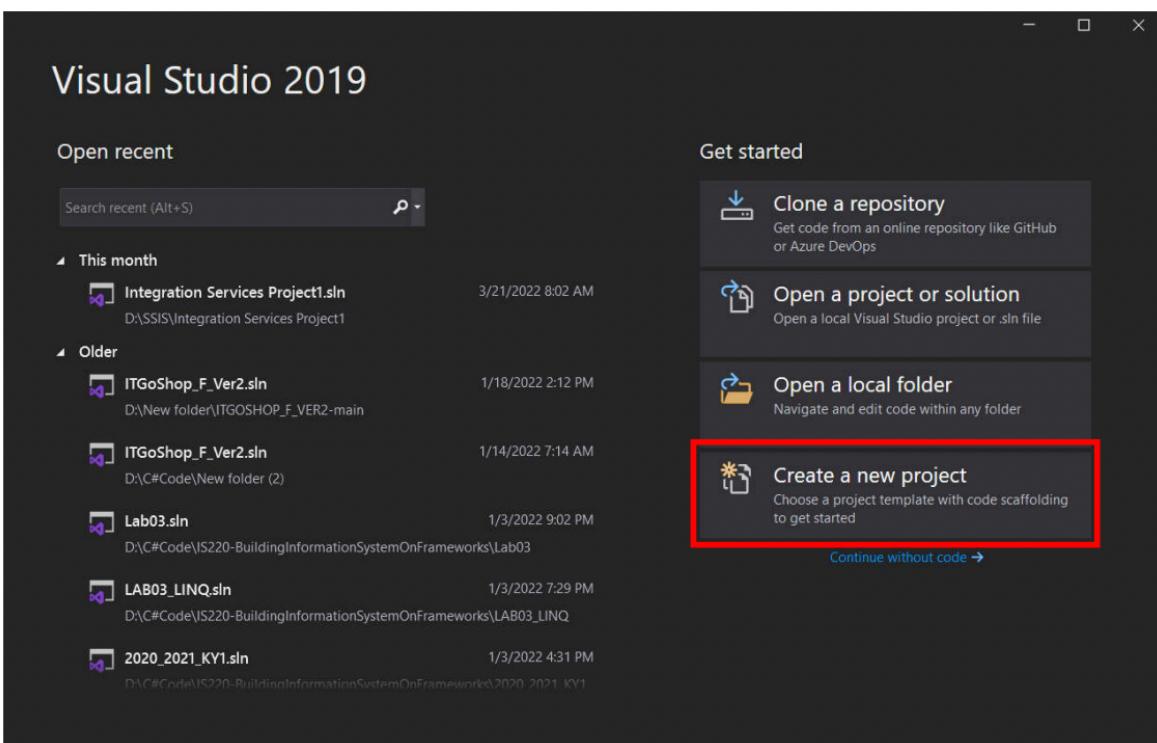
**Bước 14:** Quá trình cài đặt thành công. Chọn “Close” để hoàn tất.



Hình 176: Quá trình cài đặt hoàn tất và thành công

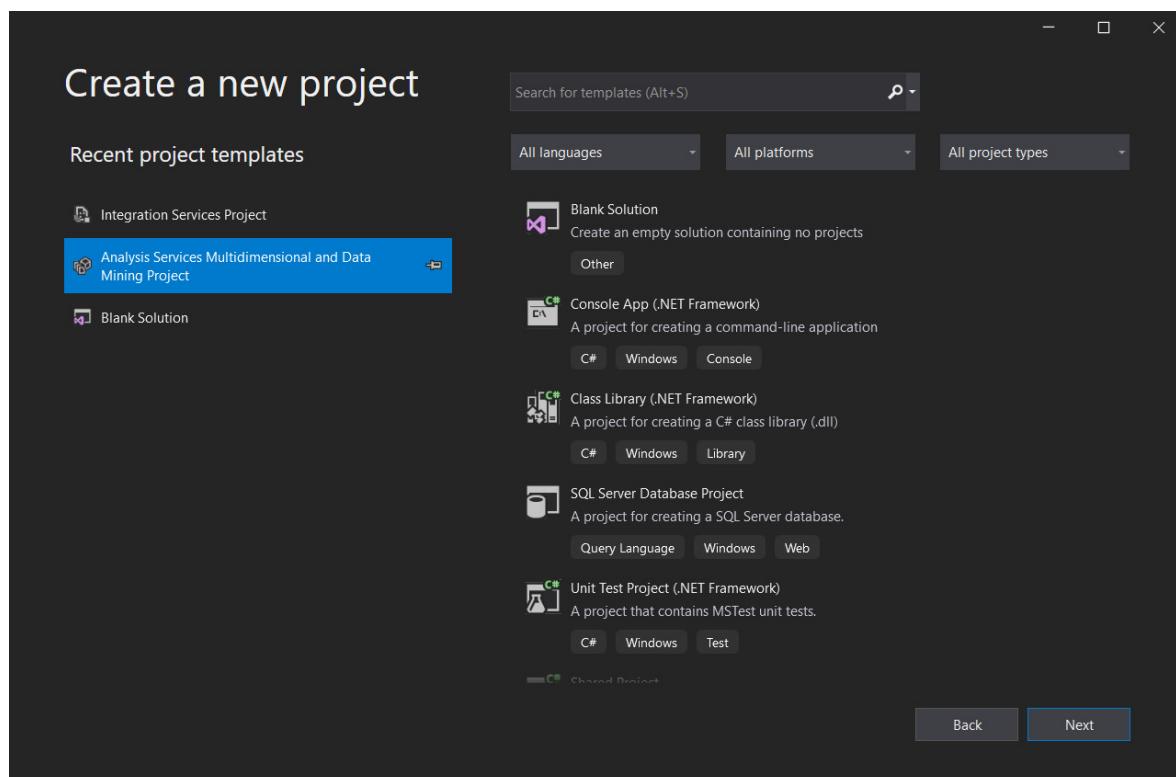
### 3.2. Tạo mới project SSAS:

Bước 1: Mở Visual Studio 2019 và chọn “Create a new project”.



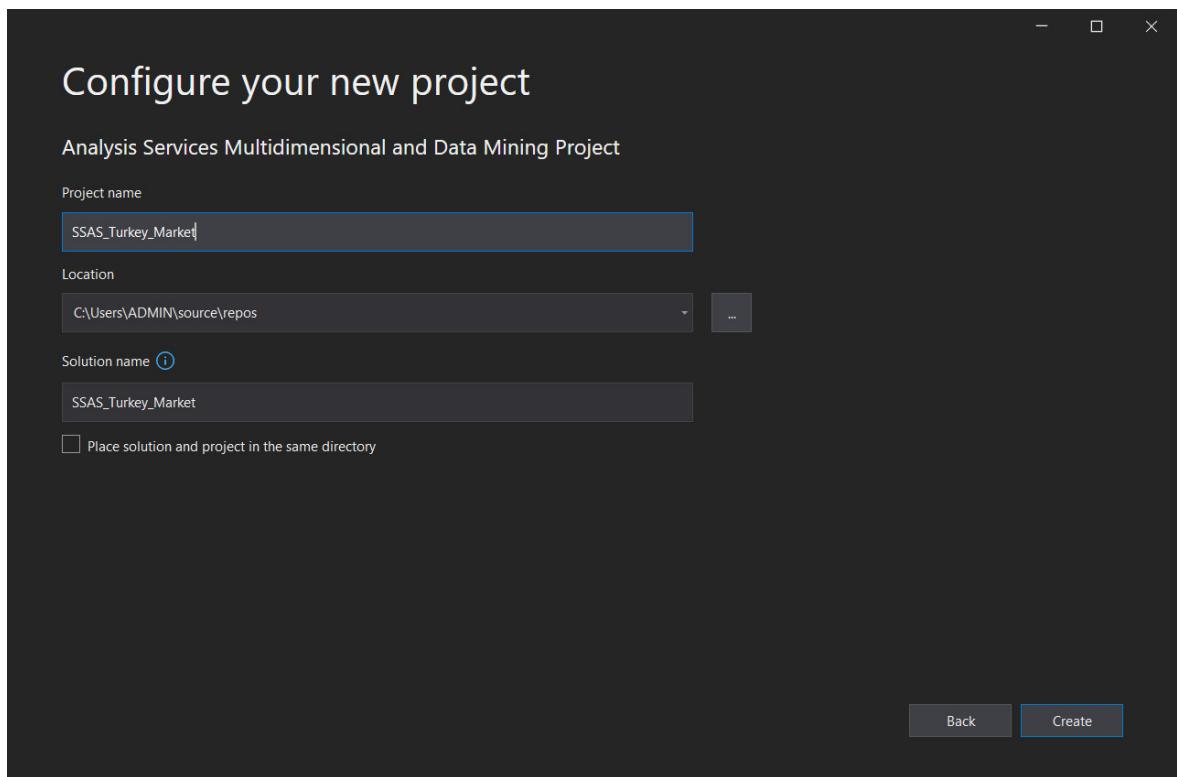
Hình 177: Khởi tạo một dự án mới

**Bước 2:** Chọn “Analysis Services Multidimensional Project” và chọn “Next”



Hình 178: Giao diện hướng dẫn chọn loại dự án SSAS

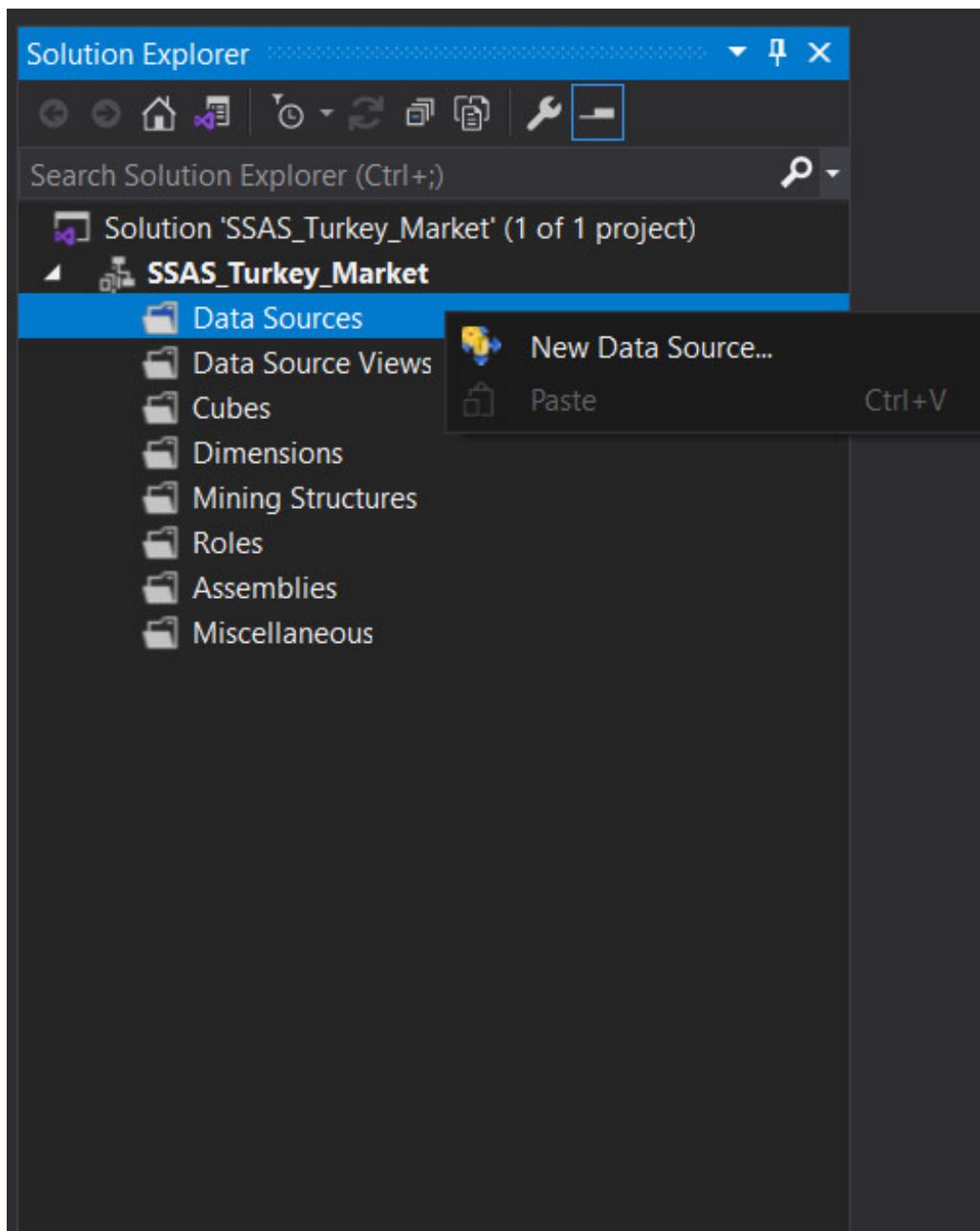
**Bước 3:** Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn “Create”.



Hình 179: Thiết lập nơi lưu trữ và đặt tên cho dự án SSAS

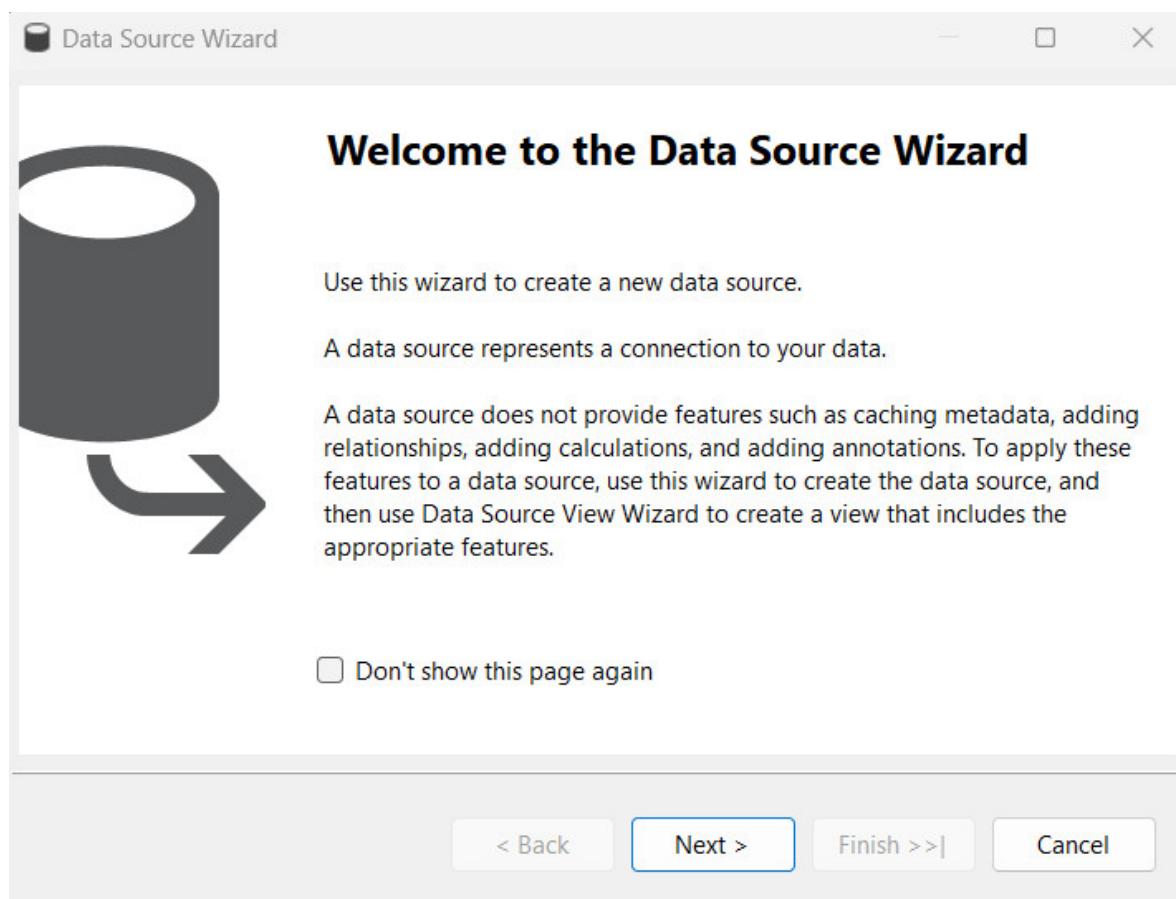
### 3.3. Xác định dữ liệu nguồn (Data Sources):

Bước 1: Tại thẻ “Solution Explorer”, click chuột phải vào thư mục “Data Sources” và chọn “New Data Source”



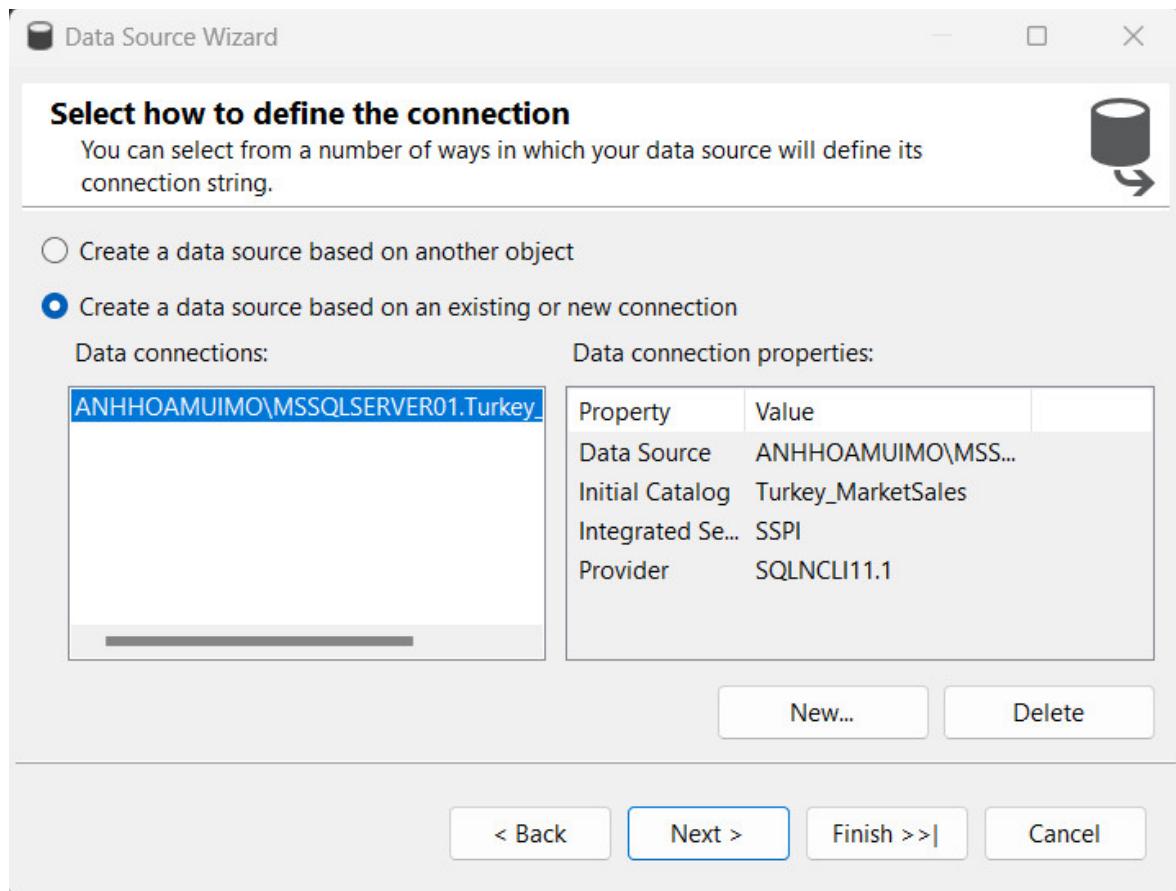
Hình 180: Giao diện tạo mới một Data Source (nguồn dữ liệu)

Bước 2: Hộp thoại “Data Source Wizard” xuất hiện, chọn “Next” để tiếp tục



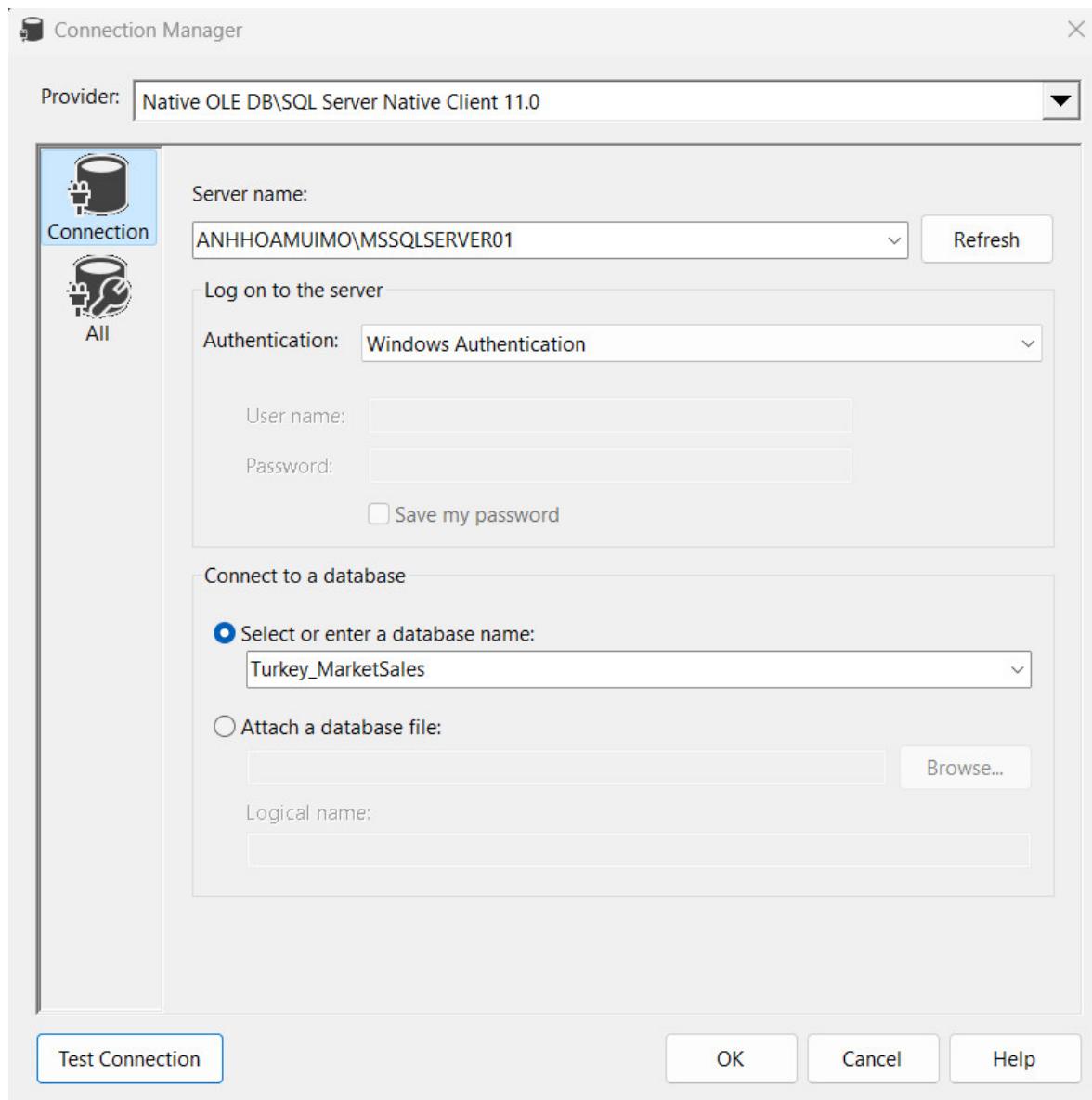
Hình 181: Hộp thoại bắt đầu quá trình tạo Data Source

**Bước 3:** Chọn “Create a data source based on an existing or new connection” sau đó chọn “New...” để tạo kết nối với cơ sở dữ liệu đã tạo trước đó trong quá trình SSIS



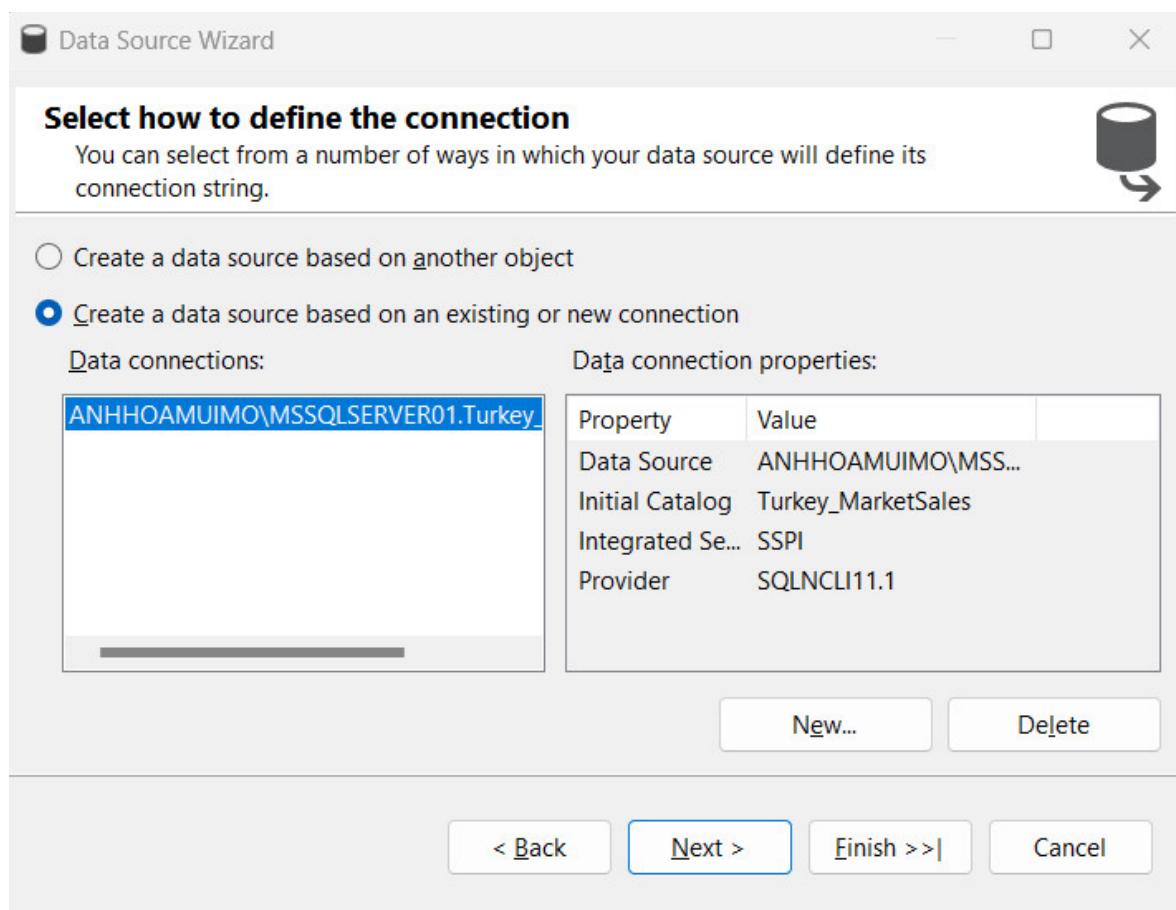
Hình 182: Tiến hành tạo mới một kết nối đến SQL Server

**Bước 4:** Hộp thoại “**Connection Manager**” xuất hiện, điền tên Server vào phần “**Server Name**” và chọn cơ sở dữ liệu đã tạo ra từ quá trình SSIS. Sau đó, click “**OK**” để hoàn tất



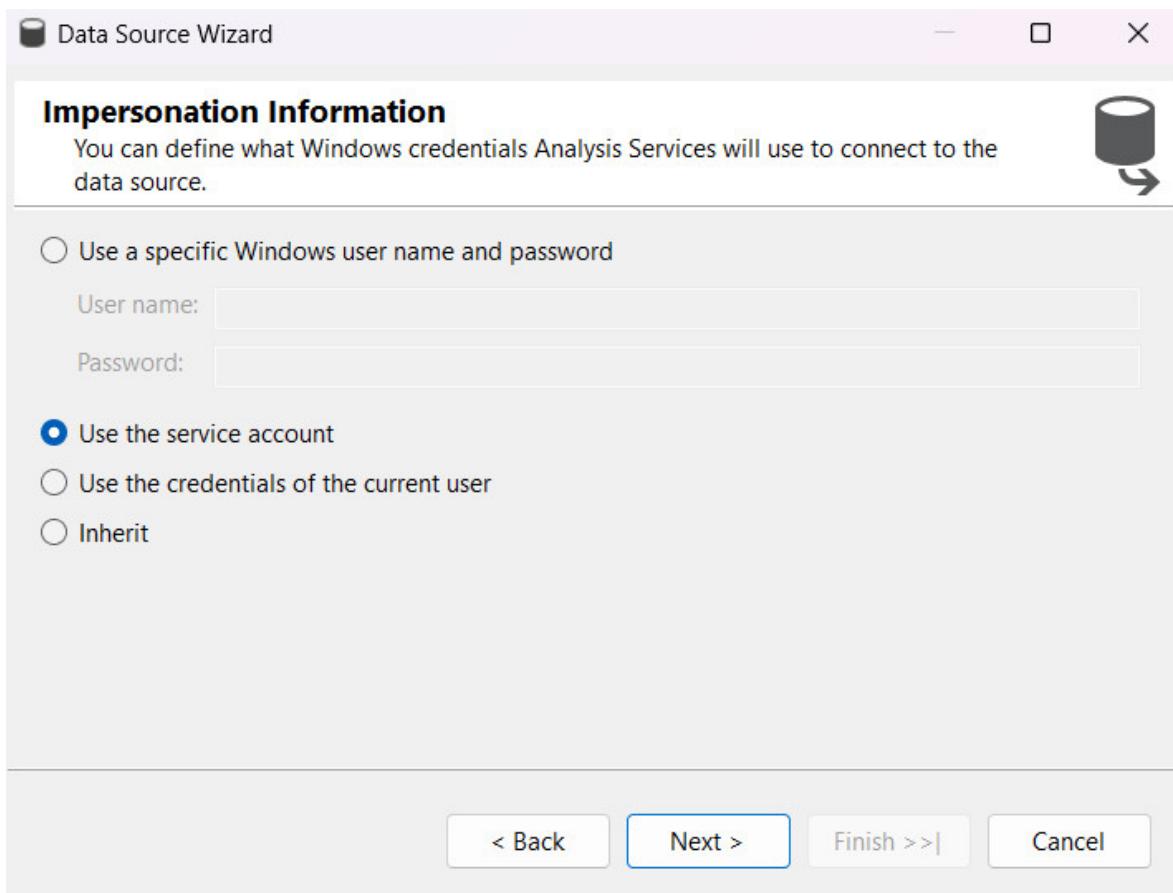
Hình 183: Tiến hành kết nối với kho dữ liệu đã tạo trong quá trình SSIS

Chọn data source vừa tạo ra và chọn “Next” để tiếp tục



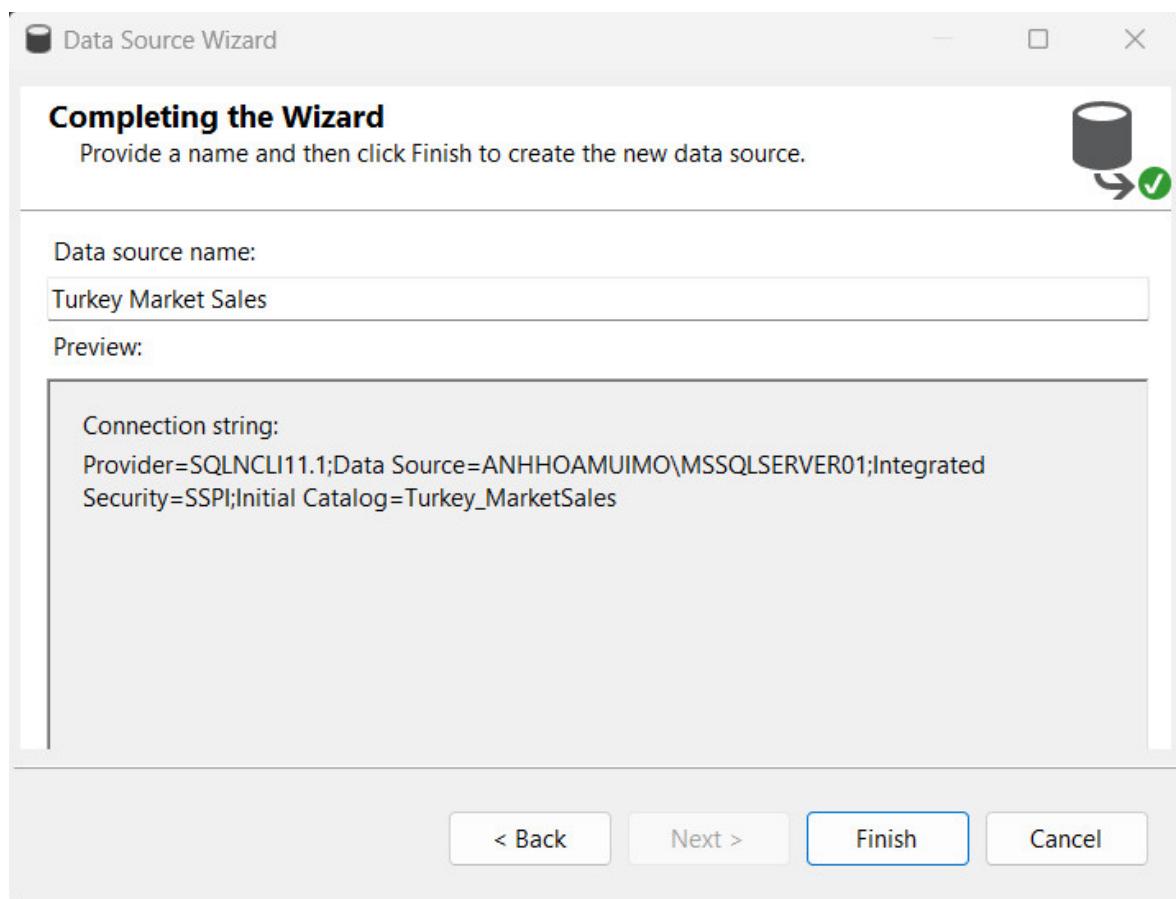
Hình 184: Tạo mới data source từ kho dữ liệu vừa kết nối

**Bước 5:** Chọn “Use the service account”, sau đó click “Next” để tiếp tục



Hình 185: Giao diện chọn tài khoản đăng nhập

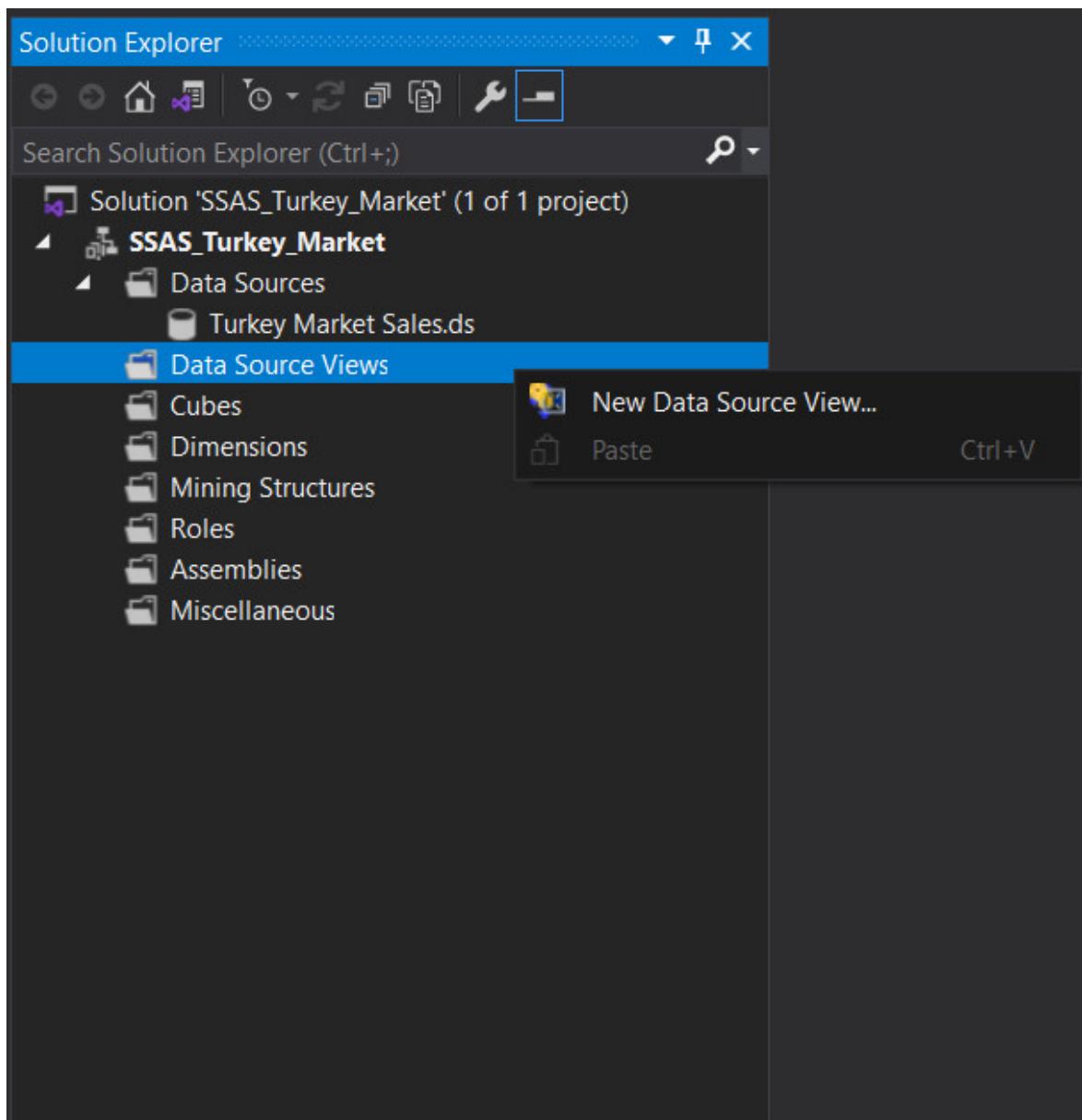
**Bước 6:** Cuối cùng ta click “Finish” để hoàn tất quy trình định nghĩa / khởi tạo nguồn dữ liệu



Hình 186: Kết quả sau khi tạo thành công data source

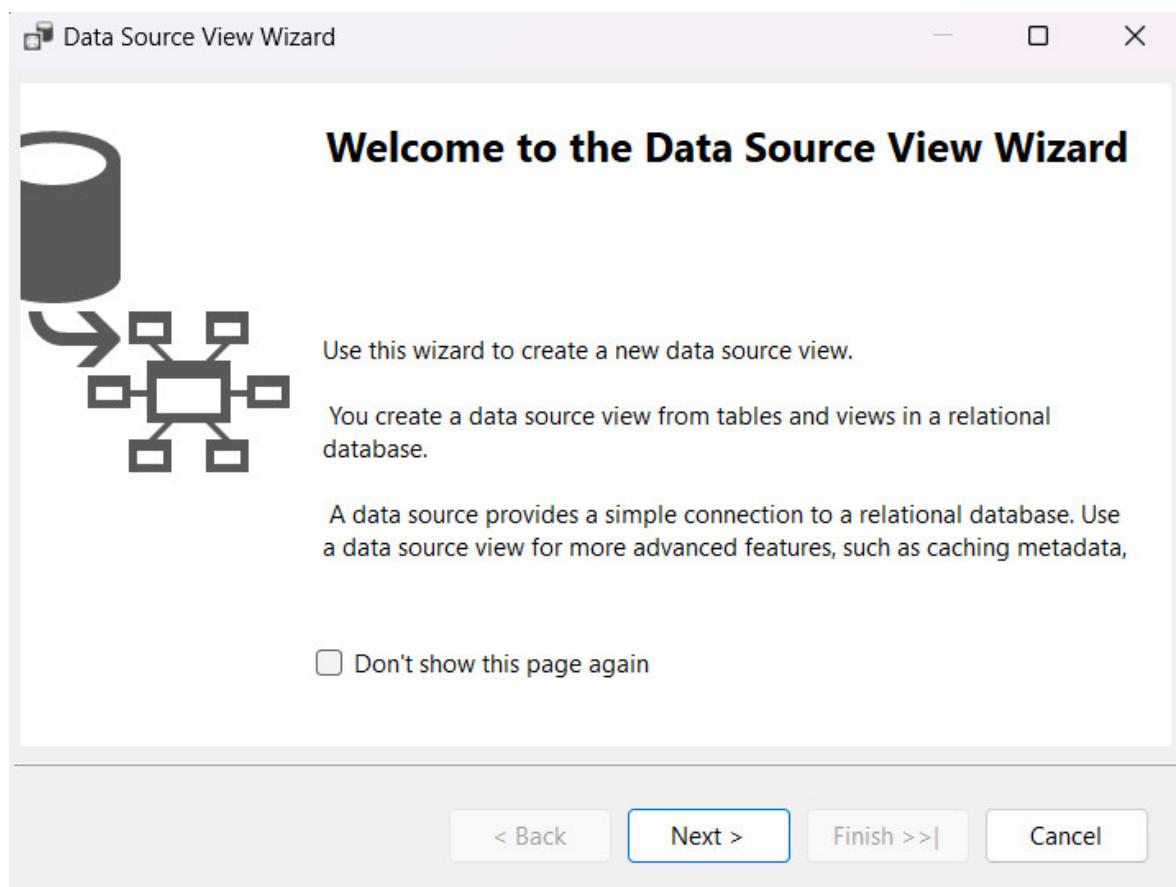
### 3.4. Xác định khung nhìn dữ liệu nguồn (Data Source Views):

Bước 1: Tại thẻ “Solution Explorer”, click chuột phải vào thư mục “Data Sources” và chọn “New Data Source View”



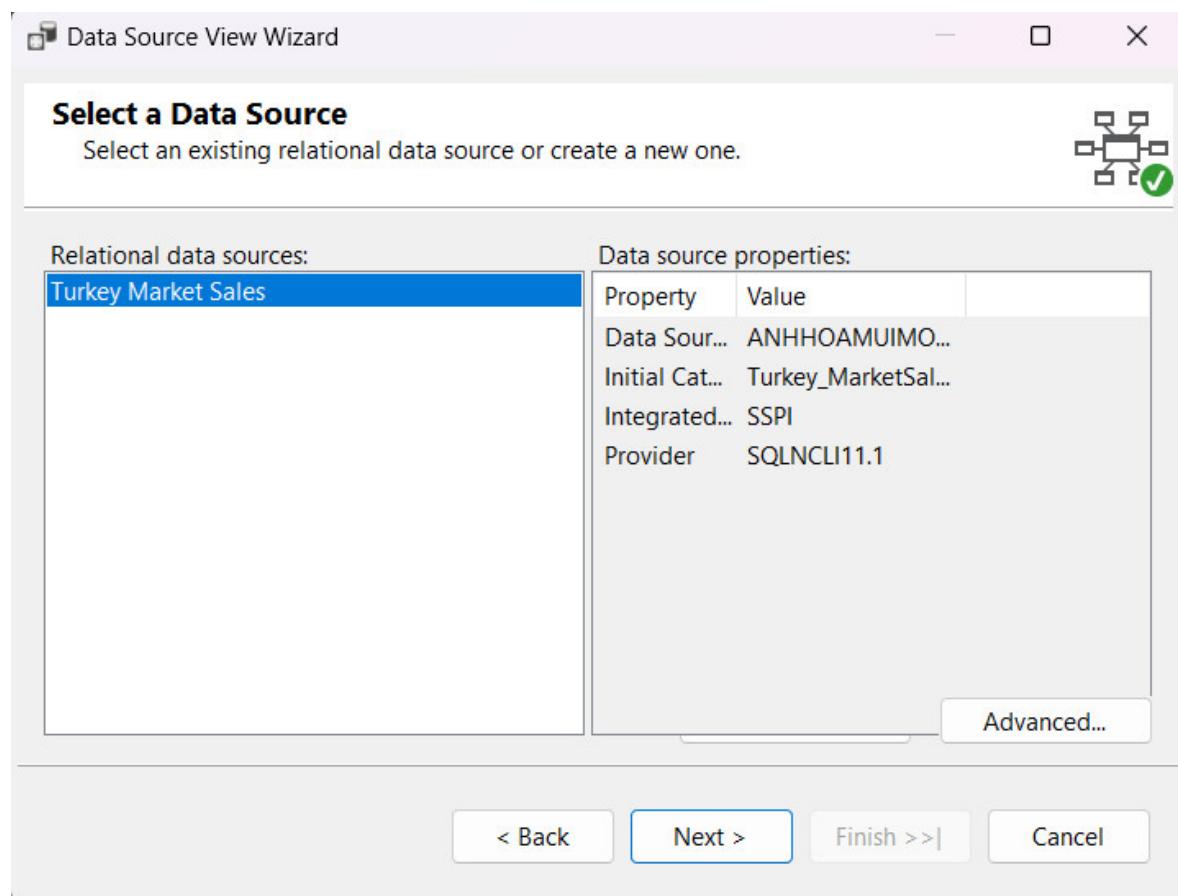
Hình 187: Giao diện tạo mới một Data Source View (khung nhìn dữ liệu)

**Bước 2:** Hộp thoại “**Data Source View Wizard**” xuất hiện, sau đó click “**Next**” để tiếp tục



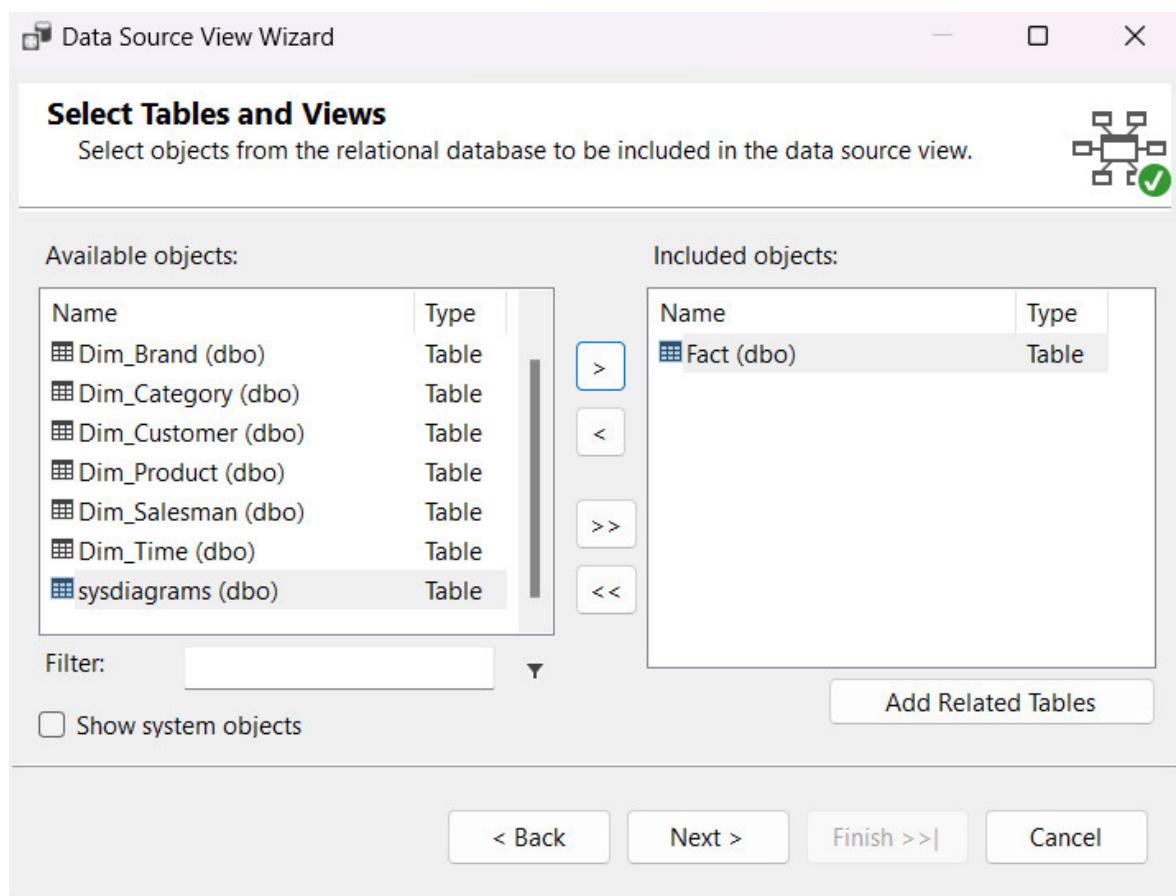
Hình 188: Hộp thoại bắt đầu quá trình tạo Data Source View

Bước 3: Chọn data source vừa tạo, sau đó click “Next” để tiếp tục



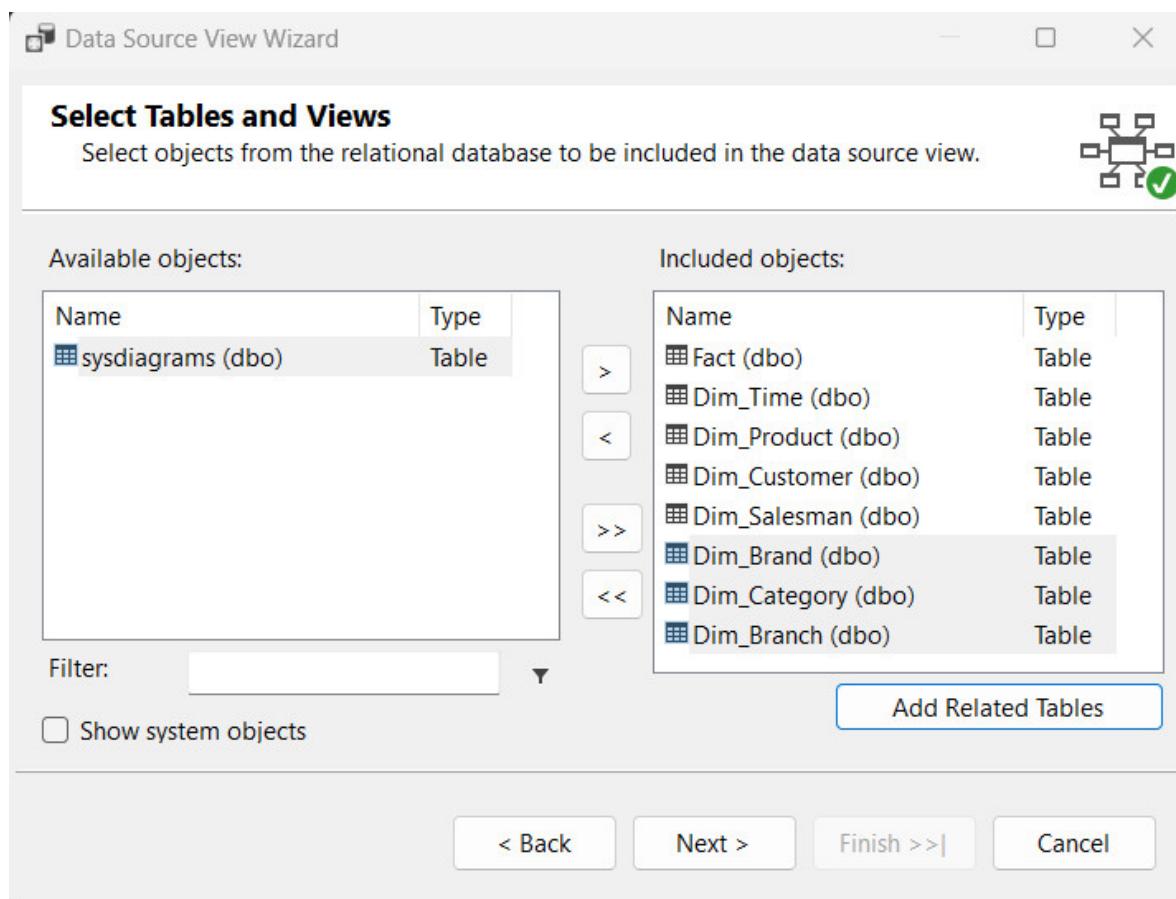
Hình 189: Chọn data source để tạo data source view

**Bước 4:** Chọn bảng “Fact”, sau đó click nút “>” để thêm bảng “Fact” vào data source view



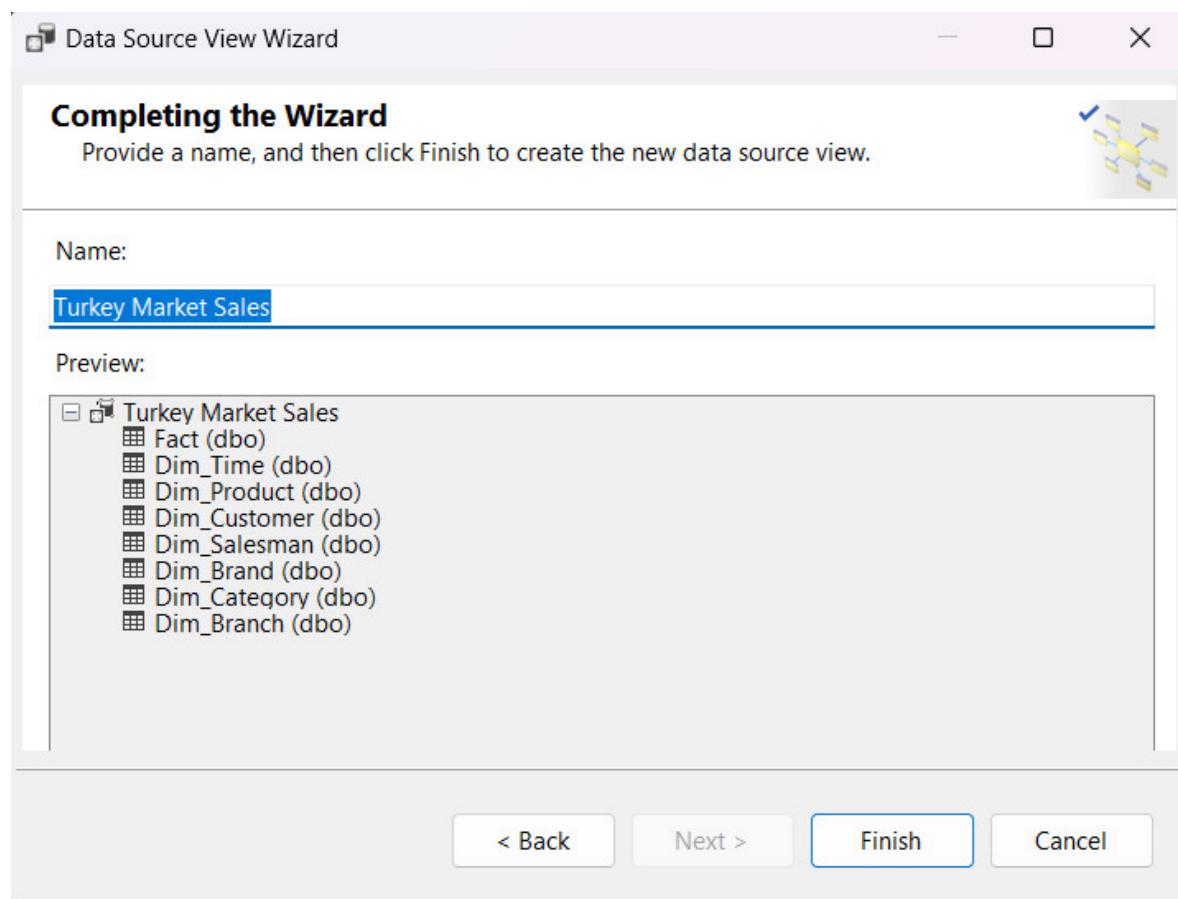
Hình 190: Thêm bảng Fact vào data source view

**Bước 5:** Tiếp theo, click “**Add Related Tables**” cho đến khi không có bảng nào được thêm vào data source view để thêm tất cả các bảng Dimension vào data source view. Sau đó click “**Next**” để tiếp tục.



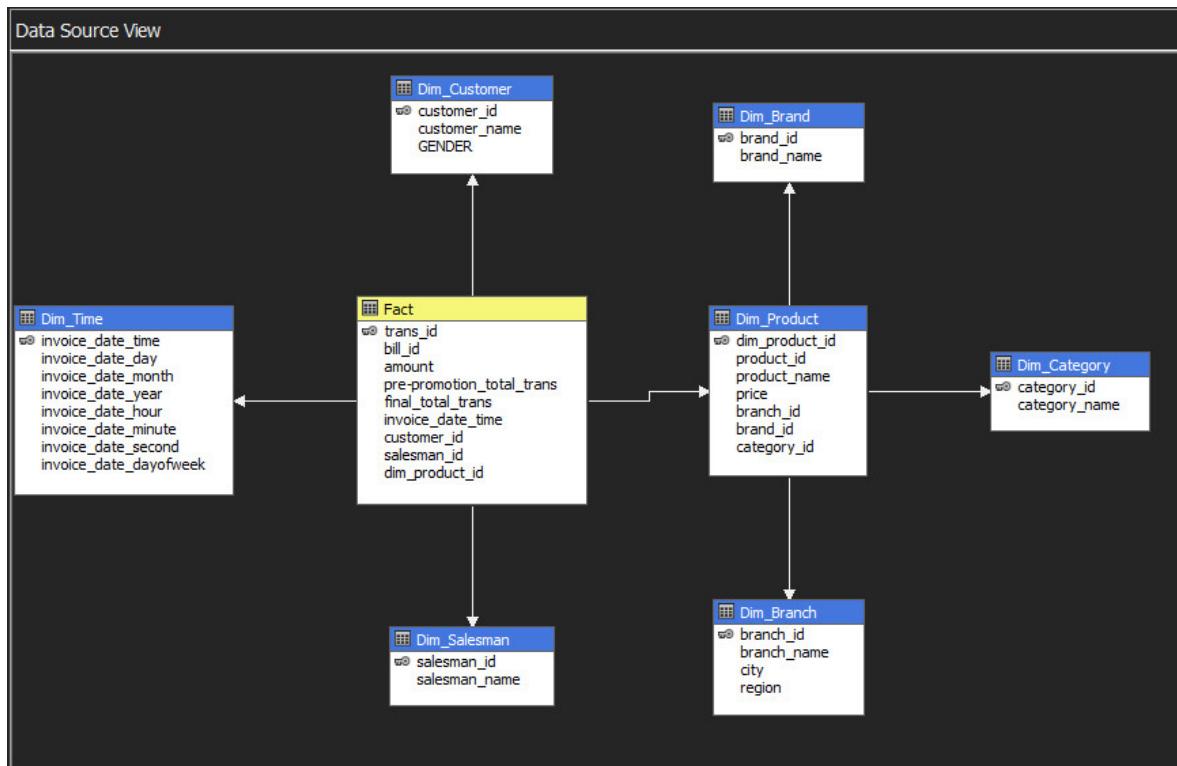
Hình 191: Thêm toàn bộ các bảng chiều còn lại vào data source view

**Bước 6:** Click “Finish” để hoàn tất quá trình xác định khung nhìn dữ liệu nguồn (Data Source View).



Hình 192: Hoàn thành quá trình tạo data source view

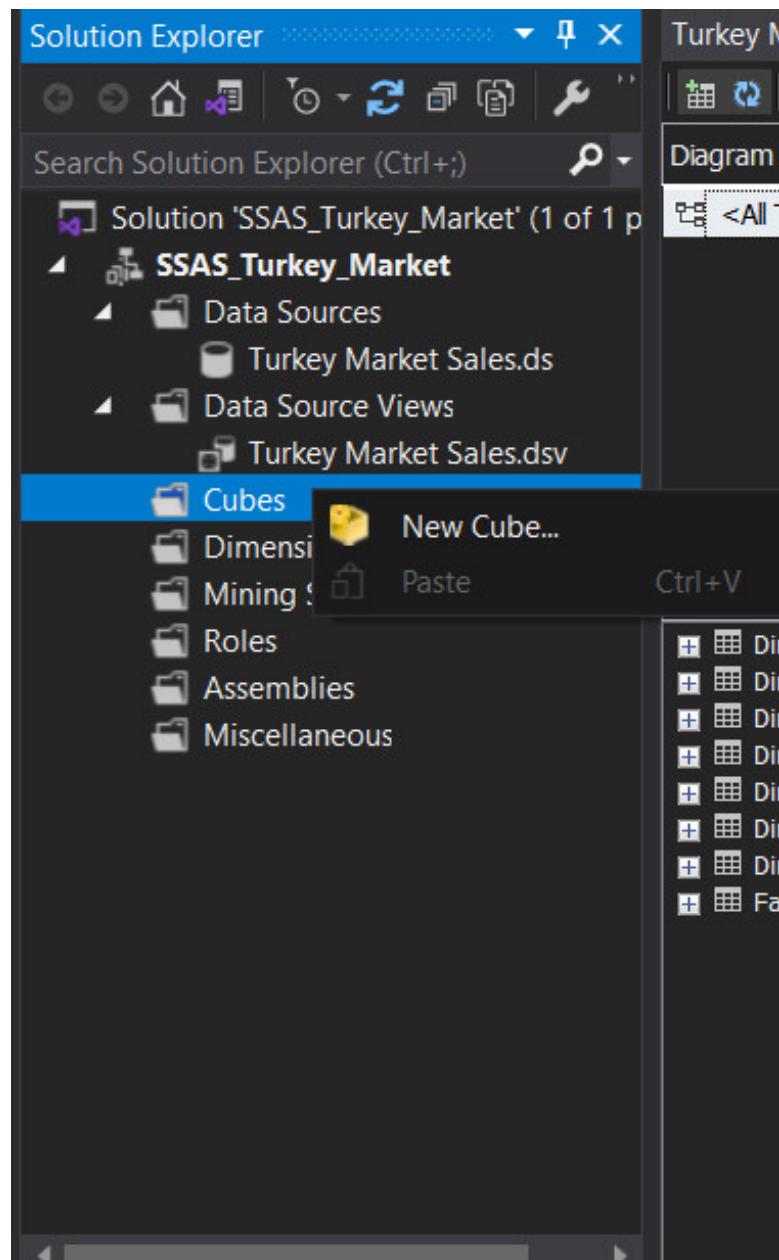
Sau khi hoàn tất quá trình này, ta thu được kết quả như sau:



Hình 193: Kết quả data source view (lược đồ)

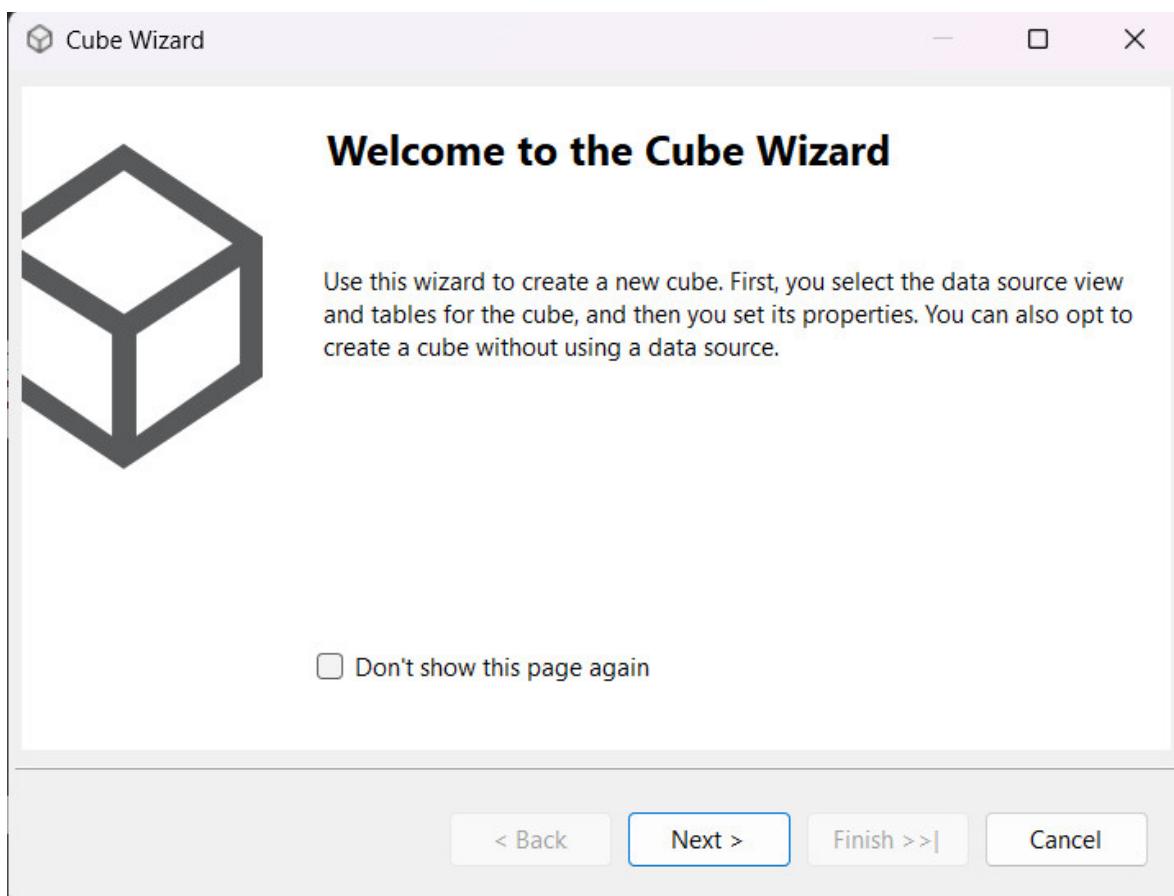
### 3.5. Xây dựng các khối (Cubes) và xác định các độ đo (Measures):

Bước 1: Tại thẻ “Solution Explorer”, click chuột phải vào thư mục “Cubes” và chọn “New Cube”



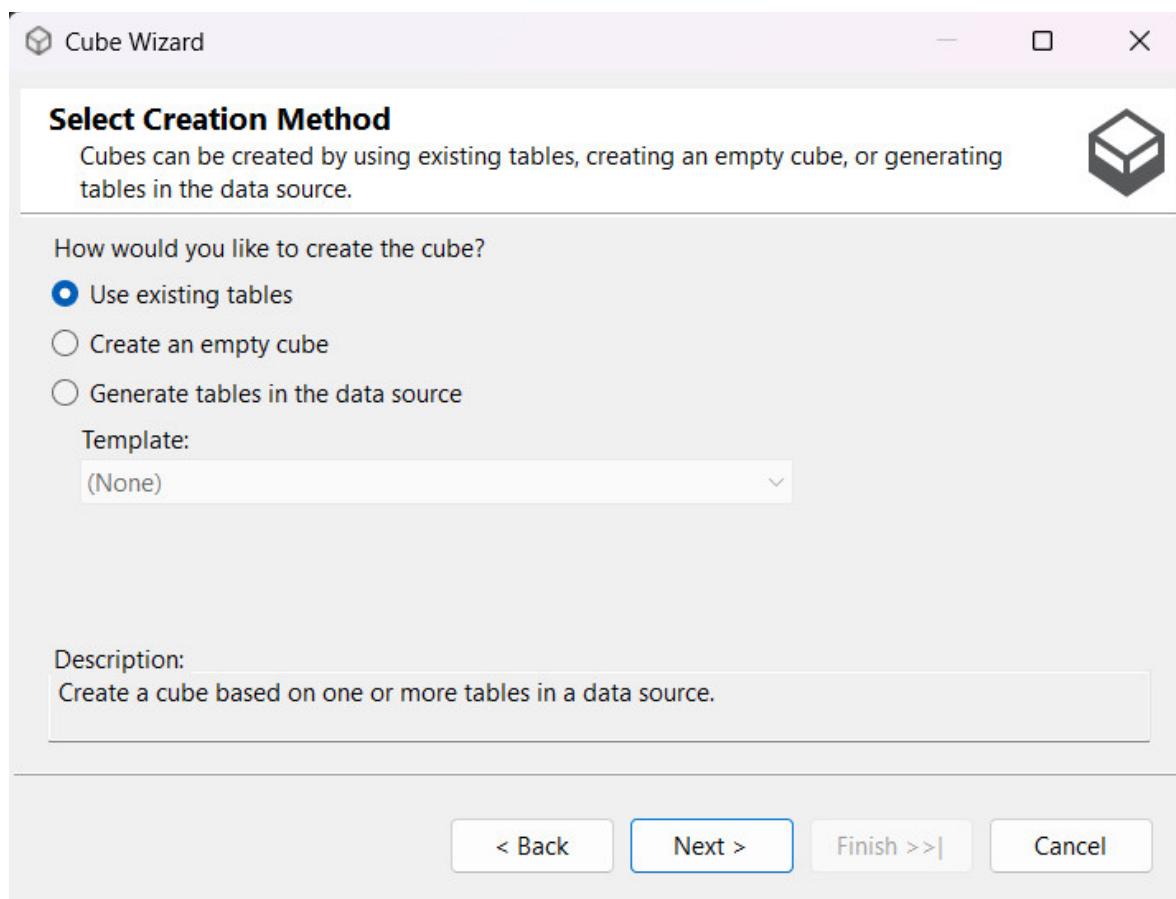
Hình 194: Giao diện tạo mới một Cube (khởi dữ liệu)

**Bước 2:** Hộp thoại “Cube Wizard” xuất hiện, click “Next” để tiếp tục



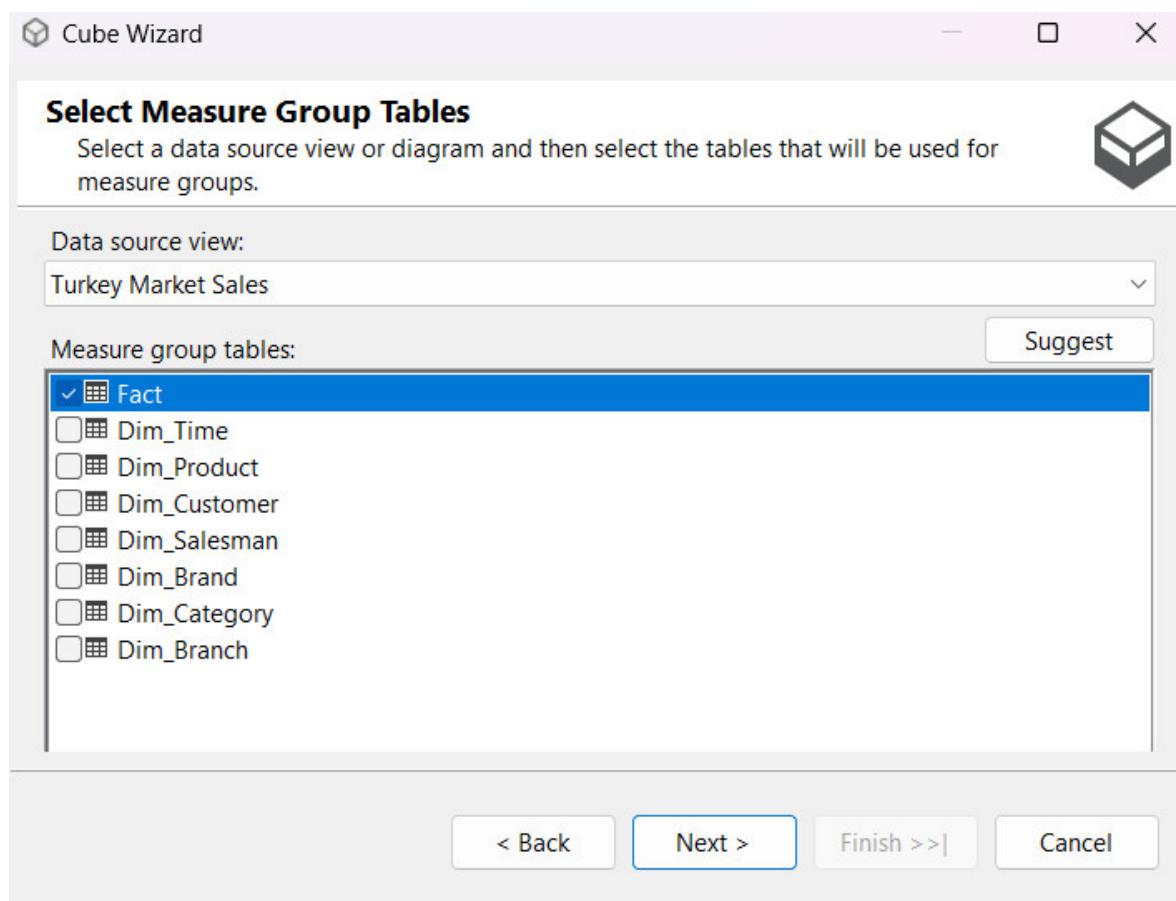
Hình 195: Hộp thoại bắt đầu quá trình tạo Cube

**Bước 3:** Chọn “Use existing table”, sau đó click “Next” để tiếp tục.



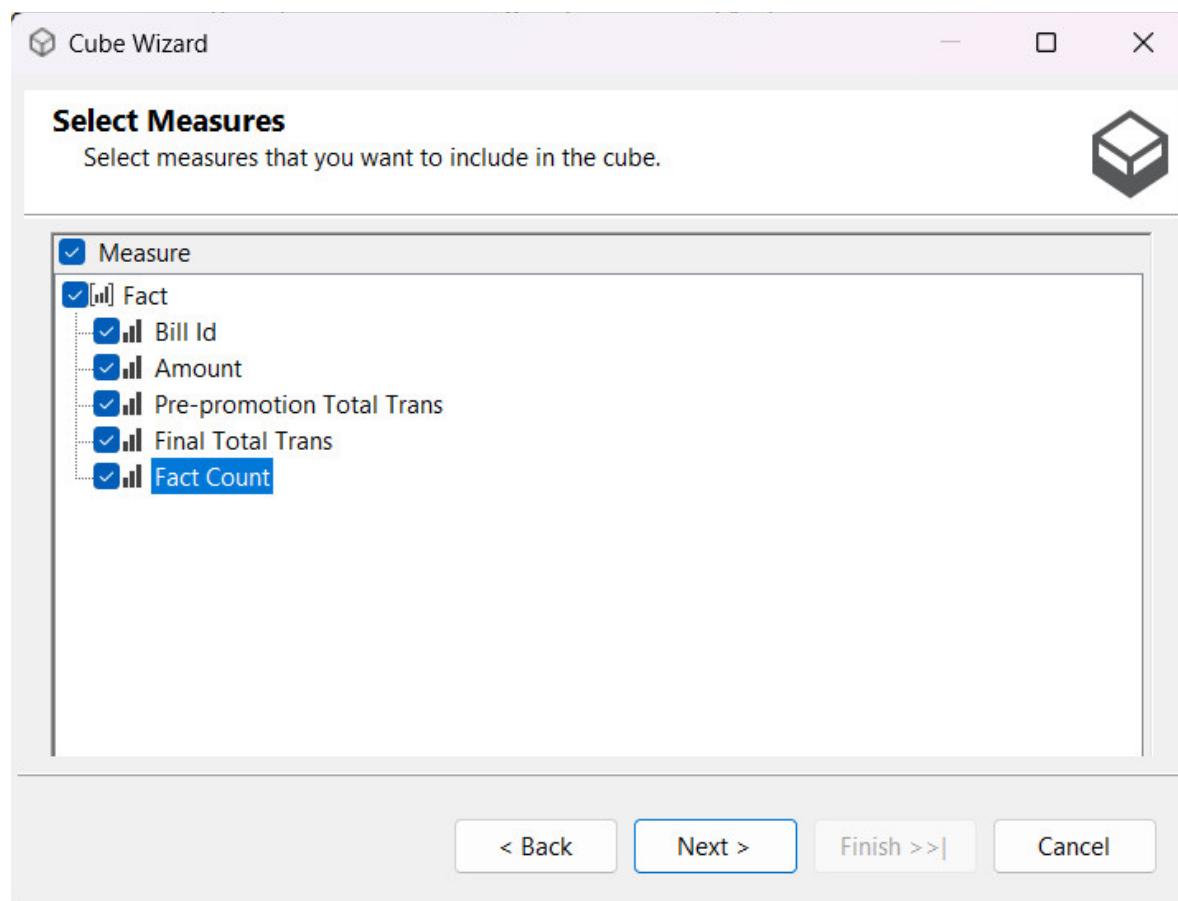
Hình 196: Giao diện chọn phương thức khởi tạo

**Bước 4:** Chọn bảng Fact để phân chia các measure group.



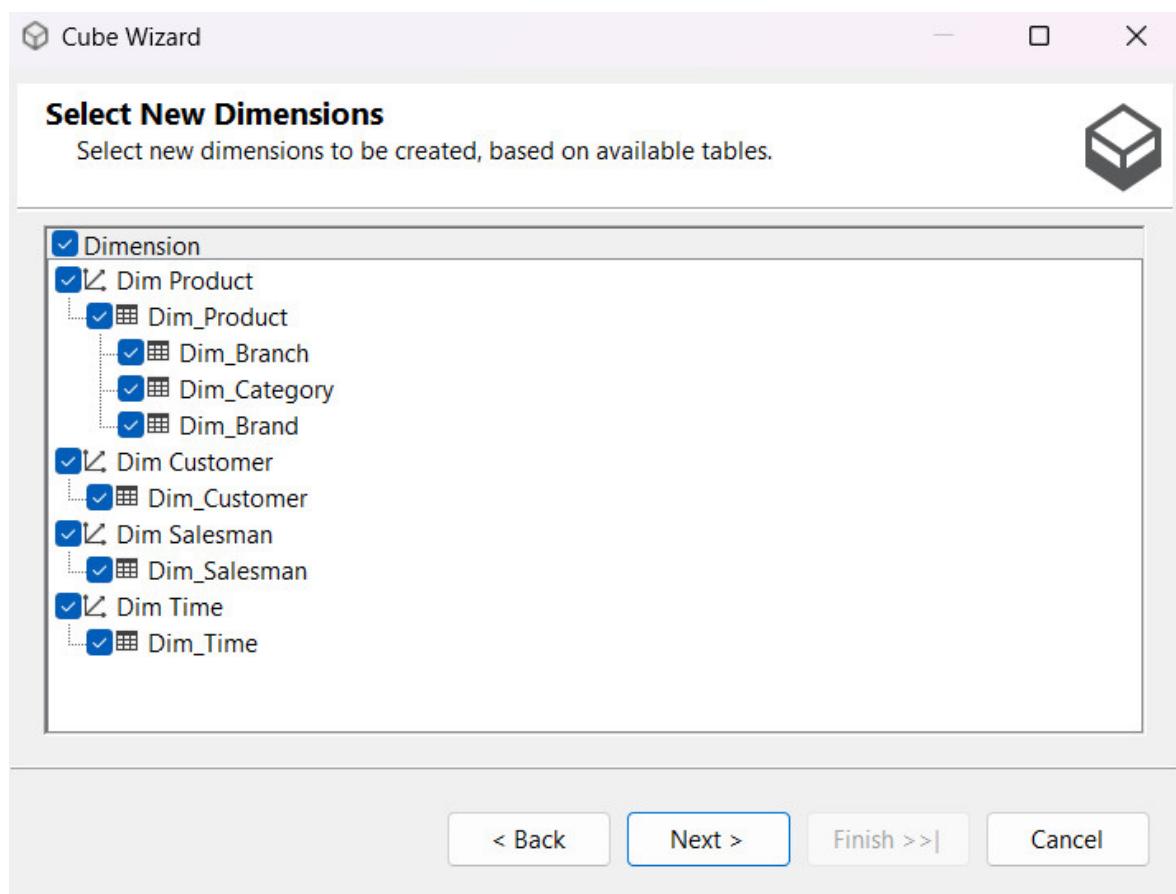
Hình 197: Giao diện phân chia các measure group

**Bước 5:** Tick chọn toàn bộ độ đo được đề xuất, sau đó click “Next” để tiếp tục.



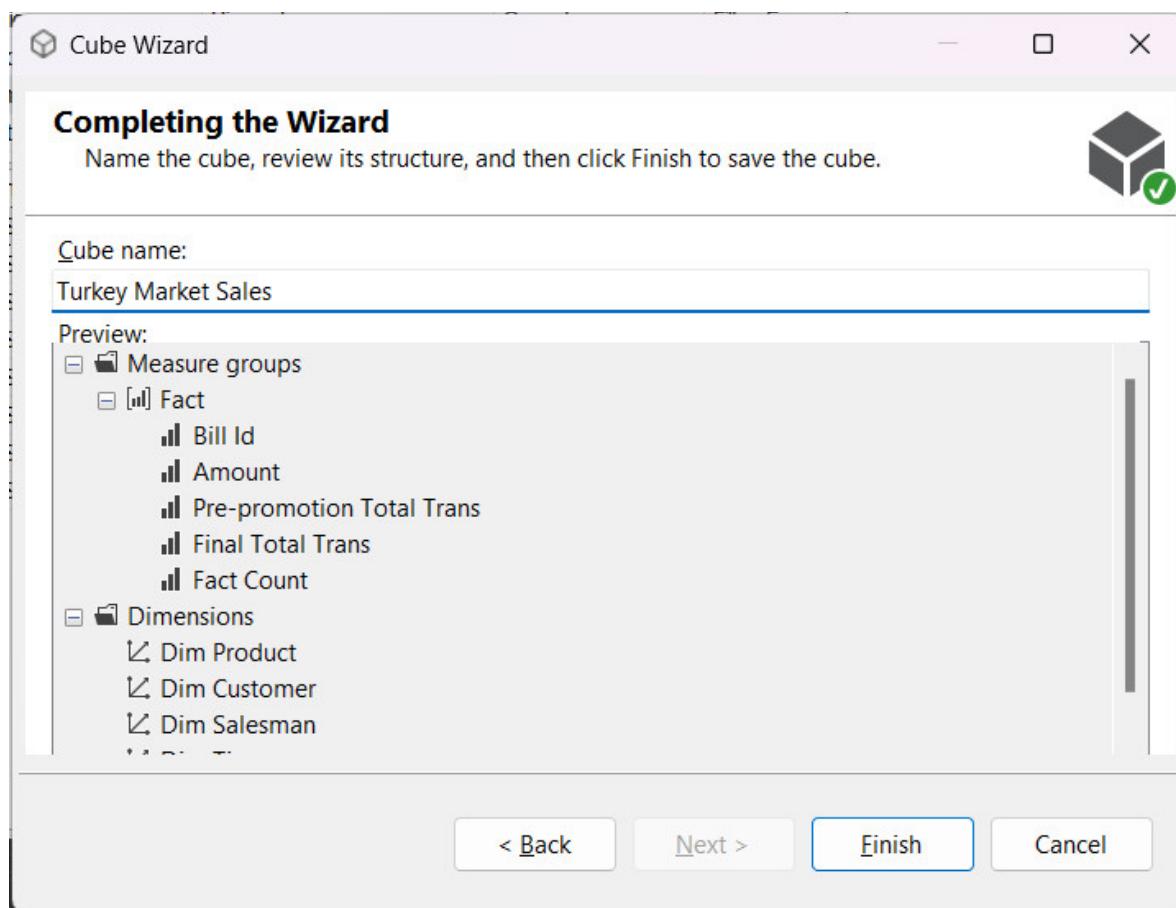
Hình 198: Giao diện để xuất một số độ đo trong bảng Fact

**Bước 6:** Chọn danh sách các bảng Dimension, sau đó click “Next” để tiếp tục.



Hình 199: Chọn các bảng Dimensions để tạo thêm chiều cho Cubes

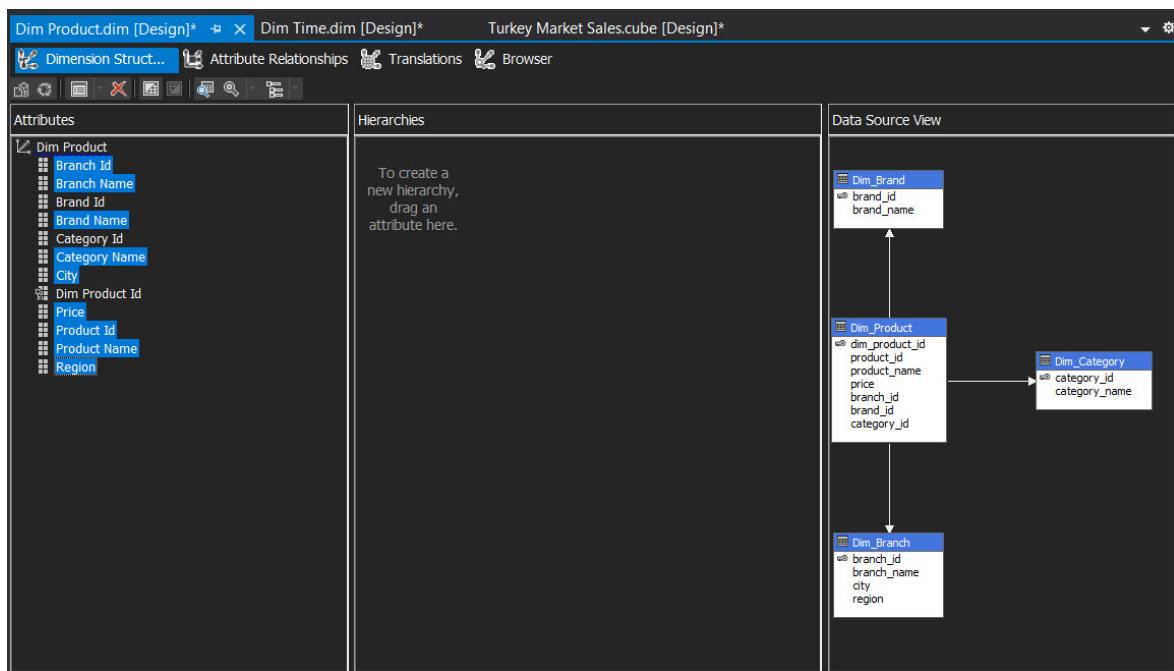
**Bước 7:** Chọn “Finish” để hoàn tất quá trình xây dựng các khối dữ liệu (Cubes) và xác định các độ đo (Measures).



Hình 200: Giao diện hoàn tất quá trình tạo Cube

### 3.6. Xác định các chiều (Dimensions):

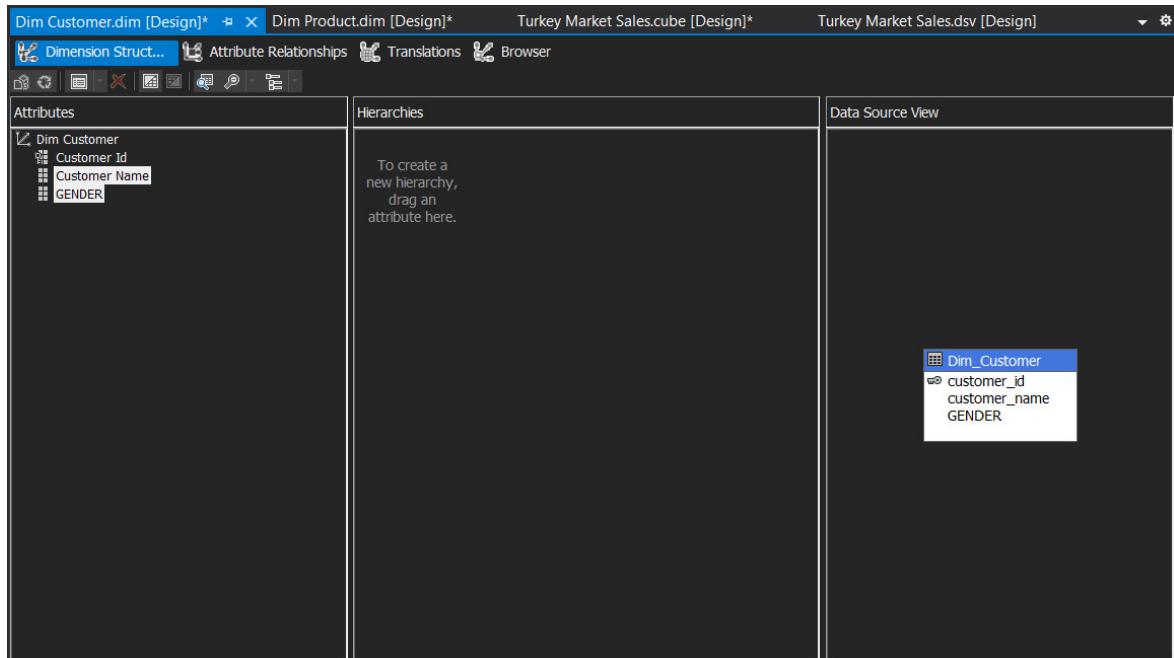
Bước 1: Tại folder Dimension trong thẻ “Solution Explorer”, ta chọn DimProduct.dim. Sau đó kéo thả các thuộc tính branch\_id, branch\_name, brand\_name, category\_name, city, region, product\_id, product\_name, price từ Data Source View vào Attributes.



Hình 201: Lựa chọn các thuộc tính đưa vào Cube (Dim\_Product)

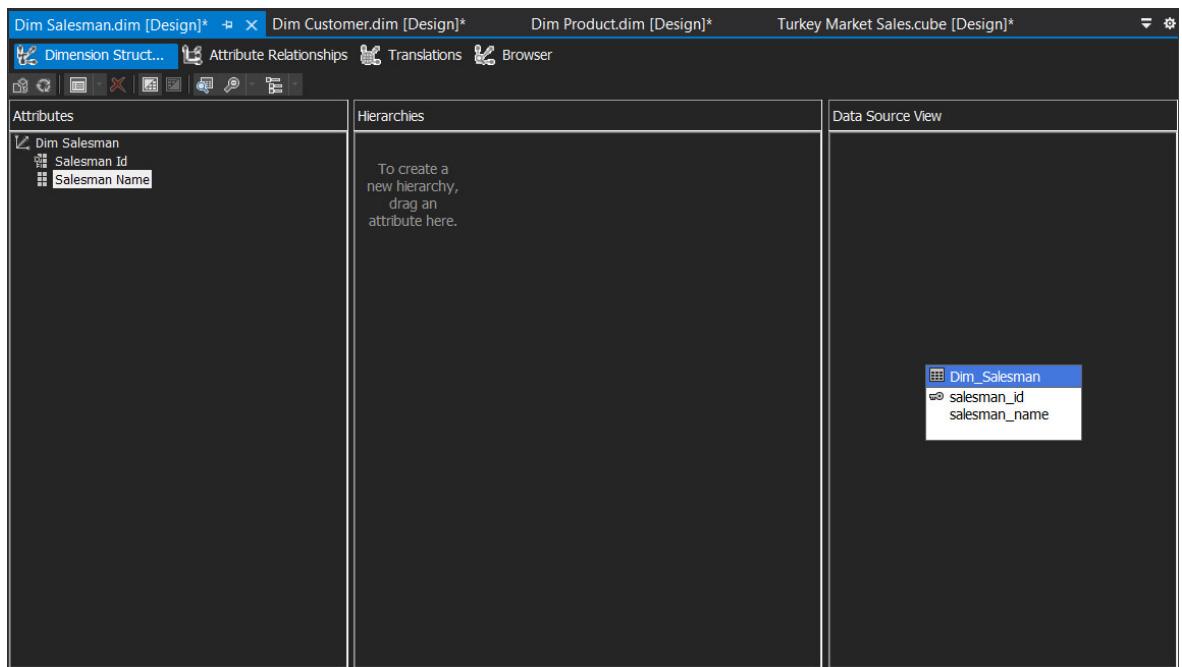
**Bước 2:** Tại folder Dimensions trong thẻ “Solution Explorer”, ta chọn

DimCustomer.dim. Sau đó, kéo thả các thuộc tính customer\_name, gender từ Data Source View vào Attributes.



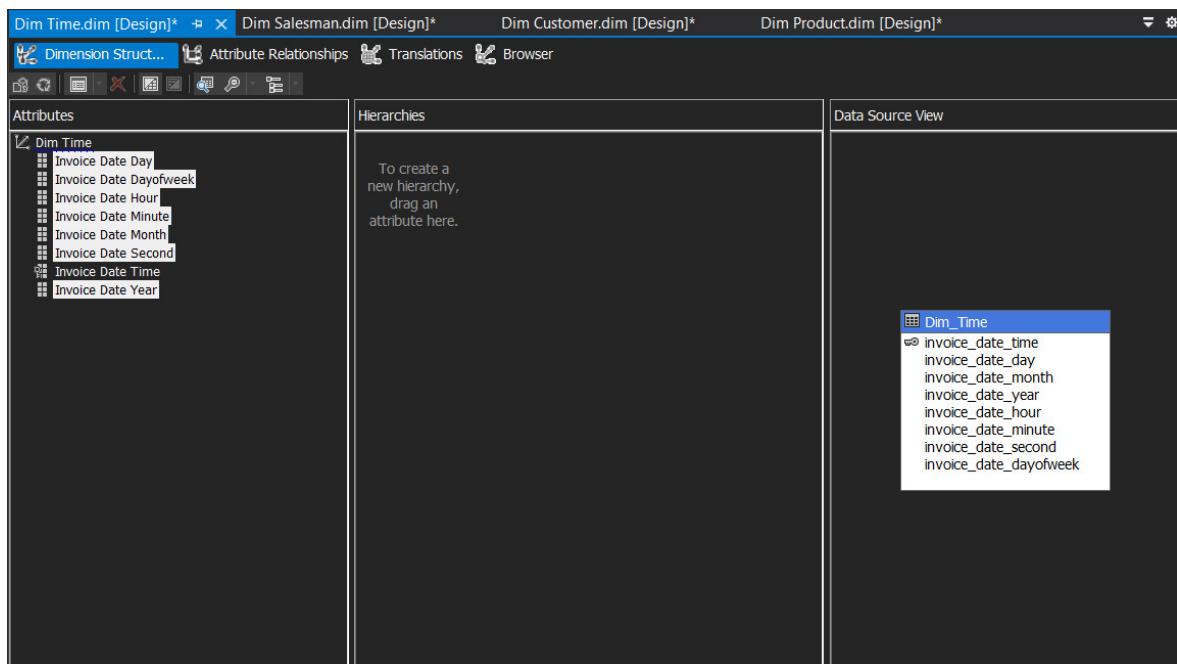
Hình 202: Lựa chọn các thuộc tính đưa vào Cube (Dim\_Customer)

**Bước 3:** Tại folder Dimensions trong thẻ “**Solution Explorer**”, ta chọn DimSalesman.dim. Sau đó, kéo thả các thuộc tính salesman\_name từ Data Source View vào Attributes



Hình 203: Lựa chọn các thuộc tính đưa vào Cube (Dim\_Salesman)

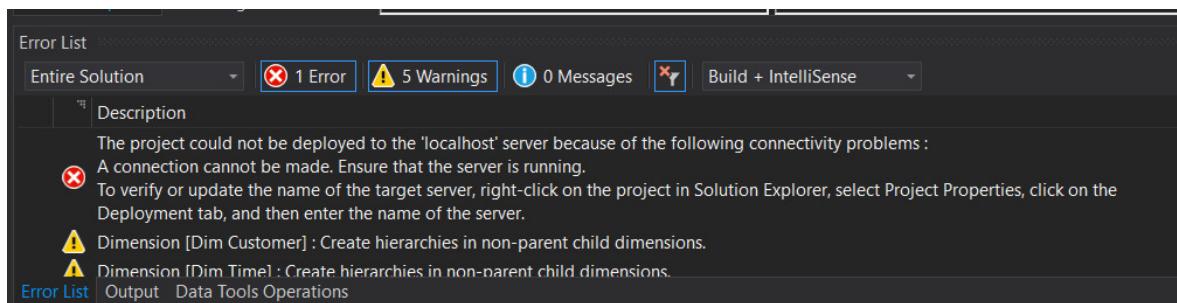
**Bước 4:** Tại folder Dimensions trong thẻ “**Solution Explorer**”, ta chọn DimTime.dim. Sau đó, kéo thả các thuộc tính invoice\_date\_year, invoice\_date\_month, invoice\_date\_day, invoice\_date\_hour, invoice\_date\_minute, invoice\_date\_second, invoice\_date\_dayofweek từ Data Source View vào Attributes



Hình 204: Lựa chọn các thuộc tính đưa vào Cube (Dim\_Time)

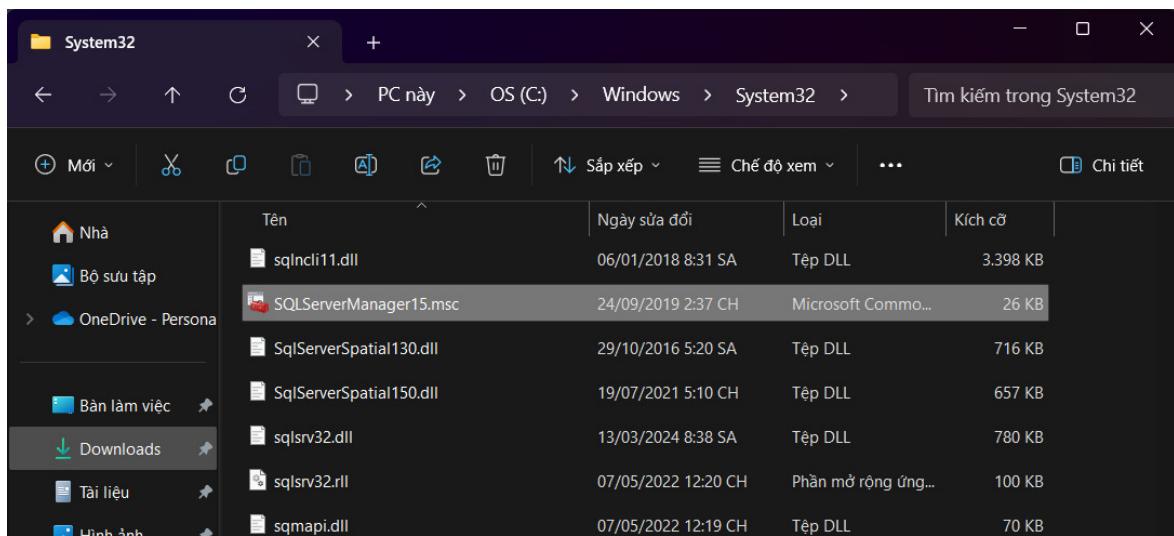
**Bước 5:** Sau khi hoàn tất, ta chọn “Start” để thực thi dự án.

Xuất hiện lỗi:



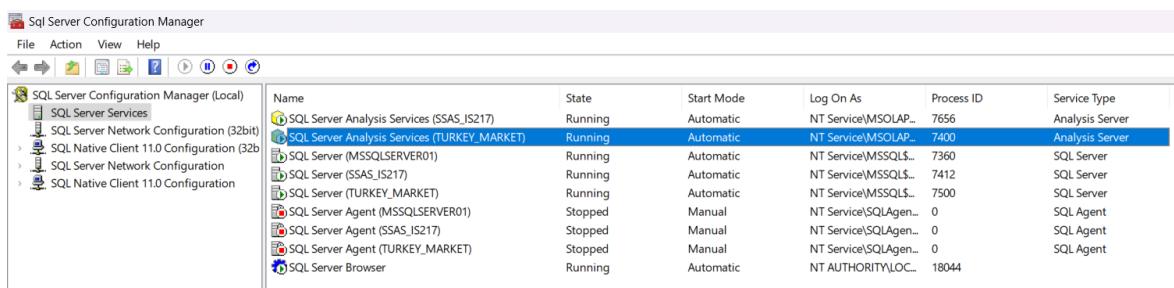
Hình 205: Hình ảnh lỗi kết nối không được thiết lập

Cách khắc phục: Truy cập vào đường dẫn “C:\Windows\System32”, tìm kiếm và mở file “SQLServerManager15.msc”



Hình 206: Vị trí chứa file SQL ServerManager15.msc

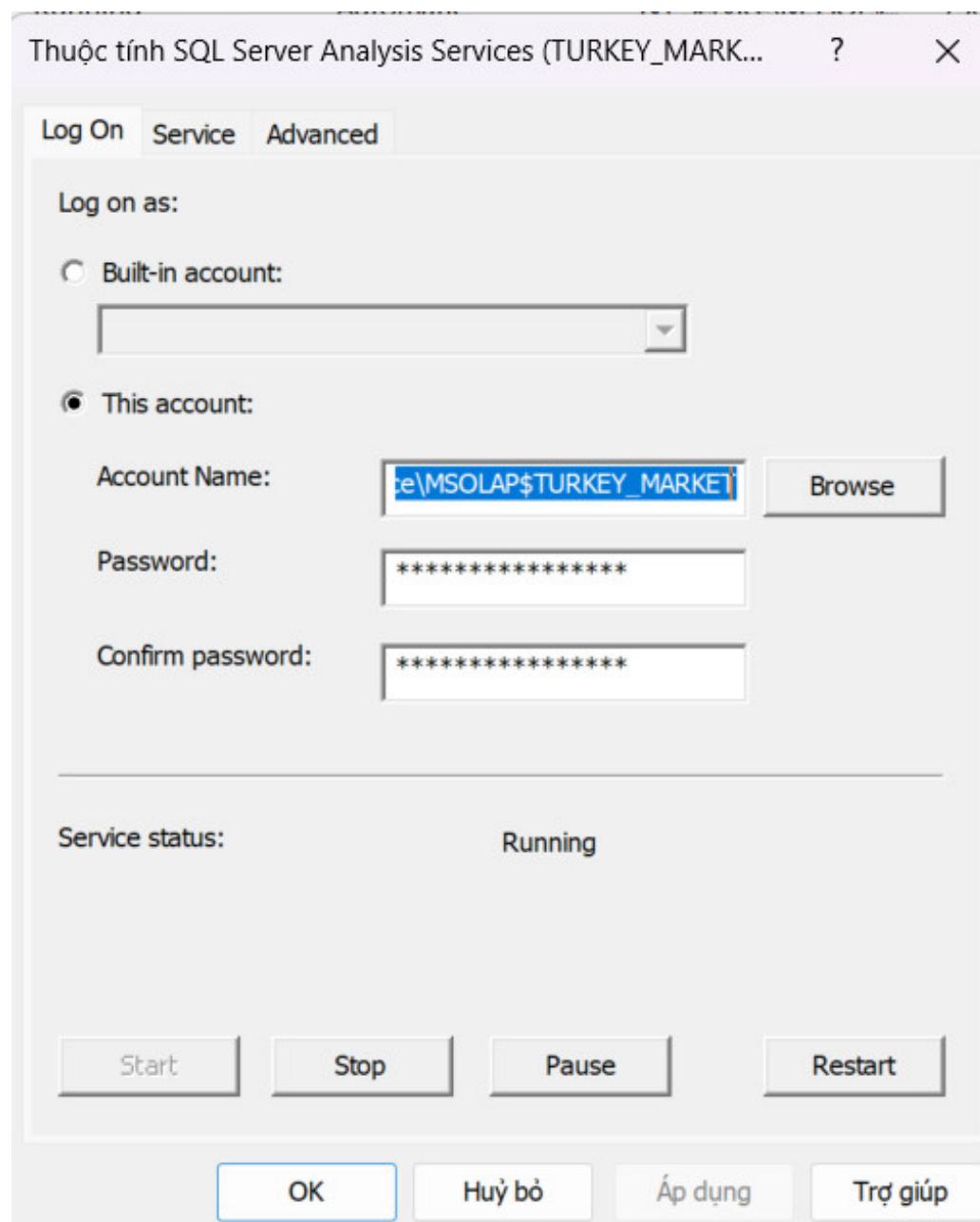
Double-click vào “"SQL Server Analysis Services (TURKEY\_MARKET)"



Hình 207: Khởi động SSAS Services

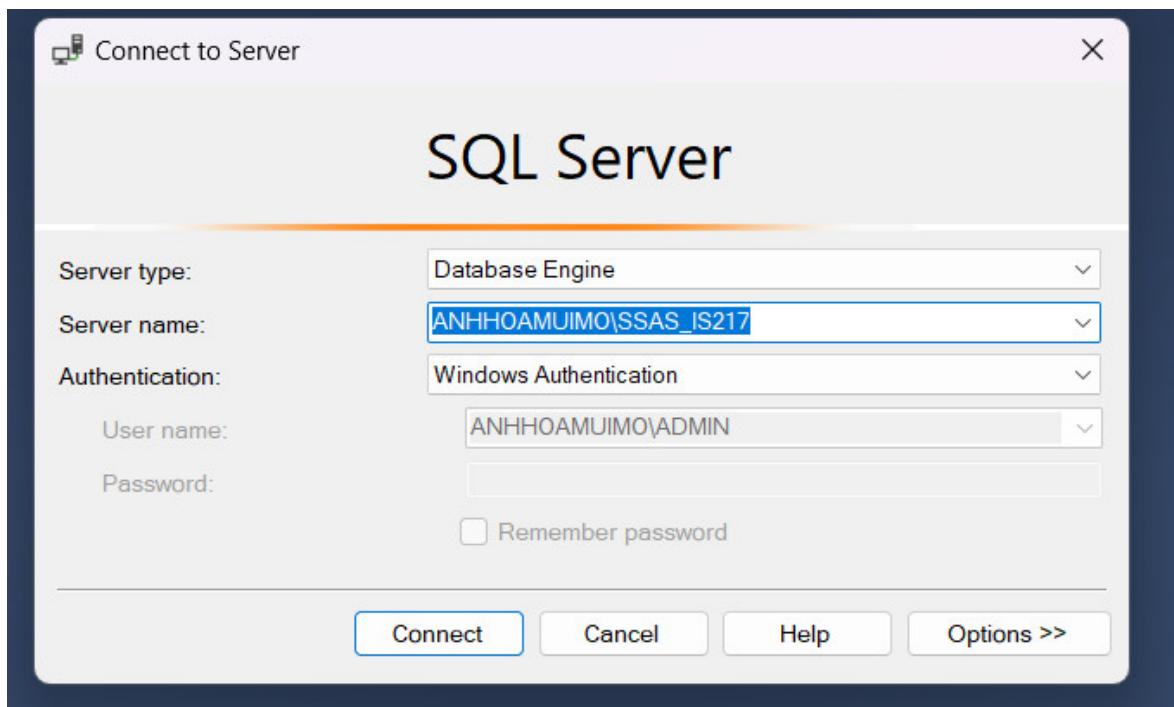
Copy toàn bộ tên tài khoản (Account Name): “NT

Service\MSOLAP\$TURKEY\_MARKET”



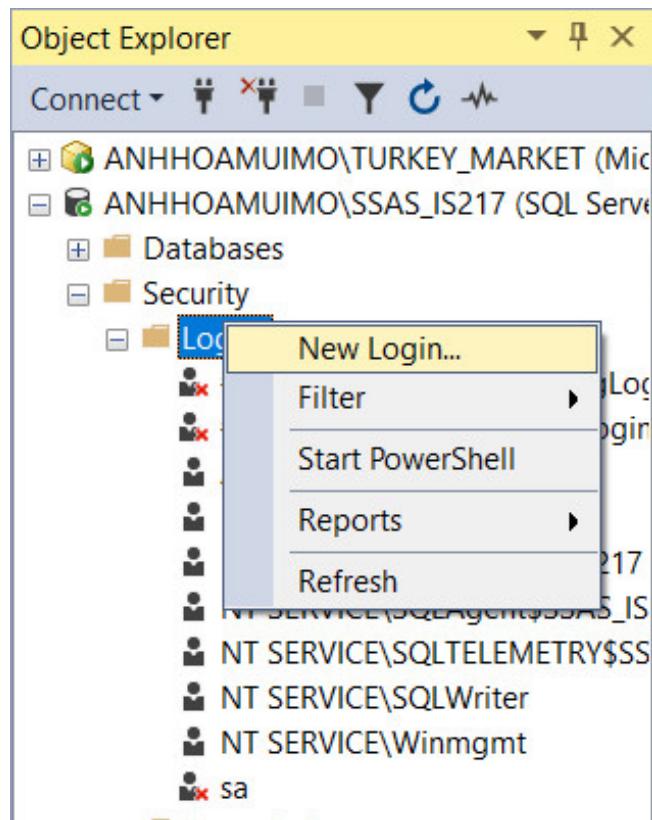
Hình 208: Giao diện hiển thị thông tin SSAS

Kết nối đến SQL Server



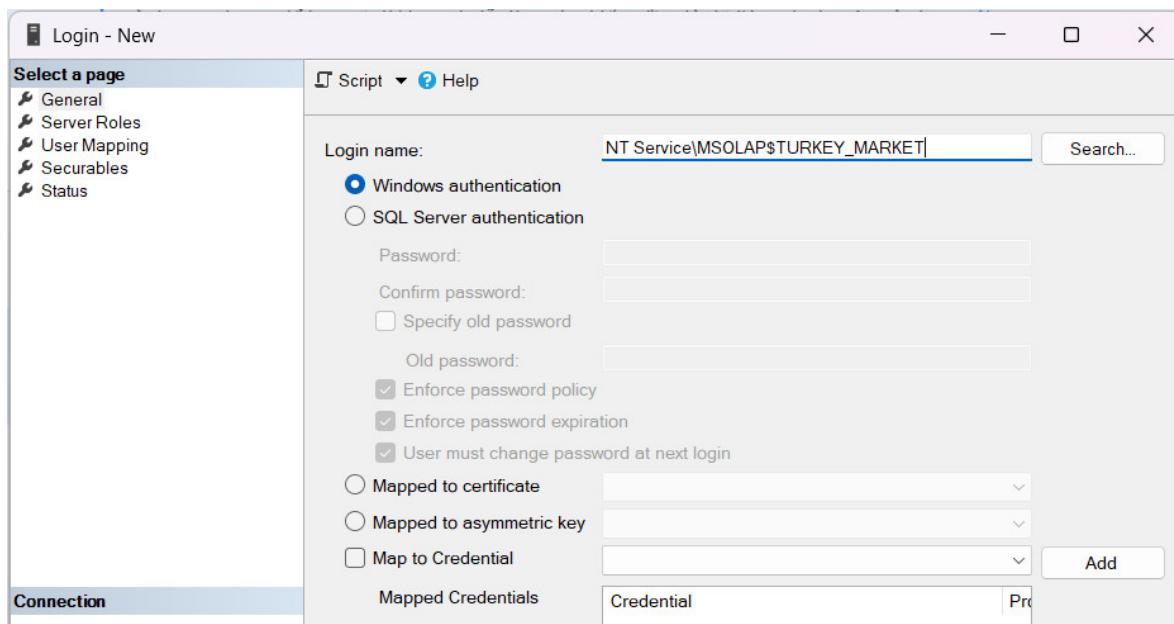
Hình 209: Kết nối vào SQL Server

Chọn “Security” -> “Login”, click chuột phải vào “Login”, chọn “New Login”



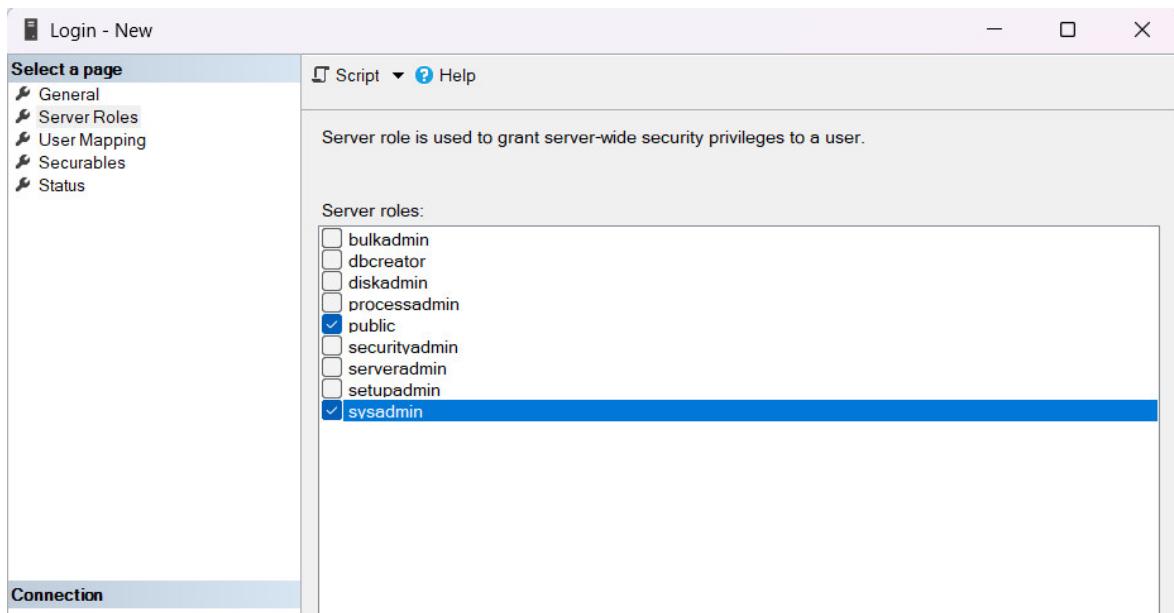
Hình 210: Tạo login mới

Lấy tên tài khoản vừa copy ở bước trên điền vào phần “Login name”



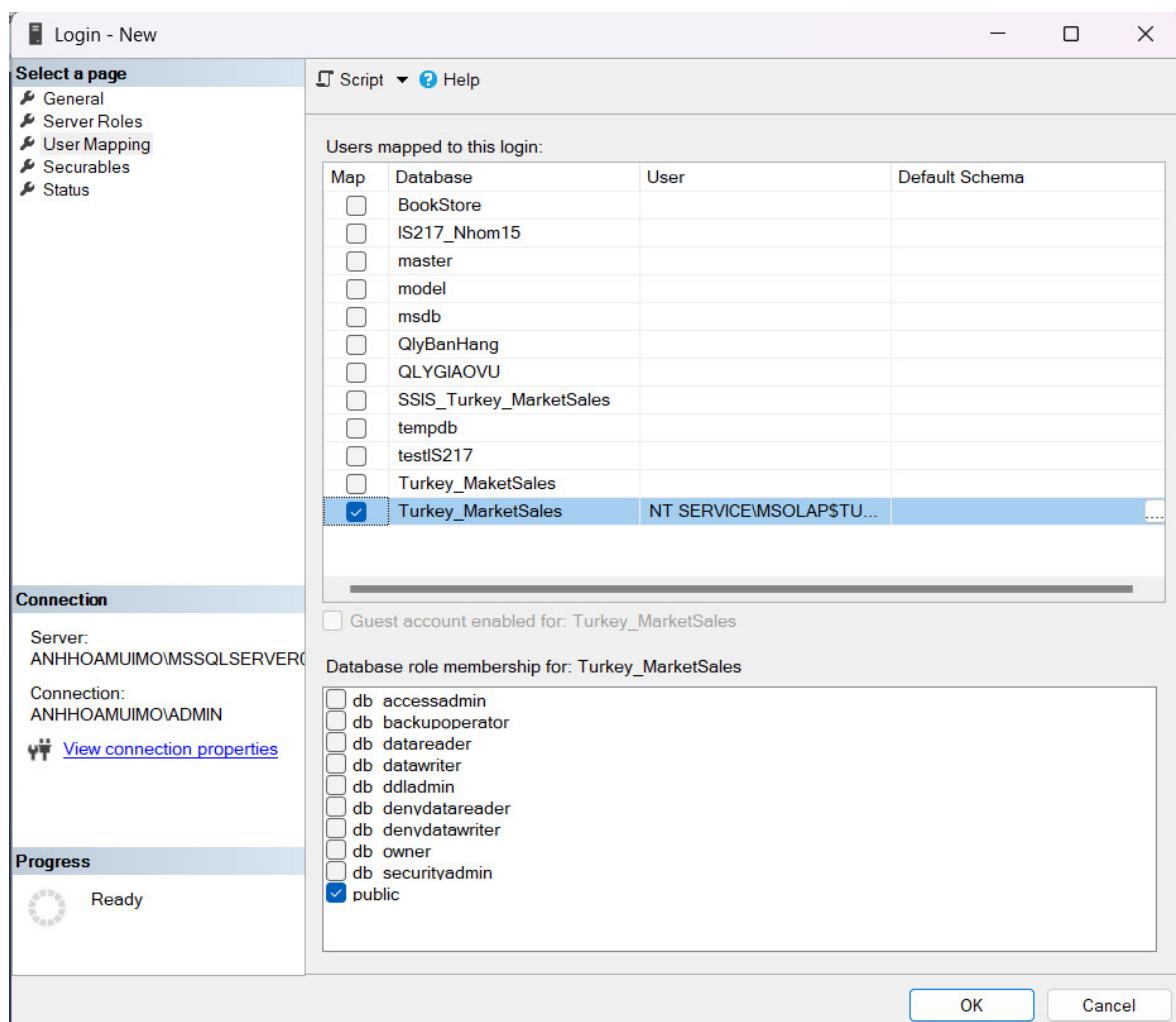
Hình 211: Giao diện tinh chỉnh các cài đặt khi tạo login

Tại tab “**Server Roles**”, click vào “**sysadmin**” để cấp toàn quyền admin cho user sắp khởi tạo



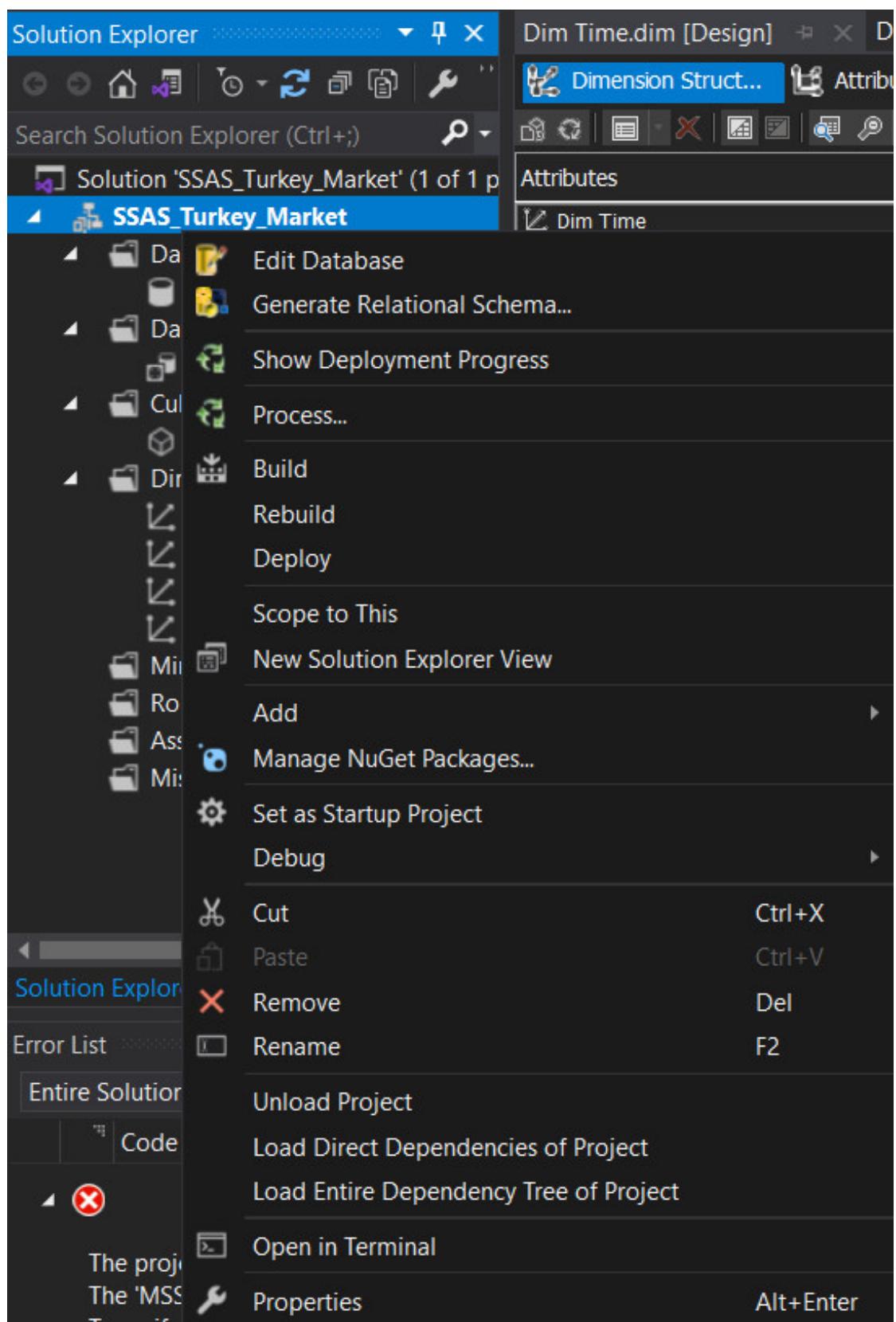
Hình 212: Giao diện cấp quyền cho login mới

Tại tab “**User Mapping**”, click chọn Turkey\_MarketSales (kho dữ liệu đang thực thi). Click “**OK**” để hoàn tất



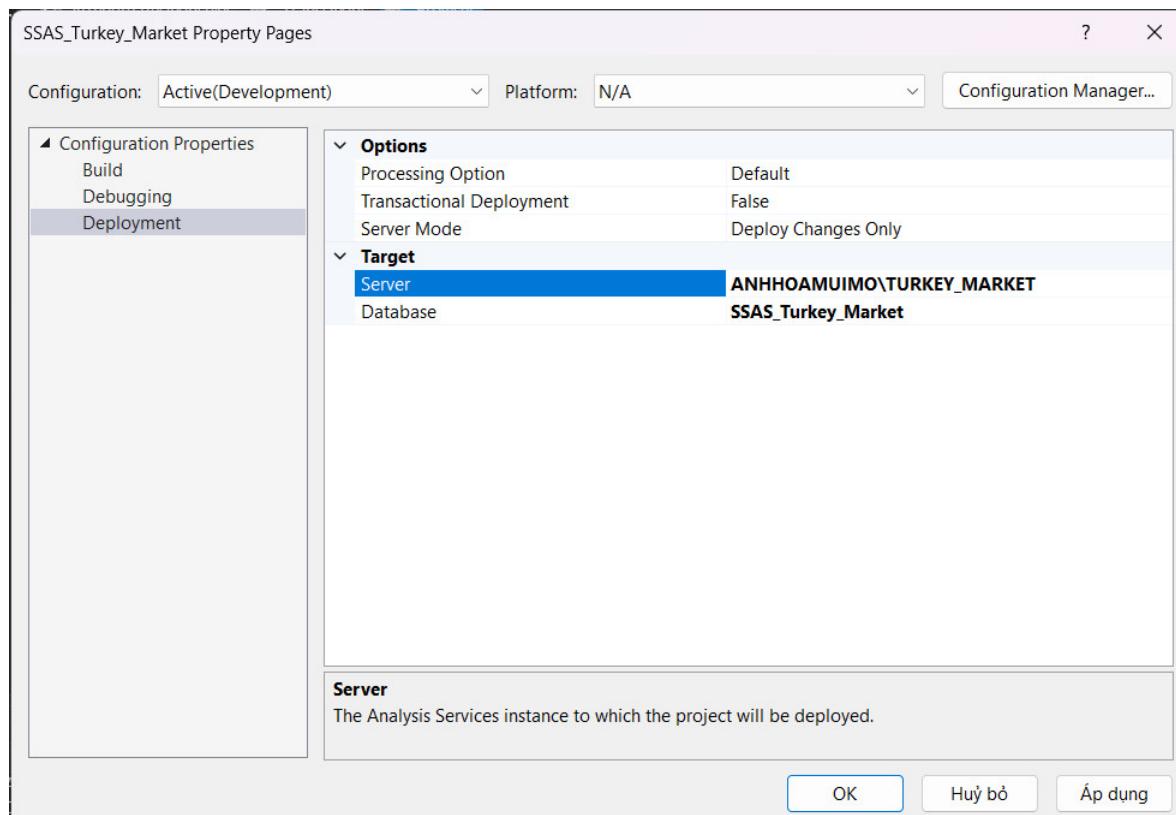
Hình 213: Trao quyền truy cập đến kho dữ liệu cho login vừa tạo

Click chuột phải vào project SSAS đang thực hiện tại thẻ “Solution Explorer”, chọn “Properties”

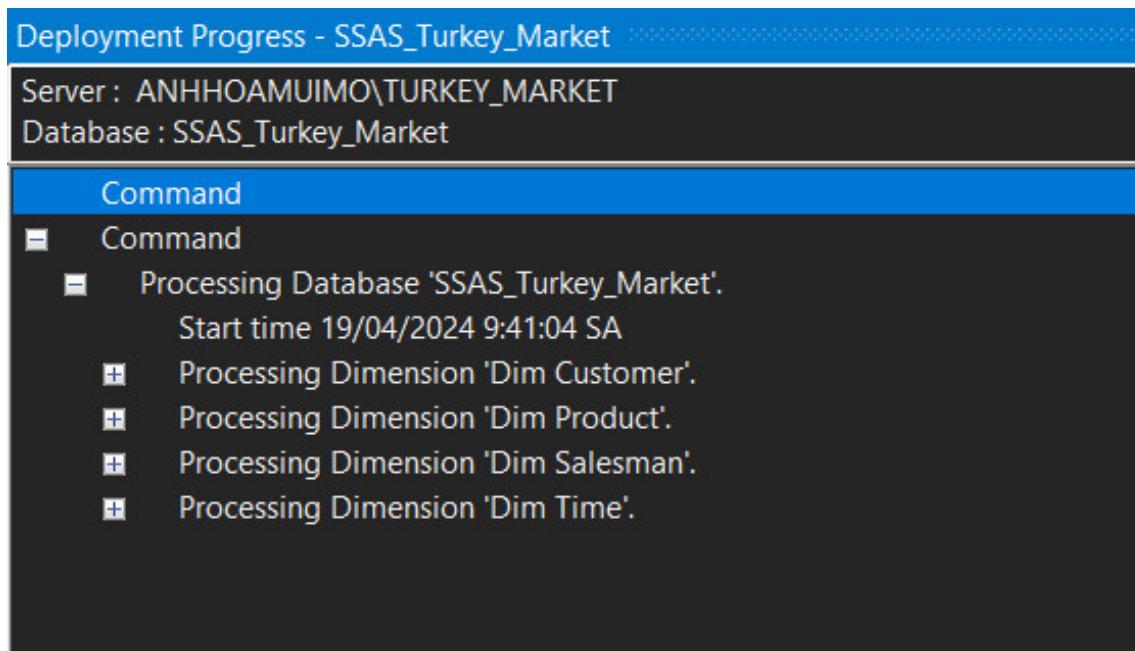


Hình 214: Điều chỉnh thông tin của dự án

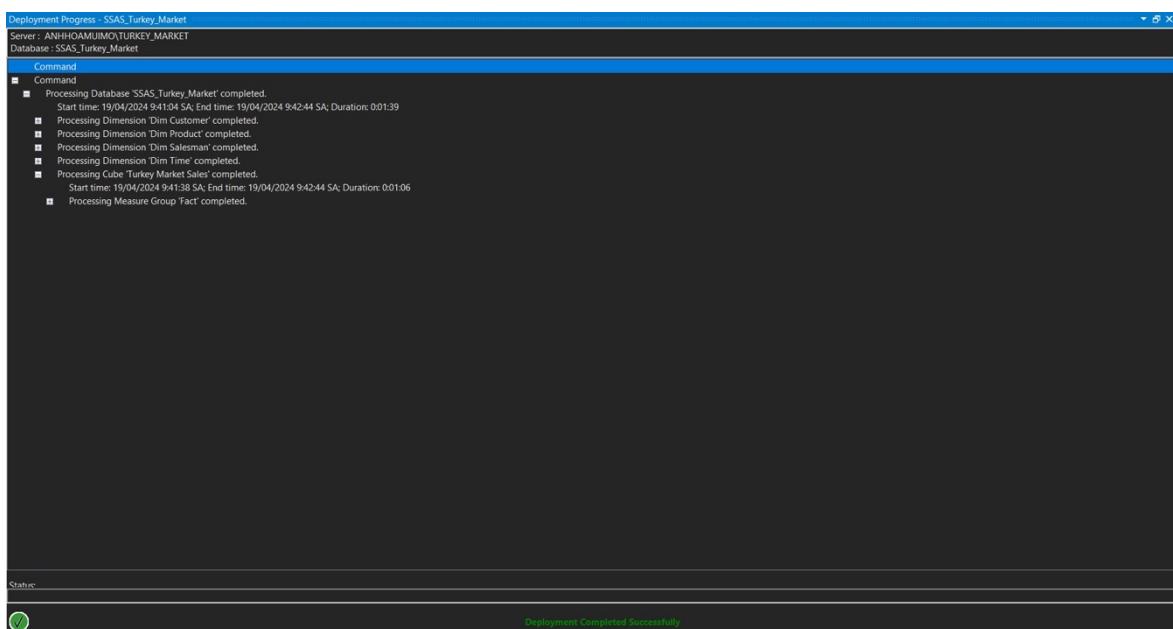
Tại thẻ “Deployment”, phần “Server” điền toàn bộ Server Name



Hình 215: Diền tên server dùng để kết nối vào SQL Server vào thẻ Deployment  
Tiến hành thực thi (deploy) lại dự án



Hình 216: Tiến hành chạy lại dự án lần nữa  
Kết quả Deploy sau khi sửa lỗi

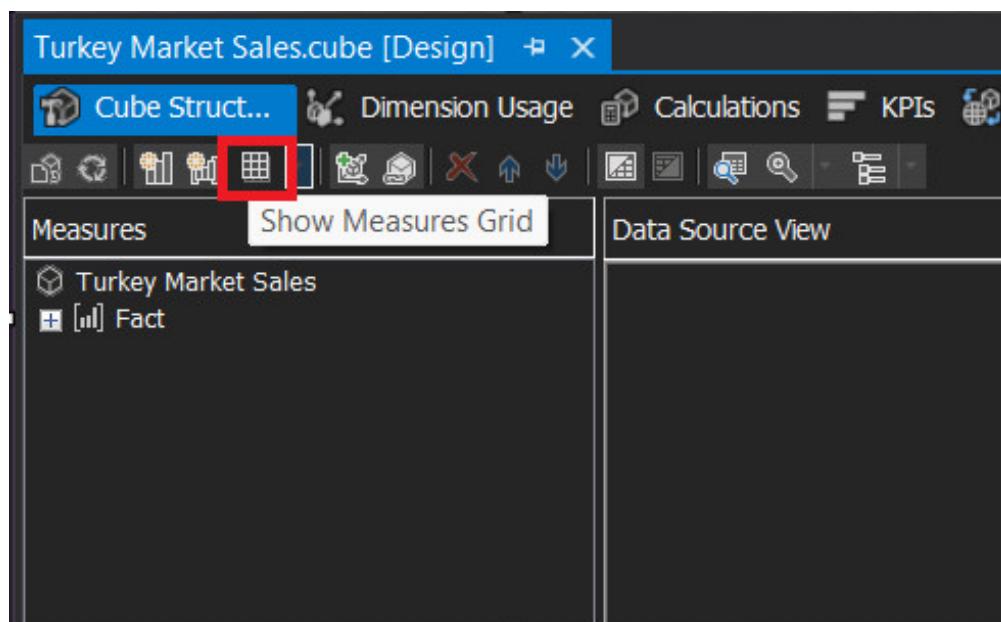


Hình 217: Dự án chạy thành công

### 3.7. Xác định các độ đo (Measures):

#### 3.7.1. Đổi tên và thuộc tính các độ đo ban đầu:

Bước 1: Tại khối vừa tạo, chọn “Show Measure Grid” để hiển thị thông tin chi tiết các độ đo được khởi tạo sẵn



Hình 218: Nút hiển thị thông tin độ đo được hệ thống gợi ý

Thông tin về các độ đo được hiển thị dưới dạng bảng, dễ dàng tương tác, chỉnh sửa cho phù hợp:

Measures				
	Name	Measure Group	Data Type	Aggregation
	Bill Id	Fact	Double	Sum
	Amount	Fact	Double	Sum
	Pre-promotion Total ...	Fact	Double	Sum
	Final Total Trans	Fact	Double	Sum
	Fact Count	Fact	Integer	Count
	Add new measure...			

Hình 219: Thông tin chi tiết về các độ đo được tạo sẵn

**Bước 2:** Ta đổi tên và thuộc tính các độ đo hiện tại theo các hàm tổng hợp aggregation như bảng sau:

Tên Độ Đo	Tên mới
Bill_id (Count)	SoLuongHoaDon
Amount(Sum)	SoLuong
Pre-promotion total trans (Sum)	GiaTruocGiamGia
Final total trans (Sum)	GiaSauGiamGia
Fact Count	FactSoLuongCTHD

Measures				
	Name	Measure Group	Data Type	Aggregation
	SoLuongHoaDon	Fact	Integer	DistinctCount
	SoLuong	Fact	Double	Sum
	GiaTruocGiamGia	Fact	Double	Sum
	GiaSauGiamGia	Fact	Double	Sum
	FactSoLuongCTHD	Fact	Integer	Count
	Add new measure...			

Hình 220: Cập nhật lại tên và kiểu dữ liệu cho phù hợp

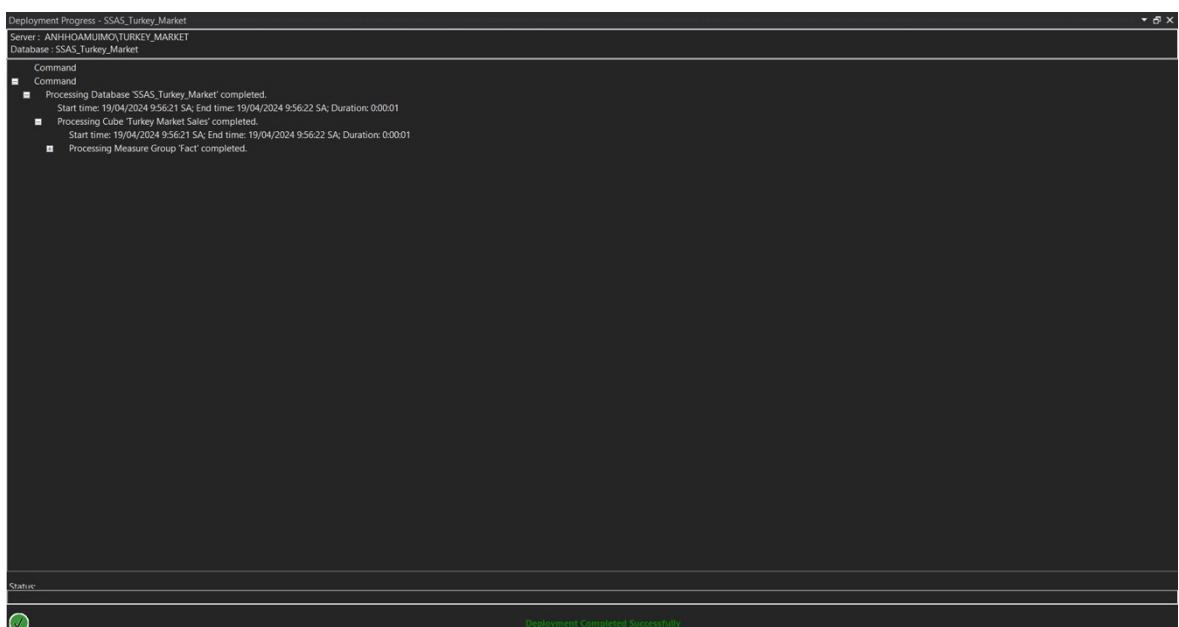
### 3.7.2. Tạo các độ đo mới:

Để tạo ra các độ đo mới, ta chọn “Add new Measure...”, sau đó tạo thêm một độ đo mới SoLuongMuaNhiuNhat bằng cách sử dụng aggregation (max)

Measures				
	Name	Measure Group	Data Type	Aggregation
SoLuongHoaDon	Fact	Integer	DistinctCount	
SoLuong	Fact	Double	Sum	
GiaTruocGiamGia	Fact	Double	Sum	
GiaSauGiamGia	Fact	Double	Sum	
FactSoLuongCTHD	Fact	Integer	Count	
SoLuongMuaNhiuNhat	Fact	Double	Max	
Add new measure...				

Hình 221: Tạo độ đo mới (SoLuongMuaNhiuNhat) bằng cách sử dụng hàm Max

Cuối cùng, ta thực thi (deploy) lại project để đảm bảo không có lỗi xảy ra



Hình 222: Thực thi lại dự án, kết quả chạy thành công

### 3.8. Phân cấp trong bảng chiều:

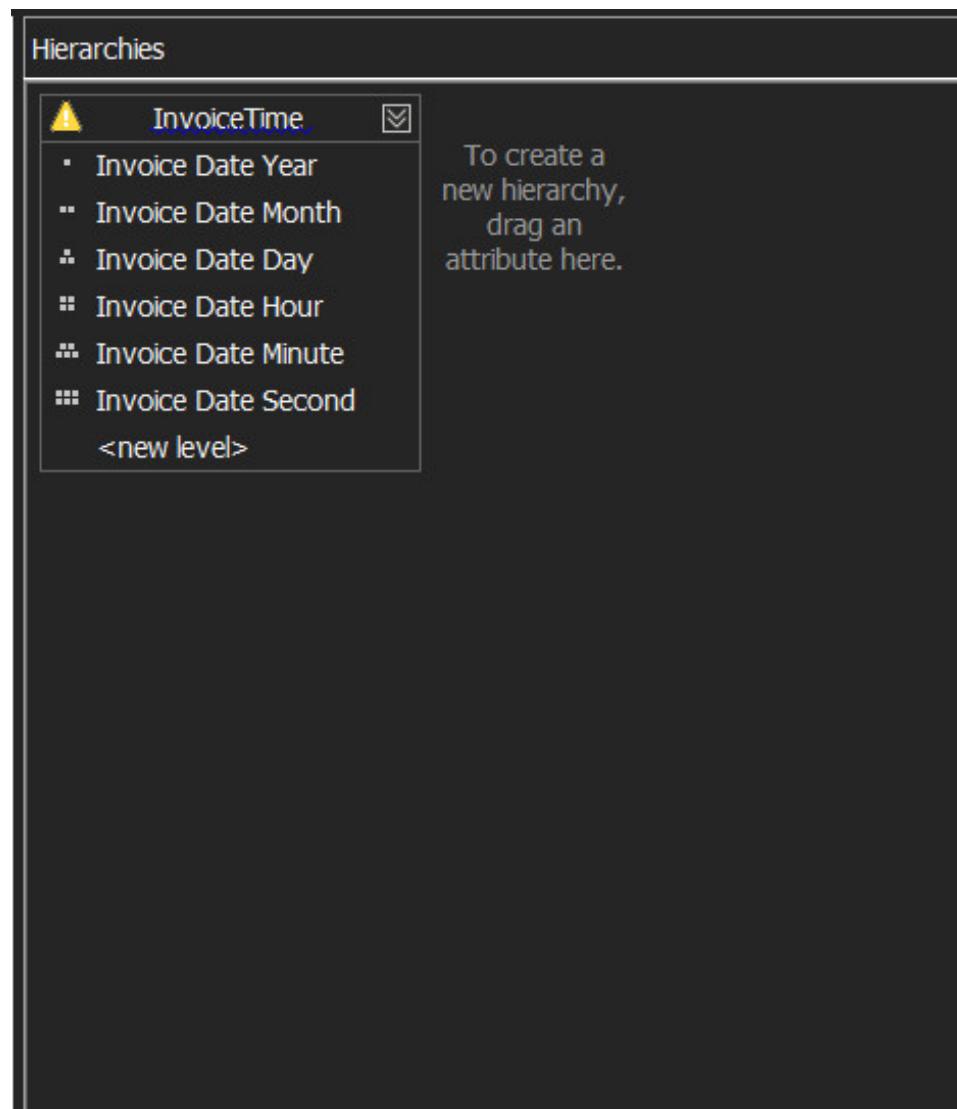
#### 3.8.1. Phân cấp bảng Dim\_Time:

**Bước 1:** Kéo toàn bộ những thuộc tính cần phân cấp qua cửa sổ “Hierarchies”

Attributes	Hierarchies	Data Source View
<p>Dim Time</p> <ul style="list-style-type: none"> <li>Invoice Date Day</li> <li>Invoice Date Dayofweek</li> <li>Invoice Date Hour</li> <li>Invoice Date Minute</li> <li>Invoice Date Month</li> <li>Invoice Date Second</li> <li>Invoice Date Time</li> <li>Invoice Date Year</li> </ul>	<p>Hierarchy</p> <ul style="list-style-type: none"> <li>Invoice Date Hour</li> <li>Invoice Date Second</li> <li>Invoice Date Minute</li> <li>Invoice Date Year</li> <li>Invoice Date Day</li> <li>Invoice Date Month</li> <li>&lt;new level&gt;</li> </ul> <p>To create a new hierarchy, drag an attribute here.</p>	<p>Dim_Time</p> <ul style="list-style-type: none"> <li>invoice_date_time</li> <li>invoice_date_day</li> <li>invoice_date_month</li> <li>invoice_date_year</li> <li>invoice_date_hour</li> <li>invoice_date_minute</li> <li>invoice_date_second</li> <li>invoice_date_dayofweek</li> </ul>

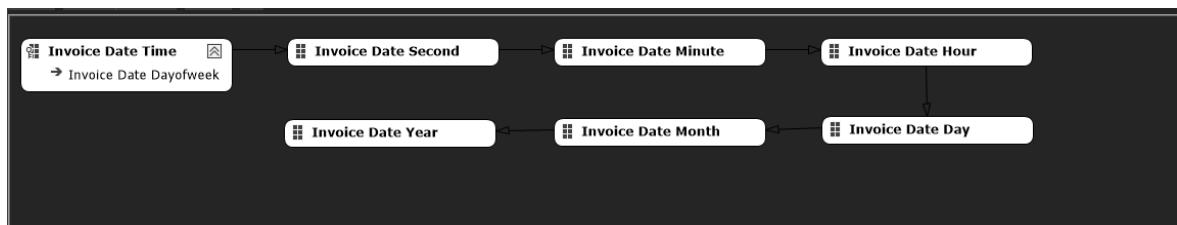
Hình 223: Đưa toàn bộ thuộc tính phân cấp của bảng Dim\_Time qua cửa sổ Hierarchies

**Bước 2:** Sắp xếp lại các thuộc tính phân cấp theo thứ tự: invoice\_date\_year, invoice\_date\_month, invoice\_date\_day, invoice\_date\_hour, invoice\_date\_minute, invoice\_date\_second và đổi tên Hierarchy thành InvoiceTime.



Hình 224: Cửa sổ sắp xếp thứ tự của các thuộc tính

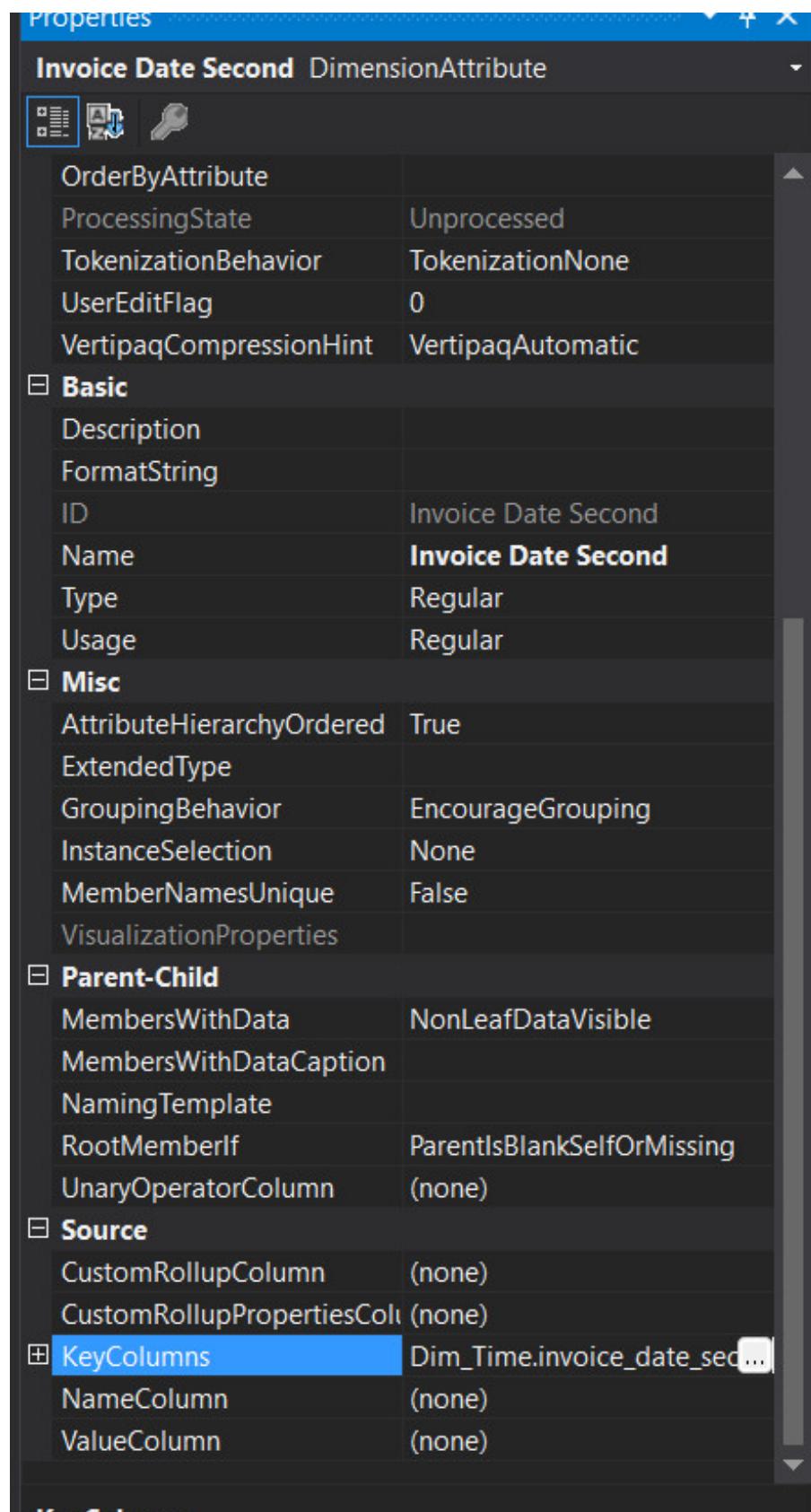
**Bước 3:** Tại panel “Attribute Relationships”, chuột phải vào các mũi tên chọn “Delete” và tạo lại mối quan hệ như sau:



Hình 225: Tạo lại mối quan hệ giữa các thuộc tính cho phù hợp

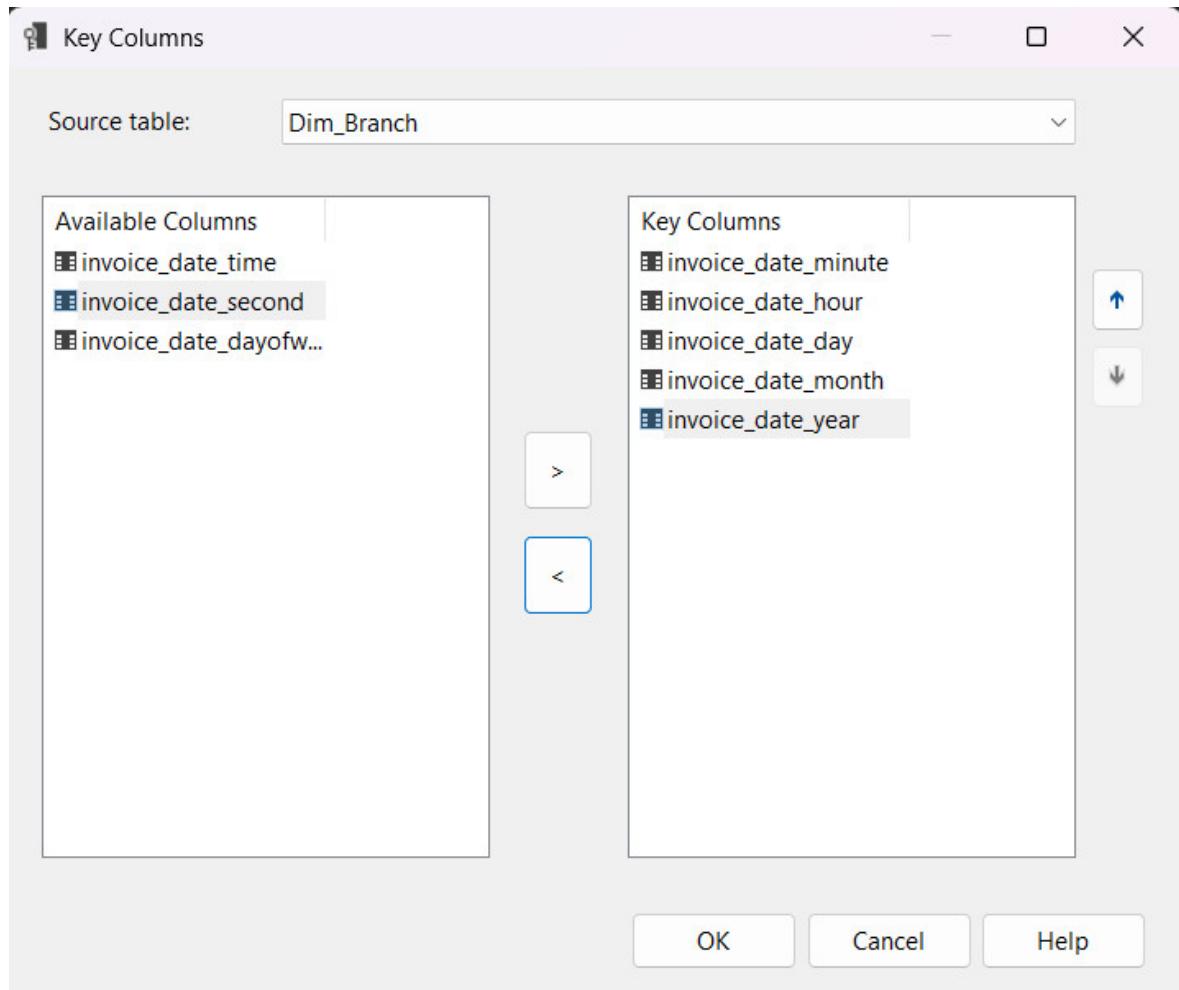
**Bước 4:** Chính khóa cột “KeyColumns” và tên cột “Name Column” của thuộc tính Invoice Date Second. Vì thuộc tính Invoice Date Second là thuộc tính cấp nhỏ nhất sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

Tại cửa sổ “**Properties**” của thuộc tính Invoice Date Second, chọn “**KeyColumns**”.



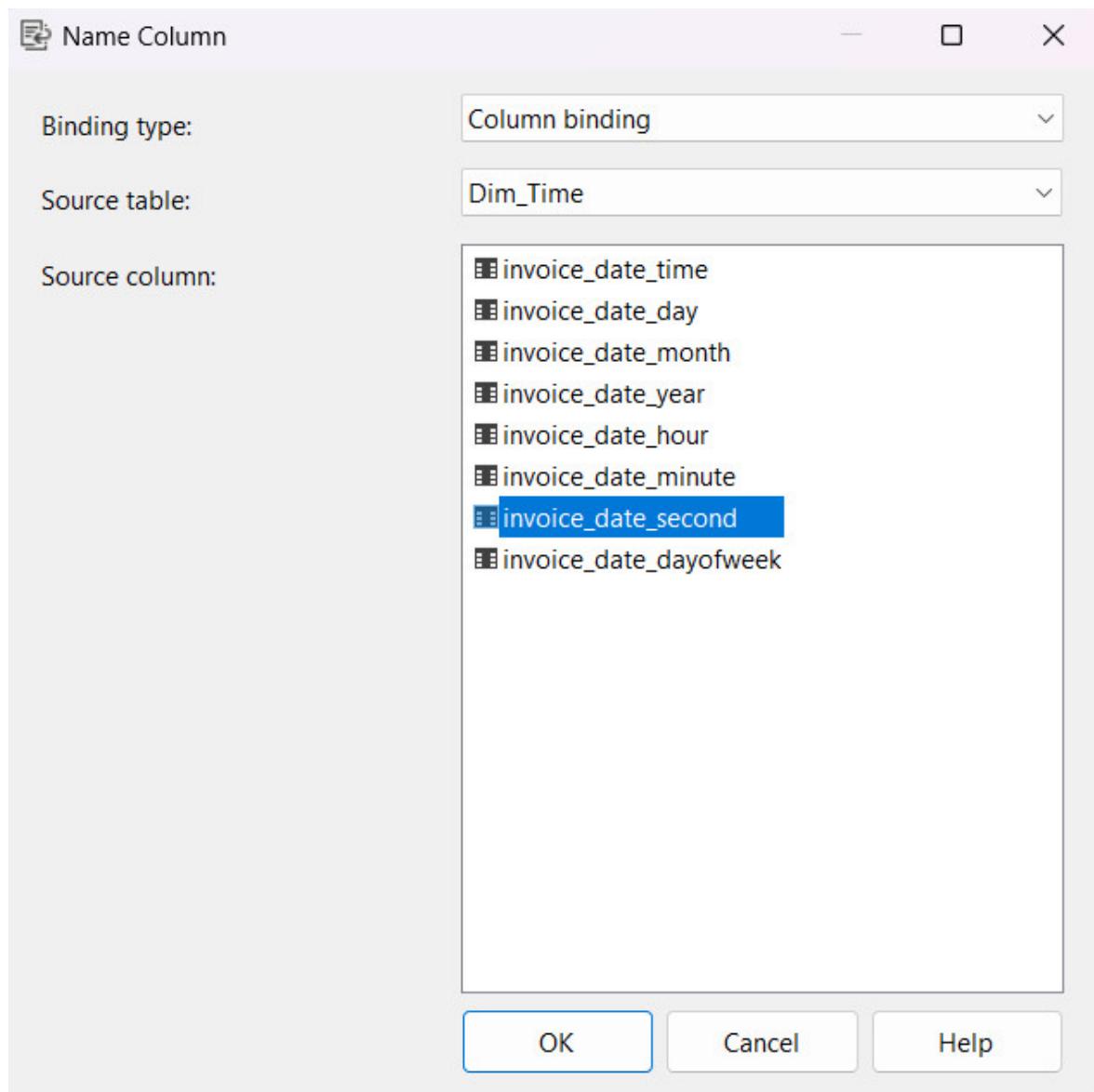
Hình 226: Chính sửa KeyColumns của thuộc tính invoice\_date\_second

Thêm các thuộc tính cấp cao hơn vào “Key Columns”, sau đó click “OK” để hoàn tất



Hình 227: Bổ sung các thuộc tính cấp cao hơn

Tại cửa sổ “Properties” của thuộc tính “Invoice Date Second”, ta chọn “Name Column” và chọn tên thuộc tính là invoice\_date\_second.



Hình 228: Chọn lại tên thuộc tính cho phù hợp

**Bước 5:** Chính khóa cột “KeyColumns” và tên cột “Name Column” của thuộc tính Invoice Date Minute. Vì thuộc tính Invoice Date Minute là thuộc tính cấp nhỏ hơn Invoice Date Hour, Invoice Date Day, Invoice Date Month, Invoice Date Year nên sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

Tại cửa sổ “Properties” của thuộc tính Invoice Date Minute, chọn “KeyColumns”.

**Properties**

**Invoice Date Minute** DimensionAttribute

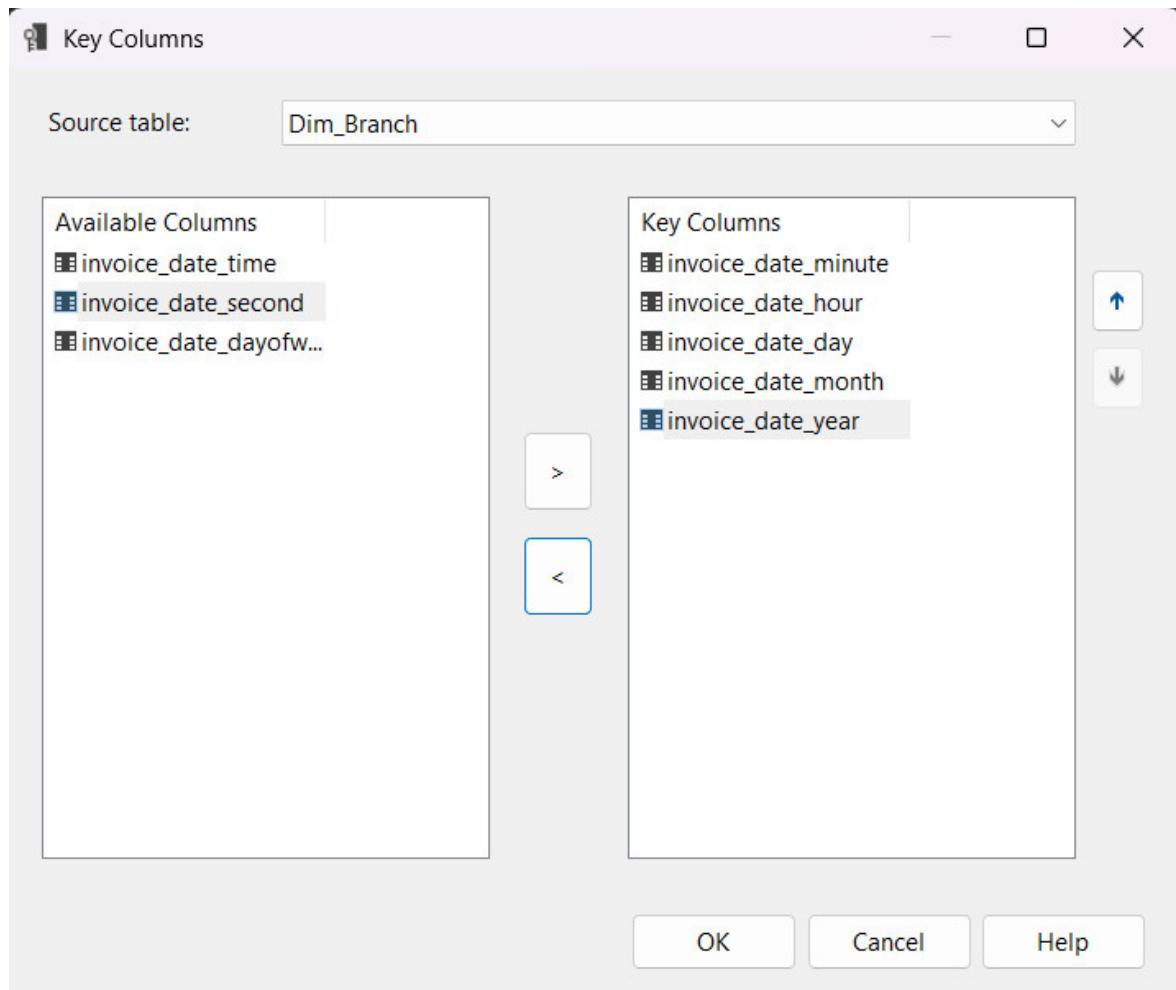
OrderByAttribute		
ProcessingState	Unprocessed	
TokenizationBehavior	TokenizationNone	
UserEditFlag	0	
VertipaqCompressionHint	VertipaqAutomatic	
<b>Basic</b>		
Description		
FormatString		
ID	Invoice Date Minute	
Name	<b>Invoice Date Minute</b>	
Type	Regular	
Usage	Regular	
<b>Misc</b>		
AttributeHierarchyOrdered	True	
ExtendedType		
GroupingBehavior	EncourageGrouping	
InstanceSelection	None	
MemberNamesUnique	False	
VisualizationProperties		
<b>Parent-Child</b>		
MembersWithData	NonLeafDataVisible	
MembersWithDataCaption		
NamingTemplate		
RootMemberIf	ParentIsBlankSelfOrMissing	
UnaryOperatorColumn	(none)	
<b>Source</b>		
CustomRollupColumn	(none)	
CustomRollupPropertiesCol	(none)	
<b>KeyColumns</b>		
Dim_Time.invoice_date_min...		
NameColumn	(none)	
ValueColumn	(none)	

**KeyColumns**  
Specifies the details of the binding to the column(s) containing the member key(s).

Hình 229: Chỉnh sửa KeyColumns của thuộc tính invoice\_date\_minute

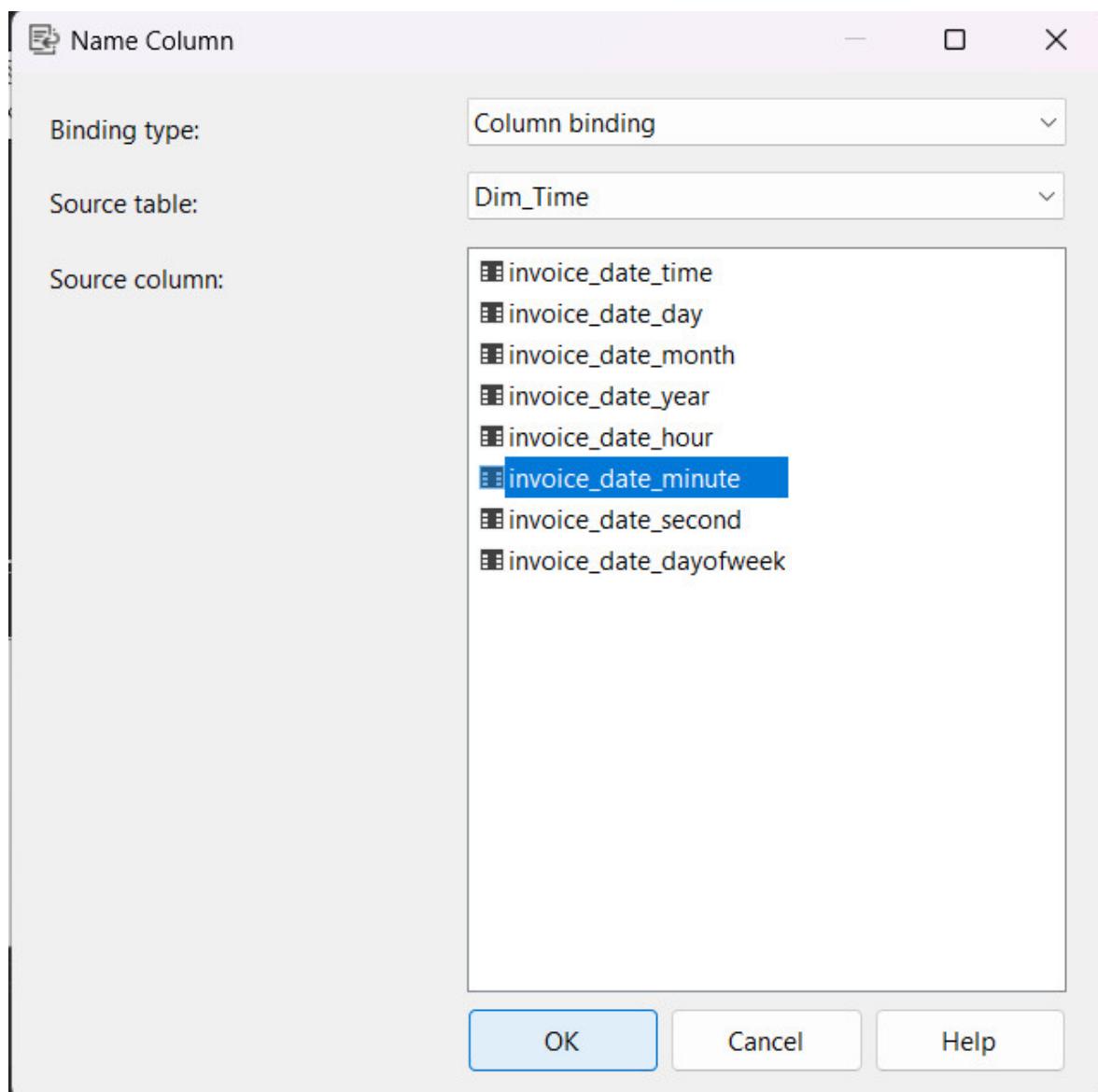
Thêm các thuộc tính cấp cao hơn vào “Key Columns”, sau đó click “OK”

để hoàn tất



Hình 230: Bổ sung các thuộc tính cấp cao hơn

Tại cửa sổ “Properties” của thuộc tính Invoice Date Minute, ta chọn “Name Column” và chọn tên thuộc tính là invoice\_date\_minute.



Hình 231: Chọn lại tên thuộc tính cho phù hợp

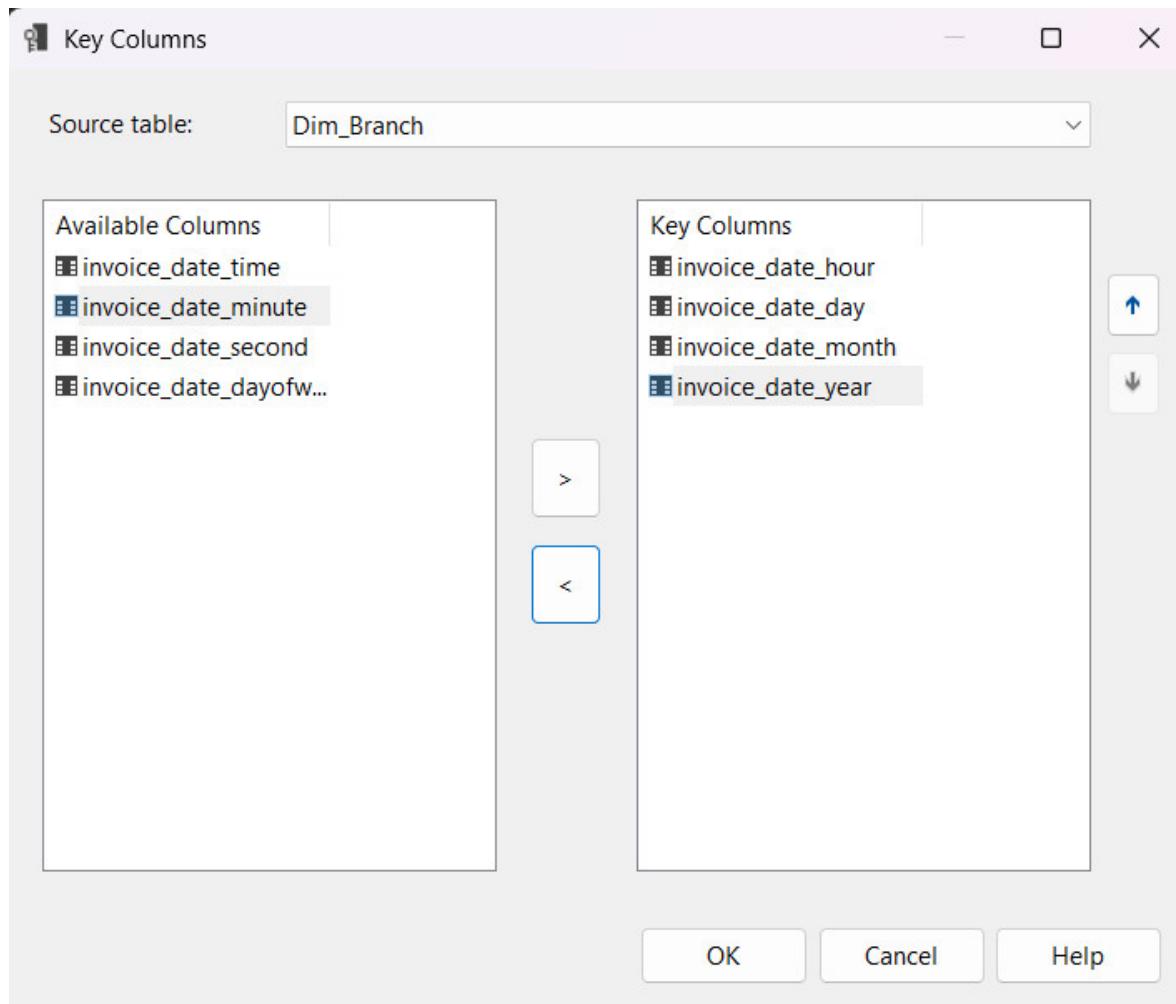
**Bước 6:** Chính khóa cột “KeyColumns” và tên cột “Name Column” của thuộc tính Invoice Date Hour. Vì thuộc tính Invoice Date Hour là thuộc tính cấp nhỏ hơn Invoice Date Day, Invoice Date Month, Invoice Date Year nên sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

Tại cửa sổ “Properties” của thuộc tính Invoice Date Hour, chọn “KeyColumns”

Properties	
Invoice Date Hour DimensionAttribute	
	OrderByAttribute
	ProcessingState
	TokenizationBehavior
	UserEditFlag
	VertipaqCompressionHint
<b>Basic</b>	
Description	
FormatString	
ID	Invoice Date Hour
Name	<b>Invoice Date Hour</b>
Type	Regular
Usage	Regular
<b>Misc</b>	
AttributeHierarchyOrdered	True
ExtendedType	
GroupingBehavior	EncourageGrouping
InstanceSelection	None
MemberNamesUnique	False
VisualizationProperties	
<b>Parent-Child</b>	
MembersWithData	NonLeafDataVisible
MembersWithDataCaption	
NamingTemplate	
RootMemberIf	ParentIsBlankSelfOrMissing
UnaryOperatorColumn	(none)
<b>Source</b>	
CustomRollupColumn	(none)
CustomRollupPropertiesCol	(none)
<b>KeyColumns</b>	Dim_Time.invoice_date_hou...
	NameColumn
	ValueColumn
<b>KeyColumns</b>	
Specifies the details of the binding to the column(s) containing the member key(s)	

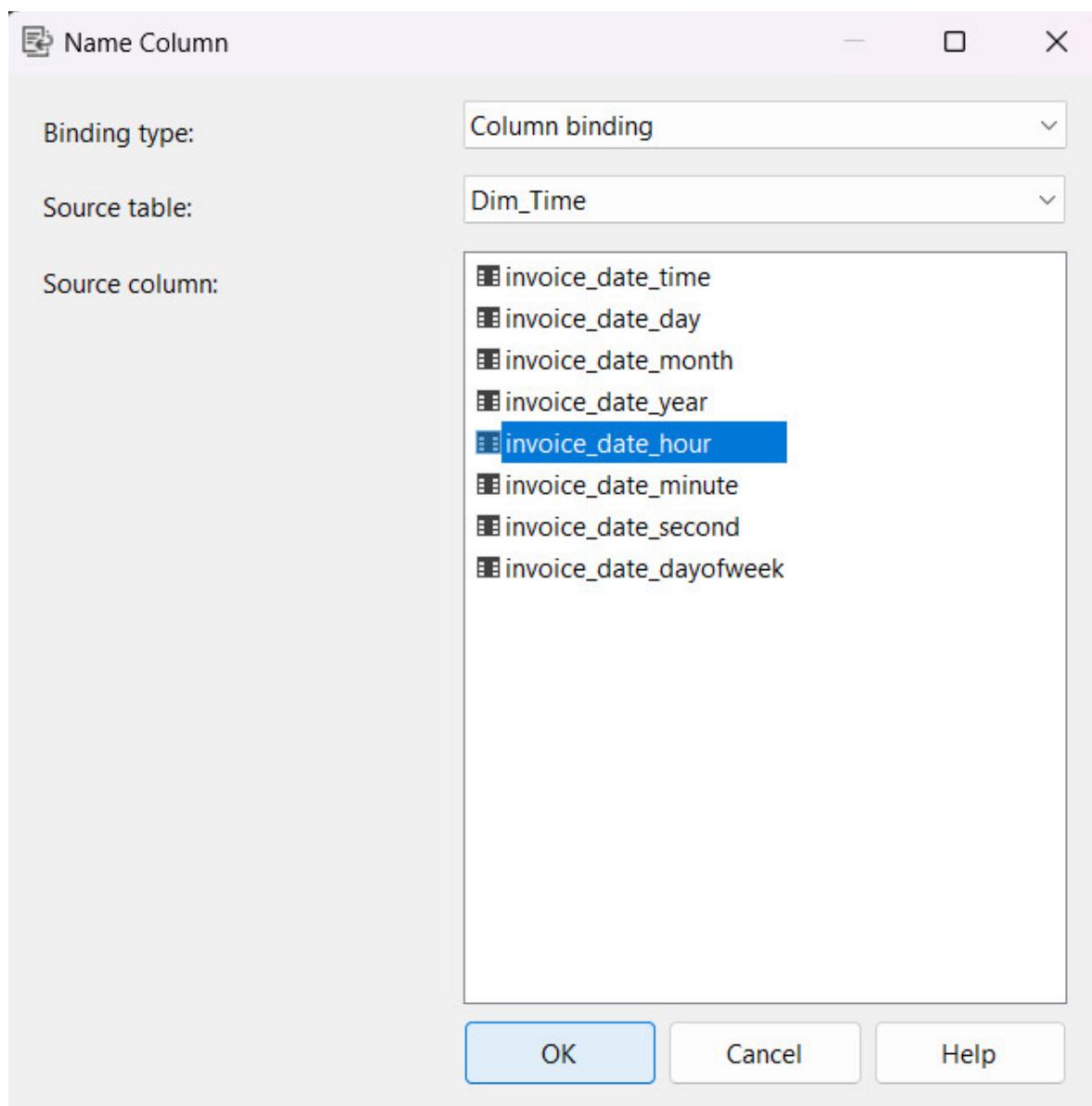
Hình 232: Chỉnh sửa KeyColumns của thuộc tính invoice\_date\_hour

Thêm các những thuộc tính cấp cao hơn vào “KeyColumns”, sau đó click “OK” để hoàn tất.



Hình 233: Bổ sung các thuộc tính cấp cao hơn

Tại cửa sổ “Properties” của thuộc tính Invoice Date Hour, ta chọn “Name Column” và chọn tên thuộc tính là invoice\_date\_hour.



Hình 234: Chọn lại tên thuộc tính cho phù hợp

**Bước 7:** Chính khóa cột “KeyColumns” và tên cột “Name Column” của thuộc tính Invoice Date Day. Vì thuộc tính Invoice Date Day là thuộc tính cấp nhỏ hơn Invoice Date Month, Invoice Date Year nên sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

Tại cửa sổ “Properties” của thuộc tính Invoice Date Day, chọn “KeyColumns”.

**Properties**

**Invoice Date Day DimensionAttribute**

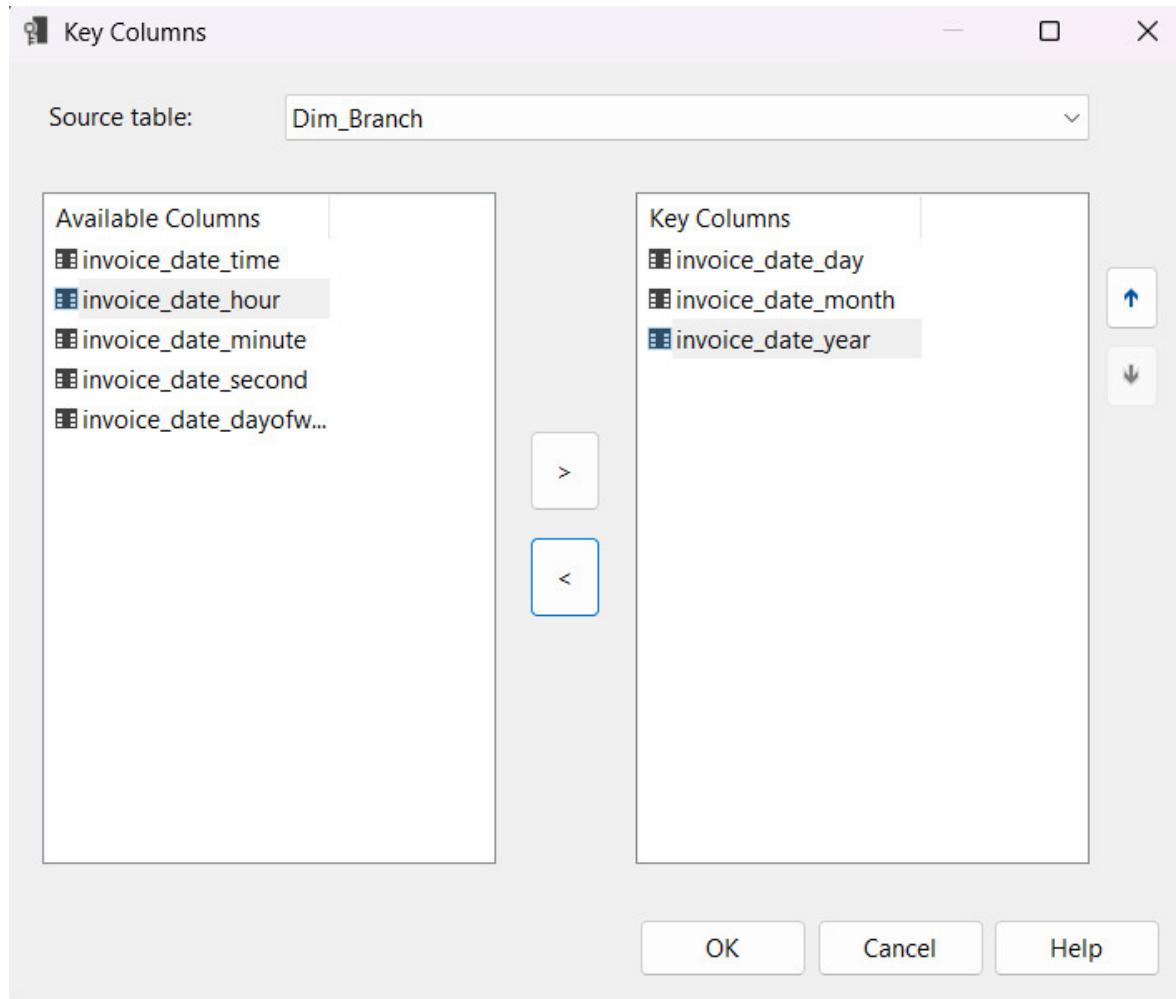
**Key Columns**

Specifies the details of the binding to the column(s) containing the member key(s).

OrderByAttribute	
ProcessingState	Unprocessed
TokenizationBehavior	TokenizationNone
UserEditFlag	0
VertipaqCompressionHint	VertipaqAutomatic
<b>Basic</b>	
Description	
FormatString	
ID	Invoice Date Day
Name	<b>Invoice Date Day</b>
Type	Regular
Usage	Regular
<b>Misc</b>	
AttributeHierarchyOrdered	True
ExtendedType	
GroupingBehavior	EncourageGrouping
InstanceSelection	None
MemberNamesUnique	False
VisualizationProperties	
<b>Parent-Child</b>	
MembersWithData	NonLeafDataVisible
MembersWithDataCaption	
NamingTemplate	
RootMemberIf	ParentIsBlankSelfOrMissing
UnaryOperatorColumn	(none)
<b>Source</b>	
CustomRollupColumn	(none)
CustomRollupPropertiesCol	(none)
<b>KeyColumns</b>	Dim_Time.invoice_date_day...
NameColumn	(none)
ValueColumn	(none)

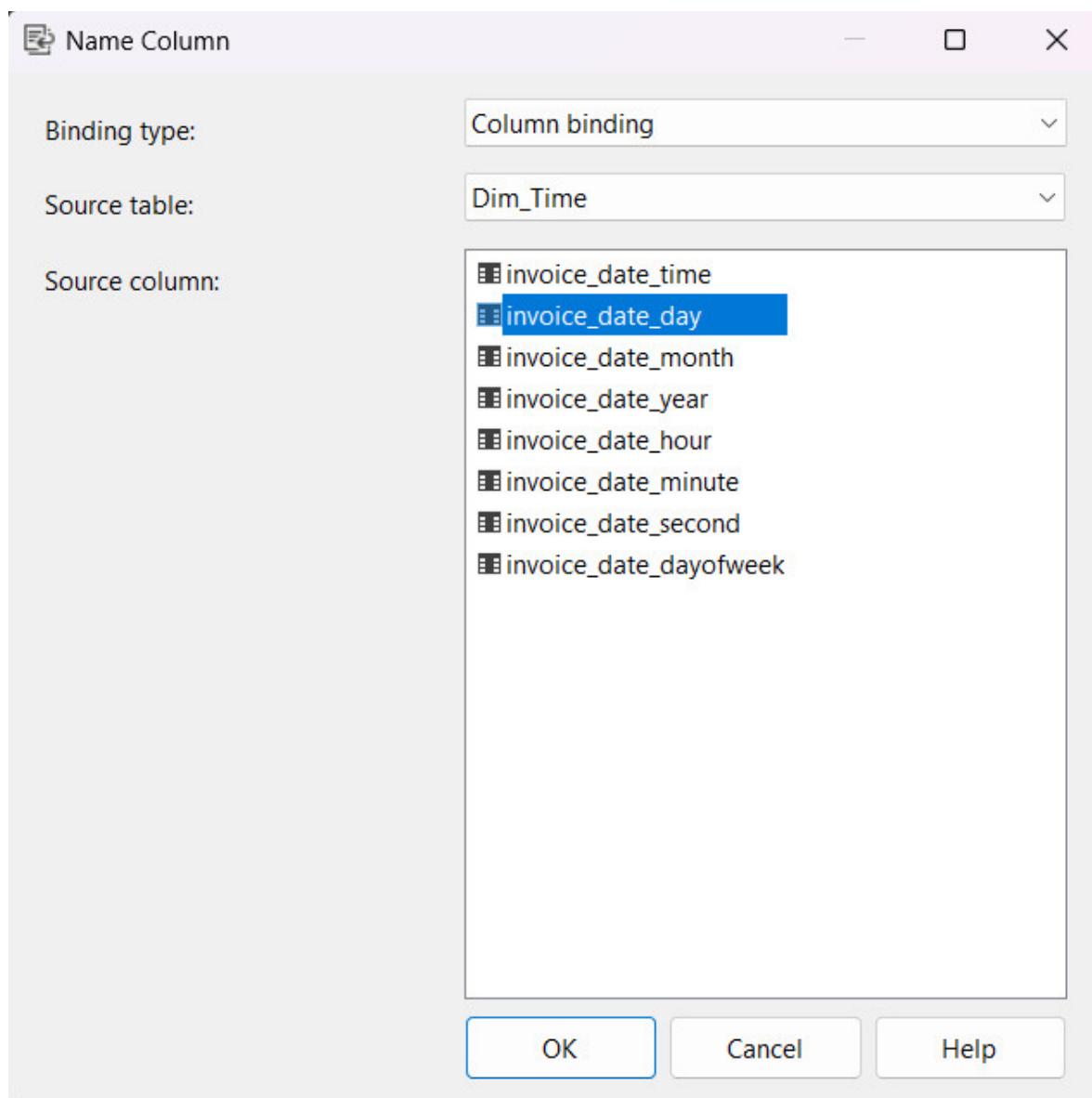
Hình 235: Chỉnh sửa KeyColumns của thuộc tính invoice\_date\_day

Thêm các những thuộc tính cấp cao hơn vào “KeyColumns”, sau đó click “OK” để hoàn tất.



Hình 236: Bổ sung các thuộc tính cấp cao hơn

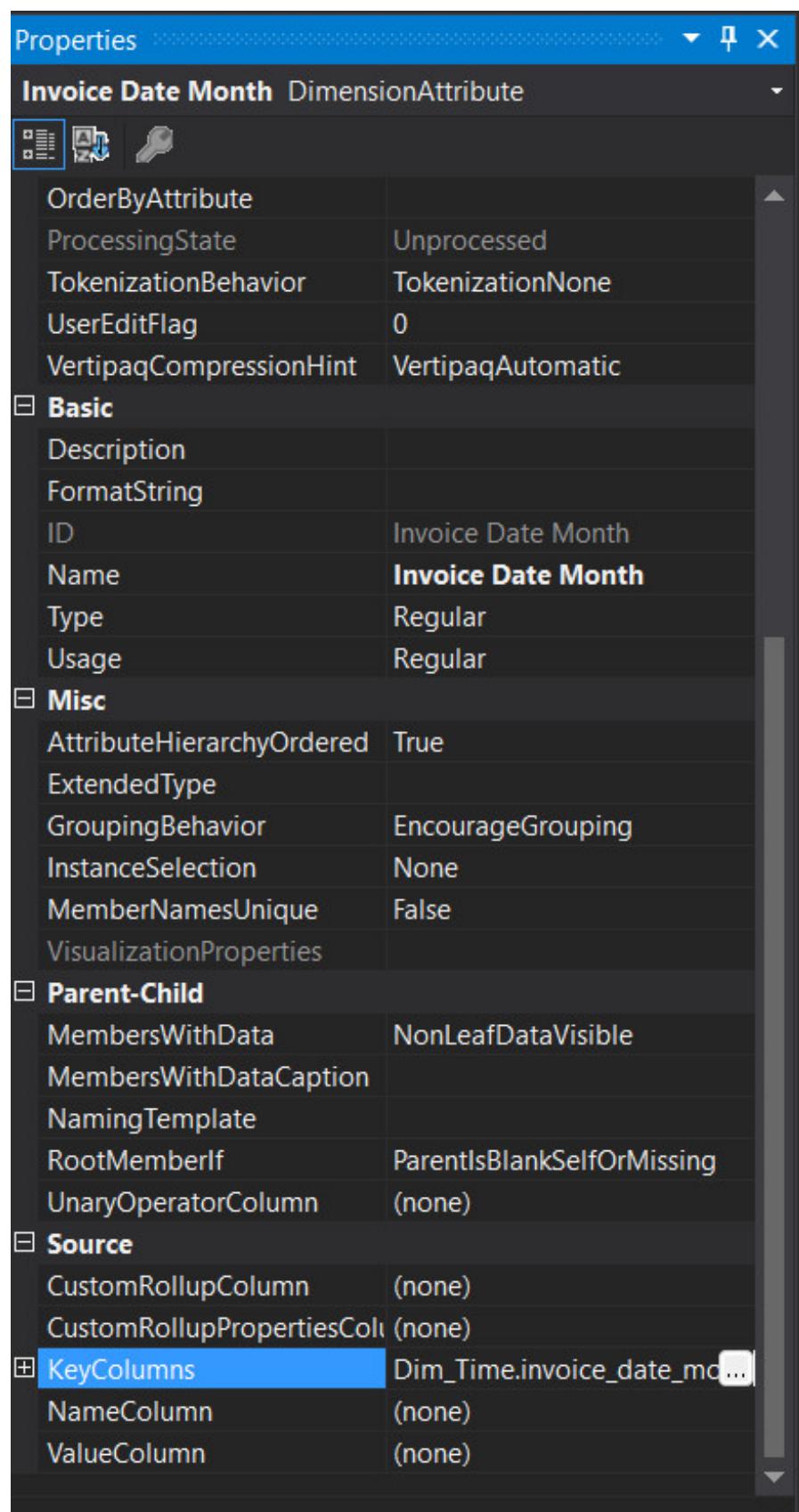
Tại cửa sổ “Properties” của thuộc tính Invoice Date Day, ta chọn “Name Column” và chọn tên thuộc tính là invoice\_date\_day.



Hình 237: Chọn lại tên thuộc tính cho phù hợp

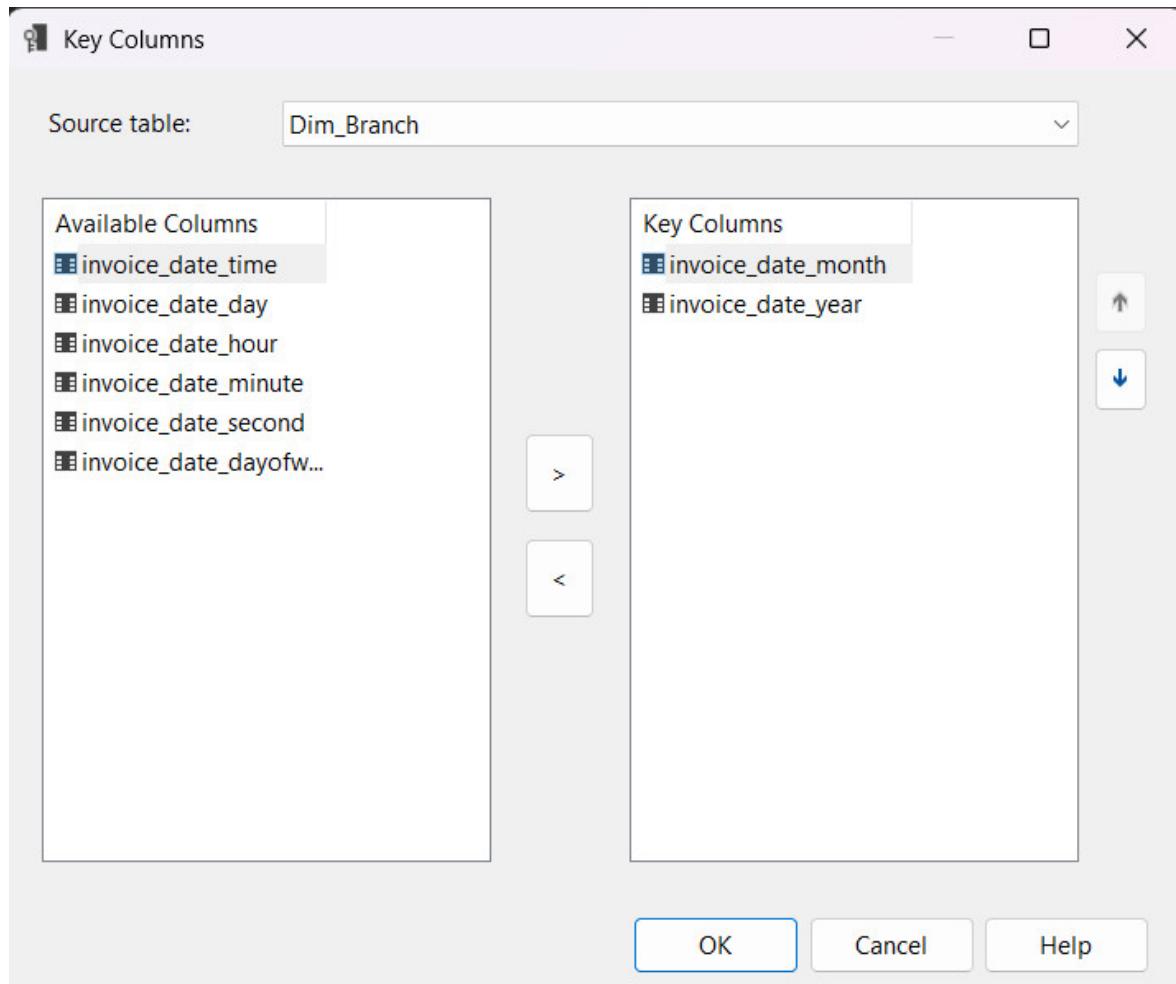
**Bước 8:** Chính khóa cột “**KeyColumns**” và tên cột “**Name Column**” của thuộc tính Invoice Date Month. Vì thuộc tính Invoice Date Month là thuộc tính cấp nhô hơn Invoice Date Year nên sẽ lấy khóa cột gồm chính nó và thuộc tính cấp cao hơn.

Tại cửa sổ “**Properties**” của thuộc tính Invoice Date Month, chọn “**KeyColumns**”.



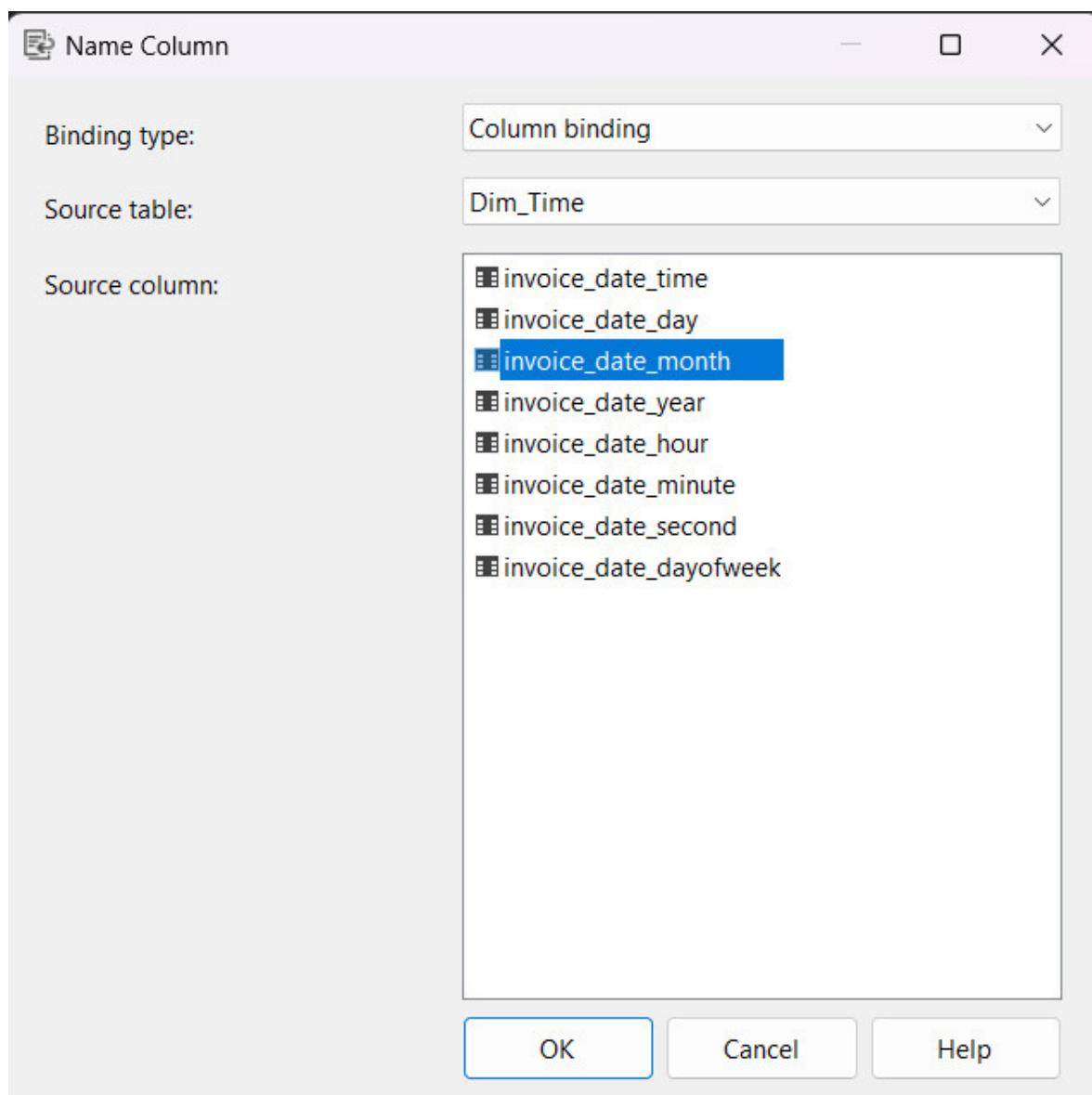
Hình 238: Chỉnh sửa KeyColumns của thuộc tính invoice\_date\_month

Thêm các những thuộc tính cấp cao hơn vào “KeyColumns”, sau đó chọn “OK” để hoàn tất.



Hình 239: Bổ sung thuộc tính cấp cao hơn

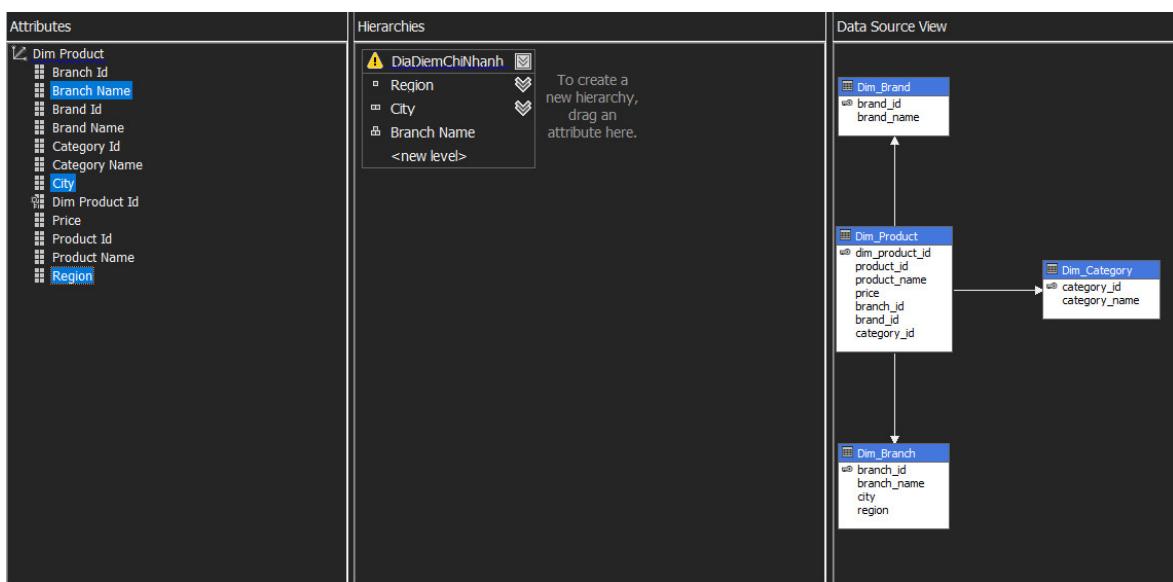
Tại cửa sổ “Properties” của thuộc tính Invoice Date Month, ta chọn “Name Column” và chọn tên thuộc tính là invoice\_date\_month.



Hình 240: Chọn lại tên thuộc tính cho phù hợp

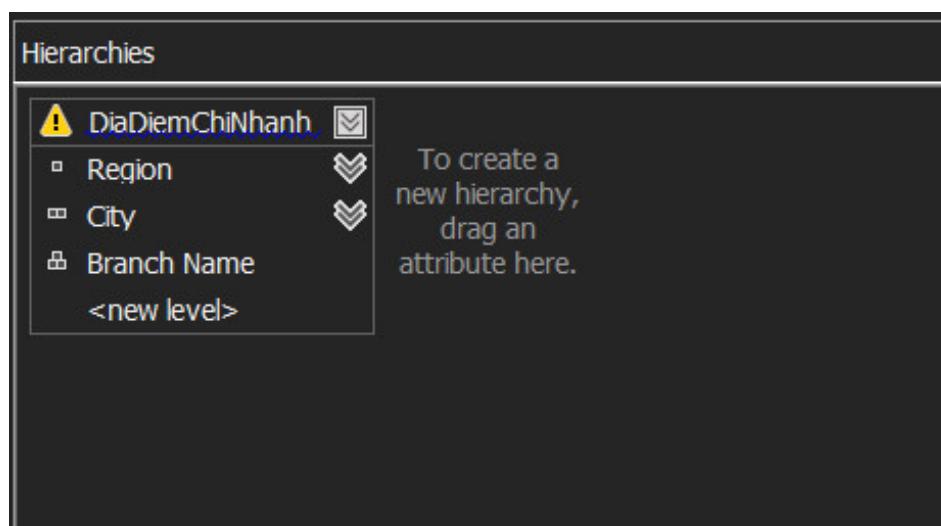
### 3.8.2. Phân cấp bảng Dim\_Product:

**Bước 1:** Kéo những thuộc tính cần phân cấp qua cửa sổ “Hierarchies”



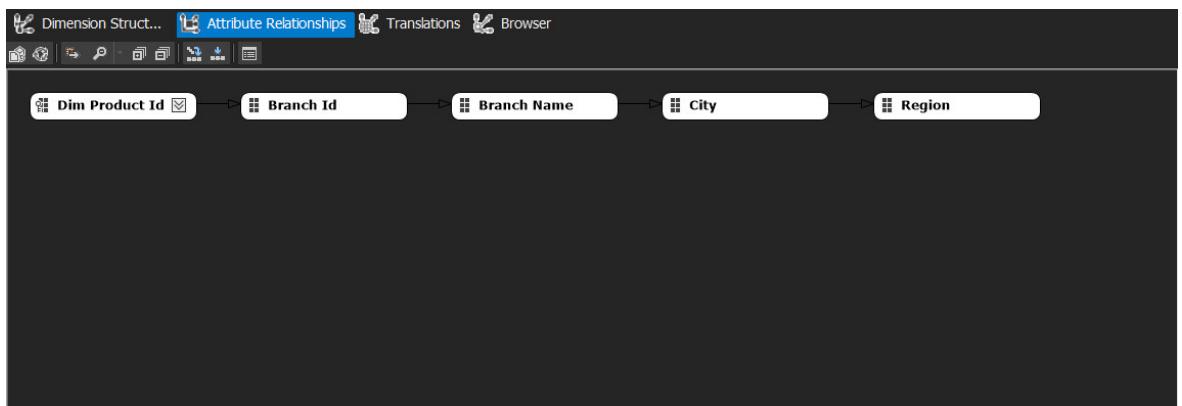
Hình 241: Đưa toàn bộ thuộc tính phân cấp của bảng Dim\_Product qua cửa sổ Hierarchies

**Bước 2:** Sắp xếp lại các thuộc tính phân cấp theo thứ tự: Region -> City -> Branch Name và đổi tên “Hierarchies” thành DiaDiemChiNhanh



Hình 242: Cửa sổ sắp xếp thứ tự của các thuộc tính

**Bước 3:** Tại panel “Attribute Relationships”, click chuột phải vào các mũi tên chọn “Delete” và tạo lại mối quan hệ như sau:



Hình 243: Tạo lại mối quan hệ giữa các thuộc tính cho phù hợp

**Bước 4:** Chính khóa cột “KeyColumns” và tên cột “Name Column” của thuộc tính Branch Name. Vì thuộc tính Branch Name là thuộc tính cấp nhỏ nhất sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

Tại cửa sổ “Properties” của thuộc tính Branch Name, chọn “KeyColumns”.

**Properties**

**Branch Name** DimensionAttribute

OrderByAttribute

ProcessingState Unprocessed

TokenizationBehavior TokenizationNone

UserEditFlag 0

VertipaqCompressionHint VertipaqAutomatic

**Basic**

Description

FormatString

ID Branch Name

Name **Branch Name**

Type Regular

Usage Regular

**Misc**

AttributeHierarchyOrdered True

ExtendedType

GroupingBehavior EncourageGrouping

InstanceSelection None

MemberNamesUnique False

VisualizationProperties

**Parent-Child**

MembersWithData NonLeafDataVisible

MembersWithDataCaption

NamingTemplate

RootMemberIf ParentIsBlankSelfOrMissing

UnaryOperatorColumn (none)

**Source**

CustomRollupColumn (none)

CustomRollupPropertiesCol (none)

**KeyColumns** Dim\_Branch.branch\_name (...)

NameColumn (none)

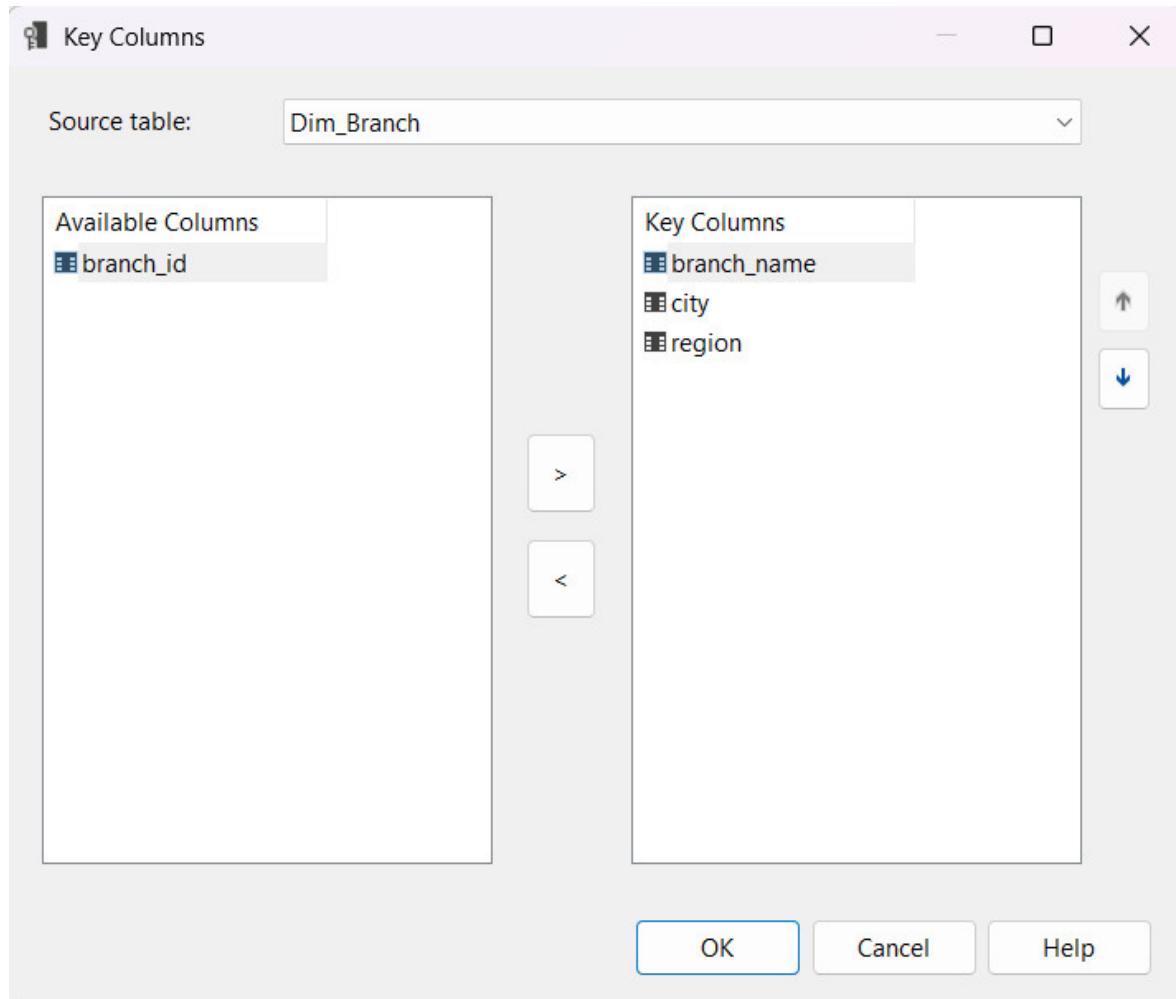
ValueColumn (none)

**KeyColumns**

Specifies the details of the binding to the column(s) containing the member key(s).

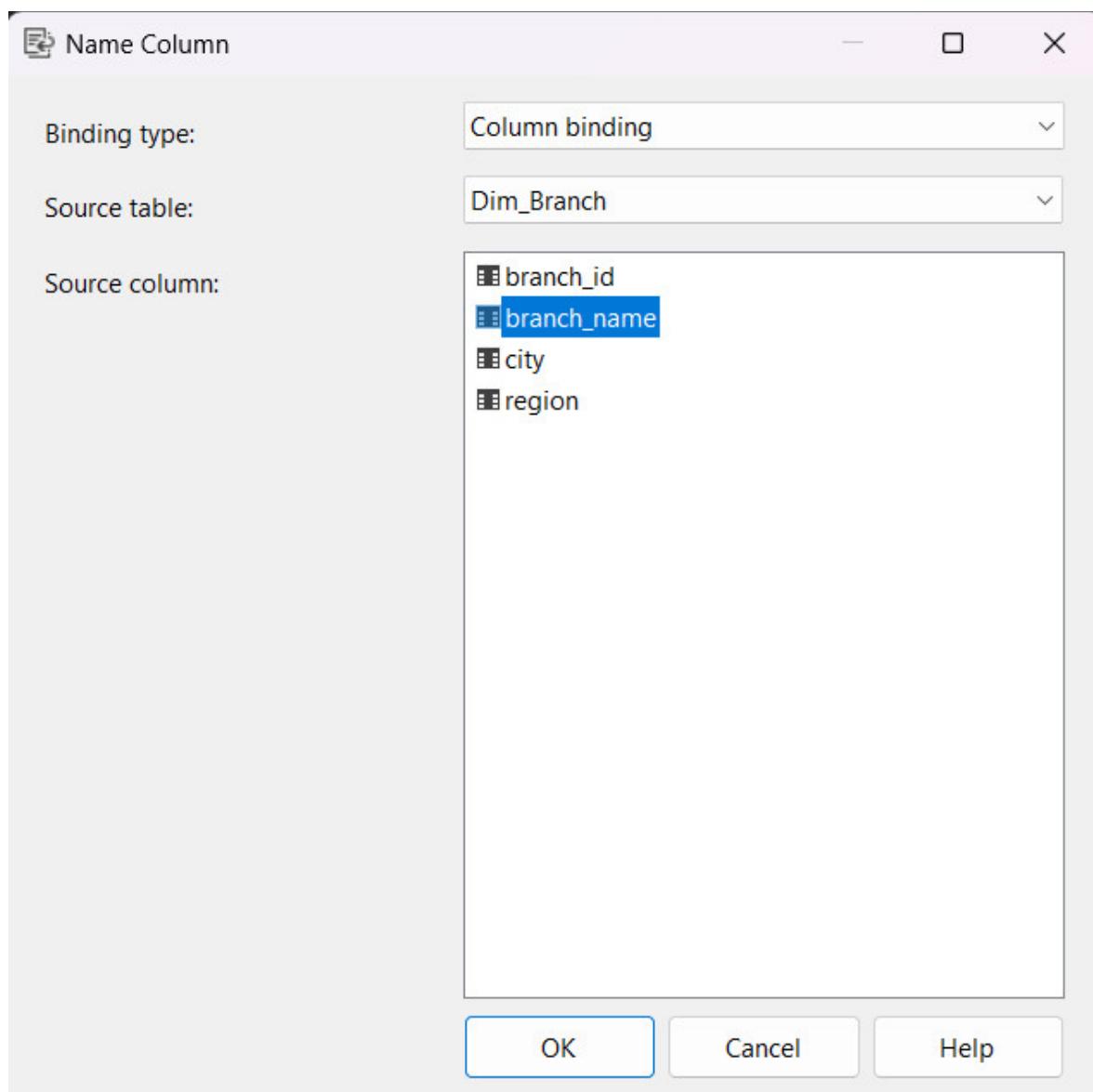
Hình 244: Chỉnh sửa KeyColumns của thuộc tính Branch\_Name

Thêm các những thuộc tính cấp cao hơn vào “KeyColumns”, sau đó chọn “OK” để hoàn tất.



Hình 245: Bổ sung các thuộc tính cấp cao hơn

Tại cửa sổ “Properties” của thuộc tính Branch Name, ta chọn “Name Column” và chọn tên thuộc tính là branch\_name.



Hình 246: Chọn lại tên thuộc tính cho phù hợp

**Bước 5:** Chính khóa cột “KeyColumns” và tên cột “Name Column” của thuộc tính City. Vì thuộc tính City là thuộc tính cấp nhỏ hơn thuộc tính Region sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

Tại cửa sổ “Properties” của thuộc tính City, chọn “KeyColumns”.

**City DimensionAttribute**

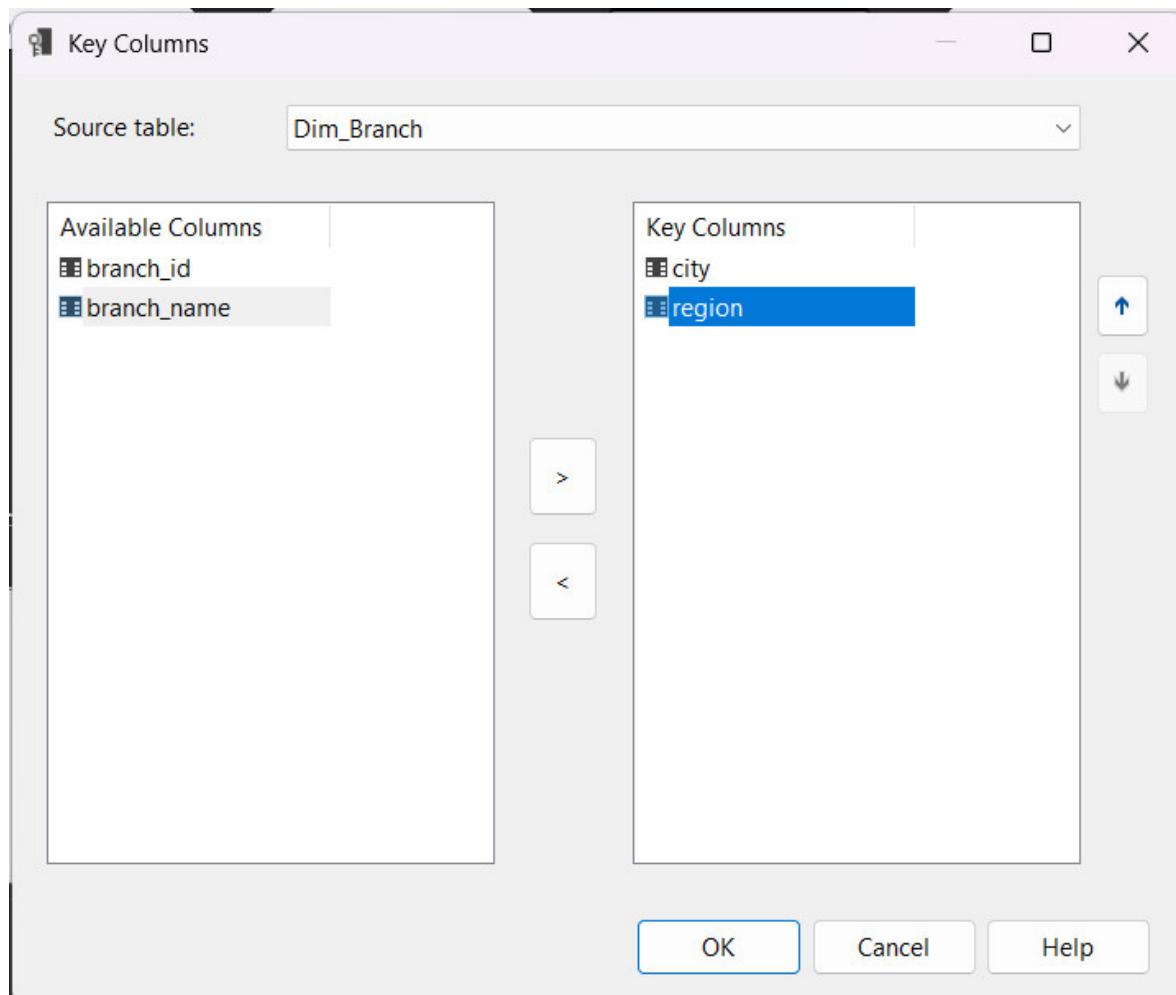
		City DimensionAttribute
OrderByAttribute		
ProcessingState	Unprocessed	
TokenizationBehavior	TokenizationNone	
UserEditFlag	0	
VertipaqCompressionHint	VertipaqAutomatic	
<b>Basic</b>		
Description		
FormatString		
ID	City	
Name	<b>City</b>	
Type	Regular	
Usage	Regular	
<b>Misc</b>		
AttributeHierarchyOrdered	True	
ExtendedType		
GroupingBehavior	EncourageGrouping	
InstanceSelection	None	
MemberNamesUnique	False	
VisualizationProperties		
<b>Parent-Child</b>		
MembersWithData	NonLeafDataVisible	
MembersWithDataCaption		
NamingTemplate		
RootMemberIf	ParentIsBlankSelfOrMissing	
UnaryOperatorColumn	(none)	
<b>Source</b>		
CustomRollupColumn	(none)	
CustomRollupPropertiesColu	(none)	
<b>KeyColumns</b>		
Dim_Branch.city (WChar)	...	
NameColumn	(none)	
ValueColumn	(none)	

**KeyColumns**

Specifies the details of the binding to the column(s) containing the member key(s).

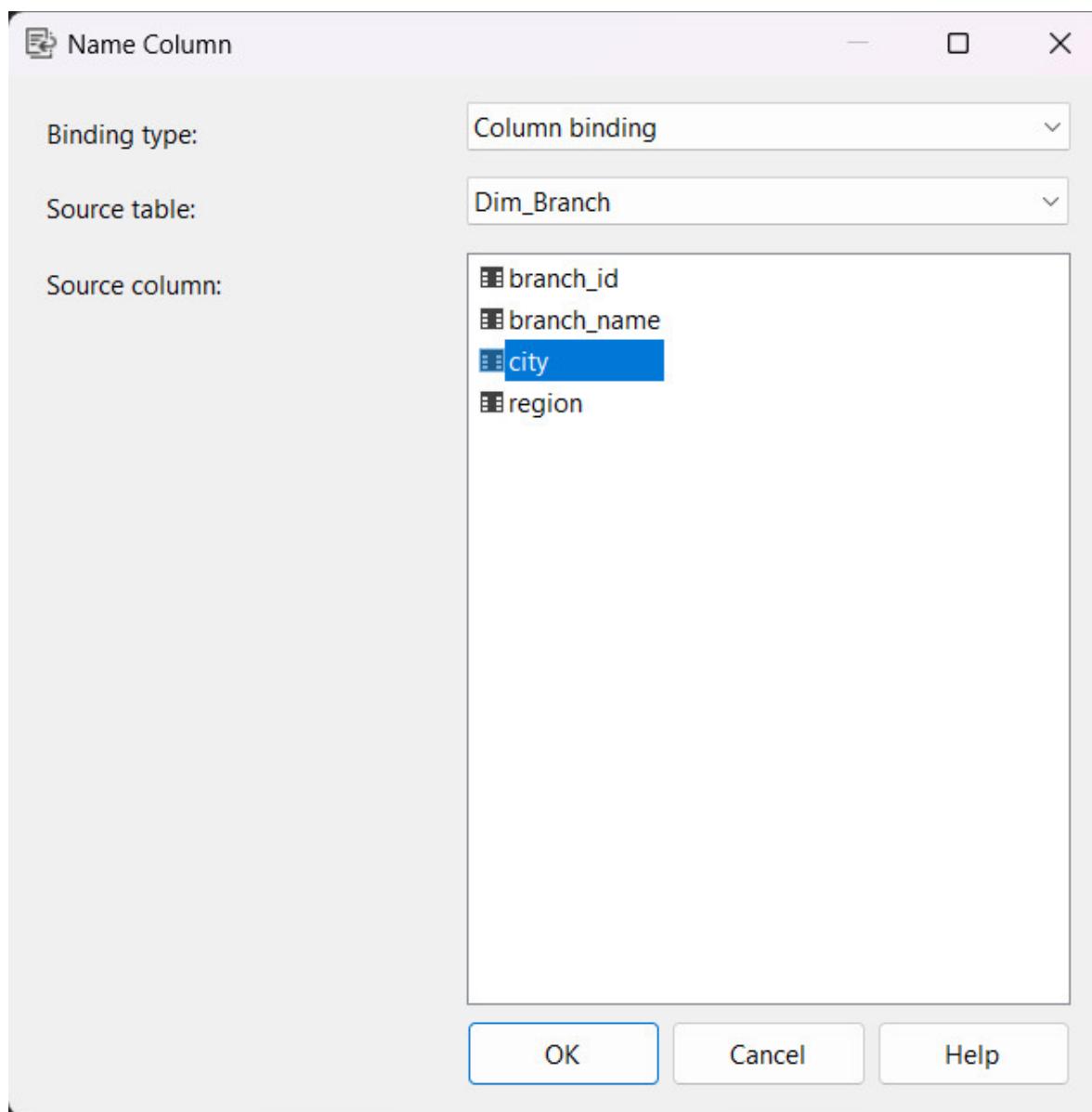
Hình 247: Chỉnh sửa KeyColumns của thuộc tính City

Thêm các những thuộc tính cấp cao hơn vào “KeyColumns”, sau đó chọn “OK” để hoàn tất.



Hình 248: Bổ sung các thuộc tính cấp cao hơn

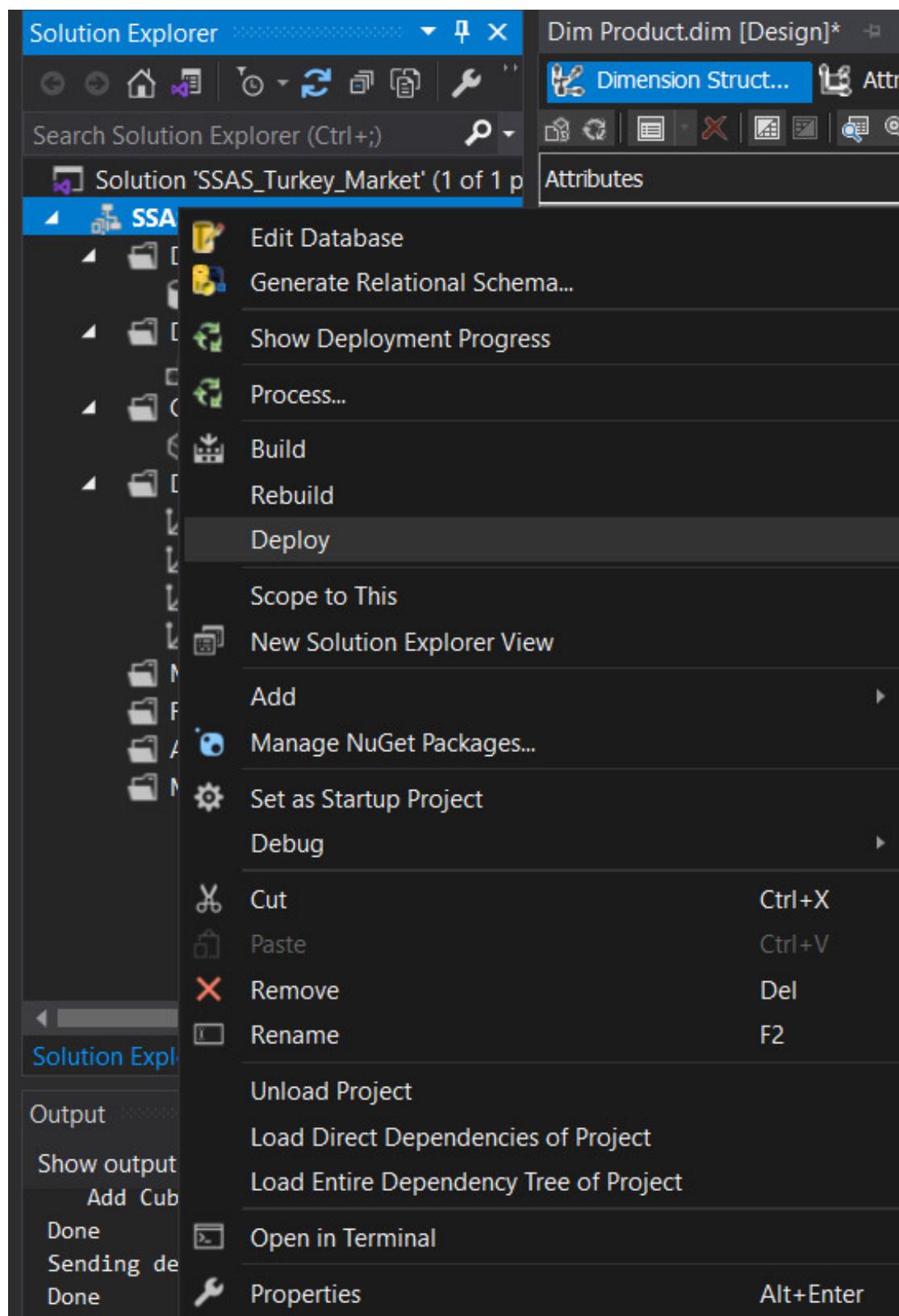
Tại cửa sổ “Properties” của thuộc tính City, ta chọn “Name Column” và chọn tên thuộc tính là City



Hình 249: Chọn lại tên thuộc tính cho phù hợp

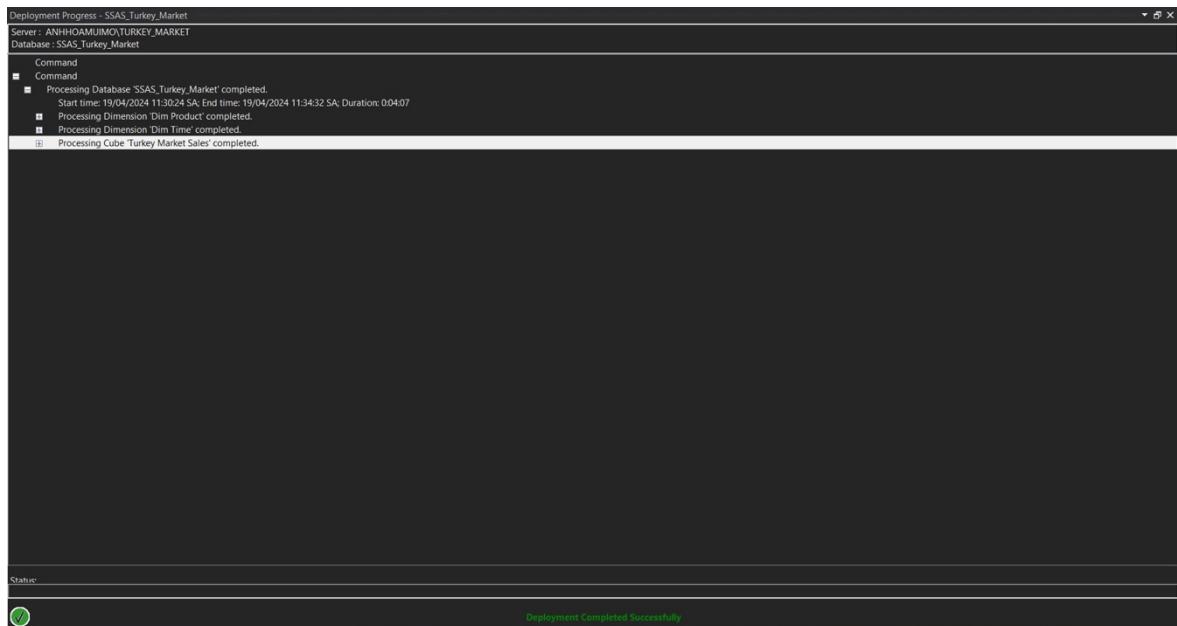
### 3.9. Chạy dự án SSAS:

Sau khi quá trình phân cấp cho các bảng chiều hoàn tất, ta thực hiện deploy project để đảm bảo không có lỗi xảy ra sau quá trình phân cấp. Click chuột phải vào tên project (SSAS\_Turkey\_Market) và nhấn Deploy như hình sau:



Hình 250: Deploy lại dự án sau khi thực hiện phân cấp

Khi deploy toàn bộ quá trình thành công, hệ thống sẽ hiển thị như hình sau và chúng ta bắt đầu thực hiện các câu truy vấn



Hình 251: Quá trình deploy thành công

### 3.10. Quá trình phân tích dữ liệu bằng thao tác tay trên các khối CUBE:

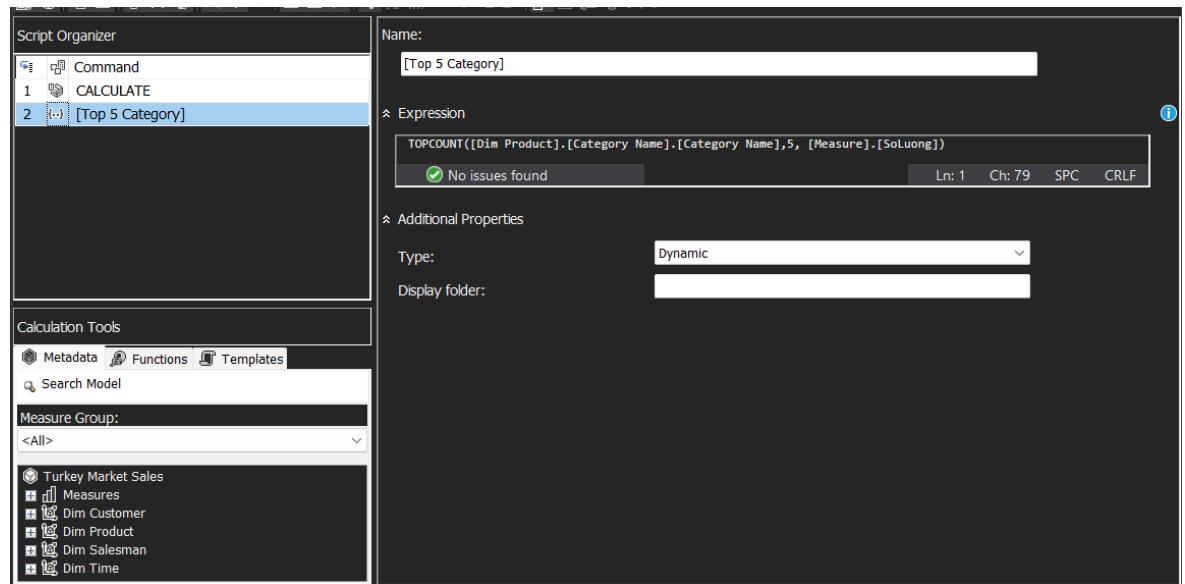
#### 3.10.1. Câu truy vấn 1:

**Nội dung câu truy vấn:** Với top 5 kho sản phẩm có doanh số bán ra cao nhất, thống kê số lượng sản phẩm bán ra được của 5 kho đó theo từng tháng.

**Thực hiện câu truy vấn:**

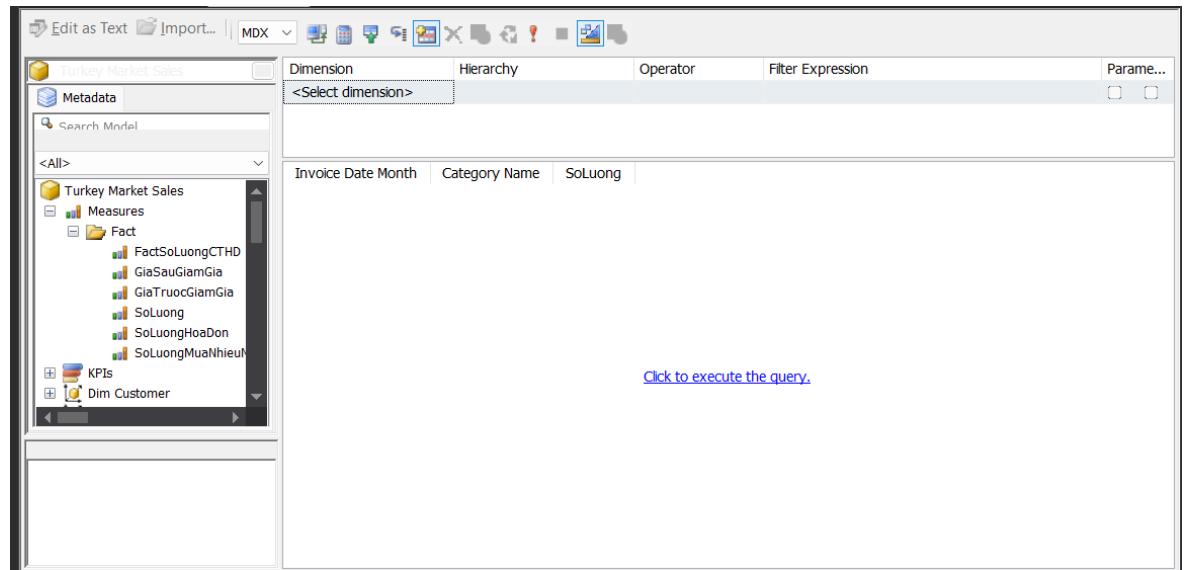
**Bước 1:** Tạo NameSet [Top 5 Category] như hình minh họa với biểu thức sau:

TOPCOUNT([Dim Product].[Category Name].[Category Name],5,  
[Measures].[SoLuong])



Hình 252: Khởi tạo NameSet [Top 5 Category]

**Bước 2:** Kéo thả thuộc tính [Invoice Date Month] trong bảng [Dim Time], [Category Name] trong bảng [Dim Product] và độ đo [SoLuong] vào cửa sổ thực thi câu truy vấn.



Hình 253: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 3:** Kéo thả Dimension [Dim Product] vào cửa sổ lọc Hierarchy [Category Name], Operator [In] và chọn NameSet [Top 5 Category] ở ô Filter Expression.

Invoice Date Month	Category Name	SoLuong
1	Beverage	31939
1	Cosmetics	5862
1	Dairy Breakfast	19488,8...
1	Detergent Clea...	6071
1	Food	59846,0...
2	Beverage	34851
2	Cosmetics	6839
2	Dairy Breakfast	21499,1...
2	Detergent Clea...	6687
2	Food	63269,2...
3	Beverage	41651
3	Cosmetics	7569
3	Dairy Breakfast	24729,4...
3	Detergent Clea...	7595
3	Food	73796,7...
4	Beverage	1375
	Formation	744

Hình 254: Thêm NameSet vào ô Filter Expression và thu được kết quả câu truy vấn 1

### 3.10.2. Câu truy vấn 2:

**Nội dung câu truy vấn:** Thống kê tổng số lượng sản phẩm bán được của các nhãn hàng “Sprite”, “Clear”, “Nestle”, “Lipton”, “Colgate” trong tháng 02 năm 2017.

**Thực hiện câu truy vấn:**

**Bước 1:** Kéo thả thuộc tính [Brand Name] trong bảng [Dim Product], độ đo [SoLuong] vào cửa sổ thực thi.

Brand Name	SoLuong
Sprite	1375
Clear	6839
Nestle	7569
Lipton	24729,4...
Colgate	73796,7...

Hình 255: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 2:** Kéo thả Dimension [Dim Product] vào dòng đầu của cửa sổ lọc với Hierarchy [Brand Name], Operator [Equal] và lọc ra các giá trị {‘Sprite’, ‘Clear’, ‘Nestle’, ‘Lipton’, ‘Colgate’} ở ô Filter Expression.

Dimension	Hierarchy	Operator	Filter Expression
Dim Product	Brand Name	Equal	{ Clear, Colgate, Lipton, Nestle, Sprite }

Brand Name	SoLuong
Clear	950
Colgate	1725
Lipton	1848
Nestle	24520
Sprite	155

Hình 256: Dưa bảng Dim\_Product và tên các nhãn hàng vào cửa sổ lọc

**Bước 3:** Kéo thả Dimension [Dim Time] vào dòng thứ hai của cửa sổ lọc Hierarchy [Invoice Date Month], Operator [Equal] và lọc ra giá trị {2} ở ô Filter Expression.

Dimension	Hierarchy	Operator	Filter Expression
Dim Product	Brand Name	Equal	{ Clear, Colgate, Lipton, Nestle, Sprite }
Dim Time	InvoiceTime	Equal	{ 2 }

Brand Name	SoLuong
Clear	309
Colgate	619
Lipton	636
Nestle	7792
Sprite	43

Hình 257: Dưa bảng Dim\_Time bổ sung điều kiện thời gian vào cửa sổ lọc và thu được kết quả câu truy vấn

### 3.10.3. Câu truy vấn 3:

**Nội dung câu truy vấn:** Với mỗi thành phố, thống kê số lượng sản phẩm bán ra được ở từng chi nhánh có số lượng sản phẩm bán ra lớn hơn 15000. (scroll up từ Branch Name lên City)

**Thực hiện câu truy vấn:**

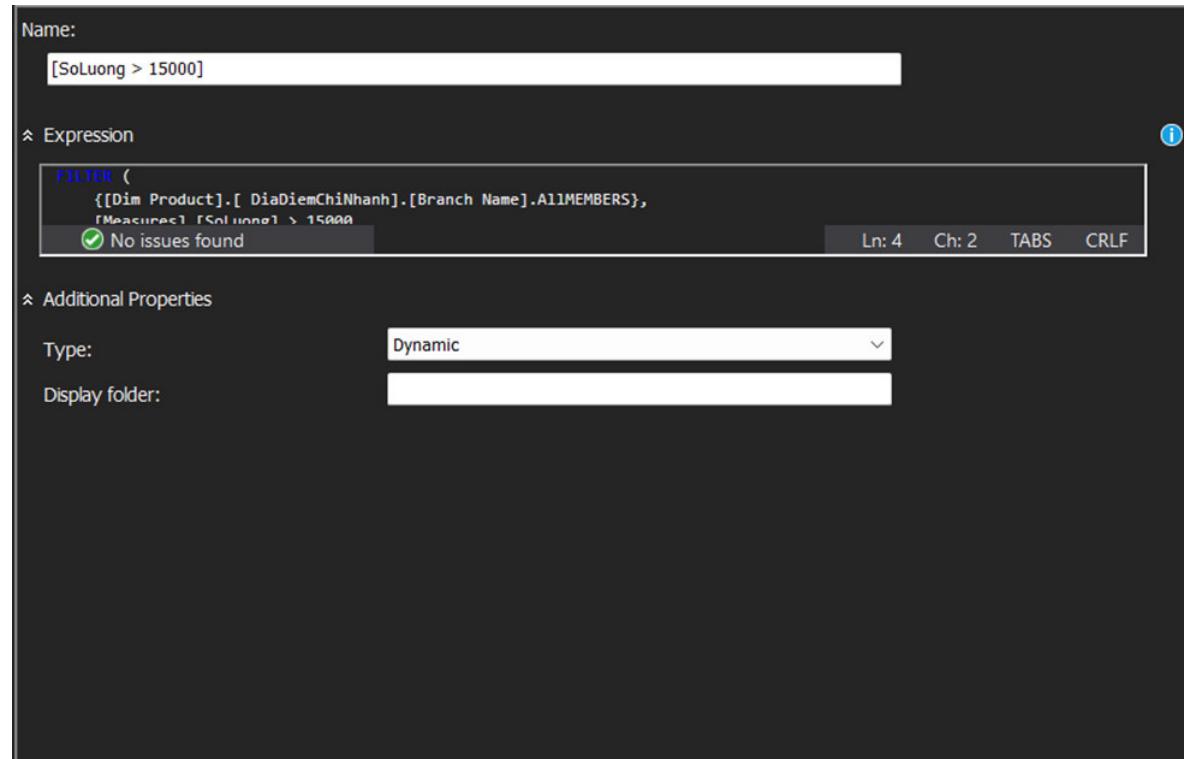
**Bước 1:** Tạo NameSet [So Luong > 15000] như hình minh họa với biểu thức sau:

FILTER (

{[Dim Product].[DiaDiemChiNhanh].[Branch Name].AllMEMBERS},

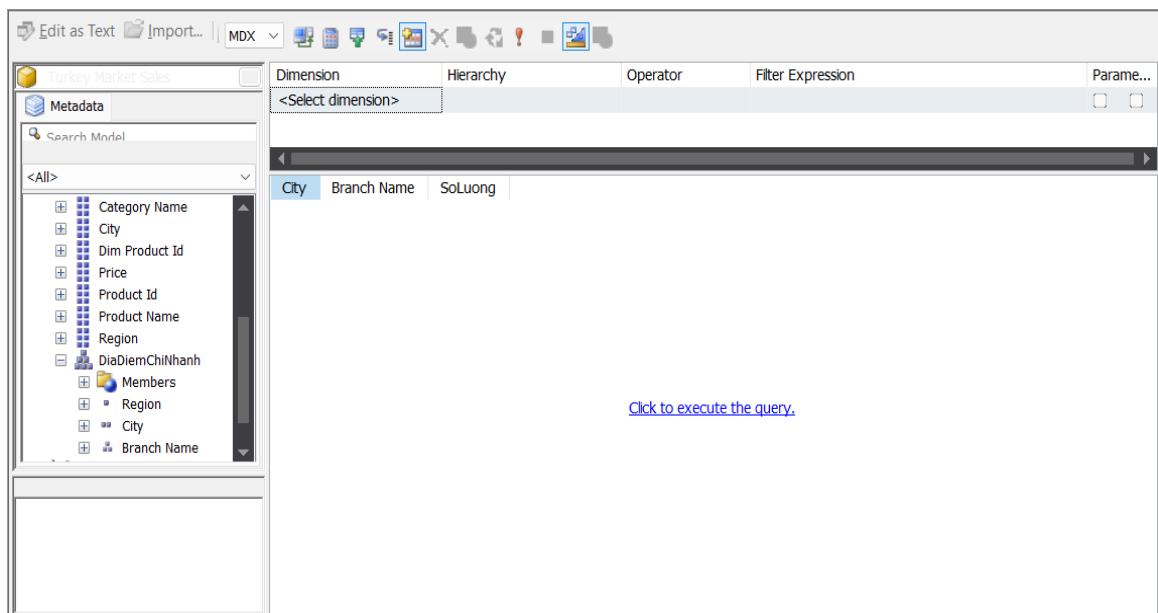
[Measures].[SoLuong] > 15000

)



Hình 258: Khởi tạo NameSet [So Luong > 15000]

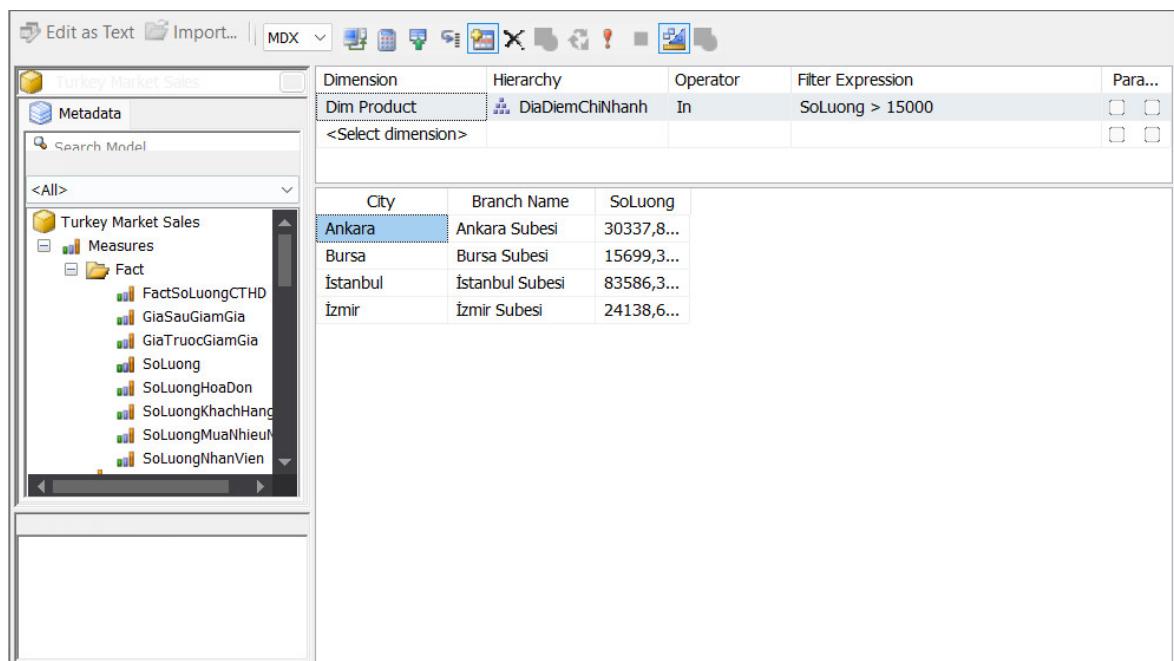
**Bước 2:** Kéo thả thuộc tính [Branch Name] trong bảng phân cấp [Dim Branch] của bảng [Dim Product], thuộc tính [City] trong bảng [Dim Product] và độ đo [So Luong] vào cửa sổ thực thi câu truy vấn.



Hình 259: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 3:** Kéo thả Dimension [Dim Product] vào cửa sổ lọc Hierarchy

[DiaDiemChiNhanh], Operator [In] và chọn NameSet [SoLuong > 15000] ở ô Filter Expression.



Hình 260: Thêm NameSet vào ô Filter Expression và thu được kết quả câu truy vấn 3

### 3.10.4. Câu truy vấn 4:

**Nội dung câu truy vấn:** Thông kê tổng doanh thu trong tháng 04 tại các chi nhánh ở vùng miền “Marmara” dựa trên nhãn hàng và loại sản phẩm.

### Thực hiện câu truy vấn:

**Bước 1:** Kéo thả thuộc tính [Brand Name] trong bảng phân cấp [Dim Brand] của bảng [Dim Product], [Category Name] trong bảng phân cấp [Dim Category] của bảng [Dim Product] và độ đo [GiaSauGiamGia] vào cửa sổ thực thi câu truy vấn.

Hình 261: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 2:** Kéo thả Dimension [Dim Product] vào dòng đầu của cửa sổ lọc với Hierarchy [DiaDiemChiNhanh], Operator [Equal] và lọc ra các giá trị {Marmara} ở ô Filter Expression.

Category Name	Brand Name	GiaSauGiamGia
Baby	Aptamil	429,21
Baby	Ari	155,66
Baby	Canbebe	1595,72
Baby	Dalin	367,23
Baby	ETİ	521,29
Baby	Evy Baby	27,04
Baby	Huggies	551,7
Baby	Molfix	1971,97
Baby	People	21,24
Baby	Prima	4372,31
Baby	Ülker	858,5
Baby	Uni	33,3
Baby	Wee	302,52
Beverage	F NEFFİS	4314,679999...
Beverage	Akdem	321,93
Beverage	Altincezve	833,12
Beverage	Arzana	425.600.000

Hình 262: Dựa thuộc tính phân cấp [DiaDiemChiNhanh] và lọc ra vùng miền 'Marmara'

**Bước 3:** Kéo thả Dimension [Dim Time] vào dòng thứ hai của cửa sổ lọc Hierarchy [Invoice Date Month], Operator [Equal] và lọc ra giá trị {4} ở ô Filter Expression.

Category Name	Brand Name	GiaSauGiamGia
Baby	Canbebe	8,85
Baby	Dalin	39,58
Baby	ETİ	3,33
Baby	Huggies	18,56
Baby	Molfix	35,42
Baby	Prima	56,66
Baby	Ülker	6,59
Baby	Uni	2,54
Beverage	F NEFFİS	59,82
Beverage	Akdem	7,35
Beverage	Aroma	4,99
Beverage	Beypazarı	2,22
Beverage	Cappy	9,86
Beverage	Çaykur	131,2
Beverage	Coca Cola	34,97
Beverage	Damlı	0,52
Beverage	Dinler	2,61

Hình 263: Dưa bảng Dim\_Time vào cửa sổ lọc để bổ sung thêm điều kiện thời gian và thu được kết quả câu truy vấn 4

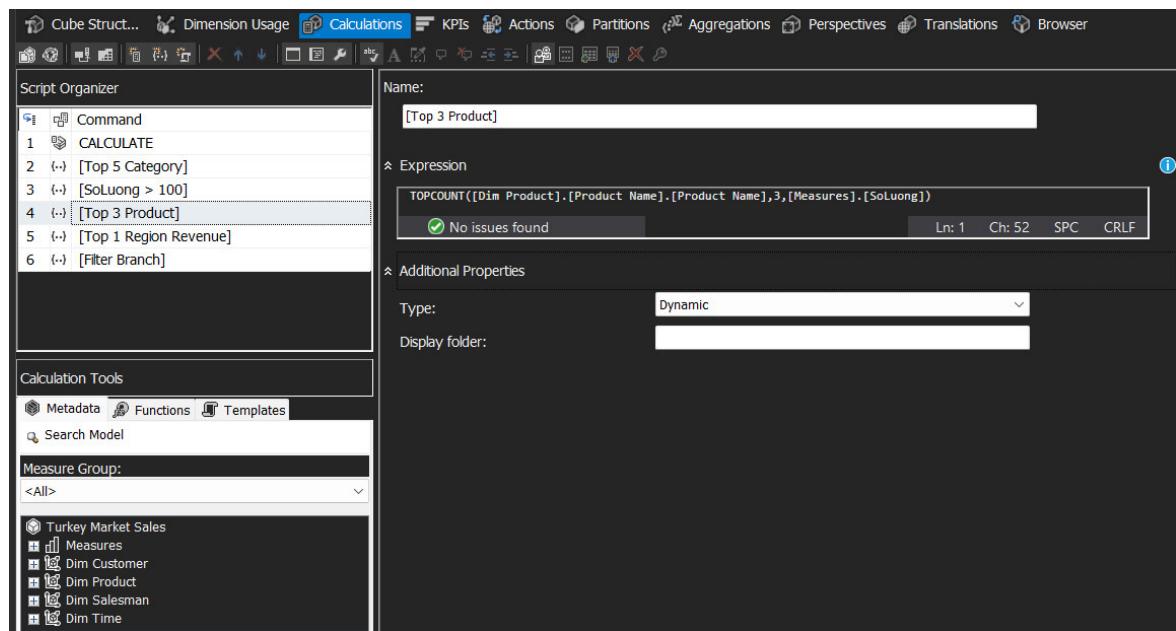
### 3.10.5. Câu truy vấn 5:

**Nội dung câu truy vấn:** Với 3 sản phẩm bán được với doanh số cao nhất theo từng ngày, tháng, năm, cho biết số lượng khách hàng đã mua 3 sản phẩm đó.

**Thực hiện câu truy vấn:**

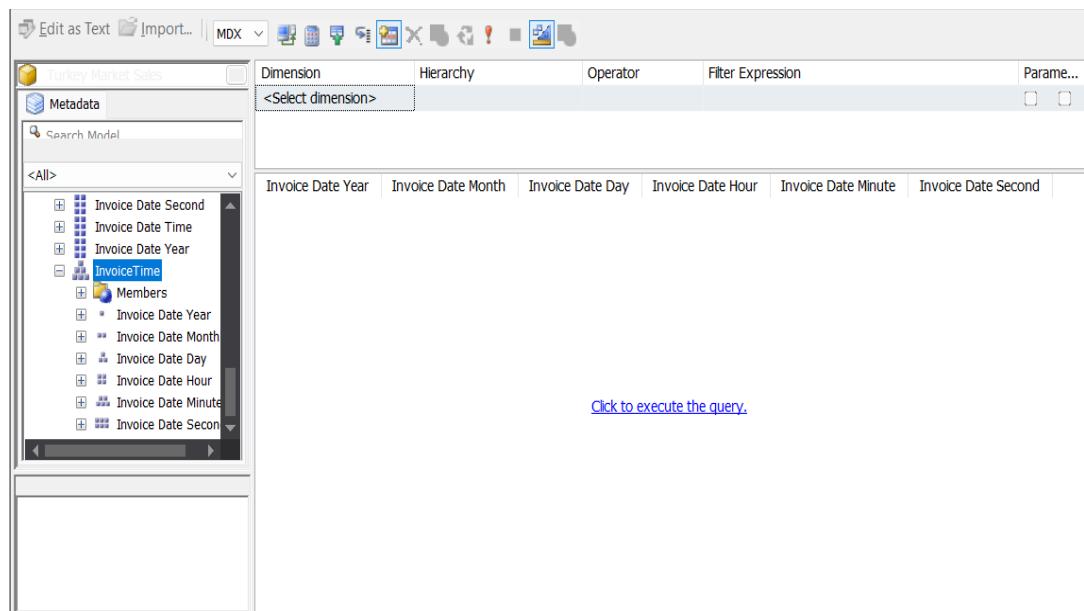
**Bước 1:** Tạo NameSet [Top 3 Product] như hình minh họa với biểu thức sau:

**TOPCOUNT([Dim Product].[Product Name].[ProductName], 3, [Measures].[SoLuong])**



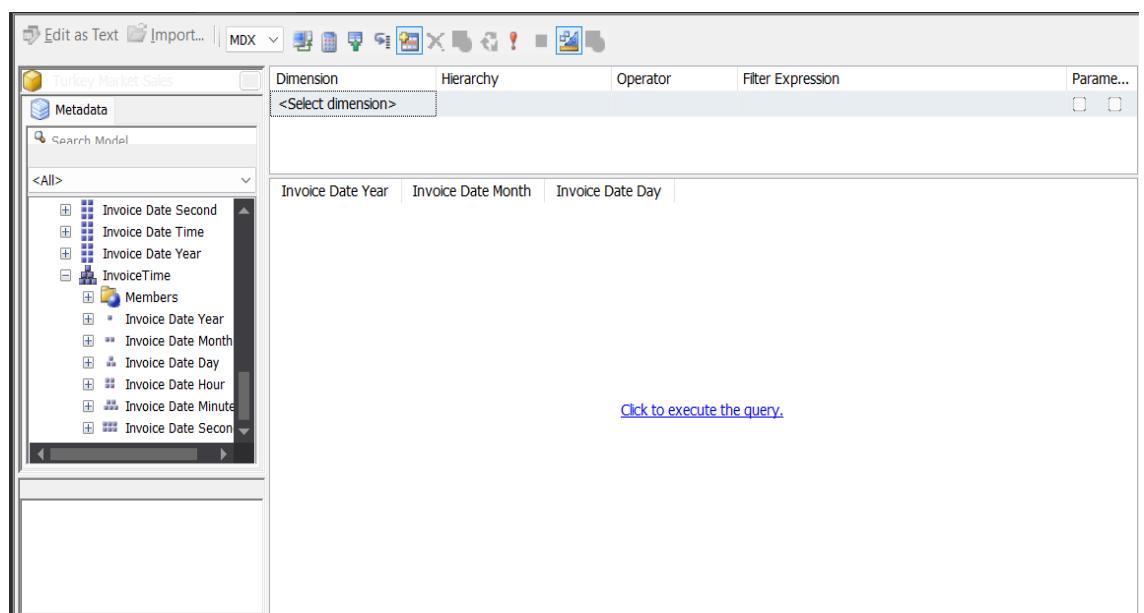
Hình 264: Khởi tạo NameSet [Top 3 Product]

**Bước 2:** Kéo thả Hierarchy [InvoiceTime] của bảng [Dim Time] vào cửa sổ thực thi câu truy vấn.



Hình 265: Kéo thả Hierarchy [InvoiceTime] vào cửa sổ thực thi câu truy vấn

**Bước 3:** Loại bỏ các thuộc tính không cần dùng [Invoice Date Hour], [Invoice Date Minute], [Invoice Date Second].



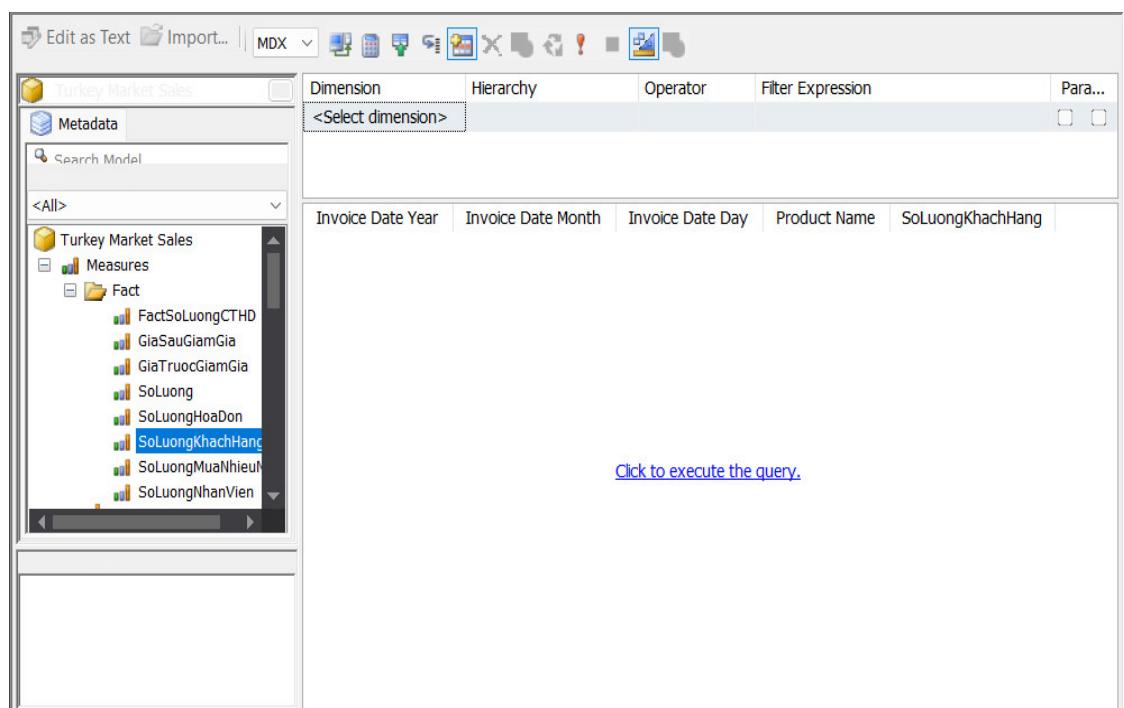
Hình 266: Loại bỏ một số thuộc tính không cần thiết

**Bước 4:** Tạo mới một độ đo [SoLuongKhachHang] trong bảng Fact.

Measures				
	Name	Measure Group	Data Type	Aggregation
Bill	SoLuongHoaDon	Fact	Integer	Count
Bill	SoLuong	Fact	Double	Sum
Bill	GiaTruocGiamGia	Fact	Double	Sum
Bill	GiaSauGiamGia	Fact	Double	Sum
Bill	FactSoLuongCTHD	Fact	Integer	Count
Bill	SoLuongMuaNhiuNhat	Fact	Double	Max
Bill	SoLuongKhachHang	Fact	Integer	DistinctCount

Hình 267: Tạo mới độ đo số lượng khách hàng (sử dụng aggregation là hàm DistinctCount)

**Bước 5:** Kéo thả thuộc tính [Product Name] của bảng [Dim Product] và độ đo [SoLuongKhachHang] vào cửa sổ thực thi câu truy vấn.



Hình 268: Kéo thả các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 6:** Kéo thả Dimension [Dim Product] vào cửa sổ lọc Hierarchy [ProductName], Operator [In] và chọn NameSet [Top 3 Product] ở ô Filter Expression.

The screenshot shows the SSAS MDX Query Editor interface after applying a filter. The 'Hierarchy' column in the query editor now shows 'Dim Product' with 'Product Name' selected. The 'Operator' column is set to 'In' and the 'Filter Expression' column contains 'Top 3 Product'. The results pane shows the following data:

Invoice Date Year	Invoice Date Month	Invoice Date Day	Product Name	SoLuongKhachHang
2017	1	10	MAGIC HANDS CIG ...	43
2017	1	10	NESCAFE 3 IN 1 18 ...	26
2017	1	10	OSMANCIK PIRINC ...	43
2017	1	11	MAGIC HANDS CIG ...	66
2017	1	11	NESCAFE 3 IN 1 18 ...	36
2017	1	11	OSMANCIK PIRINC ...	84
2017	1	12	MAGIC HANDS CIG ...	27
2017	1	12	NESCAFE 3 IN 1 18 ...	24
2017	1	12	OSMANCIK PIRINC ...	52
2017	1	13	MAGIC HANDS CIG ...	78
2017	1	13	NESCAFE 3 IN 1 18 ...	37
2017	1	13	OSMANCIK PIRINC ...	49
2017	1	14	MAGIC HANDS CIG ...	68
2017	1	14	NESCAFE 3 IN 1 18 ...	25
2017	1	14	OSMANCIK PIRINC ...	49
2017	1	15	MAGIC HANDS CIG ...	34

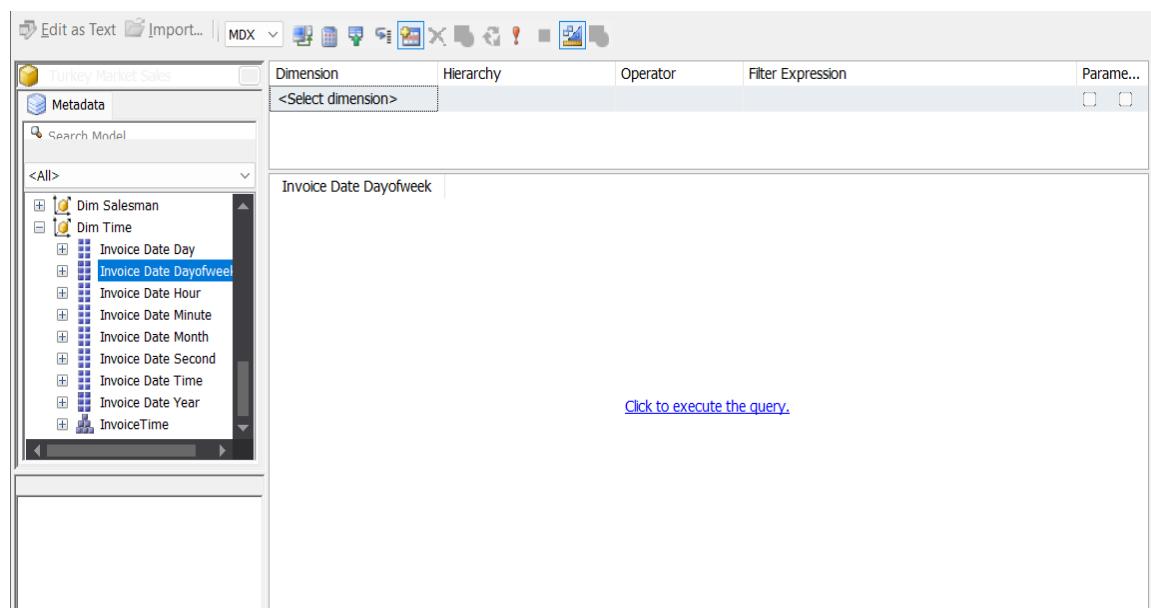
Hình 269: Thêm NameSet vào ô Filter Expression và thu được kết quả câu truy vấn 5

### 3.10.6. Câu truy vấn 6:

**Nội dung câu truy vấn:** Liệt kê tổng số lượng sản phẩm, số lượng khách hàng tại từng chi nhánh theo ngày trong tuần.

### Thực hiện câu truy vấn:

**Bước 1:** Kéo thả thuộc tính [Invoice Date Dayofweek] trong bảng [Dim Time] vào cửa sổ thực thi câu truy vấn.



Hình 270: Kéo thả thuộc tính liên quan vào cửa sổ thực thi câu truy vấn

**Bước 2:** Sửa lại độ đo [SoLuongKhachHang] bằng cách chọn giá trị “**Distinct Count**” tại mục “**Aggregation**”, các thuộc tính còn lại không được để giá trị “**Distinct Count**”.

Measures				
	Name	Measure Group	Data Type	Aggregation
SoLuongHoaDon	Fact	Integer	Count	
SoLuong	Fact	Double	Sum	
GiaTruocGiamGia	Fact	Double	Sum	
GiaSauGiamGia	Fact	Double	Sum	
FactSoLuongCTHD	Fact	Integer	Count	
SoLuongMuaNhiuNhat	Fact	Double	Max	
SoLuongKhachHang	Fact	Integer	DistinctCount	

Hình 271: Chỉnh sửa độ đo số lượng khách hàng (sử dụng aggregation là hàm DistinctCount)

**Bước 3:** Kéo thả thuộc tính [Branch Name] trong bảng phân cấp [Dim Branch] của bảng [Dim Product] và các độ đo [SoLuong], [SoLuongKhachHang] vào cửa sổ thực thi câu truy vấn và tiến hành thực thi.

The screenshot shows the SSAS MDX Editor interface. On the left, there is a navigation pane titled 'Turkey Market Sales' under 'Metadata'. It contains a tree view of measures and fact tables, including 'FactSoLuongCTHD', 'GiaSauGiamGia', 'GiaTruocGiamGia', 'SoLuong', 'SoLuongHoaDon', 'SoLuongKhachHang', 'SoLuongMuaNhiều', and 'SoLuongNhanVien'. On the right, a large grid displays the results of an MDX query. The columns are 'Invoice Date Dayofw...', 'Branch Name', 'SoLuong', 'SoLuongKhachHang', and '...'. The data includes various branch names like Adana Subesi, Adiyaman Subesi, Afyonkarahisar Subesi, Ağrı Subesi, Aksaray Subesi, Amasya Subesi, Ankara Subesi, Antalya Subesi, Ardahan Subesi, Artvin Subesi, Aydin Subesi, Balıkesir Subesi, Bartın Subesi, Batman Subesi, Bayburt Subesi, Bilecik Subesi, Bingöl Subesi, and so on, with their respective SoLuong and SoLuongKhachHang values.

Hình 272: Kéo thả độ đo liên quan và thu được kết quả câu truy vấn 6

### 3.10.7. Câu truy vấn 7:

**Nội dung câu truy vấn:** Thống kê số lượng hóa đơn được lập tại các chi nhánh nằm ở 3 thành phố Kırşehir, Nevşehir, Çankırı thuộc vùng İç Anadolu theo từng tháng.

**Thực hiện câu truy vấn:**

**Bước 1:** Kéo thả Hierarchy [DiaDiemChiNhanh] trong bảng [Dim Product] vào cửa sổ thực thi câu truy vấn.

The screenshot shows the SSAS MDX Editor interface. On the left, there is a navigation pane titled 'Turkey Market Sales' under 'Metadata'. It contains a tree view of dimensions, including 'Category Id', 'Category Name', 'City', 'Dim Product Id', 'Price', 'Product Id', 'Product Name', 'Region', 'DiaDiemChiNhanh', 'Dim Salesman', and 'Dim Time'. On the right, a grid displays the results of an MDX query. The columns are 'Region', 'City', and 'Branch Name'. Below the grid, there is a link labeled 'Click to execute the query.'

Hình 273: Kéo thả Hierarchy [DiaDiemChiNhanh] vào cửa sổ thực thi câu truy vấn

**Bước 2:** Tạo mới độ đo [SoLuongHoaDon] trong bảng Fact.

Measures				
	Name	Measure Group	Data Type	Aggregation
SoLuongHoaDon	Fact	Integer	DistinctCount	
SoLuong	Fact	Double	Sum	
GiaTruocGiamGia	Fact	Double	Sum	
GiaSauGiamGia	Fact	Double	Sum	
FactSoLuongCTHD	Fact	Integer	Count	
SoLuongMuaNhiuNhat	Fact	Double	Max	
SoLuongKhachHang	Fact	Integer	Count	

Hình 274: Tạo mới độ đo số lượng hóa đơn (sử dụng aggregation là hàm DistinctCount)

**Bước 3:** Kéo thả thuộc tính [Invoice Date Month] của bảng [Dim\_Time] và độ đo [SoLuongHoaDon] vào cửa sổ thực thi câu truy vấn.

Hình 275: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 4:** Kéo thả Dimension [Dim Product] vào cửa sổ lọc Hierarchy [DiaDiemChiNhanh], Operator [Equal] và lọc ra giá trị {Kırşehir, Nevşehir, Çankırı} ở ô Filter Expression.

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Product	Địa Điểm Chi Nhánh	Equal	{ Kirşehir, Nevşehir, Çankırı }	<input type="checkbox"/> <input type="checkbox"/>
<Select dimension>				<input type="checkbox"/> <input type="checkbox"/>

Region	City	Branch Name	Invoice Date Month	Số Lượng Hóa Đơn
İç Anadolu	Çankırı	Çankırı Subesi	1	71
İç Anadolu	Çankırı	Çankırı Subesi	2	82
İç Anadolu	Çankırı	Çankırı Subesi	3	100
İç Anadolu	Çankırı	Çankırı Subesi	4	4
İç Anadolu	Kirşehir	Kirşehir Subesi	1	93
İç Anadolu	Kirşehir	Kirşehir Subesi	2	98
İç Anadolu	Kirşehir	Kirşehir Subesi	3	120
İç Anadolu	Kirşehir	Kirşehir Subesi	4	7
İç Anadolu	Nevşehir	Nevşehir Subesi	1	118
İç Anadolu	Nevşehir	Nevşehir Subesi	2	132
İç Anadolu	Nevşehir	Nevşehir Subesi	3	137
İç Anadolu	Nevşehir	Nevşehir Subesi	4	2

Hình 276: Đưa điều kiện các thành phố được néu tên và thu được kết quả câu truy vấn 7

### 3.10.8. Câu truy vấn 8:

**Nội dung câu truy vấn:** Cho biết vùng miền có doanh thu cao nhất. Trong vùng miền này, liệt kê tên các chi nhánh có số lượng khách hàng ghé thăm ít hơn 3000.

**Thực hiện câu truy vấn:**

**Bước 1:** Tạo NameSet [Top 1 Region Revenue] để lấy được thông tin vùng miền có doanh thu cao nhất như hình minh họa với biểu thức như sau:

TOPCOUNT([Dim Product].[Region].[Region], 1, [Measures].[GiaSauGiamGia])

The screenshot shows the 'Script Organizer' window in SSAS Management Studio. The 'Name' field contains '[Top 1 Region Revenue]'. The 'Expression' field contains the DAX formula: `TOPCOUNT([Dim Product].[Region].[Region], 1, [Measures].[GiaSauGiamGia])`. A status bar at the bottom right indicates 'No issues found'. The 'Type' dropdown is set to 'Dynamic'. The 'Measure Group' dropdown is set to '<All>'. The 'Measure Group' pane on the left lists several measures: Turkey Market Sales, Measures, Dim Customer, Dim Product, Dim Salesman, and Dim Time.

Hình 277: Tạo NameSet [Top 1 Region Revenue]

**Bước 2:** Sửa lại độ đo [SoLuongKhachHang] bằng cách chọn giá trị “**Distinct Count**” tại mục “**Aggregation**”, các thuộc tính còn lại không được để giá trị “**Distinct Count**”.

Measures				
	Name	Measure Group	Data Type	Aggregation
SoLuongHoaDon	Fact	Integer	Count	
SoLuong	Fact	Double	Sum	
GiaTruocGiamGia	Fact	Double	Sum	
GiaSauGiamGia	Fact	Double	Sum	
FactSoLuongCTHD	Fact	Integer	Count	
SoLuongMuaNhiuNhat	Fact	Double	Max	
SoLuongKhachHang	Fact	Integer	DistinctCount	

Hình 278: Sửa lại độ đo số lượng khách hàng (sử dụng aggregation là hàm DistinctCount)

**Bước 3:** Tạo NameSet [Filter Branch] để lọc các chi nhánh có số lượng khách hàng ghé thăm bé hơn 3000 như hình minh họa với biểu thức sau:

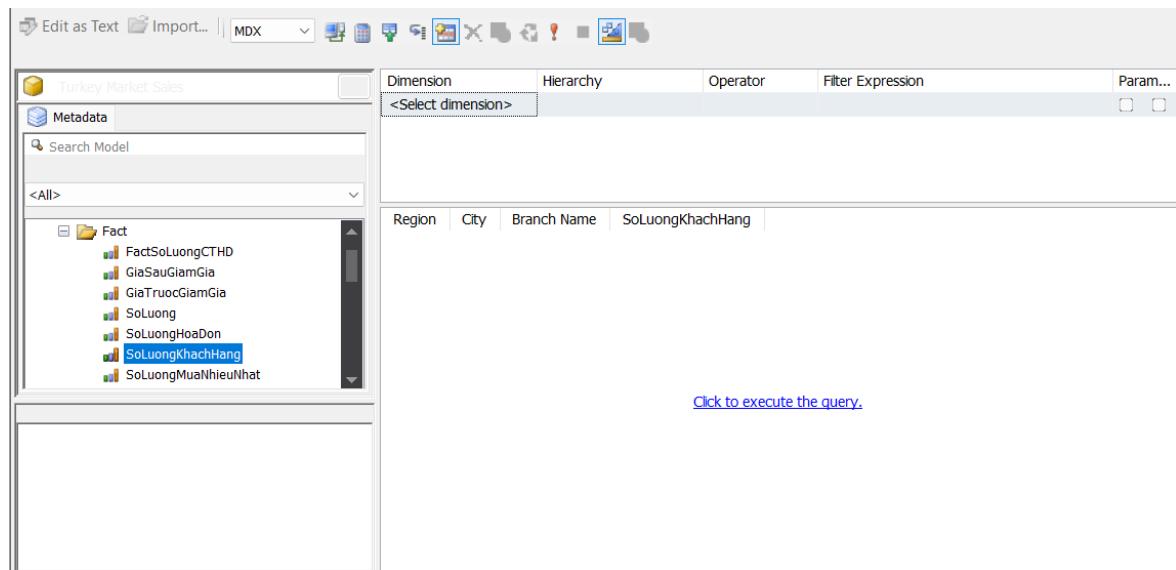
```

FILTER(
    {[Dim Product].[DiaDiemChiNhanh].[Branch Name].AllMEMBERS},
    [Measures].[SoLuongKhachHang] < 3000
)

```

Hình 279: Tạo NameSet [Filter Branch]

**Bước 4:** Ở mục Browser kéo thả các hierarchies [DiaDiemChiNhanh] của [Dim Product] và độ đo [SoLuongKhachHang] của bảng Fact vào cửa sổ thực thi câu truy vấn.



Hình 280: Dựa thuộc tính phân cấp và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 5:** Kéo thả Dimension [Dim Product] vào dòng đầu của cửa sổ lọc với Hierarchy [Region], Operator [In] và chọn NameSet [Top 1 Region Revenue] ở ô Filter Expression. Tiếp tục kéo thả Dimension [Dim Product] vào dòng thứ hai của cửa sổ lọc với Hierarchy [DiaDiemChiNhanh], Operator [In] và chọn NameSet [Filter Branch] ở ô Filter Expression.

Region	City	Branch Name	SoLuongKhachHang
Marmara	Bilecik	Bilecik Subesi	273
Marmara	Çanakkale	Çanakkale S...	618
Marmara	Edirne	Edirne Subesi	486
Marmara	Kirkareli	Kirkareli Subesi	460
Marmara	Kocaeli	Kocaeli Subesi	2349
Marmara	Sakarya	Sakarya Sub...	1233
Marmara	Tekirdağ	Tekirdağ Su...	1288
Marmara	Yalova	Yalova Subesi	310

Hình 281: Chọn 2 NameSet vừa tạo để lọc theo điều kiện và thu được kết quả câu truy vấn 8

### 3.10.9. Câu truy vấn 9:

**Nội dung câu truy vấn:** Thông kê tổng số lượng nhân viên tại từng chi nhánh trong từng vùng miền theo từng ngày.

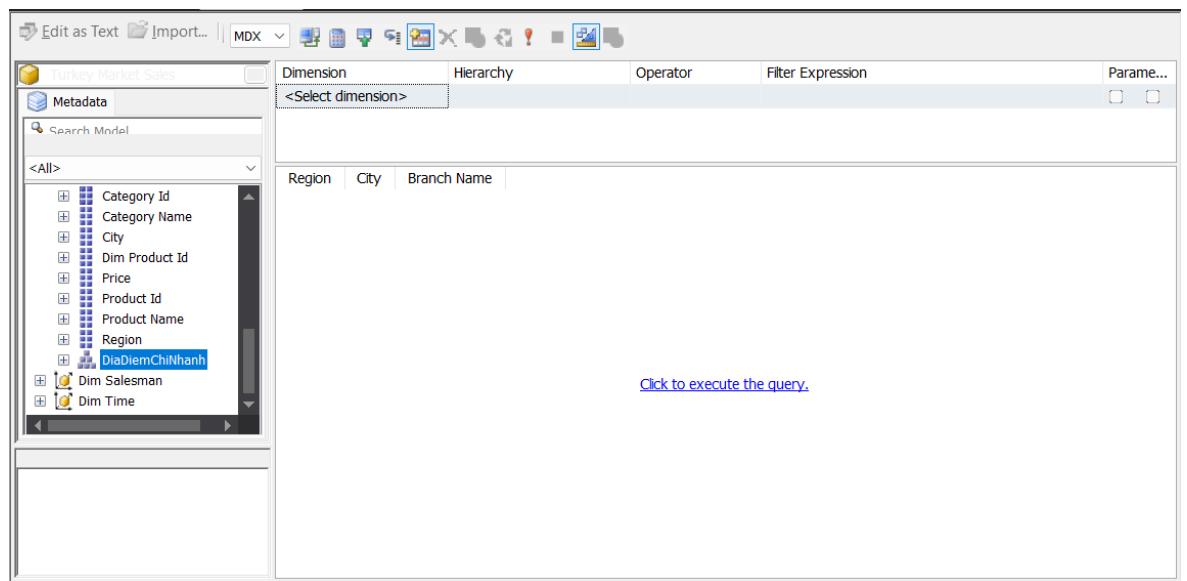
**Thực hiện câu truy vấn:**

**Bước 1:** Tạo mới độ đo [SoLuongNhanVien] trong bảng Fact.

Measures				
	Name	Measure Group	Data Type	Aggregation
SoLuongHoaDon	Fact	Integer	Count	
SoLuong	Fact	Double	Sum	
GiaTruocGiamGia	Fact	Double	Sum	
GiaSauGiamGia	Fact	Double	Sum	
FactSoLuongCTHD	Fact	Integer	Count	
SoLuongMuaNieuNhat	Fact	Double	Max	
SoLuongKhachHang	Fact	Integer	Count	
SoLuongNhanVien	Fact	Integer	DistinctCount	
Add new measure...				

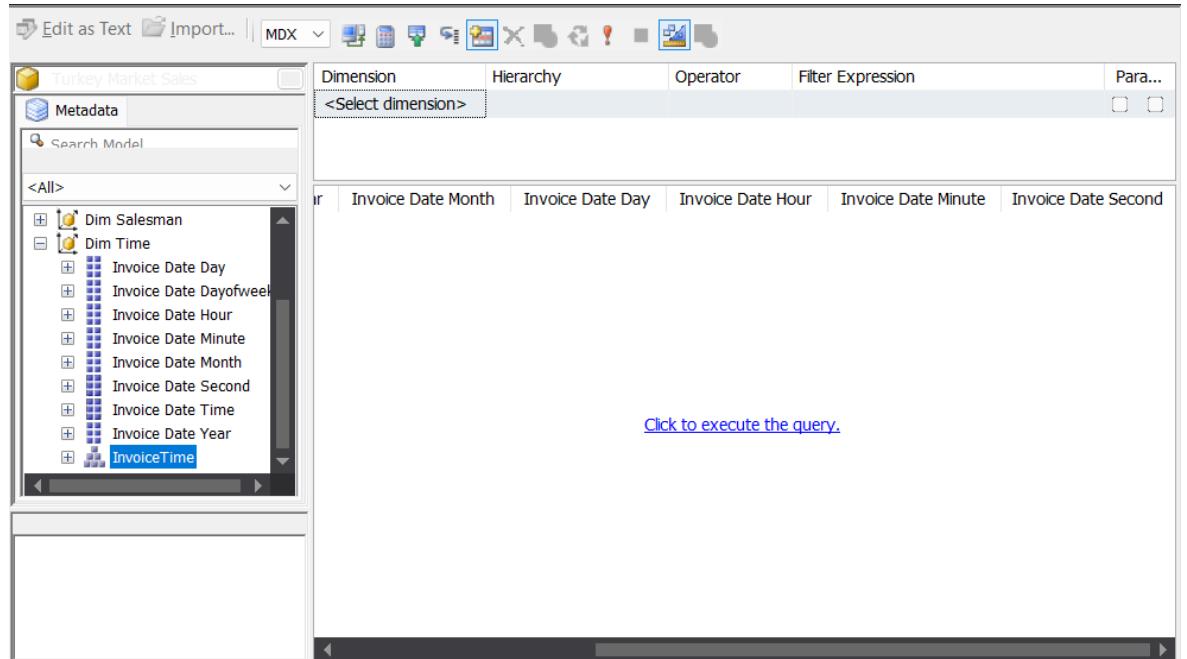
Hình 282: Tạo mới độ đo số lượng nhân viên (sử dụng aggregation là hàm DistinctCount)

**Bước 2:** Kéo thả hierarchy [DiaDiemChiNhanh] trong bảng [Dim Product] vào cửa sổ thực thi câu truy vấn.



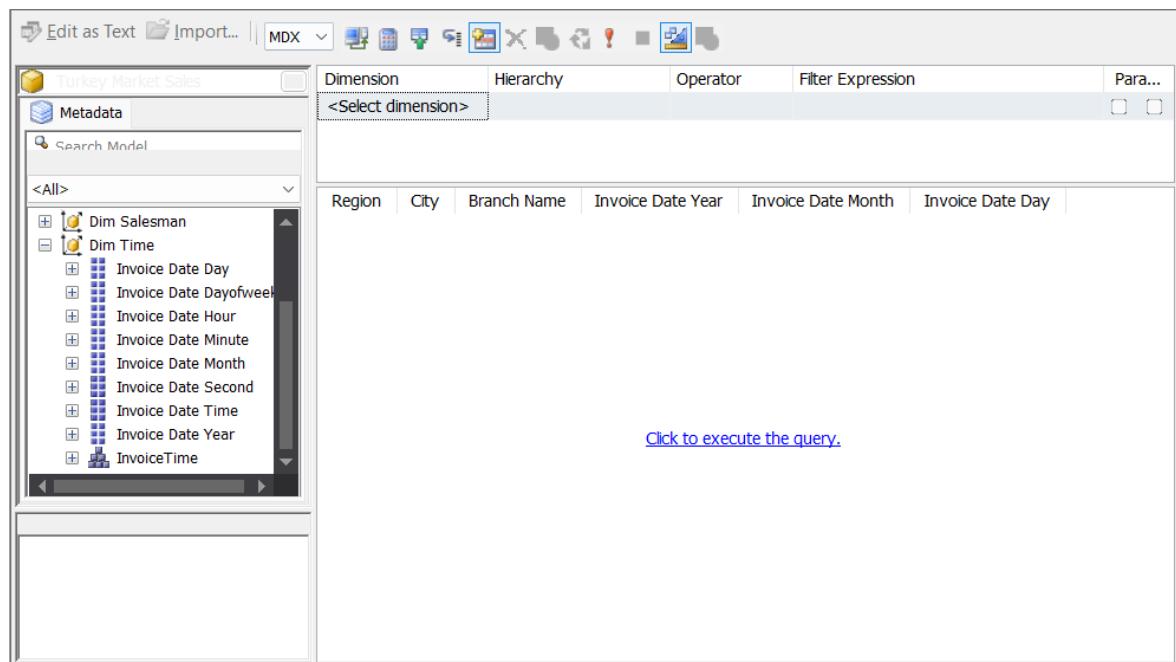
Hình 283: Kéo thả Hierarchy [DiaDiemChiNhanh] vào cửa sổ thực thi câu truy vấn

**Bước 3:** Kéo thả hierarchy [InvoiceTime] trong bảng [Dim Time] vào cửa sổ thực thi câu truy vấn.



Hình 284: Kéo thả Hierarchy [Invoice Time] vào cửa sổ thực thi câu truy vấn

**Bước 4:** Loại bỏ các thuộc tính dư thừa: [Invoice Date Hour], [Invoice Date Minute], [Invoice Date Second]



Hình 285: Loại bỏ các thuộc tính không cần thiết

**Bước 5:** Kéo thả độ đo [SoLuongNhanVien] vào cửa sổ thực thi câu truy vấn và bấm thực thi.

The screenshot shows the results of a query execution. The 'Dimension' pane now lists the 'Measures' dimension, specifically the 'Fact' folder which contains 'FactSoLuongCTHD', 'GiaSauGiamGia', 'GiaTruocGiamGia', 'SoLuong', 'SoLuongHoaDon', 'SoLuongKhachHang', 'SoLuongMuaNhiều', and 'SoLuongNhanVien'. The main workspace displays a table with the following data:

Invoic...	Invoi...	Invo...	Region	City	Branch Name	SoLuongNhanVien
2017	1	10	Akdeniz	Adana	Adana Subesi	9
2017	1	10	Akdeniz	Antalya	Antalya Subesi	8
2017	1	10	Akdeniz	Burdur	Burdur Subesi	5
2017	1	10	Akdeniz	Hatay	Hatay Subesi	3
2017	1	10	Akdeniz	Isparta	Isparta Subesi	6
2017	1	10	Akdeniz	Kahramanmaraş	Kahramanmaraş Subesi	2
2017	1	10	Akdeniz	Mersin	Mersin Subesi	5
2017	1	10	Akdeniz	Osmaniye	Osmaniye Subesi	6
2017	1	10	Doğu Anadolu	Ağrı	Ağrı Subesi	3
2017	1	10	Doğu Anadolu	Ardahan	Ardahan Subesi	2
2017	1	10	Doğu Anadolu	Bingöl	Bingöl Subesi	1
2017	1	10	Doğu Anadolu	Bitlis	Bitlis Subesi	3
2017	1	10	Doğu Anadolu	Elazığ	Elazığ Subesi	6
2017	1	10	Doğu Anadolu	Erzincan	Erzincan Subesi	1
2017	1	10	Doğu Anadolu	Erzurum	Erzurum Subesi	8
2017	1	10	Doğu Anadolu	Hakkari	Hakkari Subesi	4
2017	1	10	Doğu Anadolu	Tunceli	Tunceli Subesi	2

Hình 286: Kéo thả độ đo [SoLuongNhanVien] vào cửa sổ thực thi và thu được kết quả câu truy vấn 9

### 3.10.10. Câu truy vấn 10:

**Nội dung câu truy vấn:** Cho biết tổng doanh số bán hàng của vùng miền Güneydoğu Anadolu theo từng tháng.

**Thực hiện câu truy vấn:**

**Bước 1:** Kéo thả thuộc tính [Region] của Hierarchies có tên là [DiaDiemChiNhanh] sang cửa sổ lọc dữ liệu. Tại ô Filter Expression chọn Güneydoğu Anadolu.

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Product	DiaDiemChiNhanh	Equal	{ Güneydoğu Anadolu }	<input type="checkbox"/> <input checked="" type="checkbox"/>
<Select dimension>				

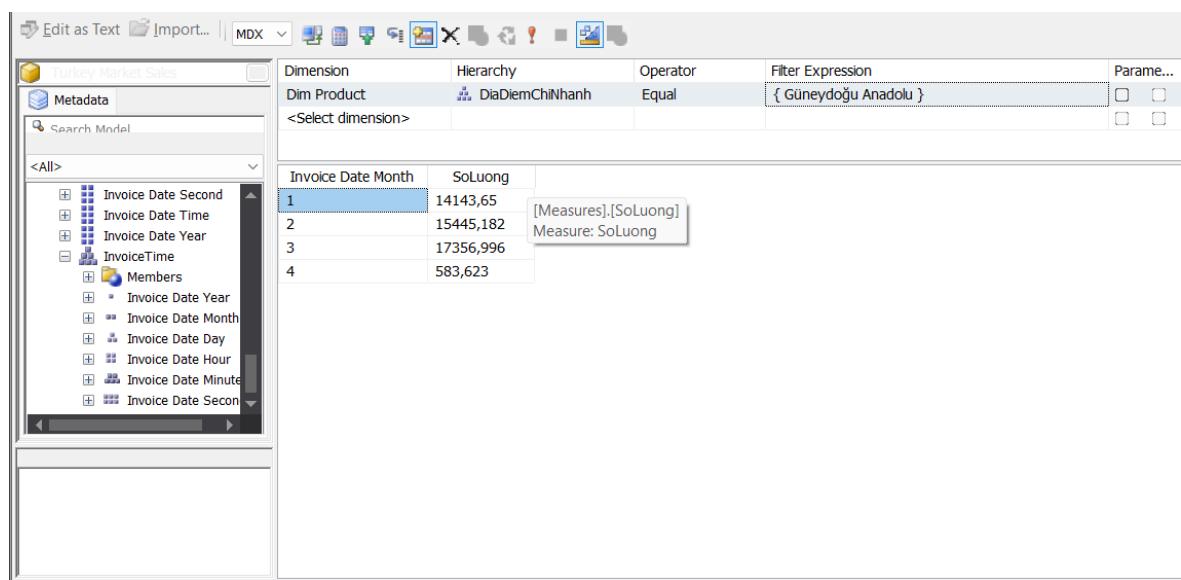
Hình 287: Dựa thuộc tính phân cấp [DiaDiemChiNhanh] và lọc theo điều kiện tại cửa sổ lọc

**Bước 2:** Kéo độ đo [SoLuong] vào cửa sổ thực thi câu truy vấn.

Dimension	Hierarchy	Operator	Filter Expression	Param...
Dim Product	DiaDiemChiNhanh	Equal	{ Güneydoğu Anadolu }	<input type="checkbox"/> <input checked="" type="checkbox"/>
<Select dimension>				

Hình 288: Dựa độ đo [SoLuong] vào cửa sổ thực thi câu truy vấn

**Bước 3:** Kéo thả thuộc tính [Invoice Date Month] của bảng [Dim Time] vào cửa sổ thực thi câu truy vấn và tiến hành thực thi.



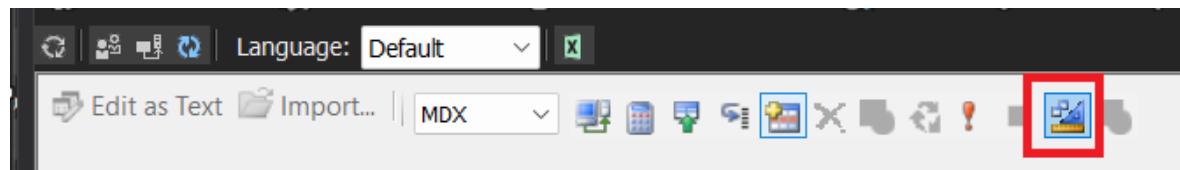
Hình 289: Đưa thuộc tính [Invoice Date Month] vào cửa sổ thực thi câu truy vấn và thu được kết quả

### 3.10.11. Câu truy vấn 11:

**Nội dung câu truy vấn:** Cho biết thông tin khách hàng (mã khách hàng, tên khách hàng, giới tính) có tổng số tiền mua lớn nhất trong từng tháng.

**Thực hiện câu truy vấn:**

**Bước 1:** Chuyển từ “Design Mode” sang “Script Mode” bằng cách click vào biểu tượng “Design Mode” như hình minh họa



Hình 290: Giao diện chuyển sang Script Mode

**Bước 2:** Thay thế câu truy vấn MDX đã chuẩn bị vào ô truy vấn

```
WITH MEMBER [Measures].[Top Customer Sales] AS
'[Measures].[GiaSauGiamGia], [Dim Customer].[Customer ID].CurrentMember'
SELECT
{[Measures].[Top Customer Sales]} ON COLUMNS,
NON EMPTY GENERATE(
[Dim Time].[Invoice Date Month].CHILDREN,
TopCount(
[Dim Time].[Invoice Date Month].CurrentMember *
[Dim Customer].[Customer ID].CHILDREN *
[Dim Customer].[Customer Name].CHILDREN *
[Dim Customer].[GENDER].CHILDREN,
1,
([Measures].[GiaSauGiamGia])
)
) ON ROWS
FROM [Dim Product]
```

Hình 291: Nhập câu truy vấn MDX đã chuẩn bị trước vào script

**Bước 3:** Click “Execute Query” để tiến hành thực thi câu truy vấn và thu được kết quả

The screenshot shows the SSAS Management Studio environment. On the left, the Script Organizer pane lists several calculated members and measures, including [PHANTRAMGIAKHUYENMAI]. In the center, the MDX Editor pane displays the following query:

```

WITH MEMBER [Measures].[Top Customer Sales] AS
  {[Measures].[GiaSauGiamGia], [Dim Customer].[Customer ID].CurrentMember}
SELECT
  {[Measures].[Top Customer Sales]} ON COLUMNS,
  NON EMPTY( GENERATE(
    [Dim Time].[Invoice Date Month].CHILDREN,
    TopCount(
      [Dim Time].[Invoice Date Month].CurrentMember *
      [Dim Customer].[Customer ID].CHILDREN *
      [Dim Customer].[Customer Name].CHILDREN *
      [Dim Customer].[GENDER].CHILDREN,
      1,
      {[Measures].[GiaSauGiamGia]}
    )
  ) ) ON ROWS
  
```

The results pane on the right shows a table with four rows of data:

Invoice Date Month	Customer Id	Customer Name	GENDER	Top Customer Sales
1	808718	Mustafa SEVİLİR	Male	1219,66
2	420299	Nazmiye AYDIL	Female	828,96
3	237685	Atakan OLGUN	Male	2012,75
4	638391	Sevim ÇANAKÇI	Female	303,51

Hình 292: Tiến hành thực thi câu truy vấn và thu được kết quả câu truy vấn 11

### 3.10.12. Câu truy vấn 12:

**Nội dung câu truy vấn:** Đưa ra phần trăm giảm giá của từng chi nhánh trong tháng 01 và tháng 04 năm 2017.

**Thực hiện câu truy vấn:**

**Bước 1:** Tạo độ đo tính toán phần trăm giảm giá khuyến mãi như hình minh họa:

$$[\text{PHANTRAMGIAKHUYENMAI}] = 1 - \frac{[\text{Measures}].[\text{GiaSauGiamGia}]}{[\text{Measures}].[\text{GiaTruocGiamGia}]}$$

The screenshot shows the SSAS Management Studio Script Organizer pane. A new measure named [PHANTRAMGIAKHUYENMAI] is being created. The expression for the measure is:

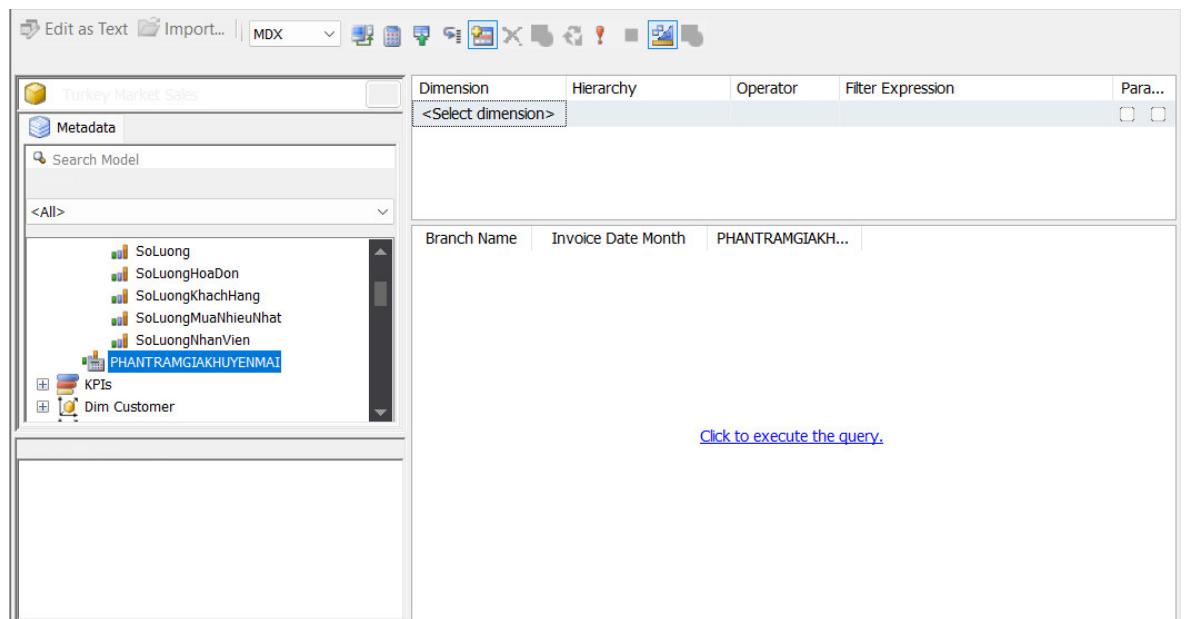
$$1 - ([\text{Measures}].[\text{GiaSauGiamGia}] / [\text{Measures}].[\text{GiaTruocGiamGia}])$$

The properties for the measure include:

- Name: [PHANTRAMGIAKHUYENMAI]
- Parent Properties: Parent hierarchy: Measures, Parent member: (empty)
- Expression: Format string: Percent, Visible: True, Non-empty behavior: (empty), Associated measure group: (Undefined), Display folder: (empty)

Hình 293: Tao mới Measure [PHANTRAMGIAKHUYENMAI]

**Bước 2:** Kéo thả thuộc tính [Branch Name] trong bảng phân cấp [Dim Branch] của bảng [Dim Product], thuộc tính [Invoice Date Month] của bảng [Dim Time], và độ đo [PHANTRAMGIAKHUYENMAI] vào cửa sổ thực thi câu truy vấn.



Hình 294: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 3:** Kéo thả Hierarchies có tên là [InvoiceTime] sang cửa sổ lọc dữ liệu, Operator [Equal] và chọn {1;4} tại ô Filter Expression.

Branch Name	Invoice Date Month	PHANTRAMGIAKH...
Adana Subesi	1	0,083578160843...
Adana Subesi	4	0,091143164262...
Adiyaman S...	1	0,087320246953...
Adiyaman S...	4	0,114898989898...
Afyonkarahis...	1	0,088762714309...
Afyonkarahis...	4	0,074375213967...
Ağrı Subesi	1	0,087311684161...
Ağrı Subesi	4	0,080992907801...
Aksaray Subesi	1	0,082151615310...
Aksaray Subesi	4	0,098503074044...
Amasya Subesi	1	0,077510677013...
Amasya Subesi	4	0,084640883977...
Ankara Subesi	1	0,086690058993...
Ankara Subesi	4	0,083453655444...
Antalya Subesi	1	0,08441611790024...

Hình 295: Đưa điều kiện các tháng được yêu cầu và thu được kết quả câu truy vấn 12

### 3.10.13. Câu truy vấn 13:

**Nội dung câu truy vấn:** Cho biết tên sản phẩm, kho sản phẩm có tổng số lượng bán ra lớn hơn 10 tại từng vùng miền.

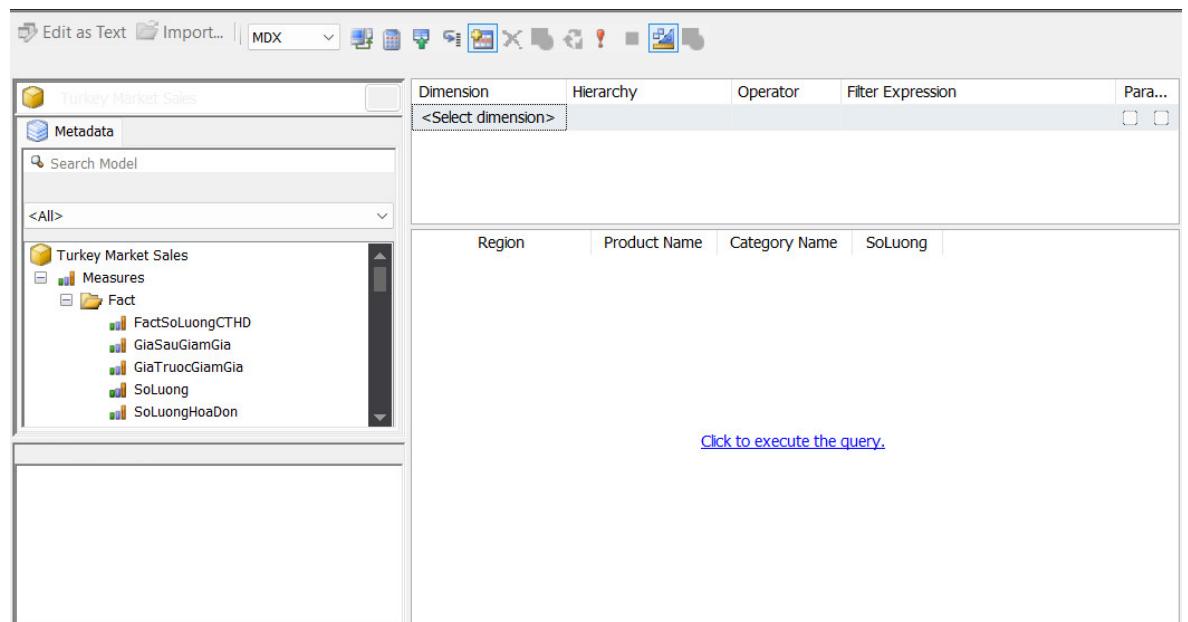
**Bước 1:** Sử dụng hàm GENERATE và CURRENTMEMBER để thực hiện câu truy vấn này:

```

GENERATE ( [Dim Product].[Region].Children, [Dim
Product].[Region].CURRENTMEMBER*
    FILTER(([Dim Product].[Dim Product ID].CHILDREN) * ([Dim
Product].[Product Name].CHILDREN) * [Dim Product].[Category
Name].CHILDREN, [Measures].[SoLuong] > 10)

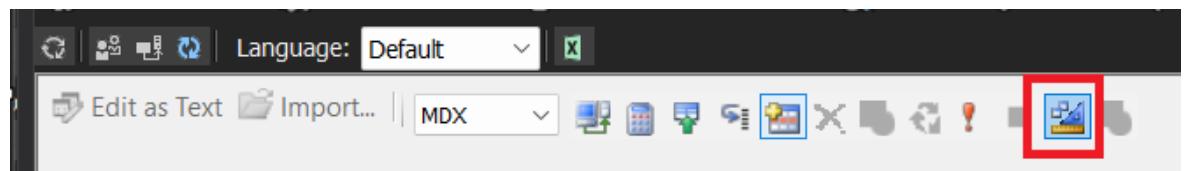
```

**Bước 2:** Kéo thả các thuộc tính [Region] trong bảng phân cấp [Dim Branch] của bảng [Dim Product], [Product Name] của bảng [Dim Product], [Category Name] trong bảng phân cấp [Dim Category] của bảng [Dim Product] và độ đo [SoLuong] vào cửa sổ thực thi câu truy vấn.



Hình 296: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 3:** Chuyển từ “Design Mode” sang “Script Mode” bằng cách click vào biểu tượng “Design Mode” như hình minh họa.



Hình 297: Giao diện chuyển sang Script Mode

**Bước 4:** Xóa bỏ dòng tô xanh như hình minh họa và dán đoạn code ở bước 1 vào vị trí vừa xóa.

```
SELECT NON EMPTY { [Measures].[SoLuong] } ON COLUMNS, NON EMPTY { ([Dim Product].[Region],
[Region].ALLMEMBERS * [Dim Product].[Product Name].[Product Name].ALLMEMBERS * [Dim Product].[Category
Name].[Category Name].ALLMEMBERS ) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME
ON ROWS FROM [Turkey Market Sales] CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR,
FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

Hình 298: Xóa đoạn code được tô đậm

```
SELECT NON EMPTY { [Measures].[SoLuong] } ON COLUMNS, NON EMPTY { GENERATE( [Dim Product],
[Region].Children, [Dim Product].[Region].CURRENTMEMBER*
FILTER(([Dim Product].[Dim Product ID]. CHILDREN)*([Dim Product].[Product
Name].CHILDREN)*[Dim Product].[Category Name].Children, [Measures].[SoLuong] > 10))
} DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [Turkey Market Sales]
CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME,
FONT_SIZE, FONT_FLAGS
```

Hình 299: Thay thế đoạn code ở bước 1 vào vị trí vừa xóa

**Bước 5:** Nhấn “Execute Query” tại cửa sổ thực thi câu truy vấn và thu được kết quả.

The screenshot shows the Analysis Services Management Studio interface. The top bar includes 'Edit as Text', 'Import...', 'MDX', and other standard toolbar icons. The main area has two panes: the left pane displays the 'Model Explorer' with objects like 'Turkey Market Sales', 'Metadata', 'Functions', and various dimension and fact tables; the right pane displays the 'Results' grid. The query in the editor is:

```
SELECT NON EMPTY { [Measures].[SoLuong] } ON COLUMNS, NON EMPTY { GENERATE( [Dim Product],
[Region].Children, [Dim Product].[Region].CURRENTMEMBER*
FILTER(([Dim Product].[Dim Product ID]. CHILDREN)*([Dim Product].[Product
Name].CHILDREN)*[Dim Product].[Category Name].Children, [Measures].[SoLuong] > 10))
} DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [Turkey Market Sales]
CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME,
FONT_SIZE, FONT_FLAGS
```

The results grid shows data from the 'Turkey Market Sales' cube, filtered by Region (Akdeniz) and Product Category (Beverage), with SoLuong values greater than 10. The data is as follows:

Region	Dim Product Id	Product Name	Category Name	SoLuong
Akdeniz	2575	CAFE CROWN FOAMED...	Beverage	22
Akdeniz	2675	SAKA WATER 1.5 LITER	Beverage	34
Akdeniz	2739	MEHMET EFENDI COFF...	Beverage	57
Akdeniz	2767	CAFE CROWN HOT CH...	Beverage	26
Akdeniz	2797	ICIM AYRAN 200 GR	Beverage	25
Akdeniz	2835	MEHMET EFENDI COFF...	Beverage	11
Akdeniz	2861	KAY WATER 5 LT	Beverage	12
Akdeniz	2907	MEHMET EFENDI COFF...	Beverage	36
Akdeniz	3001	BURC NATURAL SPRIN...	Beverage	11
Akdeniz	3050	CAFE CROWN HOT CH...	Beverage	15

Hình 300: Tiến hành thực thi câu truy vấn và thu được kết quả câu truy vấn 13

### 3.10.14. Câu truy vấn 14:

**Nội dung câu truy vấn:** Với từng vùng miền, liệt kê số lượng đơn hàng bán ra theo từng tháng ngoại trừ hai vùng miền ‘Doğu Anadolu’, ‘Karadeniz’.

**Thực hiện câu truy vấn:**

**Bước 1:** Sửa lại giá trị tại mục “Aggregation” của độ đo [SoLuongHoaDon] thành “DistinctCount”. Các thuộc tính khác chuyển từ “DistinctCount” về “Count”.

Measures				
	Name	Measure Group	Data Type	Aggregation
SoLuongHoaDon	SoLuongHoaDon	Fact	Integer	DistinctCount
SoLuong	SoLuong	Fact	Double	Sum
GiaTruocGiamGia	GiaTruocGiamGia	Fact	Double	Sum
GiaSauGiamGia	GiaSauGiamGia	Fact	Double	Sum
FactSoLuongCTHD	FactSoLuongCTHD	Fact	Integer	Count
SoLuongMuaNhiuNhat	SoLuongMuaNhiuNhat	Fact	Double	Max
SoLuongKhachHang	SoLuongKhachHang	Fact	Integer	Count

Hình 301: Sửa lại độ đo số lượng hóa đơn (sử dụng aggregation là hàm DistinctCount)

**Bước 2:** Tạo NameSet[DieuKienCau14] để loại bỏ hai vùng miền ‘Doğu Anadolu’, ‘Karadeniz’ ra khỏi kết quả câu truy vấn:

EXCEPT([Dim Product].[Region].CHILDREN, {[Dim Product].[Region]&[Doğu Anadolu], [Dim Product].[Region]&[Karadeniz]})

Hình 302: Tạo NameSet [DieuKienCau14]

**Bước 3:** Kéo thả thuộc tính [Region] trong bảng phân cấp [Dim Branch] của bảng [Dim Product], [Invoice Date Month] trong bảng [Dim Time] và độ đo [SoLuongHoaDon] vào cửa sổ thực thi câu truy vấn.

Hình 303: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 4:** Kéo thả Dimension [Dim Product] vào cửa sổ lọc Hierarchy [Region], Operator [In] và chọn NameSet [DieuKienCau14] ở ô Filter Expression.

Invoice Date Month	Region	SoLuongHoaDon
1	Akdeniz	3867
1	Ege	4458
1	Güneydoğu Anadolu	3252
1	İç Anadolu	4891
1	Marmara	9038
2	Akdeniz	4211
2	Ege	4727
2	Güneydoğu Anadolu	3596
2	İç Anadolu	5382
2	Marmara	9749
3	Akdeniz	4882
3	Ege	5571
3	Güneydoğu Anadolu	3993
3	İç Anadolu	6036
3	Marmara	11200

Hình 304: Thêm NameSet vừa tạo vào ô Filter Expression và thu được kết quả câu truy vấn 14

### 3.10.15. Câu truy vấn 15:

**Nội dung câu truy vấn:** Cho biết nhân viên thanh toán được nhiều hóa đơn nhất và ít hóa đơn nhất trong từng tháng.

**Thực hiện câu truy vấn:**

**Bước 1:** Sửa lại giá trị tại mục “**Aggregation**” của độ đo [SoLuongHoaDon] thành “**DistinctCount**”. Các thuộc tính khác chuyển từ “**DistinctCount**” về “**Count**”.

Measures				
	Name	Measure Group	Data Type	Aggregation
SoLuongHoaDon	Fact	Integer	DistinctCount	
SoLuong	Fact	Double	Sum	
GiaTruocGiamGia	Fact	Double	Sum	
GiaSauGiamGia	Fact	Double	Sum	
FactSoLuongCTHD	Fact	Integer	Count	
SoLuongMuaNhiuNhat	Fact	Double	Max	
SoLuongKhachHang	Fact	Integer	Count	

Hình 305: Sửa lại độ đo số lượng hóa đơn (sử dụng aggregation là hàm DistinctCount)

**Bước 2:** Sử dụng hàm UNION(TopCount, BottomCount) cho câu hỏi này:

Union (

TopCount({[Dim Salesman].[SalesmanId].Children \* [Dim Salesman].[Salesman Name].Children}, 1, [Measures].[SoLuongHoaDon]),

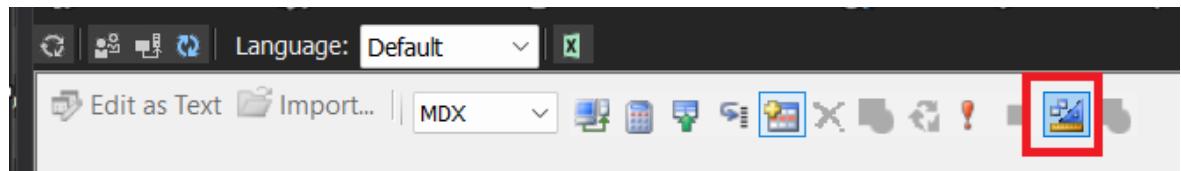
BottomCount({[Dim Salesman].[SalesmanId].Children \* [Dim Salesman].[Salesman Name].Children}, 1, not isempty([Measures].[SoLuongHoaDon])))

**Bước 3:** Kéo thả các thuộc tính [Salesman Name] của bảng [Dim Salesman] và độ đo [SoLuongHoaDon] vào cửa sổ thực thi câu truy vấn.

Salesman Name	SoLuongHoaDon
Abdullah BENGIN	520
Abdusamet TACIM	3061
Abdurrahman CAVDARLI	250
Adil ONDIN	601
Adnan YAGIZ	2962
Ahmet ELKATMIŞ	2
Aleyna ÇUKUREL	355
Aleyna ERÇETİN	278
Aleyna KASALAK	138
Ali Eymen GOKÇESU	53
Ali İLTEKİN	131
Aliye ÇÜKURÇAYIR	184
Aliye GÖKDOĞAN	1662
Aliye KOZEN	71

Hình 306: Kéo các thuộc tính và độ đo liên quan vào cửa sổ thực thi câu truy vấn

**Bước 4:** Chuyển từ “**Design Mode**” sang “**Script Mode**” bằng cách click vào biểu tượng “**Design Mode**” như hình minh họa.



Hình 307: Giao diện chuyển sang Script Mode

**Bước 5:** Xóa bỏ dòng tô xanh như hình minh họa và dán đoạn code ở bước 2 vào vị trí vừa xóa.

```
SELECT NON EMPTY { [Measures].[SoLuongHoaDon] } ON COLUMNS, NON EMPTY { ([Dim Salesman].[Salesman Name].[Salesman Name].ALLMEMBERS) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [Turkey Market Sales] CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

Hình 308: Xóa đoạn code được tô đậm

```
SELECT NON EMPTY { [Measures].[SoLuongHoaDon] } ON COLUMNS, NON EMPTY { Union ( TopCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1, [Measures].[SoLuongHoaDon]), BottomCount({[Dim Salesman].[Salesman Id].Children *[Dim Salesman].[Salesman Name].Children}, 1, not isempty([Measures].[SoLuongHoaDon]))) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [Turkey Market Sales] CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

Hình 309: Thay thế đoạn code ở bước 2 vào vị trí vừa xóa

**Bước 6:** Nhấn “**Execute Query**” tại cửa sổ thực thi câu truy vấn và thu được kết quả.

The screenshot shows the Microsoft Analysis Services MDX Editor interface. On the left, there is a navigation pane with a tree view of the cube structure. The tree includes 'Turkey Market Sales' (Measures, Fact, FactSoLuongCTHD, FactSoLuong, FactSoLuongKhachHang, FactSoLuongMuaNhanh, FactSoLuongLienVien, PHANTRAMGIAKHUYEN, KPIs, Dim Customer, Dim Product, DieukienCau14, Filter Branch, SoLuong > 100), 'Dim Time' (Invoice Date Month), 'Dim Product' (Category Name), and 'Dim Salesman' (Salesman Name). The main area contains the MDX query:

```
SELECT NON EMPTY { [Measures].[SoLuongHoaDon] } ON COLUMNS, NON EMPTY { Union ( TopCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1, [Measures].[SoLuongHoaDon]), BottomCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1, not isempty([Measures].[SoLuongHoaDon]))) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [Turkey Market Sales] CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

Below the query, the results are displayed in a table:

Salesman Id	Salesman Name	SoLuongHoaDon
1224	Yunus Emre AVUZ	3310
1266	Zübeyde PEKGÜLER	29

Hình 310: Tiến hành thực thi câu truy vấn và thu được kết quả câu truy vấn 15

### 3.11. Quá trình phân tích dữ liệu bằng ngôn ngữ MDX:

#### 3.11.1. Câu truy vấn 1:

**Nội dung câu truy vấn:** Với top 5 kho sản phẩm có doanh số bán ra cao nhất, thống kê số lượng sản phẩm bán ra được của 5 kho đó theo từng tháng.

**Câu truy vấn MDX:**

```
SELECT
NON EMPTY [Dim Time].[Invoice Date Month].[Invoice Date Month].MEMBERS ON COLUMNS,
NON EMPTY {TOPCOUNT([Dim Product].[Category Name].[Category Name],5,[Measures].[SoLuong])} ON ROWS
FROM [Turkey Market Sales]
WHERE [Measures].[SoLuong];
```

**Kết quả thực thi:**

```

--1. Với top 5 kho sản phẩm có doanh số bán ra cao nhất,
--thống kê số lượng sản phẩm bán ra được của 5 kho đó theo từng tháng
SELECT
NON EMPTY [Dim Time].[Invoice Date Month].[Invoice Date Month].MEMBERS ON COLUMNS,
NON EMPTY {TOPCOUNT([Dim Product].[Category Name].[Category Name],5,[Measures].[SoLuong])}
FROM [Turkey Market Sales]
WHERE [Measures].[SoLuong];

--2. Thống kê tổng số lượng sản phẩm bán được của
--các nhãn hàng "Sprite", "Clear", "Nestle", "Lipton", "Colgate" trong tháng 02
SELECT
NON EMPTY{[Dim Product].[Brand Name].[Sprite],
[Dim Product].[Brand Name].[Clear],
[Dim Product].[Brand Name].[Nestle],
[Dim Product].[Brand Name].[Lipton],
[Dim Product].[Brand Name].[Colgate]} ON ROWS,
NON EMPTY {[Measures].[SoLuong]} ON
COLUMNS

```

	1	2	3	4
Food	59946.0990000004	63269.2240000001	73796.747	2487.634
Beverage	31939	34851	41651	1375
Dairy Breakfast	19488.873	21499.172	24729.466	887.83
Detergent Cleaning	6071	6687	7595	294
Cosmetics	5862	6839	7569	244

Hình 311: Kết quả câu truy vấn 1

**Nhận xét:** Kết quả thực thi câu truy vấn 1 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.2. Câu truy vấn 2:

**Nội dung câu truy vấn:** Thống kê tổng số lượng sản phẩm bán được của các nhãn hàng “Sprite”, “Clear”, “Nestle”, “Lipton”, “Colgate” trong tháng 02.

**Câu truy vấn MDX:**

```

SELECT
NON EMPTY{[Dim Product].[Brand Name].[Sprite],
[Dim Product].[Brand Name].[Clear],
[Dim Product].[Brand Name].[Nestle],
[Dim Product].[Brand Name].[Lipton],
[Dim Product].[Brand Name].[Colgate]} ON ROWS,
NON EMPTY {[Measures].[SoLuong]} ON
COLUMNS

```

```

FROM [Turkey Market Sales]
WHERE [Dim Time].[Invoice Date Month].[2];

```

**Kết quả thực thi:**

```

MDXQuery1.mdx...HOAMUIMO\ADMIN*
Cube: Turkey Market Sales
Measure Group: <All>
    Turkey Market Sales
    Measures
    KPIs
    Dim Customer
    Dim Product
    Dim Salesman
    Dim Time

WHERE [Measures].[SoLuong];
--2. Thống kê tổng số lượng sản phẩm bán được
--của các nhãn hàng "Sprite", "Clear", "Nestle", "Lipton", "Colgate" trong tháng 02
SELECT
NON EMPTY {[Dim Product].[Brand Name].[Sprite],
[Dim Product].[Brand Name].[Clear],
[Dim Product].[Brand Name].[Nestle],
[Dim Product].[Brand Name].[Lipton],
[Dim Product].[Brand Name].[Colgate]} ON ROWS,
NON EMPTY {[Measures].[SoLuong]} ON COLUMNS
FROM [Turkey Market Sales]
WHERE [Dim Time].[Invoice Date Month].[2];
--3. Với mỗi thành phố, thống kê số lượng sản phẩm
--bán ra được ở các chi nhánh có số lượng sản phẩm bán ra lớn hơn 100
SELECT

```

	SoLuong
Sprite	43
Clear	309
Nestle	7792
Lipton	636
Colgate	619

Query executed successfully.

ANH HOA MU IMO \ TURKEY\_MARKET ANH HOA MU IMO \ ADMIN SSAS\_Turkey\_Market | 00:00:01

Hình 312: Kết quả câu truy vấn 2

**Nhận xét:** Kết quả thực thi câu truy vấn 2 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.3. Câu truy vấn 3:

**Nội dung câu truy vấn:** Với mỗi thành phố, thống kê số lượng sản phẩm bán ra được ở các chi nhánh có số lượng sản phẩm bán ra lớn hơn 15000.

**Câu truy vấn MDX:**

```

SELECT
NON EMPTY { [Measures].[SoLuong]} ON
COLUMNS,
FILTER (
    {[Dim Product].[Branch Name].ALLMEMBERS},
    [Measures].[SoLuong] > 15000
) ON ROWS
FROM [Turkey Market Sales];

```

**Kết quả thực thi:**

```

--3. Với mỗi thành phố, thống kê số lượng sản phẩm bán ra được ở
--các chi nhánh có số lượng sản phẩm bán ra lớn hơn 15000
SELECT
NON EMPTY { [Measures].[SoLuong]} ON
COLUMNS,
FILTER (
    {[Dim Product].[Branch Name].AllMEMBERS},
    [Measures].[SoLuong] > 15000
) ON ROWS
FROM [Turkey Market Sales];
--4. Thống kê tổng doanh thu trong tháng 04 tại các chi nhánh
--ở vùng miền “Marmara” dựa trên nhãn hàng và loại sản phẩm
SELECT NON EMPTY {[Measures].[GiaSauGiamGia]} ON COLUMNS,
NON EMPTY {[Dim Product].[Category Name].CHILDREN} ON ROWS
FROM [Turkey Market Sales]
WHERE ([Dim Product].[DiaDiemChiNhanh].[Region].&[Marmara], [Dim Time].[InvoiceTime]
--5. Với 3 sản phẩm bán được với doanh số cao nhất

```

	SoLuong
All	452285.618
Ankara Subesi	30337.849
Bursa Subesi	15699.336
Istanbul Subesi	83586.310000001
Izmir Subesi	24138.662

Hình 313: Kết quả câu truy vấn 3

**Nhận xét:** Kết quả thực thi câu truy vấn 3 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.4. Câu truy vấn 4:

**Nội dung câu truy vấn:** Thống kê tổng doanh thu trong tháng 4 tại các chi nhánh ở vùng miền “Marmara” dựa trên nhãn hàng và loại sản phẩm.

**Câu truy vấn MDX:**

```

SELECT NON EMPTY {[Measures].[GiaSauGiamGia]} ON COLUMNS,
NON EMPTY {[Dim Product].[Category Name].CHILDREN * 
    [Dim Product].[Brand Name] } ON ROWS
FROM [Turkey Market Sales]
WHERE ([Dim Product].[DiaDiemChiNhanh].[Region].&[Marmara],
[Dim Time].[InvoiceTime].[Invoice Date Month].&[4]&[2017]);

```

**Kết quả thực thi:**

```

--4. Thống kê tổng doanh thu trong tháng 04 tại các chi nhánh
--ở vùng miền "Marmara" dựa trên nhãn hàng và loại sản phẩm
SELECT NON EMPTY {[Measures].[GiaSauGiamGia]} ON COLUMNS,
NON EMPTY {[Dim Product].[Category Name].CHILDREN * [Dim Product].[Brand Name].CHILDREN} ON
FROM [Turkey Market Sales]
WHERE ([Dim Product].[DiaDiemChiNhanh].[Region].&[Marmara], [Dim Time].[InvoiceTime].[InvoiceTime])
--5. Với 3 sản phẩm bán được với doanh số cao nhất
--theo từng ngày, tháng, năm, cho biết số lượng khách hàng đã mua 3 sản phẩm đó
SELECT NON EMPTY {[Dim Product].[Product Name].CHILDREN * [Measures].[SoLuongKhachHang]} ON
NON EMPTY {DrillDownLevel(
DrillDownLevel([Dim Time].[InvoiceTime]))}
} ON ROWS
FROM (SELECT(TOPCOUNT([Dim Product].[Product Name].[Product Name],3,[Measures].[SoLuong])) O
FROM [Turkey Market Sales]);
--6. Liệt kê tổng số lượng sản phẩm, số lượng khách hàng,
--theo ngày trong tuần của từng chi nhánh
SELECT NON EMPTY {
[Measures].[SoLuong], [Measures].[SoLuongKhachHang]} ON COLUMNS.

```

		GiaSauGiamGia
Baby	Canbebe	8.85
Baby	Dalin	39.58
Baby	ETI	3.33
Baby	Huggies	18.56
Baby	Molfix	35.42
Baby	Prima	56.66
Baby	Üker	6.59
Baby	Uni	2.54
Beverage	F NEFFIS	59.82

Hình 314: Kết quả câu truy vấn 4

**Nhận xét:** Kết quả thực thi câu truy vấn 4 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.5. Câu truy vấn 5:

**Nội dung câu truy vấn:** Với 3 sản phẩm bán được với doanh số cao nhất theo từng ngày, tháng, năm, cho biết số lượng khách hàng đã mua 3 sản phẩm đó.

#### Câu truy vấn MDX:

```

SELECT NON EMPTY {[Dim Product].[Product Name].CHILDREN *
[Measures].[SoLuongKhachHang]} ON COLUMNS,
NON EMPTY {DrillDownLevel(
DrillDownLevel(
DrillDownLevel([Dim Time].[InvoiceTime])))}
} ON ROWS
FROM (SELECT(TOPCOUNT([Dim Product].[Product
Name].[Product Name],3,[Measures].[SoLuong])) ON
COLUMNS
FROM [Turkey Market Sales])

```

#### Kết quả thực thi:

The screenshot shows the SSAS\_Turkey\_Market cube in the MDXQuery1.mdx editor. The cube is named Turkey Market Sales. The MDX query retrieves data from the Turkey Market Sales fact table, filtering by Region (Marmara) and Invoice Time. It uses non-empty pivoting to find the top 3 products by sales volume per day. The results are then grouped by week and branch. The results table has three columns: MAGIC HANDS CIG MEATBALLS 200GR, NESCAFE 3 IN 1 18 GR, and OSMANCIK PIRINCI KG. The data shows sales volumes for various weeks across three branches.

	MAGIC HANDS CIG MEATBALLS 200GR	NESCAFE 3 IN 1 18 GR	OSMANCIK PIRINCI KG
SoLuongKhachHang	SoLuongKhachHang	SoLuongKhachHang	SoLuongKhachHang
All	2689	1812	3490
2017	2689	1812	3490
1	1205	633	1486
10	43	26	43
11	66	36	84
12	27	24	52
13	78	37	49
14	68	25	49
15	34	28	59
16	56	28	52

Hình 315: Kết quả câu truy vấn 5

**Nhận xét:** Kết quả thực thi câu truy vấn 5 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.6. Câu truy vấn 6:

**Nội dung câu truy vấn:** Liệt kê tổng số lượng sản phẩm, số lượng khách hàng, theo ngày trong tuần của từng chi nhánh.

**Câu truy vấn MDX:**

```
SELECT NON EMPTY {
    [Measures].[SoLuong], [Measures].[SoLuongKhachHang]
} ON COLUMNS,
NON EMPTY {[Dim Time].[Invoice Date Dayofweek].[Invoice Date
Dayofweek].MEMBERS} * [Dim Product].[Branch Name].[Branch
Name].MEMBERS} ON ROWS
FROM [Turkey Market Sales]
```

**Kết quả thực thi:**

```

--6. Liệt kê tổng số lượng sản phẩm, số lượng khách hàng,
--theo ngày trong tuần của từng chi nhánh
SELECT NON EMPTY {
    [Measures].[SoLuong], [Measures].[SoLuongKhachHang]} ON COLUMNS,
NON EMPTY {[Dim Time].[Invoice Date Dayofweek].[Invoice Date Dayofweek].MEMBERS}
* ([Dim Product].[Branch Name].[Branch Name].MEMBERS) ON ROWS
FROM [Turkey Market Sales];
--7. Thống kê số lượng hóa đơn được lập tại
--các chi nhánh nằm ở 3 thành phố Kırşehir, Nevşehir, Çankırı thuộc vùng İç Anadolu theo tháng
SELECT NON EMPTY {[Measures].[SoLuongHoaDon]} ON COLUMNS,
NON EMPTY {
    [Dim Product].[DiaDiemChiNhanh].[City].&[Nevşehir]&[İç Anadolu],
    [Dim Product].[DiaDiemChiNhanh].[City].&[Kırşehir]&[İç Anadolu],
    [Dim Product].[DiaDiemChiNhanh].[City].&[Çankırı]&[İç Anadolu]
} * [Dim Time].[Invoice Date Month].[Invoice Date Month].ALLMEMBERS ON ROWS
FROM [Turkey Market Sales]
WHERE ([Dim Product].[Region].&[İç Anadolu])
8. Cho biết vùng miền có doanh thu cao nhất. Trong vùng miền này, liệt kê tên các chi nhánh

```

	SoLuong	SoLuongKhachHang
1. Adana Subesi	1491431	406
1. Adiyaman Subesi	391801	78
1. Afyonkarahisar Subesi	445518	125
1. Ağrı Subesi	512708	106
1. Aksaray Subesi	287	72
1. Amasya Subesi	319593	73
1. Ankara Subesi	4017956	936
1. Antalya Subesi	1761403	433
1. Ardahan Subesi	89002	16
	110860	???

Hình 316: Kết quả câu truy vấn 6

**Nhận xét:** Kết quả thực thi câu truy vấn 6 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.7. Câu truy vấn 7:

**Nội dung câu truy vấn:** Thống kê số lượng hóa đơn được lập tại các chi nhánh nằm ở 3 thành phố Kırşehir, Nevşehir, Çankırı thuộc vùng İç Anadolu theo từng tháng.

#### Câu truy vấn MDX:

```

SELECT NON EMPTY {[Measures].[SoLuongHoaDon]} ON COLUMNS,
NON EMPTY
{
    [Dim Product].[DiaDiemChiNhanh].[City].&[Nevşehir]&[İç Anadolu],
    [Dim Product].[DiaDiemChiNhanh].[City].&[Kırşehir]&[İç Anadolu],
    [Dim Product].[DiaDiemChiNhanh].[City].&[Çankırı]&[İç Anadolu]
} * [Dim Time].[Invoice Date Month].[Invoice Date Month].ALLMEMBERS ON ROWS
FROM [Turkey Market Sales]

```

**WHERE ([Dim Product].[Region].&[İç Anadolu])**

**Kết quả thực thi:**

```
--7. Thông kê số lượng hóa đơn được lập tại các chi nhánh nằm ở
--3 thành phố Kırşehir, Nevşehir, Çankırı thuộc vùng İç Anadolu theo từng tháng
SELECT NON EMPTY {[Measures].[SoLuongHoaDon]} ON COLUMNS,
NON EMPTY
{
    [Dim Product].[DiaDiemChiNhanh].[City].&[Nevşehir]&[İç Anadolu],
    [Dim Product].[DiaDiemChiNhanh].[City].&[Kırşehir]&[İç Anadolu],
    [Dim Product].[DiaDiemChiNhanh].[City].&[Çankırı]&[İç Anadolu]
} * [Dim Time].[Invoice Date Month].[Invoice Date Month].ALLMEMBERS ON ROWS
FROM [Turkey Market Sales]
WHERE ([Dim Product].[Region].&[İç Anadolu]);
--8. Cho biết vùng miền có doanh thu cao nhất. Trong vùng miền này,
--liệt kê tên các chi nhánh có số lượng khách hàng ghé thăm bé hơn 3000
SELECT NON EMPTY { [Measures].[SoLuongKhachHang] } ON COLUMNS,
NON EMPTY { ([Dim Product].[Region].CHILDREN *
[Dim Product].[City].CHILDREN * }
```

SoLuongHoaDon	
Nevşehir	1: 118 2: 132 3: 137 4: 2
Kırşehir	1: 93 2: 68 3: 120 4: 7
Çankırı	1: 71 2: 82 3: 100

Hình 317: Kết quả câu truy vấn 7

**Nhận xét:** Kết quả thực thi câu truy vấn 7 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.8. Câu truy vấn 8:

**Nội dung câu truy vấn:** Cho biết vùng miền có doanh thu cao nhất. Trong vùng miền này, liệt kê tên các chi nhánh có số lượng khách hàng ghé thăm bé hơn 3000.

**Câu truy vấn MDX:**

```
SELECT NON EMPTY { [Measures].[SoLuongKhachHang] } ON
COLUMNS,
NON EMPTY { ([Dim Product].[Region].CHILDREN *
[Dim Product].[City].CHILDREN *
[Dim Product].[Branch Name].CHILDREN )} ON ROWS
FROM (
    SELECT (FILTER ({[Dim Product].[Branch
Name].ALLMEMBERS},
[Measures].[SoLuongKhachHang] < 3000)) ON COLUMNS
    FROM (
```

```

SELECT(TOPCOUNT([Dim Product].[Region].[Region], 1,
[Measures].[GiaSauGiamGia])) ON COLUMNS
FROM [Turkey Market Sales])
);

```

Kết quả thực thi:

The screenshot shows the MDXQuery1.mdx query results in the SSAS\_Turkey\_Market database. The query calculates the total number of salespeople across different regions and branch names. The results are displayed in a table titled 'SoLuongKhachHang'.

Marmara	Bilecik	Bilecik Subesi	273
Marmara	Çanakkale	Çanakkale Subesi	618
Marmara	Edirne	Edirne Subesi	486
Marmara	Kırklareli	Kırklareli Subesi	460
Marmara	Kocaeli	Kocaeli Subesi	2349
Marmara	Sakarya	Sakarya Subesi	1233
Marmara	Tekirdağ	Tekirdağ Subesi	1288
Marmara	Yalova	Yalova Subesi	310

Hình 318: Kết quả câu truy vấn 8

**Nhận xét:** Kết quả thực thi câu truy vấn 8 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.9. Câu truy vấn 9:

**Nội dung câu truy vấn:** Thông kê tổng số lượng khách hàng theo từng tháng tại từng vùng miền.

**Câu truy vấn MDX:**

```

SELECT NON EMPTY { [Dim Time].[Invoice Date Year].CHILDREN *
[Dim Time].[Invoice Date
Month].CHILDREN *
[Dim Time].[Invoice Date Day].CHILDREN}
ON ROWS,
NON EMPTY {[Measures].[SoLuongKhachHang]} ON COLUMNS
FROM [Turkey Market Sales]

```

Kết quả thực thi:

```

--9. Thống kê tổng số lượng nhân viên (count distinct salesman_id)
-- tại từng chi nhánh trong từng vùng miền trong từng ngày
SELECT NON EMPTY
{ [Dim Time].[Invoice Date Year].CHILDREN *
[Dim Time].[Invoice Date Month].CHILDREN *
[Dim Time].[Invoice Date Day].CHILDREN *
[Dim Product].[Region].CHILDREN *
[Dim Product].[City].CHILDREN *
[Dim Product].[Branch Name].CHILDREN} ON ROWS,
NON EMPTY {[Measures].[SoLuongNhanVien]} ON COLUMNS
FROM [Turkey Market Sales];
--10. Cho biết tổng doanh số bán hàng của vùng miền Güneydoğu Anadolu theo từng tháng
SELECT {[Measures].[SoLuong]} ON COLUMNS,
NON EMPTY {[Dim Time].[Invoice Date Month].MEMBERS} ON ROWS
FROM [Turkey Market Sales]
Where [Dim Product].[DiaDiemChiNhanh].[Region].&[Güneydoğu Anadolu];

```

Messages Results

			SoLuongNhanVien			
2017	1	10	Akdeniz	Adana	Adana Subesi	9
2017	1	10	Akdeniz	Antalya	Antalya Subesi	8
2017	1	10	Akdeniz	Burdur	Burdur Subesi	5
2017	1	10	Akdeniz	Hatay	Hatay Subesi	3
2017	1	10	Akdeniz	Isparta	Isparta Subesi	6
2017	1	10	Akdeniz	Kahramanmaraş	Kahramanmaraş Subesi	2
2017	1	10	Akdeniz	Mersin	Mersin Subesi	5
2017	1	10	Akdeniz	Osmaniye	Osmaniye Subesi	6
2017	1	10	Đôđu Anadolu	Ağrı	Ağrı Subesi	3
2017	1	10	Đôđu Anadolu	Ardahan	Ardahan Subesi	2
2017	1	10	Đôđu Anadolu	Bingöl	Bingöl Subesi	1

Hình 319: Kết quả câu truy vấn 9

**Nhận xét:** Kết quả thực thi câu truy vấn 9 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.10. Câu truy vấn 10:

**Nội dung câu truy vấn:** Cho biết tổng doanh số bán hàng của vùng miền Güneydoğu Anadolu theo từng tháng.

**Câu truy vấn MDX:**

```

SELECT {[Measures].[SoLuong]} ON COLUMNS,
NON EMPTY {[Dim Time].[Invoice Date Month].MEMBERS} ON ROWS
FROM [Turkey Market Sales]
Where [Dim Product].[DiaDiemChiNhanh].[Region].&[Güneydoğu
Anadolu]

```

**Kết quả thực thi:**

```

NON EMPTY {[Measures].[SoLuong]} ON COLUMNS
FROM [Turkey Market Sales];
--10. Cho biết tổng doanh số bán hàng của vùng miền Güneydoğu Anadolu theo từng t
SELECT {[Measures].[SoLuong]} ON COLUMNS,
NON EMPTY {[Dim Time].[Invoice Date Month].MEMBERS} ON ROWS
FROM [Turkey Market Sales]
Where [Dim Product].[DiaDiemChiNhanh].[Region]&[Güneydoğu Anadolu];
--11. Cho biết thông tin khách hàng (mã khách hàng, tên khách hàng, giới tính)
--có tổng số tiền mua lớn nhất trong từng tháng
WITH MEMBER [Measures].[Top Customer Sales] AS
'([Measures].[GiaSauGiamGia], [Dim Customer].[Customer ID].CurrentMember)'
SELECT
{[Measures].[Top Customer Sales]} ON COLUMNS,
NON EMPTY( GENERATE(
[Dim Time].[Invoice Date Month].CHILDREN,
TopCount(
[Dim Time].[Invoice Date Month].CurrentMember *
[Dim Customer].[Customer ID].CHILDREN *
[Dim Customer].[Customer Name].CHILDREN *
[Dim Customer].[GENDER].CHILDREN,
1,

```

Hình 320: Kết quả câu truy vấn 10

**Nhận xét:** Kết quả thực thi câu truy vấn 10 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.11. Câu truy vấn 11:

**Nội dung câu truy vấn:** Cho biết thông tin khách hàng (mã khách hàng, tên khách hàng, giới tính) có tổng số tiền mua lớn nhất trong từng tháng.

**Câu truy vấn MDX:**

```

WITH MEMBER [Measures].[Top Customer Sales] AS
'([Measures].[GiaSauGiamGia], [Dim Customer].[Customer ID].CurrentMember)'
SELECT
{[Measures].[Top Customer Sales]} ON COLUMNS,
NON EMPTY( GENERATE(
[Dim Time].[Invoice Date Month].CHILDREN,
TopCount(
[Dim Time].[Invoice Date Month].CurrentMember *
[Dim Customer].[Customer ID].CHILDREN *
[Dim Customer].[Customer Name].CHILDREN *
[Dim Customer].[GENDER].CHILDREN,
1,

```

```

([Measures].[GiaSauGiamGia])
)
) ) ON ROWS
FROM [Turkey Market Sales]
WHERE [Dim Time].[InvoiceTime].CURRENTMEMBER

```

Kết quả thực thi:

```

--11. Cho biết thông tin khách hàng (mã khách hàng, tên khách hàng, giới tính)
--có tổng số tiền mua lớn nhất trong từng tháng
WITH MEMBER [Measures].[Top Customer Sales] AS
'([Measures].[GiaSauGiamGia], [Dim Customer].[Customer ID].CurrentMember)'
SELECT
{[Measures].[Top Customer Sales]} ON COLUMNS,
NON EMPTY( GENERATE(
[Dim Time].[Invoice Date Month].CHILDREN,
TopCount(
[Dim Time].[Invoice Date Month].CurrentMember *
[Dim Customer].[Customer ID].CHILDREN *
[Dim Customer].[Customer Name].CHILDREN *
[Dim Customer].[GENDER].CHILDREN,
1,
([Measures].[GiaSauGiamGia])
)
)

```

			Top Customer Sales	
1	808718	Mustafa SEVİLİR	Male	1219.66
2	420299	Nazmiye AYDIL	Female	826.96
3	237685	Atakan OLĞUN	Male	2012.75
4	638391	Sevim ÇANAKÇI	Female	303.51

**Nhận xét:** Kết quả thực thi câu truy vấn 11 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.12. Câu truy vấn 12:

**Nội dung câu truy vấn:** Đưa ra phần trăm giảm giá của từng chi nhánh trong tháng 01 và 04 năm 2017.

Câu truy vấn MDX:

```

WITH
MEMBER [Measures].[Discount Percent] AS
(1 - ([Measures].[GiaSauGiamGia] /
[Measures].[GiaTruocGiamGia])), FORMAT_STRING = 'Percent'
SELECT
{[Measures].[Discount Percent]} ON COLUMNS,
{[Dim Product].[Branch Name].Children} *

```

```
{[Dim Time].[Invoice Date Month].&[1]&[2017], [Dim
Time].[Invoice Date Month].&[4]&[2017]} ON ROWS
FROM [Turkey Market Sales]
```

Kết quả thực thi:

```
--12. Đưa ra phần trăm giảm giá của từng chi nhánh
--trong tháng 01 và tháng 04 năm 2017
WITH MEMBER [Measures].[Discount Percent] AS
    (1 - ([Measures].[GiaSauGiamGia] / [Measures].[GiaTruocGiamGia])), FORMAT_STRING
SELECT
    {[Measures].[Discount Percent]} ON COLUMNS,
    {[Dim Product].[Branch Name].Children} *
    {[Dim Time].[Invoice Date Month].&[1]&[2017], [Dim Time].[Invoice Date Month].&[4]
FROM [Turkey Market Sales];
--13. Cho biết tên sản phẩm, kho sản phẩm có tổng số lượng bán ra
--lớn hơn 10 tại từng vùng miền
Select {[Measures].[SoLuong]} on Columns,
{ Generate(
    [Dim Product].[Region].Children,
    [Dim Product].[Region].CurrentMember *
    Filter(
        {[Dim Product].[Dim Product ID].Children} *
        {[Dim Product].[Product Name].Children} *
        {[Dim Product].[Category Name].Children},
        [Measures].[SoLuong] > 10
    )
)}
```

		Discount Percent
Adana Subesi	1	8.36%
Adana Subesi	4	9.11%
Adiyaman Subesi	1	8.73%
Adiyaman Subesi	4	11.49%
Ayionkarahisar Subesi	1	8.88%
Ayionkarahisar Subesi	4	7.44%
Ağrı Subesi	1	8.73%
Ağrı Subesi	4	8.10%
Aksaray Subesi	1	8.22%
Aksaray Subesi	4	9.85%
Amasya Subesi	1	7.75%

**Nhận xét:** Kết quả thực thi câu truy vấn 12 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.13. Câu truy vấn 13:

**Nội dung câu truy vấn:** Cho biết tên sản phẩm, loại sản phẩm và tổng số lượng sản phẩm tại từng vùng miền.

Câu truy vấn MDX:

```
Select {[Measures].[SoLuong]} on Columns,
{ Generate(
    [Dim Product].[Region].Children,
    [Dim Product].[Region].CurrentMember *
    Filter(
        {[Dim Product].[Dim Product ID].Children} *
        {[Dim Product].[Product Name].Children} *
        {[Dim Product].[Category Name].Children},
        [Measures].[SoLuong] > 10
    )
)}
```

```

        )
    )
} on Rows
From [Turkey Market Sales]

```

**Kết quả thực thi:**

The screenshot shows the SSAS Management Studio environment. On the left, the 'Cube' pane displays the 'Turkey Market Sales' cube with its dimensions (Dim Customer, Dim Product, Dim Salesman, Dim Time) and measures (Measures, KPIs). The main area contains an MDX query with three numbered comments (13, 14, 15) explaining the logic. Below the query is a results grid showing sales data for various products across different regions.

```

--13. Cho biết tên sản phẩm, kho sản phẩm có
--tổng số lượng bán ra lớn hơn 10 tại từng vùng miền
Select {[Measures].[SoLuong]} on Columns,
    { GENERATE( [Dim Product].[Region].Children, [Dim Product].[Region].CURRENTMEMBER
        FILTER(([Dim Product].[Dim Product ID]. CHILDREN)*([Dim Product].[Product N
    } on Rows
From [Turkey Market Sales];
--14. Với từng vùng miền, liệt kê số lượng đơn hàng bán ra
--theo từng tháng ngoại trừ hai vùng miền 'Đôğu Anadolu', 'Karadeniz'
Select NON EMPTY {[Dim Time].[Invoice Date Month].[Invoice Date Month]} on Columns,
    NON EMPTY {Except (
        {[Dim Product].[Region].Children},
        {[Dim Product].[Region].&[Doğu Anadolu]},
        {[Dim Product].[Region].&[Karadeniz]})} on Rows
From [Turkey Market Sales]
Where {[Measures].[SoLuongHoaDon]};
--15. Cho biết nhân viên thanh toán

```

Akdeniz	2575	CAFE CROWN FOAMED MILK COFFEE LATTE SINGLE	Beverage	22
Akdeniz	2675	SAKA WATER 1.5 LITER	Beverage	34
Akdeniz	2730	MEHMET EFENDI COFFEE 100 GR.	Beverage	57
Akdeniz	2767	CAFE CROWN HOT CHOCOLATE 230R	Beverage	26
Akdeniz	2797	ICIM AYRAN 200 GR	Beverage	25
Akdeniz	2838	MEHMET EFENDI COFFEE 100 GR.	Beverage	11
Akdeniz	2861	KAY WATER 5 LT	Beverage	12
Akdeniz	2907	MEHMET EFENDI COFFEE 100 GR.	Beverage	36
Akdeniz	3001	BURC NATURAL SPRING WATER 0.5 L	Beverage	11

**Nhận xét:** Kết quả thực thi câu truy vấn 13 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.14. Câu truy vấn 14:

**Nội dung câu truy vấn:** Với từng vùng miền, liệt kê số lượng đơn hàng bán ra theo từng tháng ngoại trừ hai vùng miền 'Đôğu Anadolu', 'Karadeniz'.

**Câu truy vấn MDX:**

```

Select NON EMPTY {[Dim Time].[Invoice Date Month].[Invoice
Date Month]} on Columns,
NON EMPTY {Except (
    {[Dim Product].[Region].Children},
    {[Dim Product].[Region].&[Doğu Anadolu]},
    {[Dim Product].[Region].&[Karadeniz]})} on Rows
From [Turkey Market Sales]
Where {[Measures].[SoLuongHoaDon]};

```

**Kết quả thực thi:**

```

--14. Với từng vùng miền, liệt kê số lượng đơn hàng bán ra
--theo từng tháng ngoại trừ hai vùng miền 'Đôğu Anadolu', 'Karadeniz'
Select NON EMPTY {[Dim Time].[Invoice Date Month].[Invoice Date Month]} on Columns,
    NON EMPTY {Except (
        {[Dim Product].[Region].Children},
        {[Dim Product].[Region].&[Đôğu Anadolu],
        [Dim Product].[Region].&[Karadeniz]})} on Rows
From [Turkey Market Sales]
Where {[Measures].[SoLuongHoaDon]};

--15. Cho biết nhân viên thanh toán nhiều hóa đơn nhất
--và ít hóa đơn nhất trong từng tháng
Select {[Measures].[SoLuongHoaDon]} on Columns,
{Union (
    TopCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1),
    BottomCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1)
)} on Rows
From [Turkey Market Sales];

```

	1	2	3	4
Akdeniz	3867	4211	4882	160
Ege	4458	4727	5571	202
Güneydoğu Anadolu	3252	3596	3993	138
İç Anadolu	4891	5382	6036	212
Marmara	9038	9749	11209	407

**Nhận xét:** Kết quả thực thi câu truy vấn 14 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.11.15. Câu truy vấn 15:

**Nội dung câu truy vấn:** Cho biết nhân viên thanh toán nhiều hóa đơn nhất và ít hóa đơn nhất trong từng tháng.

**Câu truy vấn MDX:**

```

Select {[Measures].[SoLuongHoaDon]} on Columns,
{Union (
    TopCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1,
    [Measures].[SoLuongHoaDon]),
    BottomCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1, not
    isempty([Measures].[SoLuongHoaDon])))} on Rows
From [Turkey Market Sales]

```

**Kết quả thực thi:**

```

--15. Cho biết nhân viên thanh toán nhiều hóa đơn nhất
--và ít hóa đơn nhất trong từng tháng
Select {[Measures].[SoLuongHoaDon]} on Columns,
{Union (
TopCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1)
BottomCount({[Dim Salesman].[Salesman Id].Children * [Dim Salesman].[Salesman Name].Children}, 1)
) From [Turkey Market Sales];

```

The screenshot shows the SSAS MDX Query Editor interface. On the left, the cube browser displays the 'Turkey Market Sales' cube with its dimensions and measures. The main pane contains an MDX query. Below the query, the results pane shows two rows of data:

		SoLuongHoaDon
1224	Yunus Emre AYDIN	3310
1266	Zubeyde PEKGÜLER	29

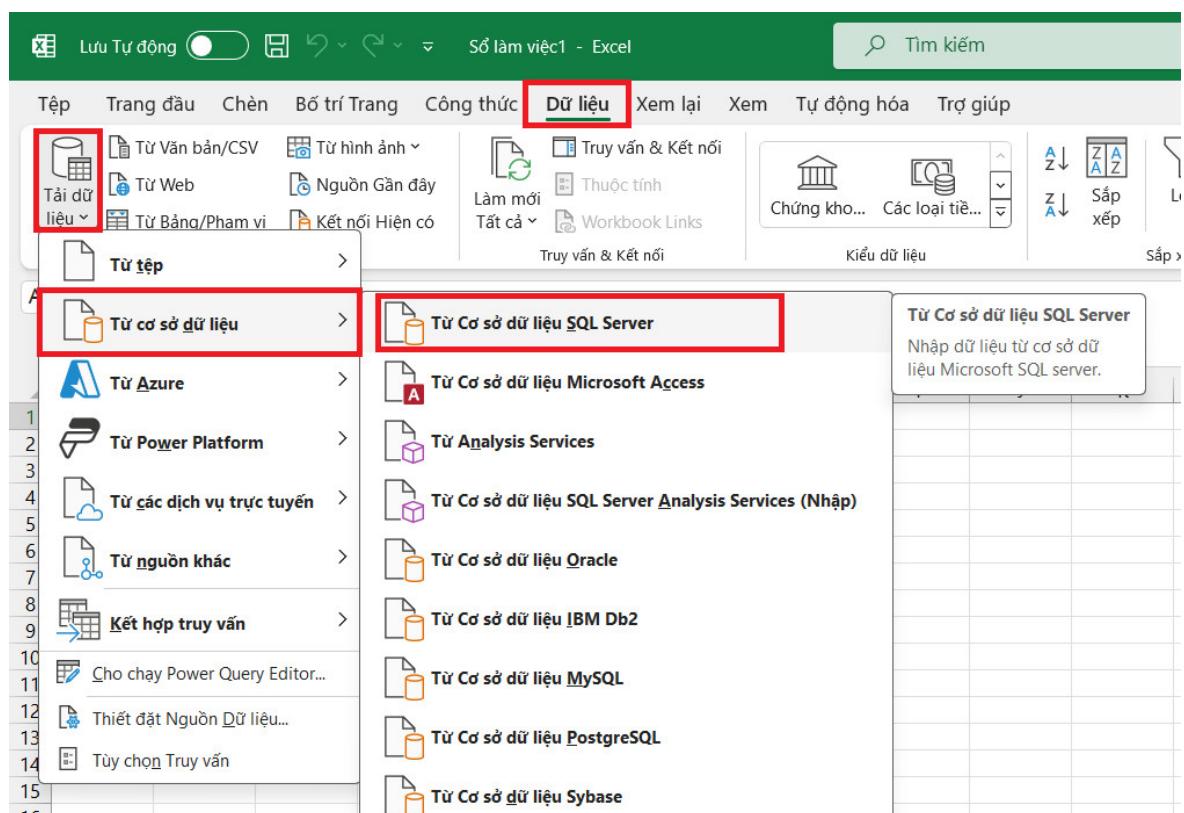
At the bottom, a message indicates the query was executed successfully.

**Nhận xét:** Kết quả thực thi câu truy vấn 15 bằng ngôn ngữ MDX tương tự như khi thao tác kéo thả trên browser của công cụ BI.

### 3.12. Quá trình lập báo cáo bằng công cụ Pivot Excel:

#### 3.12.1. Kết nối Excel với SQL Server:

**Bước 1:** Mở file excel trống lên, tại tab “Data”, chọn “Get Data”, sau đó chọn thêm “From Database”, tiếp tục chọn “From SQL Server Database”



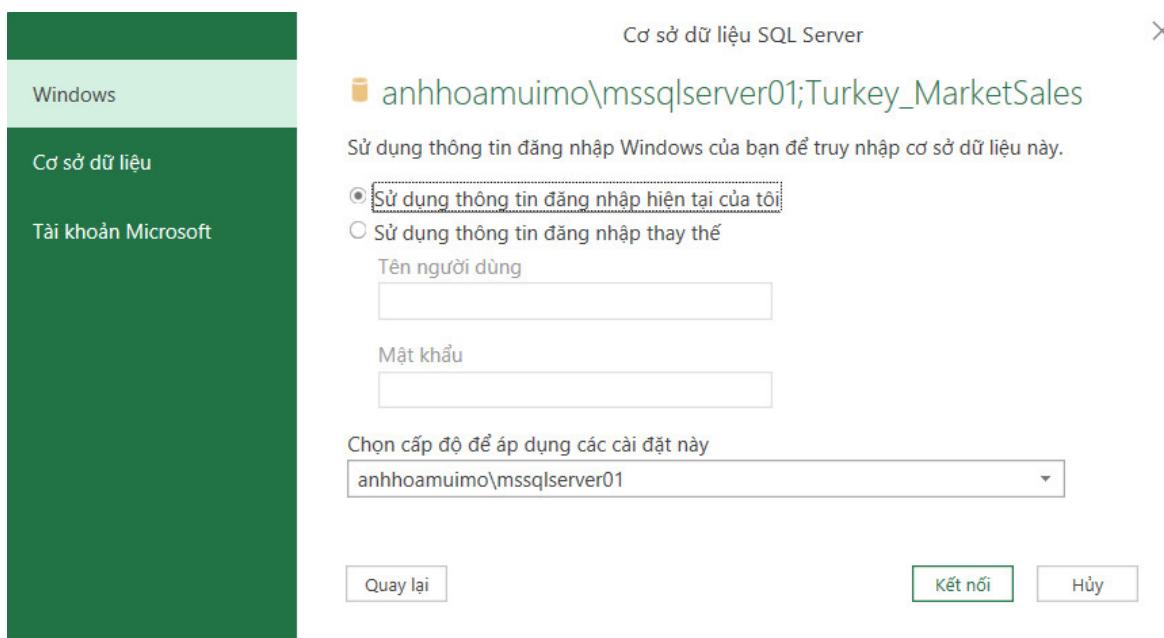
Hình 321: Giao diện hướng dẫn vào nơi kết nối SQL Server đến Excel

**Bước 2:** Lần lượt nhập tên máy chủ và tên cơ sở dữ liệu tương ứng với kho dữ liệu được lưu trữ trên SQL Server, sau đó click “OK”



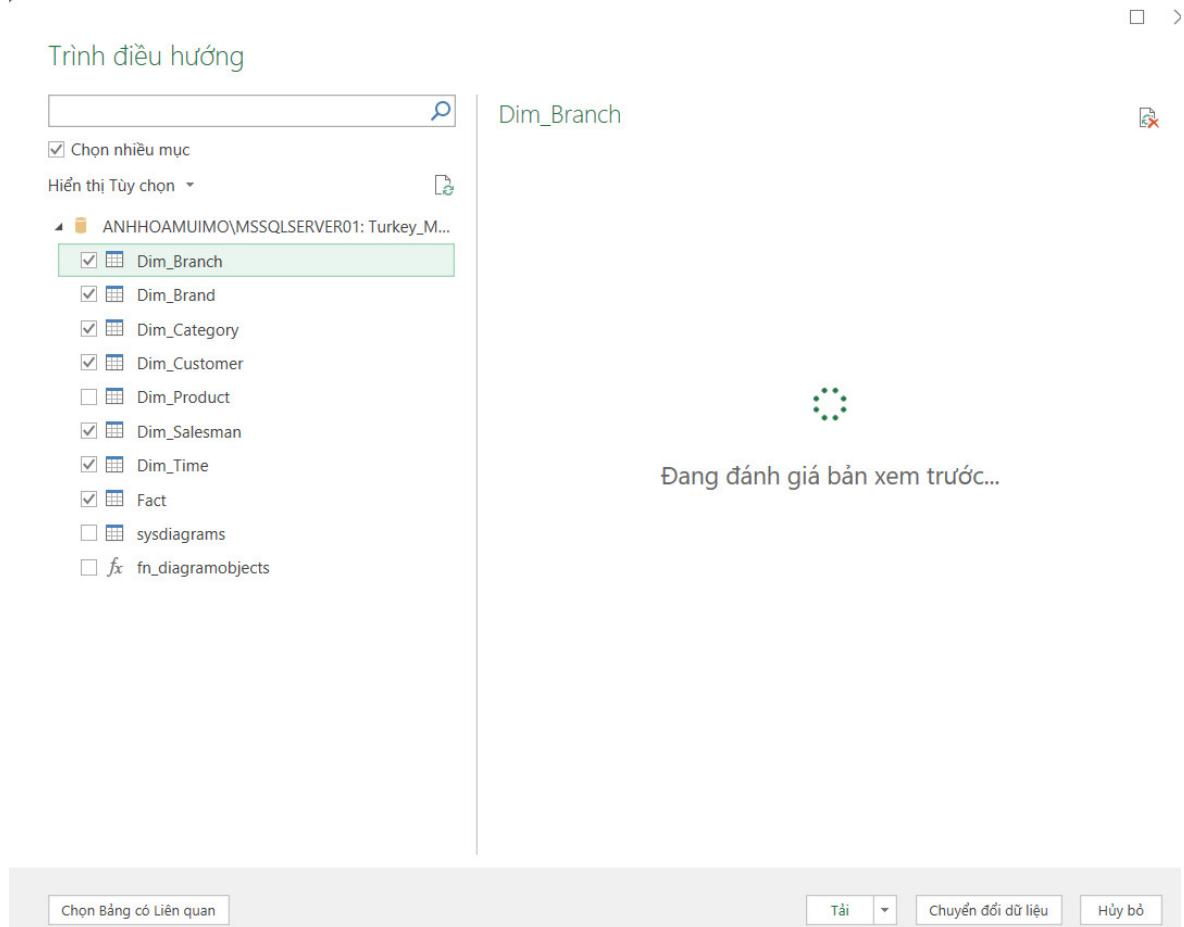
Hình 322: Hộp thoại điều thông tin Server và Database

**Bước 3:** Tiếp tục click “Kết nối”



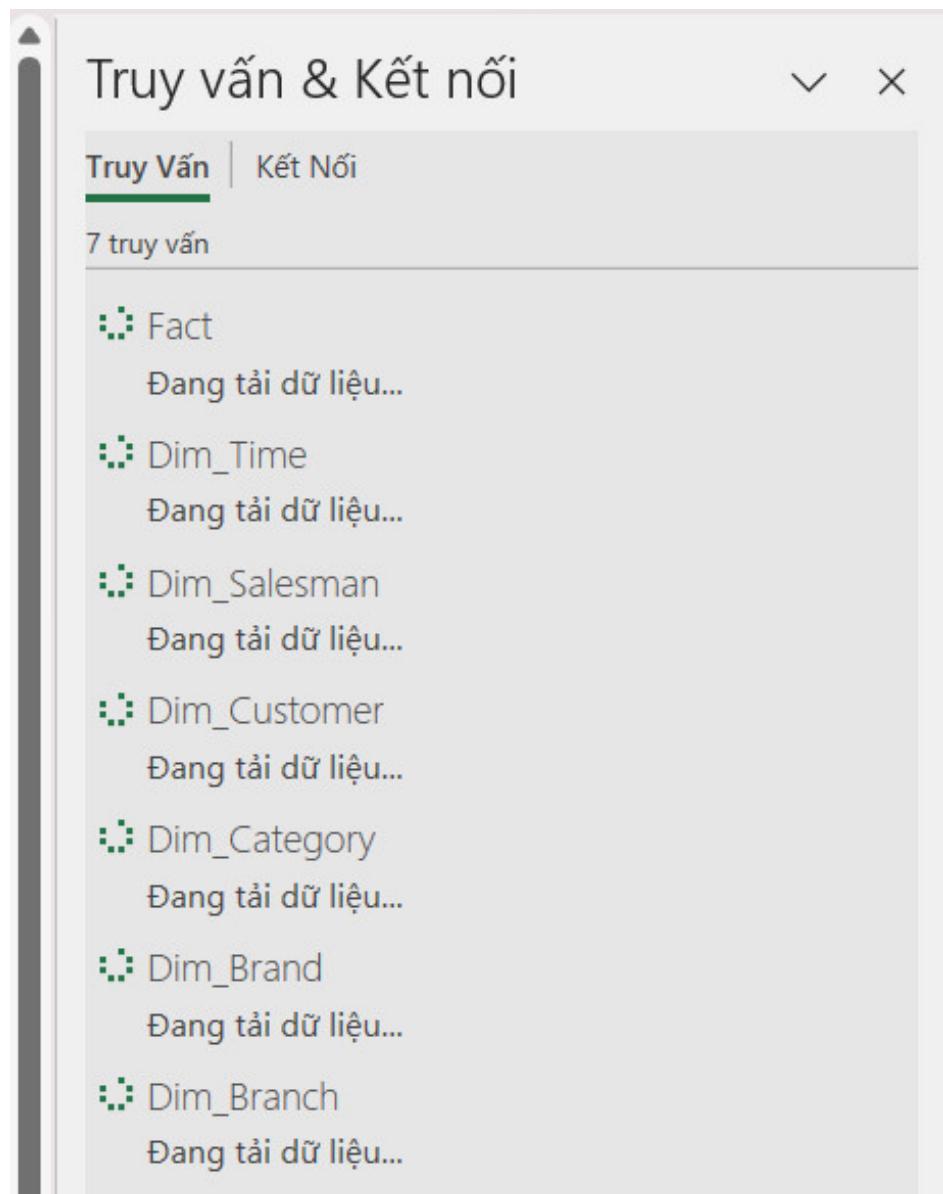
Hình 323: Kết nối đến SQL Server

**Bước 4:** Tại “**Màn hình điều hướng**”, mở cơ sở dữ liệu thông qua tên được tạo lúc đầu, click “**Chọn nhiều mục**” để lấy ra toàn bộ các bảng cần thiết, sau đó click “**Tải**” để đổ toàn bộ dữ liệu vào trong các sheet của excel



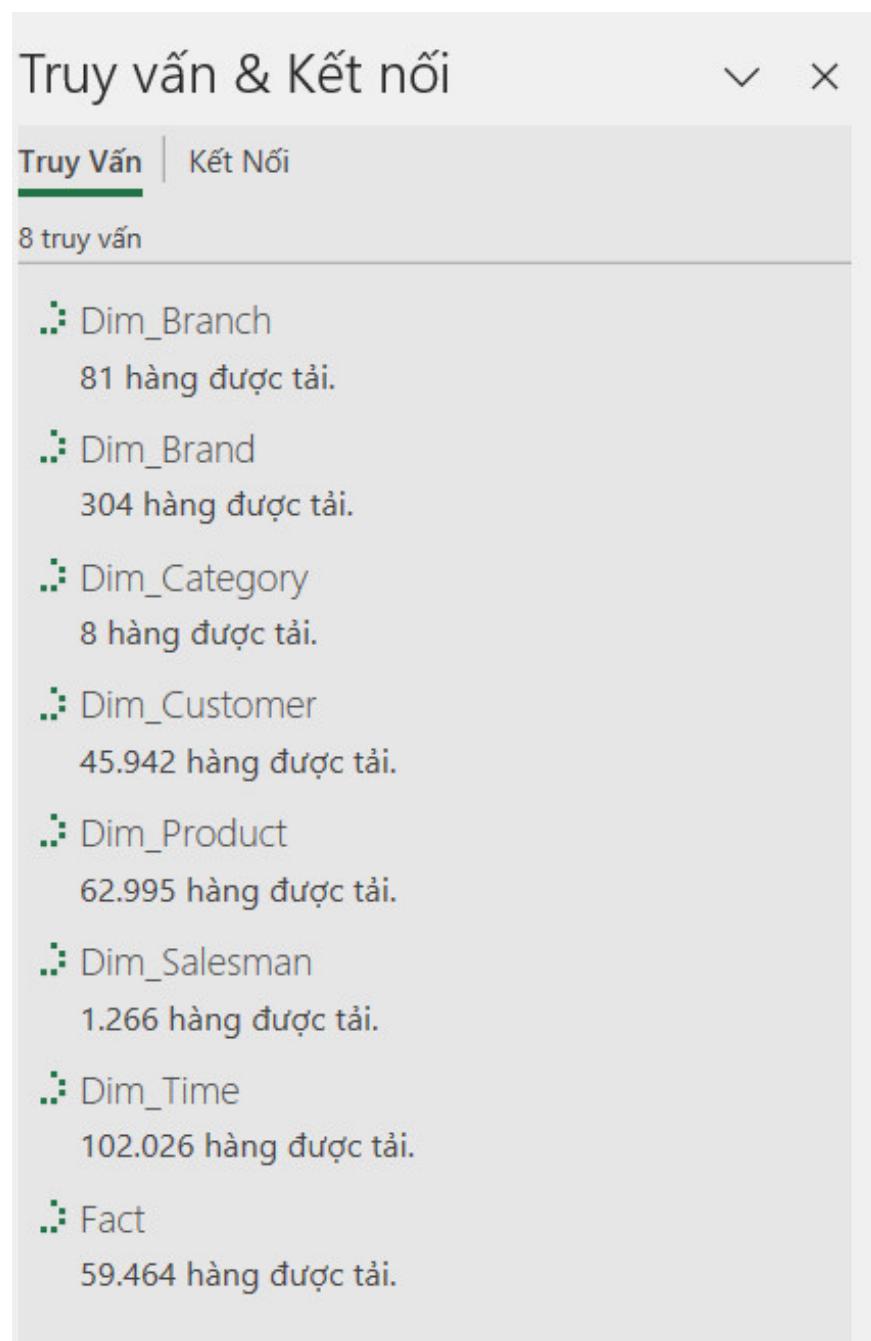
Hình 324: Giao diện chọn các bảng dữ liệu để đổ vào các sheet

**Bước 5:** Đợi Excel tiến hành quá trình đổ dữ liệu vào các sheet



Hình 325: Giao diện quá trình tải dữ liệu

**Bước 6:** Sau khi quá trình hoàn tất, kiểm tra dữ liệu tính đầy đủ và chính xác của dữ liệu trong các sheet



Hình 326: Quá trình tải dữ liệu hoàn tất

## Nhóm 15 – IS217.O21 – Kho dữ liệu và OLAP

The screenshot shows the Microsoft Power Query Editor interface with the 'Fact' table selected. The table contains 28 rows of data with columns: trans\_id, bill\_id, amount, pre\_promotion\_total\_trans, final\_total\_trans, invoice\_date\_time, and cust. The data represents transaction details over time. A 'Thiết đặt Truy vấn' (Query Settings) pane on the right shows the query name is 'Fact' and the schema is 'dbo'. The 'BƯỚC ĐƯỢC ÁP DỤNG' (Applied Step) pane shows the 'Diễn hướng' (Navigation) step.

Hình 327: Dữ liệu bảng Fact trong Excel

The screenshot shows the Microsoft Power Query Editor interface with the 'Dim\_Time' table selected. The table contains 28 rows of data with columns: invoice\_date\_time, invoice\_date\_day, invoice\_date\_month, invoice\_date\_year, invoice\_date\_hour, invoice\_date\_minute, and invoice\_date\_second. The data represents time dimensions. A 'Thiết đặt Truy vấn' (Query Settings) pane on the right shows the query name is 'Dim\_Time' and the schema is 'dbo'. The 'BƯỚC ĐƯỢC ÁP DỤNG' (Applied Step) pane shows the 'Diễn hướng' (Navigation) step.

Hình 328: Dữ liệu bảng Dim\_Time trong Excel

## Nhóm 15 – IS217.O21 – Kho dữ liệu và OLAP

The screenshot shows the Microsoft Power Query Editor interface. The main area displays the 'Dim\_Salesman' table with 28 rows of data. The columns are labeled 'salesman\_id', 'salesman\_name', and 'Fact'. The 'Fact' column contains numerical values representing the number of sales made by each salesman. The Power Query ribbon at the top includes tabs like 'Tệp', 'Trang chủ', 'Chuyển đổi', 'Thêm Cột', and 'Xem'. On the right, a 'Thiết đặt Truy vấn' (Query Settings) pane is open, showing settings for the query named 'Dim\_Salesman'. The 'BƯỚC ĐƯỢC ÁP DỤNG' (Applied Step) section is also visible.

Hình 329: Dữ liệu bảng Dim\_Salesman trong Excel

The screenshot shows the Microsoft Power Query Editor interface. The main area displays the 'Dim\_Customer' table with 56 rows of data. The columns are labeled 'customer\_id', 'customer\_name', 'GENDER', and 'Fact'. The 'Fact' column contains numerical values representing the number of purchases made by each customer. The Power Query ribbon at the top includes tabs like 'Tệp', 'Trang chủ', 'Chuyển đổi', 'Thêm Cột', and 'Xem'. On the right, a 'Thiết đặt Truy vấn' (Query Settings) pane is open, showing settings for the query named 'Dim\_Customer'. The 'BƯỚC ĐƯỢC ÁP DỤNG' (Applied Step) section is also visible.

Hình 330: Dữ liệu bảng Dim\_Customer trong Excel

Dim\_Category - Trình soạn thảo Power Query

Truy vấn [ ]

Nguyên ([Schema='dbo',Item='Dim\_Category'])[Data]

category_id	category_name	Dim_Product
1	Baby	Table
2	Beverage	Table
3	Cosmetics	Table
4	Dairy Breakfast	Table
5	Detergent Cleaning	Table
6	Food	Table
7	Home	Table
8	Meat and Poultry	Table

Thiết đặt Truy vấn

THƯỢC TÍNH

Tên: Dim\_Category

Tất cả Thuộc tính

BƯỚC ĐƯỢC ÁP DỤNG

Nguồn: Diều hướng

BẢN XEM TRƯỚC ĐƯỢC TẢI XUỐNG LÚC 1:31 SA

Hình 331: Dữ liệu bảng Dim\_Category trong Excel

Dim\_Brand - Trình soạn thảo Power Query

Truy vấn [ ]

Nguyên ([Schema='dbo',Item='Dim\_Brand'])[Data]

brand_id	brand_name	Dim_Product
1	Air Wick	Table
2	Nutella	Table
3	Ola	Table
4	Olin	Table
5	Oskar	Table
6	Pioneer	Table
7	Ötörmizel	Table
8	P.Bahçe	Table
9	Baghdad	Table
10	pakmaya	Table
11	palmolive	Table
12	pastavilla	Table
13	Penguin	Table
14	Pepsi	Table
15	peyman	Table
16	philips	Table
17	pinar	Table
18	F BAKTAT	Table
19	prenses	Table
20	Prima	Table
21	Pringles	Table
22	Raid	Table
23	Reis	Table
24	Saray	Table
25	Sarelle	Table
26	Sarahan	Table
27	Balkovan	Table
28	SEK	Table

Thiết đặt Truy vấn

THƯỢC TÍNH

Tên: Dim\_Brand

Tất cả Thuộc tính

BƯỚC ĐƯỢC ÁP DỤNG

Nguồn: Diều hướng

BẢN XEM TRƯỚC ĐƯỢC TẢI XUỐNG LÚC 1:31 SA

Hình 332: Dữ liệu bảng Dim\_Brand trong Excel

**Dim\_Branch - Trình soạn thảo Power Query**

**Thiết đặt Truy vấn**

**THUẬT TÌNH**

Tên: Dim\_Branch

Tất cả Thuộc tính

**BƯỚC ĐƯỢC ÁP DỤNG**

Nguồn: Dim\_Branch

Điều hướng

**Dim\_Branch**

branch_id	branch_name	city	region	Dim_Product
1	Adana Subesi	Adana	Akdeniz	Table
2	Adiyaman Subesi	Adiyaman	Güneydoğu Anadolu	Table
3	Afyonkarahisar Subesi	Afyonkarahisar	Ege	Table
4	Ağrı Subesi	Ağrı	Doğu Anadolu	Table
5	Aksaray Subesi	Aksaray	İç Anadolu	Table
6	Amasya Subesi	Amasya	Karadeniz	Table
7	Ankara Subesi	Ankara	İç Anadolu	Table
8	Antalya Subesi	Antalya	Akdeniz	Table
9	Ardahan Subesi	Ardahan	Doğu Anadolu	Table
10	Artvin Subesi	Artvin	Karadeniz	Table
11	Aydın Subesi	Aydın	Ege	Table
12	Balıkesir Subesi	Balıkesir	Ege	Table
13	Bartın Subesi	Bartın	Karadeniz	Table
14	Batman Subesi	Batman	Güneydoğu Anadolu	Table
15	Bayburt Subesi	Bayburt	Karadeniz	Table
16	Bilecik Subesi	Bilecik	Marmara	Table
17	Bingöl Subesi	Bingöl	Doğu Anadolu	Table
18	Bilecik Subesi	Bilecik	Marmara	Table
19	Bolu Subesi	Bolu	Karadeniz	Table
20	Burdur Subesi	Burdur	Akdeniz	Table
21	Bursa Subesi	Bursa	Marmara	Table
22	Çanakkale Subesi	Çanakkale	Marmara	Table
23	Cankırı Subesi	Cankırı	İç Anadolu	Table
24	Corum Subesi	Çorum	Karadeniz	Table
25	Denizli Subesi	Denizli	Ege	Table
26	Diyarbakır Subesi	Diyarbakır	Güneydoğu Anadolu	Table
27	Düzce Subesi	Düzce	Karadeniz	Table
28	Edirne Subesi	Edirne	Marmara	Table

5 CỘT, 81 HÀNG Quá trình định cấu hình cột dựa trên 1000 hàng trên cùng

Hình 333: Dữ liệu bảng Dim\_Branch trong Excel

**Dim\_Product - Trình soạn thảo Power Query**

**Thiết đặt Truy vấn**

**THUẬT TÌNH**

Tên: Dim\_Product

Tất cả Thuộc tính

**BƯỚC ĐƯỢC ÁP DỤNG**

Nguồn: Dim\_Product

Điều hướng

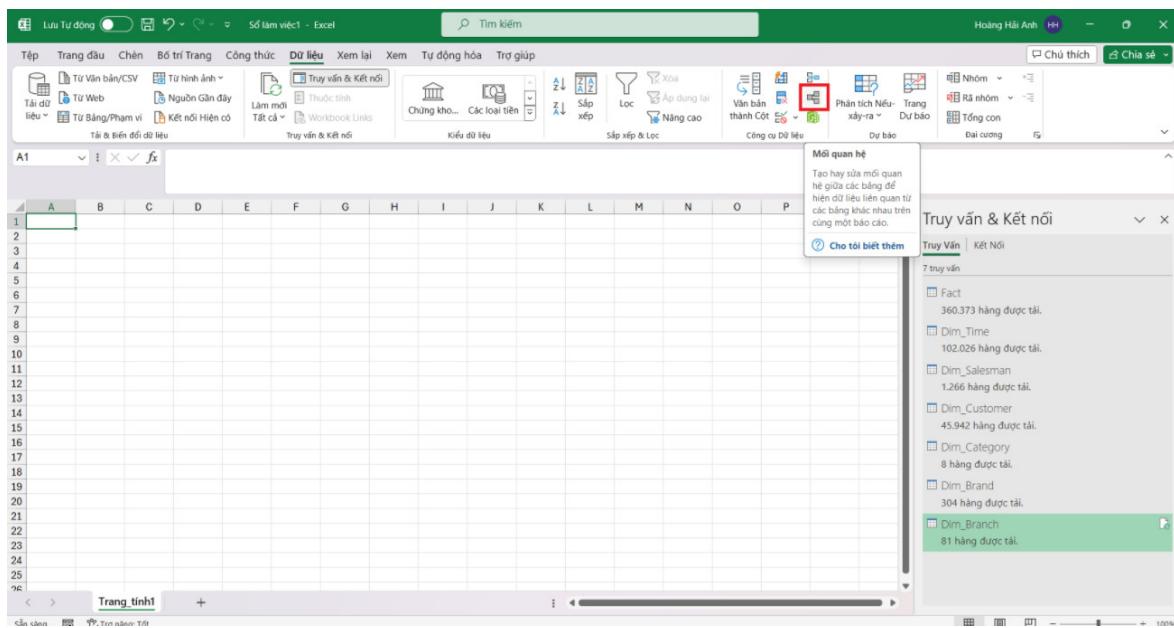
**Dim\_Product**

dim_product_id	product_id	product_name	price	branch_id	brand_id
1	7491	PRIMA ACT BABY ECONOMY 11-25 KG	19,95	47	121
2	1443	BIZIM RICE FLOUR 250G	1,95	21	146
3	23039	HERO BABY MILK 8-GRAIN 500 GR	11,5	41	146
4	23031	HERO FRUIT 8-GRAIN MILK 250GR	6,95	41	146
5	1247	HERO GOGLY RICE FLOUR PLAIN	2,99	19	146
6	23108	HERO PEACH BANANA 125 GR	1,5	19	146
7	8864	APTMIL WITH MILK BISCUIT 2*250 GR ECONOMIC	14,9	81	238
8	8869	APTMIL SEMILA WITH HONEY 250 GR	8,9	41	238
9	23391	WEE SILICONE TRANSPARENT PATTERNED PACIFIER WITH PALAT...	4,95	17	270
10	23390	WEE SILICONE PATTERNED PACIFIER WITH PALATE NO:3	4	62	270
11	3708	DALIN KIDS SAC BOCCULI SAMP.300ML TROPICAL	6,95	73	318
12	21087	ETI CICI BABY 190G BANANA BAGS	1,8	23	44
13	7488	PRIMA JUMBO MAXIPLUS 4+	25,5	21	121
14	22705	ULKER BABY BISCUIT 190GR. BANANA	1,85	41	146
15	23072	HERO PEAR-PINEAPPLE-PEACH	1,5	20	146
16	1250	HERO APPLE BANANA 125GR	1,25	21	146
17	22679	ULKER BABY BISCUIT 1KG-138-	6,95	4	146
18	22721	ULKER BABY 190GR. BANANA	1,85	67	146
19	8858	APTMIL PEARL PORT. 250GR	8,9	5	238
20	9829	CANEBEBE TWO EXTRI LARGE 28PCS	16,9	52	25
21	23377	WEE SILICONE BABY BOTTLE PIECE NO:2	2,95	77	270
22	5393	UNI EAR TUBE 100 PCS	1,4	64	275
23	3715	DALIN FABRIC SOFTENER 2LT.	9,95	41	318
24	3889	ETI CICI BABY BISCUIT 190 GR.VIT.	1,8	49	44
25	7486	PRIMA ACT BABY MEGA PACK MIDI	26,95	39	121
26	2250	HERO APPLE BANANA 125GR	1,25	7	146
27	12413	HERO GOGLY RICE FLOUR APPLE	2,99	53	146

11 CỘT, 999+ HÀNG Quá trình định cấu hình cột dựa trên 1000 hàng trên cùng

Hình 334: Dữ liệu bảng Dim\_Product trong Excel

**Bước 7:** Sau khi hoàn tất quá trình đồ dữ liệu, ta thực hiện tạo mối quan hệ giữa các sheet. Tại tab “Dữ liệu”, click chọn “Mối quan hệ”



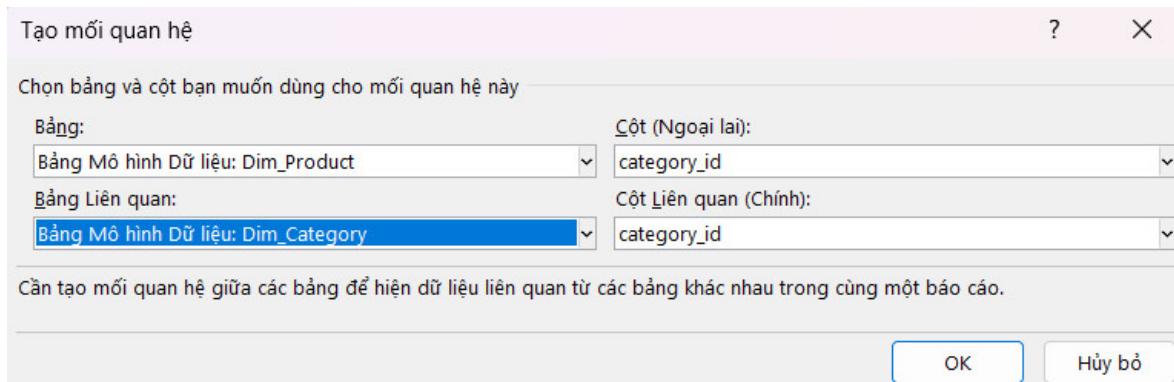
Hình 335: Giao diện hướng dẫn vào nơi tạo relationship

**Bước 8:** Tại cửa sổ “Quản lý mối quan hệ”, click “Mới” để bắt đầu quá trình tạo quan hệ

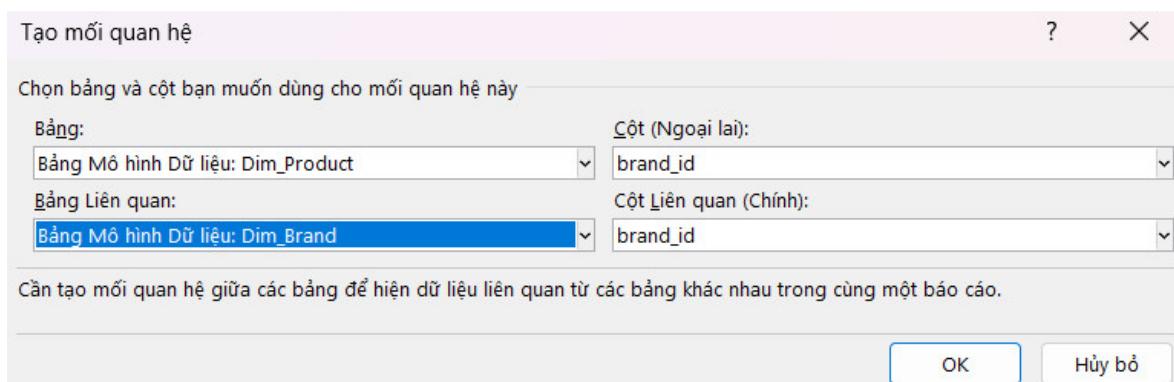


Hình 336: Tạo mới kết nối giữa các bảng

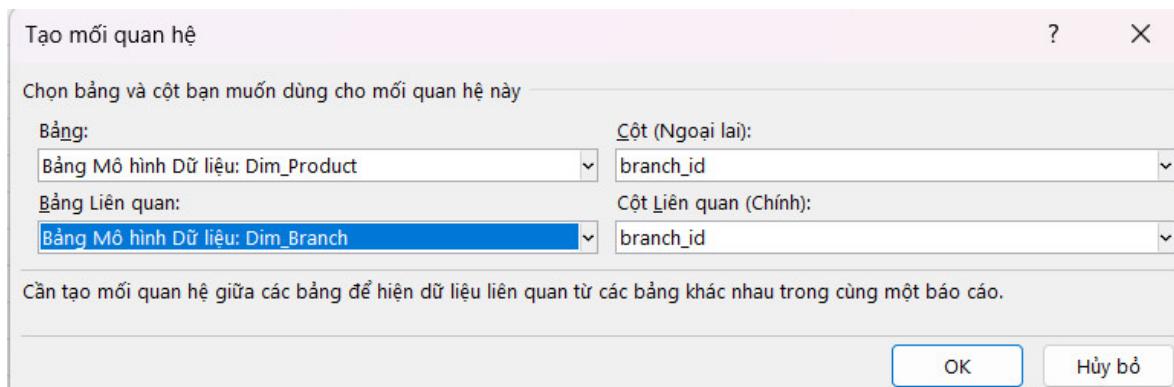
**Bước 9:** Tại cửa sổ “Tạo mối quan hệ”, chọn từng bảng và các khóa ngoại theo đúng mối quan hệ. Ví dụ: Ở đây bảng Dim\_Product có chứa khóa ngoại (Foreign Key) là thuộc tính category\_id để kết với bảng Dim\_Category với khóa chính (Primary Key) là thuộc tính category\_id. Sau đó click “OK”



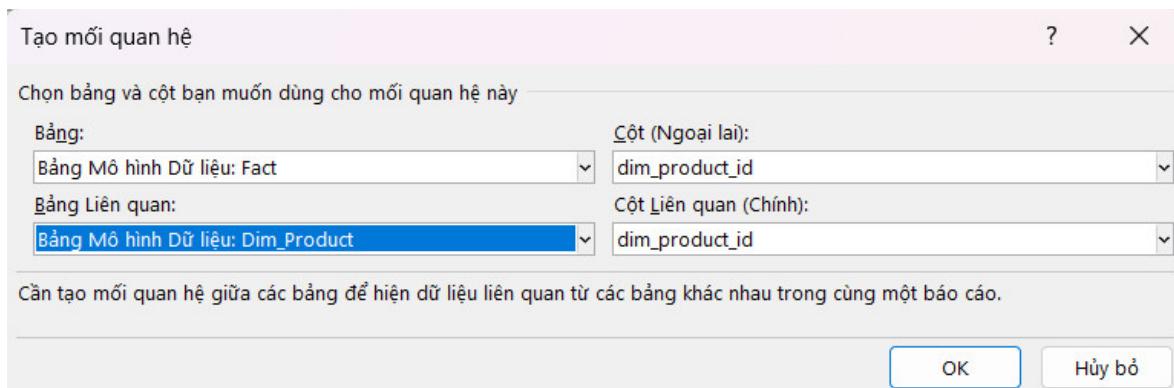
Hình 337: Tạo mối quan hệ giữa Dim\_Product và Dim\_Category



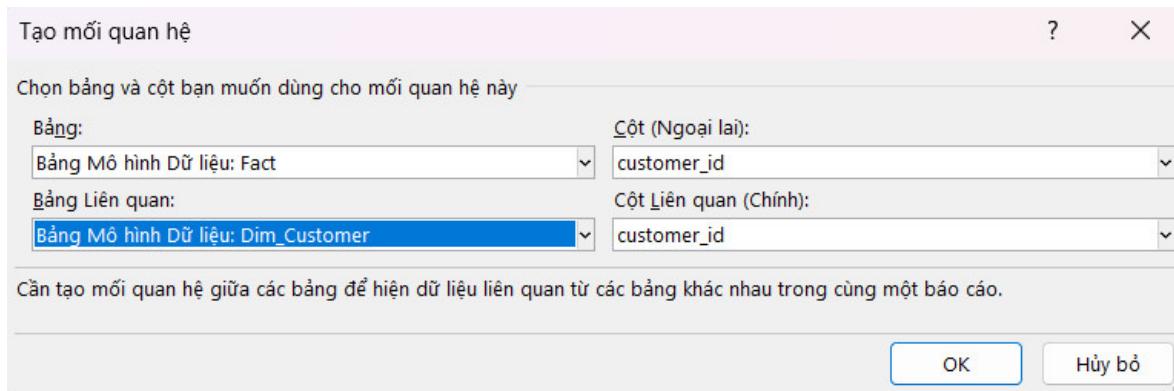
Hình 338: Tạo mối quan hệ giữa Dim\_Product và Dim\_Brand



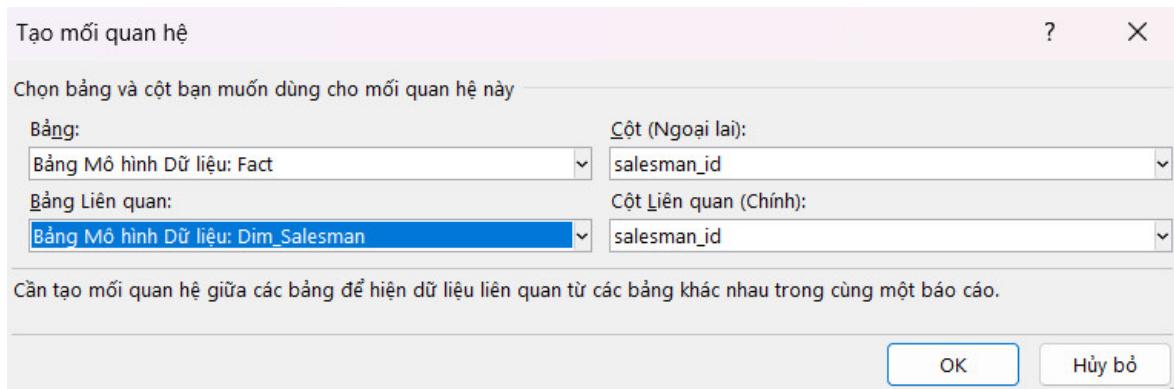
Hình 339: Tạo mối quan hệ giữa Dim\_Product và Dim\_Branch



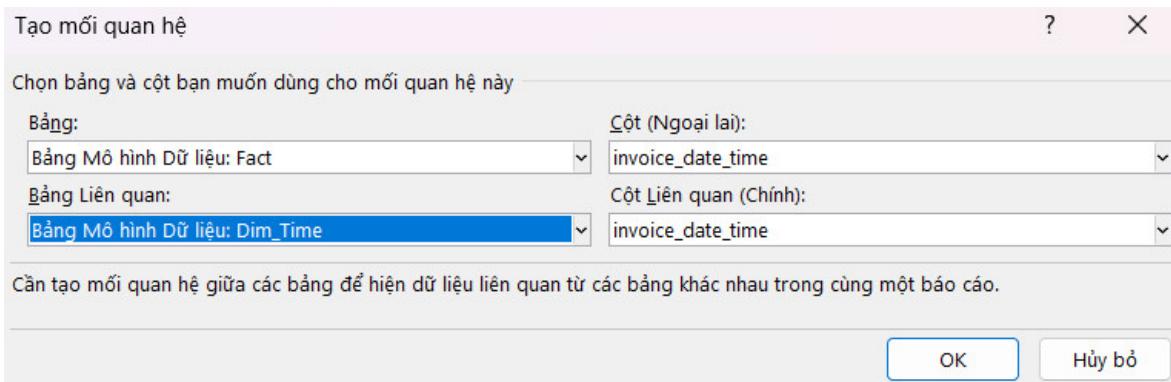
Hình 340: Tạo mối quan hệ giữa Fact và Dim\_Product



Hình 341: Tạo mối quan hệ giữa Fact và Dim\_Customer

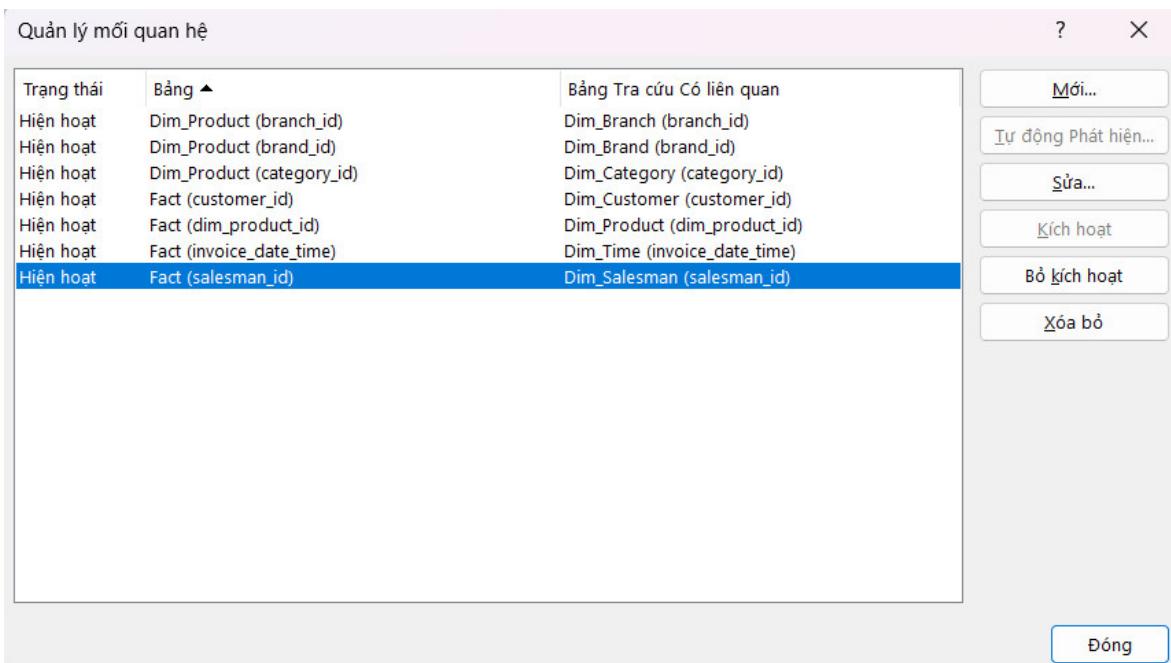


Hình 342: Tạo mối quan hệ giữa Fact và Dim\_Salesman



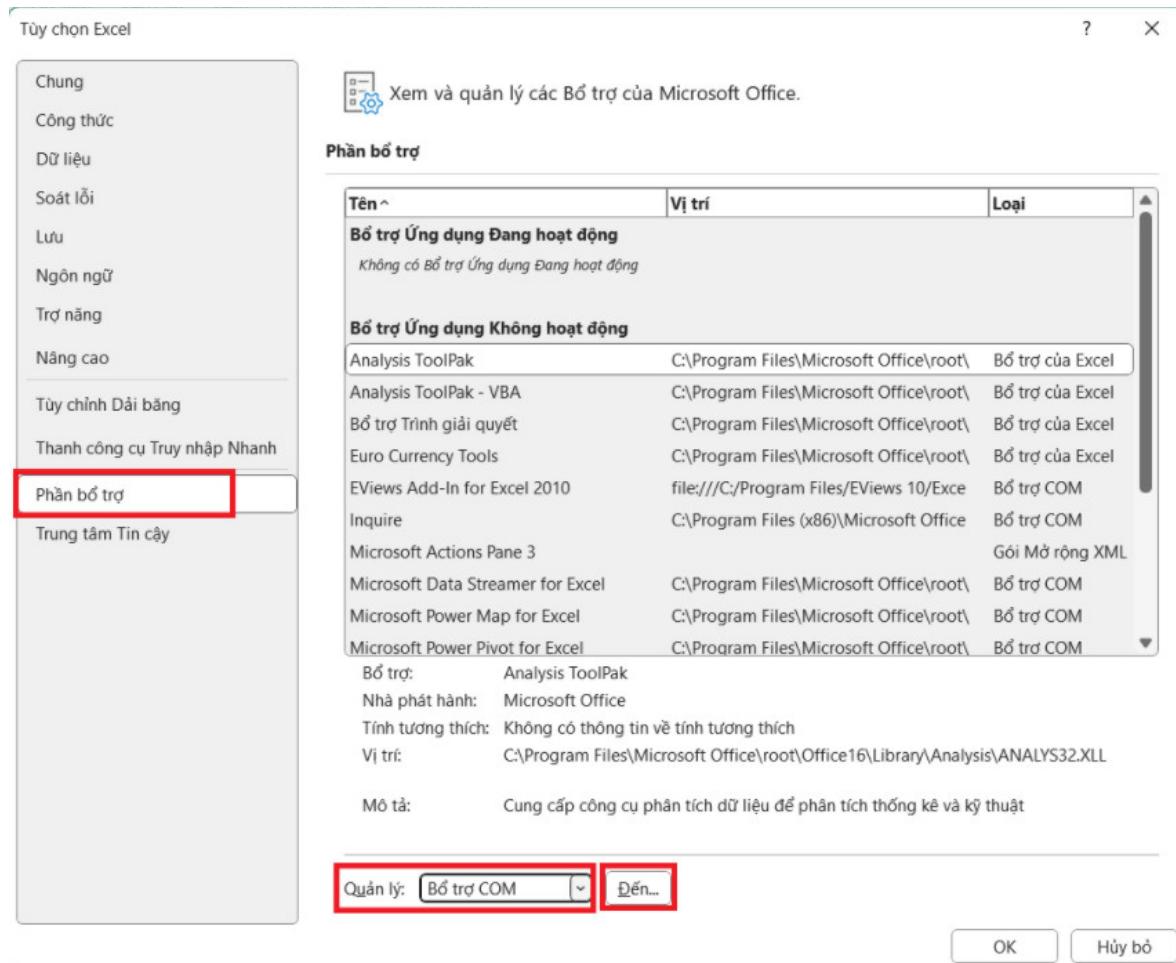
Hình 343: Tạo mối quan hệ giữa Fact và Dim\_Time

**Bước 10:** Kiểm tra lại các cài đặt mối quan hệ, sau đó nhấn “Đóng” để hoàn tất



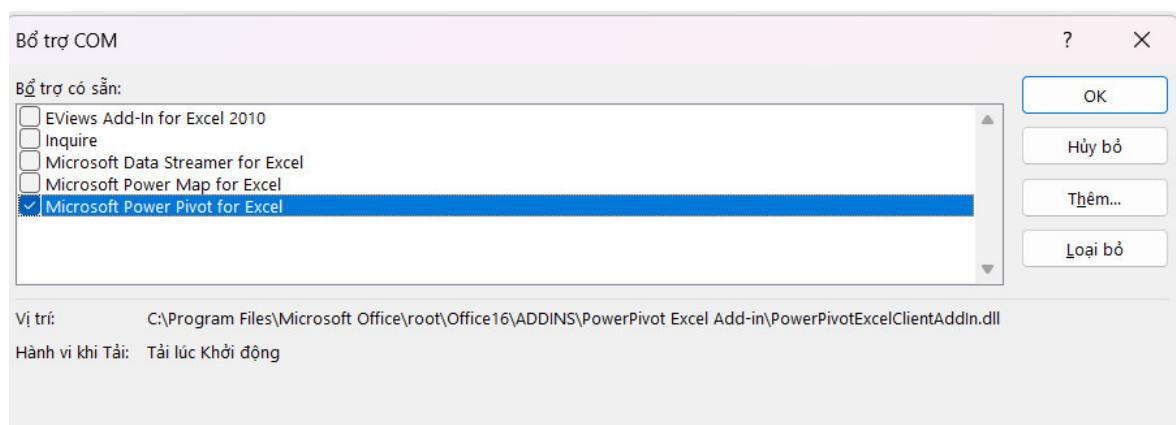
Hình 344: Giao diện kiểm tra các mối quan hệ vừa thiết lập

**Bước 11:** Click vào “Tệp” > “Tùy chọn” > “Phản bộ trợ”. Chọn “Bổ trợ COM” từ mục “Quản lý” sau đó click “Đến..” để thêm tính năng Power Pivot vào Excel



Hình 345: Giao diện hướng dẫn vào nơi mở Power Pivot

**Bước 12:** Xuất hiện hộp thoại “Bổ trợ COM”, click vào “Microsoft Power Pivot for Excel” và click “OK”

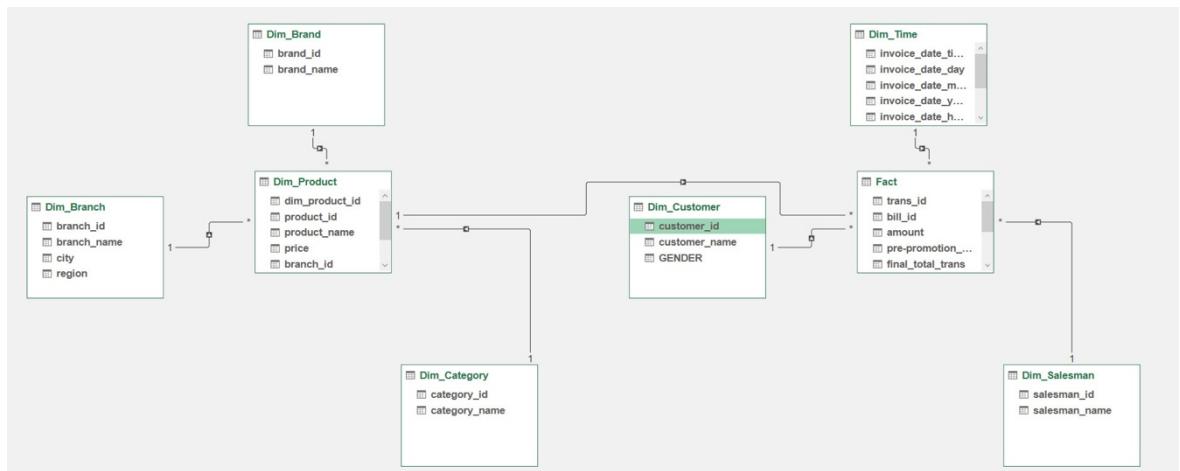


Hình 346: Hộp thoại Bổ trợ COM

**Bước 13:** Trong tab “Power Pivot”, chọn “Quản lý”, một cửa sổ mới có tên “Power Pivot for Excel” xuất hiện, nhấn chọn “Dạng xem lược đồ” để xem lược đồ dữ liệu tổng thể sau khi đã thiết lập các khóa ngoại



Hình 347: Giao diện hướng dẫn vào nơi xem lược đồ

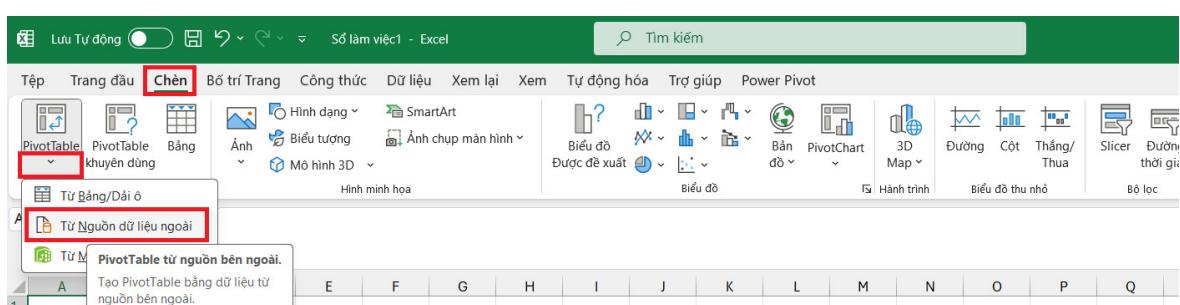


Hình 348: Lược đồ sau khi hoàn tất quá trình đổ dữ liệu vào tạo các mối quan hệ

### 3.12.2. Thiết lập môi trường để thực hiện báo biểu bằng

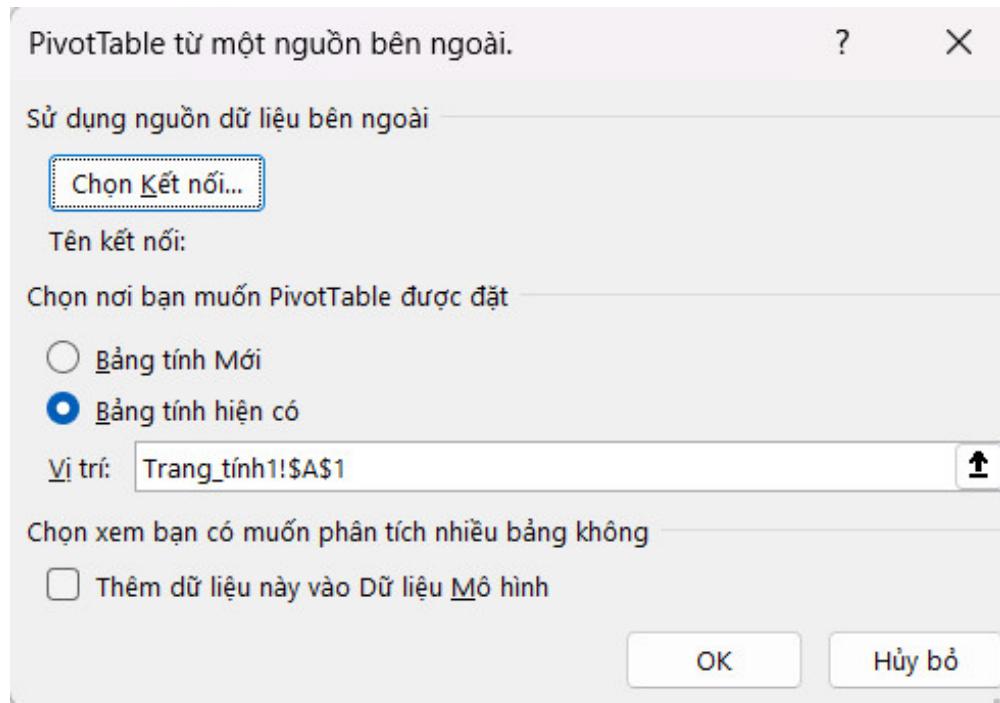
Excel:

**Bước 1:** Mở một Sheet mới, vào tab “Chèn”, chọn dấu trỏ xuống ở “PivotTable”, sau đó chọn “Từ nguồn dữ liệu ngoài”



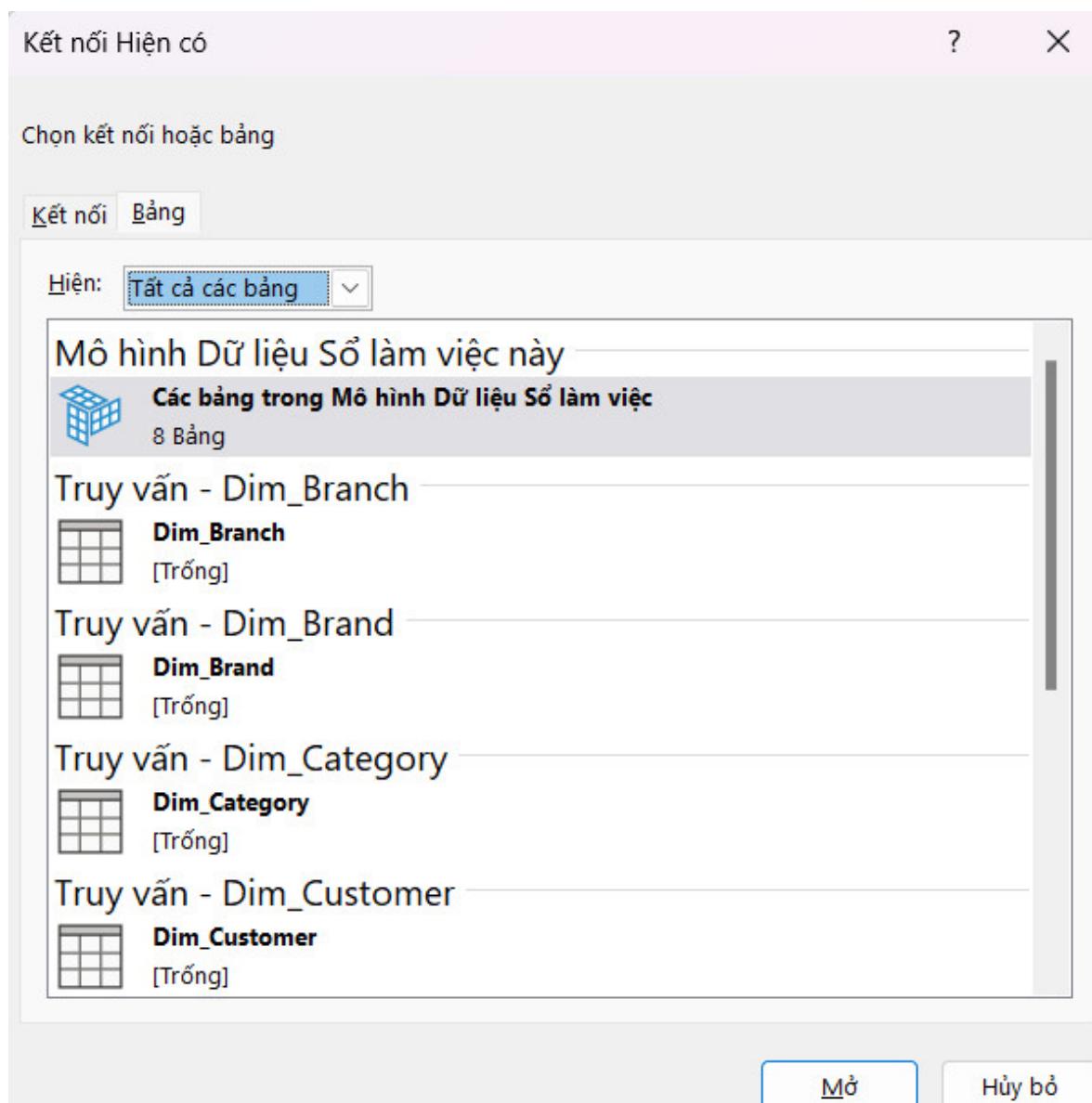
Hình 349: Giao diện hướng dẫn vào nơi tạo kết nối với nơi chứa dữ liệu ngoài

**Bước 2:** Tại cửa sổ “PivotTable từ một nguồn bên ngoài”, click “Chọn kết nối...”



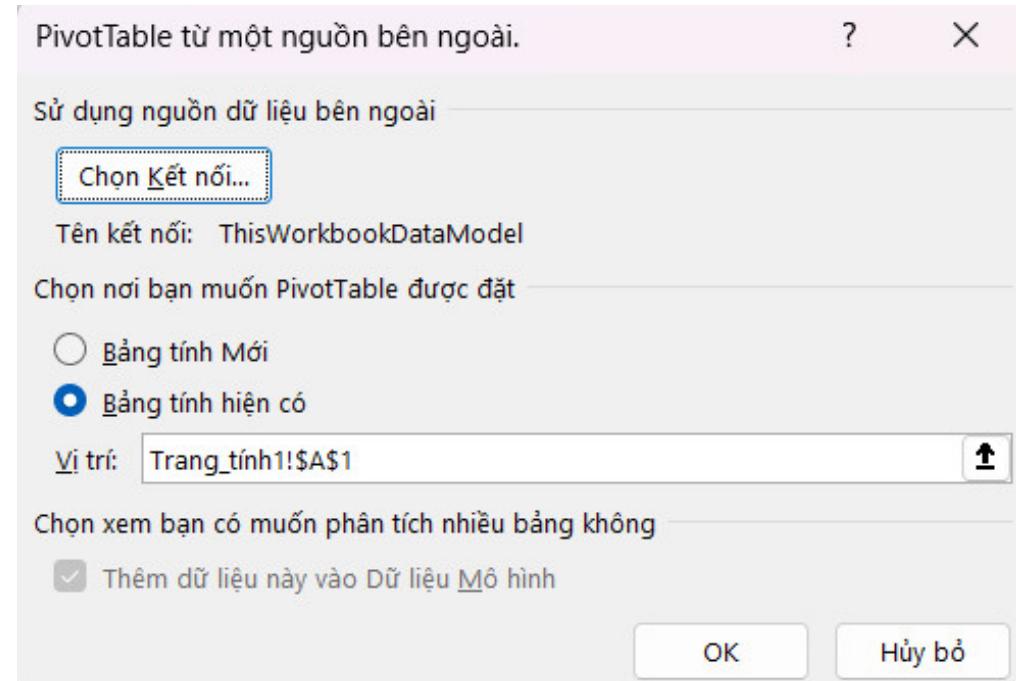
Hình 350: Hộp thoại Pivot Table xuất hiện khởi tạo kết nối mới

**Bước 3:** Chọn tab “**Bảng**”, chọn các bảng trong “**Mô hình Dữ liệu Số làm việc**”. Sau đó, nhấn “**Mở**”



Hình 351: Chọn mô hình dữ liệu sổ làm việc

Bước 4: Quay trở lại màn hình “Pivot Table từ một nguồn bên ngoài”, kiểm tra “Tên kết nối” là ThisWorkbookDataModel. Sau đó, click “OK”



Hình 352: Kiểm tra lại tên kết nối vừa thiết lập

**Bước 5:** Ở sheet mới vừa tạo đã được cài đặt, thiết lập đầy đủ môi trường để thực hiện báo biểu với đầy đủ dữ liệu và thông tin cần thiết

Lưu Tự động

Tệp Trang đầu Chèn Bố trí Trang Công thức Dữ liệu Xem lại Xem Tự động hóa Trợ giúp Power Pivot Phân tích PivotTable Thiết kế

Tên PivotTable: Trường Hiện hoạt: → Nhóm Lựa chọn  
Tùy chọn: Thiết đặt Trường Chi tiết: Lấy Tổng Nhóm Nhóm Trường  
PivotTable

A1

Để tạo báo cáo, hãy chọn các trường từ Danh sách Trường PivotTable

Trường PivotTable

Hiện hoạt: Tất cả

Chọn các trường để thêm vào báo cáo:

Tìm kiếm

Dim\_Branch  
Dim\_Brand  
Dim\_Category  
Dim\_Customer

Kéo trường giữa các vùng bên dưới:

Bộ lọc Cột  
Hàng Giá trị

Cập nhật

Hình 353: Môi trường đã được thiết lập sẵn sàng

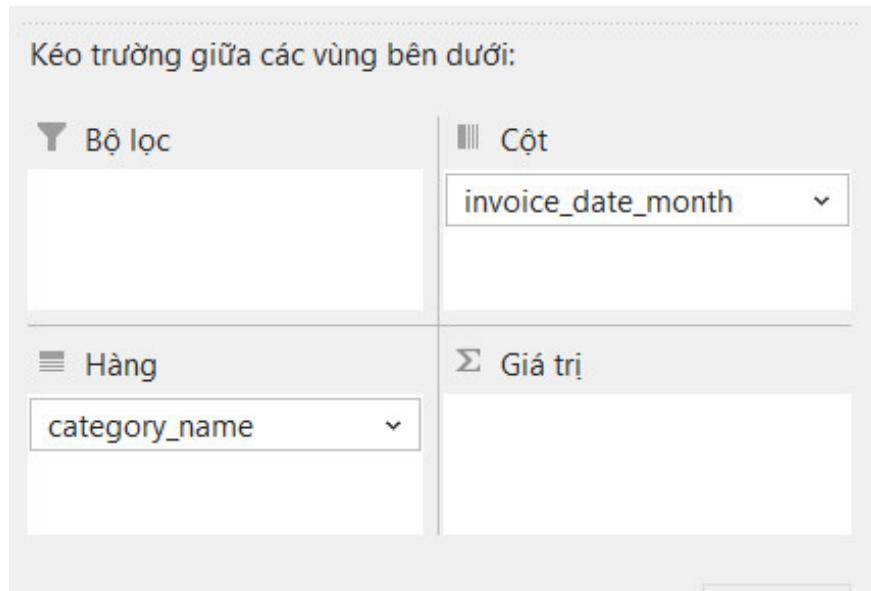
### 3.12.3. Thực hiện Report bằng Excel:

#### 3.12.3.1. Câu truy vấn 1:

**Nội dung câu truy vấn:** Với top 5 kho sản phẩm có doanh số bán ra cao nhất, thống kê số lượng sản phẩm bán ra được của 5 kho đó theo từng tháng.

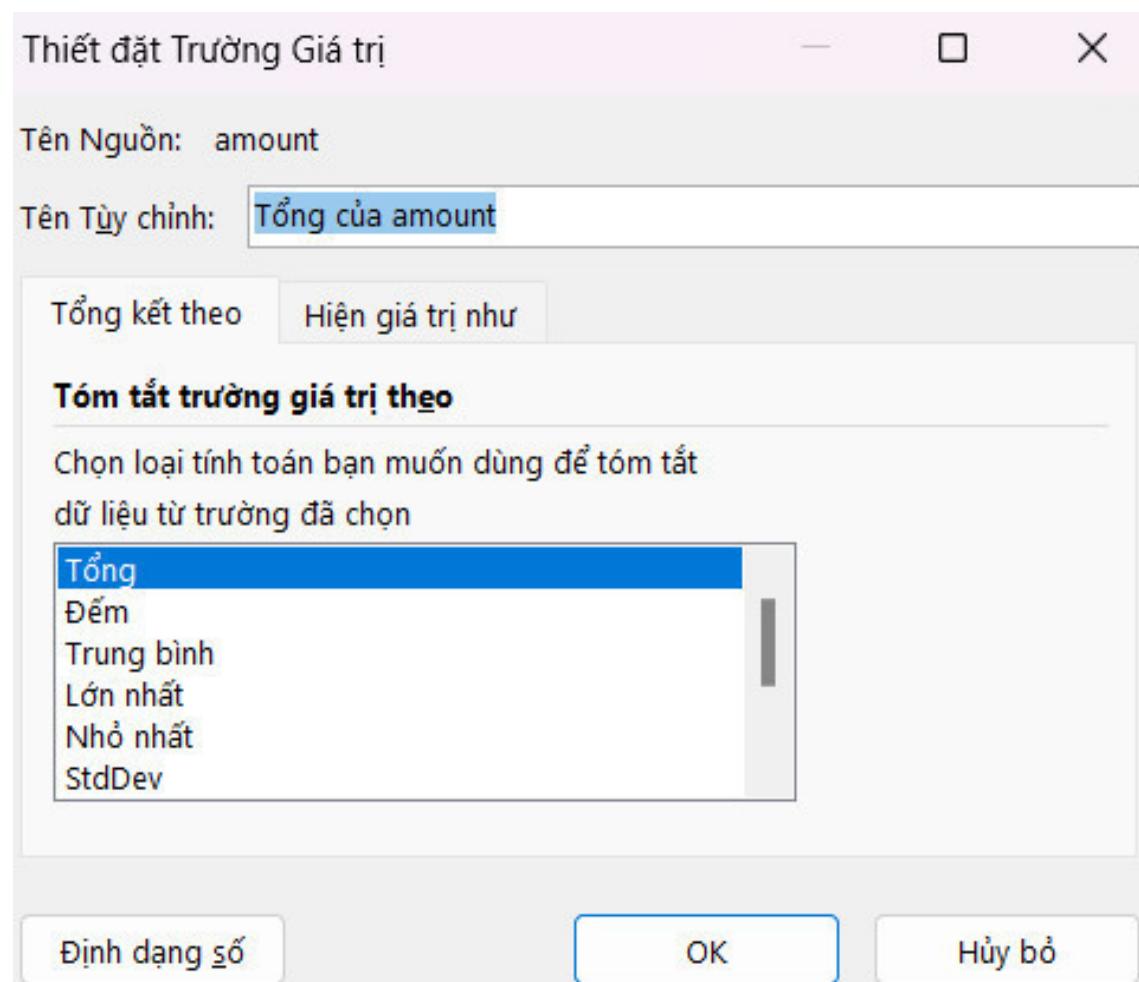
**Thực hiện Report:**

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính category\_name của bảng Dim\_Category vào mục “Cột”, invoice\_date\_month vào mục “Hàng”.



Hình 354: Kéo thả các thuộc tính thích hợp vào trường “Cột” và “Hàng” của Pivot Table

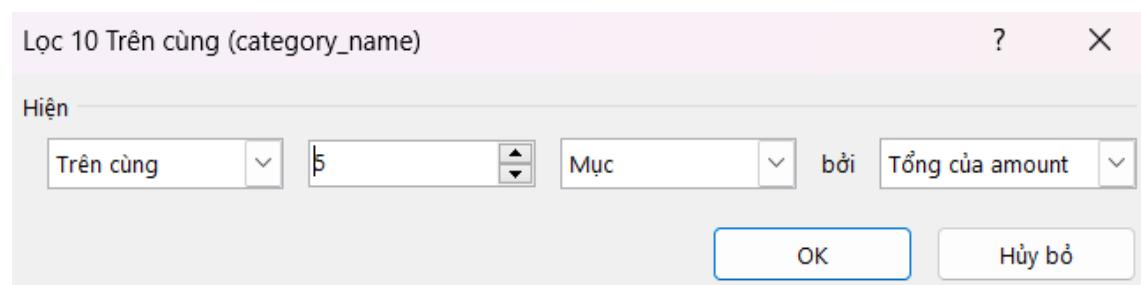
**Bước 2:** Kéo thuộc tính amount của bảng Fact vào mục “Giá trị” chọn “Tổng” cho “Tóm tắt giá trị trường theo” cho thuộc tính này.



Hình 355: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Tại “**Bộ lọc giá trị**” của thuộc tính category\_name, tiến hành lọc ra top 5 kho sản phẩm có doanh số bán ra cao nhất.

Hình 356: Chọn lọc ra top 10 doanh số theo kho sản phẩm



Hình 357: Chính sửa lại chỉ lấy top 5 theo yêu cầu của câu truy vấn 1

**Bước 4:** Theo dõi kết quả Report dưới dạng matrix.

Tổng của amount		Nhân cột			
Nhân Hàng		1	2	3	4 Tổng Cuối
Beverage	31939	34851	41651	1375	109816
Cosmetics	5862	6839	7569	244	20514
Dairy Breakfast	19488,873	21499,172	24729,466	887,83	66605,341
Detergent Cleaning	6071	6687	7595	294	20647
Food	59846,099	63269,224	73796,747	2487,634	199399,704
<b>Tổng Cuối</b>	<b>123206,972</b>	<b>133145,396</b>	<b>155341,213</b>	<b>5288,464</b>	<b>416982,045</b>

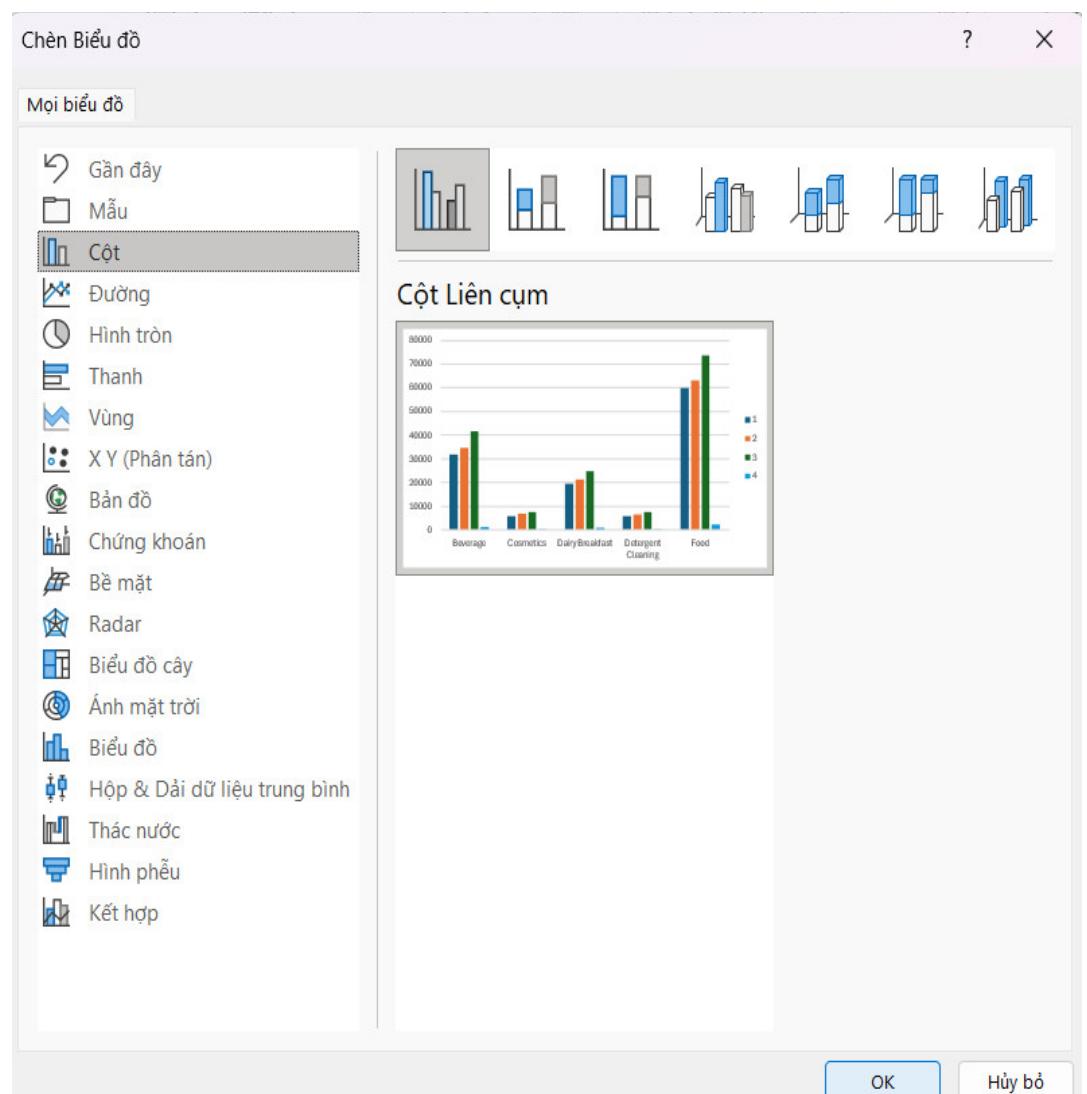
Hình 358: Kết quả hiển thị câu truy vấn 1 dưới dạng matrix

**Bước 5:** Theo dõi kết quả report dạng biểu đồ - Để hiển thị biểu đồ, tại tab “Phân biệt PivotTable”, nhấn “Pivot chart”.

The screenshot shows the Microsoft Excel ribbon with the 'Phân tích PivotTable' tab selected. Below the ribbon, a PivotTable is displayed in the worksheet area. The PivotTable has columns labeled 'Nhân Hàng' (Category), '1', '2', '3', and '4 Tổng Cuối'. The data rows include Beverage, Cosmetics, Dairy Breakfast, Detergent Cleaning, and Food, with their respective values. To the right of the PivotTable, the 'Trường PivotTable' (PivotTable Fields) pane is open, showing the fields available for charting. The 'PivotChart' icon in the 'Công cụ' (Tools) group of the ribbon is highlighted with a red box.

Hình 359: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ

Chọn loại “Chart” và nhấn “OK”.



Hình 360: Giao diện chọn loại biểu đồ

Kéo trỏng giữa các vùng bên dưới:

**Bộ lọc**

invoice\_date\_month

**Chú giải (Chuỗi)**

category\_name

**Trục (Thể loại)**

category\_name

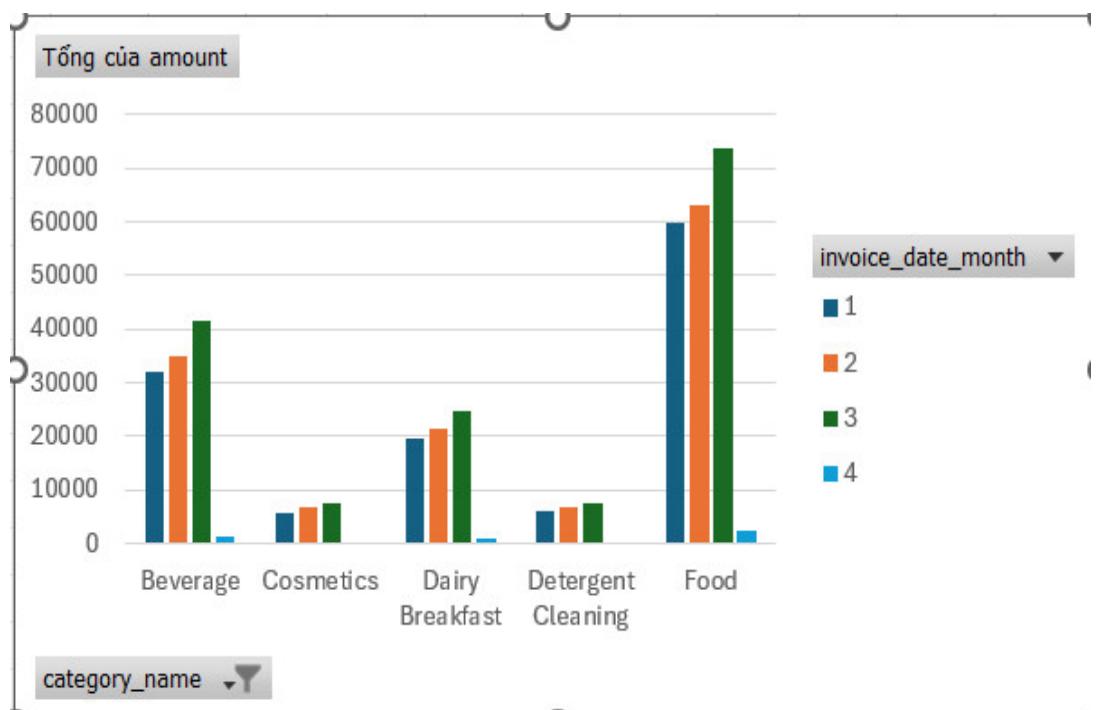
**Giá trị**

Tổng của amount

Trì hoãn Cập nhật Bổ trợ

Cập nhật

Hình 361: Điều chỉnh lại chú giải, trục và giá trị



Hình 362: Kết quả câu truy vấn 1 hiển thị dưới dạng biểu đồ

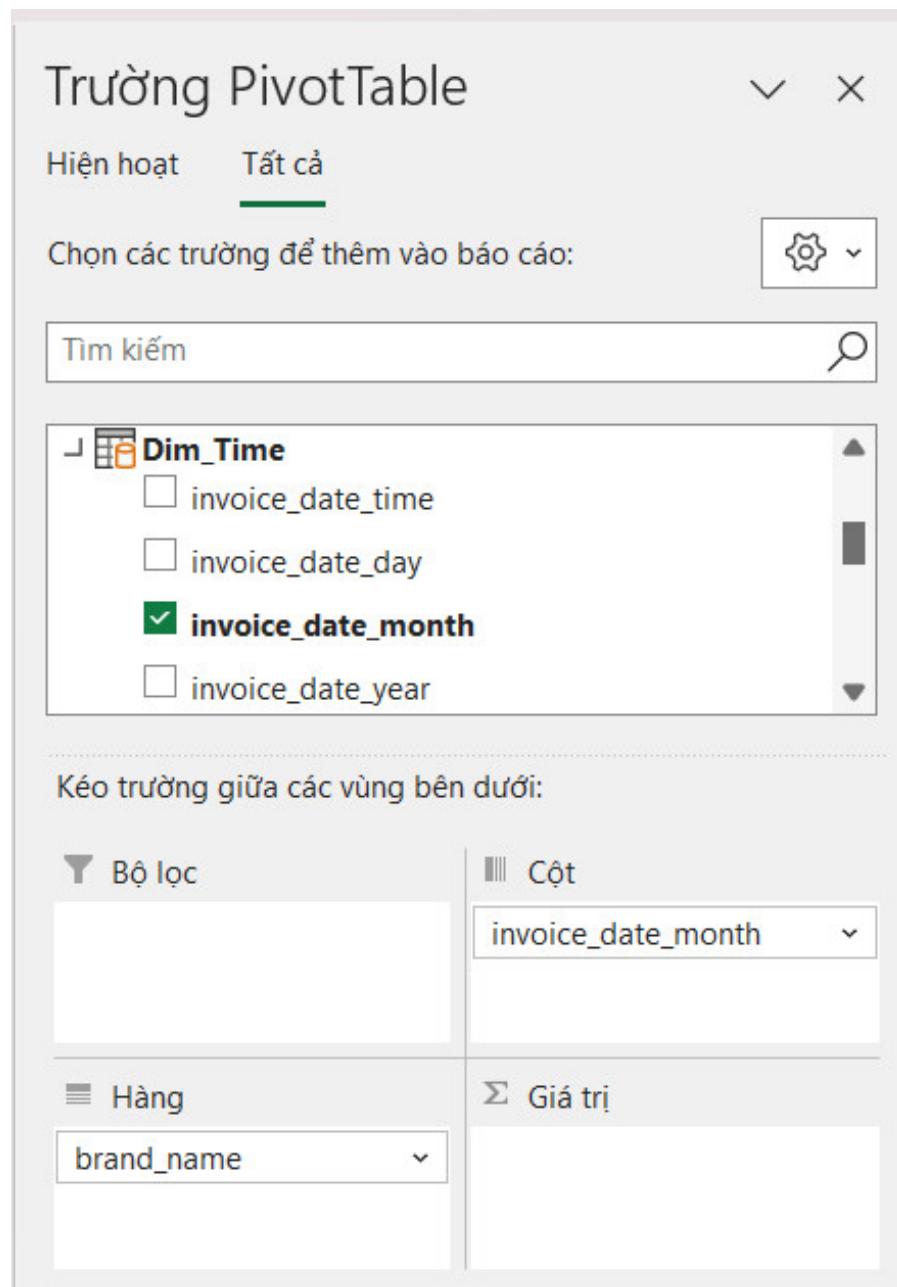
**Nhận xét:** Kết quả thực thi câu truy vấn 1 bằng ngôn ngữ Pivot Table tương tự như khi thao tác kéo thả trên browser của công cụ BI và ngôn ngữ truy vấn MDX.

### 3.12.3.2. Câu truy vấn 2:

**Nội dung câu truy vấn:** Thống kê tổng số lượng sản phẩm bán được của các nhãn hàng ‘Sprite’, ‘Clear’, ‘Nestle’, ‘Lipton’, ‘Colgate’ trong tháng 02 năm 2017.

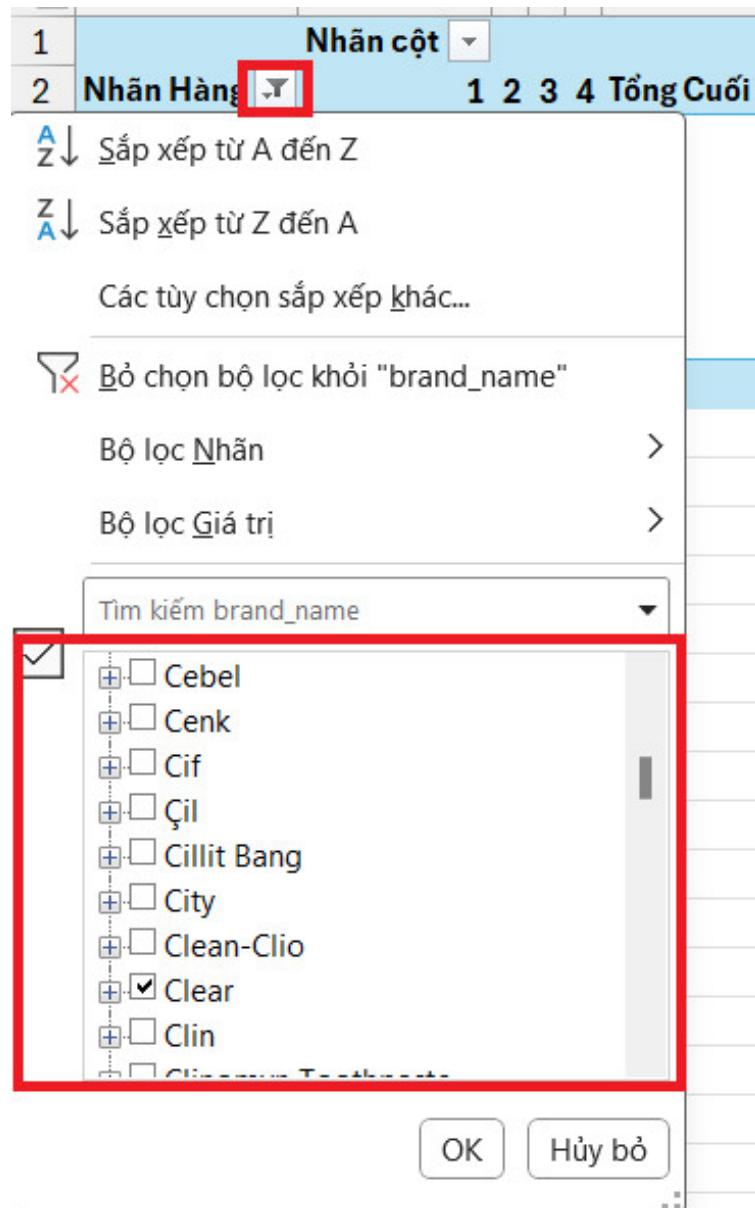
**Thực hiện Report:**

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính invoice\_date\_month của bảng Dim\_Time vào mục “Cột”, brand\_name của bảng Dim\_Brand vào mục “Hàng”.



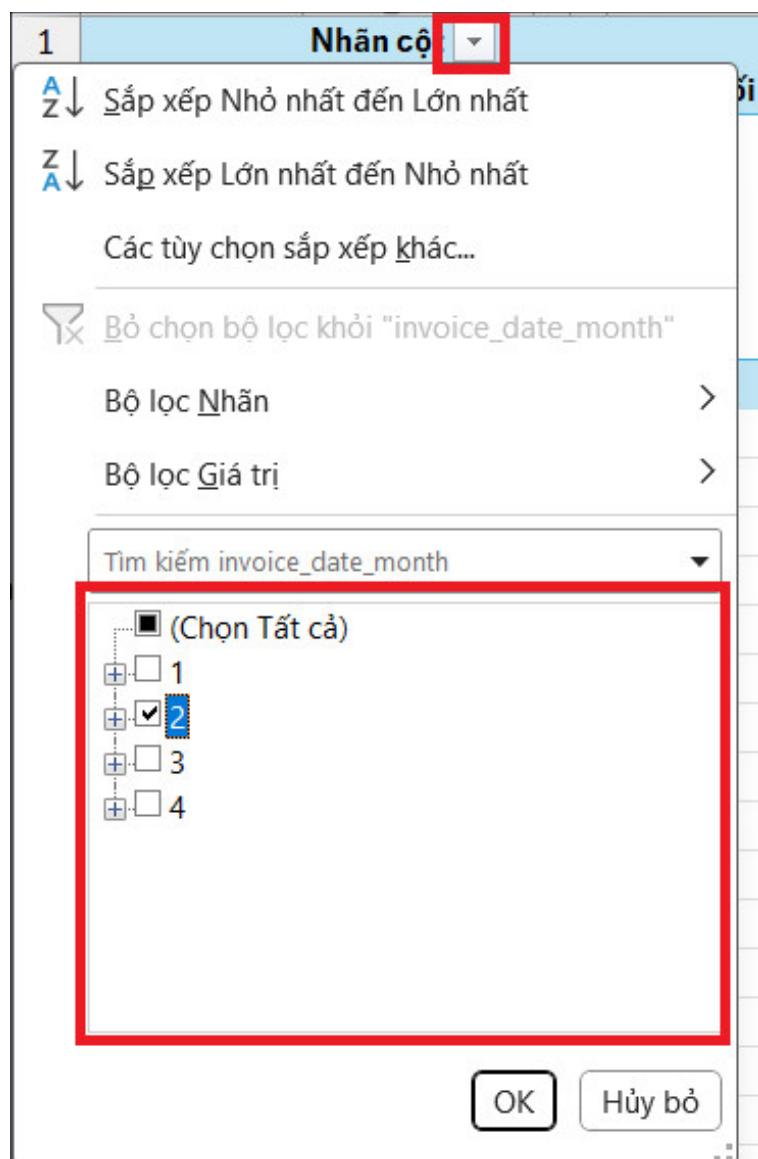
Hình 363: Kéo thả các thuộc tính thích hợp vào trường “Cột” và “Hàng” của Pivot Table

**Bước 2:** Tại “**Bộ lọc giá trị**” của thuộc tính brand\_name, tiến hành lọc ra 5 thương hiệu sản phẩm ‘Sprite’, ‘Clear’, ‘Nestle’, ‘Lipton’, ‘Colgate’.



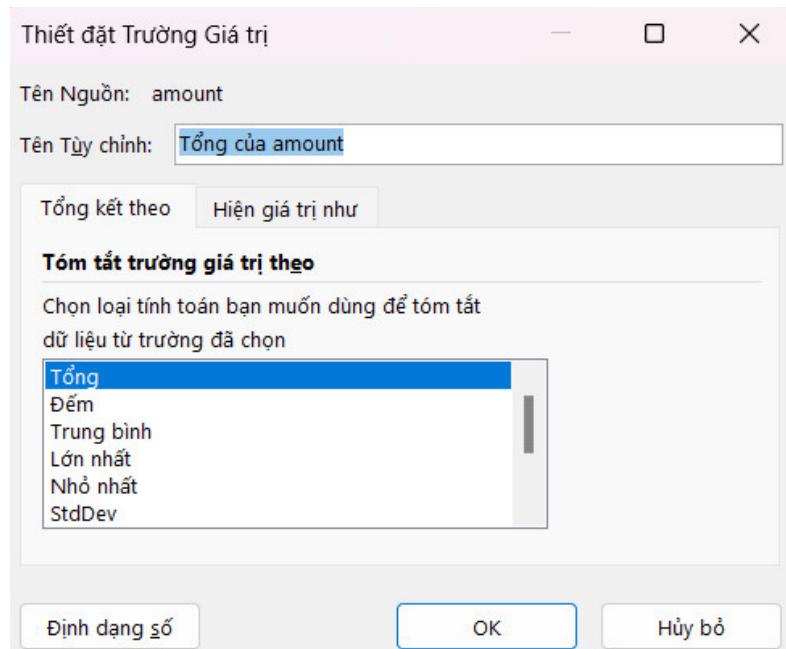
Hình 364: Lọc ra tên các thương hiệu sản phẩm được đề cập trong câu truy vấn

**Bước 3:** Tại “**Bộ lọc giá trị**” của thuộc tính invoice\_date\_month, tiến hành lọc ra tháng 2.



Hình 365: Tiến hành chọn các kết quả năm trong tháng 2

**Bước 4:** Kéo thuộc tính amount của table Fact thả vào mục “Values”, chọn “Thiết đặt giá trị trường” để chọn “Tổng” cho “Tóm tắt trường giá trị theo” của thuộc tính này.



Hình 366: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 5:** Theo dõi kết quả report dạng table.

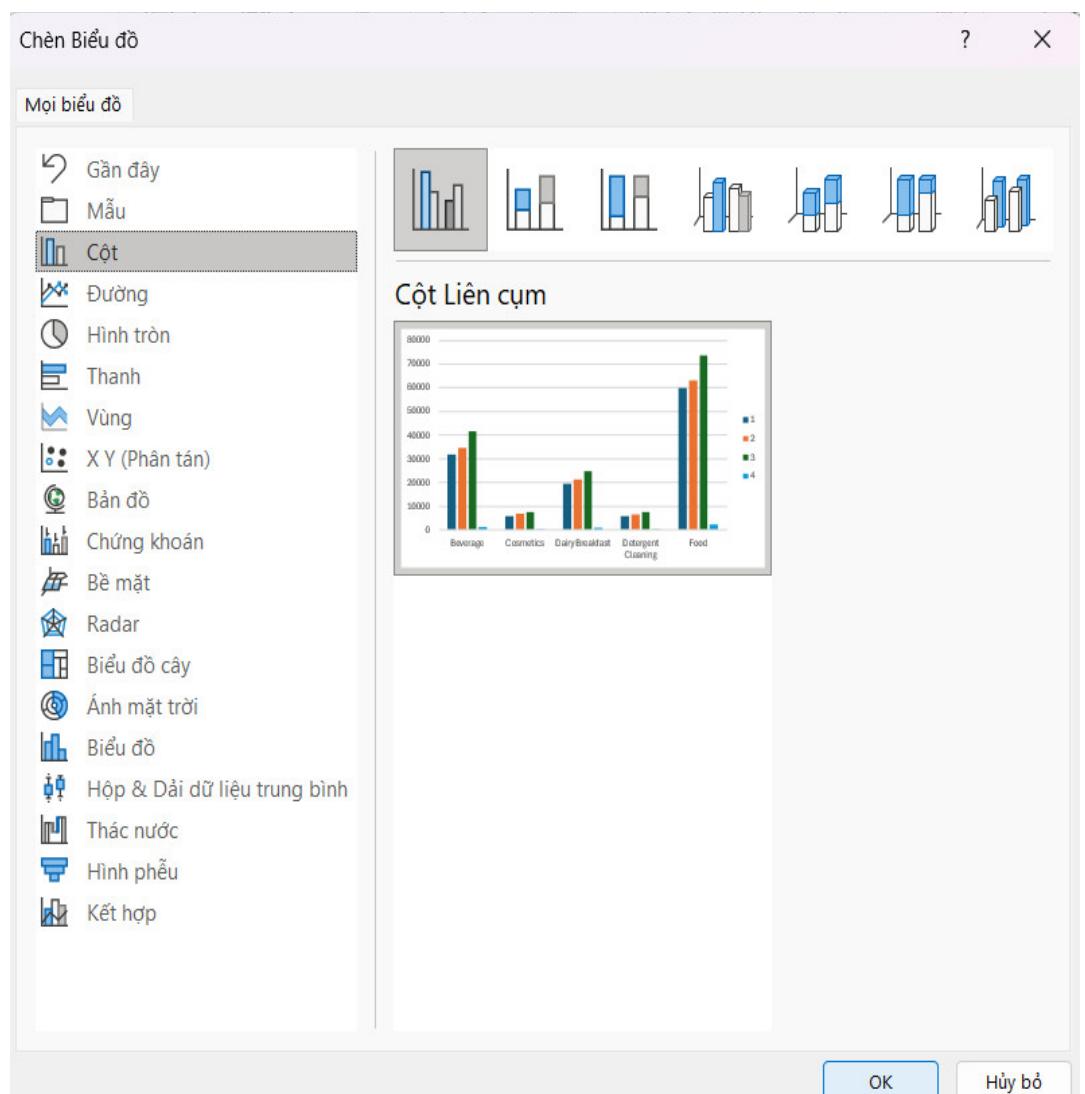
Tổng của amount Nhãn cột		
Nhãn Hàng	2 Tổng Cuối	
Clear	309	309
Colgate	619	619
Lipton	636	636
Nestle	7792	7792
Sprite	43	43
<b>Tổng Cuối</b>	<b>9399</b>	<b>9399</b>

Hình 367: Kết quả câu truy vấn 2 hiển thị dưới dạng table

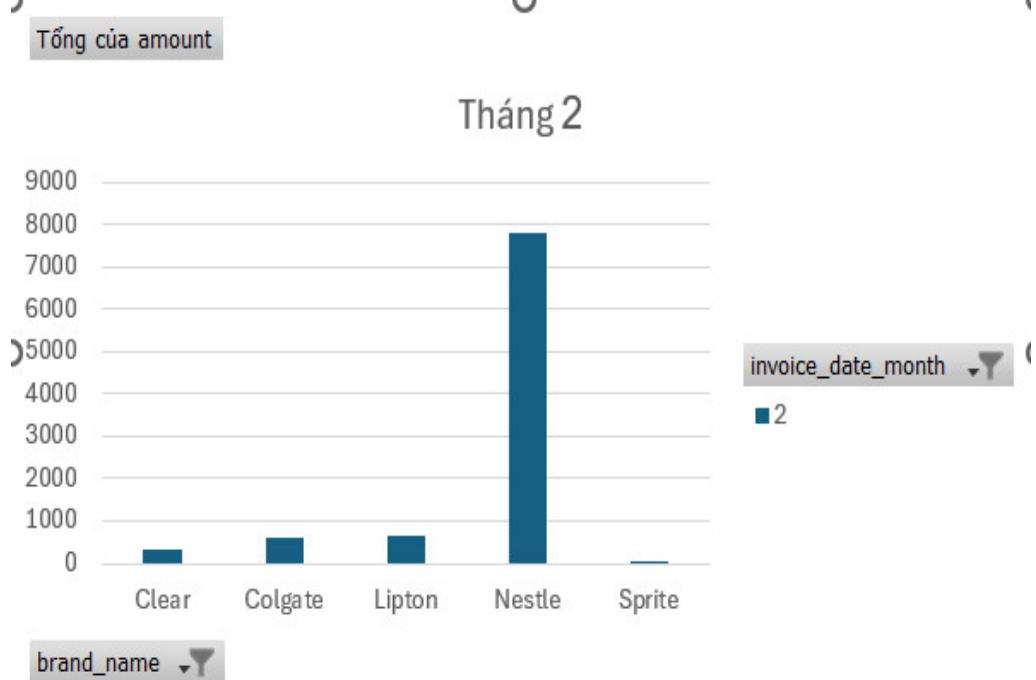
**Bước 6:** Theo dõi kết quả report dạng biểu đồ. Để hiển thị Biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

Hình 368: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại “Chart” và nhấn “OK”.



Hình 369: Giao diện chọn loại biểu đồ



Hình 370: Kết quả câu truy vấn 2 hiển thị dưới dạng biểu đồ

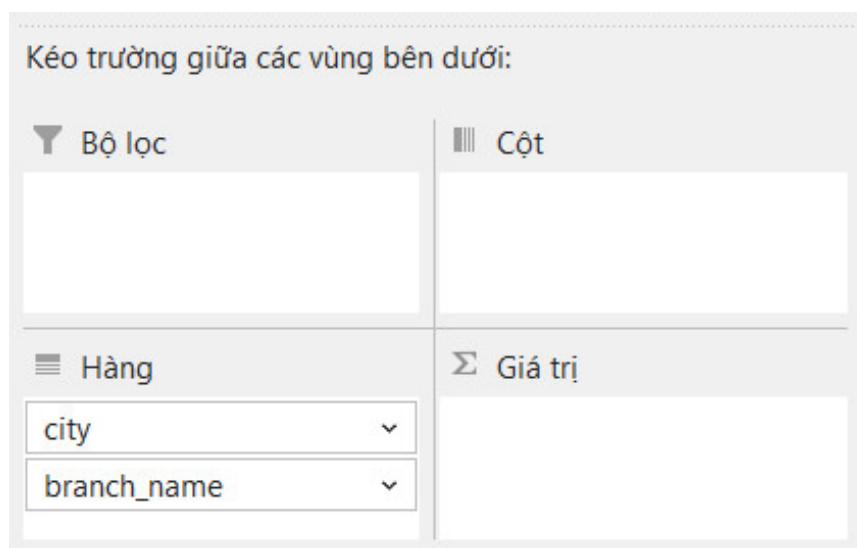
**Nhận xét:** Kết quả thực thi câu truy vấn 2 bằng ngôn ngữ Pivot Table tương tự như khi thao tác kéo thả trên browser của công cụ BI và ngôn ngữ truy vấn MDX.

### 3.12.3.3. Câu truy vấn 3:

**Nội dung câu truy vấn:** Với mỗi thành phố, thống kê số lượng sản phẩm bán ra được ở từng chi nhánh có số lượng sản phẩm bán ra lớn hơn 15000.

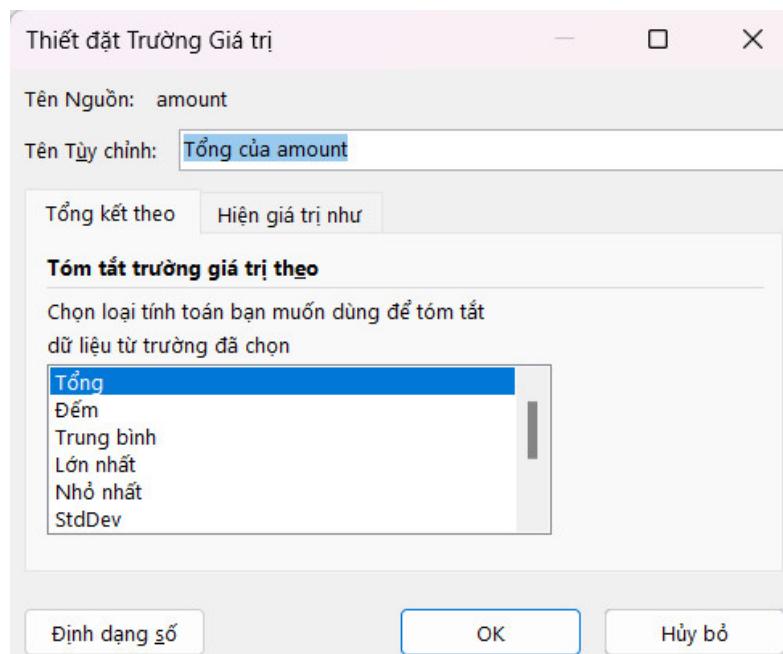
**Thực hiện Report:**

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính city của bảng Dim\_Branch vào mục “Hàng”, branch\_name của bảng Dim\_Branch vào mục “Hàng”.



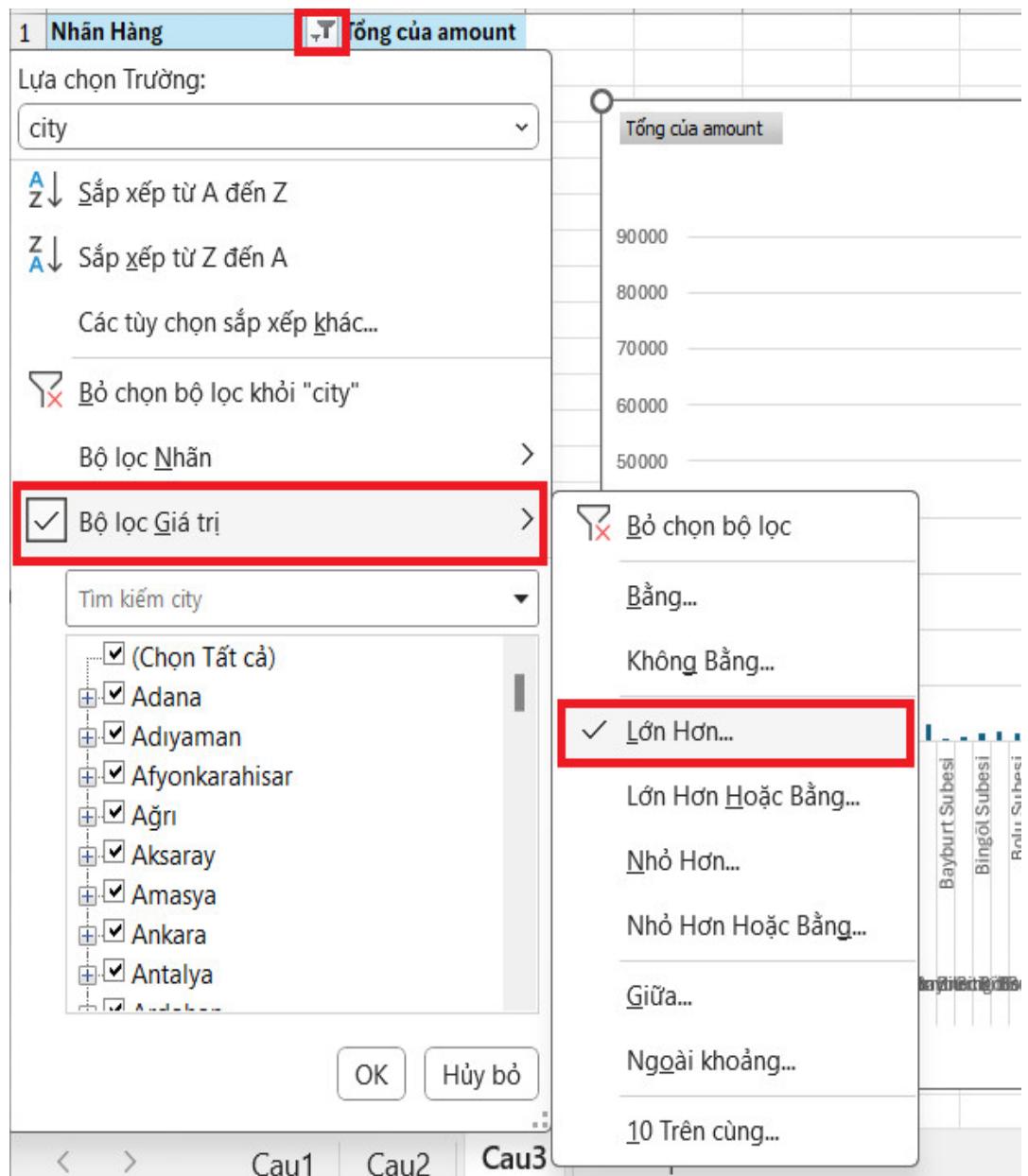
Hình 371: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính amount của bảng Fact vào mục “**Giá trị**”, chọn “**Thiết đặt giá trị trường**” để chọn “**Tổng**” cho “**Tóm tắt trường giá trị theo**” của thuộc tính này.

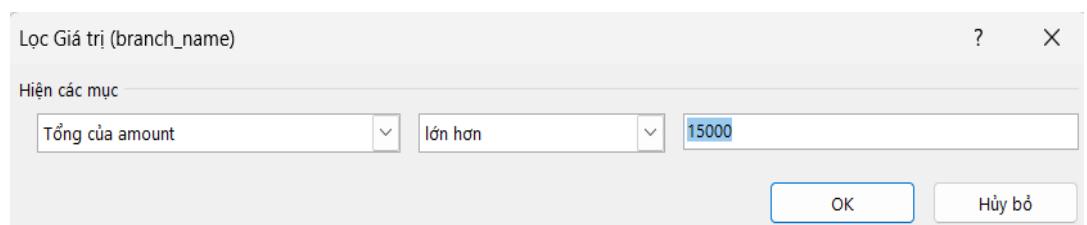


Hình 372: Kéo thả thuộc tính thích hợp vào trường “**Giá trị**” và chọn hàm tính toán

**Bước 3:** Tại “**Bộ lọc giá trị**” của thuộc tính branch\_name, tiến hành lọc ra giá trị có amount > 15000.



Hình 373: Lọc các dòng dữ liệu theo điều kiện lớn hơn



Hình 374: Thiết lập giá trị điều kiện của câu truy vấn

**Bước 4:** Theo dõi kết quả report dạng table.

Nhân Hàng		Tổng của amount
✉ Ankara		
Ankara Subesi	30337,849	
✉ Bursa		
Bursa Subesi	15699,336	
✉ İstanbul		
İstanbul Subesi	83586,31	
✉ İzmir		
İzmir Subesi	24138,662	
<b>Tổng Cuối</b>	<b>153762,157</b>	

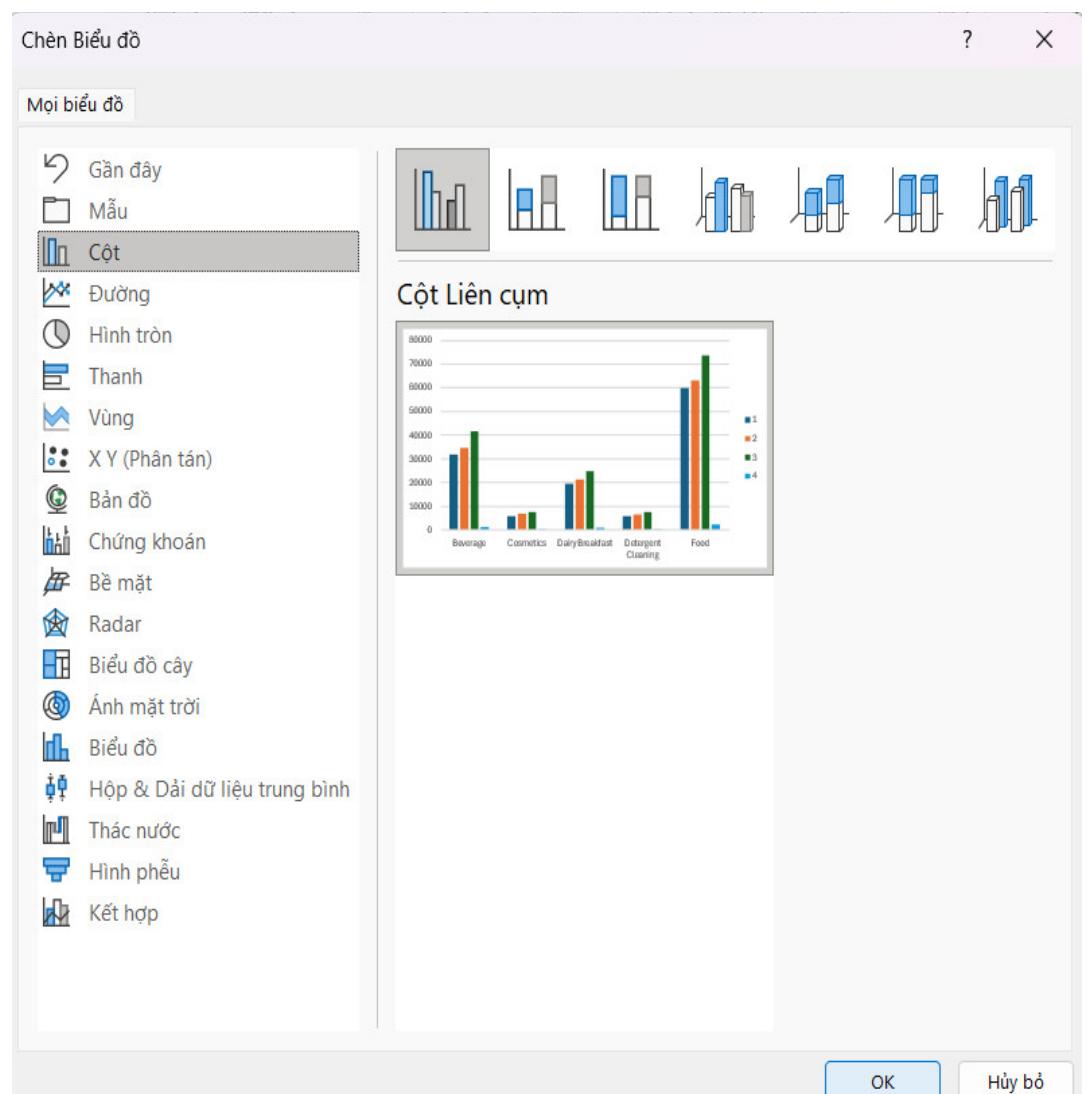
Hình 375: Kết quả câu truy vấn 3 hiển thị dưới dạng table

**Bước 5:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

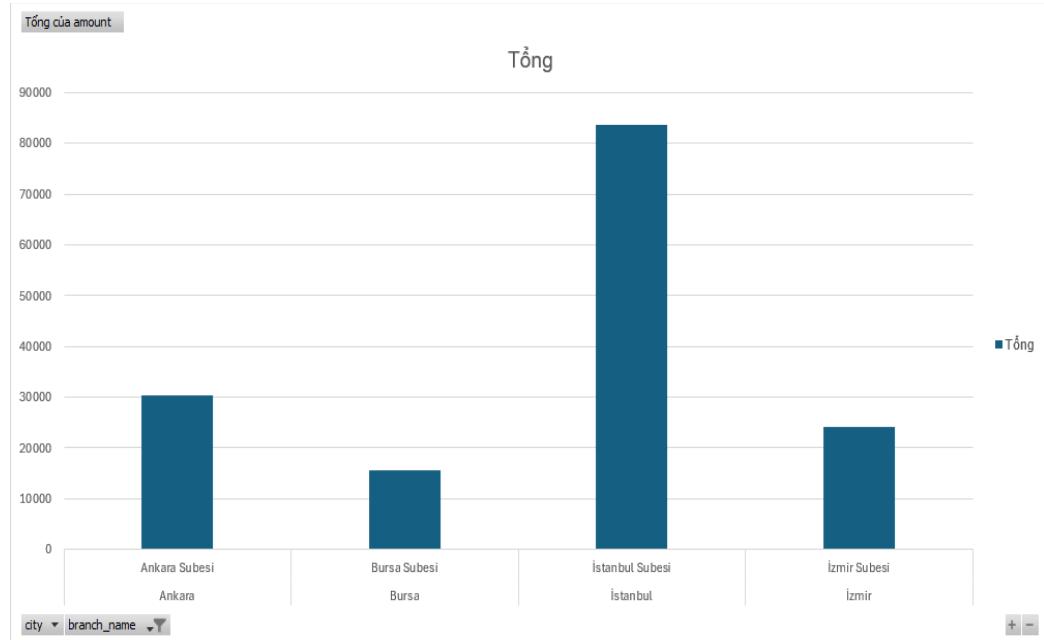


Hình 376: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 377: Giao diện chọn loại biểu đồ



Hình 378: Kết quả câu truy vấn 3 hiển thị dưới dạng biểu đồ

### 3.12.3.4. Câu truy vấn 4:

**Nội dung câu truy vấn:** Thống kê tổng doanh thu trong tháng 04 tại các chi nhánh ở vùng miền “Marmara” dựa trên nhãn hàng và loại sản phẩm.

**Thực hiện Report:**

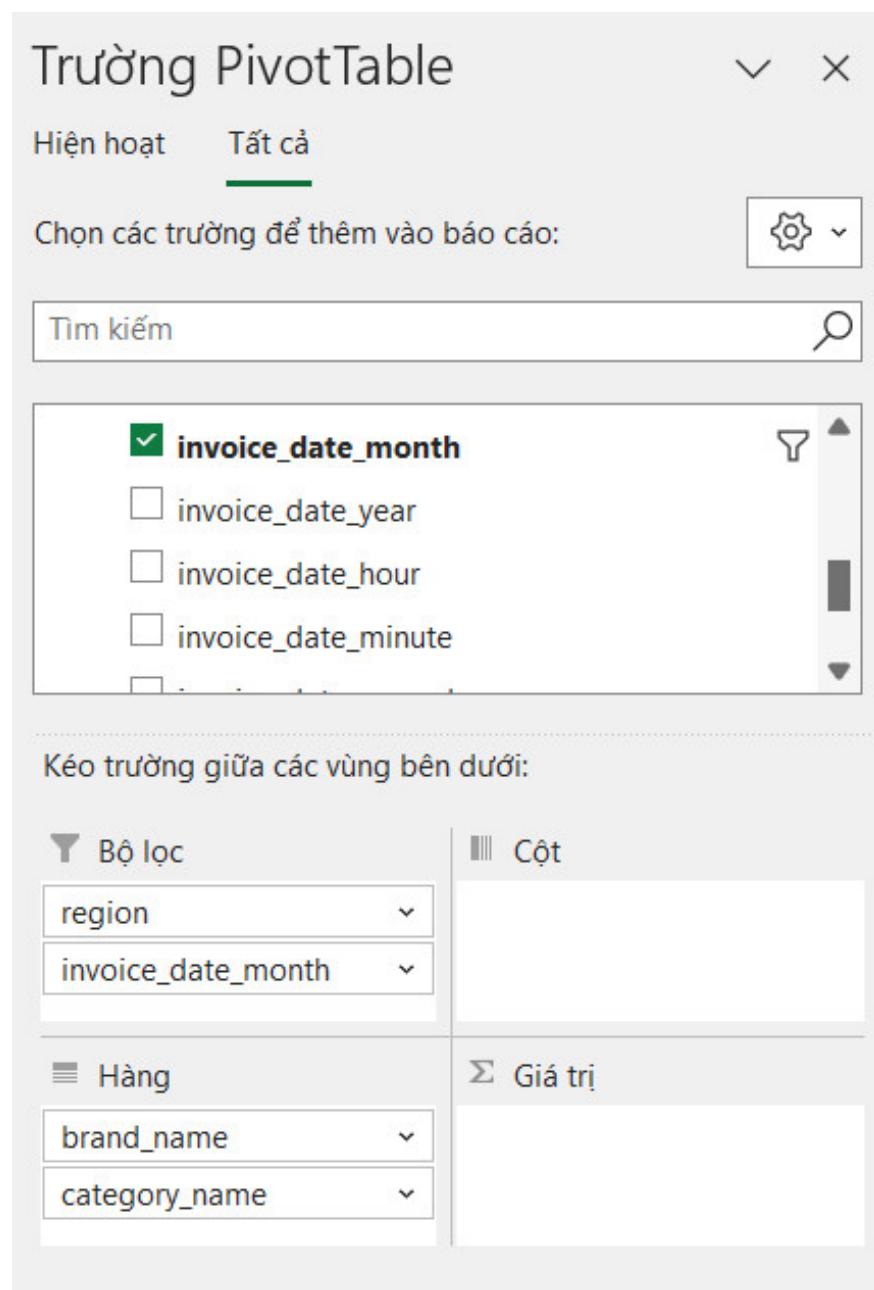
**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính category\_name của bảng Dim\_Category vào mục “Hàng”, brand\_name của bảng Dim\_Brand vào mục “Hàng”.

Hình 379: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Tiến hành kéo thả thuộc tính Region của bảng Dim\_Brand và Invoice Date Month của bảng Dim\_Time vào phần “**Bộ lọc**”.

The screenshot shows the Power BI Filter pane. At the top, there is a list of columns from the 'Dim\_Brand' table: branch\_id, branch\_name, city, and region. The 'region' column is highlighted with a green checkmark. Below this, a message says 'Kéo trêng giữa các vùng bên dưới:' (Drag between the regions below). The pane is divided into four sections: 'Bộ lọc' (Filter) containing 'region', 'Cột' (Column) empty, 'Hàng' (Row) containing 'brand\_name' and 'category\_name', and 'Giá trị' (Value) empty.

Hình 380: Thêm thuộc tính region vào bộ lọc



Hình 381: Thêm thuộc tính invoice\_date\_month vào bộ lọc

**Bước 3:** Kéo thả thuộc tính final\_total\_trans của bảng Fact vào mục “Giá trị”, chọn “Thiết đặt giá trị trường” để chọn “Tổng” cho Tóm tắt trường giá trị theo” cho thuộc tính này.

Kéo trường giữa các vùng bên dưới:

Bộ lọc

region

invoice\_date\_month

Cột

Hàng

brand\_name

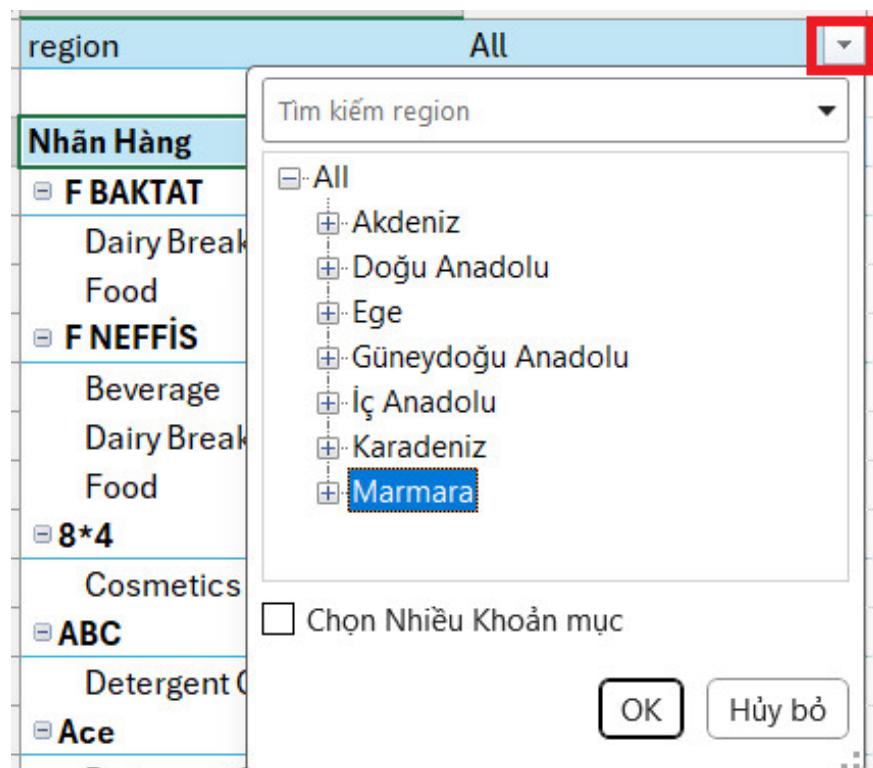
category\_name

Giá trị

Tổng của final\_total\_trans

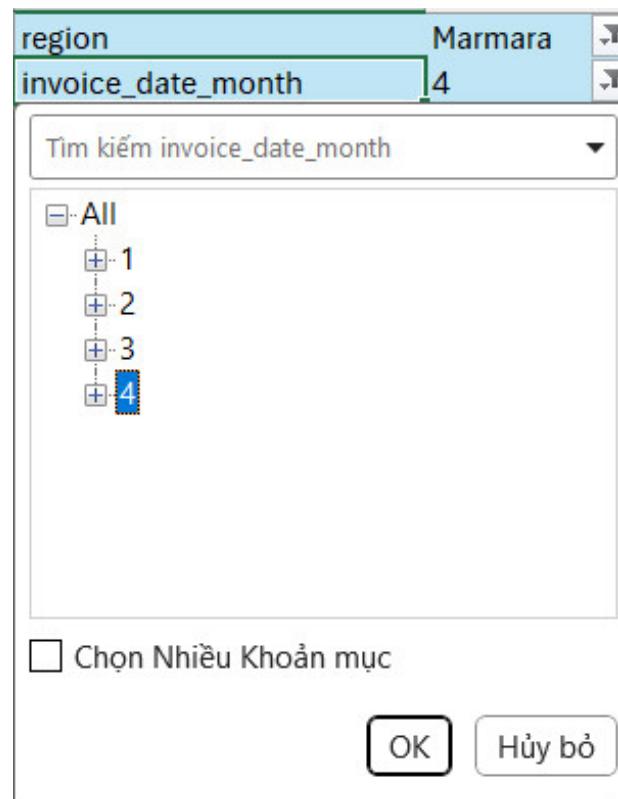
Hình 382: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 4:** Tại “Bộ lọc giá trị” của thuộc tính region, tiến hành lọc ra giá trị Marmara.



Hình 383: Lọc dữ liệu theo vùng miền Marmara

**Bước 5:** Tại “**Bộ lọc giá trị**” của thuộc tính invoice\_date\_month, tiến hành lọc ra giá trị 4.



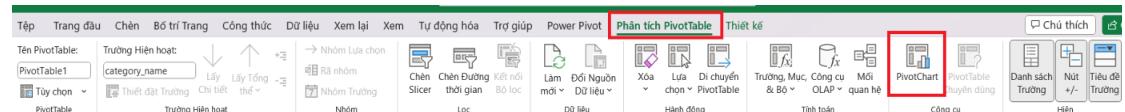
Hình 384: Lọc dữ liệu theo tháng 04

**Bước 6:** Theo dõi kết quả report dưới dạng table.

region	Marmara	
invoice_date_month	4	
<b>Nhãn Hàng</b>	<b>Tổng của final_total_trans</b>	
<b>F NEFFIS</b>		
Beverage	59,82	
Dairy Breakfast	25,93	
Food	117,84	
<b>Ace</b>		
Detergent Cleaning	8,88	
<b>Activex</b>		
Cosmetics	5,53	
<b>Air Wick</b>		
Detergent Cleaning	22,4	
<b>Akat</b>		
Cosmetics	12,63	
<b>Akdem</b>		
Beverage	7,35	
Dairy Breakfast	5,09	
Food	4,57	
<b>Algida</b>		
Dairy Breakfast	9,01	
<b>Ankara</b>		
Food	29,53	
<b>Arbella</b>		
Food	4,57	
<b>Ariel</b>		
Detergent Cleaning	27	

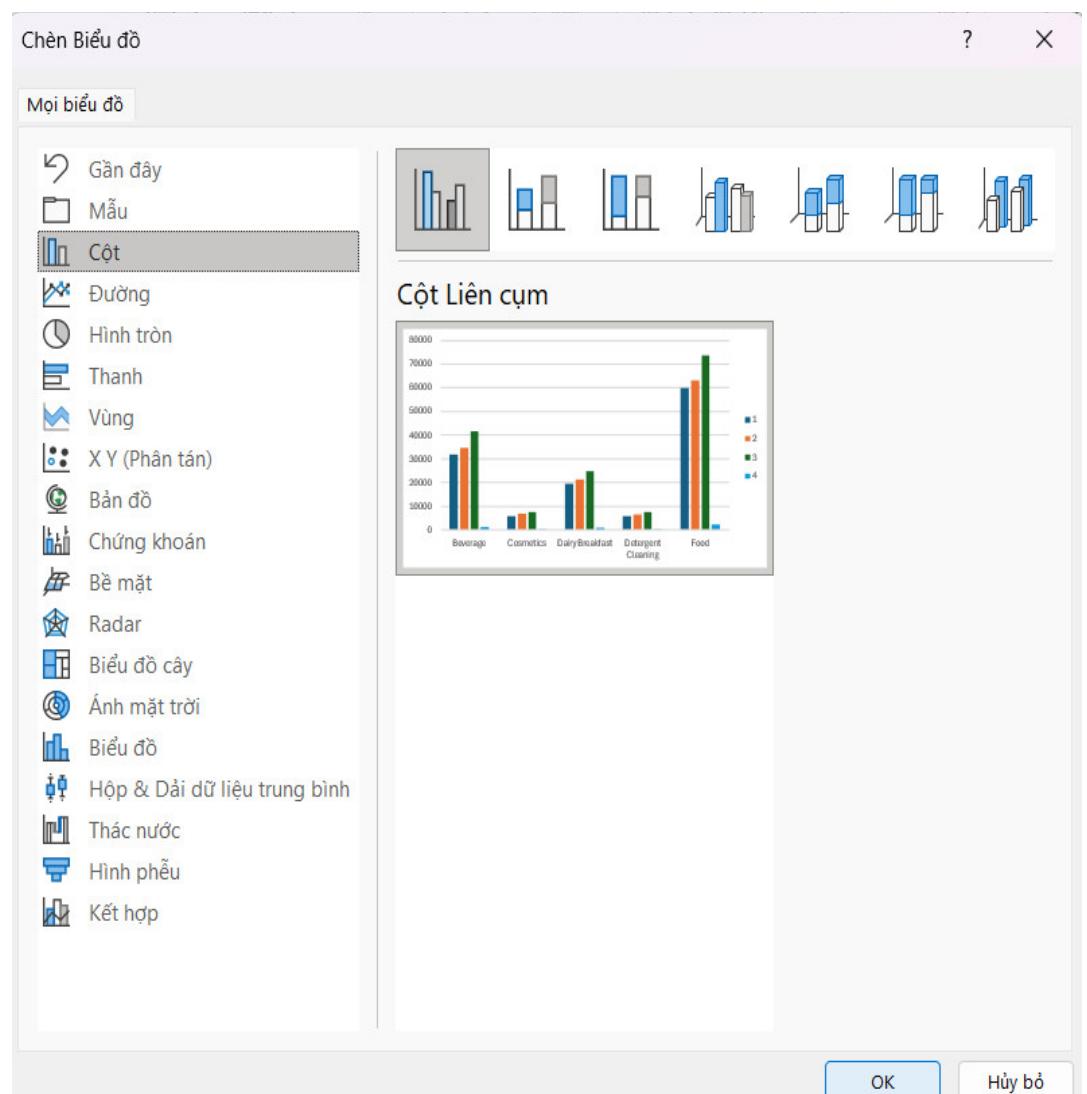
Hình 385: Kết quả câu truy vấn 4 hiển thị dưới dạng table

**Bước 7:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

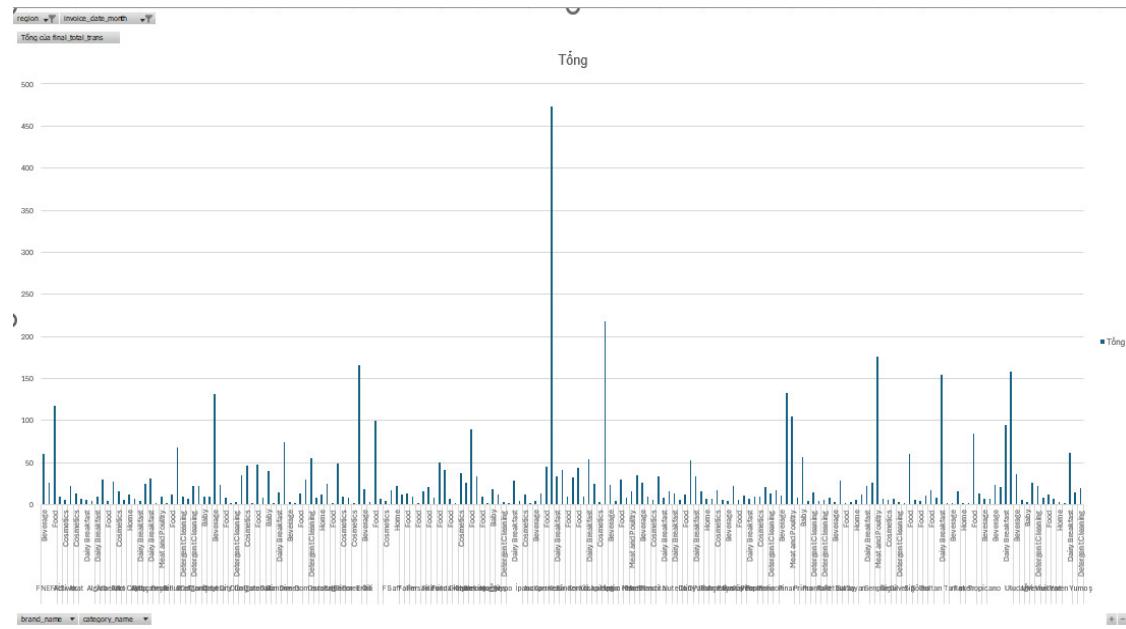


Hình 386: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 387: Giao diện chọn loại biểu đồ



Hình 388: Kết quả câu truy vấn 4 hiển thị dưới dạng biểu đồ

### 3.12.3.5. Câu truy vấn 5:

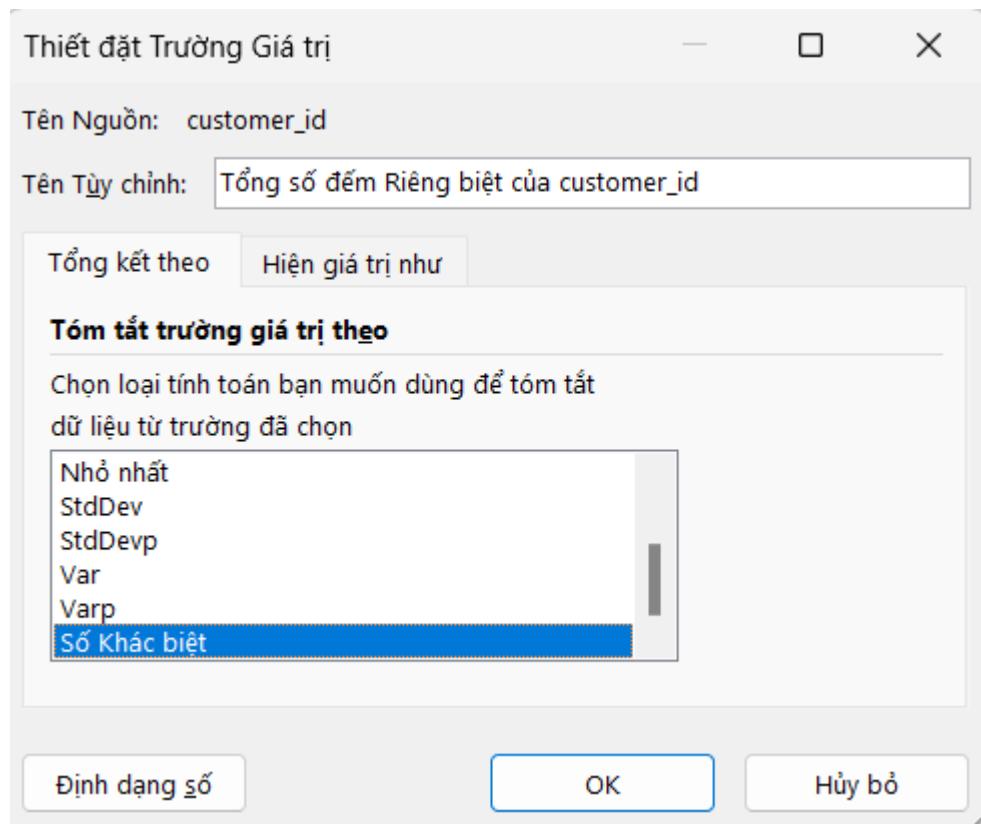
**Nội dung câu truy vấn:** Thống kê với 3 sản phẩm bán được với doanh số cao nhất. theo từng tháng, quý, năm. Cho biết số lượng khách hàng đã mua 3 loại sản phẩm đó.

#### Thực hiện Report:

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính invoice\_date\_year, invoice\_date\_month, invoice\_date\_day của bảng Dim\_Time vào mục “Hàng”.

Hình 389: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính customer\_id của bảng Fact vào mục “**Giá trị**”, chọn “**Số khác biệt**” cho “**Tóm tắt trường giá trị theo**” của thuộc tính này.



Hình 390: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Kéo thả thuộc tính product\_name của bảng Dim\_Product vào mục “**Hàng**”, và tiến hành lọc top 3 sản phẩm có số lượng khách hàng mua cao nhất.

3 Nhãn Hàng

Lựa chọn Trường:

product\_name

A ↓ Sắp xếp từ A đến Z

Z ↓ Sắp xếp từ Z đến A

Các tùy chọn sắp xếp khác...

Bỏ chọn bộ lọc khỏi "product\_name"

Bộ lọc Nhãn

Bộ lọc Giá trị

Tìm kiếm product\_name

- (Chọn Tất cả)
- AIR WICK FRESH BUBBLE PINK OR
- AKDEM FLAX SEED 100 GR.
- AKDEM GINSENG SCENT 20'S
- AKDEM LIDEN 50GR WITH LEAVE!
- ALGIDA FESTIVAL CARAMEL VAN
- ANCHOR NAIL CLIPPERS
- ARGEM RICE & DESSERT MOLD
- ARCEM SPICE BAKER

OK Hủy bỏ

Tổng số đếm của customer\_id

	Tổng số đếm của customer_id
1	1
6	6
6	6
22	22
21	21
42	42
37	37

Bỏ chọn bộ lọc

Bằng...

Không Bằng...

Lớn Hơn...

Lớn Hơn Hoặc Bằng...

Nhỏ Hơn...

Nhỏ Hơn Hoặc Bằng...

Giữa...

Ngoài khoảng...

10 Trên cùng...

Hình 391: Chọn lọc ra top 10 sản phẩm theo doanh số

Lọc 10 Trên cùng (product\_name)

Hiện

Trên cùng 3 Mục bởi Tổng của amount

OK Hủy bỏ

Hình 392: Chính sửa lại chỉ lấy top 3 theo yêu cầu của câu truy vấn 5

**Bước 4:** Theo dõi kết quả report dưới dạng table.

Nhân Hàng		Tổng số đếm Riêng biệt của customer_id
⊕	MAGIC HANDS CIG MEATBALLS 200GR	2689
⊕	NESCAFE 3 IN 1 18 GR	1812
⊕	OSMANCIK PIRINC KG.	3490
<b>Tổng Cuối</b>		<b>7357</b>

Hình 393: Kết quả câu truy vấn 5 hiển thị dưới dạng table (rút gọn)

Nhân Hàng		Tổng số đếm Riêng biệt của customer_id
⊕ MAGIC HANDS CIG MEATBALLS 200GR		
⊕	2017	
⊕	1	
3		35
4		103
5		16
6		41
7		31
8		93
9		34
10		43
11		66
12		27
13		78
14		68
15		34
16		56
17		43
18		33
19		28
20		51
21		41
22		33
23		16
24		27

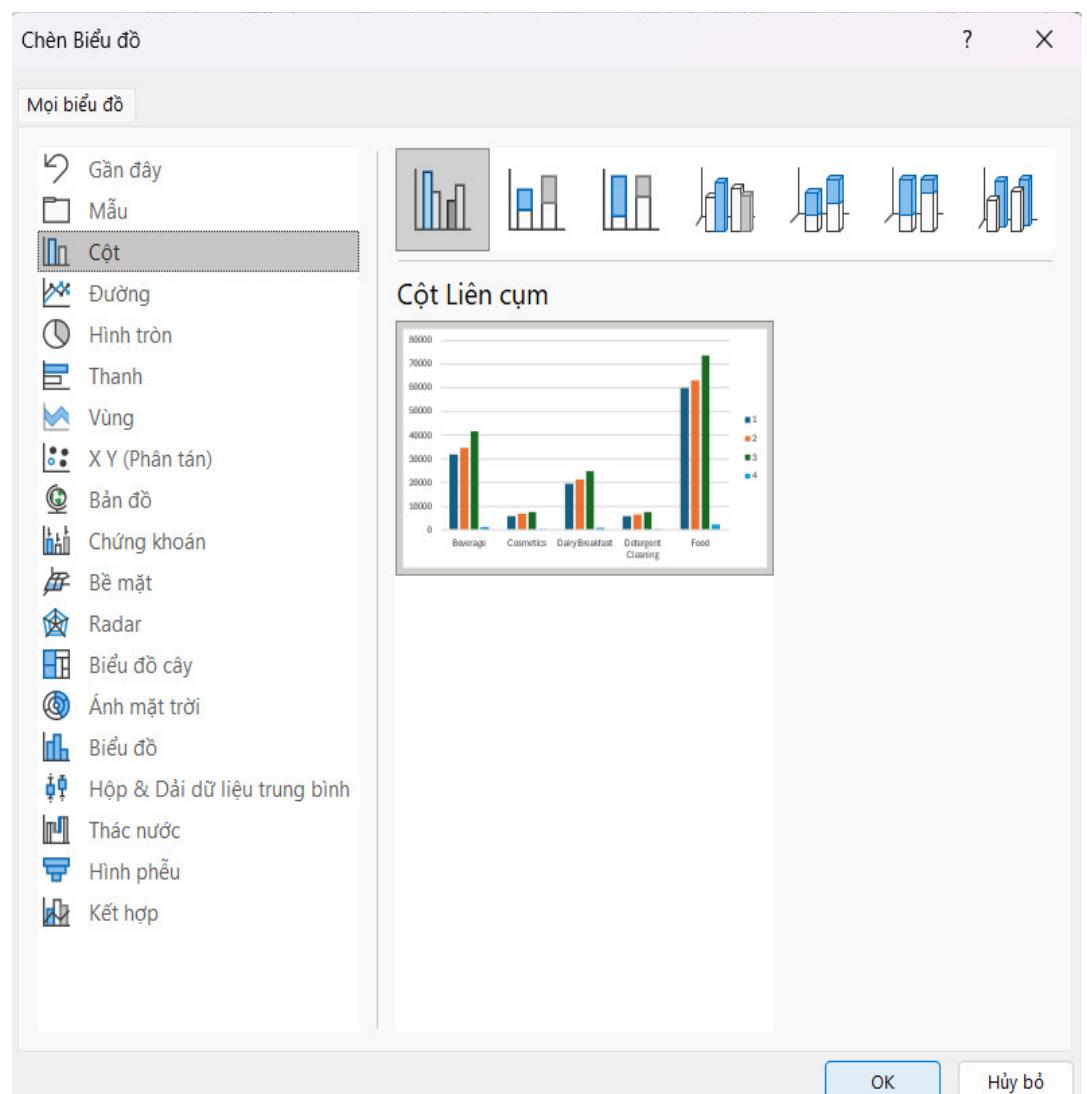
Hình 394: Kết quả câu truy vấn 5 hiển thị dưới dạng table (đầy đủ)

**Bước 5:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

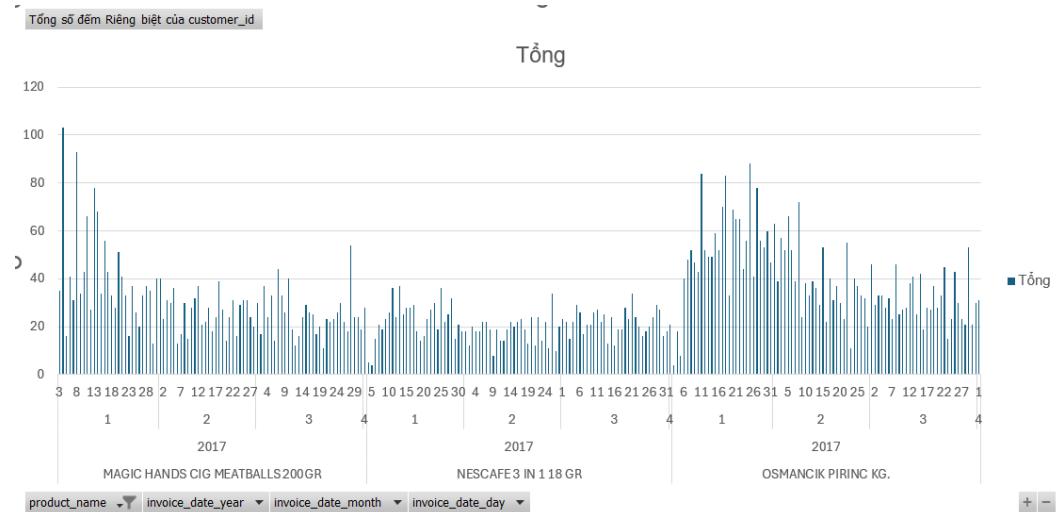


Hình 395: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

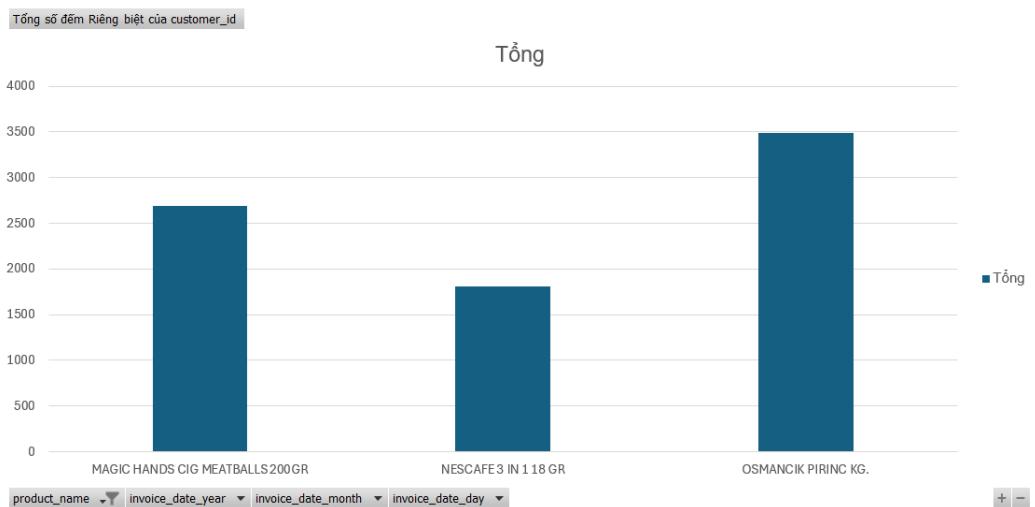
Chọn loại Chart và nhấn “OK”.



Hình 396: Giao diện chọn loại biểu đồ



Hình 397: Kết quả câu truy vấn 5 hiển thị dưới dạng biểu đồ (dày đú)



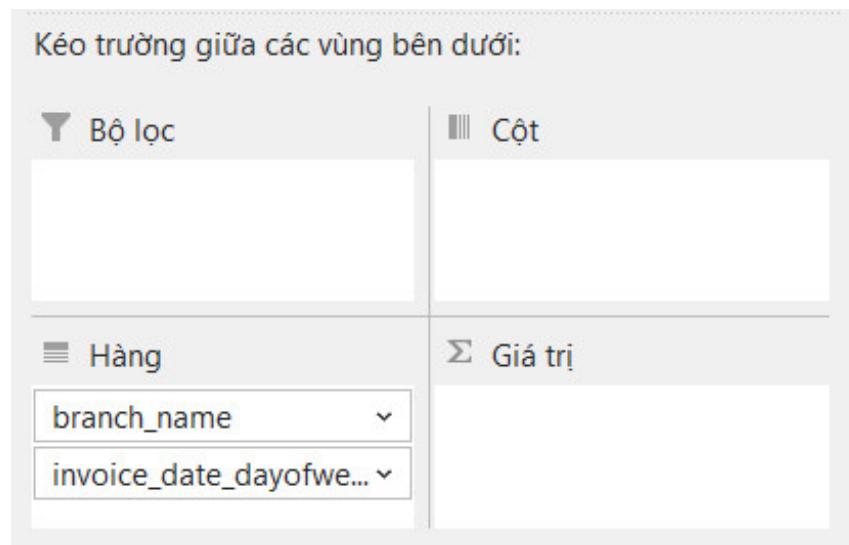
Hình 398: Kết quả câu truy vấn 5 hiển thị dưới dạng biểu đồ (rút gọn)

### 3.12.3.6. Câu truy vấn 6:

**Nội dung câu truy vấn:** Liệt kê tổng số lượng sản phẩm, số lượng khách hàng tại từng chi nhánh theo ngày trong tuần.

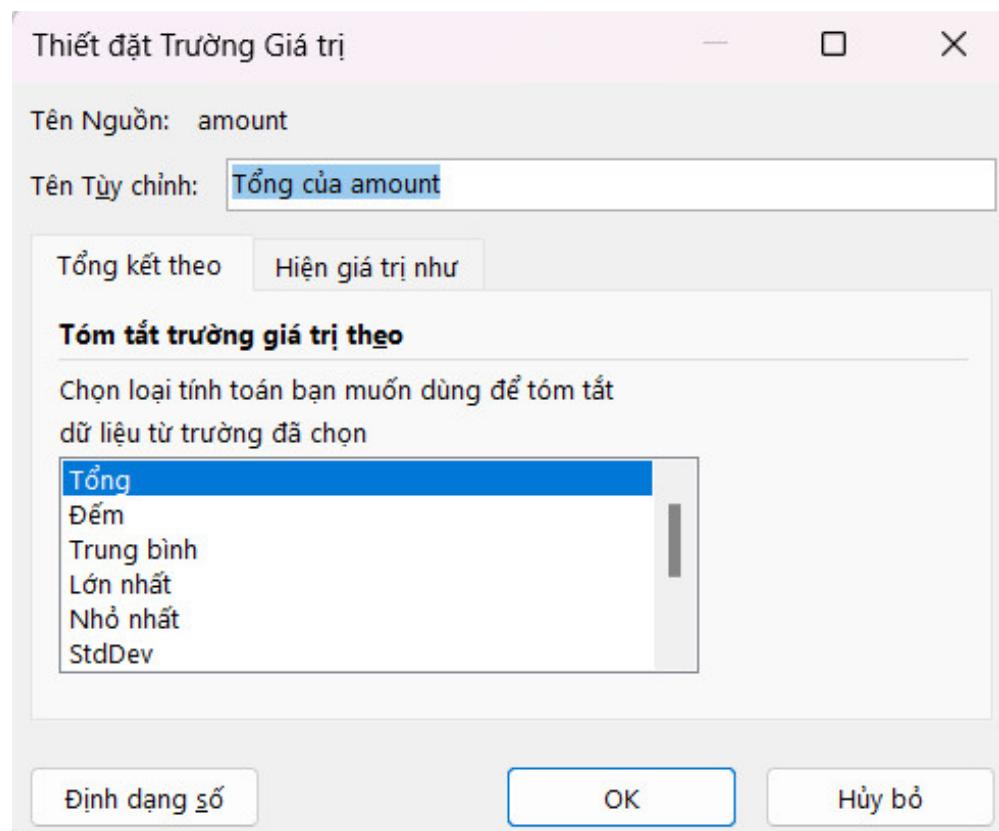
**Thực hiện Report:**

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính branch\_name của bảng Dim\_Branch, invoice\_date\_dayofweek của bảng Dim\_Time vào mục “Hàng”.



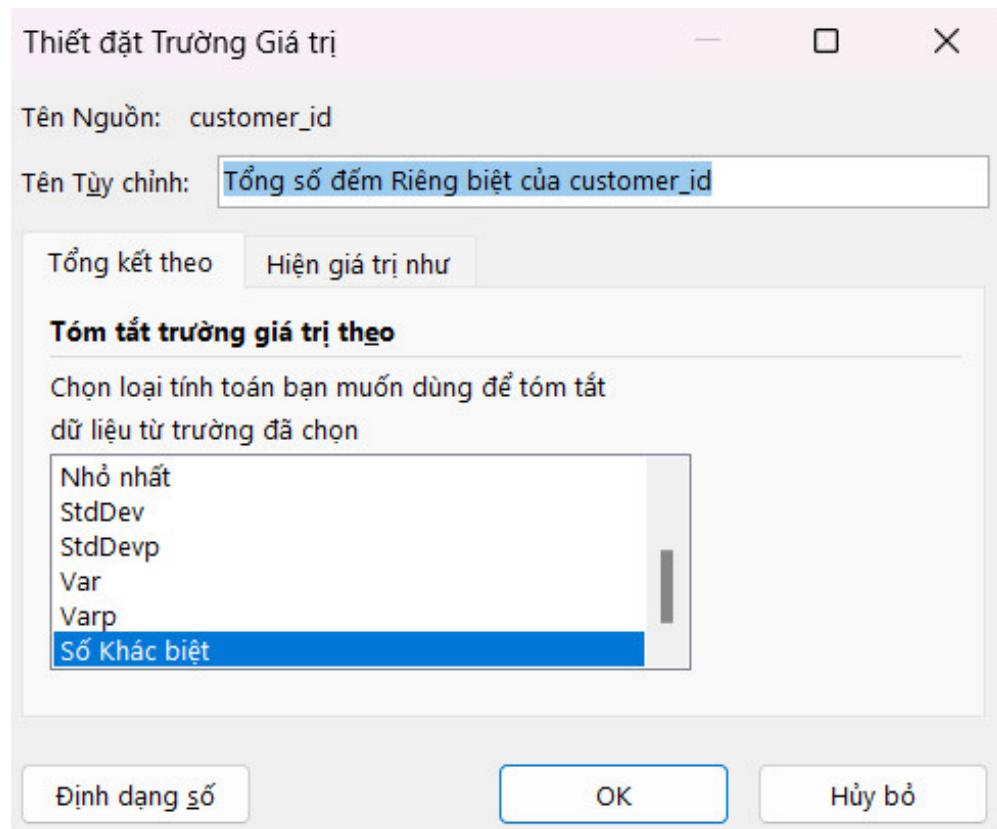
Hình 399: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính amount của bảng Fact vào mục “Giá trị”, chọn “Tổng” cho “Tóm tắt trường giá trị theo” của thuộc tính này.



Hình 400: Kéo thả thuộc tính amount vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Kéo thả thuộc tính customer\_id của bảng Fact vào mục “Giá trị”, chọn “Thiết đặt giá trị trường” để chọn “Số khác biệt” cho “Tóm tắt trường giá trị theo” của thuộc tính này.



Hình 401: Kéo thả thuộc tính customer\_id vào trường “Giá trị” và chọn hàm tính toán

**Bước 4:** Theo dõi kết quả report dưới dạng table.

Nhân Hàng		Tổng của amount	Tổng số đếm Riêng biệt của customer_id
<b>Adana Subesi</b>			
1		1491,431	406
2		1707,256	430
3		2127,501	465
4		1468,553	342
5		1765,914	424
6		1766,514	423
7		1704,452	372
<b>Adiyaman Subesi</b>			
1		391,801	78
2		409,166	107
3		565,197	128
4		462,613	84
5		543,954	112
6		472,814	110
7		513,964	120
<b>Afyonkarahisar Subesi</b>			
1		445,518	125
2		503,027	155

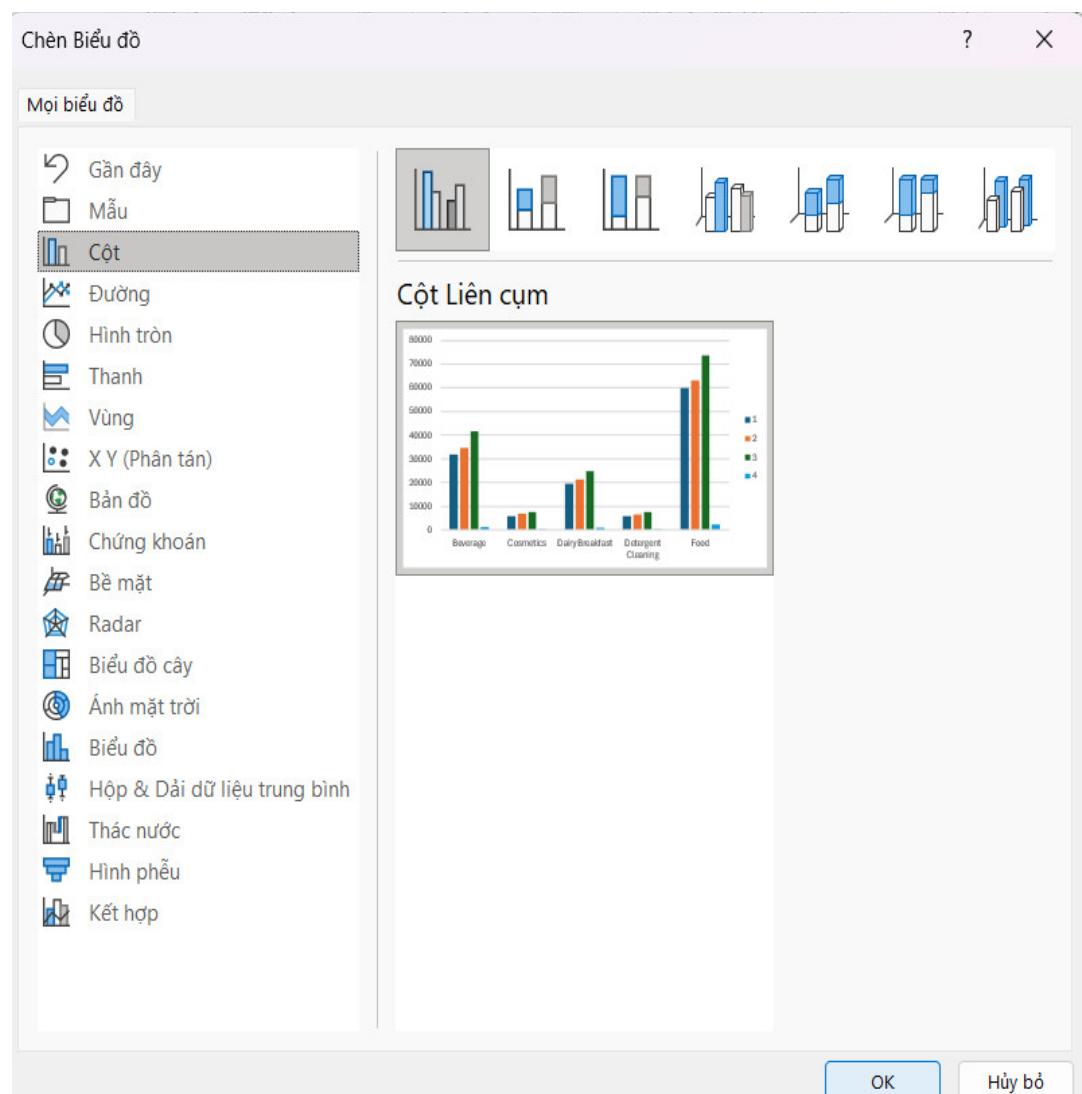
Hình 402: Kết quả câu truy vấn 6 dưới dạng table

**Bước 5:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

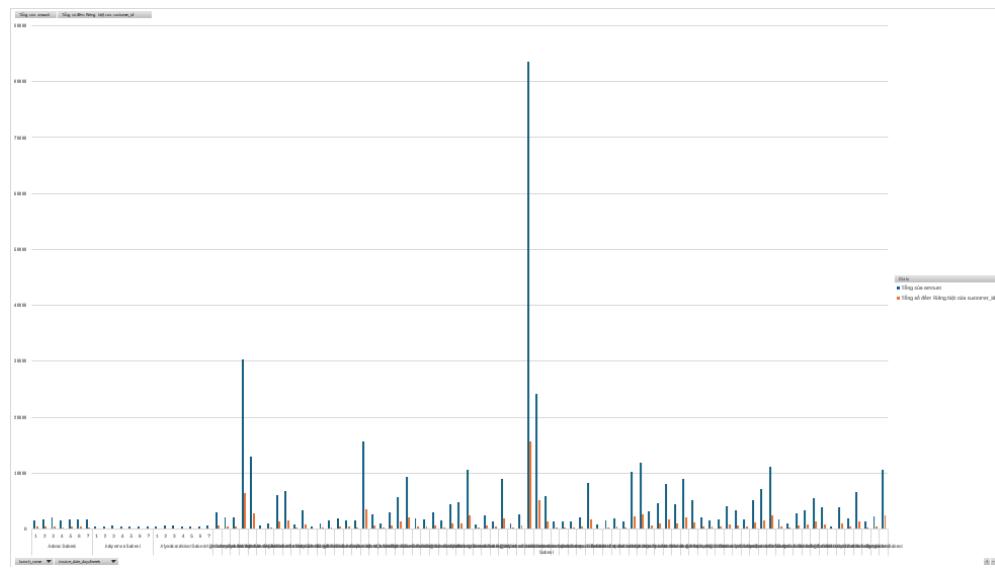


Hình 403: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 404: Giao diện chọn loại biểu đồ



Hình 405: Kết quả câu truy vấn 6 hiển thị dưới dạng biểu đồ

### 3.12.3.7. Câu truy vấn 7:

**Nội dung câu truy vấn:** Thống kê số lượng hóa đơn được thành lập tại 3 thành phố Kırşehir, Nevşehir, Çankırı thuộc vùng İç Anadolu theo từng tháng.

**Thực hiện Report:**

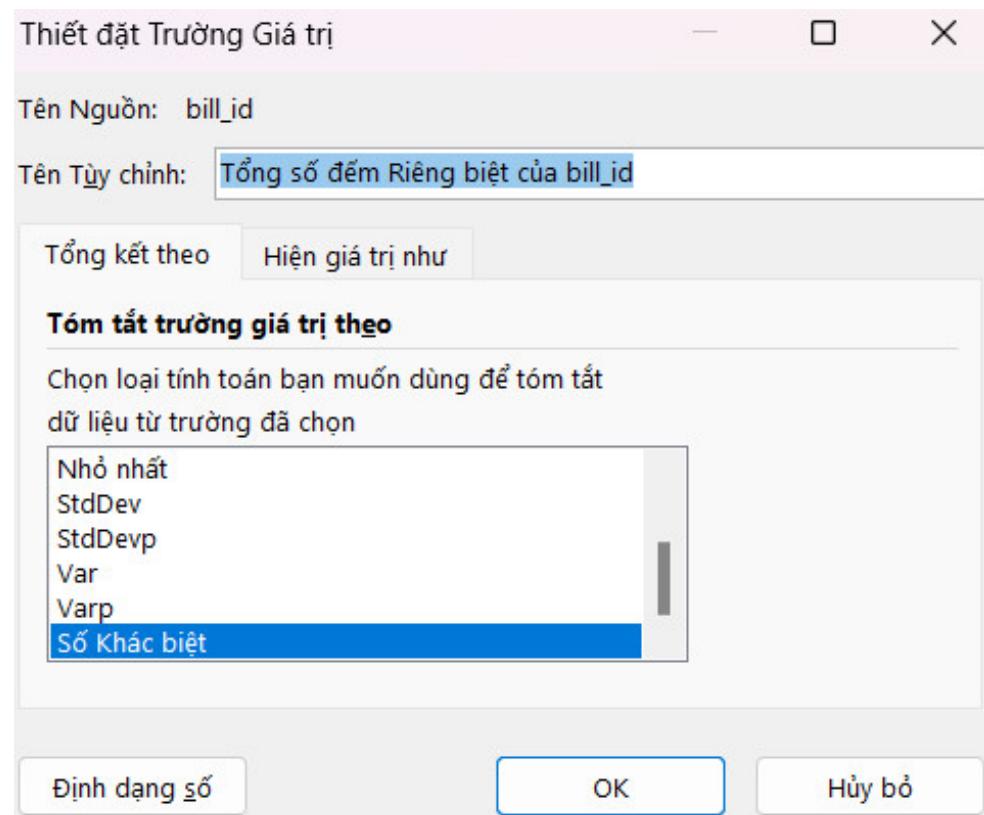
**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính city của bảng Dim\_Branch, invoice\_date\_month của bảng Dim\_Time vào mục “Hàng”.

Kéo trường giữa các vùng bên dưới:

The screenshot shows the 'Fields' pane of a Pivot Table. It has four main sections: 'Bộ lọc' (Filters), 'Cột' (Columns), 'Hàng' (Rows), and 'Giá trị' (Values). In the 'Hàng' section, 'city' and 'invoice\_date\_month' are listed under the 'Rows' dropdown. In the 'Giá trị' section, there is an empty box for entering a value field.

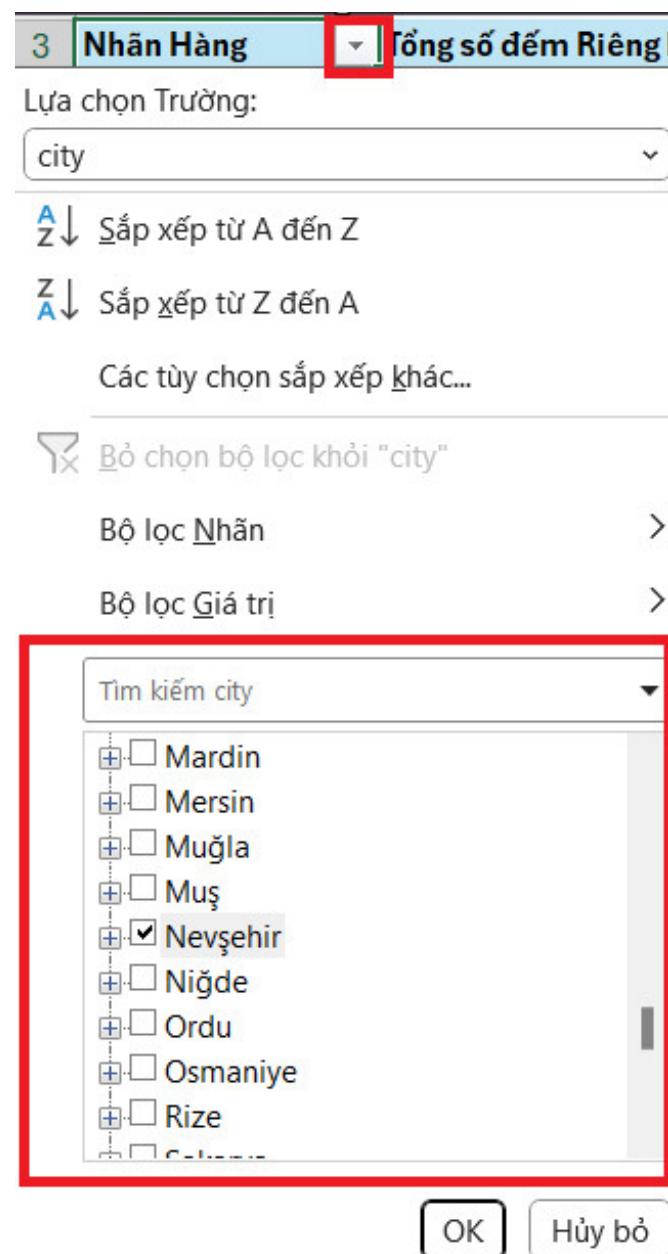
Hình 406: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính bill\_id của bảng Fact vào mục “Giá trị”, chọn “Số khác biệt” cho “Tóm tắt trường giá trị theo” của thuộc tính này.



Hình 407: Kéo thả thuộc tính `customer_id` vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Tại “**Bộ lọc giá trị**” của thuộc tính `city`, tiến hành lọc ra giá trị {  
Kırşehir, Nevşehir, Çankırı}.



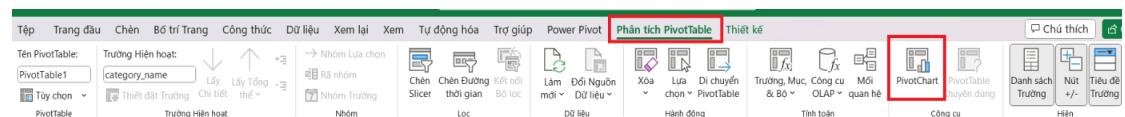
Hình 408: Lọc ra tên các thành phố được đề cập trong câu truy vấn

**Bước 4:** Theo dõi kết quả report dưới dạng table.

Nhân Hàng  Tổng số đếm Riêng biệt của bill_id	
 Çankırı	
1	71
2	82
3	100
4	4
 Kırşehir	
1	93
2	98
3	120
4	7
 Nevşehir	
1	118
2	132
3	137
4	2
<b>Tổng Cuối</b>	<b>964</b>

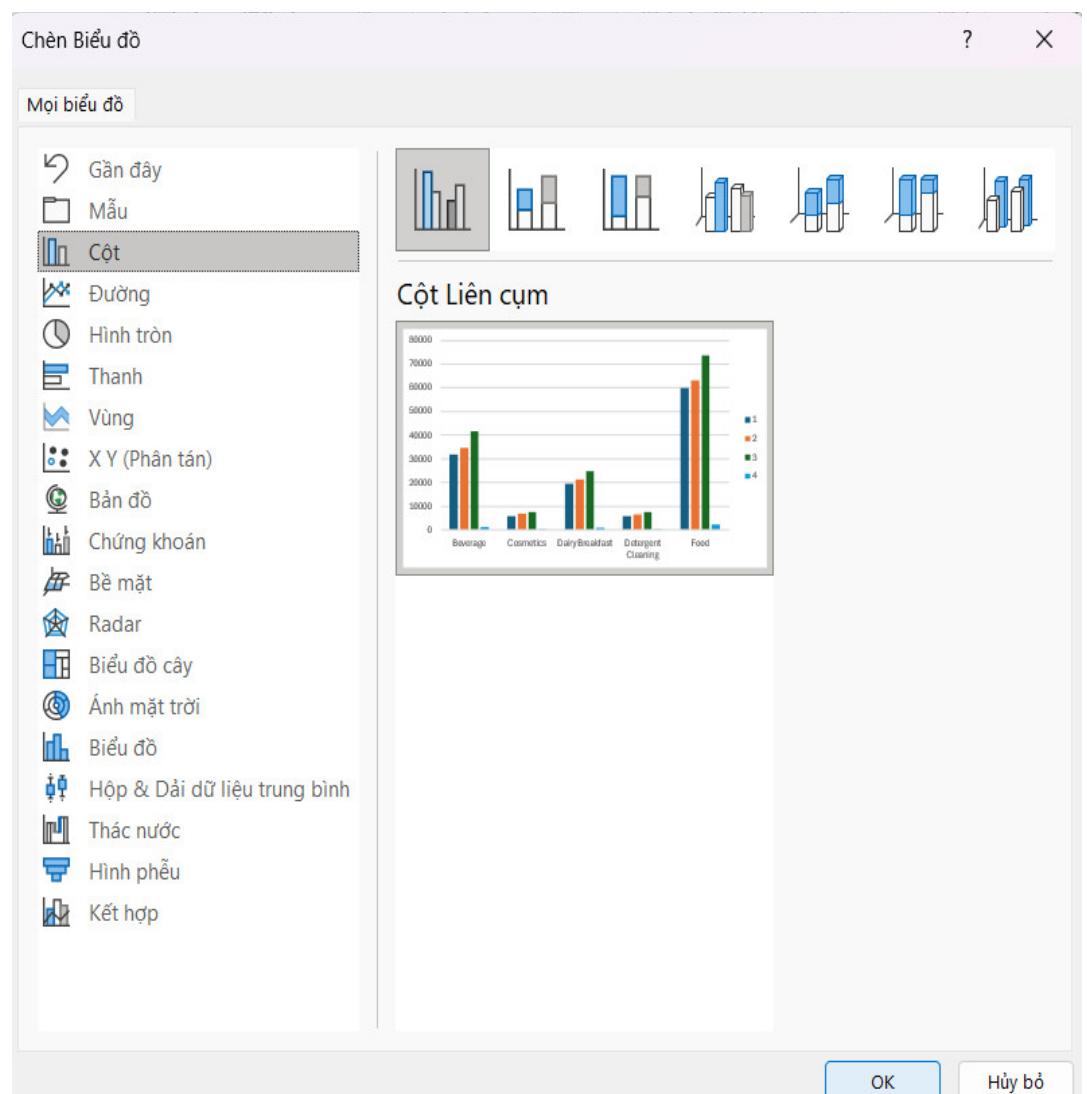
Hình 409: Kết quả câu truy vấn 7 hiển thị dưới dạng table

**Bước 6:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

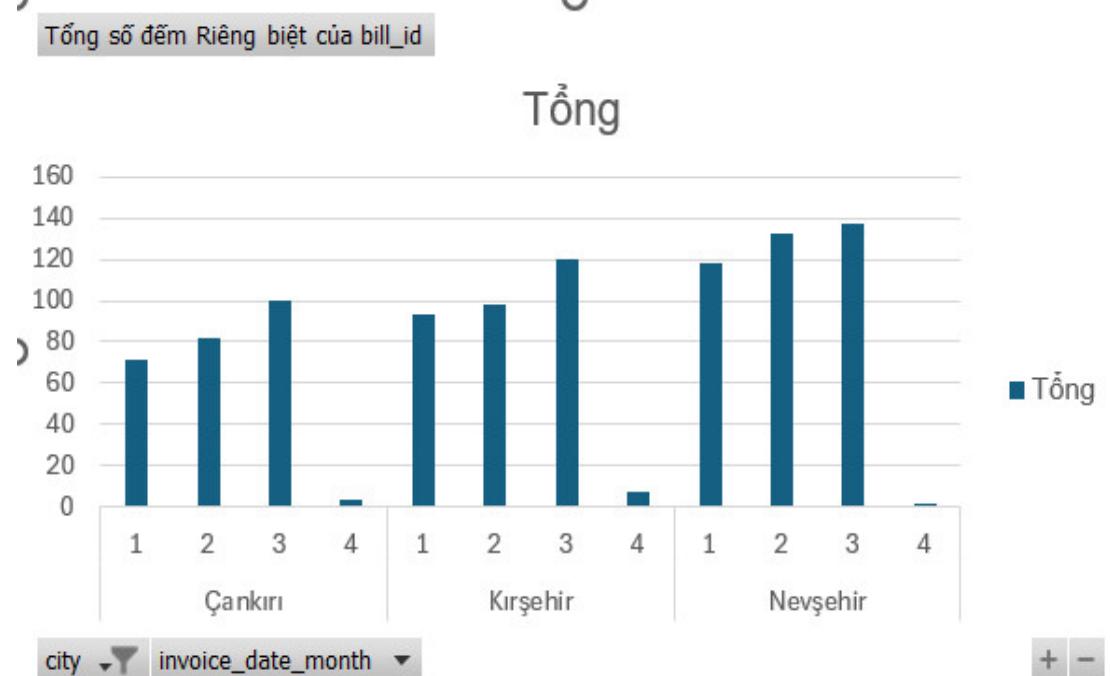


Hình 410: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 411: Giao diện chọn loại biểu đồ



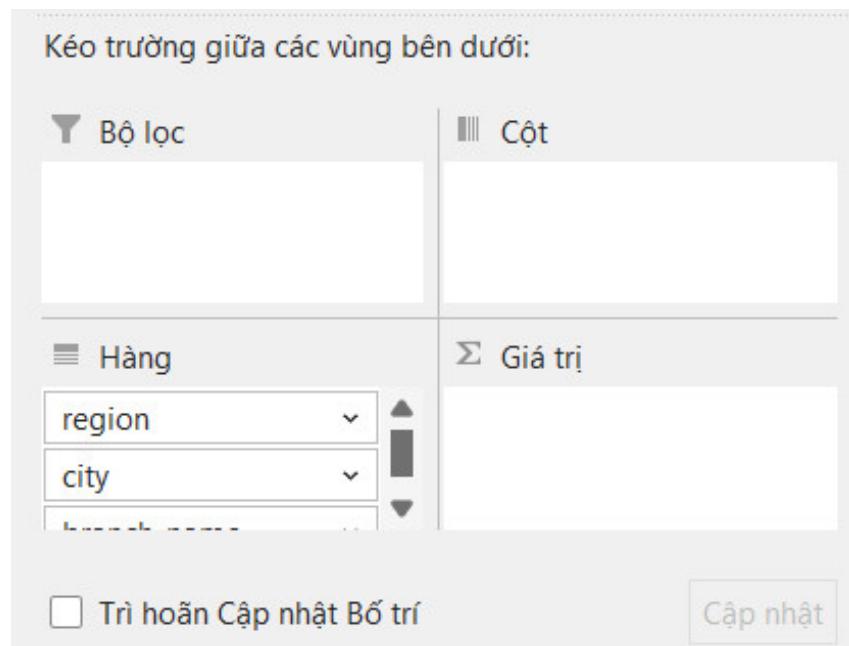
Hình 412: Kết quả câu truy vấn 7 hiển thị dưới dạng biểu đồ

### 3.12.3.8. Câu truy vấn 8:

**Nội dung câu truy vấn:** Cho biết vùng miền có doanh thu cao nhất. Trong vùng miền này, liệt kê tên các chi nhánh có số lượng khách hàng ghé thăm bé hơn 3000.

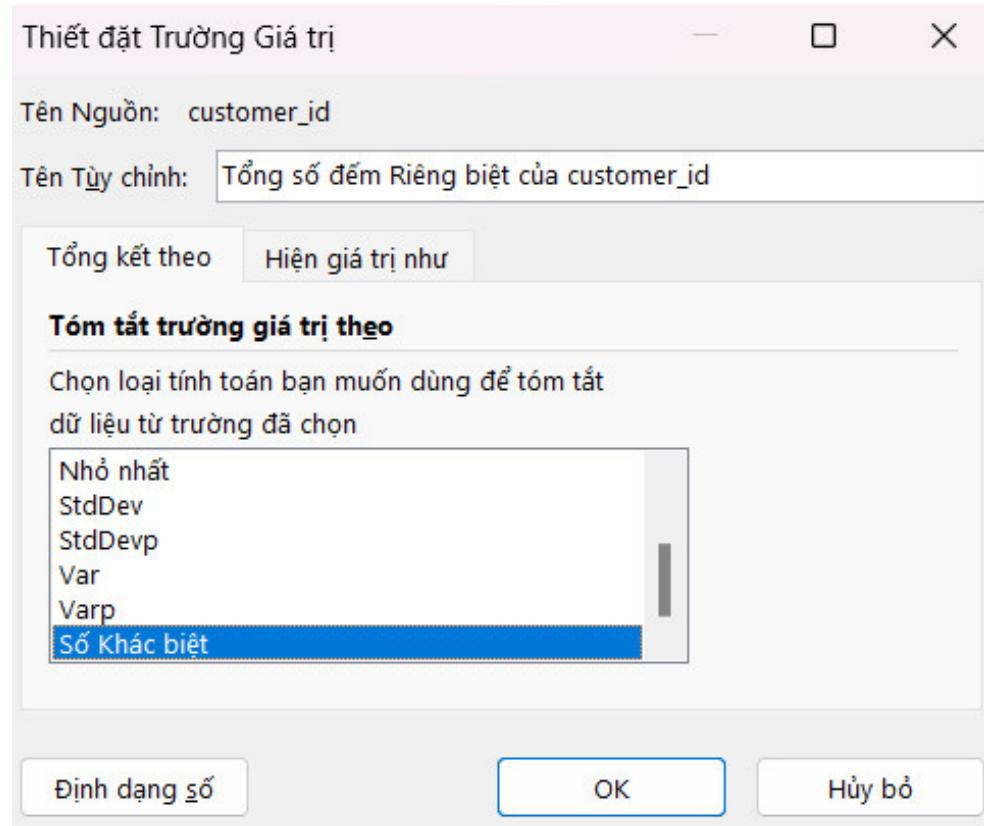
**Thực hiện Report:**

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính region, city, branch\_name của bảng Dim\_Branch vào mục “Hàng”.



Hình 413: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính customer\_id của bảng Fact vào mục “Giá trị”, chọn “Số khác biệt” cho “Tóm tắt trường giá trị theo” của thuộc tính này.



Hình 414: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Tại “**Bộ lọc giá trị**” của thuộc tính region, tiến hành lọc ra vùng miền có doanh thu cao nhất.

1 Nhãn Hàng

Lựa chọn Trường: **Tổng số đơn Riêng biệt của customer\_id**

region

Z↓ Sắp xếp từ A đến Z 278

Z↓ Sắp xếp từ Z đến A 285

Các tùy chọn sắp xếp khác... 36

Bỏ chọn bộ lọc khỏi "region" 196

Bộ lọc Nhãn >

**Bộ lọc Giá trị >**

Tìm kiếm region

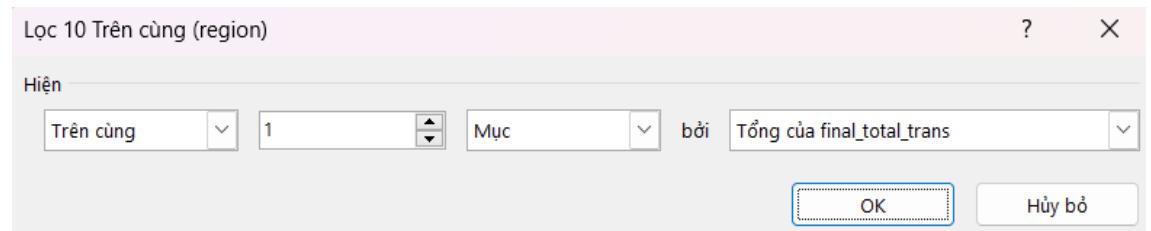
- (Chọn Tất cả)
- Akdeniz
- Doğu Anadolu
- Ege
- Güneydoğu Anadolu
- İç Anadolu
- Karadeniz
- Marmara

OK Hủy bỏ

< > Cau1 Cau2 Cau3

10 Trên cùng...

Hình 415: Lọc ra top 10 vùng miền có doanh thu cao nhất



Hình 416: Lấy ra top 1 vùng miền có doanh thu cao nhất

**Bước 4:** Tại “**Bộ lọc giá trị**” của thuộc tính branch\_name, tiến hành lọc ra chi nhánh có số lượng khác hàng nhỏ hơn 3000.

The screenshot shows a data filtering dialog box. At the top, there is a title bar with the number '1' and the text 'Nhân Hàng'. To the right of the title bar is a dropdown arrow, which is highlighted with a red box. Next to it is the text 'Tổng số đếm Riêng biệt của customer\_id'.

Below the title bar, there is a section labeled 'Lựa chọn Trưởng:' with a dropdown menu set to 'region'. To the right of this dropdown are two numerical values: 278 and 285.

Underneath the dropdown are three options:

- A ↓ Sắp xếp từ A đến Z**
- Z ↓ Sắp xếp từ Z đến A**
- Các tùy chọn sắp xếp khác...**

Below these options is a section labeled 'Bỏ chọn bộ lọc khỏi "region"' with a numerical value of 196.

There are two main sections for filtering:

- Bộ lọc Nhãn**: This section has a dropdown menu set to 'Bộ lọc Giá trị', which is also highlighted with a red box.
- Tìm kiếm region**: This section contains a list of regions with checkboxes:
  - (Chọn Tất cả)
  - Akdeniz
  - Doğu Anadolu
  - Ege
  - Güneydoğu Anadolu
  - İç Anadolu
  - Karadeniz
  - Marmara

At the bottom of the dialog box are two buttons: 'OK' and 'Hủy bỏ' (Cancel).

Below the dialog box, there is a caption: 'Hình 417: Lọc các dòng dữ liệu theo điều kiện nhỏ hơn'

This screenshot shows a 'Lọc Giá trị' (Filter Value) dialog box for the field 'branch\_name'. The title bar says 'Lọc Giá trị (branch\_name)'.

At the top right are three buttons: '?', 'X', and 'OK'.

Below the title bar is a section labeled 'Hiện các mục' (Show items). It contains a dropdown menu set to 'Tổng số đếm Riêng biệt của customer\_id' and a dropdown menu set to 'nhỏ hơn' (less than). To the right of these dropdowns is a text input field containing the value '3000'.

At the bottom right are two buttons: 'OK' and 'Hủy bỏ' (Cancel).

Hình 418: Thiết lập giá trị điều kiện của câu truy vấn

**Bước 5:** Theo dõi kết quả report dưới dạng table.

Nhân Hàng		Tổng số đếm Riêng biệt của customer_id
Marmara		
Bilecik	Bilecik Subesi	273
Çanakkale	Çanakkale Subesi	618
Edirne	Edirne Subesi	486
Kırklareli	Kırklareli Subesi	460
Kocaeli	Kocaeli Subesi	2349
Sakarya	Sakarya Subesi	1233
Tekirdağ	Tekirdağ Subesi	1288
Yalova	Yalova Subesi	310
<b>Tổng Cuối</b>		<b>6665</b>

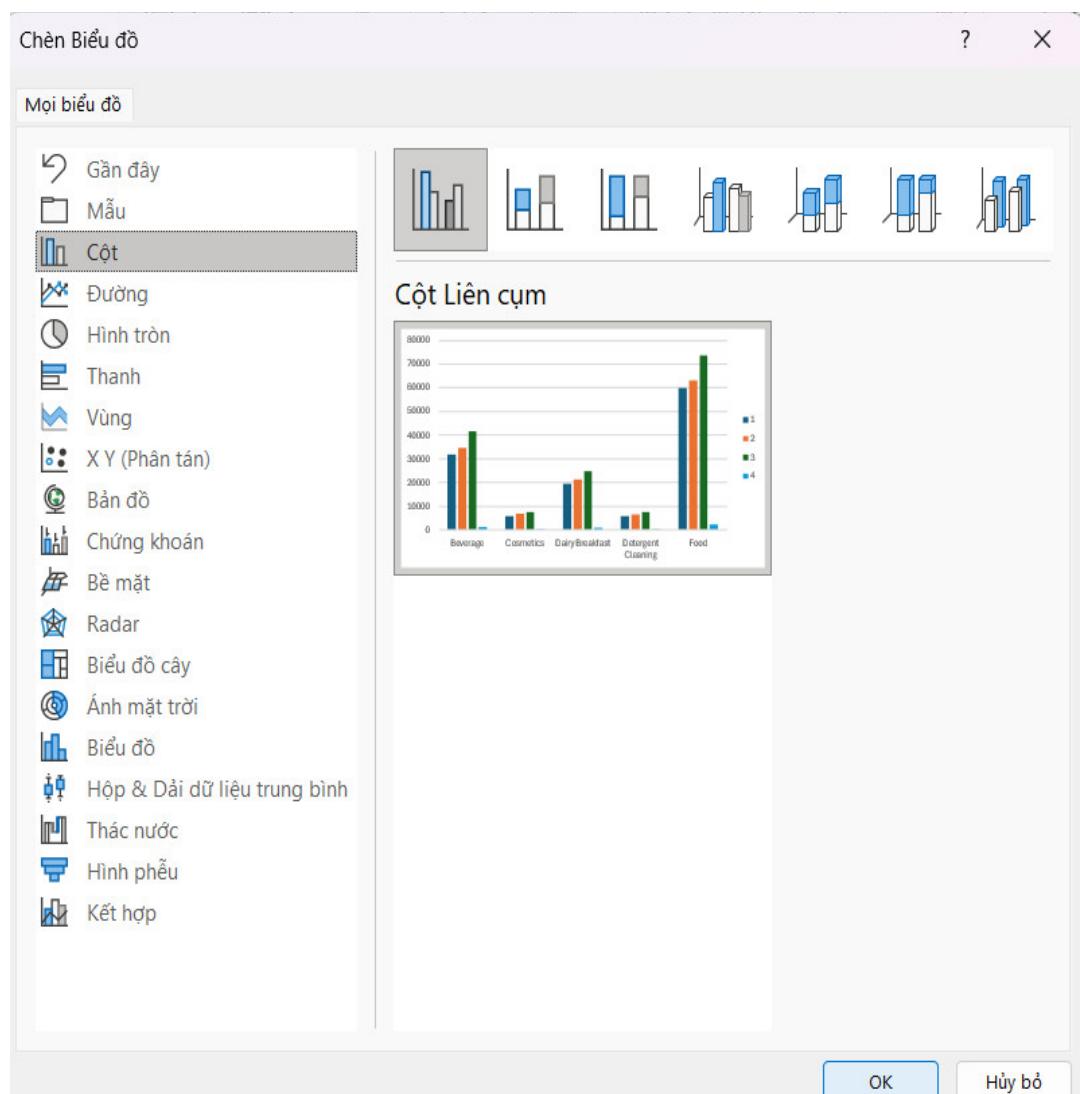
Hình 419: Kết quả câu truy vấn 8 hiển thị dưới dạng table

**Bước 6:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

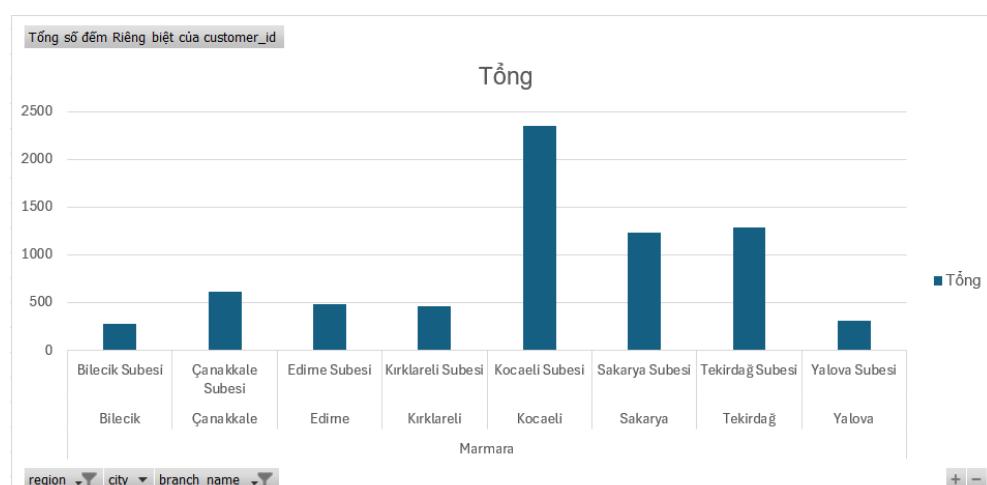


Hình 420: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 421: Giao diện chọn loại biểu đồ



Hình 422: Kết quả câu truy vấn 8 hiển thị dưới dạng biểu đồ

### 3.12.3.9. Câu truy vấn 9:

**Nội dung câu truy vấn:** Thông kê tổng số lượng nhân viên tại từng chi nhánh trong từng vùng miền theo từng ngày.

**Thực hiện Report:**

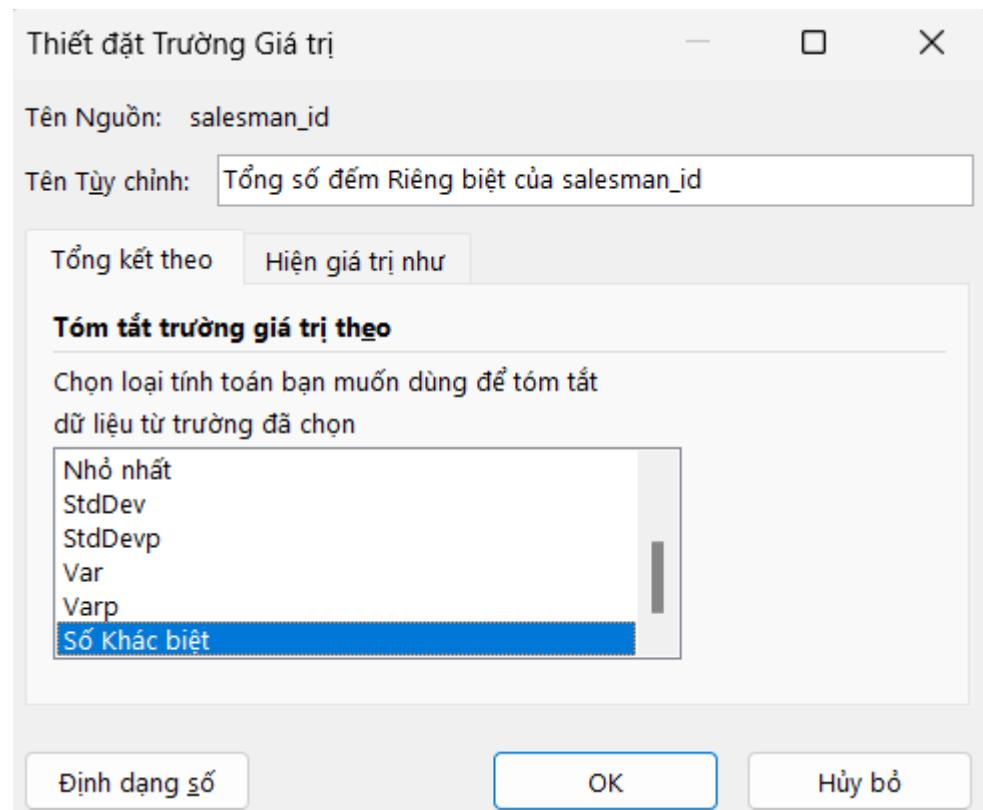
**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính region, city, branch\_name của bảng Dim\_Branch và invoice\_date\_year, invoice\_date\_month, invoice\_date\_day của bảng Dim\_Time vào mục “Hàng”.

Kéo trường giữa các vùng bên dưới:

The screenshot shows the Microsoft Power BI Pivot Table editor. At the top, there are two sections: 'Bộ lọc' (Filters) on the left and 'Cột' (Columns) on the right. Below these, the 'Rows' section is expanded, showing a list of six items under the heading 'Hàng': 'region', 'city', 'branch\_name', 'invoice\_date\_year', 'invoice\_date\_month', and 'invoice\_date\_day'. Each item has a dropdown arrow next to it. To the right of the 'Rows' section, there is a 'Giá trị' (Value) section which is currently empty.

Hình 423: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính salesman\_id của bảng Fact vào mục “Giá trị”, chọn “Số khác biệt” cho “Tóm tắt trường giá trị theo” của thuộc tính này.



Hình 424: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Theo dõi kết quả report dưới dạng table.

Nhân Hàng		Tổng số đếm Riêng biệt của salesman_id
Akdeniz		
Adana		
Adana Subesi		
2017	1	
	3	5
	4	7
	5	5
	6	10
	7	9
	8	7
	9	10
	10	9
	11	10
	12	10
	13	9
	14	10
	15	9
	16	10
	17	10
	18	10
	19	10
	20	10
	21	9
	22	10
	23	10
	24	10

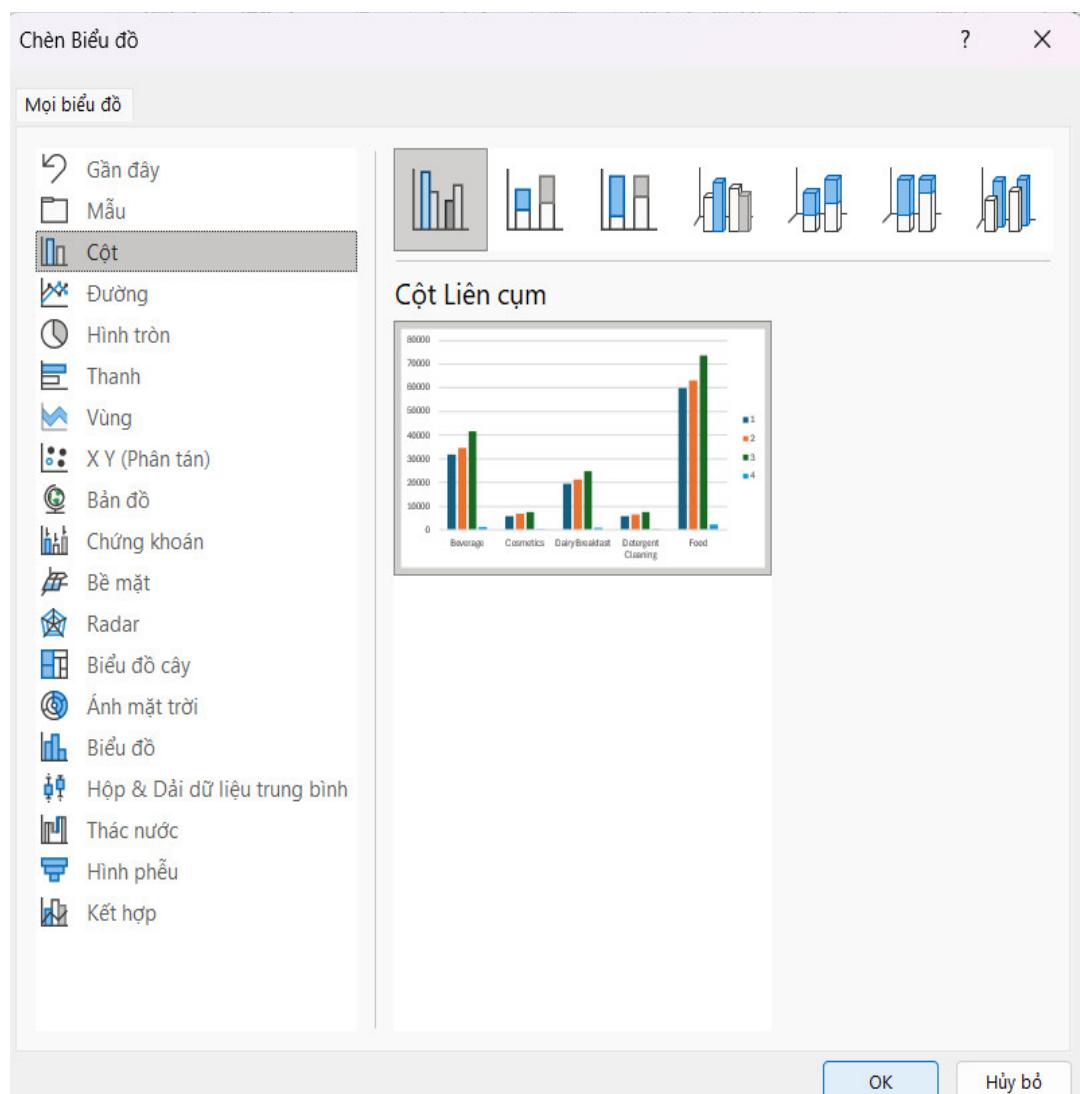
Hình 425: Kết quả câu truy vấn 9 hiển thị dưới dạng table

**Bước 4:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

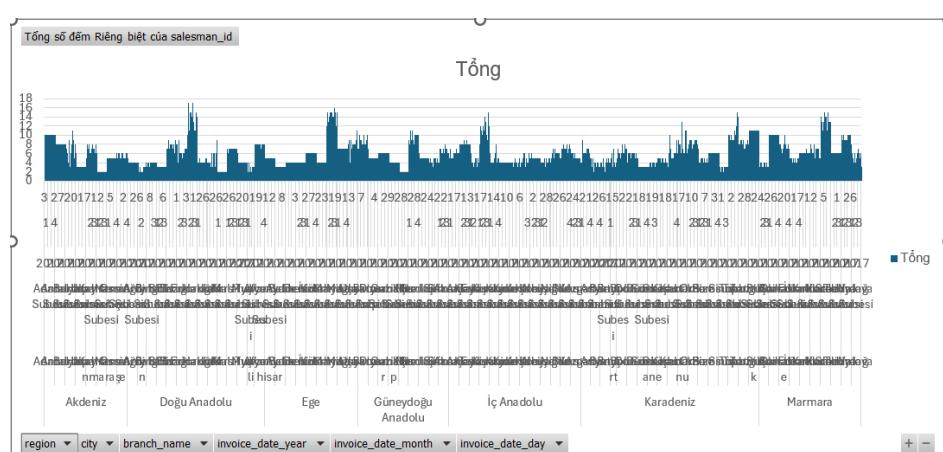


Hình 426: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 427: Giao diện chọn loại biểu đồ



Hình 428: Kết quả câu truy vấn 9 hiển thị dưới dạng biểu đồ

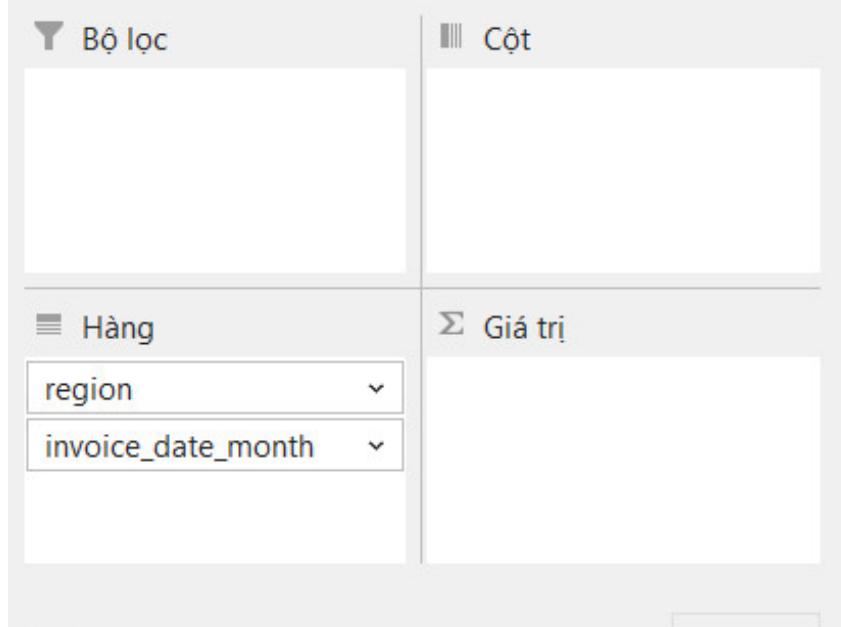
### 3.12.3.10. Câu truy vấn 10:

**Nội dung câu truy vấn:** Cho biết tổng doanh số bán hàng của vùng miền Güneydoğu Anadolu theo từng tháng.

**Thực hiện Report:**

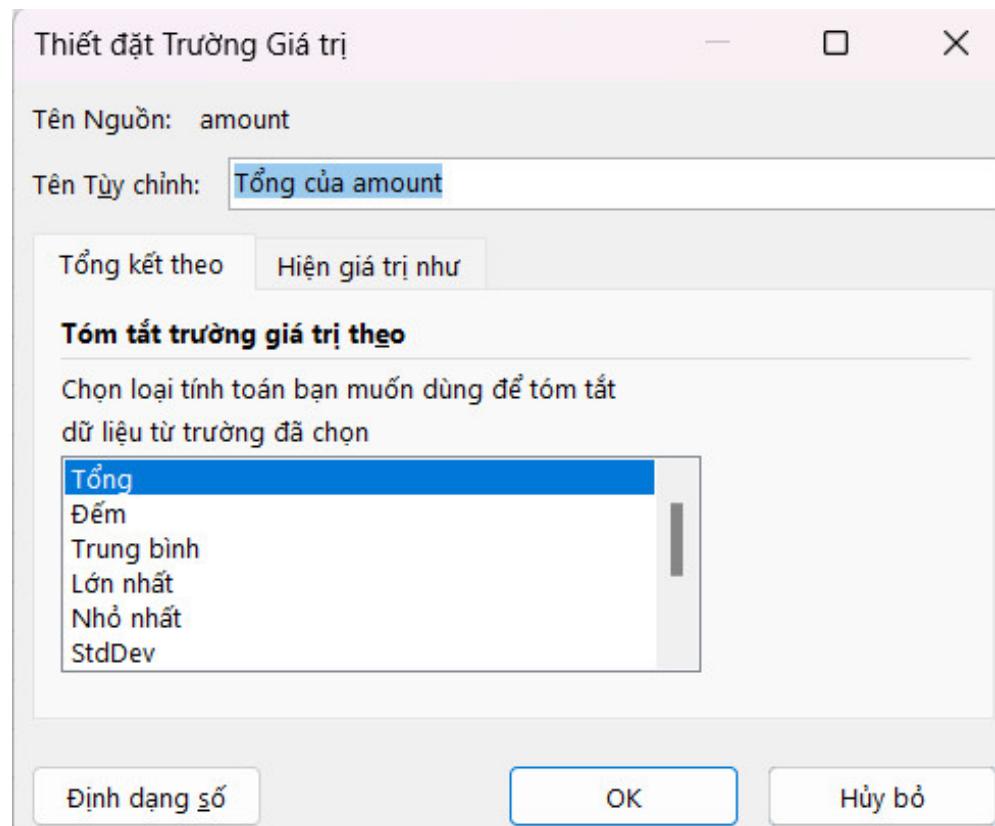
**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính region của bảng Dim\_Branch, invoice\_date\_month của bảng Dim\_Time vào mục “Hàng”.

Kéo trường giữa các vùng bên dưới:



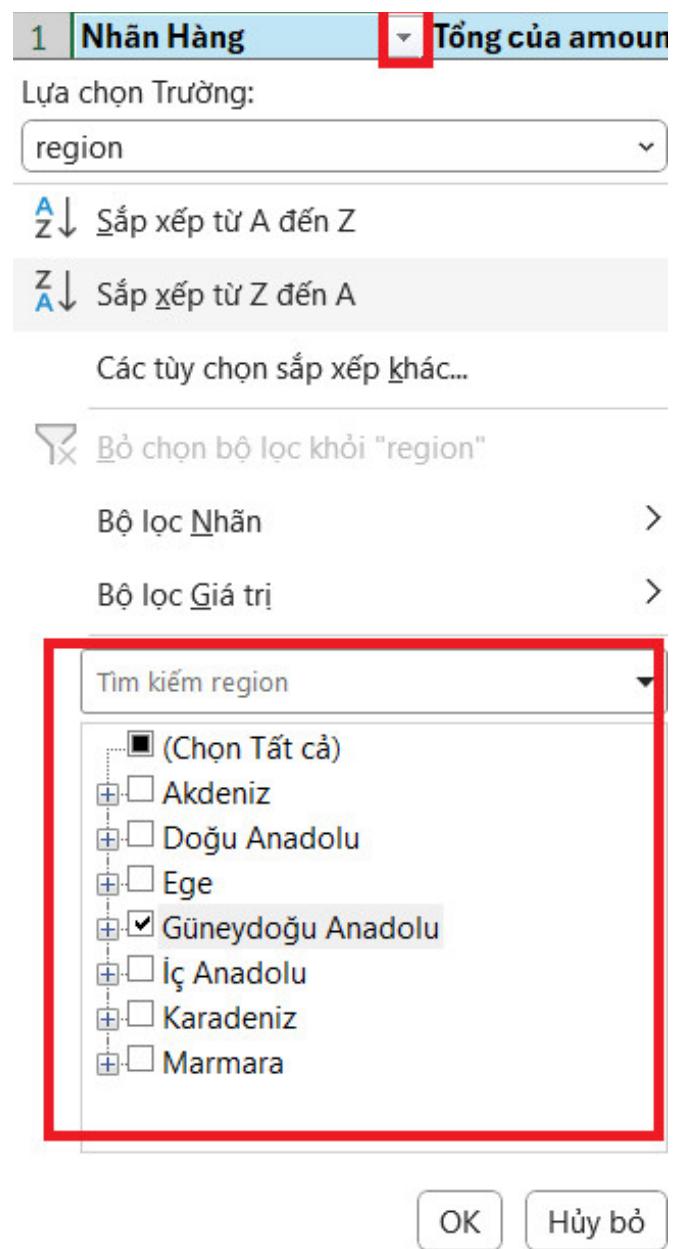
Hình 429: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính amount của bảng Fact vào mục “Giá trị”, chọn “Tổng” cho “Tóm tắt trường giá trị theo” của thuộc tính này.



Hình 430: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Tại “Bộ lọc giá trị” của thuộc tính region, tiến hành lọc ra giá trị Güneydoğu Anadolu.



Hình 431: Lọc dữ liệu theo vùng miền Güneydoğu Anadolu

**Bước 4:** Theo dõi kết quả report dưới dạng table.

Nhân Hàng		Tổng của amount
Güneydoğu Anadolu		
1		14143,65
2		15445,182
3		17356,996
4		583,623
<b>Tổng Cuối</b>		<b>47529,451</b>

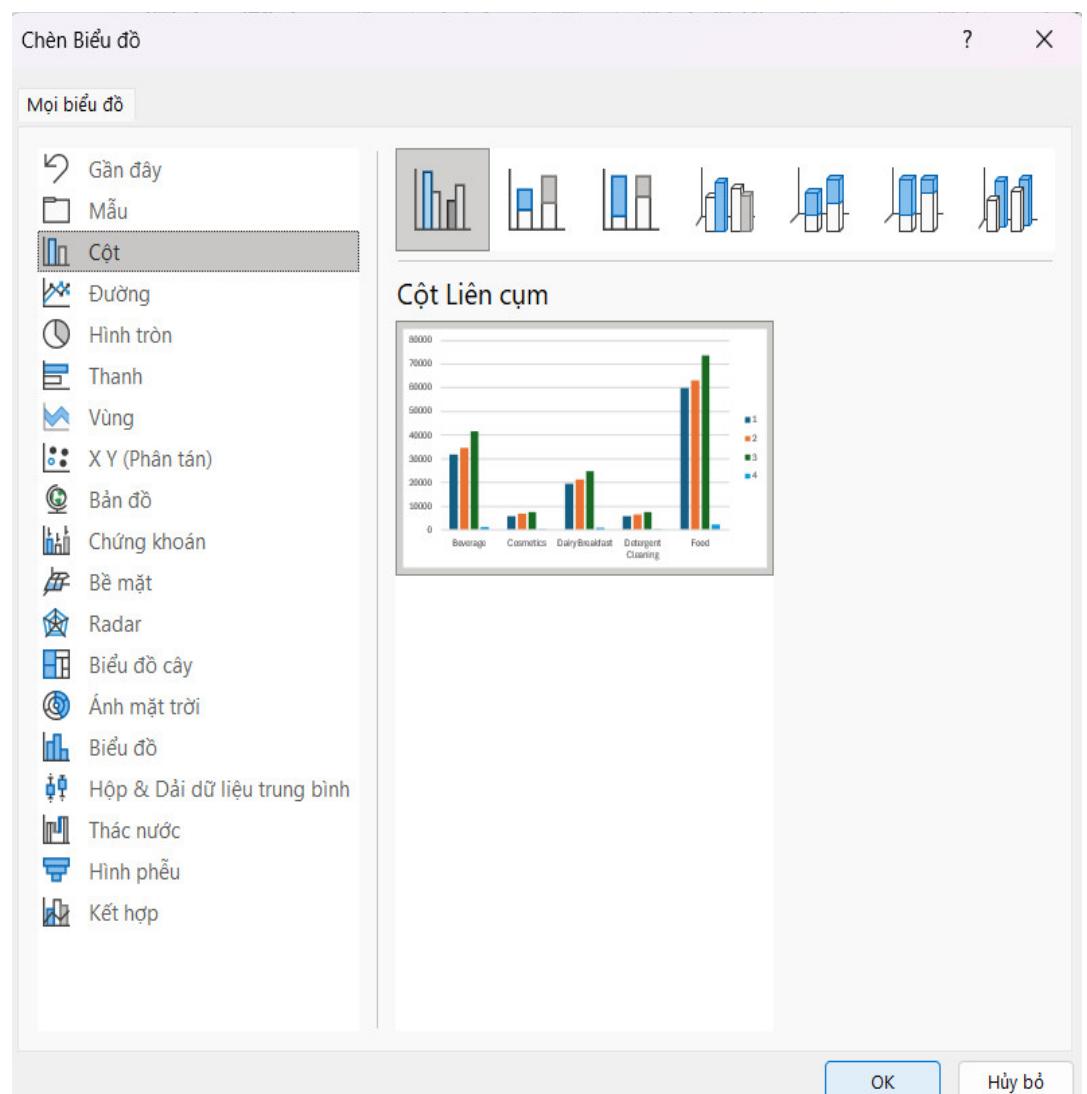
Hình 432: Kết quả câu truy vấn 10 hiển thị dưới dạng table

**Bước 5:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

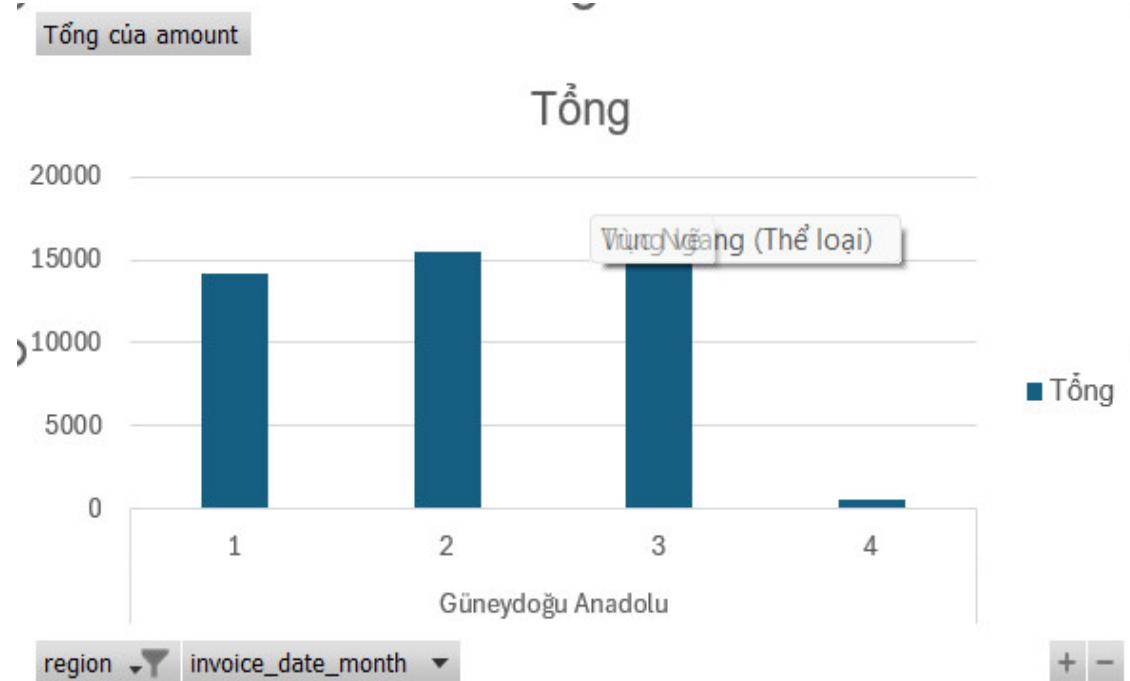


Hình 433: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 434: Giao diện chọn loại biểu đồ



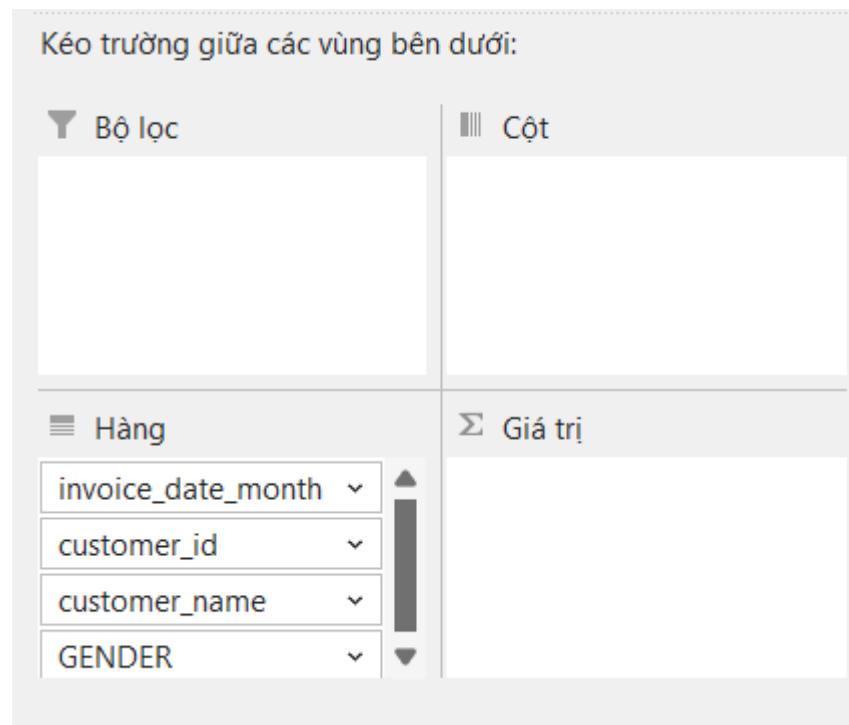
Hình 435: Kết quả câu truy vấn 10 hiển thị dưới dạng biểu đồ

### 3.12.3.11. Câu truy vấn 11:

**Nội dung câu truy vấn:** Cho biết thông tin khách hàng (mã khách hàng, tên khách hàng, giới tính) có tổng số tiền mua lớn nhất trong từng tháng.

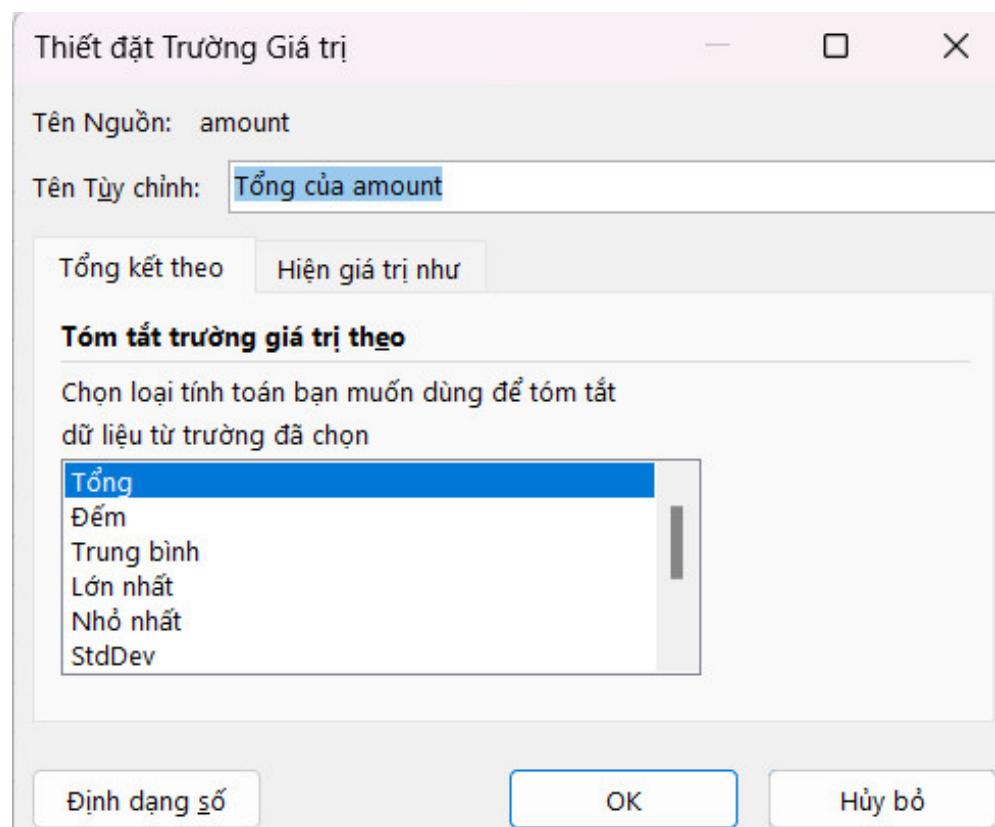
**Thực hiện Report:**

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính invoice\_date\_month của bảng Dim\_Time và customer\_id, customer\_name, GENDER của bảng Dim\_Customer vào mục “Hàng”.



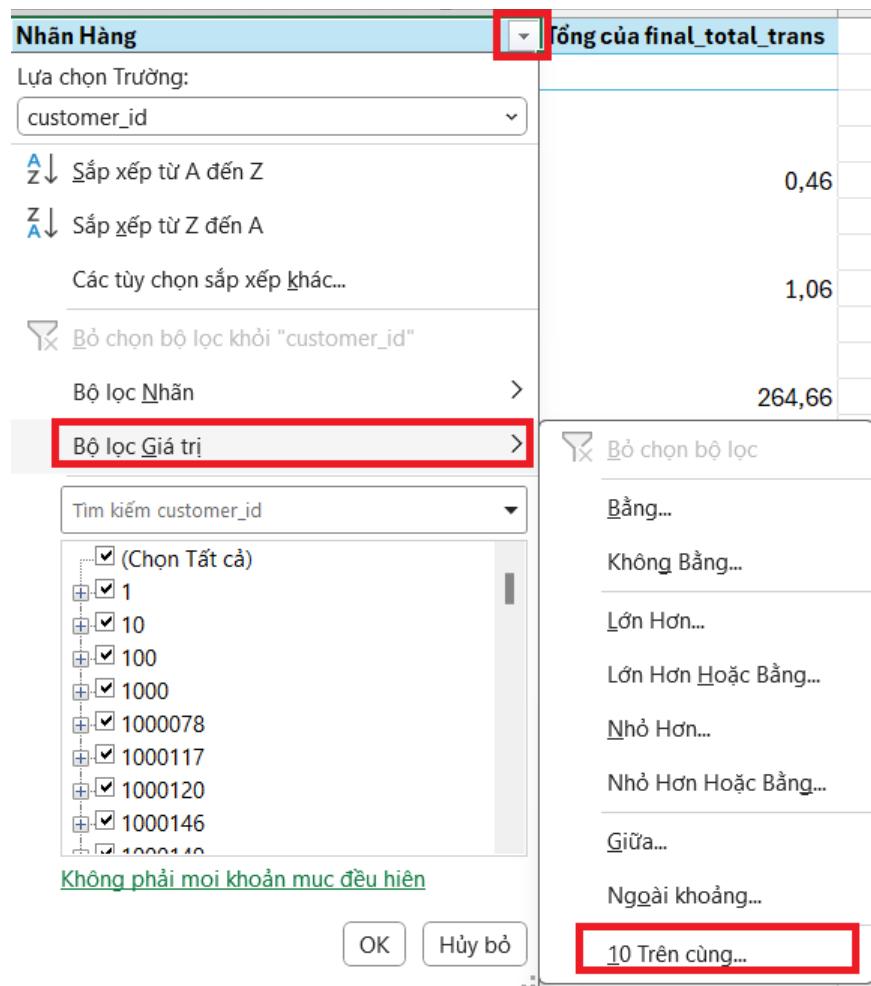
Hình 436: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính amount của bảng Fact vào mục “Giá trị”, chọn “**Tổng**” cho “**Tóm tắt trường giá trị theo**” của thuộc tính này.



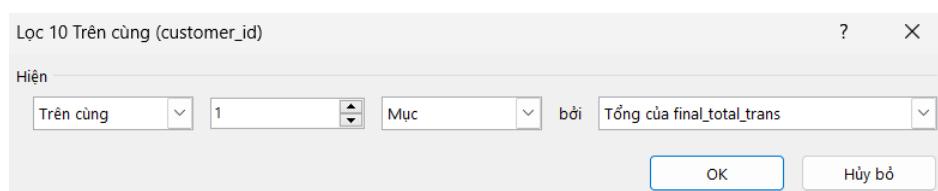
Hình 437: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Tại “**Bộ lọc giá trị**” của thuộc tính customer\_id, tiến hành chọn top 10 khách hàng.



Hình 438: Lọc dữ liệu theo customer\_id

**Bước 4:** Chọn top 1 khách hàng theo final\_total\_trans.



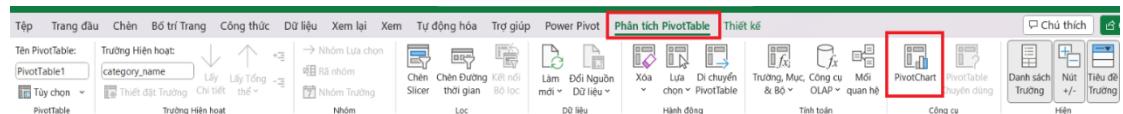
Hình 439: Chọn top 1 khách hàng theo final\_total\_trans

**Bước 5:** Theo dõi kết quả report dưới dạng table.

Nhân Hàng		Tổng của final_total_trans
1	808718	
	Mustafa SEVİLİR	
	Male	1219,66
2	420299	
	Nazmiye AYDIL	
	Female	828,96
3	237685	
	Atakan OLGUN	
	Male	2012,75
4	638391	
	Sevim ÇANAKÇI	
	Female	303,51
<b>Tổng Cuối</b>		<b>4364,88</b>

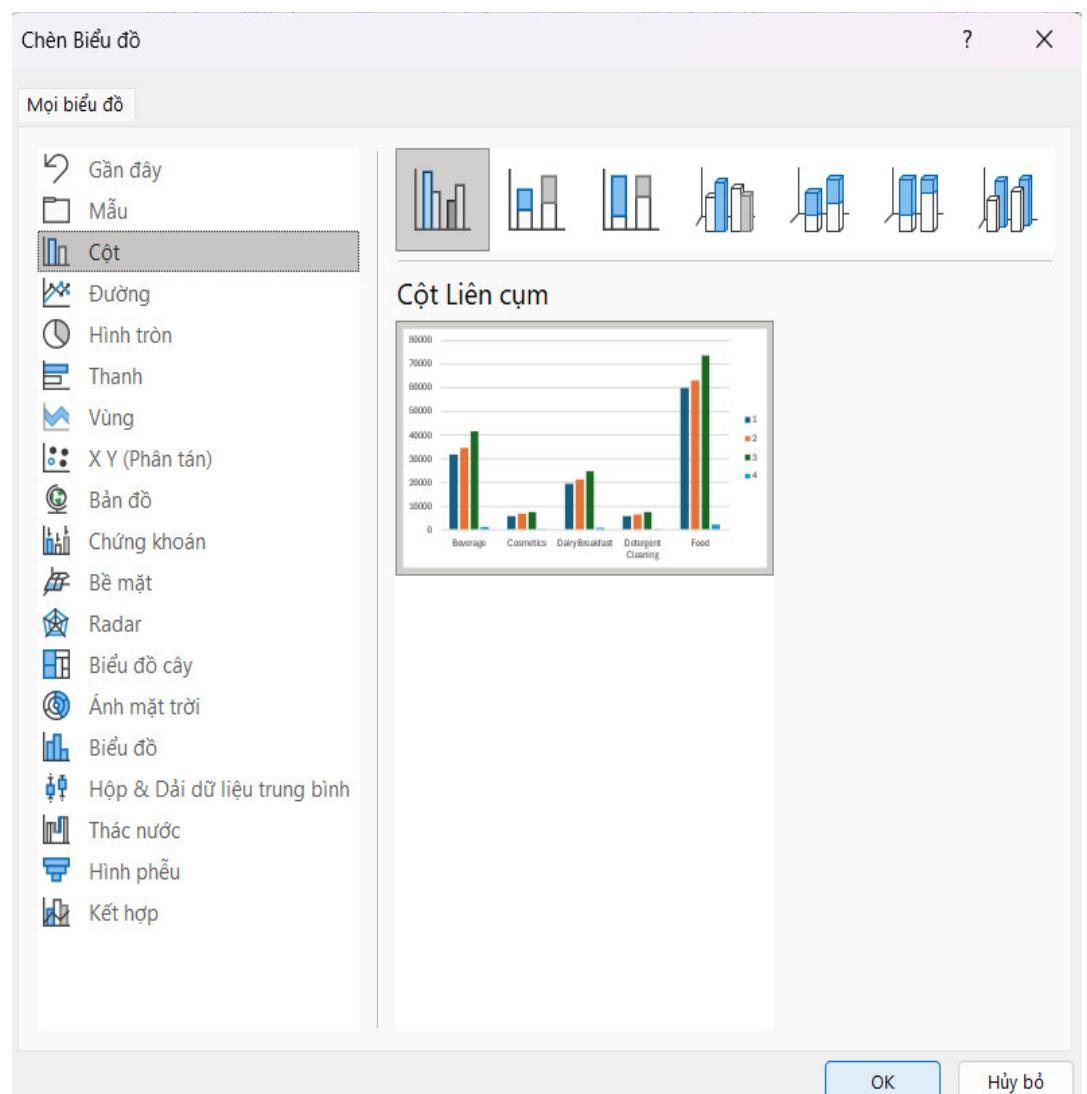
Hình 440: Kết quả câu truy vấn 11 hiển thị dưới dạng table

**Bước 6:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

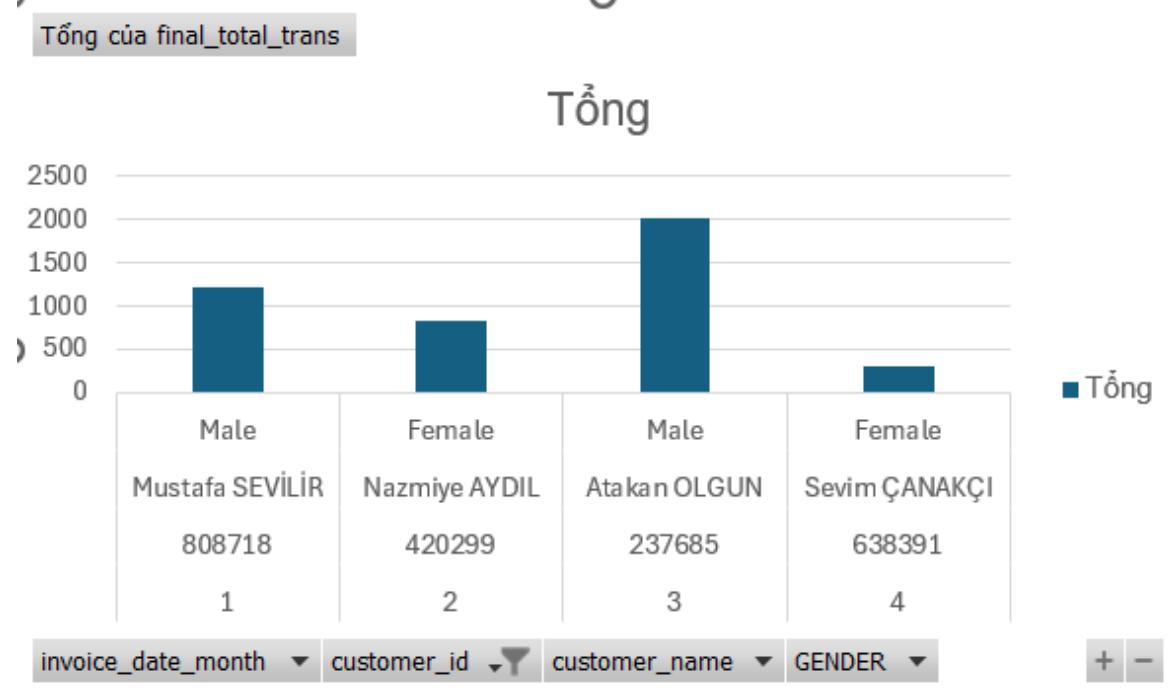


Hình 441: Giao diện hướng dẫn vào nơi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 442: Giao diện chọn loại biểu đồ



Hình 443: Kết quả câu truy vấn 11 hiển thị dưới dạng biểu đồ

### 3.12.3.12. Câu truy vấn 12:

**Nội dung câu truy vấn:** Đưa ra phần trăm giảm giá của từng chi nhánh trong tháng 01 và tháng 04 năm 2017.

**Thực hiện Report:**

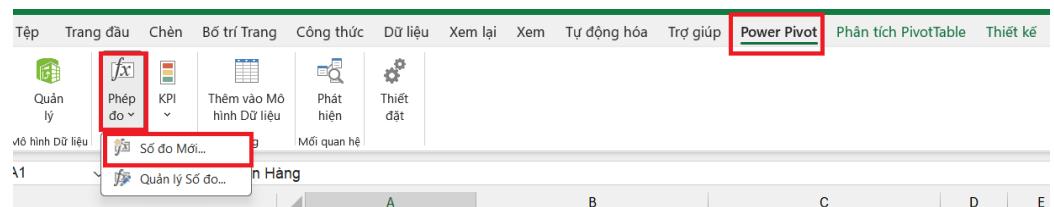
**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính branch\_name của bảng Dim\_Branch, invoice\_date\_month của bảng Dim\_Time vào mục “Hàng”.

Kéo trường giữa các vùng bên dưới:

Bộ lọc	Cột
<b>Hàng</b>	<b>Giá trị</b>
branch_name	
invoice_date_month	

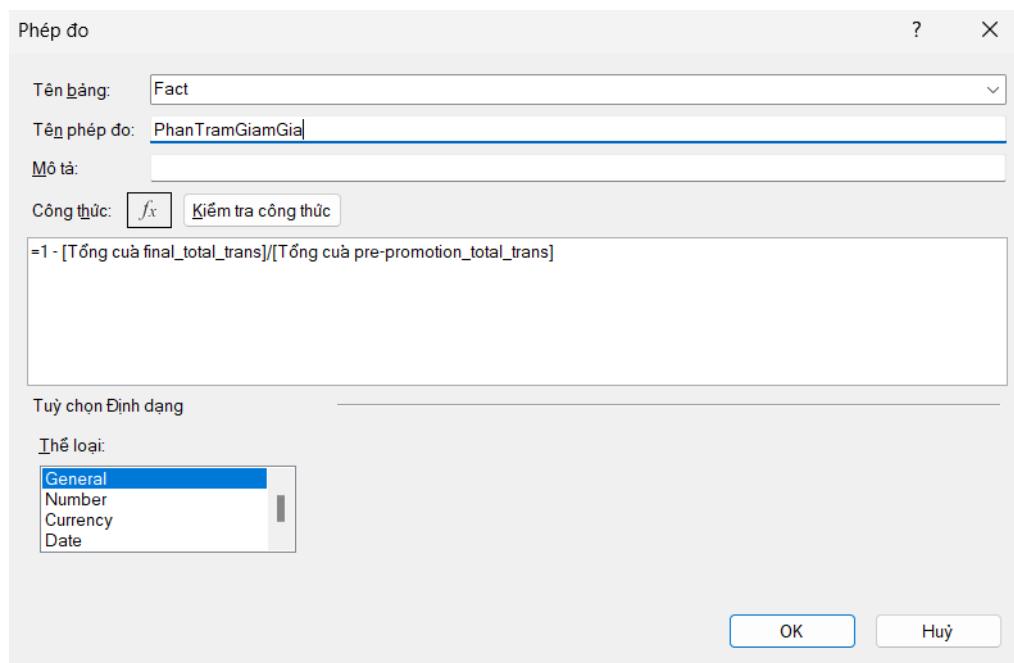
Hình 444: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

### Bước 2: Tại PivotTable, chọn Phép đo và chọn Số đo Mới....



Hình 445: Thực hiện tạo phép đo mới

### Bước 3: Đặt tên phép đo mới và điền công thức tương ứng.



Hình 446: Tạo độ đo Phản Trăm Giảm Giá

**Bước 4:** Kéo độ đo PhanTramGiamGia vừa tạo vào mục “Giá trị”.

Trường PivotTable

Hiện hoạt    Tất cả

Chọn các trường để thêm vào báo cáo:

Tìm kiếm

Invoice\_date\_time

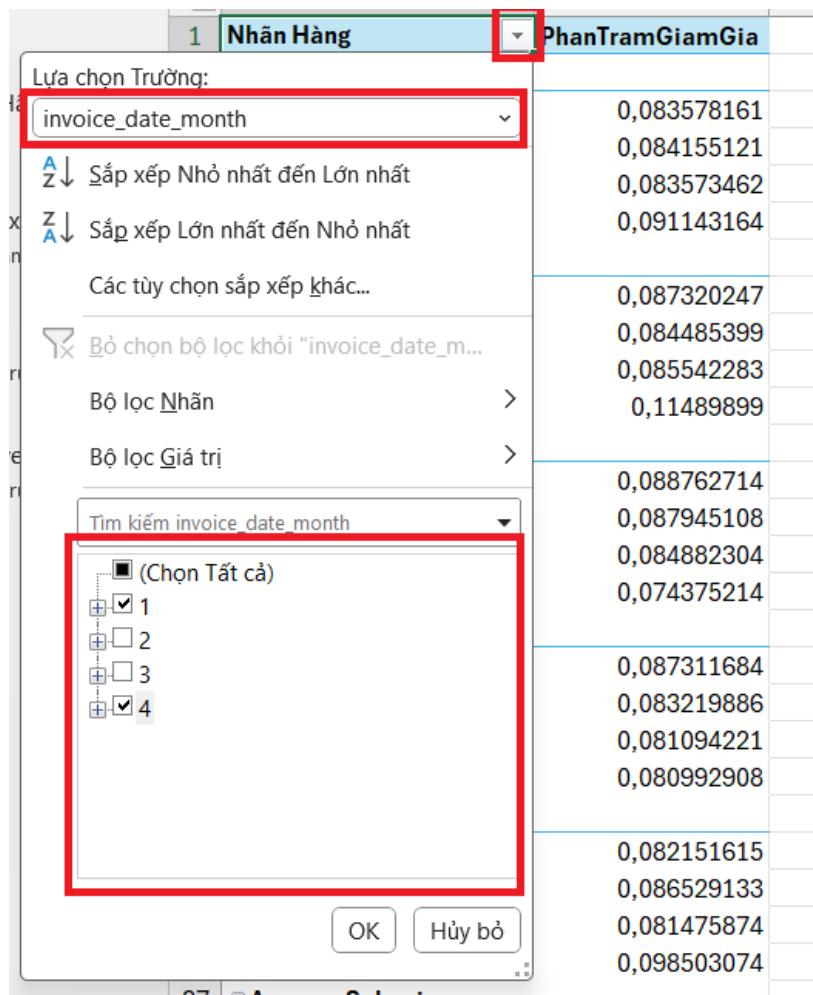
- customer\_id
- salesman\_id
- dim\_product\_id
- fx PhanTramGiamGia

Kéo trường giữa các vùng bên dưới:

Bộ lọc	Cột
☰ Hàng	Σ Giá trị
branch_name	PhanTramGiamGia
invoice_date_month	

Hình 447: Kéo thả thuộc tính vào mục “Giá trị”

**Bước 5:** Tiến hành lọc ra tháng 1 và tháng 4.



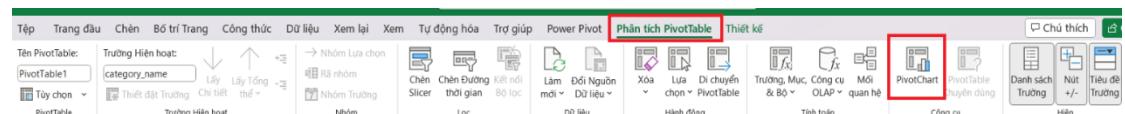
Hình 448: Lọc ra tháng 1 và tháng 4

**Bước 6:** Theo dõi kết quả ở dạng Table.

Nhân Hàng		PhanTramGiamGia
<b>Adana Subesi</b>		
1		0,083578161
4		0,091143164
<b>Adiyaman Subesi</b>		
1		0,087320247
4		0,11489899
<b>Afyonkarahisar Subesi</b>		
1		0,088762714
4		0,074375214
<b>Ağrı Subesi</b>		
1		0,087311684
4		0,080992908
<b>Aksaray Subesi</b>		
1		0,082151615
4		0,098503074
<b>Amasya Subesi</b>		
1		0,077510677
4		0,084640884
<b>Ankara Subesi</b>		
1		0,086690059
4		0,083453655
<b>Antalya Subesi</b>		
1		0,084484129
4		0,073678026
<b>Ardahan Subesi</b>		
1		0,09473151
.		

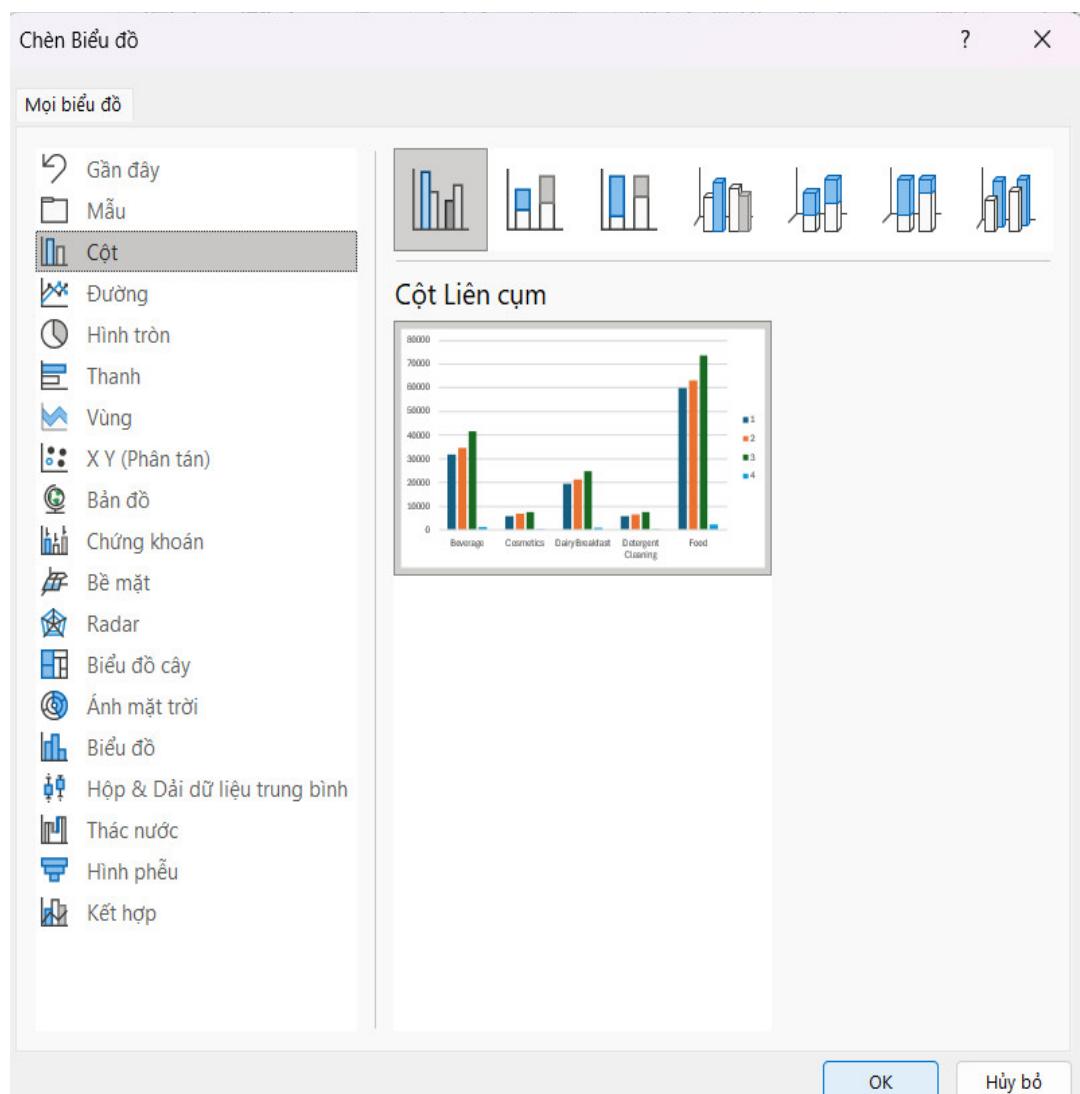
Hình 449: Kết quả câu truy vấn 12 hiển thị dưới dạng table

**Bước 7:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

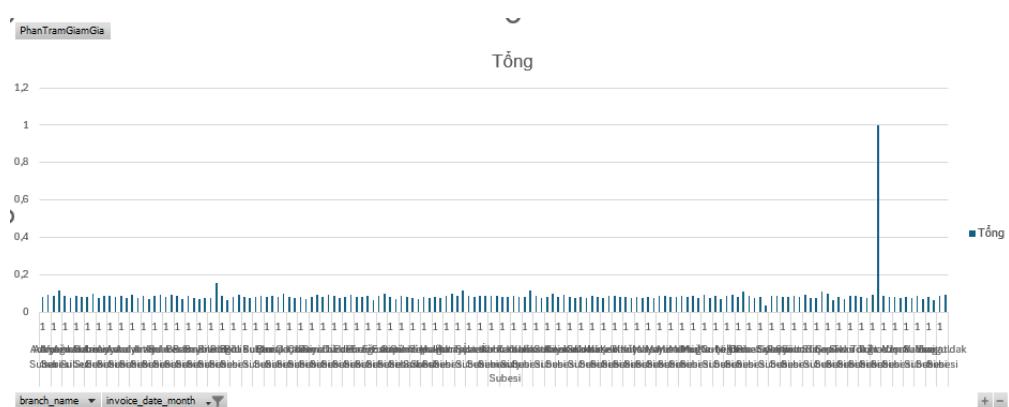


Hình 450: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 451: Giao diện chọn loại biểu đồ



Hình 452: Kết quả câu truy vấn 12 hiển thị dưới dạng biểu đồ

### 3.12.3.13. Câu truy vấn 13:

**Nội dung câu truy vấn:** Cho biết tên sản phẩm, kho sản phẩm có tổng số lượng bán ra lớn hơn 10 tại từng vùng miền.

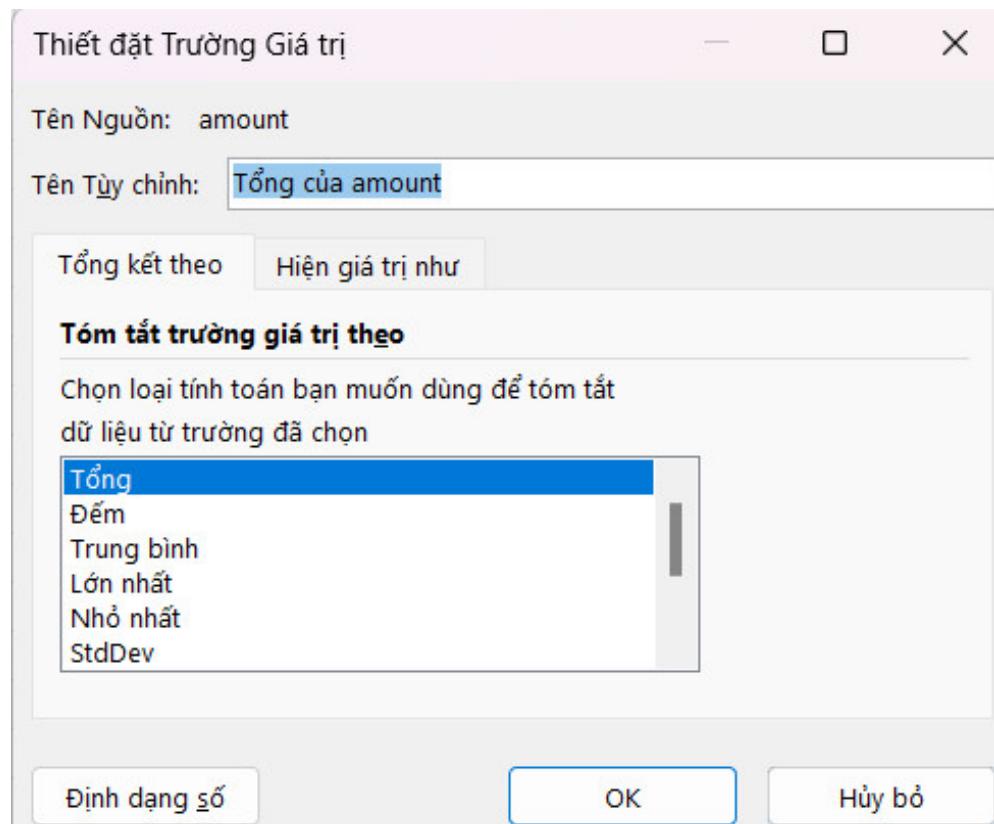
**Thực hiện Report:**

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính region của bảng Dim\_Branch và dim\_product\_id, product\_name của bảng Dim\_Product và category\_name của bảng Dim\_Category vào mục “Hàng”.

The screenshot shows the 'Rows' section of the Pivot Table editor. At the top, there is a note: "Kéo trường giữa các vùng bên dưới:". Below this, there are two columns: 'Bộ lọc' (Filter) and 'Cột' (Column). The 'Rows' section contains four dropdown menus: 'region', 'dim\_product\_id', 'product\_name', and 'category\_name'. To the right of the rows section, there is a column labeled 'Giá trị' (Value) with a summation symbol ( $\Sigma$ ) above it.

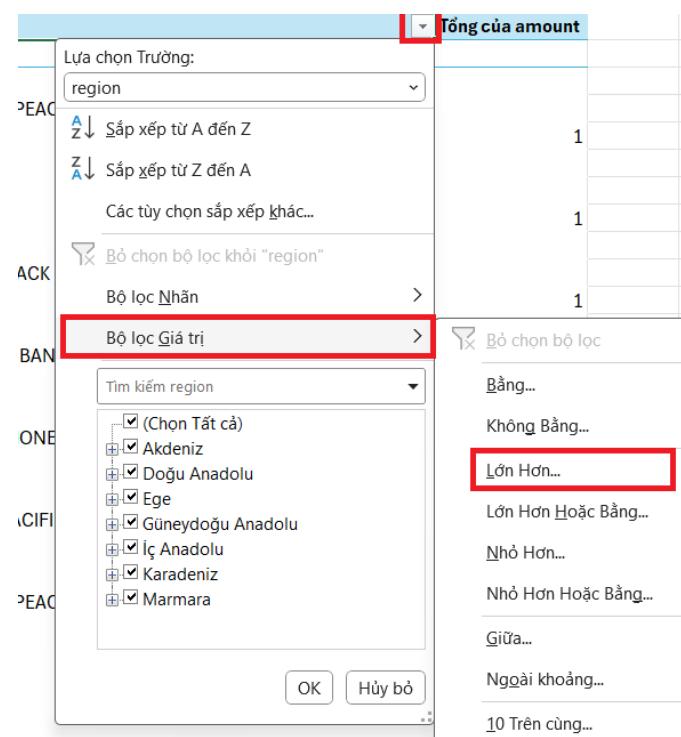
Hình 453: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính amount của bảng Fact vào mục “Giá trị”, chọn “Tổng” cho “Tóm tắt trường giá trị theo” của thuộc tính này.

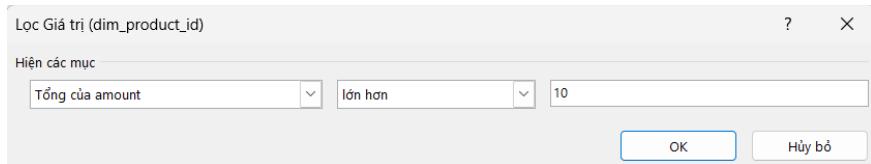


Hình 454: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Tiến hành lọc ra các sản phẩm có số lượng mua lớn hơn 10.



Hình 455: Tiến hành lọc ra sản phẩm có số lượng mua lớn hơn 10



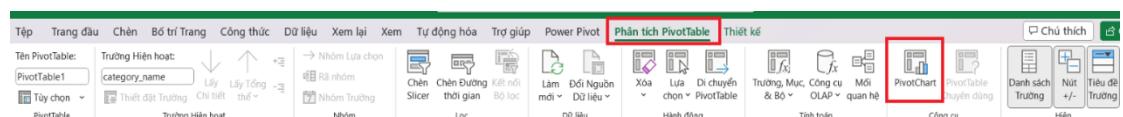
Hình 456: Tiến hành lọc ra sản phẩm có số lượng mua lớn hơn 10

**Bước 4:** Theo dõi kết quả ở dạng Table.

Nhân Hàng		Tổng của amount
✉ Akdeniz		
✉ 2575	CAFÉ CROWN FOAMED MILK COFFEE LATTE SINGLE Beverage	22
✉ 2675	SAKA WATER 1.5 LITER Beverage	34
✉ 2739	MEHMET EFENDI COFFEE 100 GR. Beverage	57
✉ 2767	CAFÉ CROWN HOT CHOCOLATE 23GR Beverage	26
✉ 2797	ICIM AYRAN 200 GR Beverage	25
✉ 2835	MEHMET EFENDI COFFEE 100 GR. Beverage	11
✉ 2861	KAY WATER 5 LT Beverage	12
✉ 2907	MEHMET EFENDI COFFEE 100 GR. Beverage	36
✉ 3001	DUZO NATURAL SPRING WATER 0.5L	

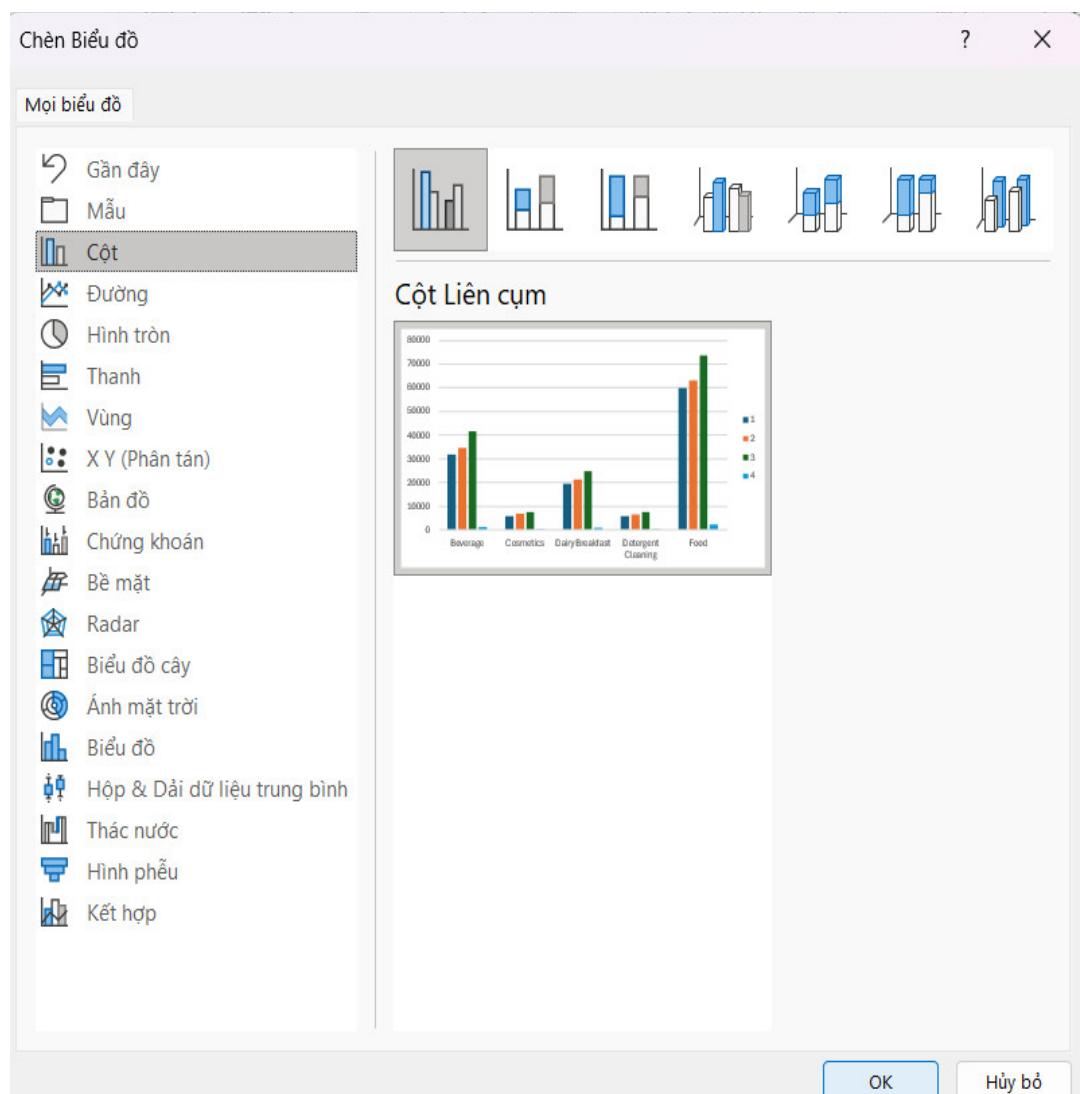
Hình 457: Kết quả câu truy vấn 13 hiển thị dưới dạng table

**Bước 5:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

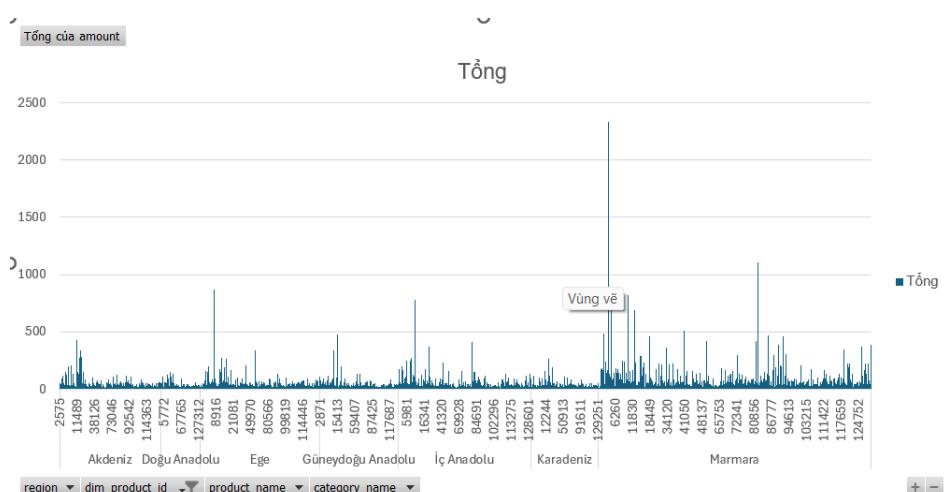


Hình 458: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 459: Giao diện chọn loại biểu đồ



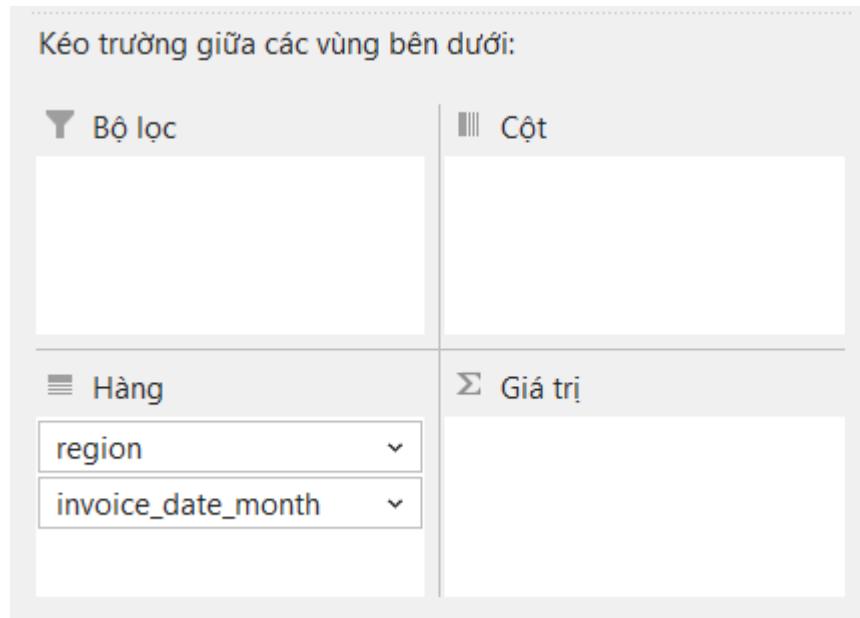
Hình 460: Kết quả câu truy vấn 13 hiển thị dưới dạng biểu đồ

### 3.12.3.14. Câu truy vấn 14:

**Nội dung câu truy vấn:** Với từng vùng miền, liệt kê số lượng đơn hàng bán ra theo từng tháng ngoại trừ hai vùng miền ‘Doğu Anadolu’, ‘Karadeniz’

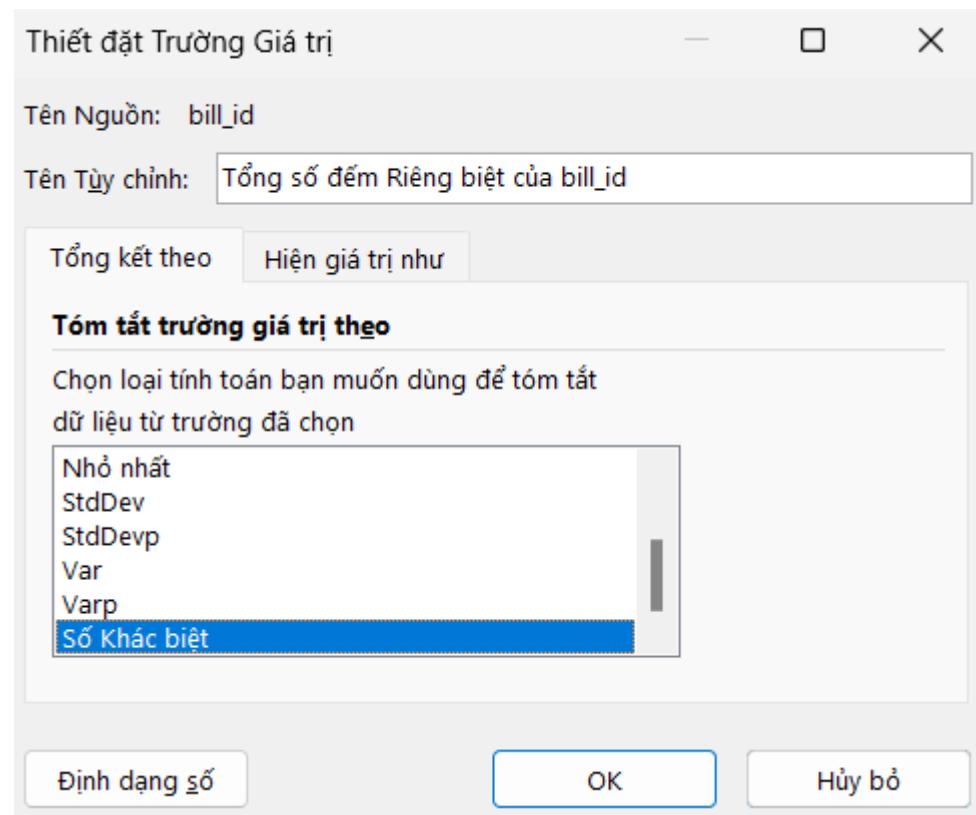
**Thực hiện Report:**

**Bước 1:** Tại tab “Trường Pivot Table”, kéo thả thuộc tính region của bảng Dim\_Branch và invoice\_date\_month của bảng Dim\_Time vào mục “Hàng”.



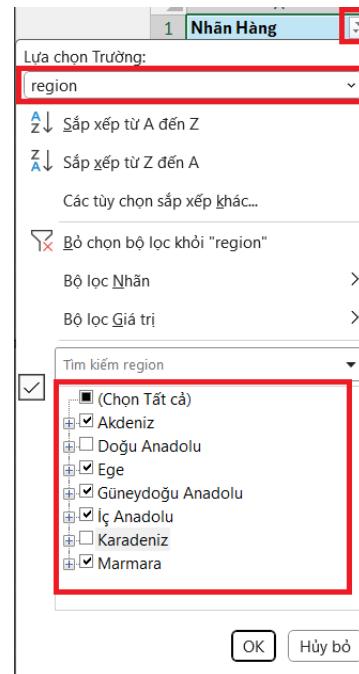
Hình 461: Kéo thả các thuộc tính thích hợp vào trường “Hàng” của Pivot Table

**Bước 2:** Kéo thả thuộc tính bill\_id của bảng Fact vào mục “Giá trị”, chọn “Số khác biệt” cho “Tóm tắt trường giá trị theo” của thuộc tính này.



Hình 462: Kéo thả thuộc tính thích hợp vào trường “Giá trị” và chọn hàm tính toán

**Bước 3:** Tiến hành lọc ra vùng miền ngoại trừ hai vùng miền ‘Doğu Anadolu’, ‘Karadeniz’



Hình 463: Tiến hành lọc ra vùng miền ngoại trừ 2 vùng miền ‘Doğu Anadolu’, ‘Karadeniz’

**Bước 4:** Theo dõi kết quả ở dạng Table.

Nhân Hàng		Tổng số đếm Riêng biệt của bill_id
Akdeniz		
1		12824
2		14323
3		16959
4		536
Ege		
1		15160
2		16306
3		19437
4		697
Güneydoğu Anadolu		
1		11241
2		12349
3		13656
4		396
İç Anadolu		
1		16538
2		18858
3		21031
4		807
Marmara		
1		29936
2		33070
3		38507
4		1330
<b>Tổng Cuối</b>		<b>293961</b>

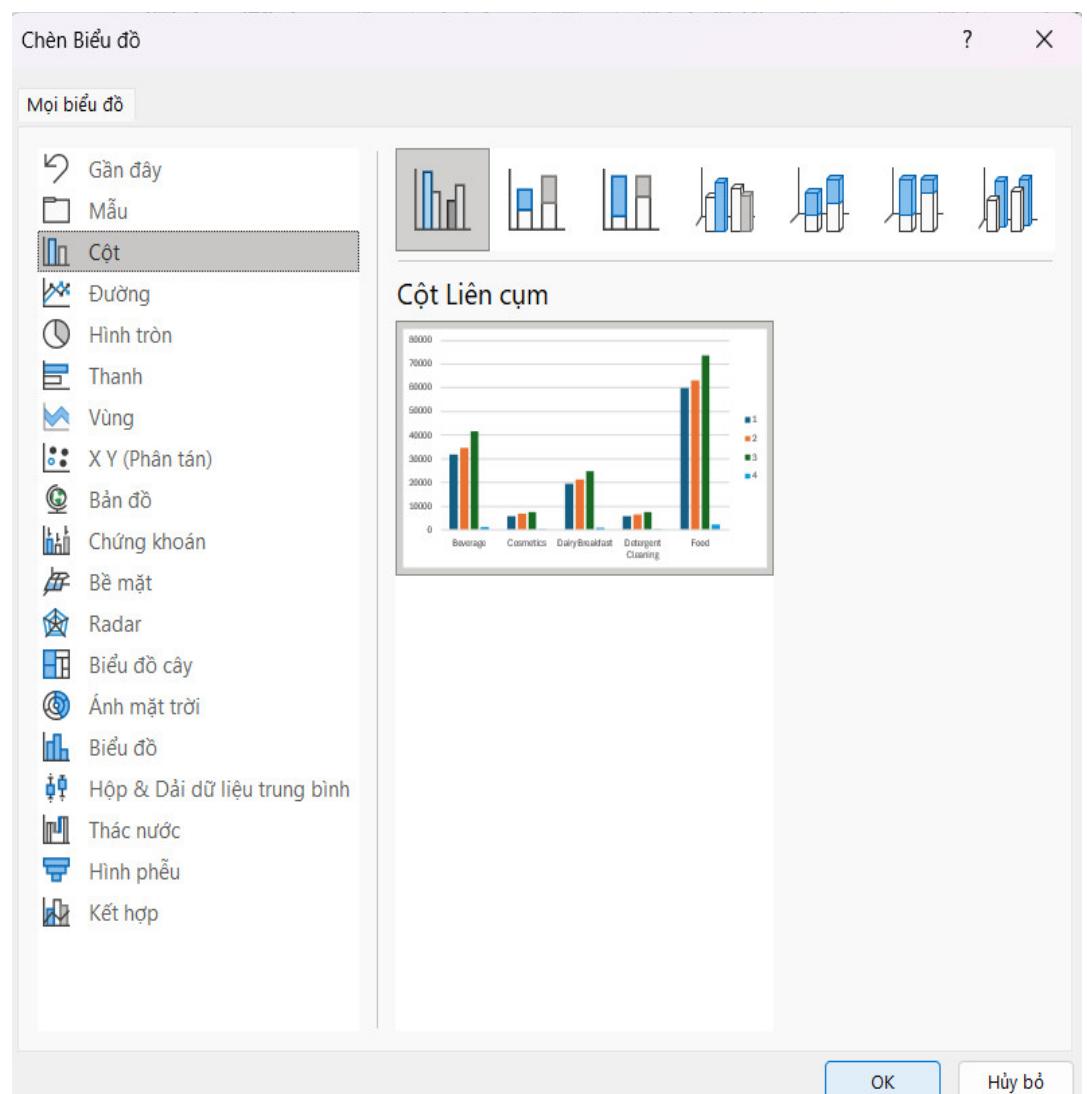
Hình 464: Kết quả câu truy vấn 14 hiển thị dưới dạng table

**Bước 5:** Theo dõi kết quả report dưới dạng biểu đồ. Để hiển thị biểu đồ, tại tab “Phân tích Pivot Table”, nhấn “Pivot chart”.

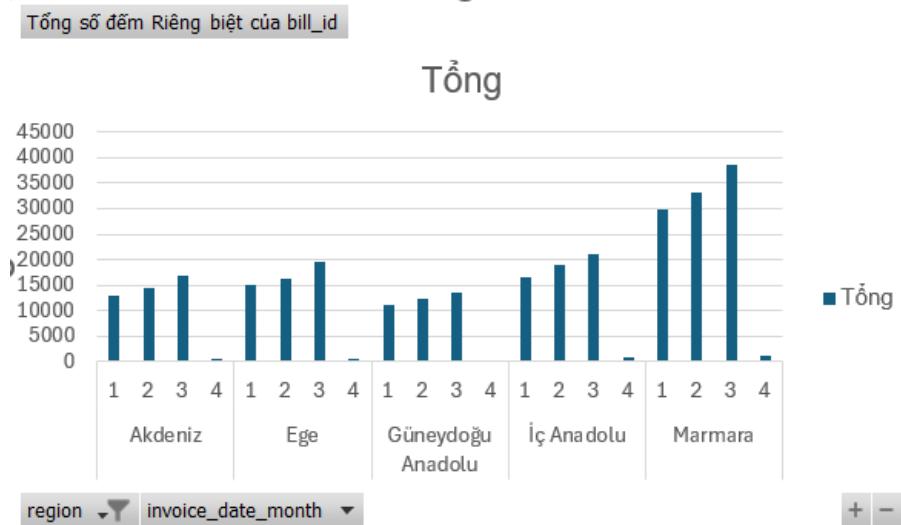


Hình 465: Giao diện hướng dẫn vào noi hiển thị kết quả dưới dạng biểu đồ

Chọn loại Chart và nhấn “OK”.



Hình 466: Giao diện chọn loại biểu đồ

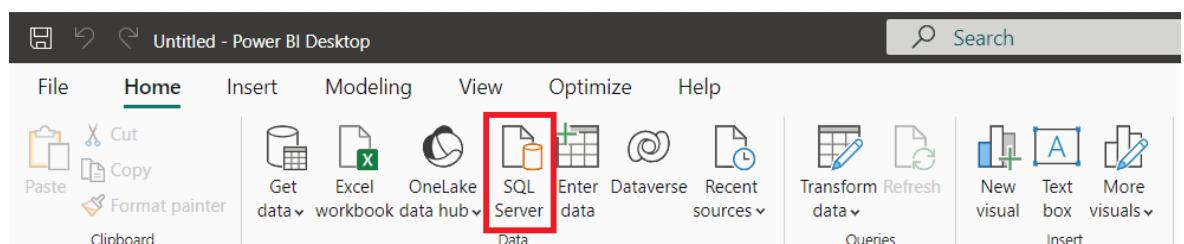


Hình 467: Kết quả câu truy vấn 14 hiển thị dưới dạng biểu đồ

### 3.13. Quá trình lập báo biểu bằng công cụ Power BI:

#### 3.13.1. Kết nối Power BI với SQL Server:

**Bước 1:** Mở Power BI. Tại mục “Data” của tab “Home”, chọn “SQL Server”.



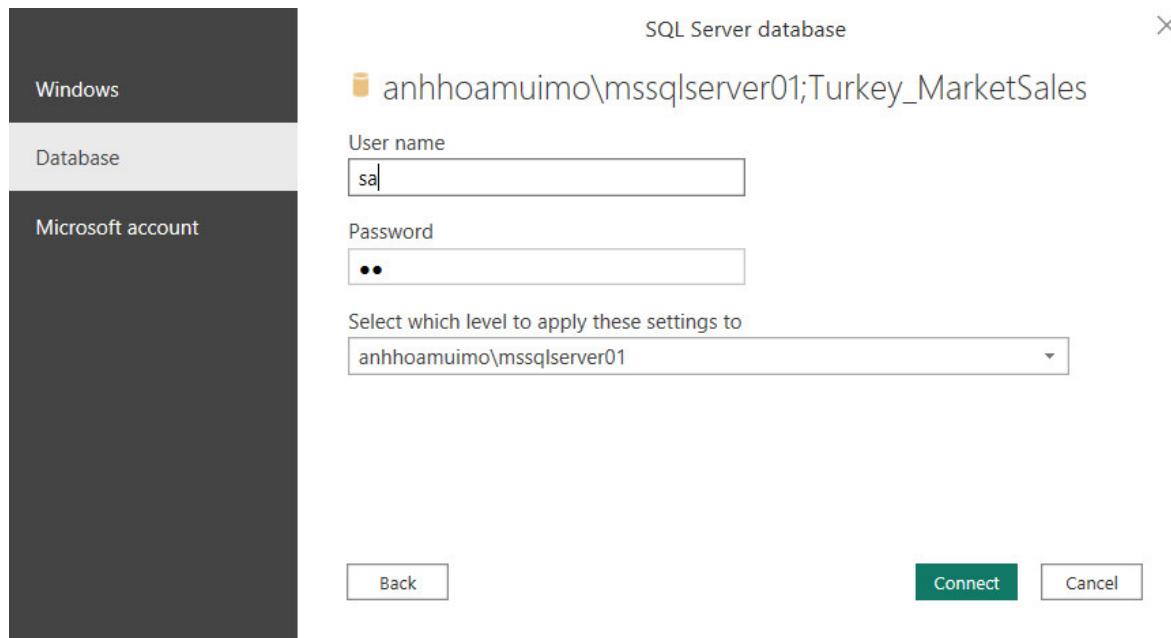
Hình 468: Giao diện hướng dẫn kết nối Excel với SQL Server

**Bước 2:** Nhập tên Server và tên Database rồi click “OK”.



Hình 469: Giao diện nhập tên Server và tên cơ sở dữ liệu (warehouse) lưu trong SQL Server

**Bước 3:** Một cửa sổ thực hiện đăng nhập vào SQL Server xuất hiện. Tại đây, ta đăng nhập bằng tài khoản sa của database rồi click “Connect”.



Hình 470: Giao diện đăng nhập tài khoản để kết nối tới SQL Server

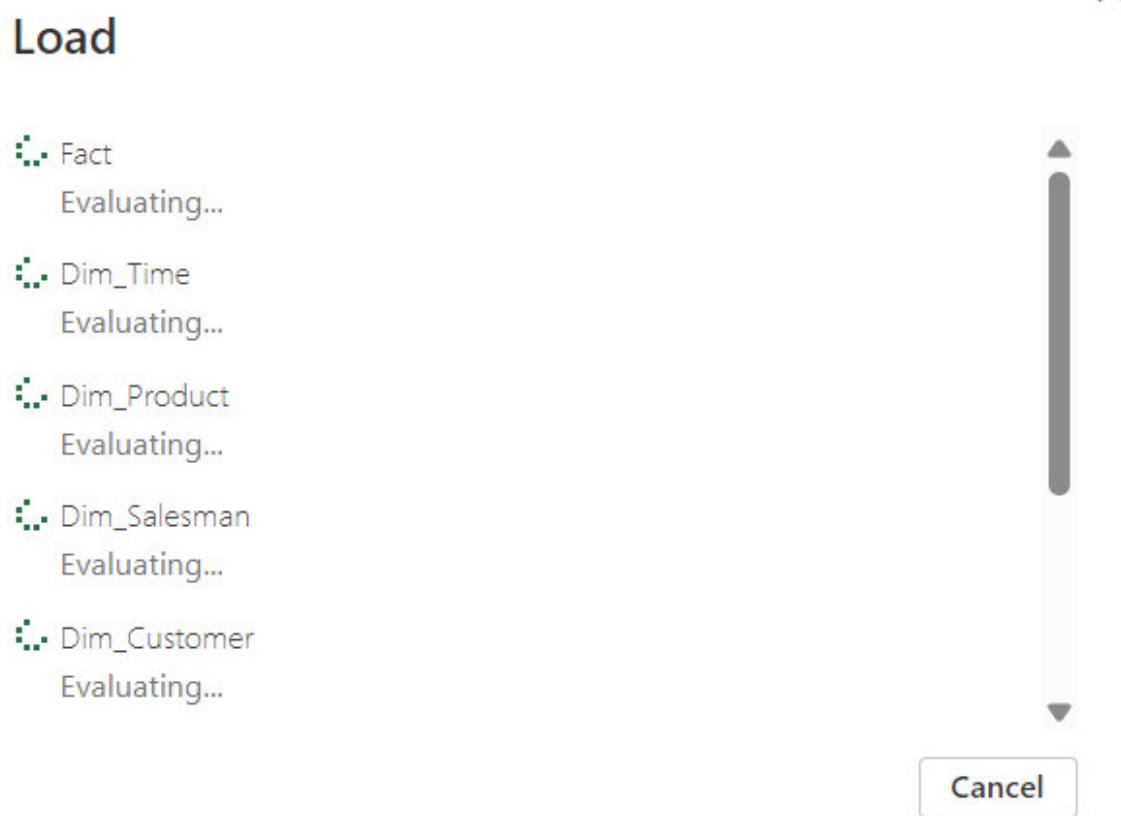
**Bước 4:** Thực hiện chọn bảng Fact và các bảng Dim và đợi quá trình load dữ liệu của power BI.

### Navigator

The screenshot shows the Microsoft Data Explorer 'Navigator' window. On the left, a tree view displays database objects under 'ANHHOAMUIMO\MSSQLSERVER01: Turkey\_M...'. The 'Dim\_Branch' table is selected, indicated by a checked checkbox next to its name. Other tables like Dim\_Brand, Dim\_Category, etc., also have checkboxes next to them. On the right, a large table titled 'Dim\_Branch' is displayed with columns: branch\_id, branch\_name, city, and region. The table contains 23 rows of data. At the bottom of the window, there are buttons for 'Select Related Tables', 'Load' (in green), 'Transform Data', and 'Cancel'.

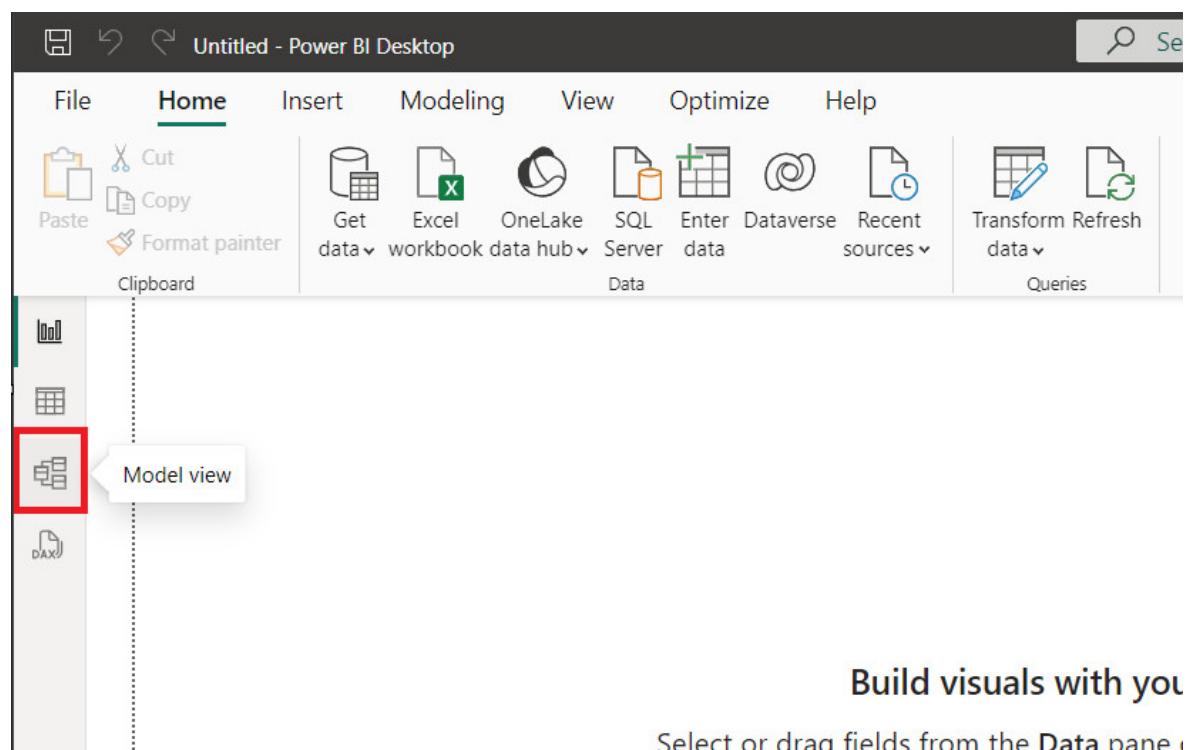
branch_id	branch_name	city	region
1	Adana Subesi	Adana	Akdeniz
2	Adiyaman Subesi	Adiyaman	Güneydoğu Anadolu
3	Afyonkarahisar Subesi	Afyonkarahisar	Ege
4	Ağrı Subesi	Ağrı	Doğu Anadolu
5	Aksaray Subesi	Aksaray	İç Anadolu
6	Amasya Subesi	Amasya	Karadeniz
7	Ankara Subesi	Ankara	İç Anadolu
8	Antalya Subesi	Antalya	Akdeniz
9	Ardahan Subesi	Ardahan	Doğu Anadolu
10	Artvin Subesi	Artvin	Karadeniz
11	Aydın Subesi	Aydın	Ege
12	Balıkesir Subesi	Balıkesir	Ege
13	Bartın Subesi	Bartın	Karadeniz
14	Batman Subesi	Batman	Güneydoğu Anadolu
15	Bayburt Subesi	Bayburt	Karadeniz
16	Bilecik Subesi	Bilecik	Marmara
17	Bingöl Subesi	Bingöl	Doğu Anadolu
18	Bitlis Subesi	Bitlis	Doğu Anadolu
19	Bolu Subesi	Bolu	Karadeniz
20	Burdur Subesi	Burdur	Akdeniz
21	Bursa Subesi	Bursa	Marmara
22	Çanakkale Subesi	Çanakkale	Marmara
23	Çankırı Subesi	Çankırı	İç Anadolu

Hình 471: Chọn tất cả các bảng dữ liệu để đổ vào Excel



Hình 472: Giao diện quá trình load dữ liệu của Power BI

**Bước 5:** Sau khi Power BI load dữ liệu xong, nhấn vào biểu tượng “Model view”, ta được relationship như sau:



Hình 473: Giao diện hướng dẫn vào xem lược đồ cơ sở dữ liệu trong Excel



Hình 474: Lược đồ mối quan hệ giữa các table

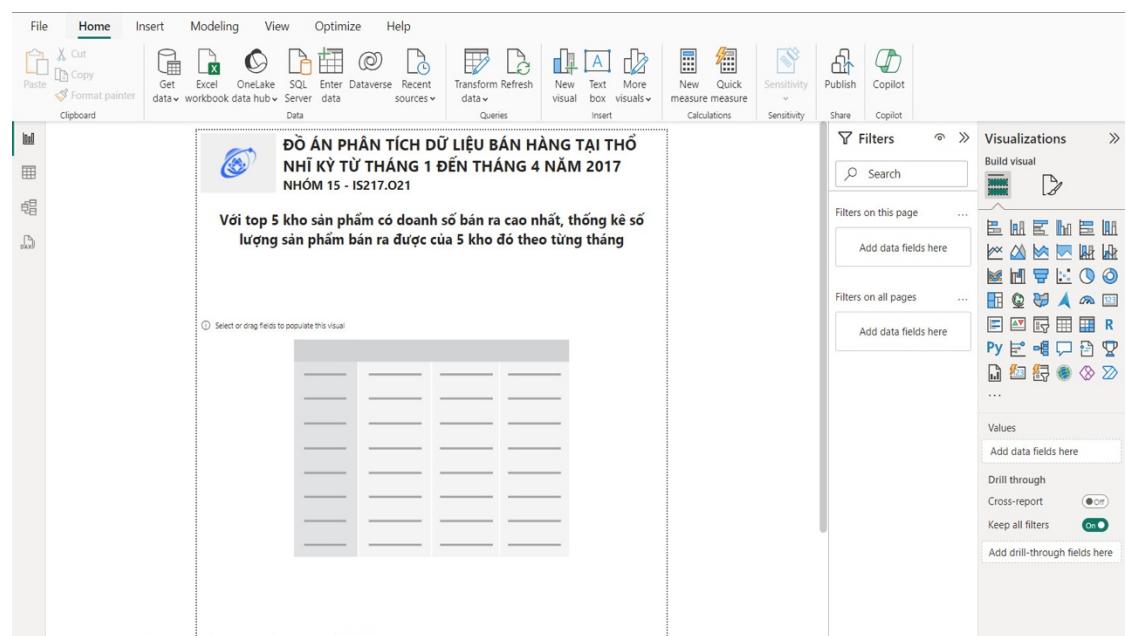
### 3.13.2. Thực hiện Report bằng Power BI:

#### 3.13.2.1. Câu truy vấn 1:

**Nội dung báo biểu:** Với top 5 kho sản phẩm có doanh số bán ra cao nhất, thống kê số lượng sản phẩm bán ra được của 5 kho đó theo từng tháng.

**Thực hiện Report:**

**Bước 1:** Tạo mẫu report và chọn “Matrix” trong tab “Visualizations”.



Hình 475: Giao diện tạo mẫu báo cáo trong Power BI

**Bước 2:** Từ tab “Data”, trong tab “Visualizations”, kéo Month trong invoice\_date\_time của bảng Dim\_Time vào “Columns”; kéo category\_name trong bảng Dim\_Category vào “Rows”; kéo amount của bảng Fact vào “Values”. Tại “Values”, nhấn chuột trái vào amount chọn “Sum”.

The screenshot shows the 'Build visual' interface in Power BI. A context menu is open under the heading 'Data'. The menu includes options like 'Remove field', 'Rename for this visual', 'Move to', 'Add a sparkline', 'Conditional formatting', 'Remove conditional formatting', and a list of aggregation functions: Sum, Average, Minimum, Maximum, Count (Distinct), Count, Standard deviation, Variance, and Median. At the bottom of the menu is a 'New quick measure' section containing several mathematical expressions such as  $\sum \text{bill_id}$ ,  $\sum \text{customer_id}$ , etc. The main area of the interface shows fields being mapped to rows, columns, and values.

**Visualizations**

**Build visual**

**Data**

- Remove field
- Rename for this visual
- Move to >
- Add a sparkline
- Conditional formatting >
- Remove conditional formatting
- ✓ Sum
- Average
- Minimum
- Maximum
- Count (Distinct)
- Count
- Standard deviation
- Variance
- Median
- Show value as >
- New quick measure

category\_name

invoice\_date\_time

Month

Sum of amount

bill\_id

customer\_id

dim\_product\_id

final\_total\_trans

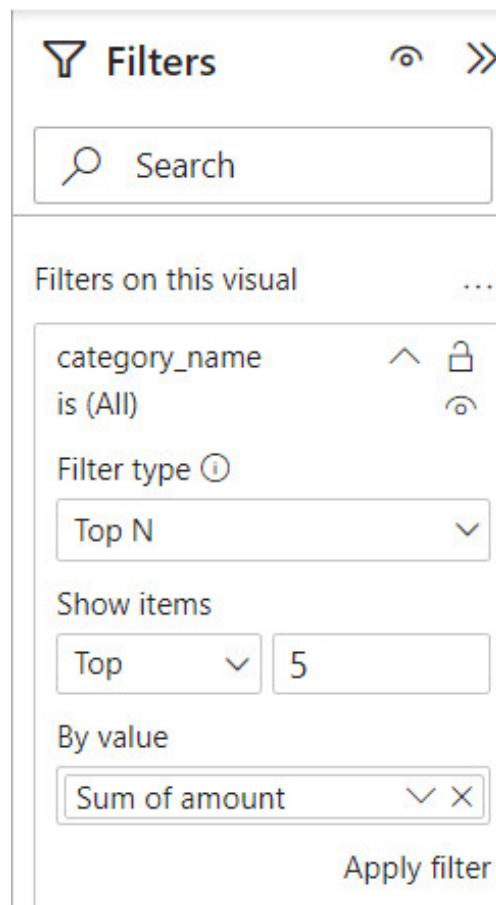
invoice\_date\_time

pre-promotion\_total\_tr...

salesman\_id

Hình 476: Kéo các thuộc tính và độ đo thích hợp vào mục “Hàng”, “Cột” và “Giá trị”

**Bước 3:** Ở tab “Filters”, ta lọc ra Top 5 kho sản phẩm có doanh số bán hàng cao nhất, sau đó click “Apply Filter”.



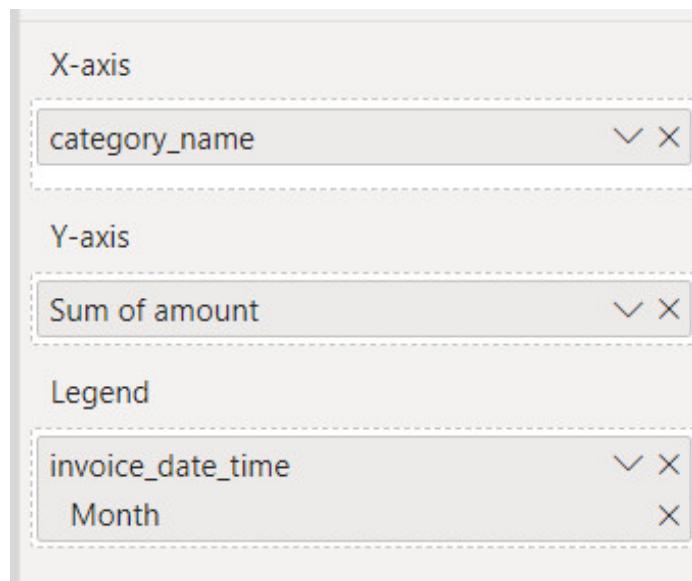
Hình 477: Lọc ra điều kiện top 5 kho sản phẩm theo doanh số

**Bước 4:** Xem kết quả câu truy vấn dưới dạng Matrix.

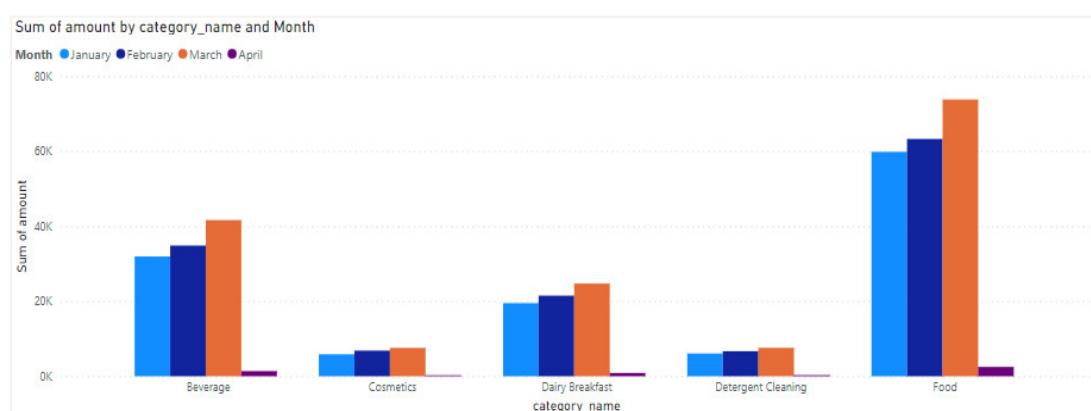
category_name	January	February	March	April	Total
Beverage	31.939,00	34.851,00	41.651,00	1.375,00	109.816,00
Cosmetics	5.862,00	6.839,00	7.569,00	244,00	20.514,00
Dairy Breakfast	19.488,87	21.499,17	24.729,47	887,83	66.605,34
Detergent Cleaning	6.071,00	6.687,00	7.595,00	294,00	20.647,00
Food	59.846,10	63.269,22	73.796,75	2.487,63	199.399,70
<b>Total</b>	<b>123.206,97</b>	<b>133.145,40</b>	<b>155.341,21</b>	<b>5.288,46</b>	<b>416.982,05</b>

Hình 478: Kết quả câu truy vấn 1 hiển thị dưới dạng ma trận

**Bước 5:** Hiển thị kết quả dưới dạng Chart, chọn dạng “**Clustered columns chart**” trong tab “**Visualizations**”. Sau đó, điều chỉnh lại Y-axis, X-axis, Legend như sau: (\*)

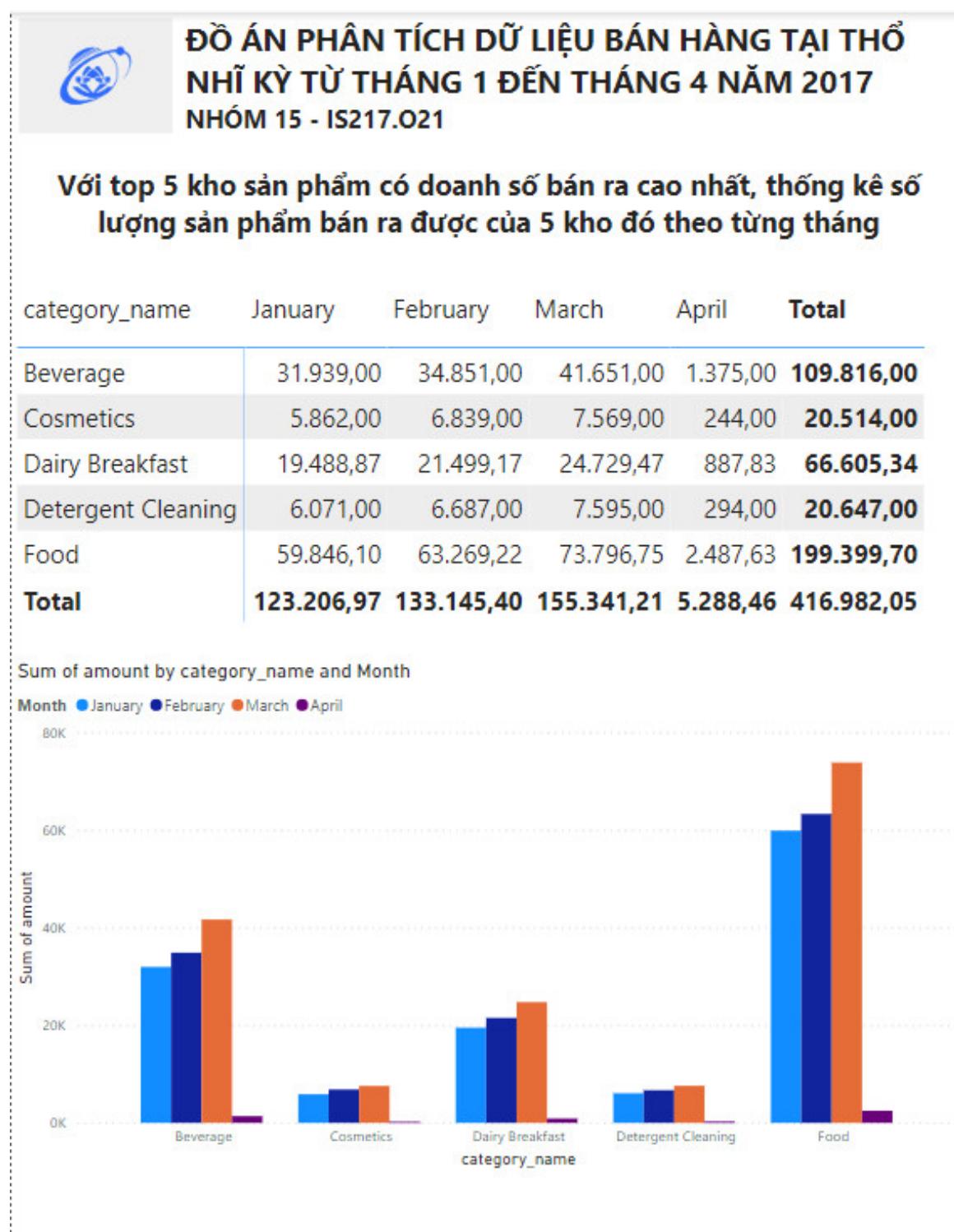


Hình 479: Tinh chỉnh cột X, cột Y và Legend của biểu đồ



Hình 480: Kết quả câu truy vấn 1 hiển thị dưới dạng biểu đồ

**Bước 6:** Tổng hợp kết quả thành Report.



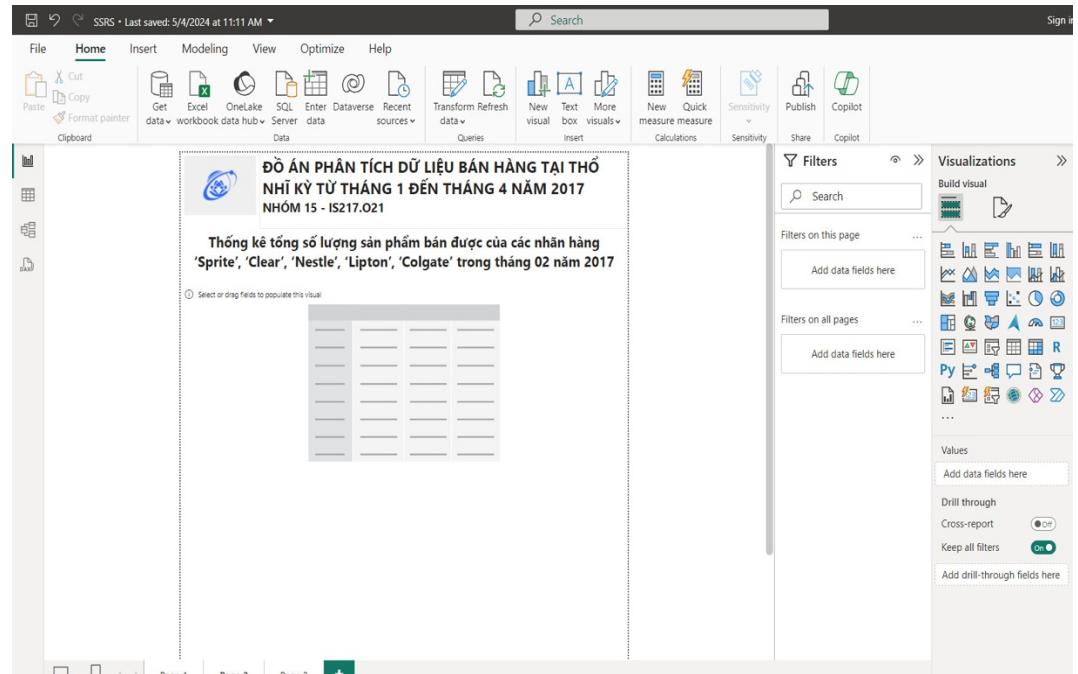
Hình 481: Report câu truy vấn 1

### 3.13.2.2. Câu truy vấn 2:

**Nội dung báo biểu:** Thống kê tổng số lượng sản phẩm bán được của các nhãn hàng ‘Sprite’, ‘Clear’, ‘Nestle’, ‘Lipton’, ‘Colgate’ trong tháng 02 năm 2017.

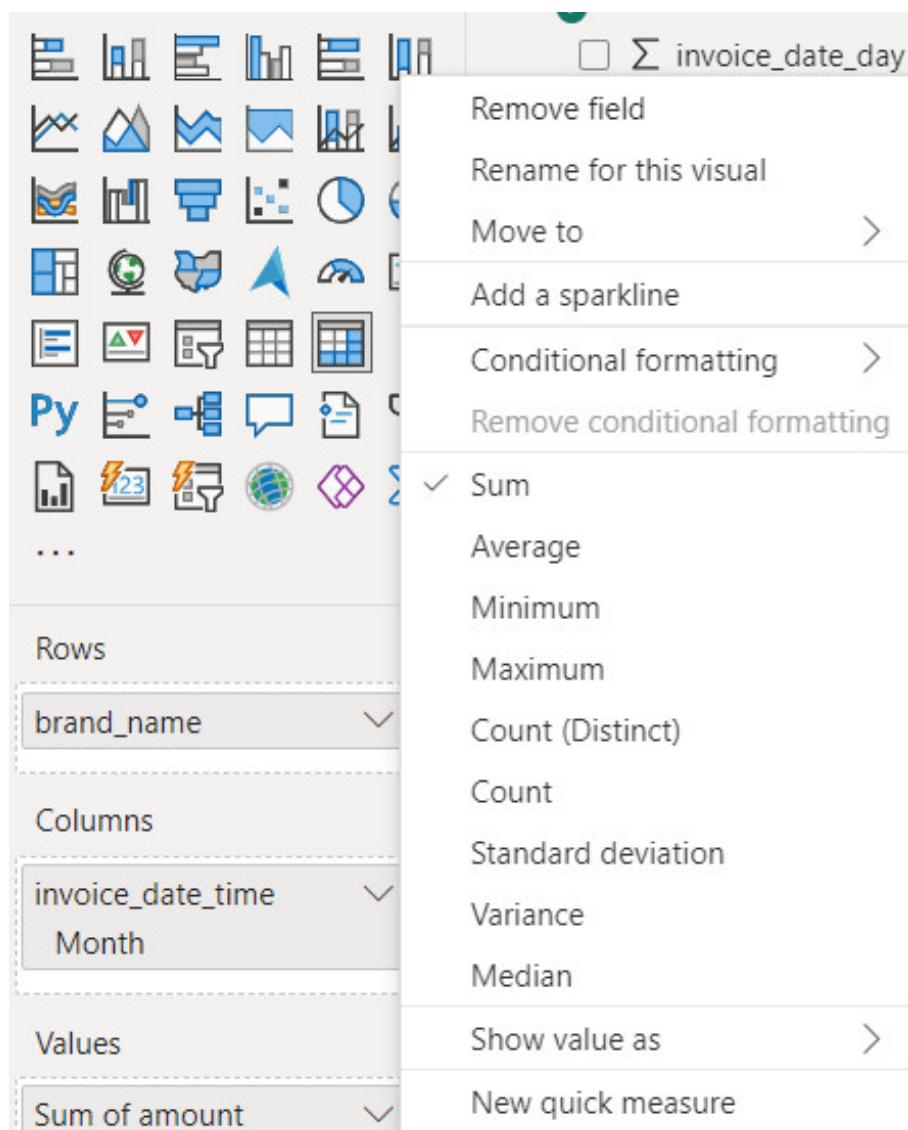
## Thực hiện Report:

Bước 1: Tạo mẫu report và chọn “Table” trong tab “Visualizations”.



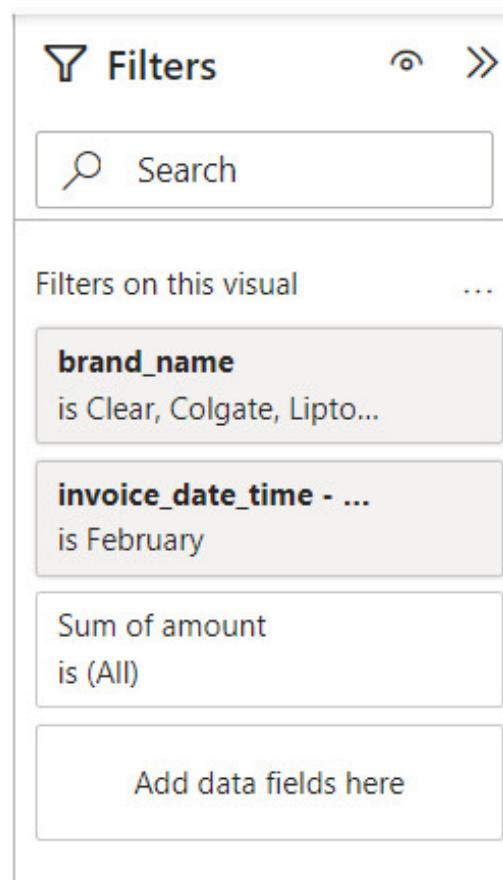
Hình 482: Giao diện tạo mẫu báo cáo trong Power BI

Bước 2: Từ tab “Data”, trong tab “Visualizations”, kéo Month trong invoice\_date\_time của bảng Dim\_Time vào “Columns”; kéo brand\_name trong bảng Dim\_Brand vào “Rows”; kéo amount của bảng Fact vào “Values”. Tại “Values”, nhấn chuột trái vào amount chọn “Sum”.



Hình 483: Kéo các thuộc tính và độ đo thích hợp vào mục “Hàng”, “Cột” và “Giá trị”

**Bước 3:** Ở tab “Filters”, ta lọc ra giá trị “2” của Month và giá trị ‘Sprite’, ‘Clear’, ‘Nestle’, ‘Lipton’, ‘Colgate’ của brand\_name, sau đó click “Apply Filter”.



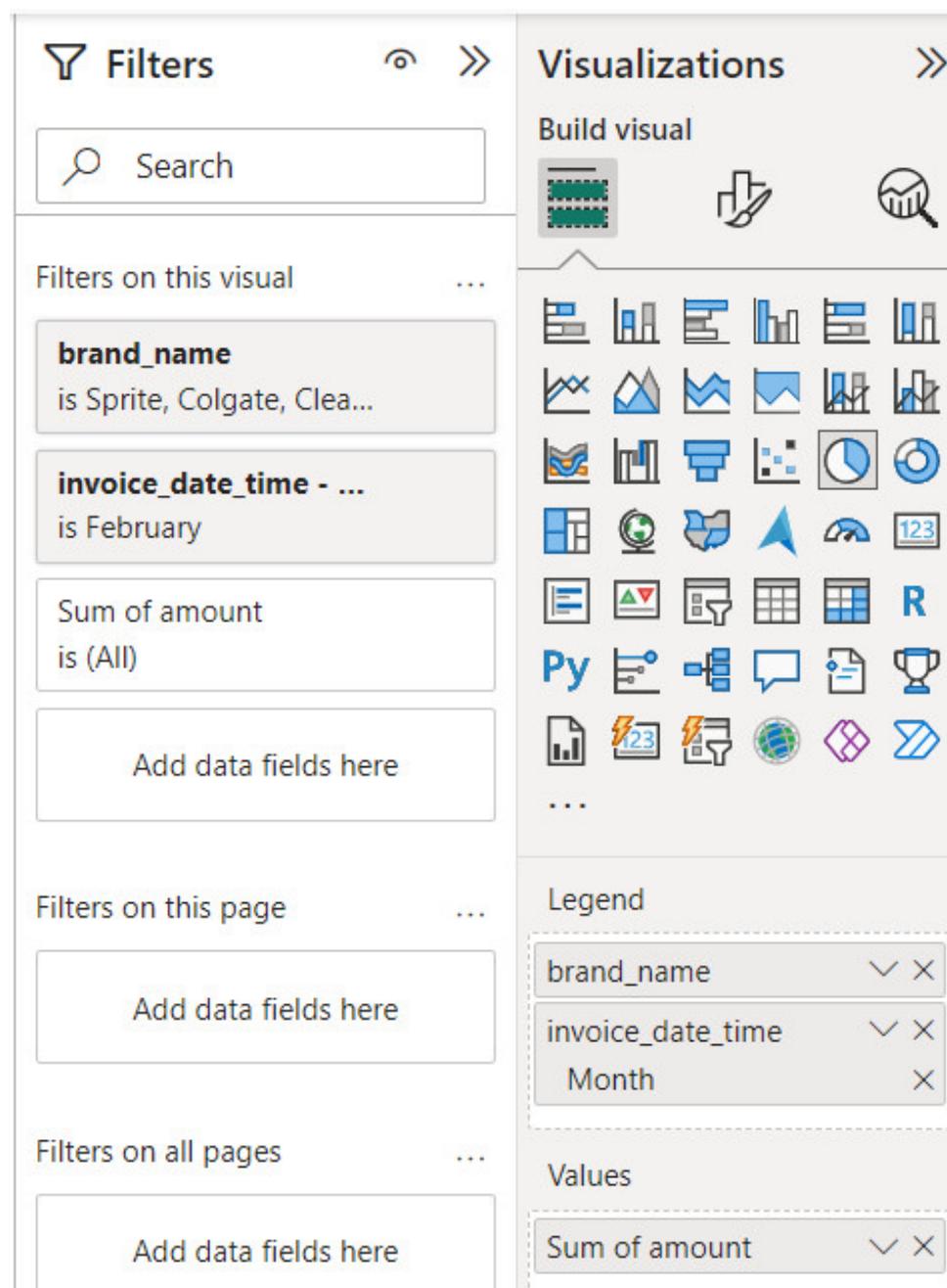
Hình 484: Lọc ra doanh số tháng 02 của các nhãn hàng trong câu truy vấn

**Bước 4:** Xem kết quả câu truy vấn dưới dạng Table.

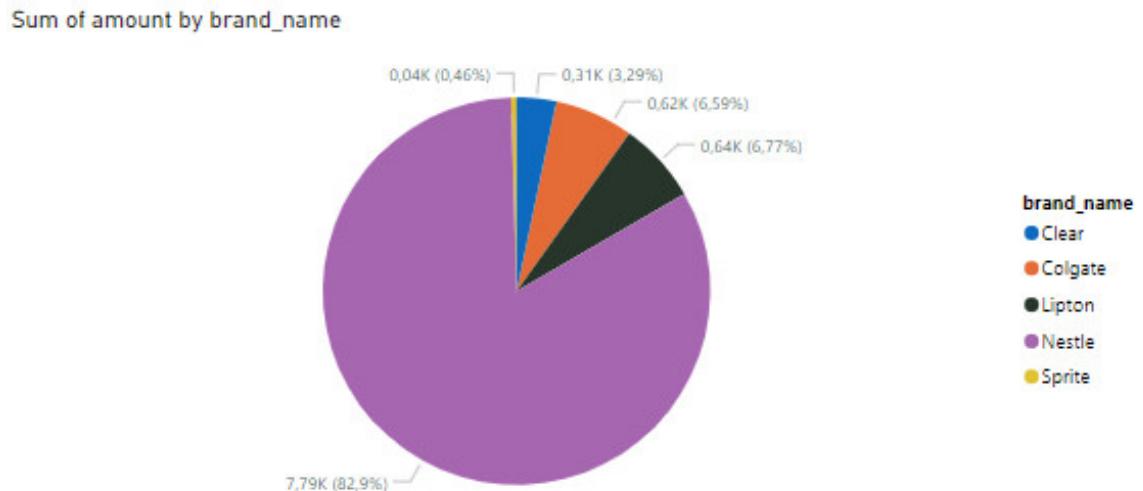
brand_name	February	Total
Clear	309,00	<b>309,00</b>
Colgate	619,00	<b>619,00</b>
Lipton	636,00	<b>636,00</b>
Nestle	7.792,00	<b>7.792,00</b>
Sprite	43,00	<b>43,00</b>
<b>Total</b>	<b>9.399,00</b>	<b>9.399,00</b>

Hình 485: Kết quả câu truy vấn 2 hiển thị dưới dạng table

**Bước 5:** Hiển thị kết quả dưới dạng Chart, chọn dạng “Pie chart” trong tab “Visualizations”. Sau đó, điều chỉnh lại Y-axis, X-axis, Legend như sau:

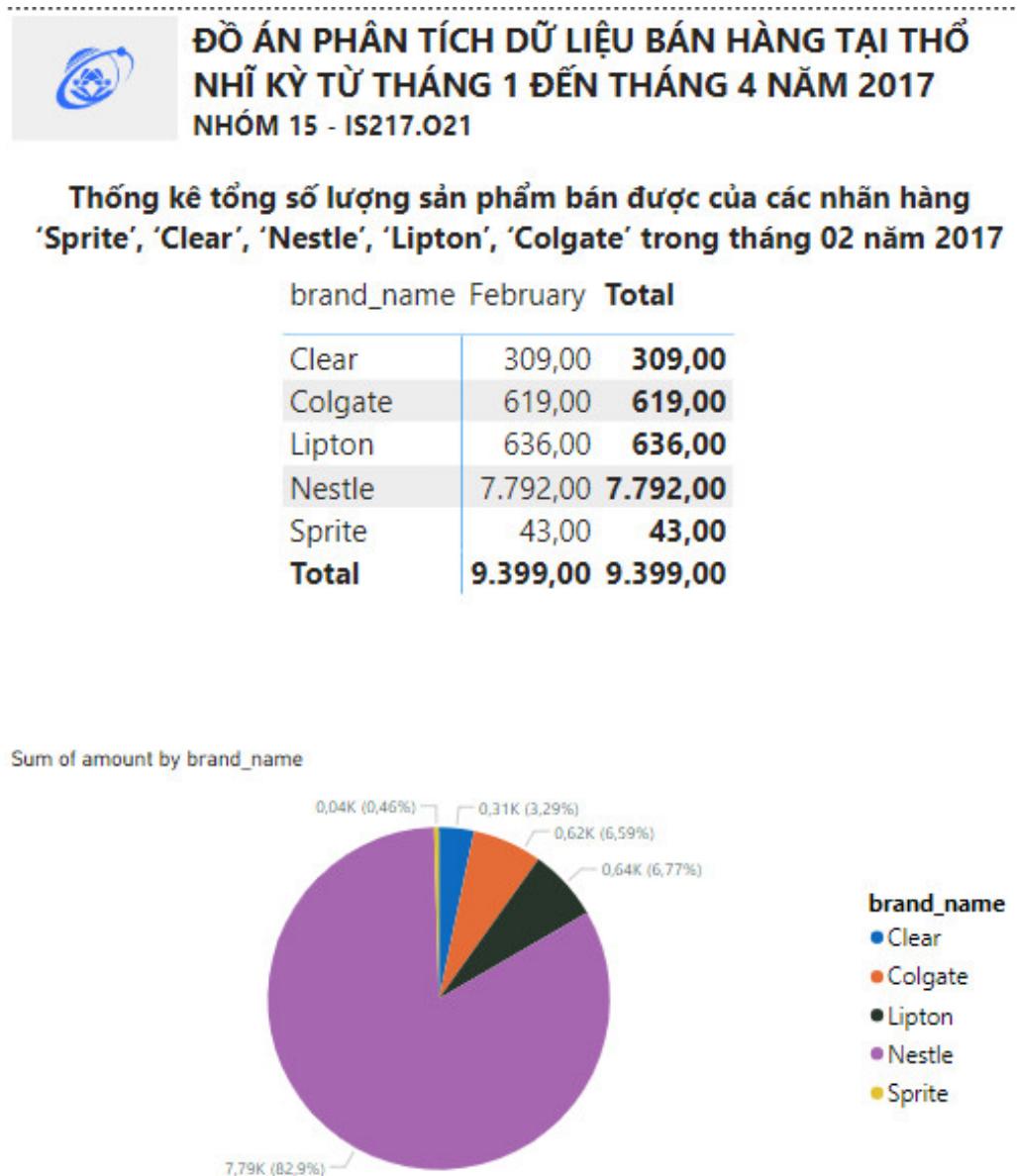


Hình 486: Chọn loại biểu đồ, tinh chỉnh cột X, cột Y và Legend của biểu đồ



Hình 487: Kết quả câu truy vấn 2 hiển thị dưới dạng biểu đồ

**Bước 6:** Tổng hợp kết quả thành Report.



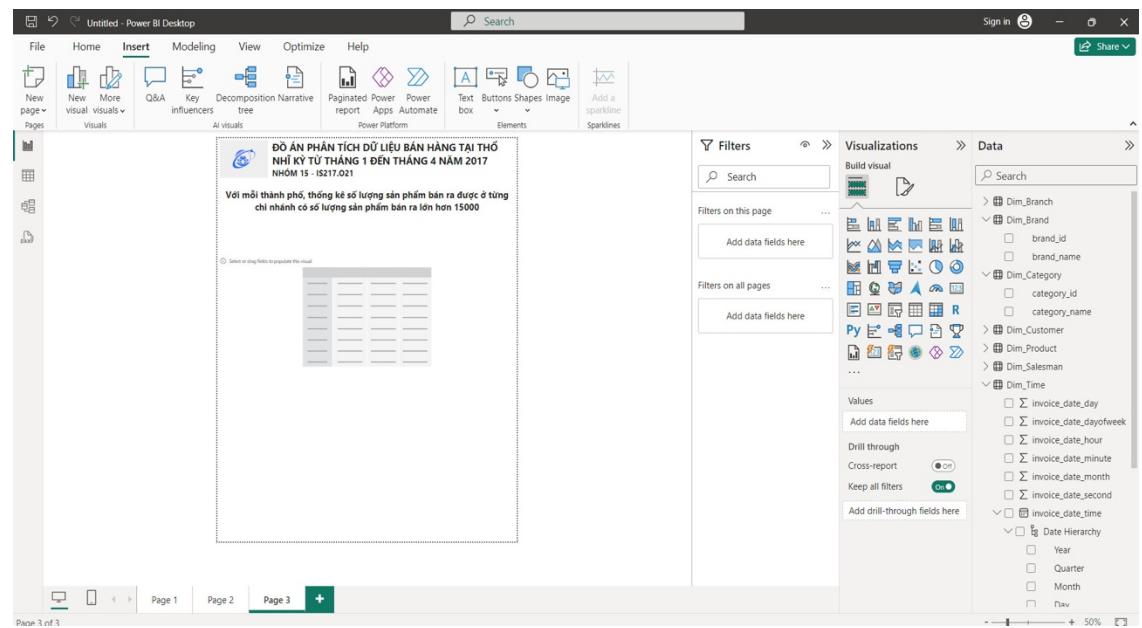
Hình 488: Report câu truy vấn 2

### 3.13.2.3. Câu truy vấn 3:

**Nội dung báo biểu:** Với mỗi thành phố, thống kê số lượng sản phẩm bán ra được ở từng chi nhánh có số lượng sản phẩm bán ra lớn hơn 15000.

**Thực hiện Report:**

**Bước 1:** Tạo mẫu report và chọn “Table” trong tab “Visualizations”.



Hình 489: Giao diện tạo mẫu báo cáo trong Power BI

**Bước 2:** Từ tab “**Data**”, trong tab “**Visualizations**”, kéo city và branch\_name trong bảng Dim\_Branch vào “**Columns**”; kéo amount của bảng Fact vào “**Values**”. Tại “**Values**”, nhấn chuột trái vào amount chọn “**Sum**”.

**Visualizations**

Build visual

Search

Remove field

Rename for this visual

Move to >

Add a sparkline

Conditional formatting >

Remove conditional formatting

✓ Sum

Average

Minimum

Maximum

Count (Distinct)

Count

Standard deviation

Variance

Median

Show value as >

New quick measure

**Data**

Search

□  $\sum$  invoice\_date\_day

□  $\sum$  invoice\_date\_dayofweek

Remove field

Rename for this visual

Move to >

Add a sparkline

Conditional formatting >

Remove conditional formatting

✓ Sum

Average

Minimum

Maximum

Count (Distinct)

Count

Standard deviation

Variance

Median

Show value as >

New quick measure

**Rows**

city

branch\_name

**Columns**

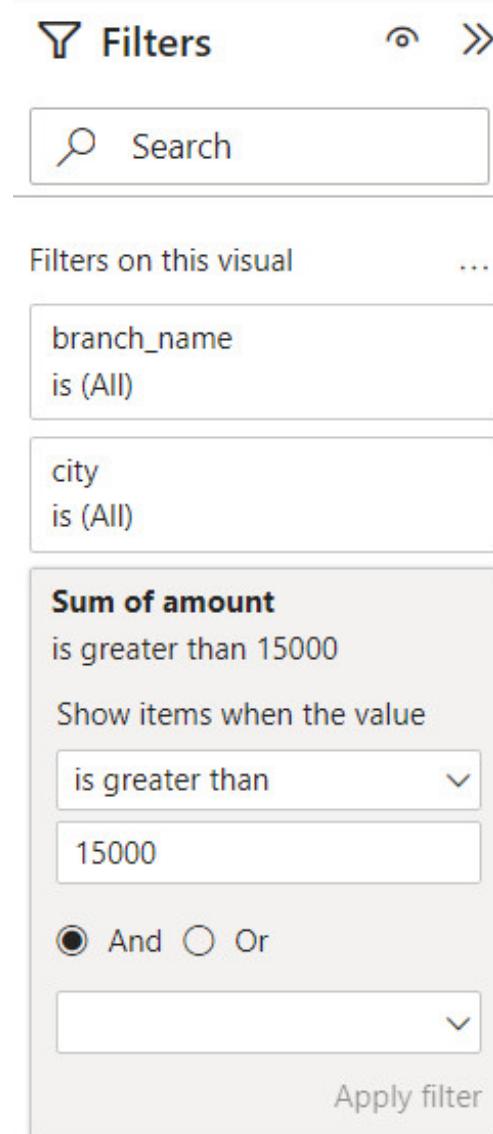
Add data fields here

**Values**

Sum of amount

Hình 490: các thuộc tính và độ đo thích hợp vào mục “Hàng”, “Cột” và “Giá trị”

**Bước 3:** Ở tab “Filters”, ta lọc lấy giá trị amount lớn hơn 15000, sau đó click “Apply Filter”.



Hình 491: Lọc ra chi nhánh có doanh số lớn hơn 15000

**Bước 4:** Xem kết quả câu truy vấn dưới dạng Table.

city	Sum of amount
■ Ankara	<b>30.337,85</b>
Ankara Subesi	30.337,85
■ Bursa	<b>15.699,34</b>
Bursa Subesi	15.699,34
■ İstanbul	<b>83.586,31</b>
İstanbul Subesi	83.586,31
■ İzmir	<b>24.138,66</b>
İzmir Subesi	24.138,66
<b>Total</b>	<b>153.762,16</b>

Hình 492: Kết quả câu truy vấn 3 hiển thị dưới dạng table

**Bước 5:** Hiển thị kết quả dưới dạng Chart, chọn dạng “**Clustered column chart**” trong tab “**Visualizations**”. Sau đó, điều chỉnh lại Y-axis, X-axis, Legend như sau:

The screenshot shows the Power BI interface with the 'Filters' and 'Visualizations' panes open.

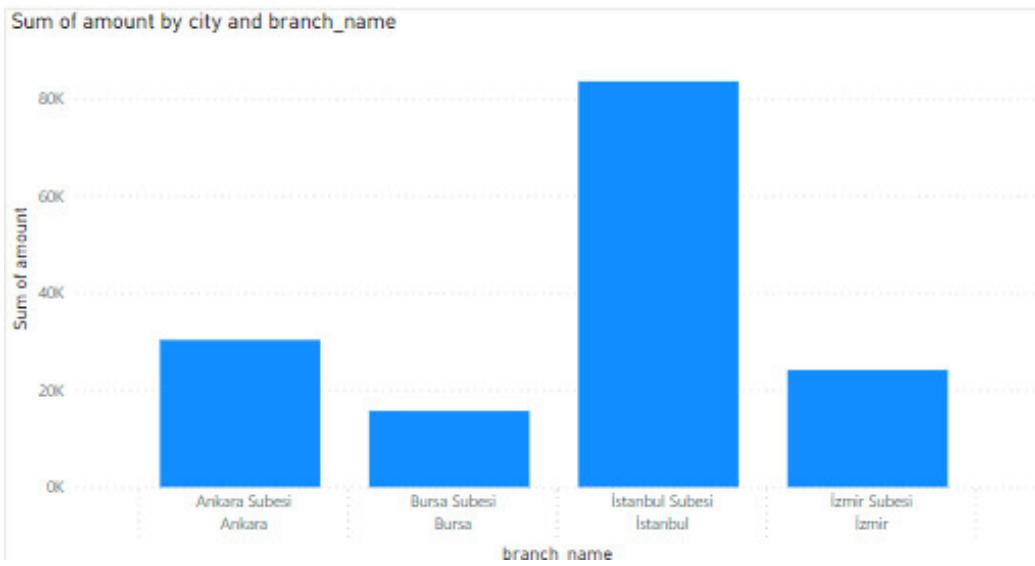
**Filters:**

- Search bar: Search
- Filters on this visual:
  - branch\_name is (All)
  - city is (All)
- Sum of amount**:
  - is greater than 15000
  - Show items when the value is greater than
  - And  Or
  -
- Add data fields here

**Visualizations:**

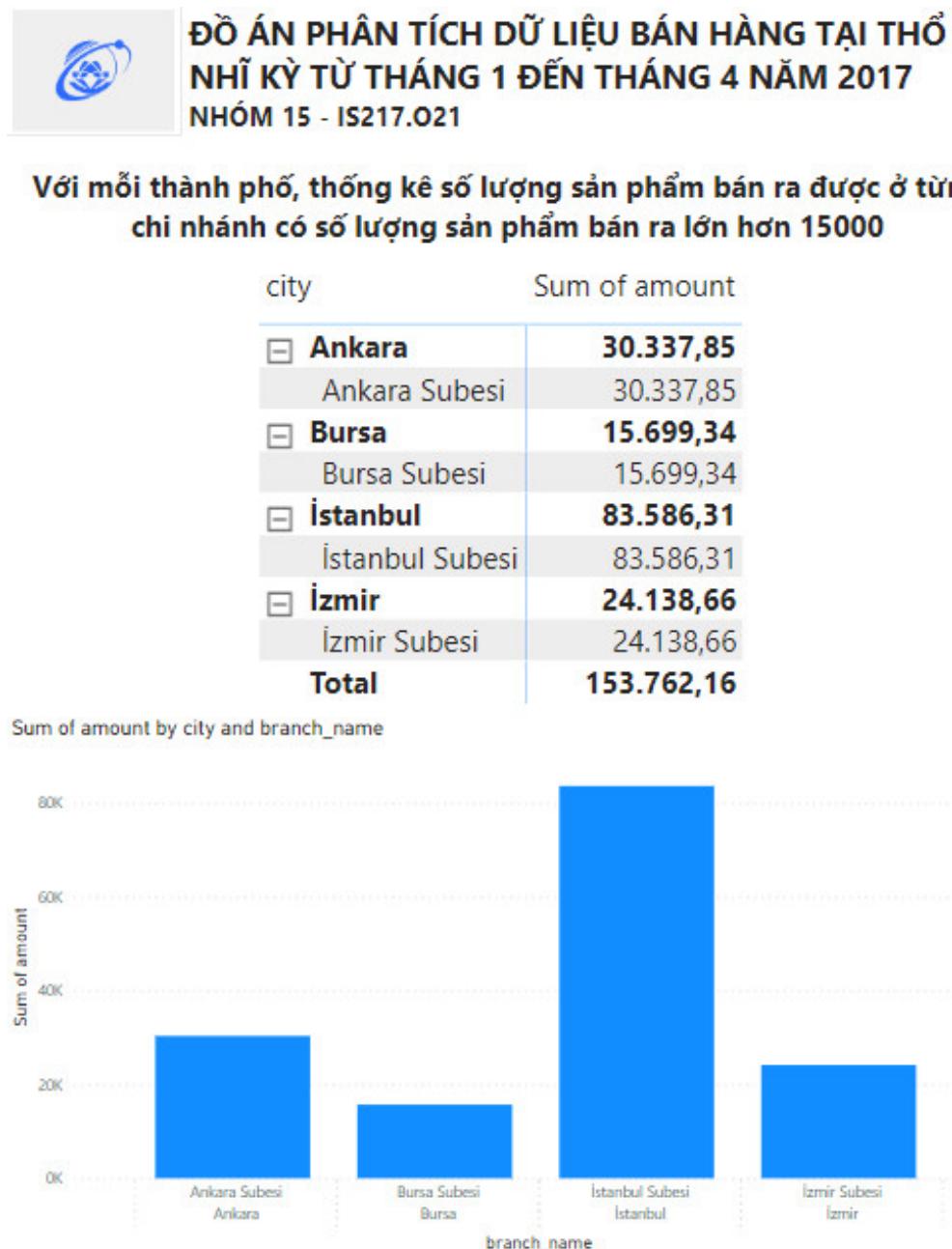
- Build visual:
  - Bar chart icon
  - Line chart icon
  - Map icon
  - Scatter plot icon
  - Donut chart icon
  - Table icon
  - Matrix icon
  - Radar chart icon
  - Pyramid chart icon
  - Gauge chart icon
  - Timeline icon
  - Icon with 123
  - Icon with green checkmark
  - Icon with red minus
  - Icon with blue arrow
  - Icon with grid
  - Icon with R
  - Icon with chart and gear
  - Icon with chart and speech bubble
  - Icon with document and gear
  - Icon with trophy
  - Icon with globe
  - Icon with lightning bolt
  - Icon with chart and lightning bolt
  - Icon with chart and globe
  - Icon with diamond
  - Icon with double arrow
- X-axis:
  - city
  - branch\_name
- Y-axis:
  - Sum of amount
- Legend:
  - Add data fields here

Hình 493: Chọn loại biểu đồ, tinh chỉnh cột X, cột Y và Legend của biểu đồ



Hình 494: Kết quả câu truy vấn 3 hiển thị dưới dạng biểu đồ

**Bước 6:** Tổng hợp kết quả thành Report.

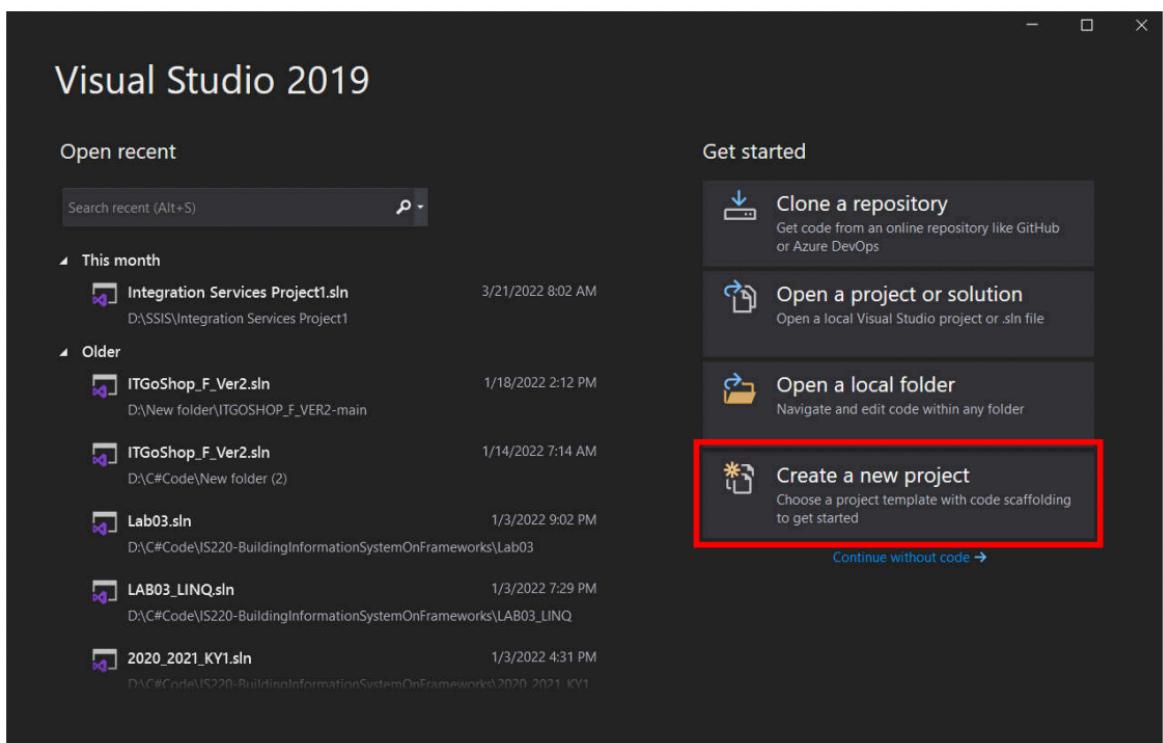


Hình 495: Report câu truy vấn 3

### 3.14. Quá trình lập báo biểu bằng Visual Studio (SSRS):

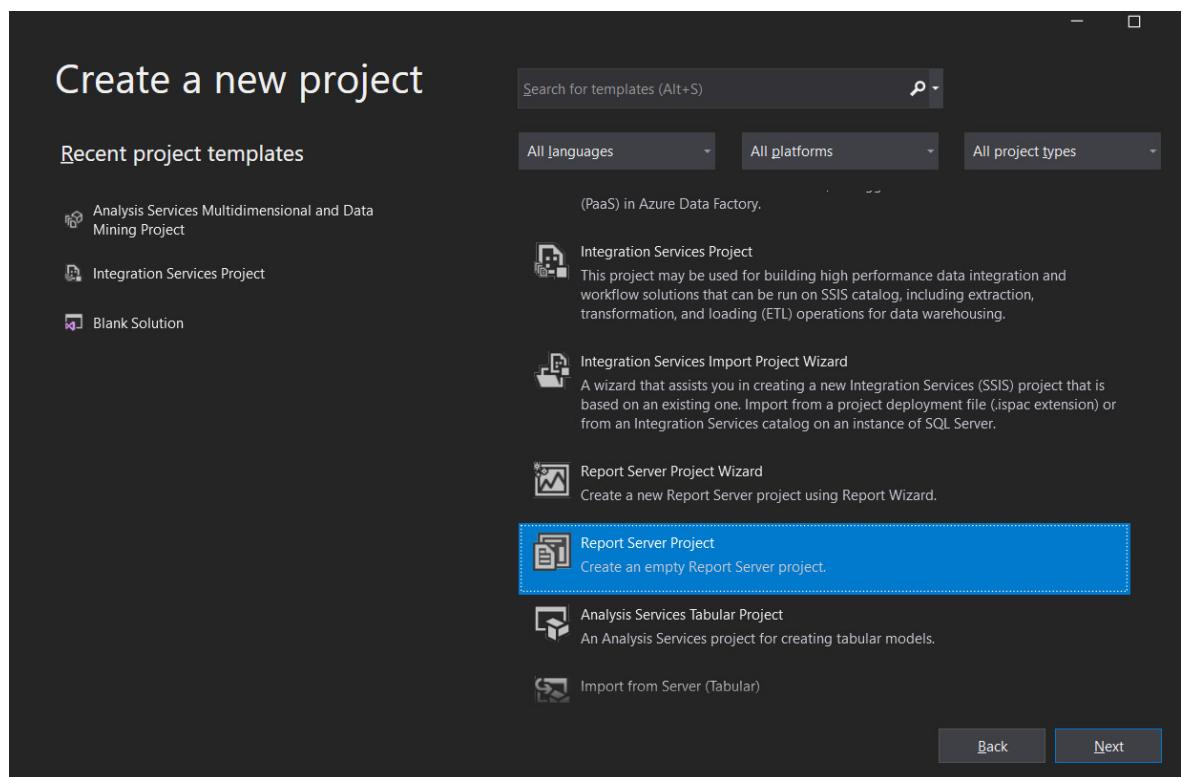
#### 3.14.1. Tạo Project SSRS, thêm data source và dataset:

Bước 1: Mở Visual Studio 2019 và chọn “Create a new project”.



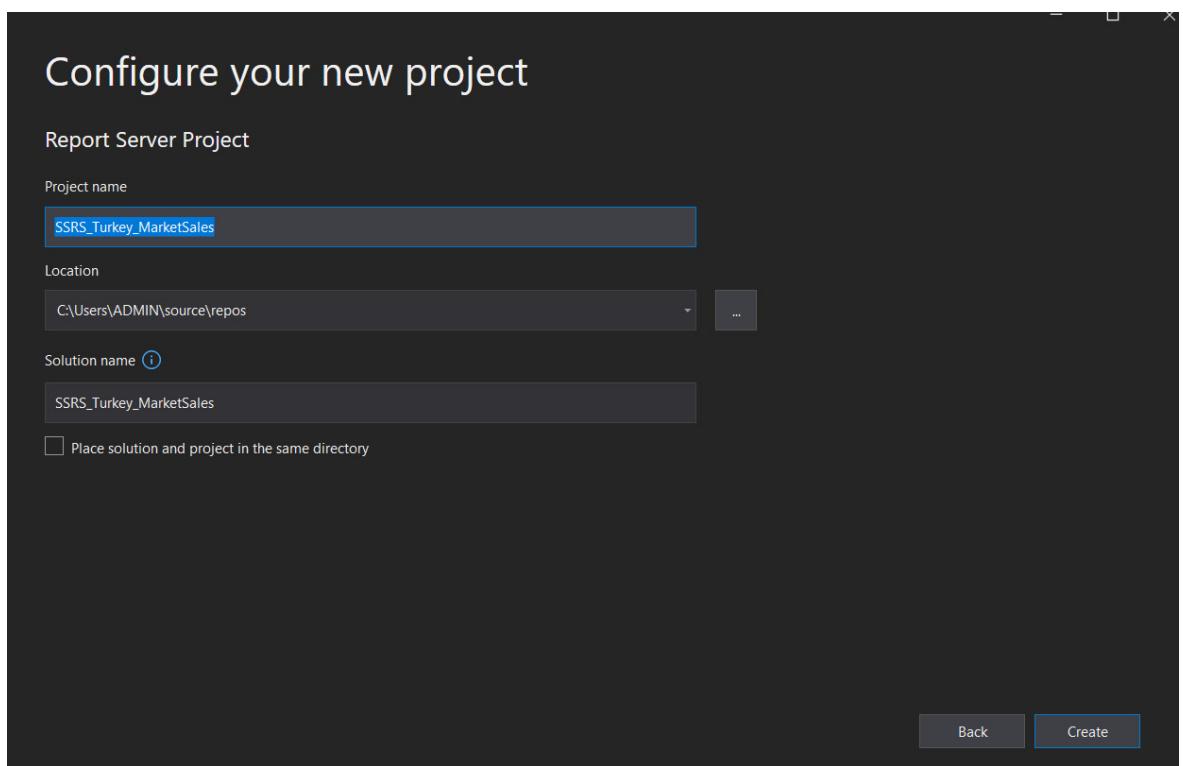
Hình 496: Khởi tạo một dự án mới

**Bước 2:** Chọn “Report Server Project” và click “Next”.



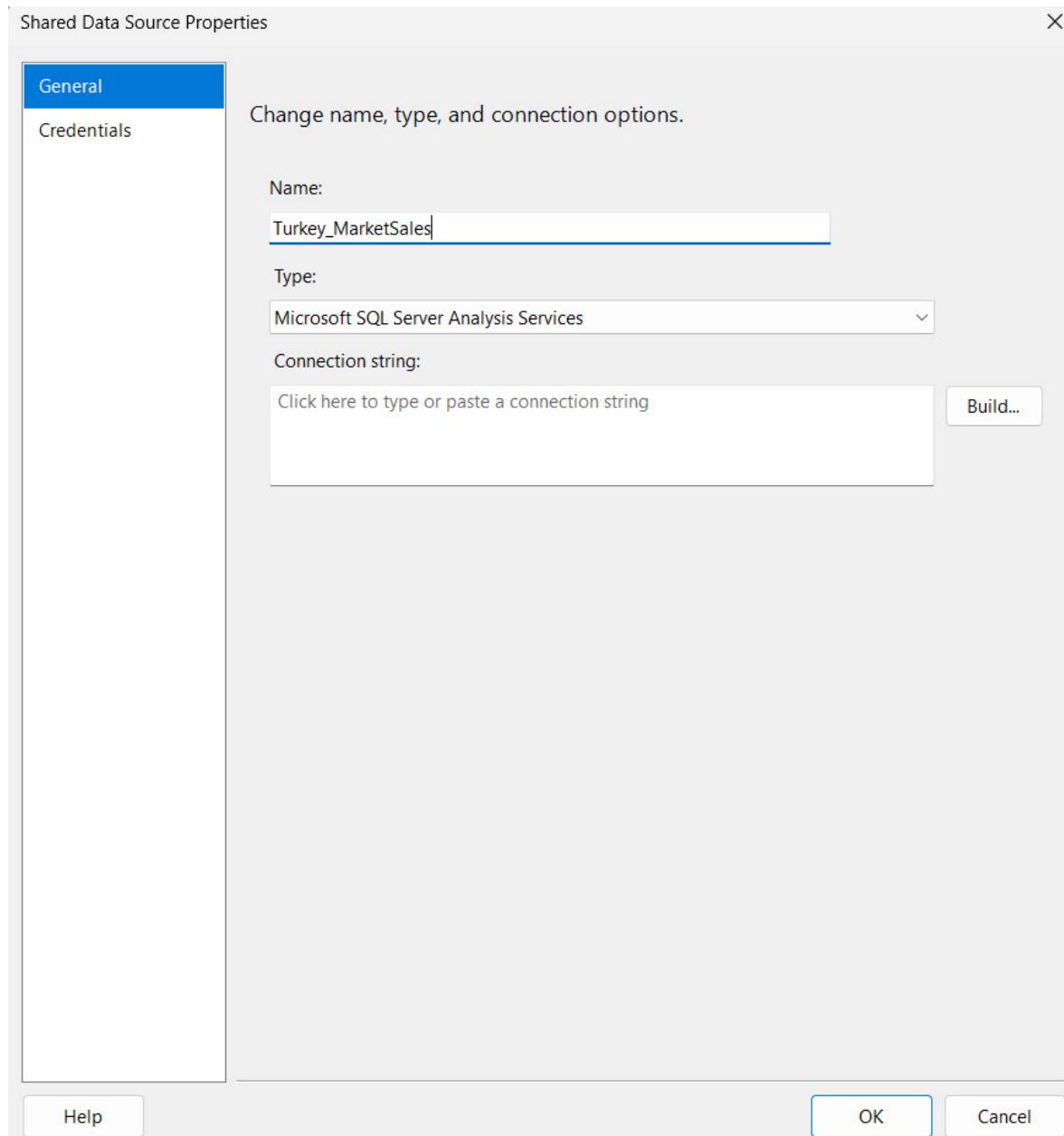
Hình 497: Giao diện hướng dẫn chọn loại dự án SSRS

**Bước 3:** Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn “Create”.



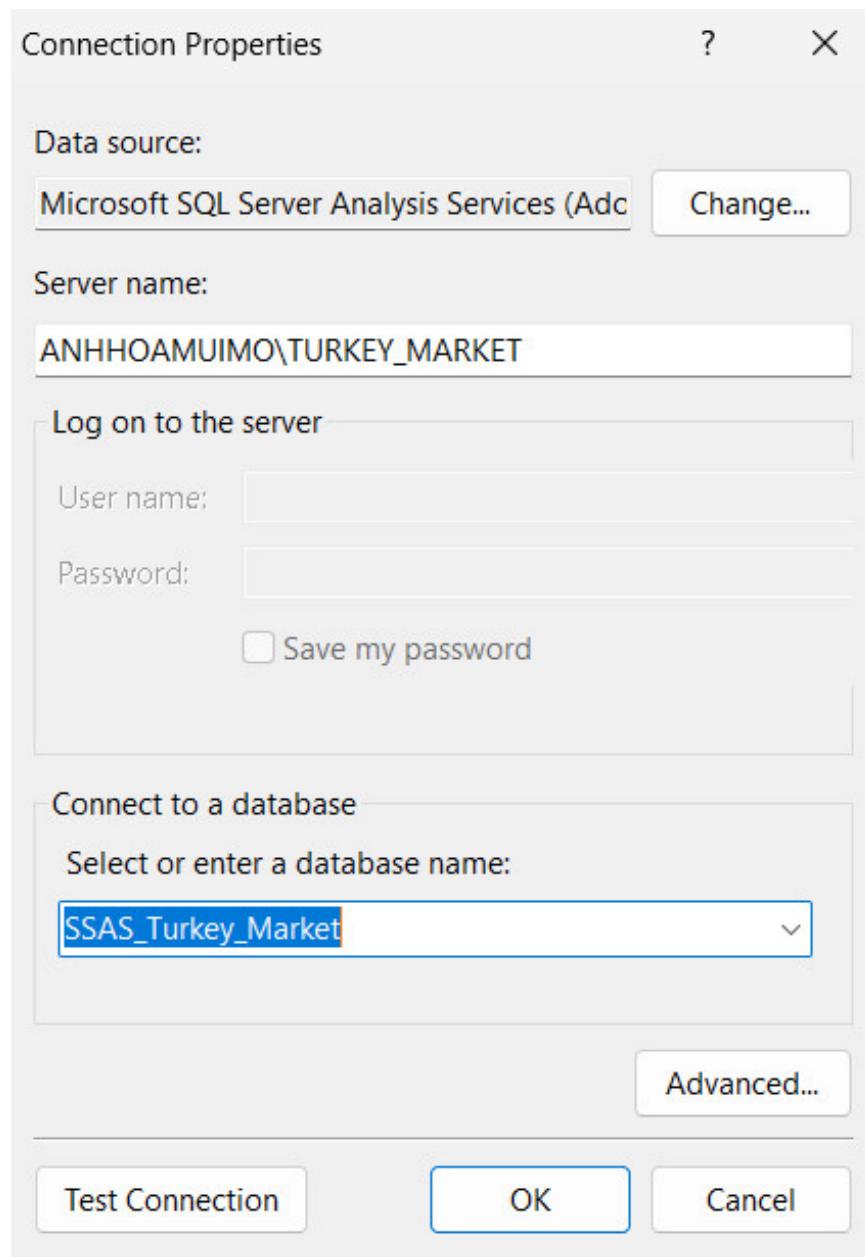
Hình 498: Thiết lập nơi lưu trữ và đặt tên cho dự án SSRS

**Bước 4:** Thêm data source cho Report.



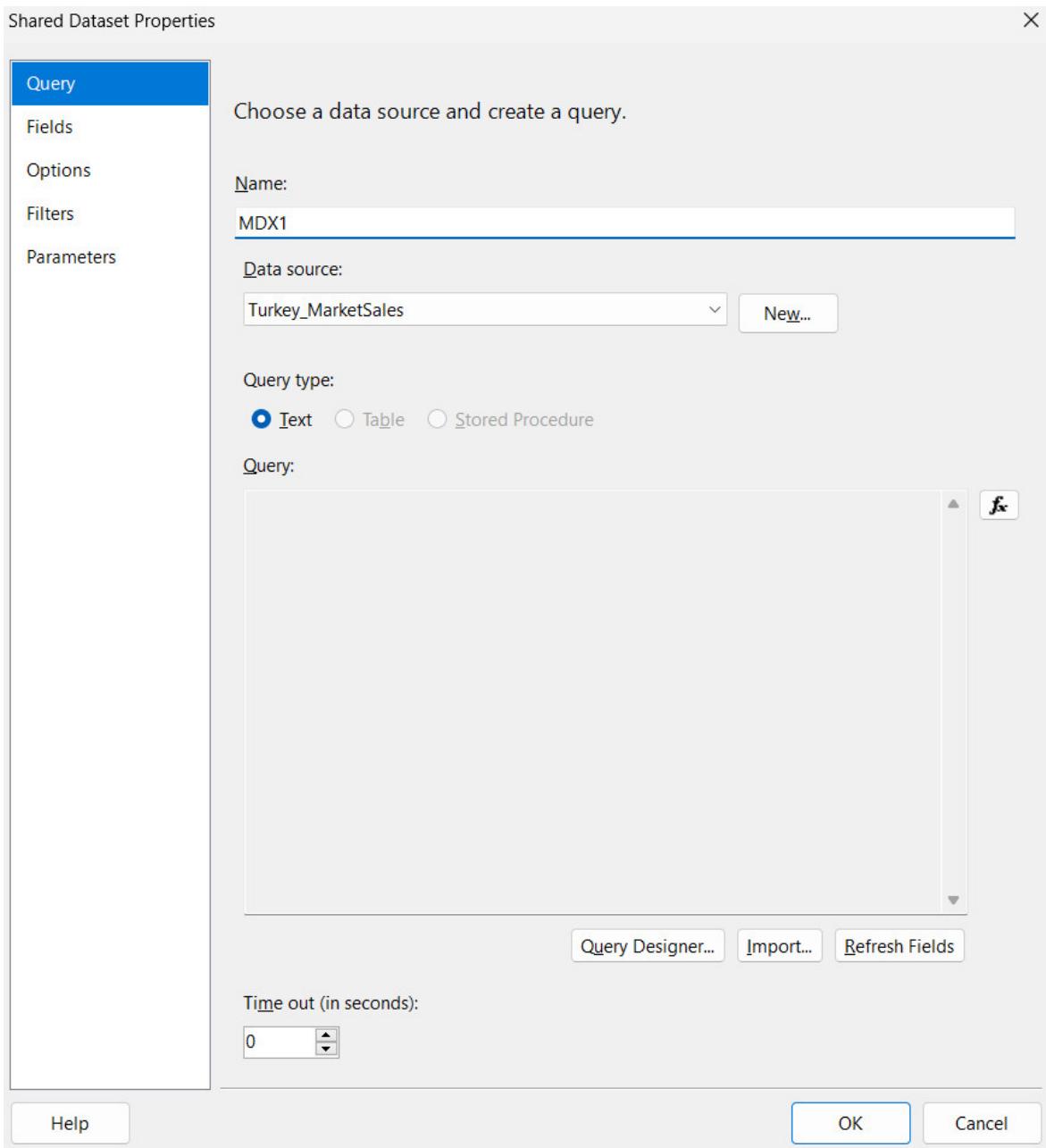
Hình 499: Thêm nguồn dữ liệu cho Report

**Bước 5:** Click “Build” để tạo mới một Connection String. Điền tên Server Name và chọn Database, sau đó click “OK”.



Hình 500: Giao diện tạo một kết nối mới đến kho dữ liệu

**Bước 6:** Tạo Dataset mới từ Data Source vừa tạo ra. Click “Query Designer...”

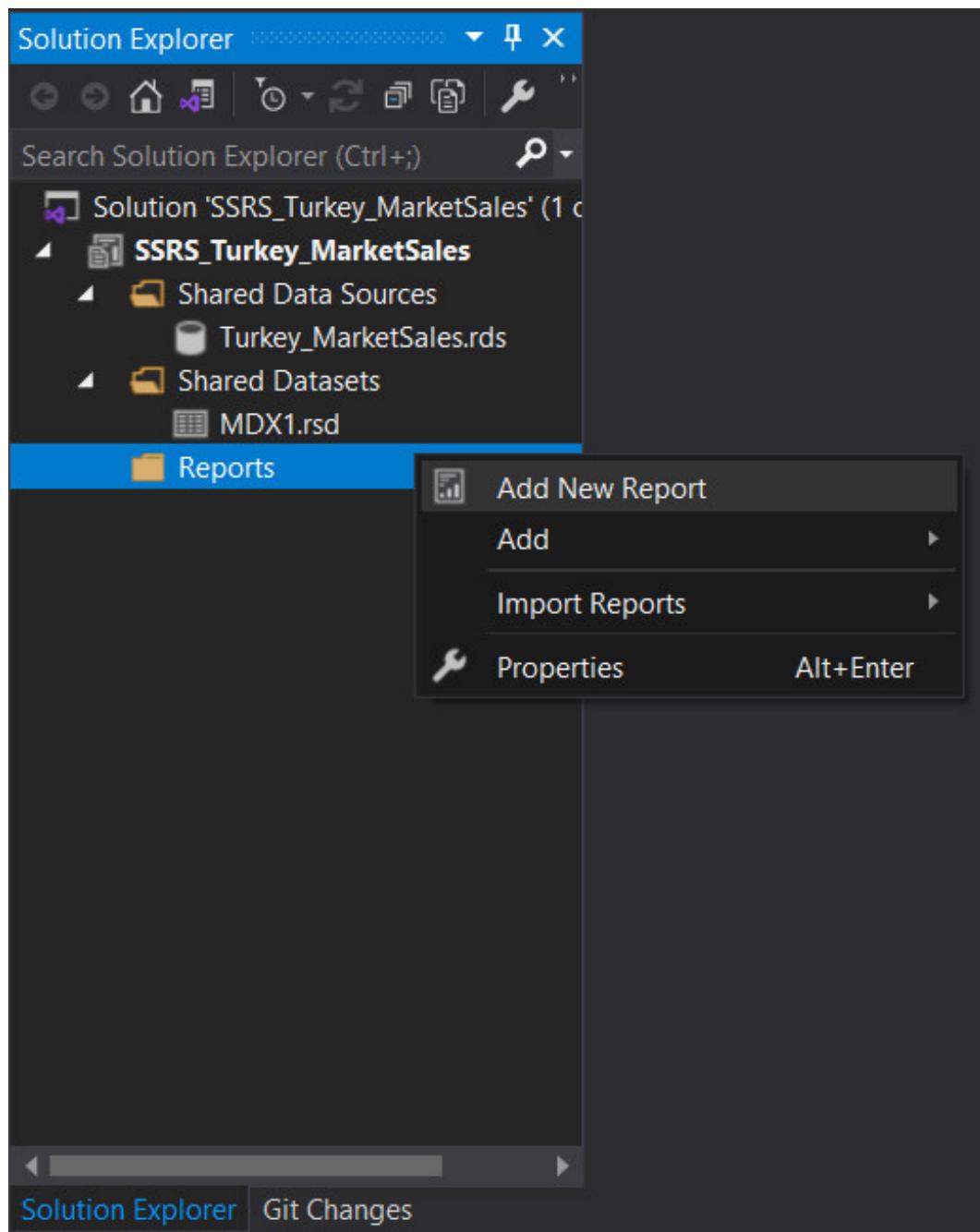


Hình 501: Tạo mới một Dataset có tên là MDX1

### 3.14.2. Tạo report bảng với Grouping với Totals:

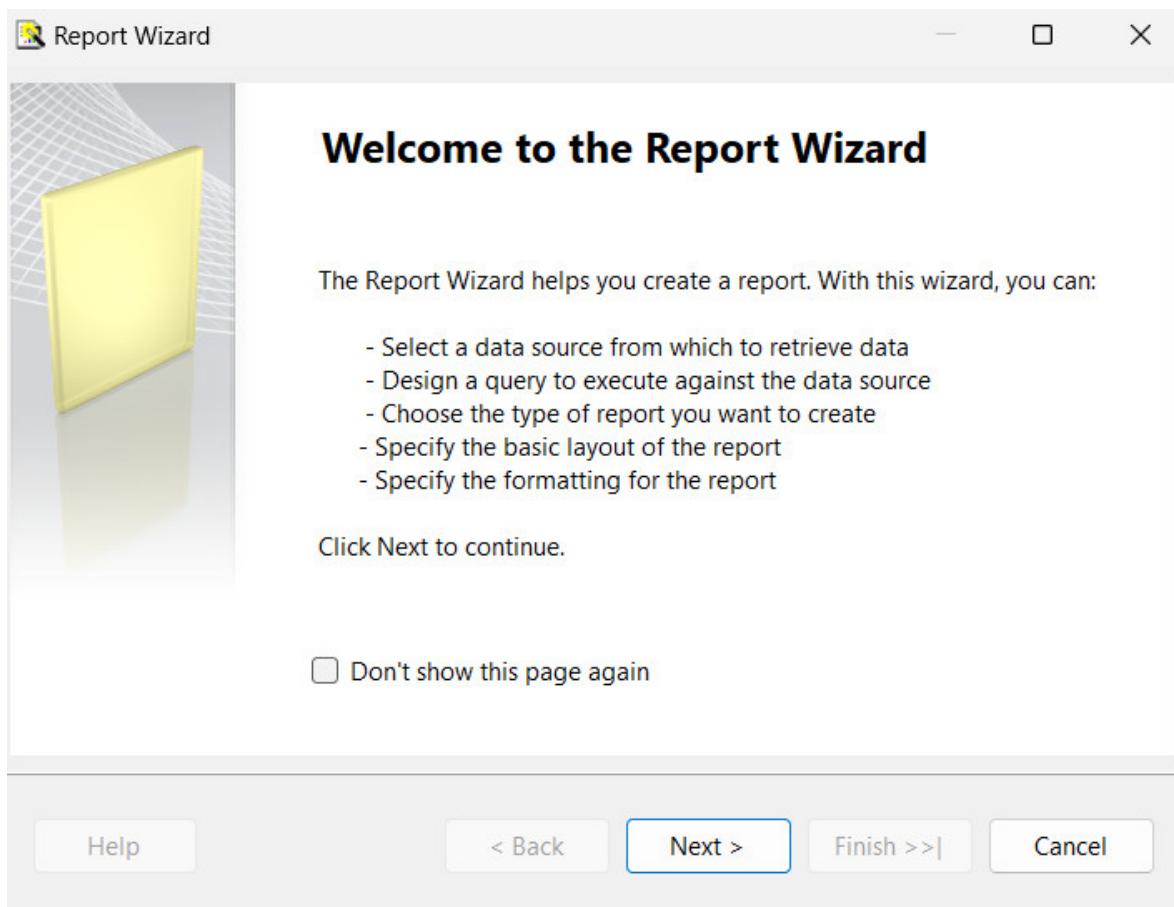
#### 3.14.2.1. Câu truy vấn 1:

Bước 1: Thêm bảng vào Report. Tại “Solution Explorer”, click chuột phải vào “Report”, chọn “Add new Report”.



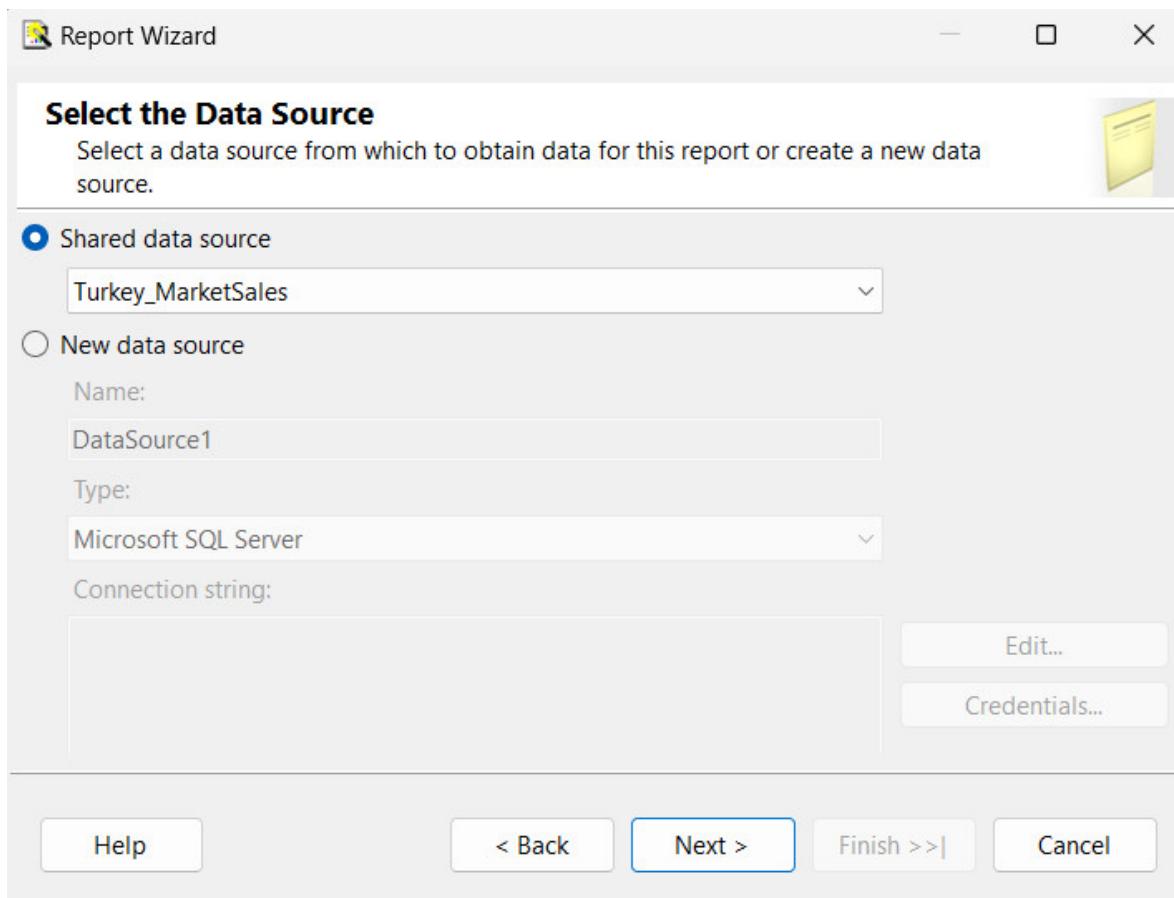
Hình 502: Giao diện tạo mới một Report

Bước 2: Click “Next”.



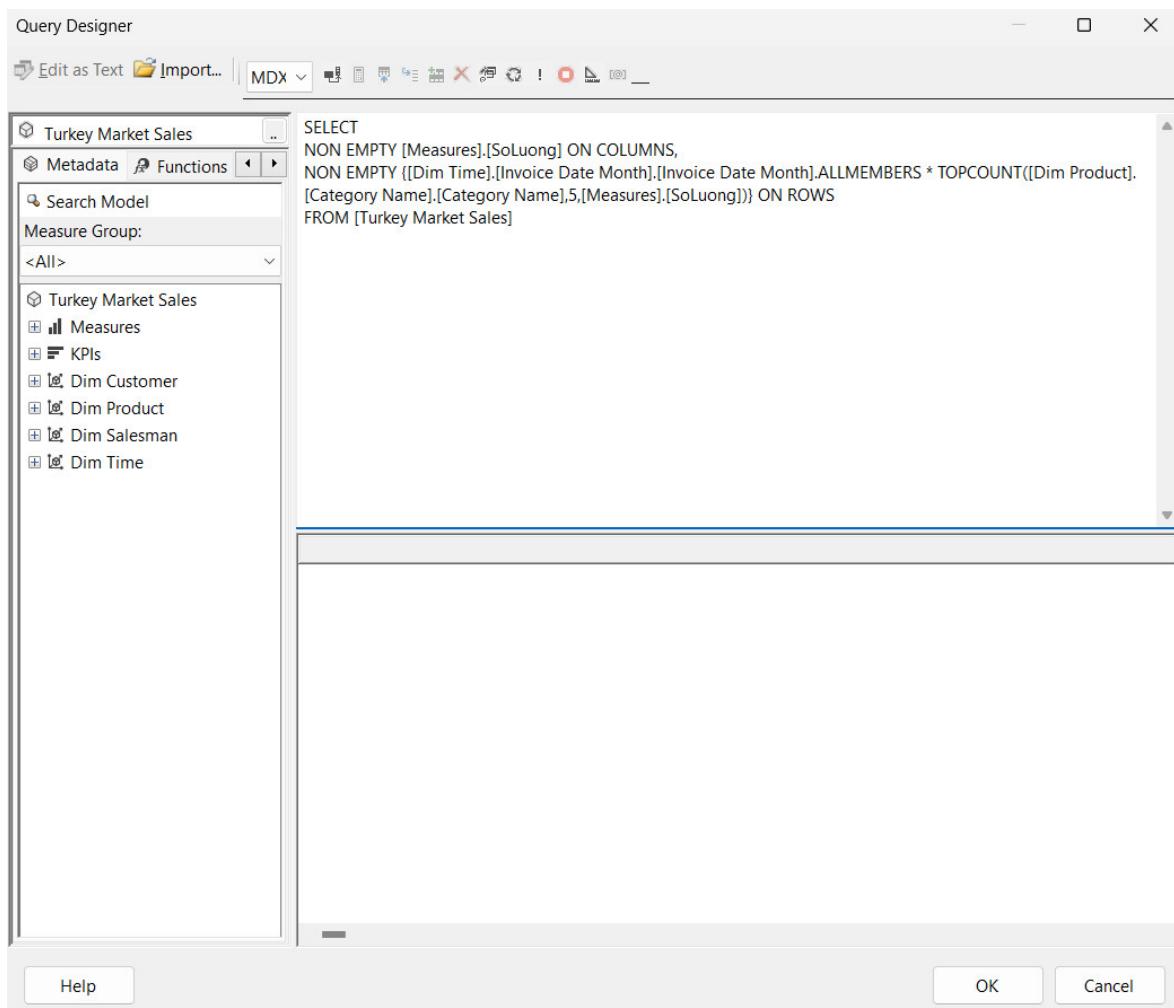
Hình 503: Hộp thoại Report Wizard xuất hiện

**Bước 3:** Tiếp tục click “Next”.



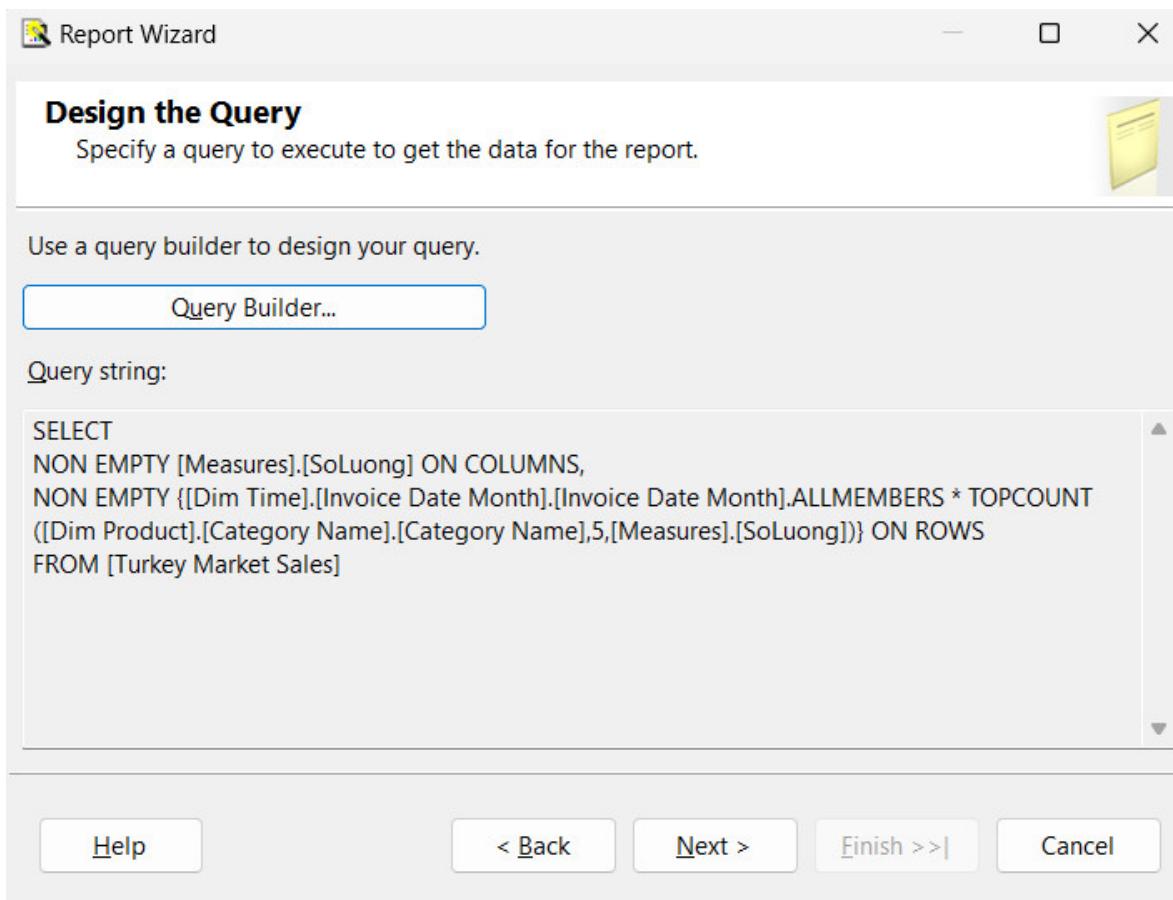
Hình 504: Giao diện chọn nguồn dữ liệu cho Report

Bước 4: Dán câu truy vấn MDX số 1 vào “Query Designer”.



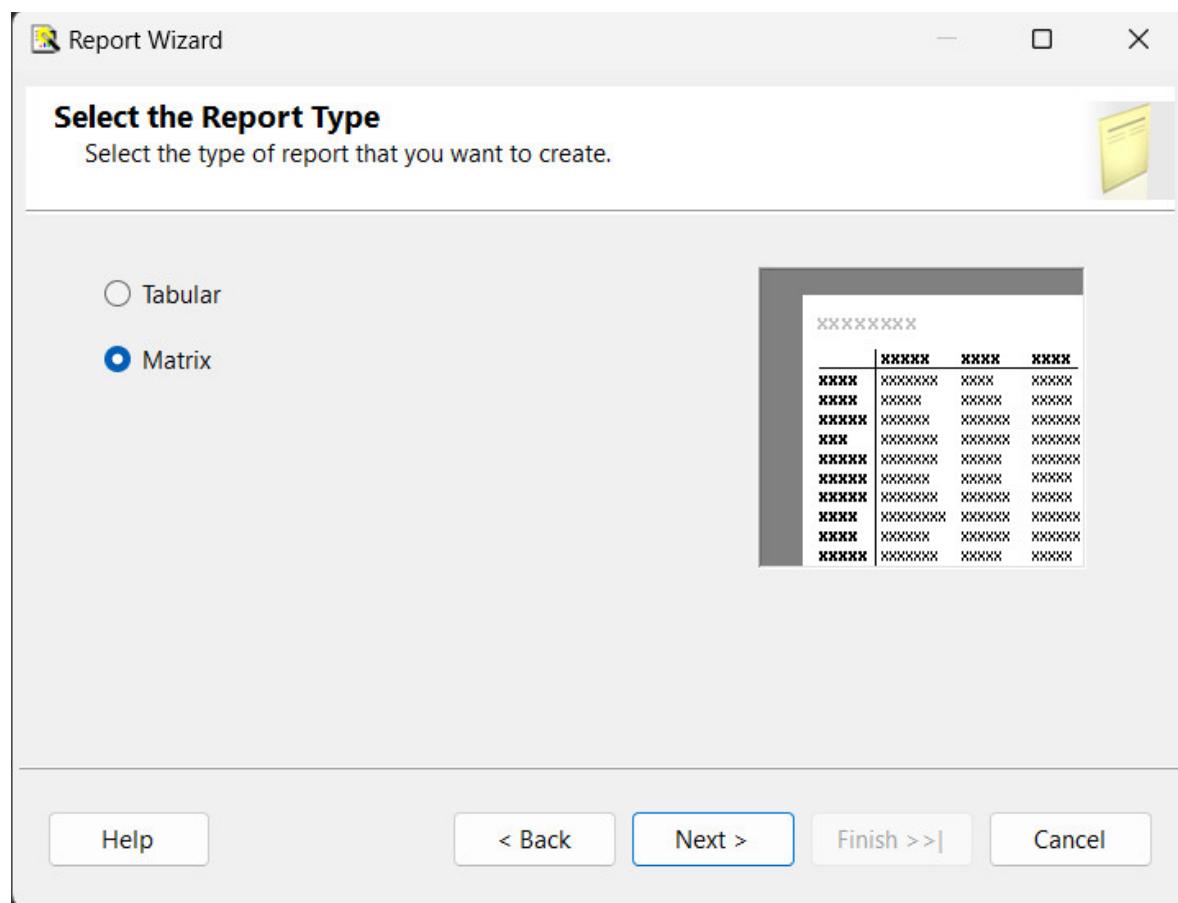
Hình 505: Dán câu truy vấn MDX của câu truy vấn 1 vào ô “Query Designer”

**Bước 5:** Click “Next”.



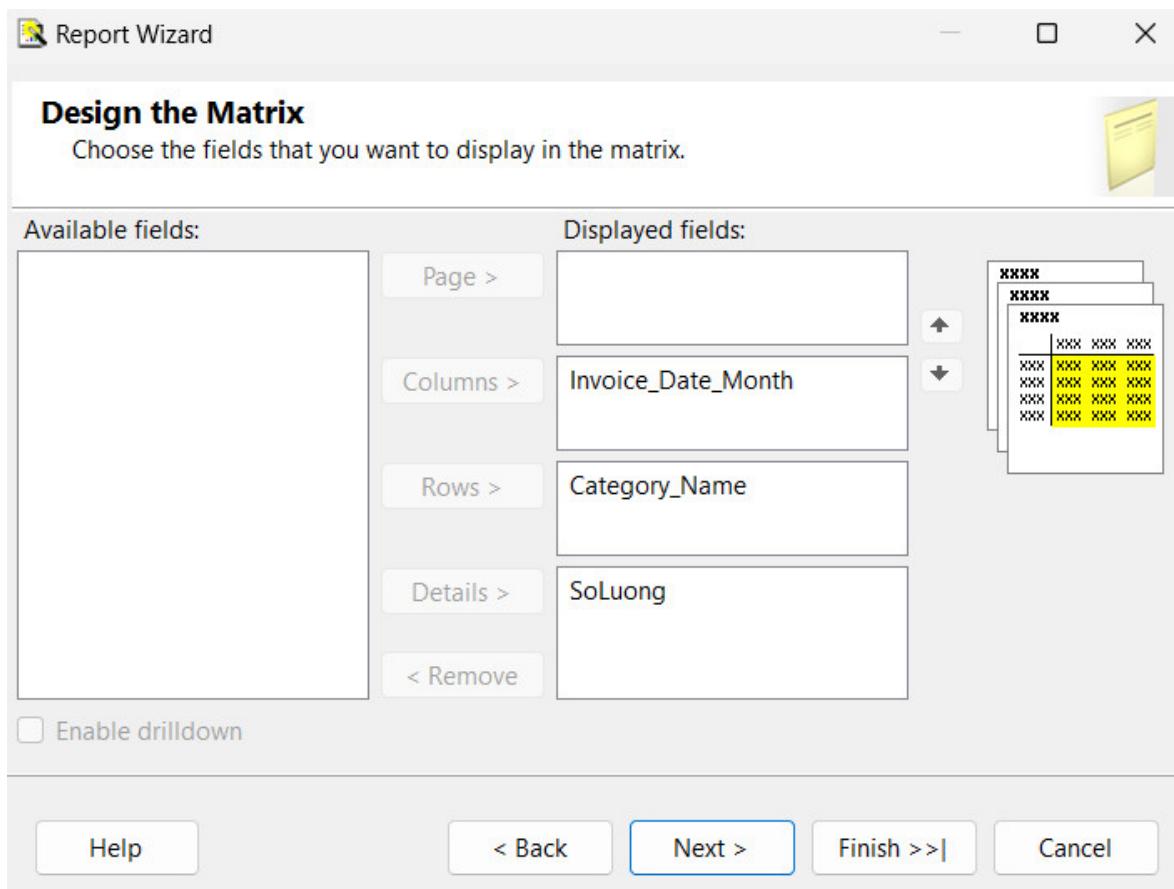
Hình 506: Kiểm tra lại câu truy vấn

Bước 6: Chọn “Matrix” và click “Next”.



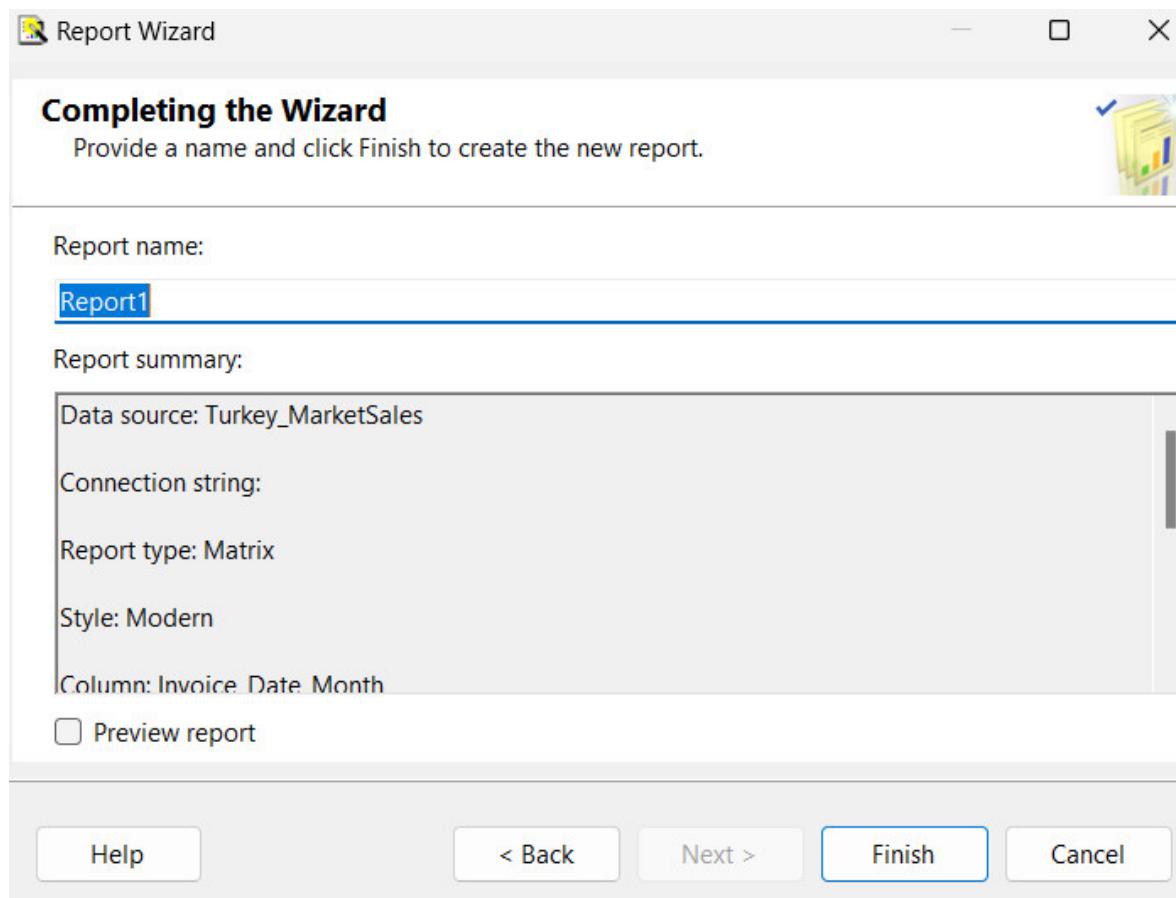
Hình 507: Giao diện chọn loại Report

**Bước 7:** Chọn thuộc tính invoice\_date\_month làm cột, category\_name làm hàng và SoLuong làm chi tiết cho ma trận.



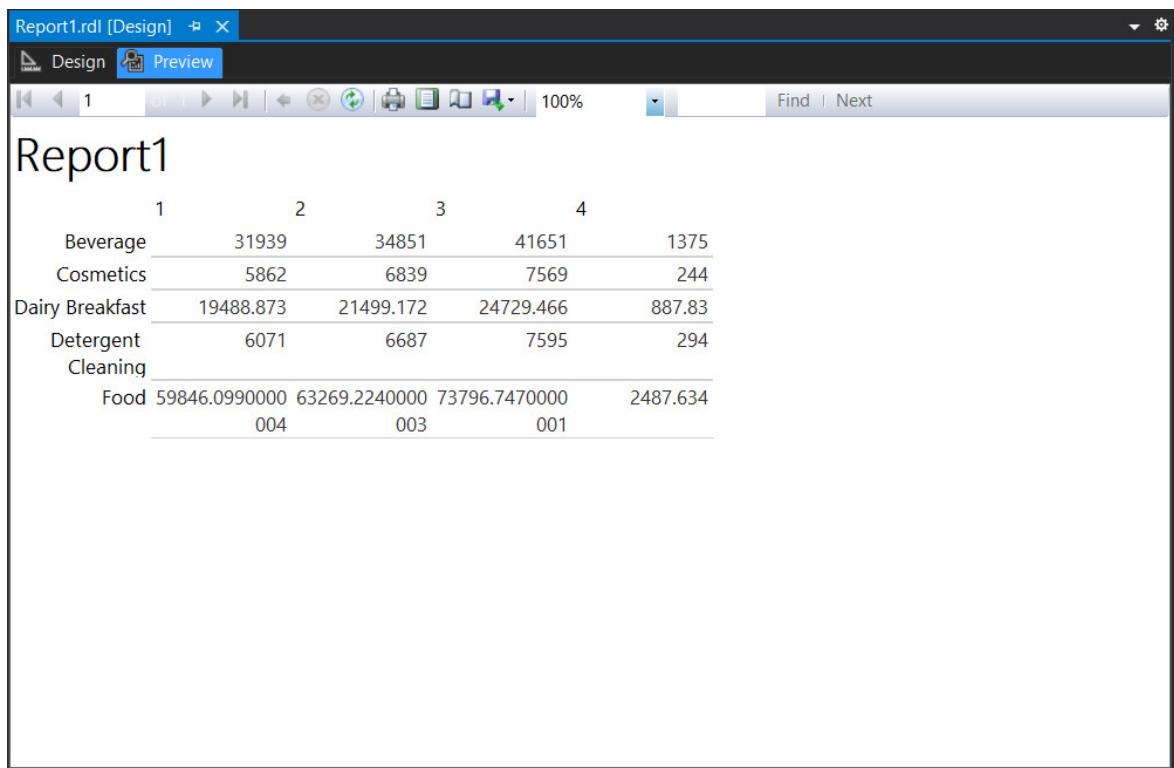
Hình 508: Chuyển các thuộc tính thích hợp vào mục “Columns”, “Rows” và “Details”

**Bước 8:** Sau đó click “Finish >>” để hoàn tất quá trình.



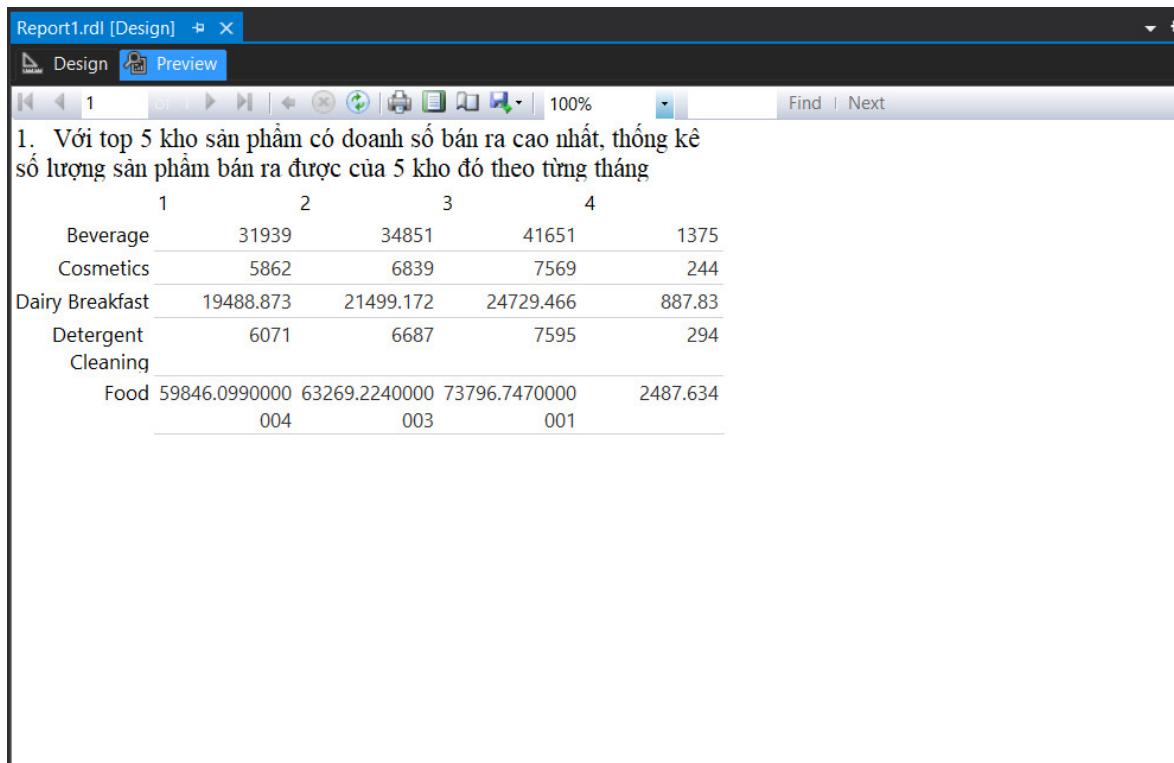
Hình 509: Giao diện đặt tên cho Report

**Bước 9:** Click “Preview” để kiểm tra lại tính đúng đắn của dữ liệu.



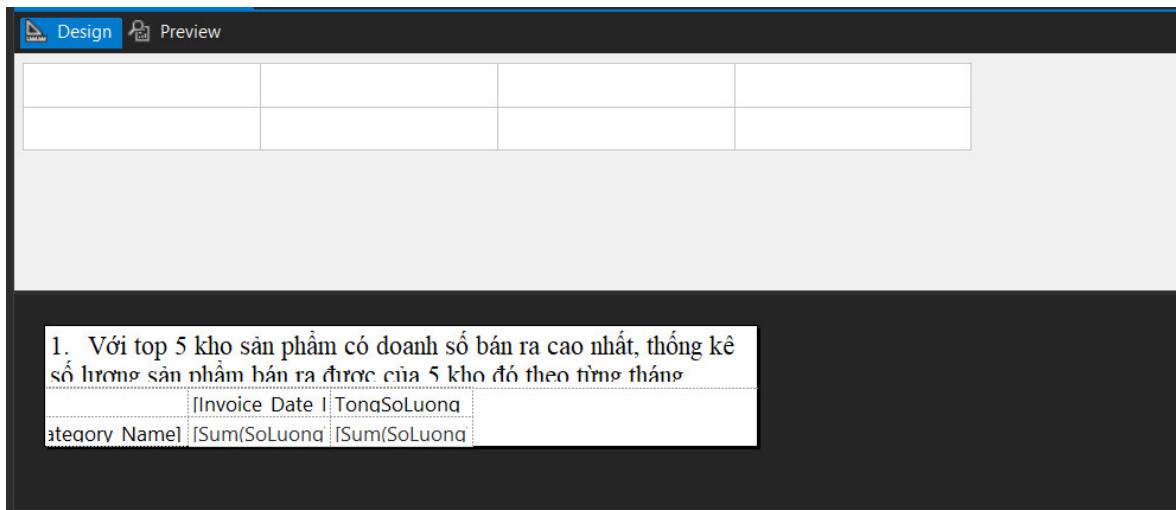
Hình 510: Kết quả câu truy vấn 1 hiển thị dưới dạng ma trận

**Bước 10:** Đổi tên Report thành nội dung câu truy vấn 1.



Hình 511: Kết quả đổi tên Report

**Bước 11:** Click chuột phải vào cột invoice\_date\_month, chọn “Add totals”, đặt tên cho Totals đó là TongSoLuong.



Hình 512: Tiến hành tạo thêm Totals có tên là TongSoLuong

Report1.rdl [Design] X

Design Preview

100% Find | Next

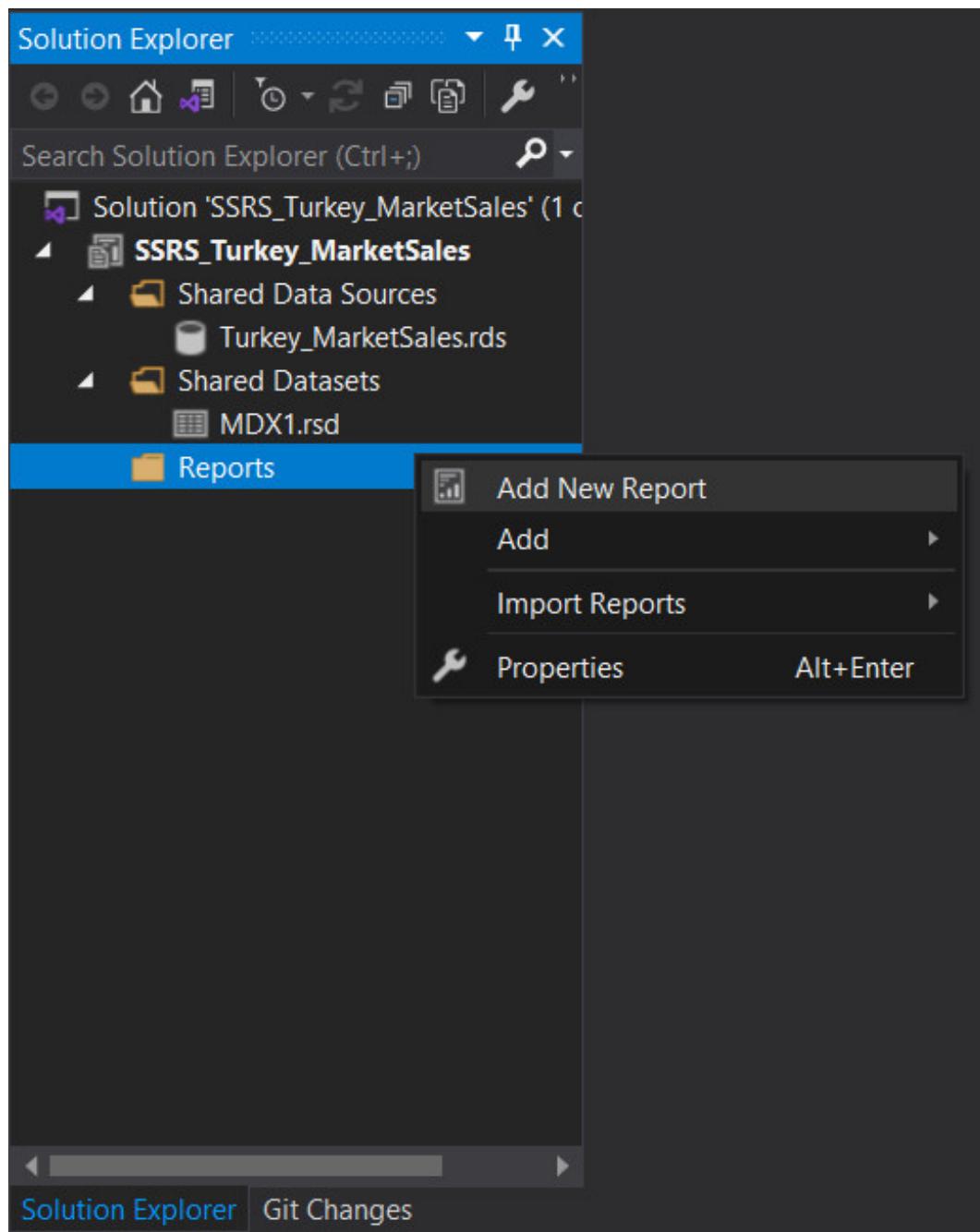
1. Với top 5 kho sản phẩm có doanh số bán ra cao nhất, thống kê số lượng sản phẩm bán ra được của 5 kho đó theo từng tháng

	1	2	3	4	TongSoLuong
Beverage	31939	34851	41651	1375	109816
Cosmetics	5862	6839	7569	244	20514
Dairy Breakfast	19488.873	21499.172	24729.466	887.83	66605.341
Detergent	6071	6687	7595	294	20647
Cleaning					
Food	59846.0990000	63269.2240000	73796.7470000	2487.634	199399.704000
	004	003	001		001

Hình 513: Report câu truy vấn 1

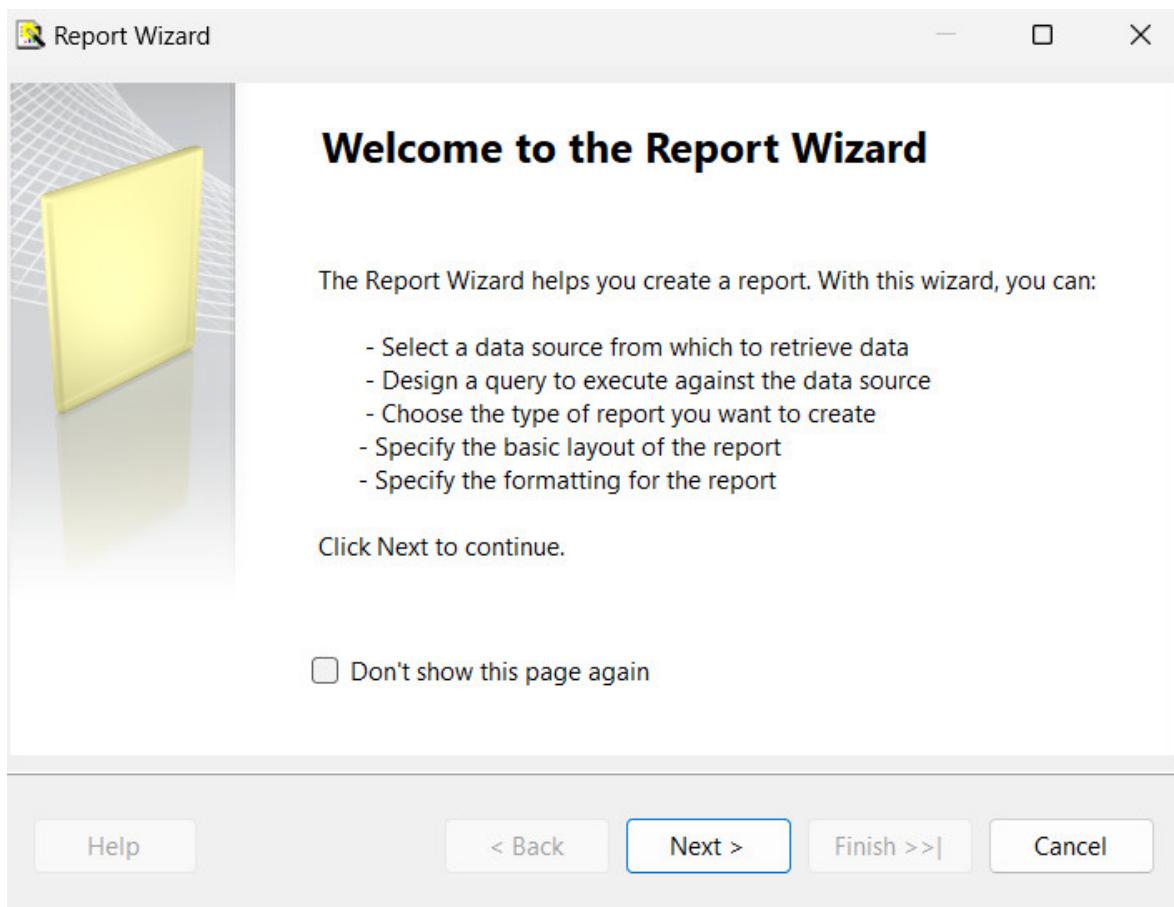
### 3.14.2.2. Câu truy vấn 2:

**Bước 1:** Thêm bảng vào Report. Tại “Solution Explorer”, click chuột phải vào “Report”, chọn “Add new Report”.



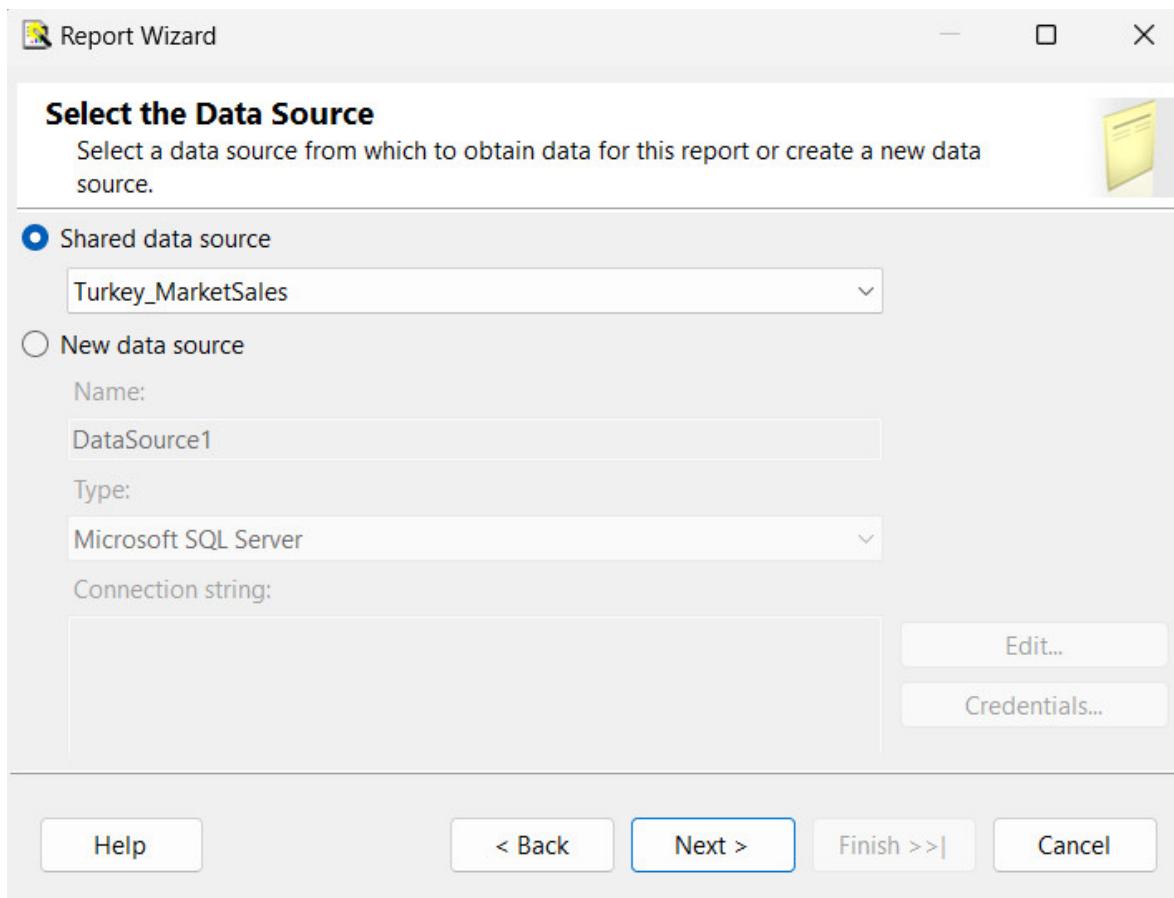
Hình 514: Giao diện tạo mới một Report

Bước 2: Click “Next”.



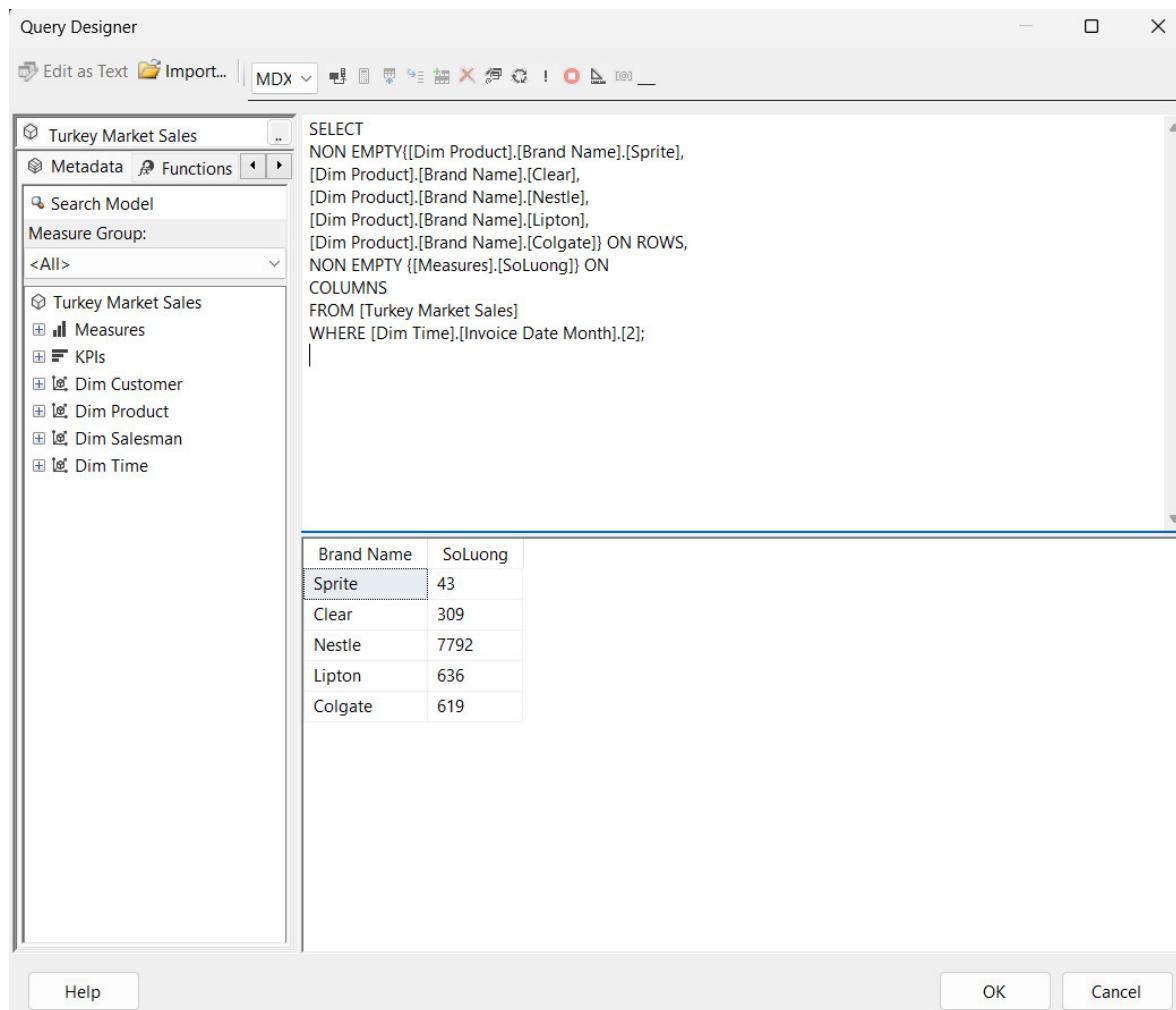
Hình 515: Hộp thoại Report Wizard xuất hiện

**Bước 3:** Tiếp tục click “Next”.



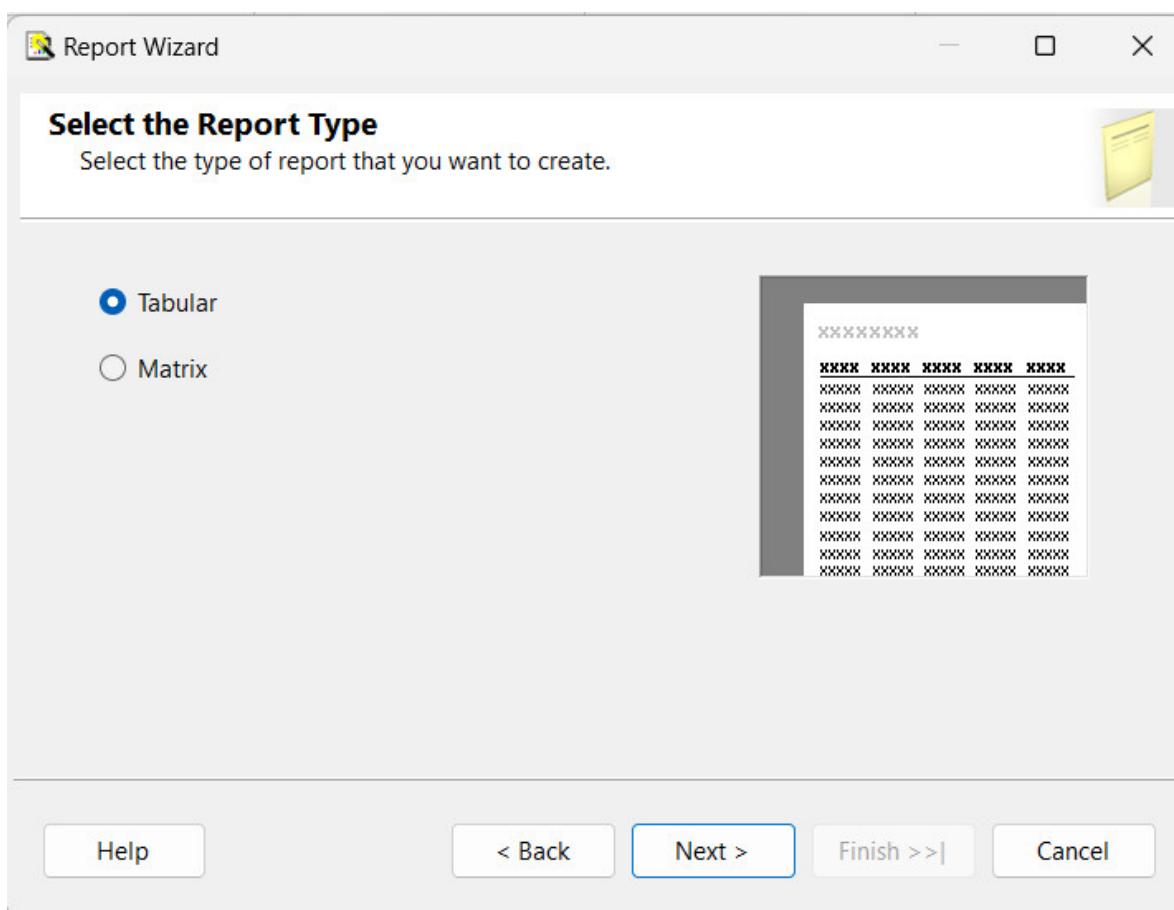
Hình 516: Giao diện chọn nguồn dữ liệu cho Report

Bước 4: Dán câu truy vấn MDX số 2 vào “Query Designer”.



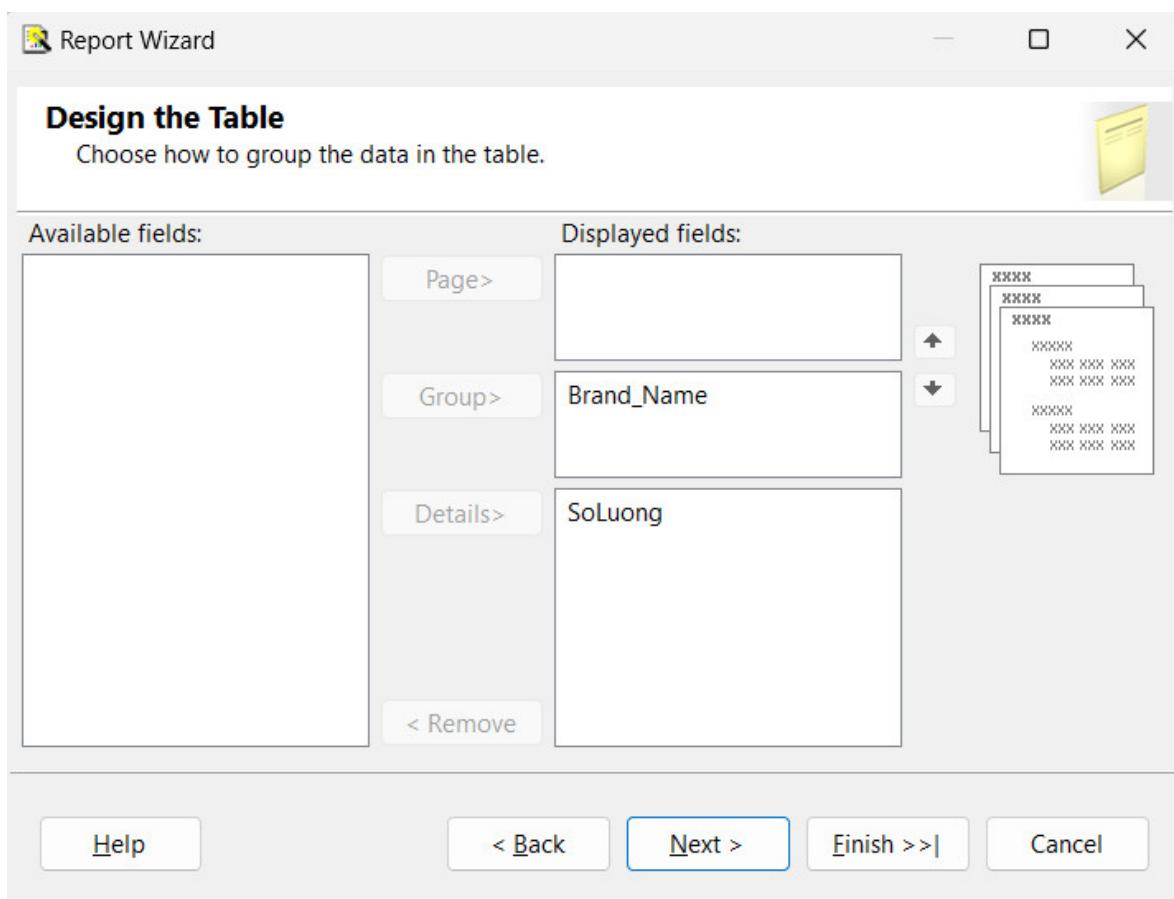
Hình 517: Dán câu truy vấn MDX của câu truy vấn số 2 vào ô “Query Designer”

**Bước 5:** Chọn “Tabular” và click “Next”.



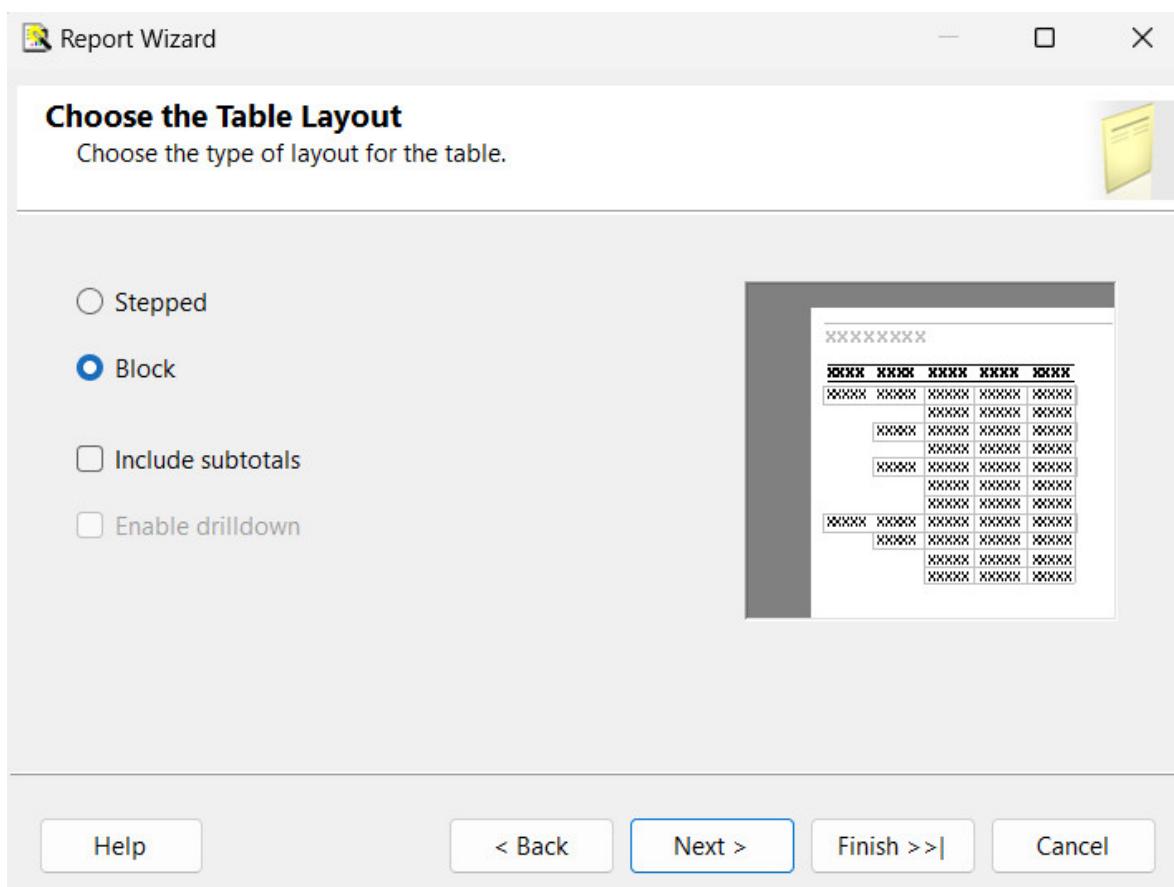
Hình 518: Giao diện chọn loại Report

**Bước 6:** Chọn thuộc tính brand\_name vào “**Group**”, SoLuong vào “**Detail**” cho bảng.



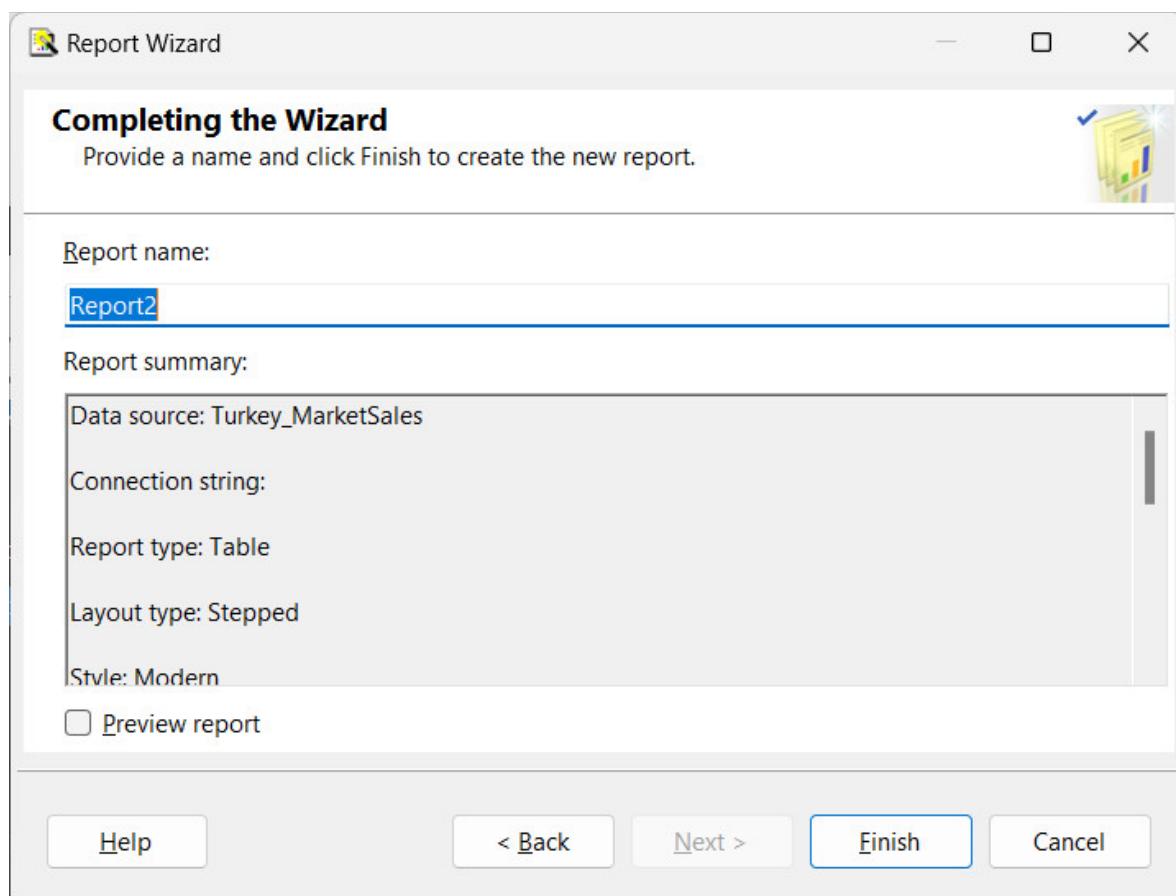
Hình 519: Chuyển các thuộc tính thích hợp vào mục “Group” và “Details”

**Bước 7:** Sau đó chọn “Block”, click “Next”.



Hình 520: Giao diện chọn Table Layout

**Bước 8:** Click “Finish >>” để hoàn tất quá trình.



Hình 521: Giao diện đặt tên cho Report

Bước 9: Click “Preview” để kiểm tra lại tính đúng đắn của dữ liệu.

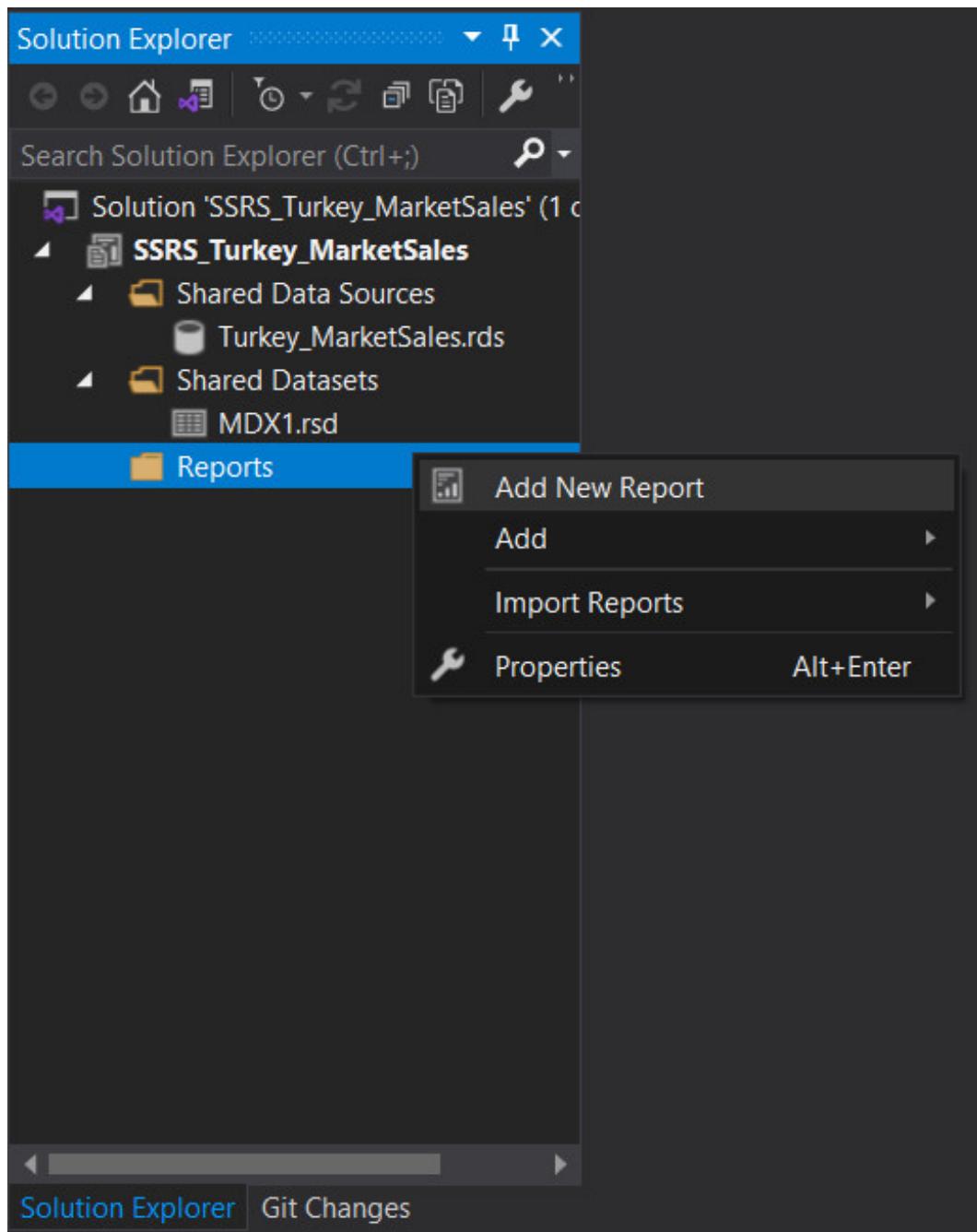
The screenshot shows the 'Report2.rdl [Design]' window with a preview tab selected. The preview area displays a table with the following data:

Brand Name	So Luong
Clear	309
Colgate	619
Lipton	636
Nestle	7792
Sprite	43

Hình 522: Report câu truy vấn 2

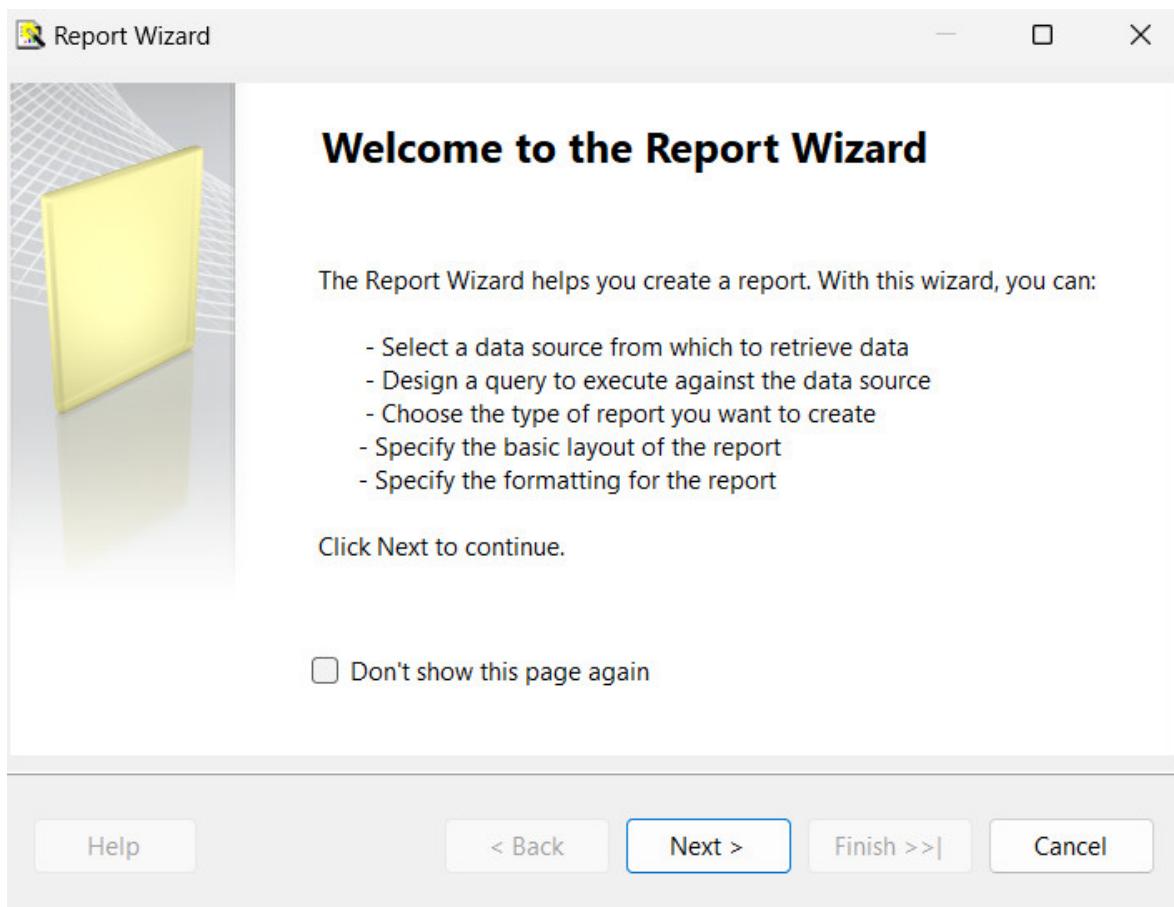
### 3.14.2.3. Câu truy vấn 3:

Bước 1: Thêm bảng vào Report. Tại “Solution Explorer”, click chuột phải vào “Report”, chọn “Add new Report”.



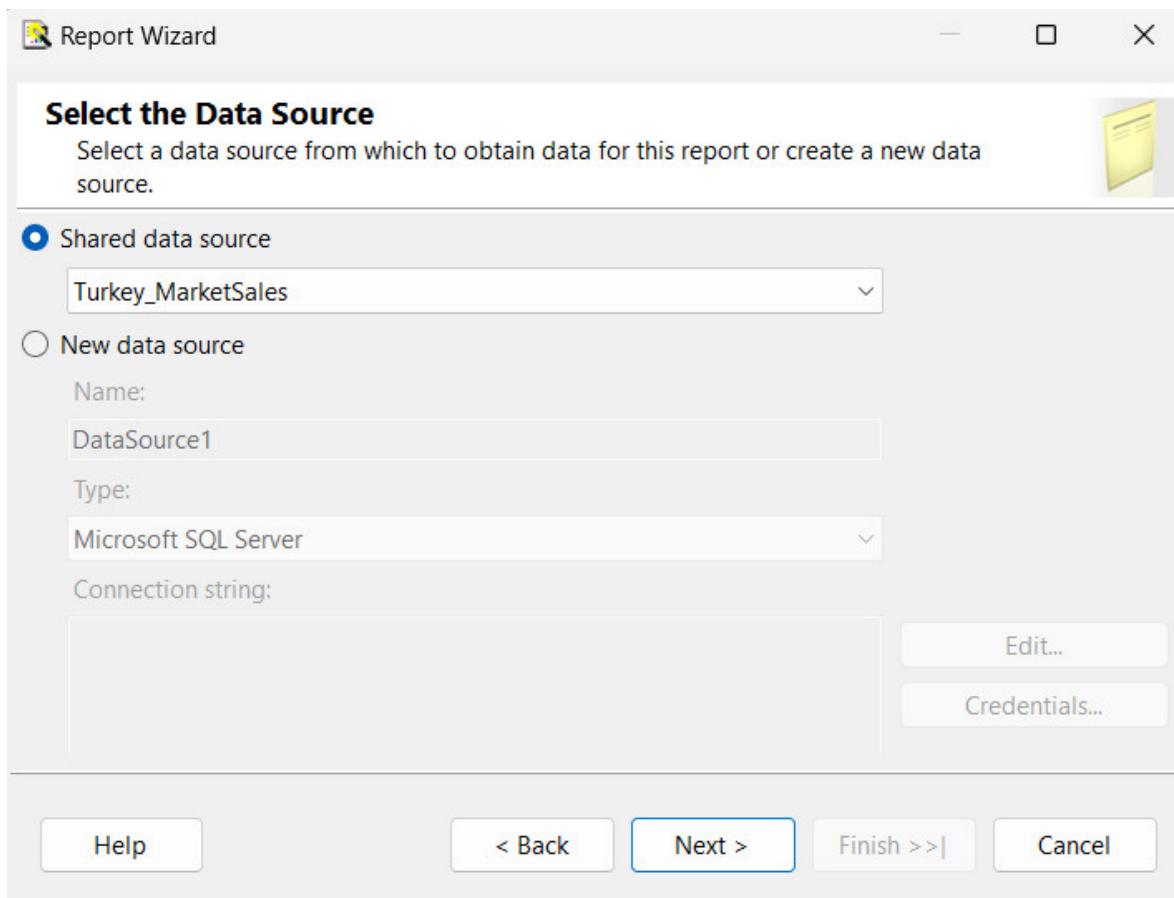
Hình 523: Giao diện tạo mới một Report

Bước 2: Click “Next”.



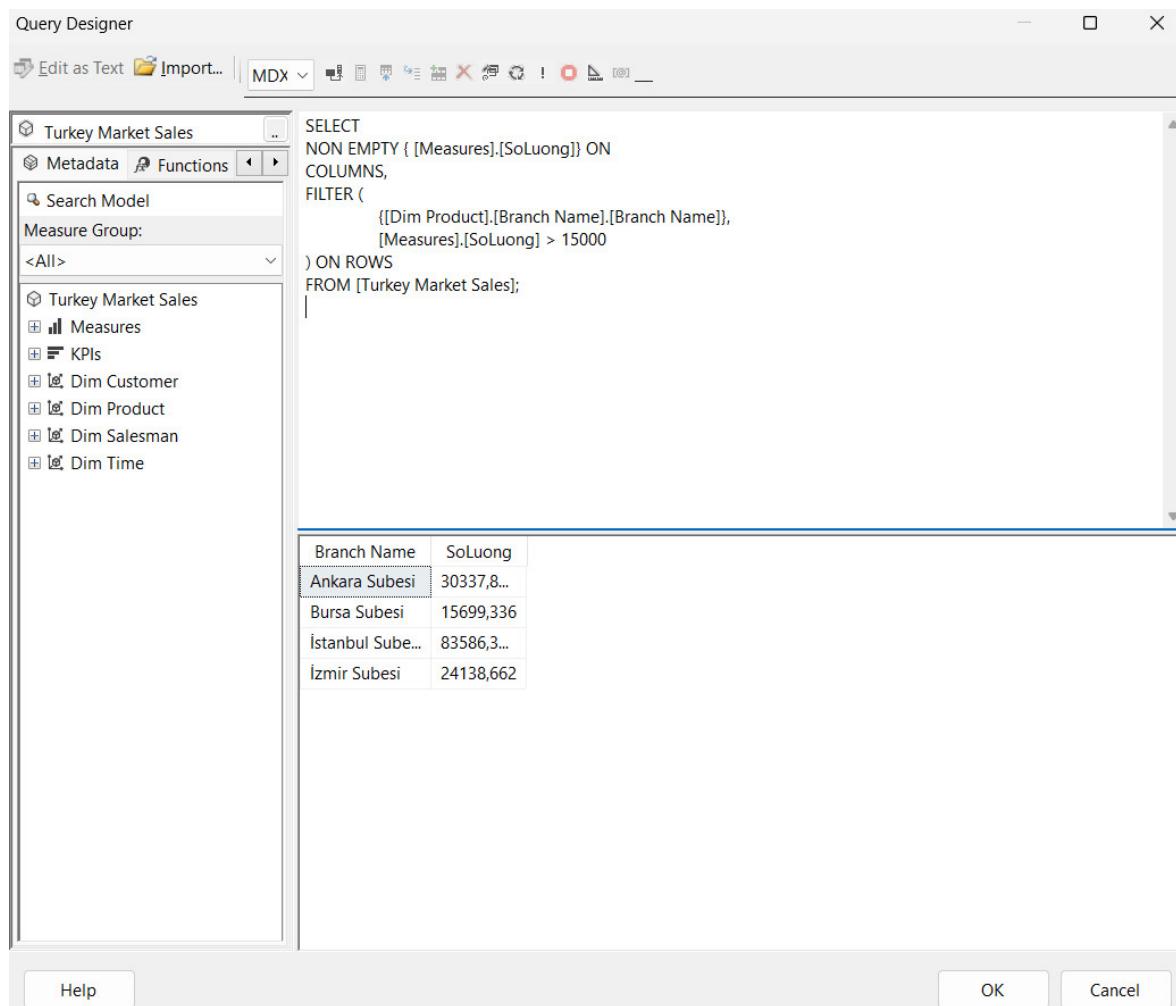
Hình 524: Hộp thoại Report Wizard xuất hiện

**Bước 3:** Tiếp tục click “Next”.



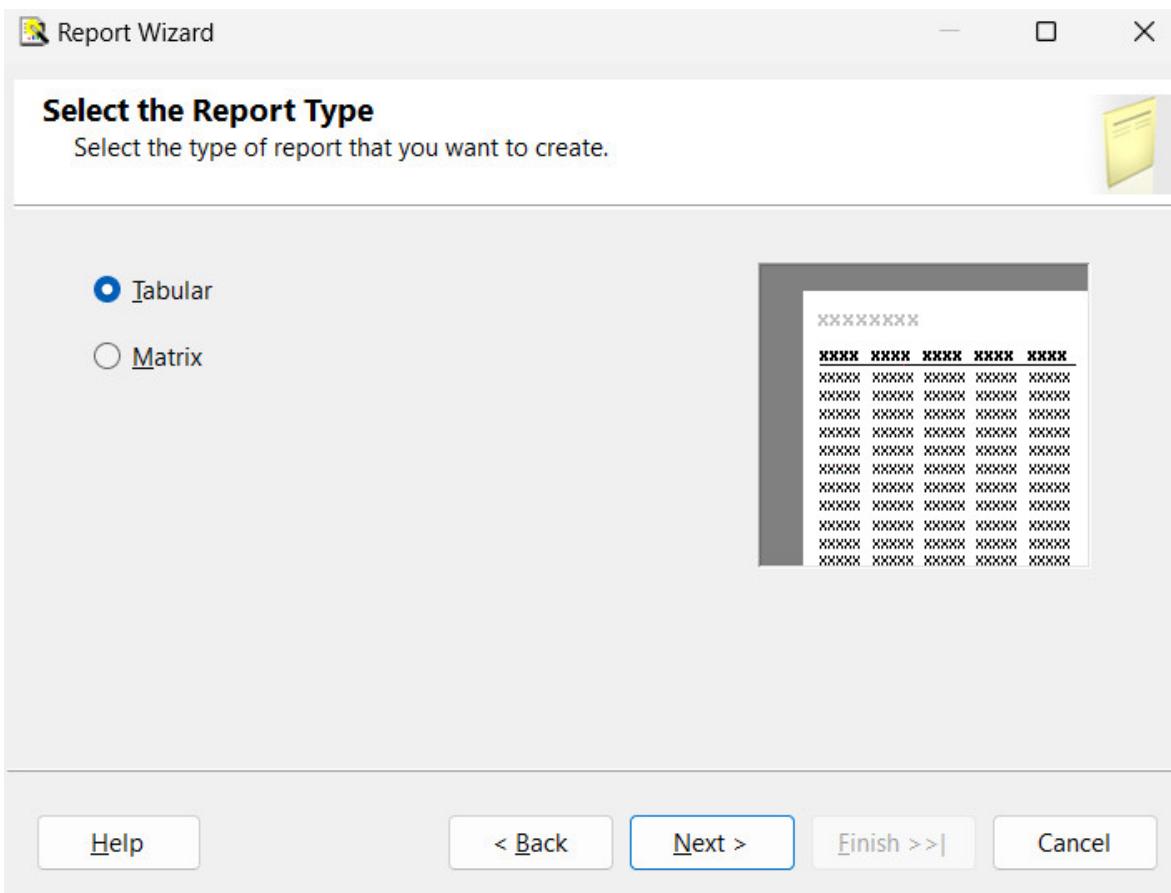
Hình 525: Giao diện chọn nguồn dữ liệu cho Report

Bước 4: Dán câu truy vấn MDX số 3 vào “Query Designer”.



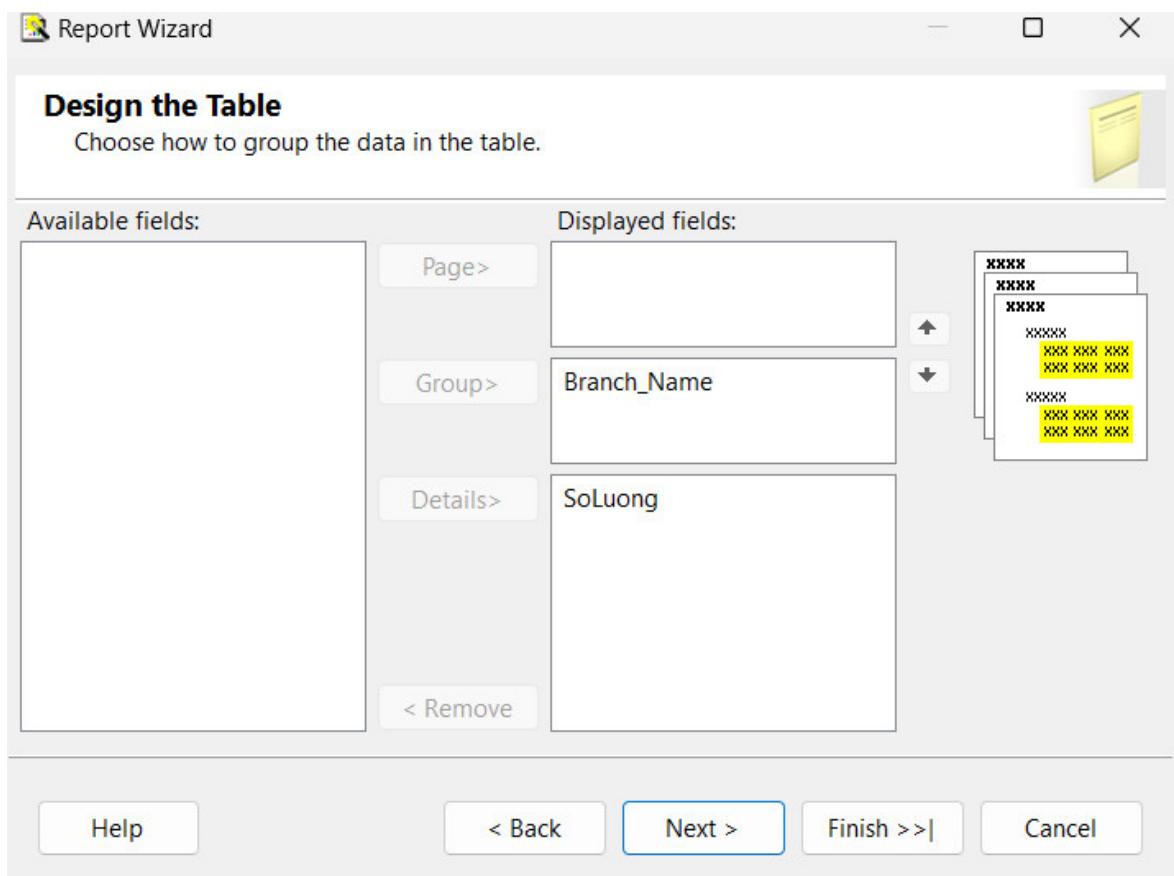
Hình 526: Dán câu truy vấn MDX của câu truy vấn 3 vào ô “Query Designer”

**Bước 5:** Chọn “Tabular” và click “Next”.



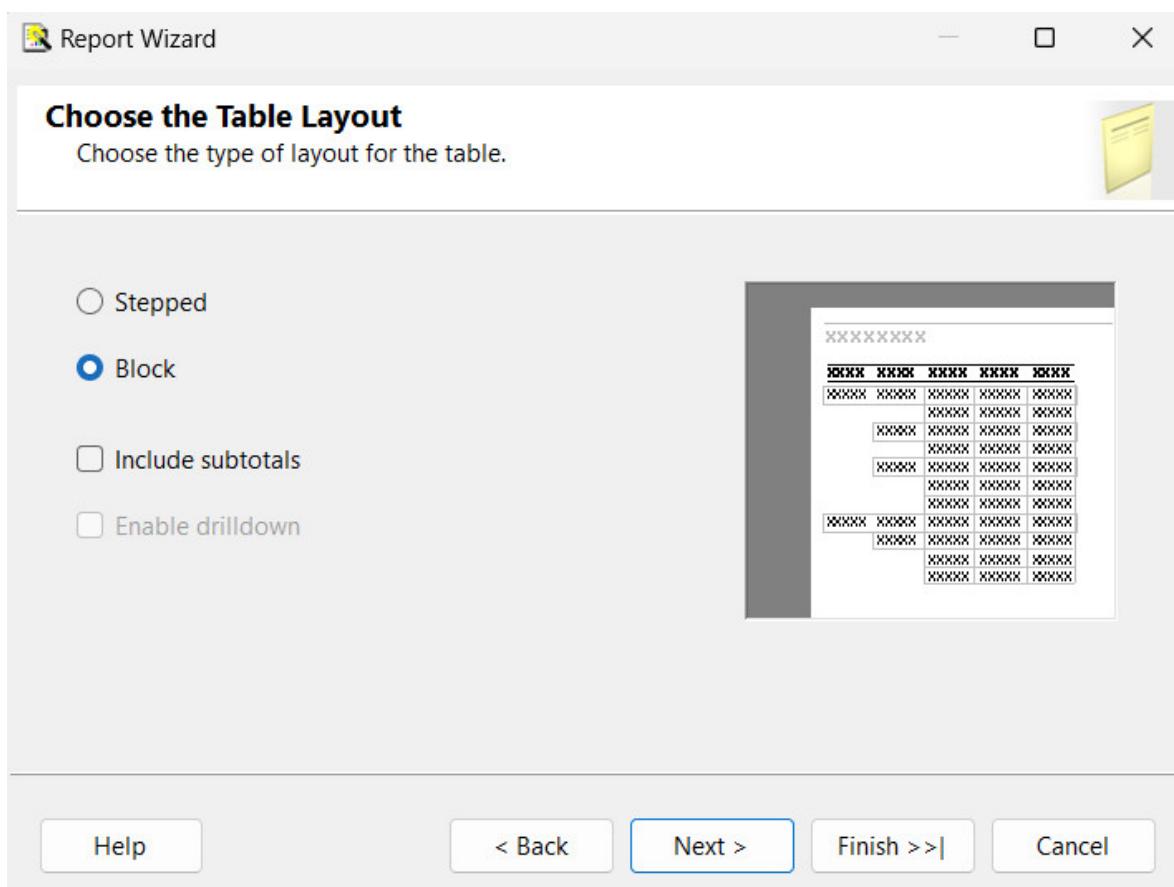
Hình 527: Giao diện chọn loại Report

**Bước 6:** Chọn thuộc tính branch\_name vào “**Group**”, SoLuong vào “**Detail**” cho bảng.



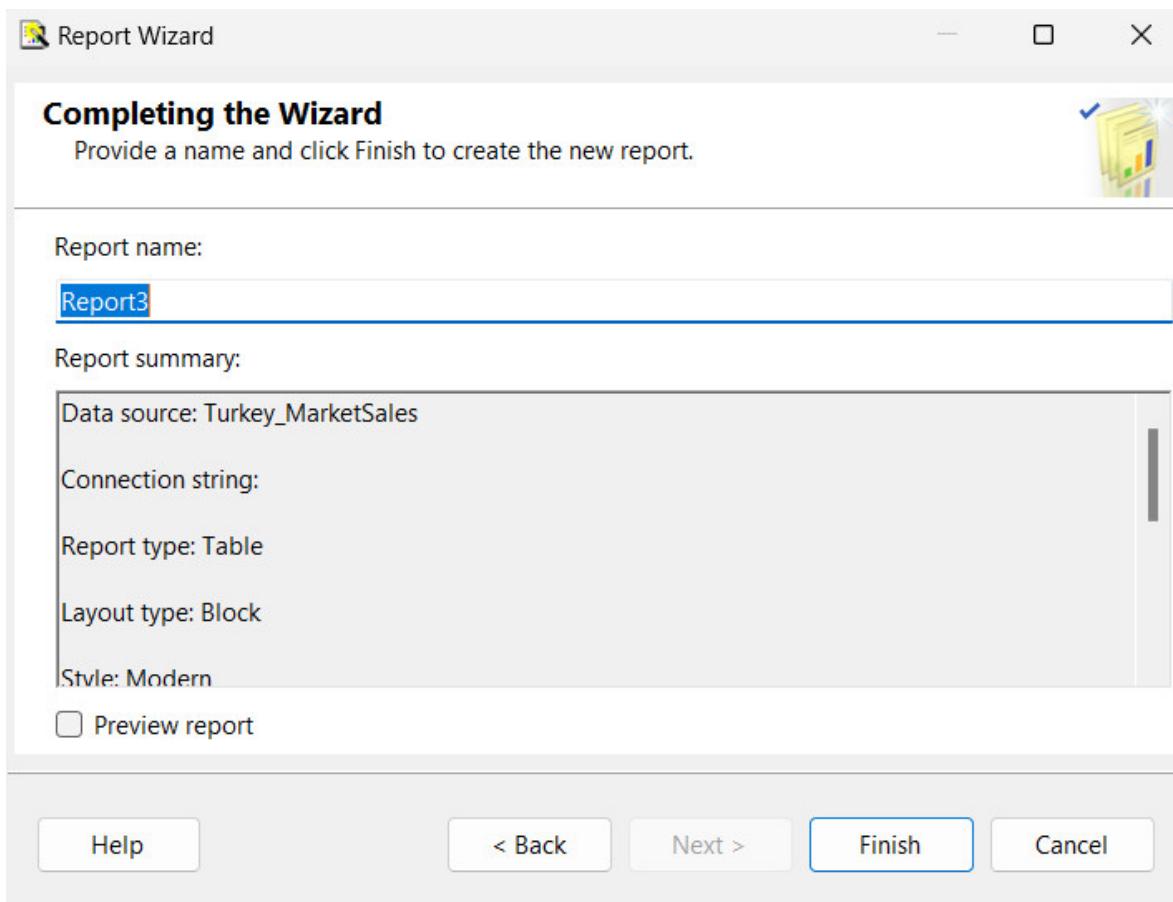
Hình 528: Chuyển các thuộc tính thích hợp vào mục “Group” và “Details”

**Bước 7:** Sau đó chọn “Block”, click “Next”.



Hình 529: Giao diện chọn Table Layout

**Bước 8:** Sau đó click “Finish >>” để hoàn tất quá trình.



Hình 530: Giao diện đặt tên cho Report

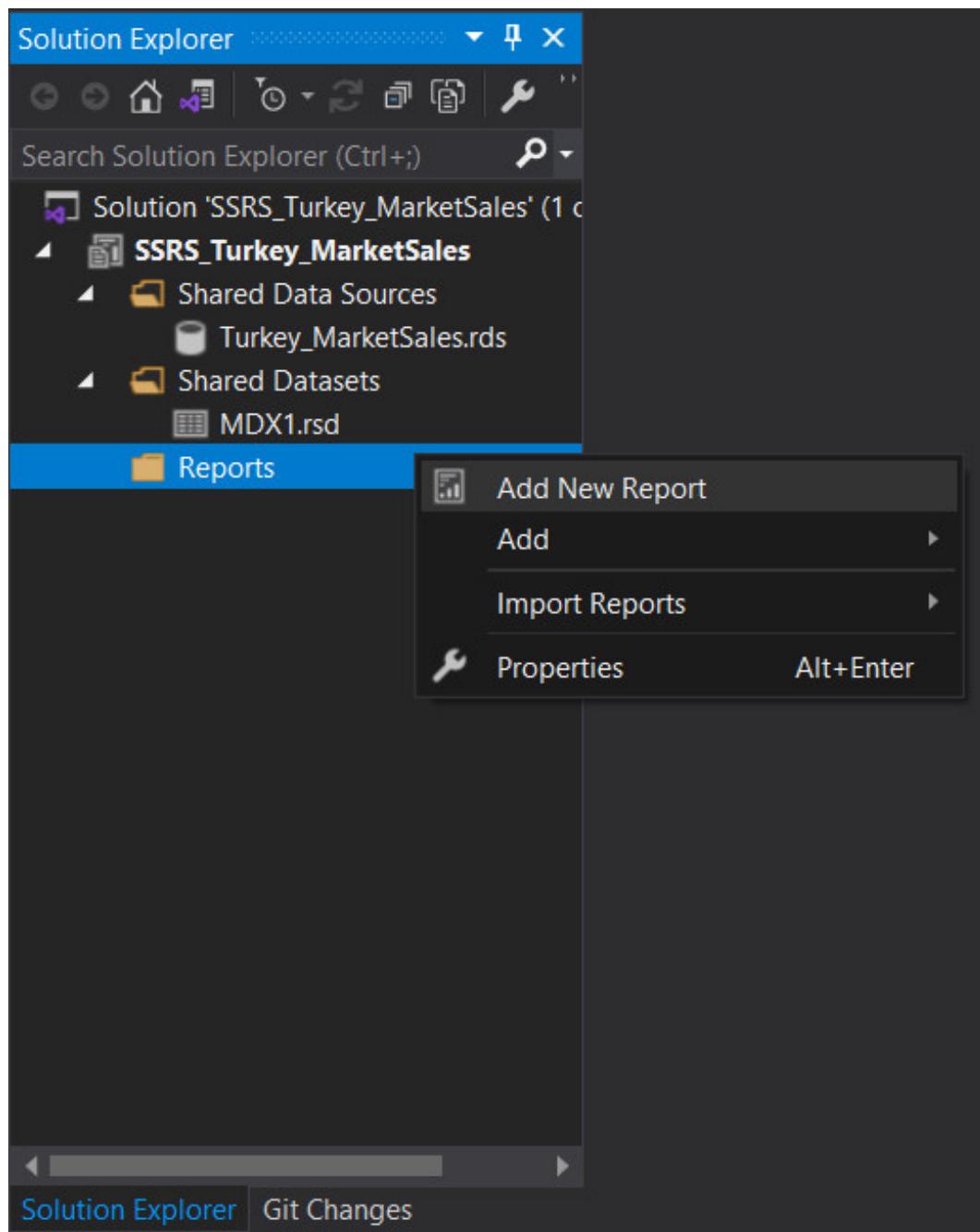
**Bước 9:** Click “Preview” để kiểm tra lại tính đúng đắn của dữ liệu.

Branch Name	So Luong
Ankara Subesi	30337.8490000 001
Bursa Subesi	15699.336
İstanbul Subesi	83586.3099999 998
İzmir Subesi	24138.662

Hình 531: Report câu truy vấn 3

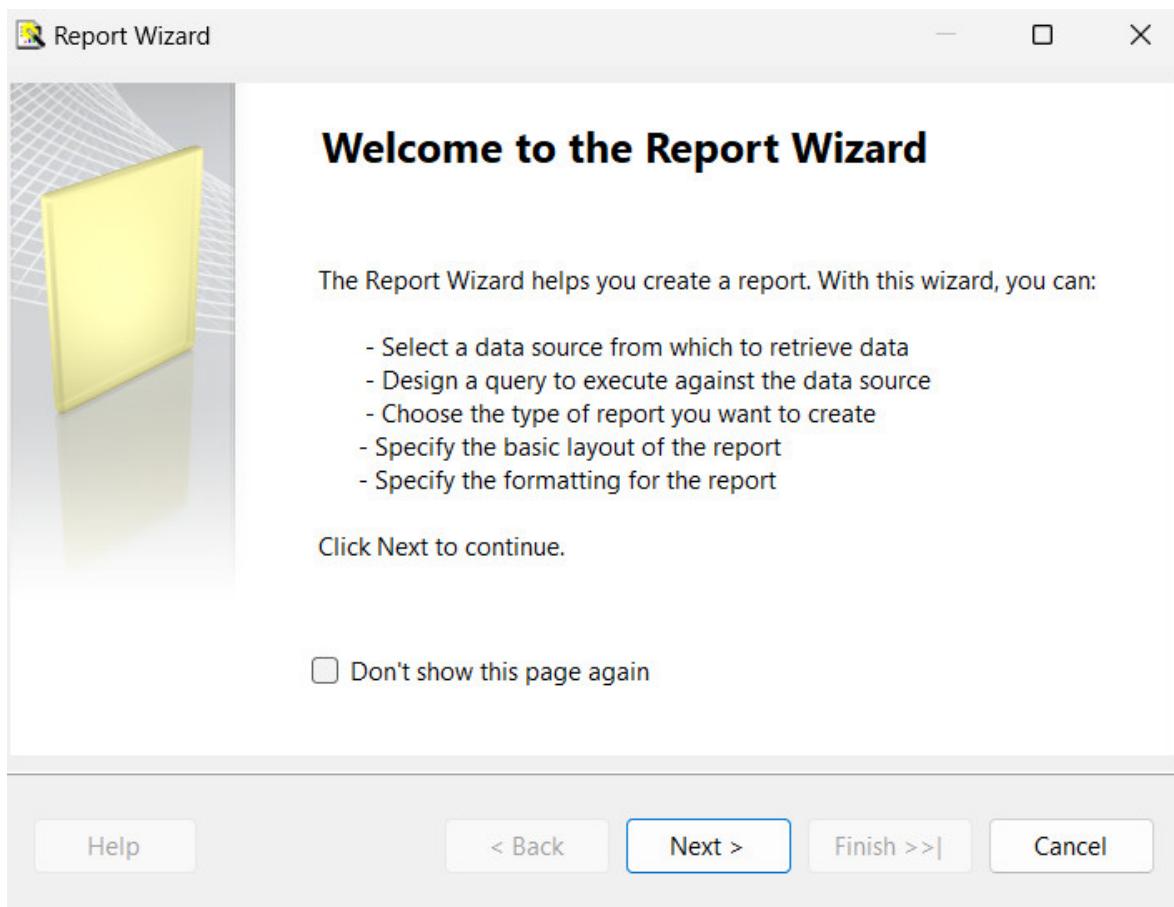
#### 3.14.2.4. Câu truy vấn 7:

**Bước 1:** Thêm bảng vào Report. Tại “Solution Explorer”, click chuột phải vào “Report”, chọn “Add new Report”.



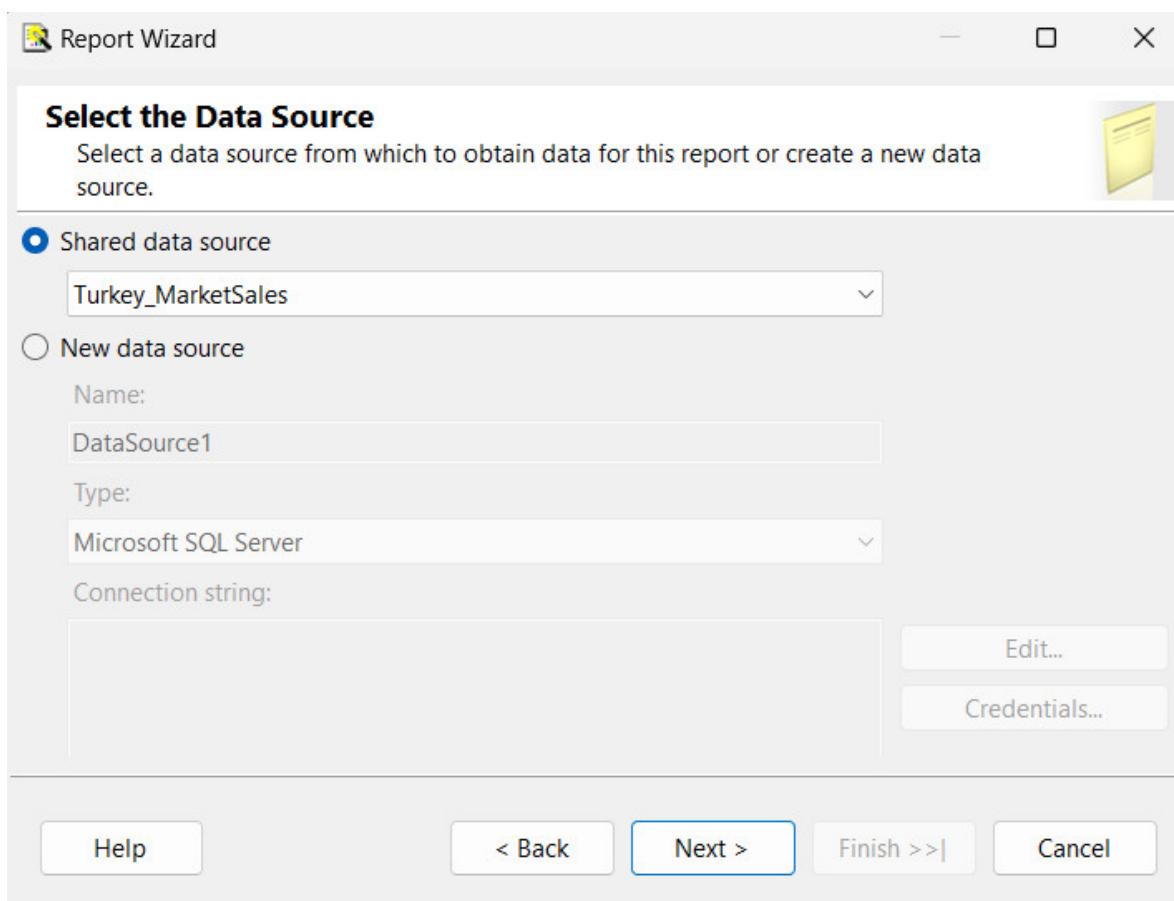
Hình 532: Giao diện tạo mới một Report

Bước 2: Click “Next”.



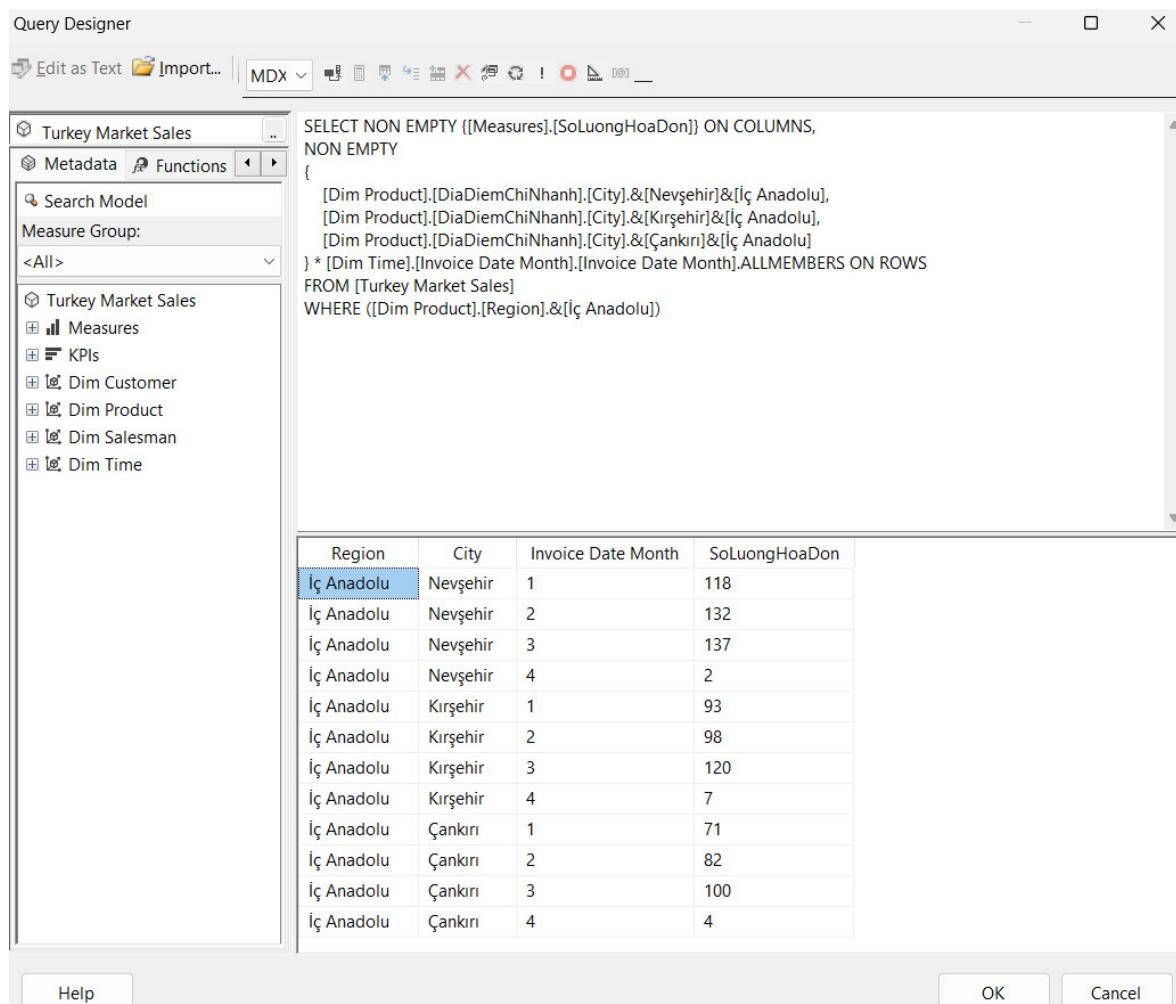
Hình 533: Hộp thoại Report Wizard xuất hiện

**Bước 3:** Tiếp tục click “Next”.



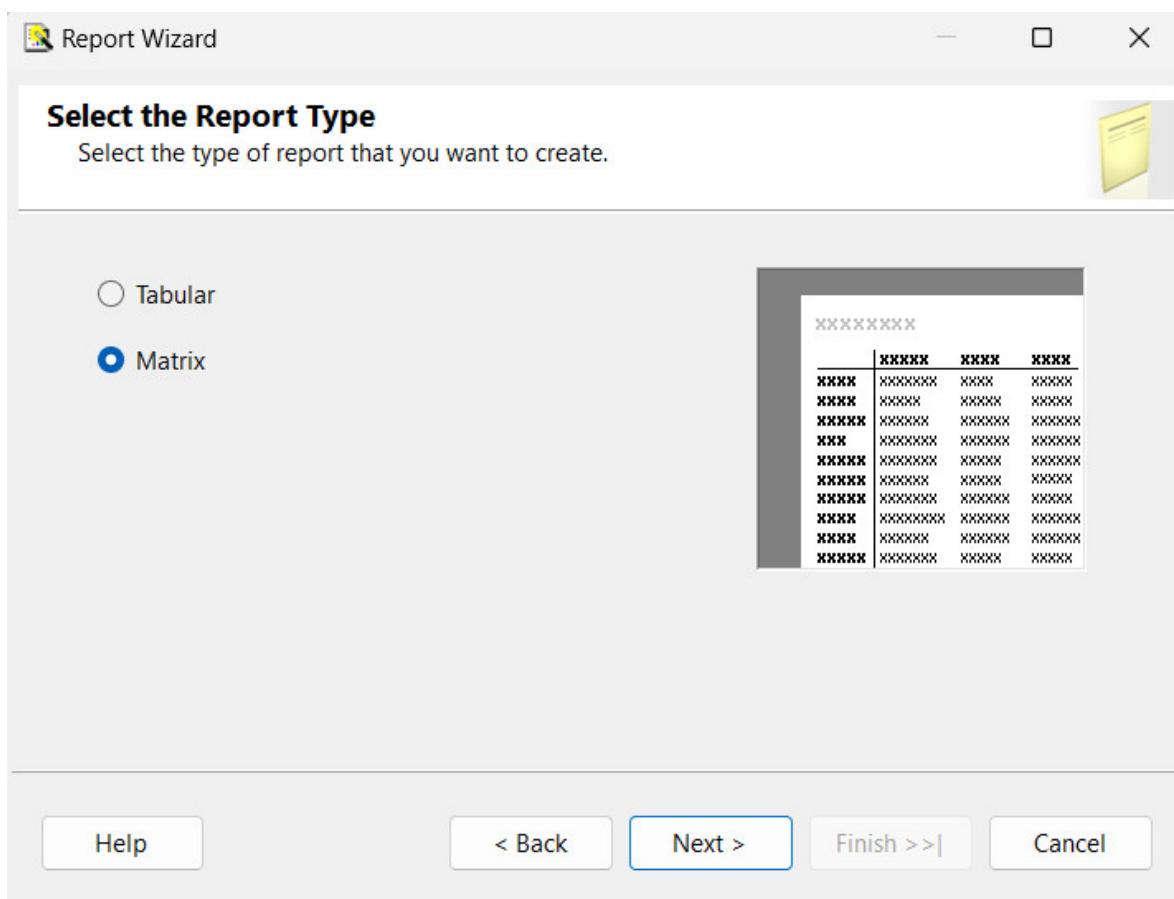
Hình 534: Giao diện chọn nguồn dữ liệu cho Report

Bước 4: Dán câu truy vấn MDX số 7 vào “Query Designer”.



Hình 535: Dán câu truy vấn MDX của câu truy vấn 7 vào ô “Query Designer”

**Bước 5:** Chọn “Matrix” và click “Next”.



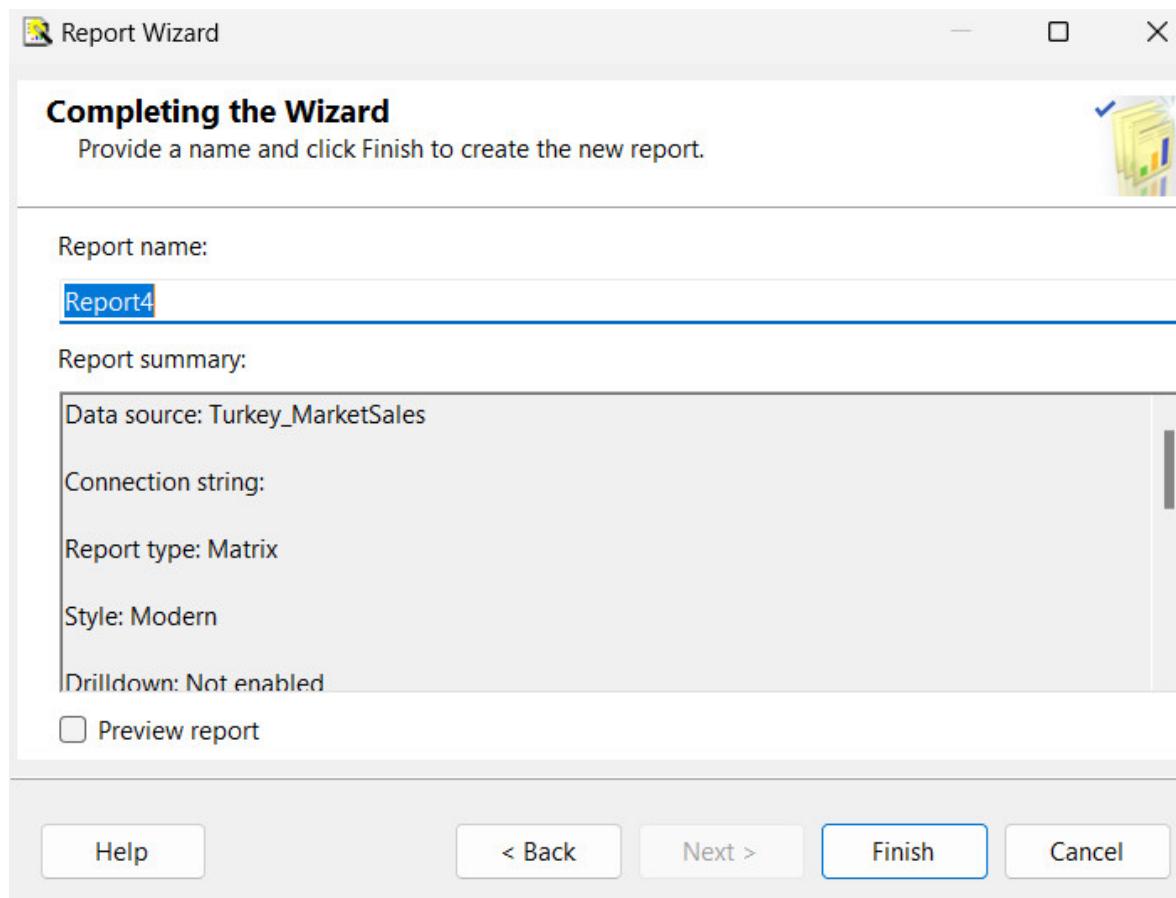
Hình 536: Giao diện chọn loại Report

**Bước 6:** Chọn thuộc tính region, city vào “**Rows**”, SoLuong vào “**Detail**”, invoice\_date\_month vào “**Columns**”.



Hình 537: Chuyển các thuộc tính thích hợp vào mục “Columns”, “Rows” và “Details”

**Bước 7:** Sau đó click “Finish >>” để hoàn tất quá trình.



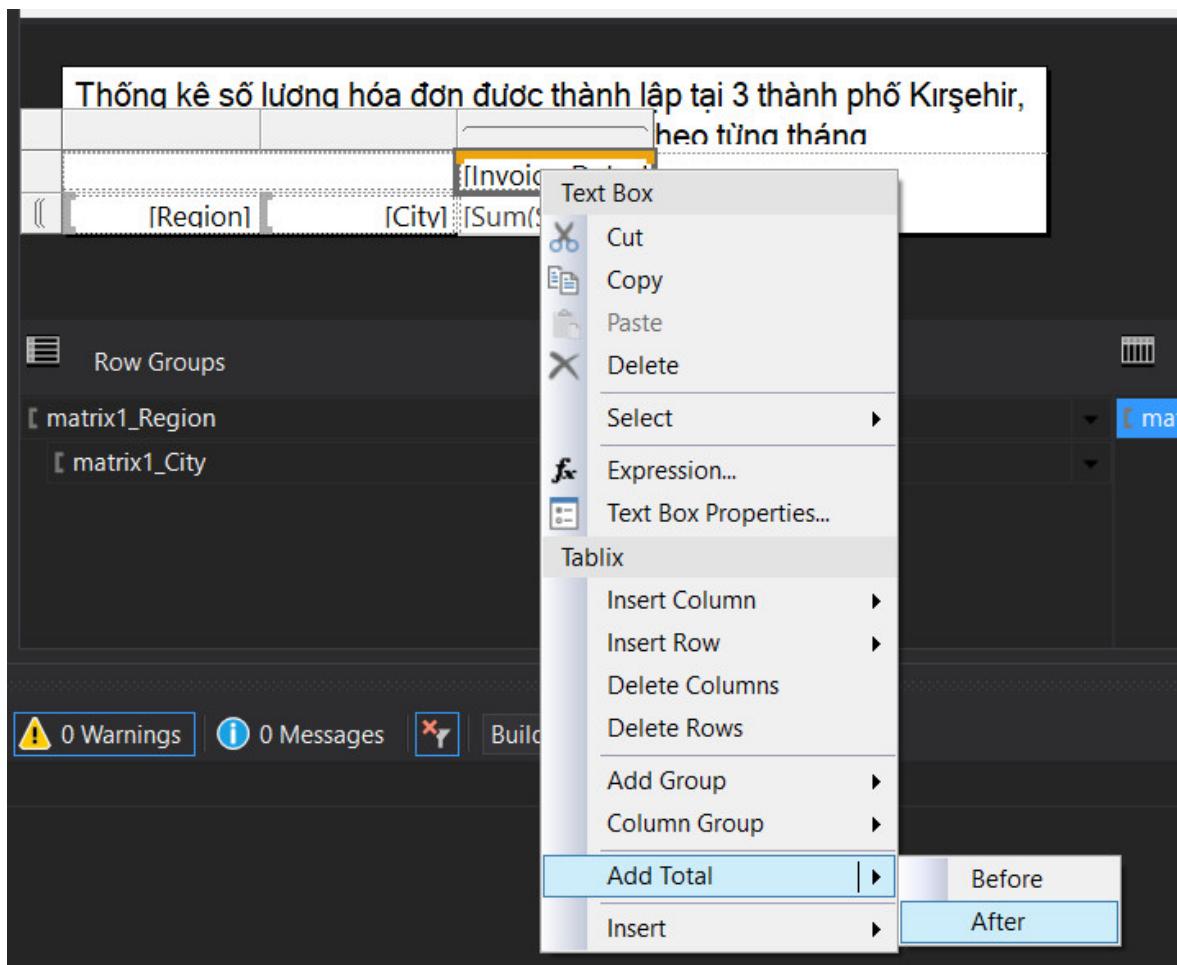
Hình 538: Giao diện đặt tên cho Report

Bước 8: Click “Preview” để kiểm tra lại tính đúng đắn của dữ liệu.

	1	2	3	4
İç Anadolu	Çankırı	71	82	100
	Kırşehir	93	98	120
	Nevşehir	118	132	137

Hình 539: Kết quả câu truy vấn 7 dưới dạng ma trận

Bước 10: Click chuột phải vào cột invoice\_date\_month, chọn “Add totals”, đặt tên cho Totals đó là TongSoLuongHoaDon.



Hình 540: Tiến hành tạo thêm Totals có tên là TongSoLuongHoaDon

The screenshot shows a report viewer window with a table. The table title is "Thống kê số lượng hóa đơn được thành lập tại 3 thành phố Kırşehir, Nevşehir, Çankırı thuộc vùng İç Anadolu theo từng tháng.". The table has columns labeled 1, 2, 3, 4, and TongSoLuongHoaDon. The data is as follows:

İç Anadolu	Çankırı	71	82	100	4	257
	Kırşehir	93	98	120	7	318
	Nevşehir	118	132	137	2	389

Hình 541: Report cho câu truy vấn 7

## 4. Data Mining:

### 4.1. Mô tả dataset:

Bộ dữ liệu được nhóm chúng em sử dụng cho quá trình Mining được thu thập tại website [Kaggle](#) – một trong những website chứa nhiều bộ dữ liệu tham khảo hữu ích về nhiều đề tài, lĩnh vực khác nhau trong cuộc sống. Bộ dữ liệu bao gồm 25 cột thuộc tính trong đó có 20 cột dữ liệu kiểu số còn lại là các thuộc tính kiểu phân loại.

Bộ dữ liệu lưu trữ toàn bộ thông tin liên quan tới hoạt động đánh giá của các khách hàng sau khi đã thực hiện thành công chuyến bay bao gồm: thông tin khách hàng, thông tin chuyến bay, các tiêu chí đánh giá về dịch vụ được cung cấp bởi hãng hàng không.

Link dataset: [Dữ liệu đánh giá chuyến bay](#)

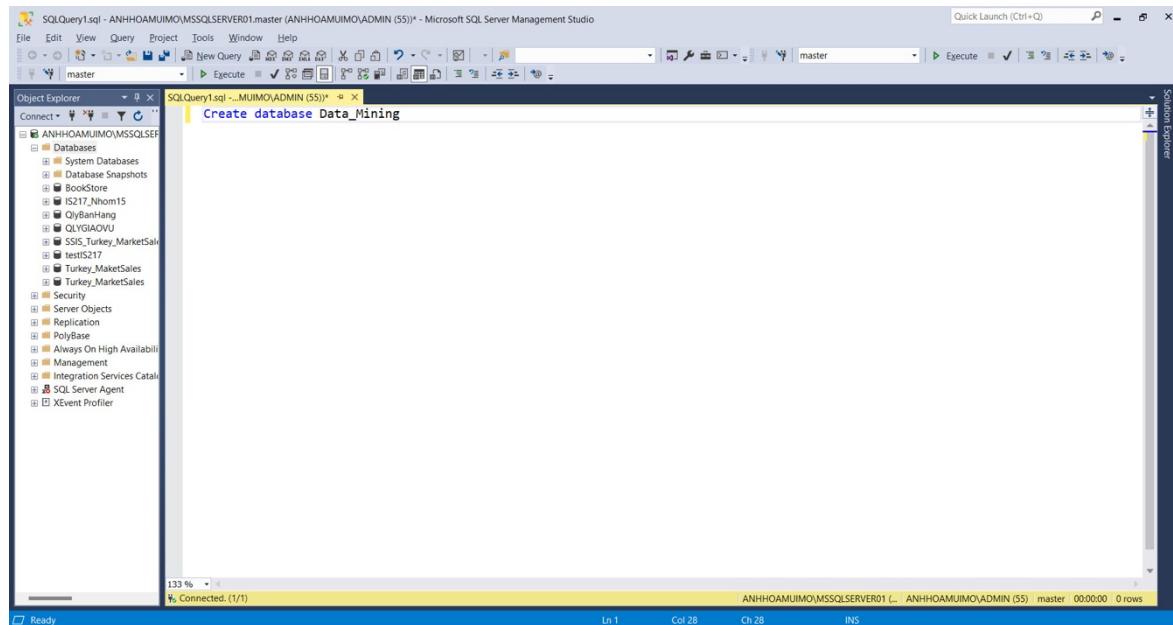
#### 4.2. Mô tả chi tiết các thuộc tính:

STT	Tên Thuộc Tính	Kiểu dữ liệu	Mô tả
1	Id	Int	Mã định danh cho bài đánh giá
2	Gender	String	Giới tính của hành khách gồm 2 giá trị Female (Nữ) và Male (Nam)
3	Customer Type	String	Loại khách hàng gồm 2 giá trị: Disloyal (khách vãng lai) và Loyal (khách hàng thân thiết)
4	Age	Int	Độ tuổi của hành khách
5	Type of Travel	String	Loại hình di chuyển gồm 2 giá trị: Business travel (đi công tác) và Personal travel (mục đích cá nhân)
6	Class	String	Hạng ghế gồm 3 giá trị: Business (hạng thương gia), Eco (hạng tiêu chuẩn) và Eco Plus (hạng cao cấp)
7	Flight Distance	Int	Khoảng cách bay
8	Inflight Wifi Service	Int	Mức độ hài lòng về dịch vụ Wifi trong suốt quá trình bay (thang điểm từ 1-5)
9	Departure/Arrival Time Convenient	Int	Mức độ hài lòng về thời gian khởi hành và thời gian đến (thang điểm từ 1-5)
10	Ease of Online Booking	Int	Mức độ hài lòng về sự thuận tiện trong việc đặt vé trực tuyến (thang điểm từ 1-5)

11	Gate location	Int	Mức độ hài lòng về cổng ra vào máy bay (thang điểm từ 1-5)
12	Food and Drink	Int	Mức độ hài lòng về chất lượng thực phẩm và nước giải khát trong suốt chuyến bay (thang điểm từ 1-5)
13	Online Boarding	Int	Mức độ hài lòng về việc làm thủ tục trực tuyến (thang điểm từ 1-5)
14	Seat Comfort	Int	Mức độ thoải mái tại chỗ ngồi trên máy bay (thang điểm từ 1-5)
15	Inflight Entertainment	Int	Mức độ hài lòng về sự thư giãn, giải trí trong suốt quá trình bay (thang điểm từ 1-5)
16	On-board Service	Int	Mức độ hài lòng về các dịch vụ được phục vụ trực tiếp trên máy bay (thang điểm từ 1-5)
17	Leg room Service	Int	Mức độ hài lòng về khoảng không gian dành cho chân của hành khách khi ngồi trên máy bay (thang điểm từ 1-5)
18	Baggage Handling	Int	Mức độ hài lòng về quy trình vận chuyển hành lý của hành khách (thang điểm từ 1-5)
19	Check-in service	Int	Mức độ hài lòng về quy trình check-in của sân bay (thang điểm từ 1-5)
20	Inflight Service	Int	Mức độ hài lòng trong suốt quá trình ngồi trên máy bay (thang điểm từ 1-5)
21	Clealiness	Int	Mức độ hài lòng về sự sạch sẽ tại sân bay và trên máy bay (thang điểm từ 1-5)
22	Departure Delay in Minutes	Int	Thời gian trễ của chuyến bay khi khởi hành (đơn vị tính là phút)
23	Arrival Delay in Minutes	Int	Thời gian trễ của chuyến bay đến so với thời gian dự kiến (đơn vị tính là phút)
24	Satisfaction	Int	Đánh giá cuối cùng của khách hàng gồm 2 giá trị: satisfied (hài lòng) hoặc dissatisfied or neutral (không hài lòng hoặc trung lập)

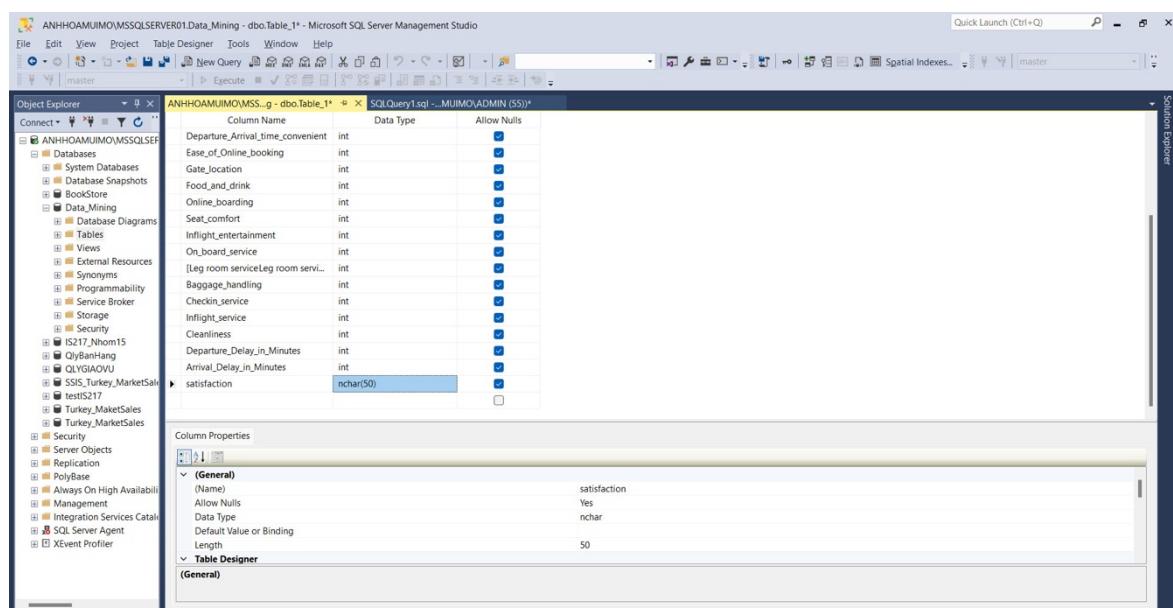
### 4.3. Quá trình đổ dữ liệu vào SQL Server bằng công cụ SSIS trên Visual Studio:

**Bước 1:** Trong SQL Server, tạo một cơ sở dữ liệu rỗng có tên là “Data\_Mining”.



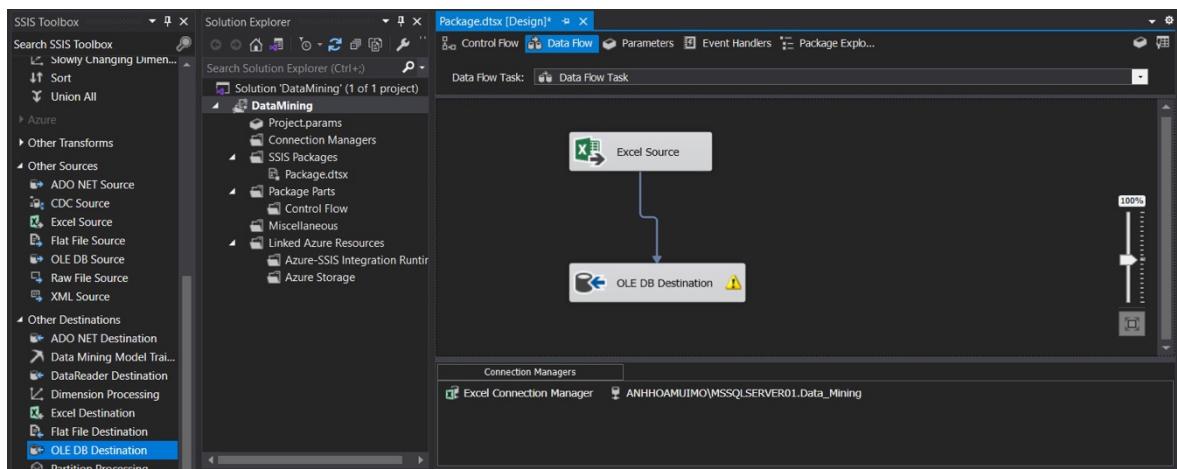
Hình 542: Khởi tạo cơ sở dữ liệu rỗng

**Bước 2:** Tạo bảng chứa dữ liệu của Dataset.



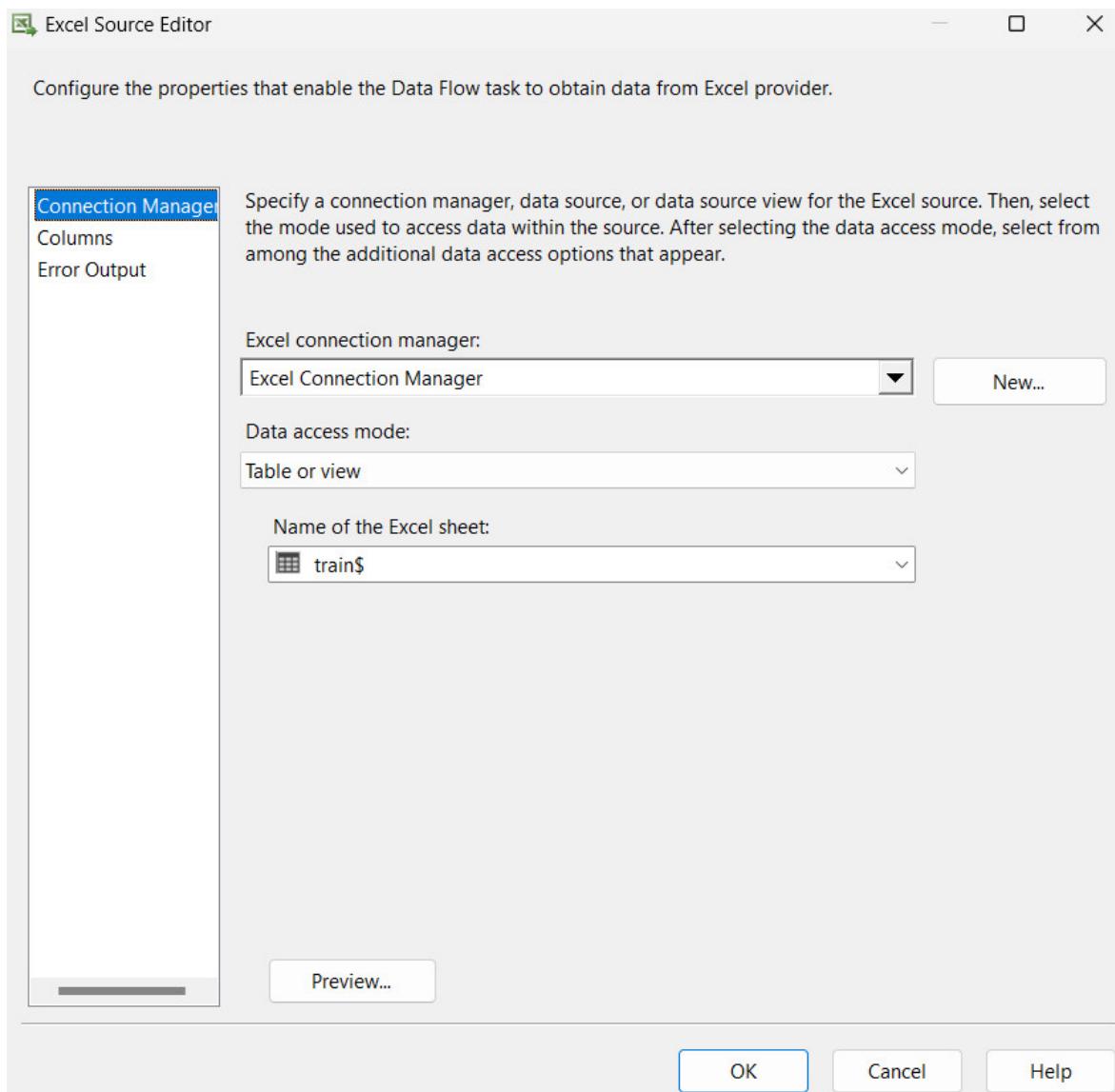
Hình 543: Tạo các bảng dữ liệu

**Bước 3:** Thực hiện đổ dữ liệu từ file Excel vào SQL Server.



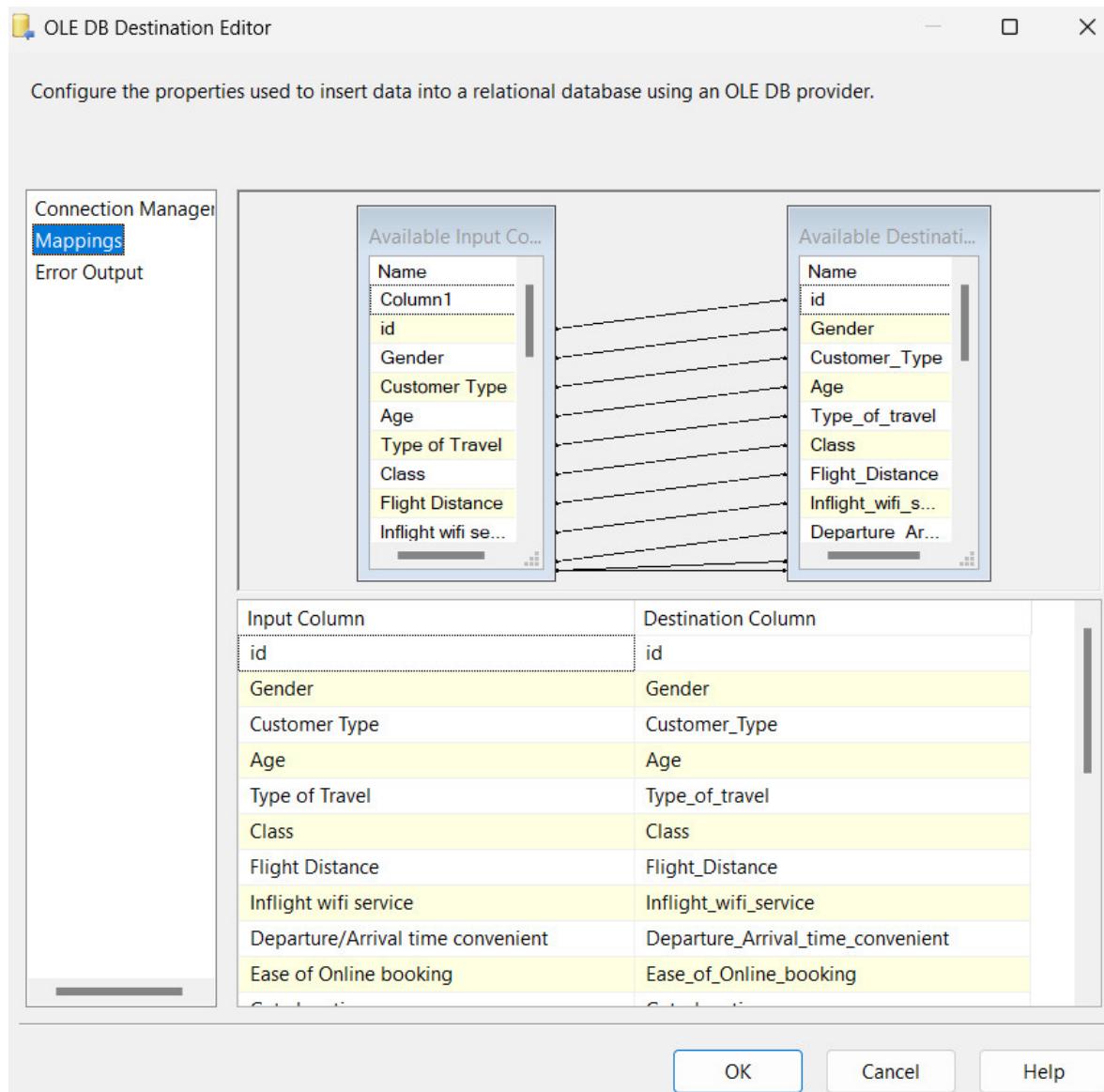
Hình 544: Tiến hành đổ dữ liệu từ file Excel bằng công cụ SSIS

**Bước 4:** Trong “Excel Source”, tạo kết nối đến file Excel chứa dataset ban đầu.



Hình 545: Tạo kết nối đến file Excel

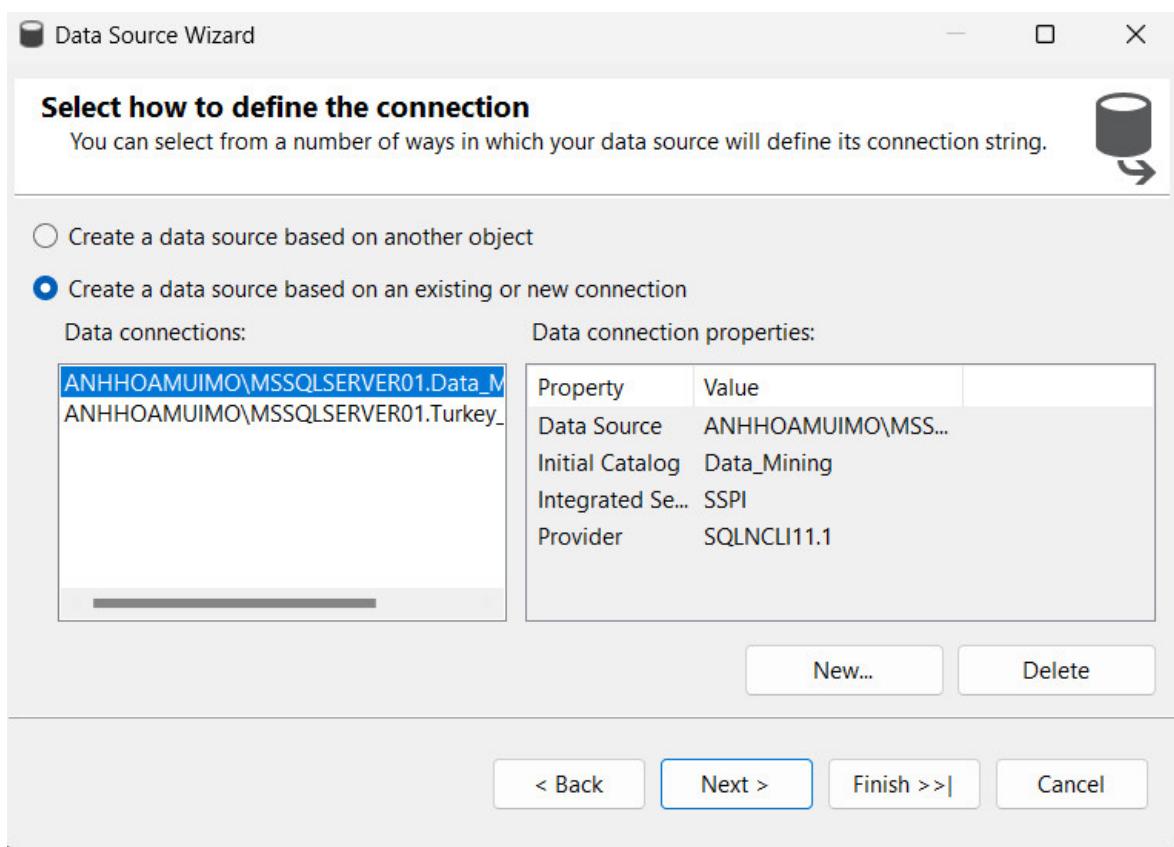
**Bước 5:** Đổ dữ liệu vào cơ sở dữ liệu rỗng đã tạo ở Bước 1. Kiểm tra lại các cột thuộc tính đã chính xác hay chưa.



Hình 546: Kiểm tra lại các cột thuộc tính

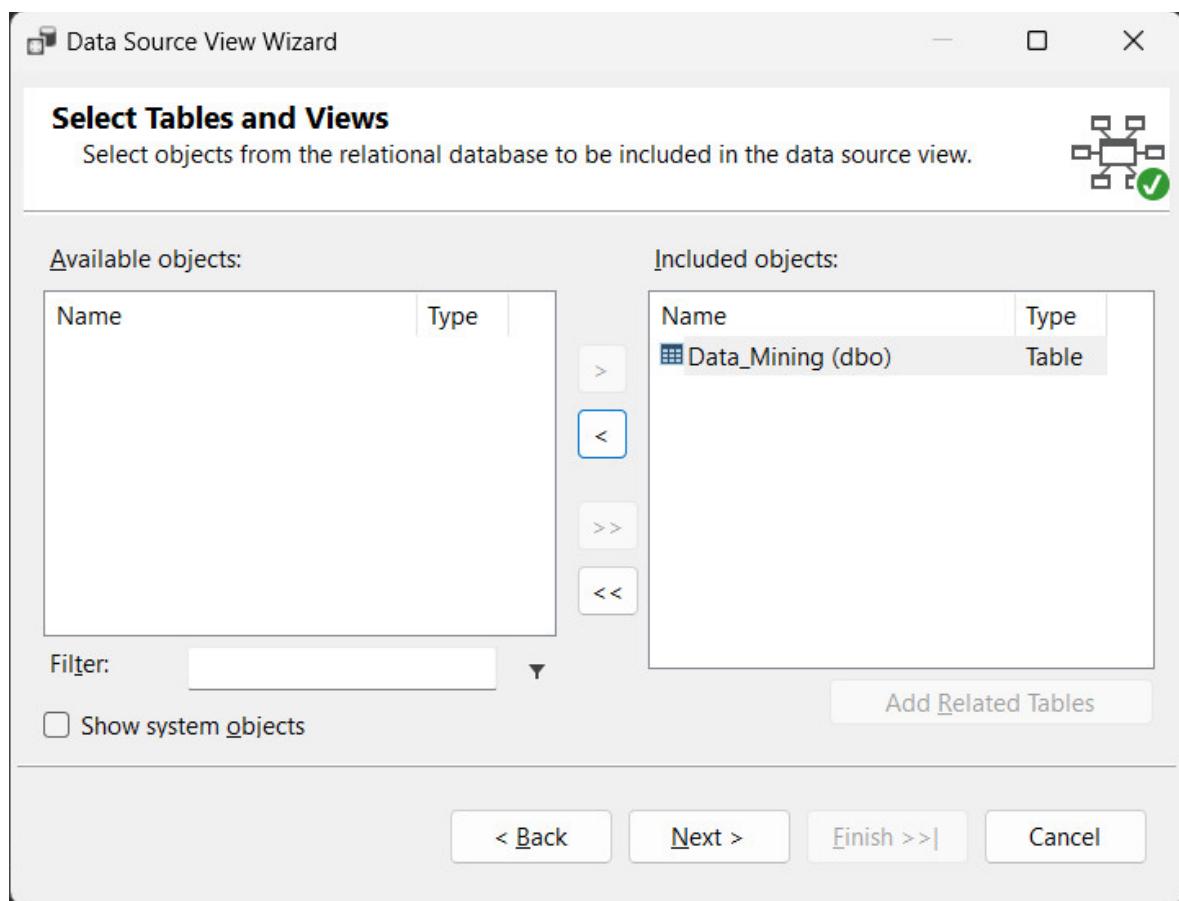
#### 4.4. Data Mining bằng công cụ SSAS trên Visual Studio:

**Bước 1:** Add Datasource và Database vừa tạo ở các bước trên.



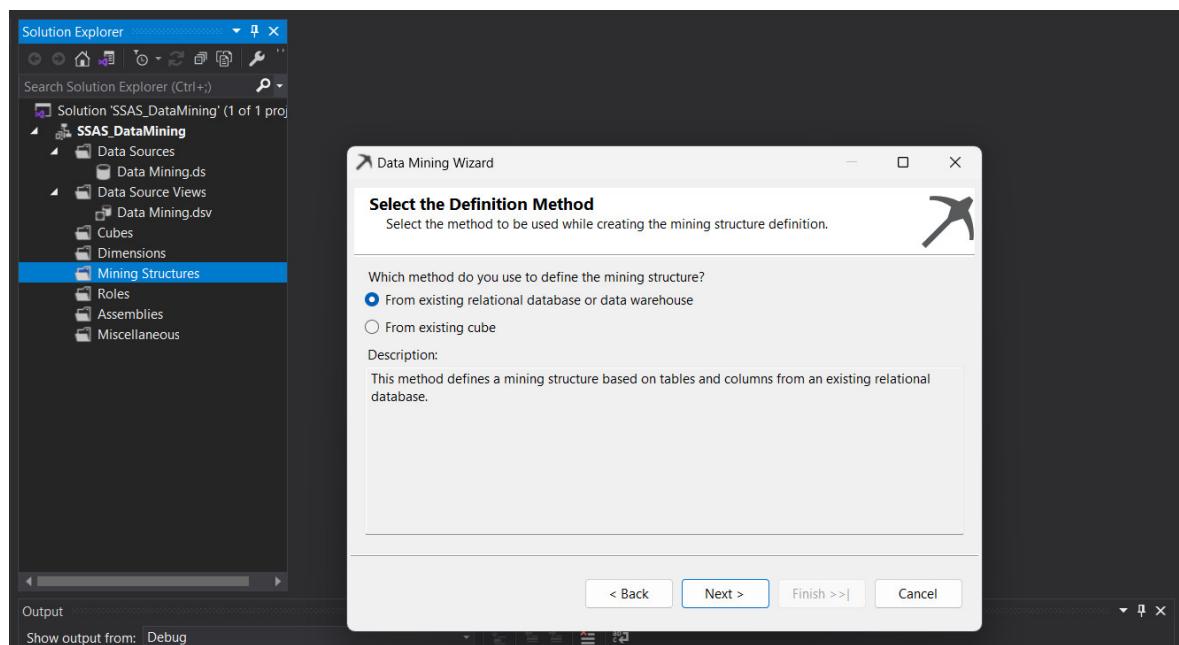
Hình 547: Giao diện thêm Datasource và Database vừa tạo

**Bước 2:** Tạo Data Source View với cơ sở dữ liệu là “Data\_Mining”.



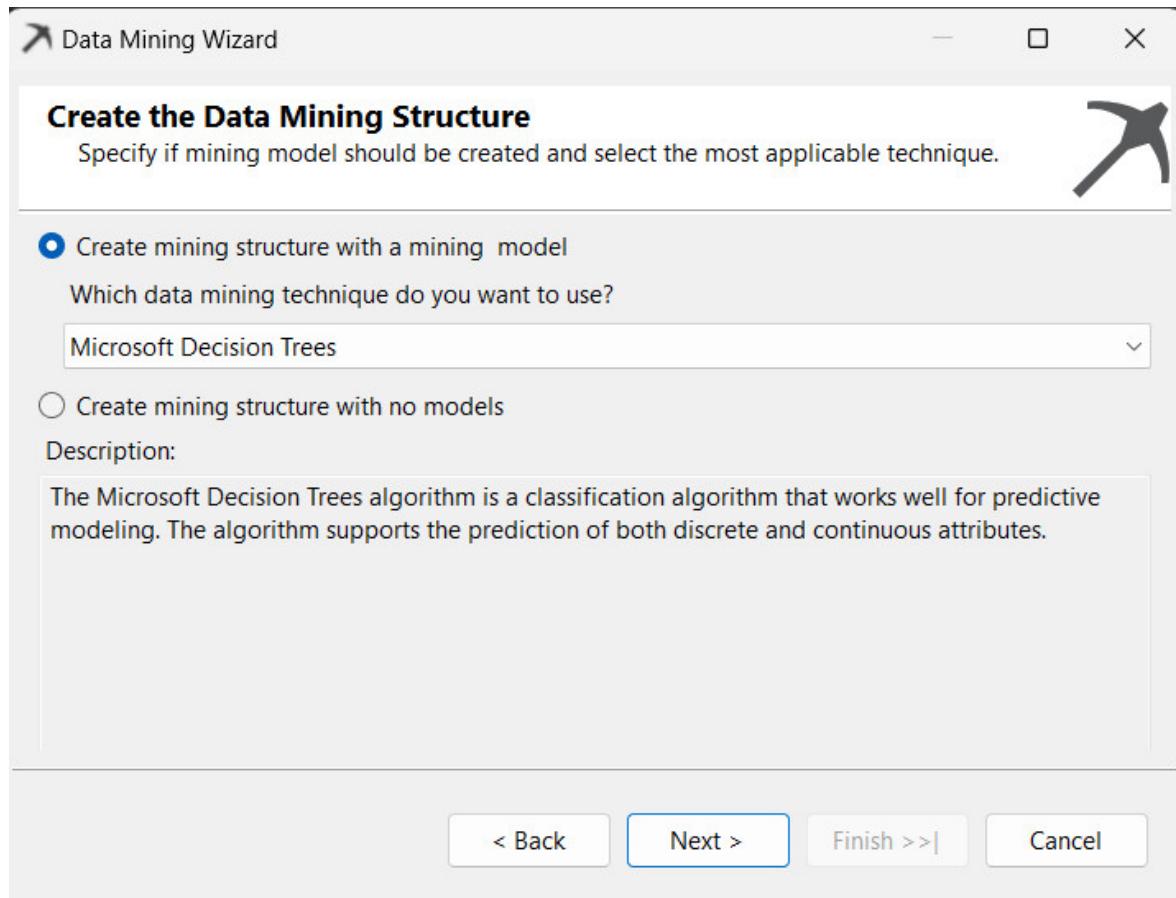
Hình 548: Tạo mới một Data Source View

### Bước 3: Tạo “Mining Structures”.



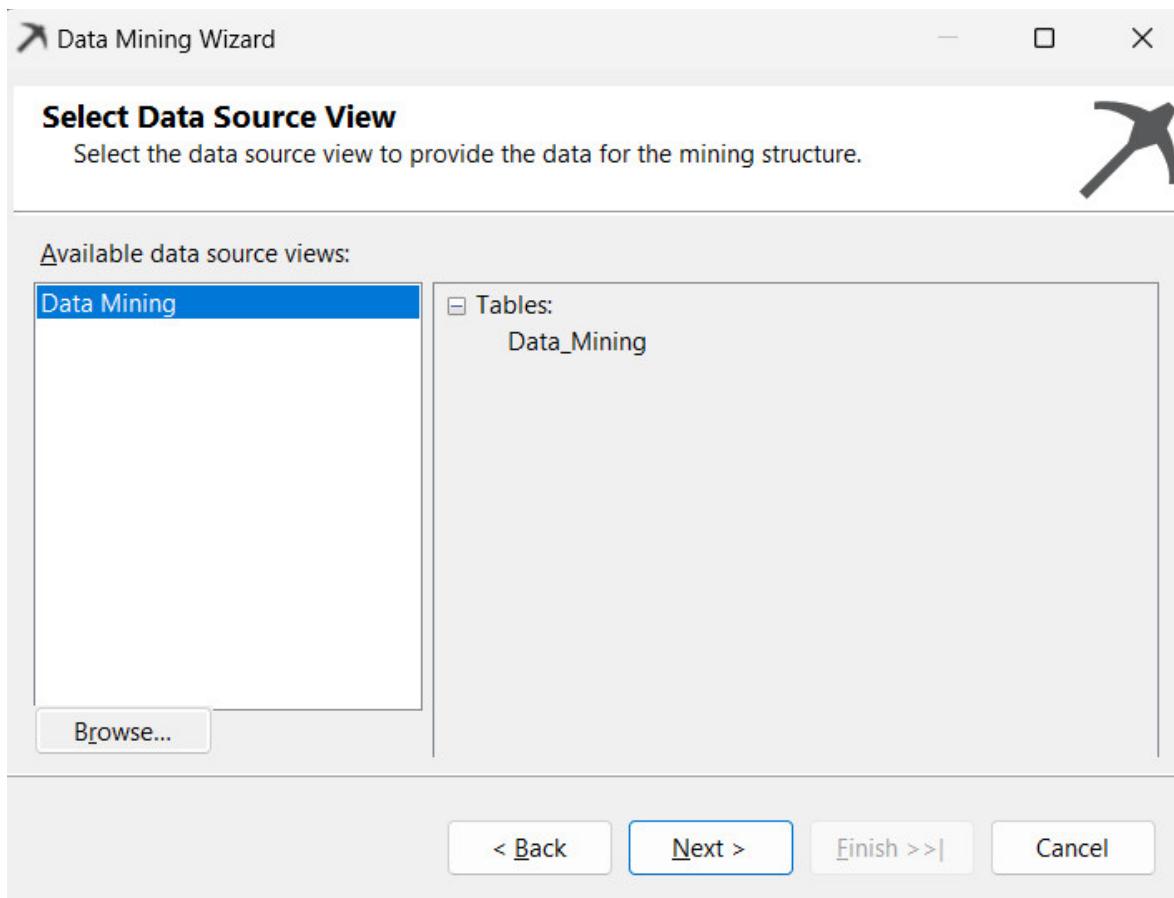
Hình 549: Tạo mới Mining Structures

**Bước 4:** Lựa chọn thuật toán để tiến hành quá trình Mining là “**Microsoft Decision Tree**” (Thuật toán cây quyết định).



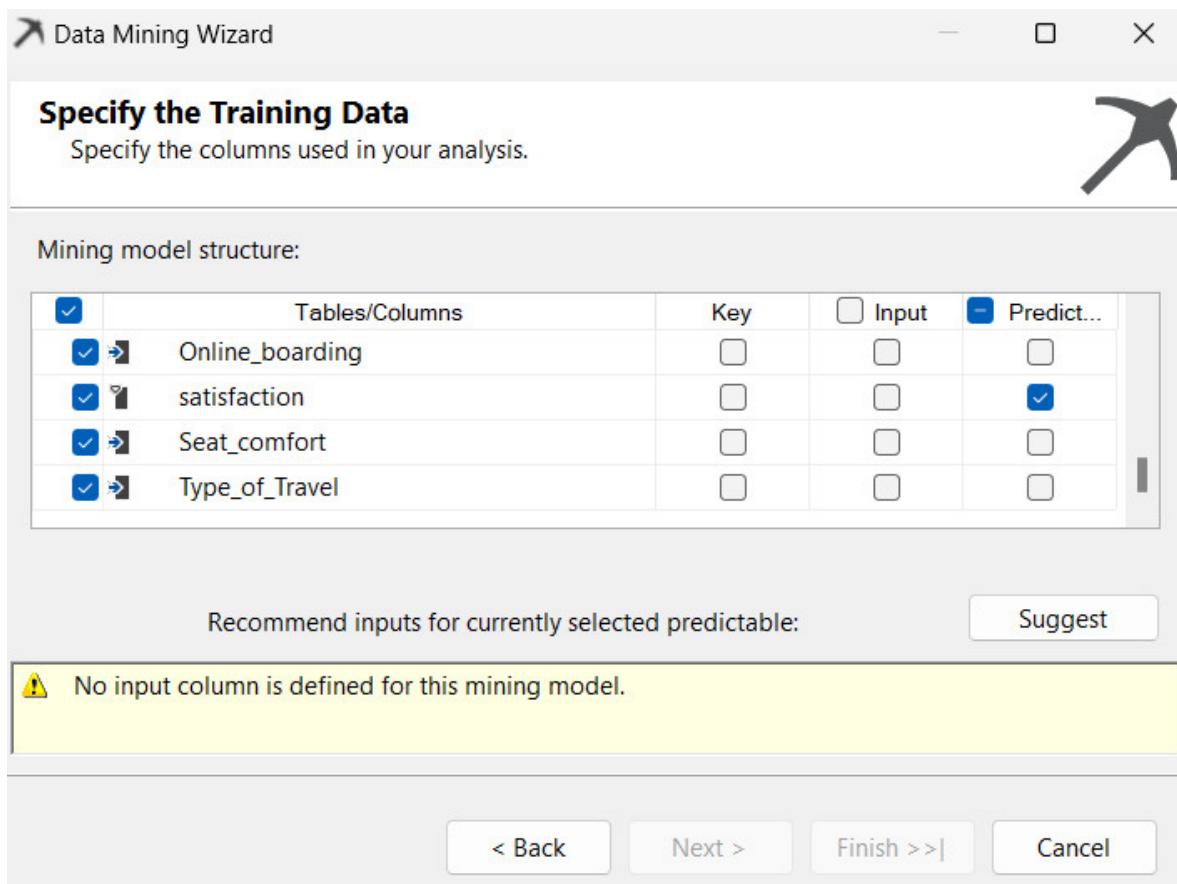
Hình 550: Chọn thuật toán Mining (thuật toán Cây quyết định)

**Bước 5:** Chọn khung nhìn dữ liệu nguồn “Data Mining” vừa tạo ở trên để tiến hành chạy thuật toán cây quyết định.



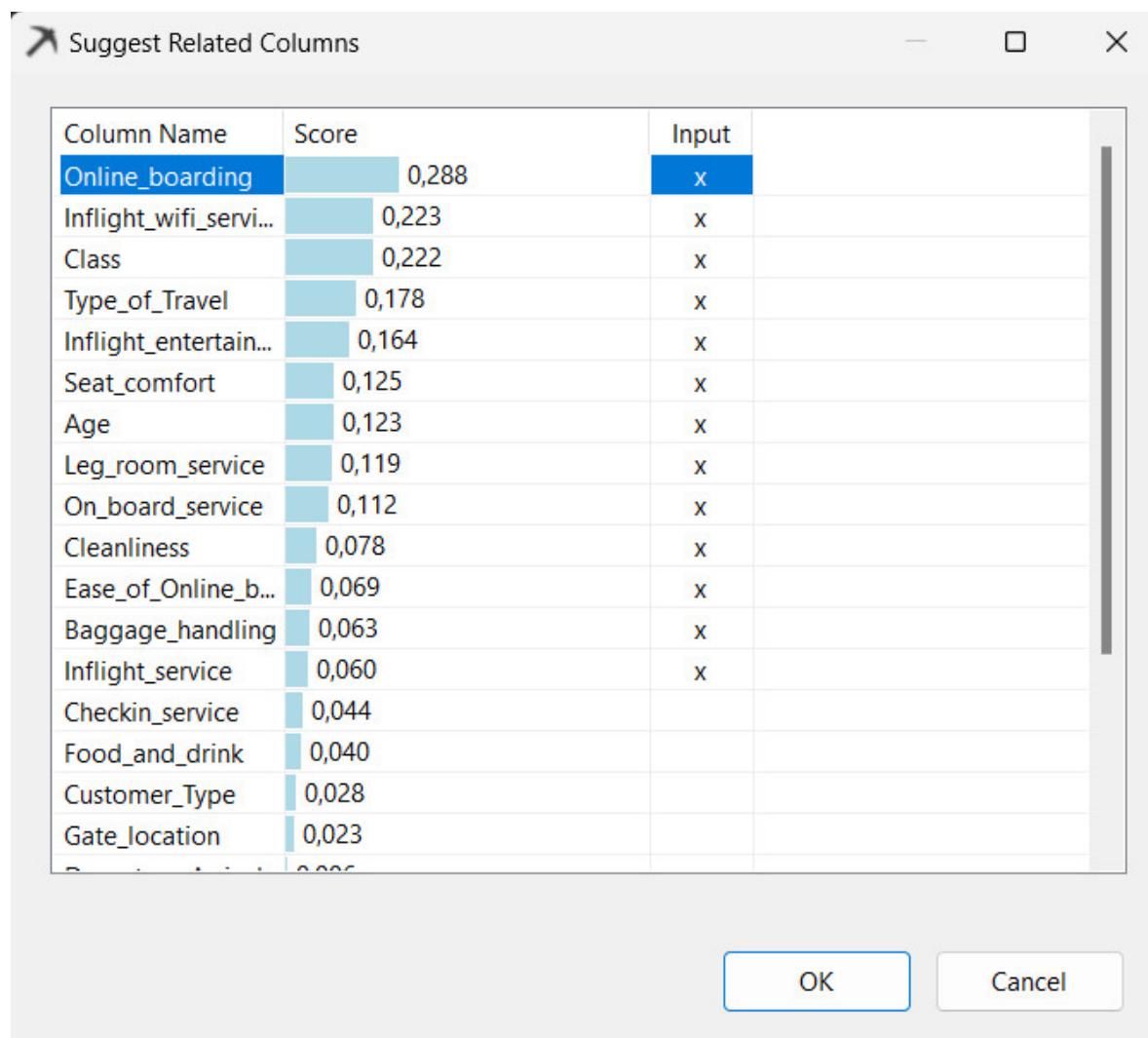
Hình 551: Chọn khung nhìn dữ liệu nguồn để chạy thuật toán

**Bước 6:** Chọn thuộc tính “key” là thuộc tính id, thuộc tính “predict” là thuộc tính satisfaction.



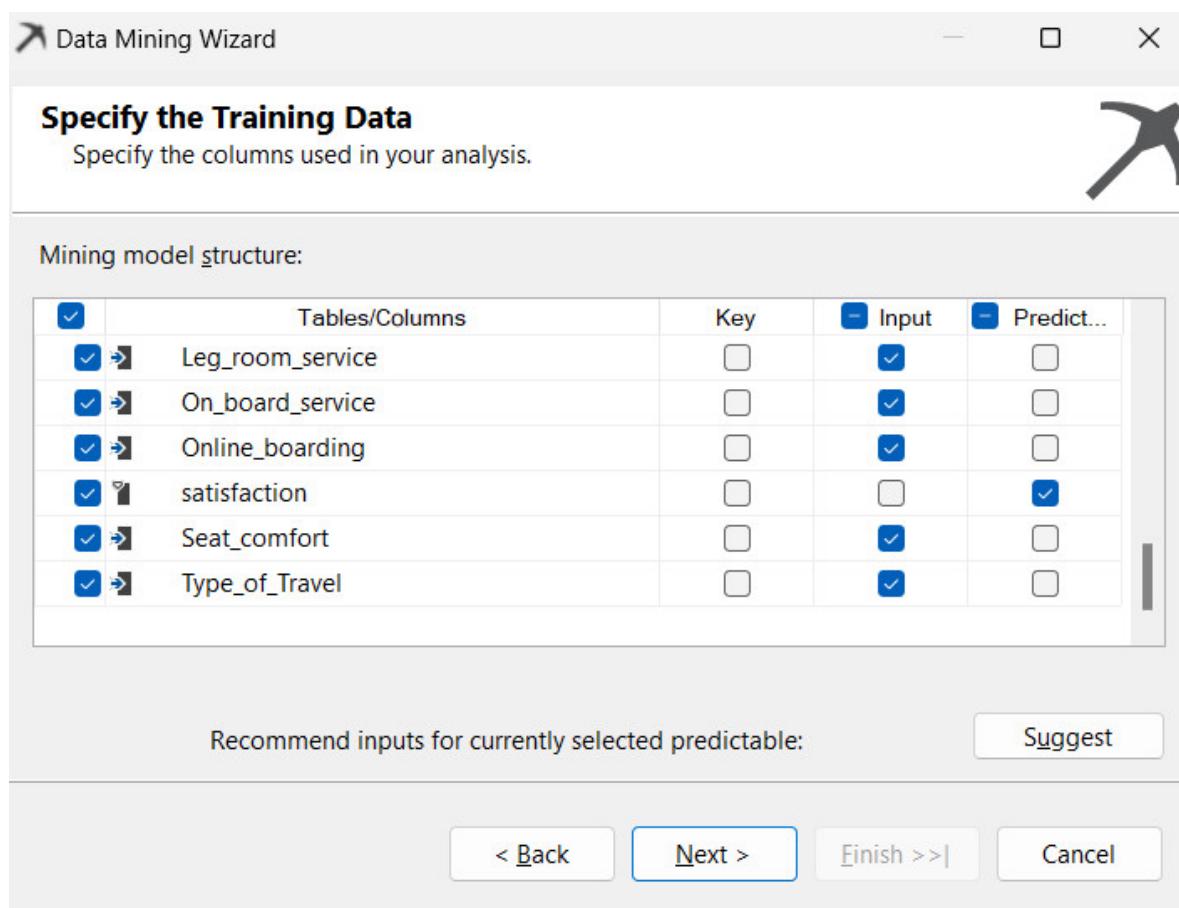
Hình 552: Lựa chọn thuộc tính khóa và thuộc tính dự đoán(mục tiêu)

**Bước 7:** Xem một số cột thuộc tính được gợi ý bởi hệ thống.



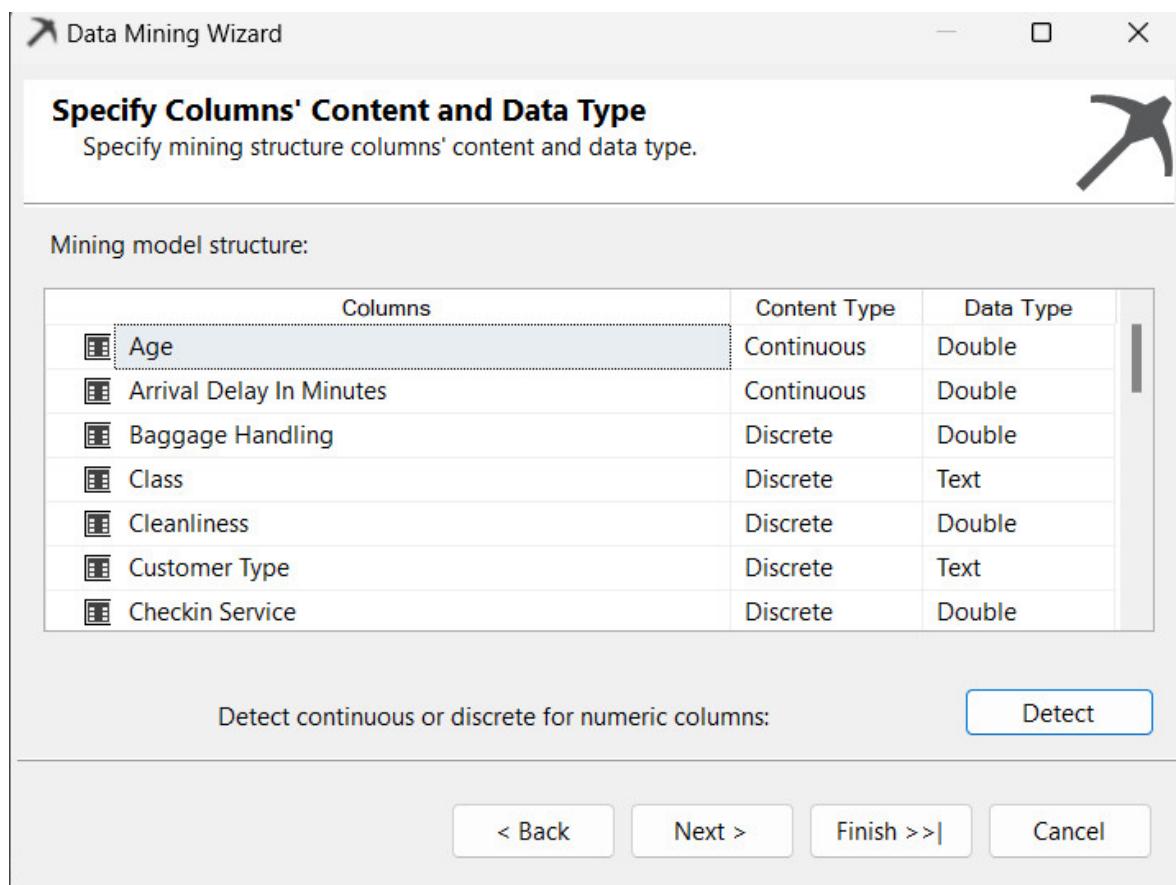
Hình 553: Một số thuộc tính được hệ thống gợi ý làm input

**Bước 8:** Chọn lọc một số thuộc tính làm input đầu vào.



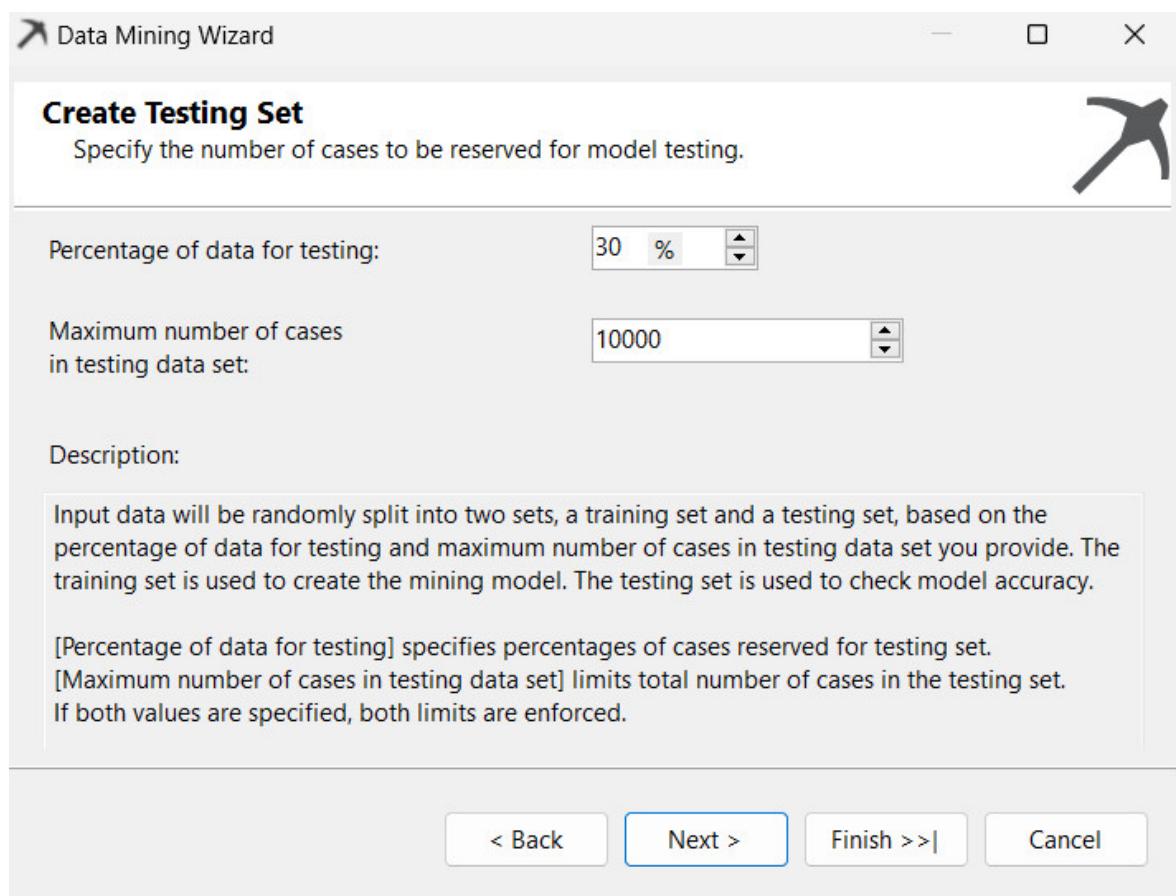
Hình 554: Lựa chọn thuộc tính làm input

**Bước 9:** Detect các trường dữ liệu.



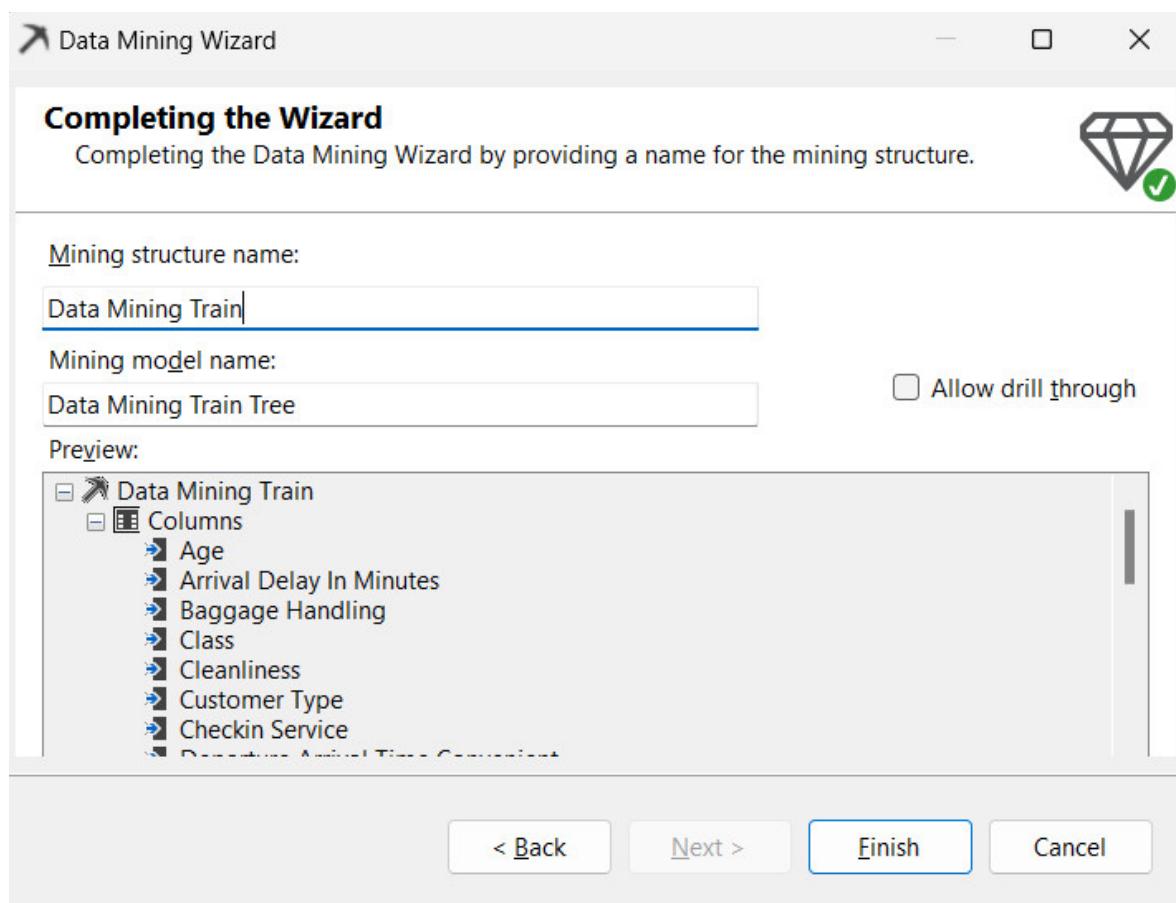
Hình 555: Detect các trường dữ liệu

**Bước 10:** Tiến hành phân chia dataset thành các tập train, tập test và số lượng dữ liệu tối đa của tập test.



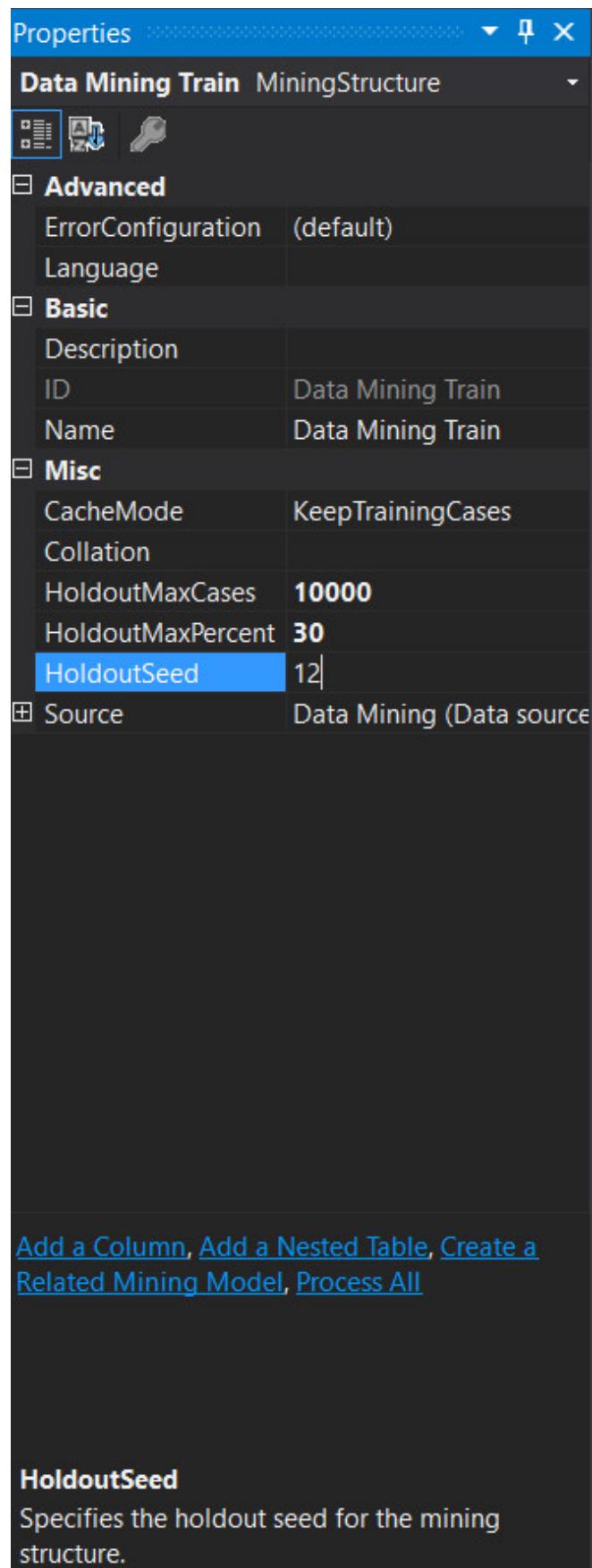
Hình 556: Phân chia bộ dữ liệu thành training set và test set

**Bước 11:** Đặt tên cho Model Training là Data Mining Train Tree.



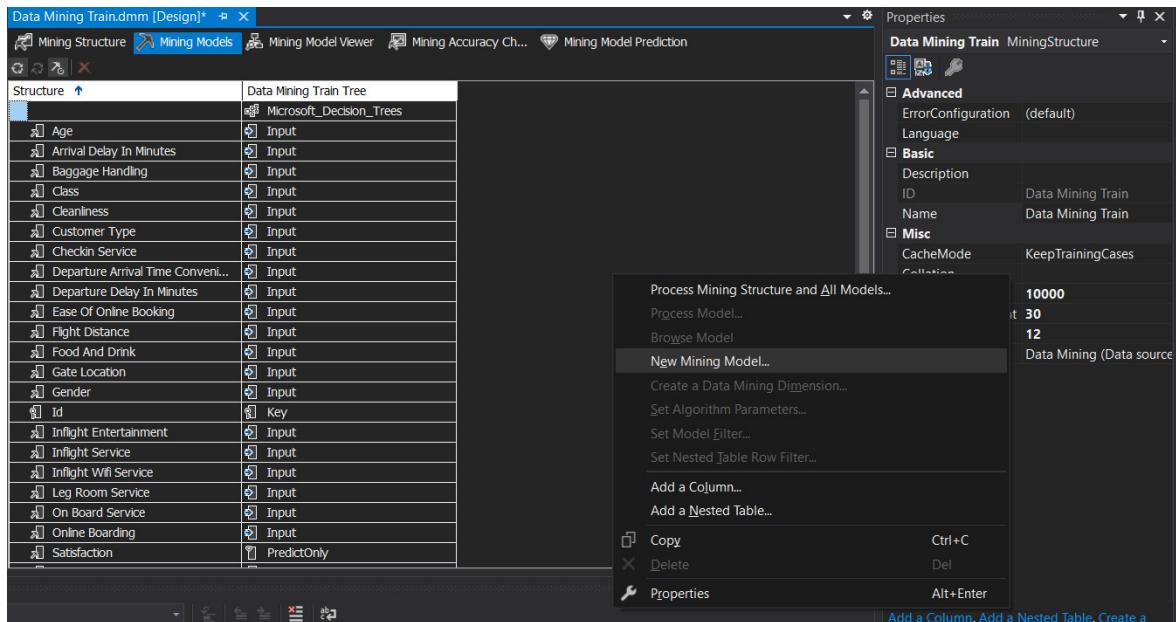
Hình 557: Đặt tên cho Model

Bước 12: Đặt “Holdout Seed” là 12.



Hình 558: Thiết lập giá trị Holdoutseed

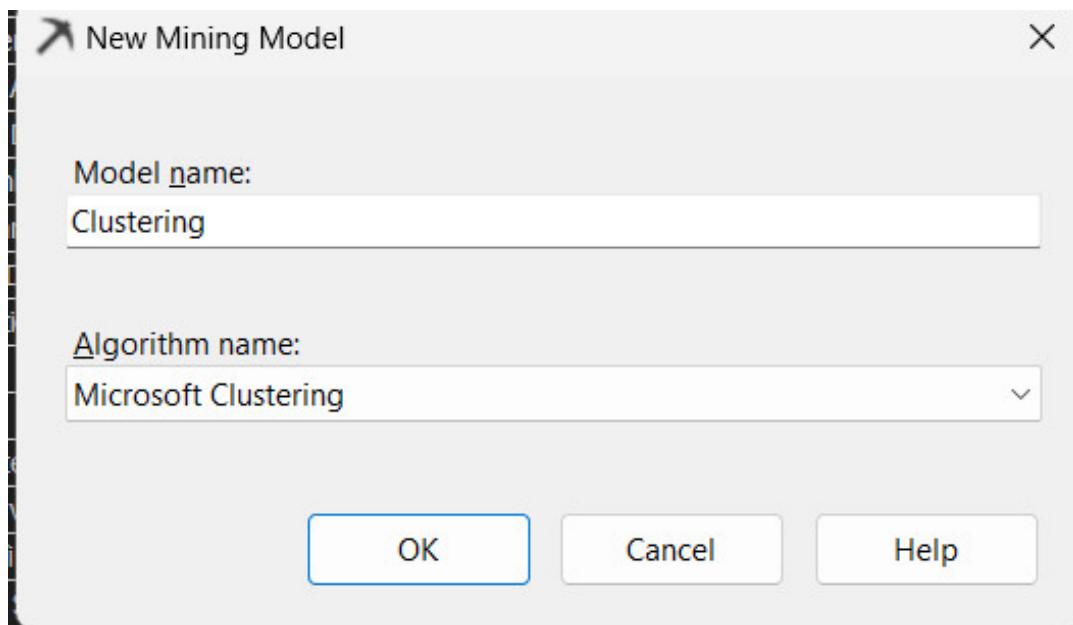
**Bước 13:** Tạo một mô hình data mining mới.



Hình 559: Tạo mới một mô hình datamining khác

**Bước 14:** Lựa chọn thuật toán mới để tiến hành quá trình Mining là

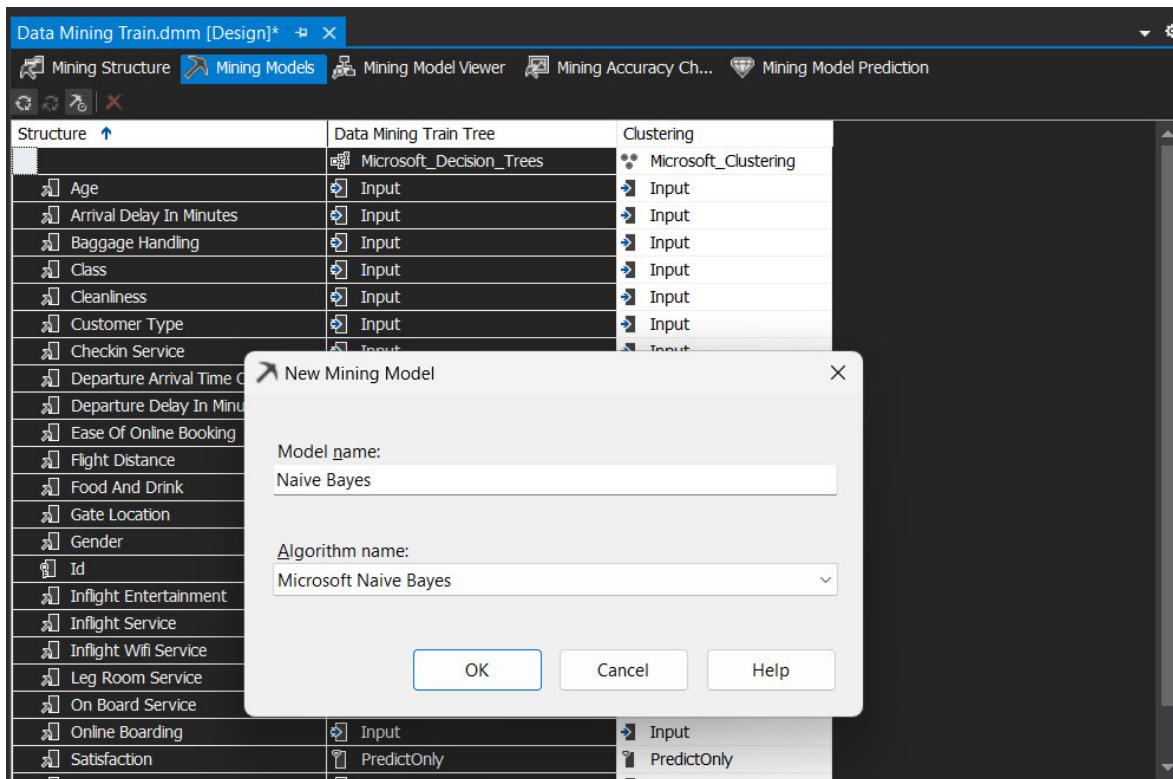
“Microsoft Clustering” (Thuật toán phân cụm).



Hình 560: Chọn thuật toán Mining (thuật toán phân cụm)

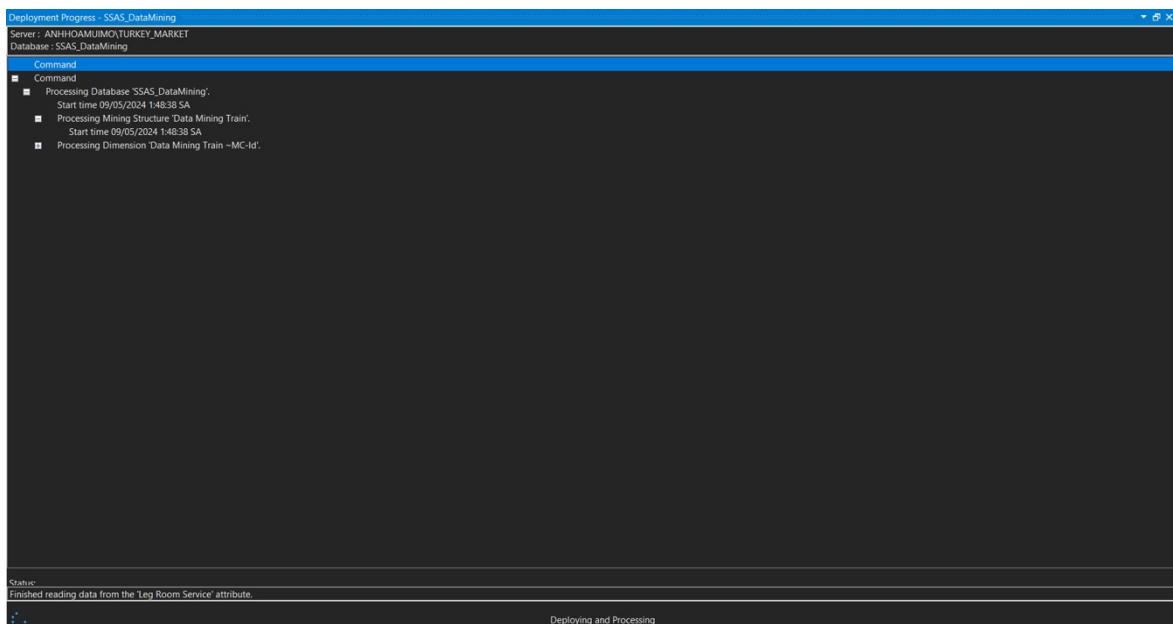
**Bước 15:** Lựa chọn thuật toán mới để tiến hành quá trình Mining là

“Microsoft Naïve Bayes” (Thuật toán Naïve Bayes).



Hình 561: Chọn thuật toán Mining (thuật toán Naïve Bayes)

**Bước 16:** Run toàn bộ process để bắt đầu quá trình Mining.

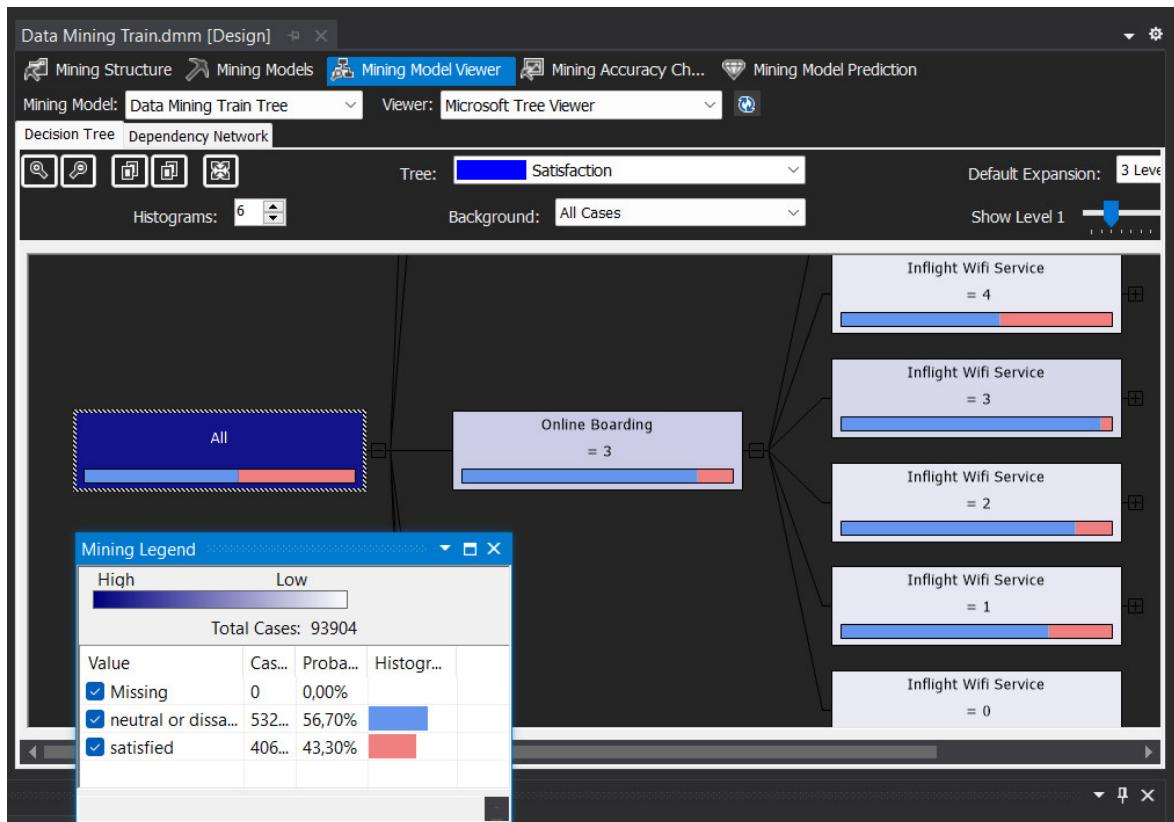


Hình 562: Khởi động quá trình Mining

#### 4.5. Nhận xét kết quả:

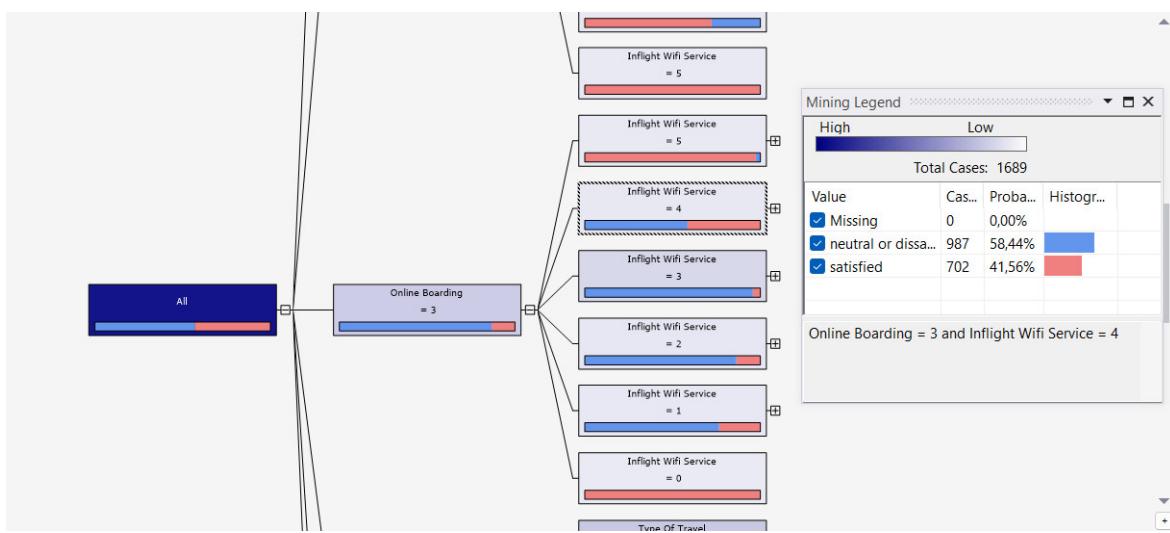
##### 4.5.1. Thuật toán Microsoft Decision Tree:

Xem kết quả thông qua tính năng “**Mining Model Viewer**”, với thuộc tính dự đoán là satisfaction. Tỷ lệ khách hàng thỏa mãn (satisfied) đạt đến 43,30 % trong khi đó tỷ lệ khách hàng trung lập (neutral) hoặc không hài lòng (dissatisfied) đạt 56,70%.



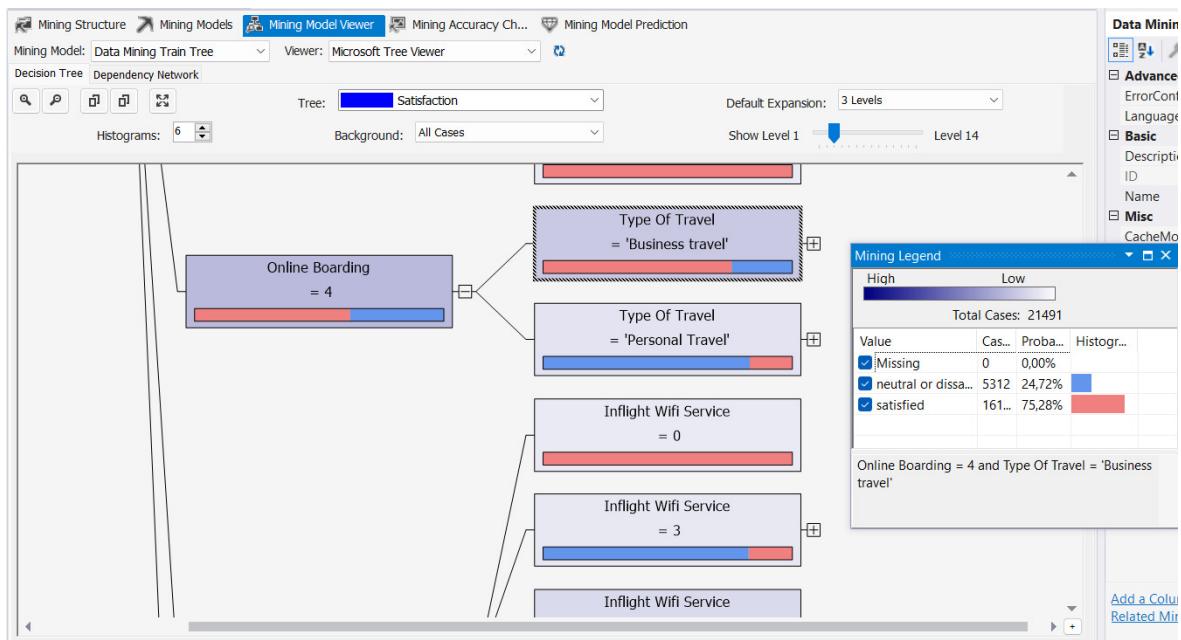
Hình 563: Tỷ lệ đánh giá của khách hàng ban đầu

Với những đánh giá có tiêu chí “Online Boarding” đạt 3 điểm và tiêu chí “Inflight Wifi Service” được 4 điểm, tỉ lệ đánh giá không hài lòng đạt 58,44%.



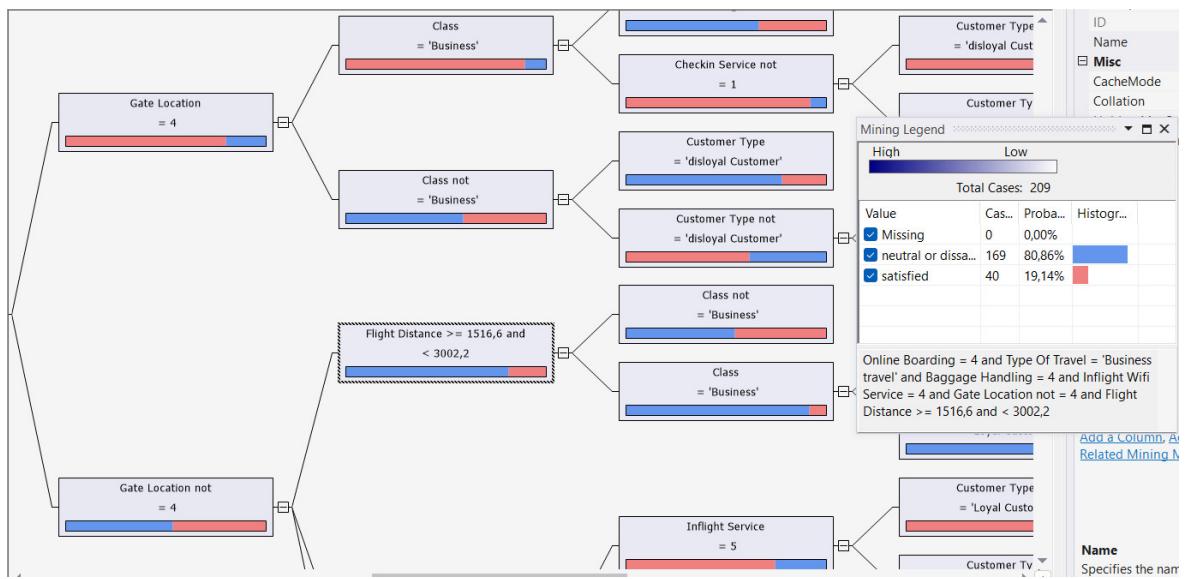
Hình 564: Tỷ lệ đánh giá của khách hàng dựa trên “Online Boarding” và “Inflight Wifi Service”

Với những chuyến đi công tác, tỉ lệ hài lòng tăng lên đến 75,28% khi tiêu chí “Online Boarding” được đánh giá 4 điểm.



Hình 565: Tỷ lệ đánh giá của khách hàng dựa trên “Online Boarding” và “Type of Travel”

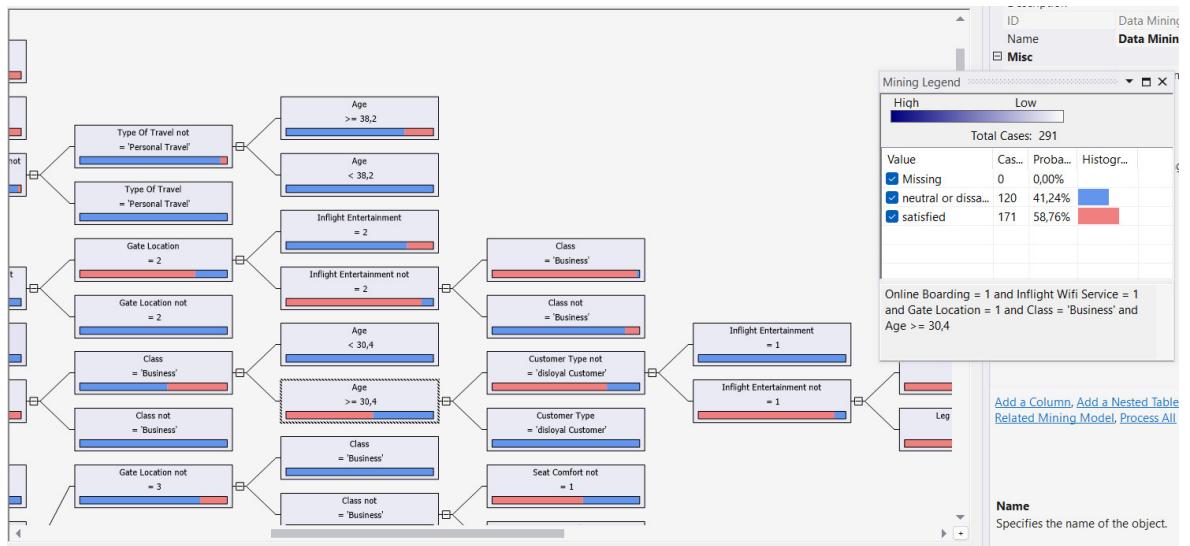
Với những chuyến bay cá nhân có khoảng cách lớn hơn 1516,6 km và nhỏ hơn 3002,2 km và có tiêu chí đánh giá “Baggage Handling”, “Gate Location” và “Inflight Wifi Service” cùng bằng 4 điểm, tỉ lệ không hài lòng đạt cao đến 80,86%.



Hình 566: Tỷ lệ đánh giá của khách hàng dựa trên “Online Boarding”, “Type of Travel”, “Baggage Handling”, “Inflight Wifi Service”, “Gate Location” và “Flight Distance”

Với chuyến đi của các khách hàng lớn hơn 30 tuổi sử dụng hạng ghế

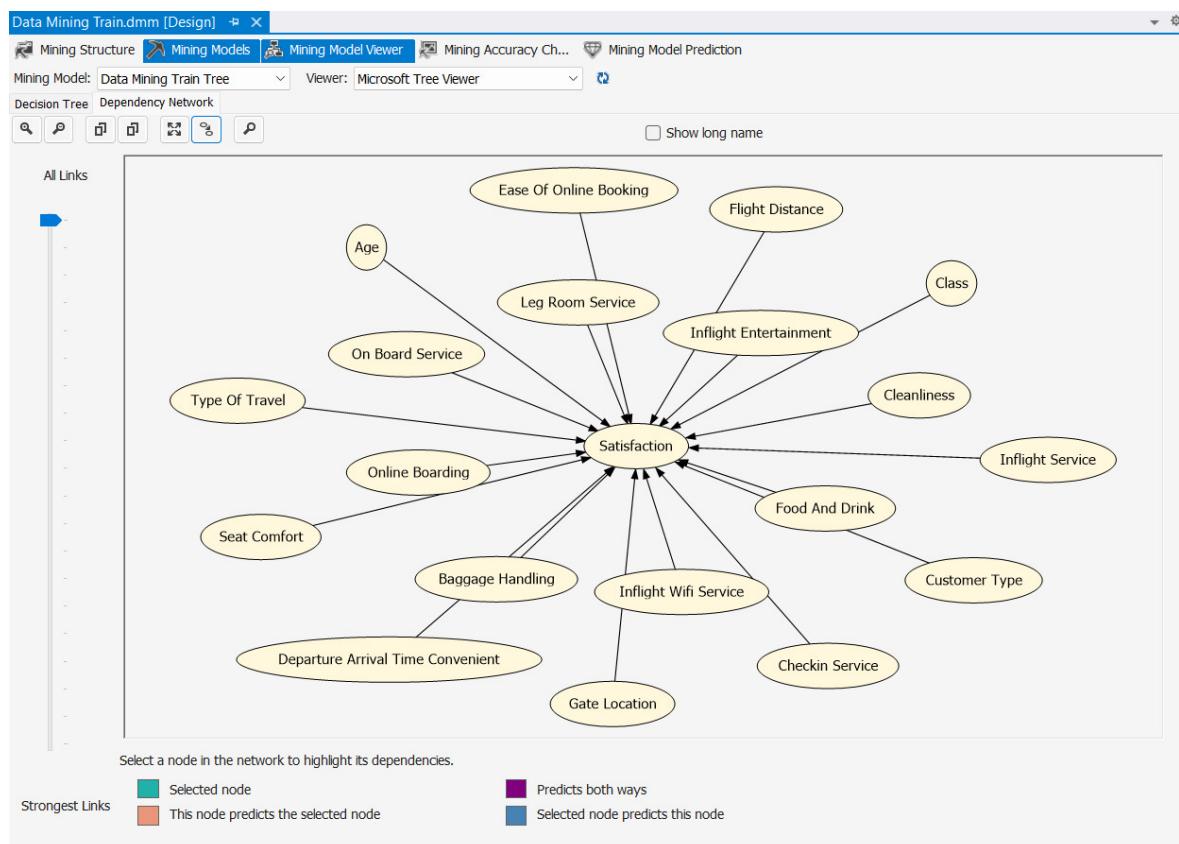
Thương gia, các tiêu chí đánh giá “Online Boarding”, “Inflight Wifi Service”, “Gate Location” đều đạt 1 điểm, tỉ lệ hài lòng vẫn ở mức tương đối ổn khoảng 58,76%.



Hình 567: Tỷ lệ đánh giá dựa trên “Online Boarding”, “Inflight Wifi Service”, “Gate Location”, “Class” và “Age”

Dependency Network của thuật toán Microsoft Decision Tree cho thấy những thuộc tính có ảnh hưởng đến việc đánh giá trải nghiệm chuyến bay. Những thuộc tính có ảnh hưởng theo mức độ giảm dần là:

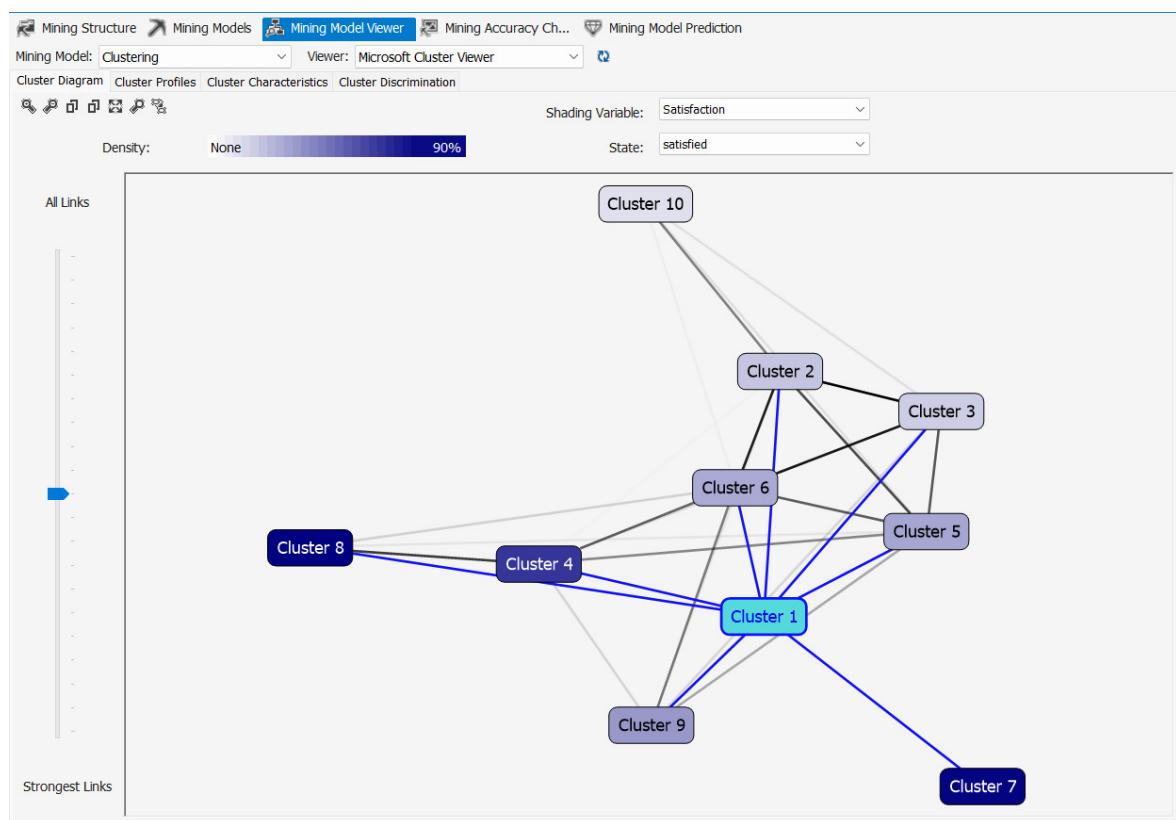
Online Boarding -> Type of Travel -> Inflight Wifi Service -> Baggage Handling -> Customer Type -> Gate Location -> Depature Arrival Time Convenient -> Class -> Inflight Service -> Inflight Entertainment -> Age -> Check-in Service -> Flight Distance -> Seat Comfort -> Leg Room Service -> Food And Drink -> Cleanliness -> On Board Service -> Ease of Online Booking



Hình 568: Sơ đồ những thuộc tính ảnh hưởng đến kết quả đánh giá của khách hàng

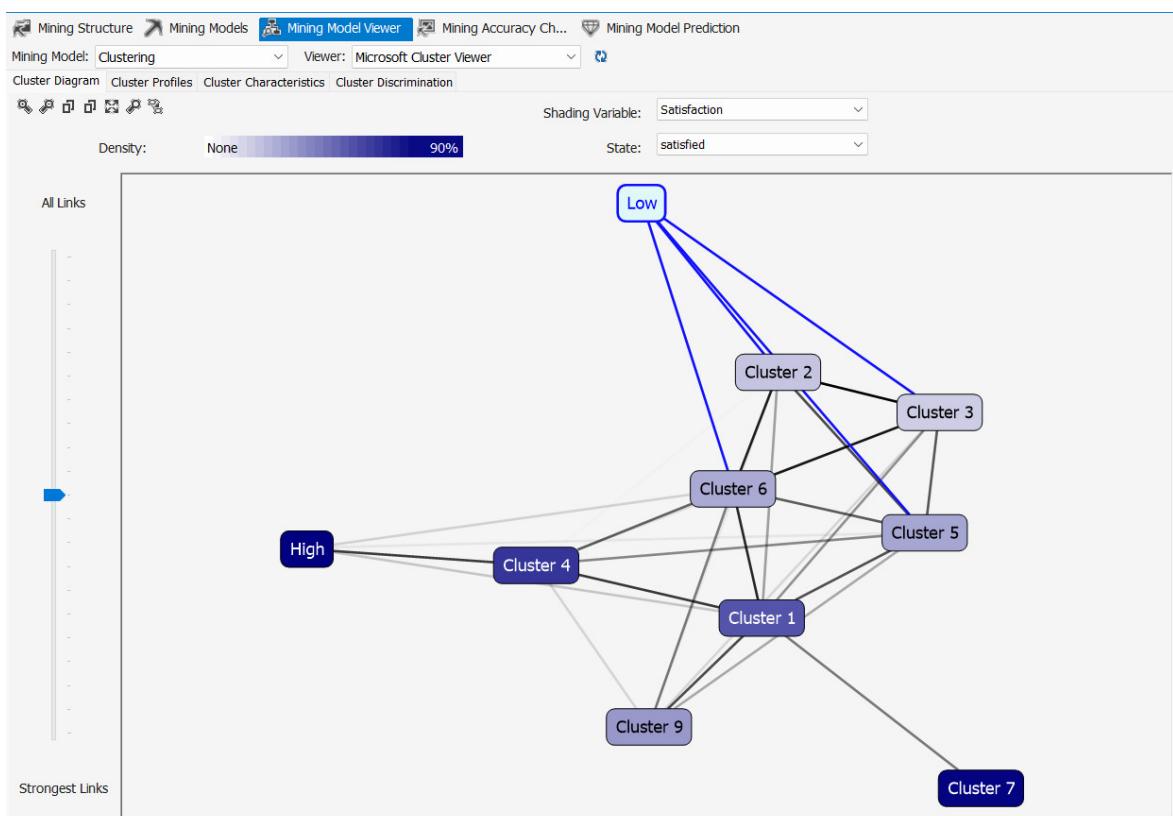
#### 4.5.2. Thuật toán Microsoft Clustering:

Cluster Diagram của thuật toán với Shading Variable là Satisfaction, value là satisfied (hài lòng).



Hình 569: Sơ đồ Cluster của thuật toán

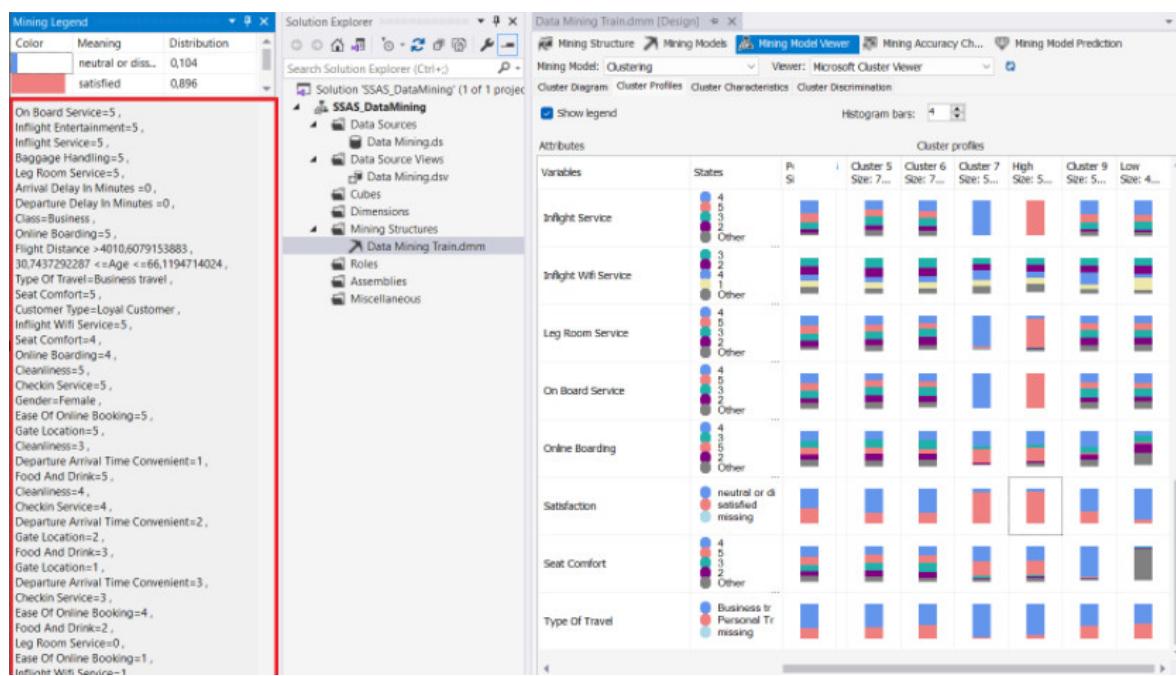
Đổi tên Cluster đậm màu nhất thành High (Khả năng hài lòng cao nhất). Đổi tên Cluster nhạt màu nhất thành Low (Khả năng trung lập hoặc không hài lòng cao nhất).



Hình 570: Đổi tên các cluster có độ hài lòng cao nhất và thấp nhất thành “High” và “Low”

Với Cluster Profiles của Cluster High cho thấy tập lục của những đánh giá hài lòng đạt tỉ lệ 89,6%.

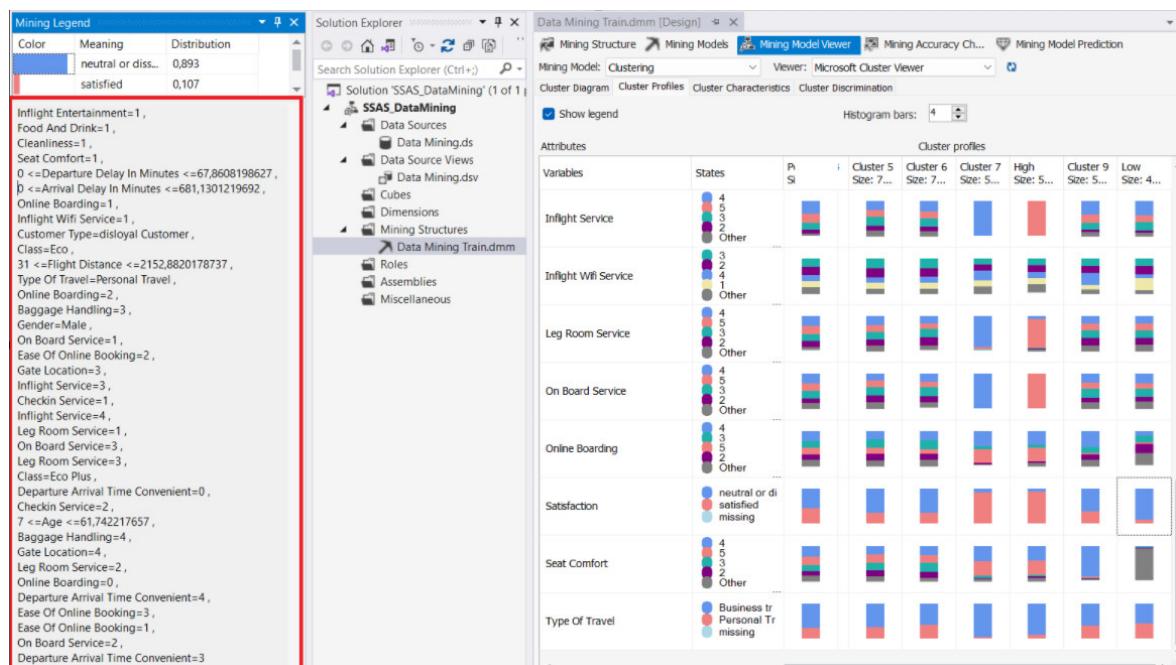
Tập lục: Được hiển thị trong khung màu đỏ của hình bên dưới.



Hình 571: Tập lục và kết quả của cluster có độ hài lòng cao nhất

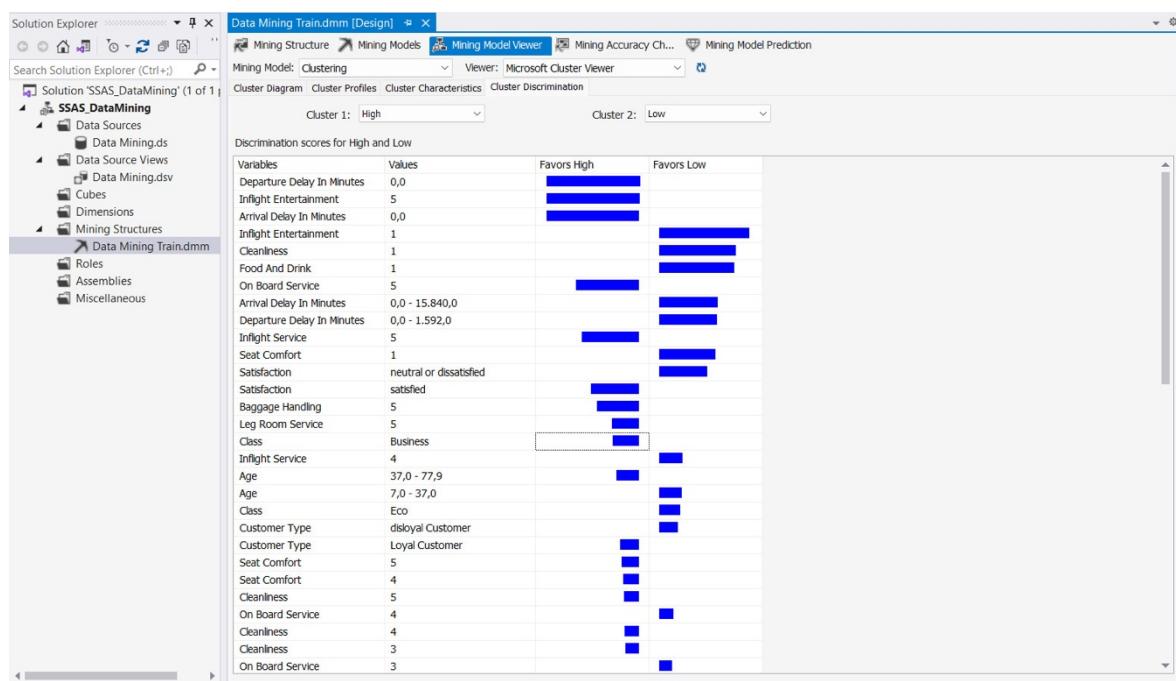
Với Cluster Profiles của Cluster Low cho thấy tập lục của những đánh giá trung lập hoặc không hài lòng đạt 89,3%.

Tập lục: Được hiển thị trong khung màu đỏ trong hình bên dưới.



Hình 572: Tập lục và kết quả của cluster có đánh giá trung lập hoặc không hài lòng cao nhất

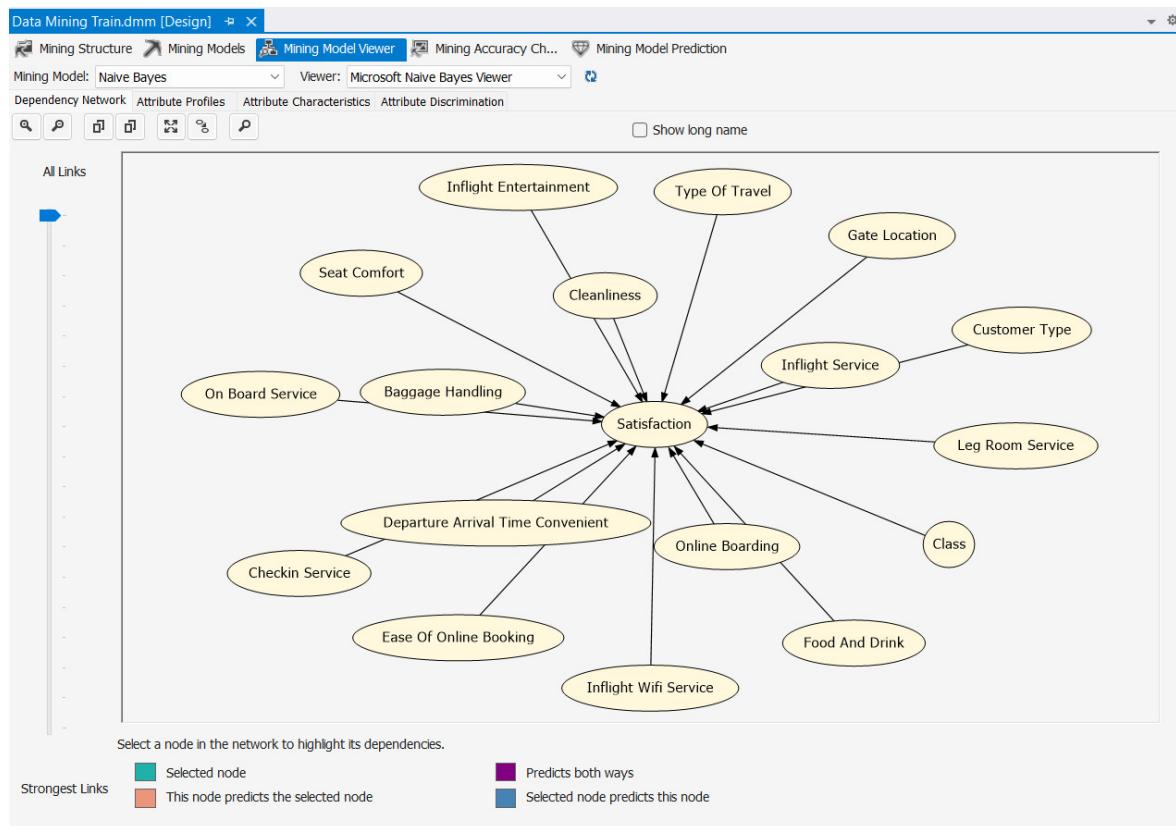
Cluster Discrimination của thuật toán cho ta thấy sự tương quan giữa hai Cluster High và Low.



Hình 573: Mối tương quan giữa hai cluster “High” và “Low”

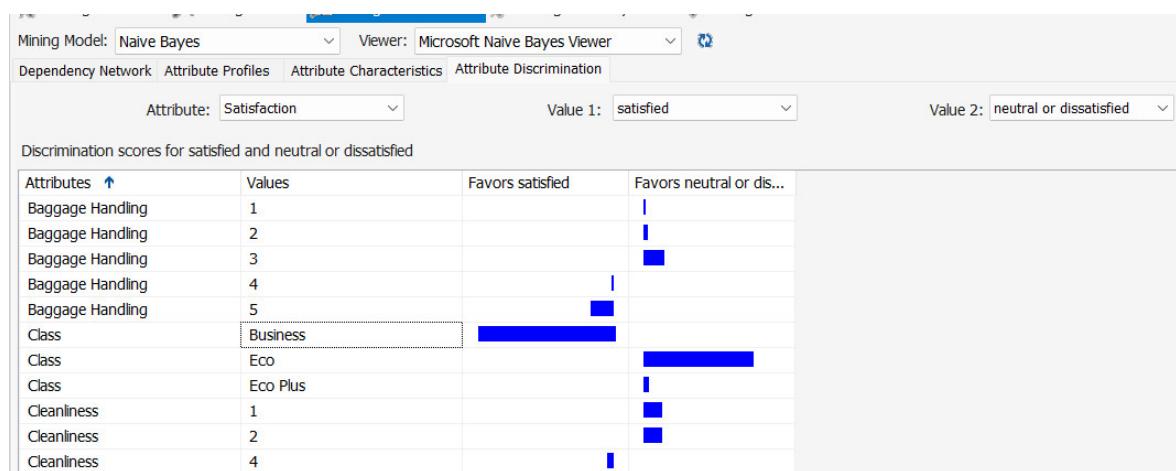
#### 4.5.3. Thuật toán Microsoft Naïve Bayes:

Dependency Network của thuật toán Microsoft Naïve Bayes cho ta thấy các thuộc tính ảnh hưởng tới kết quả đánh giá chuyến bay.



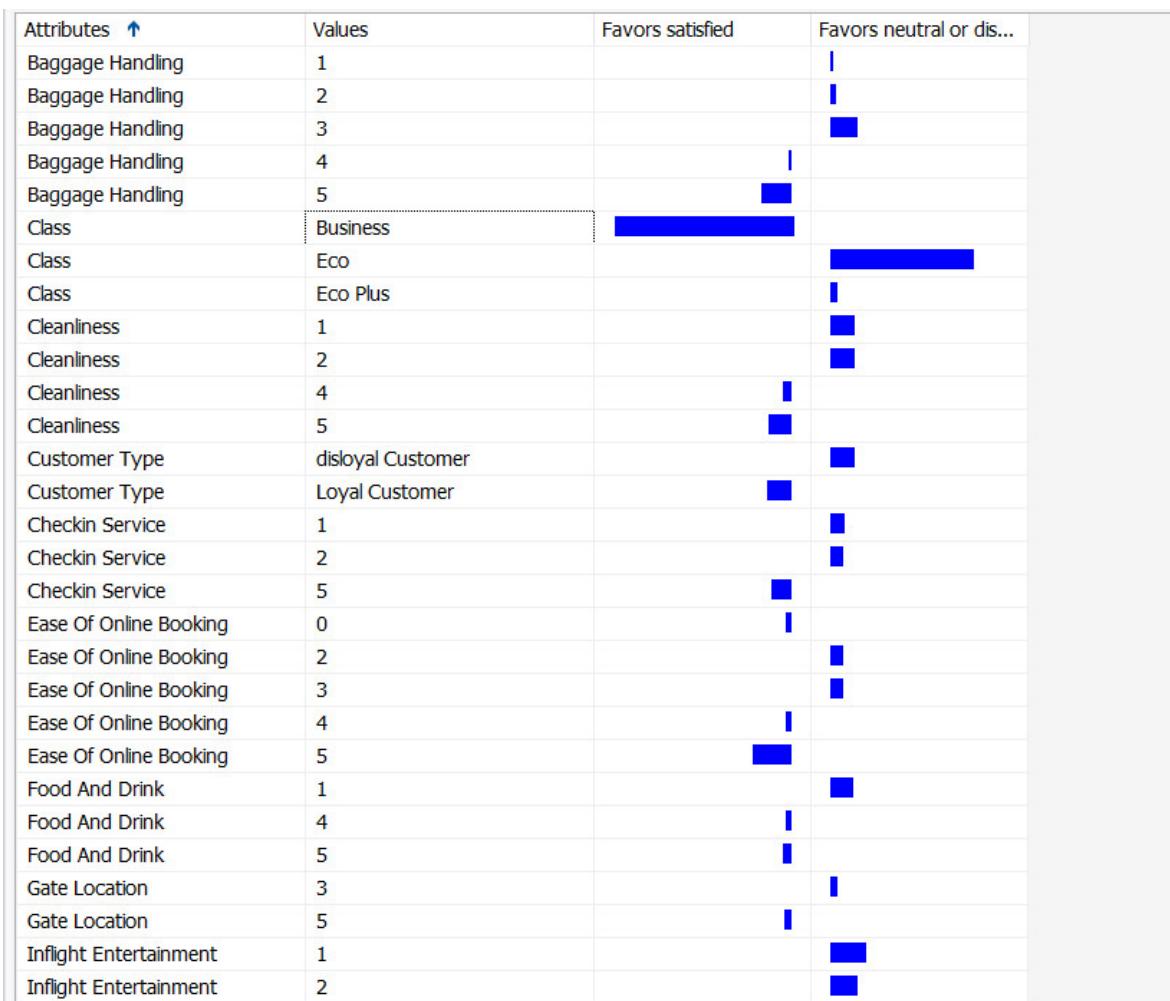
Hình 574: Sơ đồ những thuộc tính ảnh hưởng đến kết quả đánh giá của khách hàng

Trong ba hạng ghế, hạng ghế Thương Gia đạt được tỉ lệ hài lòng cao nhất.



Hình 575: Giá trị của thuộc tính “Class” ảnh hưởng đến kết quả đánh giá của hành khách

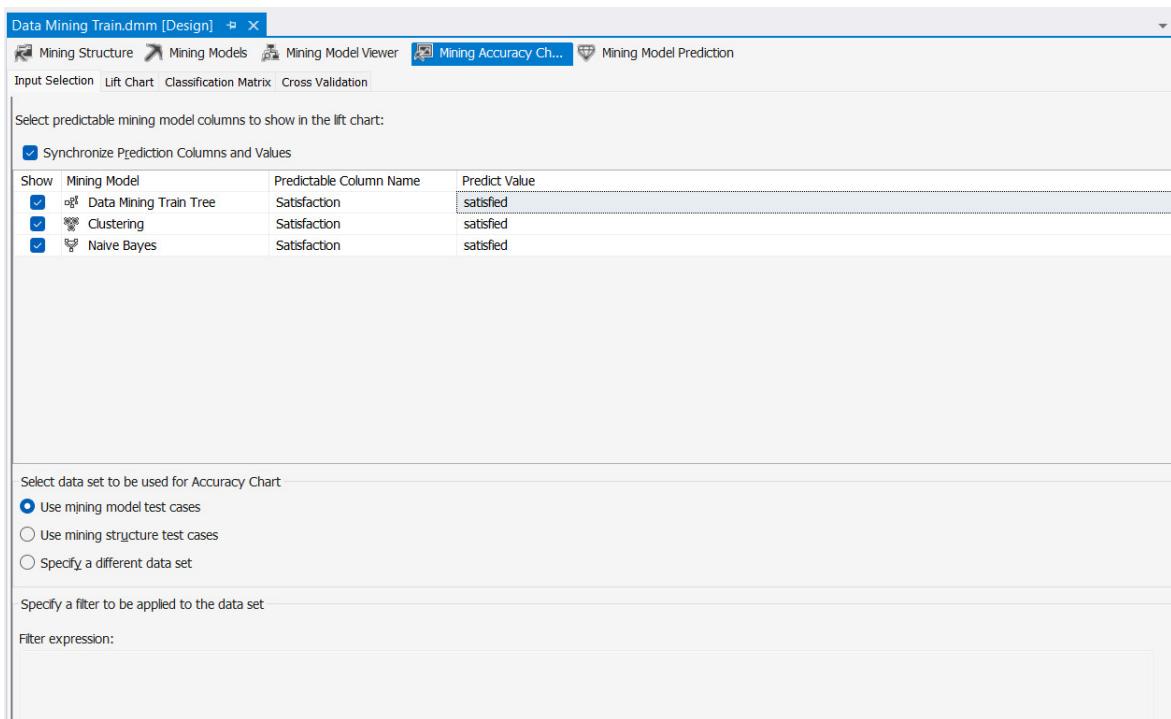
Các tiêu chí còn lại, điểm số đánh giá càng cao, tỉ lệ hài lòng càng lớn và ngược lại.



Hình 576: Một số thuộc tính còn lại ảnh hưởng đến kết quả đánh giá của hành khách

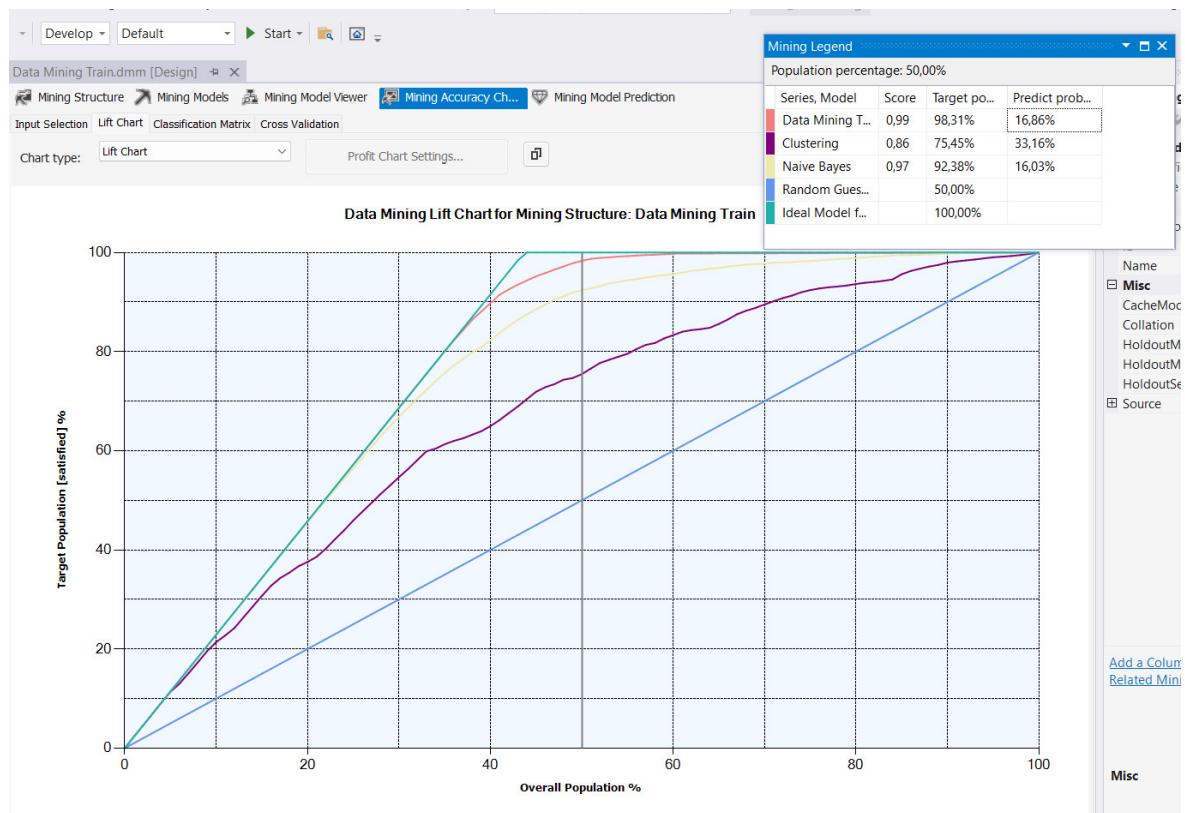
#### 4.5.4. Đánh giá thuật toán bằng Mining Accuracy Chart:

Thiết lập thông số Input Selection với các chuyến bay được đánh giá hài lòng (satisfied).



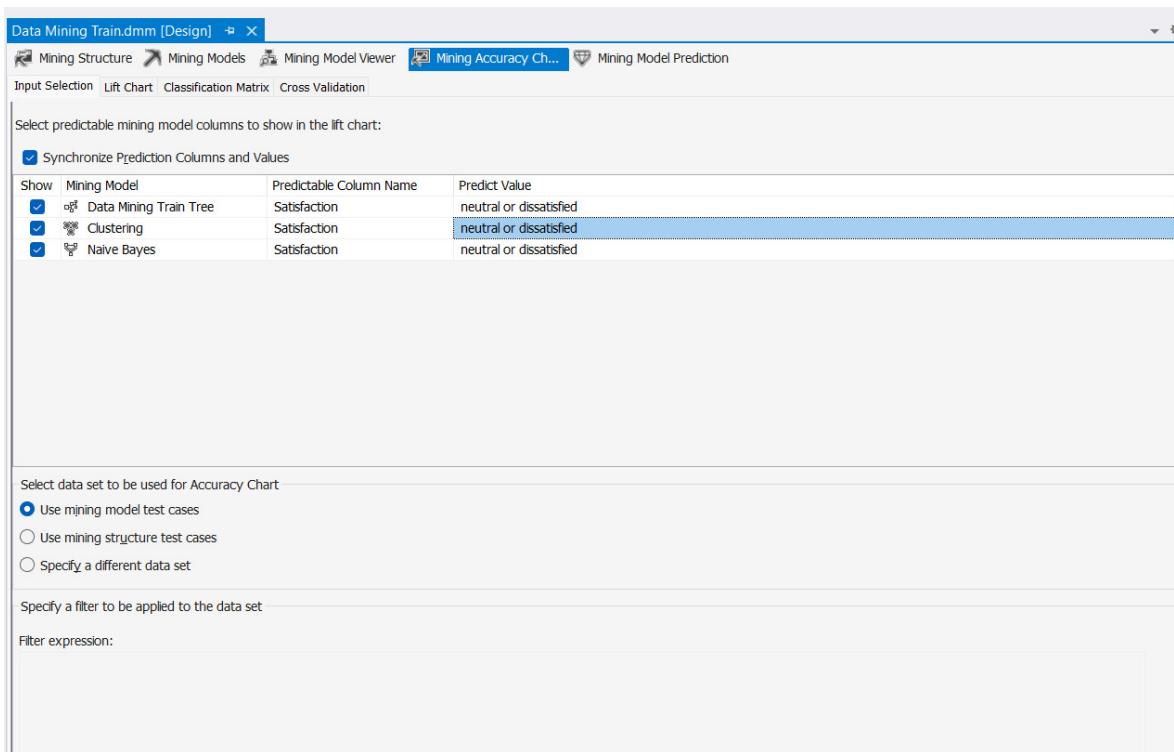
Hình 577: Thiết lập giá trị thông số Input Selection

Line chart cho thấy tỷ lệ chính xác của 3 thuật toán khi mining trong trường hợp chuyến bay được đánh giá hài lòng: Microsoft Decision Tree (0.99 Score), Microsoft Clustering (0.86 Score) và Microsoft Naïve Bayes (0.77 Score).



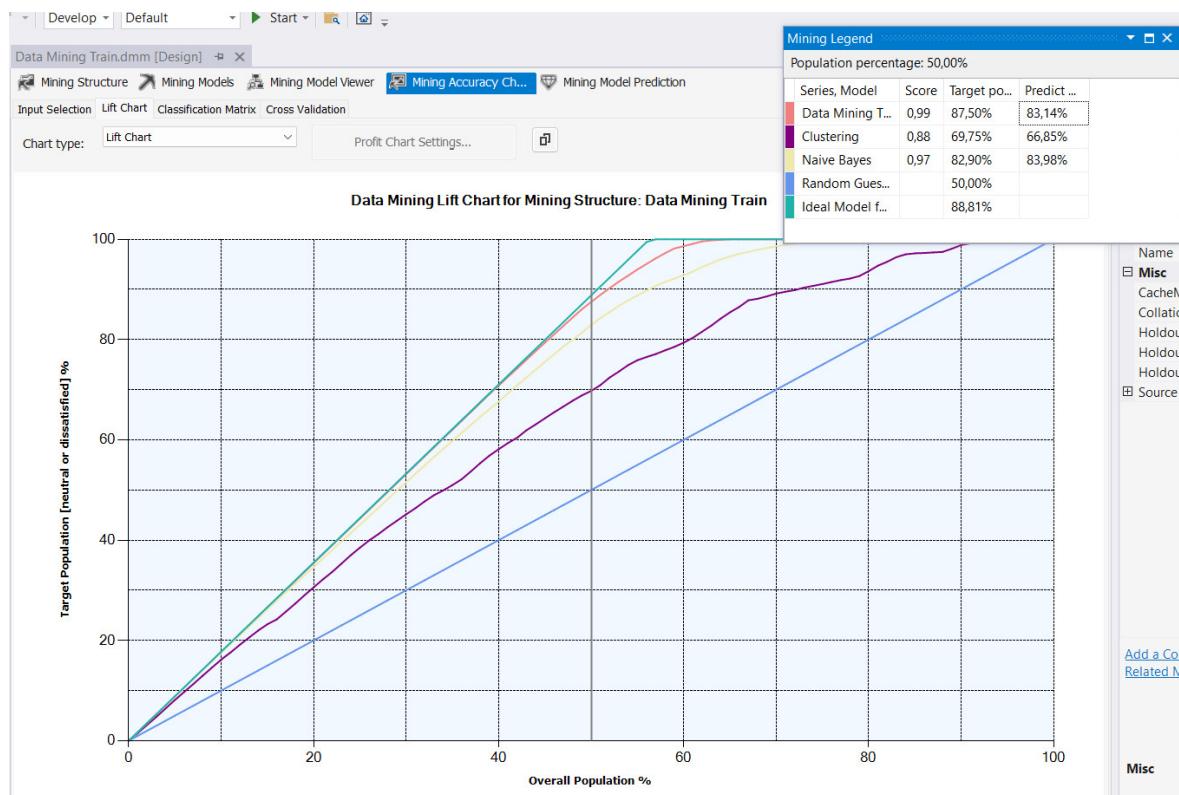
Hình 578: So sánh kết quả dự đoán của ba thuật toán thông qua line chart

Thiết lập thông số Input Selection với các chuyến bay được đánh giá trung lập hoặc không hài lòng.



Hình 579: Thiết lập giá trị thông số Input Selection

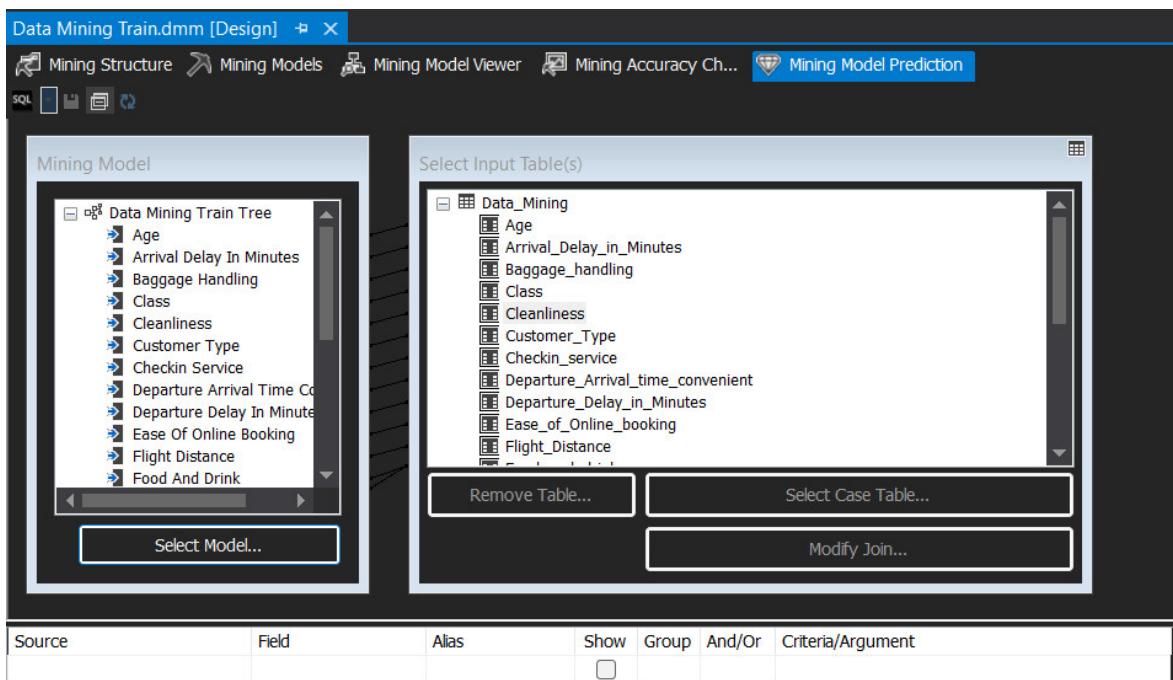
Line chart cho thấy tỷ lệ chính xác của 3 thuật toán khi mining trong trường hợp chuyến bay được đánh giá trung lập hoặc không hài lòng: Microsoft Decision Tree (0.99 Score), Microsoft Clustering (0.88 Score) và Microsoft Naïve Bayes (0.97 Score).



Hình 580: So sánh kết quả dự đoán của ba thuật toán thông qua line chart

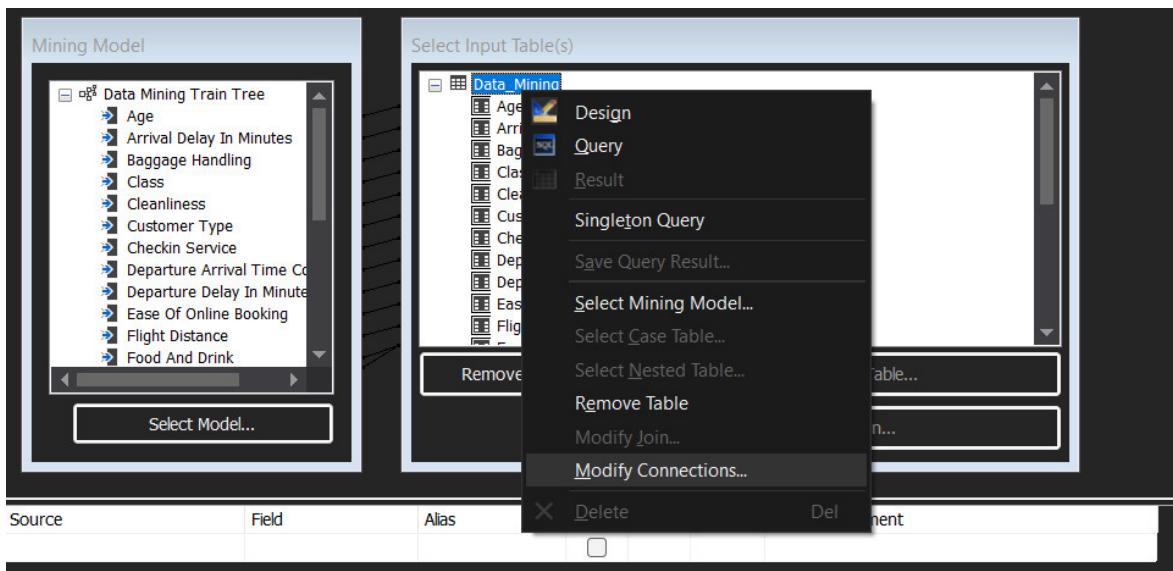
#### 4.6. Dự đoán:

**Bước 1:** Ta chọn “**Mining Model Prediction**” sẽ xuất hiện giao diện để tùy chỉnh.



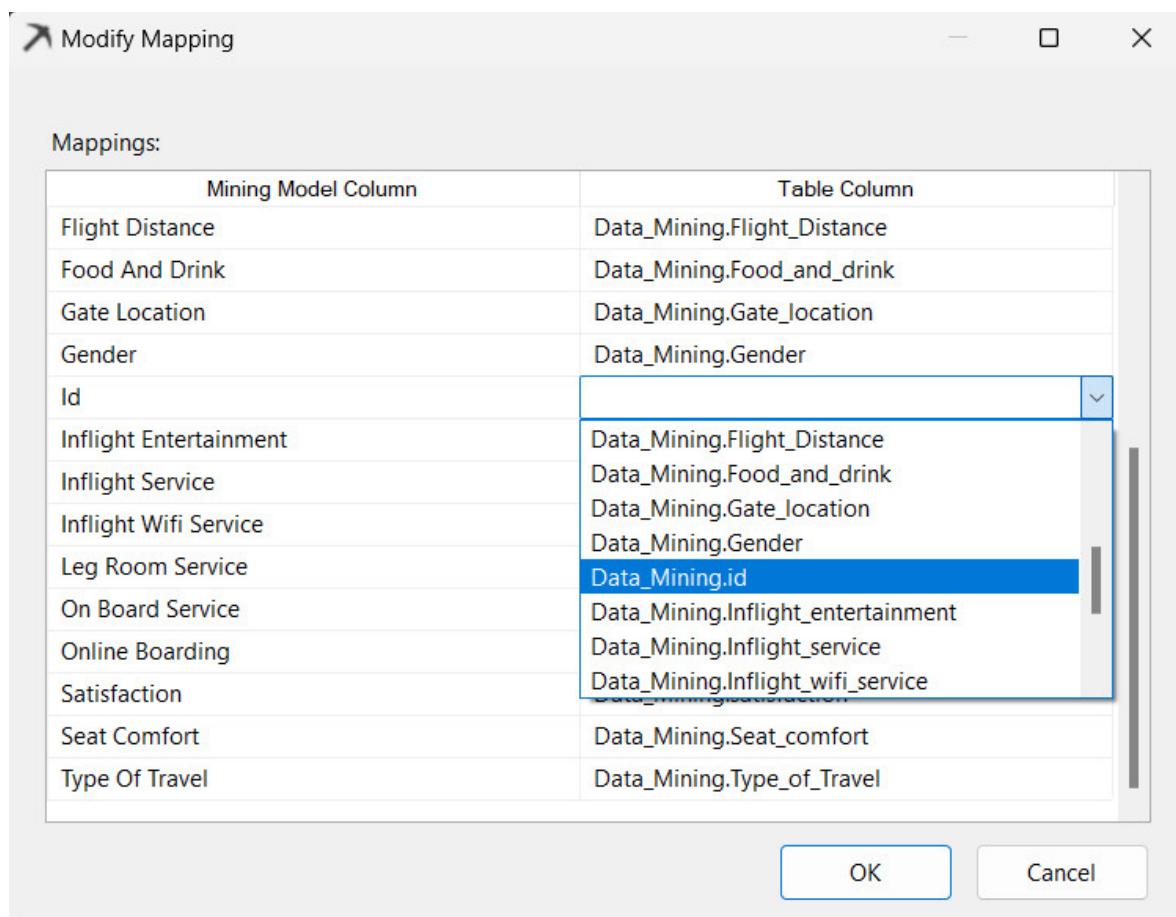
Hình 581: Giao diện Mining Model Prediction

**Bước 2:** Ta quan sát các liên kết cột dữ liệu của mô hình có khớp với cột của bảng dữ liệu chưa, ta có thể chỉnh sửa bằng cách click chuột phải vào một liên kết bất kỳ, chọn “Modify Connections”



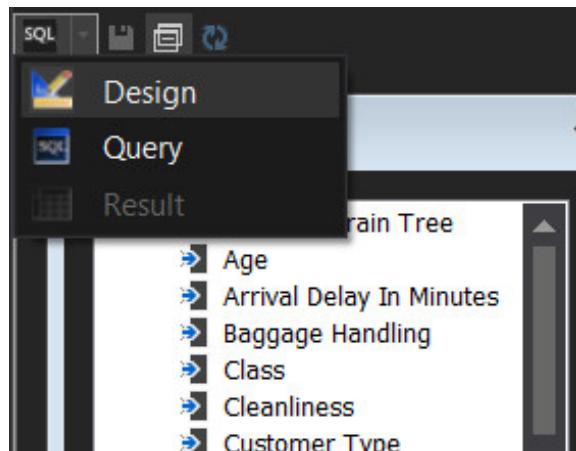
Hình 582: Kiểm tra ánh xạ các thuộc tính

**Bước 3:** Cửa sổ “Modify Mapping” hiện ra, ta thực hiện chọn các liên kết chính xác.



Hình 583: Giao diện cửa sổ Modify Mapping

**Bước 4:** Chọn mục “Design” tại nút nhấp vào nằm trên “Mining Models”. Tạo lệnh bằng Design giúp ta tiết kiệm thời gian viết lệnh SQL.



Hình 584: Chọn mục Design để kéo thả các thuộc tính và mô hình phục vụ cho việc dự đoán

**Bước 5:** Các hàng tùy chỉnh nằm bên dưới Mining Models, ta chọn “Source”, “Field” trong dropdown để lấy chức năng “Predict Probability”, đặt Alias là

“Probability”. Sau đó, kéo thả thuộc tính muốn dự đoán vào mục “Criteria/Argument”.

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Prediction Function	PredictProbability	ResultProbability	<input checked="" type="checkbox"/>			[Data Mining Train Tree].[Satisfaction]
			<input type="checkbox"/>			

Hình 585:Lựa chọn các thuộc tính và trường tính toán thích hợp

**Bước 6:** Tạo điều kiện cho câu lệnh Where với mô hình Cây quyết định và cột satisfaction trong dropdown “Source”, “Field”, “Criteria/Argument” là satisfied.

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Prediction Function	PredictProbability	ResultProbability	<input checked="" type="checkbox"/>			[Data Mining Train Tree].[Satisfaction]
Data Mining Train Tree	Satisfaction		<input checked="" type="checkbox"/>			=‘satisfied’
			<input type="checkbox"/>			

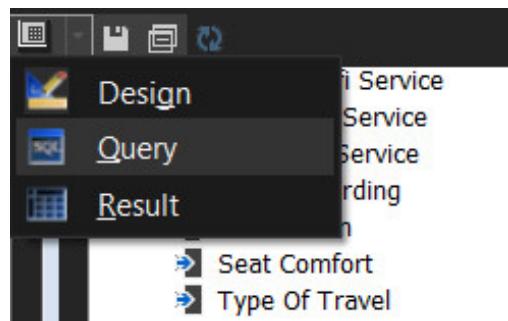
Hình 586:Thêm điều kiện giá trị cho biến satisfaction

**Bước 7:** Thêm các cột thông tin sẽ được truy vấn, phải có ít nhất cột khóa chính, Source lấy tên bảng từ dropdown.

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
Prediction Function	PredictProbability	ResultProbability	<input checked="" type="checkbox"/>			[Data Mining Train Tree].[Satisfaction]
Data Mining Train Tree	Satisfaction		<input checked="" type="checkbox"/>			=‘satisfied’
Data_Mining	id		<input checked="" type="checkbox"/>			
Data_Mining	Online_boarding		<input checked="" type="checkbox"/>			
Data_Mining	Type_of_Travel		<input checked="" type="checkbox"/>			
Data_Mining	Inflight_wifi_ser...		<input checked="" type="checkbox"/>			
Data_Mining	Baggage_hand...		<input checked="" type="checkbox"/>			

Hình 587: Thêm các thuộc tính phục vụ cho câu truy vấn phải có ít nhất cột khóa chính

**Bước 8:** Thực hiện kiểm tra câu lệnh truy vấn được tạo tự động trong mục “Query”.

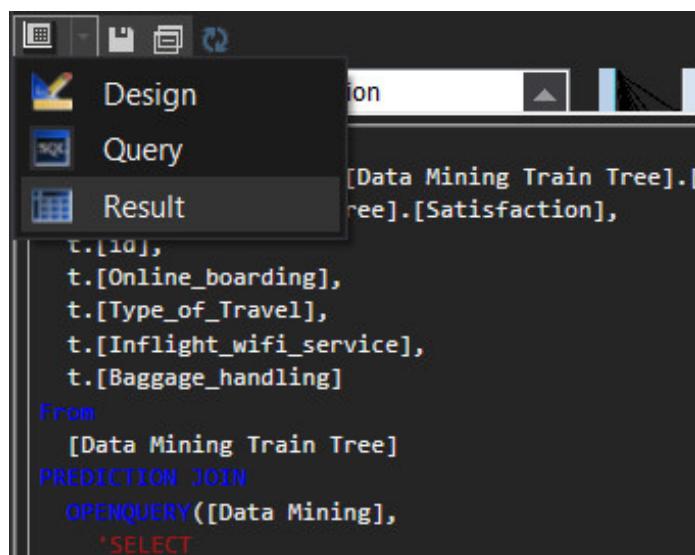


Hình 588: Chọn mục Query để kiểm tra câu truy vấn

```
 SELECT  
    (PredictProbability([Data Mining Train Tree].[Satisfaction])) as [ResultProbability],  
    [Data Mining Train Tree].[Satisfaction],  
    t.[id],  
    t.[Online_boarding],  
    t.[Type_of_Travel],  
    t.[Inflight_wifi_service],  
    t.[Baggage_handling]  
From  
    [Data Mining Train Tree]  
PREDICTION JOIN  
OPENQUERY([Data Mining],  
    'SELECT  
        [id],  
        [Online_boarding],  
        [Type_of_Travel],  
        [Inflight_wifi_service],  
        [Baggage_handling],  
        [Gender],  
        [Customer_Type],  
        [Age],  
        [Class],  
        [Flight_Distance],  
        [Departure_Arrival_time_convenient],  
        [Ease_of_Online_booking],  
        [Gate_location],  
        [Food_and_drink],  
        [Seat_comfort],  
        [Inflight_entertainment],  
        [On_board_service],  
        [Leg_room_service],  
        [Checkin_service],  
        [Inflight_service].
```

Hình 589: Kiểm tra câu truy vấn đã đúng hay chưa

**Bước 9:** Xem kết quả truy vấn bằng cách chọn “Result”.



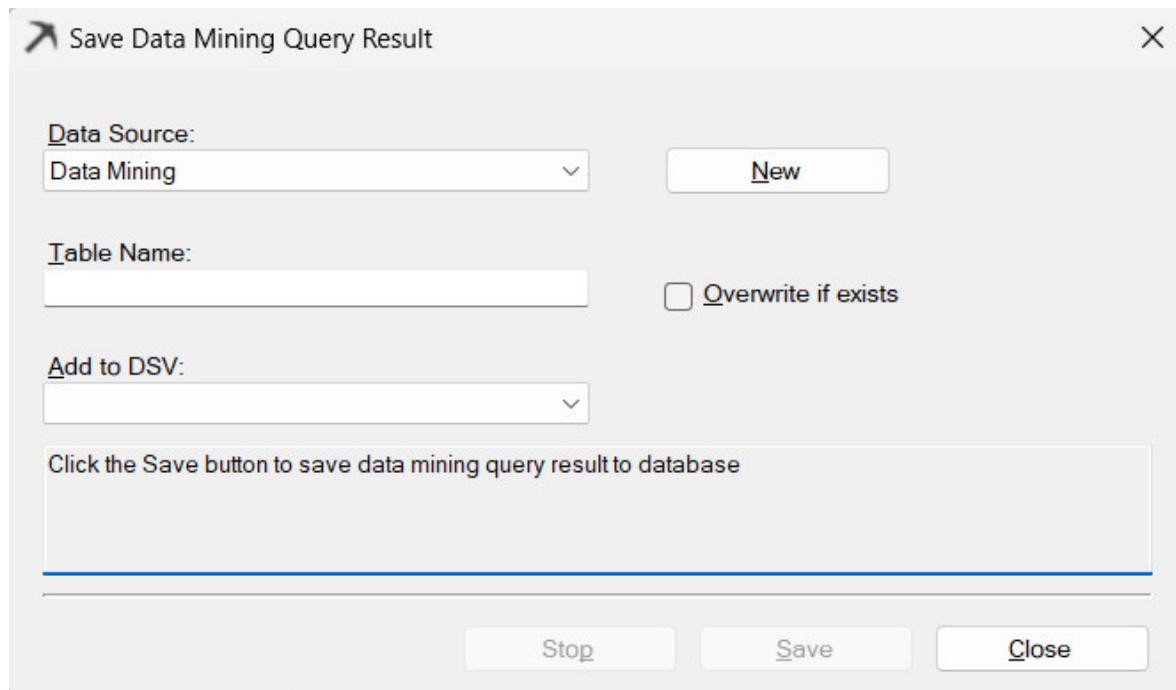
Hình 590: Chon mục “Result” để kiểm tra kết quả dự đoán

Ta nhận thấy kết quả truy vấn dự đoán một đánh giá hài lòng đạt 99% đối với những đánh giá thực sự hài lòng. Kết quả này được xem như mô hình dự đoán tương đối tốt.

ResultProba...	Satisfaction	id	Online_boar...	Type_of_Tr...	Inflight_wifi_...	Baggage_ha...
0,9996282...	satisfied	2	5	Business tra...	2	5
0,9996282...	satisfied	3	5	Business tra...	4	3
0,9999992...	satisfied	4	4	Business tra...	2	5
0,9996282...	satisfied	5	5	Business tra...	3	3
0,9996282...	satisfied	6	5	Business tra...	4	4
0,9999992...	satisfied	7	4	Business tra...	3	5
0,9996282...	satisfied	10	5	Business tra...	2	5
0,9701030...	satisfied	25	4	Business tra...	3	5
0,9999999...	satisfied	26	4	Business tra...	4	4
0,9999992...	satisfied	27	4	Business tra...	3	5
0,9996282...	satisfied	28	5	Business tra...	0	3
0,9999992...	satisfied	29	4	Business tra...	4	5
0,9207918...	satisfied	30	4	Business tra...	4	4
0,9996282...	satisfied	34	5	Business tra...	0	5
0,9999992...	satisfied	36	4	Business tra...	5	5
0,9996282...	satisfied	37	5	Business tra...	4	5
0,9701030...	satisfied	38	4	Business tra...	4	5
0,9996282...	satisfied	39	5	Business tra...	3	4
0,9999987...	satisfied	40	4	Business tra...	2	4
0,9999987...	satisfied	41	4	Business tra...	2	4
0,9996282...	satisfied	42	5	Business tra...	4	5
n 9999987	catisfied	43	4	Routine tra...	2	4

Hình 591: Kết quả dự đoán của mô hình Cây quyết định

**Bước 10:** Ta có thể lưu kết quả câu truy vấn bằng cách click chuột phải vào kết quả hiển thị và cửa sổ “Save Data Mining Query Result” được tạo ra. Click “Save” để lưu.



Hình 592: Lưu kết quả dự đoán

## 4.7. Phân tích Dataset gốc:

### 4.7.1. Thống kê mô tả:

Tính toán đại lượng thống kê mô tả: Count, Min, Max, Mean, Median, Quantile, Range, Mode và Variance trên tập dữ liệu.

**Bước 1:** Import các thư viện cần thiết và đọc dữ liệu từ file .csv.

```

import pandas as pd
import numpy as np
[1]: ✓ 0.5s
[2]: ✓ 0.2s
Airlane_DF = pd.read_csv('Airlane_Passenger_Satisfaction.csv')
Airlane_DF = Airlane_DF.loc[:, ~Airlane_DF.columns.str.contains('^\u0302named')]

Airlane_DF
[3]: ✓ 0.2s

```

	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	... entertainment	Inflight service	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness
0	19556	Female	Loyal Customer	52	Business travel	Eco	160	5	4	3	...	5	5	5	5	2	5	
1	90035	Female	Loyal Customer	36	Business travel	Business	2863	1	1	3	...	4	4	4	4	3	4	
2	12360	Male	disloyal Customer	20	Business travel	Eco	192	2	0	2	...	2	4	1	3	2	2	
3	77959	Male	Loyal Customer	44	Business travel	Business	3377	0	0	0	...	1	1	1	1	3	1	
4	36875	Female	Loyal Customer	49	Business travel	Eco	1182	2	3	4	...	2	2	2	2	4	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
129875	94171	Female	disloyal Customer	23	Business travel	Eco	192	2	1	2	...	2	3	1	4	2	3	
129876	73097	Male	Loyal Customer	49	Business travel	Business	2347	4	4	4	...	5	5	5	5	5	5	
129877	68825	Male	disloyal Customer	30	Business travel	Business	1995	1	1	1	...	4	3	2	4	5	5	
129878	54173	Female	disloyal Customer	22	Business travel	Eco	1000	1	1	1	...	1	4	5	1	5	4	
129879	62567	Male	Loyal Customer	27	Business travel	Business	1723	1	3	3	...	1	1	1	4	4	3	

129880 rows × 24 columns

Hình 593: Thêm dữ liệu và các thư viện cần thiết

**Bước 2:** Lọc ra các cột dữ liệu số để tính toán.

```

numeric_cols = Airlane_DF.select_dtypes(include='number').columns
numeric_cols
[3]: ✓ 0.0s

```

```

... Index(['id', 'Age', 'Flight Distance', 'Inflight wifi service',
       'Departure\Arrival time convenient', 'Ease of Online booking',
       'Gate location', 'Food and drink', 'Online boarding', 'Seat comfort',
       'Inflight entertainment', 'On-board service', 'Leg room service',
       'Baggage handling', 'Checkin service', 'Inflight service',
       'Cleanliness', 'Departure Delay in Minutes',
       'Arrival Delay in Minutes'],
      dtype='object')

```

Hình 594: Lọc ra các cột dữ liệu kiểu số

**Bước 3:** Sử dụng phương thức `describe()` để tính toán thống kê cơ bản về số lượng, trung bình, độ lệch chuẩn, giá trị tối thiểu, giá trị tối đa và các phân vị cho tất cả các cột kiểu số.

	count	mean	std	min	25%	50%	75%	max
id	129880.0	64940.500000	37493.270818	1.0	32470.75	64940.5	97410.25	129880.0
Age	129880.0	39.427957	15.119360	7.0	27.00	40.0	51.00	85.0
Flight Distance	129880.0	1190.316392	997.452477	31.0	414.00	844.0	1744.00	4983.0
Inflight wifi service	129880.0	2.728696	1.329340	0.0	2.00	3.0	4.00	5.0
Departure/Arrival time convenient	129880.0	3.057599	1.526741	0.0	2.00	3.0	4.00	5.0
Ease of Online booking	129880.0	2.756876	1.401740	0.0	2.00	3.0	4.00	5.0
Gate location	129880.0	2.976925	1.278520	0.0	2.00	3.0	4.00	5.0
Food and drink	129880.0	3.204774	1.329933	0.0	2.00	3.0	4.00	5.0
Online boarding	129880.0	3.252633	1.350719	0.0	2.00	3.0	4.00	5.0
Seat comfort	129880.0	3.441361	1.319289	0.0	2.00	4.0	5.00	5.0
Inflight entertainment	129880.0	3.358077	1.334049	0.0	2.00	4.0	4.00	5.0
On-board service	129880.0	3.383023	1.287099	0.0	2.00	4.0	4.00	5.0
Leg room service	129880.0	3.350878	1.316252	0.0	2.00	4.0	4.00	5.0
Baggage handling	129880.0	3.632114	1.180025	1.0	3.00	4.0	5.00	5.0
Checkin service	129880.0	3.306267	1.266185	0.0	3.00	3.0	4.00	5.0
Inflight service	129880.0	3.642193	1.176669	0.0	3.00	4.0	5.00	5.0
Cleanliness	129880.0	3.286326	1.313682	0.0	2.00	3.0	4.00	5.0
Departure Delay in Minutes	129880.0	14.713713	38.071126	0.0	0.00	0.0	12.00	1592.0
Arrival Delay in Minutes	129487.0	15.091129	38.465650	0.0	0.00	0.0	13.00	1584.0

Hình 595: Thống kê sơ bộ một số giá trị cơ bản của các cột kiểu số

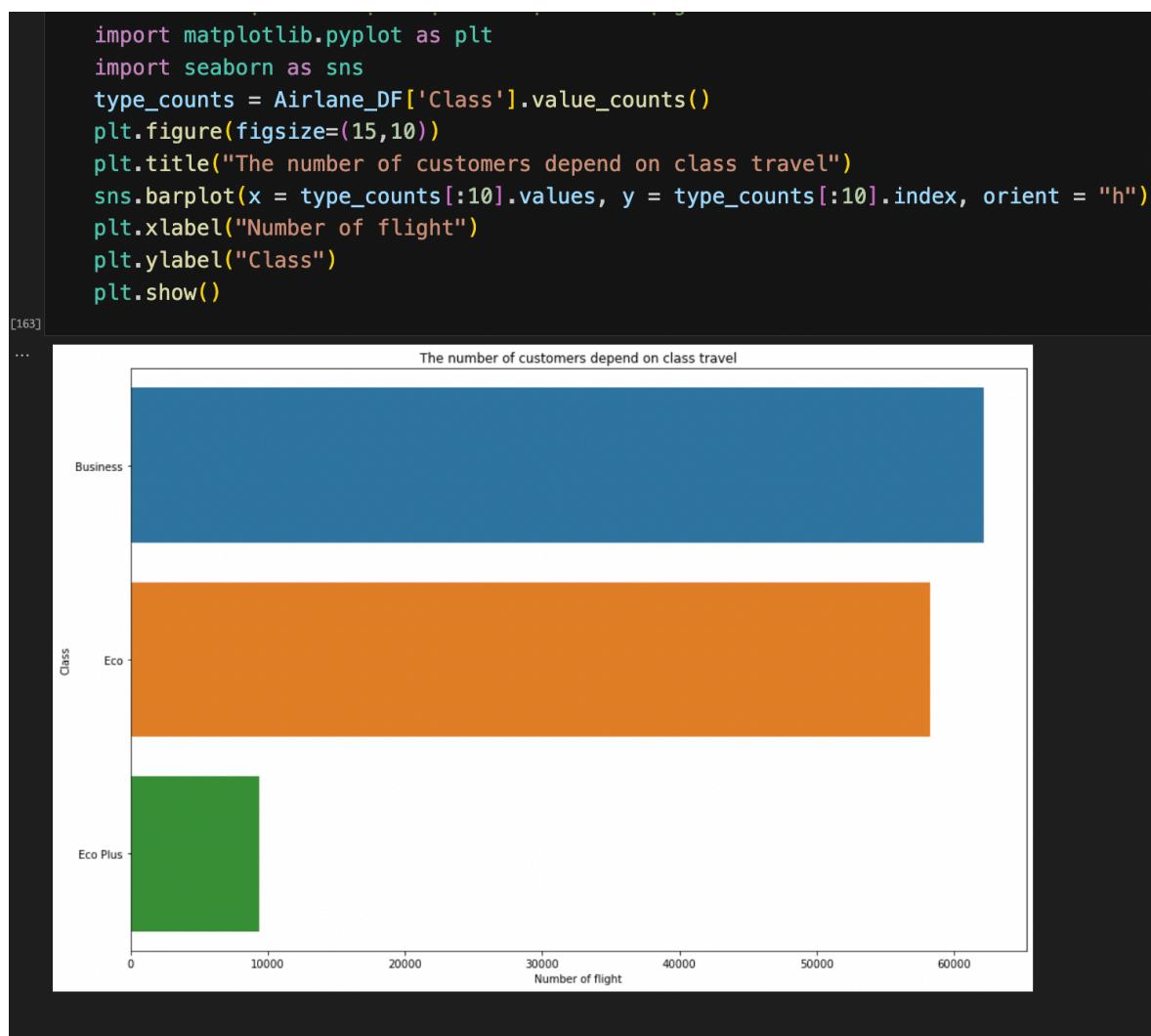
**Bước 4:** Tính toán một vài giá trị mode, range và variance cho các cột kiểu số.

	count	mean	std	min	25%	50%	75%	max	mode	range	variance
id	129880.0	64940.500000	37493.270818	1.0	32470.75	64940.5	97410.25	129880.0	1.0	129879.0	1.405745e+09
Age	129880.0	39.427957	15.119360	7.0	27.00	40.0	51.00	85.0	39.0	78.0	2.285950e+02
Flight Distance	129880.0	1190.316392	997.452477	31.0	414.00	844.0	1744.00	4983.0	337.0	4952.0	9.949114e+05
Inflight wifi service	129880.0	2.728696	1.329340	0.0	2.00	3.0	4.00	5.0	2.0	5.0	1.767145e+00
Departure/Arrival time convenient	129880.0	3.057599	1.526741	0.0	2.00	3.0	4.00	5.0	4.0	5.0	2.330940e+00
Ease of Online booking	129880.0	2.756876	1.401740	0.0	2.00	3.0	4.00	5.0	3.0	5.0	1.964874e+00
Gate location	129880.0	2.976925	1.278520	0.0	2.00	3.0	4.00	5.0	3.0	5.0	1.634613e+00
Food and drink	129880.0	3.204774	1.329933	0.0	2.00	3.0	4.00	5.0	4.0	5.0	1.768722e+00
Online boarding	129880.0	3.252633	1.350719	0.0	2.00	3.0	4.00	5.0	4.0	5.0	1.824441e+00
Seat comfort	129880.0	3.441361	1.319289	0.0	2.00	4.0	5.00	5.0	4.0	5.0	1.740523e+00
Inflight entertainment	129880.0	3.358077	1.334049	0.0	2.00	4.0	4.00	5.0	4.0	5.0	1.779687e+00
On-board service	129880.0	3.383023	1.287099	0.0	2.00	4.0	4.00	5.0	4.0	5.0	1.656625e+00
Leg room service	129880.0	3.350878	1.316252	0.0	2.00	4.0	4.00	5.0	4.0	5.0	1.732519e+00
Baggage handling	129880.0	3.632114	1.180025	1.0	3.00	4.0	5.00	5.0	4.0	4.0	1.392458e+00
Checkin service	129880.0	3.306267	1.266185	0.0	3.00	3.0	4.00	5.0	4.0	5.0	1.603225e+00
Inflight service	129880.0	3.642193	1.176669	0.0	3.00	4.0	5.00	5.0	4.0	5.0	1.384550e+00
Cleanliness	129880.0	3.286326	1.313682	0.0	2.00	3.0	4.00	5.0	4.0	5.0	1.725761e+00
Departure Delay in Minutes	129880.0	14.713713	38.071126	0.0	0.00	0.0	12.00	1592.0	0.0	1592.0	1.449411e+03
Arrival Delay in Minutes	129487.0	15.091129	38.465650	0.0	0.00	0.0	13.00	1584.0	0.0	1584.0	1.479606e+03

Hình 596: Kết quả thống kê mô tả dữ liệu

#### 4.7.2. Trực quan hóa dữ liệu:

Trong ba dạng ghép, dạng ghép nào được ưa chuộng nhất.



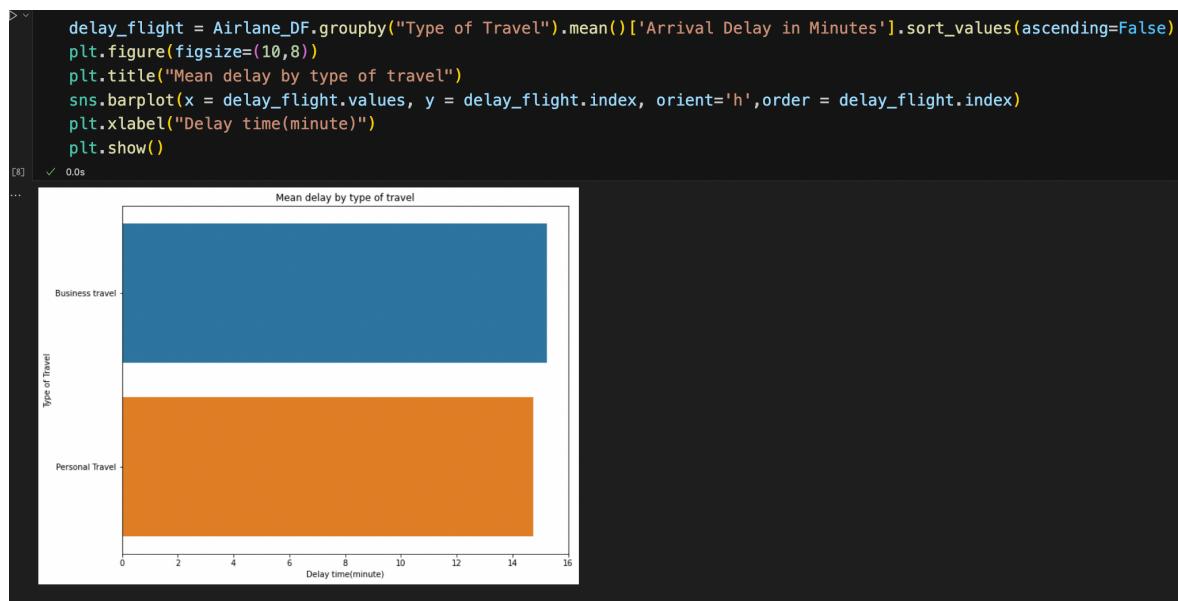
Hình 597: Trục quan hóa dữ liệu cột Class

Nhìn vào biểu đồ, ta có thể quan sát được hạng ghế thương gia (Business class) được nhiều khách hàng sử dụng nhất với 62160 lượt khách sử dụng (chiếm khoảng 47,86%). Vị trí thứ hai thuộc về hạng phổ thông với 58309 lượt khách sử dụng (chiếm khoảng 44,89%). Đứng cuối cùng là hạng ghế phổ thông đặc biệt (Eco Plus Class) với số lượt khách sử dụng chỉ đạt 9411 lượt (chiếm 7,25%).

Qua đó, ta có thể kết luận rằng, hạng ghế thương gia chiếm ưu thế do nhu cầu đi lại của các quan chức cấp cao, tầng lớp thượng lưu, các tỉ phú tăng cao. Họ mong muốn được trải nghiệm các dịch vụ cao cấp đặc biệt dành riêng cho họ với tốc độ nhanh nhất không cần phải chờ đợi. Hạng ghế phổ thông đứng thứ hai phần lớn do nhu cầu đi du lịch, về quê, đi công tác của mọi người. Hạng ghế này chiếm

số lượng tương đối lớn do yếu tố quan trọng nhất là giá thành rẻ và đảm bảo được nhu cầu cơ bản trong việc di chuyển xa của bộ phận khách hàng.

### Tính toán trung bình thời gian trễ dựa trên loại hình di chuyển



Hình 598: Trực quan hóa trung bình thời gian trễ của hai loại hình di chuyển

Nhìn qua biểu đồ, ta nhận thấy không có sự cách biệt quá lớn trong thời gian trễ giữa hai loại hình dịch vụ. Thời gian trễ ở cả hai loại hình dịch vụ trung bình là 15 phút.

## 4.8. Tiền xử lý dữ liệu:

### 4.8.1. Kiểm tra mối tương quan giữa các thuộc tính:

Nhóm trình bày ma trận tương quan giữa tất cả các thuộc tính, dưới dạng biểu đồ nhiệt. Từ đó, chúng ta có thể xem mối liên hệ giữa chúng. Nếu thuộc tính nào có mối tương quan thấp so với các thuộc tính còn lại, nhóm sẽ xem xét loại bỏ chúng ra khỏi tập dữ liệu.



Hình 599: Ma trận tương quan giữa các thuộc tính trong bộ dữ liệu

Từ ma trận này, chúng ta có thể quan sát được thuộc tính “Inflight wifi service”, “Departure/Arrival time convenient”, “Ease of Online Booking”, “Gate location” có mối tương quan cao.

Hơn nữa, hai thuộc tính “Departure Delay in Minutes” tỷ lệ thuận với “Arrival Delay in Minutes”, vì vậy, ta có thể cân nhắc loại bỏ một trong hai thuộc tính này.

Nhóm chỉ xóa đi một thuộc tính duy nhất là thuộc tính id vì chúng không ảnh hưởng đến việc đánh giá trải nghiệm của khách hàng. Số lượng cột thuộc tính còn lại là 23 thuộc tính.

```

X = Airlane_DF.drop('id', axis = 1)
X.info()

[12] ✓ 0.0s

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 129880 entries, 0 to 129879
Data columns (total 23 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Gender          129880 non-null   object 
 1   Customer Type  129880 non-null   object 
 2   Age             129880 non-null   int64  
 3   Type of Travel 129880 non-null   object 
 4   Class            129880 non-null   object 
 5   Flight Distance 129880 non-null   int64  
 6   Inflight wifi service 129880 non-null   int64  
 7   Departure/Arrival time convenient 129880 non-null   int64  
 8   Ease of Online booking 129880 non-null   int64  
 9   Gate location    129880 non-null   int64  
 10  Food and drink  129880 non-null   int64  
 11  Online boarding 129880 non-null   int64  
 12  Seat comfort     129880 non-null   int64  
 13  Inflight entertainment 129880 non-null   int64  
 14  On-board service 129880 non-null   int64  
 15  Leg room service 129880 non-null   int64  
 16  Baggage handling 129880 non-null   int64  
 17  Checkin service  129880 non-null   int64  
 18  Inflight service 129880 non-null   int64  
 19  Cleanliness      129880 non-null   int64  
 20  Departure Delay in Minutes 129880 non-null   int64  
 21  Arrival Delay in Minutes 129487 non-null   float64 
 22  satisfaction      129880 non-null   object 

dtypes: float64(1), int64(17), object(5)
memory usage: 22.8+ MB

```

Hình 600: Chọn ra các cột thuộc tính để tiến hành quá trình Mining

#### 4.8.2. Xử lý giá trị khuyết:

Qua thông tin của Hình 546, ta có thể nhận thấy chỉ duy nhất cột Arrival Delay in Minutes có giá trị khuyết và số lượng cụ thể là 393 giá trị, chiếm tỉ lệ rất

nhỏ chỉ 0,3% số lượng dữ liệu. Vì vậy, nhóm quyết định xóa toàn bộ các hàng có dữ liệu bị khuyết tại cột này.

Hình 601: Bộ dữ liệu sau khi xóa bỏ các dòng dữ liệu trống

#### 4.8.3. Chuẩn hóa dữ liệu:

Nhóm sử dụng thư viện Scikit-learn với hàm hỗ trợ LabelEncoder() để mã hóa toàn bộ các thuộc tính phân loại về dạng số.

```

from sklearn.preprocessing import LabelEncoder
encoder = LabelEncoder()

[14] ✓ 0.0s

original_values = {}

for column in (variable) column: str
    if X[column].dtype == 'object':
        original_values[column] = X[column].copy()

# Fit và transform dữ liệu của cột
X[column] = encoder.fit_transform(X[column])

```

Hình 602: Mã hóa các cột thuộc tính phân loại

Nhóm sử dụng thư viện Scikit-learn với hàm hỗ trợ StandardScaler() để chuẩn hóa các giá trị của cột thuộc tính kiểu số về chung một miền giá trị.

```

from sklearn.preprocessing import StandardScaler
std_scaler = StandardScaler()

[16]: ✓ 0.0s
features_to_scale = ['Age','Flight Distance', 'Departure Delay in Minutes', 'Arrival Delay in Minutes']
X[features_to_scale] = std_scaler.fit_transform(X[features_to_scale])
X.head(5)

[17]: ✓ 0.0s

```

	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness
0	0	0	0.831567	0	1	-1.032734	5	4	3	4	...	5	5	5	5	2	5	5
1	0	0	-0.226807	0	0	1.676886	1	1	3	1	...	4	4	4	4	3	4	5
2	1	1	-1.285180	0	1	-1.000655	2	0	2	4	...	2	4	1	3	2	2	2
3	1	0	0.302380	0	0	2.192145	0	0	0	2	...	1	1	1	1	3	1	4
4	0	0	0.633122	0	1	-0.008231	2	3	4	3	...	2	2	2	2	4	2	4

5 rows × 23 columns

Hình 603: Dữ liệu các cột phân loại sau khi được mã hóa

Bộ dữ liệu thu được sau khi tiền xử lý và chuẩn hóa để bắt đầu đưa vào mô hình huấn luyện.

	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	...	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness
0	0	0	0.831567	0	1	-1.032734	5	4	3	4	...	5	5	5	5	2	5	5
1	0	0	-0.226807	0	0	1.676886	1	1	3	1	...	4	4	4	4	3	4	5
2	1	1	-1.285180	0	1	-1.000655	2	0	2	4	...	2	4	1	3	2	2	2
3	1	0	0.302380	0	0	2.192145	0	0	0	2	...	1	1	1	1	3	1	4
4	0	0	0.633122	0	1	-0.008231	2	3	4	3	...	2	2	2	2	4	2	4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
129875	0	1	-1.086735	0	1	-1.000655	2	1	2	3	...	2	3	1	4	2	3	5
129876	1	0	0.633122	0	0	1.159622	4	4	4	4	...	5	5	5	5	5	5	5
129877	1	1	-0.623697	0	0	0.806760	1	1	1	3	...	4	3	2	4	5	5	5
129878	0	1	-1.152884	0	1	-0.190676	1	1	1	5	...	1	4	5	1	5	4	5
129879	1	0	-0.822142	0	0	0.534094	1	3	3	3	...	1	1	1	4	4	4	3

129487 rows × 23 columns

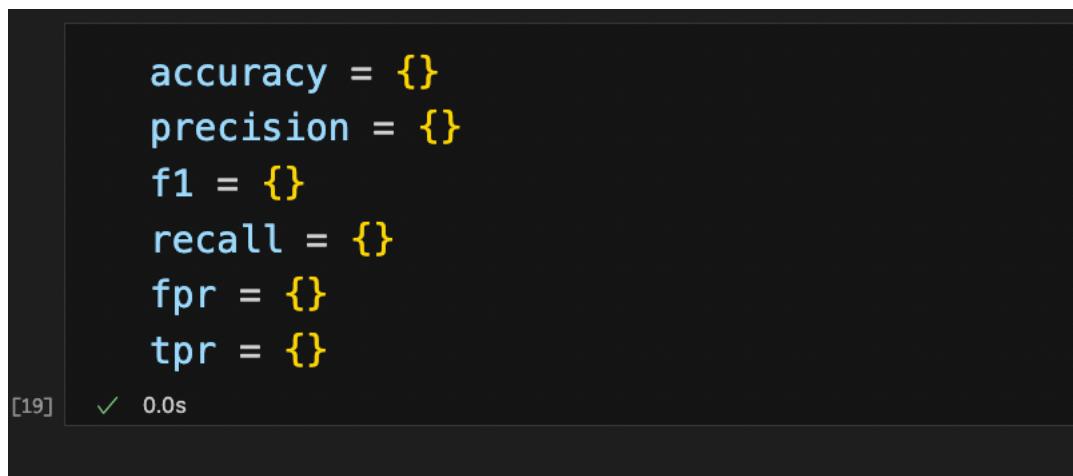
Hình 604: Chuẩn hóa các cột dữ liệu kiểu số

#### 4.9. Ứng dụng mô hình thuật toán khai thác dữ liệu:

Khai báo 6 từ điển (dictionary) trống:

- + Accuracy: Lưu trữ kết quả độ chính xác (accuracy) của mô hình trên từng lớp.
- + Precision: Lưu trữ kết quả độ chính xác dự báo (precision) của mô hình trên từng lớp.
- + Recall: Lưu trữ kết quả độ phủ sóng (recall) của mô hình trên từng lớp.
- + F1: Lưu trữ kết quả độ đo F1-Score của mô hình trên từng lớp.
- + FPR: Lưu trữ tỉ lệ dự báo sai (false positive rate) của mô hình trên từng lớp.
- + TPR: Lưu trữ tỉ lệ dự báo đúng (true positive rate) của mô hình trên từng lớp.

Mỗi từ điển được sử dụng để lưu trữ kết quả đánh giá cho một lớp (class) trong bài toán phân loại đa lớp (multi-classification). Các giá trị sẽ được cập nhật trong quá trình đánh giá mô hình (evaluation).



```
accuracy = {}
precision = {}
f1 = {}
recall = {}
fpr = {}
tpr = {}

[19]    ✓  0.0s
```

Hình 605: Tạo từ điển lưu trữ kết quả các độ đo đánh giá mô hình

#### 4.9.1. Chia dữ liệu trước khi xây dựng mô hình thuật toán:

Sau quá trình tiền xử lý dữ liệu, chúng ta thu được DataFrame cuối cùng X với các thuộc tính phân loại được mã hóa.

DataFrame X được chia thành các tập lần lượt Train – Validate – Test.

**Bước 1:** Chia DataFrame X thành hai phần: tập huấn luyện / đánh giá (training set) / tập kiểm tra (test\_set).



```
X, X_test = train_test_split(X, test_size = 0.2, random_state=42)
print(X.shape, X_test.shape)

[21]    ✓  0.0s
...   (103589, 23) (25898, 23)
```

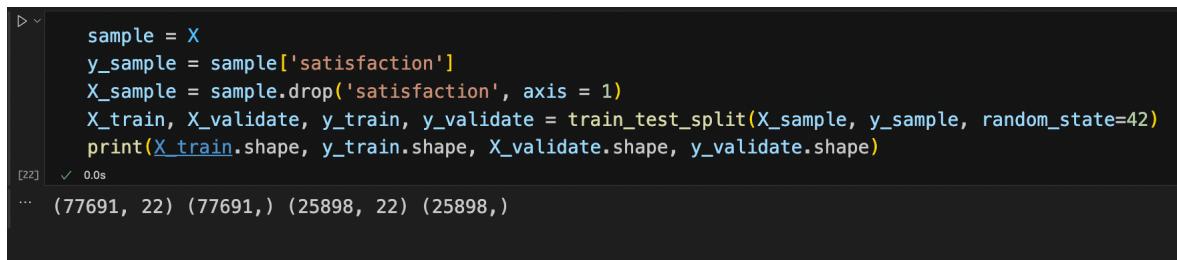
Hình 606: Phân chia bộ dữ liệu thành hai tập (test set và validate & training set)

Sử dụng hàm `train_test_split()` từ thư viện scikit-learn để thực hiện việc tách bộ dữ liệu. Thiết lập tham số `test_size = 0.2`, có nghĩa là 20% dữ liệu sẽ được dùng cho tập kiểm tra và 80% còn lại sẽ sử dụng cho việc huấn luyện và đánh giá.

Thiết lập tham số `random_state = 42` để đảm bảo việc chia có thể tái lập.

In ra kích thước của hai tập kết quả sử dụng thuộc tính `shape` của mỗi DataFrame, trả về một tuple chứa số hàng và số cột (theo thứ tự đó).

**Bước 2:** Gán DataFrame X vào một biến mới là `sample`. Chia DataFrame `sample` thành hai phần: tập huấn luyện (`X_train` và `y_train`) và tập đánh giá (`X_validate` và `y_validate`), sử dụng hàm `train_test_split()` từ thư viện scikit-learn.



```

sample = X
y_sample = sample['satisfaction']
X_sample = sample.drop('satisfaction', axis = 1)
X_train, X_validate, y_train, y_validate = train_test_split(X_sample, y_sample, random_state=42)
print(X_train.shape, y_train.shape, X_validate.shape, y_validate.shape)

[22]:    0.0s
... (77691, 22) (77691,) (25898, 22) (25898,)

```

Hình 607: Tiếp tục phân chia thành validate set và training set

Trước hết, gán DataFrame X vào một biến mới là sample.

Trích xuất thuộc tính mục tiêu satisfaction từ sample và gán cho y\_sample, còn lại các thuộc tính được gán cho X\_sample.

Sử dụng hàm train\_test\_split() để chia X\_sample và y\_sample thành tập huấn luyện và tập đánh giá. Thiết lập tham số random\_state = 42 để đảm bảo việc chia có thể tái lập.

Cuối cùng, in ra kích thước của hai tập kết quả sử dụng thuộc tính shape của mỗi DataFrame hoặc Series.

#### 4.9.2. Decision Tree:

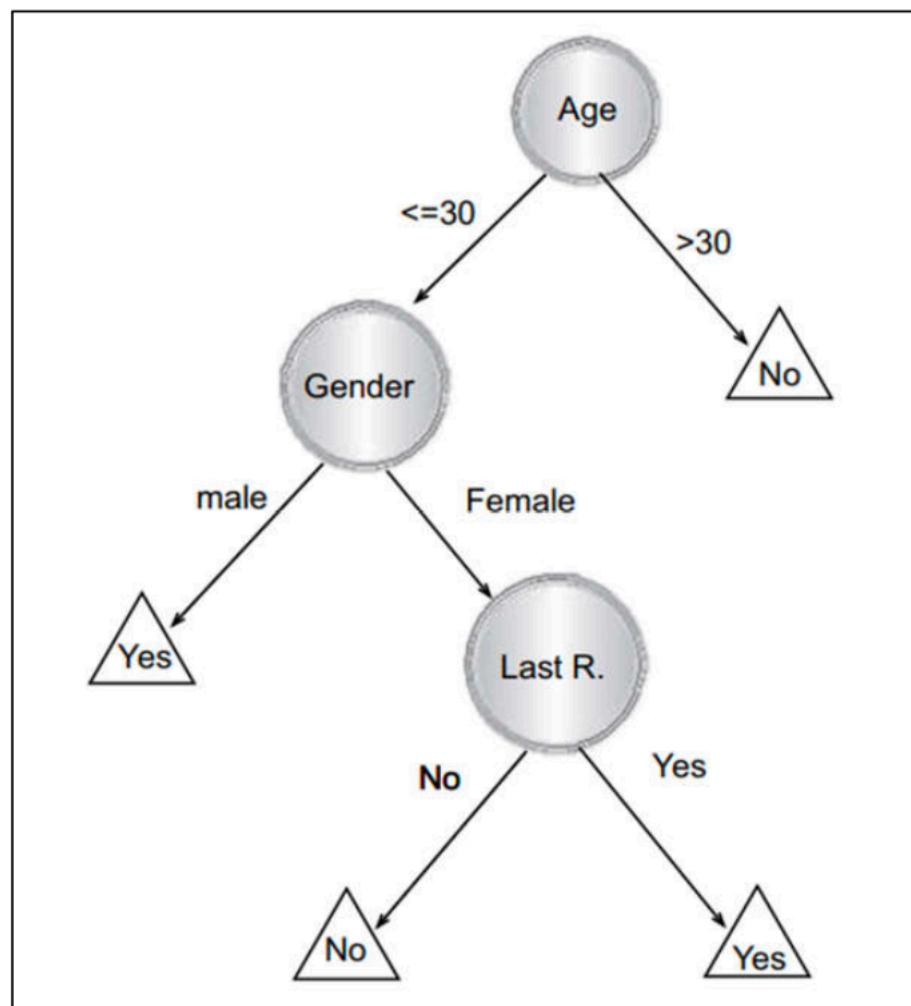
##### 4.9.2.1. Giới thiệu thuật toán:

Cây quyết định là một bộ phân loại được biểu diễn dưới dạng phân chia đệ quy của không gian các ví dụ. Cây quyết định bao gồm các nút tạo thành một cây gốc, có nghĩa là nó là một cây có gốc, với một nút được gọi là "gốc" không có cạnh vào. Tất cả các nút khác đều có đúng một cạnh vào. Một nút với các cạnh ra được gọi là nút nội hoặc nút kiểm tra. Tất cả các nút khác được gọi là lá (còn được gọi là nút cuối hoặc quyết định). Trong một cây quyết định, mỗi nút nội chia không gian ví dụ thành hai hoặc nhiều phân không gian theo một hàm rời rạc cụ thể của các giá trị thuộc tính đầu vào. Trong trường hợp đơn giản và thường gặp nhất, mỗi kiểm tra xem xét một thuộc tính duy nhất, sao cho không gian ví dụ được phân chia theo giá trị của thuộc tính. Trong trường hợp thuộc tính số, điều kiện áp dụng cho một khoảng giá trị.

Mỗi lá được gán cho một lớp đại diện cho giá trị mục tiêu phù hợp nhất. Theo cách khác, lá có thể chứa một vector xác suất chỉ ra xác suất của

thuộc tính mục tiêu có một giá trị cụ thể. Các ví dụ được phân loại bằng cách điều hướng chúng từ gốc cây xuống lá, theo kết quả của các kiểm tra trên đường đi.

Ví dụ, hình ảnh này mô tả một cây quyết định dùng để suy luận xem một khách hàng tiềm năng có phản hồi với thư email trực tiếp hay không. Các nút nội được biểu diễn dưới dạng các hình tròn, trong khi các lá được ký hiệu là các tam giác.



Hình 608: Cây quyết định khách hàng có phản hồi email hay không

#### 4.9.2.2. Xây dựng mô hình:

**Bước 1:** Tạo mô hình cây quyết định (dtc) sử dụng DecisionTreeClassifier từ sklearn.tree, với tham số random\_state được đặt thành 42. Tạo danh sách các tham số để thử trong tìm kiếm dạng lướt. Danh sách này bao gồm hai tham số:

criterion (gồm hai phương pháp gini index và entropy đo độ tách chât của nút) và max\_depth (độ sâu tối đa của cây). Tạo đối tượng tìm kiếm dạng lưới với mô hình cây quyết định ‘dtc’, các tham số để thử ‘parameters’, ‘verbose = 5’ để in kết quả tìm kiếm dạng lưới và n\_jobs = -1 để sử dụng tất cả các lõi CPU có sẵn. Huấn luyện mô hình trên tập huấn luyện ‘X\_train’ và ‘y\_train’, đồng thời thực hiện tìm kiếm dạng lưới để tìm các tham số tốt nhất cho mô hình. Sau đó, in ra các tham số tốt nhất, điểm số trên tập huấn luyện và điểm số trên tập hợp lẻ của mô hình tốt nhất được tìm thấy bằng tìm kiếm lưới.

```

[26]   from sklearn.tree import DecisionTreeClassifier
        from sklearn.metrics import confusion_matrix
[26]   ✓ 0.1s      + Code + Markdown

        dtc = DecisionTreeClassifier(random_state=42)
        parameters = [{"criterion": ["gini", "entropy"], "max_depth": [5, 10, 15, 30]}]
        grid = GridSearchCV(dtc, parameters, verbose=5, n_jobs=-1)
        grid.fit(X_train, y_train)
[27]   ✓ 5.8s

... Fitting 5 folds for each of 8 candidates, totalling 40 fits

</>   ▶      GridSearchCV
      ▶ estimator: DecisionTreeClassifier
          ▶ DecisionTreeClassifier

```

```

[28]   print("Best parameters scores:", grid.best_params_)
        print("Train score:", grid.score(X_train, y_train))
        print("Validation score:", grid.score(X_validate, y_validate))
[28]   ✓ 0.0s

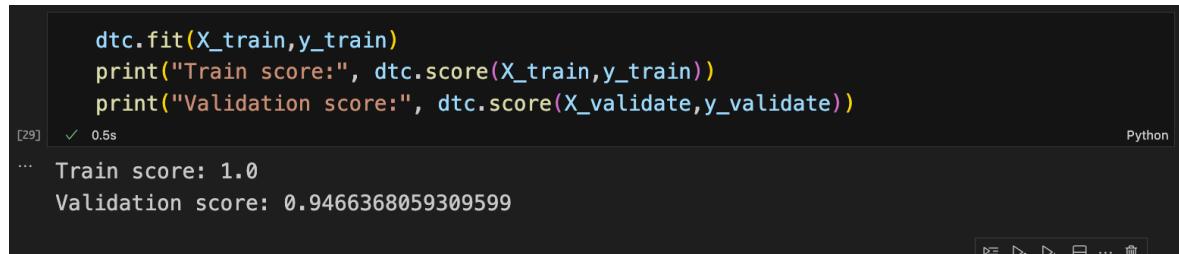
... Best parameters scores: {'criterion': 'entropy', 'max_depth': 15}
      Train score: 0.972043093794648
      Validation score: 0.9534326975055989

```

Hình 609: Tìm kiếm bộ tham số thích hợp nhất với mô hình Cây quyết định

Đầu ra có nghĩa là mô hình Cây quyết định đã được huấn luyện và cho bộ tham số tốt nhất đối với bộ dữ liệu huấn luyện là “criterion: gini” và “max\_depth: 10”. Điểm chính xác trên dữ liệu huấn luyện (train data) xấp xỉ 0.97, cho thấy mô hình đã dự đoán chính xác nhãn lớp cho 97% mẫu huấn luyện. Điểm chính xác trên dữ liệu kiểm định (validation data) xấp xỉ 0.95, cho biết mô hình đã dự đoán chính xác nhãn lớp cho 95% tập kiểm định.

**Bước 2:** Khớp mô hình (Fit model) cây quyết định (dtc) vào tập huấn luyện ( $X_{train}$ ,  $y_{train}$ ) nhằm tìm ra các tham số của mô hình sao cho nó có khả năng dự đoán tốt trên dữ liệu mới và in điểm đánh giá hiệu suất mô hình dựa trên dữ liệu huấn luyện và dữ liệu kiểm định cho các siêu tham số mặc định.



```

[29] 0.5s
dtc.fit(X_train,y_train)
print("Train score:", dtc.score(X_train,y_train))
print("Validation score:", dtc.score(X_validate,y_validate))

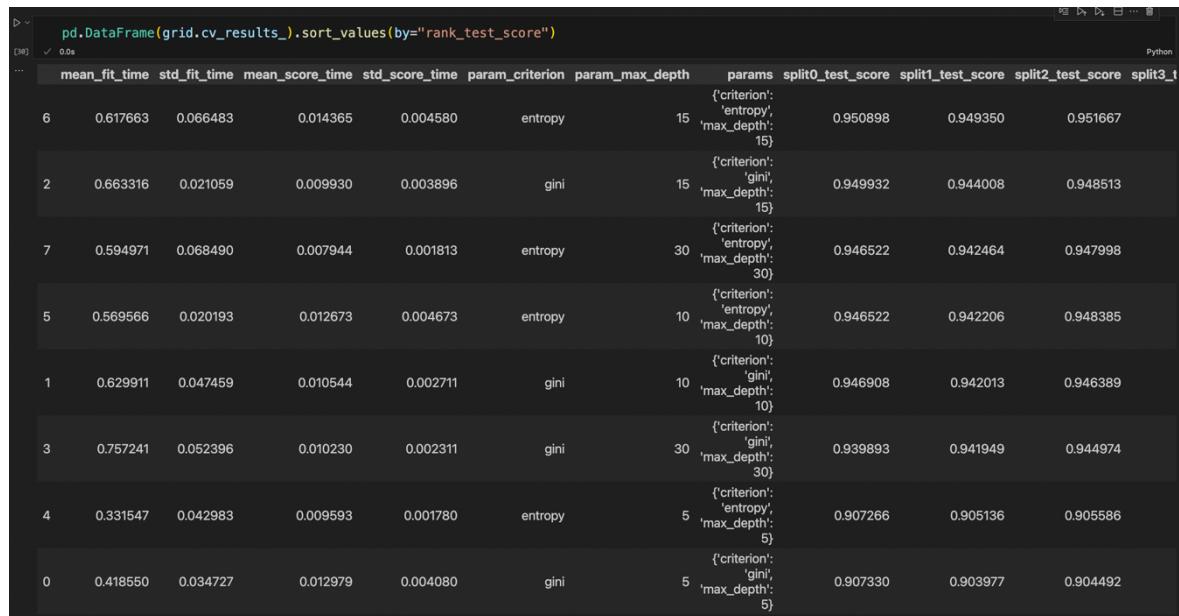
```

... Train score: 1.0  
Validation score: 0.9466368059309599

Hình 610: Điểm đánh giá hiệu suất mô hình

Điểm huấn luyện đạt điểm tuyệt đối 1.0 cho thấy mô hình dự đoán đúng toàn bộ nhãn của dữ liệu huấn luyện. Điểm kiểm chứng xấp xỉ 0.94 cho thấy mô hình đã dự đoán đúng 94% mẫu dữ liệu kiểm chứng.

**Bước 3:** Tạo DataFrame từ kết quả tìm kiếm dạng lưới (grid search) và sắp xếp kết quả theo thứ tự tăng dần của điểm xác thực.



	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_criterion	param_max_depth	params	split0_test_score	split1_test_score	split2_test_score	split3_t
6	0.617663	0.066483	0.014365	0.004580	entropy	15	{'criterion': 'entropy', 'max_depth': 15}	0.950898	0.949350	0.951667	
2	0.663316	0.021059	0.009930	0.003896	gini	15	{'criterion': 'gini', 'max_depth': 15}	0.949932	0.944008	0.948513	
7	0.594971	0.068490	0.007944	0.001813	entropy	30	{'criterion': 'entropy', 'max_depth': 30}	0.946522	0.942464	0.947998	
5	0.569566	0.020193	0.012673	0.004673	entropy	10	{'criterion': 'entropy', 'max_depth': 10}	0.946522	0.942206	0.948385	
1	0.629911	0.047459	0.010544	0.002711	gini	10	{'criterion': 'gini', 'max_depth': 10}	0.946908	0.942013	0.946389	
3	0.757241	0.052396	0.010230	0.002311	gini	30	{'criterion': 'gini', 'max_depth': 30}	0.939893	0.941949	0.944974	
4	0.331547	0.042983	0.009593	0.001780	entropy	5	{'criterion': 'entropy', 'max_depth': 5}	0.907266	0.905136	0.905586	
0	0.418550	0.034727	0.012979	0.004080	gini	5	{'criterion': 'gini', 'max_depth': 5}	0.907330	0.903977	0.904492	

Hình 611: DataFrame kết quả tìm kiếm dạng lưới theo thứ tự tăng dần của điểm xác thực

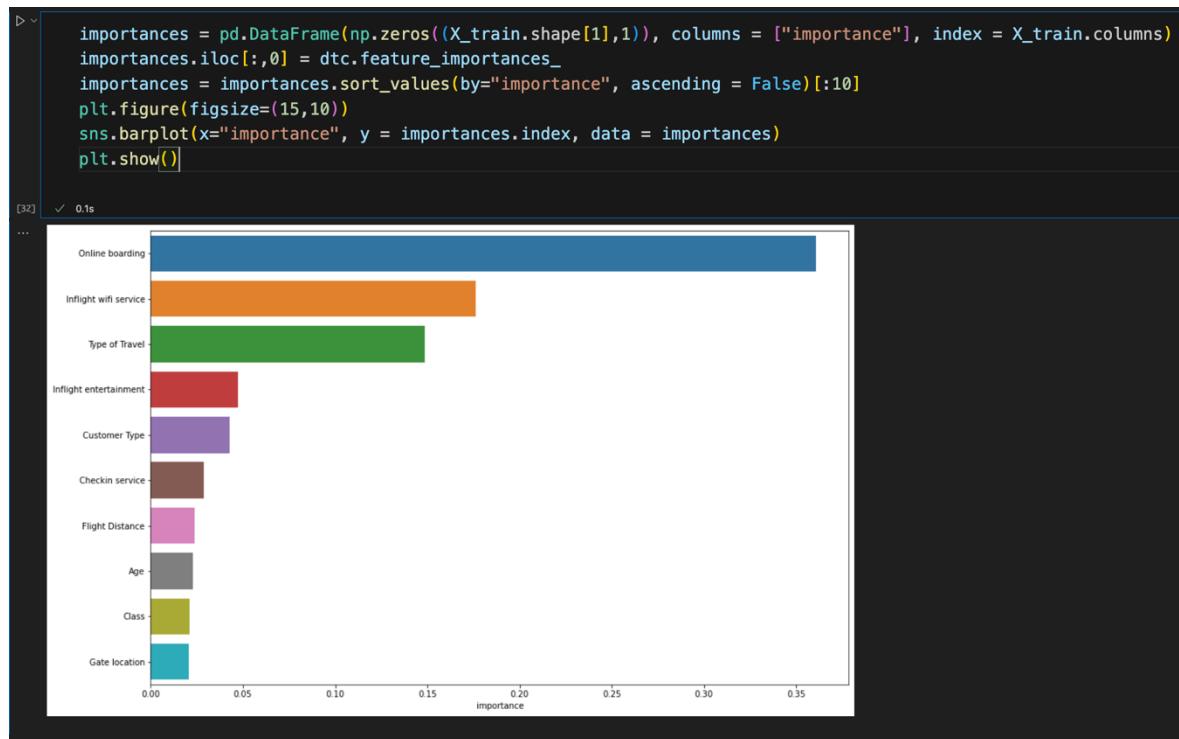
**Bước 4:** Dự đoán các nhãn của bộ xác thực ( $X_{validate}$ ) bằng cách sử dụng mô hình cây quyết định (dtc) đã được đào tạo. Tính toán ma trận nhầm lẫn bằng cách so sánh các nhãn thực ( $y_{validate}$ ) với các nhãn dự đoán ( $y_{pred}$ ). Tạo ma trận nhầm lẫn để trực quan hóa hiệu suất của mô hình. Ma trận nhầm lẫn được lưu

trữ trong khung dữ liệu Pandas và được hiển thị bằng cách sử dụng biểu đồ nhiệt từ thư viện seaborn.



Hình 612: Ma trận nhầm lẫn khi dự đoán tập validate bằng mô hình Cây quyết định

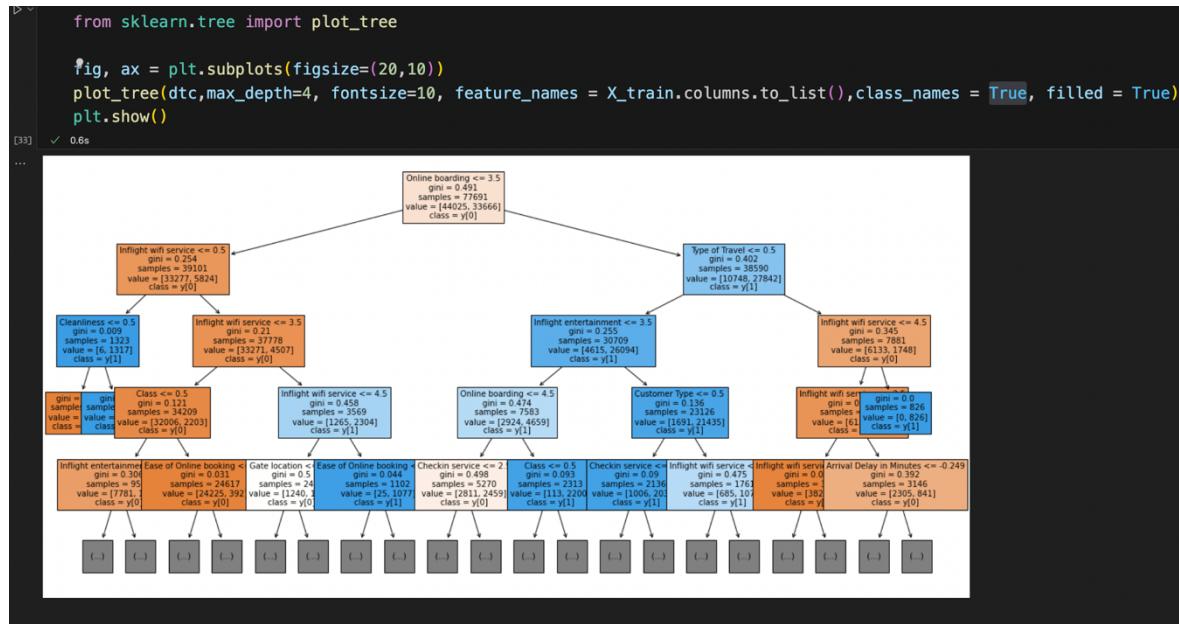
**Bước 5:** Tìm ra 10 thuộc tính quan trọng hàng đầu bằng cách tạo DataFrame có số hàng bằng số cột trong X\_train và một cột có tên là “quan trọng”, sử dụng tên thuộc tính trong X\_train làm chỉ mục.



Hình 613: 10 thuộc tính quan trọng ảnh hưởng đến biến satisfaction (mô hình Cây quyết định)

**Bước 6:** Tạo trực quan hóa mô hình Cây quyết định ‘dtc’ bằng cách sử dụng

hàm plot\_tree() từ sklearn.tree, với max\_depth = 4.



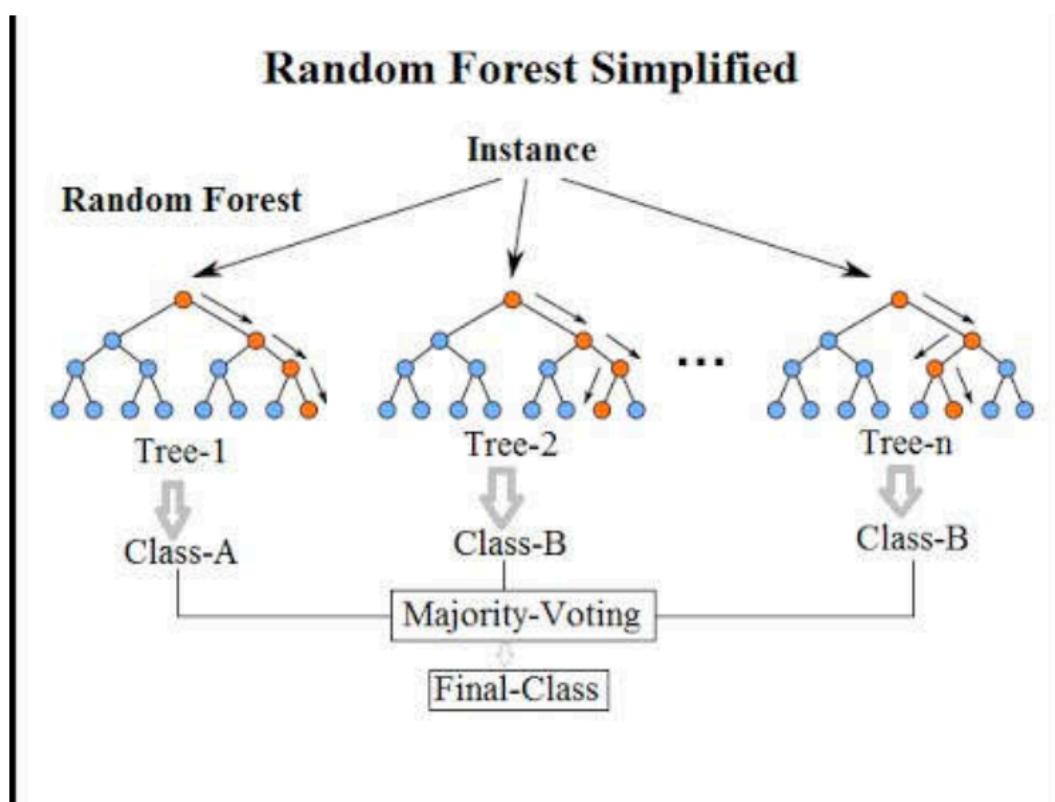
Hình 614: Trực quan hóa cây quyết định

### 4.9.3. Random Forest:

#### 4.9.3.1. Giới thiệu thuật toán:

Nhằm khắc phục tình trạng overfitting của mô hình cây quyết định, mô hình Random forest được giới thiệu lần đầu bởi Breiman (2001), tác giả đã đề xuất tạo một tập hợp các cây quyết định được xây dựng dựa trên một tập con ngẫu nhiên của các mẫu dữ liệu với các thuộc tính quyết định cũng được lựa chọn ngẫu nhiên. Dữ liệu của mỗi cây được lấy ngẫu nhiên và có thể trùng lặp(Bootstrapping) để đem lại các kết quả dự báo từ bộ dữ liệu huấn luyện độc lập nhau. Từ đó, kết quả dự báo là giá trị trung bình của tập hợp cây quyết định.

Xem minh họa cấu trúc rừng ngẫu nhiên:



Hình 615: Minh họa cấu trúc rừng ngẫu nhiên

Trong phạm vi đồ án môn học IS217, nhóm sử dụng hàm Random Forest được cài đặt sẵn trong thư viện Scikit-learn với các cây được tạo nên từ mô hình cây quyết định Classification and Regression Trees – CART.

#### 4.9.3.2. Xây dựng mô hình:

**Bước 1:** Thực hiện bộ phân loại Random Forest: Đoạn mã khởi tạo một bộ phân loại Random Forest (rfc) bằng cách sử dụng lớp Random Forest Classifier. Đối tượng bộ phân loại Random Forest được khởi tạo với n\_jobs và random\_state.

Các siêu tham số cần được điều chỉnh được chỉ định trong một từ điển được gọi là parameters. Một đối tượng GridSearchCV được tạo ra và phù hợp với dữ liệu huấn luyện. Các siêu tham số tốt nhất được xuất ra cùng với các điểm số độ chính xác của mô hình dữ liệu huấn luyện và dữ liệu xác thực.

```

[36] ✓ 0.0s

from sklearn.ensemble import RandomForestClassifier

[37] ✓ 3m 56.9s
rfc = RandomForestClassifier(n_jobs = -1, random_state=42)
parameters = [{"n_estimators": [50,100,200,500], "max_depth": [5,10,20,30]}]
grid = GridSearchCV(rfc, parameters, verbose=5, n_jobs=-1)
grid.fit(X_train, y_train)

... Fitting 5 folds for each of 16 candidates, totalling 80 fits

</>
    ▶      GridSearchCV
    ▶ estimator: RandomForestClassifier
        ▶ RandomForestClassifier

```

```

[38] ✓ 2.6s
print("Best parameters scores: ", grid.best_params_)
rfc.fit(X_train, y_train)
print("Train score:", grid.score(X_train,y_train))
print("Validation score:", grid.score(X_validate,y_validate))

... Best parameters scores: {'max_depth': 30, 'n_estimators': 500}
Train score: 1.0
Validation score: 0.9625067572785543

```

Hình 616: Tìm kiếm bộ tham số thích hợp nhất với mô hình rừng ngẫu nhiên

Kết quả đầu ra cho thấy mô hình bộ phân loại Random Forest đã được huấn luyện và đánh giá bằng cách sử dụng GridSearchCV với 5-fold cross-validation trên 2 tố hợp siêu tham số (`n_estimators` và `max_depth`).

Điểm số độ chính xác trên dữ liệu đạt tuyệt đối 1.0, cho thấy rằng mô hình đã dự đoán đúng nhãn lớp cho toàn bộ các ví dụ trong tập huấn luyện. Điểm số độ chính xác trên dữ liệu xác thực là 0.96, cho biết mô hình đã dự đoán đúng nhãn lớp 96% các ví dụ trong tập xác thực.

**Bước 2:** Sử dụng mô hình bộ phân loại Random Forest (rfc) để phù hợp với tập dữ liệu huấn luyện (`X_train` và `y_train`) và in ra các điểm số huấn luyện và xác thực với các siêu tham số mặc định.

```

rfc.fit(X_train, y_train)
print("Train score:", rfc.score(X_train,y_train))
print("Validation score:", rfc.score(X_validate,y_validate))

```

[38] ✓ 2.2s

... Train score: 1.0  
Validation score: 0.9621978531160708

+ Code + Markdown

Hình 617: Điểm đánh giá hiệu suất mô hình

Điểm số huấn luyện là 1.0, cho thấy mô hình dự đoán chính xác nhẫn của các mẫu huấn luyện. Điểm số xác thực là 0.96 cho biết mô hình dự đoán đúng 96.21% các mẫu dữ liệu xác thực.

**Bước 3:** Tạo một DataFrame Pandas chứa kết quả của cross-validation lưới tìm kiếm và sắp xếp theo rank\_test\_score.

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_max_depth	param_n_estimators	params	split0_test_score	split1_test_score	split2_test_score
15	44.337844	2.826300	0.676969	0.289450	30	500	{'max_depth': 30, 'n_estimators': 500}	0.962675	0.959454	0.960806
14	24.438216	0.678690	0.488621	0.067582	30	200	{'max_depth': 30, 'n_estimators': 200}	0.962353	0.958811	0.960870
11	61.070286	0.342051	1.586315	0.051887	20	500	{'max_depth': 20, 'n_estimators': 500}	0.962288	0.958875	0.960291
13	12.873259	0.170498	0.157346	0.093390	30	100	{'max_depth': 30, 'n_estimators': 100}	0.961967	0.957073	0.959969
12	6.623099	0.149491	0.043679	0.004694	30	50	{'max_depth': 30, 'n_estimators': 50}	0.961902	0.957266	0.960098
10	24.437888	0.344015	0.449066	0.156543	20	200	{'max_depth': 20, 'n_estimators': 200}	0.961516	0.958875	0.959132
9	12.696067	0.090186	0.184816	0.066851	20	100	{'max_depth': 20, 'n_estimators': 100}	0.961581	0.958231	0.959519
8	6.548365	0.107849	0.040570	0.002597	20	50	{'max_depth': 20, 'n_estimators': 50}	0.961709	0.956365	0.958103
7	46.673548	0.348557	0.838769	0.079512	10	500	{'max_depth': 10, 'n_estimators': 500}	0.947423	0.943815	0.945617
6	18.853372	0.095948	0.163218	0.083576	10	200	{'max_depth': 10, 'n_estimators': 200}	0.947101	0.942785	0.943944

Hình 618: DataFrame kết quả tìm kiếm dạng lưới theo thứ tự tăng dần của điểm xác thực

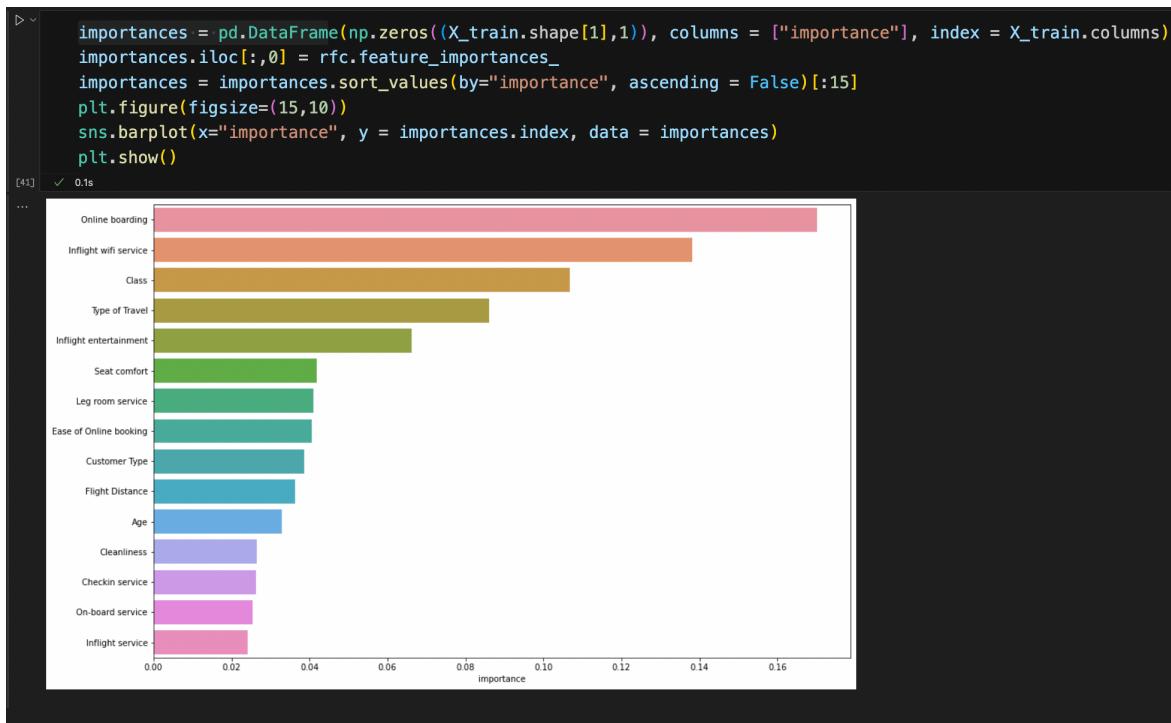
Kết quả DataFrame chứa một hàng cho mỗi tổ hợp siêu tham số và các cột cho các số liệu đánh giá khác nhau, bao gồm điểm số trung bình trên tập kiểm tra, độ lệch chuẩn của điểm số kiểm tra và các giá trị của siêu tham số.

**Bước 4:** Tạo ma trận nhầm lẫn để trực quan hóa hiệu suất của mô hình. Ma trận nhầm lẫn được lưu trữ trong một DataFrame Pandas và được hiển thị bằng cách sử dụng biểu đồ nhiệt từ thư viện seaborn.



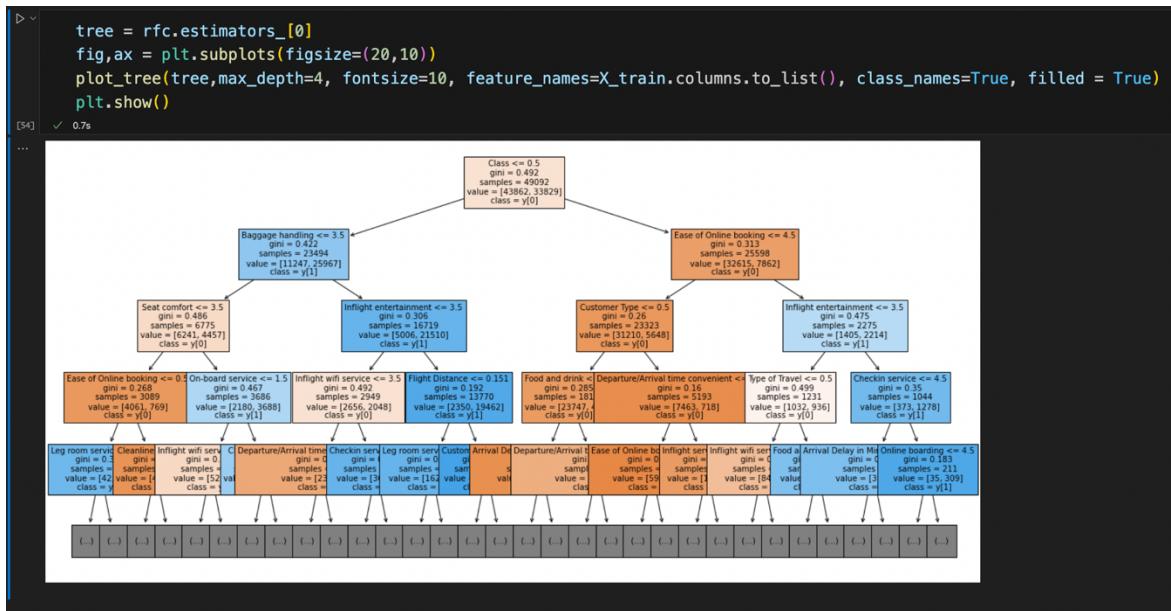
Hình 619: Ma trận nhầm lẫn khi dự đoán tập validate bằng mô hình Rừng ngẫu nhiên

**Bước 5:** Tạo biểu đồ cột để đại diện cho độ quan trọng của thuộc tính, được xác định bởi bộ phân loại Random Forest cho 15 thuộc tính hàng đầu. Điểm số độ quan trọng của một thuộc tính càng cao thì thuộc tính đó càng ảnh hưởng đến việc dự đoán với mô hình Random Forest.



Hình 620: 15 thuộc tính quan trọng ảnh hưởng đến biến satisfaction (mô hình Rừng ngẫu nhiên)

**Bước 6:** Vẽ cây quyết định. Trong đó, ta sử dụng thuộc tính estimators\_ của mô hình Random Forest để lấy ra cây quyết định đầu tiên. Sau đó, ta sử dụng hàm plot\_tree để vẽ cây quyết định với các tham số tương tự như khi sử dụng với Decission Tree. Cuối cùng, ta sử dụng hàm show() của matplotlib để hiển thị cây quyết định.



Hình 621: Trực quan hóa cây quyết định đầu tiên trong rừng ngẫu nhiên

#### 4.9.4. Đánh giá thuật toán:

##### 4.9.4.1. Các độ đo dùng để đánh giá thuật toán:

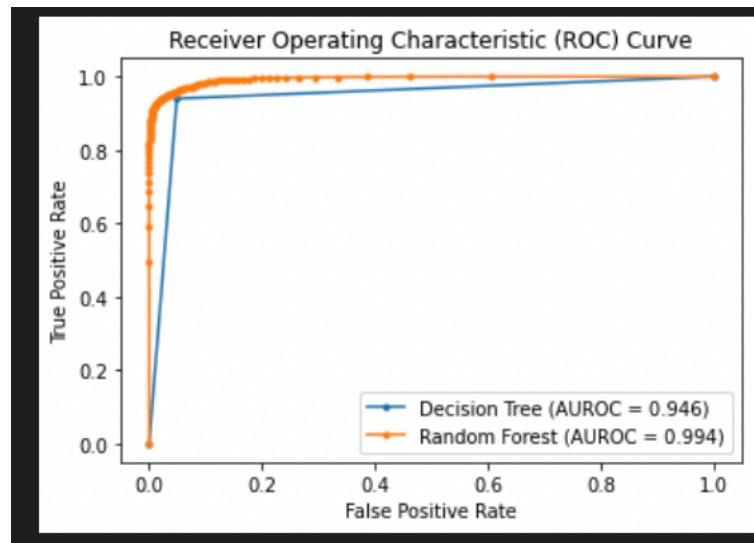
Đánh giá mô hình phân loại trong máy học, ta sử dụng các chỉ số: TP, FP, TN, FN. Trong đó, TP (True Positive) là tổng số trường hợp dự báo khớp với nhãn Positive, TN (True Negative) là tổng số trường hợp dự báo khớp với nhãn Negative, FP (False Positive): Tổng số trường hợp dự báo các quan sát thuộc nhãn Negative thành Positive và FN (False Negative) là tổng số trường hợp dự báo các quan sát thuộc nhãn Positive thành Negative.

Ở bài toán của nhóm, nhãn Negative tương ứng với nhãn satisfied (thỏa mãn) và nhãn neutral or dissatisfied tương ứng với nhãn Positive. Những chỉ số trên là cơ sở để tính toán các độ đo như accuracy, f1-score và vẽ được đường cong ROC.

##### 4.9.4.2. Đánh giá, so sánh các mô hình thuật toán:

	Random Forest	Decision Tree
Accuracy (Tập train)	0,9622	0,9466
Accuracy (Tập Test)	0.9608	0,9462
F1 – Score (Tập Train)	0,9613	0,9456
F1 – Score (Tập Test)	0.9549	0.9391

Hiển thị đồ thị ROC curve để đánh giá hiệu suất của mô hình phân loại nhị phân trên bộ dữ liệu test để đưa ra quyết định thông minh về việc chọn ngưỡng (threshold).



Hình 622: Đường cong ROC so sánh hiệu suất giữa hai mô hình

**Kết luận:** Dựa trên số liệu và đồ thị, ta có thể kết luận rằng thuật toán Random Forest phù hợp hơn thuật toán Decision Tree khi áp dụng đối với bộ dữ liệu này vì các chỉ số độ đo đánh giá (Accuracy, F1-Score và giá trị AUEOC) của thuật toán Random Forest đều cao hơn thuật toán Decision Tree.

## TÀI LIỆU THAM KHẢO

- [1] [SQL Server 2012 Tutorials – Analysis Services Data Mining – Microsoft](#)
- [2] [Quá trình trích xuất dữ liệu, biến đổi và nạp dữ liệu vào kho dữ liệu \(quá trình SSIS\)](#)
- [3] [Xử lý phân tích dữ liệu trực tuyến từ kho dữ liệu \(quá trình SSAS\)](#)
- [4] [SSRS Tutorials - SQL Server Reporting Service \(quá trình SSRS\)](#)
- [5] [TUTORIAL - Introduction to MDX](#)
- [6] [Tutorial: Get started creating in the Power BI service](#)
- [7] [Các đô đo sử dụng đánh giá mô hình phân loại trong Machine Learning](#)

## BẢNG PHÂN CÔNG CÔNG VIỆC

Công việc	21521750	21521819
Tìm kiếm chủ đề lựa chọn dataset trên Kaggle	x	x
Tiền xử lý dữ liệu	x	Hỗ trợ
Thiết kế và xây dựng mô hình kho dữ liệu, liệt kê các câu truy vấn mở đầu	x	x
Tiến hành cài đặt và minh họa trực tiếp quá trình đổ dữ liệu vào kho dữ liệu đã thiết kế (SSIS)	Hỗ trợ	x
Tiến hành cài đặt và thiết lập kết nối giữa kho dữ liệu với công cụ BI hỗ trợ việc truy vấn và phân tích dữ liệu (SSAS)	Hỗ trợ	x
Lên ý tưởng và hiện thực một số câu truy vấn liên quan đến việc mua bán, kinh doanh của siêu thị (sử dụng ngôn ngữ MDX và browser của Visual Studio 2019)	x	Hỗ trợ
Sử dụng Pivot Table trên nền tảng Microsoft Excel để truy vấn dữ liệu, tổng hợp, thống kê một số thuộc tính quan trọng trong dataset	Hỗ trợ	x
Thực hiện lập báo biểu bằng công cụ SSRS và Power BI	x	x
Quá trình Data Mining bộ dữ liệu đánh giá chất lượng dịch vụ hàng không trên công cụ SSAS và Jupyter Notebook	x	x
<b>Mức độ hoàn thành các công việc được phân công (%)</b>	<b>100%</b>	<b>100%</b>
<b>Mức độ đóng góp cho kết quả của đồ án (tổng cộng = 100%)</b>	<b>50%</b>	<b>50%</b>

**Thời gian thực hiện: 11 tuần.**

**-Hết-**