

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



MÔN HỌC: KHO DỮ LIỆU VÀ OLAP
BÁO CÁO ĐỒ ÁN
ĐỀ TÀI: HỆ THỐNG QUẢN LÝ VÀ PHÂN TÍCH DỮ
LIỆU KHÁCH HÀNG QUA CÁC CHIẾN DỊCH
MARKETING

GVHD : Ths.Nguyễn Thị Kim Phụng

LỚP: IS217.N22.HTCL

SINH VIÊN THỰC HIỆN:

Nguyễn Thị Cẩm Vân 20522145

Hồ Bảo An 20520876

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành và sự tri ân sâu sắc đối với các Thầy Cô của **Trường Đại học Công nghệ Thông tin – Đại học Quốc Gia TP.HCM**, đặc biệt là quý **Thầy Cô khoa Hệ Thống Thông Tin** của Trường đã giúp cho chúng em trang bị các kiến thức cơ bản, các kỹ năng thực tế và tạo điều kiện để chúng em có thể hoàn thành đồ án môn học của mình.

Đặc biệt nhóm xin gửi lời cảm ơn chân thành tới **Cô Nguyễn Thị Kim Phụng – Giảng viên lý thuyết môn Kho dữ liệu và OLAP** đã tận tình giúp đỡ, trực tiếp chỉ bảo, hướng dẫn nhóm trong suốt quá trình làm đồ án môn học. Nhờ đó, nhóm em đã tiếp thu được nhiều kiến thức bổ ích trong việc vận dụng cũng như kỹ năng làm đồ án. Một lần nữa, nhóm xin chân thành cảm ơn Cô.

Xuất phát từ mục đích học tập tìm hiểu về kho dữ liệu, phương pháp xây dựng kho dữ liệu và phân tích dữ liệu trên kho dữ liệu **MarketingData**. Dựa trên những kiến thức được Cô cung cấp trên trường kết hợp với việc tự tìm hiểu những công cụ và kiến thức mới, nhóm đã cố gắng thực hiện đồ án một cách tốt nhất. Trong thời gian một học kỳ thực hiện đề tài, nhóm đã vận dụng những kiến thức nền tảng đã tích lũy đồng thời kết hợp với việc học hỏi và nghiên cứu những kiến thức mới. Từ đó, em vận dụng tối đa những gì đã thu thập được để hoàn thành một báo cáo đồ án tốt nhất. Tuy nhiên, trong quá trình thực hiện, nhóm không tránh khỏi những thiếu sót. Chính vì vậy, em rất mong nhận được những sự góp ý từ phía Cô nhằm hoàn thiện những kiến thức mà em đã học tập và là hành trang để em thực hiện tiếp các đề tài khác trong tương lai.

NHẬN XÉT CỦA GIẢNG VIÊN

MỤC LỤC

LỜI CẢM ƠN.....	2
NHẬN XÉT CỦA GIẢNG VIÊN	3
MỤC LỤC.....	4
CHƯƠNG 1: MÔ TẢ DỮ LIỆU BAN ĐẦU	7
1.1 Tổng quan về dữ liệu.....	7
1.1.1 Lý do chọn đề tài.....	7
1.1.2 Mô tả về dữ liệu	7
1.1.3 Thuộc tính kho dữ liệu	8
1.1.4 Kho dữ liệu đã thực hiện tiền xử lý	10
1.2 Xây dựng kho dữ liệu.....	14
1.2.1 Lược đồ kho dữ liệu (Star Schema)	14
1.2.2 Chi tiết các bảng.....	14
CHƯƠNG 2: XÂY DỰNG KHO DỮ LIỆU – QUÁ TRÌNH SSIS	16
2.1 Tạo project SSIS và thiết lập kết nối.....	16
2.2 Xây dựng kho dữ liệu và thiết lập kết nối.....	18
2.2.1 Xây dựng kho dữ liệu	18
2.2.2 Thiết lập kết nối	20
2.3 Mô hình SSIS	21
2.3.1 Làm sạch và load dữ liệu	21
2.3.2 Định nghĩa và nạp dữ liệu vào các bảng Dimension	36
2.3.3 Định nghĩa và nạp dữ liệu vào bảng FACT	54
2.3.4 Định nghĩa và set các khóa ngoại cho bảng FACT	68
2.4 Thực thi Package	71
2.5 Dữ liệu sau khi hoàn thành	72
CHƯƠNG 3: PHÂN TÍCH DỮ LIỆU TRỰC TUYẾN – QUÁ TRÌNH SSAS	75
3.1 Tạo project SSAS và thiết lập kết nối	75
3.1.1 Tạo project trong Visual Studio.....	75
3.1.2 Xác định Data Source	77
3.1.3 Xác định Data Source Views	80
3.1.4 Xây dựng các Cube và các Measures.....	81
3.1.5 Phân cấp trong bảng Dimension.....	84

3.1.5.1 Bảng Dim_Customer	84
3.1.5.2 Bảng Dim_Time	85
3.1.5.3 Bảng Dim_Country	88
3.1.6 Process Cube	88
3.2 Quá trình phân tích dữ liệu bằng ngôn ngữ MDX	90
3.2.1 Thống kê thu nhập hàng năm của Khách hàng đã chấp nhận 4 chiến dịch Marketing	90
3.2.2 Thống kê Top 10 Khách hàng có thu nhập hàng năm lớn nhất	92
3.2.3 Thống kê số tiền đã mua hàng của top 10 Khách hàng có thu nhập hàng năm lớn hơn 30000\$.....	94
3.2.4 Thống kê những Khách hàng đã chi nhiều hơn 150\$ cho mặt hàng bánh kẹo nhưng số lượng trẻ em trong nhà nhỏ hơn 2.....	96
3.2.5 Thống kê Top5 Khách hàng có tổng số đơn mua hàng nhiều nhất trong năm 2014 ...	98
3.2.6 Thống kê Top5 Quốc gia có số lượng mua hàng ở Store ít nhất	100
3.2.7 Thống kê ngày có lượng truy cập vào Web ít nhất trong tháng 8 năm 2012	102
3.2.8 Thống kê những Khách hàng vừa mua hàng ở Web vừa mua hàng ở Store	104
3.2.9 Thống kê những Khách hàng có tổng số đơn hàng cao nhất tại các Quốc gia	106
3.2.10 Tính trung bình số tiền đã chi cho tất cả các mặt hàng của Top5 Khách hàng có thu nhập hàng năm cao nhất theo từng Quốc gia.....	108
3.2.11 Thống kê số lượng đơn hàng đã thanh toán qua catalog và thu nhập hàng năm của Top10 Khách hàng độc thân và mới tốt nghiệp (Tăng dần theo số lượng đơn hàng)	110
3.2.12 Truy vấn DrillDown các tháng của năm 2012 có lượng tiêu thụ các mặt hàng về rượu nhiều hơn 2000\$.....	112
CHƯƠNG 4: QUÁ TRÌNH REPORT – POWER BI.....	114
4.1 Tạo Project vs Power BI	114
4.2 Report Power BI	116
4.2.1 Thống kê thu nhập hàng năm của Khách hàng đã chấp nhận 4 chiến dịch Marketing	116
4.2.2 Thống kê Top 10 Khách hàng có thu nhập hàng năm lớn nhất	118
4.2.3 Thống kê số tiền đã mua hàng của top 10 Khách hàng có thu nhập hàng năm lớn hơn 30000\$.....	119
4.2.4 Thống kê những Khách hàng đã chi nhiều hơn 150\$ cho mặt hàng bánh kẹo nhưng số lượng trẻ em trong nhà nhỏ hơn 2.....	121
4.2.5 Thống kê Top5 Khách hàng có tổng số đơn mua hàng nhiều nhất trong năm 2014 .	123
4.2.6 Thống kê Top5 Quốc gia có số lượng mua hàng ở Store ít nhất	124
4.2.7 Thống kê ngày có lượng truy cập vào Web ít nhất trong tháng 8 năm 2012	126

4.2.8	Thống kê những Khách hàng vừa mua hàng ở Web vừa mua hàng ở Store	127
4.2.9	Thống kê những Khách hàng có tổng số đơn hàng cao nhất tại các Quốc gia	129
4.2.10	Tính trung bình số tiền đã chi cho tất cả các mặt hàng của Top5 Khách hàng có thu nhập hàng năm cao nhất theo từng Quốc gia.....	130
4.2.11	Thống kê số lượng đơn hàng đã thanh toán qua catalog và thu nhập hàng năm của Top10 Khách hàng độc thân và mới tốt nghiệp (Tăng dần theo số lượng đơn hàng)	131
4.2.12	Truy vấn DrillDown các tháng của năm 2012 có lượng tiêu thụ các mặt hàng về rượu nhiều hơn 2000\$.....	133
CHƯƠNG 5: QUÁ TRÌNH DATAMINING	135
5.1	Xử lý dữ liệu	135
5.2	Phân lớp dữ liệu	138
5.3	Đưa dữ liệu và mô hình Random Forest và Decision Tree.....	140
5.3.1	Mô hình Random Forest.	140
5.3.1.1	Test độ ảnh hưởng của các nhóm đã phân so với output.	140
5.3.1.2	Test các thuộc tính của nhóm có độ ảnh hưởng cao nhất do với output.	141
5.3.2	Mô hình Decision Tree.	142
5.3.2.1	Test độ ảnh hưởng của các nhóm đã phân so với output.	142
5.3.2.2	Test các thuộc tính của nhóm có độ ảnh hưởng cao nhất do với output.	143
5.4	So sánh và Nhận Xét.....	144
5.4.1	So sánh	144
5.4.2	Nhận xét	145
5.4.3	Vẽ biểu đồ đặc trưng.....	145
TÀI LIỆU THAM KHẢO	146

CHƯƠNG 1: MÔ TẢ DỮ LIỆU BAN ĐẦU

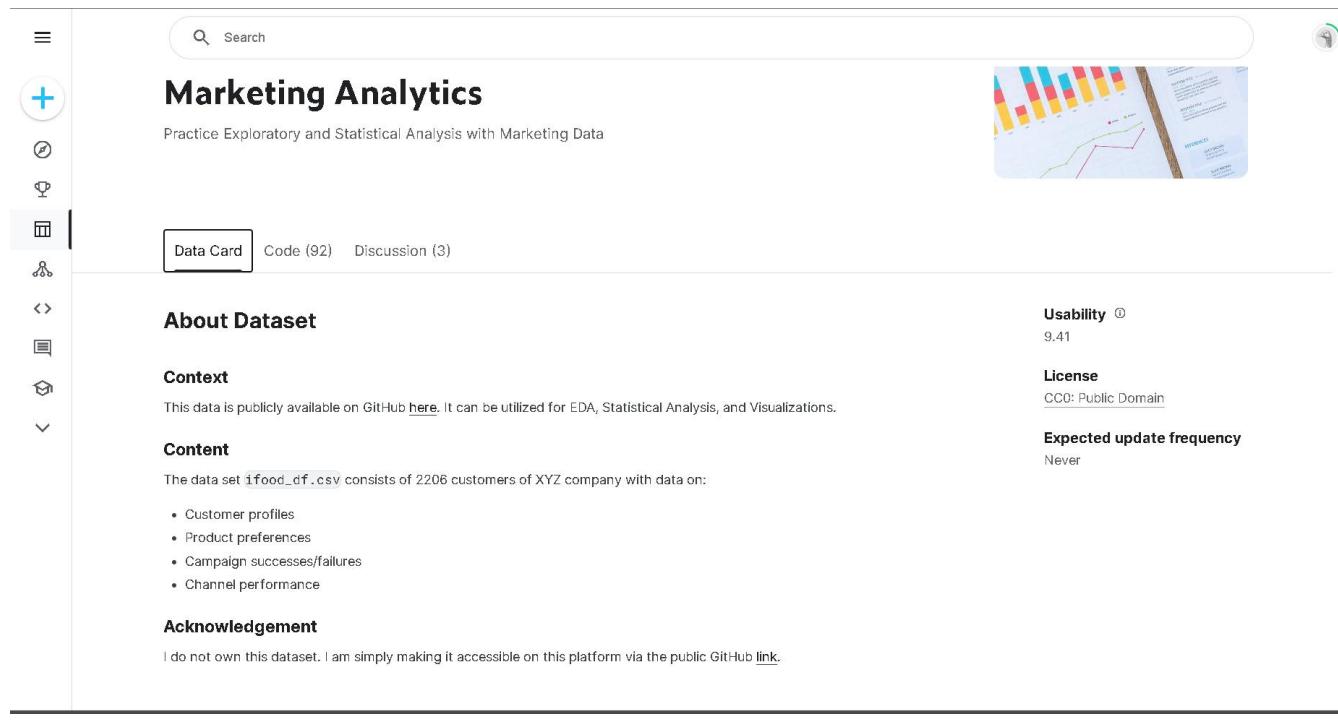
1.1 Tổng quan về dữ liệu

1.1.1 Lý do chọn đề tài

Bên cạnh tốc độ phát triển vượt bậc về công nghệ hiện nay, nhiều lĩnh vực kinh tế được thúc đẩy mạnh, thị trường dần xuất hiện đa dạng các doanh nghiệp, cá nhân buôn bán với quy mô lớn nhỏ khác nhau. Cùng với đó, nhiều thử thách được đặt ra với các doanh nghiệp hoạt động trong lĩnh vực buôn bán và thương mại, đặc biệt nhất có thể kể tới là việc dẫn đầu thị trường từ việc đẩy mạnh Marketing trên nhiều nền tảng khác nhau nhằm tối đã hóa được lợi nhuận.

Xác định vấn đề chính có thể giúp doanh nghiệp tăng lợi nhuận một cách đáng kể, nhóm nhận thấy việc tổ chức xây dựng một hệ thống giúp quản lý và phân tích dữ liệu là vô cùng cần thiết. Từ hệ thống này, doanh nghiệp có thể đưa ra được cái nhìn tổng thể, phân loại được nhóm khách hàng có khả năng phản hồi tích cực với các chiến dịch marketing. Từ đó, doanh nghiệp có thể giảm thiểu chi phí bằng cách thu hẹp phạm vi marketing, lựa chọn được những khách hàng có tiềm năng cao dựa trên tập dữ liệu và các thuật toán phân loại.

1.1.2 Mô tả về dữ liệu



The screenshot shows a dataset titled "Marketing Analytics" from a platform. The interface includes a sidebar with various icons for navigation, a search bar at the top, and a main content area. The main content area features a title "Marketing Analytics", a subtitle "Practice Exploratory and Statistical Analysis with Marketing Data", and a preview image showing various charts and graphs. Below the title, there are tabs for "Data Card", "Code (92)", and "Discussion (3)". The "Data Card" tab is selected. The card contains sections for "About Dataset", "Context", "Content", "Acknowledgement", and "Usability". The "About Dataset" section includes a brief description and a link to GitHub. The "Content" section lists customer profiles, product preferences, campaign successes/failures, and channel performance. The "Usability" section shows a rating of 9.41. Other sections visible include "License" (CC0: Public Domain) and "Expected update frequency" (Never). The footer of the card states: "I do not own this dataset. I am simply making it accessible on this platform via the public GitHub link."

Hình 1. Mô tả Marketing Analytics

Dữ liệu về các chiến dịch Marketing được public vào tháng 12 năm 2020 và thu nhập trong 2 năm (từ năm 2012 – 2014)

Thông qua dữ liệu, người dùng có thể biết được thông tin của khách hàng, số tiền đã chi trả cho từng loại hàng hóa, số hóa đơn đã đặt theo các hình thức khác nhau, phản hồi của khách hàng,...

Dữ liệu gồm 28 cột, 2240 dòng.

Link dữ liệu: [Marketing Analytics | Kaggle](#)

1.1.3 Thuộc tính kho dữ liệu

STT	Thuộc tính	Kiểu dữ liệu	Ý nghĩa	Giá trị
1	ID	Int	Mã khách hàng	Random number
2	Year_Birth	Int	Năm sinh của khách hàng	Trong khoảng [1893 – 1996]
3	Education	Varchar(255)	Trình độ học vấn	1. Graduation 2. Master 3. PhD 4. 2n Cycle 5. Basic
4	Marital_Status	Varchar(255)	Tình trạng hôn nhân	1. Married 2. Single 3. Widow 4. Divorced 5. Together 6. Alone 7. YoLo 8. Absurd
5	Income	Object	Thu nhập hằng năm	Trong khoảng [1730 – 666666]
6	Kidhome	Int	Số lượng trẻ em trong nhà	Trong khoảng [0 – 2]
7	Teenhome	Int	Số lượng trẻ vị thành niên trong nhà	Trong khoảng [0 -2]
8	Dt_Customer	Date	Ngày ghi danh vào hệ thống	Trong khoảng [2012 – 2014]
9	Recency	Int	Số lượng ngày kể từ khi khách hàng mua	Trong khoảng [0 – 99]

			hàng lần cuối	
10	MntWines	Int	Tổng số tiền khách hàng chi trả cho mặt hàng về rượu (trong 2 năm gần nhất)	Trong khoảng [0 – 1493]
11	MntFruits	Int	Tổng tiền khách hàng chi trả cho mặt hàng về hoa quả (trong 2 năm gần nhất)	Trong khoảng [0 – 199]
12	MntMeatProducts	Int	Tổng tiền khách hàng chi trả cho mặt hàng về thịt (trong 2 năm gần nhất)	Trong khoảng [0 – 1725]
13	MntFishProducts	Int	Tổng tiền khách hàng chi trả cho mặt hàng về cá (trong 2 năm gần nhất)	Trong khoảng [0 – 259]
14	MntSweetProducts	Int	Tổng tiền khách hàng chi trả cho mặt hàng về kẹo (trong 2 năm gần nhất)	Trong khoảng [0 – 262]
15	MntGoldProducts	Int	Tổng tiền khách hàng chi trả cho mặt hàng về vàng (trong 2 năm gần nhất)	Trong khoảng [0 – 321]
16	NumDealsPurchases	Int	Số lượng đơn hàng có khuyến mãi đã thanh toán	Trong khoảng [0 – 15]
17	NumWebPurchases	Int	Số lượng đơn hàng đã thanh toán thông qua website bán hàng	Trong khoảng [0 – 27]
18	NumCatalogPurchases	Int	Số lượng đơn hàng đã thanh toán thông qua catalouge	Trong khoảng [0 – 28]
19	NumStorePurchases	Int	Số lượng đơn hàng đã thanh toán trực tiếp tại cửa hàng	Trong khoảng [0 – 13]
20	NumWebVisitsMonth	Int	Số lần mà khách hàng	Trong khoảng

			ghé thăm website mua hàng của cửa hàng (tháng gần nhất)	[1 – 19]
21	AcceptedCmp1	Int	Khách hàng phản hồi Marketing Campaign lần 1	0: Không chấp nhận 1. Chấp nhận
22	AcceptedCmp2	Int	Khách hàng phản hồi Marketing Campaign lần 2	0: Không chấp nhận 1. Chấp nhận
23	AcceptedCmp3	Int	Khách hàng phản hồi Marketing Campaign lần 3	0: Không chấp nhận 1. Chấp nhận
24	AcceptedCmp4	Int	Khách hàng phản hồi Marketing Campaign lần 4	0: Không chấp nhận 1. Chấp nhận
25	AcceptedCmp5	Int	Khách hàng phản hồi Marketing Campaign lần 5	0: Không chấp nhận 1. Chấp nhận
26	Response	Int	Khách hàng phản hồi Marketing Campaign lần cuối (lần gần nhất)	0: Không phàn nàn 1. Phàn nàn
27	Complain	Int	Phản hồi của khách hàng	0: Không phàn nàn 1. Phàn nàn
28	Country	Varchar(255)	Quốc gia	SA: South Africa CA: Canada, SP: Spain GER: Germany IND: India US: United States

1.1.4 Kho dữ liệu đã thực hiện tiền xử lý

Nhóm đã thực hiện các bước như đổi tên một số cột

Sau khi xử lý dữ liệu, ta thu được 2240 dòng và 28 cột

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	User_ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Date_Enroll	Recency	MntWines	MntFruits	Mnt_Meats	Mnt_Fishes	Mnt_Sweets	Mnt_Golds	Num_DealsPur	Num_WebPur	Num_Catn
2	6862	1971	Graduation	Divorced	1730	0	0	5/18/2014	65	1	1	3	1	1	1	15	0	0
3	5376	1979	Graduation	Married	2447	1	0	1/6/2013	42	1	1	1725	1	1	1	15	0	28
4	11110	1973	Graduation	Single	3502	1	0	4/13/2013	56	2	1	1	0	0	1	0	0	0
5	9931	1963	PhD	Married	4023	1	1	6/23/2014	29	5	0	1	1	1	1	15	0	0
6	10311	1969	Graduation	Married	4428	0	1	10/5/2013	0	16	4	12	2	4	321	0	25	0
7	3955	1965	Graduation	Divorced	4861	0	0	6/22/2014	20	2	1	1	1	0	1	0	0	0
8	9303	1976	Graduation	Married	5305	0	1	7/30/2013	12	12	4	7	1	3	5	0	1	0
9	5899	1950	PhD	Together	5648	1	1	3/13/2014	68	28	0	6	1	1	13	1	1	0
10	4246	1982	Master	Single	6560	0	0	12/12/2013	2	67	11	26	4	3	262	0	1	0
11	4303	1957	PhD	Together	6835	0	1	12/8/2012	76	107	2	12	2	2	12	0	0	0
12	6237	1966	PhD	Single	7144	0	2	12/7/2013	92	81	4	33	5	2	291	0	23	1
13	387	1976	Basic	Married	7500	0	0	11/13/2012	59	6	16	11	11	1	16	1	2	0
14	2139	1975	Master	Married	7500	1	0	10/2/2013	19	3	1	10	3	2	12	3	2	0
15	9909	1996	2n Cycle	Married	7500	0	0	11/9/2012	24	3	18	14	15	22	50	3	3	1
16	7297	1973	Graduation	Single	7500	1	0	12/26/2013	54	5	3	10	12	7	20	4	3	1
17	456	1986	2n Cycle	Married	7500	1	0	2/7/2013	96	1	11	5	4	6	9	2	2	0
18	4692	1976	Graduation	Married	7500	1	0	8/1/2012	19	7	0	12	13	7	32	5	4	1
19	9553	1987	Graduation	Married	7500	0	0	1/9/2013	94	0	2	3	4	1	5	1	0	0
20	10710	1979	Graduation	Married	7500	0	1	8/29/2012	61	5	2	3	3	0	5	1	1	0
21	4136	1992	Basic	Single	7500	1	0	12/3/2012	63	10	17	18	8	26	40	4	3	2
22	10641	1978	Graduation	Single	7500	1	1	12/27/2012	5	6	5	4	13	4	25	4	2	1
23	10001	1985	2n Cycle	Together	7500	1	0	8/1/2012	98	5	17	17	13	14	34	4	2	1
24	10659	1979	2n Cycle	Together	7500	1	0	5/7/2013	7	2	8	11	3	8	21	4	3	2

Hình 2. Dữ liệu sau xử lý

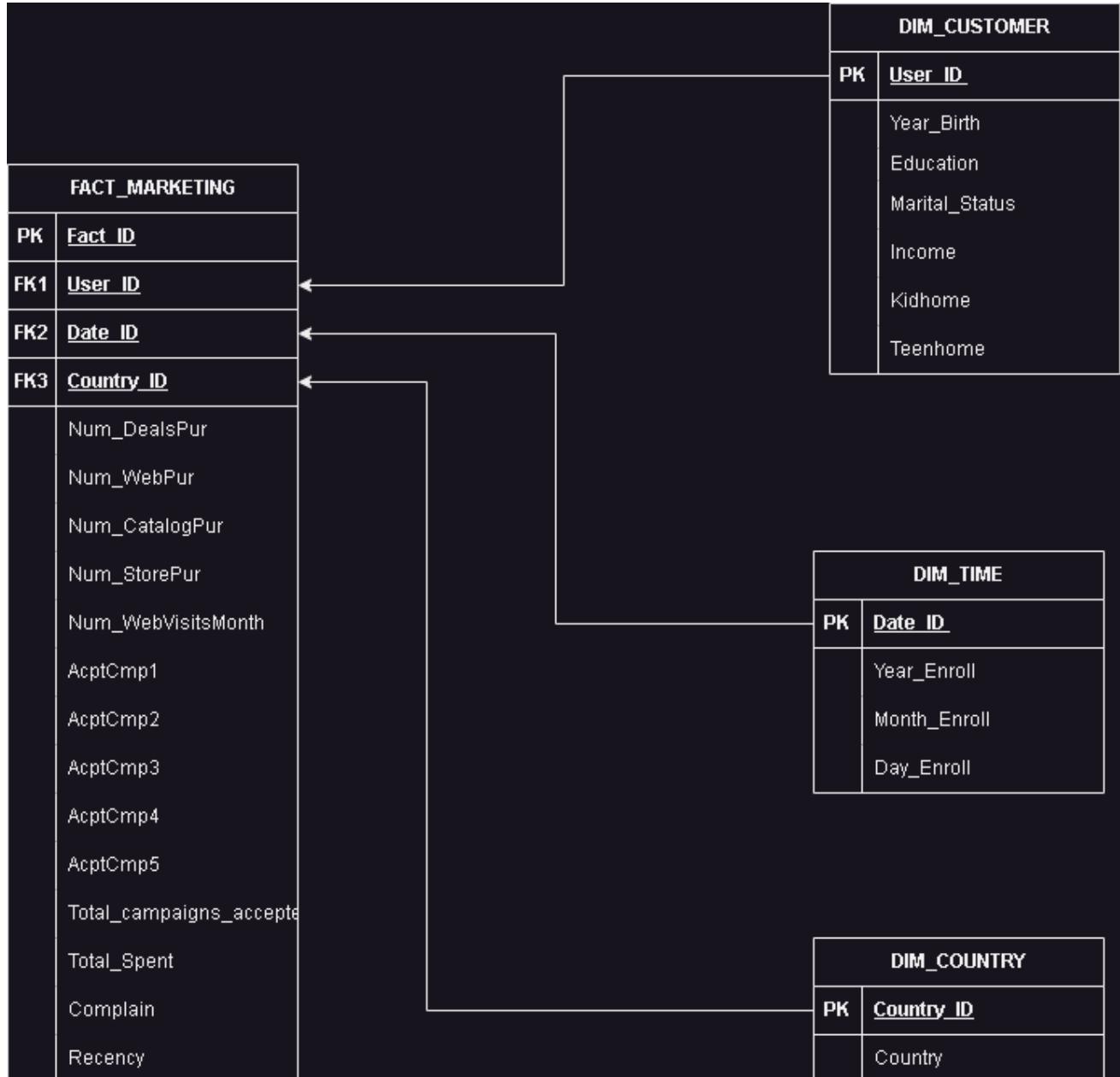
STT	Thuộc tính	Kiểu dữ liệu	Ý nghĩa	Giá trị
1	ID	Int	Mã khách hàng	Random number
2	Year_Birth	Int	Năm sinh của khách hàng	Trong khoảng [1893 – 1996]
3	Education	Varchar(255)	Trình độ học vấn	1. Graduation 2. Master 3. PhD 4. 2n Cycle 5. Basic
4	Marital_Status	Varchar(255)	Tình trạng hôn nhân	1. Married 2. Single 3. Widow 4. Divorced 5. Together 6. Alone 7. YoLo 8. Absurd
5	Income	Object	Thu nhập hàng năm	Trong khoảng [1730 – 666666]
6	Kidhome	Int	Số lượng trẻ em trong nhà	Trong khoảng [0 – 2]
7	Teenhome	Int	Số lượng trẻ vị thành niên trong nhà	Trong khoảng [0 - 2]
8	Data_Enroll	Date	Ngày ghi danh vào hệ thống	Trong khoảng [2012 – 2014]

9	Recency	Int	Số lượng ngày kể từ khi khách hàng mua hàng lần cuối	Trong khoảng [0 – 99]
10	MntWines	Float	Tổng số tiền khách hàng chi trả cho mặt hàng về rượu (trong 2 năm gần nhất)	Trong khoảng [0 – 1493]
11	MntFruits	Float	Tổng tiền khách hàng chi trả cho mặt hàng về hoa quả (trong 2 năm gần nhất)	Trong khoảng [0 – 199]
12	Mnt_Meat	Float	Tổng tiền khách hàng chi trả cho mặt hàng về thịt (trong 2 năm gần nhất)	Trong khoảng [0 – 1725]
13	Mnt_Fish	Float	Tổng tiền khách hàng chi trả cho mặt hàng về cá (trong 2 năm gần nhất)	Trong khoảng [0 – 259]
14	Mnt_Sweet	Float	Tổng tiền khách hàng chi trả cho mặt hàng về kẹo (trong 2 năm gần nhất)	Trong khoảng [0 – 262]
15	Mnt_Gold	Float	Tổng tiền khách hàng chi trả cho mặt hàng về vàng (trong 2 năm gần nhất)	Trong khoảng [0 – 321]
16	Num_DealsPur	Int	Số lượng đơn hàng có khuyến mãi đã thanh toán	Trong khoảng [0 – 15]
17	Num_WebPur	Int	Số lượng đơn hàng đã thanh toán thông qua website bán hàng	Trong khoảng [0 – 27]
18	Num_CatalogPur	Int	Số lượng đơn hàng đã thanh toán thông qua catalouge	Trong khoảng [0 – 28]
19	Num_StorePur	Int	Số lượng đơn hàng đã thanh toán trực tiếp tại cửa hàng	Trong khoảng [0 – 13]

20	Num_WebVisitsMonth	Int	Số lần mà khách hàng ghé thăm website mua hàng của cửa hàng (tháng gần nhất)	Trong khoảng [1 – 19]
21	AcptCmp1	Int	Khách hàng phản hồi Marketing Campaign lần 1	0: Không chấp nhận 1. Chấp nhận
22	AcptCmp2	Int	Khách hàng phản hồi Marketing Campaign lần 2	0: Không chấp nhận 1. Chấp nhận
23	AcptCmp3	Int	Khách hàng phản hồi Marketing Campaign lần 3	0: Không chấp nhận 1. Chấp nhận
24	AcptCmp4	Int	Khách hàng phản hồi Marketing Campaign lần 4	0: Không chấp nhận 1. Chấp nhận
25	AcptCmp5	Int	Khách hàng phản hồi Marketing Campaign lần 5	0: Không chấp nhận 1. Chấp nhận
26	Response	Int	Khách hàng phản hồi Marketing Campaign lần cuối (lần gần nhất)	0: Không phàn nàn 1. Phàn nàn
27	Complain	Int	Phản hồi của khách hàng	0: Không phàn nàn 1. Phàn nàn
28	Country	Varchar(255)	Quốc gia	SA: South Africa CA: Canada, SP: Spain GER: Germany IND: India US: United States

1.2 Xây dựng kho dữ liệu

1.2.1 Lược đồ kho dữ liệu (Star Schema)



Hình 3. Star Schema

1.2.2 Chi tiết các bảng

+BẢNG FACT

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
Fact_ID	Int	Khóa chính

User_ID	Int	Khóa ngoại
Date_ID	Date	Khóa ngoại
Country_ID	Int	Khóa ngoại
Num_DealsPur	Int	
Num_WebPur	Int	
Num_CatalogPur	Int	
Num_StorePur	Int	
Num_WebVisitsMonth	Int	
AcptCmp1	Int	
AcptCmp2	Int	
AcptCmp3	Int	
AcptCmp4	Int	
AcptCmp5	Int	
Total_campaigns_accepted	Int	
Total_Spent	Int	
Complain	Int	
Recency	Int	

✚ BẢNG DIM_TIME

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
Date_ID	Date	Khóa chính
Year_Enroll	Int	
Month_Enroll	Int	
Day_Enroll	Int	

✚ BẢNG DIM_CUSTOMER

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
User_ID	Int	Khóa chính
Year_Birth	Int	
Education	Varchar(255)	
Marital_Status	Varchar(255)	
Income	Int	
Kidhome	Int	
Teenhome	Int	

💠 **BẢNG DIM_COUNTRY**

Tên thuộc tính	Kiểu dữ liệu	Ràng buộc
Country_ID	Int	Khóa chính
Country	Varchar(255)	

CHƯƠNG 2: XÂY DỰNG KHO DỮ LIỆU – QUÁ TRÌNH SSIS

2.1 Tạo project SSIS và thiết lập kết nối

Bước 1: Mở visual studio 2019 -> Create a new project

Visual Studio 2019

Open recent

- ◀ Older
- SSAS_1.sln 4/4/2023 8:41 AM D:\MONHOC\SQL2017_OLAP\SSAS\SSAS_1
- SSAS.sln 4/4/2023 8:31 AM D:\MONHOC\SQL2017_OLAP\SSAS\SSAS
- SSIS.sln 3/21/2023 3:55 PM D:\MONHOC\SQL2017_OLAP\SSIS\SSIS
- OLAP_SSIS.sln 3/21/2023 1:14 PM C:\Users\LENOVO\source\repos\OLAP_SSIS
- Factttt.sln 3/21/2023 10:09 AM C:\Users\LENOVO\source\repos\Factttt
- OLAP.sln 3/21/2023 9:33 AM C:\Users\LENOVO\source\repos\OLAP

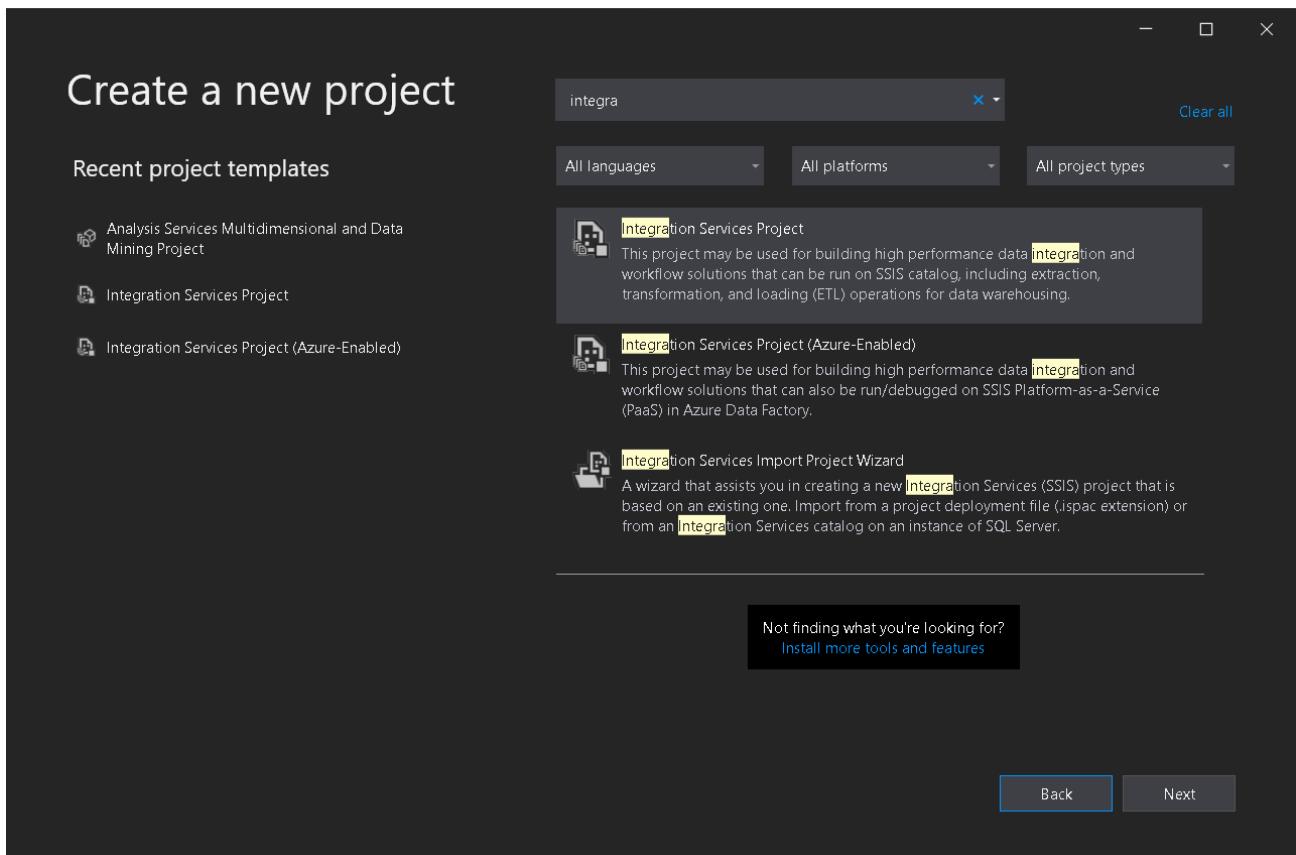
Get started

- Clone a repository Get code from an online repository like GitHub or Azure DevOps
- Open a project or solution Open a local Visual Studio project or .sln file
- Open a local folder Navigate and edit code within any folder
- Create a new project Choose a project template with code scaffolding to get started

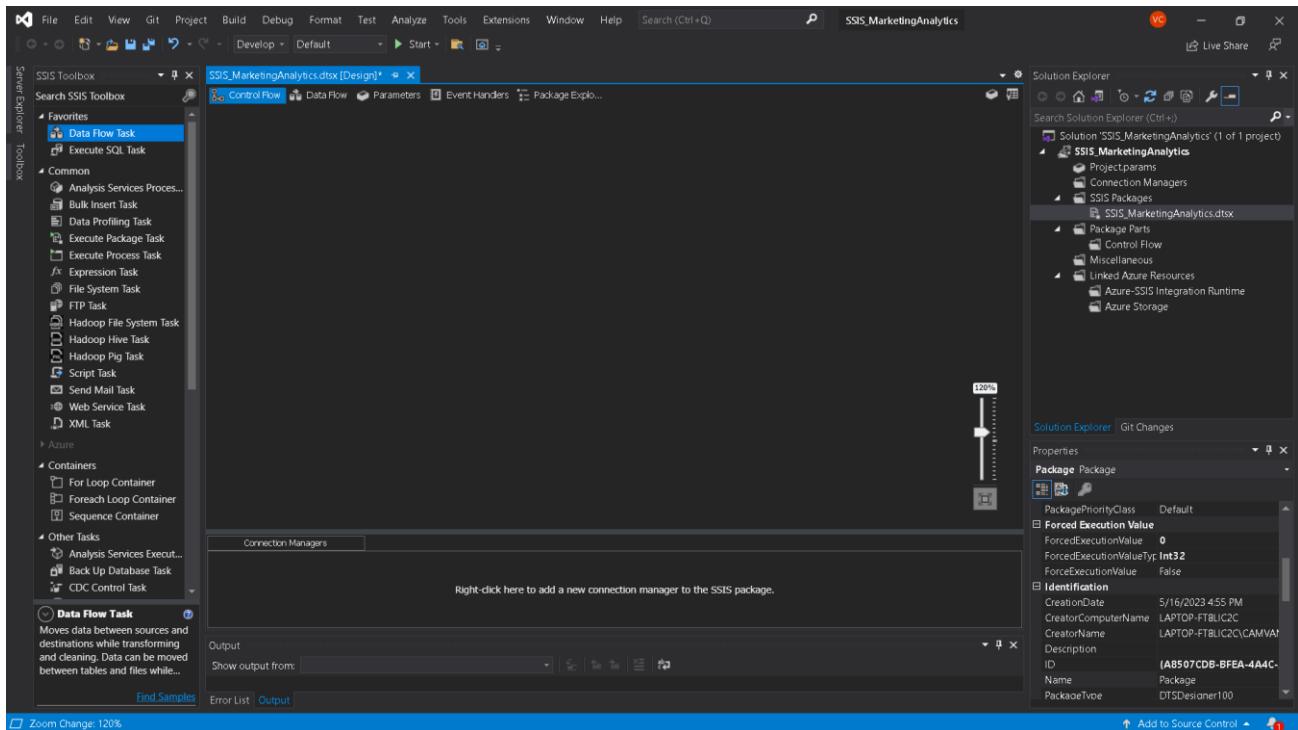
[Continue without code →](#)

Hình 4. Visual Studio 2019

Bước 2: Chọn Integrationa Services Project -> Click chọn và nhấn Next



Hình 5. Tạo SSIS trong Visual Studio 2019

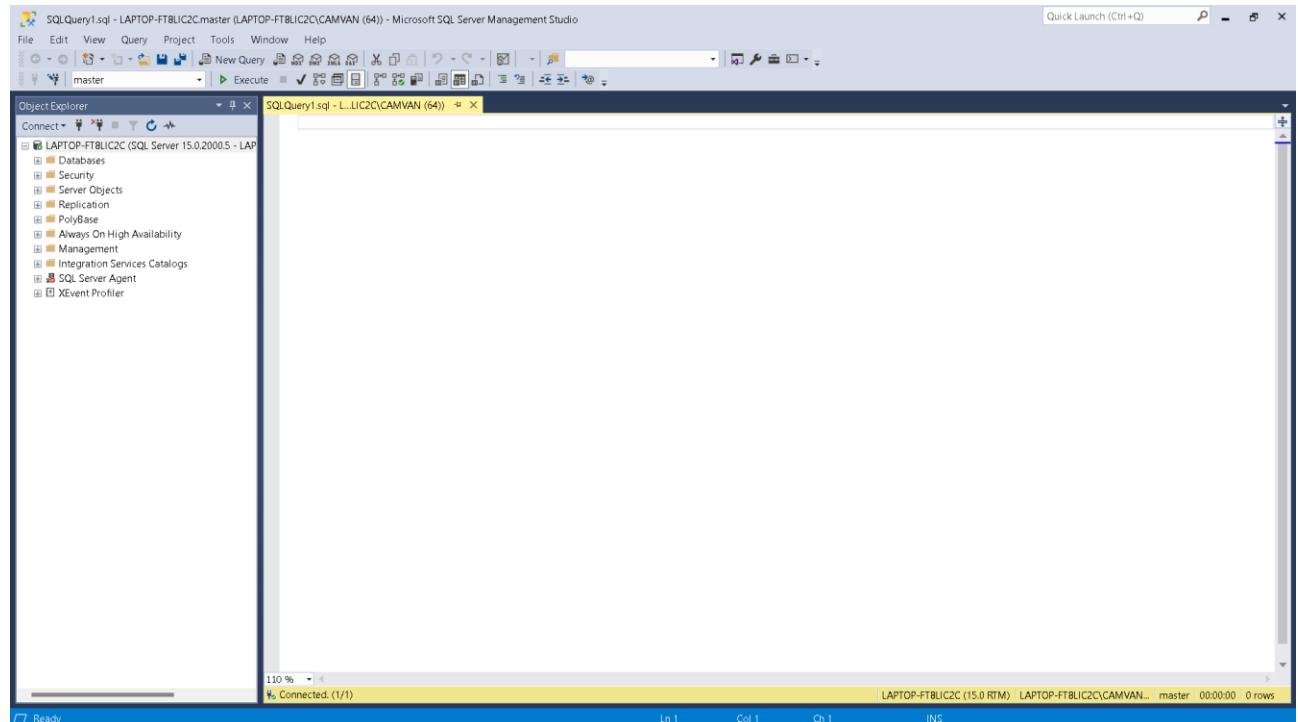


Hình 6. Giao diện công cụ BI trong quá trình SSIS

2.2 Xây dựng kho dữ liệu và thiết lập kết nối

2.2.1 Xây dựng kho dữ liệu

Bước 1: Vào Microsoft SQL Server Management Studio -> Tạo New Query



Hình 7. Giao diện Microsoft SQL Server Management Studio

Bước 2: Viết lệnh tạo database cần thiết -> Execute

The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer on the left, a tree view of databases is shown, including LAPTOP-FTBLIC2C (SQL Server 15.0.2000.5 - L), System Databases, BENHVIEN, DANGKYHOCPHAN, DATA_WAREHOUSE, DATALAKE_DEMO, DATAWAREHOUSE_DEMO, DATAWAREHOUSE_MARKETING, DEMO, OLAP_DEMO, PHANTAN, QBAIHAT, QBANHANG, QBHAT, QLQIAOVU, QLSINHVIEU, QLVEDAP, Test, THEVANHOI, Security, Server Objects, Replication, PolyBase, Always On High Availability, Management, Integration Services Catalogs, SQL Server Agent, and XEvent Profiler. The central pane displays the following SQL command:

```
CREATE DATABASE MARKETING_WAREHOUSE;
```

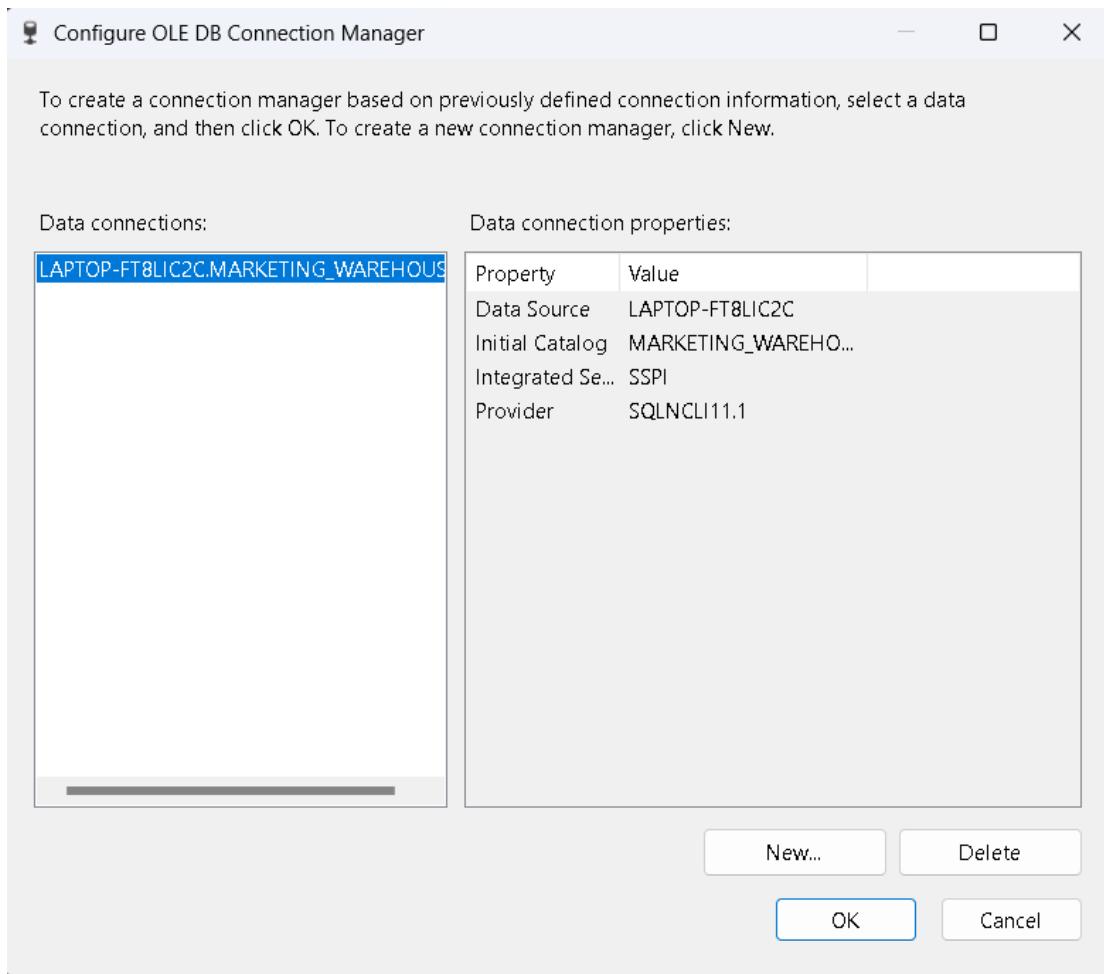
Below the command, the status bar shows "Commands completed successfully." and "Completion time: 2023-05-16T17:01:39.3387216+07:00". The bottom status bar also indicates "Query executed successfully." and provides details about the session: LAPTOP-FTBLIC2C (15.0 RTM) | LAPTOP-FTBLIC2C\CAMVAN... | DATA_WAREHOUSE | 00:00:00 | 0 rows.

Hình 8. Xây dựng kho dữ liệu MARKETING_WAREHOUSE

MARKETING_WAREHOUSE: một kho dữ liệu, chứa các table theo đúng thiết kế Star Schema ở mục Xây dựng lược đồ kho dữ liệu.

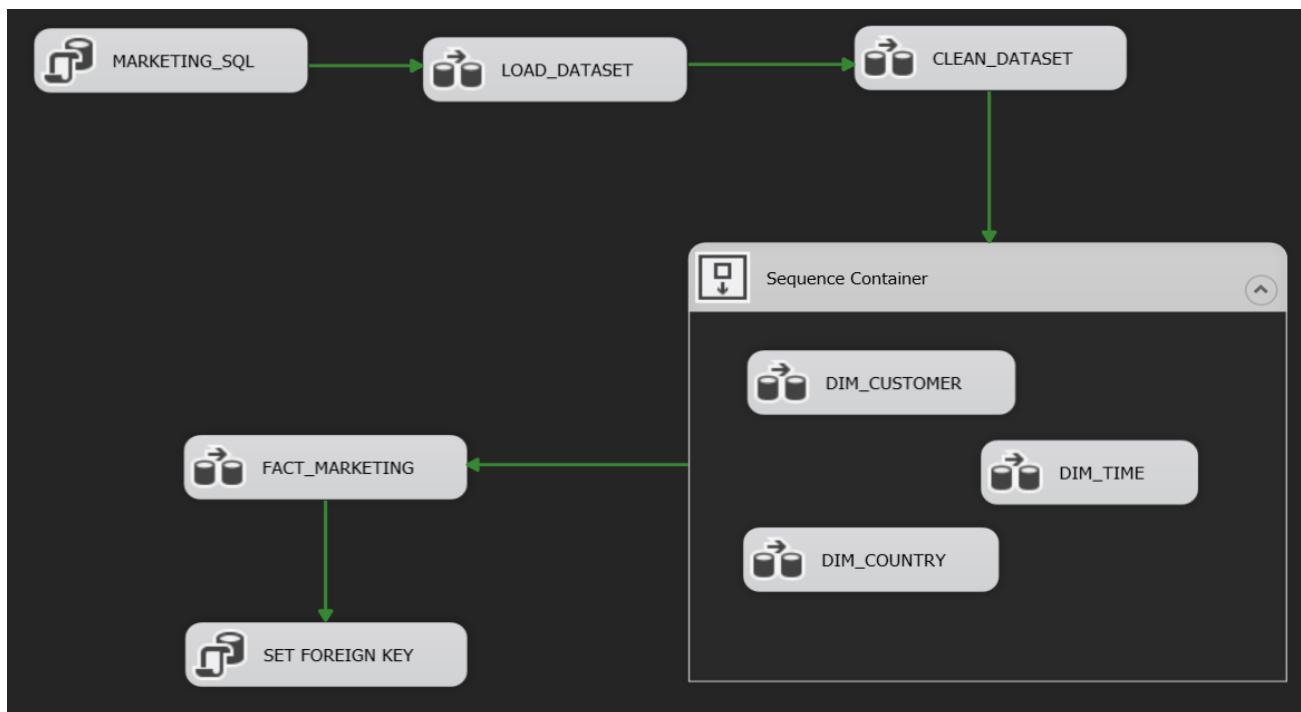
- **Original_Data**: Là table chứa dữ liệu gốc, dữ liệu chưa qua các bước tiền xử lý
- **Marketing_Data**: Là Table chứa dữ liệu đã trải qua các bước tiền xử lý (Lọc NULL...)
- **Null_Data**: Là Table chứa những dòng dữ liệu có giá trị NULL được lọc từ bảng original_data
- **Dim_Customer**: là Table chứa dữ liệu thông tin khách hàng
- **Dim_Time**: là Table chứa dữ liệu thông tin về ngày tháng năm
- **Dim_Country**: Là Table chứa dữ liệu thông tin về địa điểm
- **Fact_Marketing**: Table chứa thông tin và các thuộc tính dùng để tính toán

2.2.2 Thiết lập kết nối



Hình 9. Kết nối đến MARKETING_WAREHOUSE

2.3 Mô hình SSIS



Hình 10. Mô hình SSIS

Bước 1: Load and Clean Dataset

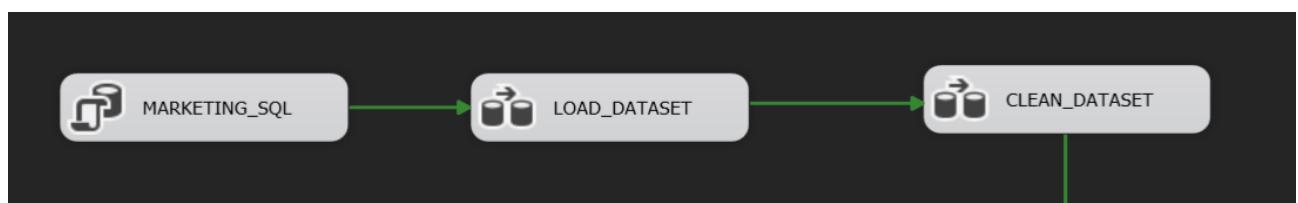
Bước 2: Load Dimension Tables

Bước 3: Load Fact Table

Bước 4: Set Foreign Keys

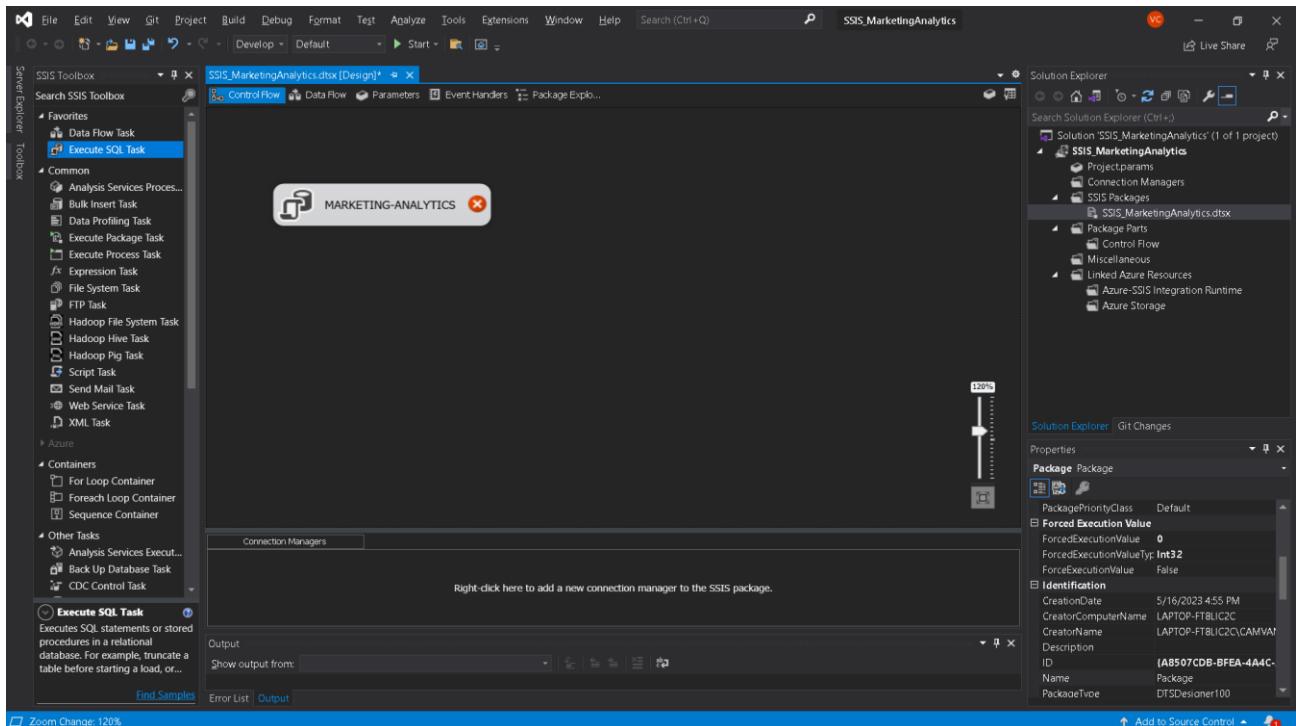
2.3.1 Làm sạch và load dữ liệu

- ❖ **Mục tiêu:** Xóa tất cả dữ liệu đã được insert vào 3 bảng (**Original_Data**, **Marketing_Data**, **Null_Data**) , tiến hành đổ dữ liệu vào bảng Original_Data, làm sạch dữ liệu và đổ dữ liệu vào 2 bảng **Marketing_Data** và **Null_Data** tương ứng.



Hình 11. Load and Clean Dataset

Bước 1: Sử dụng **Execute SQL Task** -> Đổi tên thành **MARKETING_ANALYTICS**



Hình 12. MARKETING_ANALYTICS

Bước 2: Edit MARKETING_ANALYTICS

Chỉnh kết nối của Connection và SQL Statement

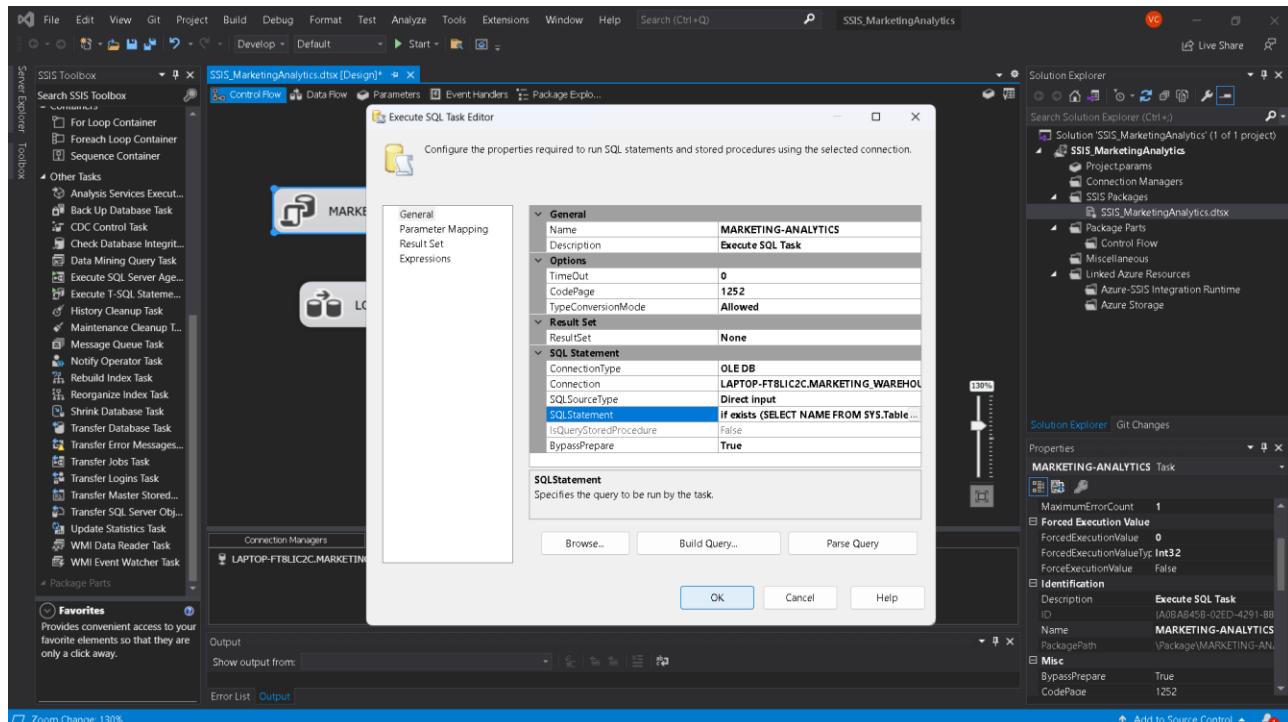
- **Connection:** đến database SQL MARKETING_WAREHOUSE
- **SQL Statement:** Câu lệnh xóa các bảng nếu đã tồn tại để tránh xảy ra lỗi khi load lại dữ liệu

```

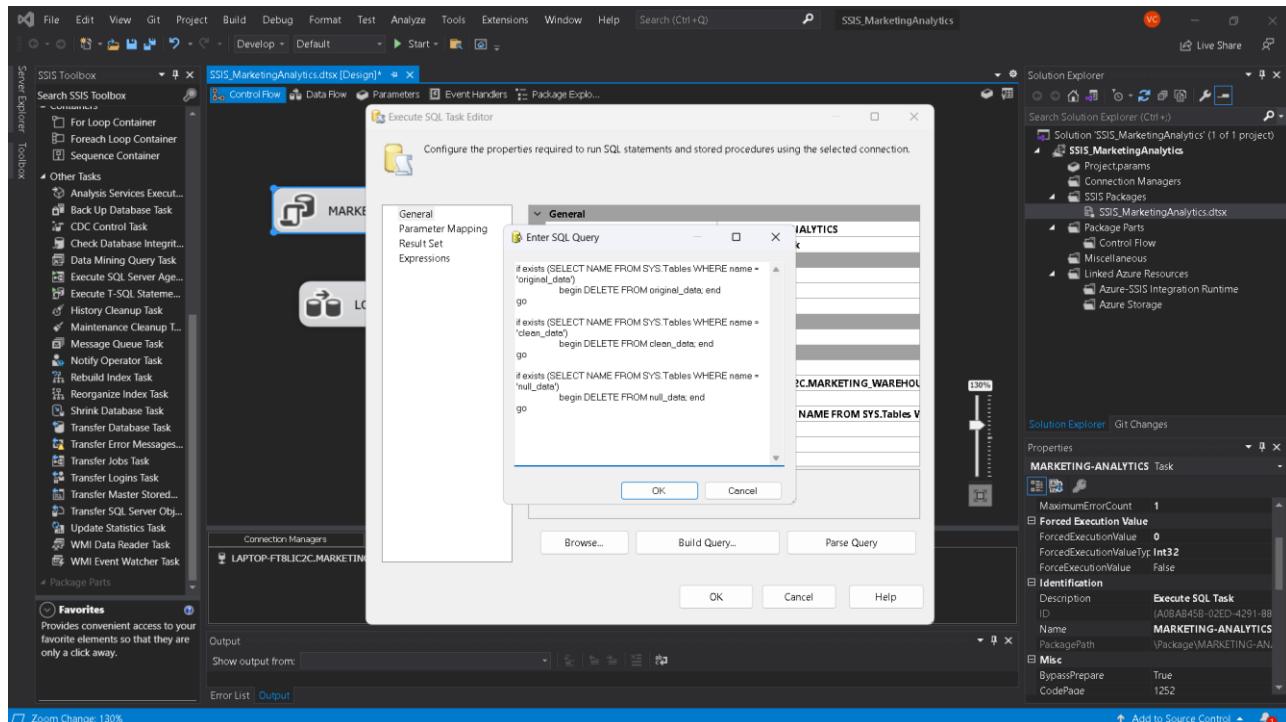
IF EXISTS (SELECT NAME FROM SYS.Tables WHERE NAME = 'ORIGINAL_DATA')
    BEGIN DELETE FROM ORIGINAL_DATA; END GO
IF EXISTS (SELECT NAME FROM SYS.Tables WHERE NAME= 'MARKETING_DATA')
    BEGIN DELETE FROM MARKETING_DATA; END GO
IF EXISTS (SELECT NAME FROM SYS.Tables WHERE NAME = 'NULL_DATA')
    BEGIN DELETE FROM NULL_DATA; END GO
IF EXISTS (SELECT NAME FROM sys.foreign_keys WHERE NAME = 'FK_FACT_TIME')
    BEGIN
        ALTER TABLE FACT_MARKETING DROP CONSTRAINT FK_FACT_TIME END GO
IF EXISTS (SELECT NAME FROM sys.foreign_keys WHERE NAME = 'FK_FACT_COUNTRY')
    BEGIN
        ALTER TABLE FACT_MARKETING DROP CONSTRAINT FK_FACT_COUNTRY END GO
IF EXISTS (SELECT NAME FROM sys.foreign_keys WHERE NAME = 'FK_FACT_CUS')
    BEGIN
        ALTER TABLE FACT_MARKETING DROP CONSTRAINT FK_FACT_CUS END GO
IF EXISTS (SELECT NAME FROM SYS.Tables WHERE name = 'DIM_TIME')
    BEGIN DELETE FROM DIM_TIME; END GO
IF EXISTS (SELECT NAME FROM SYS.Tables WHERE name = 'DIM_CUSTOMER')
    BEGIN DELETE FROM DIM_CUSTOMER; END GO
IF EXISTS (SELECT NAME FROM SYS.Tables WHERE name = 'DIM_COUNTRY')
    BEGIN DELETE FROM DIM_COUNTRY; END GO
IF EXISTS (SELECT NAME FROM SYS.Tables WHERE name = 'FACT_MARKETING')
    BEGIN DELETE FROM FACT_MARKETING; END GO

```

Hình 13. SQL Statement

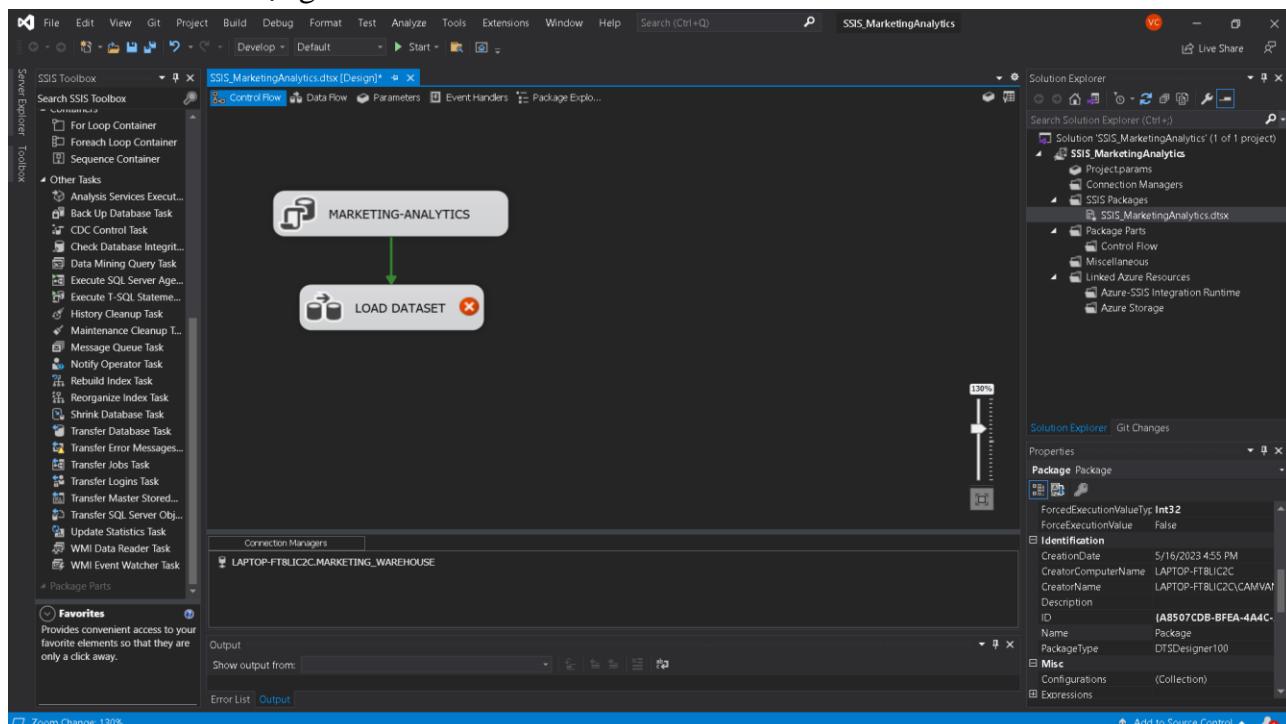


Hình 14. Edit MARKETING_ANALYTICS



Hình 15. Edit SQL Statement

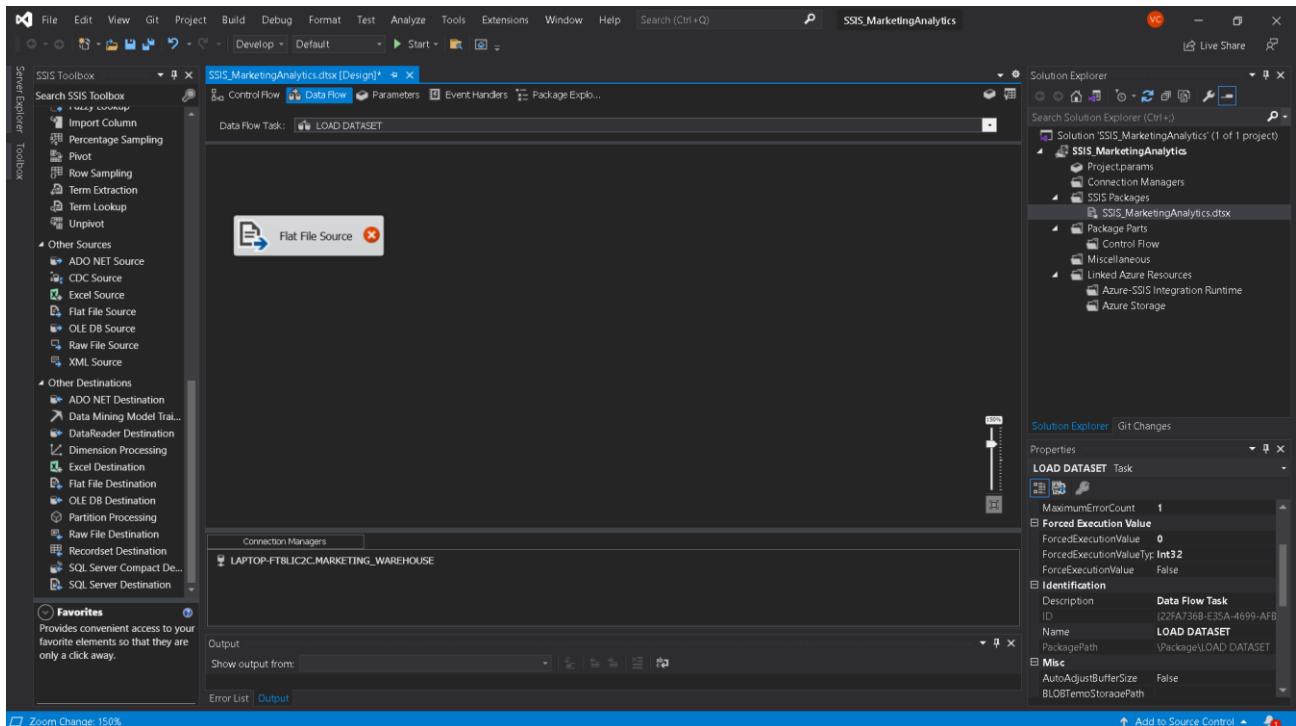
Bước 3: Sử dụng Data Flow Task -> Đổi tên thành LOAD DATASET



Hình 16. LOAD DATASET

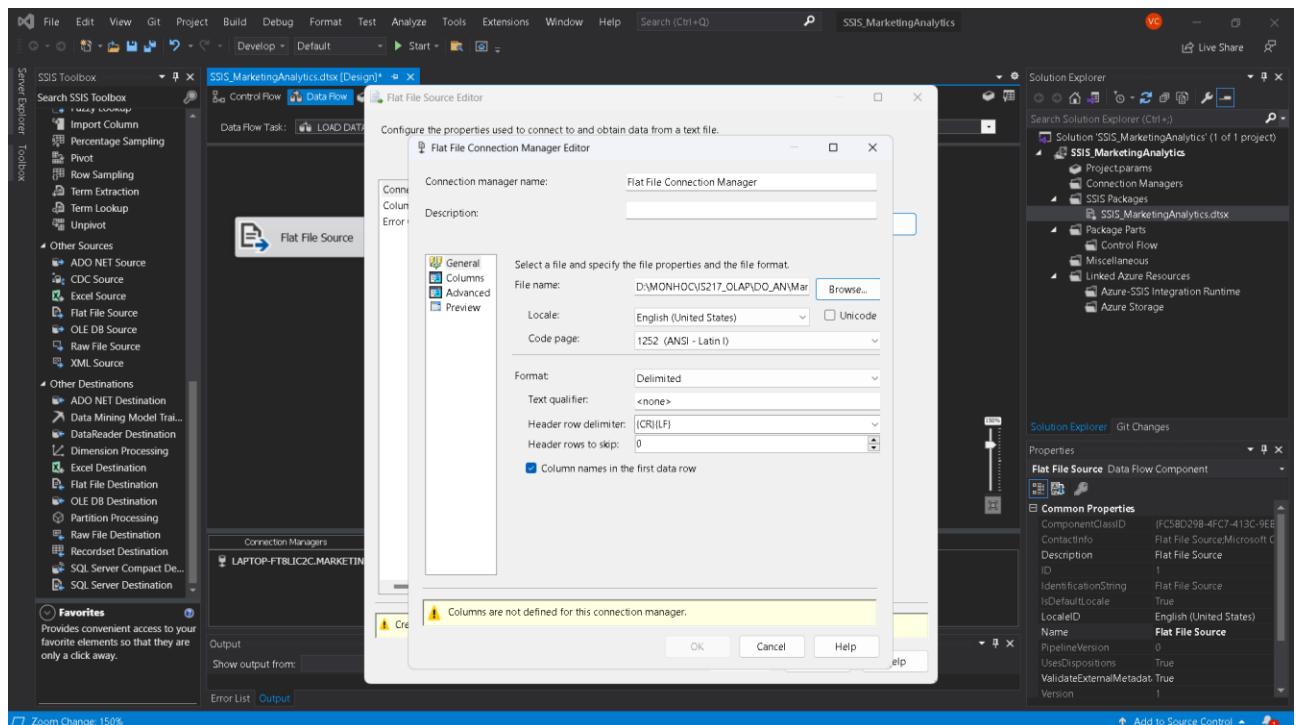
Bước 4: Edit LOAD DATASET

Sử dụng Flat File Source để định nghĩa nguồn dữ liệu



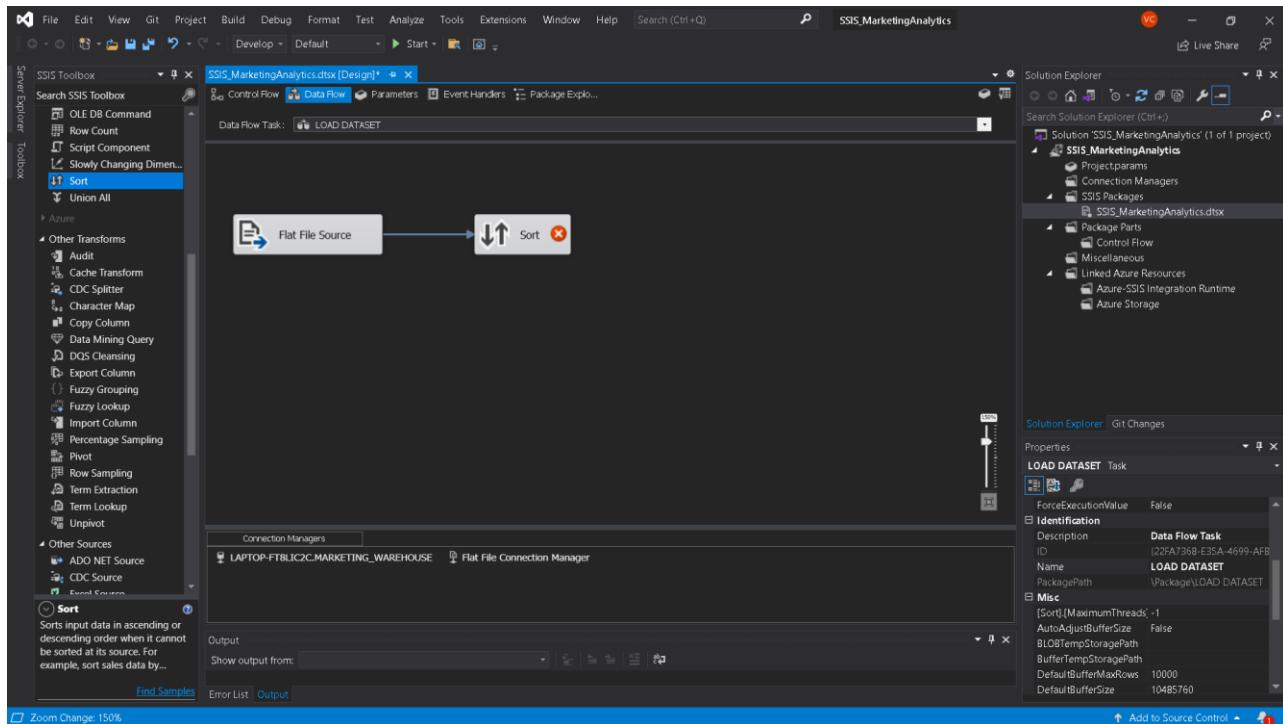
Hình 17. Flat File Source để định nghĩa nguồn dữ liệu

Edit Flat File Source -> Browse đến nơi chứa file dữ liệu với định dạng csv.



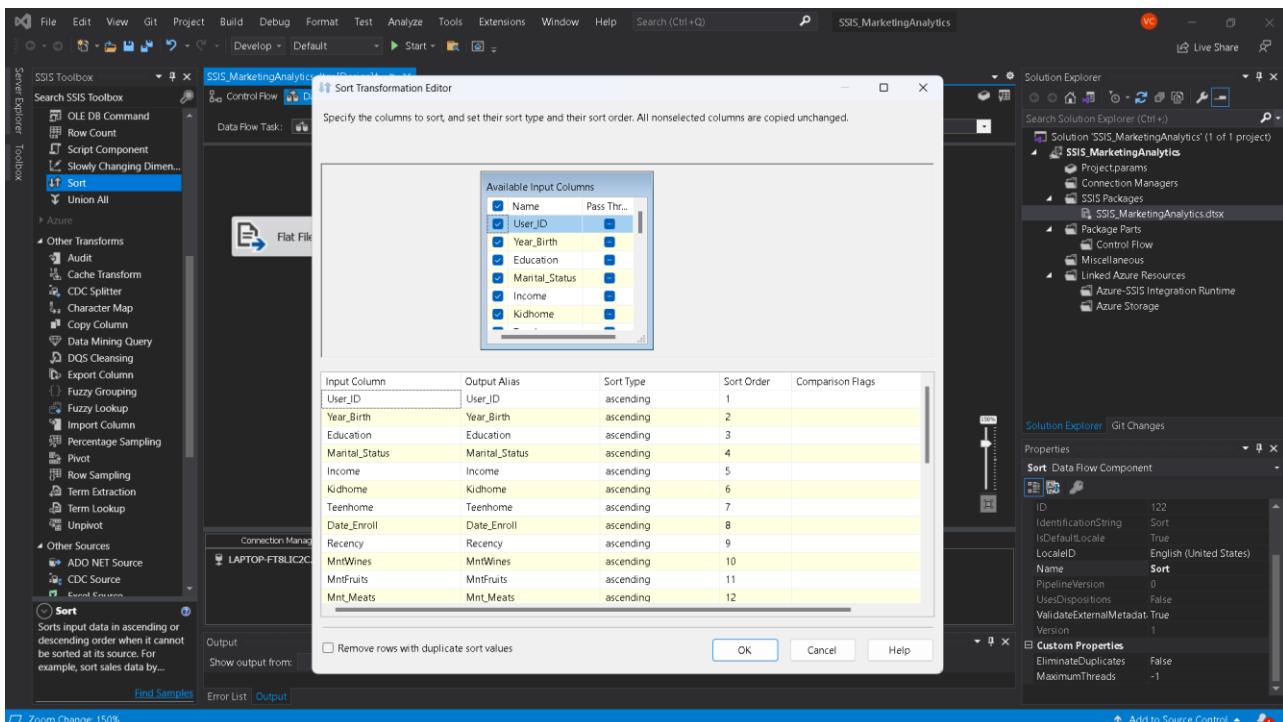
Hình 18. Connection đến file dữ liệu csv

Sử dụng Sort để sắp xếp lại dữ liệu



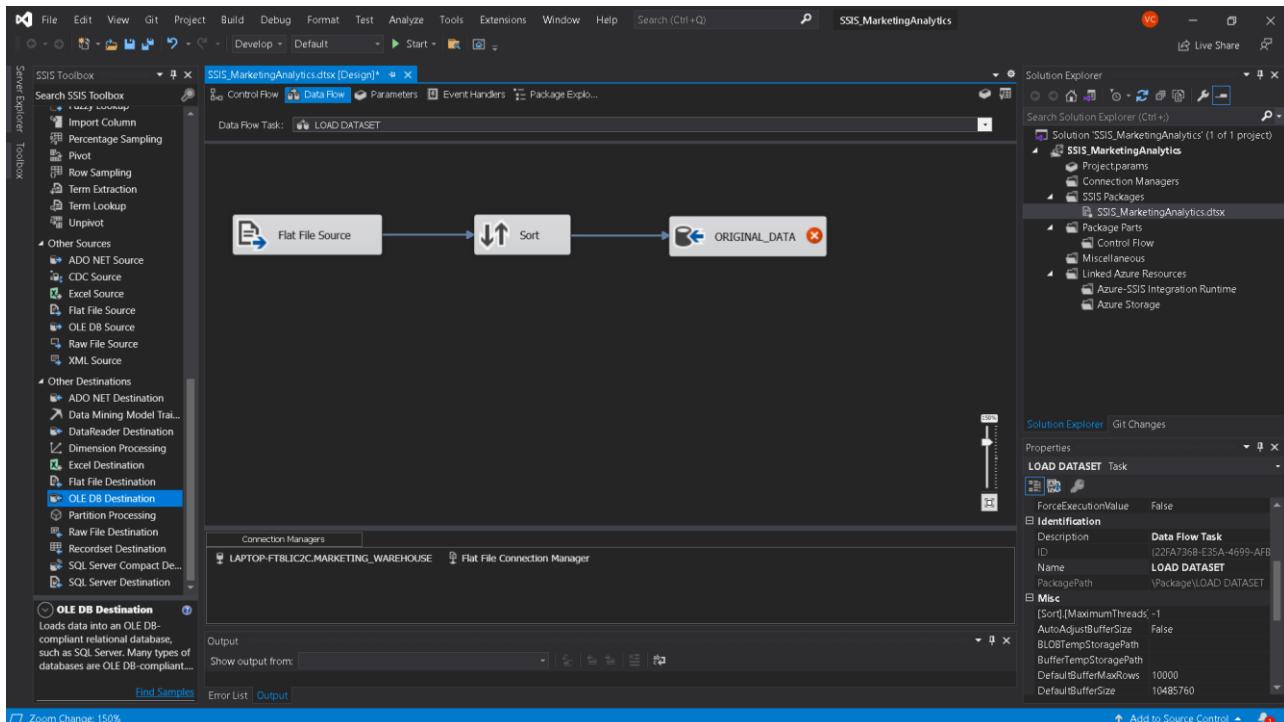
Hình 19. Sort sắp xếp dữ liệu

Edit Sort -> Chọn muôn sắp xếp theo cột dữ liệu nào



Hình 20. Chọn cột để sắp xếp dữ liệu

Sử dụng OLE DB Destination để đổ dữ liệu vào bảng -> Đổi tên thành
ORIGINAL_DATA

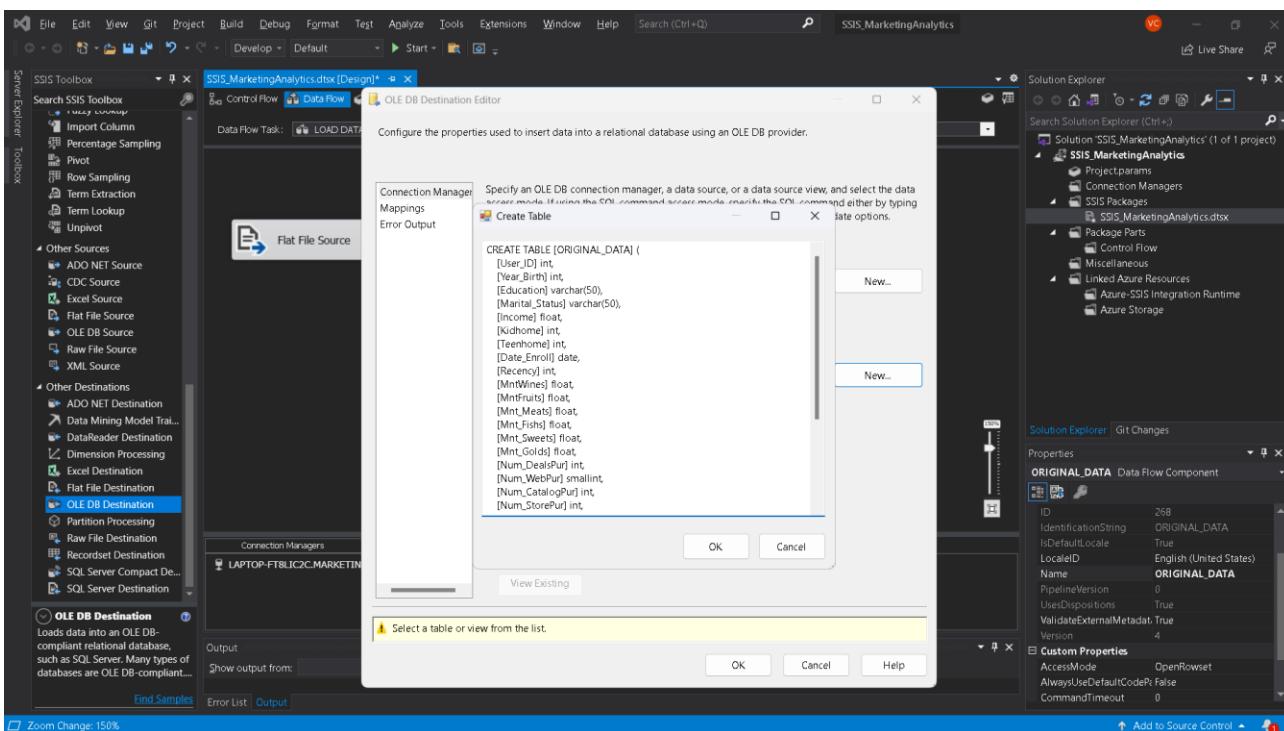


Hình 21. Tạo ORIGINAL_DATA để đỗ dữ liệu vào

Edit ORIGINAL_DATA

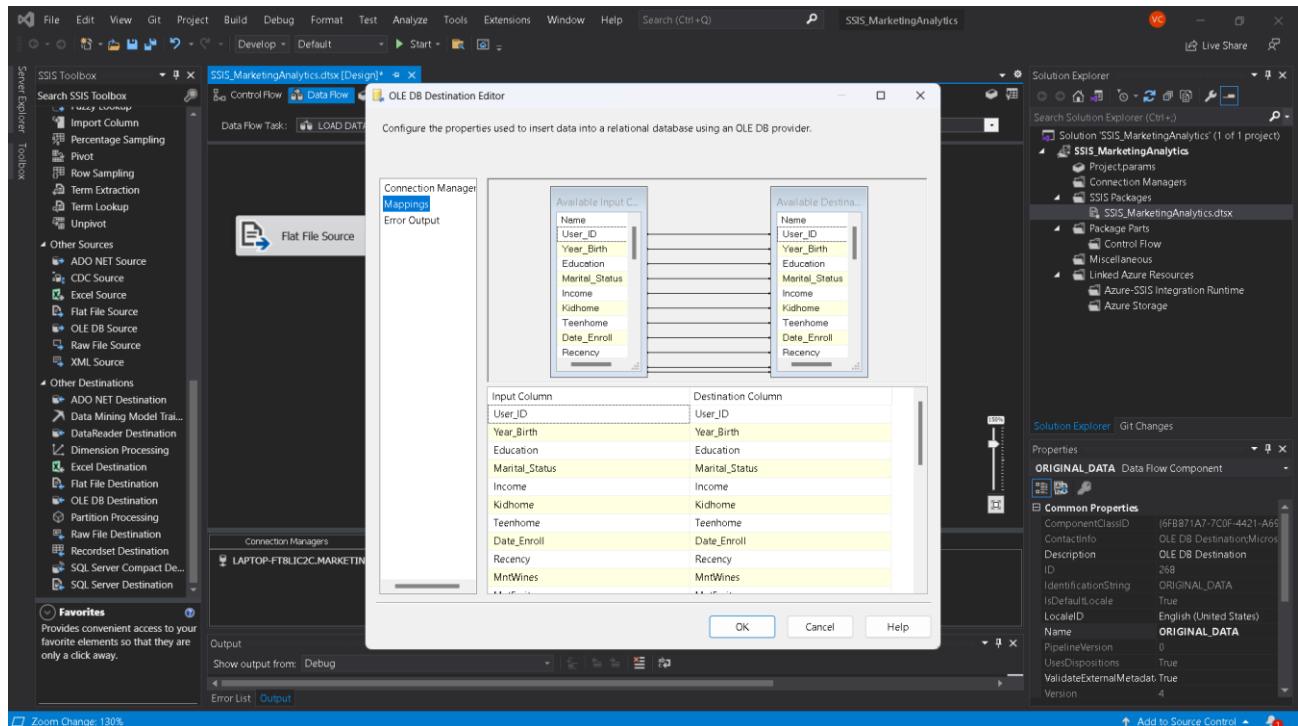
Chọn Connection đến MARKETING_WAREHOUSE

Chọn New để SSIS tự động sinh ra đoạn mã insert data



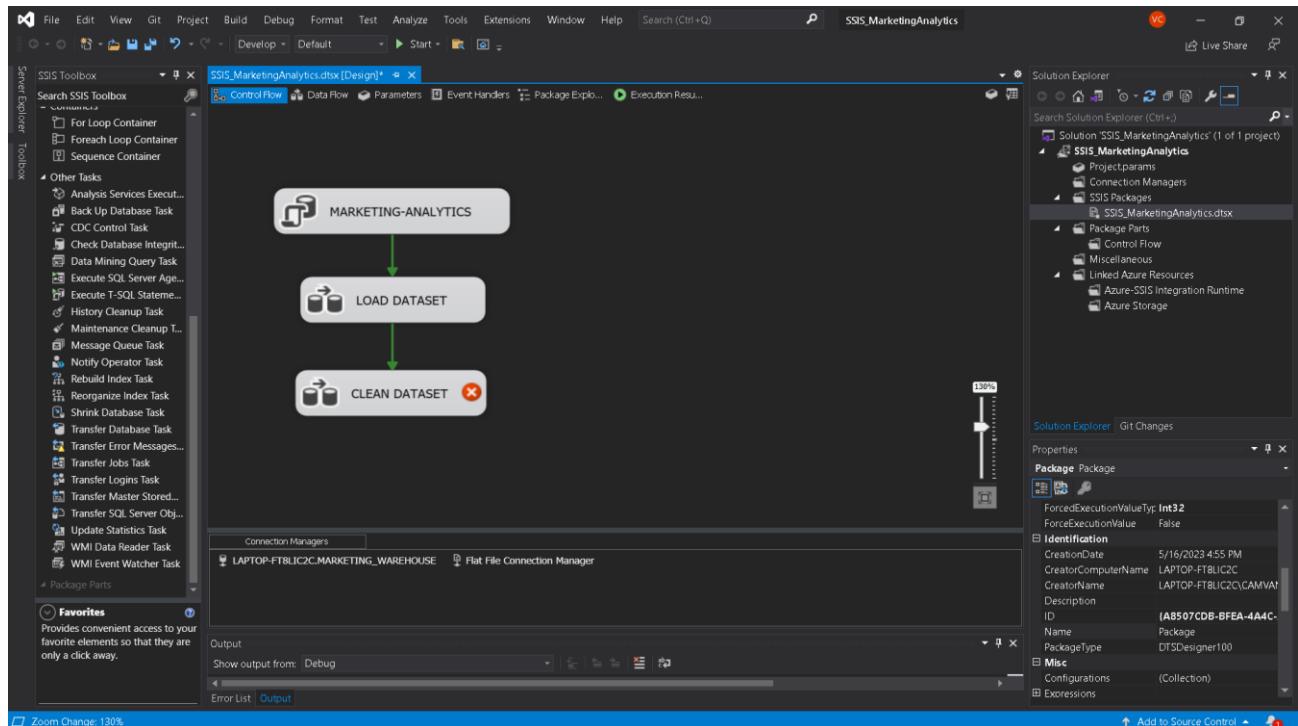
Hình 22. Tạo Bảng ORIGINAL_DATA để đỗ dữ liệu vào

Chọn Mappings -> Kiểm tra columns đã map với nhau đúng hay chưa



Hình 23. Kiểm tra Mappings

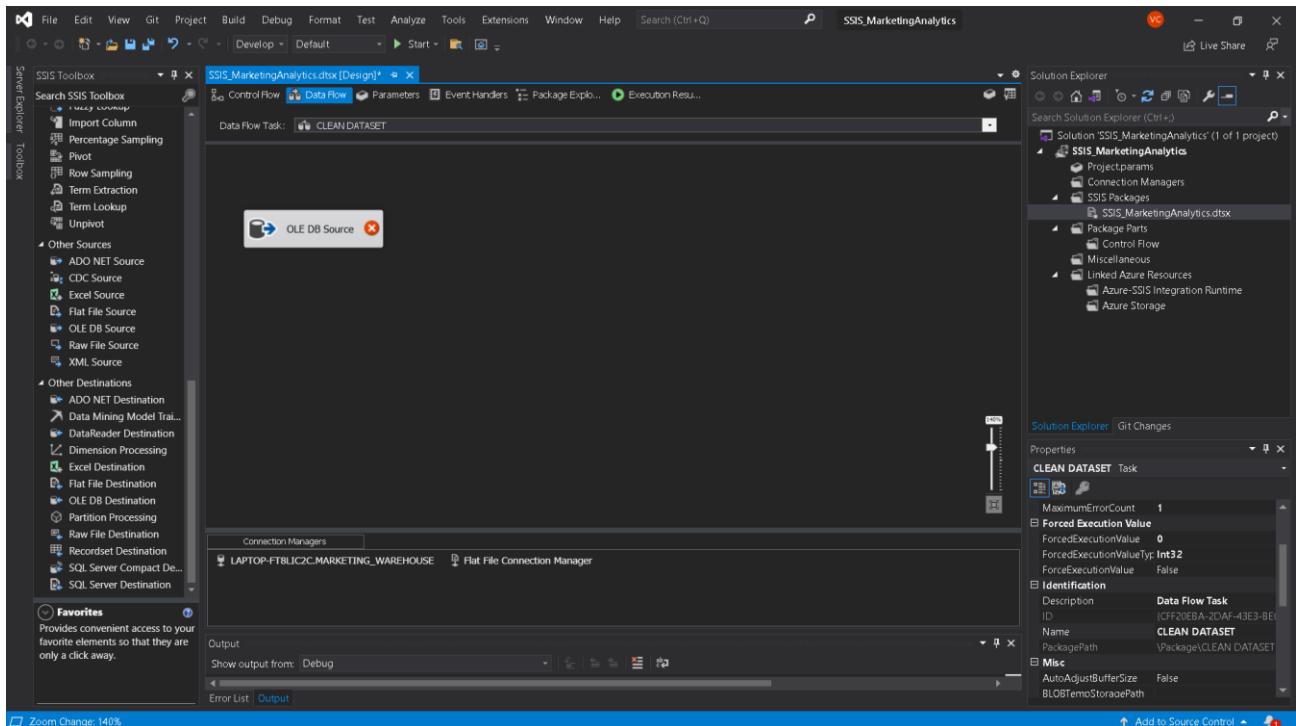
Bước 5: Sử dụng Data Flow Task → Đổi tên thành CLEAN DATASET



Hình 24. CLEAN DATASET

Bước 6: Edit CLEAN DATASET

Sử dụng OLE DB Source để định nghĩa nguồn dữ liệu

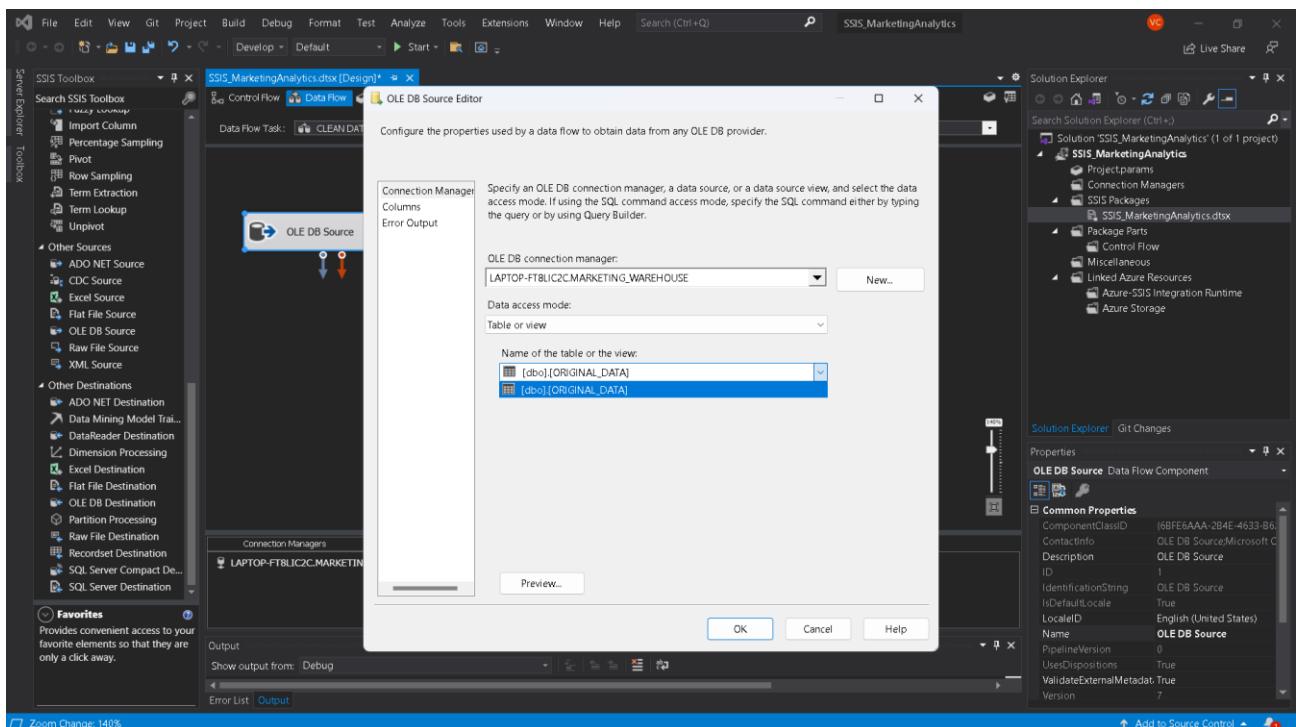


Hình 25. OLE DB Source định nghĩa nguồn dữ liệu

Edit OLE DB Source

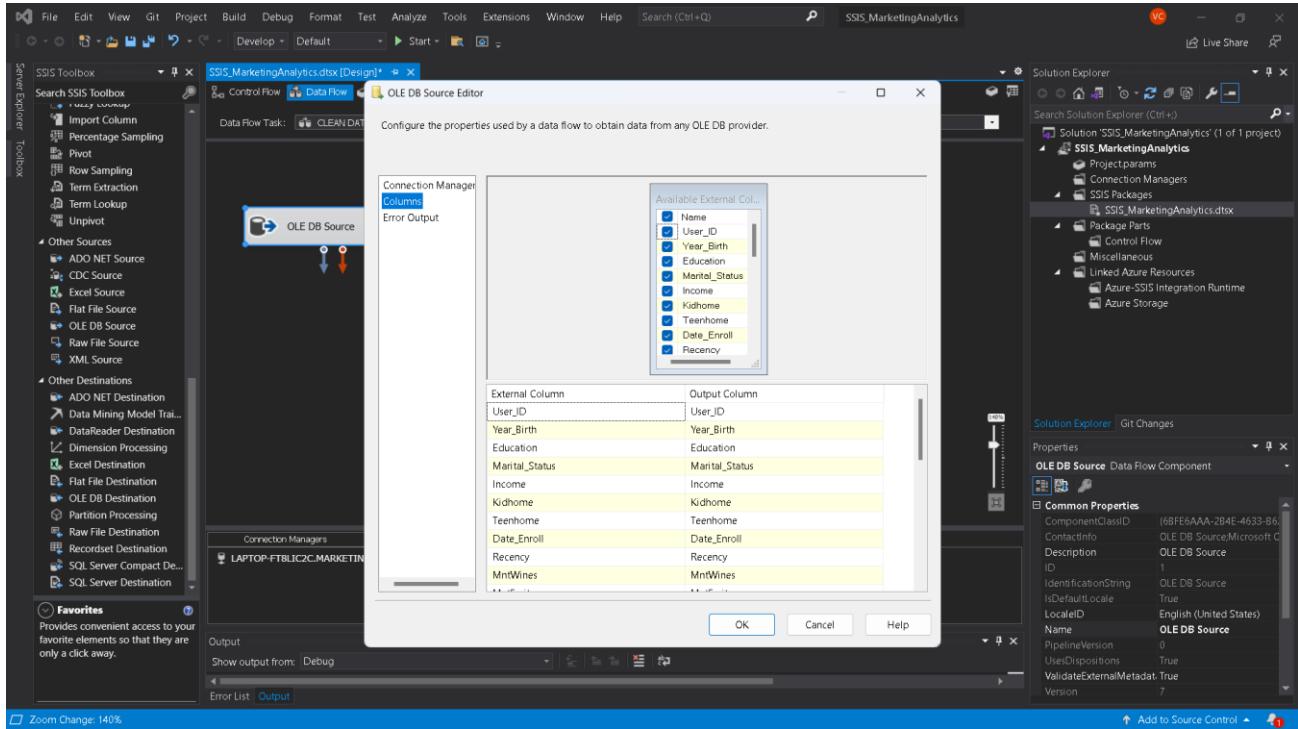
Chọn Connection đến MARKETING_WAREHOUSE

Chọn Table ORIGINAL_DATA



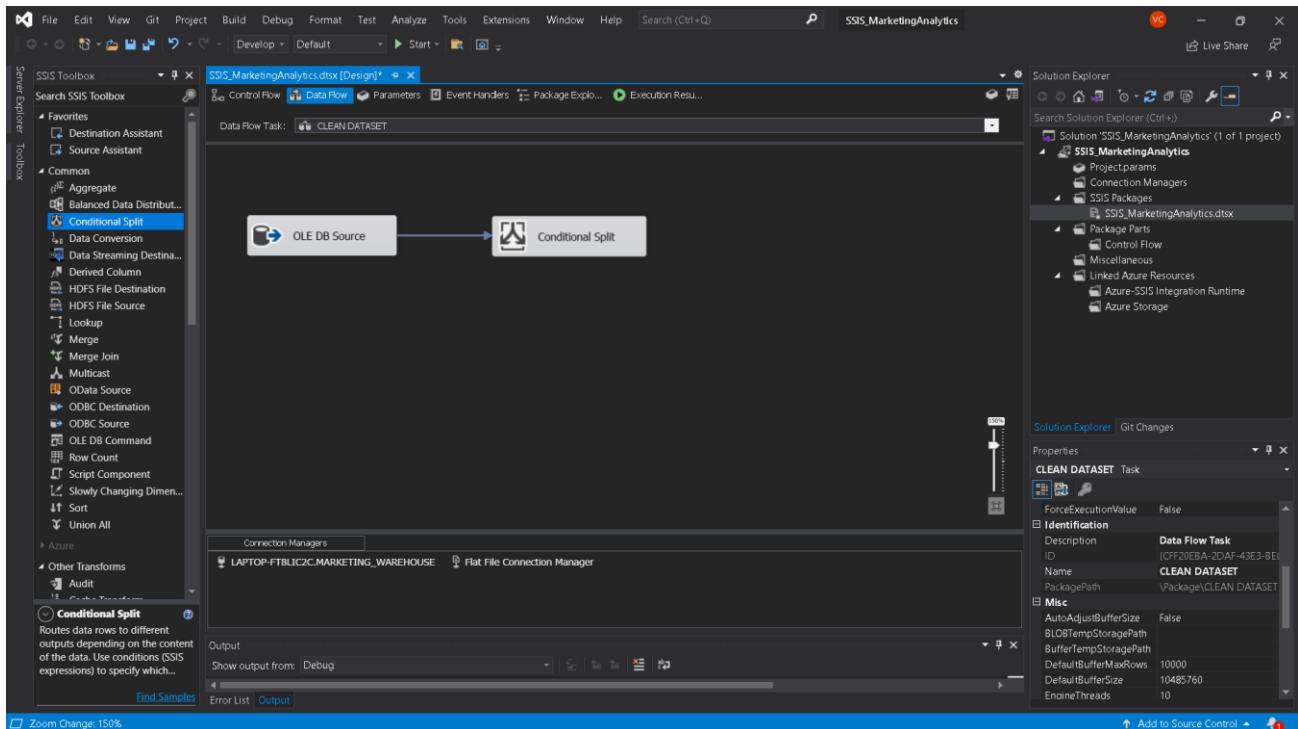
Hình 26. Chọn nguồn dữ liệu là ORIGINAL_DATA

Chọn Columns -> Chọn những thuộc tính cần thiết



Hình 27. Chọn thuộc tính

Sử dụng Conditional Split để lọc dữ liệu

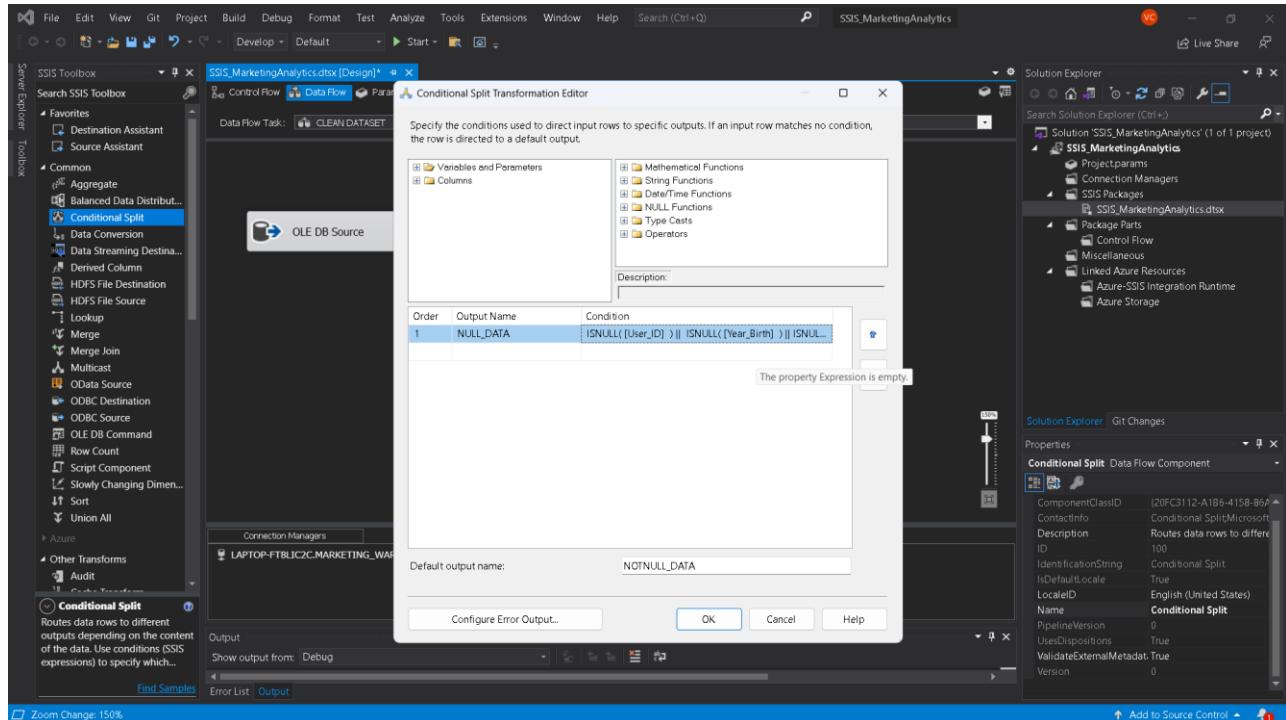


Hình 28. Lọc dữ liệu NULL

Edit Conditional Split

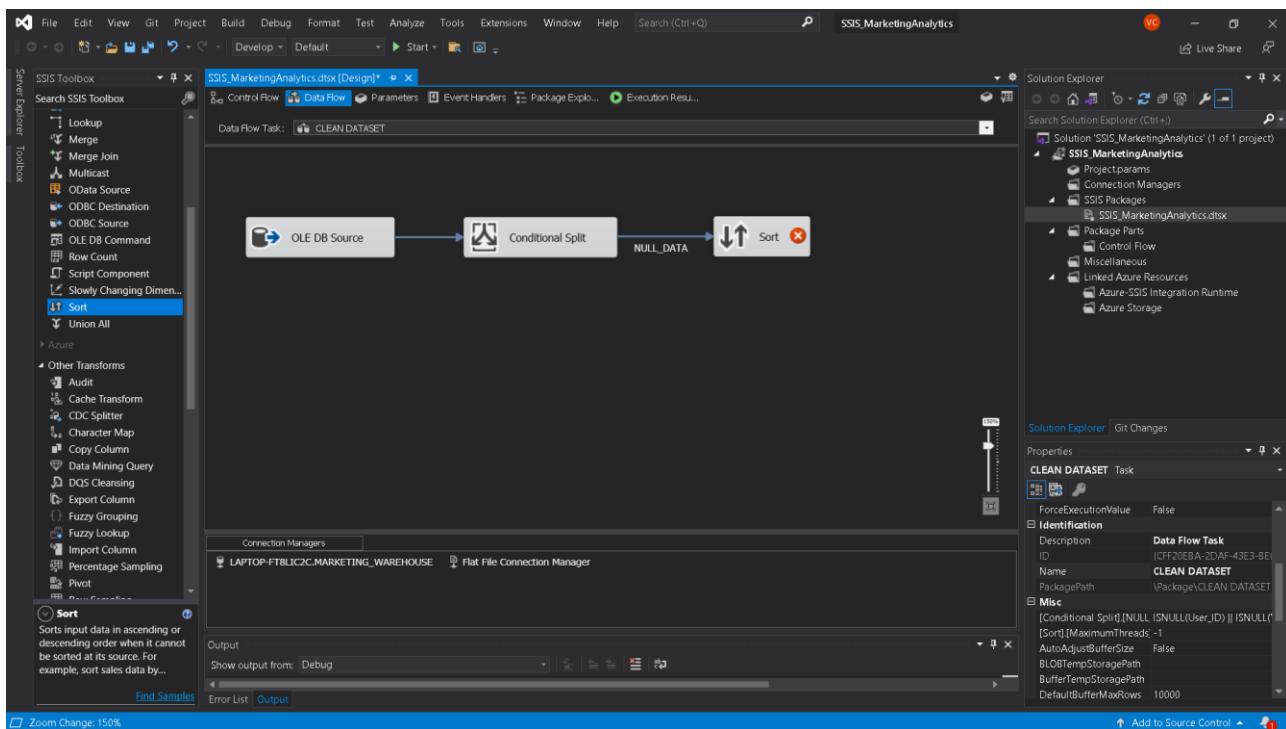
Chỉnh sửa Output Name thành **NULL_DATA** -> Thêm điều kiện lọc dữ liệu Null

Chỉnh sửa Default OutPut Name thành **NOTNULL_DATA**



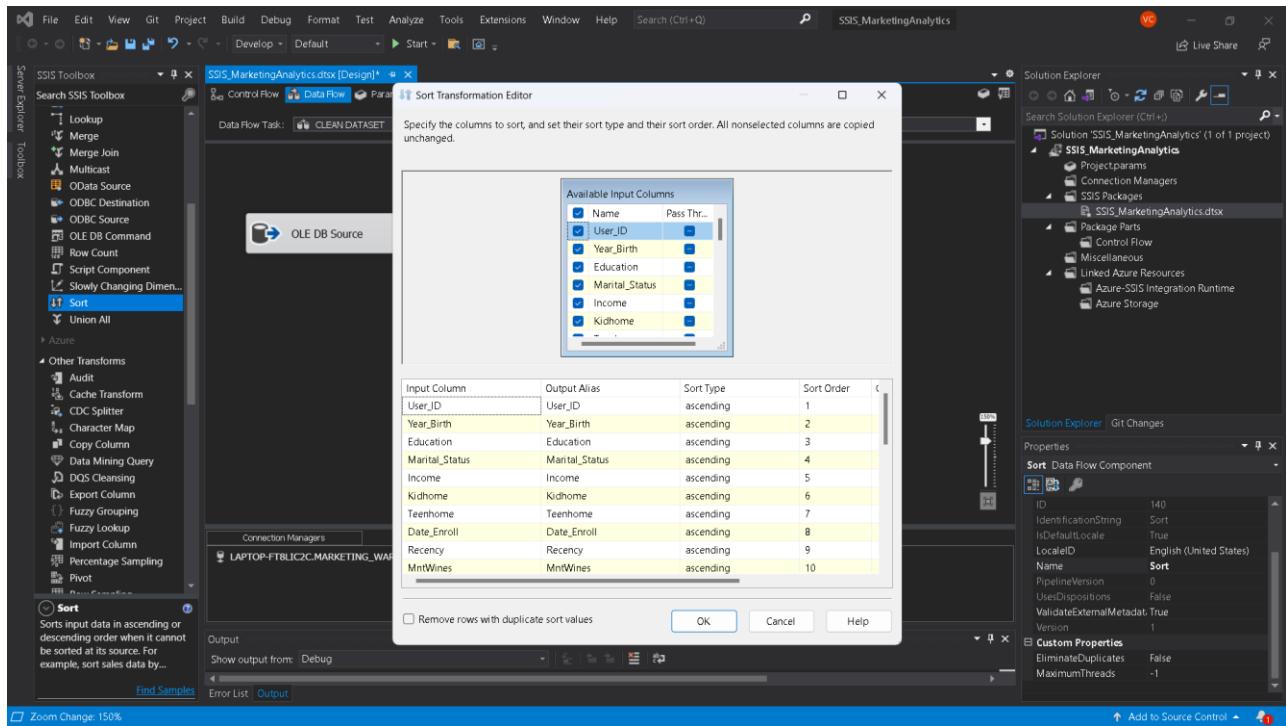
Hình 29. Điều kiện để lọc các dữ liệu NULL

Sử dụng Sort để sắp xếp dữ liệu



Hình 30. Sắp xếp dữ liệu

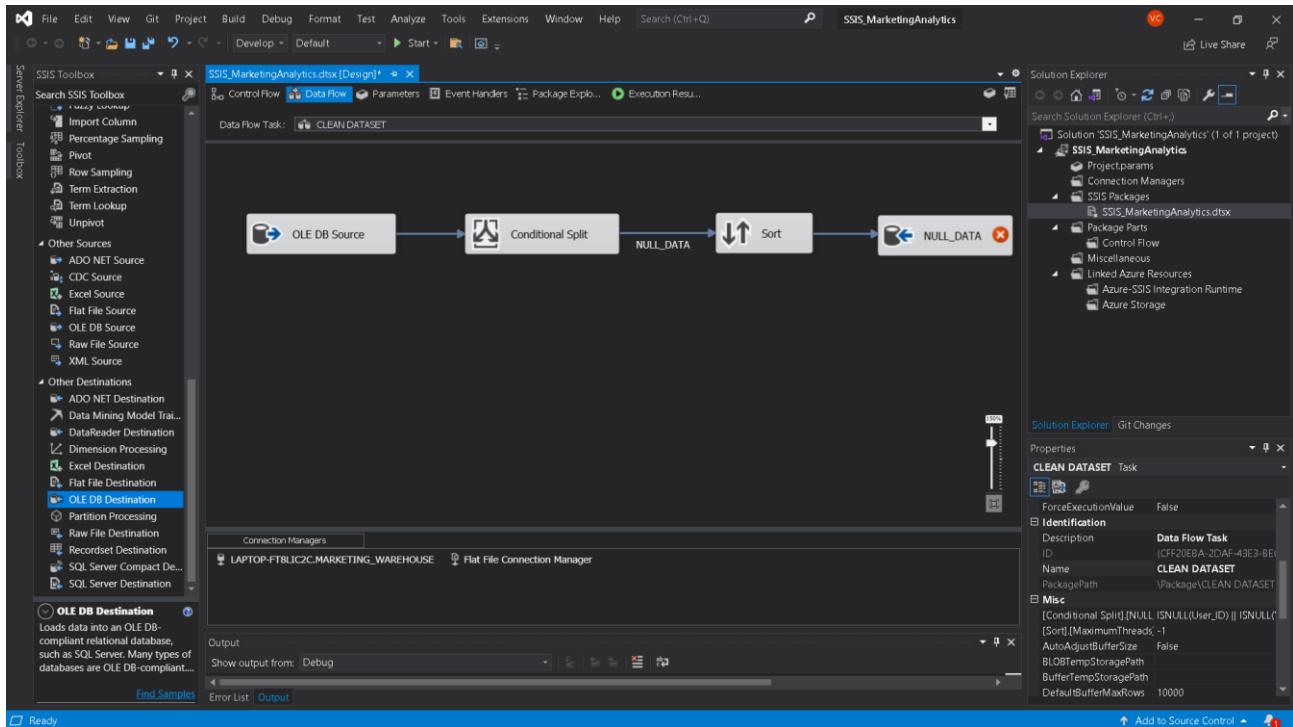
Edit Sort -> Chọn muôn sắp xếp theo thuộc tính nào



Hình 31. Sắp xếp dữ liệu theo thuộc tính

Sử dụng OLE DB Destination để đỗ dữ liệu NULL vào bảng -> Đổi tên thành

NUL_DATA

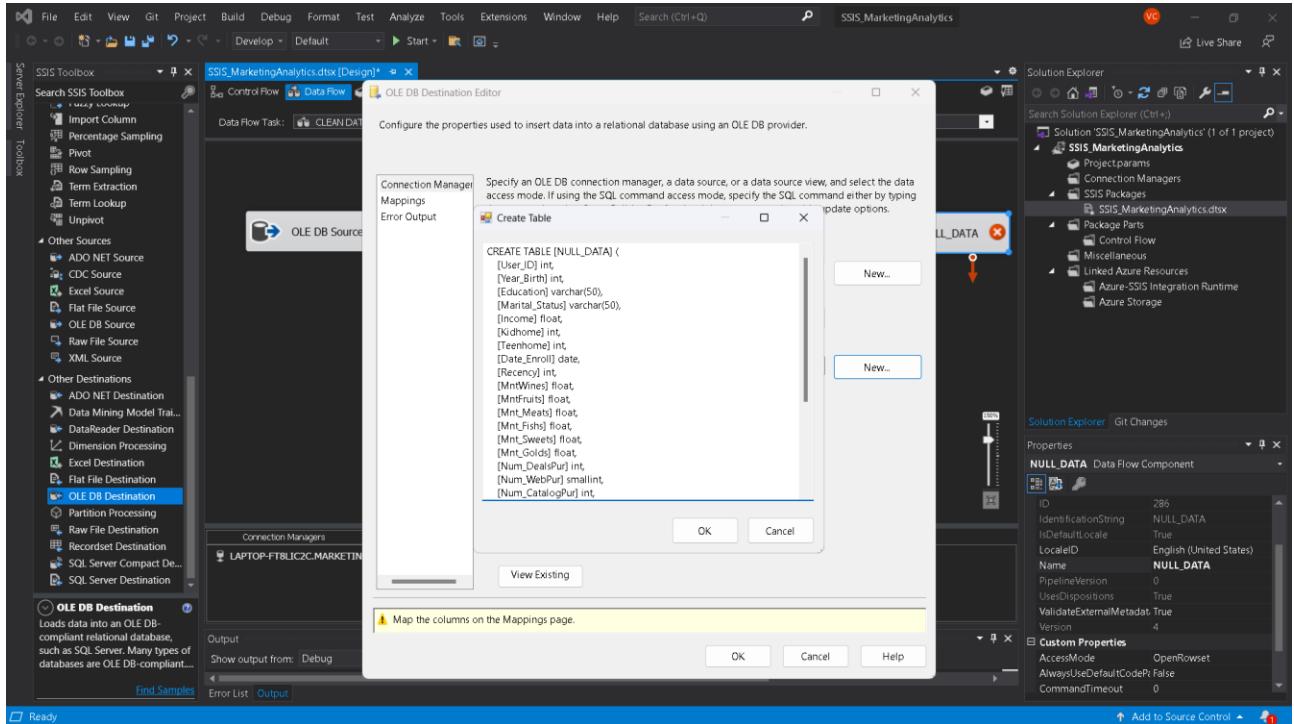


Hình 32. Đỗ dữ liệu NULL vào bảng NUL_DATA

Edit NUL_DATA

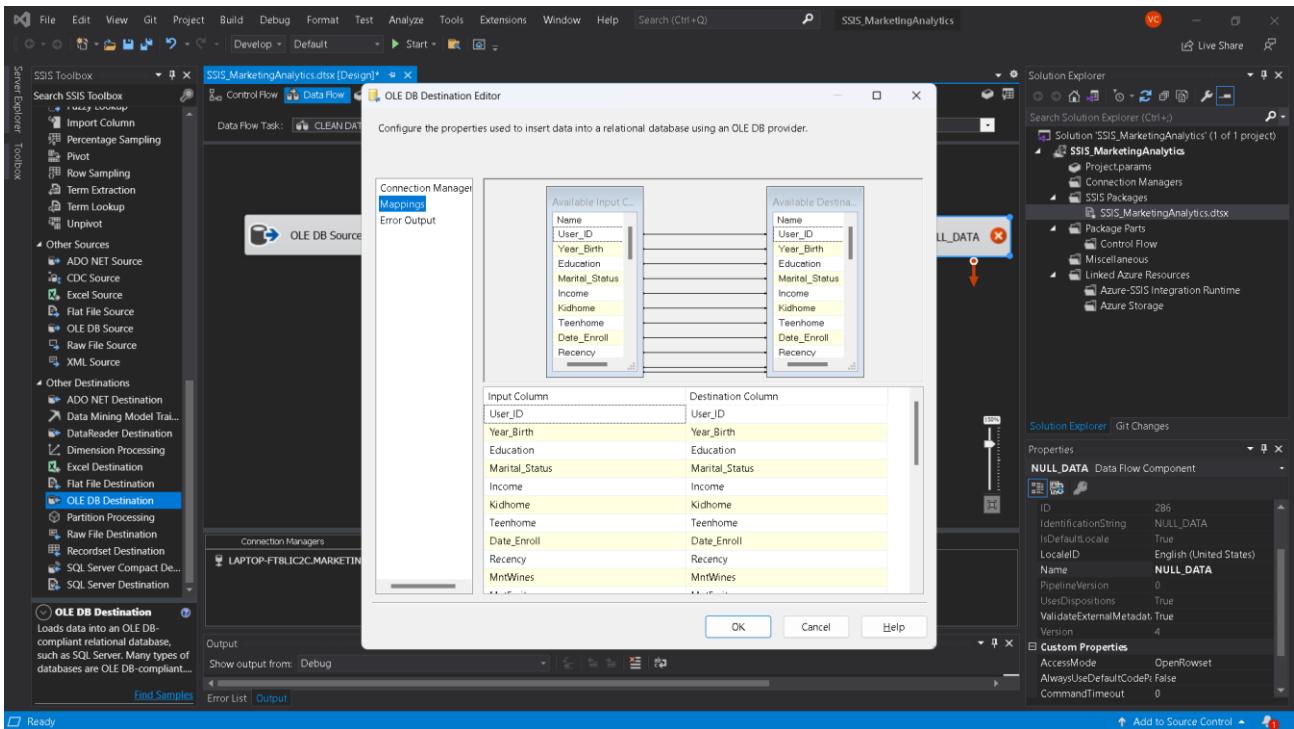
Chọn connection đến MARKETING_WAREHOUSE

Chọn New để SSIS tự động sinh ra đoạn mã insert data



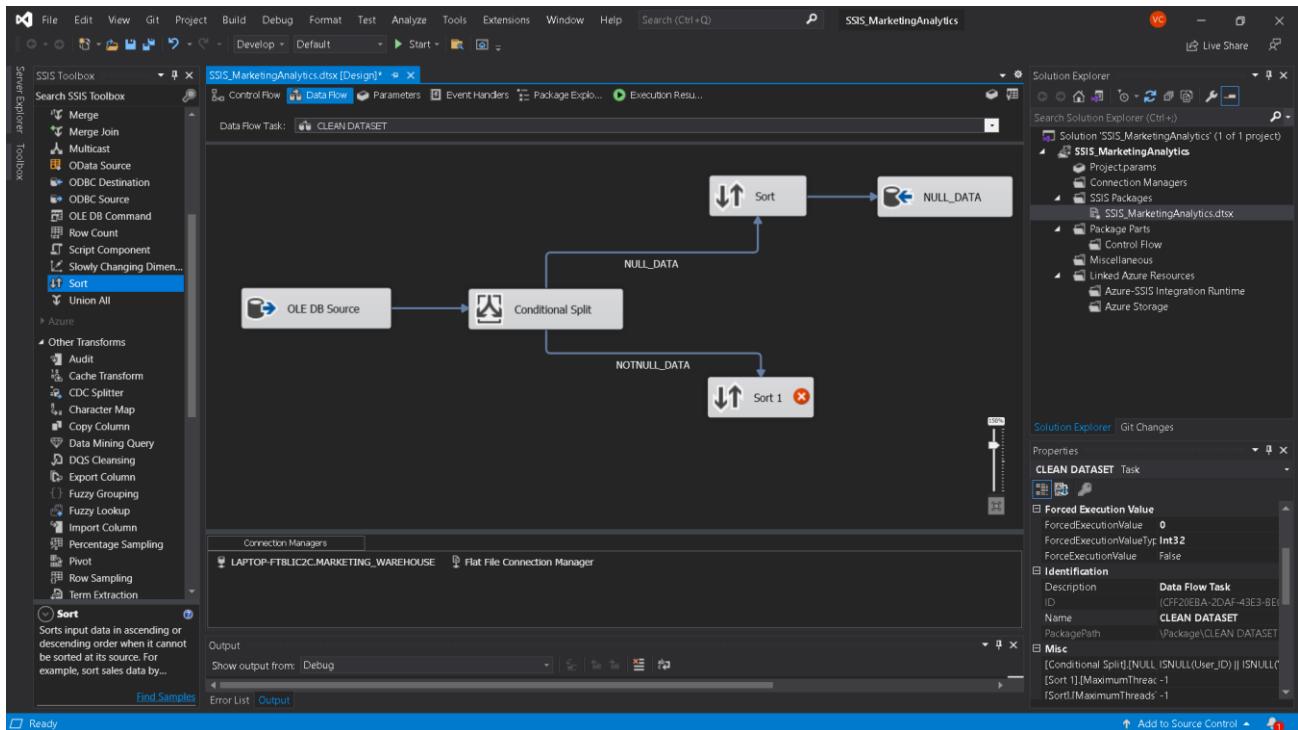
Hình 33. Tạo bảng NULL_DATA để đỗ dữ liệu NULL vào

Chọn Mappings -> Kiểm tra columns đã được map đúng hay chưa



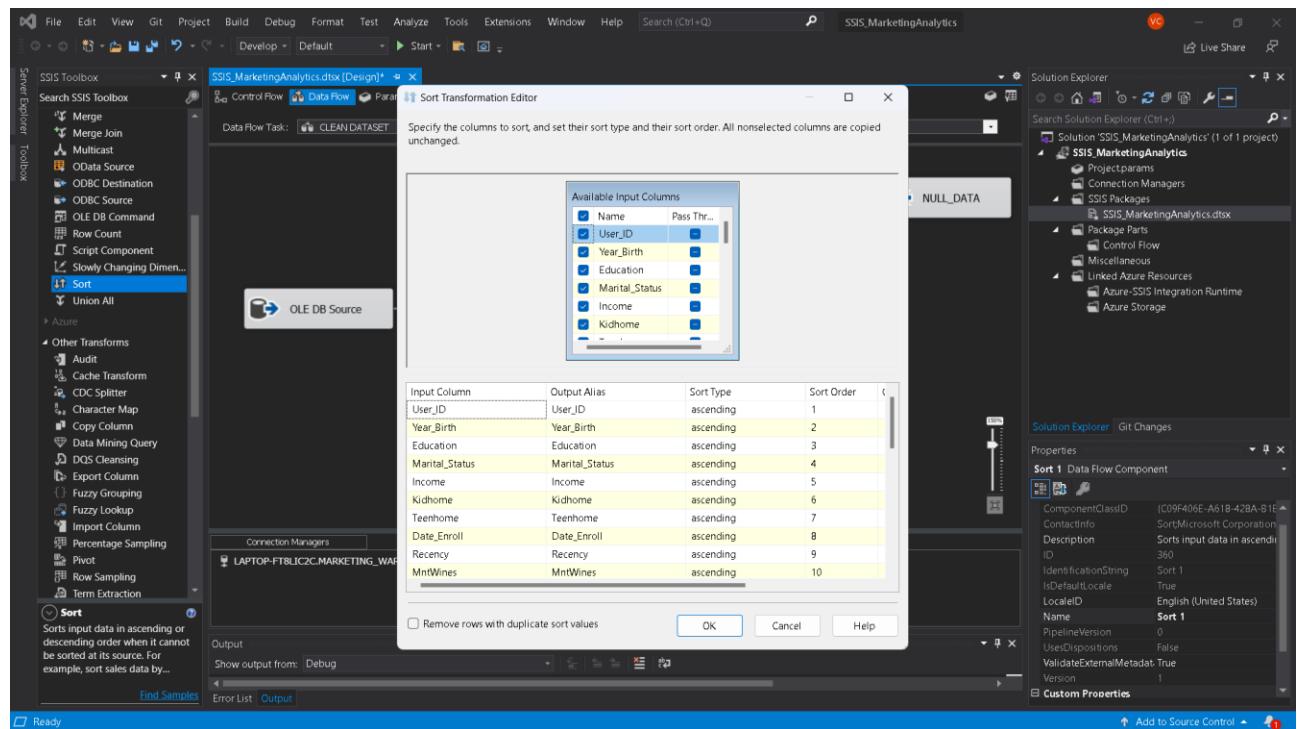
Hình 34. Kiểm tra Mappings

Sử dụng Sort để sắp xếp lại các dòng dữ liệu



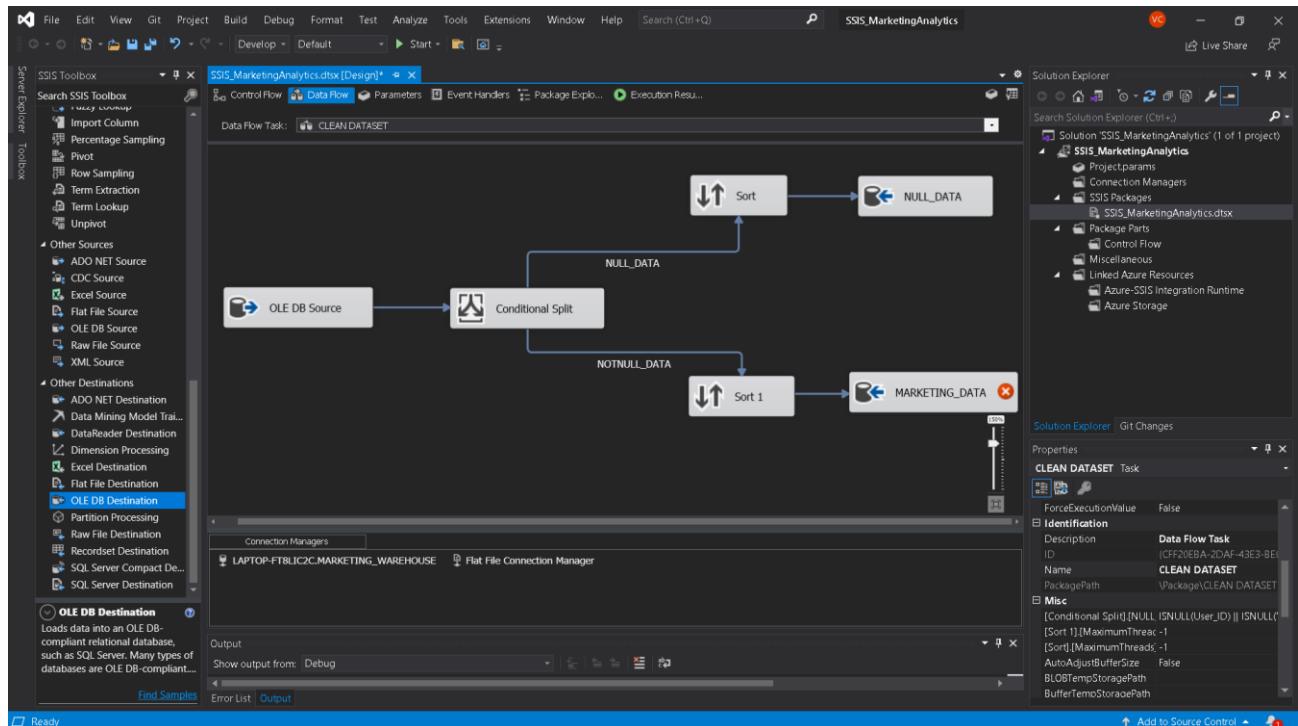
Hình 35. Sort sắp xếp dữ liệu

Edit Sort -> Chọn muôn sắp xếp theo thuộc tính nào



Hình 36. Sắp xếp dữ liệu theo thuộc tính

Sử dụng OLE DB Destination để đổ dữ liệu đã Clean vào bảng -> Đổi tên thành MARKETING_DATA

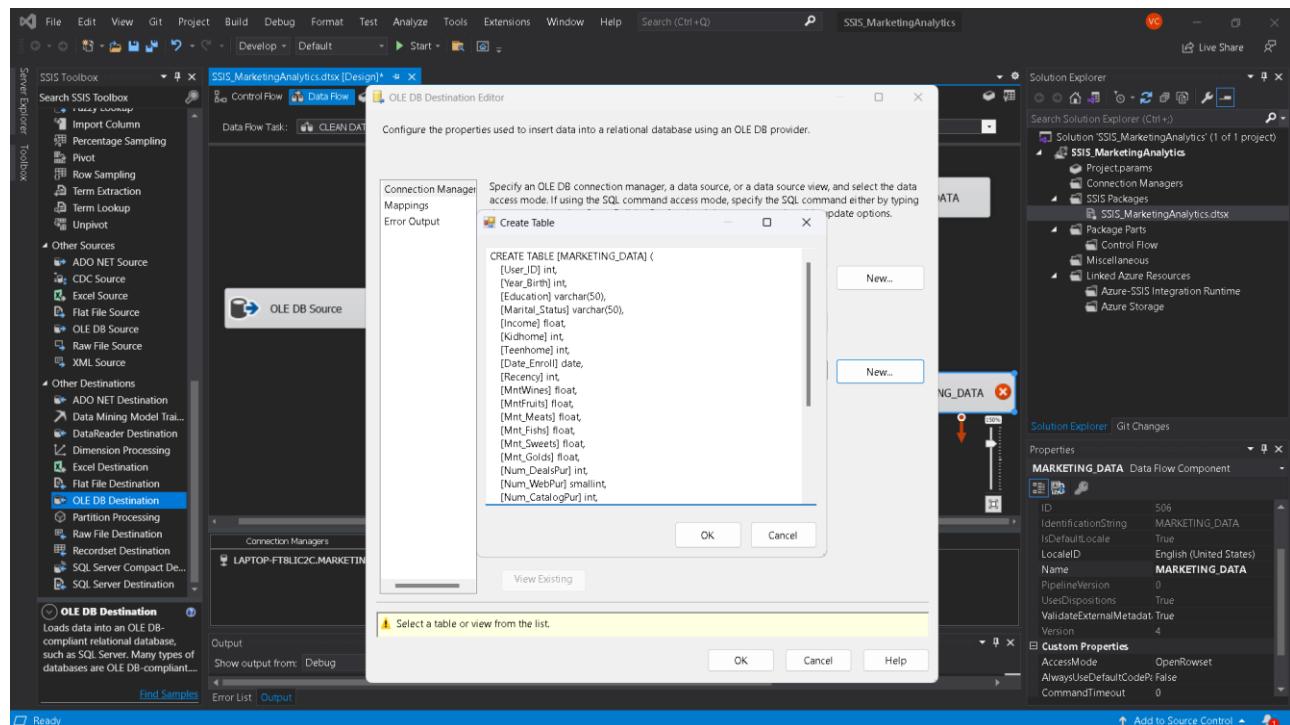


Hình 37. Đỗ dữ liệu đã Clean vào bảng MARKETING_DATA

Edit MARKETING_DATA

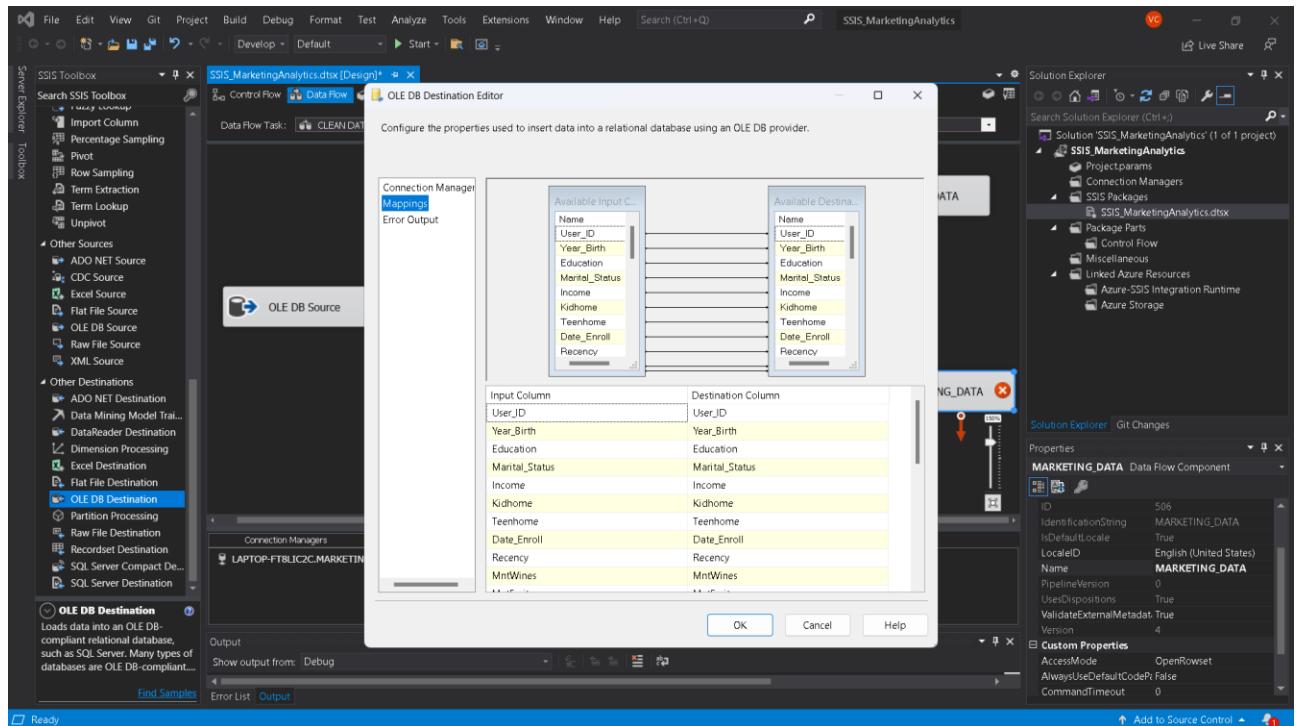
Chọn connection đến MARKETING_WAREHOUSE

Chọn New để SSIS tự động sinh ra đoạn mã insert data



Hình 38. Tạo bảng MARKETING_DATA để đỗ dữ liệu đã clean vào

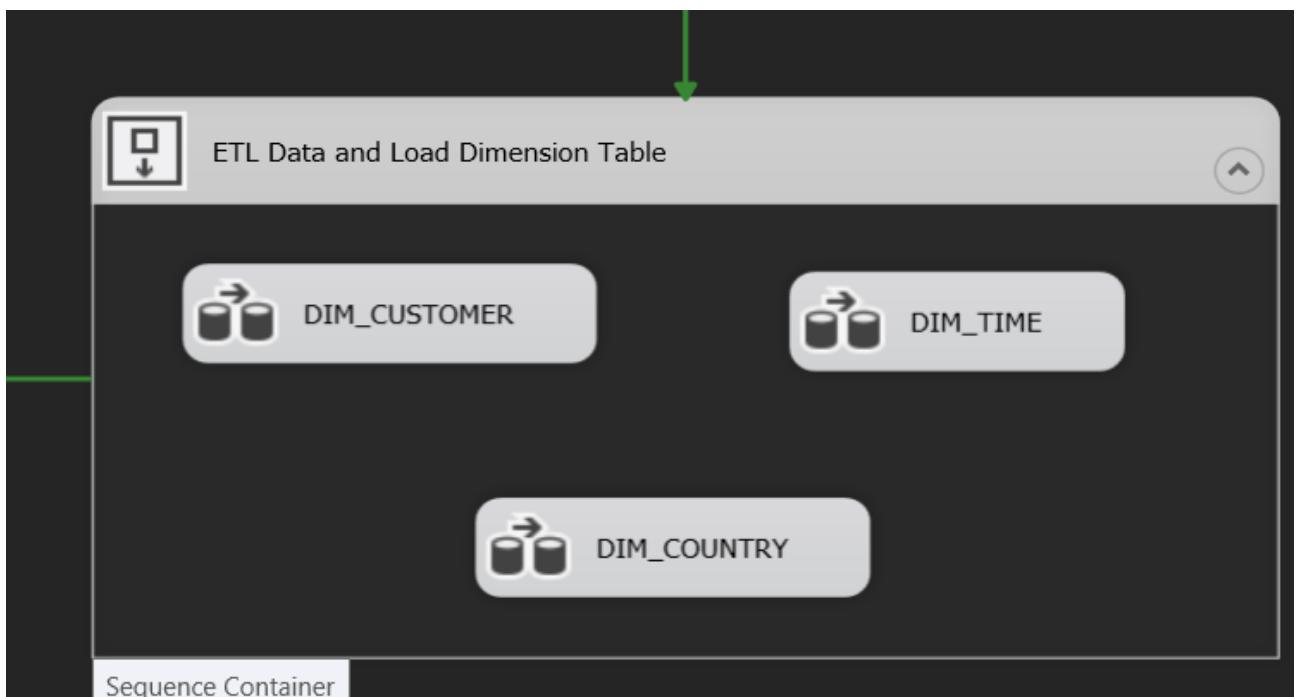
Chọn Mappings -> Kiểm tra columns đã được map đúng hay chưa



Hình 39. Kiểm tra Mappings

2.3.2 Định nghĩa và nạp dữ liệu vào các bảng Dimension

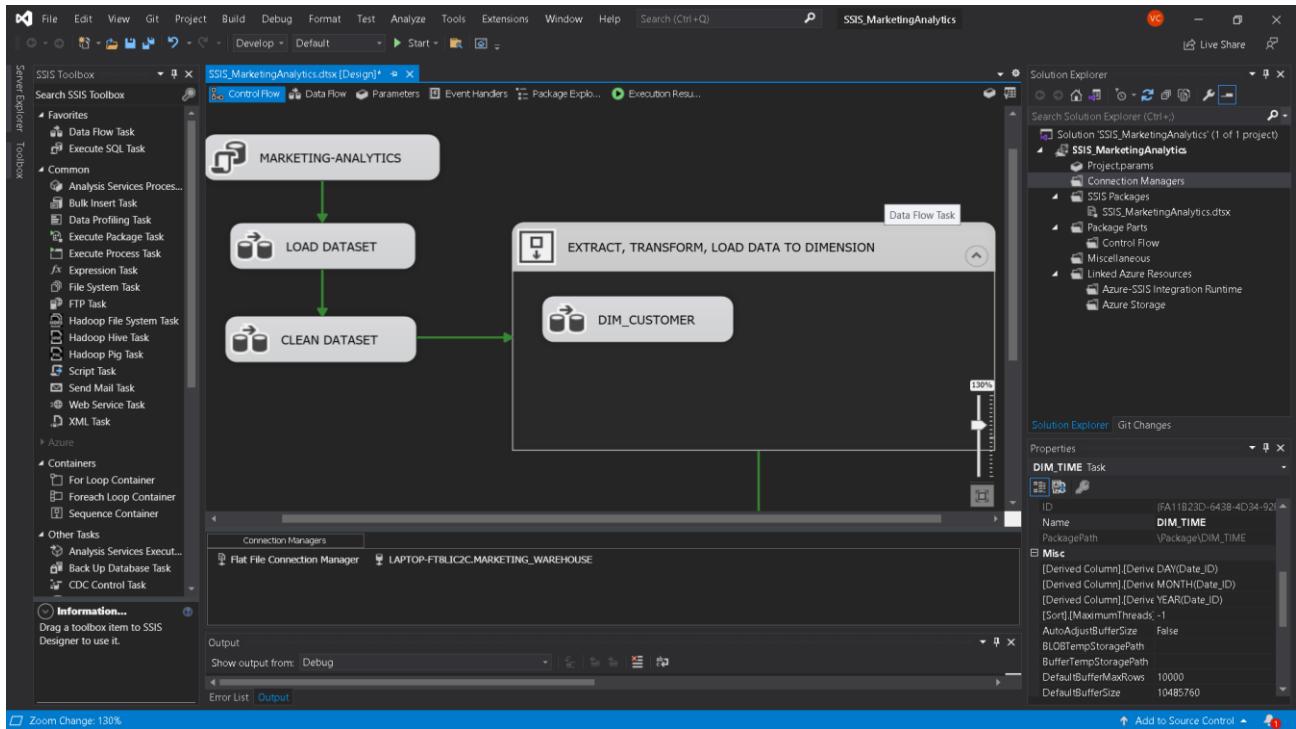
- ❖ **Mục tiêu:** Nạp dữ liệu đã được tiền xử lý nằm trong bảng MARKETING_DATA vào các bảng Dimension tương ứng theo đúng thiết kế lược đồ kho dữ liệu



Hình 40. ETL Data and Load Dimension Tables

Kéo thả để tạo mới **Sequence Container** từ thanh công cụ SSIS Toolbox -> Đổi tên thành “**EXTRACT, TRANSFORM, LOAD DATA TO DIMENTSION**”

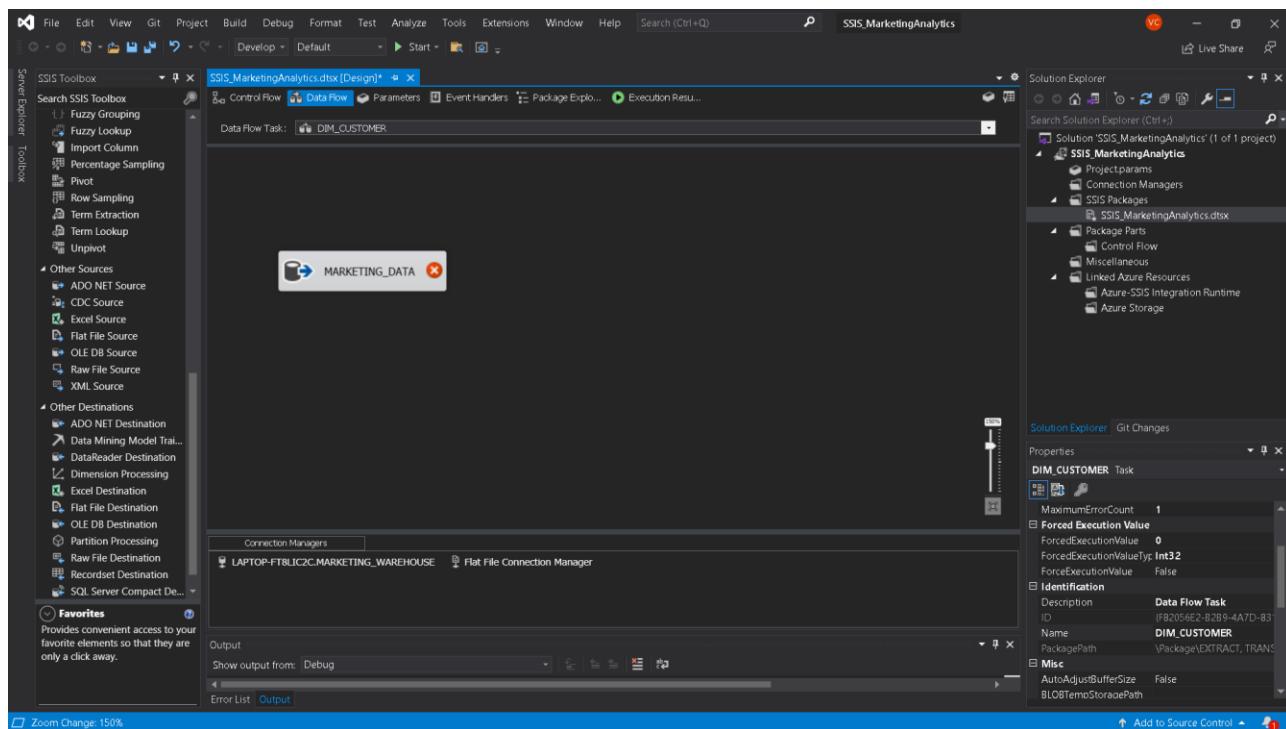
Bước 1: Sử dụng **Data Flow Task** vào Container -> Đổi tên thành **DIM_CUSTOMER**



Hình 41. Tạo Data Flow Task **DIM CUSTOMER**

Bước 2: Edit DIM_CUSTOMER

Sử dụng **OLE DB Source** để định nghĩa dữ liệu nguồn -> Đổi tên thành **MARKETING_DATA**

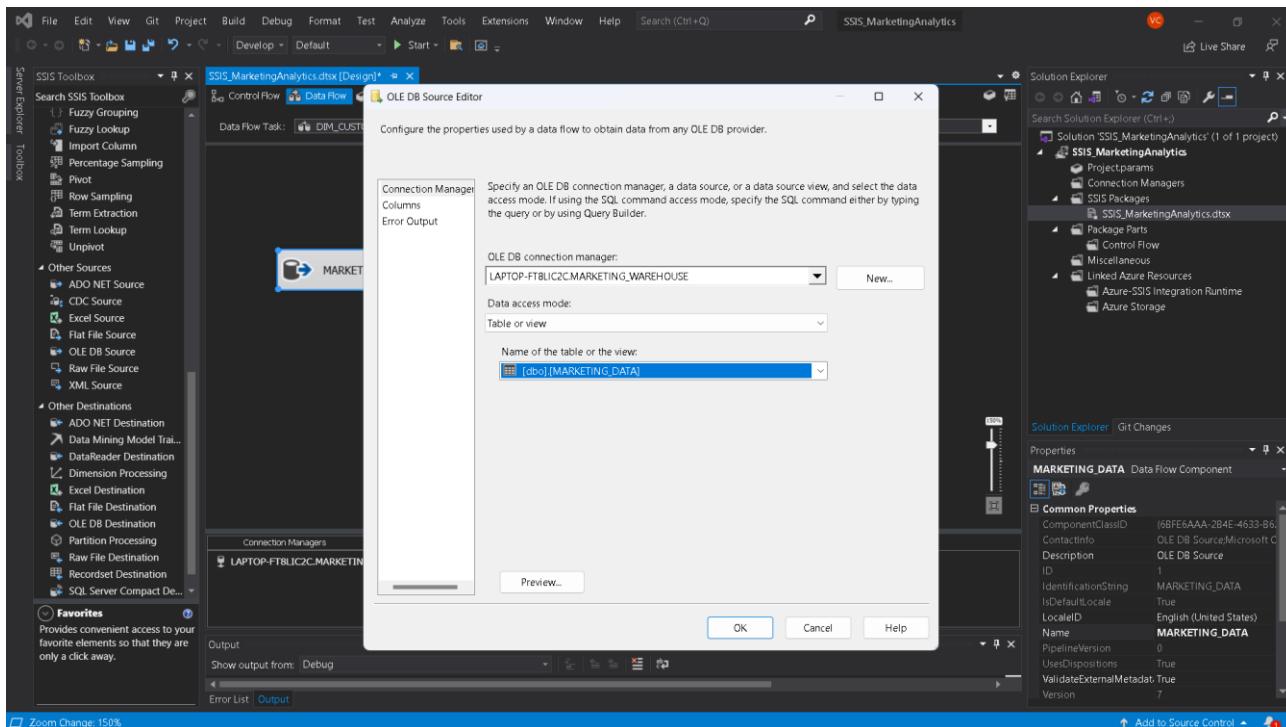


Hình 42. OLE DB Source định nghĩa dữ liệu nguồn

Edit MARKETING_DATA

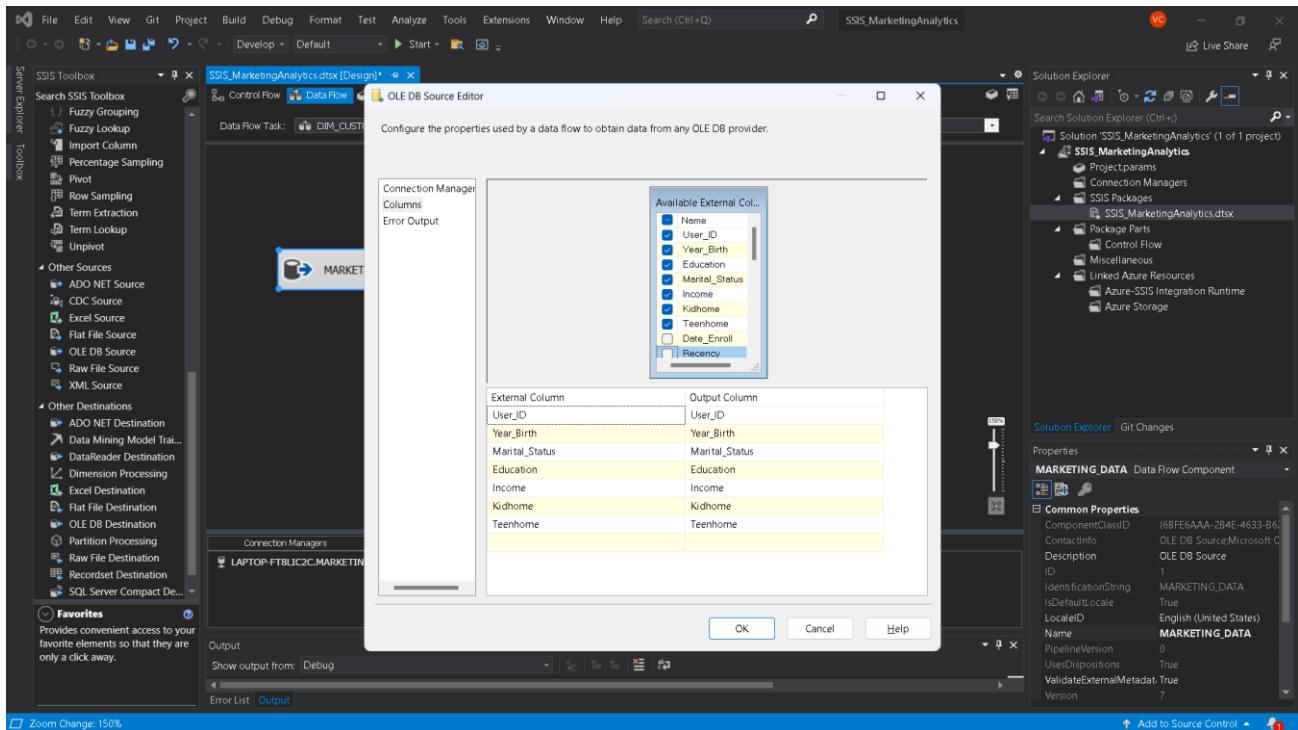
Chọn connection đến MARKETING_WAREHOUSE

Chọn table MARKETING_DATA



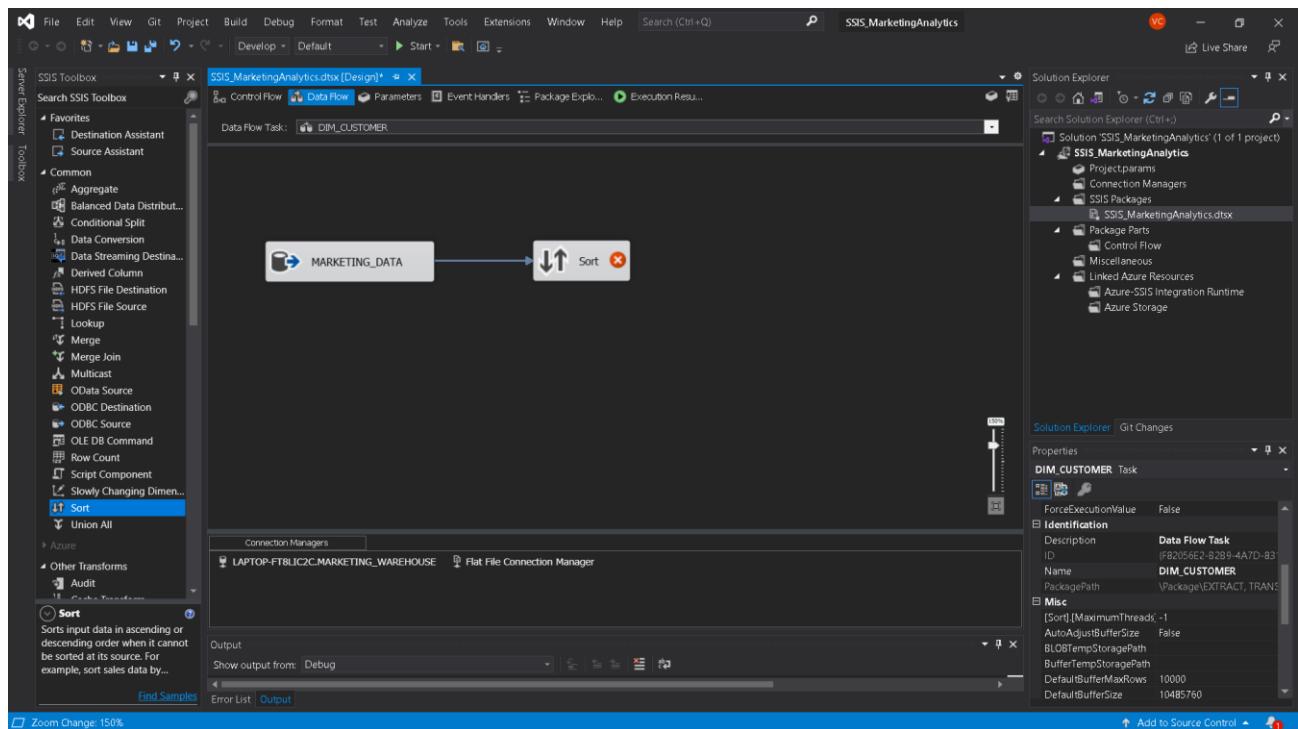
Hình 43. Định nghĩa dữ liệu nguồn là MARKETING_DATA

Chọn Columns -> Chọn những thuộc tính cần thiết cho DIM_CUSTOMER



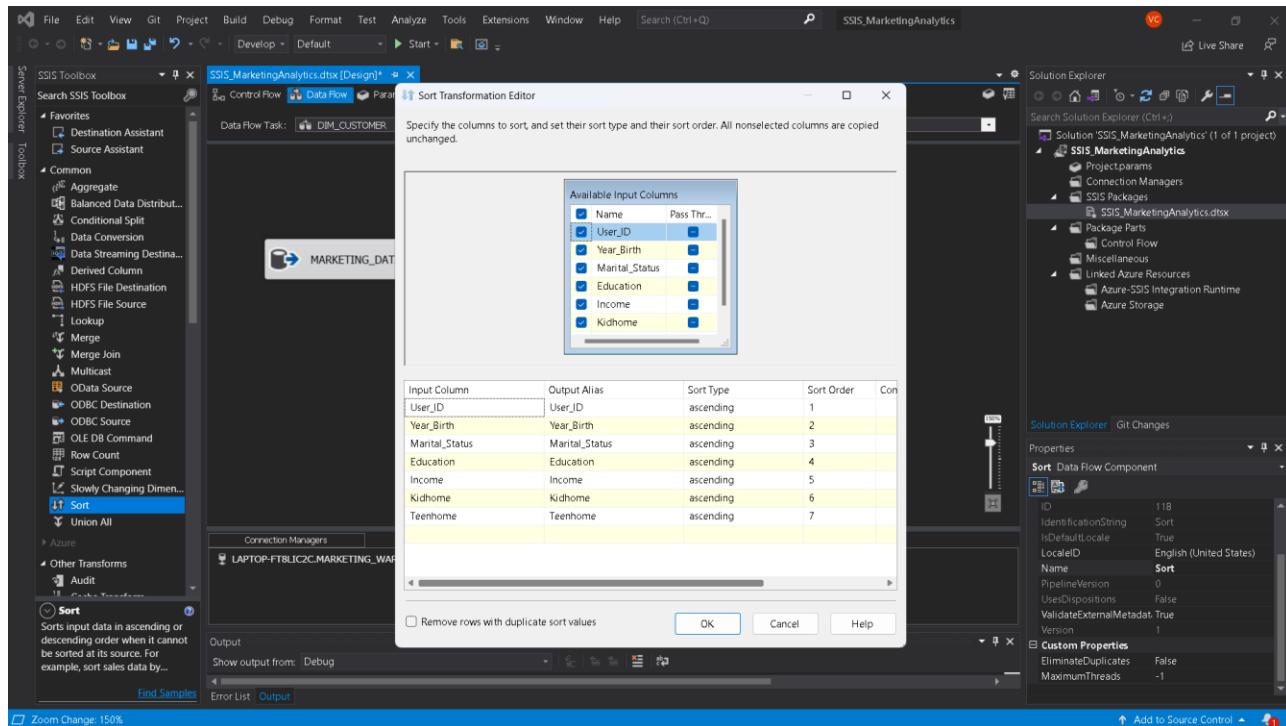
Hình 44. Chọn thuộc tính cho DIM_CUSTOMER

Sử dụng Sort để sắp xếp lại các dòng dữ liệu



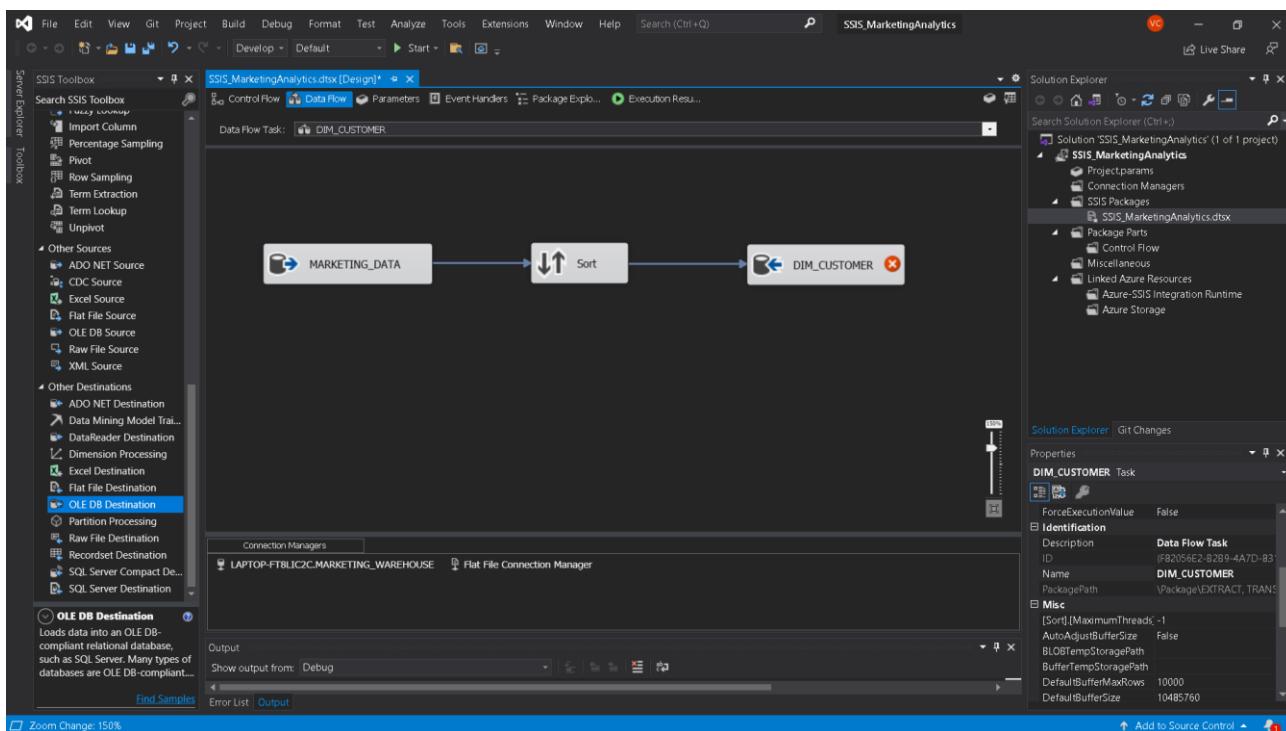
Hình 45. Sort sắp xếp dữ liệu

Edit Sort -> Chọn muốn sắp xếp theo thuộc tính nào



Hình 46. Sắp xếp dữ liệu theo thuộc tính

Sử dụng **OLE DB Destination** để đổ dữ liệu vào bảng -> Đổi tên thành
DIM_CUSTOMER

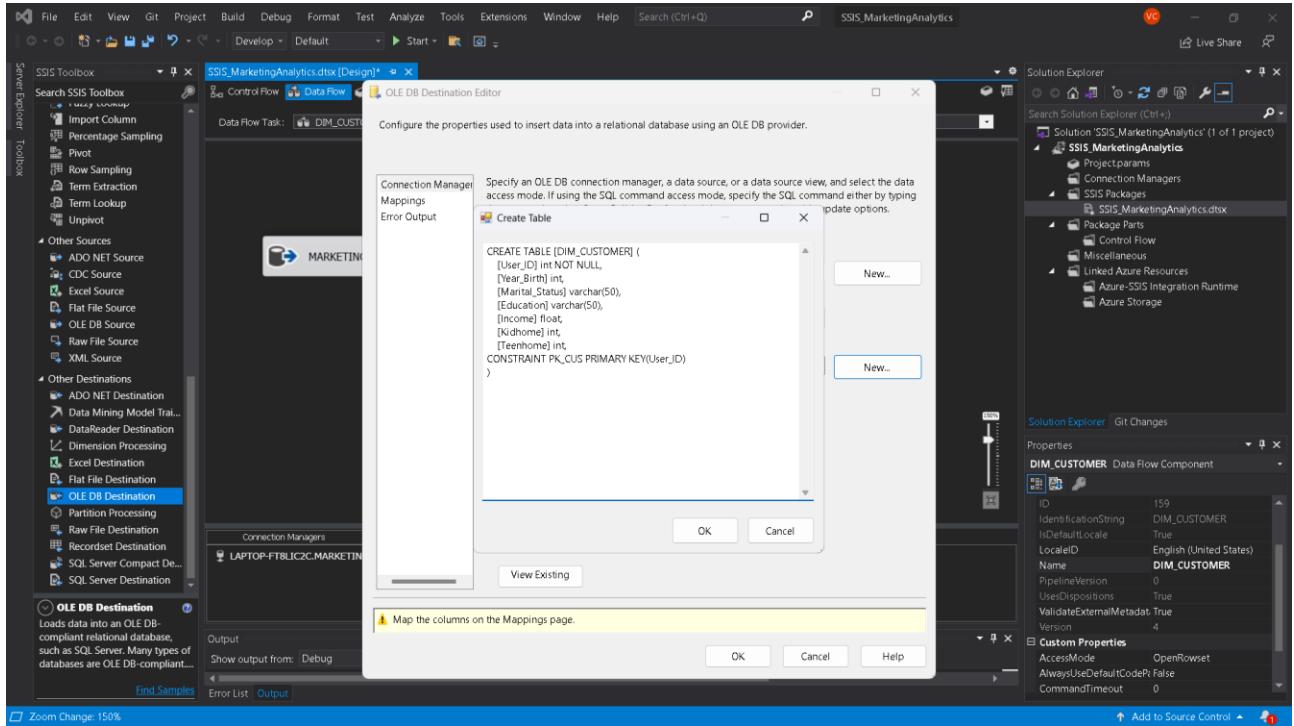


Hình 47. Đổ dữ liệu vào bảng DIM_CUSTOMER

Edit **DIM_CUSTOMER**

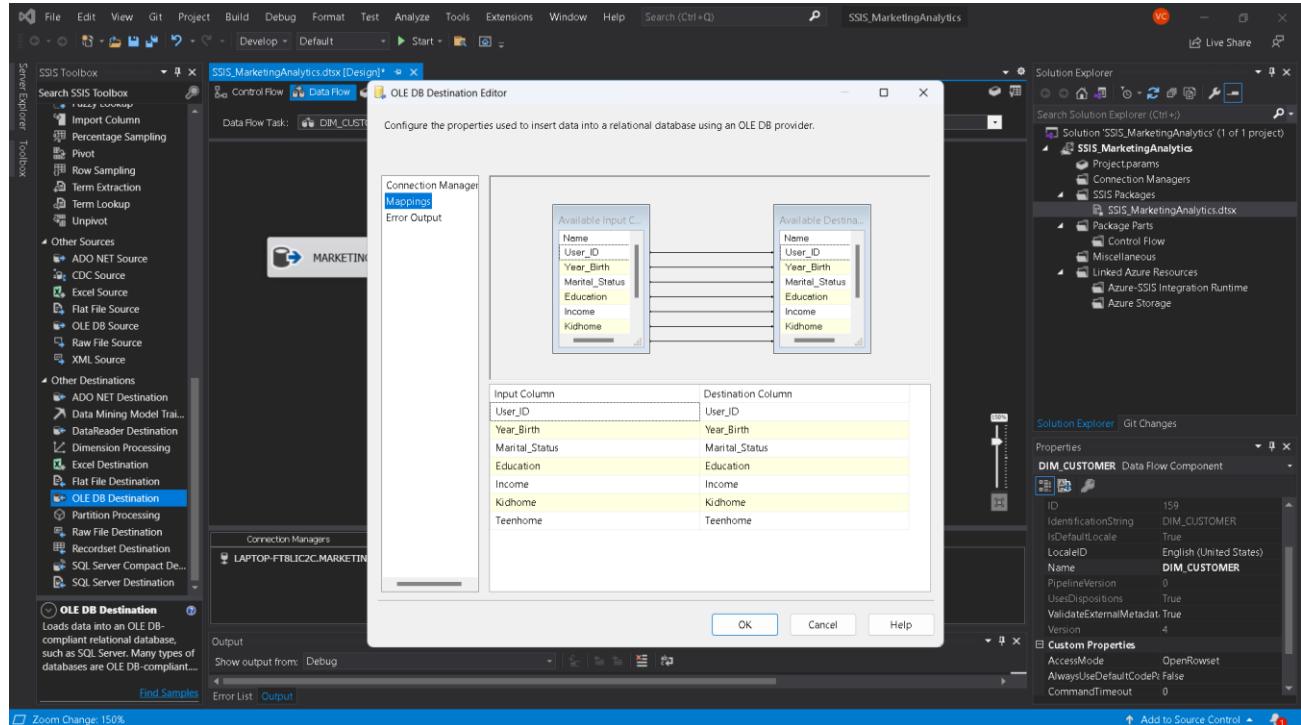
Chọn **Connection** đến **MARKETING_WAREHOUSE**

Chọn New để SSIS tự sinh ra đoạn mã insert data



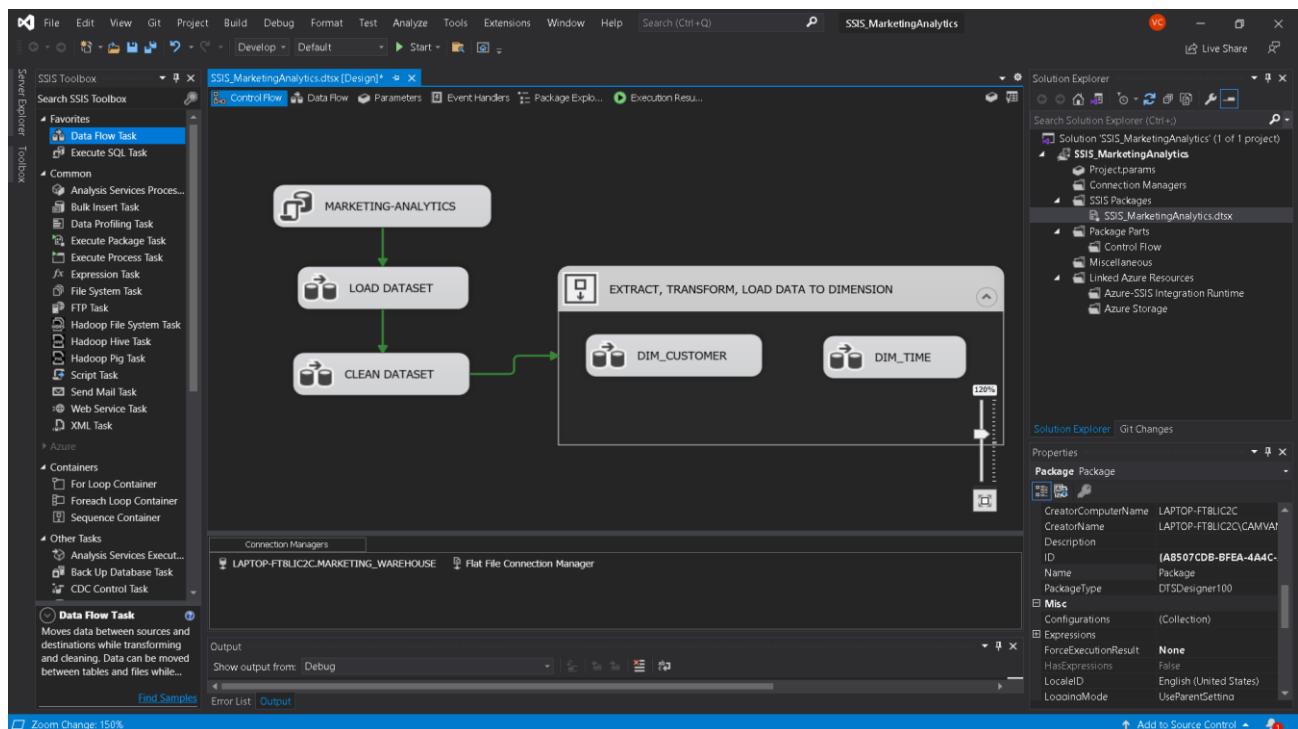
Hình 48. Tạo bảng DIM_CUSTOMER để đỗ dữ liệu vào

Chọn -> Mappings Kiểm tra xem columns đã được map đúng hay chưa



Hình 49. Kiểm tra Mappings

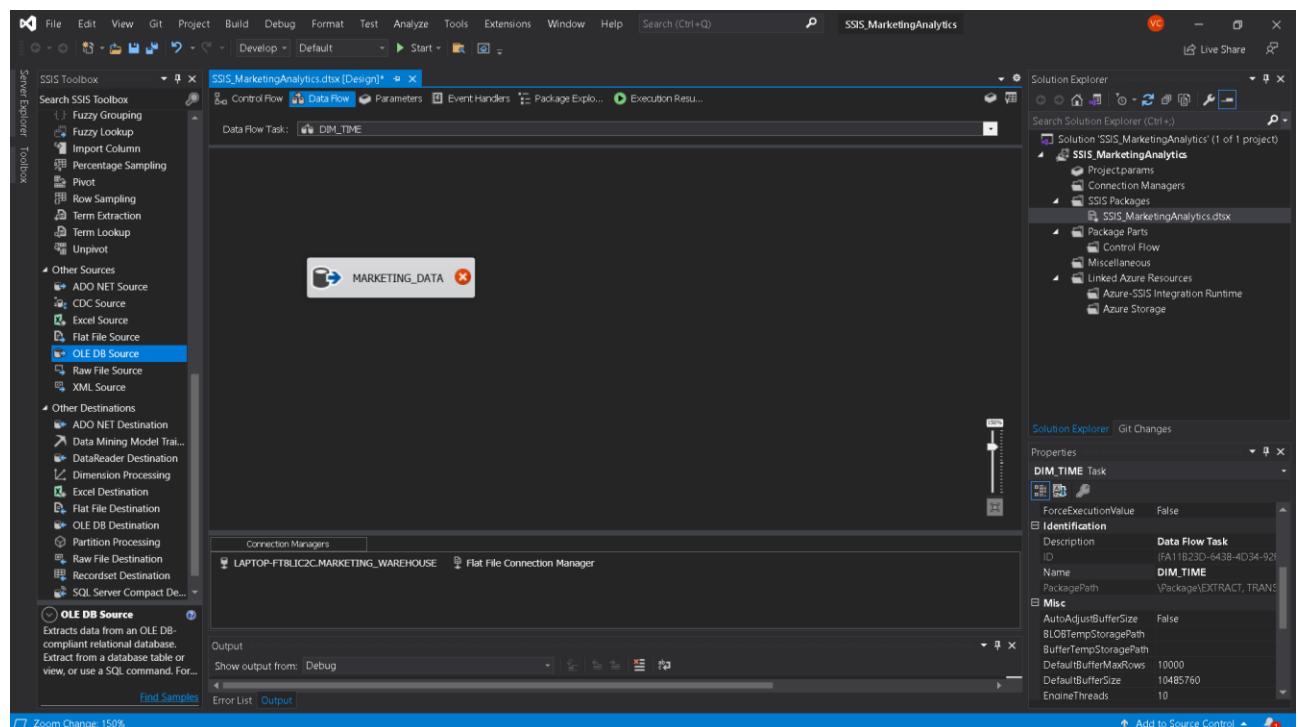
Bước 3: Sử dụng Data Flow Task vào Container -> Đổi tên thành DIM_TIME



Hình 50. Tạo Data Flow Task **DIM_TIME**

Bước 4: Edit **DIM_TIME**

Sử dụng **OLE DB Source** để định nghĩa dữ liệu nguồn -> Đổi tên thành **MARKETING_DATA**

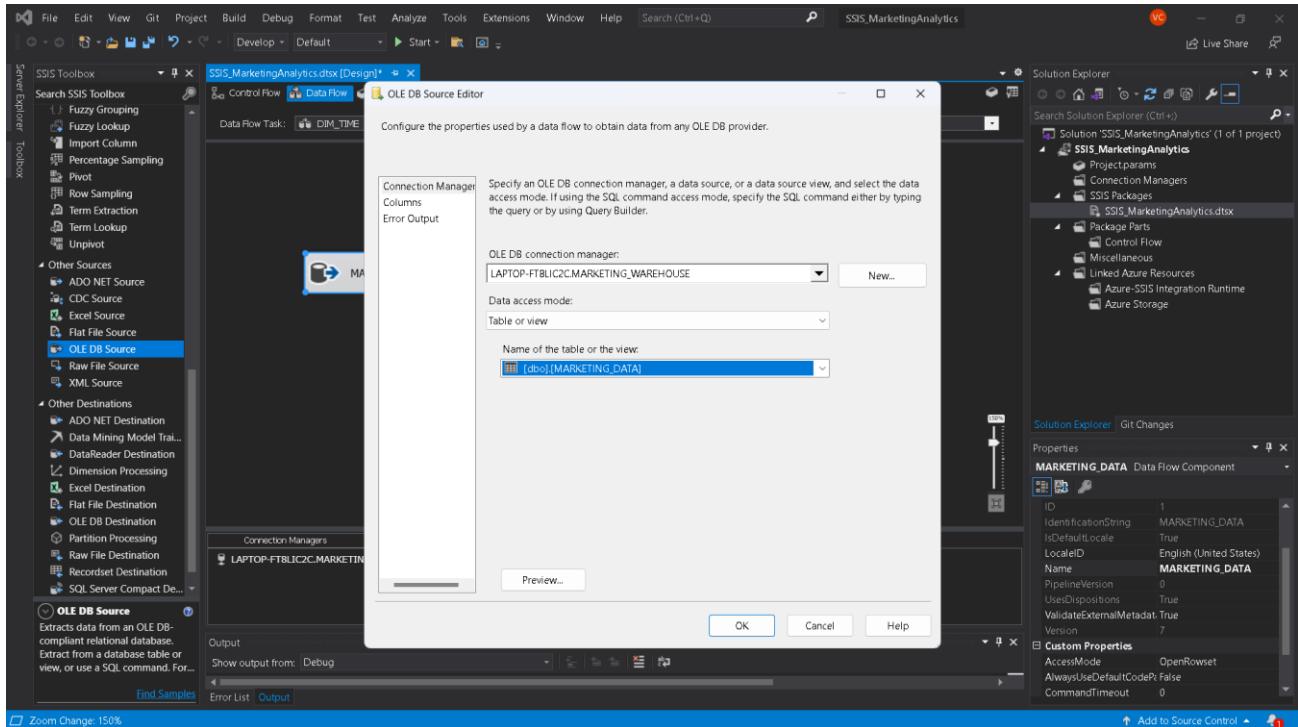


Hình 51. OLE DB Source định nghĩa dữ liệu nguồn

Edit **MARKETING_DATA**

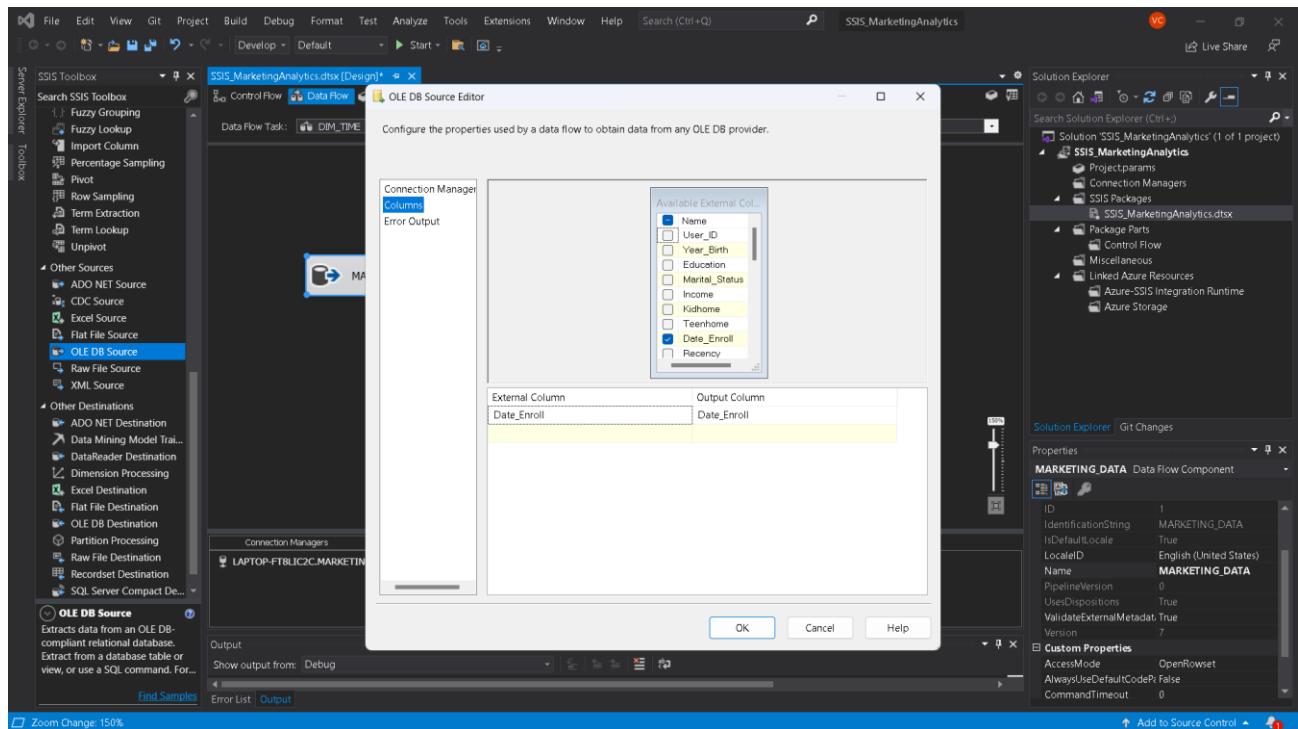
Chọn connection đến MARKETING_WAREHOUSE

Chọn table MARKETING_DATA



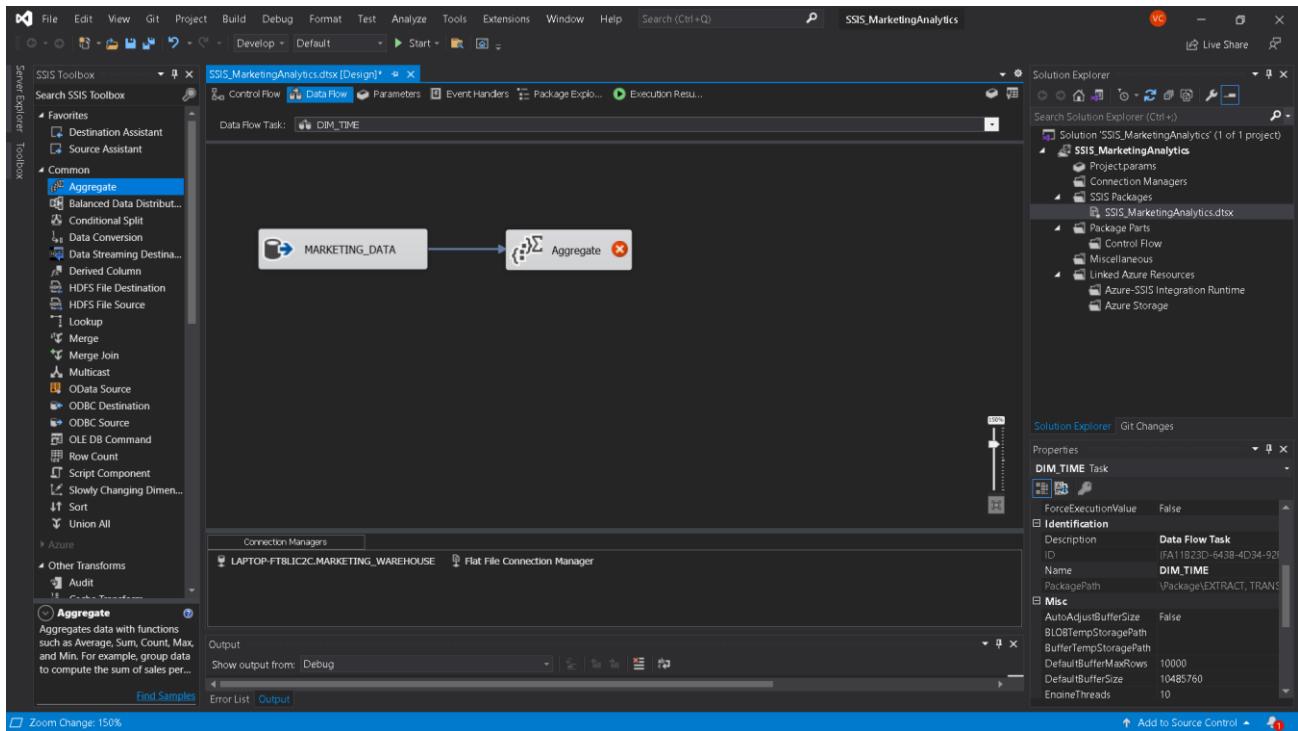
Hình 52. Định nghĩa dữ liệu nguồn là MARKETING_DATA

Chọn Columns -> Chọn những thuộc tính cần thiết cho “DIM_TIME”



Hình 53. Chọn thuộc tính cho DIM_TIME

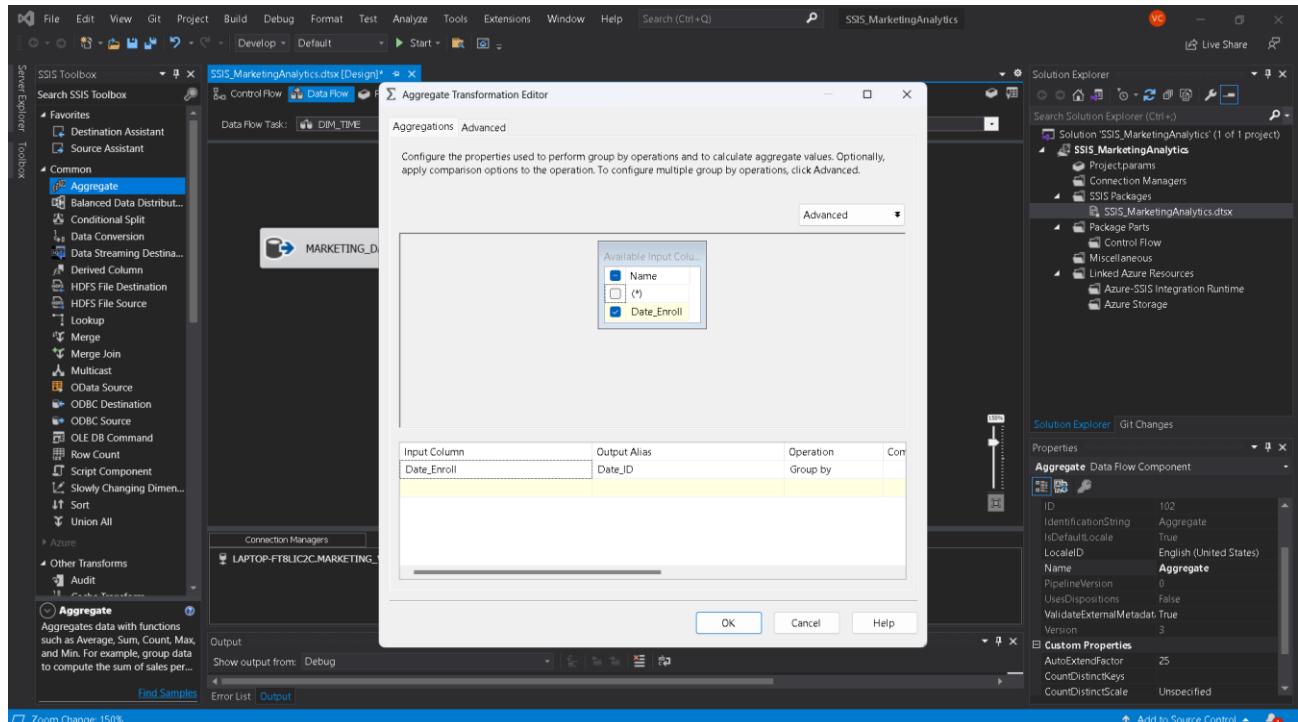
Sử dụng Aggregate để Group by dữ liệu, xóa những dữ liệu bị trùng



Hình 54. Agggregate Group By dữ liệu

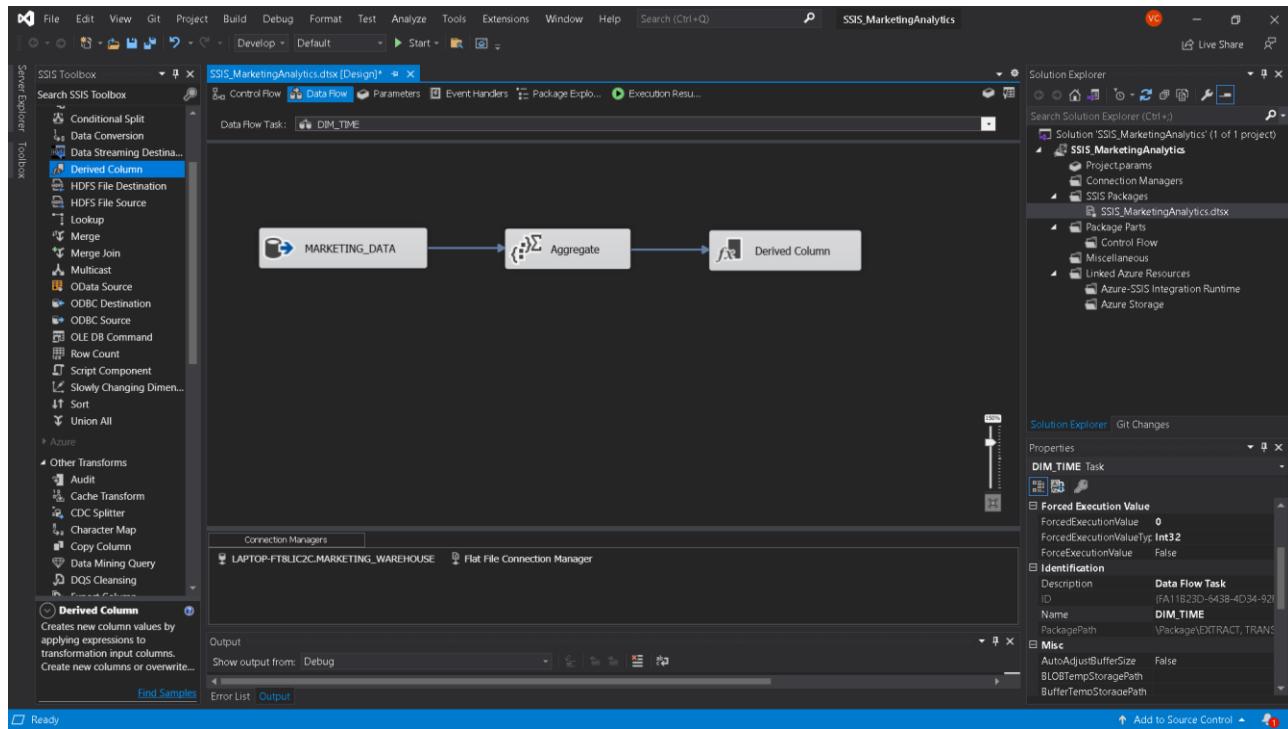
Edit Agggregate

Chọn GrouBy theo cột Date_Enroll -> Đổi tên Output Alias thành Date_ID



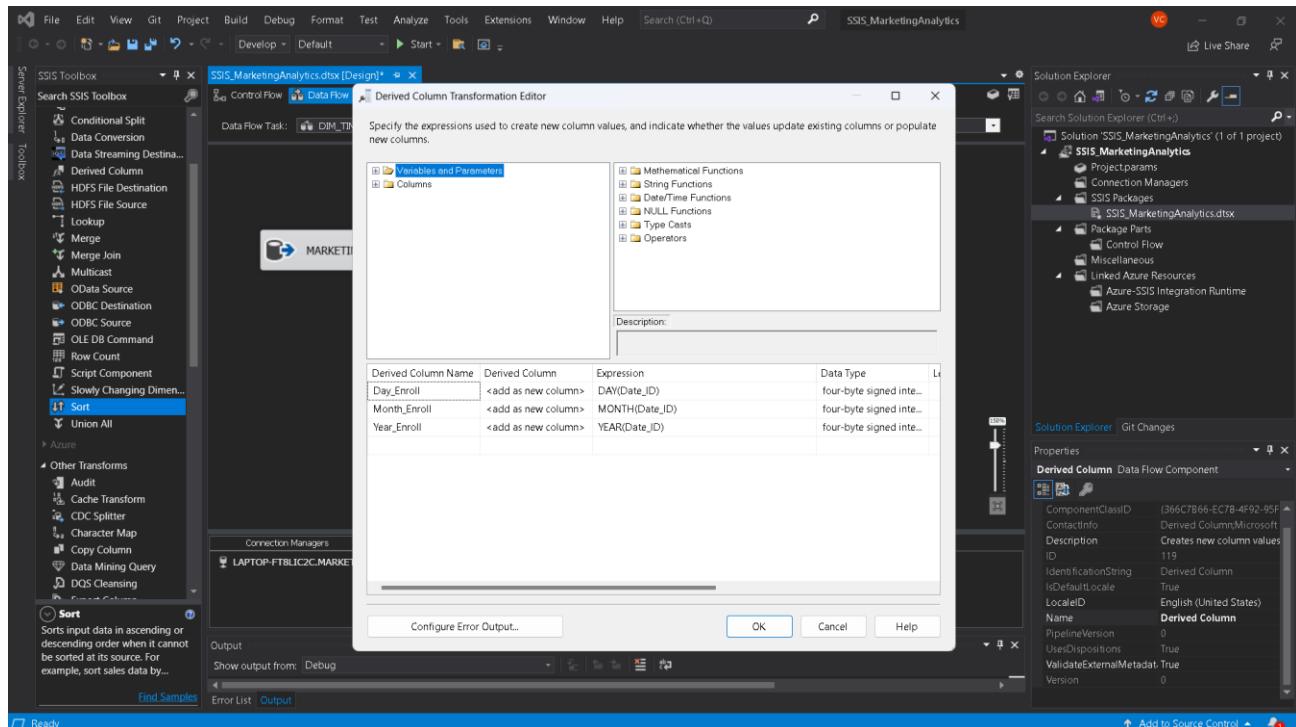
Hình 55. Group By dữ liệu thành Date_ID

Sử dụng Derived Column để tách dữ liệu Date_Enroll



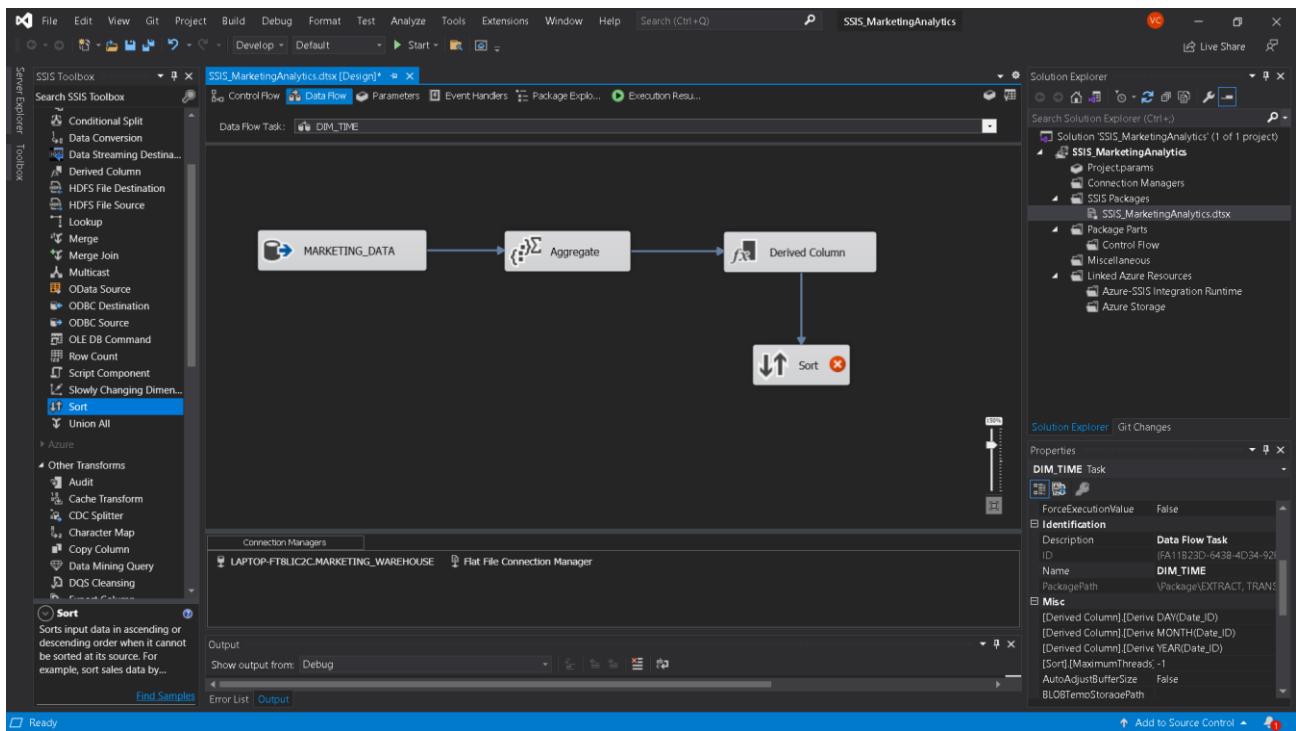
Hình 56. Derived Column để tách dữ liệu

Edit Derived Column -> Thêm các function cần thiết để split Date_Enroll



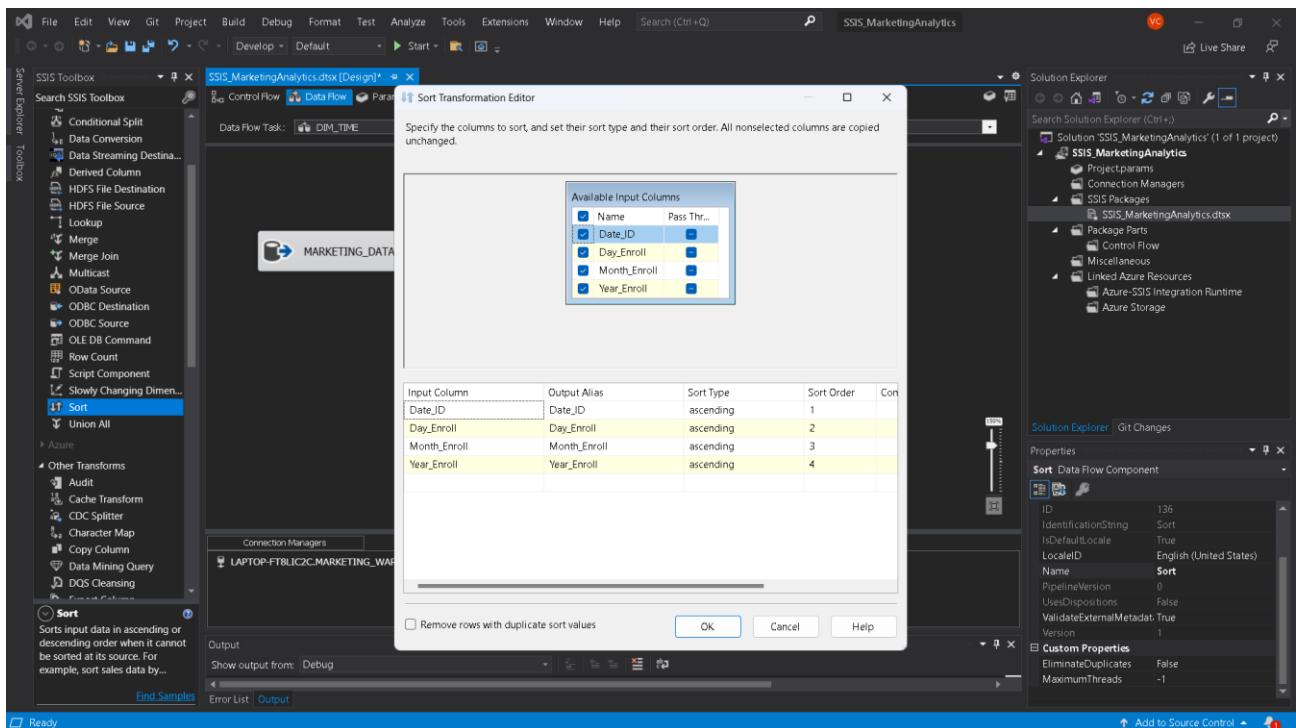
Hình 57. Tách dữ liệu Date_Enroll

Sử dụng Sort để sắp xếp lại các dòng dữ liệu



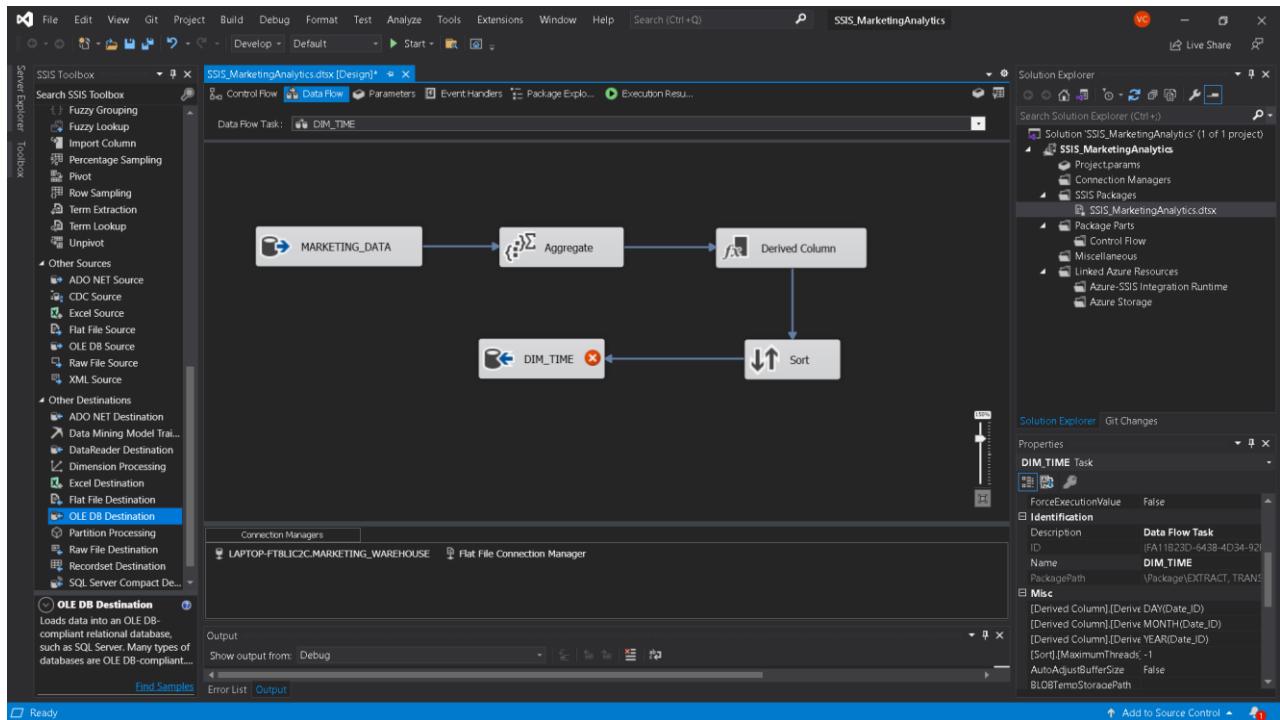
Hình 58. Sort sắp xếp dữ liệu

Edit Sort -> Chọn muôn sắp xếp theo thuộc tính nào



Hình 59. Sắp xếp dữ liệu theo thuộc tính

Sử dụng **OLE DB Destination** để đổ dữ liệu vào bảng -> Đổi tên thành **DIM_TIME**

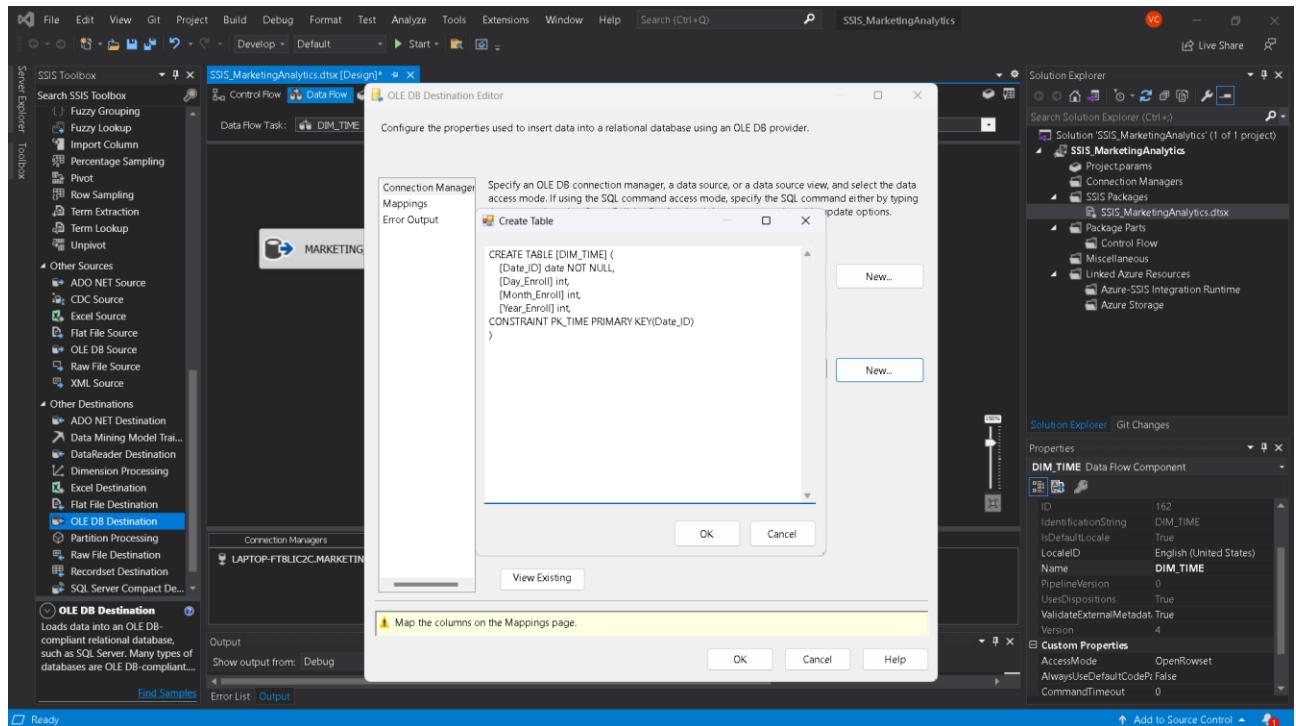


Hình 60. Đổ dữ liệu vào bảng **DIM_TIME**

Edit **DIM_TIME**

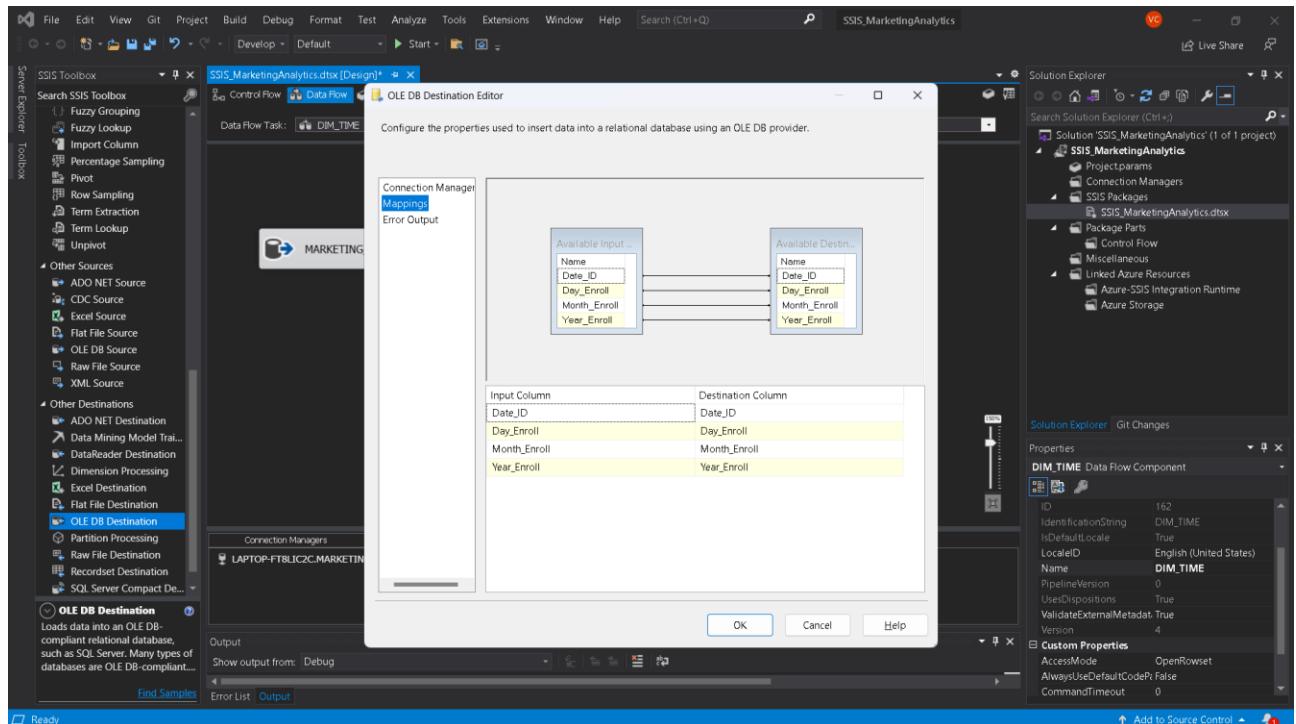
Chọn **Connection** đến **MARKETING_WAREHOUSE**

Chọn **New** để SSIS tự sinh ra đoạn mã insert data



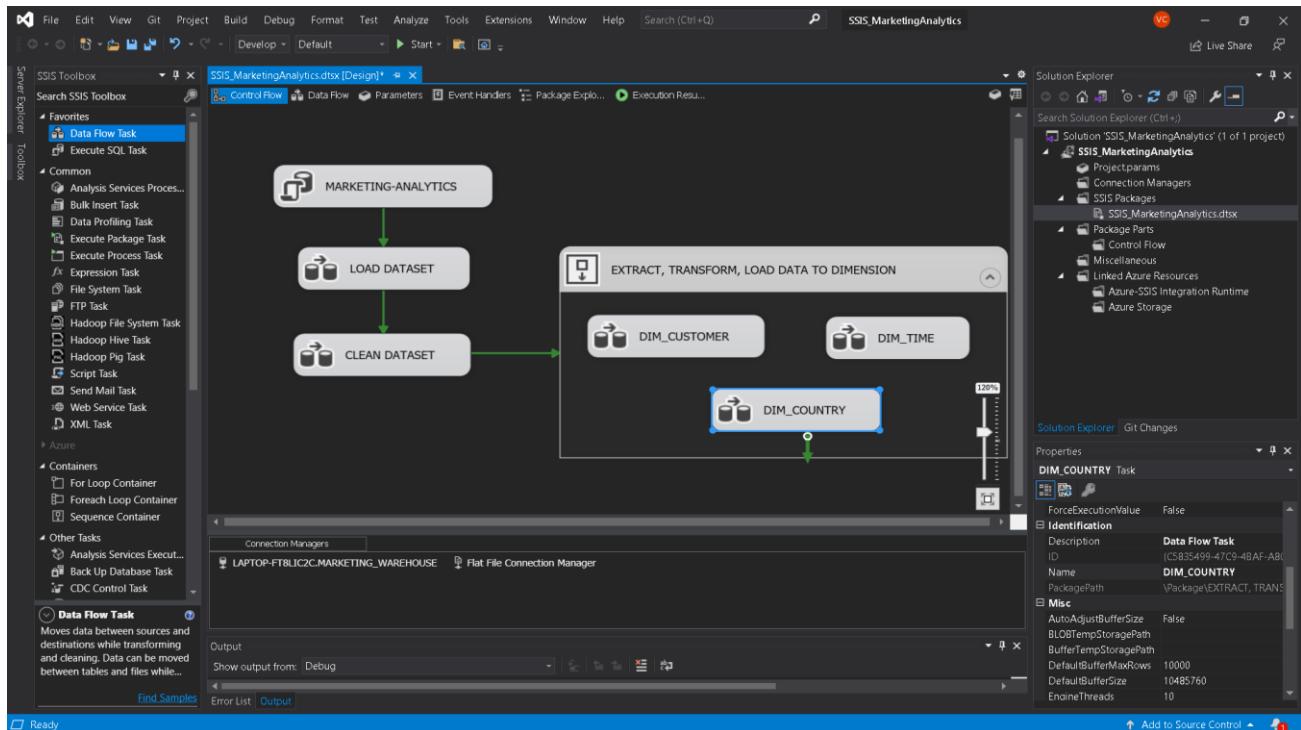
Hình 61. Tạo bảng DIM_TIME để đỗ dữ liệu vào

Chọn Mappings -> Kiểm tra columns đã được map đúng hay chưa



Hình 62. Kiểm tra Mappings

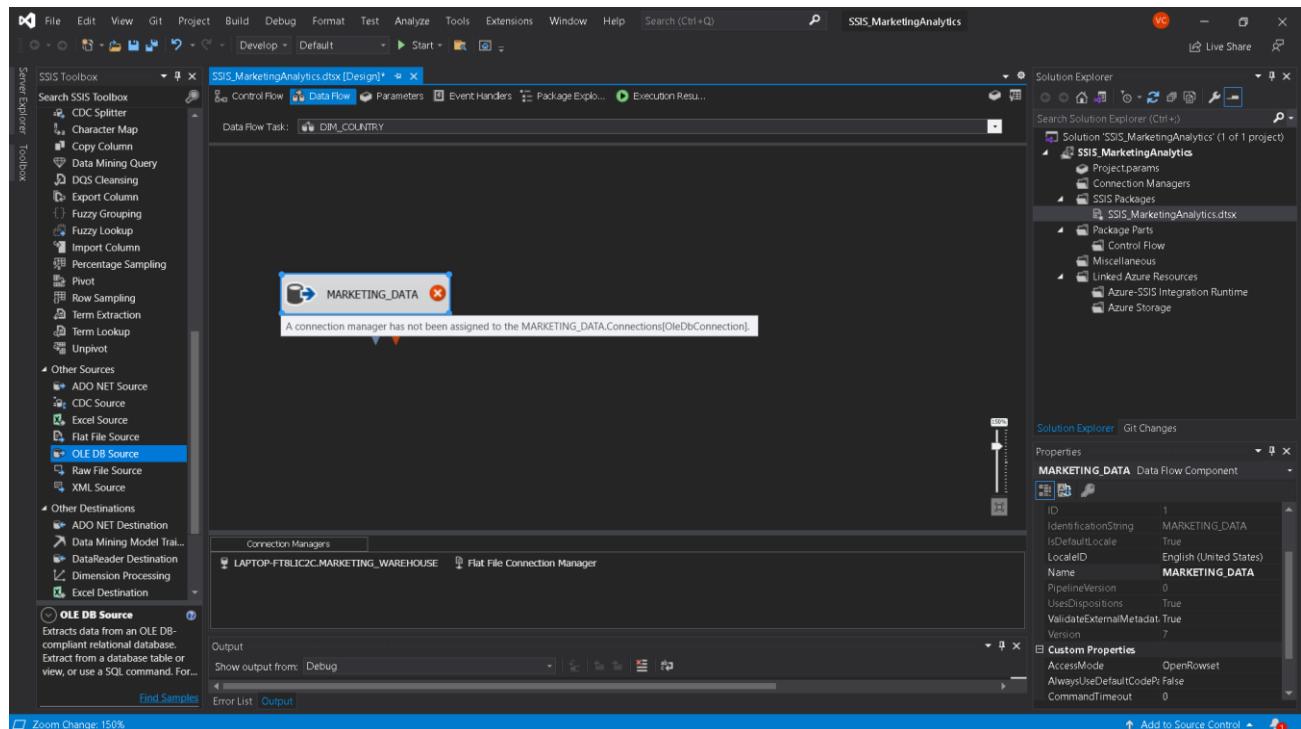
Bước 5: Sử dụng Data Flow Task vào Container -> Đổi tên thành DIM_COUNTRY



Hình 63. Tạo Data Flow Task DIM_COUNTRY

Bước 6: Edit DIM_COUNTRY

Sử dụng OLE DB Source để định nghĩa dữ liệu nguồn -> Đổi tên thành MARKETING_DATA

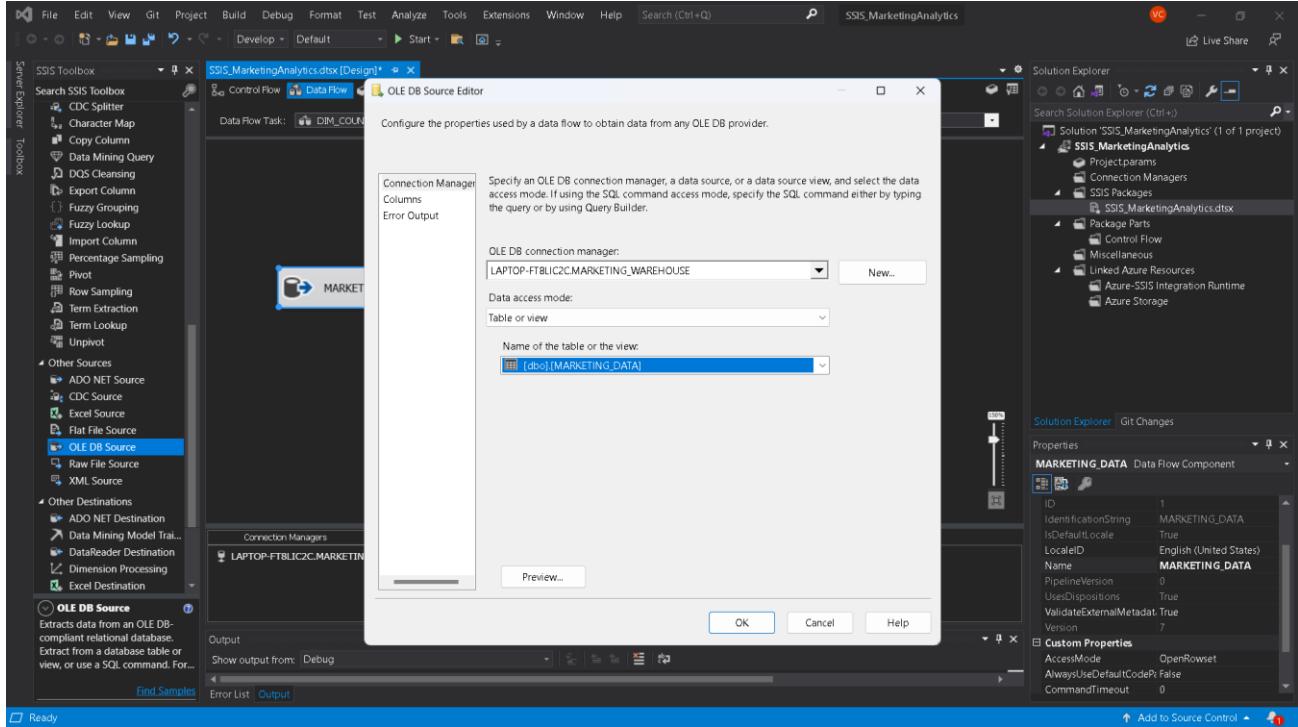


Hình 64. OLE DB Source định nghĩa dữ liệu nguồn

Edit MARKETING_DATA

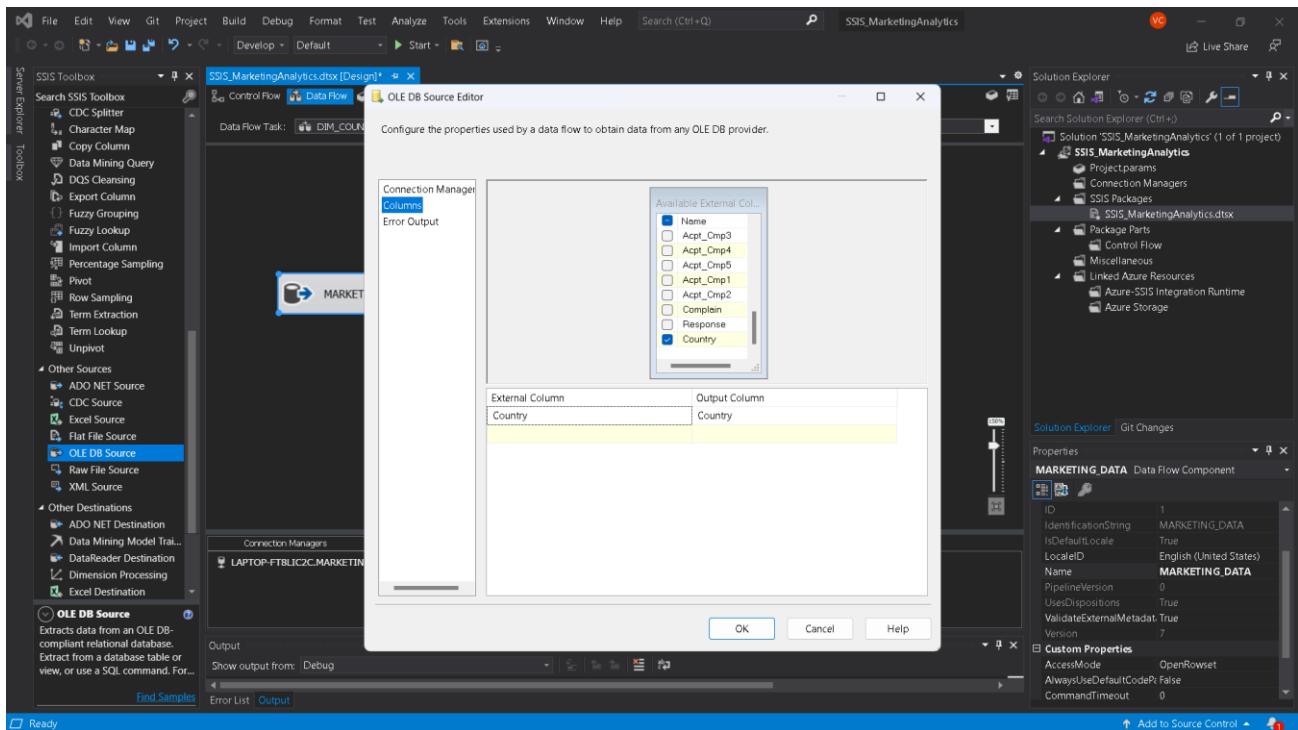
Chọn connection đến MARKETING_WAREHOUSE

Chọn table MARKETING_DATA



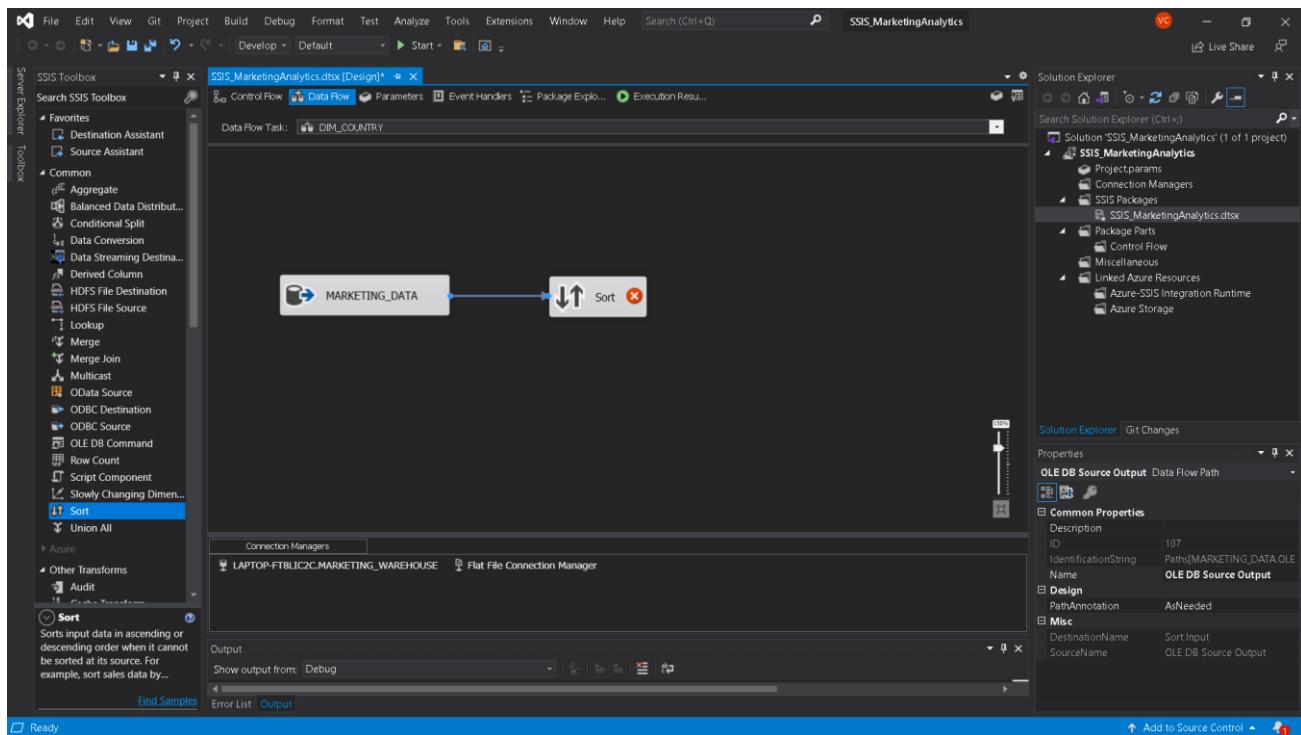
Hình 65. Định nghĩa dữ liệu nguồn là MARKETING_DATA

Chọn Columns -> Chọn những thuộc tính cần thiết cho DIM_COUNTRY



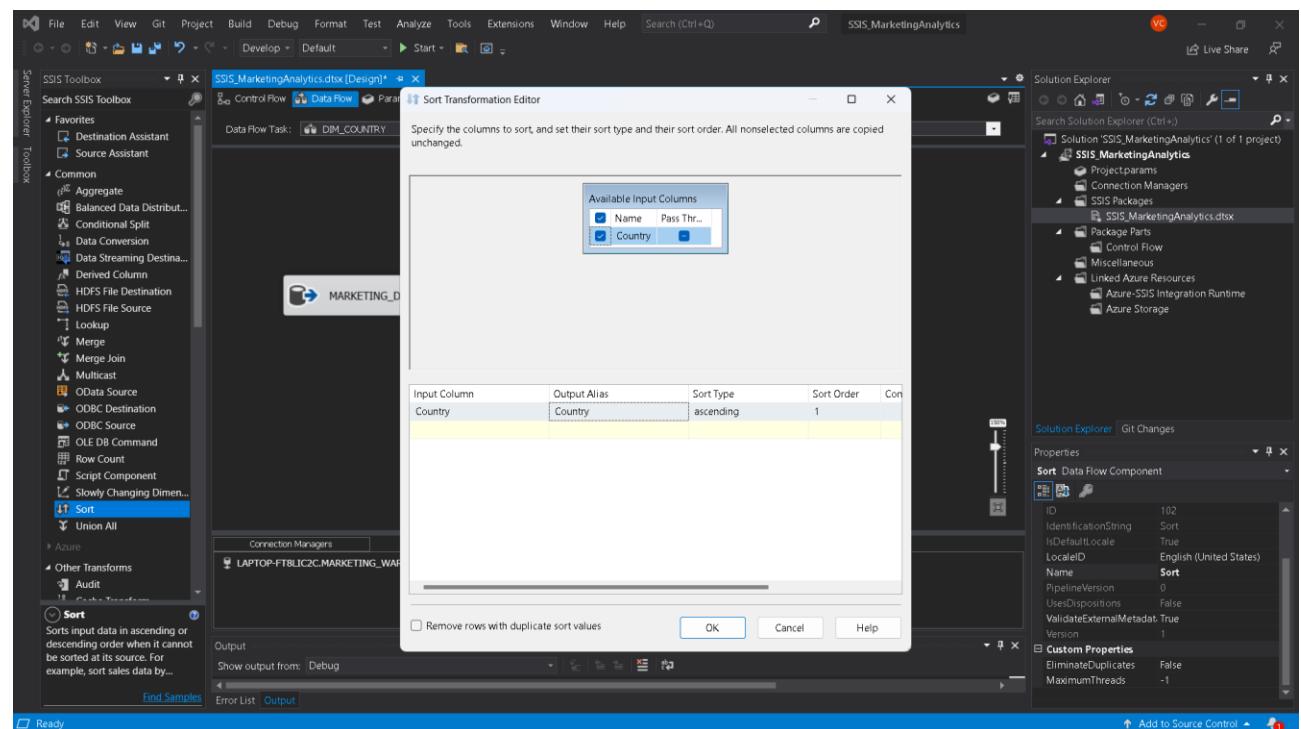
Hình 66. Chọn thuộc tính cho DIM_COUNTRY

Sử dụng Sort để sắp xếp lại các dòng dữ liệu



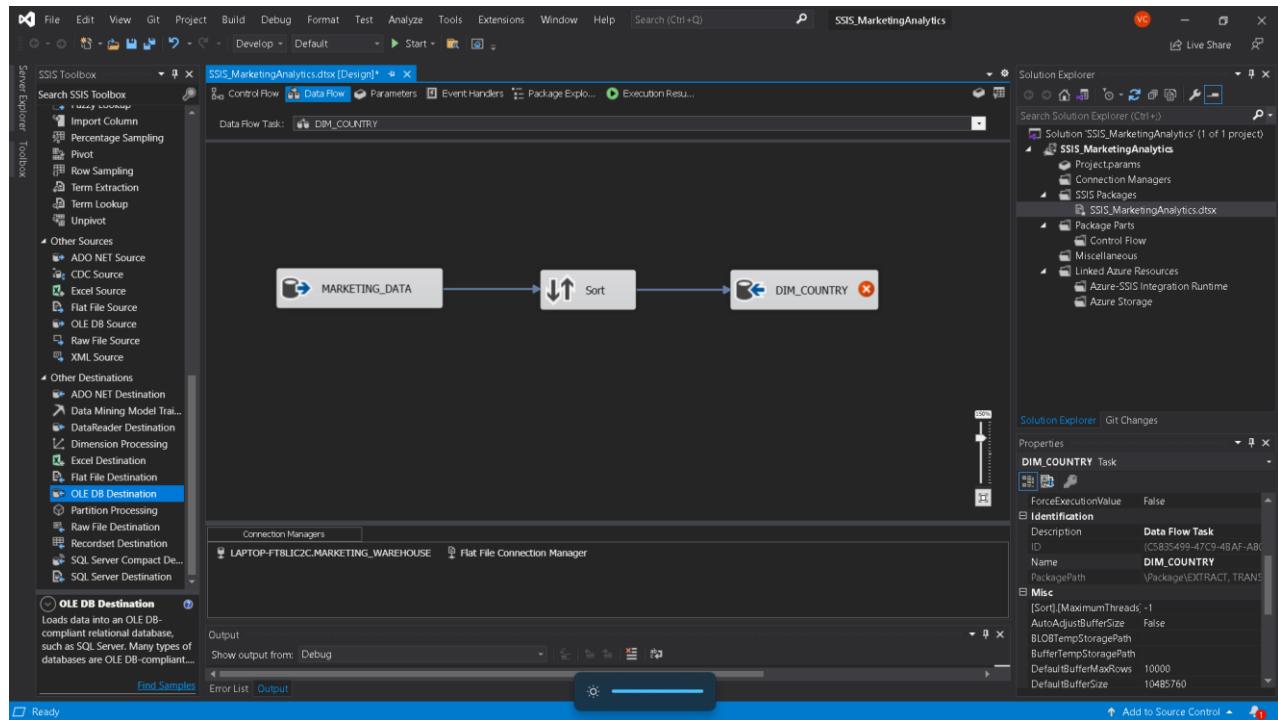
Hình 67. Sort sắp xếp dữ liệu

Edit Sort -> Chọn muôn sắp xếp theo thuộc tính nào



Hình 68. Sắp xếp dữ liệu theo thuộc tính

Sử dụng OLE DB Destination để đổ dữ liệu vào bảng -> Đổi tên thành DIM_COUNTRY

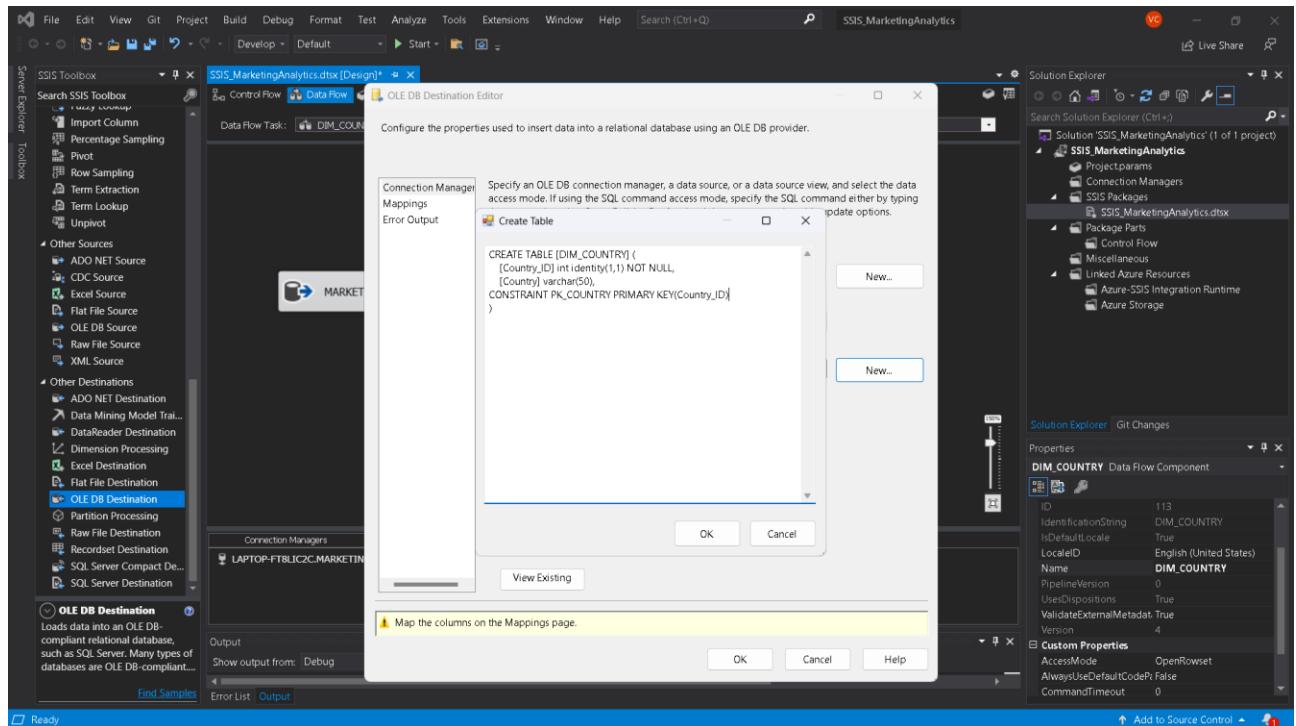


Hình 69. Đổ dữ liệu vào bảng DIM_COUNTRY

Edit DIM_COUNTRY

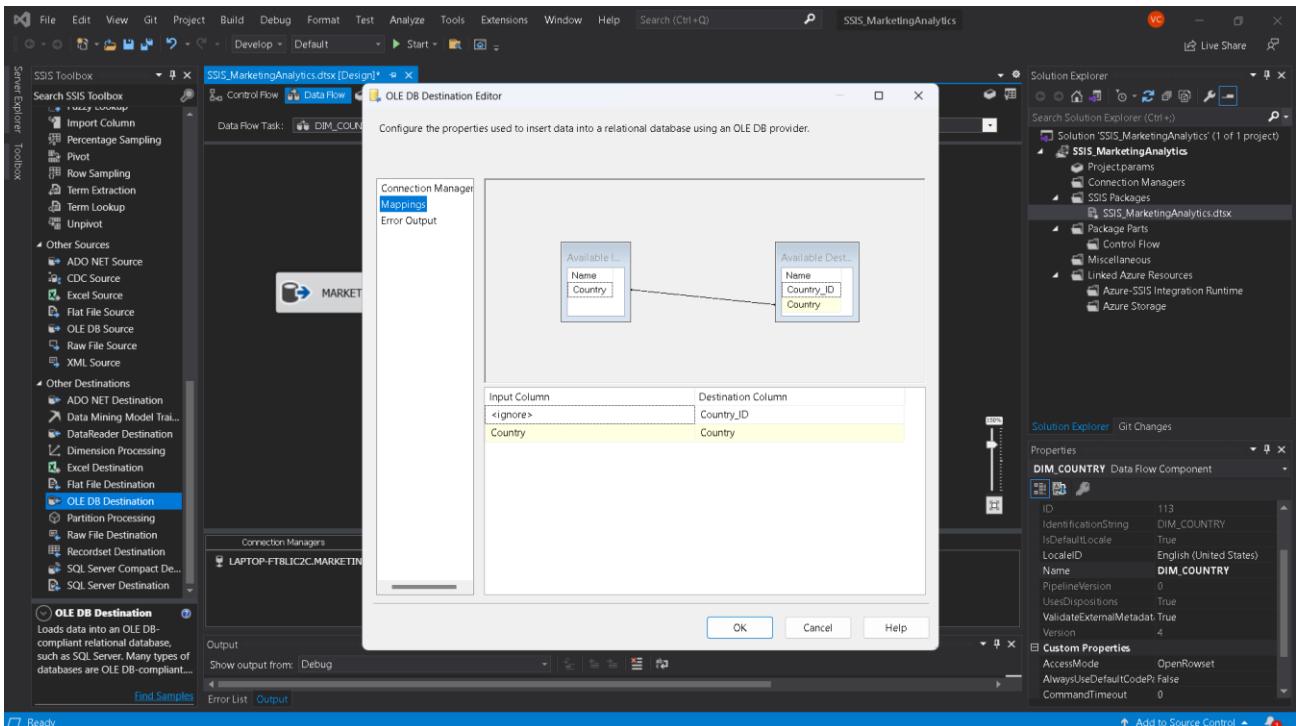
Chọn Connection đến MARKETING_WAREHOUSE

Chọn New để SSIS tự sinh ra đoạn mã insert data



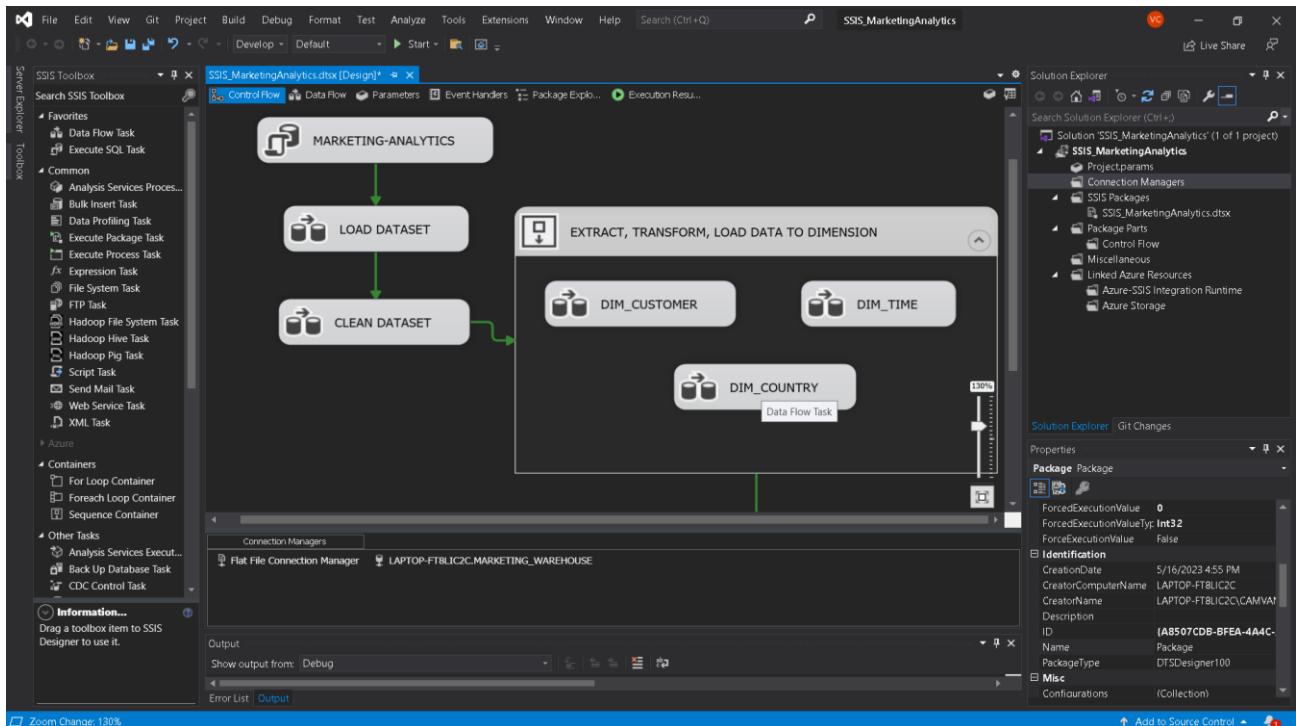
Hình 70. Tạo bảng DIM_COUNTRY để đỗ dữ liệu vào

Chọn Mappings -> Kiểm tra columns đã được map đúng hay chưa



Hình 71. Kiểm tra Mappings

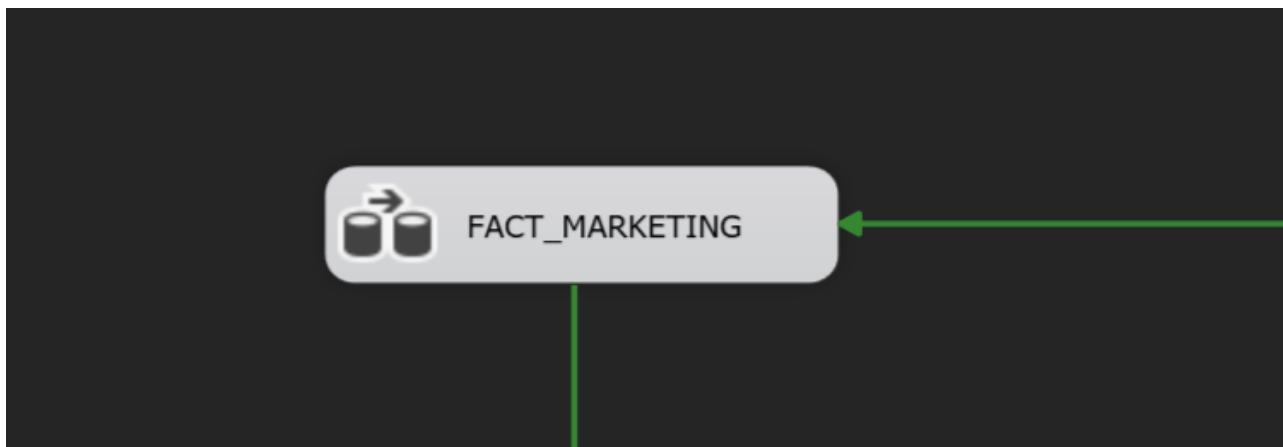
Kết quả thu được sau khi hoàn thành định nghĩa các **Data Flow Task** cho Sequence Container “**EXTRACT,TRANSFORM, LOAD DATA TO DIMENSION**”



Hình 72. EXTRACT, TRANSFORM, LOAD DATA TO DIMENSION

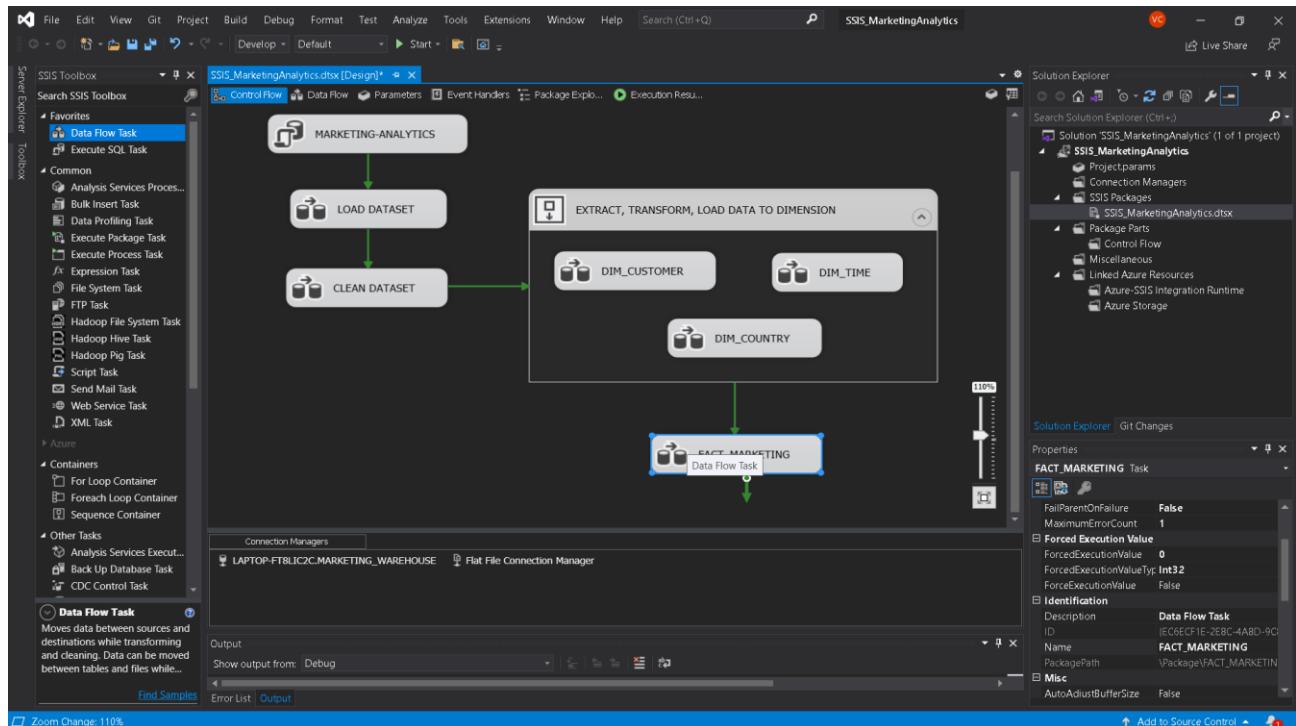
2.3.3 Định nghĩa và nạp dữ liệu vào bảng FACT

- ❖ **Mục tiêu:** Nạp dữ liệu đã được tiền xử lý nằm trong bảng MARKETING_DATA vào bảng FACT_MARKETING tương ứng theo đúng thiết kế lược đồ kho dữ liệu



Hình 73. Load Fact Table

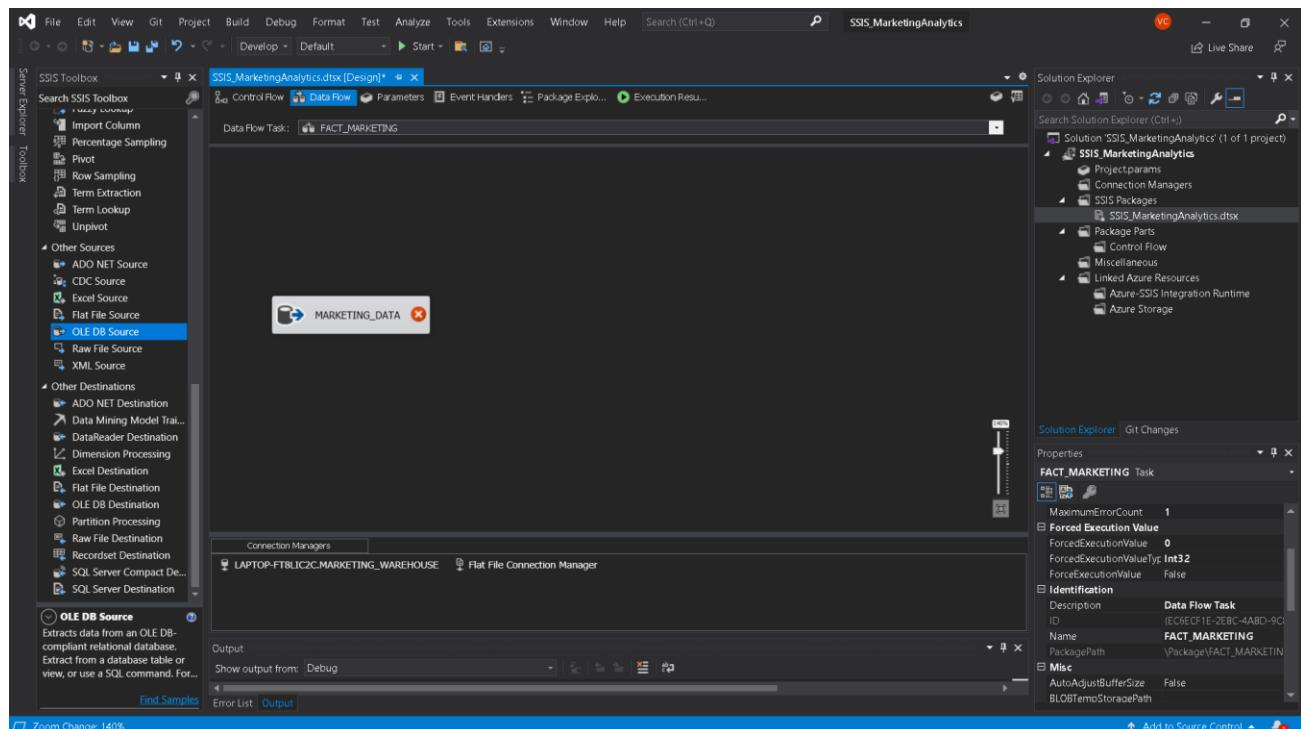
Bước 1: Sử dụng Data Flow Task -> Đổi tên thành FACT_MARKETING



Hình 74. Data Flow Task FACT_MARKETING

Bước 2: Edit FACT_MARKETING

Sử dụng OLE DB Source để định nghĩa dữ liệu nguồn -> Đổi tên thành MARKETING_DATA

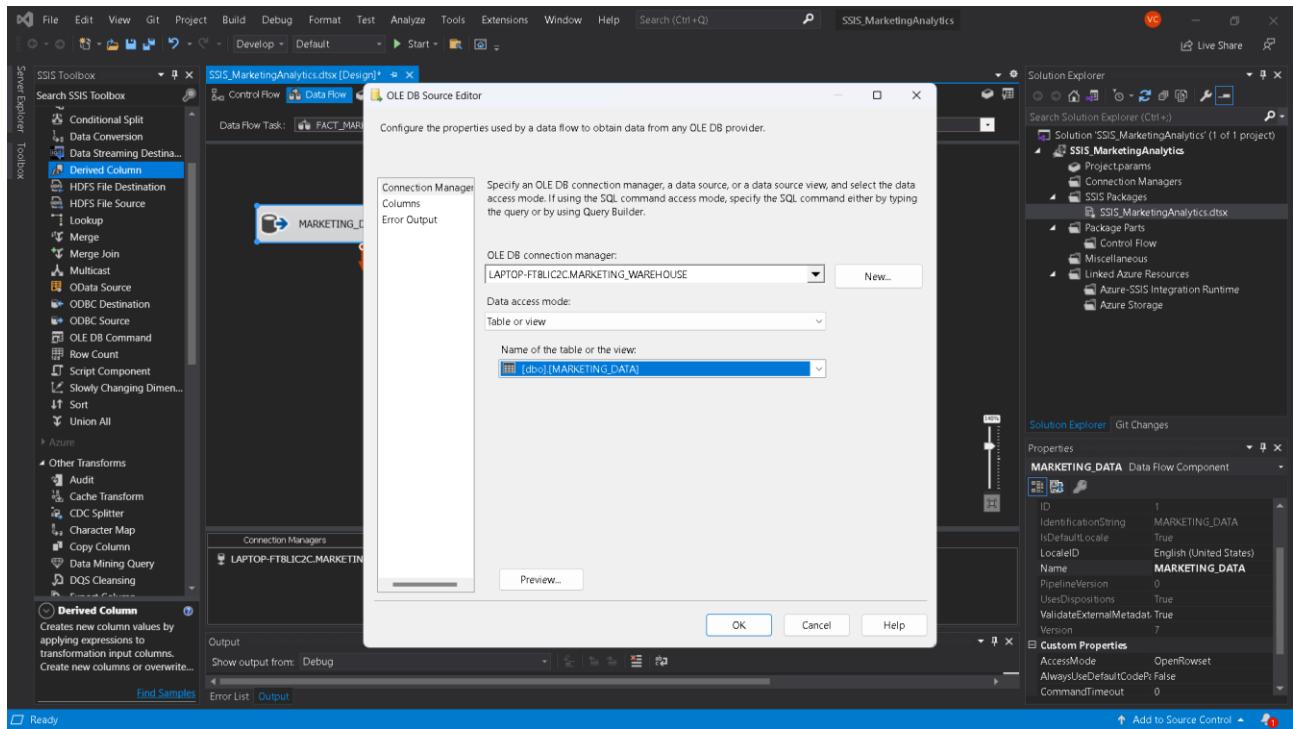


Hình 75. OLE DB Source định nghĩa nguồn dữ liệu

Bước 3: Edit OLE DB Source

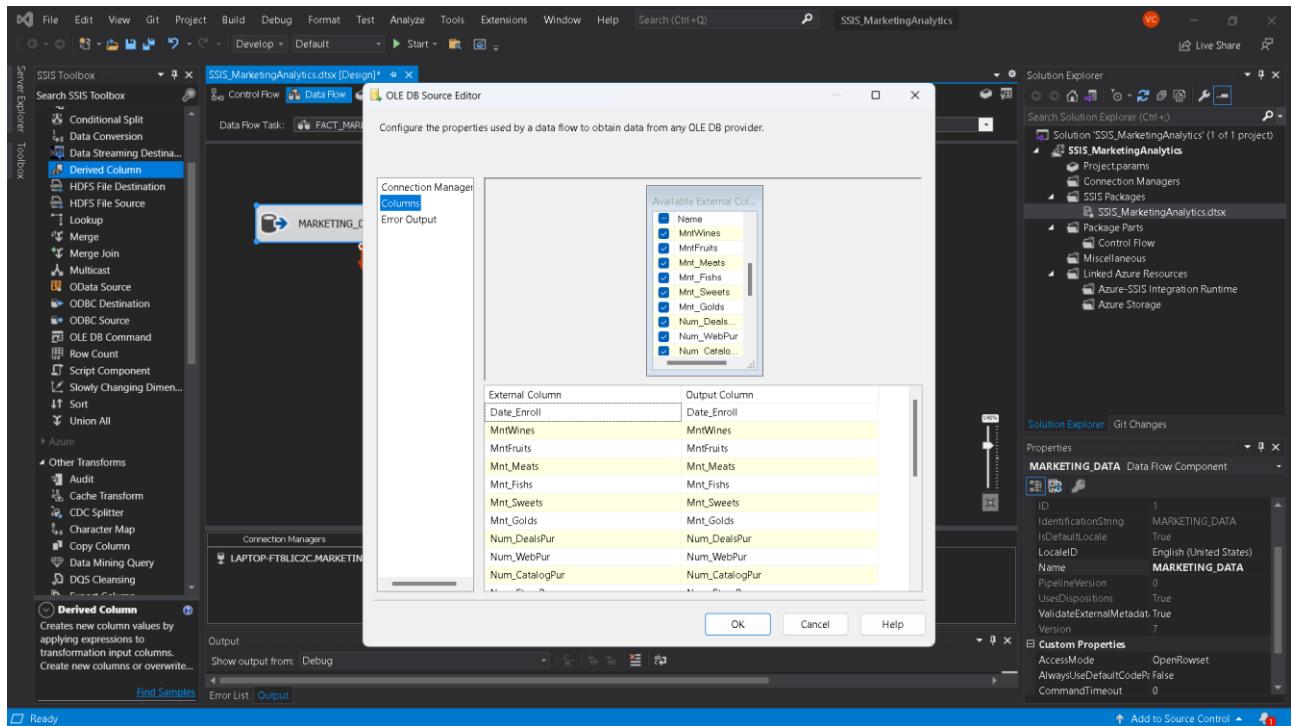
Chọn connection đến MARKETING_WAREHOUSE

Chọn table MARKETING_DATA



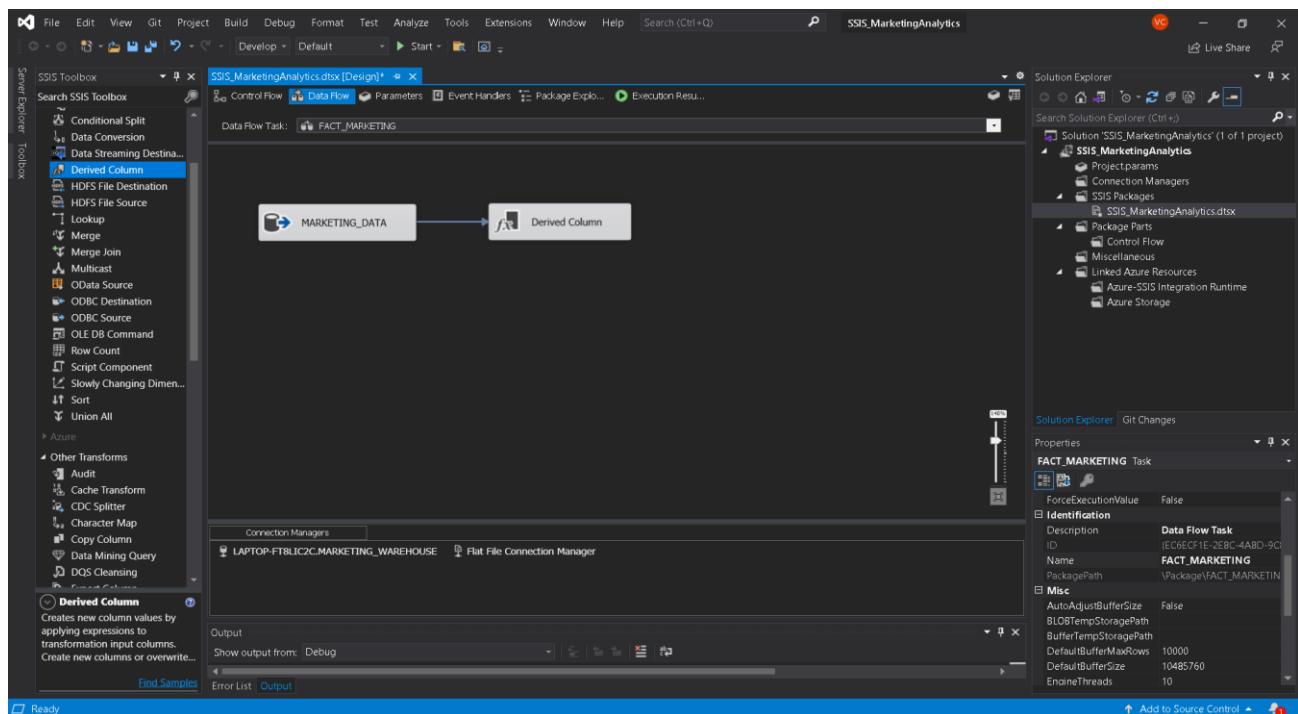
Hình 76. Định nghĩa nguồn dữ liệu là MARKETING_DATA

Chọn **Columns** -> Chọn những thuộc tính cần thiết cho FACT_MARKETING



Hình 77. Chọn thuộc tính cho FACT_MARKETING

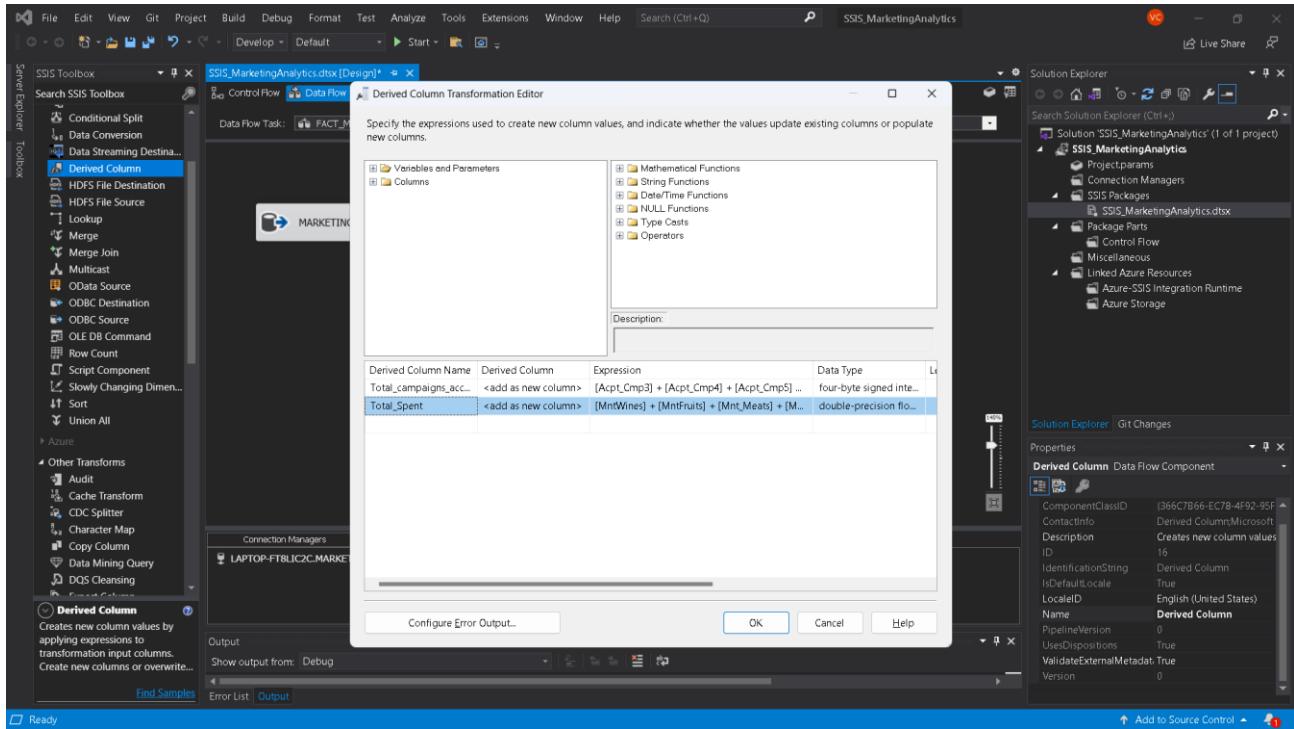
Bước 4: Sử dụng Derived Column để thêm 2 cột dữ liệu Total_Campaigns_Accepted và Total_Spent



Hình 78. Derived Column để thêm dữ liệu

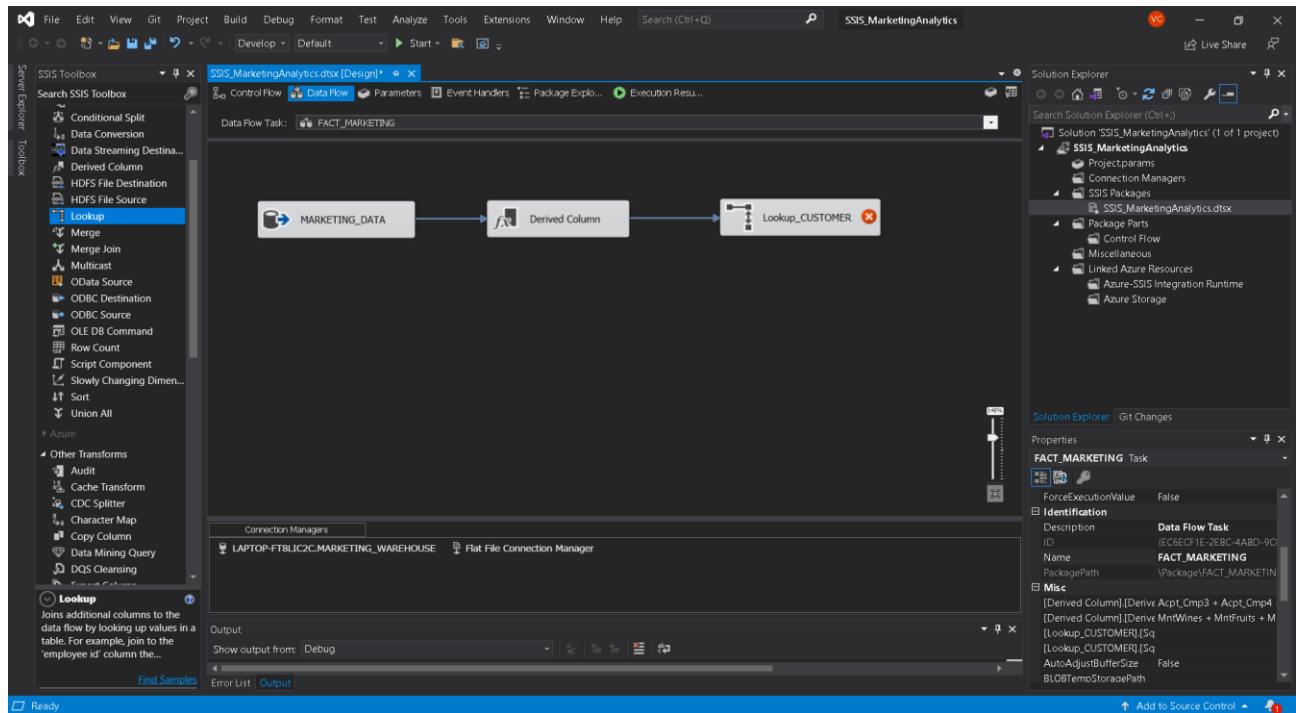
Bước 5: Edit Derived Column -> Thêm các function cần thiết để thêm 2 cột

Total_Campaigns_Accepted và Total_Spent



Hình 79. Thêm 2 cột Total_Campaigns_Accepted và Total_Spent

Bước 6: Kéo thả Lookup để link bảng FACT và bảng DIM_CUSTOMER -> Đổi tên thành Lookup_CUSTOMER



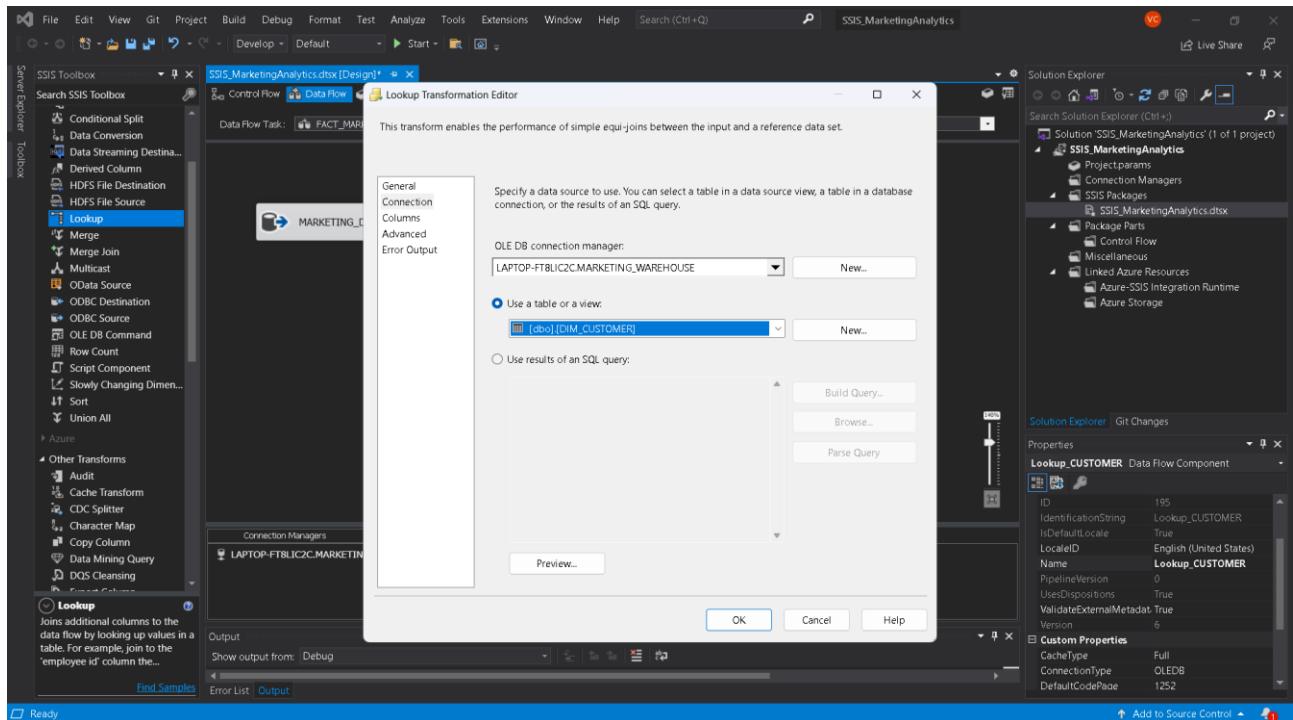
Hình 80. Link bảng FACT và bảng DIM_CUSTOMER

Bước 7: Edit Lookup_CUSTOMER

Chuyển **Specify how to handle rows** thành **Redirect rows to no match output**

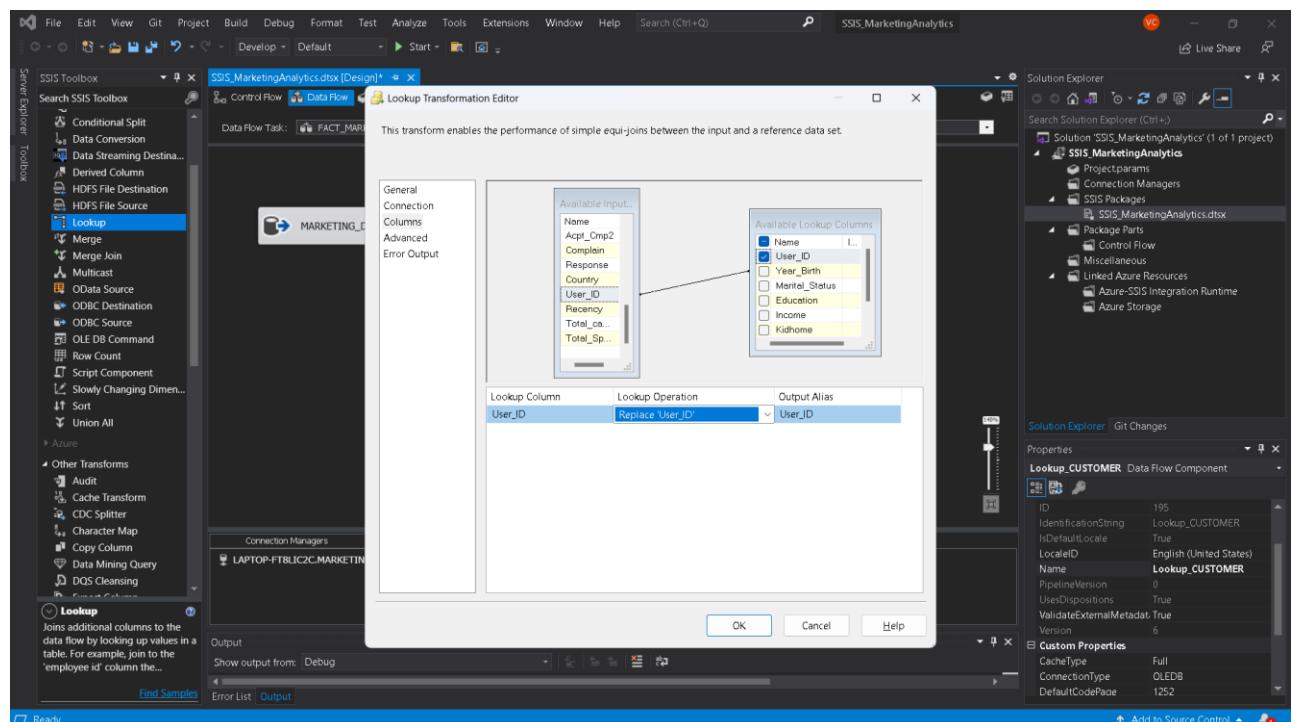
Chọn **Connection** -> Đổi connection đến **MARKETING_WAREHOUSE**

Chọn table **DIM_CUSTOMER**



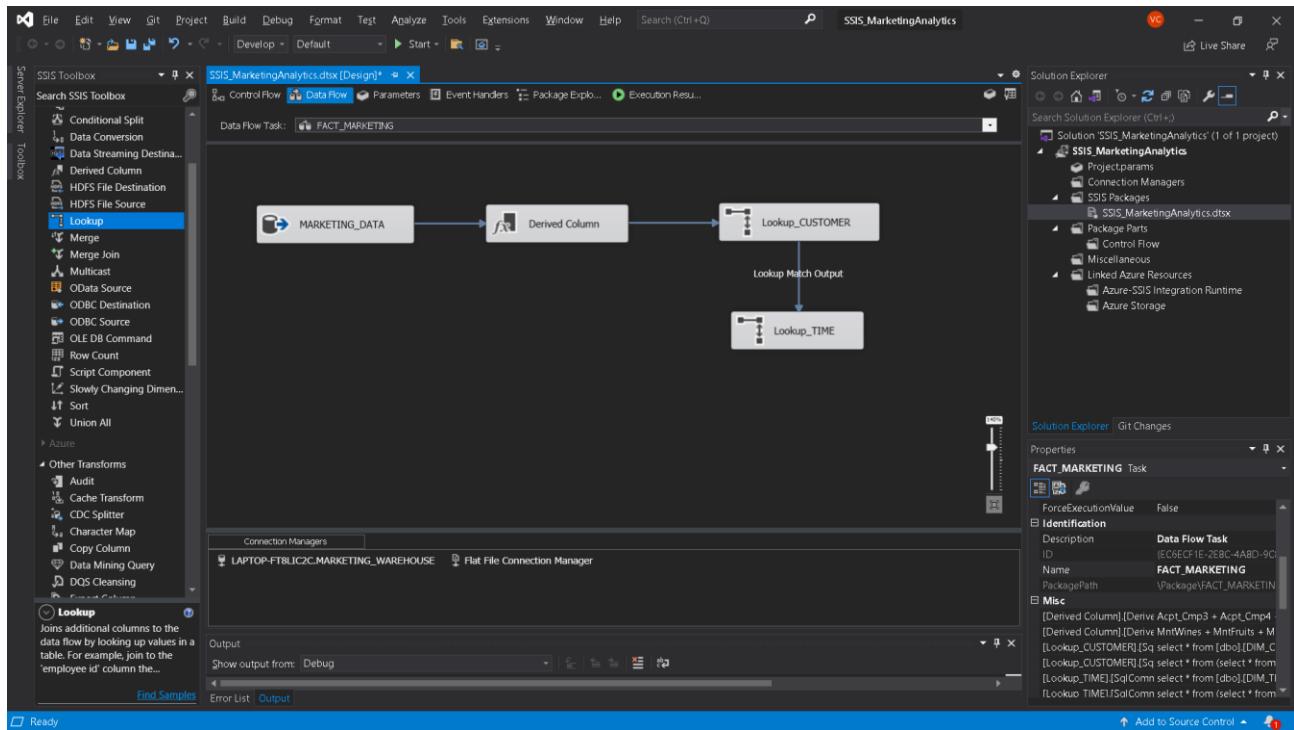
Hình 81. Chọn Table DIM_CUSTOMER

Chọn columns -> Nối cột User_ID



Hình 82. Nối cột User_ID

Bước 8: Kéo thả **Lookup** để link bảng **FACT** và bảng **DIM_TIME** -> Đổi tên thành **Lookup_TIME**



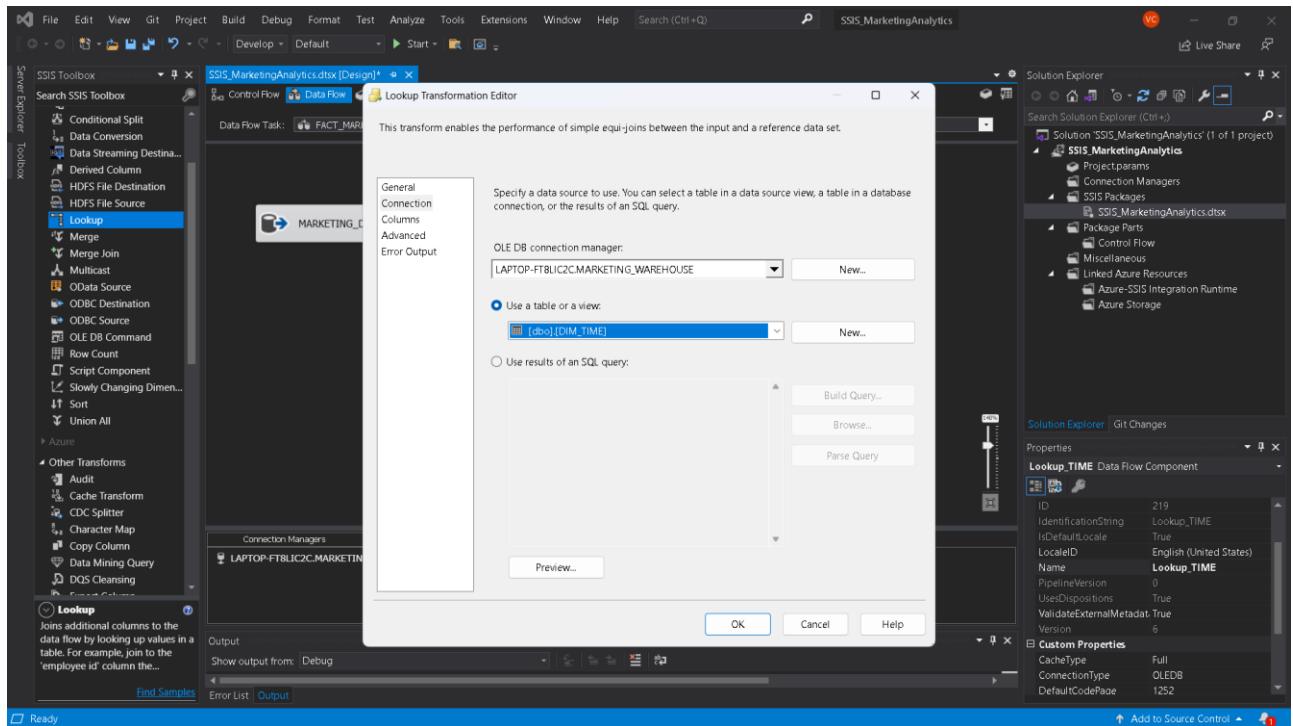
Hình 83. Link bảng FACT và bảng DIM_TIME

Bước 9: Edit Lookup_TIME

Chuyển **Specify how to handle rows** thành **Redirect rows to no match output**

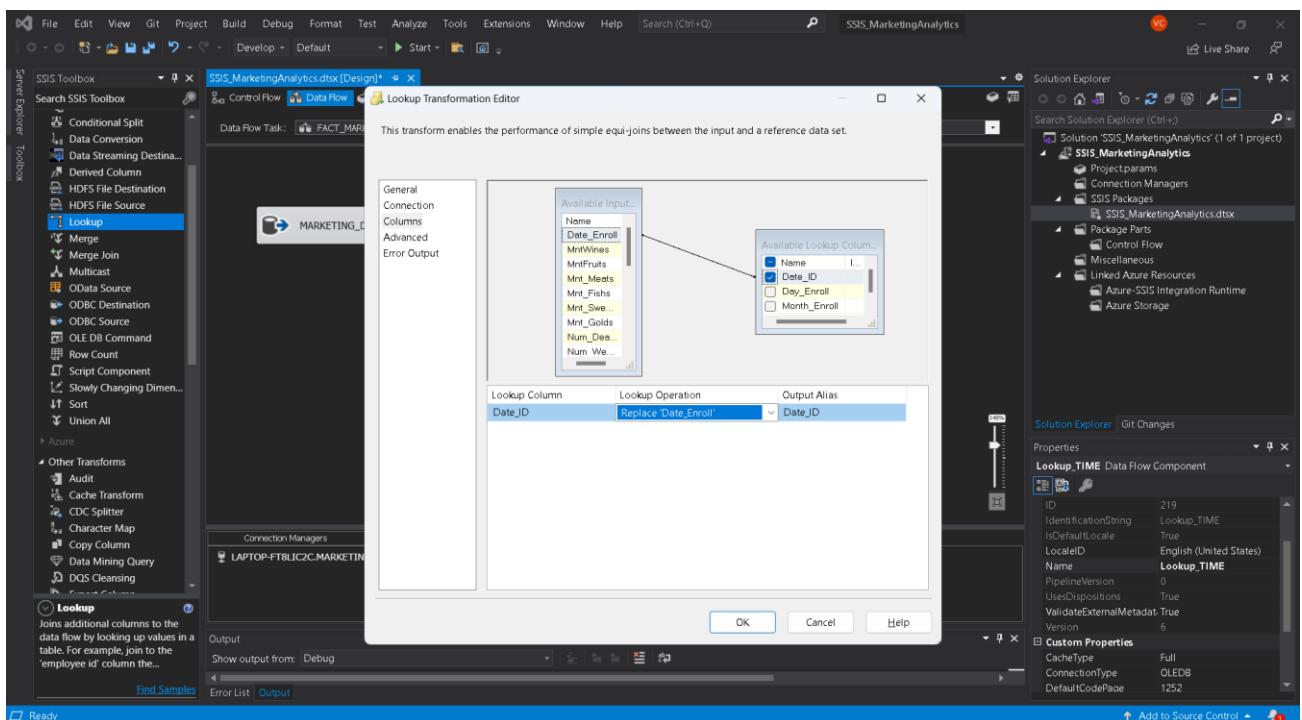
Chọn **Connection** -> Đổi connection đến **MARKETING_WAREHOUSE**

Chọn table **DIM_DATE**



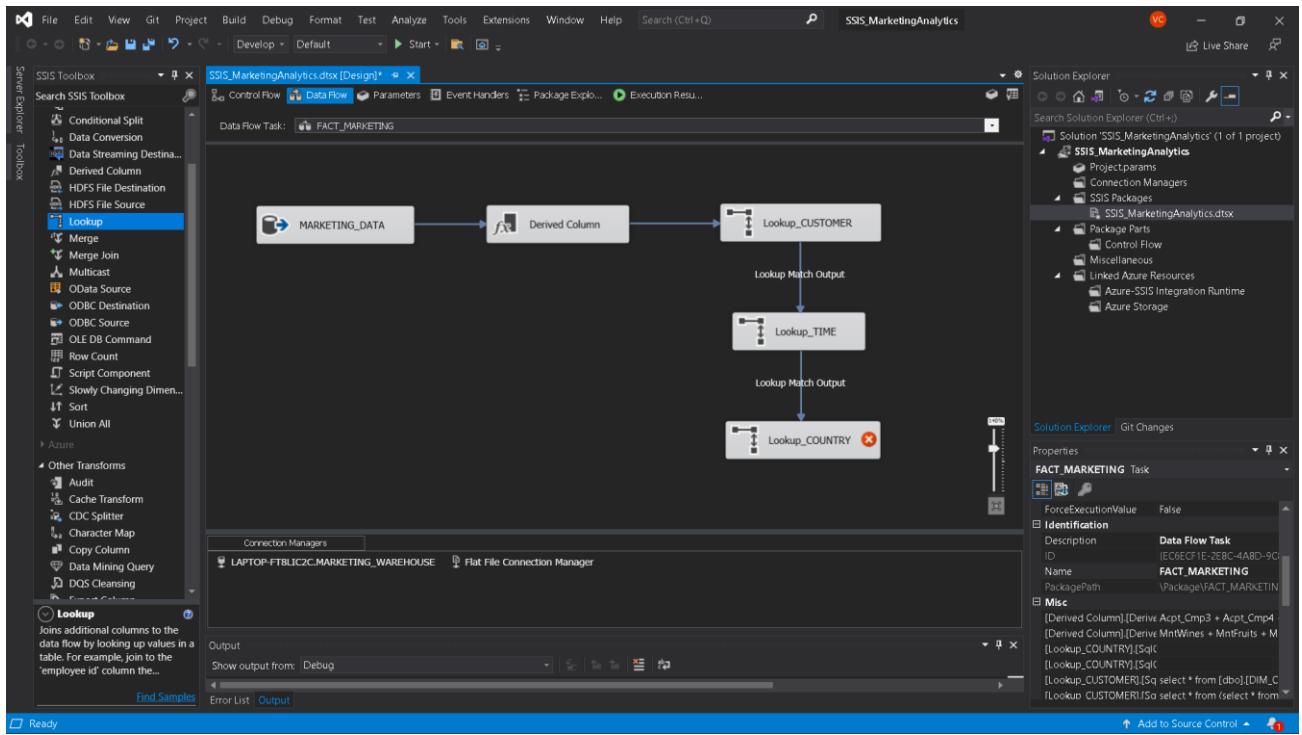
Hình 84. Chọn bảng DIM_TIME

Chọn columns -> Nối cột Date_ID



Hình 85. Nối cột Date_ID

Bước 10: Kéo thả Lookup để link bảng FACT và bảng DIM_COUNTRY -> Đổi tên thành **Lookup_COUNTRY**



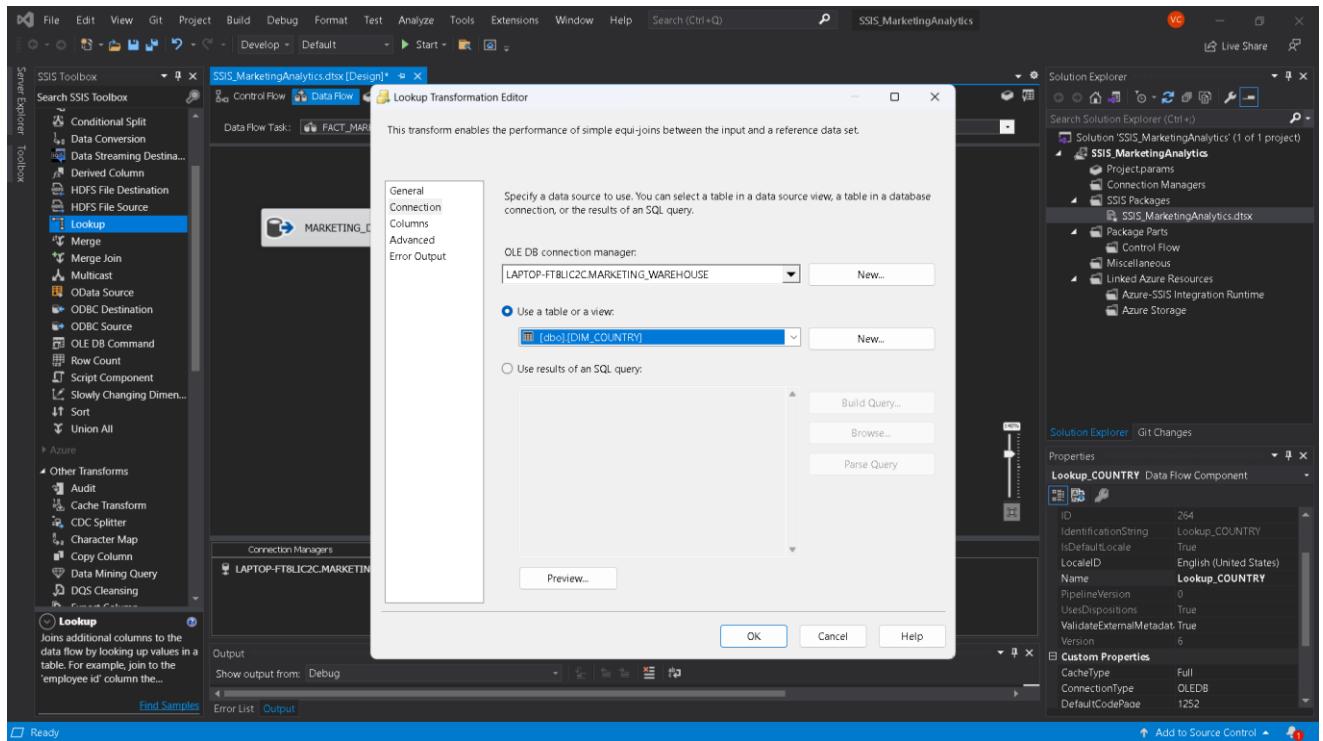
Hình 86. Link bảng FACT và bảng DIM_COUNTRY

Bước 11: Edit Lookup_COUNTRY

Chuyển **Specify how to handle rows** thành **Redirect rows to no match output**

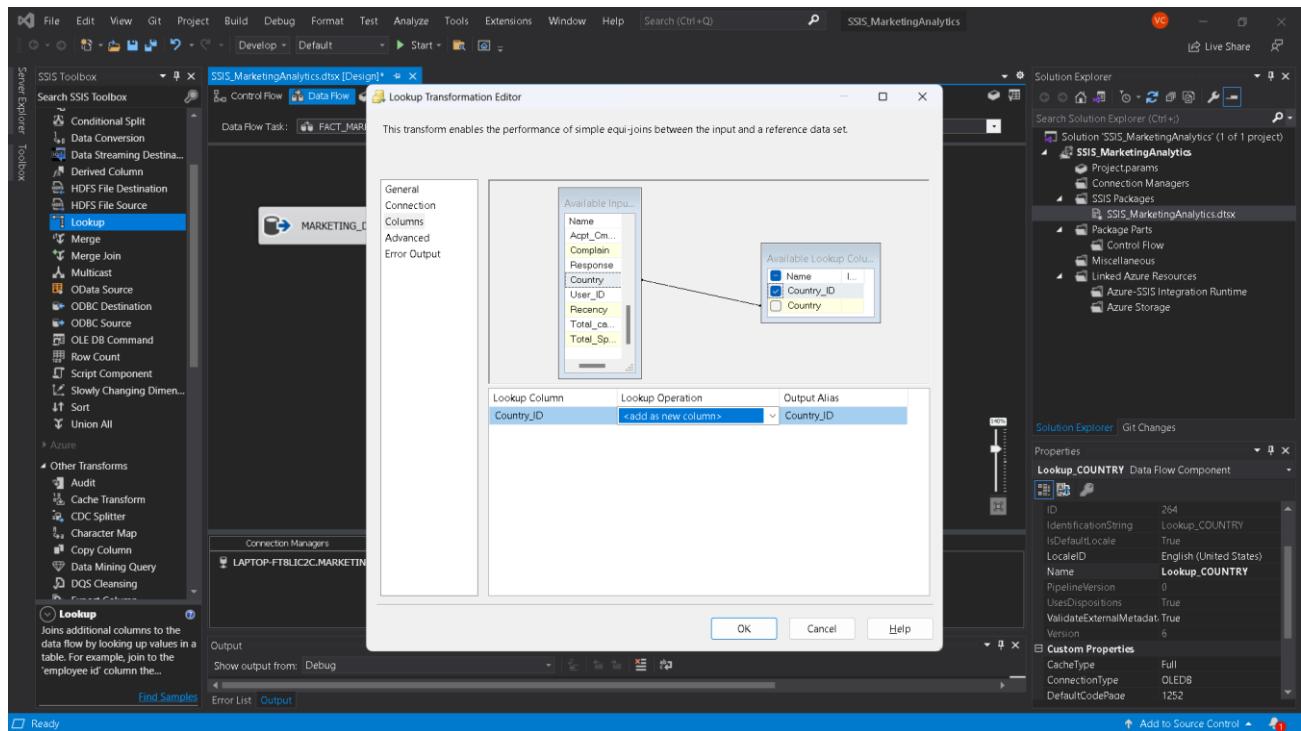
Chọn **Connection** -> Đổi connection đến **MARKETING_WAREHOUSE**

Chọn table **DIM_COUNTRY**



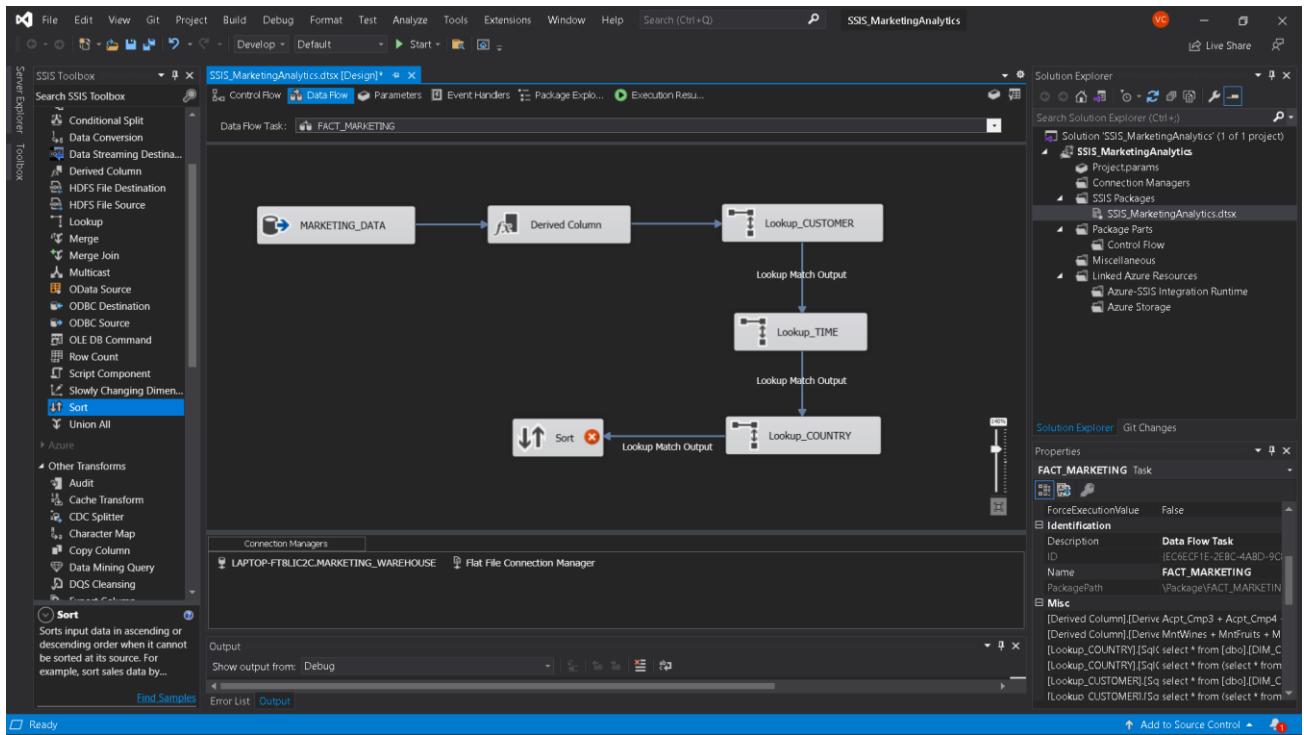
Hình 87. Chọn bảng DIM_COUNTRY

Chọn columns -> Nối cột Country



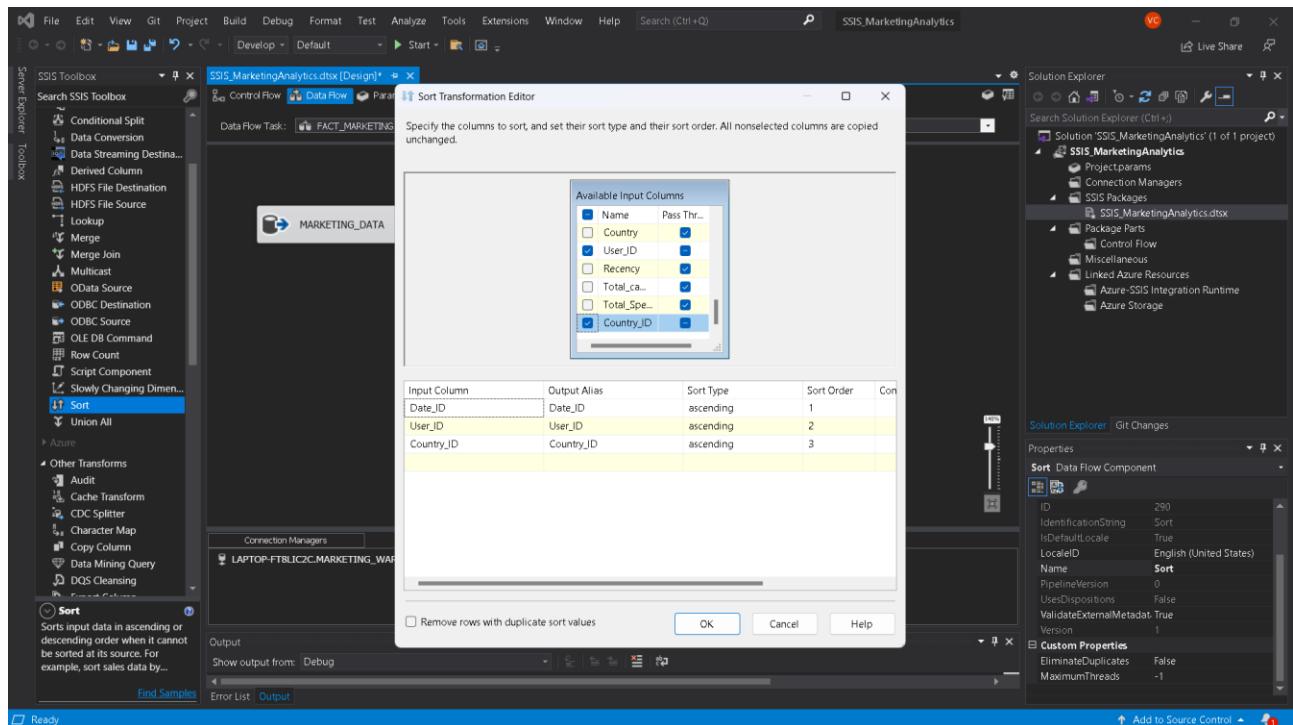
Hình 88. Nối cột Country

Bước 12: Sử dụng Sort để sắp xếp lại các dòng dữ liệu



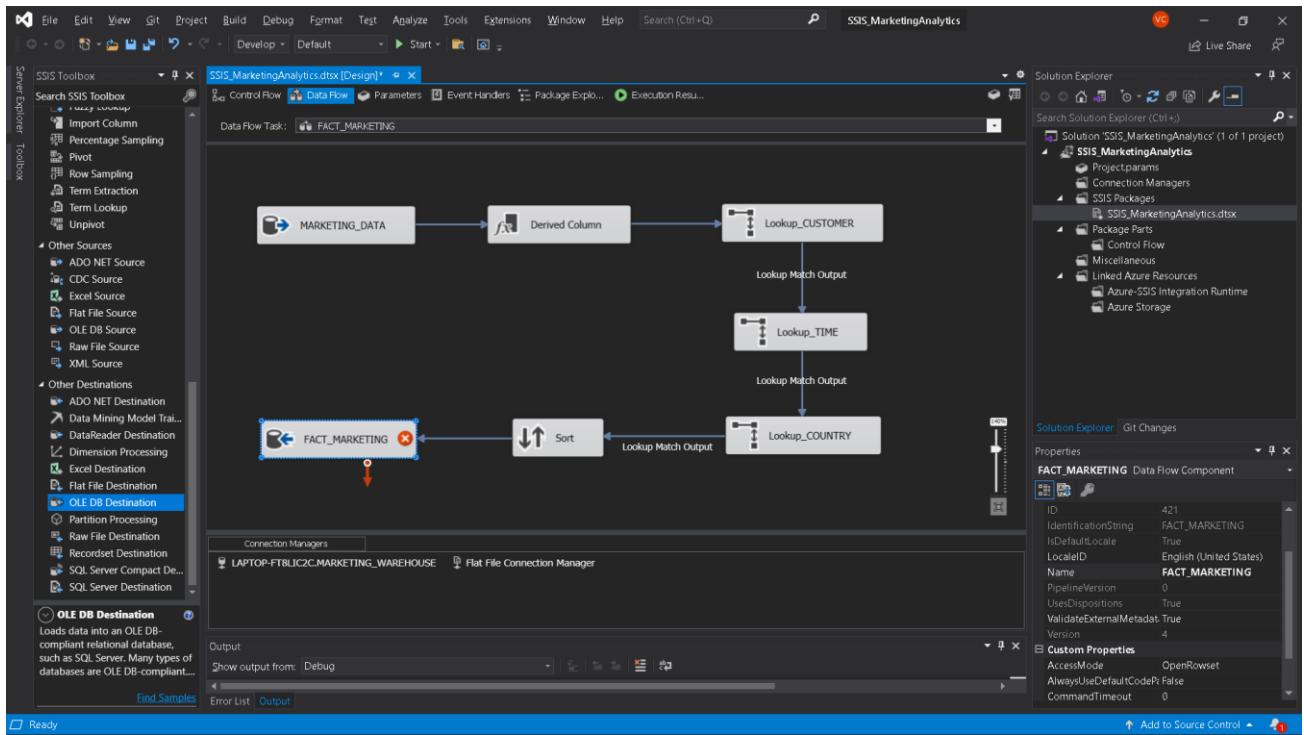
Hình 89. Sort sắp xếp dữ liệu

Bước 13: Edit Sort -> Chọn sắp xếp dữ liệu theo User_ID, Date_ID , Country_ID



Hình 90. Sắp xếp dữ liệu theo Date_ID, User_ID, Country_ID

Bước 14: Sử dụng OLE DB Destination để đổ dữ liệu vào bảng -> Đổi tên thành FACT_MARKETING

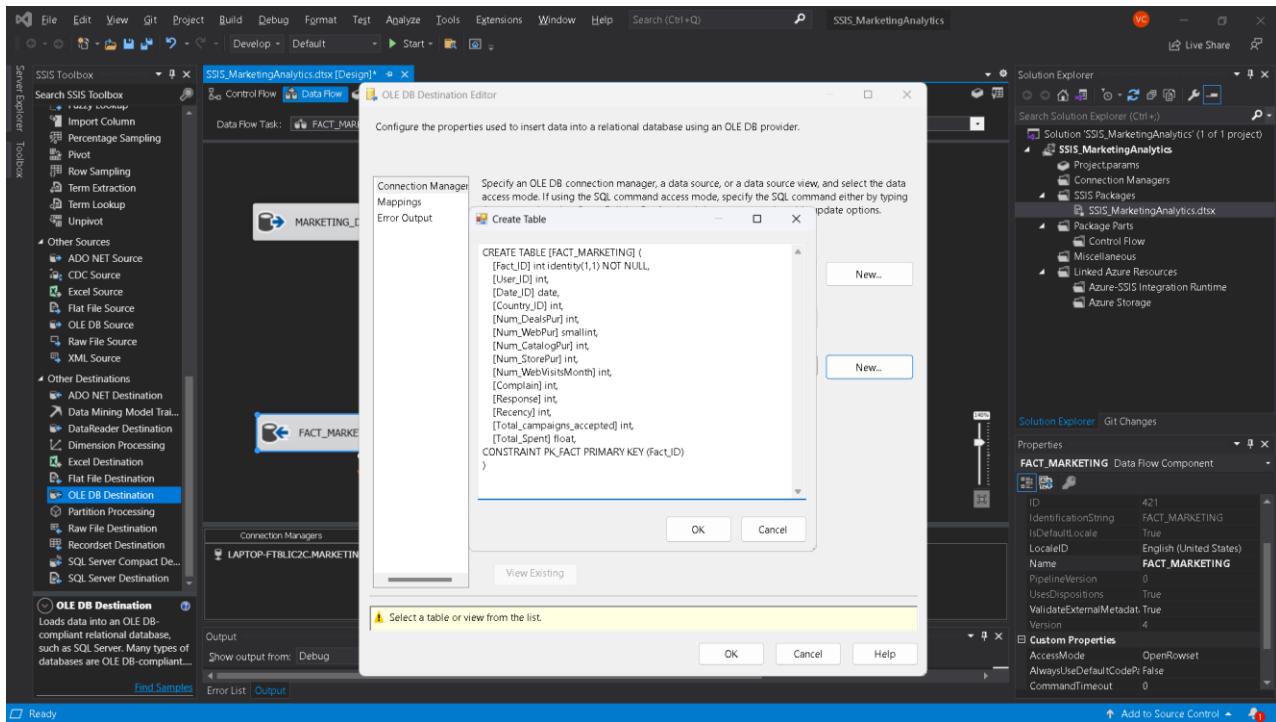


Hình 91. Đỗ dữ liệu vào bảng FACT_MARKETING

Bước 15: Edit FACT_MARKETING

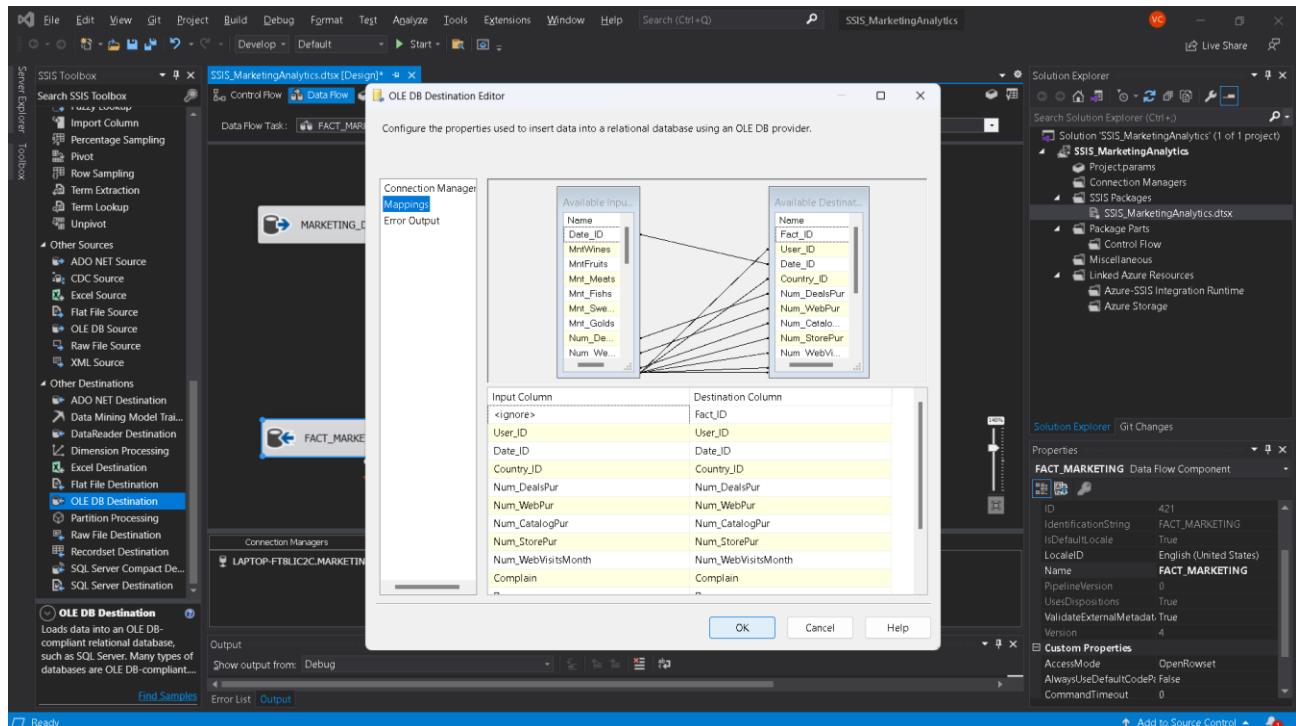
Chọn connection đến MARKETING_WAREHOUSE

Chọn New để SSIS tự sinh ra đoạn mã insert data



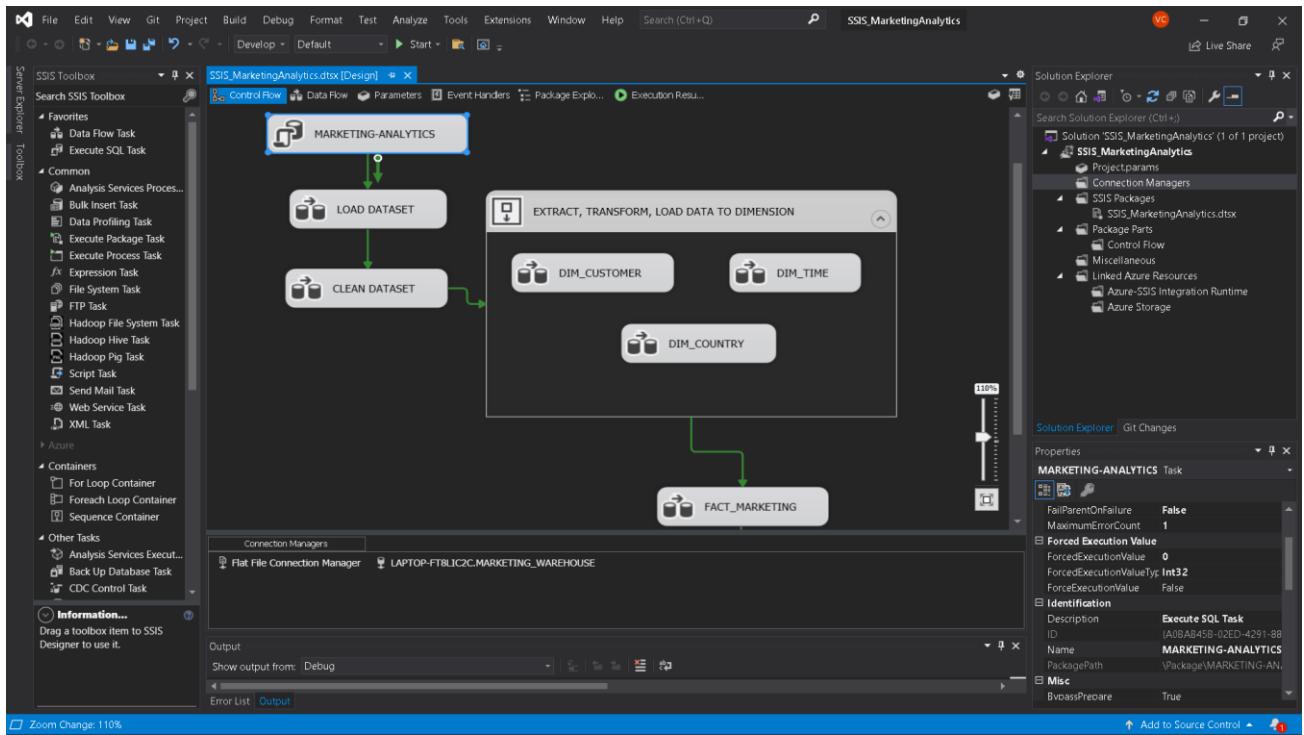
Hình 92. Tạo bảng FACT_MARKETING để dữ liệu vào

Chọn Mappings -> Kiểm tra columns đã được map đúng hay chưa



Hình 93. Kiểm tra Mappings

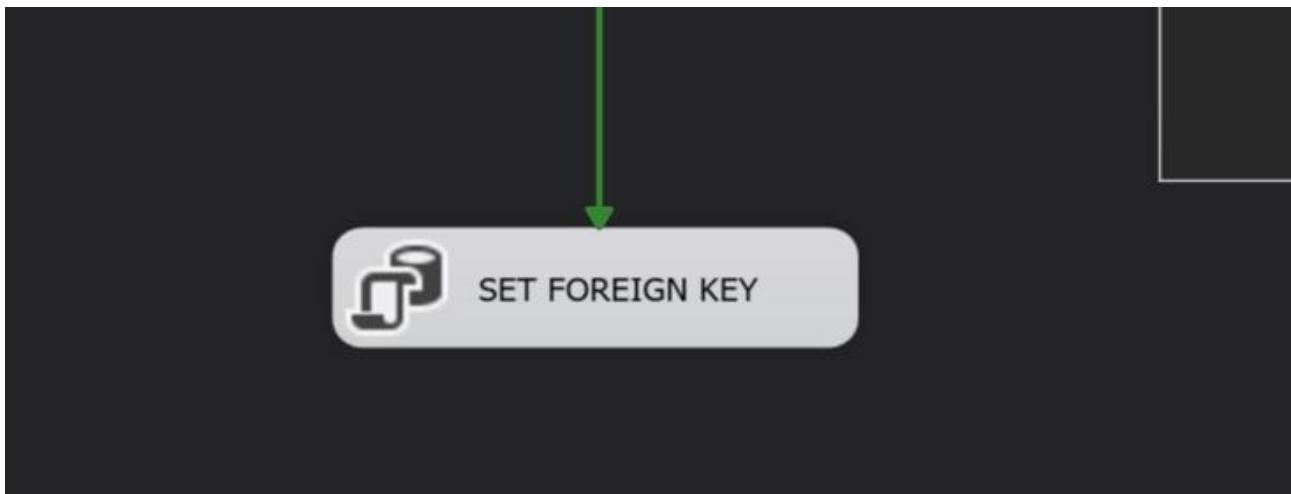
Kết quả thu được sau khi hoàn thành định nghĩa FACT_MARKETING



Hình 94. FACT_MARKETING

2.3.4 Định nghĩa và set các khóa ngoại cho bảng FACT

- ❖ **Mục tiêu:** Sau khi đã nạp dữ liệu vào bảng Fact, tiến hành tạo các khóa ngoại theo đúng thiết kế



Hình 95. Set Foreign Keys

```

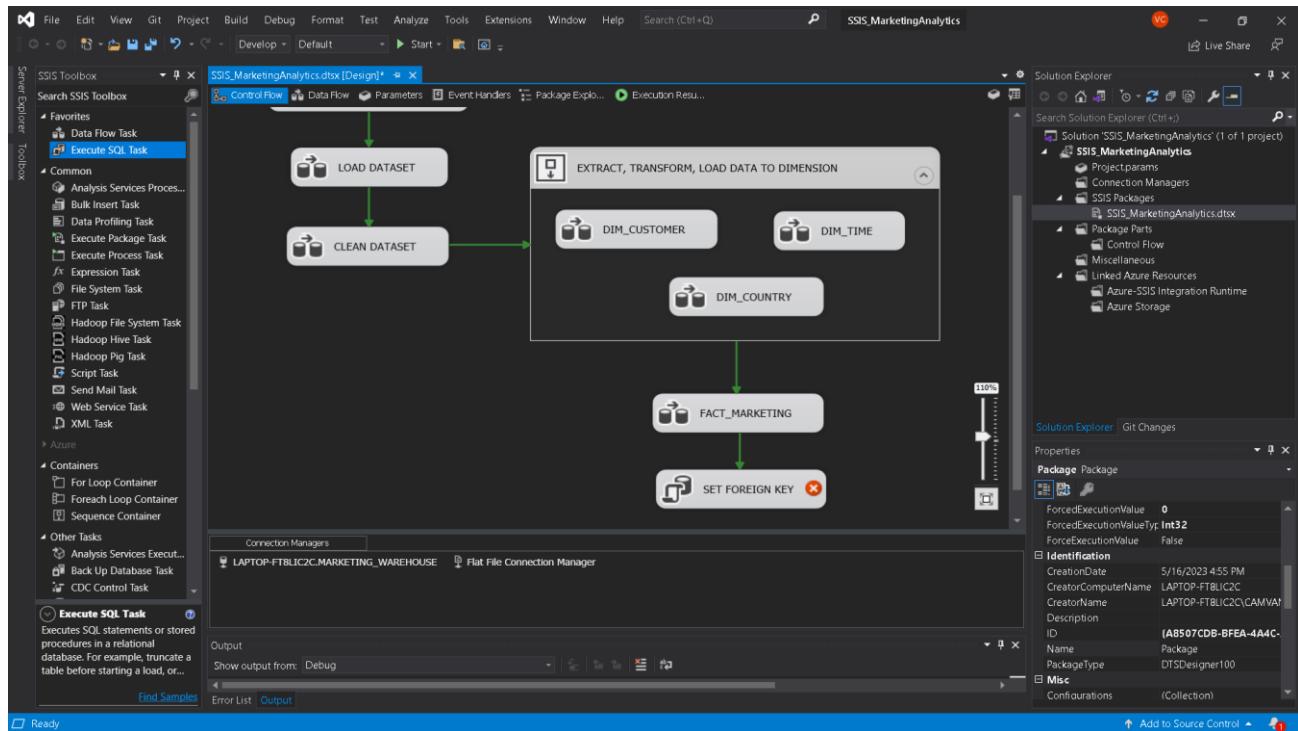
IF NOT EXISTS (SELECT NAME FROM sys.foreign_keys WHERE NAME = 'FK_FACT_CUS')
BEGIN
    ALTER TABLE FACT_MARKETING
        ADD CONSTRAINT FK_FACT_CUS FOREIGN KEY(User_ID)
        REFERENCES DIM_CUSTOMER(User_ID);
END
GO

IF NOT EXISTS (SELECT NAME FROM sys.foreign_keys WHERE NAME = 'FK_FACT_TIME')
BEGIN
    ALTER TABLE FACT_MARKETING
        ADD CONSTRAINT FK_FACT_TIME
        FOREIGN KEY(Date_ID) REFERENCES DIM_TIME(Date_ID);
END
GO

IF NOT EXISTS (SELECT NAME FROM sys.foreign_keys WHERE NAME = 'FK_FACT_COUNTRY')
BEGIN
    ALTER TABLE FACT_MARKETING
        ADD CONSTRAINT FK_FACT_COUNTRY
        FOREIGN KEY(Country_ID) REFERENCES Dim_Country(Country_ID);
END
GO

```

Bước 1: Sử dụng Execute SQL Task -> Đổi tên thành SET FOREIGN KEY

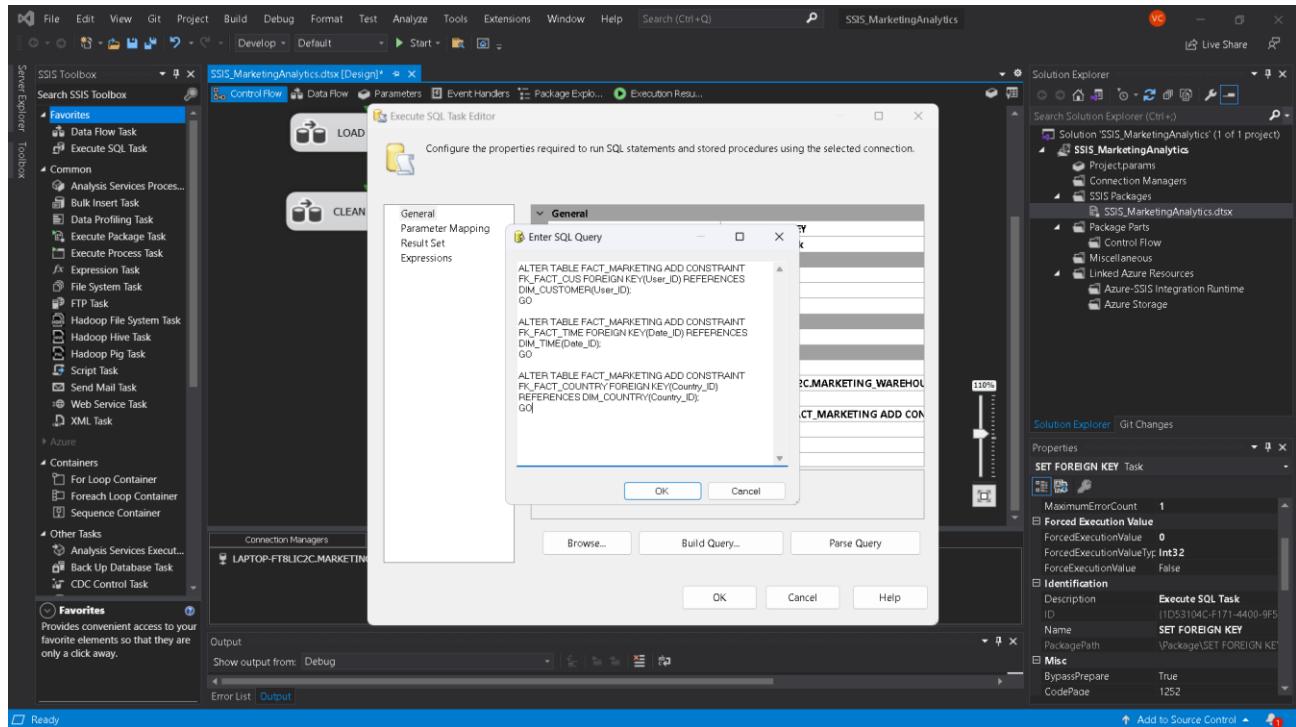


Hình 96. Set Foreign Key

Bước 2: Edit SET FOREIGN KEY

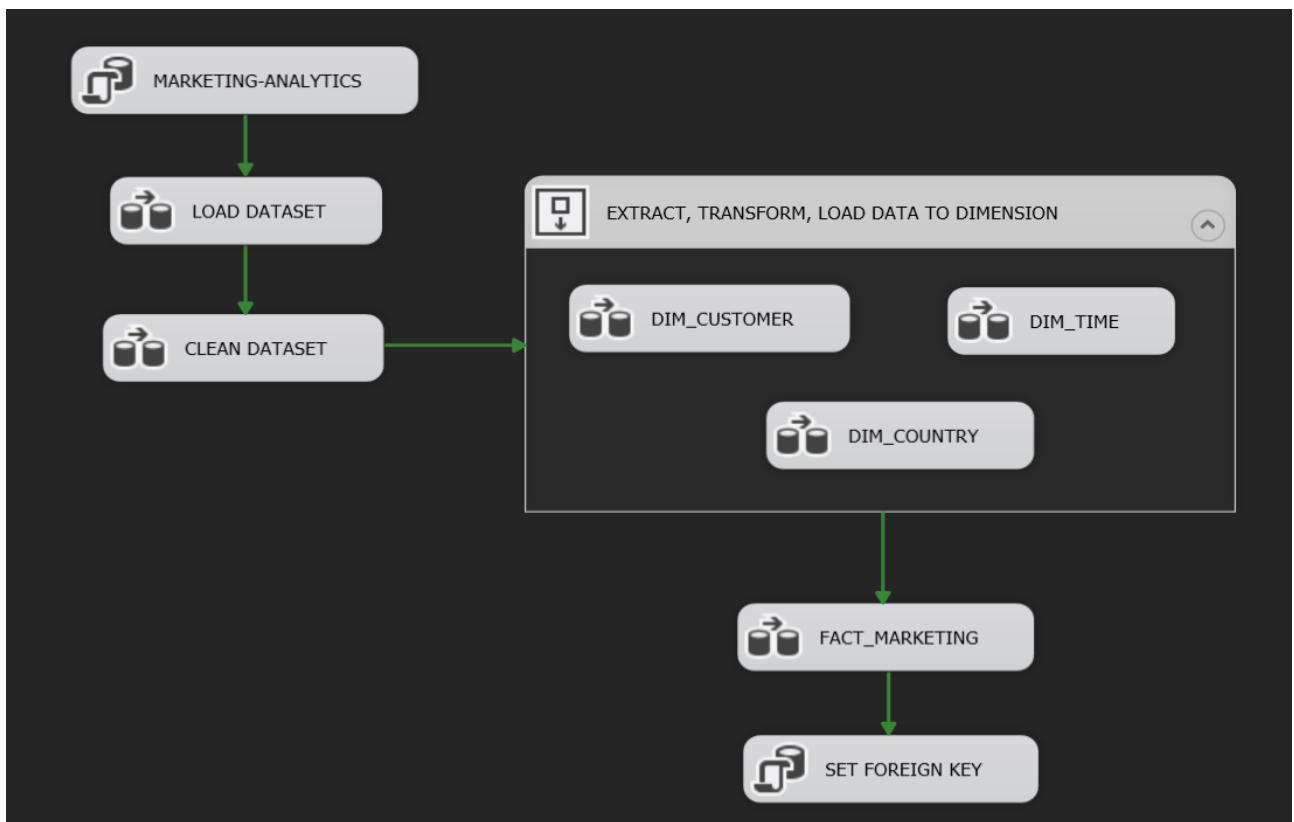
Định nghĩa Connection là tới MARKETING_WAREHOUSE

Chỉnh sửa SQLStatement



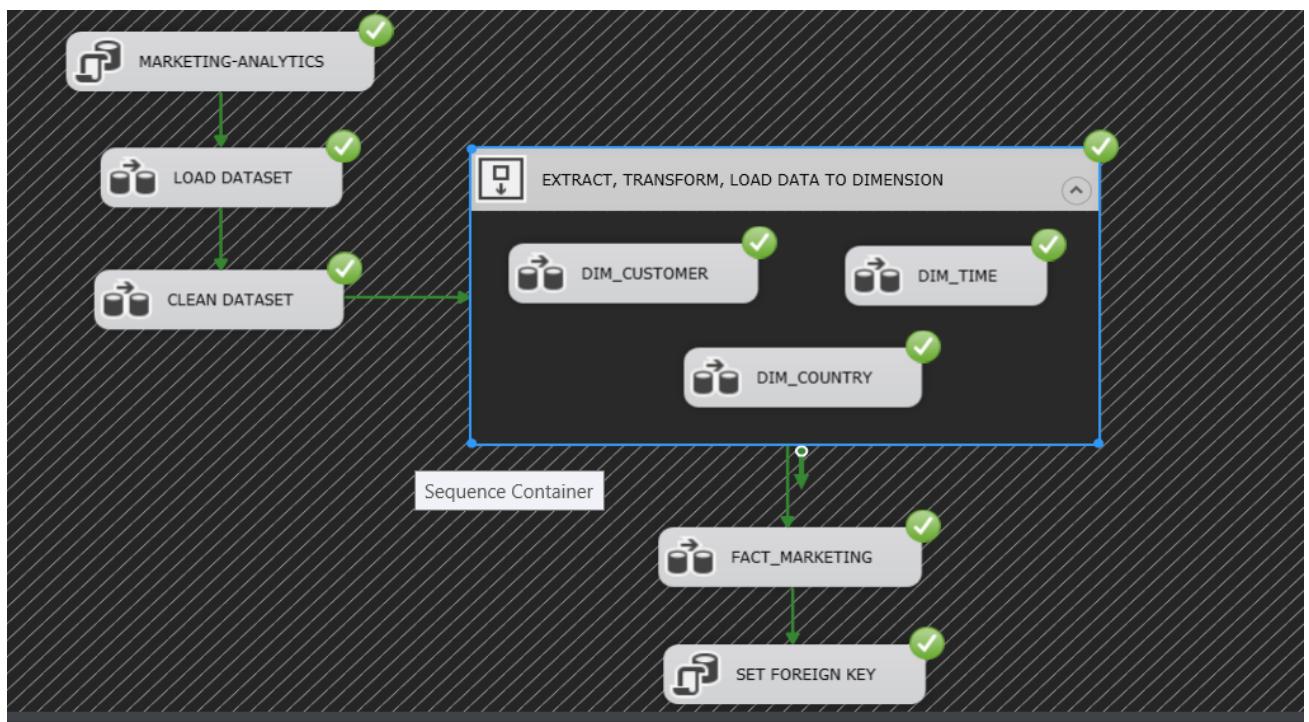
Hình 97. Chỉnh sửa SQL Statement

2.4 Thực thi Package



Hình 98. Thực thi Packagea

CONTROL FLOW



Hình 99. Control Flow

2.5 Dữ liệu sau khi hoàn thành

SQLQuery1.sql - LAPTOP-FTBLIC2C\MARKETING_WAREHOUSE (LAPTOP-FTBLIC2C\CAMVAN (52))- Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Object Explorer Results Messages

CREATE DATABASE MARKETING_WAREHOUSE;

USE MARKETING_WAREHOUSE;

SELECT * FROM ORIGINAL_DATA;

Results

User_ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Date_Enroll	Recency	MinFives	MinFruits	MinMeals	Min_Fishes	Min_Sweets	Min_Golds	Num_DealsPur	Num_WebPur	Num_CatalogPur	Num_StorePur	Num_WebVisitsMonth	
1	1295	1983	Graduation	Married	70981	0	0	2013-05-04	98	239	10	554	254	87	54	1	3	4	9	1
2	1812	1981	PhD	Single	57101	0	0	2014-09-15	98	494	5	64	7	0	36	4	3	2	2	9
3	9	1975	Master	Single	46998	1	1	2013-09-18	98	57	0	27	0	0	37	1	7	3	7	5
4	13	1947	PhD	Widow	25259	0	1	2013-07-22	57	19	0	5	0	0	8	2	1	0	3	6
5	17	1971	PhD	Married	60491	0	1	2013-09-01	81	637	47	237	12	19	76	4	6	11	7	5
6	20	1965	2n Cycle	Married	46891	0	1	2013-09-01	91	43	12	23	29	15	81	1	2	1	4	4
7	22	1976	Graduation	Divorced	46310	1	0	2012-12-03	99	185	2	88	15	5	14	2	6	1	5	8
8	24	1960	Master	Together	17144	1	1	2014-02-15	96	18	2	19	0	2	6	5	3	0	4	7
9	25	1958	Graduation	Married	65148	0	1	2012-11-18	9	468	35	422	33	12	153	2	6	6	7	4
10	35	1987	Graduation	Married	25545	1	0	2012-09-04	35	32	1	64	16	12	85	3	2	2	3	6
11	48	1964	Graduation	Together	55781	0	1	2014-04-24	97	136	1	12	0	3	32	2	4	1	3	6
12	49	1970	Graduation	Single	20587	1	0	2014-05-11	39	2	3	6	4	1	9	1	1	1	2	7
13	55	1983	Graduation	Together	56253	0	1	2012-12-07	83	509	0	65	7	11	5	4	7	2	9	6
14	67	1972	Master	Single	46423	1	1	2013-09-18	6	68	0	16	0	0	8	3	2	0	4	7
15	73	1953	PhD	Single	51411	1	2	2013-05-05	81	14	0	3	0	0	1	1	0	0	3	5
16	75	1982	2n Cycle	Together	41039	0	0	2013-07-24	32	34	11	137	179	61	45	1	8	1	5	8
17	78	1889	Graduation	Married	25259	1	0	2013-09-15	51	15	0	11	0	2	9	1	1	1	2	8
18	97	1981	2n Cycle	Married	27733	0	1	2013-06-28	16	0	7	5	26	2	17	2	2	0	3	7
19	99	1975	2n Cycle	Married	22212	1	0	2013-08-28	49	5	0	20	5	2	21	2	2	0	4	6
20	92	1988	Graduation	Alone	34176	1	0	2014-05-12	12	5	7	24	19	14	20	4	3	0	4	6
21	113	1951	Graduation	Married	20833	1	1	2013-07-18	33	11	0	7	3	1	6	2	1	0	3	5
22	115	1966	Master	Single	43456	0	1	2013-03-01	0	275	11	88	25	7	7	3	5	1	8	5
23	123	1951	Graduation	Widow	67948	0	1	2013-01-04	92	544	35	133	23	151	107	2	4	5	4	2
24	125	1958	2n Cycle	Together	53083	1	1	2013-05-15	65	215	7	33	3	0	13	6	5	1	5	7
25	143	1970	Graduation	Single	61299	0	0	2013-08-25	73	466	0	224	119	49	99	1	5	3	4	2
26	146	1960	PhD	Single	76045	0	0	2013-11-15	78	766	53	400	17	93	0	1	4	5	11	2
27	153	1953	Graduation	Divorced	23272	0	0	2013-10-17	29	19	3	6	4	2	12	1	1	0	3	4
28	158	1945	PhD	Together	71604	0	0	2013-11-17	3	345	53	528	98	75	97	1	8	3	5	4
29	164	1977	PhD	Married	28973	0	0	2012-10-13	59	206	0	46	3	2	12	2	5	1	5	8
30	175	1988	Graduation	Divorced	71952	1	0	2013-01-10	93	656	80	455	52	107	93	2	8	4	8	4
31	176	1980	2n Cycle	Single	67506	0	1	2014-01-02	73	90	38	67	33	38	67	1	4	1	6	3

Query executed successfully.

Laptop-FTBLIC2C (15.0 RTM) | LAPTOP-FTBLIC2C\CAMVAN... | MARKETING_WAREHOUSE | 00:00:00 | 2,240 rows

Ready Ln 4 Col 29 Ch 29 INS

Hình 100. TABLE ORIGINAL_DATA

SQLQuery1.sql - LAPTOP-FTBLIC2C\MARKETING_WAREHOUSE (LAPTOP-FTBLIC2C\CAMVAN (52))- Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Object Explorer Results Messages

CREATE DATABASE MARKETING_WAREHOUSE;

USE MARKETING_WAREHOUSE;

SELECT * FROM NULL_DATA;

Results

User_ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Date_Enroll	Recency	MinFives	MinFruits	MinMeals	Min_Fishes	Min_Sweets	Min_Golds	Num_DealsPur	Num_WebPur	Num_CatalogPur	Num_StorePur	Num_WebVisitsMonth	Avg	
1	1295	1983	Graduation	Married	NULL	0	1	2013-08-11	96	231	65	196	38	71	124	1	6	5	7	4	0
2	1812	1981	PhD	Single	NULL	1	0	2010-05-31	82	23	0	15	0	2	7	2	3	0	3	6	0
3	1948	1983	Graduation	Married	NULL	0	1	2013-06-01	11	6	5	6	0	2	1	1	0	2	7	0	
4	2477	1989	Graduation	Married	NULL	0	0	2013-06-03	69	661	138	461	60	30	61	1	6	5	12	3	0
5	2883	1970	Graduation	Single	NULL	1	2	2013-08-23	67	738	20	172	52	50	20	0	2	3	10	7	0
6	2902	1989	Graduation	Together	NULL	1	1	2013-09-13	87	19	4	12	2	6	1	1	0	3	5	0	
7	3117	1985	Graduation	Single	NULL	0	1	2013-10-19	95	264	0	21	12	6	147	3	6	1	5	7	0
8	3769	1972	PhD	Single	NULL	1	0	2014-03-02	17	25	1	13	0	0	3	1	1	0	3	7	0
9	4345	1964	2n Cycle	Single	NULL	1	1	2014-01-12	49	5	1	9	2	0	4	1	1	0	2	7	0
10	5079	1971	Graduation	Married	NULL	1	1	2013-03-03	82	71	1	16	0	0	9	4	2	1	3	8	0
11	5250	1943	Master	Widow	NULL	0	0	2013-10-19	75	532	126	490	164	126	1	5	5	11	1	0	
12	5255	1986	Graduation	Single	NULL	1	0	2013-02-20	19	5	1	3	3	263	362	0	27	0	1	0	
13	5798	1973	Master	Together	NULL	0	0	2013-11-23	87	445	37	359	98	28	18	1	2	4	8	1	0
14	7187	1989	Master	Single	NULL	1	1	2013-05-18	52	375	42	48	94	66	98	7	4	10	4	3	0
15	7244	1951	Graduation	Single	NULL	2	1	2014-01-01	96	48	5	48	6	10	7	3	2	1	4	6	0
16	7281	1959	PhD	Single	NULL	0	0	2013-11-08	80	81	11	50	3	2	39	1	1	3	4	2	0
17	8288	1981	PhD	Married	NULL	0	1	2013-07-11	23	352	0	27	10	0	15	3	6	1	7	6	0
18	8557	1982	Graduation	Single	NULL	1	0	2013-06-17	57	11	3	22	2	6	2	2	0	3	6	0	
19	8720	1978	2n Cycle	Together	NULL	0	0	2012-08-12	53	32	2	1607	12	4	22	0	0	0	1	0	0
20	8897	1987	PhD	Married	NULL	2	1	2013-11-19	4	230	42	192	49	37	53	12	7	2	8	9	0
21	1295	1987	Graduation	Single	NULL	1	1	2014-04-06	46	7	0	6	2	0	1	1	0	2	7	0	
22	1339	1954	Master	Together	NULL	0	1	2013-06-23	83	161	0	22	0	8	24	2	4	1	4	6	0
23	10475	1970	Master	Together	NULL	0	1	2013-04-01	39	187	5	65	26	20	14	2	4	2	6	5	0
24	10629	1973	2n Cycle	Married	NULL	1	0	2012-09-14	25	25	3	43	17	4	17	3	3	0	3	8	0

Query executed successfully.

Laptop-FTBLIC2C (15.0 RTM) | LAPTOP-FTBLIC2C\CAMVAN... | MARKETING_WAREHOUSE | 00:00:00 | 24 rows

Ready Ln 4 Col 25 Ch 25 INS

Hình 101. TABLE NULL_DATA

MARKETING_DATA

SQLQuery2.sql - LAPTOP-FTBLIC2C\MARKETING_WAREHOUSE (LAPTOP-FTBLIC2C\CAMVAN (52)) - Microsoft SQL Server Management Studio

```
CREATE DATABASE MARKETING_WAREHOUSE;
USE MARKETING_WAREHOUSE;
SELECT * FROM MARKETING_DATA;
```

Results Messages

User_ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Date_Enroll	Reency	MntWines	MntFruits	MntMeats	MntFishes	MntSweets	MntGolds	Num_DealPur	Num_WebPur	Num_CatalogPur	Num_StorePur	Num_WebVisitsMonth
0	1985	Graduation	Married	70951	0	0	2013-05-06	66	238	10	554	254	87	54	1	3	9	1	1
1	1961	Graduation	Single	57991	0	0	2014-06-15	0	484	5	64	7	0	37	1	7	3	5	5
3	1975	Master	Single	46098	1	1	2012-08-18	86	57	0	27	0	0	36	4	3	2	2	8
4	13	1947	PhD	25358	0	1	2013-09-22	57	19	0	5	0	0	8	2	1	0	3	6
5	17	1971	Married	60491	0	1	2013-09-01	91	637	47	237	12	19	76	4	6	11	7	5
6	20	1965	Zn Cycle	Married	46891	0	1	2013-09-01	91	43	12	23	20	15	61	1	2	1	4
7	22	1976	Graduation	Divorced	46310	1	0	2012-12-03	99	185	2	68	15	5	14	2	6	1	5
8	24	1969	Master	Together	17144	1	1	2014-10-15	98	18	2	19	0	2	6	5	3	0	7
9	25	1958	Graduation	Married	65149	0	1	2012-11-16	9	480	35	422	33	12	153	2	6	6	7
10	35	1987	Graduation	Married	25545	1	0	2012-09-08	35	32	1	64	16	12	85	3	2	2	8
11	48	1964	Graduation	Together	55781	0	1	2014-04-24	97	136	1	12	0	3	32	2	4	1	3
12	49	1970	Graduation	Single	20587	1	0	2014-05-11	39	2	3	6	4	1	9	1	1	1	2
13	55	1963	Graduation	Together	56253	0	1	2012-12-07	83	509	0	65	7	11	5	4	7	2	9
14	67	1972	Master	Single	46423	1	1	2013-09-18	6	68	0	16	0	0	8	3	2	0	4
15	73	1953	PhD	Single	51411	1	2	2013-05-05	81	14	0	3	0	0	1	0	0	3	5
16	75	1982	Zn Cycle	Together	41039	0	0	2013-07-26	32	34	11	137	179	61	45	1	8	1	5
17	78	1969	Graduation	Married	25293	1	0	2013-09-15	51	15	0	11	0	2	9	1	1	1	2
18	87	1981	Zn Cycle	Married	27733	1	0	2013-08-26	16	0	7	5	26	2	17	2	2	0	3
19	89	1975	Zn Cycle	Married	22212	1	0	2013-12-09	49	5	9	20	6	8	21	2	2	0	4
20	92	1988	Graduation	Alone	34176	1	0	2014-05-12	12	5	7	24	19	14	20	4	3	0	4
21	113	1951	Graduation	Married	30633	1	1	2013-07-18	33	11	0	7	3	1	6	2	1	0	3
22	115	1966	Master	Single	43456	0	1	2013-03-26	0	275	11	68	25	7	7	3	5	1	5
23	123	1951	Graduation	Widow	67048	0	1	2013-01-01	92	544	35	133	23	151	107	2	4	5	2
24	125	1960	Zn Cycle	Together	53083	1	1	2013-05-15	65	215	7	33	0	0	13	6	5	1	5
25	137	1970	Graduation	Divorced	61520	0	0	2013-09-10	70	465	0	224	119	49	59	1	5	3	4
26	146	1960	PhD	Single	76045	0	0	2013-11-15	78	760	53	400	17	93	60	1	4	5	11
27	153	1953	Graduation	Divorced	23272	0	0	2013-10-17	29	19	3	6	4	2	12	1	1	0	3
28	158	1945	PhD	Together	71604	0	0	2013-11-17	3	345	53	528	98	75	97	1	8	3	5
29	164	1977	PhD	Married	28973	0	0	2012-10-13	59	208	0	46	3	2	12	2	5	1	5
30	175	1986	Graduation	Married	71952	1	0	2013-01-10	93	656	80	455	52	107	93	2	8	4	8
31	176	1980	Zn Cycle	Single	67506	0	1	2014-01-02	73	90	38	67	33	36	67	1	4	1	6

Query executed successfully.

Hình 102. TABLE MARKETING_DATA

SQLQuery2.sql - LAPTOP-FTBLIC2C\MARKETING_WAREHOUSE (LAPTOP-FTBLIC2C\CAMVAN (52)) - Microsoft SQL Server Management Studio

```
CREATE DATABASE MARKETING_WAREHOUSE;
USE MARKETING_WAREHOUSE;
SELECT * FROM DIM_CUSTOMER;
```

Results Messages

User_ID	Year_Birth	Marital_Status	Education	Income	Kidhome	Teenhome
0	1985	Married	Graduation	70951	0	0
1	1961	Single	Graduation	57991	0	0
3	1975	Single	Master	46098	1	1
4	13	1947	Widow	PhD	25358	0
5	17	1971	Married	PhD	60491	0
6	20	1965	Zn Cycle	Married	46891	0
7	22	1976	Divorced	Graduation	46310	1
8	24	1960	Together	Master	17144	1
9	25	1958	Married	Graduation	65148	0
10	35	1987	Married	Graduation	25545	1
11	48	1964	Together	Graduation	55781	0
12	49	1970	Single	Graduation	20587	0
13	55	1953	Together	Graduation	61299	0
14	57	1972	Single	Married	46423	1
15	73	1953	Zn Cycle	PhD	51411	1
16	75	1982	Together	Zn Cycle	25533	0
17	78	1969	Married	Graduation	25293	1
18	87	1981	Zn Cycle	Married	27733	1
19	89	1975	Married	Zn Cycle	22212	1
20	92	1988	Alone	Graduation	34176	0
21	113	1951	Married	Graduation	30933	1
22	115	1966	Single	Master	43456	0
23	123	1951	Widow	Graduation	67048	0
24	125	1958	Zn Cycle	Together	53083	1
25	143	1970	Single	Graduation	61299	0
26	146	1960	Single	PhD	76045	0
27	153	1953	Divorced	Graduation	23272	0
28	158	1945	Together	PhD	71604	0
29	164	1977	Married	PhD	28973	0
30	175	1986	Married	Graduation	71952	1
31	176	1980	Single	Zn Cycle	67506	0
32	178	1956	Married	Graduation	62503	1

Query executed successfully.

Hình 103. TABLE DIM_CUSTOMER

DIM_TIME

File Edit View Query Project Tools Window Help

Object Explorer | MARKETING_WAREHOUSE | Connect | SQLQuery1.sql - LAPTOP-FTBLIC2C\MARKETING_WAREHOUSE (LAPTOP-FTBLIC2C\CAMVAN (52)) - Microsoft SQL Server Management Studio

SQLQuery2.sql - L:\LIC2C\CAMVAN (73) | SQLQuery1.sql - L:\LIC2C\CAMVAN (52)*

```
CREATE DATABASE MARKETING_WAREHOUSE;

USE MARKETING_WAREHOUSE;
SELECT * FROM DIM_TIME;
```

Results Messages

Date_Enroll	Day_Enroll	Month_Enroll	Year_Enroll
2012-07-30	30	7	2012
2012-07-31	31	7	2012
2012-08-01	1	8	2012
2012-08-02	2	8	2012
2012-08-03	3	8	2012
2012-08-04	4	8	2012
2012-08-05	5	8	2012
2012-08-06	6	8	2012
2012-08-07	7	8	2012
2012-08-08	8	8	2012
2012-08-09	9	8	2012
2012-08-10	10	8	2012
2012-08-11	11	8	2012
2012-08-12	12	8	2012
2012-08-13	13	8	2012
2012-08-14	14	8	2012
2012-08-15	15	8	2012
2012-08-16	16	8	2012
2012-08-17	17	8	2012
2012-08-18	18	8	2012
2012-08-19	19	8	2012
2012-08-20	20	8	2012
2012-08-21	21	8	2012
2012-08-22	22	8	2012
2012-08-23	23	8	2012
2012-08-24	24	8	2012
2012-08-25	25	8	2012
2012-08-26	26	8	2012
2012-08-27	27	8	2012
2012-08-28	28	8	2012
2012-08-29	29	8	2012
2012-08-30	30	8	2012

Query executed successfully.

Ln 4 Col 1 Ch 1 INS

Hinh 104. DIM_TIME

DIM_COUNTRY

File Edit View Query Project Tools Window Help

Object Explorer | MARKETING_WAREHOUSE | Connect | SQLQuery1.sql - LAPTOP-FTBLIC2C\MARKETING_WAREHOUSE (LAPTOP-FTBLIC2C\CAMVAN (52)) - Microsoft SQL Server Management Studio

SQLQuery2.sql - L:\LIC2C\CAMVAN (73) | SQLQuery1.sql - L:\LIC2C\CAMVAN (52)*

```
CREATE DATABASE MARKETING_WAREHOUSE;

USE MARKETING_WAREHOUSE;
SELECT * FROM DIM_COUNTRY;
```

Results Messages

Country_ID	Country
1	Australia
2	Australia
3	Australia
4	Australia
5	Australia
6	Australia
7	Australia
8	Australia
9	Australia
10	Australia
11	Australia
12	Australia
13	Australia
14	Australia
15	Australia
16	Australia
17	Australia
18	Australia
19	Australia
20	Australia
21	Australia
22	Australia
23	Australia
24	Australia
25	Australia
26	Australia
27	Australia
28	Australia
29	Australia
30	Australia
31	Australia
32	Australia

Query executed successfully.

Ln 5 Col 2 INS

Hinh 105. DIM_COUNTRY

FACT_MARKETING

SQLQuery1.sql - LAPTOP-FTBLIC2C\MARKETING_WAREHOUSE (LAPTOP-FTBLIC2C\CAMVAN (81)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

MARKETING_WAREHOUSE

Object Explorer

Connect -> MARKETING_WAREHOUSE

SQLQuery1.sql - L:\LIC2C\CAMVAN (81)* - Microsoft SQL Server Management Studio

Execute

SELECT * FROM DIM_TIME;
SELECT * FROM FACT_MARKETING;

Results Messages

FactID	User_ID	Date_ID	Country_ID	Num_DealPur	Num_WebPur	Num_CatalogPur	Num_StorePur	Num_WebVisitsMonth	Complain	Response	Recency	Total_campaigns_accepted	Total_Spent
1	4472	0	2013-05-04	13297	1	3	4	9	1	0	66	0	1198
2	4473	1	2014-06-15	13456	1	7	3	7	5	0	1	0	577
3	4474	9	2012-08-18	14321	4	3	2	2	8	0	0	86	0
4	4475	13	2013-07-22	13987	2	1	0	3	6	0	0	57	0
5	4476	17	2013-09-06	14321	4	6	11	7	5	0	0	81	0
6	4477	20	2013-09-09	13456	1	2	1	4	4	0	0	91	0
7	4478	22	2012-12-03	13938	2	6	1	5	8	0	0	99	0
8	4479	24	2013-01-19	13456	5	3	0	4	7	0	0	96	0
9	4480	25	2012-11-16	14321	2	6	6	7	4	0	0	9	0
10	4481	35	2013-09-09	13987	3	2	2	3	6	0	1	35	0
11	4482	48	2014-04-20	14321	2	4	1	3	6	0	0	97	1
12	4483	49	2014-05-11	14321	1	1	1	2	7	0	0	39	0
13	4484	55	2012-12-07	14321	4	7	2	9	6	0	0	83	0
14	4485	67	2013-09-18	14321	3	2	0	4	7	0	0	6	0
15	4486	73	2013-05-26	14321	1	0	0	3	5	0	0	81	0
16	4487	75	2013-07-25	14321	1	8	1	5	8	0	0	32	0
17	4488	78	2013-09-15	13719	1	1	1	2	8	0	0	51	0
18	4489	87	2013-08-26	13987	2	2	0	3	7	0	0	16	0
19	4490	89	2013-12-09	14321	2	2	0	4	6	0	0	49	0
20	4491	92	2014-05-12	14321	4	3	0	4	6	0	0	12	0
21	4492	113	2013-07-16	13456	2	1	0	3	5	0	0	33	0
22	4493	115	2013-03-26	14321	3	5	1	8	5	0	0	0	393
23	4494	123	2013-01-10	14321	2	4	5	4	2	0	0	92	0
24	4495	125	2013-05-19	13456	6	5	1	5	7	0	0	65	1
25	4496	143	2013-09-05	13987	1	5	3	4	2	0	0	73	0
26	4497	146	2013-11-15	14321	1	4	5	11	2	0	0	78	1
27	4498	153	2013-10-17	14321	1	1	0	3	4	0	0	29	0
28	4499	158	2013-11-17	13719	1	8	3	5	4	0	1	3	1
29	4500	164	2012-10-13	13456	2	5	1	5	8	0	0	59	0
30	4501	175	2013-01-10	14321	2	8	4	8	4	0	0	93	2
31	4502	176	2014-01-02	14321	1	4	1	6	3	0	0	73	0
32	4503	178	2013-02-18	13987	4	8	3	11	5	0	0	40	0
33	4504	182	2013-05-11	13456	4	6	2	5	8	0	0	62	0
34	4505	193	2014-02-17	13719	1	1	0	2	5	0	0	81	1
35	4506	194	2014-06-09	13987	1	1	0	3	6	0	0	55	0
36	4507	195	2012-08-26	14321	4	5	1	4	8	0	1	21	1

Query executed successfully.

LAPTOP-FTBLIC2C (15.0 RTM) | LAPTOP-FTBLIC2C\CAMVAN... | MARKETING_WAREHOUSE | 00:00:00 | 2,216 rows

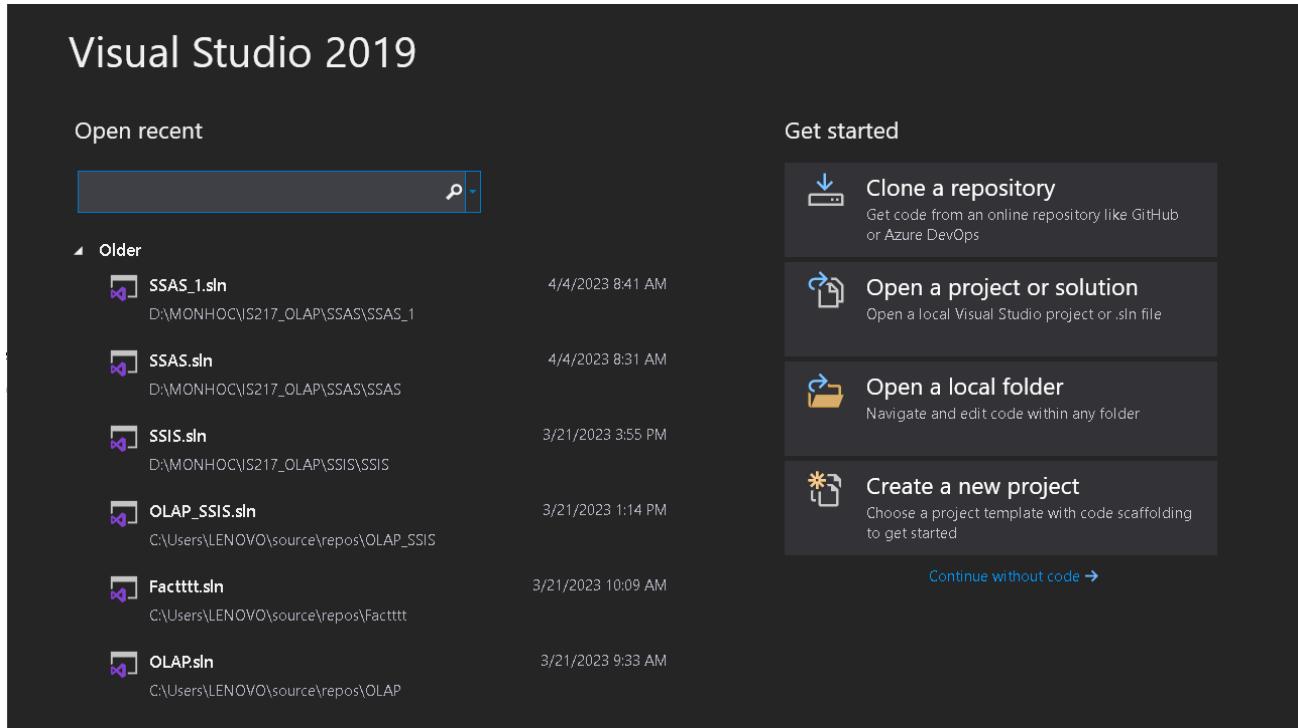
Hình 106. TABLE FACT_MARKETING

CHƯƠNG 3: PHÂN TÍCH DỮ LIỆU TRỰC TUYẾN – QUÁ TRÌNH SSAS

3.1 Tạo project SSAS và thiết lập kết nối

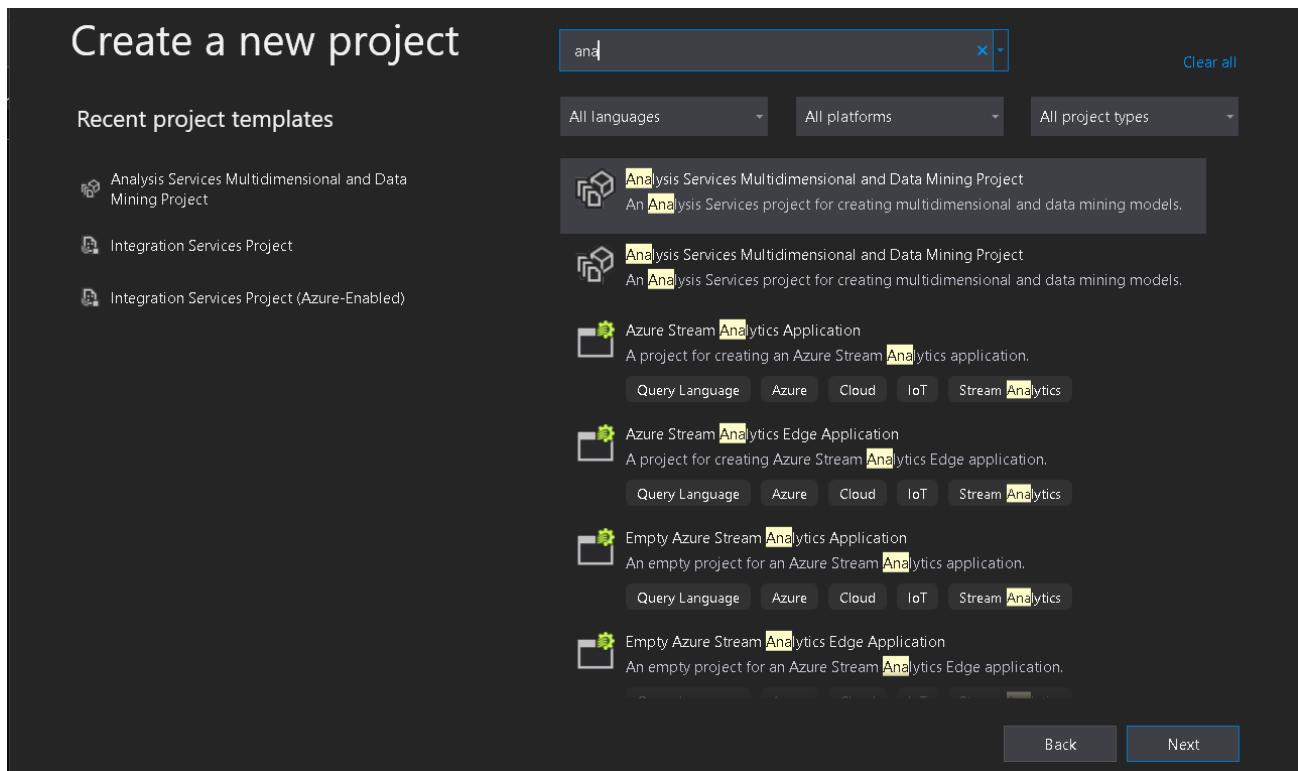
3.1.1 Tạo project trong Visual Studio

Mở visual studio 2019 -> Create a new project

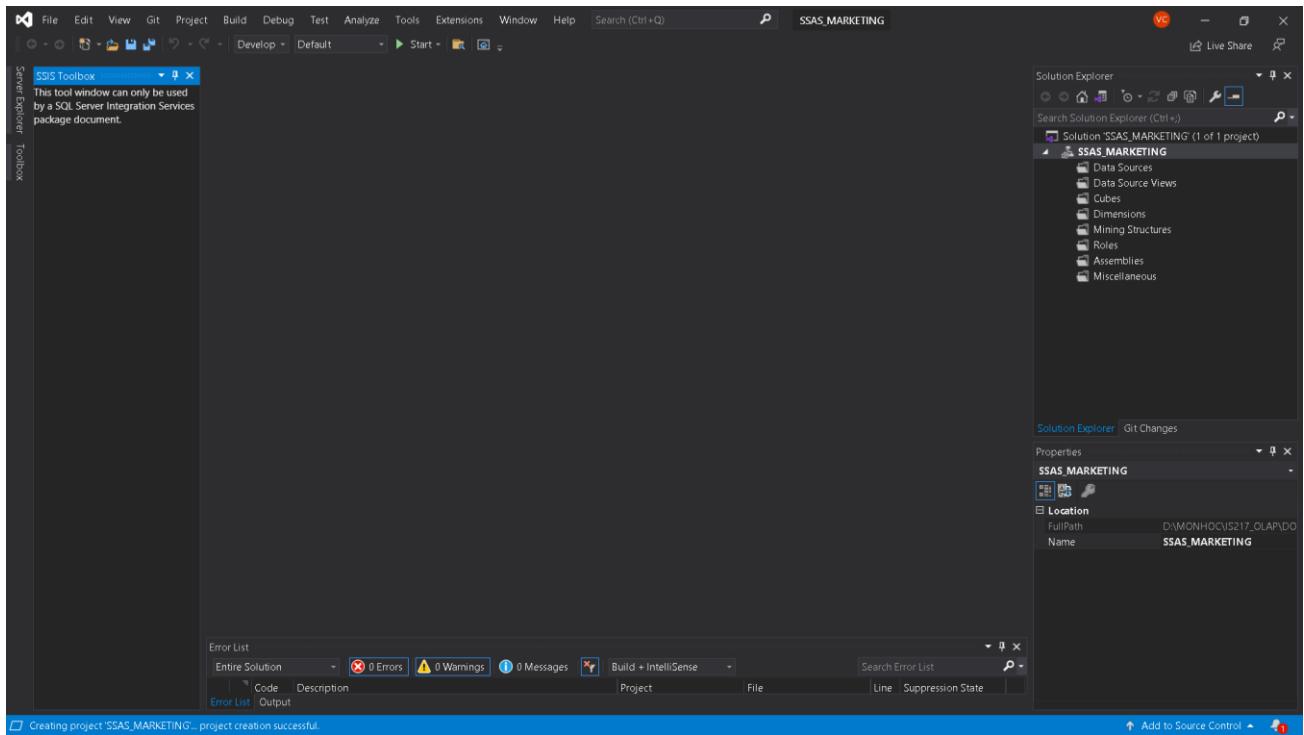


Hình 107. Visual Studio 2019

Chọn **Analysis Service Multidimensional and Data Mining Project** -> Click chọn và nhấn **Next**



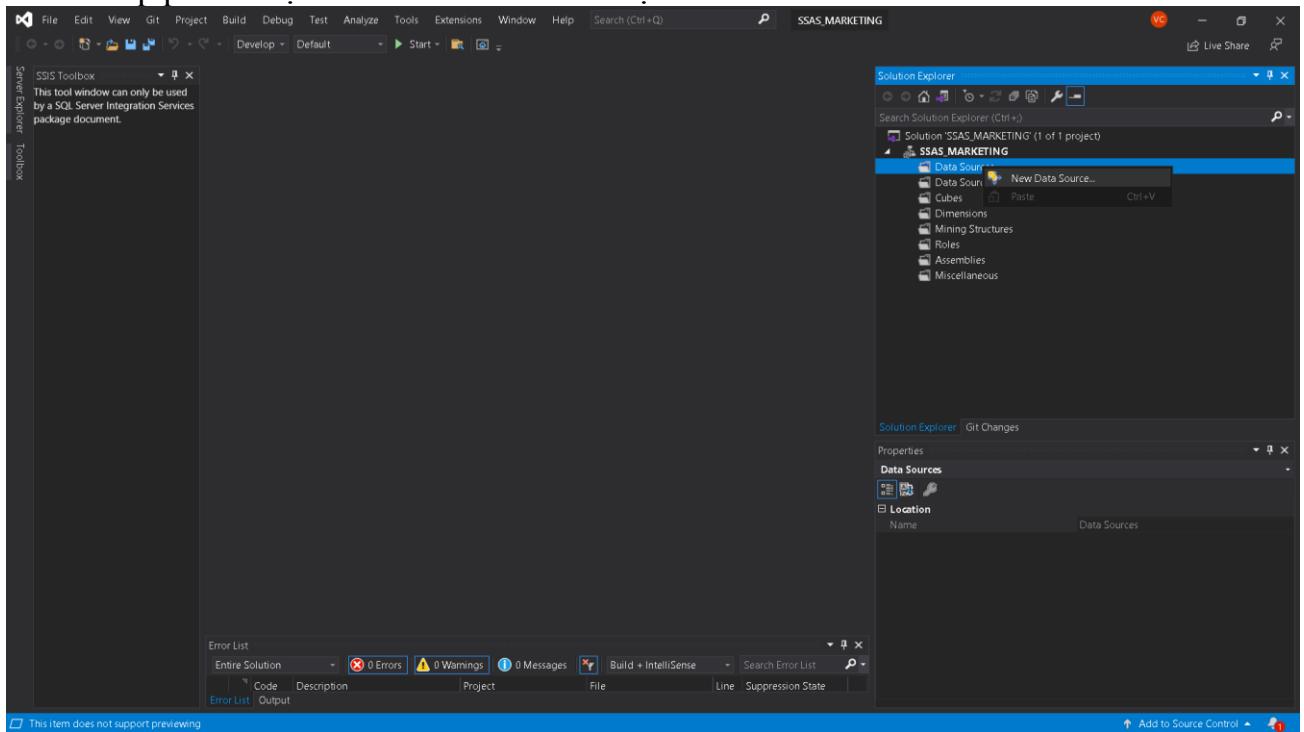
Hình 108. Cấu hình Project



Hình 109. Giao diện SSAS

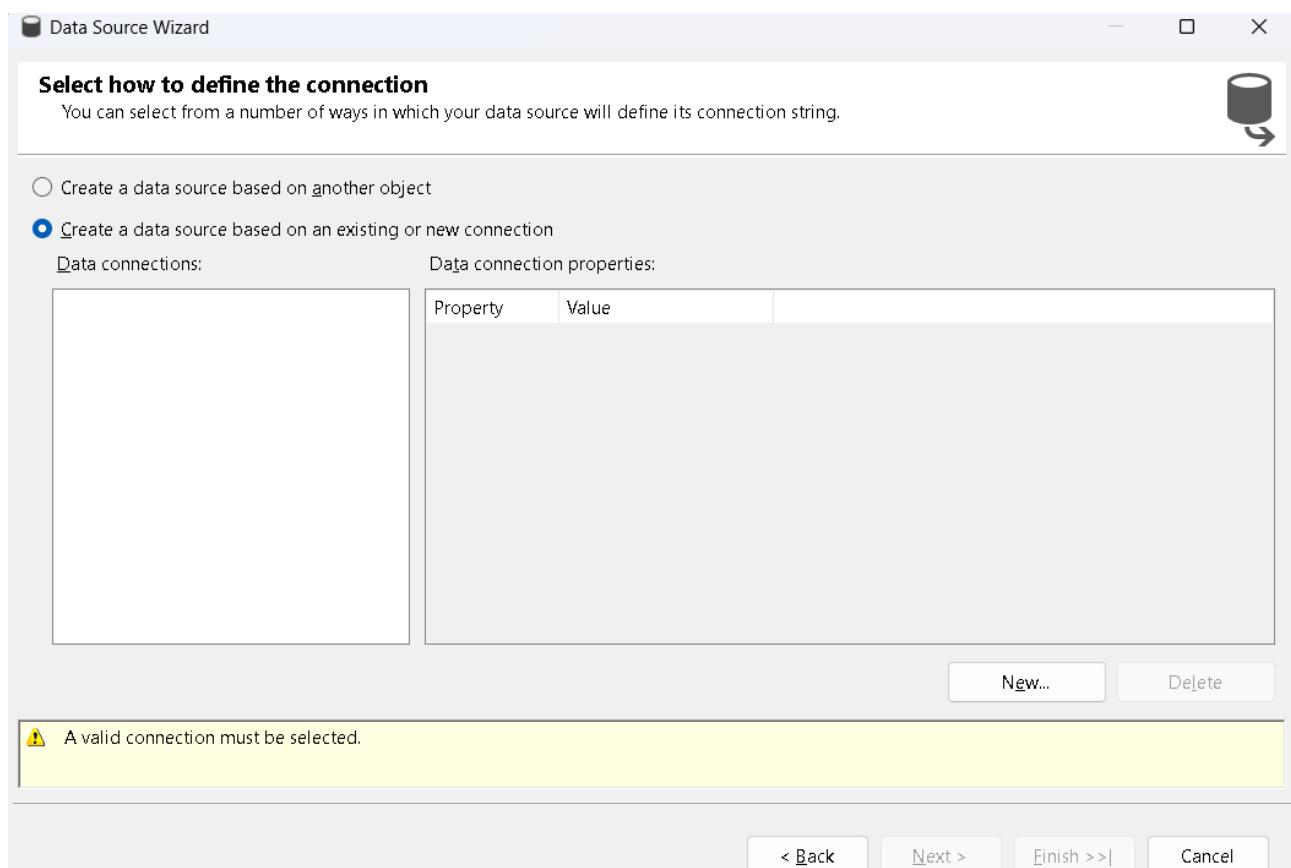
3.1.2 Xác định Data Source

Nhấp phải chuột vào Data Source -> Chọn New Data Source



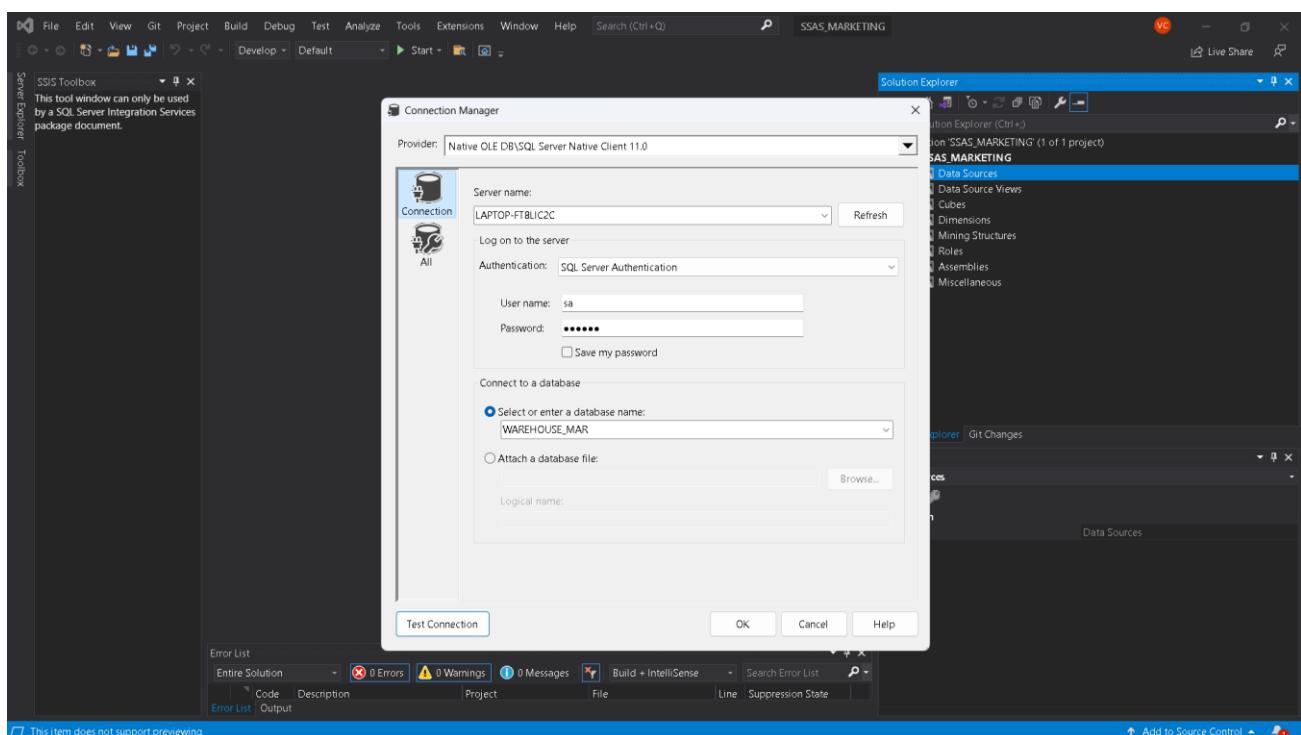
Hình 110. Chọn New Data Source

Chọn Create a data source based on an existing or new connection



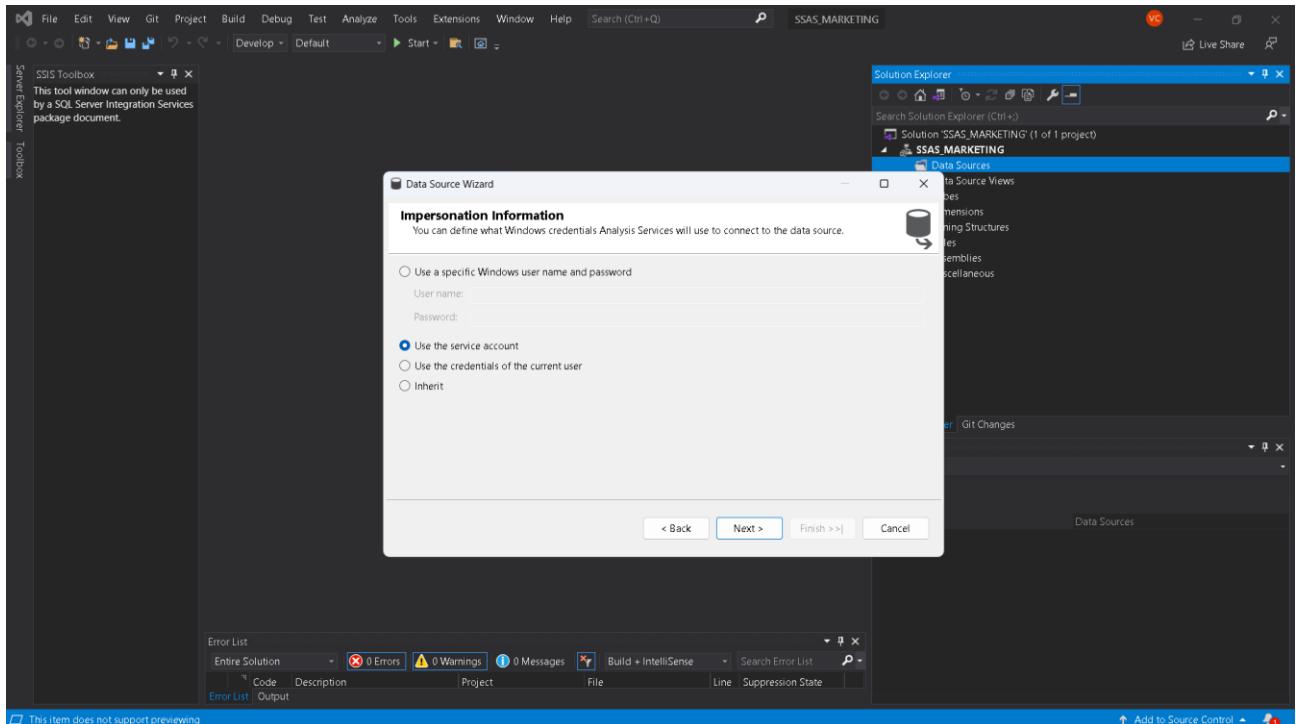
Hình 111. Chọn Connection

Chọn Database đã tạo từ tiến trình SSIS -> Next



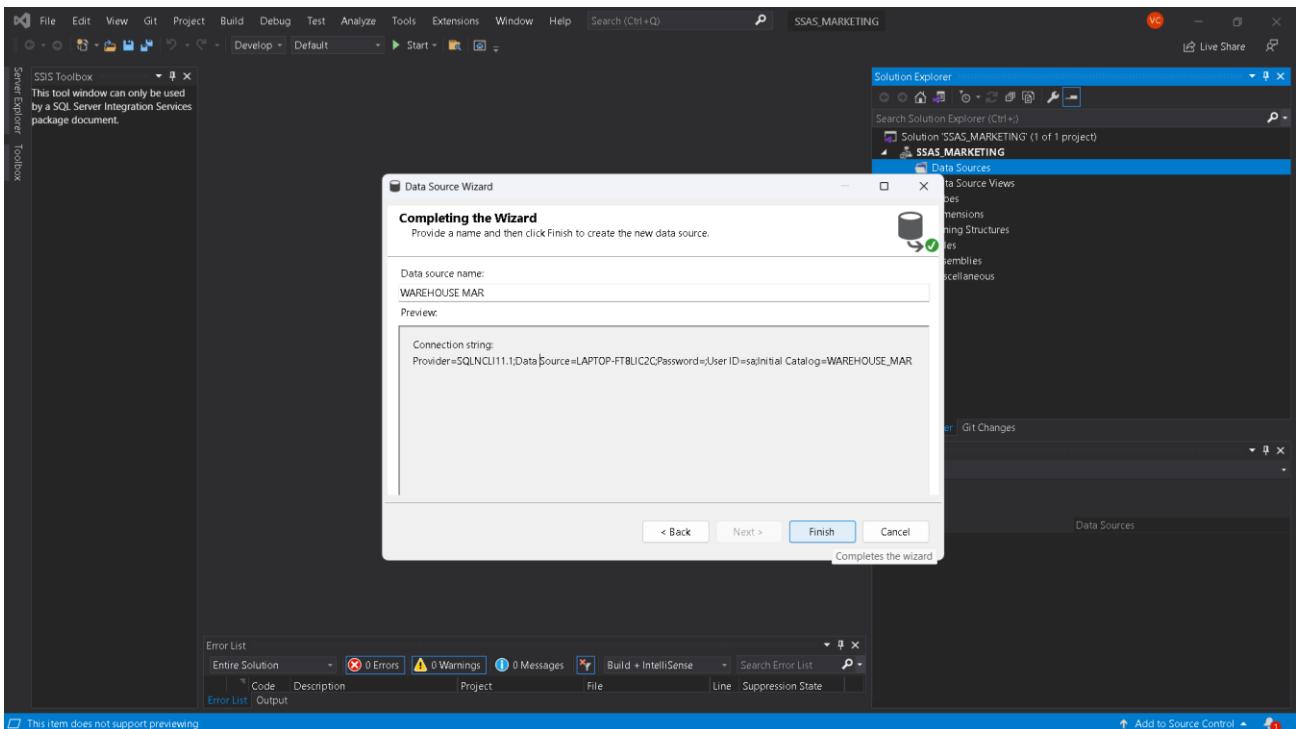
Hình 112. Chọn Database từ tiến trình SSIS

Chọn Use the service account -> Next



Hình 113. Chọn tài khoản User

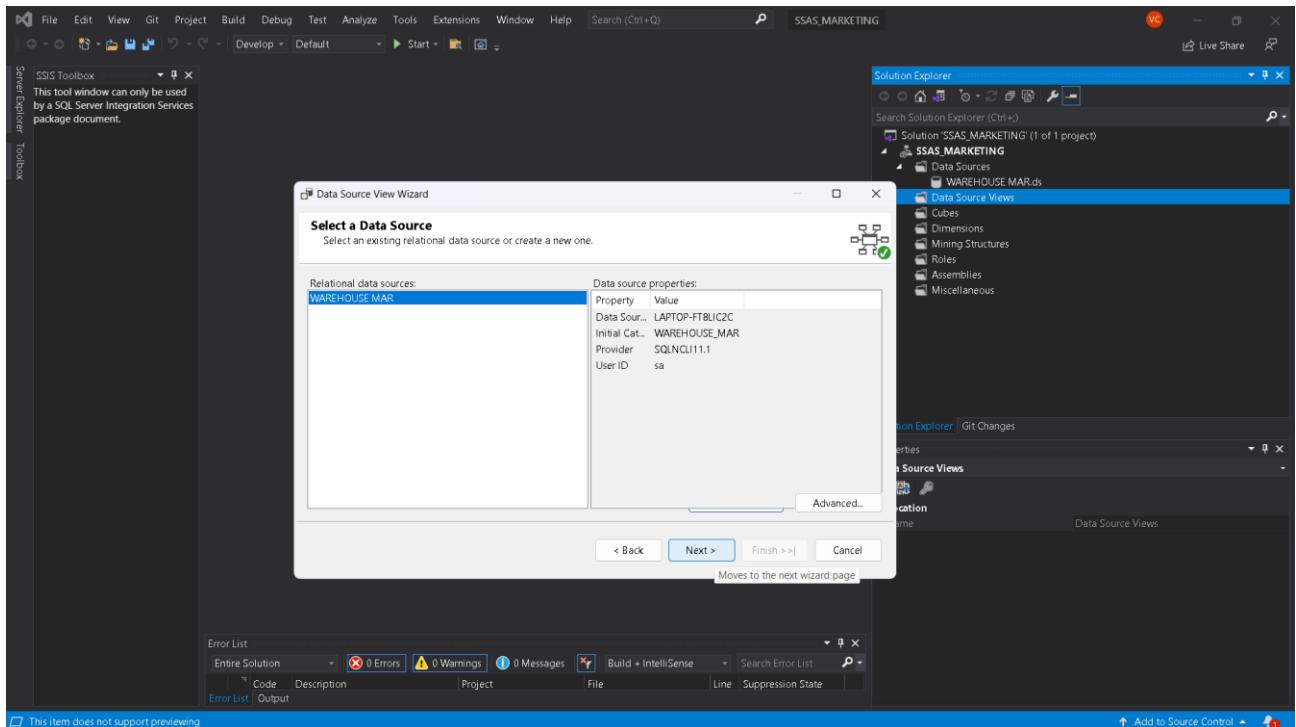
Thêm Data Source thành công -> Nhấn Finish



Hình 114. Connection thành công

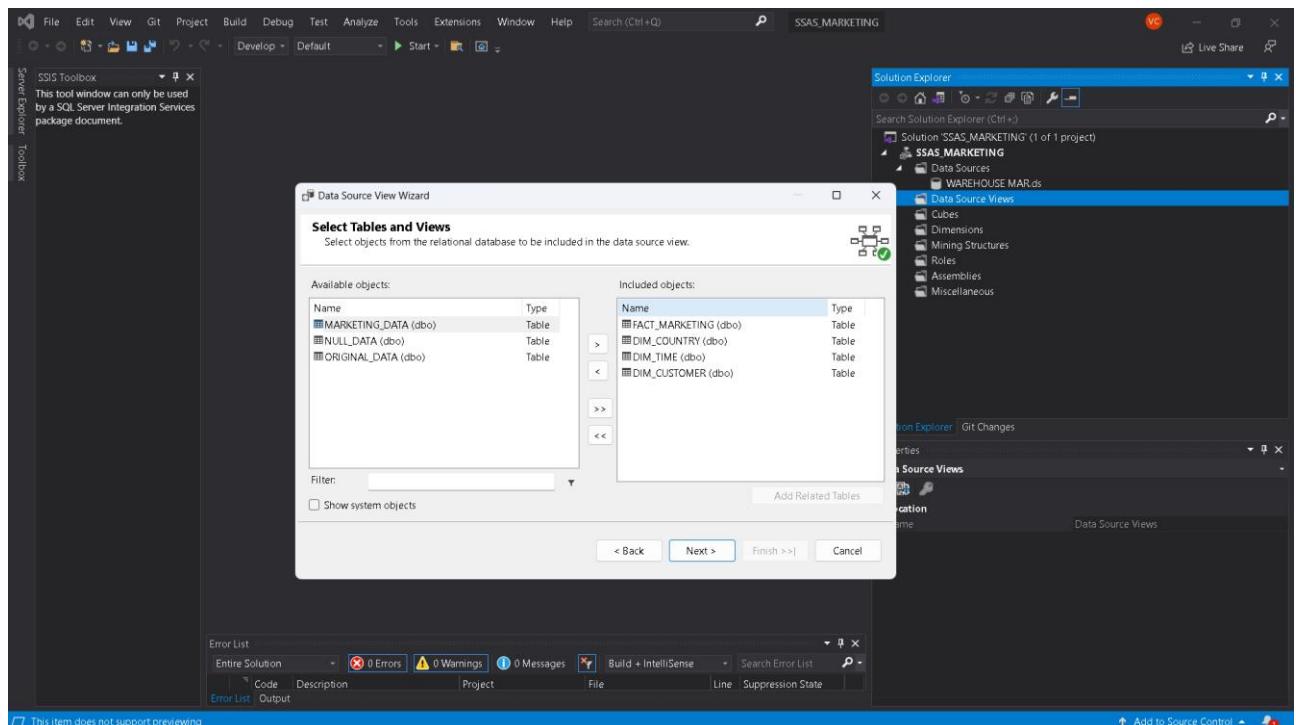
3.1.3 Xác định Data Source Views

Nhấp phải chuột vào Data Source Views -> Chọn New Data Source Views



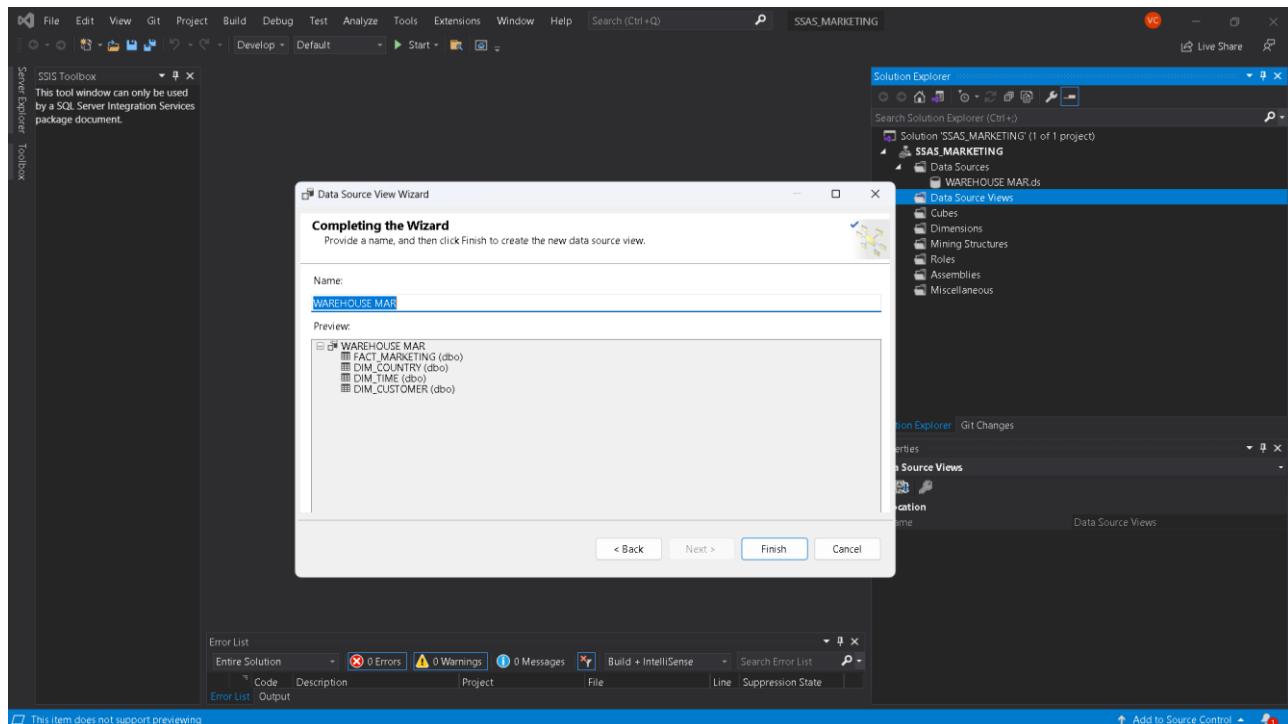
Hình 115. Chọn New Data Source Views

Chọn bảng FACT và bảng DIM cho quá trình phân tích



Hình 116. Chọn bảng FACT và DIM

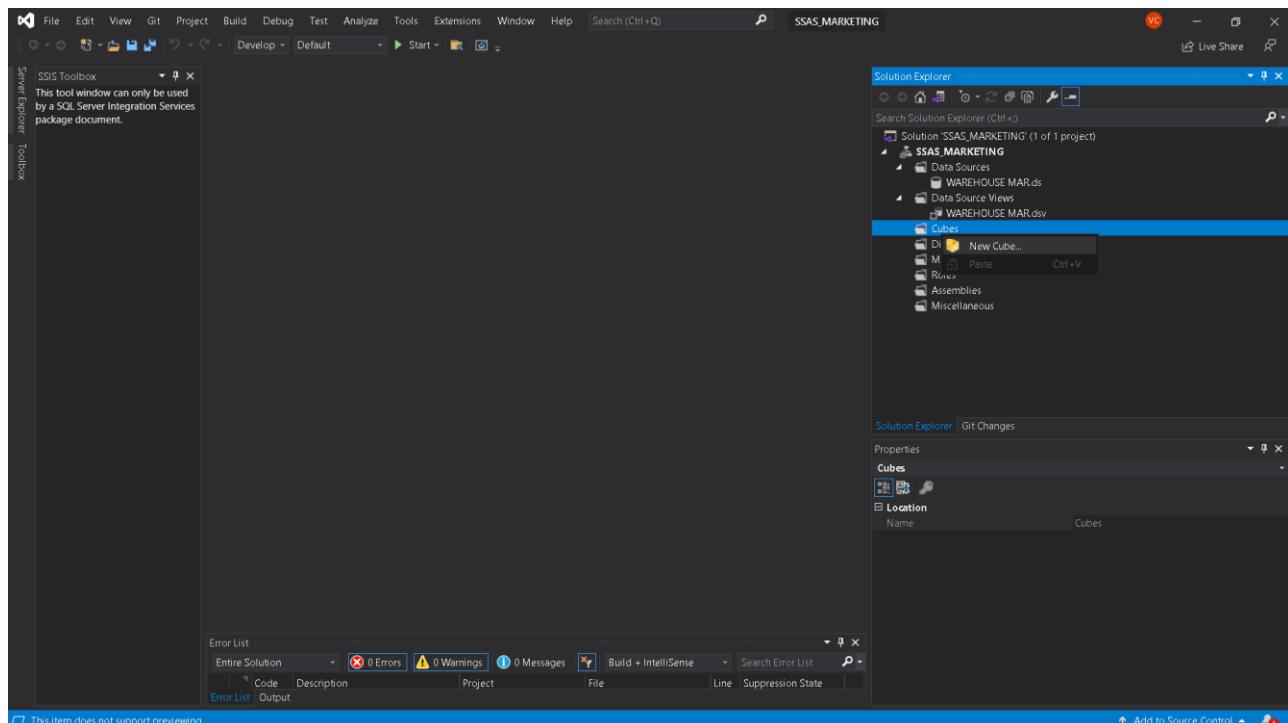
Nhấn Finish để kết thúc quá trình tạo Data Source Views



Hình 117. Kết thúc quá trình tạo New Data Source Views

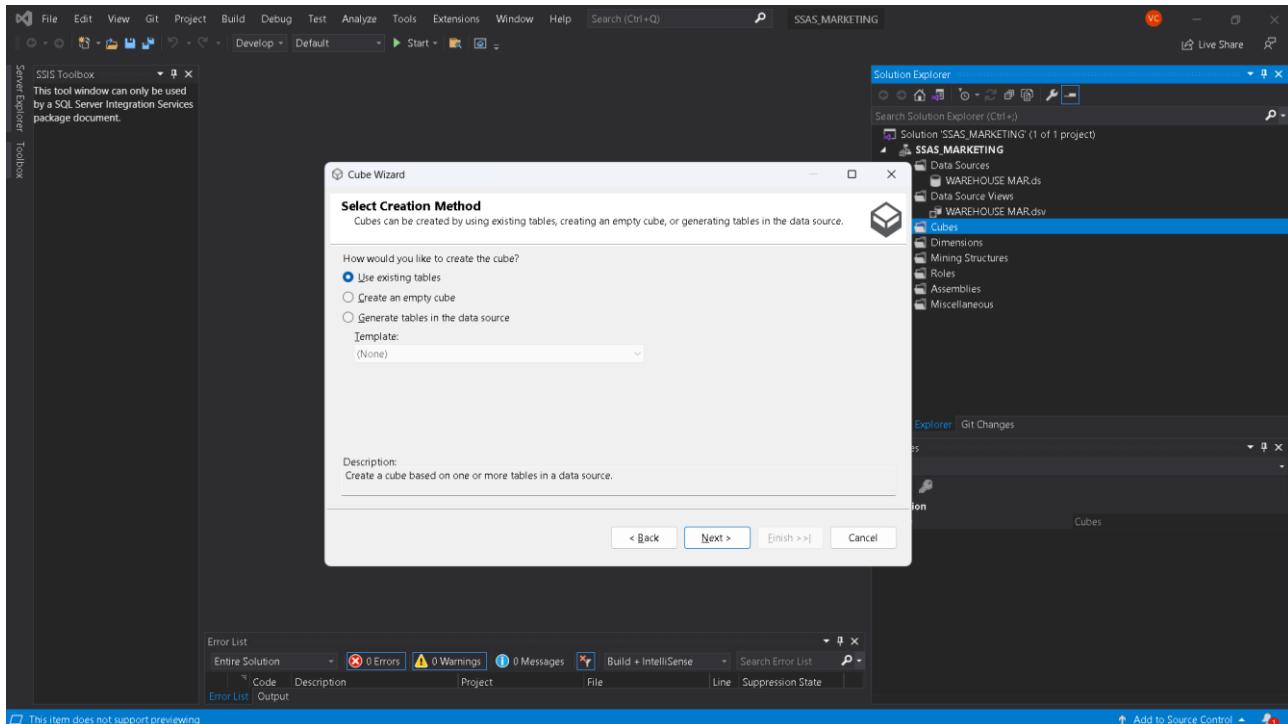
3.1.4 Xây dựng các Cube và các Measures

Nhấp phải chuột vào Cube -> Chọn New Cube



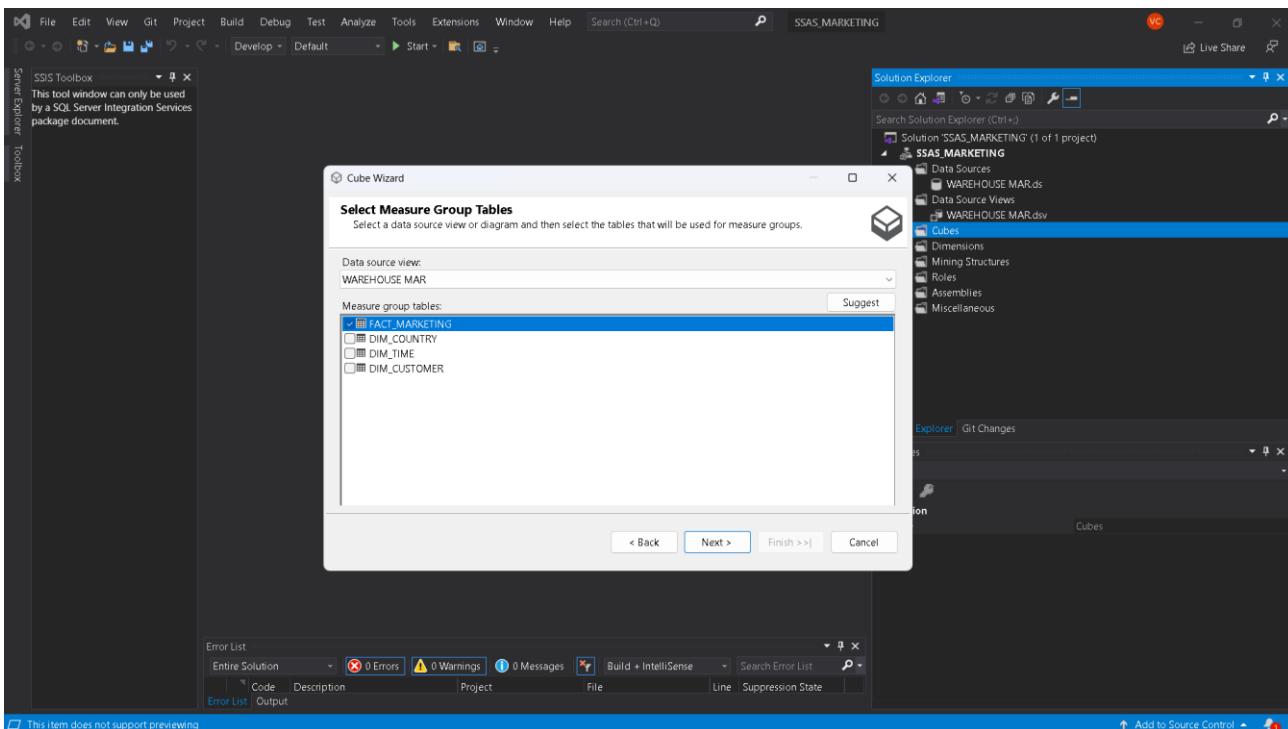
Hình 118. Chọn New Cube

Chọn Use existing tables -> Nhấn Next

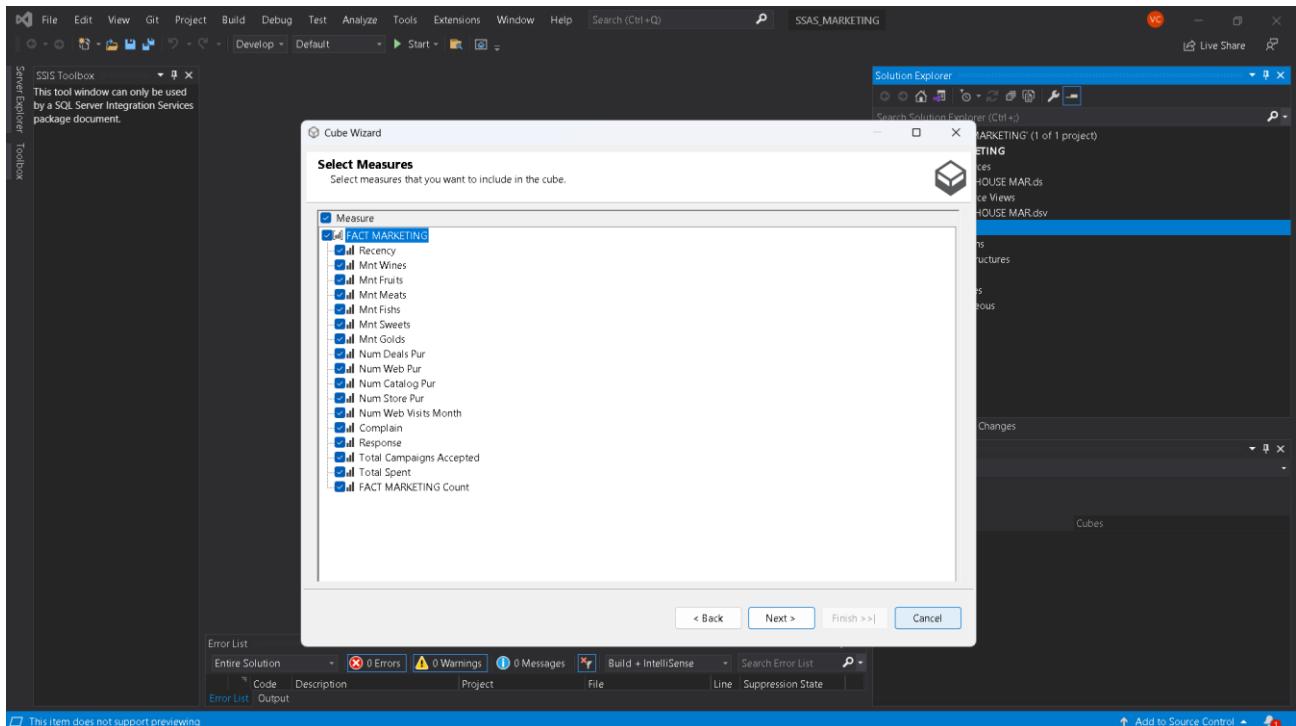


Hình 119. Chọn Use existing tables

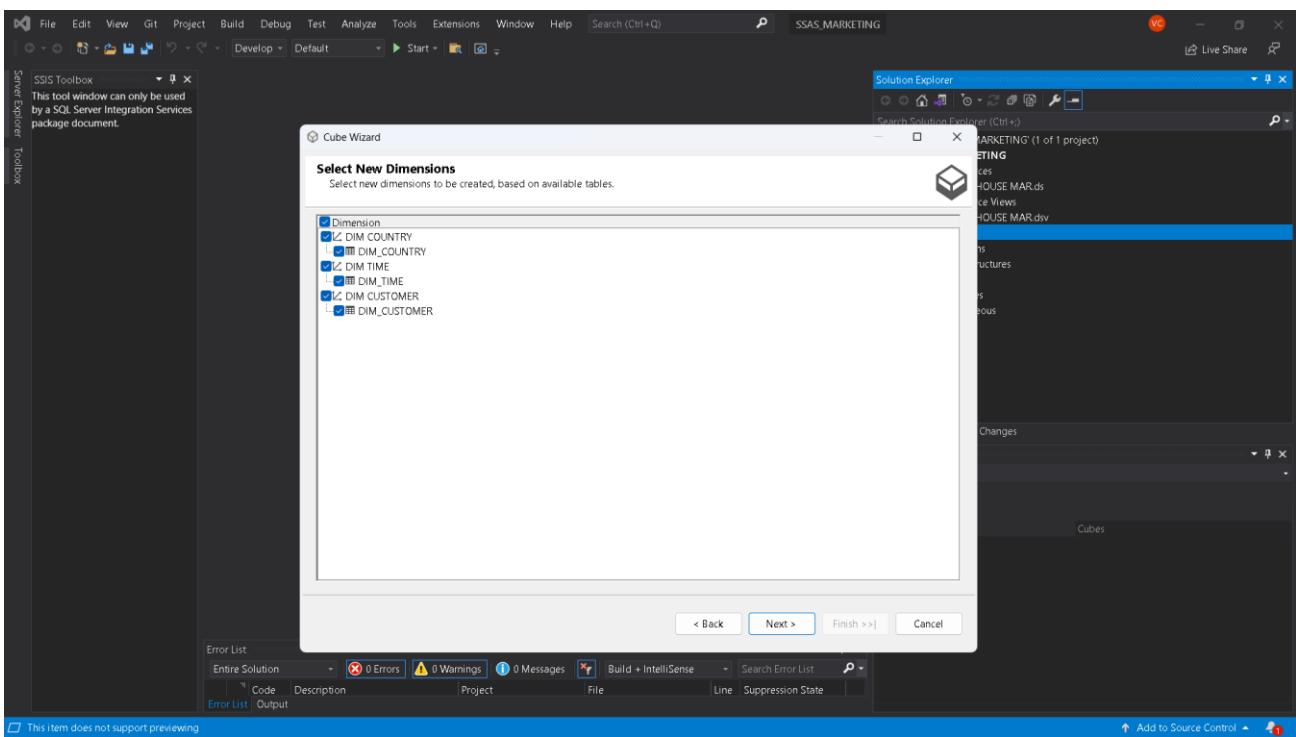
Click chọn bảng FACT -> Nhấn Next



Hình 120. Chọn bảng FACT

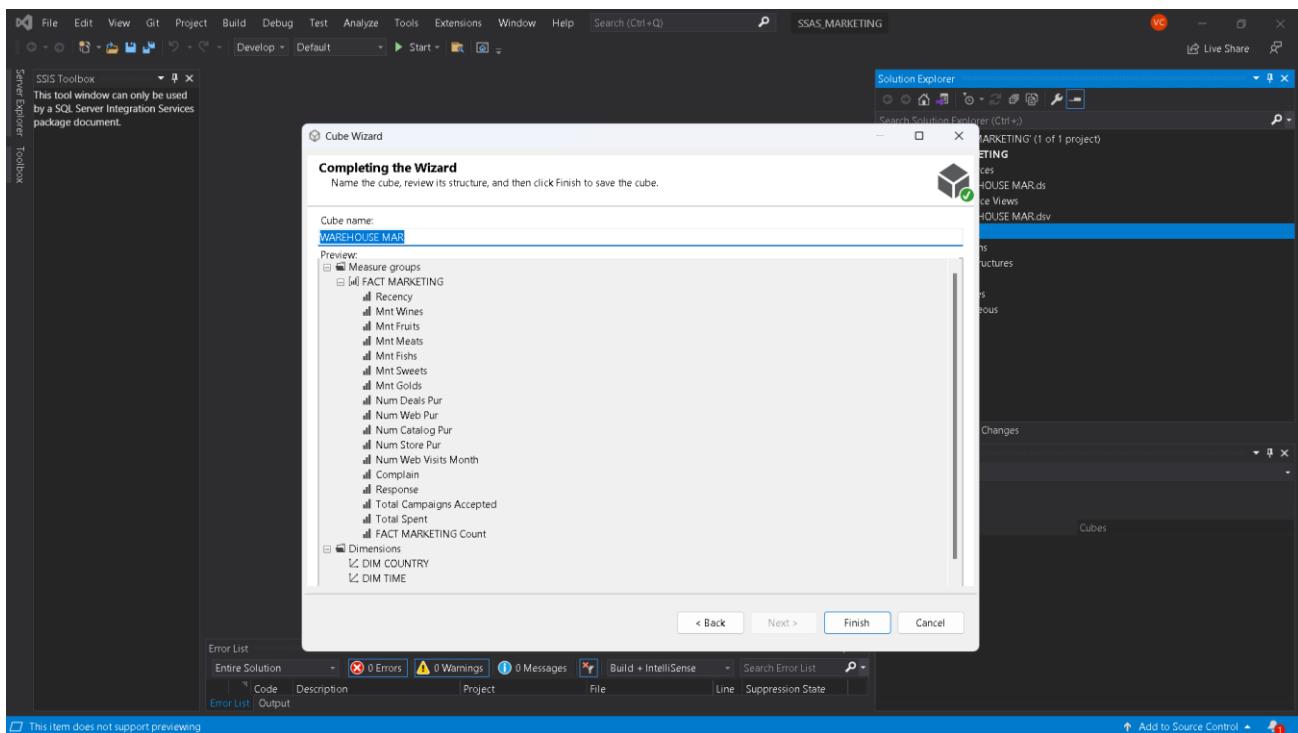


Hình 121. Chọn các Measures

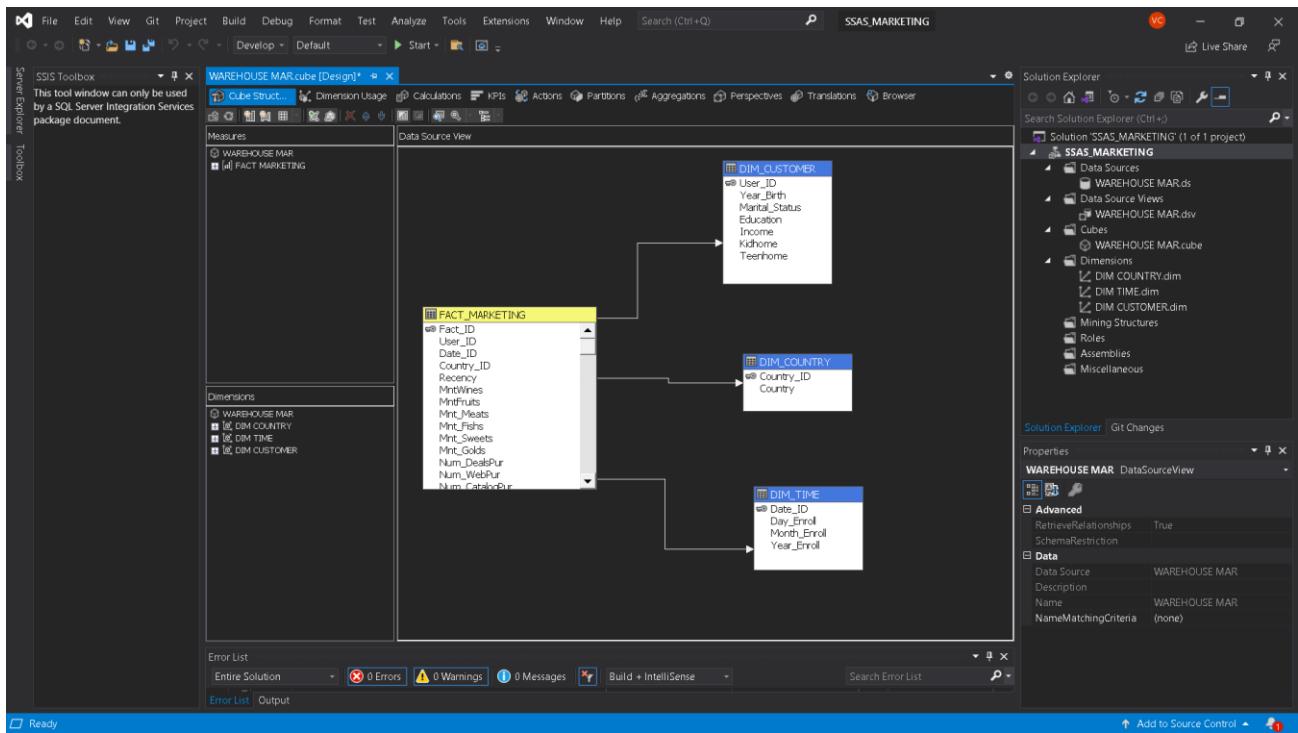


Hình 122. Chọn các Dimentions

Nhấn Finish để hoàn thành quá trình tạo Cube



Hình 123. Hoàn thành tạo Cube

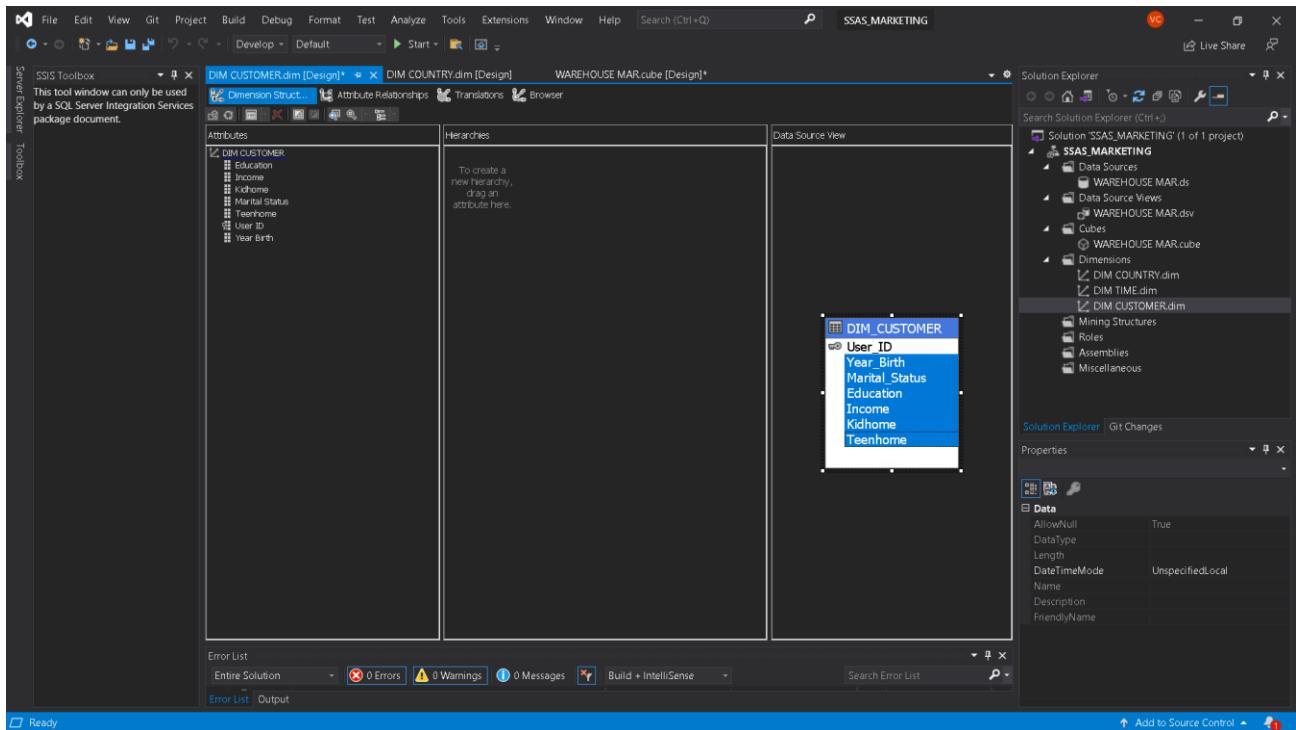


Hình 124. Cấu trúc của Cube sau khi tạo

3.1.5 Phân cấp trong bảng Dimension

3.1.5.1 Bảng Dim_Customer

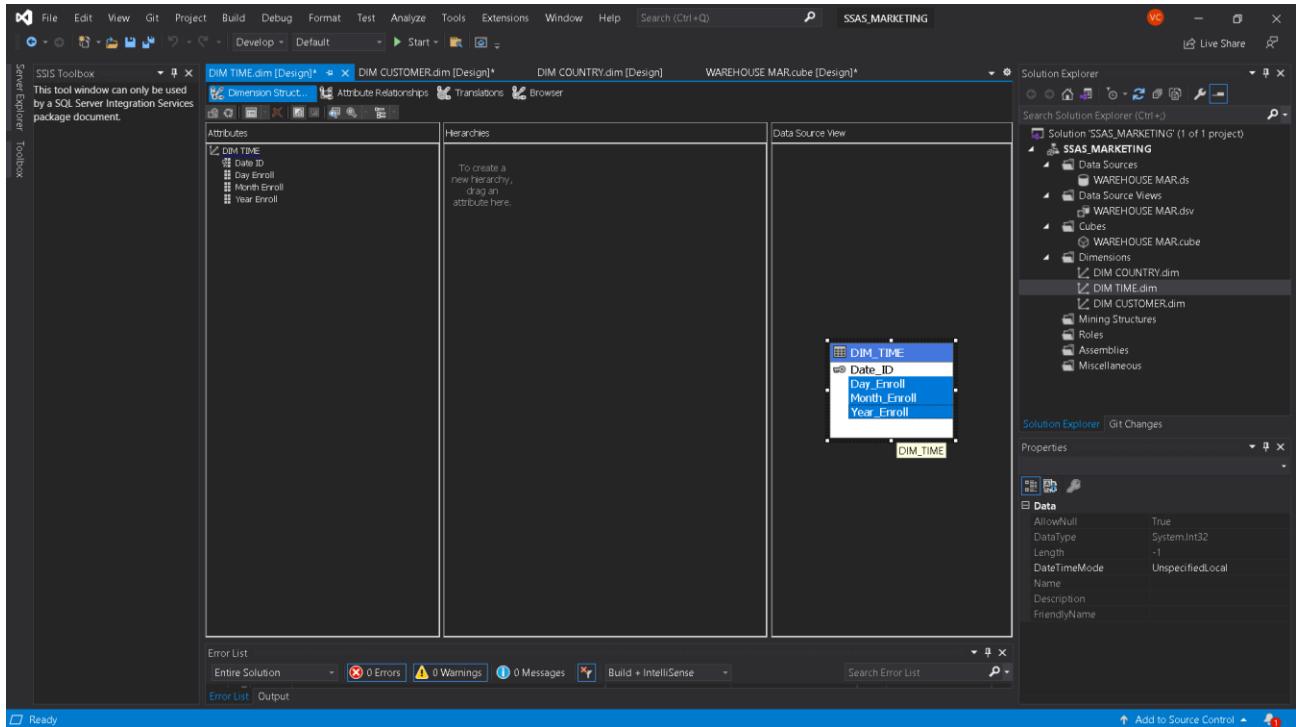
Kéo những thuộc tính cần phân cấp qua cửa sổ hierachies



Hình 125. DIM_CUSTOMER

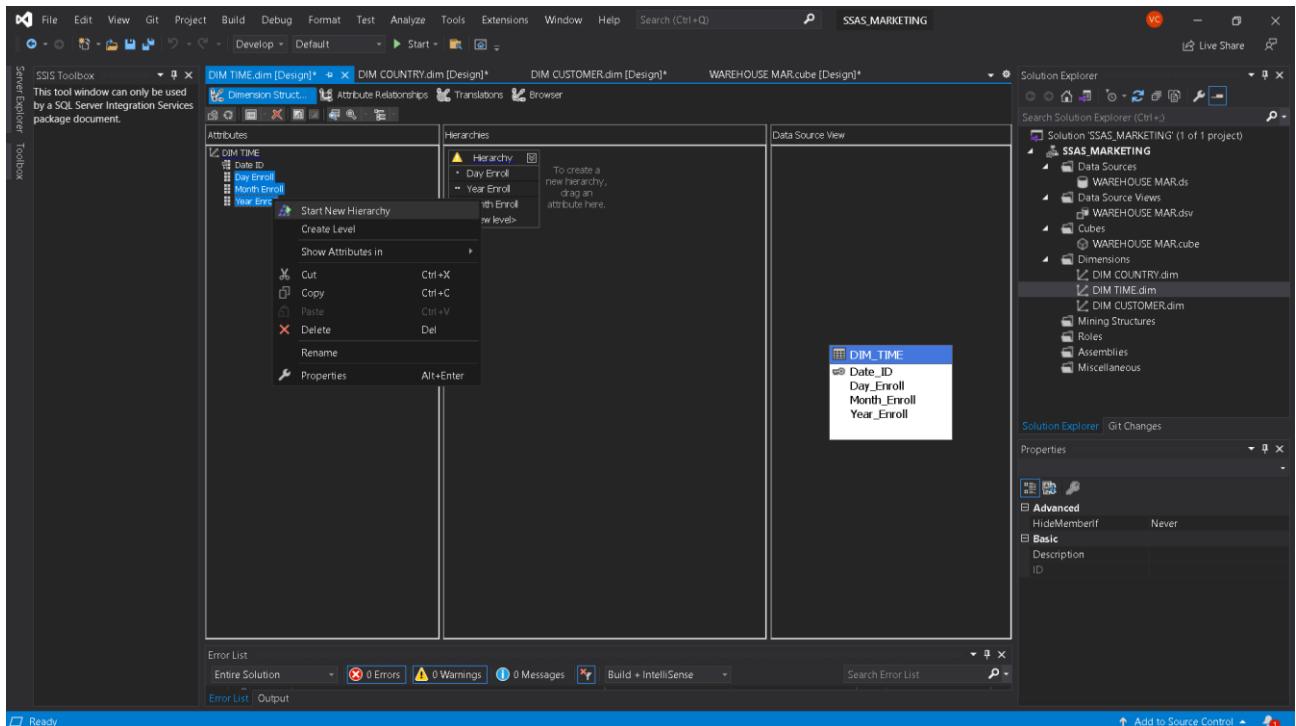
3.1.5.2 Bảng Dim_Time

Kéo những thuộc tính cần phân cấp qua cửa sổ hierachies



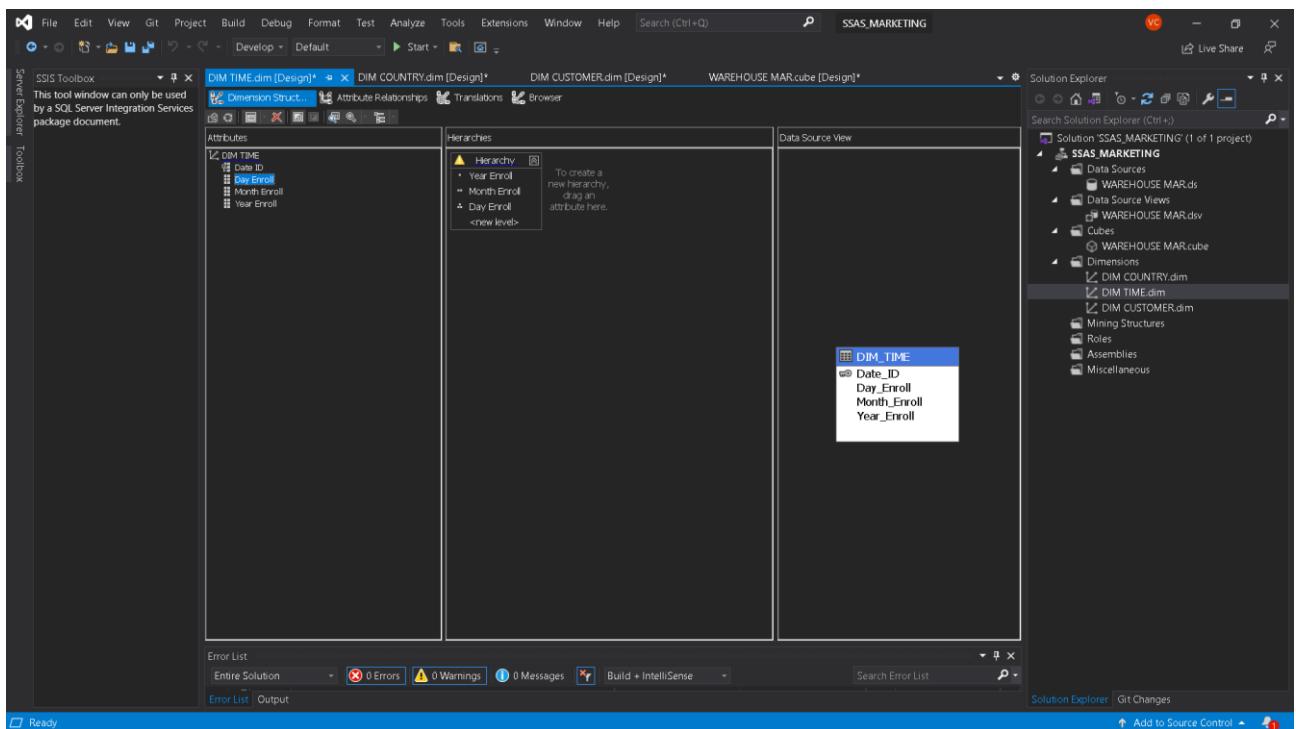
Hình 126. DIM_TIME

Nhấp chuột phải vào bảng Attributes -> Chọn Start New Hierarchy



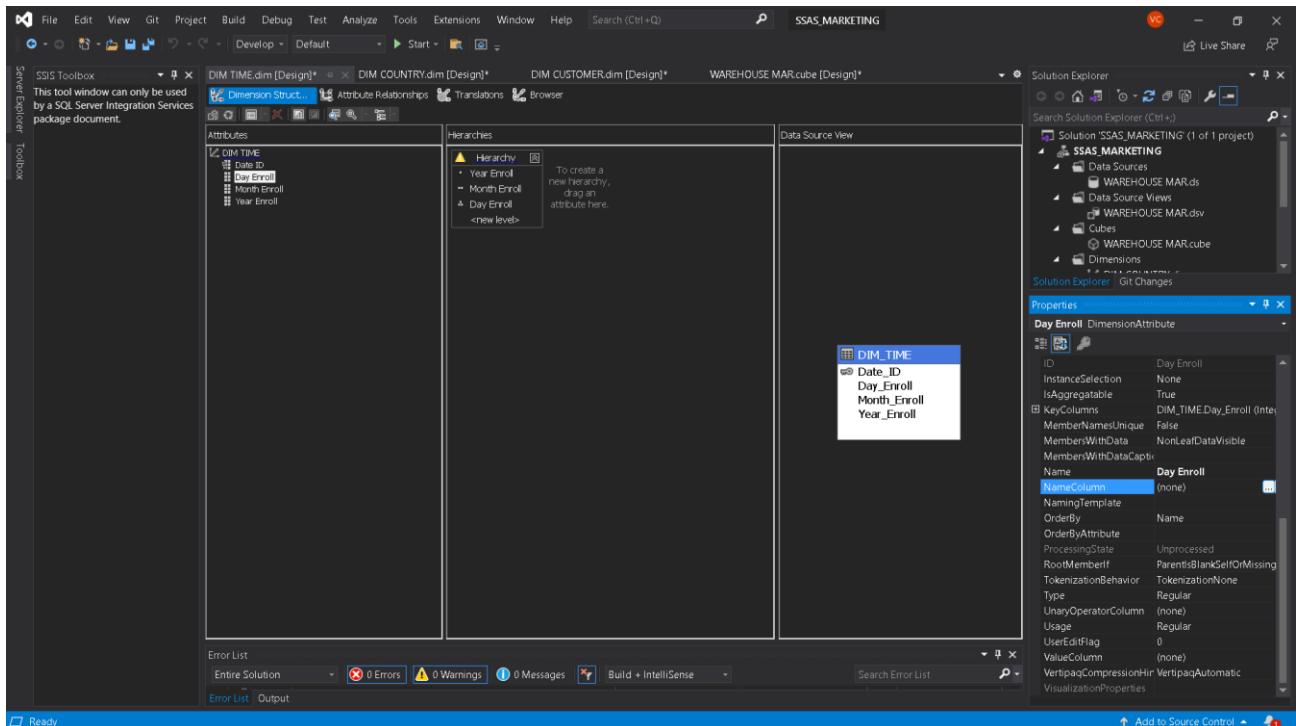
Hình 127. Start New Hierarchy

Chỉnh các thuộc tính theo Level mong muốn: **Year_Enroll -> Month_Enroll -> Day_Enroll**



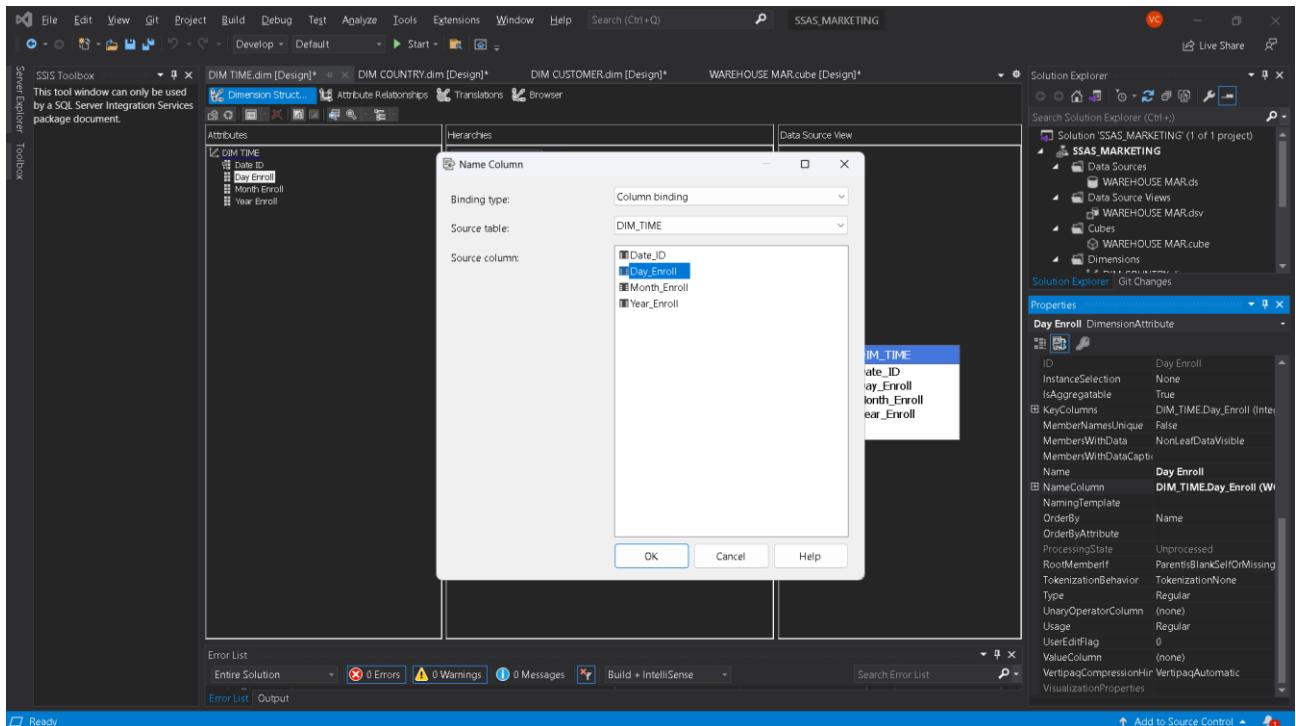
Hình 128. Chỉnh sửa Level các thuộc tính

Nhấp chuột phải vào thuộc tính **Day_Enroll -> Chọn Properties -> Nhập vào ... ở cuối hàng NameColumn**



Hình 129. Chỉnh sửa NameColumn/

Chọn Day_Enroll -> OK, Tiếp tục ở Order By chọn Key

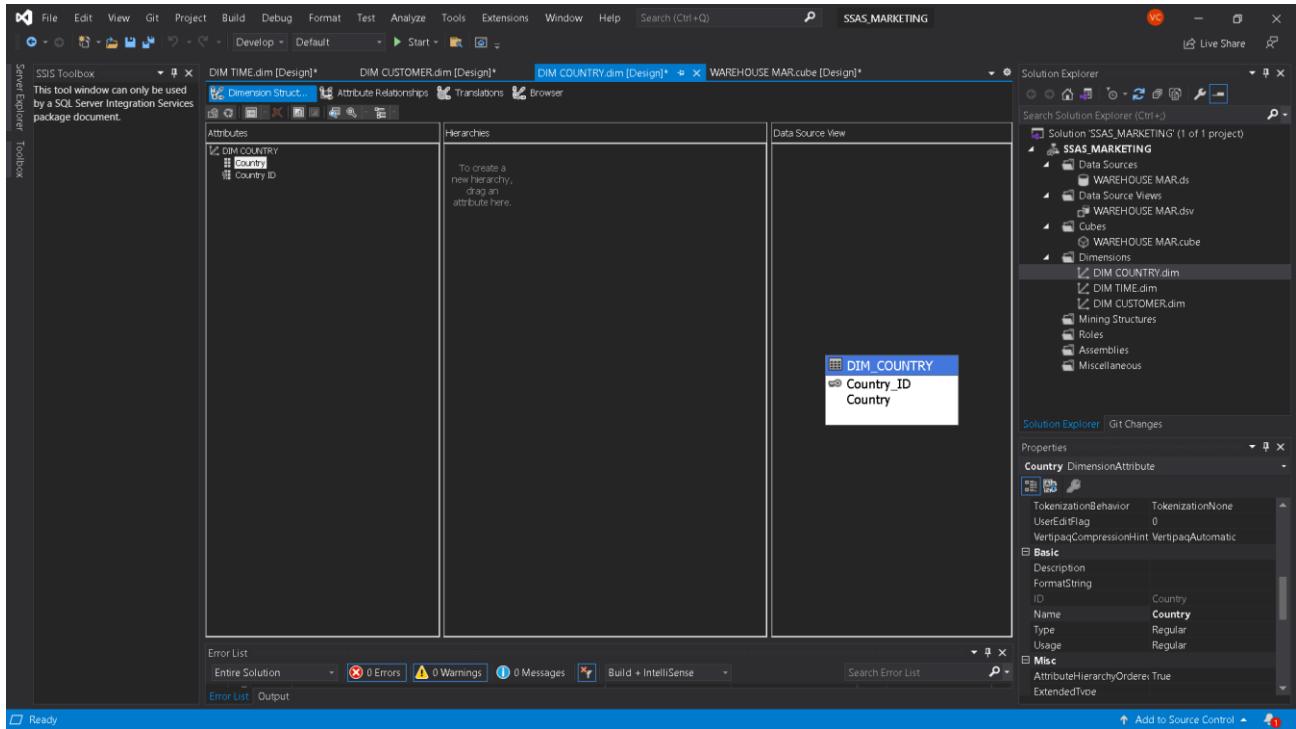


Hình 130. Set Key

Làm tiếp tục với Month_Enroll và Year_Enroll

3.1.5.3 Bảng Dim_Country

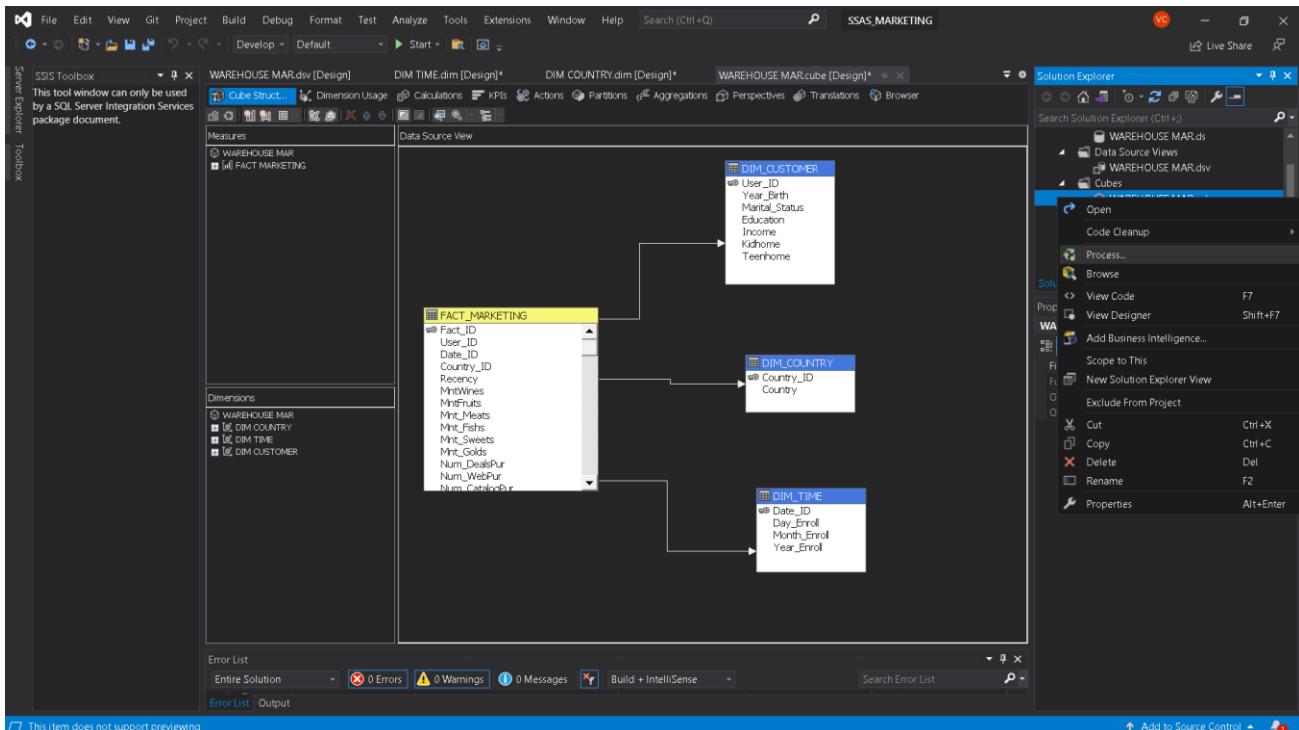
Kéo những thuộc tính cần phân cấp qua cửa sổ hierachies



Hình 131. DIM_COUNTRY

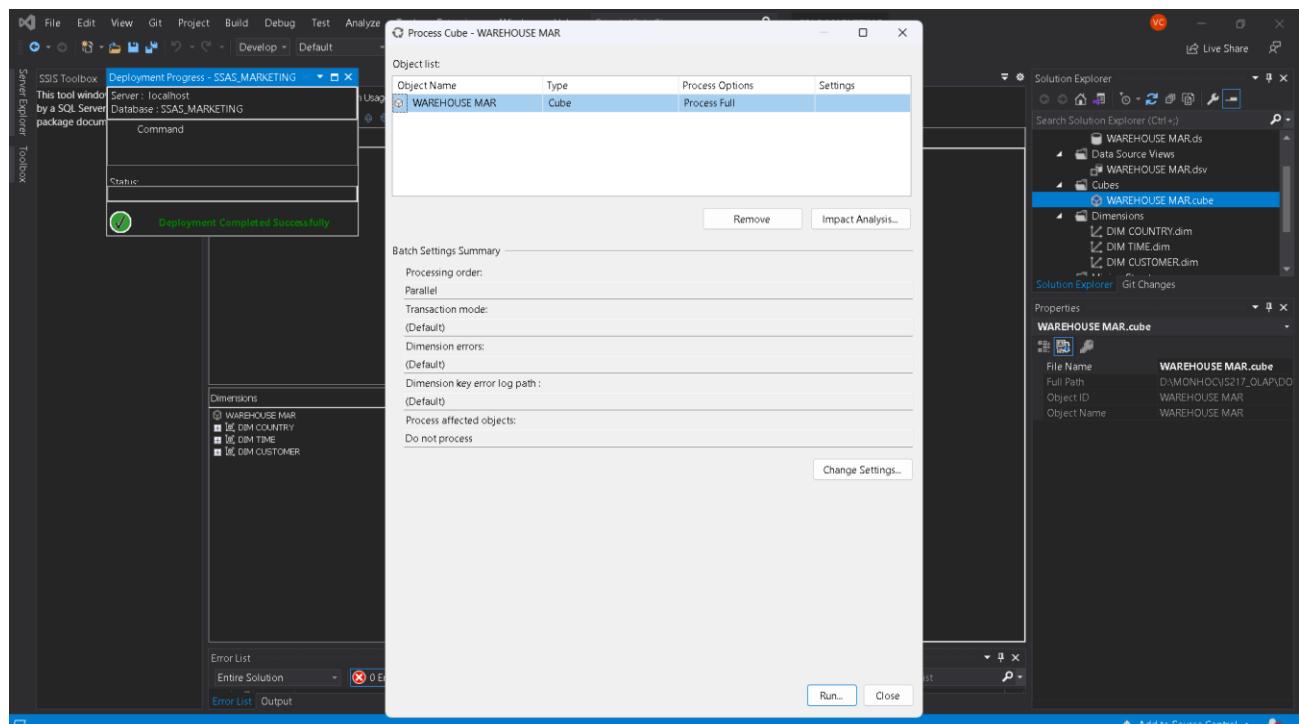
3.1.6 Process Cube

Nhấp phải chuột vào **WAREHOUSE MAR.cube** -> Chọn **Process** -> Xuất hiện thông báo ta chọn **Yes**



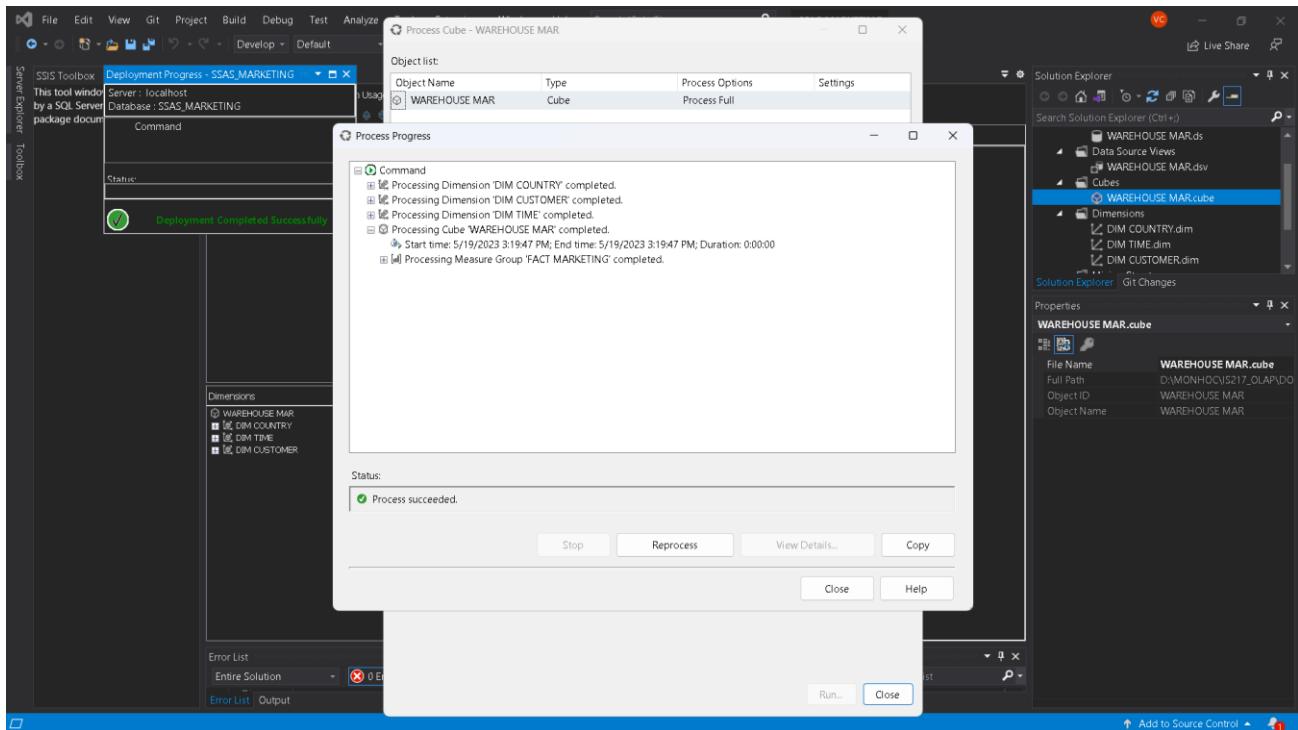
Hình 132. Process Cube

Ta nhấn Run để tiến hành chạy Process



Hình 133. Run Process

Sau khi chạy thành công màn hình sẽ hiển thị kết quả như dưới đây



Hình 134. Chạy Process thành công

3.2 Quá trình phân tích dữ liệu bằng ngôn ngữ MDX

3.2.1 Thông kê thu nhập hàng năm của Khách hàng đã chấp nhận 4 chiến dịch Marketing

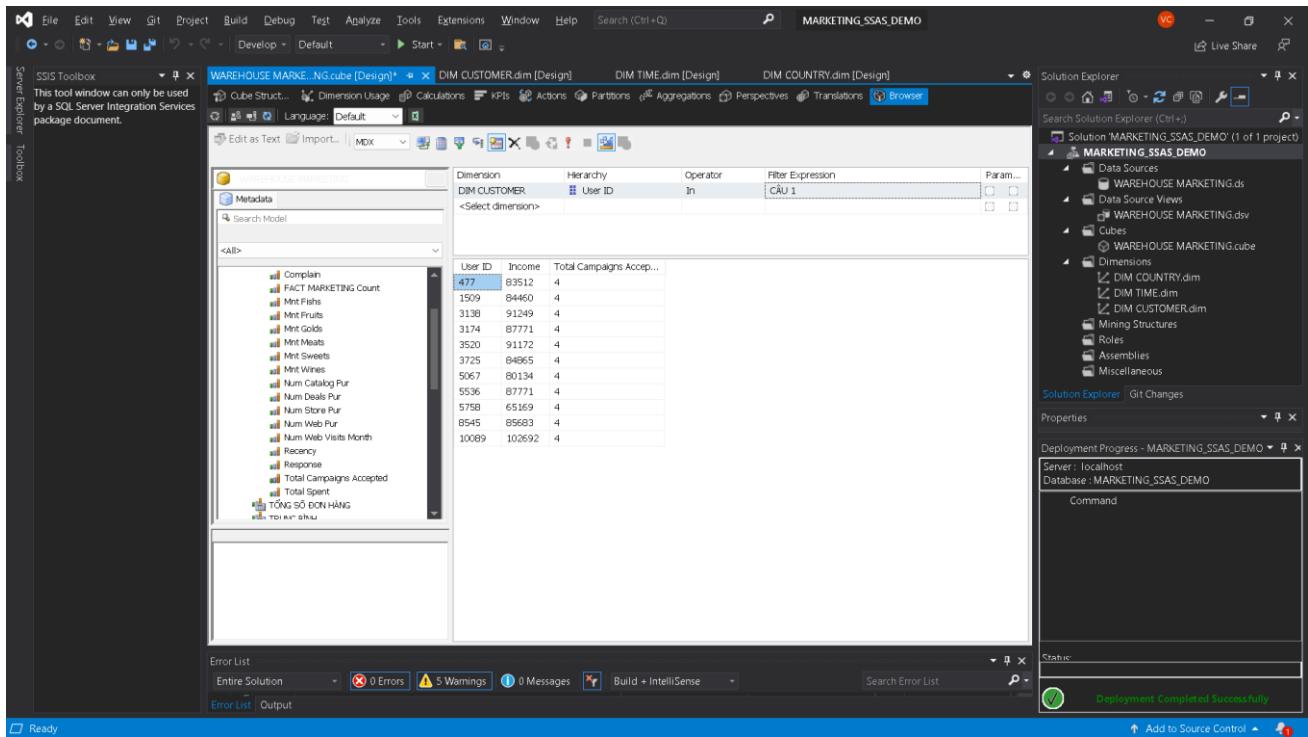
Truy vấn bằng Visual Studio 2019

```

[Name: [CÂU 1]
Expression
FILTER([DIM CUSTOMER].[User ID].[User ID], [Measures].[Total Campaigns Accepted]=4)
No issues found
Ln: 1 Ch: 43 SPC CRLF]

```

Hình 135. Câu truy vấn 1



Hình 136. Truy vấn Visual Studio 2019

✚ Truy vấn bằng MDX

The screenshot shows Microsoft SQL Server Management Studio (SSMS). The title bar reads "OLAP.mdx - (local)\MARKETING_SSAS_DEMO (LAPTOP-FTB1C2C\CAMVAN)* - Microsoft SQL Server Management Studio". The left pane is the "Object Explorer" showing the cube "WAREHOUSE MARKETING". The main pane displays an MDX query:

```
// CÂU 1: THỐNG KÊ THU NHẬP HÀNG NĂM CỦA KHÁCH HÀNG ĐÃ CHẤP NHẬN 4 CHIẾN DỊCH MARKETING
SELECT {[Measures].[Income],
          [Measures].[Total Campaigns Accepted]} ON 0,
       FILTER ([DIM CUSTOMER].[User ID].[User ID],
               [Measures].[Total Campaigns Accepted] = 4) ON 1
FROM [WAREHOUSE MARKETING];
```

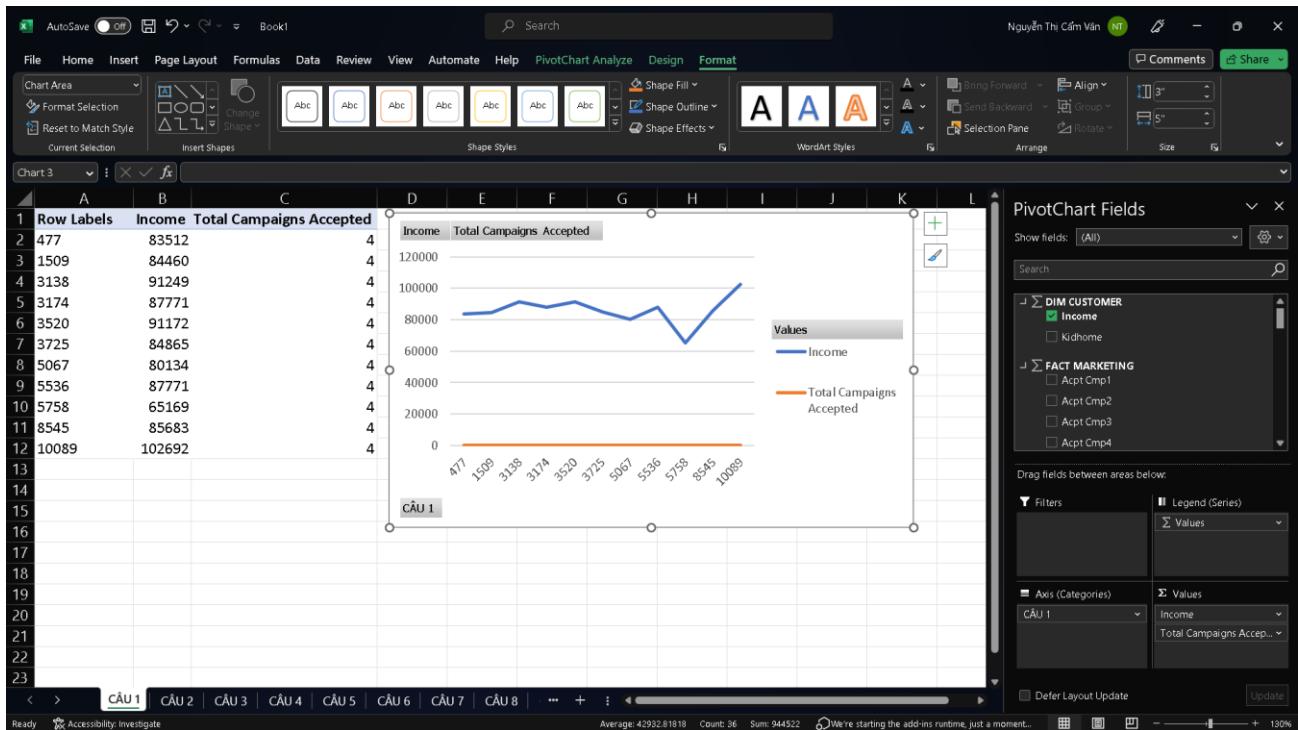
The results pane shows a table with two columns: "Income" and "Total Campaigns Accepted". The data is as follows:

User ID	Income	Total Campaigns Accepted
477	83512	4
1509	84460	4
3138	91249	4
3174	87771	4
3520	91172	4
3725	84865	4
5067	80134	4
5536	87771	4
5758	65169	4
8545	85683	4
10089	102692	4

The status bar at the bottom indicates "Query executed successfully".

Hình 137. Truy vấn MDX SQL Server

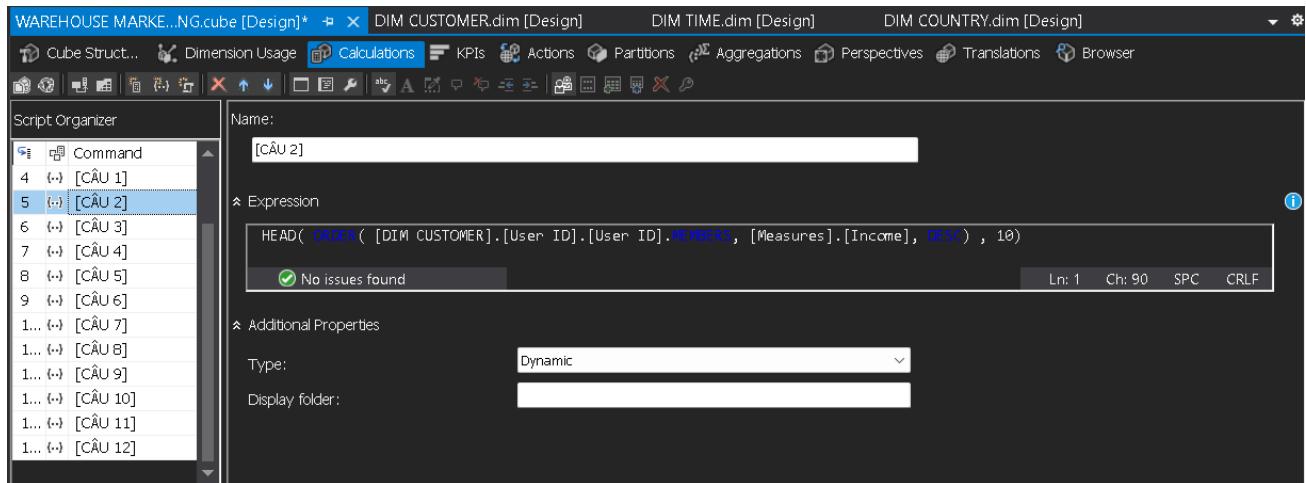
✚ Truy vấn bằng Pivot Excel



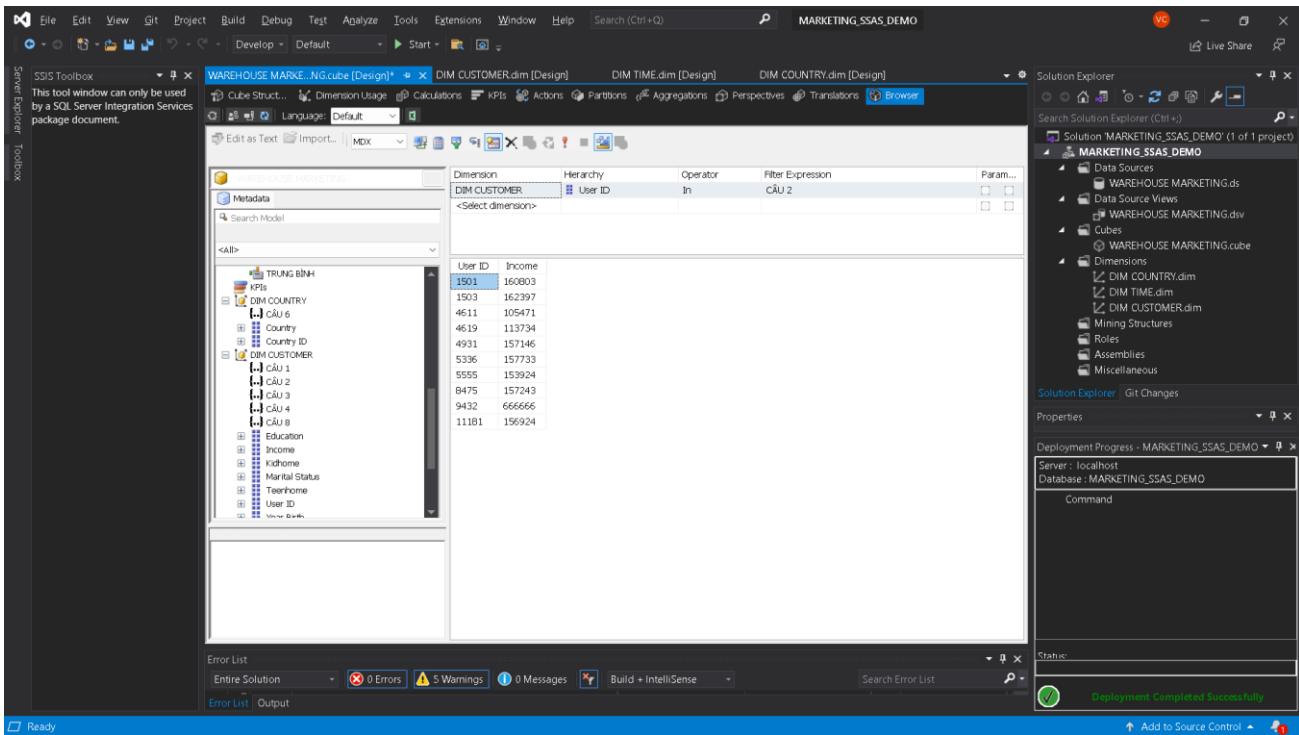
Hình 138. Truy vấn Pivot Excel

3.2.2 Thống kê Top 10 Khách hàng có thu nhập hàng năm lớn nhất

Truy vấn bằng Visual Studio 2019



Hình 139. Câu truy vấn 2



Hình 140. Truy vấn Visual Studio 2019

✚ Truy vấn bằng MDX – SQL Server

```

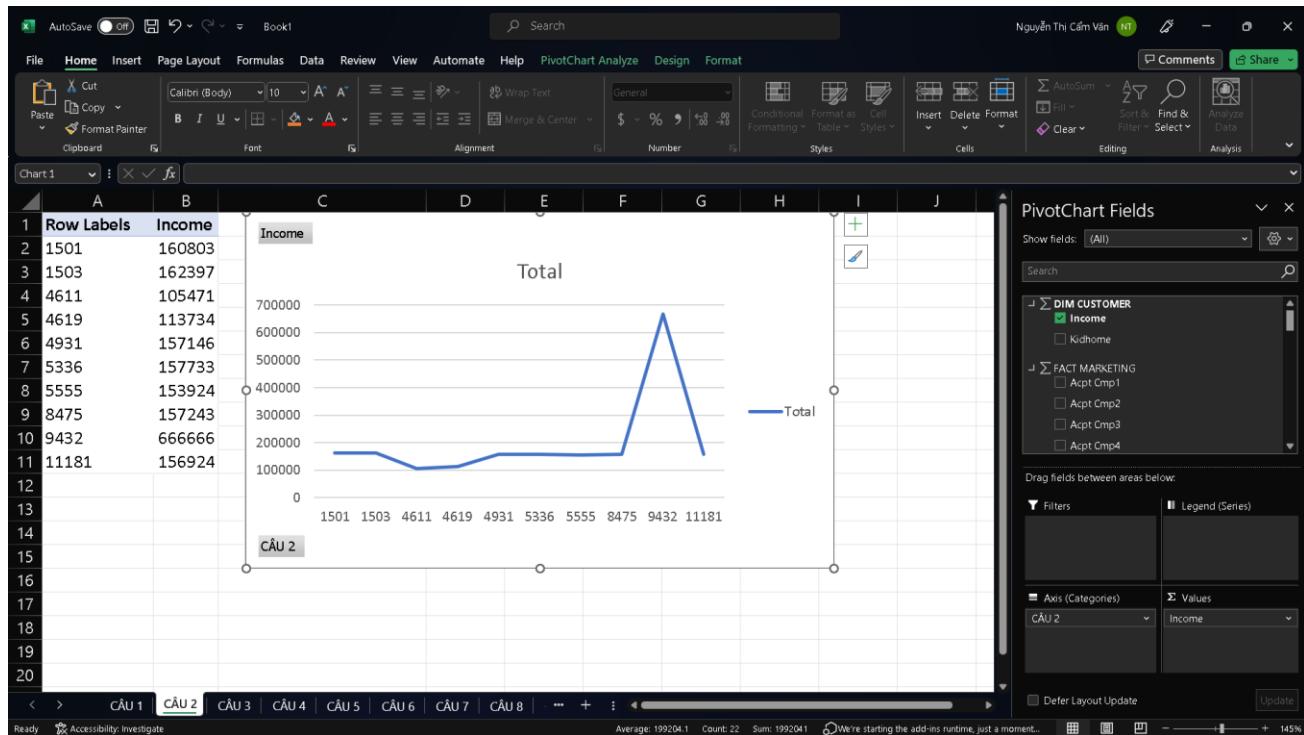
//CÂU 2: THỐNG KÊ TOP 10 KHÁCH HÀNG CÓ THU NHẬP HÀNG NĂM LỚN NHẤT
SELECT [Measures].[Income] ON 0,
HEAD (
    ORDER ([DIM CUSTOMER].[User ID].[User ID].MEMBERS,
           [Measures].[Income], DESC), 10) ON 1
FROM [WAREHOUSE MARKETING];

```

User ID	Income
9432	160803
1503	162397
1501	160803
5336	157733
9475	157243
4931	157146
11181	156924
5555	153924
4619	113734
4611	105471

Hình 141. Truy vấn MDX SQL Server

✚ Truy vấn bằng Pivot Excel



Hình 142. Truy vấn Pivot Excel

3.2.3 Thống kê số tiền đã mua hàng của top 10 Khách hàng có thu nhập hàng năm lớn hơn 30000\$

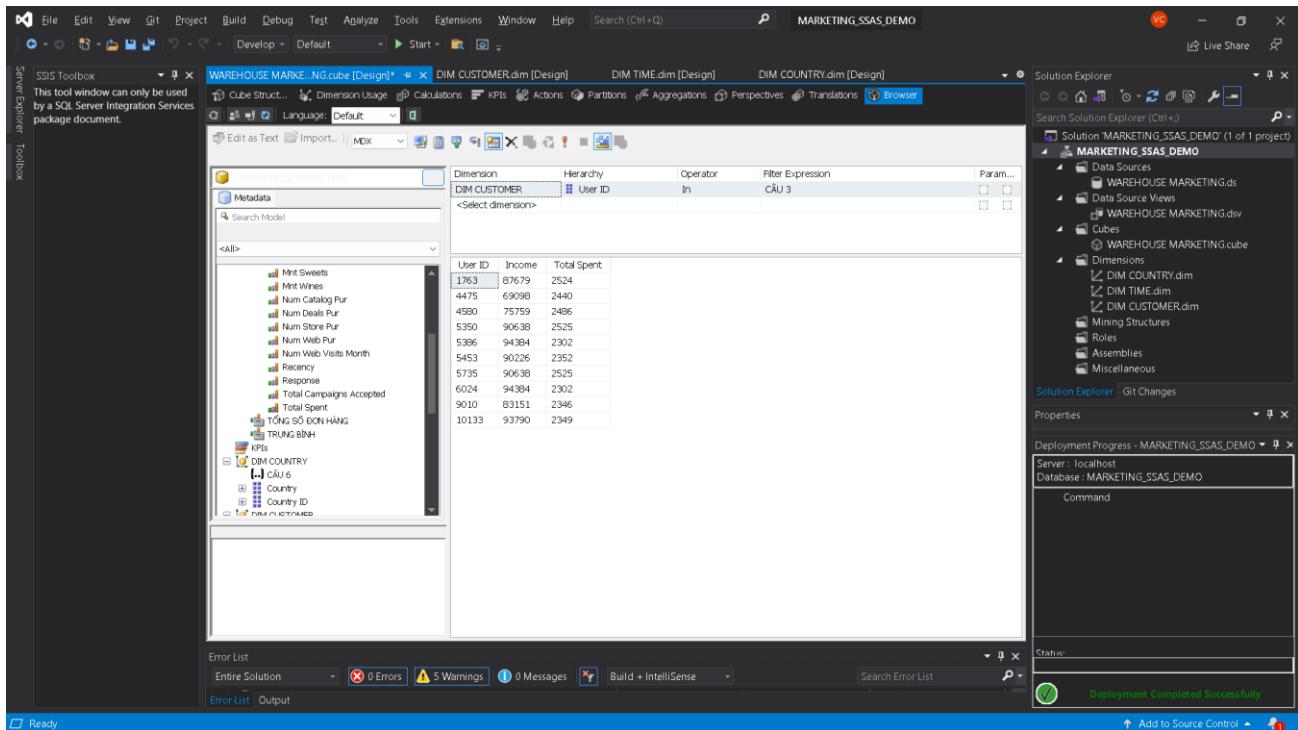
Truy vấn bằng Visual Studio 2019

The screenshot shows the Microsoft Data Tools interface within Visual Studio 2019. The top navigation bar includes tabs for 'WAREHOUSE MARKETING.cube [Design]', 'DIM CUSTOMER.dim [Design]', 'DIM TIME.dim [Design]', and 'DIM COUNTRY.dim [Design]'. Below the tabs, there are several toolbars and panes. The 'Script Organizer' pane on the left lists queries such as '[CÂU 1]', '[CÂU 2]', '[CÂU 3]', '[CÂU 4]', '[CÂU 5]', '[CÂU 6]', '[CÂU 7]', '[CÂU 8]', '[CÂU 9]', '[CÂU 10]', '[CÂU 11]', and '[CÂU 12]'. The main workspace shows the 'Calculations' tab selected. A query named '[CÂU 3]' is currently open, displaying the following DAX code:

```
HEAD({ FILTER( {ORDER([DIM CUSTOMER].[User ID].[User ID], [Measures].[Total Spent], DESC)}, [Measures].[Income] >= 30000)},10)
```

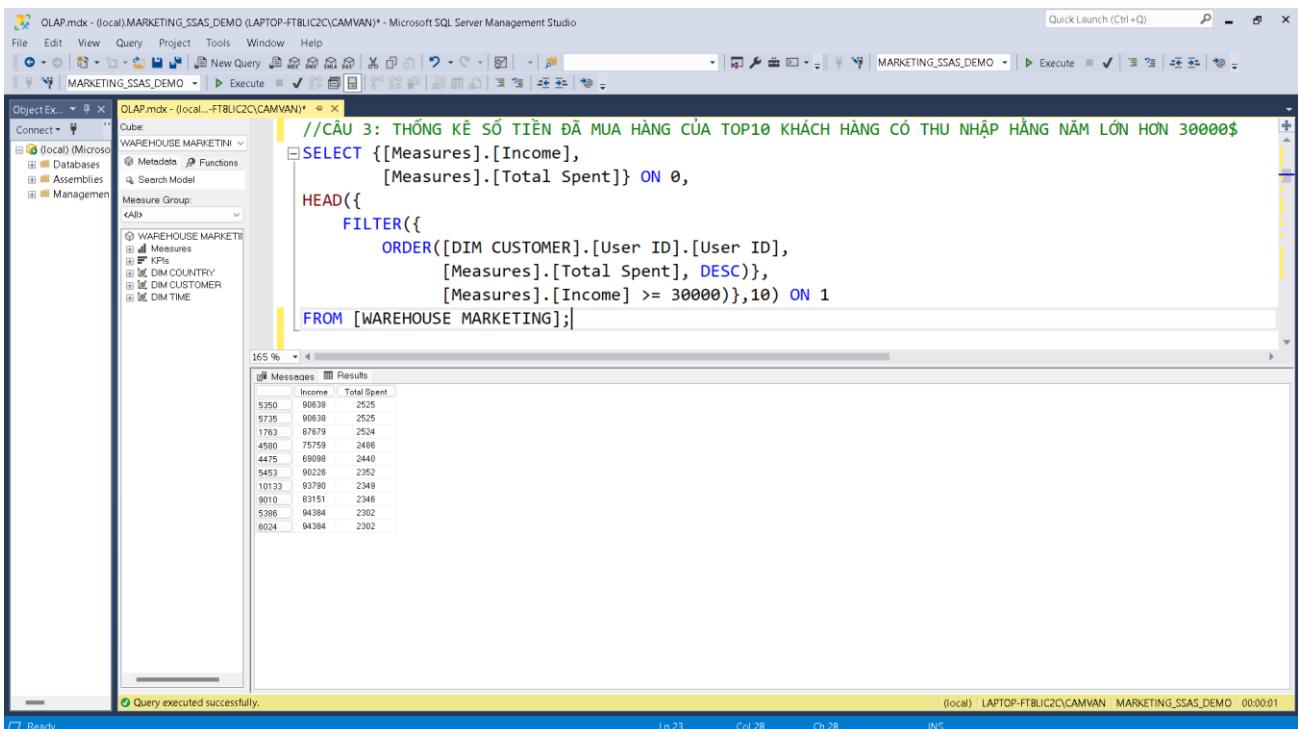
The code uses the HEAD function to return the top 10 customers based on their total spent, filtered by an income of at least 30,000. The 'Expression' pane also shows a note: 'No issues found'. Other visible tabs include 'Dimension Usage', 'KPIs', 'Actions', 'Partitions', 'Aggregations', 'Perspectives', 'Translations', and 'Browser'.

Hình 143. Câu truy vấn 3



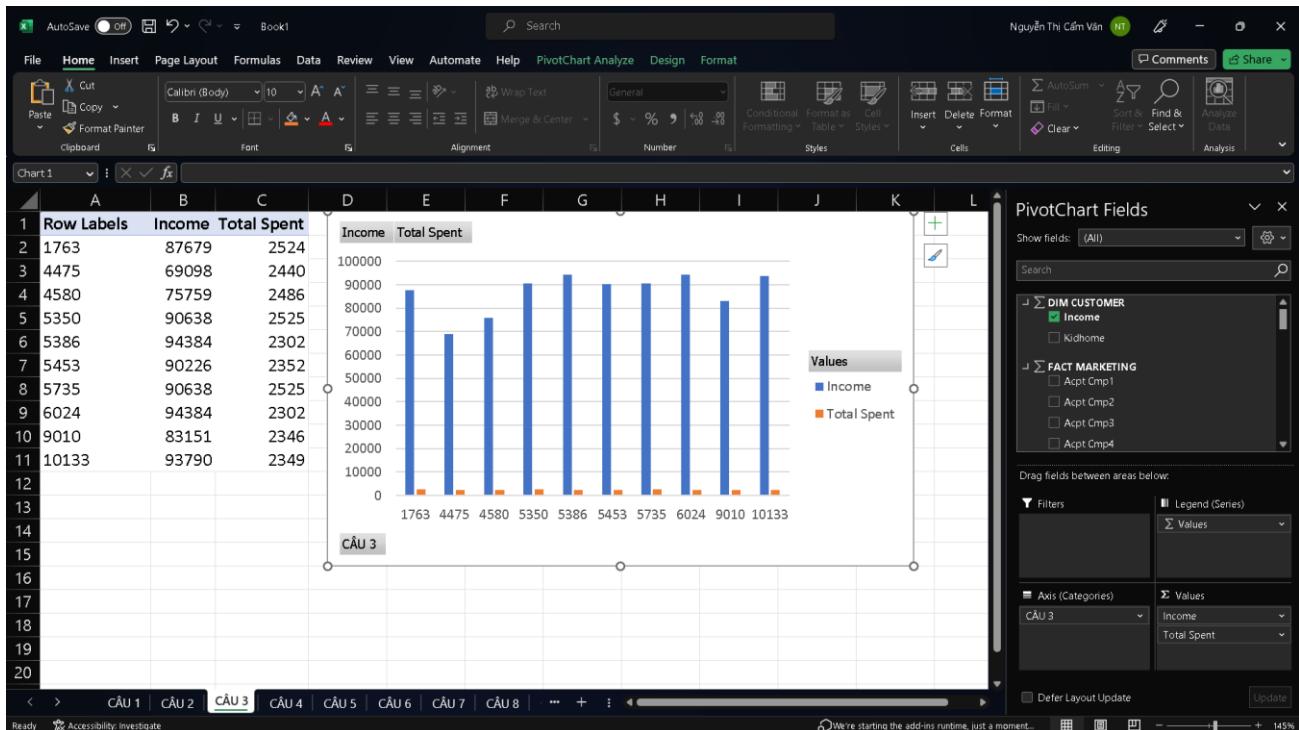
Hình 144. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server



Hình 145. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



Hình 146. Truy vấn Pivot Excel

3.2.4 Thống kê những Khách hàng đã chi nhiều hơn 150\$ cho mặt hàng bánh kẹo nhưng số lượng trẻ em trong nhà nhỏ hơn 2

Truy vấn bằng Visual Studio 2019

```

WAREHOUSE MARKETING.cube [Design] * DIM CUSTOMER.dim [Design] DIM TIME.dim [Design] DIM COUNTRY.dim [Design]
Cube Struct... Dimension Usage Calculations KPIs Actions Partitions Aggregations Perspectives Translations Browser

Name: [CÂU 4]

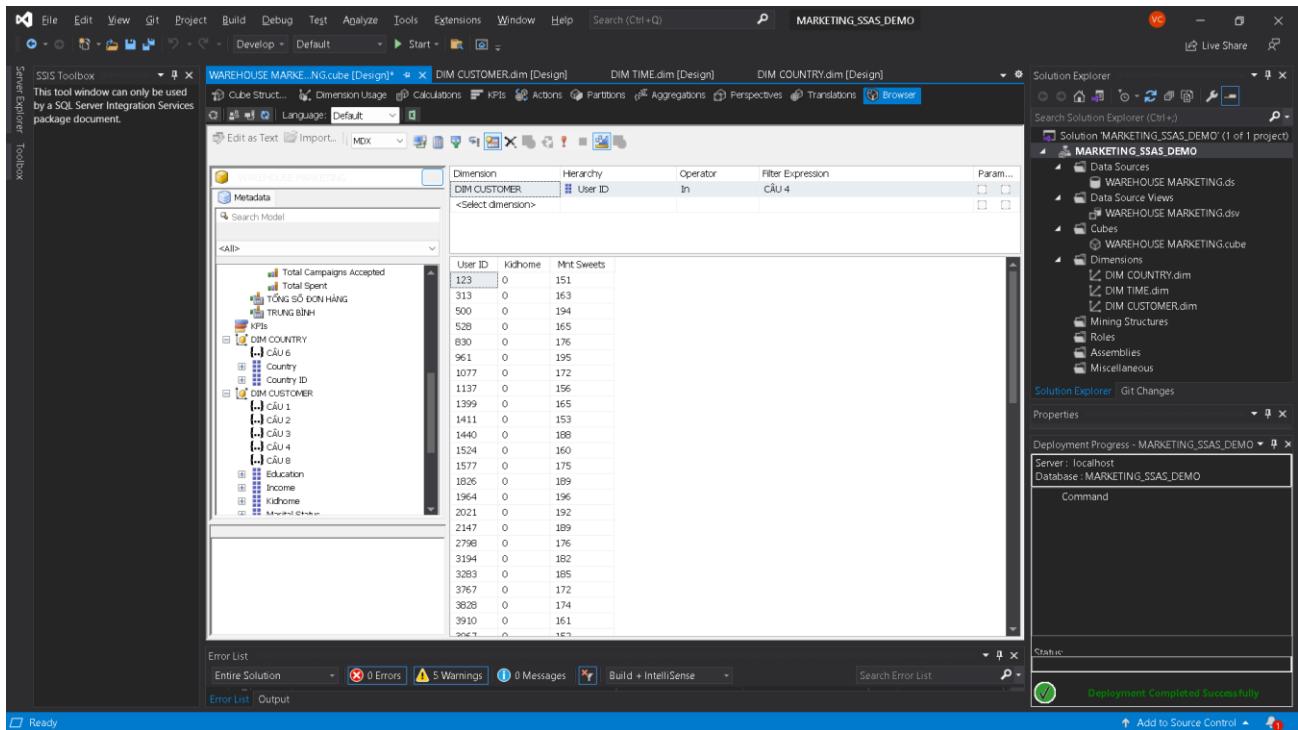
Expression
EXCEPT( {FILTER( [DIM CUSTOMER].[User ID].CHILDREN , [Measures].[Mnt Sweets] > 150)},
{FILTER( [DIM CUSTOMER].[User ID].CHILDREN , [Measures].[Kidhome] = 2)})  

No issues found
Ln: 2 Ch: 53 SPC CRLF

Additional Properties
Type: Dynamic
Display folder:

```

Hình 147. Câu truy vấn 4



Hình 148. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server

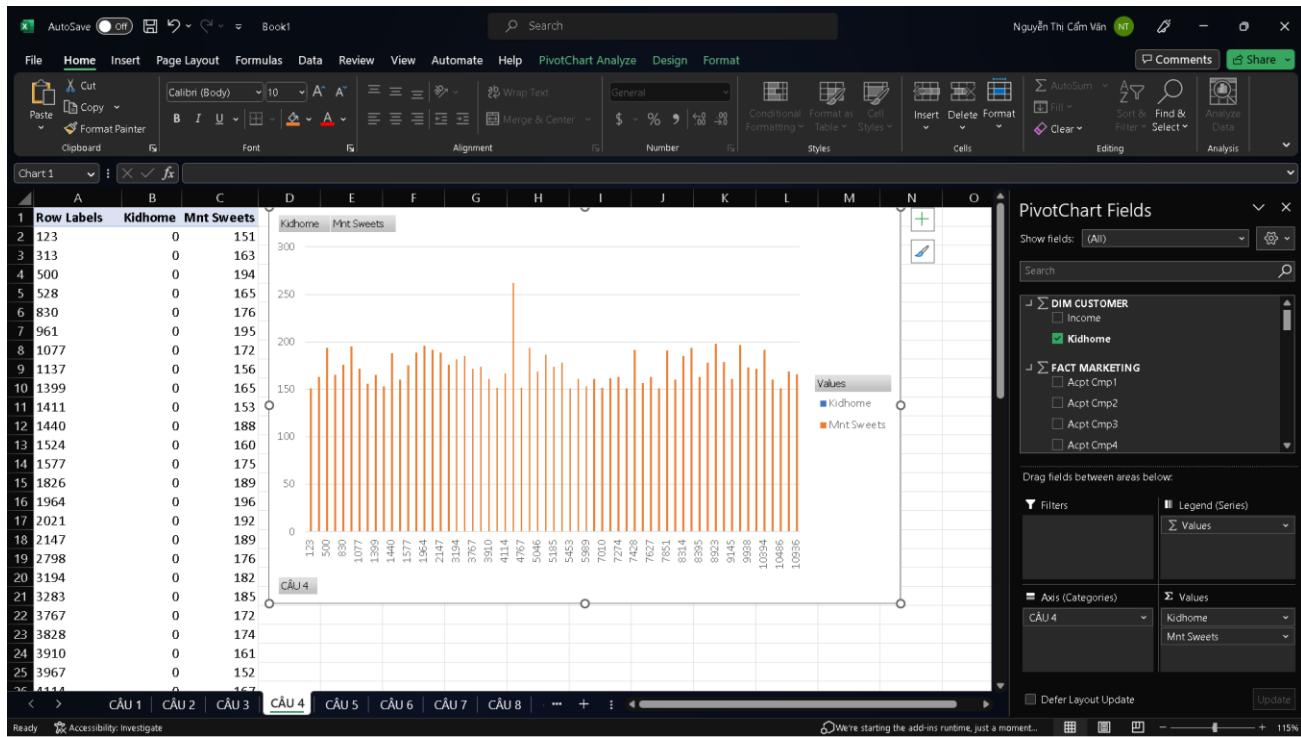
```

//CÂU 4: THỐNG KÊ NHỮNG KHÁCH HÀNG ĐÃ CHI NHÌU HƠN 150$ CHO MẶT HÀNG BÁNH KẸO NHƯNG SỐ LƯỢNG TRÈ
SELECT {[Measures].[Kidhome], [Measures].[Mnt Sweet]} ON 0,
EXCEPT (
    {FILTER ([DIM CUSTOMER].[User ID].CHILDREN ,
    [Measures].[Mnt Sweet] > 150)},
    {FILTER ([DIM CUSTOMER].[User ID].CHILDREN ,
    [Measures].[Kidhome] = 2)}) ON 1
FROM [WAREHOUSE MARKETING];
  
```

Kidhome	Mnt Sweet
123	151
313	163
500	194
528	165
830	176
961	195
1077	172
1137	156
1399	165
1411	153
1440	188
1524	160
1577	175
1826	189
1964	196
2021	192
2147	189
2798	176
3194	182
3283	185
3767	172
3828	174
3910	161
3957	152

Hình 149. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



Hình 150. Truy vấn Pivot Excel

3.2.5 Thông kê Top5 Khách hàng có tổng số đơn mua hàng nhiều nhất trong năm 2014

Truy vấn bằng Visual Studio 2019

The screenshot shows the Visual Studio 2019 interface with the 'Script Organizer' window open. The 'Name:' field contains '[CÂU 5]'. The 'Expression' section contains the following DAX code:

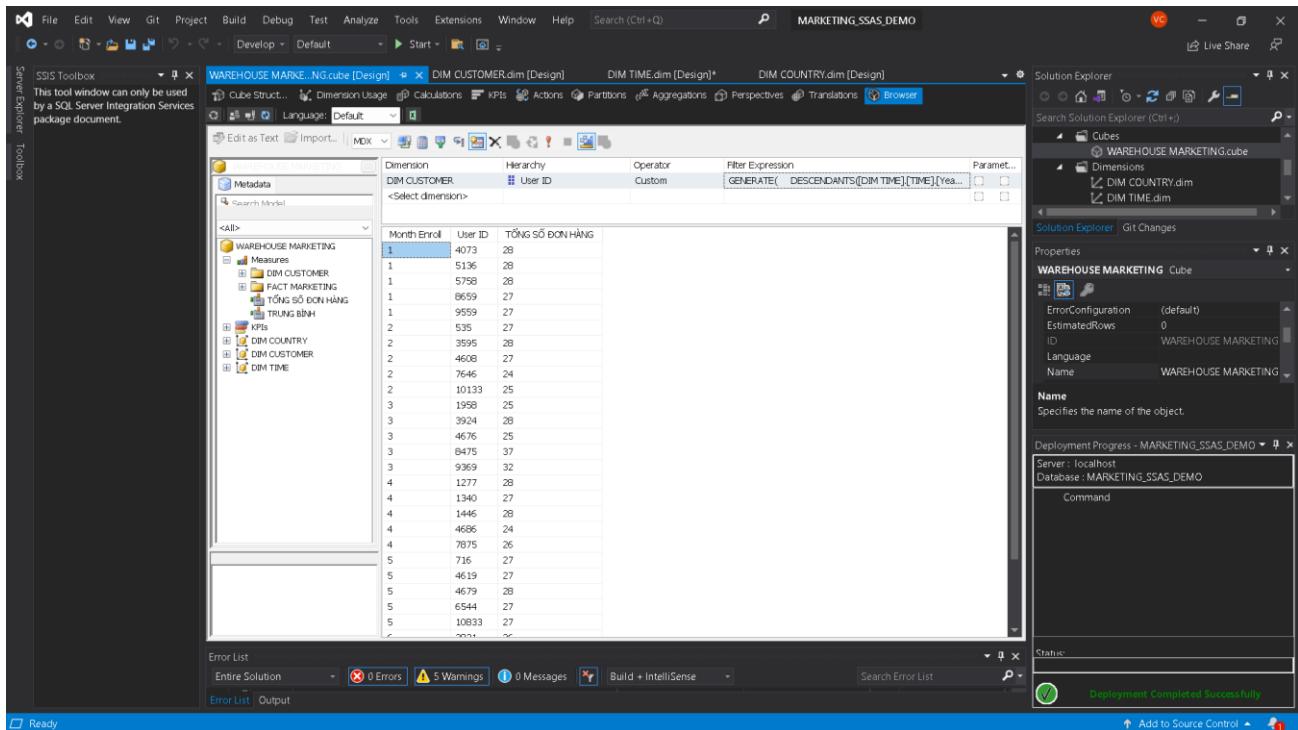
```

GENERATE(
    DESCENDANTS([DIM TIME].[TIME].[Year Enroll].&[2014].CHILDREN,[DIM TIME].[TIME].[Month Enroll]),
    TOPCOUNT([DIM TIME].[Month Enroll].CURRENTMEMBER*
    [DIM CUSTOMER].[User ID].CHILDREN,5,[Measures].[TỔNG SỐ ĐƠN HÀNG]))

```

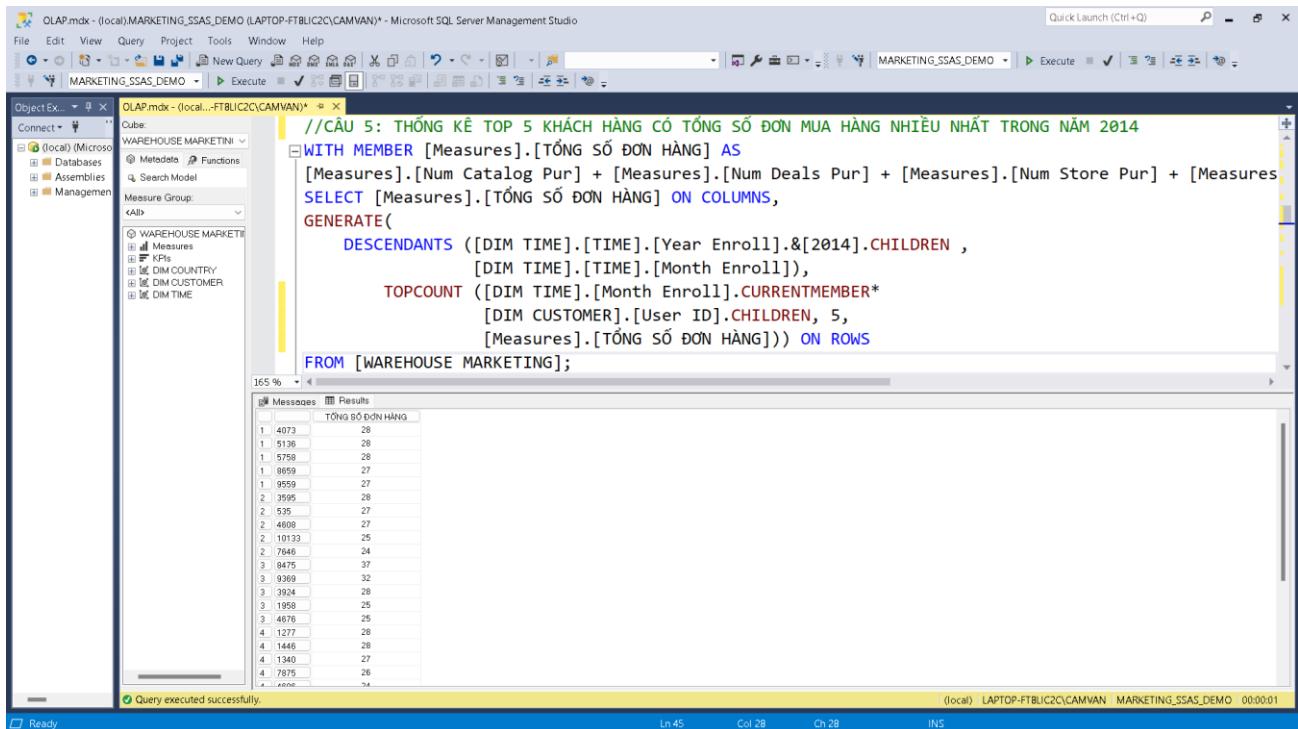
The 'Additional Properties' section shows 'Type: Dynamic' and 'Display folder: query'. The status bar at the bottom right shows 'Ln: 9 Ch: 1 SPC CRLF'.

Hình 151. Câu truy vấn 5



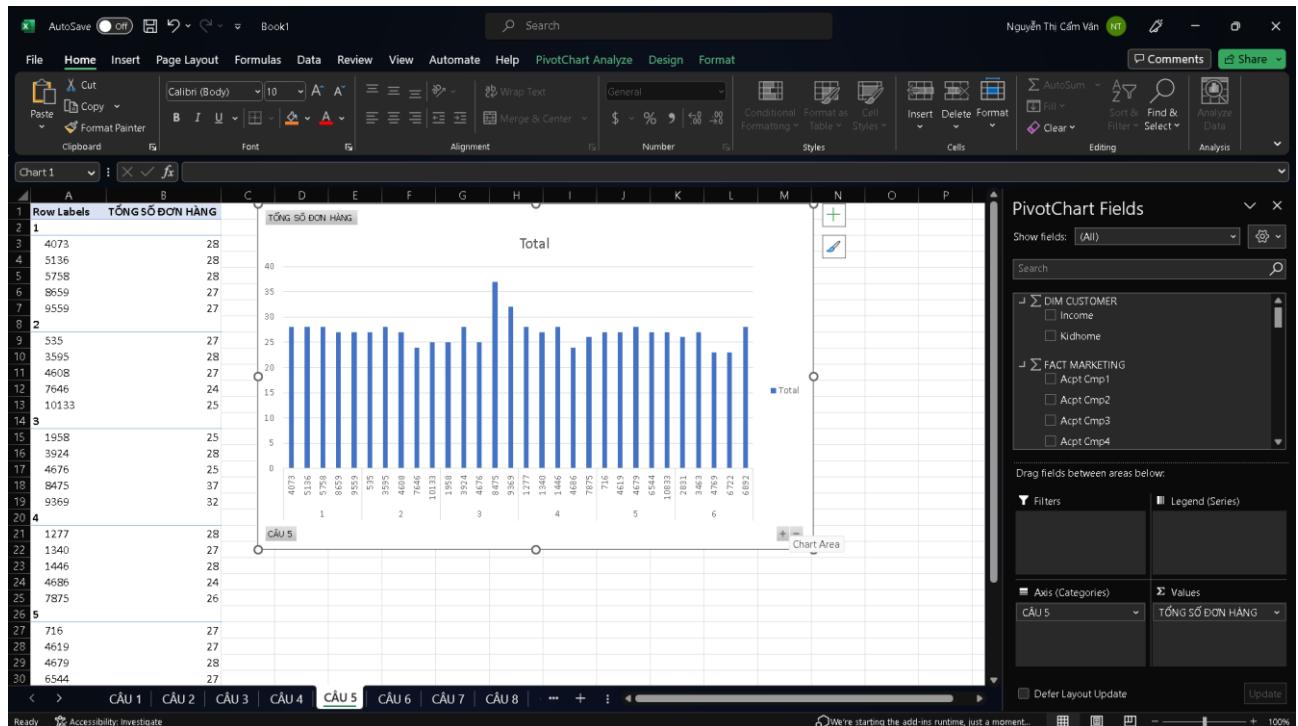
Hình 152. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server



Hình 153. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



Hình 154. Truy vấn Pivot Excel

3.2.6 Thống kê Top5 Quốc gia có số lượng mua hàng ở Store ít nhất

Truy vấn bằng Visual Studio 2019

The screenshot shows the Microsoft Analysis Services Scripting Editor in Visual Studio 2019. The title bar indicates the project is 'WAREHOUSE MARKETNG.cube [Design]'. The main area displays a DAX query:

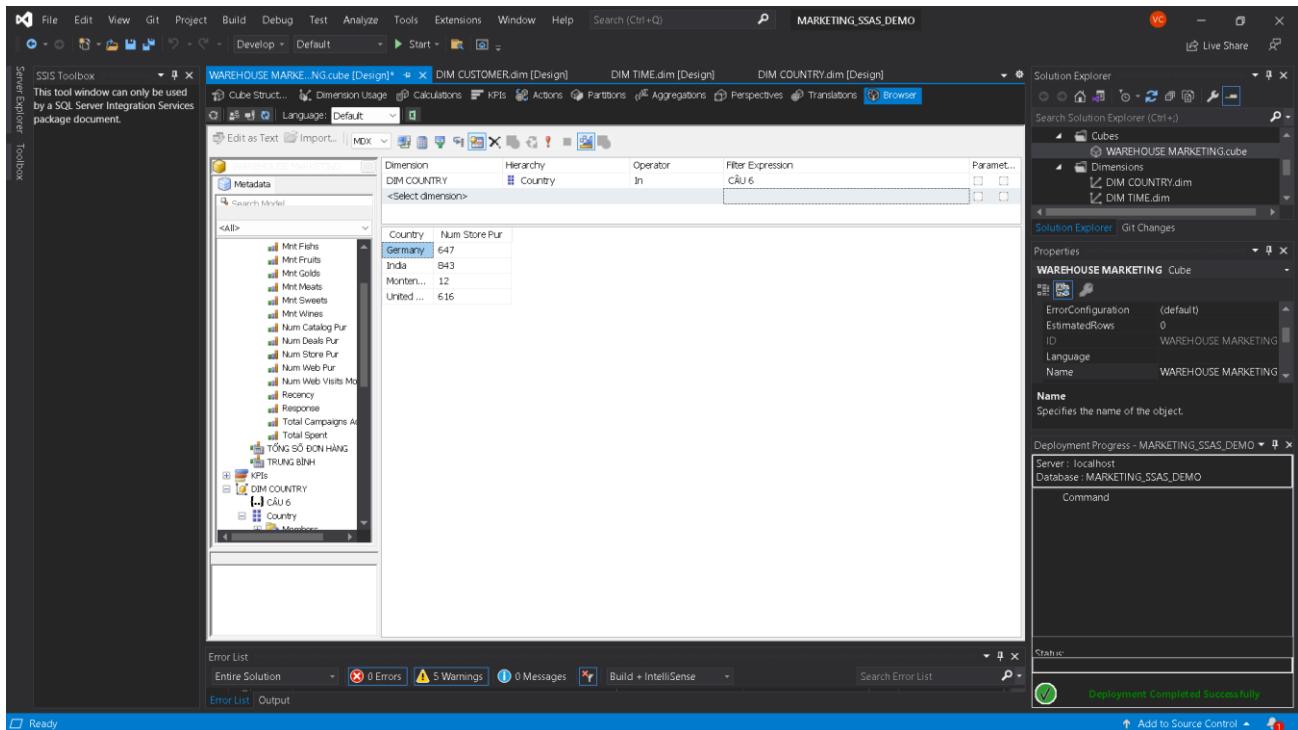
```

Name: [CÂU 6]
Expression:
BOTTOMCOUNT([DIM COUNTRY].[Country].[Country].MEMBERS,5,[Measures].[Num Store Pur])
No issues found
Ln: 1 Ch: 84 SPC CRLF

```

The 'Script Organizer' on the left shows a list of previous queries, with '[CÂU 6]' currently selected. The 'Additional Properties' section shows 'Type: Dynamic' and 'Display folder:'.

Hình 155. Câu truy vấn 6



Hình 156. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server

```

[Measures].[TỔNG SỐ ĐƠN HÀNG]) ON ROWS
FROM [WAREHOUSE MARKETING];

//CÂU 6: TOP5 QUỐC GIA CÓ SỐ LƯỢNG MUA HÀNG Ở STORE ÍT NHẤT
SELECT [Measures].[Num Store Pur] ON 0,
BOTTOMCOUNT( [DIM COUNTRY].[Country].[Country].MEMBERS, 5,
[Measures].[Num Store Pur]) ON 1
FROM [WAREHOUSE MARKETING];

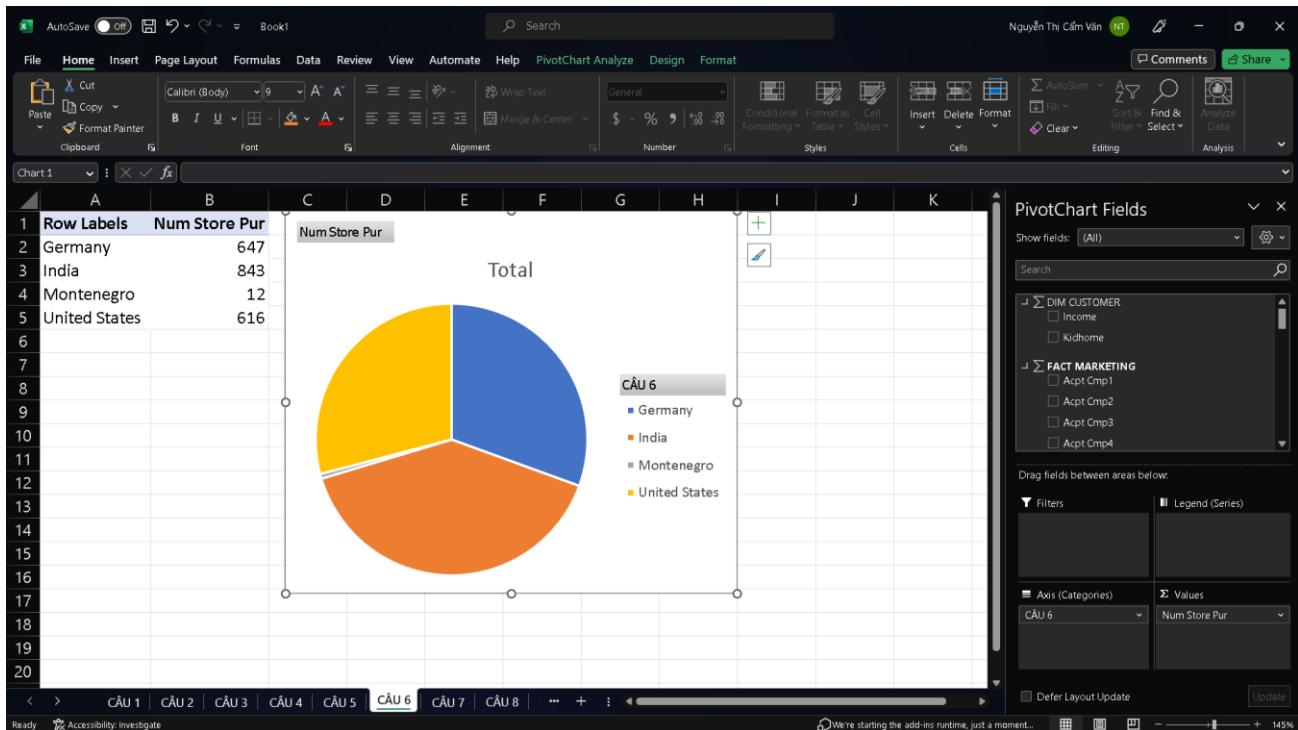
//CÂU 7: THỐNG KÊ NGÀY CÓ LƯỢNG TRUY CẬP VÀO WEBSITE ÍT NHẤT TRONG THÁNG 8 NĂM 2012
SELECT [Measures].[Num Web Visits Month] ON 0,

```

Country	Num Store Pur
Germany	647
India	843
Monten...	12
United ...	616

Hình 157. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



Hình 158. Truy vấn Pivot Excel

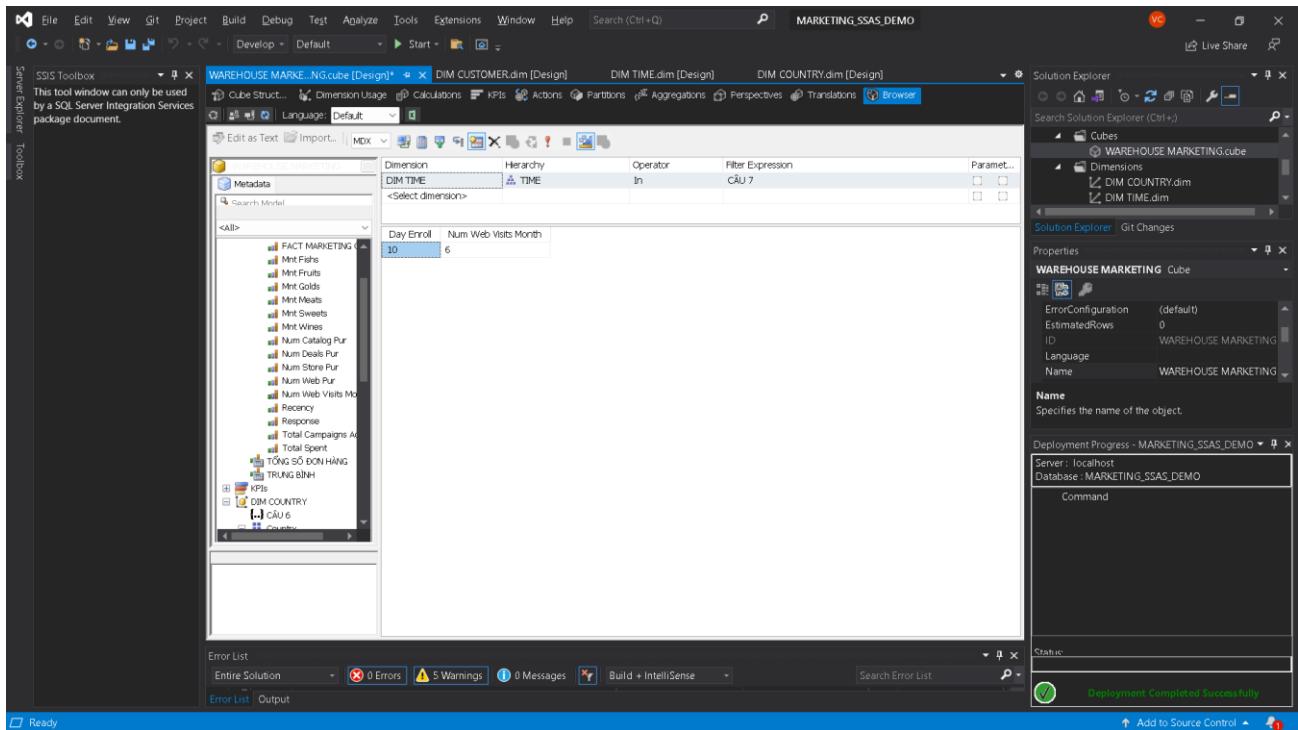
3.2.7 Thống kê ngày có lượng truy cập vào Web ít nhất trong tháng 8 năm 2012

Truy vấn bằng Visual Studio 2019

```

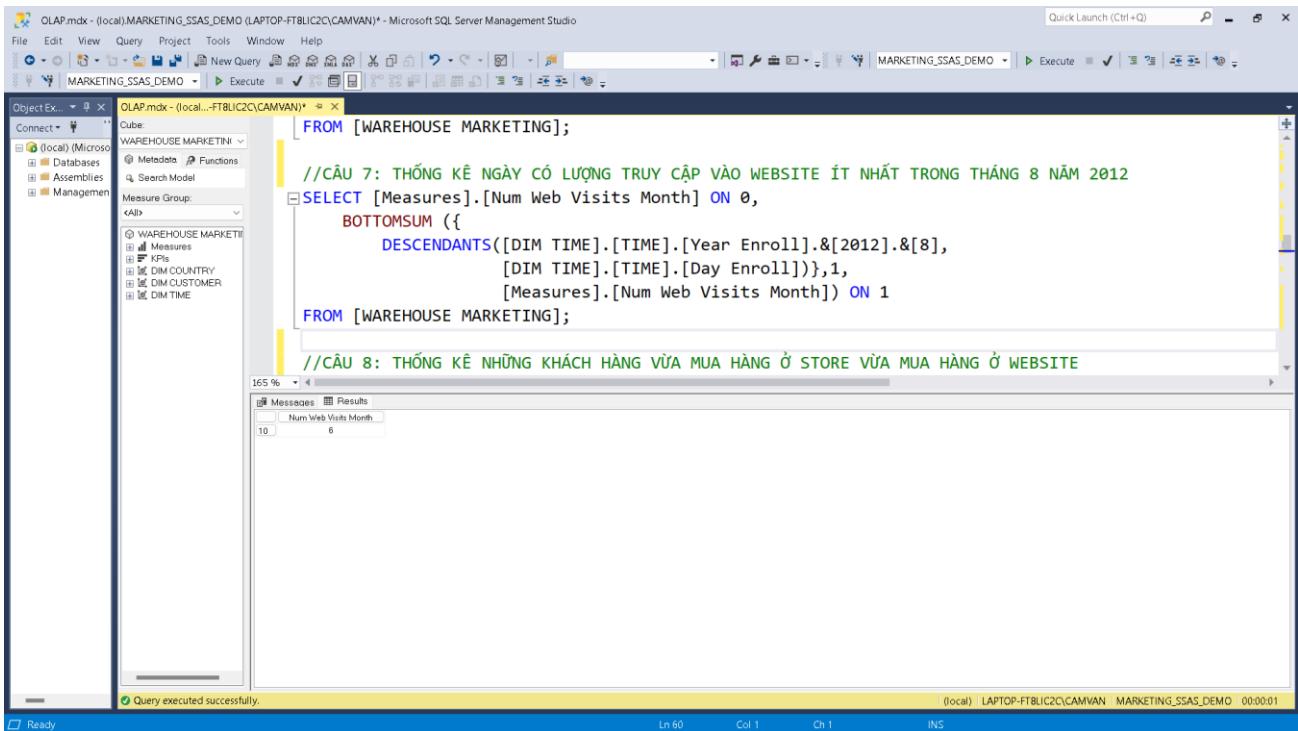
    BEGIN
        BOTTOMSUM(
            DESCENDANTS([DIM TIME].[TIME].[Year Enroll].&[2012].&[8],
            [DIM TIME].[TIME].[Day Enrol]), 1,
            [Measures].[Num Web Visits Month])
    END
    No issues found
    Ln: 6 Ch: 1 SPC CRLF
  
```

Hình 159. Truy vấn câu 7



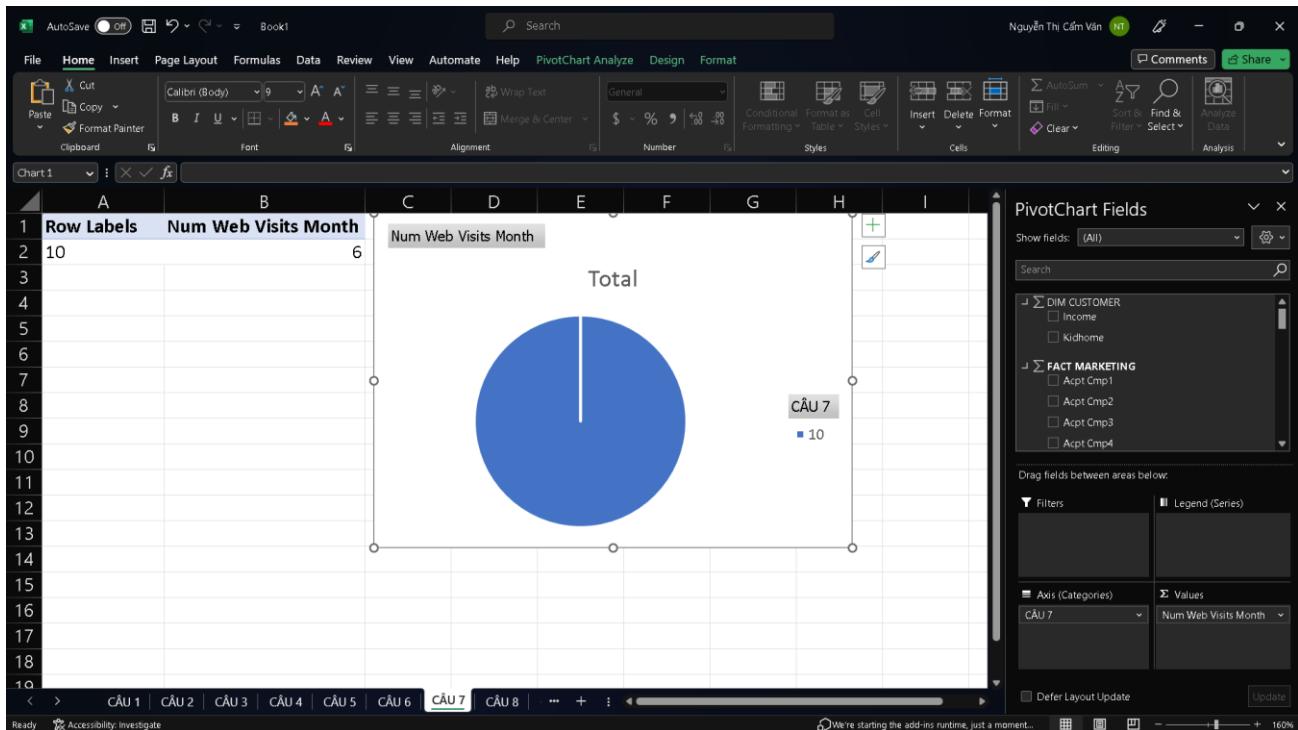
Hình 160. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server



Hình 161. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



Hình 162. Truy vấn Pivot Excel

3.2.8 Thông kê những Khách hàng vừa mua hàng ở Web vừa mua hàng ở Store

Truy vấn bằng Visual Studio 2019

```

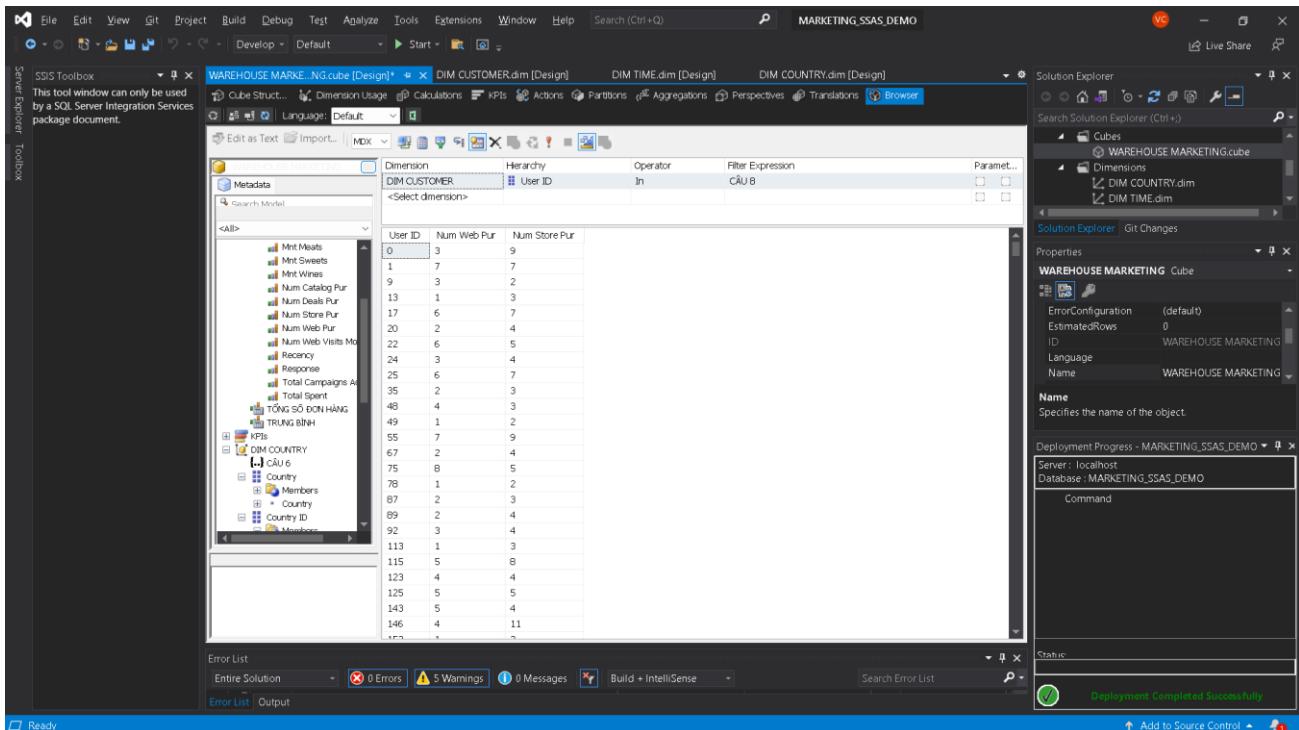
Name: [CÂU 8]
Expression:
INTERSECT(
    { FILTER([DIM CUSTOMER].[User ID].[User ID],
        [Measures].[Num Store Pur] > 0)},
    { FILTER([DIM CUSTOMER].[User ID].[User ID],
        [Measures].[Num Web Pur] > 0)})

No issues found
Ln: 5 Ch: 13 SPC CRLF

```

The screenshot shows the Microsoft Analysis Services Scripting Editor interface within Visual Studio 2019. The top navigation bar includes tabs for WAREHOUSE MARKE..NG(cube [Design]), DIM CUSTOMER.dim [Design], DIM TIME.dim [Design], and DIM COUNTRY.dim [Design]. The ribbon menu has tabs like Cube Struct., Dimension Usage, Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser. The left sidebar contains a 'Script Organizer' with a tree view of calculated members, and a 'Calculation Tools' section with buttons for Metadata, Functions, Templates, and Search Model. The main workspace displays the script for [CÂU 8], which filters users who have made purchases both online and in-store. A status bar at the bottom indicates 'No issues found'.

Hình 163. Câu truy vấn 8



Hình 164. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server

```

//CÂU 8: THỐNG KÊ NHỮNG KHÁCH HÀNG VỪA MUA HÀNG Ở STORE VỪA MUA HÀNG Ở WEBSITE
SELECT {[Measures].[Num Web Pur], [Measures].[Num Store Pur]} ON 0,
INTERSECT (
    {FILTER ([DIM CUSTOMER].[User ID].[User ID],
    [Measures].[Num Store Pur] > 0)},
    {FILTER ([DIM CUSTOMER].[User ID].[User ID],
    [Measures].[Num Web Pur] > 0)}) ON 1
FROM [WAREHOUSE MARKETING] ;

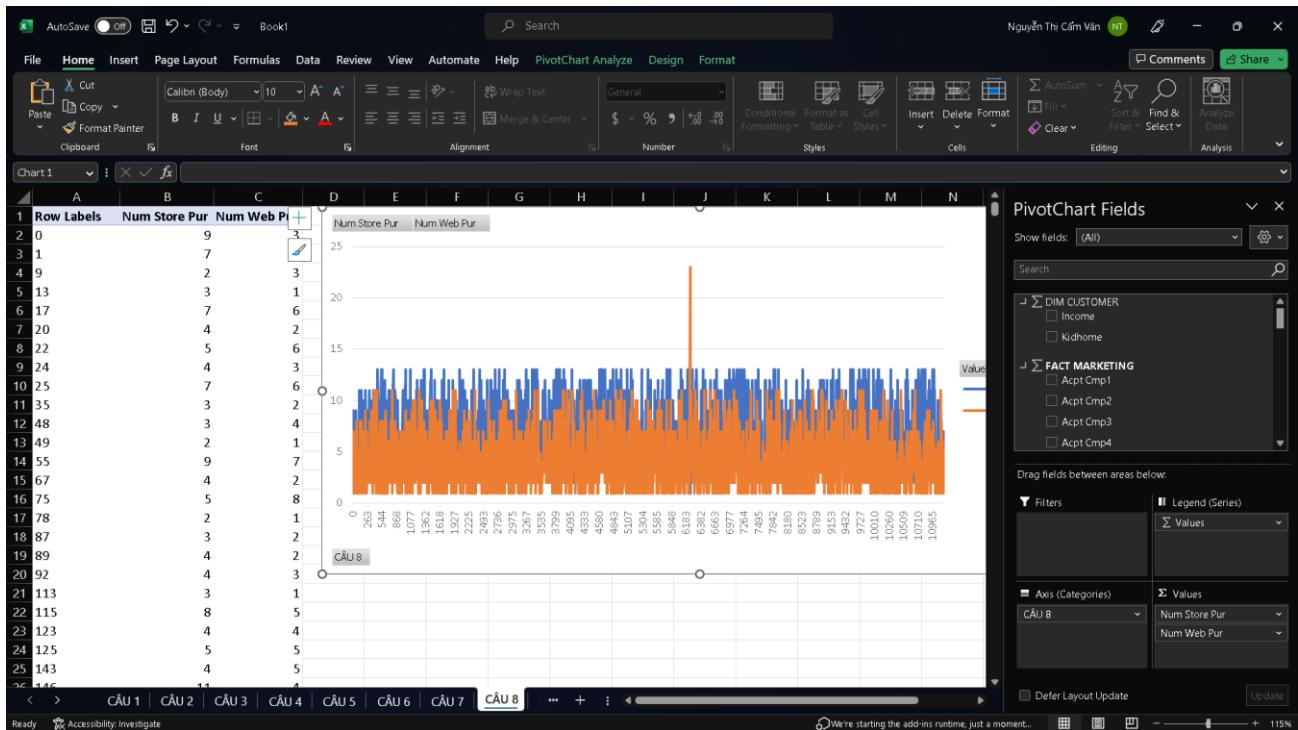
// CÂU 9: THỐNG KÊ NHỮNG KHÁCH HÀNG CÓ TỔNG SỐ ĐƠN HÀNG CAO NHẤT TẠI CÁC QUỐC GIA

```

User ID	Num Web Pur	Num Store Pur
0	3	9
1	7	7
9	3	2
13	1	3
17	6	7
20	2	4
22	6	5
24	3	4
25	6	7
35	2	3
48	4	3
49	1	2
55	7	9
67	2	4
75	8	5
78	1	2
87	2	3
89	2	4
92	3	4
113	1	3
115	5	8
123	4	4
125	5	5
143	5	4
146	4	11

Hình 165. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



Hình 166. Truy vấn Pivot Excel

3.2.9 Thống kê những Khách hàng có tổng số đơn hàng cao nhất tại các Quốc gia

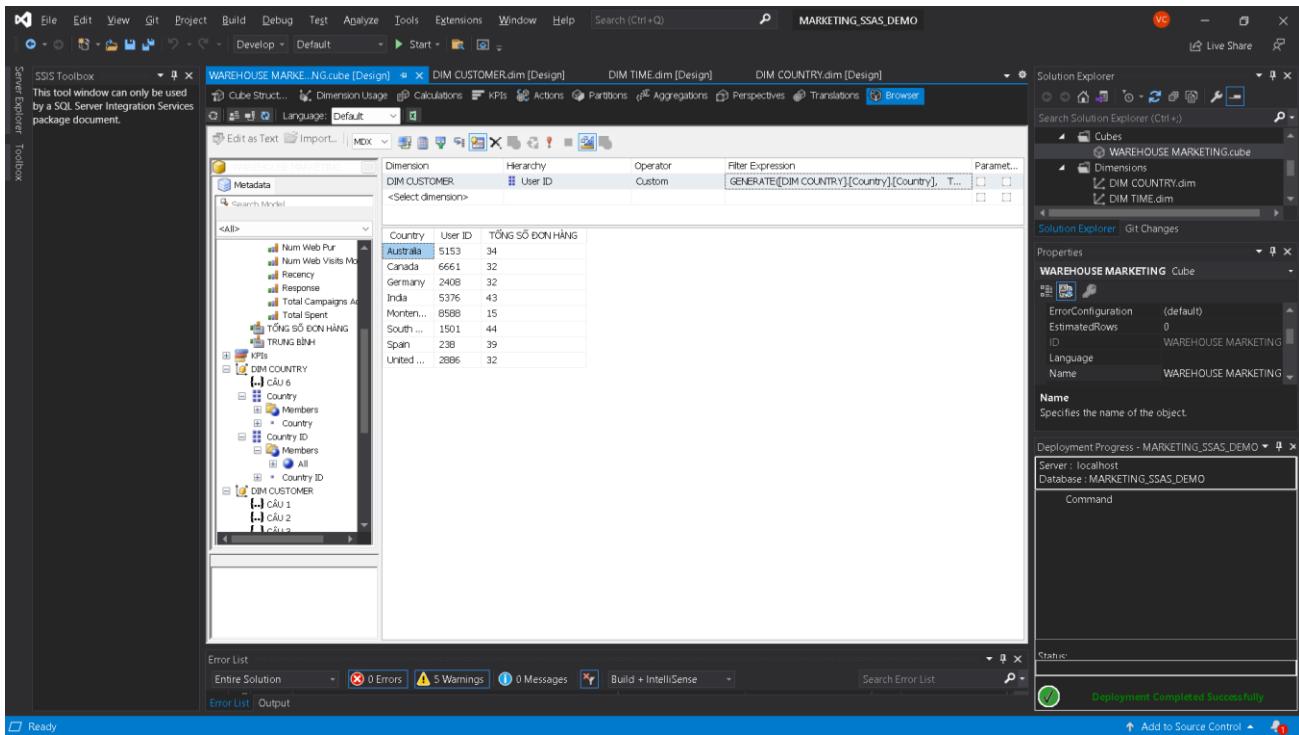
Truy vấn bằng Visual Studio 2019

```

Name: [CẤU 9]
Expression:
GENERATE([DIM COUNTRY].[Country].[Country],
TOPCOUNT([DIM COUNTRY].[Country], CURRENTMEMBER*[DIM CUSTOMER].[User ID].CHILDREN, 1,
[Measures].[TỔNG SỐ ĐƠN HÀNG]))
No issues found
Ln: 3 Ch: 15 SPC CRLF
Type: Dynamic
Display folder:

```

Hình 167. Câu truy vấn 9



Hình 168. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. The title bar reads "OLAP.mdx - (local)\MARKETING_SSAS_DEMO (LAPTOP-FTBLIC2C\CAMVAN)* - Microsoft SQL Server Management Studio". The main area contains an MDX query window with the following code:

```
// CÂU 9: THỐNG KÊ NHỮNG KHÁCH HÀNG CÓ TỔNG SỐ ĐƠN HÀNG CAO NHẤT TẠI CÁC QUỐC GIA
WITH MEMBER [Measures].[TỔNG SỐ ĐƠN HÀNG] AS
[Measures].[Num Catalog Pur] + [Measures].[Num Deals Pur] + [Measures].[Num Store Pur] + [Measures]
SELECT [Measures].[TỔNG SỐ ĐƠN HÀNG] ON 0,
GENERATE([DIM COUNTRY].[Country].[Country],
TOPCOUNT
([DIM COUNTRY].[Country].CURRENTMEMBER*)
,[DIM CUSTOMER].[User ID].CHILDREN, 1,
[Measures].[TỔNG SỐ ĐƠN HÀNG])) ON 1
FROM [WAREHOUSE MARKETING];
```

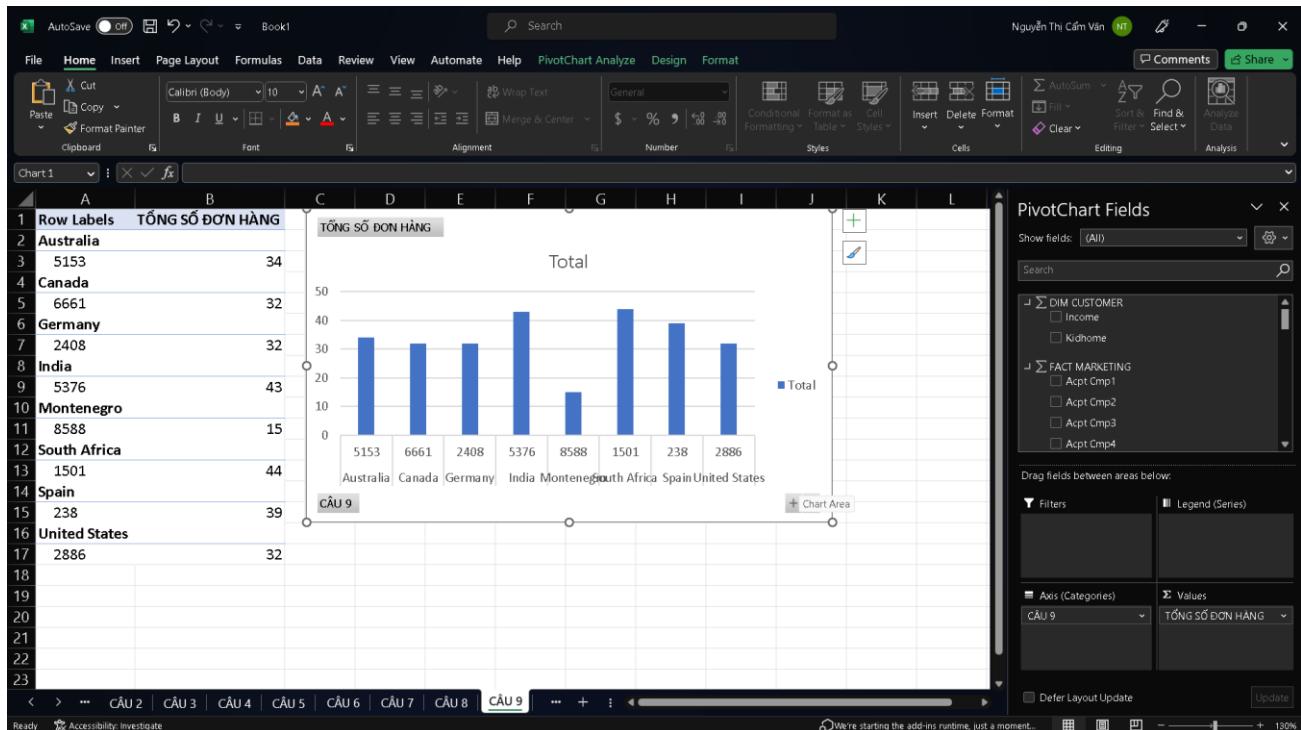
Below the query window is a results grid showing the following data:

Country	User ID	TỔNG SỐ ĐƠN HÀNG
Australia	5153	34
Canada	6661	32
Germany	2408	32
India	5376	43
Montenegro	8588	15
South Africa	1501	44
Spain	238	39
United States	2686	32
Unknown	0	(null)

At the bottom of the screen, a status bar indicates "Query executed successfully." and "Ln 76 Col 49 Ch 43 INS".

Hình 169. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



Hình 170. Truy vấn Pivot Excel

3.2.10 Tính trung bình số tiền đã chi cho tất cả các mặt hàng của Top5 Khách hàng có thu nhập hàng năm cao nhất theo từng Quốc gia

Truy vấn bằng Visual Studio 2019

The screenshot shows the Microsoft Analysis Services Scripting Editor in Visual Studio 2019. The top menu bar includes 'WAREHOUSE MARKE..NG(cube [Design])', 'DIM CUSTOMER.dim [Design]', 'DIM TIME.dim [Design]', and 'DIM COUNTRY.dim [Design]'. The main area displays the following DAX script:

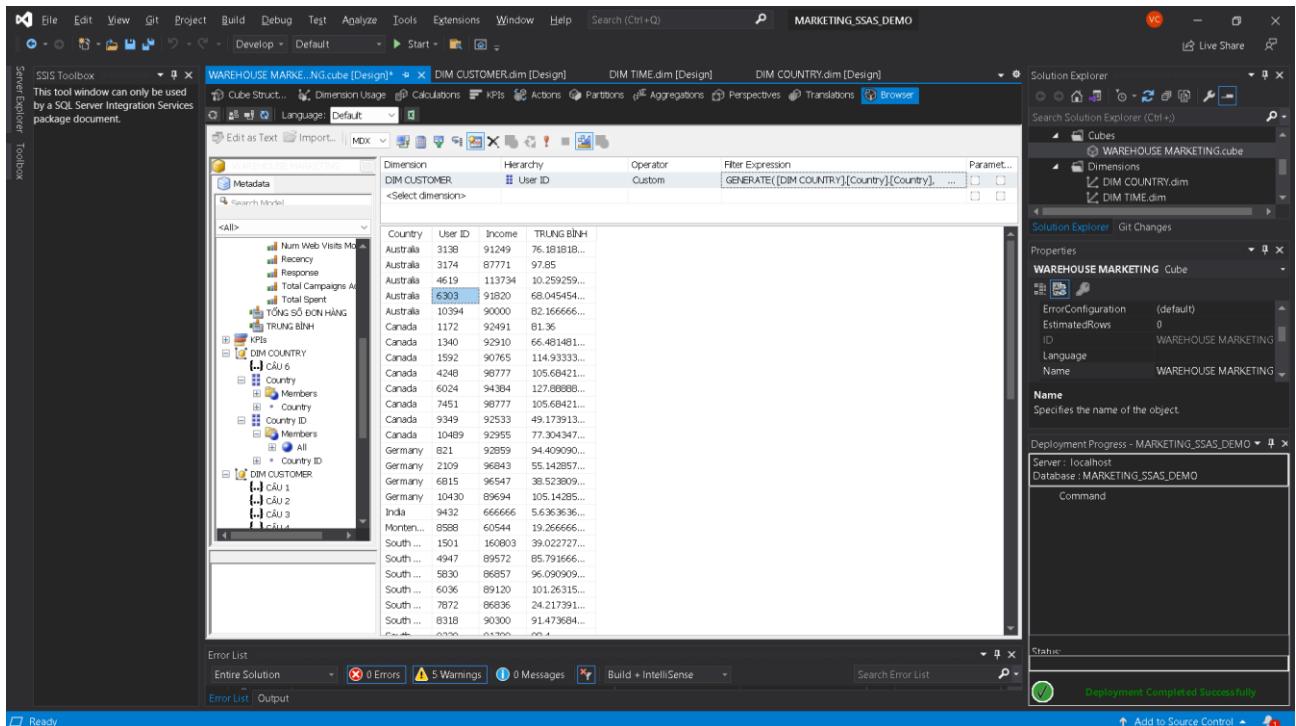
```

Name: [CÂU 10]
Expression:
GENERATE([DIM COUNTRY].[Country].[Country],
TOPPERCENT([DIM COUNTRY].[Country].CURRENTMEMBER*
FILTER([DIM CUSTOMER].[User ID].[User ID],
[Measures].[TỔNG SỐ ĐƠN HÀNG] > 0)},5,
[Measures].[Income]))

```

The 'Script Organizer' pane on the left shows a list of scripts, with '[CÂU 10]' currently selected. The bottom navigation bar includes 'Calculation Tools', 'Metadata', 'Functions', and 'Templates'.

Hình 171, Câu truy vấn câu 10



Hình 172. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server

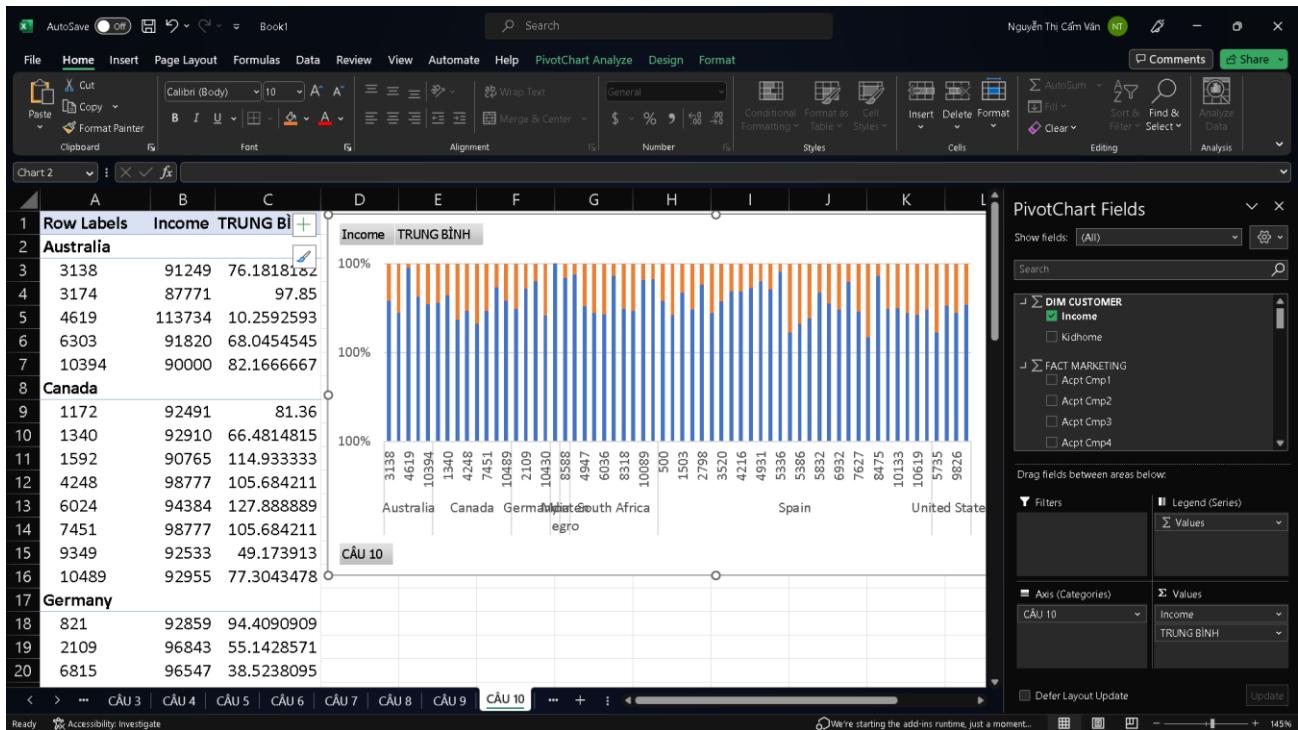
```

//CÂU 10: TÍNH TRUNG BÌNH SỐ TIỀN ĐÃ CHI CHO TẤT CẢ CÁC MẶT HÀNG CỦA TOP5 KHÁCH HÀNG CÓ THU NHẬP HÀNG ĐẦU
WITH MEMBER [Measures].[TRUNG BÌNH] AS [Measures].[Total Spent]/[Measures].[TỔNG SỐ ĐƠN HÀNG]
SELECT {[Measures].[Income], [Measures].[TRUNG BÌNH]} ON 0,
    GENERATE([DIM COUNTRY].[Country].[Country],
        TOPPERCENT([DIM COUNTRY].[Country].CURRENTMEMBER*
            FILTER([DIM CUSTOMER].[User ID].[User ID],
                [Measures].[TỔNG SỐ ĐƠN HÀNG] <> 0)),5,
            [Measures].[Income])) ON 1
FROM [WAREHOUSE MARKETING];
//CÂU 11: THỐNG KÊ SỐ LƯỢNG ĐƠN HÀNG ĐÃ THANH TOÁN QUA CATALOG VÀ THU NHẬP HẰNG NĂM CỦA TOP 15 KHÁCH HÀNG ĐẦU
  
```

Country	User ID	Income	TRUNG BÌNH
Australia	3138	91249	76.181818...
Australia	3174	87771	97.85
Australia	4619	113734	10.259259...
Australia	6303	91820	68.045454...
Australia	10394	90000	82.166666...
Canada	1172	92491	81.36
Canada	1340	92910	66.461461...
Canada	1592	90765	114.9333...
Canada	4248	98777	105.68421...
Canada	6024	94384	127.68889...
Canada	7451	98777	105.68421...
Canada	9349	92533	49.173913...
Canada	10489	92955	77.304347...
Germany	821	92859	94.409090...
Germany	2109	96843	55.142857...
Germany	6815	96547	38.523009...
Germany	10430	98964	105.14285...
India	9432	666666	5.6363636...
Montenegro	8588	60544	19.266666...
South...	1501	160800	39.022727...
South...	4947	69572	85.791666...
South...	5830	86857	96.090909...
South...	6036	89120	101.26315...
South...	7872	86836	24.217391...
South...	8318	90300	91.473684...
South...	9220	91290	99.4...

Hình 173. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



Hình 174. Truy vấn Pivot Excel

3.2.11 Thống kê số lượng đơn hàng đã thanh toán qua catalog và thu nhập hàng năm của Top10 Khách hàng độc thân và mới tốt nghiệp (Tăng dần theo số lượng đơn hàng)

Truy vấn bằng Visual Studio 2019

```

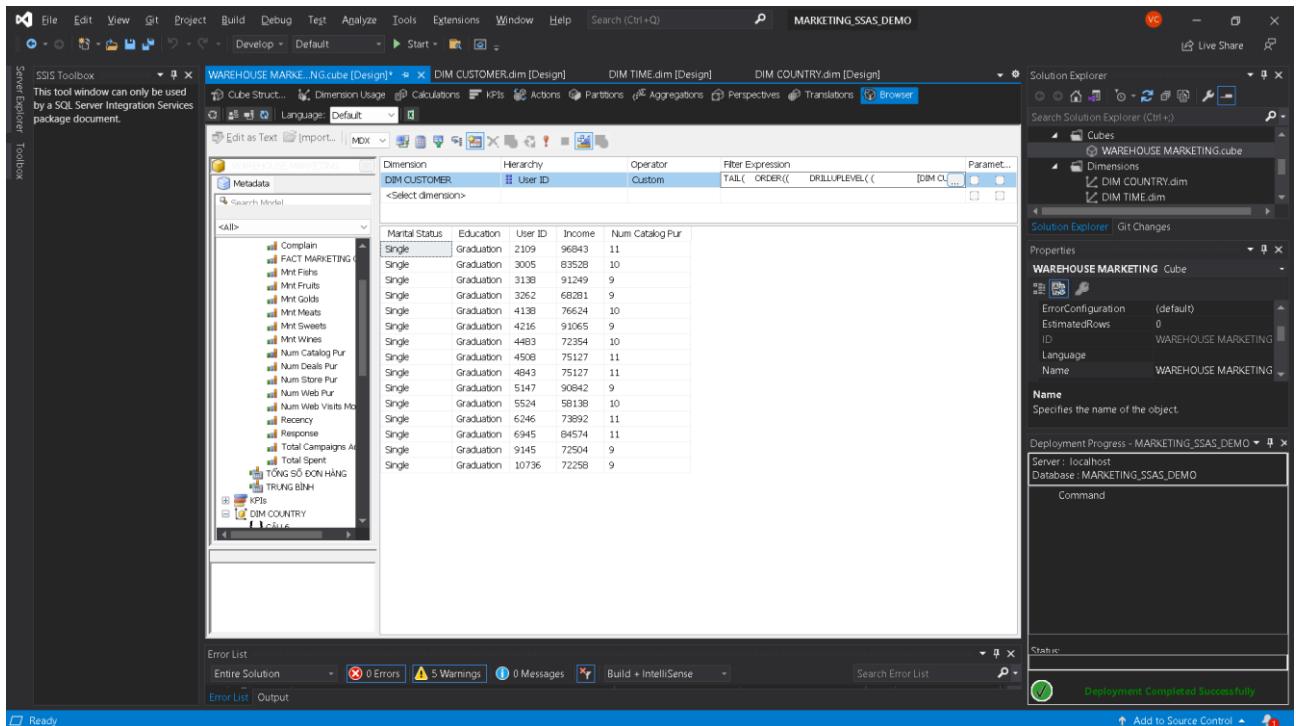
Name: [CÂU 11]

Expression:
TAIL(
    ORDER(
        DRILLUPLEVEL( (
            [DIM CUSTOMER].[Marital Status].&[Single],
            [DIM CUSTOMER].[Education].&[Graduation]),
            [DIM CUSTOMER].[Marital Status].[Marital Status])*
            [DIM CUSTOMER].[User ID].[User ID]),[Measures].[Num Catalog Pur],ASC),15)

```

The screenshot shows the Microsoft Analysis Services Scripting Editor in Visual Studio 2019. The left sidebar shows a 'Script Organizer' with a tree view of calculated members, including '[CÂU 1]', '[CÂU 2]', '[CÂU 3]', '[CÂU 4]', '[CÂU 5]', '[CÂU 6]', '[CÂU 7]', '[CÂU 8]', '[CÂU 9]', '[CÂU 10]', '[CÂU 11]' (which is currently selected), and '[CÂU 12]'. The main pane shows the script for '[CÂU 11]'. The script uses DAX functions like TAIL, ORDER, and DRILLUPLEVEL to calculate the top 10 customers based on their marital status and education level, ordered by the number of catalog purchases. The editor also includes tabs for Calculations, KPIs, Actions, Partitions, Aggregations, Perspectives, Translations, and Browser. The status bar at the bottom shows 'Ln: 1 Ch: 1 SPC CRLF'.

Hình 175. Câu truy vấn 11



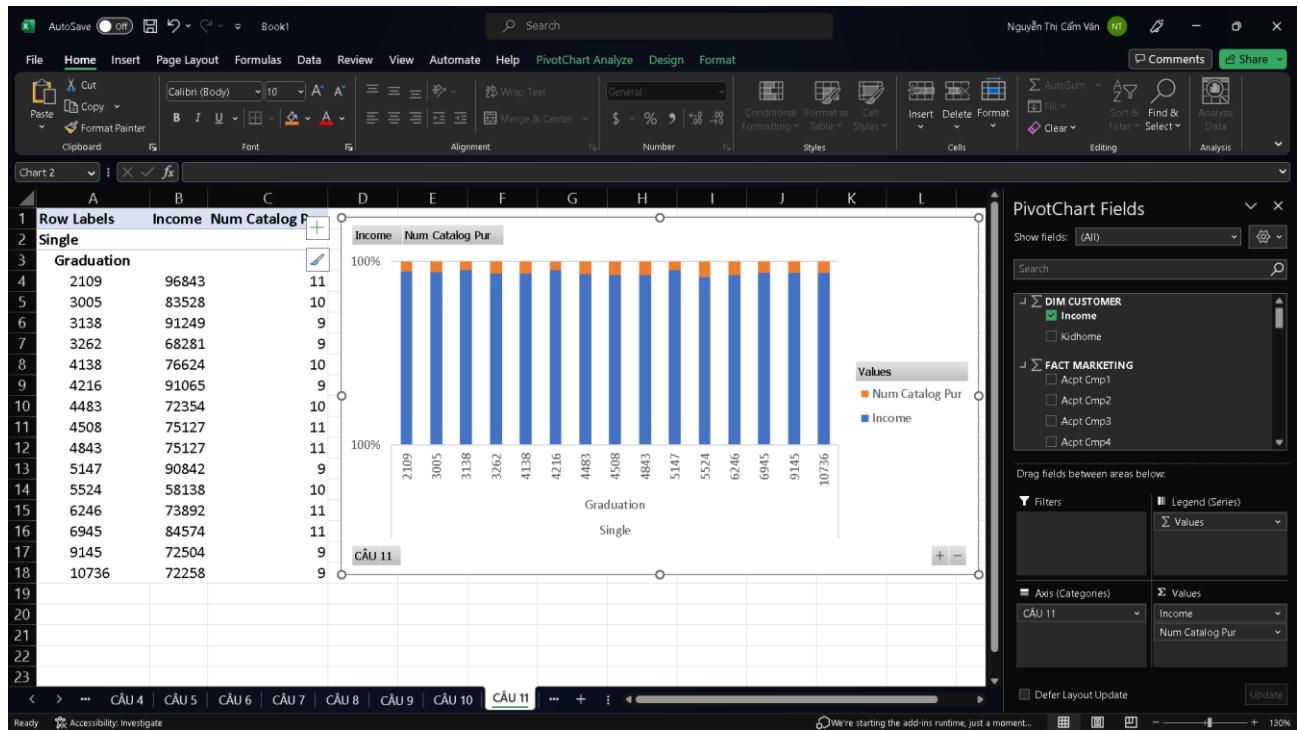
Hình 176. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server

```
-- (TẤNG ĐẦN THEO SỐ LƯỢNG ĐƠN HÀNG)
SELECT {[Measures].[Income]}, {[Measures].[Num Catalog Pur]} ON 0,
      TAIL (
      ORDER (
      (DRILLUPLEVEL(([DIM CUSTOMER].[Marital Status].&[Single],
      [DIM CUSTOMER].[Education].&[Graduation]),
      [DIM CUSTOMER].[Marital Status].[Marital Status]),
      [DIM CUSTOMER].[User ID].[User ID]),
      [Measures].[Num Catalog Pur],ASC),15) ON 1
FROM [WAREHOUSE MARKETING];
```

Hình 177. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel



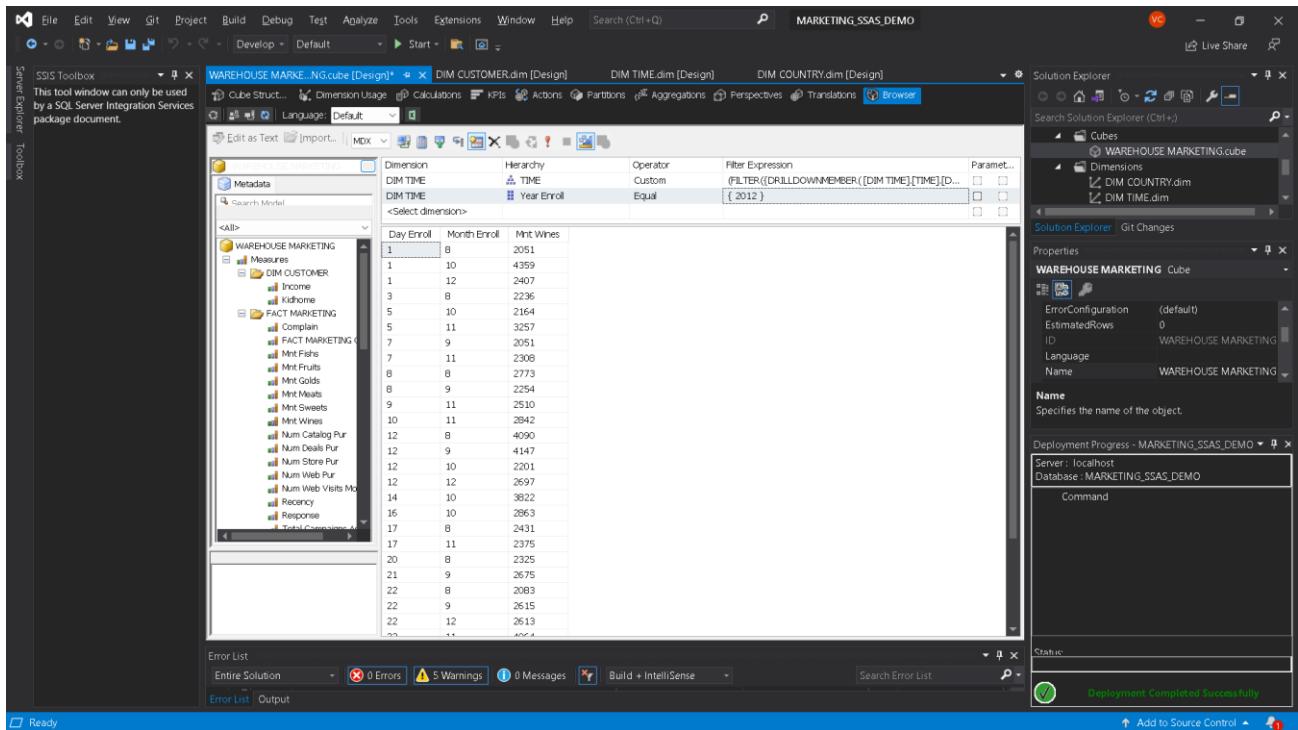
Hình 178. Truy vấn Pivot Excel

3.2.12 Truy vấn DrillDown các tháng của năm 2012 có lượng tiêu thụ các mặt hàng về rượu nhiều hơn 2000\$

Truy vấn bằng Visual Studio 2019

```
[CÂU 12]
[DIM TIME].[Month Enroll].[Month Enroll]* 
(FILTER({DRILLDOWNMEMBER( [DIM TIME].[TIME].[Day Enroll], [DIM TIME].[TIME].[Month Enroll].MEMBERS)}, 
[Measures].[Mnt Wines] >= 2000 ))
```

Hình 179. Câu truy vấn 12



Hình 180. Truy vấn Visual Studio 2019

Truy vấn bằng MDX – SQL Server

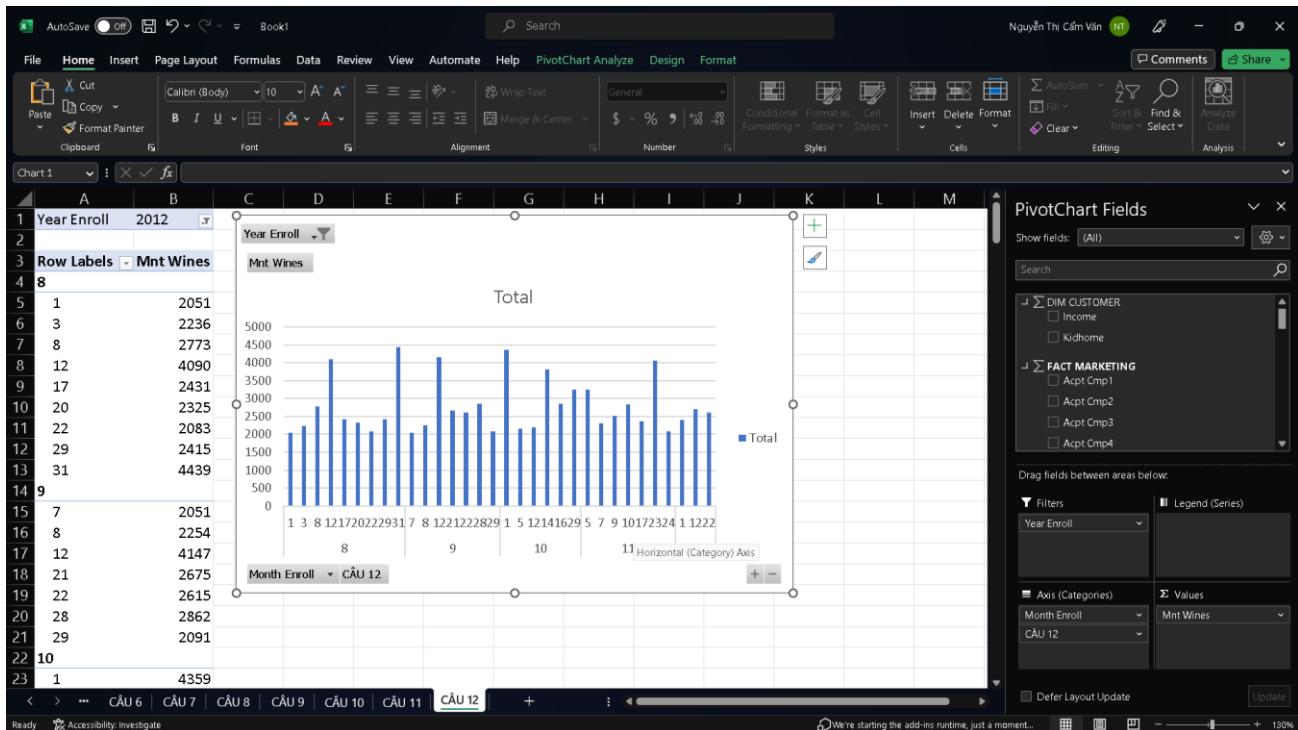
```

//CÂU 12: TRUY VẤN DRILLDOWN CÁC THÁNG CỦA NĂM 2012 CÓ LƯỢNG TIÊU THỤ CÁC MẶT HÀNG VỀ RƯỢU NHIỀU
SELECT [Measures].[Mnt Wines] ON 0,
[DIM TIME].[Month Enroll].[Month Enroll]*
(FILTER (
    [DRILLDOWNMEMBER([DIM TIME].[TIME].[Day Enroll],
    [DIM TIME].[TIME].[Month Enroll].MEMBERS)},
    [Measures].[Mnt Wines] >= 2000)) ON 1
FROM [WAREHOUSE MARKETING]
WHERE [DIM TIME].[Year Enroll].&[2012];

```

Hình 181. Truy vấn MDX SQL Server

Truy vấn bằng Pivot Excel

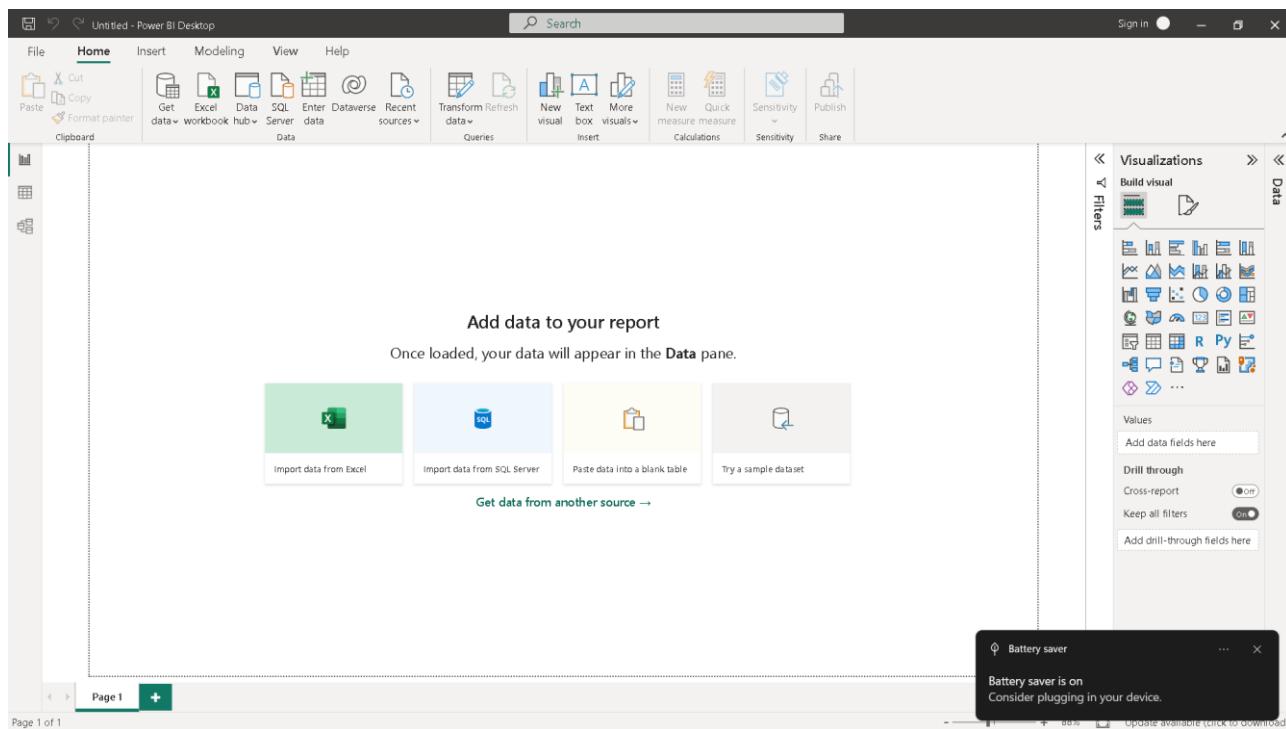


Hình 182. Truy vấn Pivot Excel

CHƯƠNG 4: QUÁ TRÌNH REPORT – POWER BI

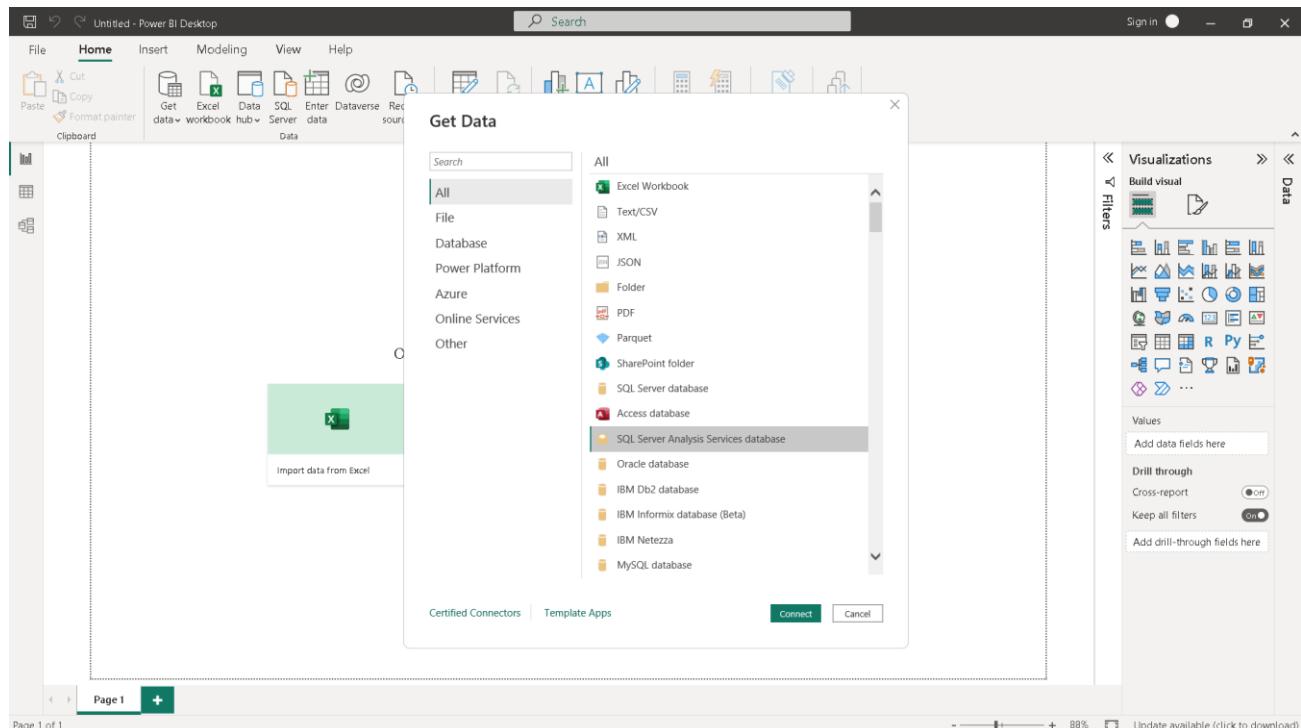
4.1 Tạo Project vs Power BI

Mở giao diện chính của Power BI



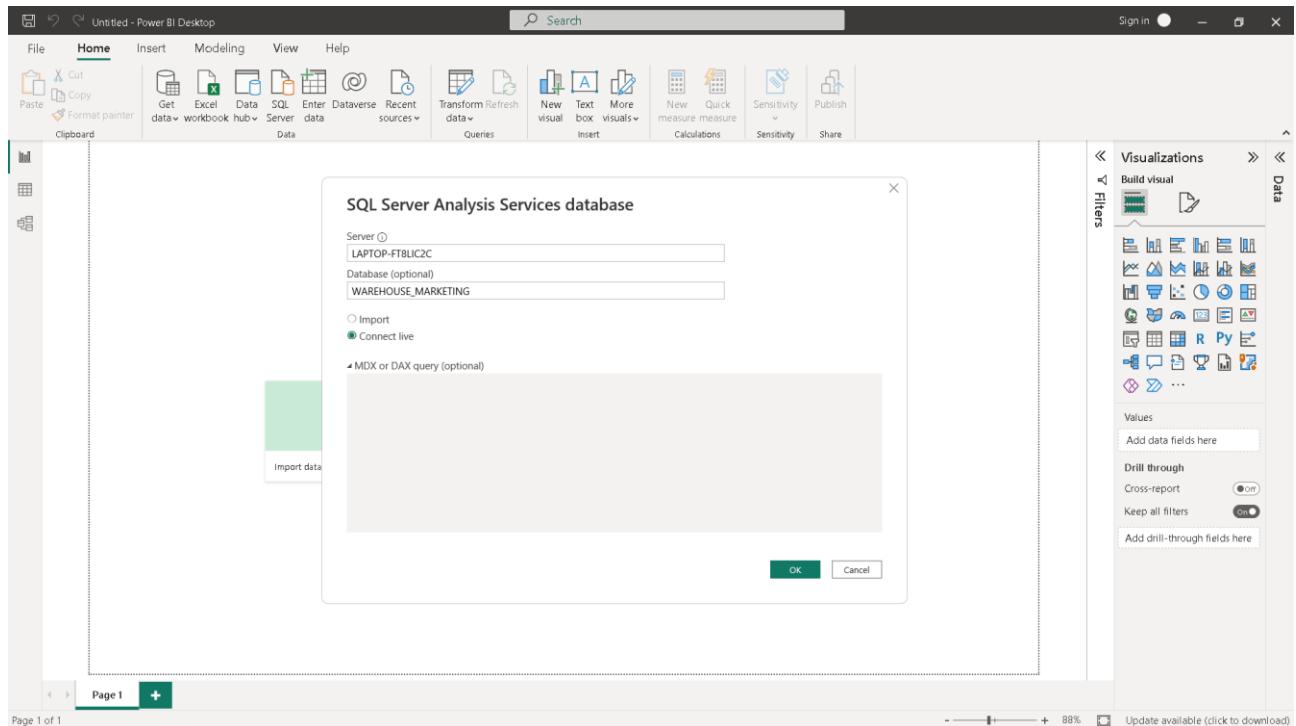
Hình 183. Giao diện Power BI

Chọn Get Data -> Chọn SQL Server Analysis Services Database



Hình 184. SQL Server Analysis Services DataBase

Điền tên Server máy -> Chọn Import -> Điền câu lệnh MDX tương ứng

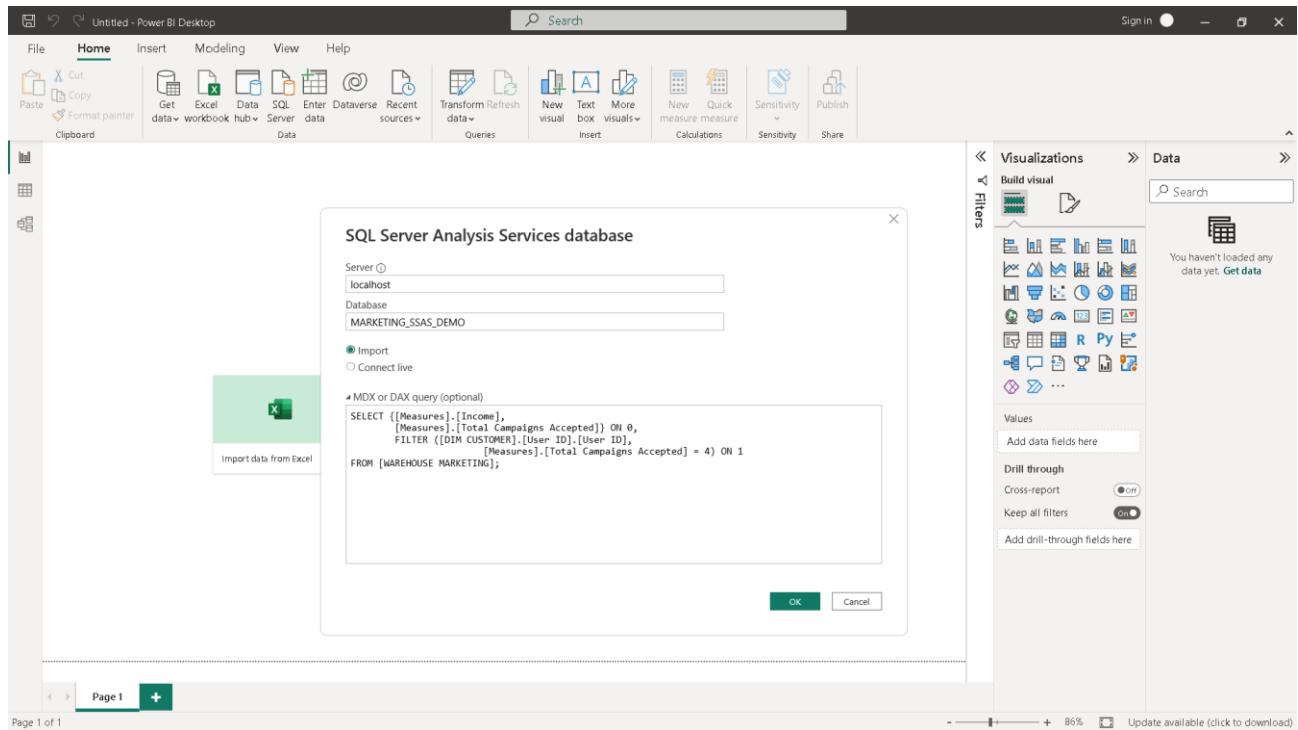


Hình 185. Import câu lệnh MDX

4.2 Report Power BI

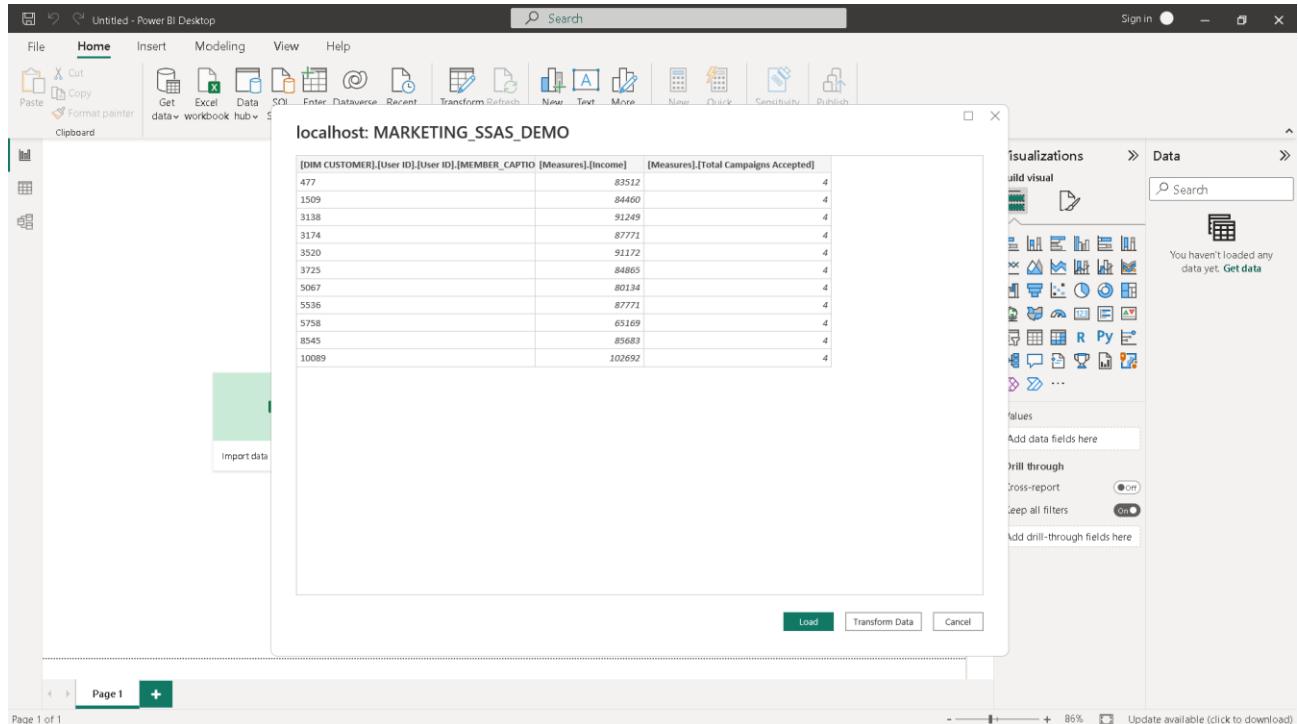
4.2.1 Thống kê thu nhập hàng năm của Khách hàng đã chấp nhận 4 chiến dịch Marketing

Bước 1: Điene câu truy vấn MDX



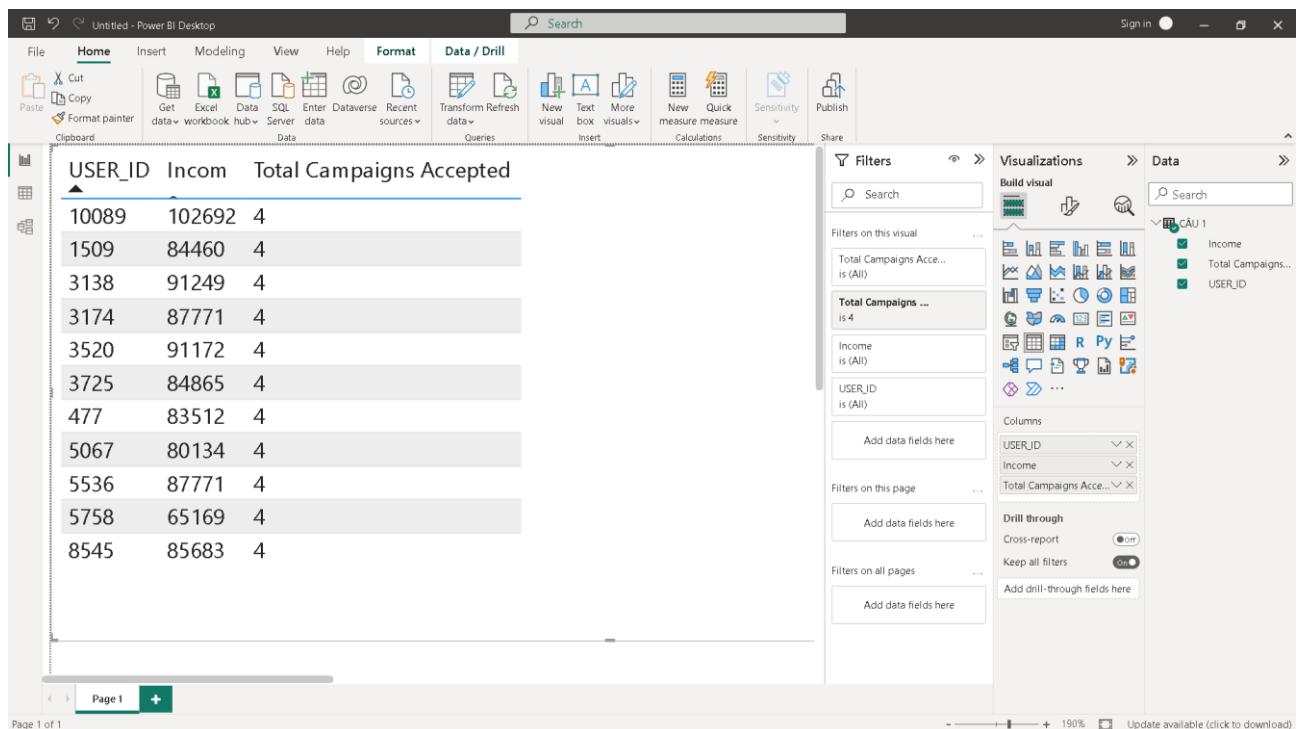
Hình 186. Điện câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu



Hình 187. Transform Data

Bước 3: Chọn Table Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn



Hình 188. Report Câu truy vấn 1

4.2.2 Thống kê Top 10 Khách hàng có thu nhập hằng năm lớn nhất

Bước 1: Điền câu truy vấn MDX

Server ⓘ
localhost

Database
MARKETING_SSAS_DEMO

Import
 Connect live

MDX or DAX query (optional)

```
SELECT [Measures].[Income] ON 0,
HEAD (
    ORDER ([DIM CUSTOMER].[User ID].[User ID].MEMBERS,
    [Measures].[Income], DESC), 10) ON 1
FROM [WAREHOUSE MARKETING];
```

OK **Cancel**

Hình 189. Điền câu truy vấn MDX

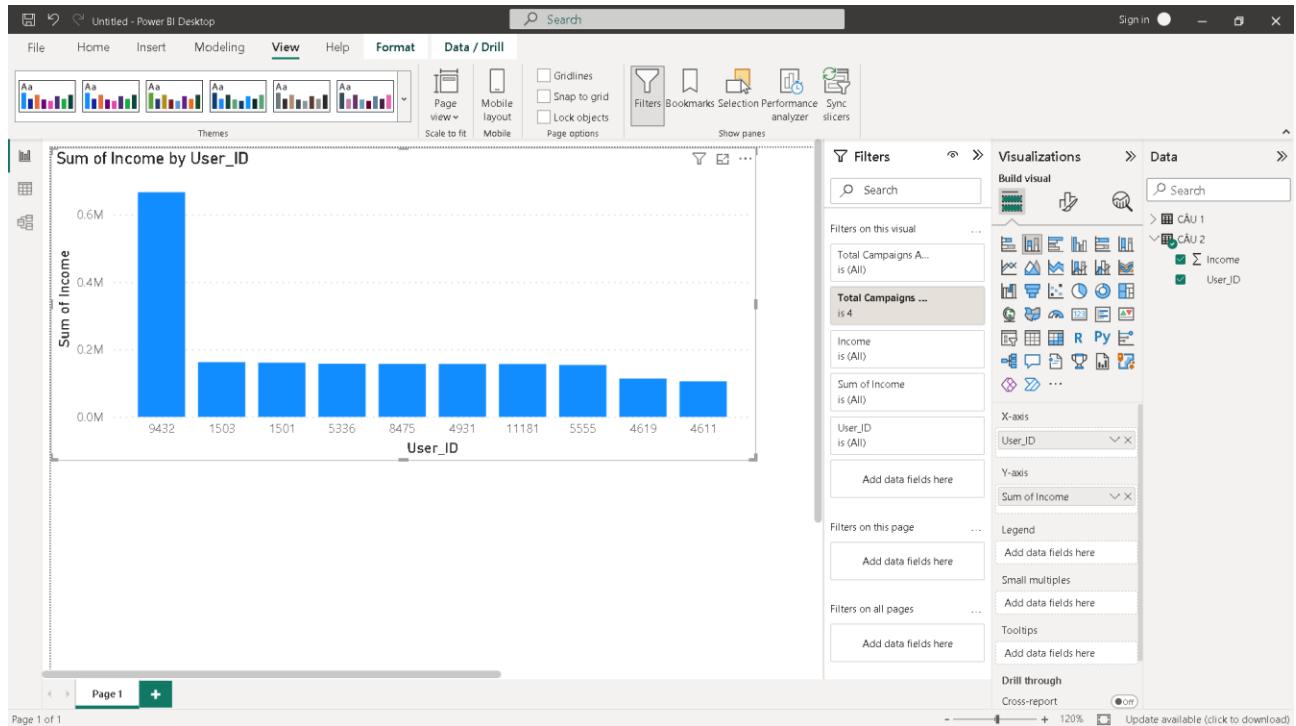
Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

localhost: MARKETING_SSAS_DEMO

[DIM CUSTOMER].[User ID].[User ID].[MEMBER_CAPTION]	[Measures].[Income]
9432	666666
1503	162397
1501	160803
5336	157733
8475	157243
4931	157146
11181	156924
5555	153924
4619	113734
4611	105471

Hình 190. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn



Hình 191. Report câu truy vấn 2

**4.2.3 Thống kê số tiền đã mua hàng của top 10 Khách hàng có thu nhập
hàng năm lớn hơn 30000\$**

Bước 1: Điền câu truy vấn MDX

- Import
 Connect live

▲ MDX or DAX query (optional)

```

SELECT {[Measures].[Income],
          [Measures].[Total Spent]} ON 0,
HEAD({
    FILTER({
        ORDER([DIM CUSTOMER].[User ID].[User ID],
              [Measures].[Total Spent], DESC),
        [Measures].[Income] >= 30000}),10) ON 1
FROM [WAREHOUSE MARKETING];
  
```

OK

Cancel

Hình 192. Điện câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

localhost: MARKETING_SSAS_DEMO

[DIM CUSTOMER].[User ID].[User ID].[MEMBER_CAPTION]	[Measures].[Income]	[Measures].[Total Spent]
5350	90638	2525
5735	90638	2525
1763	87679	2524
4580	75759	2486
4475	69098	2440
5453	90226	2352
10133	93790	2349
9010	83151	2346
5386	94384	2302
6024	94384	2302

Hình 193. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn

The screenshot shows the Power BI Desktop interface. On the left, there is a table visual titled "User ID Income Total Spent" containing the following data:

User ID	Income	Total Spent
5386	94384	2302
6024	94384	2302
9010	83151	2346
10133	93790	2349
5453	90226	2352
4475	69098	2440
4580	75759	2486
1763	87679	2524
5350	90638	2525
5735	90638	2525

The right side of the screen displays the "Filters", "Visualizations", and "Data" panes. The "Filters" pane shows filters applied to the report. The "Visualizations" pane lists various chart types. The "Data" pane shows the data fields used in the report: CÁU 1, CÁU 2, CÁU 3, Income, Total Spent, and User ID.

Hình 194. Report câu truy vấn 3

4.2.4 Thông kê những Khách hàng đã chi nhiều hơn 150\$ cho mặt hàng bánh kẹo nhưng số lượng trẻ em trong nhà nhỏ hơn 2

Bước 1: Điene câu truy vấn MDX

The screenshot shows the MDX Editor dialog box. It includes fields for "Server" (localhost), "Database" (MARKETING_SSAS_DEMO), and connection options ("Import" selected). Below these, there is a section for "MDX or DAX query (optional)" containing the following MDX code:

```

SELECT {[Measures].[Kidhome], [Measures].[Mnt Sweets]} ON 0,
EXCEPT (
    {FILTER ([DIM CUSTOMER].[User ID].CHILDREN ,
        [Measures].[Mnt Sweets] > 150)},
    {FILTER ([DIM CUSTOMER].[User ID].CHILDREN ,
        [Measures].[Kidhome] = 2)}) ON 1
FROM [WAREHOUSE MARKETING];

```

At the bottom right of the dialog box are "OK" and "Cancel" buttons.

Hình 195. Điene câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

localhost: MARKETING_SSAS_DEMO

[DIM CUSTOMER].[User ID].[User ID].[MEMBER_CAPTION]	[Measures].[Kidhome]	[Measures].[Mnt Sweets]
123	0	151
313	0	163
500	0	194
528	0	165
830	0	176
961	0	195
1077	0	172
1137	0	156
1399	0	165
1411	0	153
1440	0	188
1524	0	160
1577	0	175
1826	0	189
1964	0	196
2021	0	192
2147	0	189
2798	0	176
3194	0	182
3283	0	185

The data in the preview has been truncated due to size limits.

Hình 196. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn

The screenshot shows the Power BI desktop application interface. On the left, there is a table visualization titled "User ID Income Total Spent" with the following data:

User ID	Income	Total Spent
4475	69098	2440
4580	75759	2486
9010	83151	2346
1763	87679	2524
5453	90226	2352
5350	90638	2525
5735	90638	2525
10133	93790	2349
5386	94384	2302
6024	94384	2302

The top ribbon menu includes options like Cut, Copy, Format painter, Get data, Data, Enter data, Recent sources, Transform Refresh data, New visual, Text box, More visuals, Quick measure, Calculations, Sensitivity, and Publish. To the right of the table, there are sections for Filters, Visualizations, and Data, each containing various filter and visualization tools.

Hình 197. Report câu truy vấn 4

4.2.5 Thống kê Top5 Khách hàng có tổng số đơn mua hàng nhiều nhất trong năm 2014

Bước 1: Dièn câu truy vấn MDX

Server ①
localhost

Database
MARKETING_SSAS_DEMO

Import
 Connect live

▲ MDX or DAX query (optional)

```
WITH MEMBER [Measures].[TỔNG SỐ ĐƠN HÀNG] AS
[Measures].[Num Catalog Pur] + [Measures].[Num Deals Pur] + [Measures].[Num Store Pur] + [Measur
SELECT [Measures].[TỔNG SỐ ĐƠN HÀNG] ON COLUMNS,
GENERATE(
    DESCENDANTS ([DIM TIME].[TIME].[Year Enroll].&[2014].CHILDREN ,
        [DIM TIME].[TIME].[Month Enroll]),
    TOPCOUNT ([DIM TIME].[Month Enroll].CURRENTMEMBER*
        [DIM CUSTOMER].[User ID].CHILDREN, 5,
        [Measures].[TỔNG SỐ ĐƠN HÀNG])) ON ROWS
FROM [WAREHOUSE MARKETING];|
```

< **>**

OK **Cancel**

Hình 198. Dièn câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

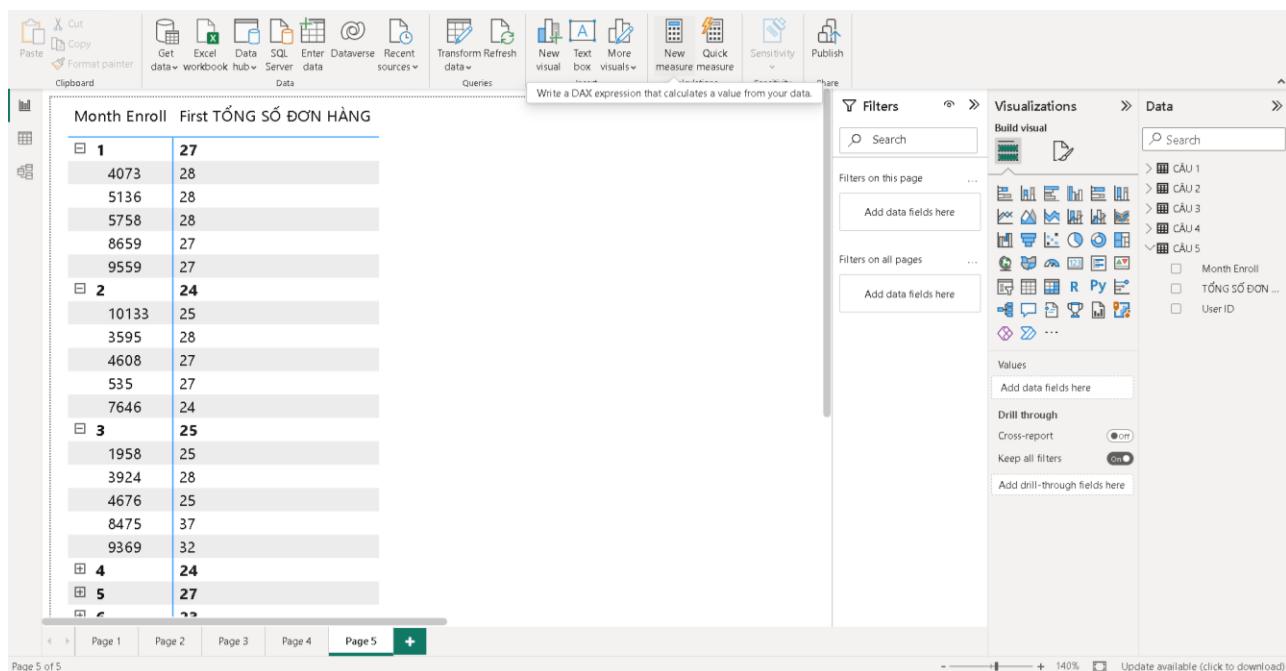
localhost: MARKETING_SSAS_DEMO

	[DIM TIME].[Month Enroll].[Month Enroll].[MEMBER_C]	[DIM CUSTOMER].[User ID].[User ID].[MEMBER_CAPTION]	[Measures].[TỔNG SỐ ĐƠN HÀNG]
1		4073	28
1		5136	28
1		5758	28
1		8659	27
1		9559	27
2		3595	28
2		535	27
2		4608	27
2		10133	25
2		7646	24
3		8475	37
3		9369	32
3		3924	28
3		1958	25
3		4676	25
4		1277	28
4		1446	28
4		1340	27
4		7875	26
4		4686	24

The data in the preview has been truncated due to size limits.

Hình 199. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn



Hình 200. Report câu truy vấn 5

4.2.6 Thống kê Top5 Quốc gia có số lượng mua hàng ở Store ít nhất

Bước 1: Điền câu truy vấn MDX

Server ⓘ
localhost

Database
MARKETING_SSAS_DEMO

Import
 Connect live

▲ MDX or DAX query (optional)

```
SELECT [Measures].[Num Store Pur] ON 0,
       BOTTOMCOUNT( [DIM COUNTRY].[Country].[Country].MEMBERS, 5,
                     [Measures].[Num Store Pur]) ON 1
FROM [WAREHOUSE MARKETING];
```

OK

Cancel

Hình 201. Điện câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

localhost: MARKETING_SSAS_DEMO

[DIM COUNTRY].[Country].[Country].[MEMBER_CAPTION]	[Measures].[Num Store Pur]
Unknown	null
Montenegro	12
United States	616
Germany	647
India	843

Hình 202. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn

The screenshot shows the Power BI Desktop interface. On the left, there is a table visual titled "Num Store Pur" with columns "Country" and "Num Store Pur". The data includes Germany (647), India (843), Montenegro (12), United States (616), and Unknown. The "Column tools" ribbon tab is selected, showing options like "Format", "Summarization", "Sort by column", and "Data category". To the right, there are several toolbars: "Filters", "Visualizations", "Data", and "Columns". The "Visualizations" toolbar contains icons for various chart types. The "Data" toolbar shows a tree view of data fields, including "Country" and "Num Store Pur". The bottom of the screen shows a navigation bar with "Page 6" highlighted.

Hình 203. Report câu truy vấn 6

4.2.7 Thống kê ngày có lượng truy cập vào Web ít nhất trong tháng 8 năm 2012

Bước 1: Điền câu truy vấn MDX

Server ⓘ
localhost

Database
MARKETING_SSAS_DEMO

Import
 Connect live

MDX or DAX query (optional)

```
SELECT [Measures].[Num Web Visits Month] ON 0,
       BOTTOMSUM ({}
                  DESCENDANTS([DIM TIME].[TIME].[Year Enroll].&[2012].&[8],
                             [DIM TIME].[TIME].[Day Enroll])),1,
                  [Measures].[Num Web Visits Month]) ON 1
FROM [WAREHOUSE MARKETING];
```

OK

Cancel

Hình 204. Điền câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

localhost: MARKETING_SSAS_DEMO

[DIM TIME].[TIME].[Year Enroll].[MEMBER_CAPTION]	[DIM TIME].[TIME].[Month Enroll].[MEMBER_CAPTION]	[DIM TIME].[TIME].[Day Enroll].[MEMBER_CAPTION]
2012	8	10

Hình 205. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn

The screenshot shows the Power BI Desktop interface. On the left, there's a table visualization with four columns: Year Enroll, Month Enroll, Day Enroll, and Num Web Visits Month. The data row shows 2012, 8, 10, and 6 respectively. To the right of the visualization are several panels: 'Filters' containing 'Add data fields here' and 'Filters on all pages' sections; 'Visualizations' showing a list of visualizations including 'CẤU 6' through 'CẤU 7'; 'Data' showing a list of values including 'Day Enroll', 'Month Enroll', 'Num Web Visits ...', and 'Year Enroll'; and 'Drill through' and 'Cross-report' options. At the bottom, there are navigation buttons for pages 1 through 7, and the status bar indicates 'Page 7 of 7'.

Hình 206. Report câu truy vấn 7

4.2.8 Thống kê những Khách hàng vừa mua hàng ở Web vừa mua hàng ở Store

Bước 1: Điền câu truy vấn MDX

Server ⓘ
localhost

Database
MARKETING_SSAS_DEMO

Import
 Connect live

▲ MDX or DAX query (optional)

```
SELECT {[Measures].[Num Web Pur], [Measures].[Num Store Pur]} ON 0,
INTERSECT (
    {FILTER ([DIM CUSTOMER].[User ID].[User ID],
        [Measures].[Num Store Pur] > 0)},
    {FILTER ([DIM CUSTOMER].[User ID].[User ID],
        [Measures].[Num Web Pur] > 0)}) ON 1
FROM [WAREHOUSE MARKETING] ;
```

OK

Cancel

Hình 207. Điền câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

localhost: MARKETING_SSAS_DEMO

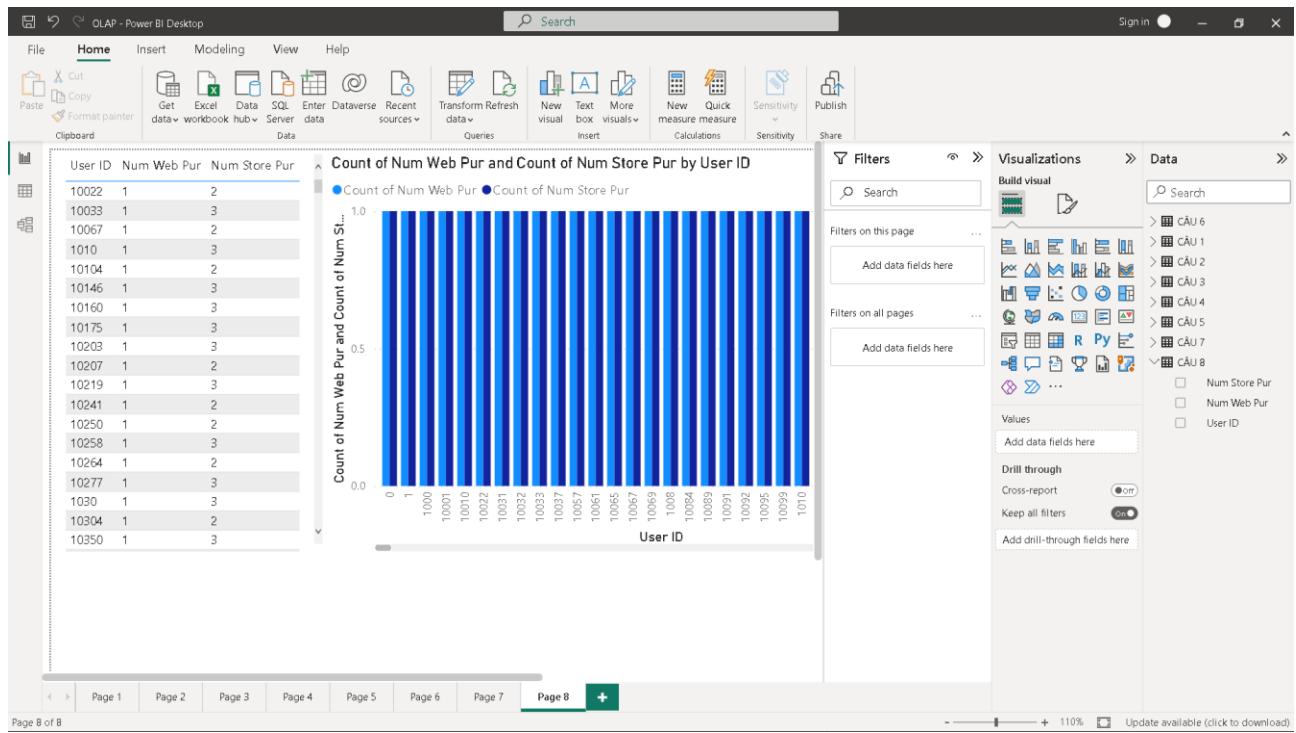
[DIM CUSTOMER].[User ID].[User ID].[MEMBER_CAPTION]	[Measures].[Num Web Pur]	[Measures].[Num Store Pur]
0	3	9
1	7	7
9	3	2
13	1	3
17	6	7
20	2	4
22	6	5
24	3	4
25	6	7
35	2	3
48	4	3
49	1	2
55	7	9
67	2	4
75	8	5
78	1	2
87	2	3
89	2	4
92	3	4
113	1	3

 ⓘ The data in the preview has been truncated due to size limits.

Hình 208. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính

theo mong muốn



Hình 209. Report Câu truy vấn 8

4.2.9 Thông kê những Khách hàng có tổng số đơn hàng cao nhất tại các Quốc gia

Bước 1: Điền câu truy vấn MDX

Database

MARKETING_SSAS_DEMO

Import

Connect live

▼ MDX or DAX query (optional)

```
SELECT [Measures].[TỔNG SỐ ĐƠN HÀNG] ON 0,
GENERATE([DIM COUNTRY].[Country].[Country],
TOPCOUNT
    ([DIM COUNTRY].[Country].CURRENTMEMBER*
    [DIM CUSTOMER].[User ID].CHILDREN, 1,
    [Measures].[TỔNG SỐ ĐƠN HÀNG])) ON 1
FROM [WAREHOUSE MARKETING];
```

OK **Cancel**

Hình 210. Điền câu truy vấn MDX

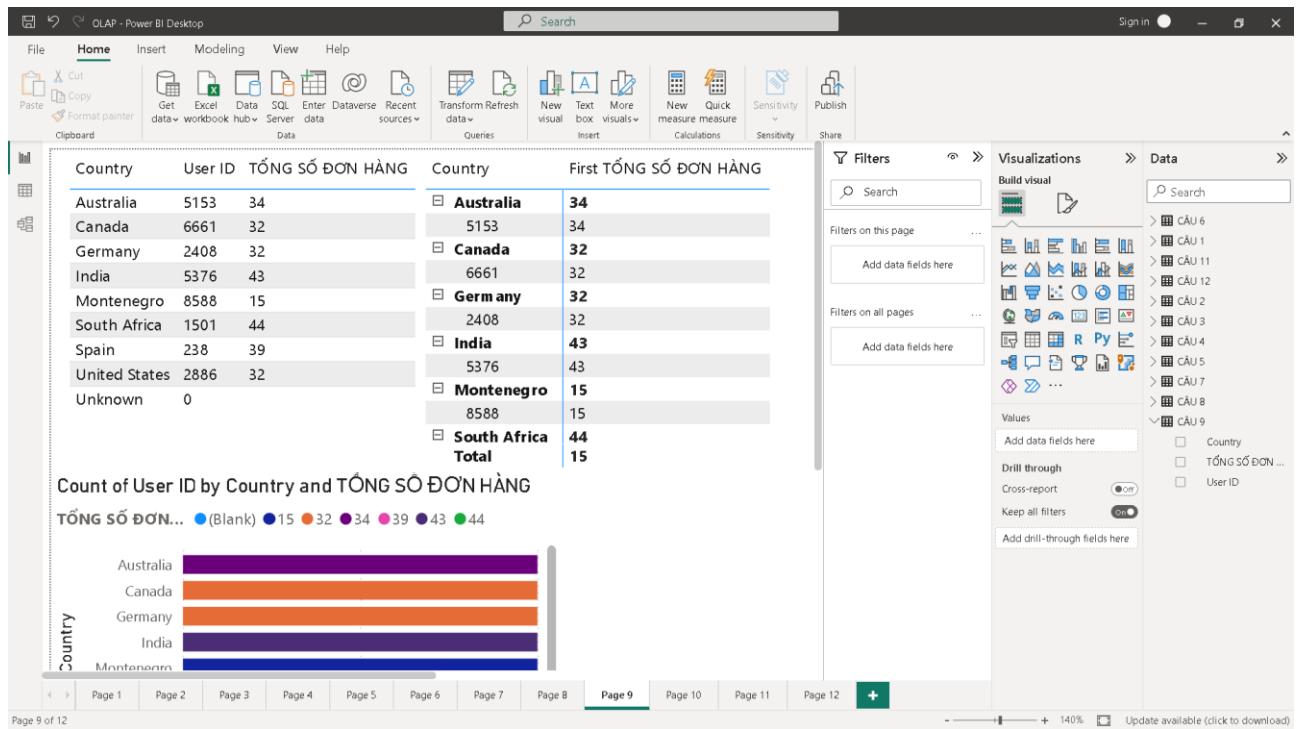
Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

localhost: MARKETING_SSAS_DEMO

[DIM COUNTRY].[Country].[Country].[MEMBER_CAPTION]	[DIM CUSTOMER].[User ID].[User ID].[MEMBER_CAPTION]	[Measures].[TỔNG SỐ ĐƠN HÀNG]
Australia	5153	34
Canada	6661	32
Germany	2408	32
India	5376	43
Montenegro	8588	15
South Africa	1501	44
Spain	238	39
United States	2886	32
Unknown	0	null

Hình 211. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn



Hình 212. Report câu truy vấn 9

4.2.10 Tính trung bình số tiền đã chi cho tất cả các mặt hàng của Top5 Khách hàng có thu nhập hàng năm cao nhất theo từng Quốc gia

Bước 1: Dièn câu truy vấn MDX

Server ⓘ
localhost

Database
MARKETING_SSAS_DEMO

Import
 Connect live

▲ MDX or DAX query (optional)

```
WITH MEMBER [Measures].[TRUNG BÌNH] AS [Measures].[Total Spent]/[Measures].[TỔNG SỐ ĐƠN HÀNG]
    SELECT {[Measures].[Income], [Measures].[TRUNG BÌNH]} ON 0,
        GENERATE([DIM COUNTRY].[Country].[Country],
            TOPPERCENT({[DIM COUNTRY].[Country].CURRENTMEMBER*
                FILTER([DIM CUSTOMER].[User ID].[User ID],
                    [Measures].[TỔNG SỐ ĐƠN HÀNG] <> 0)},5,
            [Measures].[Income])) ON 1
FROM [WAREHOUSE MARKETING];
```

OK Cancel

Hình 213. Điền câu truy vấn MDX

4.2.11 Thông kê số lượng đơn hàng đã thanh toán qua catalog và thu nhập hàng năm của Top10 Khách hàng độc thân và mới tốt nghiệp (Tăng dần theo số lượng đơn hàng)

Bước 1: Điền câu truy vấn MDX

Database
MARKETING_SSAS_DEMO

Import
 Connect live

▲ MDX or DAX query (optional)

```
SELECT {[Measures].[Income], [Measures].[Num Catalog Pur]} ON 0,
    TAIL (
        ORDER (
            (DRILLUPLEVEL(([DIM CUSTOMER].[Marital Status]&[Single],
                [DIM CUSTOMER].[Education]&[Graduation]),
                [DIM CUSTOMER].[Marital Status].[Marital Stat
                [DIM CUSTOMER].[User ID].[User ID]),
                [Measures].[Num Catalog Pur],ASC),15) ON 1
FROM [WAREHOUSE MARKETING];
```

◀ ▶

OK Cancel

Hình 214. Điền câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

localhost: MARKETING_SSAS_DEMO

[DIM CUSTOMER].[Marital Status].[Marital Status].[MEN] [DIM CUSTOMER].[Education].[Education].[MEMBER_CAT] [DIM CUSTOMER].[User ID].[User ID].[MEMBER_CAT]		
Single	Graduation	3138
Single	Graduation	3262
Single	Graduation	4216
Single	Graduation	5147
Single	Graduation	9145
Single	Graduation	10736
Single	Graduation	3005
Single	Graduation	4138
Single	Graduation	4483
Single	Graduation	5524
Single	Graduation	2109
Single	Graduation	4508
Single	Graduation	4843
Single	Graduation	6246
Single	Graduation	6945

Hình 215. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn

The screenshot shows the Power BI Desktop interface. On the left, a table displays data from a query:

Marital Status	Education	User ID	Income	Num	Marital Status	First Income
Single	Graduation	5524	58138	10	Single	58138
Single	Graduation	3262	68281	9	Graduation	58138
Single	Graduation	10736	72258	9	10	58138
Single	Graduation	4483	72354	10	5524	58138
Single	Graduation	9145	72504	9	4483	72354
Single	Graduation	6246	73892	11	4138	76624
Single	Graduation	4508	75127	11	3005	83528
Single	Graduation	4843	75127	11	9	68281
Single	Graduation	4138	76624	10	3262	68281
Single	Graduation	3005	83528	10	10736	72258
Single	Graduation	6945	84574	11	9145	72504
Single	Graduation	5147	90842	9	5147	90842
Single	Graduation	4216	91065	9	4216	91065
Single	Graduation	3138	91249	9	3138	91249
Single	Graduation	2109	96843	11	11	73892
					Total	58138

On the right, the Visualizations pane is open, showing various chart and table options. A drill-through filter is applied to the 'First Income' column.

Hình 216. Report câu truy vấn 11

4.2.12 Truy vấn DrillDown các tháng của năm 2012 có lượng tiêu thụ các mặt hàng về rượu nhiều hơn 2000\$

Bước 1: Dièn câu truy vấn MDX

SQL Server Analysis Services database

Server ⓘ
localhost

Database
MARKETING_SSAS_DEMO

Import
 Connect live

▪ MDX or DAX query (optional)

```
SELECT [Measures].[Mnt Wines] ON 0,
[DIM TIME].[Month Enroll].[Month Enroll]*
(FILTER (
    {DRILLDOWNMEMBER([DIM TIME].[TIME].[Day Enroll],
    [DIM TIME].[TIME].[Month Enroll].MEMBERS)},
    [Measures].[Mnt Wines] >= 2000)) ON 1
FROM [WAREHOUSE MARKETING]
WHERE [DIM TIME].[Year Enroll].&[2012];
```

OK **Cancel**

Hình 217. Dièn câu truy vấn MDX

Bước 2: Chọn Transform Data để chỉnh sửa kiểu dữ liệu

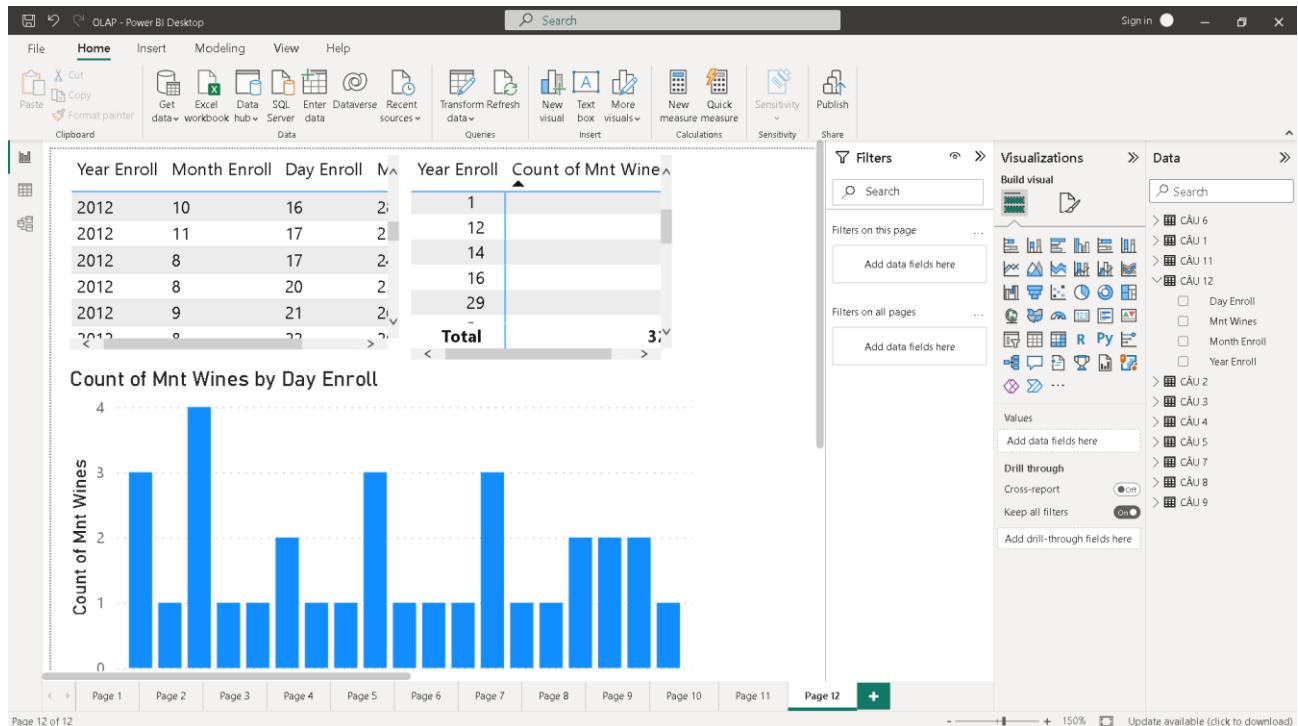
localhost: MARKETING_SSAS_DEMO

[DIM TIME].[Month Enroll].[Month Enroll].[MEMBER_C	[DIM TIME].[TIME].[Year Enroll].[MEMBER_CAPTION]	[DIM TIME].[TIME].[Month Enroll].[MEMBER_CA
8	2012	8
8	2012	8
8	2012	8
8	2012	8
8	2012	8
8	2012	8
8	2012	8
8	2012	8
9	2012	9
9	2012	9
9	2012	9
9	2012	9
9	2012	9
10	2012	10
10	2012	10
10	2012	10
10	2012	10

i The data in the preview has been truncated due to size limits.

Hình 218. Transform Data

Bước 3: Chọn Visualization -> Chọn các thuộc tính cần thiết -> Stack các thuộc tính theo mong muốn



Hình 219. Report câu truy vấn 12

CHƯƠNG 5: QUÁ TRÌNH DATAMINING

5.1 Xử lý dữ liệu

Bước 1: Quan sát dữ liệu

❖ Sau khi quan sát và kiểm tra dữ liệu, thì chúng ta có một vài điểm cần lưu ý:

Cần chuyển dữ liệu từ kiểu object sang số, quan sát dữ liệu thì chúng ta thấy thuộc tính **Education, Marital_status** và **Country**.

Education	Marital_Status	Country
Graduation	Single	Spain
Graduation	Single	Canada
Graduation	Together	United States
Graduation	Together	Australia
PhD	Married	Spain
Master	Together	Spain
Graduation	Divorced	Germany
PhD	Married	Spain
PhD	Together	United States

Hình 220. Thuộc tính Education - Marital_status - Country

Cần xoá các dữ liệu null để tránh bị lỗi khi đưa các tập dữ liệu và model.

Bước 2: Chuyển từ kiểu dữ liệu object sang số

Kiểm tra các giá trị có trong 3 thuộc tính **Education, Marital_status** và **Country**.

```
# Kiểm tra các giá trị có trong Education
df.Education.value_counts()

✓ 0.0s

Graduation    1127
PhD           486
Master         370
2n Cycle      203
Basic          54
Name: Education, dtype: int64
```

Hình 221. Kiểm tra Thuộc tính Education

```
# Kiểm tra các giá trị có trong Marital_Status
df.Marital_Status.value_counts()

✓ 0.0s

Married      864
Together     580
Single       480
Divorced     232
Widow        77
Alone         3
Absurd        2
YOLO          2
Name: Marital_Status, dtype: int64
```

Hình 222. Kiểm tra Thuộc tính Marital_status

```
# Kiểm tra các giá trị có trong Country
df.Country.value_counts()

✓ 0.0s

Spain        1095
South Africa    337
Canada        268
Australia      160
India          148
Germany        120
United States    109
Montenegro      3
Name: Country, dtype: int64
```

Hình 223. Kiểm tra Thuộc tính Country

Tiến hành chuyển đổi sang dạng số.

```
# Chuyển đổi sang dạng số
mar_type_Education = {'Graduation': 0, 'PhD':1, 'Master':2 , '2n Cycle':3, 'Basic':4}
mar_type_Marital_Status = {'Married': 0, 'Together':1, 'Single':2 , 'Divorced':3, 'Widow':4, 'Alone':5, 'Absurd':6, 'YOLO':7}
mar_Country = {'Spain': 0, 'South Africa':1, 'Canada':2 , 'Australia':3, 'India':4, 'Germany':5, 'United States':6, 'Montenegro':7}
df.Education = df.Education.map(mar_type_Education)
df.Marital_Status = df.Marital_Status.map(mar_type_Marital_Status)
df.Country = df.Country.map(mar_Country)

✓ 0.0s
```

Hình 224. Tiến hành chuyển đổi từ object sang dạng số

Bước 3: Xoá các giá trị null

Kiểm tra và đếm số lượng null của từng thuộc tính.

```
# Đếm số lượng null của từng thuộc tính  
df.isnull().sum()  
✓ 0.0s  
  
User_ID          0  
Year_Birth        0  
Education         0  
Marital_Status   0  
Income            24  
Kidhome           0  
Teenhome           0  
Date_Enroll       0  
Recency            0  
MntWines          0  
MntFruits         0  
Mnt_Meats          0  
Mnt_Fishs          0  
Mnt_Sweets          0  
Mnt_Golds          0  
Num_DealsPur       0  
Num_WebPur          0  
Num_CatalogPur     0  
Num_StorePur       0  
Num_WebVisitsMonth 0  
Acpt_Cmp3          0  
Acpt_Cmp4          0  
Acpt_Cmp5          0  
Acpt_Cmp1          0  
Acpt_Cmp2          0  
Complain           0  
Response            0  
Country             0  
dtype: int64
```

Hình 225. Kiểm tra số lượng Null của tất cả thuộc tính

Tiến hành xoá những dữ liệu null và kiểm tra lại số lượng null sau khi xoá. Sau khi kiểm tra, chúng ta thấy dữ liệu null đã được xoá hết.

```
# Xóa các hàng có giá null của từng thuộc tính  
df.dropna(axis=0, how='any', inplace = True)
```

✓ 0.0s

```
# Kiểm tra lại số lượng các giá trị null  
df.isnull().sum().sum()
```

✓ 0.0s

Hình 226. Xóa và kiểm tra lại các giá trị Null

5.2 Phân lớp dữ liệu

Bước 1: Xác định dữ liệu đầu ra để tìm các đặc trưng

Chọn thuộc tính **Response** là dữ liệu đầu ra

```
# Lưu Response vào output  
output = df['Response']
```

✓ 0.0s

Hình 227. Xác định dữ liệu đầu ra

Tách 4 nhóm dữ liệu theo các thuộc tính liên quan gồm:

✚ **Nhóm Mnt :** Tổng số tiền khách hàng phải chi trả cho các mặt hàng trong 2 năm gần nhất.

```
# Tách theo nhóm Mnt  
df_Mnt = df[df.columns[8:13]]  
df_Mnt.head(5)
```

✓ 0.0s

	MntWines	MntFruits	Mnt_Meats	Mnt_Fishs	Mnt_Sweets
0	635.0	88.0	546.0	172.0	88.0
1	11.0	1.0	6.0	2.0	1.0
2	426.0	49.0	127.0	111.0	21.0
3	11.0	4.0	20.0	10.0	3.0
4	173.0	43.0	118.0	46.0	27.0

Hình 228. Tách nhóm Mnt

✚ **Nhóm Num:** Số lượng đơn hàng mà khách hàng đã thanh toán.

```
# Tách theo nhóm Num
df_Num = df[df.columns[14:18]]
df_Num.head(5)
```

✓ 0.0s

	Num_DealsPur	Num_WebPur	Num_CatalogPur	Num_StorePur
0	3	8	10	4
1	2	1	1	2
2	1	8	2	10
3	2	2	0	4
4	5	5	3	6

Hình 229. Tách nhóm Num

✚ Nhóm Accepted: Khách hàng phản hồi cho Marketing Campaign

```
# Tách theo nhóm Acpt
df_Acpt = df[df.columns[19:24]]
df_Acpt.head(5)
```

✓ 0.0s

	Acpt_Cmp3	Acpt_Cmp4	Acpt_Cmp5	Acpt_Cmp1	Acpt_Cmp2
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0

Hình 230. Tách nhóm Acpt

✚ Nhóm Info: Nhóm dữ liệu còn lại là thông tin khách hàng

```
# Tách theo nhóm các thông tin còn lại
df_info = df.copy()
df_info.drop(df.columns[8:24], axis=1, inplace=True)
df_info.drop(['User_ID', 'Response'], axis=1, inplace=True)
df_info.head(5)
```

✓ 0.0s

	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Recency	Complain	Country
0	1957	0	2	58138.0	0	0	58	0	0
1	1954	0	2	46344.0	1	1	38	0	2
2	1965	0	1	71613.0	0	0	26	0	6
3	1984	0	1	26646.0	1	0	26	0	3
4	1981	1	0	58293.0	1	0	94	0	0

Hình 231. Tách nhóm info

5.3 Đưa dữ liệu và mô hình Random Forest và Decision Tree

Tạo tập train và test cho các nhóm đã phân.

```
# Tạo tập train và test cho các nhóm

X_train_Mnt, X_test_Mnt, y_train_Mnt, y_test_Mnt = train_test_split(df_Mnt, output, test_size=0.2, random_state=100)
X_train_Num, X_test_Num, y_train_Num, y_test_Num = train_test_split(df_Num, output, test_size=0.2, random_state=100)
X_train_Acpt, X_test_Acpt, y_train_Acpt, y_test_Acpt = train_test_split(df_Acpt, output, test_size=0.2, random_state=100)
X_train_Info, X_test_Info, y_train_Info, y_test_Info = train_test_split(df_Info, output, test_size=0.2, random_state=100)

✓ 0.0s
```

Hình 232. Tạo tập Train và Test

5.3.1 Mô hình Ramdom Forest.

5.3.1.1 Test độ ảnh hưởng của các nhóm đã phân so với output.

Đưa tập train vào mô hình.

```
# Đưa tập train của các nhóm vào mô hình

rf_model_Mnt = RandomForestClassifier(n_estimators=100, random_state=0)
rf_model_Num = RandomForestClassifier(n_estimators=100, random_state=0)
rf_model_Acpt = RandomForestClassifier(n_estimators=100, random_state=0)
rf_model_Info = RandomForestClassifier(n_estimators=100, random_state=0)

rf_model_Mnt.fit(X_train_Mnt, y_train_Mnt)
rf_model_Num.fit(X_train_Num, y_train_Num)
rf_model_Acpt.fit(X_train_Acpt, y_train_Acpt)
rf_model_Info.fit(X_train_Info, y_train_Info)

✓ 0.4s
```

▼ RandomForestClassifier
RandomForestClassifier(random_state=0)

Hình 233. Đưa tập Train vào mô hình RF để huấn luyện

Độ chính xác của từng nhóm theo mức độ giảm dần.

```

# Độ chính xác theo từng nhóm cho mô hình Random Forest theo mức độ giảm dần

rf_accuracy_Mnt = rf_model_Mnt.score(X_test_Mnt, y_test_Mnt)
rf_accuracy_Num = rf_model_Num.score(X_test_Num, y_test_Num)
rf_accuracy_Acpt = rf_model_Acpt.score(X_test_Acpt, y_test_Acpt)
rf_accuracy_info = rf_model_info.score(X_test_info, y_test_info)

rf_accuracies = [
    ("Accuracy Mnt:", rf_accuracy_Mnt),
    ("Accuracy Num:", rf_accuracy_Num),
    ("Accuracy Acpt:", rf_accuracy_Acpt),
    ("Accuracy info:", rf_accuracy_info)
]

sorted_accuracies = sorted(rf_accuracies, key=lambda x: x[1], reverse=True)
for rf_accuracy in sorted_accuracies:
    print(rf_accuracy[0], rf_accuracy[1])

```

✓ 0.0s

Accuracy Acpt: 0.8918918918918919
 Accuracy info: 0.8828828828828829
 Accuracy Mnt: 0.8648648648648649
 Accuracy Num: 0.8220720720720721

Hình 234. Độ chính xác của từng nhóm cho mô hình RF

➤ Nhận xét: Ta thấy mức độ ảnh hưởng so với output của nhóm Acpt là cao nhất

5.3.1.2 Test các thuộc tính của nhóm có độ ảnh hưởng cao nhất do với output.

Đưa các thuộc tính của nhóm Acpt vào mô hình

```

X_train_Acpt_Cmp1, X_test_Acpt_Cmp1, y_train_Acpt_Cmp1, y_test_Acpt_Cmp1 = train_test_split(Acpt_Cmp1, output, test_size=0.2, random_state=100)
X_train_Acpt_Cmp2, X_test_Acpt_Cmp2, y_train_Acpt_Cmp2, y_test_Acpt_Cmp2 = train_test_split(Acpt_Cmp2, output, test_size=0.2, random_state=100)
X_train_Acpt_Cmp3, X_test_Acpt_Cmp3, y_train_Acpt_Cmp3, y_test_Acpt_Cmp3 = train_test_split(Acpt_Cmp3, output, test_size=0.2, random_state=100)
X_train_Acpt_Cmp4, X_test_Acpt_Cmp4, y_train_Acpt_Cmp4, y_test_Acpt_Cmp4 = train_test_split(Acpt_Cmp4, output, test_size=0.2, random_state=100)
X_train_Acpt_Cmp5, X_test_Acpt_Cmp5, y_train_Acpt_Cmp5, y_test_Acpt_Cmp5 = train_test_split(Acpt_Cmp5, output, test_size=0.2, random_state=100)

✓ 0.0s Python

```

```

rf_model_Acpt_Cmp1 = RandomForestClassifier(n_estimators=3, random_state=0)
rf_model_Acpt_Cmp2 = RandomForestClassifier(n_estimators=3, random_state=0)
rf_model_Acpt_Cmp3 = RandomForestClassifier(n_estimators=3, random_state=0)
rf_model_Acpt_Cmp4 = RandomForestClassifier(n_estimators=3, random_state=0)
rf_model_Acpt_Cmp5 = RandomForestClassifier(n_estimators=3, random_state=0)

rf_model_Acpt_Cmp1.fit(X_train_Acpt_Cmp1, y_train_Acpt_Cmp1)
rf_model_Acpt_Cmp2.fit(X_train_Acpt_Cmp2, y_train_Acpt_Cmp2)
rf_model_Acpt_Cmp3.fit(X_train_Acpt_Cmp3, y_train_Acpt_Cmp3)
rf_model_Acpt_Cmp4.fit(X_train_Acpt_Cmp4, y_train_Acpt_Cmp4)
rf_model_Acpt_Cmp5.fit(X_train_Acpt_Cmp5, y_train_Acpt_Cmp5)

```

✓ 0.0s Python

```

    RandomForestClassifier
    RandomForestClassifier(n_estimators=3, random_state=0)

```

Hình 235. Đưa các thuộc tính vào mô hình RF để huấn luyện

Độ chính xác của từng thuộc tính theo mức độ giảm dần

```
# Độ chính xác theo từng thuộc tính của nhóm Acpt cho mô hình Random Forest theo mức độ giảm dần

rf_accuracy_Acpt_Cmp1 = rf_model_Acpt_Cmp1.score(X_test_Acpt_Cmp1, y_test_Acpt_Cmp1)
rf_accuracy_Acpt_Cmp2 = rf_model_Acpt_Cmp2.score(X_test_Acpt_Cmp2, y_test_Acpt_Cmp2)
rf_accuracy_Acpt_Cmp3 = rf_model_Acpt_Cmp3.score(X_test_Acpt_Cmp3, y_test_Acpt_Cmp3)
rf_accuracy_Acpt_Cmp4 = rf_model_Acpt_Cmp4.score(X_test_Acpt_Cmp4, y_test_Acpt_Cmp4)
rf_accuracy_Acpt_Cmp5 = rf_model_Acpt_Cmp5.score(X_test_Acpt_Cmp5, y_test_Acpt_Cmp5)

rf_accuracies_Acpt = [
    ("Accuracy Acpt.Cmp1:", rf_accuracy_Acpt_Cmp1),
    ("Accuracy Acpt.Cmp2:", rf_accuracy_Acpt_Cmp2),
    ("Accuracy Acpt.Cmp3:", rf_accuracy_Acpt_Cmp3),
    ("Accuracy Acpt.Cmp4:", rf_accuracy_Acpt_Cmp4),
    ("Accuracy Acpt.Cmp5:", rf_accuracy_Acpt_Cmp5)
]

sorted_accuracies_Acpt = sorted(rf_accuracies_Acpt, key=lambda x: x[1], reverse=True)
for rf_accuracy_Acpt in sorted_accuracies_Acpt:
    print([rf_accuracy_Acpt[0], rf_accuracy_Acpt[1]])

✓ 0.0s

Accuracy Acpt.Cmp2: 0.8693693693693694
Accuracy Acpt.Cmp1: 0.8671171171171171
Accuracy Acpt.Cmp3: 0.8648648648648649
Accuracy Acpt.Cmp4: 0.8648648648648649
Accuracy Acpt.Cmp5: 0.8648648648648649
```

Hình 236. Độ chính xác của từng thuộc tính

- **Nhận xét:** Thuộc tính Acpt.Cmp2 có mức độ ảnh hưởng cao nhất trong toàn bộ tập dữ liệu so với output.

5.3.2 Mô hình Decision Tree.

5.3.2.1 Test độ ảnh hưởng của các nhóm đã phân so với output.

Đưa tập train vào mô hình.

```
# Đưa tập train của các nhóm vào mô hình

dt_model_Mnt = DecisionTreeClassifier(random_state=0)
dt_model_Num = DecisionTreeClassifier(random_state=0)
dt_model_Acpt = DecisionTreeClassifier(random_state=0)
dt_model_Info = DecisionTreeClassifier(random_state=0)

dt_model_Mnt.fit(X_train_Mnt, y_train_Mnt)
dt_model_Num.fit(X_train_Num, y_train_Num)
dt_model_Acpt.fit(X_train_Acpt, y_train_Acpt)
dt_model_Info.fit(X_train_Info, y_train_Info)

✓ 0.0s
```

Python

Hình 237. Đưa tập Train vào mô hình DT để huấn luyện

Độ chính xác của từng nhóm theo mức độ giảm dần.

```

# Độ chính xác theo từng nhóm cho mô hình Decision Tree

dt_accuracy_Mnt = dt_model_Mnt.score(X_test_Mnt, y_test_Mnt)
dt_accuracy_Num = dt_model_Num.score(X_test_Num, y_test_Num)
dt_accuracy_Acpt = dt_model_Acpt.score(X_test_Acpt, y_test_Acpt)
dt_accuracy_info = dt_model_info.score(X_test_info, y_test_info)

dt_accuracies = [
    ("Accuracy Mnt:", dt_accuracy_Mnt),
    ("Accuracy Num:", dt_accuracy_Num),
    ("Accuracy Acpt:", dt_accuracy_Acpt),
    ("Accuracy info:", dt_accuracy_info)
]

sorted_accuracies = sorted(dt_accuracies, key=lambda x: x[1], reverse=True)
for dt_accuracy in sorted_accuracies:
    print(dt_accuracy[0], dt_accuracy[1])

```

✓ 0.0s

```

Accuracy Acpt: 0.8918918918918919
Accuracy Num: 0.831081081081081
Accuracy info: 0.8175675675675675
Accuracy Mnt: 0.8063063063063063

```

Hình 238. Độ chính xác của từng nhóm

➤ **Nhận xét:** Ta thấy mức độ ảnh hưởng so với output của nhóm Acpt là cao nhất.

5.3.2.2 Test các thuộc tính của nhóm có độ ảnh hưởng cao nhất do với output.

Đưa các thuộc tính của nhóm Acpt vào mô hình.

```

dt_model_Acpt_Cmp1 = DecisionTreeClassifier(random_state=0)
dt_model_Acpt_Cmp2 = DecisionTreeClassifier(random_state=0)
dt_model_Acpt_Cmp3 = DecisionTreeClassifier(random_state=0)
dt_model_Acpt_Cmp4 = DecisionTreeClassifier(random_state=0)
dt_model_Acpt_Cmp5 = DecisionTreeClassifier(random_state=0)

dt_model_Acpt_Cmp1.fit(X_train_Acpt_Cmp1, y_train_Acpt_Cmp1)
dt_model_Acpt_Cmp2.fit(X_train_Acpt_Cmp2, y_train_Acpt_Cmp2)
dt_model_Acpt_Cmp3.fit(X_train_Acpt_Cmp3, y_train_Acpt_Cmp3)
dt_model_Acpt_Cmp4.fit(X_train_Acpt_Cmp4, y_train_Acpt_Cmp4)
dt_model_Acpt_Cmp5.fit(X_train_Acpt_Cmp5, y_train_Acpt_Cmp5)

```

✓ 0.0s

```

▼ DecisionTreeClassifier
DecisionTreeClassifier(random_state=0)

```

Hình 239. Đưa các thuộc tính vào mô hình DT để huấn luyện

Độ chính xác của từng nhóm theo mức độ giảm dần.

```

# Độ chính xác theo từng thuộc tính của nhóm Acpt cho mô hình Decision Trê theo mức độ giảm dần

dt_accuracy_Acpt_Cmp1 = dt_model_Acpt_Cmp1.score(X_test_Acpt_Cmp1, y_test_Acpt_Cmp1)
dt_accuracy_Acpt_Cmp2 = dt_model_Acpt_Cmp2.score(X_test_Acpt_Cmp2, y_test_Acpt_Cmp2)
dt_accuracy_Acpt_Cmp3 = dt_model_Acpt_Cmp3.score(X_test_Acpt_Cmp3, y_test_Acpt_Cmp3)
dt_accuracy_Acpt_Cmp4 = dt_model_Acpt_Cmp4.score(X_test_Acpt_Cmp4, y_test_Acpt_Cmp4)
dt_accuracy_Acpt_Cmp5 = dt_model_Acpt_Cmp5.score(X_test_Acpt_Cmp5, y_test_Acpt_Cmp5)

dt_accuracies_Acpt = [
    ("Accuracy Acpt.Cmp1:", dt_accuracy_Acpt_Cmp1),
    ("Accuracy Acpt.Cmp2:", dt_accuracy_Acpt_Cmp2),
    ("Accuracy Acpt.Cmp3:", dt_accuracy_Acpt_Cmp3),
    ("Accuracy Acpt.Cmp4:", dt_accuracy_Acpt_Cmp4),
    ("Accuracy Acpt.Cmp5:", dt_accuracy_Acpt_Cmp5)
]

sorted_accuracies_Acpt = sorted(dt_accuracies_Acpt, key=lambda x: x[1], reverse=True)
for dt_accuracy_Acpt in sorted_accuracies_Acpt:
    print(dt_accuracy_Acpt[0], dt_accuracy_Acpt[1])

```

✓ 0.0s

Accuracy Acpt.Cmp5: 0.8806306306306306
Accuracy Acpt.Cmp2: 0.8693693693693694
Accuracy Acpt.Cmp1: 0.8671171171171171
Accuracy Acpt.Cmp3: 0.8648648648648649
Accuracy Acpt.Cmp4: 0.8648648648648649

Hình 240. Độ chính xác của từng thuộc tính

- **Nhận xét:** Thuộc tính Acpt.Cmp5 có mức độ ảnh hưởng cao nhất trong toàn bộ tập dữ liệu so với output.

5.4 So sánh và Nhận Xét

5.4.1 So sánh

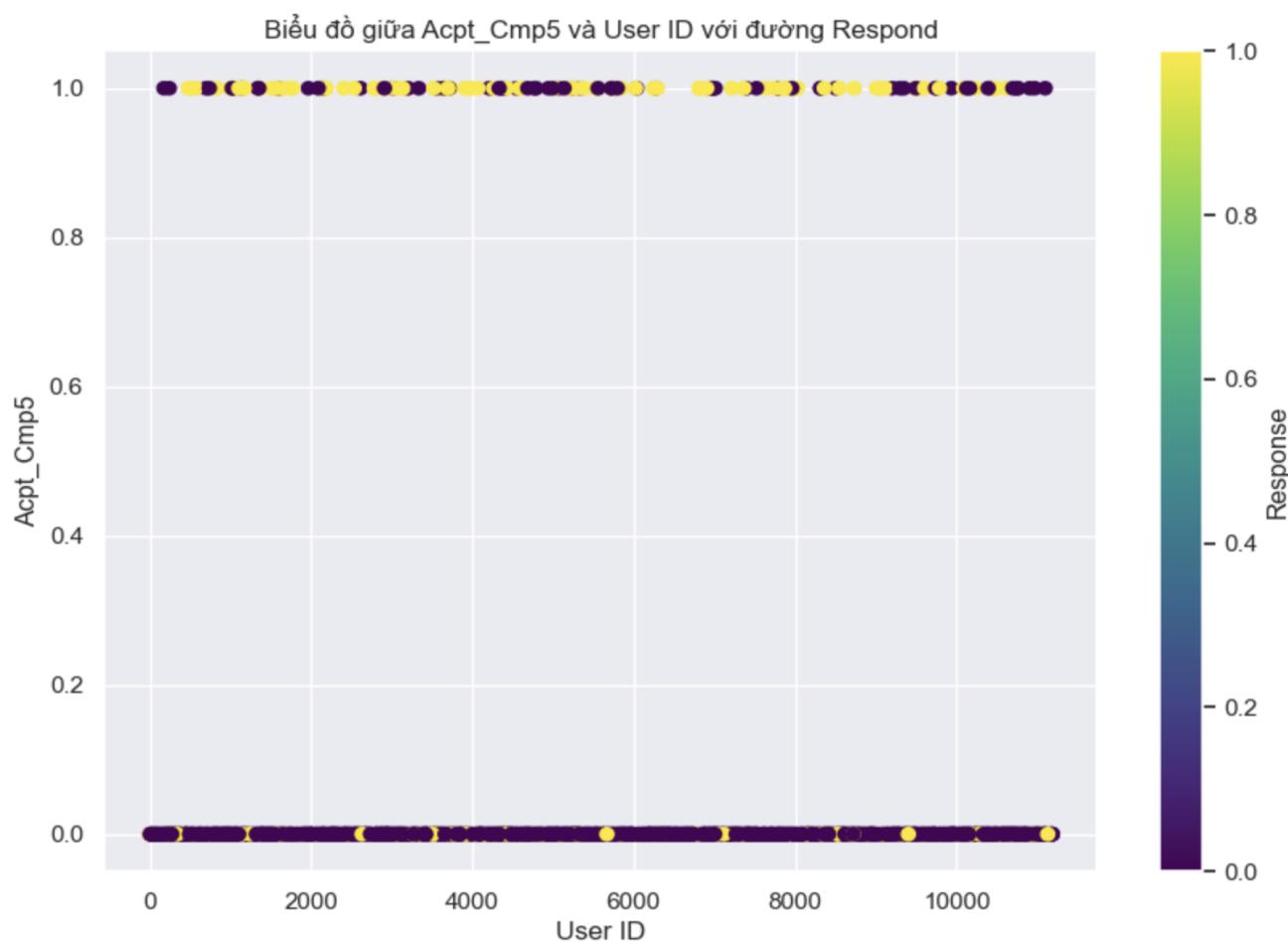
	Random Forest	Decision Tree
Nhóm dữ liệu có độ ảnh hưởng lớn nhất đến output	Acpt	Acpt
Mức độ ảnh hưởng	89%	89%
Thuộc tính của nhóm dữ liệu có độ ảnh hưởng lớn nhất đến output	Acpt.Cmp2	Acpt.Cmp5
Mức độ ảnh hưởng	87%	88%

5.4.2 Nhận xét

Hai mô hình **Random Forest** và **Decision Tree** đều cho ra một kết quả về **nhóm có độ ảnh hưởng cao nhất** đến sự phản hồi của người dùng là **nhóm Acpt** (nhóm khách hàng phản hồi cho Marketing Campaign) lên đến 89%.

Tuy nhiên, khi kiểm tra **độ ảnh hưởng của từng thuộc tính trong nhóm Acpt** thì mô hình **Random Forest** đã cho ra kết quả là **Acpt.Cmp2** (khách hàng phản hồi cho Marketing Campaign lần 2) là cao nhất với kết quả 87%. **Mô hình Decision Tree** lại cho ra kết quả **Acpt.Cmp5** (khách hàng phản hồi cho Marketing Campaign lần 5) là cao nhất với 88%.

5.4.3 Vẽ biểu đồ đặc trưng



Hình 241. Biểu đồ đặc trưng



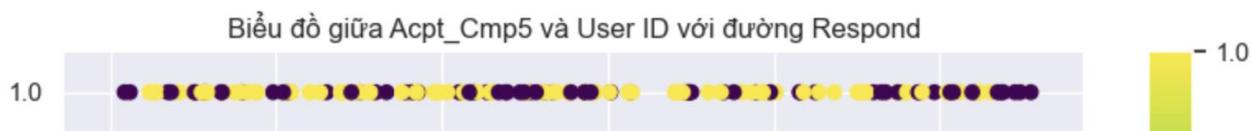
Response : Phản hồi của khách hàng cho cả chiến dịch Marketing.



Acpt_Cmp5 : Lần phản hồi thứ 5 của khách hàng về Marketing Campaign.

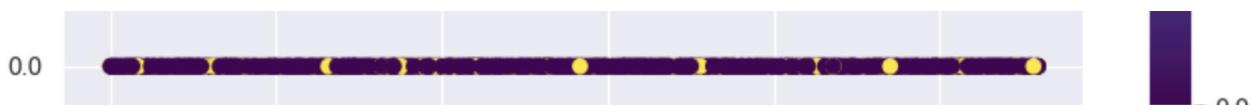
Quan sát biểu đồ ta thấy:

- Đối với nhóm khách hàng có lần phản hồi thứ 5 của khách hàng về **Acpt_Cmp5** (Marketing Campaign) là 1 (chấp thuận) thì **Response** của nhóm khách hàng đó là 1 rất nhiều, chiếm đến 89% theo như số liệu đã khai thác. Càng nhiều **Acpt_Cmp5** chấp thuận thì sẽ có càng nhiều **Response** chấp thuận.



Hình 242. Mối tương quan giữa Acpt cmp5 và User ID với Respond

- Ngược lại Càng nhiều Acpt_Cmp5 không chấp thuận thì sẽ có càng ít Response chấp thuận.



Hình 243. Mối tương quan giữa Acpt_cmp5 và User ID với Response

TÀI LIỆU THAM KHẢO

1. Tài liệu thiết kế kho dữ liệu
https://repository.vnu.edu.vn/bitstream/VNU_123/6559/1/V_L0_01959.pdf
 2. Slide bài giảng và hướng dẫn thực hành do giảng viên cung cấp.
 3. SQL Server 2012 Tutorials - Analysis Services Multidimensional Modeling.pdf.
 4. MDX Query Fundamentals (Analysis Services)”, Microsoft,
[MDX Query Fundamentals \(Analysis Services\) | Microsoft Learn](#)